

# Metodologia Automatizada para Descrever a Cobertura de Sinal Móvel em Rodovias: Estudo de Caso Considerando Incerteza de Dados

Leonardo L. Röpke, Marcia Pasin  
*Programa de Pós-Graduação em Ciência da Computação*  
*Universidade Federal de Santa Maria*  
Santa Maria, Brasil  
leonardoropke@gmail.com, marcia@inf.ufsm.br

Lucas M. Schnorr  
*Instituto de Informática*  
*Universidade Federal do Rio Grande do Sul*  
Porto Alegre, Brasil  
schnorr@inf.ufrgs.br

**Resumo**—Devido à grande distribuição geográfica ou à existência de ampla infraestrutura rodoviária, é desafiador garantir que operadoras de dados móveis forneçam um serviço adequado ao longo das rodovias. Uma metodologia automatizada para descrever e avaliar a cobertura de sinal de rede móvel em rodovias pode colaborar para que governos e operadoras planejem com eficiência ações para melhorar a qualidade do serviço provido pelas operadoras. Neste trabalho, é proposta uma metodologia automatizada para descrever e permitir a avaliação da cobertura de sinal de rede móvel em rodovias usando dados de rastreamento de veículos. A abordagem proposta agrega, usando um critério espaço-temporal, dados obtidos através de um sistema de rastreamento de veículos. A saída resultante deste processo de agregação é um mapa das rodovias com um esquema de cores gradiente, para indicar a disponibilidade do sinal e a confiança desta informação. Como o tráfego é dinâmico e os dados de tráfego não seguem uma distribuição uniforme, a medida de incerteza associada ao mapa indica a confiança da informação provida. Como estudo de caso, foi usado um conjunto de dados relativos a viagens no sul do Brasil coletados através de um serviço privado de rastreamento de veículos.

**Palavras-chave**—mapa de cobertura, disponibilidade do sinal, rodovias, rastreamento de veículos.

## I. INTRODUÇÃO

O rastreamento de veículos permite maior confiabilidade, eficiência e segurança para os serviços de transporte rodoviário. Para operar corretamente, este serviço requer suporte adequado de infraestrutura de rede de comunicação ao longo das rodovias. Normalmente essa infraestrutura é fornecida pelas operadoras de telefonia móvel, cabendo ao governo verificar a qualidade do serviço prestado. A qualidade do sinal de telefonia móvel pode ser facilmente medida por um aparelho e avaliada por um humano. No entanto, em geral, a avaliação manual é passível de erros e de difícil aplicação em territórios com áreas abrangentes. Além disso, a avaliação manual está intimamente relacionado à visão do especialista em um curto período de tempo. Quando este serviço é aplicado a territórios de maior abrangência, como é o caso das medições ao longo de rodovias, pode envolver deslocamento de pessoal e equipamentos, o que gera custos adicionais. Acidentes geográficos, como montanhas, prédios e túneis, podem comprometer a

eficiência do serviço de comunicação e, conseqüentemente, a resposta do serviço de avaliação da disponibilidade do sinal. Assim, avaliar e descrever a disponibilidade do serviço de comunicação em rodovias não é tarefa trivial.

Uma solução mais adequada seria ter uma metodologia automatizada para descrever e permitir avaliar o serviço da rede móvel em rodovias, onde a cobertura resultante pudesse ser observada em termos de espaço e tempo plotados em um mapa. Este trabalho contribui neste sentido. Como estudo de caso para demonstrar os resultados de nossa proposta, foi utilizado um conjunto de dados de propriedade privada relativo a viagens de rastreamento de veículos coletadas no sul do Brasil. Os dados consumidos neste trabalho foram gerados a partir de dispositivos que usam a tecnologia 2G. Entretanto, a metodologia aqui proposta pode ser aplicada a outras tecnologias.

A metodologia automatizada consome dados de sensores, que podem estar nos veículos/caminhões em circulação nas rodovias, desde que possuam dispositivos embarcados com capacidade de comunicação e localização. Os dados coletados são armazenados em um banco de dados centralizado (em uma empresa) e são analisados e agregados a partir de uma perspectiva espaço-temporal. Como o esquema de agregação reduz a quantidade de dados de entrada, é definido o grau de incerteza para descrever a qualidade dos dados resultantes que são plotados no mapa. Nesse contexto, as características do próprio conjunto de dados devem ser observadas. Em particular, são destacados os seguintes problemas que podem colaborar para adicionar ruído indesejado no mapa resultante:

- uma vez que uma característica intrínseca do tráfego é a falta de distribuição uniforme, existem estradas para as quais corre maior coleta de dados de comunicação e estradas com menos dados de comunicação disponíveis. Assim, a distribuição de dados no *dataset* carece de uniformidade geográfica e temporal, e esse ruído impacta no resultado do processo de agregação. Então, deve ser calculado um grau de incerteza no resultado. Assim, dada uma certa quantidade de dados com menor variação de valores (i.e., baixo desvio padrão), a chance de incerteza

dos dados é baixa; caso contrário, alta. Neste trabalho, representamos esses valores como um gradiente de cores no mapa, buscando prover informação com mais fidelidade;

- em especial, em algum veículo, o serviço de rastreamento pode ter sido desativado por motivos não especificados e os dados podem não ter sido obtidos adequadamente apesar da disponibilidade do serviço de comunicação. Portanto, veículos que apresentem baixo número de observações coletados no intervalo de tempo considerado não devem ser incluídos no processo de agregação;
- nem todas as estradas ou segmentos rodoviários têm infraestrutura de comunicação disponível. Nesse caso, devem ser considerados que os veículos também podem coletar dados de forma *offline*, e que esses dados serão transferidos quando o equipamento do veículo reconectar a rede móvel;
- uma vez que pode haver muitos veículos em circulação nas rodovias e, potencialmente, muitos dados podem ser gerados, e esses dados precisam ser tratados adequadamente antes de serem plotados em um mapa. Assim, o mapa é dividido em quadrantes e dados coletados são agregados considerando os quadrantes. Como o esquema de agregação reduz a quantidade de dados, o grau de incerteza descreve a qualidade dos quadrantes medidos no mapa resultante. Um esquema de cores gradiente é usado para representar a incerteza. Este esquema de gradiente garante uma descrição dos dados sobrepostos.

Finalmente, nem todas as estradas têm veículos que passam por elas com recursos de rastreamento e comunicação. Esta é uma limitação deste trabalho; estradas que não possuem veículos rastreados passando por elas não são contabilizadas. Os dados devem estar disponíveis para que a cobertura do sinal seja estimada. Esta é uma limitação intrínseca de qualquer metodologia automatizada baseada em dados obtidos pelo rastreamento de veículos.

Resumidamente, o objetivo deste trabalho é propor uma metodologia automatizada para avaliar a qualidade da rede de sinal móvel em rodovias e outras vias pavimentadas. Para validar a metodologia proposta, é usado um conjunto de dados coletados por rastreadores disponibilizados por meio de um sistema de rastreamento usando suporte de comunicação veículo para infraestrutura. Aqui, a qualidade do sinal na rede móvel é avaliada em termos de espaço e tempo, com especial atenção à incerteza dos dados.

Mais especificamente, a metodologia aqui proposta abrange:

- a construção de um índice de disponibilidade de sinal nas rodovias estaduais e federais, que contemple a incerteza dos dados;
- o mapeamento da disponibilidade da rede de sinal móvel nas rodovias estaduais e federais do Sul do Brasil.

A metodologia automatizada permite avaliar a qualidade do serviço de rede móvel disponível nas rodovias estaduais e federais, e também permite que tanto o governo quanto as empresas responsáveis planejem ações adicionais para melhorar a qua-

lidade do sinal da cobertura da rede móvel nas rodovias. Em especial, a metodologia automatizada pode ajudar a verificar se o Projeto de Lei 1721/2019<sup>1</sup> que estabelece a obrigatoriedade de oferta de serviços de telefonia móvel e de Internet móvel ao longo de rodovias federais está sendo obedecido.

O restante deste trabalho está organizado da seguinte forma. Seção II apresenta trabalhos relacionados. Seção III detalha e exemplifica a metodologia aqui proposta. Por fim, a Seção IV apresenta as conclusões.

## II. TRABALHOS RELACIONADOS

Os desafios em transformar dados em informação no contexto da mobilidade humana foram descritos anteriormente [1]. Devido a acidentes de terreno, clima, mudanças de rede, prédios, túneis, intensidade do sinal, falhas de dispositivos, etc., o processo de aquisição de dados possivelmente está sujeito a ruídos. Os dados podem estar ausentes, atrasados, distribuídos de forma heterogênea, o que impacta na qualidade do processo de aquisição dos dados. Além disso, a própria agregação de dados é um gargalo, em termos de qualidade da solução e tempo de processamento, e pode introduzir restrições importantes no resultado, pois um grupo de observações precisa ser reduzido a um único valor. Assim, a agregação de dados aplicada para construir os mapas de cobertura precisa considerar a existência de ruído de dados.

O mapeamento da disponibilidade e da qualidade do sinal móvel em rodovias continua sendo um problema em aberto. Em geral, trabalhos anteriores se concentram em descrever a existência ou a força do sinal em mapas de cobertura, considerando regiões em mapas. No texto que segue, trabalhos relacionados são brevemente comentados.

Com o foco em cobertura do sinal da rede celular, Lauridsen et al. [2] compararam diferentes tecnologias IoT para ver qual oferece melhor cobertura de sinal. Experimentos foram realizados através da leitura da quantidade de sinal móvel em um misto de áreas urbanas e rurais da Dinamarca. Rago et al. [3] buscaram identificar padrões de utilização de tráfego móvel através aprendizado não supervisionado. A análise apresentada é baseada em conjuntos de dados com informações do nível de sinal disponível. Já Adarsh et al. [4] utilizaram métricas de Qualidade de Serviço (QoS) e Quality of Experience (QoE) para medir o desempenho de redes Long Term Evolution (LTE) em regiões tribais, rurais e urbanas.

Com foco em avaliação de sinal de banda larga móvel, Khatouni et al. [5] utilizaram a plataforma aberta MONROE para realizar testes de velocidade em redes 3G/4G. Os autores destacam os desafios da realização de medições em Redes de Banda Larga Móvel (MBB) devido a fatores como mobilidade dos usuários, deficiências físicas, diversidade de tecnologia,

<sup>1</sup>Câmara dos Deputados. Projeto de Lei PL 1721/2019 e seus apensados. Altera a Lei no. 9.472, de 16 de julho de 1997, que "dispõe sobre a organização dos serviços de telecomunicações, a criação e funcionamento de um órgão regulador e outros aspectos institucionais, nos termos da Emenda Constitucional no. 8, de 1995", para estabelecer a obrigatoriedade de oferta de serviços de telefonia móvel e de Internet móvel ao longo de rodovias federais. <https://www.camara.leg.br/propostas-legislativas/2195292>

configurações de operadoras e outros. Siedner et al. [6] analisou o sinal da rede móvel na zona rural de Uganda para calcular a frequência de falhas e compor mapas de cobertura. Marina et al. [7] estudaram o impacto do contexto ambiente interno/externo do usuário na medição da qualidade do sinal, utilizando a força do sinal como métrica e usando um grande conjunto de dados obtidos através de *crowdsourcing* os autores demonstram que o contexto interior/exterior tem um impacto significativo nos resultados das medições. Wang et al. [8] utilizaram a intensidade do sinal recebido para descobrir a localização de antenas instaladas em uma cidade, utilizando um mapa dividido na forma de grade utilizando quadrantes de 50m por 50m, onde em cada grade o método de centroide ponderado modificado é empregado para realizar a agregação de dados e obter a localização.

Com foco em mapas de cobertura, Liu et al. [9] apresentam em sua pesquisa uma visão geral sobre os problemas de localização utilizando redes sem fio e de sensores. Os autores fornecem uma visão geral das abordagens de localização existentes e discutem os desafios de pesquisa implícitos. Freschi et al. [10] apresentam uma abordagem de redução e agregação de dados para uso em uma aplicação de *crowdsensing* de sensoriamento veicular destinada a monitorar a rugosidade de superfícies de estradas. A metodologia proposta pelos autores permitiu reduzir o tamanho dos dados em aplicações geoespaciais de forma bastante eficiente. Bokani et al. [11] avaliaram um conjunto de dados com repetidas medições realizadas em redes 3G e 4G sob condições de condução veicular sobre rotas pré-definidas. Em sua análise, os autores encontram implicações importantes sobre as flutuações da largura de banda que podem ocorrer em um ambiente urbano real. Alimpertis et al. [12] desenvolveram uma estrutura de previsão baseada em florestas aleatórias para a criação de mapas de cobertura com dados limitados. Em Fida et al. [13], apresentaram a construção de mapas de cobertura com a Indicação de Intensidade do Sinal Recebido (RSSI) usando um conjunto de dados com tamanho reduzido, para isso são utilizados algoritmos de clusterização como DBSCAN e HCLUST. Em Lutu et al. [14], a cobertura móvel nas ferrovias na Noruega foi avaliada em relação à perspectiva do usuário, incluindo desempenho de download HTTP e taxa de perda de pacotes, foi construída uma grade de cobertura dividindo as rotas em quadrantes e analisando a cobertura de duas operadoras distintas em cada segmento. Fida et al. [15] realizaram um estudo sobre o impacto do uso de dispositivos heterogêneos no resultado de mapas de disponibilidade do sinal. Os experimentos indicam que os mapas de cobertura móvel baseados em medições de diversos dispositivos são bastante confiáveis. Abrahamsson et al. [16] usaram dados de banda larga móvel da plataforma MONROE para realizar um estudo de medição multiacesso em veículos. Uma requisição analisa a disponibilidade de três operadoras e, para cada transação, escolhe a operadora com melhor tecnologia de acesso e melhor qualidade de sinal. Os experimentos do trabalho indicam que pode-se melhorar significativamente a disponibilidade do serviço e os tempos de transação em comparação com o uso de serviços de operadoras

individuais. Xavier et al. [17] apresentaram uma abordagem para visualização e análise de dados georreferenciados de uso de redes móveis. Engiz et al. [18] compararam a qualidade do sinal móvel 2G, 3G e 4G de três operadoras dentro de um campus universitário da Turquia. Os resultados do trabalho indicam que as operadoras não entregam o que exigem os órgãos regulatórios. Liu et al. [19] propuseram um método de triangulação adaptativa para dividir as regiões de interesse em triângulos. Os resultados dos experimentos evidenciam que o método apresentou um bom desempenho na construção de mapas de cobertura e tem um custo significativamente menor e maior eficiência do que algoritmos tradicionais. Jarvis et al. [20] desenvolvem um método de avaliação da qualidade do sinal em estradas através da divisão do mapa em segmentos de rota, tornando a visualização dos dados mais eficiente. Madariaga et al. [21] propuseram a melhoria da agregação da intensidade de sinal em cenários de *crowdsourcing*, evitando problemas comuns relacionados ao manuseio incorreto de valores de sinal em escala logarítmica e pela proposta de um novo método de agregação baseado em interpolação. Khatouni et al. [22] utilizaram dados de rede móvel coletados pela plataforma MONROE para analisar fluxos TCP e UDP e demonstrar o desempenho de serviços da Web.

Kousias et al. [23] e Schwind et al. [24] utilizaram dados provenientes de *crowdsourcing* para avaliar a experiência do usuário utilizando internet móvel, utilizando-se de métricas de desempenho, como taxa de *download/upload*, latência e intensidade do sinal.

Em contrapartida, neste trabalho, é proposta uma metodologia automatizada que descreve a existência da cobertura da rede móvel em estradas e a incerteza (ruído) desta informação. Os trabalhos anteriores não descrevem a incerteza dos resultados. Um conjunto de dados real é usado para avaliar a metodologia proposta. Como resultado, dados são plotados em um mapa usando um esquema de cores de gradiente. O esquema de cores descreve a incerteza da informação de disponibilidade, do mais alto (incorreto) ao mais baixo (correto).

### III. METODOLOGIA

Esta seção descreve com mais detalhes a metodologia automatizada para avaliar a qualidade do sinal da infraestrutura de rede móvel ao longo das rodovias, com informações sobre o grau de incerteza. A metodologia consome dados obtidos da infraestrutura de comunicação. Os dados são coletados usando um conjunto de veículos. Durante a viagem de cada veículo, o dispositivo de rastreamento instalado coleta dados que são enviados e armazenados em uma central. Portanto, os veículos participantes possuem dispositivos de rastreamento e recursos de comunicação disponíveis.

Em cada veículo, um dispositivo rastreador registra o veículo *id*, a existência ou não do sinal, e as coordenadas de latitude LAT e longitude LNG usando um suporte de GPS. Os dados coletados são enviados através de comunicação V2I para um serviço central, que armazena os dados em um banco de dados. Esses dados são então filtrados e usados como *dataset*

para gerar um mapa da região rastreada com a indicação da disponibilidade do sinal da rede móvel.

Basicamente, o serviço da unidade central filtra os dados armazenados, completa os dados faltantes usando o suporte de serviços abertos de terceiros, agrega e apresenta a informação final de forma adequada em um mapa, no contexto do espaço e do tempo. Para permitir a construção do mapa, uma série de filtros é aplicada. Cada filtro é implementado por um módulo de processamento diferente. Um esquema com a arquitetura da metodologia proposta, e seus módulos, é apresentado na Fig. 1. A seguir, são apresentados mais detalhes sobre os módulos que fazem parte da arquitetura.

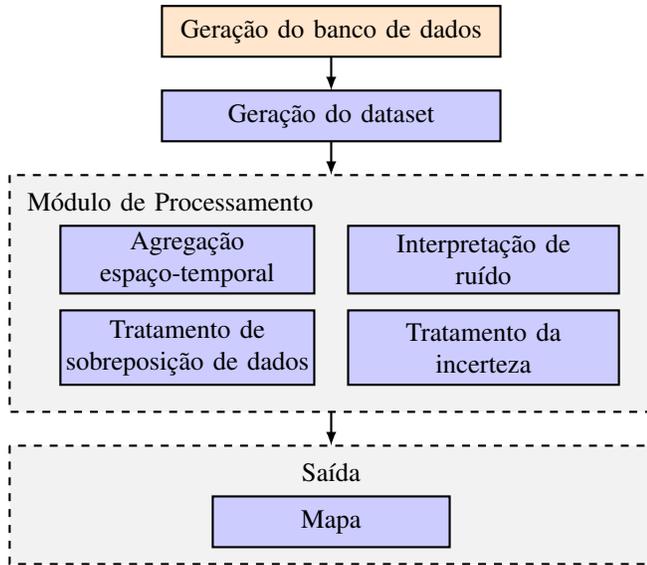


Figura 1. Esquema da metodologia automatizada proposta.

#### A. Geração de dados para o banco de dados

Na metodologia proposta, dados são coletados pelos veículos participantes. Esses dados são enviados de forma síncrona para o banco de dados centralizado por meio de comunicação V2I seguindo intervalos  $\Delta t$  durante a comunicação *online*. Se a conexão entre o veículo e o banco de dados centralizado não pode ser estabelecida, e a mensagem não pode ser enviada imediatamente após ser gerada, o veículo armazena os dados em uma memória local e os transmite, de forma assíncrona, quando a conexão retorna. Neste caso, o sistema considera que os dados foram gerados de forma *offline*.

Ao receber as mensagens emitidas pelos veículos que estavam *online* ou *offline*, a unidade central consolida os dados em uma tabela com as seguintes colunas:

- veículo *id*;
- valores de LAT e LNG coletados via GPS;
- *online*, que é um atributo binário para descrever se os dados foram transmitidos no mesmo momento em que foram gerados (*online*, sem atraso) ou se os dados foram transmitidos *offline* (ou seja, a mensagem tem um atraso de envio e foi enviada quando a ligação foi restabele-

cida). Valores possíveis são 1, que é usado em caso de comunicação *online*, 0 caso contrário;

- *date + time*, que é um timestamp atribuído pelo veículo emissor da mensagem.

Desde que os dados podem ser coletados ao longo dos anos, o banco de dados resultante pode ser usado como entrada para a metodologia avaliar a disponibilidade do sinal da rede móvel nas rodovias. Aqui, o banco de dados foi gerado por uma empresa privada de rastreamento, e foi utilizado  $\Delta t = 60s$ . A base de dados original era de 225 milhões de observações. Alternativamente, os dados poderiam ter sido coletados por meio de um aplicativo de *crowdsourcing*.

#### B. Geração do dataset

A próxima etapa consiste em filtrar e organizar os dados armazenados no banco de dados para formar um *dataset*.

Inicialmente, os dados são filtrados, em uma abordagem de duas etapas:

- definição do período de dados – nesta fase, é definido um período de tempo durante o qual a disponibilidade do sinal é avaliada; os dados que não estão nesse período são excluídos;
- anonimização dos dados de identificação – para manter o anonimato dos veículos rastreados, o campo do banco de dados *id\_object* é mascarado.

A seguir, os registros armazenados no banco de dados são completados com informações adicionais e um período de tempo (ou seja, um conjunto de anos) é selecionado para avaliar a disponibilidade do sinal da rede móvel nas rodovias.

Dados ausentes no banco de dados são incluídos usando o suporte de serviços abertos de terceiros. Para cada registro armazenado no banco de dados, os dados de estradas ausentes são inseridos usando dados de latitude e longitude. Dados rodoviários são obtidos de acordo com as informações e classificação disponíveis via APIs de Overpass<sup>2</sup> e OpenStreetMap<sup>3</sup>, e incluem:

- *type*: tipo da via, que pode ser *motorway*, *motorway\_link*, *primary*, *primary\_link*, *secondary*, *secondary\_link*, *tertiary*, *tertiary\_link*, *trunk* e *trunk\_link*;
- *name*: nome da via, como apresentado pela OpenStreetMap API.

Para completar os dados ausentes, os seguintes passos são executados para cada item no banco de dados, considerando os valores de LAT e LNG:

- aplicar a API Overpass com a seguinte consulta: `[out: json]; way (around: 25, LAT, LNG); out tags;`, que busca em um raio de 25 metros todas as vias disponíveis e grava em um arquivo com formato JSON;
- usando o arquivo JSON, extrair os elementos *name* e/ou *ref* para obter o nome da via e o elemento *highway* para obter o tipo da via. Se o processamento no arquivo JSON retornar mais de uma instância, o nível mais alto da via será usado.

<sup>2</sup><http://overpass-api.de>

<sup>3</sup><https://www.openstreetmap.org>

Após este tratamento, é aplicada uma série de métodos de limpeza dos dados armazenados no banco de dados, nomeadamente:

- seleção de dados que pertencem apenas a uma determinada região – uma região em um mapa é selecionada para avaliar a qualidade do sinal. Os dados que não estão relacionados a esta região são removidos. Neste trabalho, foram utilizados dados relativos ao estado do Rio Grande do Sul (RS);
- seleção de dados que pertencem apenas às vias principais, descartando ruas e avenidas, dado que o estudo aqui apresentado foca em rodovias;
- retirada de dados de veículos que se comunicavam apenas em modo *offline* – provavelmente este comportamento indica que o equipamento do veículo está com defeito durante o período de tempo avaliado;
- remoção de pontos emissores que ficaram *offline* por mais de 48 horas. Quando um emissor pára de funcionar, em até 48 horas o técnico é avisado. O equipamento é então levado ao laboratório para investigação e seus dados são armazenados no banco de dados como *offline*. Esses dados *offline* podem ser devido a um chip 2G defeituoso. Nesse caso, os dados *offline* relacionados a esse dispositivo com defeito são selecionados e removidos do banco de dados.

Por fim, os dados são armazenados como arquivo CSV, o que resulta em um conjunto de dados a ser utilizado para avaliar a disponibilidade do sinal rede móvel nas rodovias.

1) *Características do Dataset*: Os registros do banco de dados foram coletados durante o período de janeiro de 2017 a dezembro de 2020. No total, são 225 milhões de observações armazenadas (antes da filtragem), usando  $\Delta t = 60s$  como intervalo de tempo entre as observações coletadas. Este período temporal permite evitar o processamento dispendioso de dados, ao mesmo tempo que oferece uma quantidade volumosa de dados para validar a metodologia. A entrada resultante para os experimentos consiste em um conjunto de dados no formato CSV com  $\approx 27$  milhões de registros armazenados ( $\approx 3$  GBytes). O banco de dados original tem  $\approx 40$  milhões de registros armazenados. Um resumo das características do conjunto de dados é mostrado na Tab. I. As colunas na tabela indicam o *Ano*, o número de *Total de dispositivos* usados para coletar os dados e o número *Total de observações* coletadas naquele ano.

Tabela I  
CARACTERÍSTICAS DO DATASET

Ano	Total de dispositivos	Total de observações
2017	382	5.054.007
2018	495	6.816.778
2019	565	7.465.821
2020	455	8.045.743

No total, a empresa provedora dos dados disponibilizou dados de 861 dispositivos de rastreamento. Entretanto, neste

trabalho, para cada ano apenas um conjunto desses dispositivos foi usado para coletar os valores descritos na Tab. I.

2) *Configurações para os experimentos*: Nesta seção, são descritos os experimentos realizados no contexto deste trabalho. Para validar a metodologia, foi implementado uma aplicação em nível de servidor para executar um conjunto de experimentos usando o *dataset* privado como entrada. A aplicação utiliza da linguagem R v.3.6.3, e da IDE RStudio<sup>4</sup> v.1.2.5033.

Os principais pacotes R utilizados neste trabalho são *geosphere* para o tratamento de dados envolvendo geolocalização, *ggmap* para plotagem de dados em mapas e *shiny* para geração de páginas *Web* que permitem a visualização das informações. O servidor de aplicação está atualmente hospedado em um computador comum. Na borda, os modelos de dispositivos de rastreamento são ST340, ST300HD e ST310U do fabricante Suntech. Os dispositivos de rastreamento utilizam o chip MT2503 da MediaTek como padrão para transmissão de dados, que possui um modem 2G integrado.

A aplicação desenvolvida plota dados em mapas gerados por diferentes dispositivos de rastreamento, de uma empresa privada de rastreamento. A empresa de rastreamento tem clientes concentrados principalmente na região noroeste do estado do RS, no Sul do Brasil. Para este trabalho, são utilizados dados de 861 veículos rastreados. Os clientes da empresa atuam nas mais diversas áreas: transportadores de produtos perecíveis, grãos, máquinas, equipes comerciais, assistência técnica, fornecimento de energia, setor público, máquinas pesadas, coleta de resíduos, etc.

### C. Módulo de processamento

Após filtrar e organizar o conjunto de dados, o próximo passo é realizado pelo módulo de processamento, que gera o mapa de cobertura. Para isso, é usado o suporte da linguagem R<sup>5</sup>. O módulo de processamento é dividido em quatro partes que são descritas no texto a seguir.

1) *Agregação espaço-temporal*: A agregação de dados considera que os dados do *dataset* pertencem a diferentes veículos, diferentes estradas e foram coletados em diferentes períodos de tempo. Para resolver isso, dados são agregados no espaço e no tempo. Para agregação espacial, basicamente o mapa é dividido em quadrantes e, então, os pontos são agregados por região. Este procedimento é melhor descrito neste texto juntamente com o tratamento de pontos sobrepostos, no texto que segue. Para a agregação temporal, é avaliada a disponibilidade do sinal ao longo dos anos, para verificar se houve melhora no provimento deste serviço. Para verificar se ocorreu a melhora do serviço ao longo dos anos, a informação é mostrada no mapa por ano.

2) *Tratamento de dados sobrepostos*: Dado um conjunto de pontos a ser plotado em um mapa, o ponto  $d_i$  pode sobrepor o ponto  $d_{i-1}$  anteriormente plotado, simplesmente porque  $d_i$

<sup>4</sup><https://rstudio.com>

<sup>5</sup><https://www.r-project.org>

foi o último ponto a ser plotado no mapa. Isso pode dar a impressão de que o sinal está frequentemente disponível quando na verdade não está, ou vice-versa. O último ponto a ser plotado pode ocultar os valores previamente plotados.

Fig. 2 apresenta dois exemplos de mapas resultantes para o *dataset*, considerando pontos *online* e *offline*, em um recorte do estado do RS, sem usar esquema de agregação. Ambos os mapas descrevem dados para o mesmo período. Para a plotagem dos mapas, foram usadas diferentes quantidades de pontos. O gráfico representado na Fig. 2 (a) usa 50k pontos enquanto que na Fig. 2 (b) o número de pontos é o dobro. Uma observação nos mapas apresentados nesta figura pode resultar em uma interpretação incorreta dos dados. As imagens da Fig. 2 não são idênticas, apesar dos dados em ambos os mapas representarem exatamente o mesmo período de tempo.

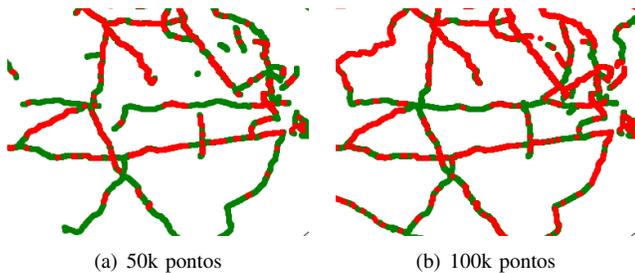


Figura 2. Plotagem com dados de disponibilidade de sinal com 50k pontos (a) e 100k pontos (b). Mapas gerados com python v. 3.7.6 e Geopandas; vermelho significa dados *offline* – portanto, nenhum sinal de rede disponível, e verde significa dados *online* transferidos – ou seja, sinal de rede disponível.

Para corrigir esse problema de sobreposição, dados devem ser agregados antes de serem plotados no mapa, e um grau de incerteza foi calculado para medir o ruído gerado pela agregação de dados. Por exemplo, em um conjunto de dados com  $n$  observações, um grupo de  $n_i$  observações, com  $n_i < n$ , pode ser usado para calcular a média de observações *online*, que pode ser plotada para representar a disponibilidade do sinal. Além disso, dado que a quantidade de observações agregadas pode ter um desvio padrão variável, por exemplo, muitos pontos em vermelho e poucos em verde resultando em um ponto vermelho e todos os pontos em vermelho também resultando em um ponto vermelho, um grau de (in)certeza deve ser definido para esses dados plotados.

Neste trabalho, como o mapa possui duas dimensões, é realizado um esquema de discretização de coordenadas em quadrantes com área configurável e então, com base nessas coordenadas, é realizada uma agregação espacial para unir todos os pontos que cabem em um quadrante, delimitado por um par com coordenadas de latitude e longitude. Assim, a agregação foi aplicada em termos de quadrantes do mapa. Uma grade foi gerada usando uma granularidade pré-definida  $\delta$ . A Fig. 3 mostra um exemplo genérico para a aplicação do esquema de divisão de uma estrada em quadrantes de mesmo tamanho, dado um valor  $\delta$ . Mais especificamente, neste exemplo, é realizada a identificação (1-15) e associação dos pontos amostrais à grade, partindo do quadrante do canto sudoeste (1) até o último quadrante (15). Os pontos são associados um

a um ao quadrante correspondente. A seguir, os pontos são plotados na grade inicialmente gerada, reunindo os pontos nos quadrantes correspondentes e calculando a disponibilidade do sinal, dada pela média das observações *online* em relação ao número de observações pertencentes ao quadrante.

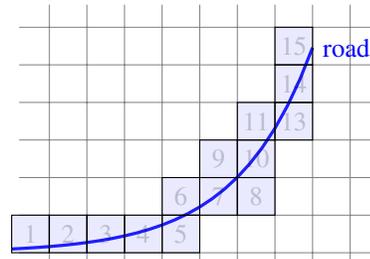


Figura 3. Uma estrada traçada na abordagem de agregação de dados.

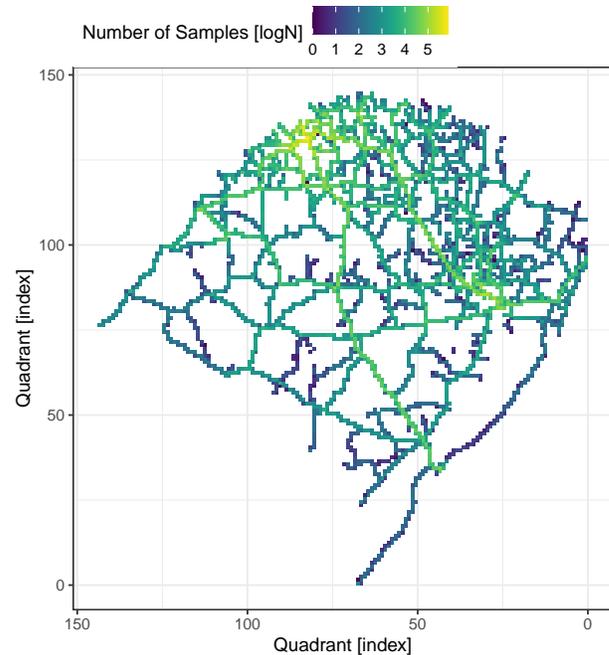


Figura 4. Mapa gerado com R para a cobertura móvel do sinal no RS, no Sul do Brasil, com  $\delta=5\text{km}$ . Na legenda, N indica o número de observações por quadrante.

A Fig. 4 ilustra o mapa resultante do processo de agregação de dados aplicado a dados reais coletados de janeiro de 2017 a dezembro de 2020, para o estado do RS. Em particular, o mapa apresenta dados plotados em escala logarítmica para destacar o impacto da quantidade de observações usadas para gerar cada quadrante.

De fato, a distribuição das observações coletadas pelos veículos é muito irregular na área geográfica considerada. Não há uma quantidade regular de caminhos percorrendo o RS no intervalo de tempo em que a coleta de dados foi aplicada. Neste exemplo,  $\approx 41$  milhões de observações resultaram em 5 mil quadrantes plotados no mapa. A abordagem de agregação de dados foi aplicada em um mapa de grade usando  $\delta = 5\text{km}$ . A

agregação reduz o número de observações e organiza os pontos a serem plotados no mapa. No mapa, a cor azul significa que poucos pontos foram usados como entrada para o processo de agregação e amarelo significa a existência de muitos pontos nos dados de entrada originais. Há duas áreas em destaque no mapa, a leste e a noroeste, com predominância da cor amarela indicando o grande número de observações nos dados de entrada. Isso significa que, nessas áreas, o grau de confiança nas informações fornecidas é maior.

3) *Tratamento da certeza/incerteza dos dados*: Como efeito colateral do processo de agregação, dados podem ser perdidos, pois um conjunto de amostras é transformado em um único valor resultante. Assim, é necessário o tratamento da certeza/incerteza dos dados.

Para representar esta limitação, calculamos a disponibilidade do sinal e o grau de certeza/incerteza associado a essa informação. Para calcular a disponibilidade do sinal, é selecionado o número de pontos *online*  $n$  em um vetor  $o = \{o_1, o_2, \dots, o_n\}$ , em um determinado quadrante, limitado por LAT ( $x_1, y_1$ ) e LNG ( $x_2, y_2$ ). Assim, a disponibilidade do sinal  $\rho$  (*Availability*) é a fração da soma dos pontos *online* em relação ao número total de pontos em determinado quadrante e pode ser dada por:

$$\rho = \frac{\sum_{i=1}^n o_i}{n_i}. \quad (1)$$

O valor resultante é um número entre 1 e 0. No extremo, 0 significa que a disponibilidade do sinal é ruim, e 1 a disponibilidade é boa. Então, para descrever a disponibilidade do sinal  $\rho$  no mapa, foi um esquema de gradiente de cores, entre azul e vermelho:

- a cor azul, representada por 1, indica a disponibilidade do sinal (ou seja, a maioria das observações está *online*);
- a cor vermelha representada por 0, indica a indisponibilidade do sinal (ou seja, a maioria das observações está *offline*).

Com relação à certeza/incerteza, foi estimada a incerteza  $u$  (*Uncertainty*), calculada pelo desvio padrão  $sd$  da existência de pontos *online/offline* no mesmo quadrante, que é dado por:

$$u = 2 \cdot \frac{sd(o_i)}{\sqrt{n_i}}. \quad (2)$$

Como um ponto no mapa de disponibilidade do sinal está associado a várias observações de dados, ou seja, as observações foram agregadas, também é gerado outro mapa gradiente com o grau de certeza/incerteza dos dados. O esquema de cores usado neste mapa é descrito por:

- a cor mais escura indica incerteza  $u$ , e é representada por 1;
- a cor mais clara indica certeza, e é representada por 0.

A Fig. 5 descreve a aplicação de abordagem de agregação de dados usando as duas métricas, disponibilidade de sinal  $\rho$ , calculada pela Eq. 1, onde 1 indica sinal *online* e 0 indica sinal *offline*, e a incerteza  $u$ , dada pela Eq. 2, onde os resultantes 0 significam forte confiança de que os dados estão corretos, 1 a

informação pode estar incorreta (já que poucas amostras foram usadas no processo de agregação). Na figura, o esquema de cores usado no mapa destaca o que está ruim. Pontos em azul e vermelho forte têm incerteza muito grande. Pontos claros indicam baixa incerteza.

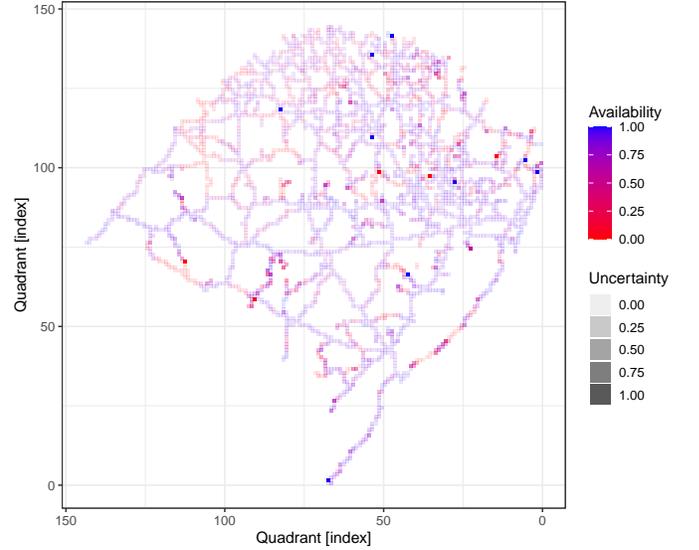


Figura 5. Disponibilidade e incerteza do sinal para rodovias do estado do RS, no Sul do Brasil, utilizando todas as observações no período total considerado.

Na metodologia proposta, é imposta uma alta incerteza sempre que o número de observações da entrada do processo de agregação é abaixo de um limite. Em particular, os experimentos conduzidos no escopo deste trabalho indicam que abaixo de 30 observações em um determinado quadrante refletem uma medição de disponibilidade do sinal de má qualidade.

#### IV. CONCLUSÕES

Neste trabalho, foi proposta uma metodologia para descrever a cobertura de sinal de rede móvel em rodovias usando dados de rastreamento de veículos. Como estudo de caso, foi utilizado um conjunto de dados relativos às viagens de veículos rastreados no estado do RS, Brasil. Uma vez que uma característica natural do tráfego é a falta de distribuição uniforme, e para descrever os dados que foram perdidos no processo de agregação de dados, é considerada a medida de incerteza juntamente com a cobertura do sinal. Ao contrário de trabalhos anteriores, a metodologia aqui apresentada considera a incerteza dos dados, medida pela quantidade de dados que é utilizada no processo de agregação. A metodologia proposta pode ser um ponto de partida para permitir que governos e empresas responsáveis planejem de forma mais eficientes ações para melhorar a cobertura do sinal da rede nas rodovias. Neste estudo, foi avaliado o sinal 2G, mas outros sinais podem ser avaliados com a mesma metodologia. De forma geral, pode-se constatar, pelos achados neste estudo, que a disponibilidade do sinal 2G no estado do RS ainda não é

adequada, mas predomina a disponibilidade do sinal (com destaque para a metade leste do estado).

Como trabalho futuro, pode ser explorada a divisão de estradas não por quadrantes de mesmo tamanho, mas usando segmentos de estrada, como em Jarvis et al. [20]. Isso permitirá descrever melhor a rodovia, sem considerar entroncamentos. Outro trabalho futuro é agregar dados por regiões de similaridade como em Fida et al. [13], o que permite ter uma visão macro do serviço de cobertura. Essas duas formas alternativas de agregação permitirão uma comparação direta de resultados com os referidos trabalhos relacionados. Outra possibilidade de continuação deste trabalho é um estudo que contemple a variação do parâmetro  $\delta$ , que já está em andamento. Alternativamente, os dados do conjunto de dados podem ser associados a cidades, o que possibilita observar a disponibilidade do sinal de uma perspectiva diferente.

## REFERÊNCIAS

- [1] L. Rosa, F. Silva, and C. Analide, "Mobile networks and internet of things infrastructures to characterize smart human mobility," *Smart Cities*, vol. 4, p. 894–918, 2021.
- [2] M. Lauridsen, H. Nguyen, B. Vejlgard, I. Kovacs, P. Mogensen, and M. Sorensen, "Coverage comparison of GPRS, NB-IoT, LoRa, and SigFox in a 7800 km<sup>2</sup> area," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, 06 2017, pp. 1–5.
- [3] A. Rago, G. Piro, H. D. Trinh, G. Boggia, and P. Dini, "Unveiling radio resource utilization dynamics of mobile traffic through unsupervised learning," in *2019 Network Traffic Measurement and Analysis Conference (TMA)*, 2019, pp. 209–214.
- [4] V. Adarsh, M. Nekrasov, U. Paul, T. Mangla, A. Gupta, M. Vigil-Hayes, E. Zegura, and E. Belding, "Coverage is not binary: Quantifying mobile broadband quality in urban, rural, and tribal contexts," in *2021 International Conference on Computer Communications and Networks (ICCCN)*, 07 2021, pp. 1–9.
- [5] A. S. Khatouni, M. Mellia, M. A. Marsan, S. Alfredsson, J. Karlsson, A. Brunstrom, O. Alay, A. Lutu, C. Midoglu, and V. Mancuso, "Speedtest-like measurements in 3g/4g networks: The monroe experience," in *2017 29th International Teletraffic Congress (ITC 29)*, vol. 1, 2017, pp. 169–177.
- [6] M. J. Siedner, A. Lankowski, D. Musinga, and et al., "Optimizing network connectivity for mobile health technologies in sub-Saharan Africa," *PLoS ONE* 7(9): e45643, September 2012.
- [7] M. K. Marina, V. Radu, and K. Balampekos, "Impact of indoor-outdoor context on crowdsourcing based mobile coverage analysis," in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, 2015, pp. 45–50.
- [8] H. Wang, S. Xie, and M. Li, K. and Ahmad, "Big data-driven cellular information detection and coverage identification," *Sensors*, vol. 19, no. 4, pp. 937–, 2019.
- [9] Y. Liu, Z. Yang, X. Wang, and L. Jian, "Location, localization, and localizability," *Journal of Computer Science and Technology*, vol. 25, no. 2, pp. 274–297, 2010.
- [10] V. Freschi, S. Delpriori, L. C. Klopfenstein, E. Lattanzi, G. Luchetti, and A. Bogliolo, "Geospatial data aggregation and reduction in vehicular sensing applications: The case of road surface monitoring," in *2014 International Conference on Connected Vehicles and Expo (ICCVE)*, 2014, pp. 711–716.
- [11] A. Bokani, M. Hassan, S. S. Kanhere, J. Yao, and G. Zhong, "Comprehensive mobile bandwidth traces from vehicular networks," in *Proceedings of the 7th International Conference on Multimedia Systems*, ser. MMSys '16. New York, NY, USA: Association for Computing Machinery, 2016.
- [12] E. Alimpertis, A. Markopoulou, C. Butts, and K. Psounis, "City-wide signal strength maps: Prediction with random forests," in *The World Wide Web Conference*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 2536–2542.
- [13] M. Fida, A. Lutu, M. K. Marina, and Özgü. Alay, "Zipweave: Towards efficient and reliable measurement based mobile coverage maps," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9.
- [14] A. Lutu, Y. R. Siwakoti, Özgü. Alay, D. Baltrūnas, and A. Elmokashfi, "The good, the bad and the implications of profiling mobile broadband coverage," *Computer Networks*, vol. 107, pp. 76–93, 2016.
- [15] M.-R. Fida and M. K. Marina, "Impact of device diversity on crowd-sourced mobile coverage maps," in *2018 14th International Conference on Network and Service Management (CNSM)*, 2018, pp. 348–352.
- [16] H. Abrahamsson, F. B. Abdesslem, B. Ahlgren, A. Brunstrom, I. Marsh, and M. Björkman, "Connected vehicles in cellular networks: Multi-access versus single-access performance," in *2018 Network Traffic Measurement and Analysis Conference (TMA)*, 2018, pp. 1–6.
- [17] W. Z. Xavier, F. H. Z. Xavier, and H. T. Marques-Neto, "Visualizing and analyzing georeferenced workloads of mobile networks," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2017, pp. 306–310.
- [18] B. Engiz and C. Kurnaz, "Comparison of signal strengths of 2G/3G/4G services on a university campus," *International Journal of Applied Mathematics, Electronics and Computers (IJAMEC)*, pp. 37–42, December 2016.
- [19] Y. Liu, W. Huangfu, H. Zhang, and K. Long, "Multi-criteria coverage map construction based on adaptive triangulation-induced interpolation for cellular networks," *IEEE Access*, vol. 7, pp. 80 767–80 777, 2019.
- [20] C. Jarvis, C. Midoglu, A. Lutu, and O. Alay, "Visualizing mobile coverage from repetitive measurements on defined trajectories," in *2018 Network Traffic Measurement and Analysis Conference (TMA)*, 2018, pp. 1–6.
- [21] D. Madariaga, J. Madariaga, J. Bustos, and B. Bustos, "Improving signal-strength aggregation for mobile crowdsourcing scenarios," *Sensors*, vol. 21, p. 1084, 02 2021.
- [22] A. S. Khatouni, M. Trevisan, D. Giordano, M. Rajiullah, S. Alfredsson, A. Brunstrom, C. Midoglu, and O. Alay, "An open dataset of operational mobile networks," in *Proceedings of the 18th ACM Symposium on Mobility Management and Wireless Access*, ser. MobiWac '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 83–90. [Online]. Available: <https://doi.org/10.1145/3416012.3424619>
- [23] K. Kousias, C. Midoglu, O. Alay, A. Lutu, A. Argyriou, and M. Riegler, "The same, only different: Contrasting mobile operator behavior from crowdsourced dataset," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–6.
- [24] A. Schwind, F. Wamser, T. Hossfeld, S. Wunderer, E. Tarnvik, and A. Hall, "Crowdsourced network measurements in germany: mobile internet experience from end user perspective," in *Broadband Coverage in Germany; 14. ITG Symposium*, 2020, pp. 1–7.