

UNIVERSIDADE FEDERAL DO PARANÁ

TAYANI BELLUCCI PETERLINI

ANÁLISE DA QUALIDADE DA REGRESSÃO LOGÍSTICA DE CONTRIBUIÇÕES
ACUMULADAS NO OPENSTREETMAP: ESTUDO DE CASO NOS MUNICÍPIOS
BRASILEIROS

CURITIBA

2024

TAYANI BELLUCCI PETERLINI

ANÁLISE DA QUALIDADE DA REGRESSÃO LOGÍSTICA DE CONTRIBUIÇÕES
ACUMULADAS NO OPENSTREETMAP: ESTUDO DE CASO NOS MUNICÍPIOS
BRASILEIROS

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Geodésicas (PPGCG), Setor de Ciências da Terra, da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Ciências Geodésicas.

Orientador: Prof. Dr. Marcio Augusto Reolon Schmidt

Coorientador: Prof. Dr. Elias Nasr Naim Elias

CURITIBA

2024

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

Peterlini, Tayani Bellucci

Análise da qualidade da regressão logística de contribuições acumuladas no OpenStreetMap: estudo de caso nos municípios brasileiros / Tayani Bellucci Peterlini. – Curitiba, 2024.

1 recurso on-line : PDF.

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-Graduação em Ciências Geodésicas.

Orientador: Marcio Augusto Reolon Schmidt

Coorientador: Elias Nasr Naim Elias

1. Análise de regressão. 2. Mapas topográficos. 3. Dados Geoespaciais. 4. OpenStreetMap. I. Universidade Federal do Paraná. II. Programa de Pós-Graduação em Ciências Geodésicas. III. Schmidt, Marcio Augusto Reolon. IV. Elias, Elias Nasr Naim. V. Título.

Bibliotecário: Elias Barbosa da Silva CRB-9/1894



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS DA TERRA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS
GEODÉSICAS - 40001016002P6

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **TAYANI BELLUCCI PETERLINI** intitulada: **ANÁLISE DA QUALIDADE DA REGRESSÃO LOGÍSTICA DE CONTRIBUIÇÕES ACUMULADAS NO OPENSTREETMAP: ESTUDO DE CASO EM MUNICÍPIOS BRASILEIROS**, sob orientação do Prof. Dr. MARCIO AUGUSTO REOLON SCHMIDT, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestra está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 15 de Julho de 2024.

Assinatura Eletrônica

15/07/2024 14:19:15.0

MARCIO AUGUSTO REOLON SCHMIDT

Presidente da Banca Examinadora

Assinatura Eletrônica

16/07/2024 15:50:28.0

SILVANA PHILIPPI CAMBOIM

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

16/07/2024 11:38:54.0

JULIA CELIA MERCEDES STRAUCH

Avaliador Externo (ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS)

AGRADECIMENTOS

É com imensa gratidão e alegria que dedico este espaço para expressar meus agradecimentos. Primeiramente, a Deus, cujo direcionamento, proteção e graça foram fontes constantes de conforto e inspiração. Em momentos desafiadores, Sua presença foi meu refúgio, proporcionando força e discernimento para superar obstáculos e seguir em frente.

À minha família, meu alicerce, expresso meu profundo agradecimento. Vocês foram meu apoio constante, encorajando-me nos dias difíceis e celebrando comigo nos momentos de conquista. Minha mãe, Sandra, meu pai, Luiz Henrique e meus irmãos, Deivid e Maria Thereza, vocês dividiram comigo este sonho; a paciência, sensibilidade e amor que recebi de cada um de vocês foram essenciais para completar esta jornada acadêmica.

Minha gratidão aos meus admiráveis orientadores; obrigada pela generosidade em compartilhar seus conhecimentos e experiências. Prof. Dr. Marcio Schmidt minha sincera gratidão pela sua compreensão, paciência e confiança, sem a sua sabedoria e fraternidade a me orientar, isso não teria sido possível. Ao meu coorientador Prof. Dr. Elias Nasr Naim Elias, por sua amizade, paciência e contribuições, que foram inestimáveis, vocês enriqueceram esta pesquisa e tornaram esta experiência ainda mais significativa.

À UFPR, à CAPES pelo fomento e a todos os professores e colaboradores do PPGCG. Aos colegas que passaram por esta jornada comigo, em especial minha amiga Bruna Zambrano, cuja amizade singular transformou esta vivência, sua presença e apoio fizeram toda a diferença, uma conexão que levarei para vida.

Aos membros da banca examinadora, Prof. Dra. Julia Strauch, Prof. Dr. Caio dos Anjos Paiva e Prof. Dra. Silvana Philippi Camboim, por dedicarem seu tempo e conhecimento na avaliação do meu trabalho, com suas sugestões e observações valiosas. Agradeço por ter tido a chance de apresentar meu trabalho diante de uma banca tão respeitável.

Por fim, expresso meus sinceros agradecimentos a todos que, de alguma forma, contribuíram para a concretização deste momento. Cada gesto de apoio, palavra de incentivo e colaboração deixaram uma marca valiosa. Que a generosidade de cada um seja retribuída multiplicadamente, meu profundo reconhecimento por toda a ajuda e encorajamento recebidos.

“Estuda. Estuda com empenho. Se tens de ser sal e luz, necessitas de ciência, de idoneidade. Ou julgas que, por seres preguiçoso e comodista, hás de receber ciência infusa?”

(S. Josemaria Escrivá, Caminho 340)

RESUMO

Em países em desenvolvimento, como o Brasil, onde os mapas topográficos e outras fontes cartográficas oficiais são muitas vezes desatualizados ou incompletos, a VGI emergiu como uma importante fonte de informações espaciais. No entanto, melhorias na qualidade desses dados podem ampliar sua utilidade em diversas aplicações. Este estudo explora a importância crescente das contribuições voluntárias de Dados Geoespaciais (VGI), com foco no contexto do *OpenStreetMap* (OSM) no Brasil, analisando o comportamento dos parâmetros da curva de regressão logística, aplicada na avaliação da qualidade dos dados VGI, e sua relação com as características das contribuições, incluindo tipo de dados e tamanho da célula de 15 cidades brasileiras de diferentes portes populacionais, variando desde grandes centros urbanos até municípios menores, com o objetivo de avaliar a influência do tipo de dados, para estabelecer os limites de aplicação do modelo utilizado em ferramentas de análise, como Elias (2022) trouxe em sua pesquisa, trazendo melhorias em relação a confiabilidade das decisões baseadas em dados voluntários. A partir das análises realizadas, verificou-se que a adaptação do tamanho da célula de acordo com a área urbana é fundamental para garantir a precisão dos resultados, uma vez que células grandes em municípios pequenos geram erros na modelagem. O estudo identificou que as células de 3 km eram adequadas para cidades de médio a grande porte, enquanto cidades menores exigiram células de tamanho reduzido. Além disso, foi observada a ocorrência de erros em células com altas taxas de inserção de dados em períodos curtos, o que compromete a eficácia da regressão logística, especialmente em análises anuais. Já em análises realizadas em períodos de 5 a 10 anos, a ferramenta se mostrou eficaz em identificar padrões de contribuição e na elaboração dos gráficos de regressão logística, fornecendo informações confiáveis para a tomada de decisões baseadas em dados VGI. Este estudo, ao preencher uma lacuna de conhecimento sobre a eficácia dos parâmetros e modelos utilizados em ferramentas de análise de dados voluntários, contribui para o avanço do uso dessas ferramentas em análises geoespaciais, com maior confiabilidade em diversas áreas de aplicação.

Palavras-chave: Parâmetros Intrínsecos 1. Dados colaborativos 2. Mapeamento colaborativo 3.

ABSTRACT

In developing countries like Brazil, where topographic maps and other official cartographic sources are often outdated or incomplete, Volunteered Geographic Information (VGI) has emerged as a significant source of spatial data. However, improvements in the quality of these data can enhance their utility across various applications. This study explores the growing importance of voluntary geospatial data contributions (VGI), with a focus on the OpenStreetMap (OSM) context in Brazil. It analyzes the behavior of logistic regression curve parameters, applied in the assessment of VGI data quality, and their relationship with contribution characteristics, including data type and cell size, across 15 Brazilian cities of different population sizes, ranging from large urban centers to smaller municipalities. The objective is to evaluate the influence of data type and establish the limits of the model's application in analysis tools, as discussed by Elias (2022), who proposed improvements to the reliability of decisions based on voluntary data. The analyses revealed that adapting cell size according to urban area is crucial to ensuring result accuracy, as large cells in small municipalities generate modeling errors. The study found that 3 km cells were suitable for medium to large cities, while smaller cities required reduced cell sizes. Additionally, errors were observed in cells with high data insertion rates over short periods, which compromised the efficacy of logistic regression, particularly in annual analyses. However, in analyses conducted over periods of 5 to 10 years, the tool proved effective in identifying contribution patterns and generating reliable logistic regression graphs, providing trustworthy information for decision-making based on VGI data. By addressing a knowledge gap regarding the efficacy of parameters and models used in voluntary data analysis tools, this study contributes to the advancement of these tools' application in geospatial analyses, ensuring greater reliability in various areas of application.

Keywords: Logistic regression 1. Collaborative data 2. Collaborative mapping 3.

LISTA DE FIGURAS

Figura 1 - Etapas da avaliação quantitativa no VGI.....	26
Figura 3 - Figura do mapa de localização dos municípios da área de estudo.....	46
Figura 4 - Parâmetros e zonas da curva logística.....	55
Figura 5 - Grades de 3 km aplicadas à área total dos municípios de São Paulo e Curitiba.....	61
Figura 6 - Grades de 3 km aplicadas à área urbana dos municípios de São Paulo e Curitiba.....	63
Figura 7 - Exemplo de grade limitada a condição do modelo utilizado na cidade de Fortaleza.....	68
Figura 8 - Grades nas áreas de São Paulo e Curitiba com limites a condição do modelo utilizado.....	71
Figura 9 - Grades limitadas a condição do modelo utilizado nos municípios de São Paulo e Curitiba.....	72
Figura 10 - Grades 3x3 da cidade de Curitiba no período de 10, 5 e 1 ano, com e sem soluções do modelo utilizado.....	74
Figura 11 - Grades 3x3 da cidade de São Paulo no período de 10, 5 e 1 ano, com e sem soluções do modelo utilizado.....	75
Figura 12 - Exemplo de valores de inclinação da curva de regressão por grade.....	76
Figura 13 - Exemplo de modelagem inconsistente.....	77
Figura 14 - Dispersão do n° de contribuintes x IDH.....	81

LISTA DE QUADROS

Quadro 1 - Municípios selecionadas para área de estudo.....	42
---	----

LISTA DE TABELAS

Tabela 1 - Valor médios de RMSE por município.....	64
Tabela 2 - Valor médios de RMSE da área urbana por município.....	66
Tabela 3 - Relação de nº de grades atingidas por dados em massa por município...	68
Tabela 4 - Relação de nº de grades área urbana atingidas por dados em massa por município.....	69
Tabela 5 - Número de usuários por intervalo de tempo em cada município.....	78
Tabela 6 - IDH dos municípios analisadas.....	80

LISTA DE ABREVIATURAS OU SIGLAS

API	- Interfaces de Programação de Aplicativos
ET-CQDG	-Especificação Técnica para Controle de Qualidade de Dados Geoespaciais
ET-PCDG	-Especificação. Técnicas dos Produtos dos Conjuntos de Dados Geoespaciais
IBGE	- Instituto Brasileiro de Geografia e Estatística
INDE	- Infraestrutura Nacional de Dados Espaciais
ISO	- <i>International Organization for Standardization</i>
GNSS	- Sistema Global de Navegação por Satélite
OS	- Nome por extenso
OSM	- <i>OpenStreetMap</i>
OSMF	- <i>OpenStreetMap Foundation</i>
SIG	- Sistema de Informação Geográfica
VGI	- Informação Geográfica Voluntária

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 JUSTIFICATIVA.....	17
1.2 OBJETIVOS.....	18
1.2.1 Objetivo Geral.....	18
1.2.2 Objetivos Específicos.....	19
2 REVISÃO DE LITERATURA.....	20
2.1 MAPEAMENTO COLABORATIVO.....	20
2.1.1 VGI e Qualidade dos Dados Geoespaciais.....	21
2.2 MÉTODOS DE MODELAGEM DE PARÂMETROS INTRÍNSECOS.....	35
2.3 ERRO MÉDIO QUADRÁTICO (RMSE).....	36
3 MATERIAL E MÉTODOS.....	39
3.1 DEFINIÇÃO DA ÁREA DE ESTUDO.....	41
3.2 OBTENÇÃO DOS DADOS PARA DESENVOLVIMENTO DAS ANÁLISES... 48	
3.2.1 Software Utilizado.....	49
3.2.2 Desenvolvimento de aplicações.....	49
3.3 TAMANHO DAS CÉLULAS.....	50
3.4 ERRO MÉDIO QUADRÁTICO.....	53
3.5 PARÂMETROS DA CURVA.....	54
4 RESULTADOS E DISCUSSÕES.....	60
4.1 DEFINIÇÃO DO TAMANHO DA CÉLULA.....	60
4.2 CÁLCULO DO RMSE.....	64
4.3 DADOS EM MASSA.....	67
4.4 USUÁRIOS OSM.....	78
5 CONCLUSÃO.....	82
6 RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	84
REFERÊNCIAS.....	86

1 INTRODUÇÃO

Quando se observa o mapeamento em países em desenvolvimento, como o Brasil, onde os mapas são muitas vezes antigos ou incompletos, a relevância da contribuição da VGI surge como uma importante fonte de informações espaciais, especialmente em contextos onde os mapeamentos sistemáticos apresentam lacunas ou não acompanham o ritmo de evolução das geotecnologias. No caso do Brasil, a cobertura total para mapeamentos na escala de 1:25.000 não passa de 5% (Silva e Camboim, 2021). Nesse contexto, os avanços tecnológicos, como o uso de dados de GPS, imagens de satélites e a ampla disseminação de informações geoespaciais, abriram novas perspectivas para o uso e a produção de mapas. Com essa propagação, novas possibilidades surgiram para a obtenção de dados geoespaciais, a exemplo do mapeamento voluntário colaborativo.

O mapeamento colaborativo ou informações geográficas voluntárias (VGI), ocorre por meio de plataformas, como o *OpenStreetMap* (OSM), fundado em 2004, o qual utiliza uma licença *Open Database License*, produzindo e disponibilizando dados geoespaciais atualizados e abertos na internet, apresentando-se como uma alternativa relevante para ser utilizada em conjunto com o mapeamento de referência e métodos de análise geoespacial, os dados abertos também podem ser integrados em qualquer software de Sistemas de Informações Geográficas (SIG). Embora o QGIS ofereça *plugin* que facilita o acesso a esses dados, outras ferramentas e métodos também permitem seu uso. Uma das grandes vantagens do VGI é sua capacidade de obter dados geoespaciais de uma ampla gama de fontes de forma rápida e econômica, e com alta frequência de atualização. Com a crescente disponibilidade de *smartphones* e outros dispositivos móveis, tem se tornado progressivamente mais acessível para os indivíduos coletar e compartilhar dados, porém quando relacionamos dados geoespaciais com informações geográficas voluntárias (VGI), é necessário conhecer a qualidade dos dados, possibilitando mensurar o quanto seu uso pode ser viável em situações específicas. Atualmente, os avanços tecnológicos viabilizaram a qualquer indivíduo que tenha computador ou *smartphone* com acesso à internet a geração de dados geoespaciais (Ganapati, 2011).

A avaliação da qualidade é um dos elementos-chave da produção, sempre analisada no contexto do cumprimento das especificações técnicas. A ISO (*International Organization for Standardization*) 19157 (ISO, 2013) é a responsável pelos indicadores de avaliação da qualidade de dados geoespaciais, onde são estabelecidos os critérios para a avaliação da qualidade de dados geoespaciais, sendo: Acurácia Posicional, Acurácia Temática, Consistência Lógica, Completude e Usabilidade.

Um dos desafios significativos associados ao VGI é a mensuração da qualidade dos dados, conforme discutido por Goodchild (2007), ao introduzir o conceito de VGI, e Haklay (2010), que investigou a precisão dos dados em comparação com fontes oficiais. No Brasil, essa avaliação ocorre de forma comparativa por meio da Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG, 2016) que especifica os indicadores de verificação. Esses indicadores são complementados por uma série de métodos de amostragem pelo Manual Técnico para Avaliação da Qualidade de Dados Geoespaciais do Instituto Brasileiro de Geografia e Estatística (IBGE, 2019), as normas que regem os conjuntos de dados espaciais definem mais de uma dezena de indicadores, tanto quantitativos quanto qualitativos, para descrever suas características.

Como as informações geoespaciais voluntárias são coletadas por indivíduos profissionais e não profissionais com uma diversidade de equipamentos, geralmente dispositivos móveis ou contribuições em massa por empresas e instituições, a precisão e a confiabilidade dos dados podem variar significativamente. Portanto, é essencial desenvolver métodos eficazes para a qualidade do VGI antes de usá-lo em processos de tomada de decisão. Pesquisas como (Haklay 2010; Girres e Touya 2010; Zhou 2018) apontam as preocupações em relação à qualidade. A importância da qualidade dos dados nos negócios e na ciência é bem reconhecida e amplamente descrita por (Redman, 1997; Loshin, 2010; Bielecka, 2015).

Um *software* amplamente manipulado tende a se aprimorar com o tempo, devido à contribuição colaborativa de seus usuários (Torvalds, 2001). De maneira análoga, Raymond (1997), ao propor a análise do parâmetro intrínseco, destacava que um dado que é frequentemente atualizado e revisado também passa por um processo contínuo de melhoria. Essa base conceitual ressalta a importância de

observar o parâmetro intrínseco para avaliar a qualidade dos dados, considerando que a interação constante dos usuários pode contribuir para o refinamento da informação. Nos últimos anos, há uma tendência na concentração das pesquisas na análise intrínseca da qualidade de dados VGI. Paiva (2021) utilizou um modelo de inferência espacial e conseguiu prever índices de acurácia posicional medindo a variação de um conjunto de elementos amostrais, permitindo inferir informações importantes sem a necessidade de comparações diretas com dados de referência (acurácia posicional e completude). Pesquisadores como Sehra, Singh e Rai (2017), Brückner et al. (2021) e Elias (2022) têm desenvolvido uma série de ferramentas projetadas para aprimorar a avaliação desses dados, considerando tanto parâmetros extrínsecos quanto intrínsecos. Elias (2022) focou na mensuração da heterogeneidade dos dados utilizando uma medida de qualidade intrínseca que considera o acúmulo de contribuições ao longo do tempo e parâmetros derivados de modelagem matemática para análise de qualidade.

No entanto, apesar dos avanços nessas ferramentas e de seus parâmetros de uso, uma preocupação se destaca com relação às limitações de análise em si para ser possível estimar sua eficácia e confiabilidade. Embora tenham surgido várias ferramentas para avaliação de dados voluntários, ainda não existe uma metodologia consolidada para avaliá-las, compreender suas limitações e determinar até que ponto elas contribuem efetivamente para a mensuração de qualidade de dados voluntários. Essa ausência de avaliações levanta questões sobre a qualidade das análises realizadas e destaca a necessidade premente de estudo nesta área, como a integridade dos resultados, a aplicabilidade em diferentes cenários, possíveis prejuízos devido a análises falhas, entre outros.

Em especial, a análise da componente temporal de contribuições tem sido foco nas pesquisas como em Gröchenig, Brunauer e Rehr (2014), Arsanjani et al. (2015), Viana, Encalada e Rocha (2019), Brückner et al. (2021), Paiva (2021), Grinberger et al. (2021) e Elias (2022), por considerar que os dados têm frequência e número de contribuições variáveis no decorrer do tempo e por razões que geralmente expressam a melhoria contínua de algum dos parâmetros da ISO 19157, bem como as abordagens de modelagem destas contribuições.

Assim, esta pesquisa adotou como hipótese que é possível estabelecer os limites de aplicação de ferramentas de avaliação de qualidade de dados espaciais, por meio da análise, como a propostas por a partir de modelos matemáticos de regressão e correlação dos parâmetros com as características das contribuições, a fim de auxiliar que decisões baseadas em dados voluntários, beneficiando diversos campos de estudo e aplicações práticas.

O estudo de Elias (2022) com foco na análise temporal das contribuições, a pesquisa desenvolveu uma metodologia baseada em Regressão Logística para modelar e avaliar padrões espaço-temporais de contribuições utilizando células de 1x1 km. Os resultados demonstraram que a inclinação da curva de contribuições pode ser usada como um indicador de qualidade intrínseca dos dados, revelando a completude e a acurácia posicional. O estudo reforça a hipótese de que é possível aplicar ferramentas de avaliação de qualidade em dados espaciais voluntários de maneira confiável, trazendo benefícios para diversos campos e aplicações práticas.

1.1 JUSTIFICATIVA

A contribuição do VGI tem se tornado cada vez mais relevante, especialmente em países em desenvolvimento, onde a escassez de dados geoespaciais tradicionais impede o avanço em diversas áreas. A adoção de dados geográficos voluntários é uma solução promissora, no entanto, é essencial considerar a qualidade desses dados para auxiliar sua viabilidade em diferentes contextos de uso.

Esta pesquisa baseia-se na crescente importância das análises de parâmetros utilizados em ferramentas de análise de dados voluntários, que têm sido utilizadas em diversos campos de estudo e aplicações práticas. Embora pesquisas recentes tenham se concentrado na análise da qualidade dos dados voluntários e no desenvolvimento de métodos para aprimorar essa análise, ainda há preocupações e limitações a serem consideradas. Uma dessas limitações refere-se à falta de uma metodologia consolidada para avaliar a eficácia e a confiabilidade, levantando questões importantes sobre a qualidade das análises baseadas em dados voluntários.

Portanto, a pesquisa explorou limitações das ferramentas específicas, utilizando a regressão logística conforme proposto por Brückner et al. (2021) e a medida de qualidade intrínseca introduzida por Elias (2022). Isso foi realizado por meio da análise e correlação dos parâmetros desses modelos com as características das contribuições dos usuários ao mapeamento colaborativo, visando que as decisões baseadas em dados voluntários sejam confiáveis, contribuindo para melhorias nas análises geoespaciais.

Ao preencher essa lacuna de conhecimento sobre a eficácia e a confiabilidade de dados voluntários, tem-se impacto na capacidade de utilização destes dados, não apenas aprimorando a qualidade, mas também nas análises de qualidade de bases cartográficas e atualizações integradas com dados oficiais e VGI.

O cenário atual, caracterizado pelo desenvolvimento de novas ferramentas destinadas a avaliar a qualidade de dados geográficos colaborativos, é promissor, conferindo valorização desses dados e ampliando suas possibilidades de aplicação. No entanto, é imperativo reconhecer a importância de avaliar até que ponto esses dados são confiáveis. A crescente utilização de ferramentas pode, inadvertidamente, prejudicar o trabalho com dados se não forem reconhecidas e compreendidas as suas limitações. Portanto, é fundamental estabelecer até que ponto podemos confiar nessas ferramentas, de forma que sua utilização não comprometa a integridade e a confiabilidade das análises baseadas em dados voluntários. Este aspecto crítico abordado nesta pesquisa, contribui para o avanço do conhecimento e que as ferramentas destinadas a fortalecer o valor dos dados geográficos colaborativos sejam benéficas.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Determinar a influência dos parâmetros da curva de regressão logística aplicada na avaliação da qualidade dos dados VGI e sua relação com as características das contribuições no contexto do *OpenStreetMap* (OSM).

1.2.2 Objetivos Específicos

- a) Analisar o comportamento da curva de regressão logística, em diversos municípios brasileiros, incorporando uma avaliação das tendências observadas;
- b) Avaliar o modelo matemático de regressão logística para diferentes variações temporais e regionais;
- c) Avaliar as informações provenientes da progressão das curvas em diferentes períodos temporais, com dados do *OpenStreetMap*.

2 REVISÃO DE LITERATURA

2.1 MAPEAMENTO COLABORATIVO

O termo "mapeamento colaborativo" é usado para descrever a natureza colaborativa do compartilhamento de informações geográficas em plataformas da *web 2.0* (Rouse et al., 2007; Perkins, 2008; See et al., 2016). Essa expressão implica no fortalecimento de comunidades e cidadãos que anteriormente não tinham participação no processo de criação de informações geográficas (Rouse et al., 2007). O mapeamento colaborativo é um novo contexto de uso dos mapas que demanda atenção dos pesquisadores em Cartografia (Griffin et al., 2017).

No início da década de 2000, a crescente demanda pela utilização de tecnologias da informação instigou o desenvolvimento de soluções colaborativas em diversas aplicações *web* (O'reilly, 2007; Jarret, 2008; West et al., 2012; Jamieson, 2016). Estes autores consideraram a possibilidade dos usuários gerarem seus próprios conteúdos, para compartilhar suas experiências e, voluntariamente, promover a pluralidade e diversidade de informações disponíveis na rede (Cormode e Krishnamurthy, 2008; Jarret, 2008). Por definição, os sistemas idealizados sob o conceito da *web 2.0* são aqueles que permitem que haja maior interação entre usuários e as aplicações *web*, quando comparados com os pioneiros sites da *web 1.0* (O'reilly, 2007; Cormode e Krishnamurthy, 2008; Oxley, 2009; West et al., 2012). Neste caso, é comum que tais soluções tecnológicas incorporem diversos tipos de indivíduos à tarefa de produção do conteúdo (Haklay et al., 2008; West et al., 2012; Jamieson, 2016; Proferes, 2016). Antoniou e Skopeliti (2017) afirmam que a plataforma de mapeamento colaborativo do *OpenStreetMap* (OSM) evidencia a *web 2.0*. A edição e alteração contínuas das feições permitiram que estas evoluíssem de um conjunto de dados com várias camadas e recursos físicos para um conjunto de dados extremamente detalhado.

O termo *geoweb* foi originalmente concebido como conteúdo da *web* referenciado espacialmente e o uso dessa informação geográfica como base para organizar a *web* (Scharl e Tochtermann, 2007), mas agora é usado de forma mais ampla para novas mídias espaciais, novas políticas de conhecimento referenciam novas formas de dados espaciais como 'informações geográficas voluntárias'

(Goodchild, 2007); novas práticas de curadoria e interação com conteúdo locacional na *web*; e os objetos de hardware/software, dispositivos e padrões técnicos que os facilitam (Elwood e Leszczynski, 2011). Ou seja, a *geoweb* é constituída por novas formas de conteúdo, novas práticas de dados e novas mídias espaciais.

Comparado ao mapeamento tradicional, os méritos do mapeamento na *web* podem ser resumidos como: mantém o papel tradicional de fornecer informações sobre padrões e relações espaciais; os usuários finais não são mais analistas de SIG predominantemente profissionais, mas usuários públicos que não possuem conhecimento específico de SIG (Skarlatidou e Haklay, 2006); pode ser acessado quando e onde a internet estiver disponível; o conteúdo do mapa pode ser uma variedade de recursos de dados dispersos pela Internet; e permite que os usuários interajam com mapas online.

O mapeamento na *web* integra e visualiza dados e serviços espaciais com base em bibliotecas de funções *online* gratuitas, que oferecem funções por Interfaces de Programação de Aplicativos (APIs) (Bildirici e Ulugtekin, 2010). Há um crescente corpo de pesquisa nas novas formas de conteúdo da *geoweb*, técnicas de validá-los ou analisá-los e o significado de novas práticas de dados, como *crowdsourcing* (Elwood, Goodchild e Sui, 2012), essa prática é considerada uma atividade participativa de natureza aberta, na qual qualquer indivíduo tem a oportunidade de propor tarefas específicas de coleta de dados, que podem ser executadas por várias pessoas (Papapesios et al., 2019).

2.1.1 VGI e Qualidade dos Dados Geoespaciais

Em 2007, o termo *Volunteered Geographic Information* (VGI) começou a ser amplamente empregado para representar o mapeamento colaborativo, conforme destacado por Goodchild (2007). Descreve o fornecimento de informações geográficas de forma voluntária por parte de indivíduos de dados geoespaciais, destacando a importância das plataformas online desempenham um papel crucial ao possibilitar que cidadãos compartilhem conteúdo, formando assim uma espécie de mosaico global composto por diversas partes. É também referido como *crowdsourcing* o uso de ferramentas digitais para coletar, analisar e disseminar informações geográficas fornecidas por indivíduos (Ferster et al., 2020); ou então

colaborativamente coletadas, se tornou um dos tópicos de pesquisa mais importantes na ciência que engloba os dados geoespaciais, onde sua importância foi reconhecida não só acadêmicos, mas também por profissionais (Yan, et al., 2020).

De acordo com Goodchild (2007), o surgimento das atividades ligadas às plataformas VGI foi catalisado por evoluções tecnológicas e sociais no início do século XXI. A popularização dos mapas online e da conectividade global da internet democratizou o acesso à informação geográfica. Simultaneamente, a transição para a *web 2.0*, que enfatiza a interatividade e colaboração, permitiu aos usuários não só consumir, mas também contribuir com dados geoespaciais. Neste cenário, o VGI emergiu como uma ponte entre cidadãos comuns e o mapeamento tradicional, ampliando significativamente o espectro de contribuições geográficas.

Howe (2006) relata que o crescimento das plataformas de *crowdsourcing* desempenhou um papel crucial nesse cenário. Estas plataformas são reconhecidas como fontes de obtenção de dados, permitindo a coleta de dados por meio de sistemas informatizados. É importante salientar que nelas, indivíduos, independentemente de sua formação técnica, têm a capacidade de contribuir com informações.

O desafio reside não só em motivar contribuições nas plataformas VGI, mas também em integrar essas informações, muitas vezes simples, com dados mais complexos mantidos por instituições, sejam elas públicas ou privadas (Davis et al., 2009). Sem uma integração eficaz, os dados coletados podem permanecer fragmentados e subutilizados, enquanto uma boa integração pode potencializar seu valor, permitindo análises mais completas e decisões mais informadas em diversos campos de aplicação.

As informações cartográficas passaram a ser parte do cotidiano da sociedade, tais como a utilização de mapas e imagens orbitais, disponíveis em diversas plataformas, como *Google Earth*, *Google Maps*, *Bing Maps*, *OpenStreetMap* e outros; de receptores GNSS (*Global Navigation Satellite System*), disponíveis em celulares e de sistemas de roteamento e navegação terrestres para automóveis. Atualmente a informação espacial pode ser produzida e apresentada por usuários comuns que tenham acesso a um computador ou *smartphone* e à internet (Bravo e Sluter, 2015).

Goodchild (2007) descreveu este tipo de voluntários: “são na maioria destreinados e suas ações são quase sempre voluntárias, e os resultados podem ou não ser precisos. Mas coletivamente, eles representam uma inovação dramática que certamente teve impactos profundos nos Sistemas de Informação Geográfica (SIG) da época”. Ou seja, embora os dados geográficos voluntários possam ter limitações individuais, seu resultado no aspecto coletivo é significativo, podem trazer informações valiosas para os SIGs, ampliando a quantidade e a diversidade de dados disponíveis. Mesmo que sua precisão possa variar, a agregação desses dados ainda assim pode proporcionar uma visão mais abrangente e detalhada do ambiente geográfico.

Também é pertinente reconhecer as limitações associadas ao uso de dados geográficos voluntários, é necessário considerar sua dependência de contribuições voluntárias, o que pode resultar em lacunas ou discrepâncias nos dados, especialmente em regiões com níveis reduzidos de participação comunitária ou conhecimento técnico (Goodchild, 2007). Apesar disso, continuam sendo uma fonte valiosa e eficiente de dados para a análise de informações, fornecendo uma vasta gama de dados geoespaciais que podem ser utilizados para aprimorar a compreensão da distribuição espacial.

Pesquisas têm sido realizadas em relação à avaliação da qualidade global dos dados produzidos pelos cidadãos, com o intuito de utilizá-los na atualização do mapeamento de referência em vários países (Touya et al., 2017; Johnson e Sieber, 2012). Essas pesquisas incluem a compatibilização do mapeamento colaborativo com o mapeamento de referência urbano (Touya et al., 2017; Johnson e Sieber, 2012; Bearden, 2007), bem como a sua utilização em países, como o Brasil (Sluter e Camboim, 2009; Estes e Mooneyhan, 1994).

Ao escolher um sistema de mapeamento baseado na *web* de acesso gratuito para navegação, serviços baseados em localização ou informações gerais, existem três serviços dominantes: *Google Maps*, *Bing Maps* e *OpenStreetMap*. Essas plataformas de mapeamento na *web* são as fontes de mapas populares para uso em serviços baseados em localização com ênfase específica na navegação de pedestres, aplicativos de guia turístico e outros aplicativos de pesquisa baseados em localização (Ciepluch et al., 2010).

Em meio à necessidade de compreensão do espaço geográfico e de suas representações, as técnicas para manipulação de dados geoespaciais têm sido bastante aprimoradas ao longo dos últimos anos, impulsionadas pela necessidade de criação e aperfeiçoamento de mecanismos que auxiliassem as tomadas de decisão sobre o território (Câmara et al., 2001; Rocha, 2003; Rosa, 2005; Longley, 2009).

Devillers e Jeansoulin (2006) apontaram a existência de problemas relacionados com a avaliação da qualidade em dados geoespaciais e destacaram a intensificação destes a partir dos avanços tecnológicos e a disponibilização de informações espaciais em plataformas *web*, além das diferentes aplicações relacionadas ao uso dos dados espaciais em meio digital. Anteriormente, a qualidade de um dado geoespacial era determinada principalmente pela acurácia espacial dos dados coletados, ou seja, o quão precisas eram as coordenadas geográficas de certas feições. Porém, o conceito de qualidade vem sendo associado a uma ampla variedade de parâmetros, desde o processo de aquisição e gerenciamento dos dados até a comunicação e uso dos dados geoespaciais (Câmara, 2001).

Visando simplificar as avaliações relacionadas à qualidade dos dados espaciais, diretrizes nacionais, como a ET-CQDG e internacionais, como a ISO:19157, buscam categorizá-los. Essas diretrizes e especificações fornecem métodos para atribuir crédito às informações, atribuindo-lhes diferentes níveis de confiabilidade. Em sua maioria, essas diretrizes avaliam a confiabilidade comparando o produto em análise com amostras da realidade ou com outros produtos considerados mais confiáveis, geralmente sendo bases de referência oficiais (ISO 19.157, 2013).

A ISO formou um Comitê Técnico voltado para as informações geográficas, o que viabilizou o desenvolvimento da série ISO 19.100, que estabelece padrões abrangentes para a definição, estruturação e gerenciamento de dados geoespaciais e metadados, garantindo interoperabilidade, qualidade e consistência na troca de informações geográficas (Albakri, 2012). A partir desta série é possível citar três padrões que estão diretamente relacionados com procedimentos para a avaliação da qualidade de dados, sendo estas: A ISO 19113 – *Quality Principles* (ISO, 2002),

ISO 19114 – *Quality Evaluation Procedures* (ISO, 2003) e ISO:19157 - *Data Quality* (ISO, 2013).

A qualidade atrelada às informações geográficas, segundo a ISO:19157, é decomposta em dois tipos: qualidade posicional e qualidade semântica. Portanto, pode-se dizer que a qualidade posicional é mensurada por meio da averiguação da acurácia posicional associada à feição cartográfica. Nesse sentido, a ISO:19157 (ISO, 2013) define:

- Acurácia Posicional: acurácia da posição das feições num determinado sistema de referência espacial.

Adicionalmente, tem-se os elementos indicadores da qualidade semântica num outro campo, cinco parâmetros de análise a ela associados. Suas componentes e respectivas definições, segundo a ISO:19157 (ISO, 2013) são as que seguem:

- Consistência Lógica: grau de aderência às regras lógicas da estrutura de dados, atribuição e relacionamentos;
- Usabilidade: é baseada nos requisitos dos usuários e na aderência, com as informações às necessidades destes;
- Completude: presença ou falta de feições, seus atributos ou relacionamentos;
- Acurácia Temática: acurácia dos atributos quantitativos, quão corretos são os não quantitativos e as classificações das feições e seus relacionamentos.
- Qualidade Temporal: qualidade dos atributos temporais e dos relacionamentos temporais entre feições.

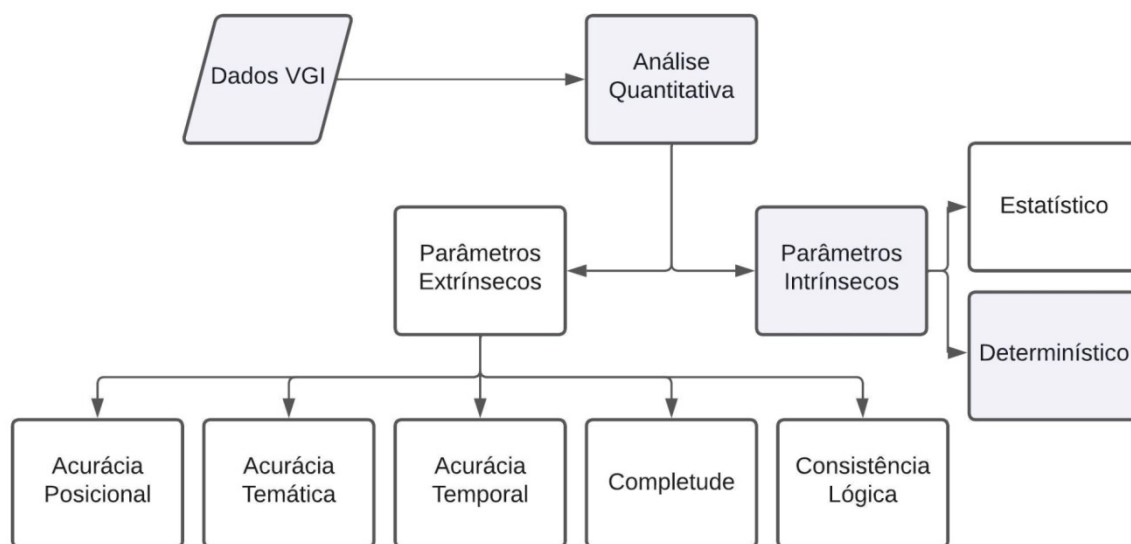
O plano de ação da Infraestrutura Nacional de Dados Espaciais (INDE) prevê uma série de normas, padrões e especificações para mapeamento. Uma dessas especificações que auxilia é a Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG).

A ET-EDGV estabelece padrões para a produção de produtos cartográficos vetoriais e matriciais (BRASIL, 2010), que se complementam com os padrões nacionais voltados à qualidade de dados geoespaciais (DSG, 2015). No contexto brasileiro, a ET-CQDG desempenha um papel crucial na definição de padrões para a

avaliação da qualidade de dados geoespaciais, alinhando-se com as diretrizes da ISO:19157 (ISO, 2013). Além disso, o Manual Técnico em Geociências do IBGE, que aborda a Avaliação da Qualidade de Dados Geoespaciais, apresenta uma abordagem abrangente, com sua primeira versão desenvolvida em 2017 e uma segunda edição em 2019. Ambos os padrões citados são fundamentados nas definições de indicadores estabelecidos pela ISO:19157, evidenciando a coesão e a convergência entre as normativas brasileiras e internacionais.

As características originadas da Informação Geográfica Voluntária (VGI) utilizadas para os dados de referência, estão sendo relacionadas à estudos em âmbito global com o intuito de avaliar a qualidade desses dados e estabelecer parâmetros que possibilitem sua integração eficaz com mapeamentos oriundos de outras fontes de dados. Este processo de avaliação da qualidade no VGI, conforme já citado e demonstrado a seguir no fluxograma da Figura 1, quando analisado por parâmetros intrínsecos, pode ocorrer por meio de abordagens estatísticas ou determinísticas, está segunda opção foi a qual esta pesquisa conduziu suas análises, como indicado no caminho em azul do fluxograma.

Figura 1 - Etapas da avaliação quantitativa no VGI



Fonte: Autor (2024)

A qualidade é um aspecto importante a ser considerado nos dados geoespaciais, pois permite determinar sua adequação para uso para fins específicos. Conforme já dito, no VGI essa questão torna-se ainda mais relevante

porque os dados possuem aspectos heterogêneos, pois podem variar conforme a localização, características das colaborações e perfil dos colaboradores. Portanto, os procedimentos de avaliação da qualidade no VGI são descritos a partir de abordagens extrínsecas, por meio de parâmetros estabelecidos pela ISO:19157 (ISO, 2013) (Brovelli e Zamboni, 2018; Zhang e Malczewski, 2017; Haklay, 2010) e na intrínseca (Sehra, Singh e Rai, 2017), caracterizada pelo histórico de edições, pelo número de colaboradores e contribuições; ou ainda combinando parâmetros extrínsecos e intrínsecos (Nasiri et al., 2018).

Um dos estudos pioneiros associados à análise de dados colaborativos foi conduzido por Haklay (2010), no qual a acurácia posicional e a completude dos dados da plataforma VGI do *OpenStreetMap* (OSM) foram avaliadas em relação ao mapeamento de referência do Reino Unido (*Ordnance Survey*), antes disso ainda Flanagan e Metzger (2008) conduziram uma análise sobre a credibilidade vinculada às informações voluntariamente geradas pelo usuário (VGI), identificaram avanços e melhorias nas informações geográficas e ressaltaram as vantagens decorrentes da disponibilidade de dados geoespaciais, enfatizando seu papel fundamental no aprofundamento do entendimento de aspectos específicos de um local determinado. Além disso, outras pesquisas foram realizadas em diferentes partes do mundo, abordando questões semelhantes, como Wagle e Archarya (2020), visando contrastar abordagens anteriores e futuras para a atualização dos Mapas Topográficos de Base do Nepal, foi proposto o uso do VGI como uma tendência global na atualização cartográfica, devido à frequência consistente na adição de características geográficas e às vantagens econômicas associadas a essa abordagem.

No contexto das discussões sobre a utilização de informações provenientes do VGI, o tópico central de debate concentra-se principalmente na qualidade desses dados. As investigações enfatizam geralmente a relevância da avaliação do VGI, uma vez que, como mencionado, essas informações não obedecem aos mesmos padrões técnicos rigorosos encontrados no mapeamento de referência. Além dos já citados, são abordados em trabalhos como Goodchild e Glennon (2010), AlBakri (2012), Goodchild e Li (2012), Neis e Zipf (2012), Elwood, Goodchild e Sui (2012), Bégin, Devillers e Roche (2013), Senaratne et al., (2017), Seto, Kanasugi e Nishimura (2020).

Como previamente mencionado, uma vez que as características provenientes do VGI não aderem aos mesmos critérios de produção utilizados para dados de referência, pesquisas estão em andamento em várias partes do mundo, pois em algumas situações, os dados de referência podem estar indisponíveis devido a custos elevados ou restrições de licenciamento como apontado por Zhou (2018). Como resultado, algumas pesquisas têm se voltado para a avaliação da qualidade dos dados de VGI sem considerar fatores externos, ou seja, de forma intrínseca.

Pesquisas como Dror, Doytsher e Daltoy (2021), Nasiri et al. (2018) e Zhou (2018) têm se concentrado em analisar a qualidade do VGI ao considerar tanto parâmetros intrínsecos e extrínsecos, visando compreender as peculiaridades do VGI e determinar se a integração de diversos fatores resulta em uma melhoria na qualidade. Barron, Neis e Zipf (2014) também analisaram indicadores intrínsecos e extrínsecos de qualidade para avaliar um conjunto de dados OSM, indicando a viabilidade do uso dos indicadores intrínsecos, principalmente quando o conceito de qualidade não está rigidamente ligado a parâmetros oficiais de avaliação.

2.1.1.1 Parâmetros extrínsecos

Enquanto o Parâmetro Intrínseco se concentra nos atributos fundamentais dos dados, como precisão e confiabilidade, o Parâmetro Extrínseco examina como esses dados se encaixam no contexto mais amplo e atendem às necessidades dos usuários e aplicações. Esta dimensão abrange aspectos como a relevância, atualidade e interoperabilidade dos dados colaborativos geográficos. Para compreender plenamente o valor dos dados VGI, é essencial analisar não apenas sua qualidade intrínseca, mas também como eles se alinham com as demandas do mundo real e com outros sistemas de informação geográfica.

Haklay (2008) enfatiza que a relevância é crucial na avaliação da qualidade extrínseca dos dados colaborativos, pois os dados devem ser pertinentes para as aplicações em que são utilizados. Ele destaca a importância da relevância dos dados colaborativos geográficos para as aplicações do mundo real, argumentando que a qualidade dos parâmetros extrínsecos está diretamente relacionada à capacidade dos dados de atender às necessidades dos usuários finais. Assim como Goodchild e Glennon (2010) abordaram a confiabilidade dos dados VGI em

contextos científicos e de pesquisa, destacando a essencialidade de garantir que os dados colaborativos geográficos sejam considerados confiáveis o suficiente para serem utilizados em estudos acadêmicos e pesquisas de alta qualidade.

Pesquisas no Brasil também têm investigado os parâmetros extrínsecos dos dados do OSM. Por exemplo, Almeida et al. (2021) avaliaram a qualidade posicional dos dados do OSM na região central de Viçosa, MG, destacando a precisão dos dados rodoviários em relação aos dados oficiais. Santos e Bernardes (2020) examinaram a qualidade dos dados do OSM no contexto de planejamento urbano em Campinas/SP, identificando desafios e oportunidades na integração desses dados com sistemas de informações geográficas municipais. Outra pesquisa realizada por Santos et al. (2017) focou na qualidade posicional dos dados do OSM na região de Viçosa/MG, apontando a necessidade de metodologias específicas para garantir a precisão e a confiabilidade dos dados colaborativos geográficos. Tardin et al. (2019) discutiram a aplicabilidade dos dados do OSM em projetos de planejamento urbano, enfatizando como o parâmetro extrínseco dos dados pode ser medido pela capacidade desses dados de serem integrados com sistemas de informações existentes. Pacheco e Oliveira (2020) analisaram a acurácia e a completude dos dados do OSM para usos específicos em sistemas de transporte urbano, evidenciando a necessidade de avaliações contínuas para garantir a qualidade dos dados utilizados.

Além disso, Elias e Fernandes (2021) avaliaram a qualidade dos dados geoespaciais do OSM para os indicadores de acurácia posicional, acurácia temática e completude, destacando a importância desses parâmetros para a utilização prática dos dados em diversas aplicações. Em uma avaliação preliminar da qualidade do OSM em uma grande metrópole no nordeste do Brasil, De Souza e Schmidt (2023) focaram na precisão dos dados geoespaciais para eixos rodoviários, demonstrando a viabilidade do uso desses dados em contextos urbanos.

A aplicabilidade dos dados VGI também é discutida em estudos como o de Miller (2013), que destaca a importância da atualidade dos dados em contextos de transporte e mobilidade. Ele observa que, em sistemas de navegação baseados em colaboração, como o *Waze*, a qualidade extrínseca é evidenciada pela rapidez com que as informações de tráfego são atualizadas e disponibilizadas aos motoristas.

Teixeira e Schmidt (2020) realizou uma análise detalhada da acurácia posicional dos dados colaborativos do OSM em Uberlândia/MG, especificamente para eixos viários de bairro, ressaltando a relevância da precisão posicional para a confiabilidade dos dados em aplicações locais. Seguindo a mesma linha de pesquisas, Elias et al. (2020) também avaliaram a qualidade dos dados de eixos rodoviários utilizando a plataforma colaborativa OSM. Verificaram a completude e acurácia temática, concluindo que os dados colaborativos complementam a cartografia de referência em diversas análises. Isso reforça a importância da qualidade extrínseca dos dados onde a capacidade de integração com sistemas de informações existentes é crucial para a tomada de decisões.

2.1.1.2 Parâmetros intrínsecos

Os parâmetros intrínsecos podem ser entendidos como ferramentas de avaliação da qualidade VGI que não usam métodos comparativos para retratar a confiabilidade dos dados colaborativos (Mooney e Corcoran, 2014), esses parâmetros, que incluem precisão, exatidão, confiabilidade, completude, atualidade e relevância. Os parâmetros intrínsecos caracterizados pelo número de contribuições, contribuidores e histórico de edições foram baseados na Lei Linus. Esta lei aplicada ao contexto VGI, conforme abordado por Haklay et al. (2010), afirma que à medida que aumenta a quantidade de contribuintes em uma determinada unidade espacial, maior e mais conhecida é a sua qualidade. A análise destes parâmetros não apenas amplia a credibilidade das informações geográficas geradas por VGI, mas também desempenha um papel crucial na determinação do seu potencial de uso.

No que concerne aos parâmetros intrínsecos para a avaliação da qualidade dos dados VGI, informações relativas à quantidade de entidades geográficas representadas, os períodos nos quais as últimas modificações foram efetuadas, a quantidade de edições realizadas e o número de colaboradores envolvidos, podem fornecer aspectos significativos em relação ao comportamento espacial das atualizações desencadeadas, permitindo a análise de sua correlação com outros fatores relevantes, como pode ser observado nos estudos realizados, por exemplo, por Tian, Zhou e Fu (2019) e Hecht, Kunze e Hahmann (2013).

A avaliação de parâmetros intrínsecos é ainda mais relevante em países em desenvolvimento, onde a crônica falta de recursos para cartografia muitas vezes resulta na falta de dados atualizados para fornecer uma comparação. Além disso, essa situação torna os dados provenientes do VGI ainda mais necessários para complementar os mapeamentos de referência existentes (Camboim, Bravo e Sluter, 2015). No início da pesquisa sobre os parâmetros intrínsecos, a variável única considerada era a data da última edição dos dados, no entanto, Paiva e Camboim (2021) demonstraram que essa abordagem isolada não é suficiente para explicar integralmente o comportamento da qualidade na VGI. Esse achado está em consonância com estudos recentes que se concentram na análise de padrões temporais, visando a aprimorar a compreensão da distribuição espacial da qualidade dos dados geoespaciais na VGI, como observado em pesquisas de Le Guilcher, Olteanu-Raimond e Balde (2022), Grinberger et al. (2021), Brückner et al. (2021), Witt, Loos e Zipf (2021), Arsanjani e outros (2015) e Gröching, Brunauer e Rehr (2014).

Autores como Goodchild (2007) e Haklay et al. (2008) também destacaram a importância de avaliar a qualidade intrínseca dos dados colaborativos, enfatizando a necessidade de desenvolver métodos e métricas específicas para abordar essa questão. Zielstra e Zipf (2010) enfatizam a importância da confiabilidade dos dados colaborativos em projetos de mapeamento participativo, como o *OpenStreetMap* (OSM). Ele destaca que a qualidade intrínseca pode ser afetada por erros humanos e falta de padronização na coleta de dados, exigindo abordagens avançadas de validação e correção

Um exemplo notável de pesquisa que aborda a qualidade intrínseca dos dados é o estudo de Neis e Zielstra (2014), que examinou a precisão das informações de localização coletadas por meio do projeto *OpenStreetMap* (OSM). Eles analisaram a qualidade intrínseca dos dados, identificando desafios e oportunidades na avaliação e melhoria da precisão dos dados VGI. Além disso, a obra de Steinmann et al. (2013) contribuiu para o desenvolvimento de métricas e metodologias para avaliar a confiabilidade dos dados em contextos específicos, como mapeamento participativo.

No estudo de Westerholt et al. (2016) sobre a qualidade dos dados de VGI, os autores ressaltam que a qualidade intrínseca dos dados colaborativos pode variar conforme o tipo de informação geográfica. Soares (2018) conduziu pesquisas sobre a completude dos dados VGI em relação à infraestrutura urbana. Seu trabalho demonstrou que, em alguns casos, os dados colaborativos podem ser incompletos ou tendenciosos em termos de representação geográfica, sublinhando a importância de considerar a qualidade intrínseca para identificar potenciais lacunas nos dados.

Bégin, Devillers e Roche (2013) estudaram o comportamento de mapeamento dos contribuidores de VGI para avaliar a completude. Concluíram que a análise dos processos de mapeamento de apenas alguns contribuidores poderia representar a maioria dos conjuntos de dados. Budhathoki e Haythornthwaite (2013) introduziram como medidas o número de artigos editados, os dias de contribuição e a longevidade das contribuições como indicadores da experiência dos contribuidores, utilizando questionários para investigar as motivações e características das contribuições do OSM. Mooney e Corcoran (2012) concentraram suas análises em recursos do OSM que foram editados mais de 15 vezes e argumentaram que o histórico de edição dos dados do OSM é a base para avaliar sua qualidade.

Além disso, nota-se que a heterogeneidade no VGI tem feito com que os estudos mais recentes investiguem as características intrínsecas destas informações associadas com o histórico de edições, dinâmica e padrões de contribuições e quantidade dos usuários no intuito de entender como estes aspectos influenciam a qualidade extrínseca do VGI.

Como pesquisas recentes nesta temática, podemos citar Paiva e Camboim (2022) que avaliaram a qualidade posicional do OSM com base em parâmetros intrínsecos, Silva e Camboim (2021) e Machado e Camboim (2019), que realizaram uma compatibilização semântica de estruturação dos dados colaborativos da plataforma OSM em relação a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV) e Elias (2023), que realizou estudo da relação entre os parâmetros obtidos na modelagem matemática (regressão logística) de feições e a completude dos dados no OSM.

De Oliveira et al. (2022) investigaram a avaliação intrínseca dos dados do OSM e os aspectos espaço-temporais no caso do rompimento da barragem de

Brumadinho destacando a importância da acurácia e da completude dos dados para aplicações de monitoramento de desastres e evidenciando como dados colaborativos podem ser críticos em situações de emergência. Elias et al. (2021) exploraram padrões espaço temporais das contribuições em áreas urbanas heterogêneas, demonstrando como a densidade de dados e a variabilidade temporal podem influenciar na qualidade e utilidade das informações geográficas coletadas, demonstrando a necessidade de entender as dinâmicas de contribuição para melhorar a precisão dos dados e sua aplicabilidade em projetos urbanos. Além disso, Elias et al. (2021) em outro estudo preliminar avaliaram os padrões espaço temporais na qualidade dos dados colaborativos de mapeamento, reforçando a importância de análises detalhadas para identificar e mitigar problemas de qualidade, especialmente em áreas com alta heterogeneidade de informações.

Estas análises não apenas avaliam as informações geográficas geradas por VGI, mas também são cruciais para ampliar o seu potencial de uso em várias aplicações, incluindo planejamento urbano, gestão de recursos naturais e tomada de decisões. Portanto, desempenham papel fundamental na compreensão e aplicação eficaz dos dados colaborativos geográficos.

2.1.2 *OpenStreetMap* (OSM)

OpenStreetMap é exemplo representativo de tecnologia de mapeamento colaborativo/VGI, destaca-se como uma iniciativa inovadora que visa criar e distribuir dados geográficos gratuitos para o mundo. Foi desenvolvido no ano de 2004 visando se criar um mapa livre e editável do mundo construído por voluntários, e lançado com uma licença de conteúdo aberto (OSM, 2023). O Reino Unido foi a sede inicial do mapeamento pelo OSM, já que o *Ordnance Survey* (OS), agência oficial de mapeamento da Grã-Bretanha, produziu conjuntos de dados geoespaciais, mas, a exemplo de outras agências no mundo, não os distribuía gratuitamente (Machado, 2020).

Uma das características distintivas do OSM é constituído por dados abertos, são disponibilizados gratuitamente e podem ser utilizados, editados e compartilhados livremente por qualquer pessoa. Isso permite que desenvolvedores, pesquisadores, bem como qualquer usuário, tenham a liberdade de acessar,

modificar e adaptar os dados de mapeamento de acordo com suas necessidades específicas. Embora sua qualidade seja heterogênea, estudos têm demonstrado que ela é satisfatória para muitos usos. Por exemplo, Touya et al., (2017) avaliou a qualidade do OSM ao longo do tempo a partir da combinação de dados de referência, histórico de edições e relações espaciais. Johnson e Sieber (2012), que analisaram processo de coleta de VGI para enfrentar os desafios relacionados aos custos, precisão e jurisdição, permitindo que os governos adotassem efetivamente essa fonte colaborativa de dados geográficos, entre outros como Camboim e Sluter (2009); Bearden (2007); Estes e Mooneyhan (1994).

Devido à natureza colaborativa do projeto, o *OpenStreetMap* atrai uma comunidade ativa de voluntários que se envolvem na coleta e atualização de dados geográficos. Essa participação colaborativa contribui para um rápido crescimento e enriquecimento contínuo dos mapas do OSM. Além disso, oferece uma ampla gama de aplicações, desde navegação até análise de dados geoespaciais. Possui ainda um significativo potencial para atrair voluntários de regiões menos desenvolvidas do planeta, em que a obtenção de dados pode ser mais difícil para a maioria das empresas de mapeamento comercial (Sehra, Singh e Rai, 2014).

A flexibilidade e personalização oferecidas pelo OSM permite a criação de aplicações e soluções específicas, adaptadas às necessidades do usuário. Desde aplicativos de navegação até análise de dados geoespaciais, o OSM tem sido aplicado em diversas áreas, demonstrando sua versatilidade e utilidade em várias disciplinas e setores. Em muitos casos, pesquisadores investigaram os dados e as informações dos colaboradores do projeto *OpenStreetMap*, um dos projetos VGI de maior sucesso nos últimos anos (Goetz, 2012; Mooney, Corcoran e Winstanley, 2010; Neis, Zielstra e Zipf, 2011), que também tem sido frequentemente citado na comunidade SIG (Goodchild, 2007, Budhathoki e Haythornthwaite, 2013).

Atualmente, o projeto VGI ativo mais difundido é o *OpenStreetMap*, pois conta com uma comunidade numerosa de usuários, conectando mais de 10,5 milhões de usuários registrados que contribuem para o projeto em algum nível, incluindo cerca de 20,2 bilhões de pontos GPS carregados, 8,4 bilhões de nós, mais de 945 milhões de caminhos e 11,0 milhões de tags e relações (OSM, 2023). Conforme já dito, todos os mapas, dados e metadados ofertados pelo OSM são

abertos, disponíveis sob uma licença *Open Database License*, e são formalmente operados pela *OpenStreetMap Foundation* (OSMF), em nome da comunidade de mapeadores.

2.2 MÉTODOS DE MODELAGEM DE PARÂMETROS INTRÍNSECOS

Nasiri et al. (2018) apontam que analisar, bem como avaliar qualidade, pode ocorrer a partir de diferentes tipos de informações. O comportamento da função logística, quando relacionado às edições em plataformas de VGI, pode ser delineado nas seguintes etapas: o início, o crescimento e a saturação.

Um exemplo que ilustra a aplicação de análises e avaliações semelhantes é o estudo realizado por Grinberger et al. (2021). Nesse estudo, os pesquisadores monitoraram as contribuições ao longo do tempo no contexto do *OpenStreetMap* a fim de identificar eventos e tendências significativas. Para modelar o comportamento dos dados, eles adotaram uma abordagem baseada na função logística, estabelecendo uma relação entre as características da curva resultante e os princípios estabelecidos por Gröchenig, Brunauer e Rehr (2014), os quais propuseram uma metodologia para identificar variações regionais e temporais associadas à evolução do mapeamento no OSM. O modelo desenvolvido por Grinberger et al. permitiu a classificação dos diferentes estágios de atividade nas contribuições realizadas no OSM. Brückner et al. (2021) estimou a completude das lojas de varejo no OSM na Alemanha, usando a regressão logística.

Elias (2022) extraiu o histórico da quantidade de feições entre 2008 e 2022, procedendo com a modelagem matemática dos dados a partir da Regressão Logística. A adequação do modelo foi verificada para a contagem acumulada de feições pontuais, lineares e poligonais mapeadas em uma determinada área e a medida do menor tamanho possível do retângulo circunvizinho. Elias (2022) desenvolveu uma metodologia para modelar e avaliar os padrões espaço-temporais dos indicadores de qualidade intrínseca dos dados da plataforma OSM e sua relação com os indicadores de qualidade tradicionais; utilizou a regressão logística em uma resolução espacial de 1x1 km, obtendo resultados que destacaram a diferenciação de regiões com contribuições significativas, a sinergia entre o OSM e os dados

oficiais e a relação entre a completude dos dados e a saturação da curva, especialmente nas análises dos eixos viários.

Nestes casos citados, os autores nortearam a pesquisa na quantidade acumulada de contribuições ao longo do tempo e utilizaram o modelo de regressão logística em seus dados, uma abordagem que permite descrever o comportamento dos dados por meio de uma curva sigmoidal “S”. Essa curva pode ser associada ao padrão de contribuições em uma área específica, onde inicialmente há poucas contribuições, seguidas por um crescimento gradual que eventualmente se estabiliza com o tempo. Além disso, Brückner et al. (2021) também exploraram outros modelos de regressão, incluindo a regressão logística de três parâmetros e a hipérbole retangular. Optaram por curvas sigmóides, como a função logística de três e quatro parâmetros, associadas a processos de mapeamento em três fases. Além disso, consideraram curvas de crescimento não logísticas para representar processos de mapeamento sem a fase inicial de crescimento lento. A escolha da função logística baseou-se em critérios de ajuste e confiabilidade, considerando o declínio no crescimento como critério fundamental.

2.3 ERRO MÉDIO QUADRÁTICO (RMSE)

A aplicação do RMSE na avaliação da qualidade de VGI, para compreender a confiabilidade e a acurácia dos dados geográficos coletados por voluntários (Neis e Zielstra, 2014). Neis e Zielstra (2014) aplicaram o RMSE para comparar dados do *OpenStreetMap* com fontes oficiais, destacando a importância desta métrica na detecção de erros espaciais. Por meio da comparação entre os dados do OSM, o RMSE permite quantificar a discrepância espacial entre as informações fornecidas pelos voluntários e a realidade observada (Haklay e Weber, 2008). Assim, ao integrar o RMSE na análise da qualidade de VGI, os pesquisadores e os gestores podem identificar áreas de melhorias, entender os padrões de erro e incerteza, e promover aprimoramentos contínuos nos dados colaborativos (Goodchild, 2007).

É importante ressaltar que o RMSE não deve ser utilizado isoladamente como única medida de qualidade, mas sim como parte de uma abordagem multidimensional que considere diferentes aspectos da qualidade de dados, como completude, precisão posicional e atualidade (Senaratne et al., 2017). Além disso,

Chen et al. (2013) enfatizaram a utilização do RMSE para penalizar grandes discrepâncias entre valores previstos e observados, destacando sua aplicabilidade em situações específicas de análise de dados VGI. No entanto, é crucial ter em mente que o RMSE pode ser sensível a valores extremos, sendo necessário avaliar sua robustez em diferentes contextos de aplicação (López-de-Ipiña et al., 2011).

Fonte et al. (2017) empregaram regressão logística para analisar dados do OSM, demonstrando como variáveis como densidade populacional e acesso à internet afetam a precisão dos dados geográficos. Eles mostraram como a regressão logística pode ser uma ferramenta para compreender os fatores que influenciam a qualidade dos dados geográficos colaborativos, permitindo melhorias direcionadas.

Zielstra e Hochmair (2011) combinaram RMSE e regressão logística para avaliar a qualidade dos dados VGI em áreas urbanas e rurais, identificando padrões específicos de erro e sugerindo melhorias, assim como Antoniou et al. (2016) que investigaram a precisão posicional dos dados VGI utilizando regressão logística, considerando a experiência dos voluntários e a intensidade das contribuições como fatores cruciais. Eles destacaram a importância de considerar não apenas os dados em si, mas também o contexto e o perfil dos contribuidores ao avaliar a qualidade dos dados geográficos colaborativos.

O RMSE é empregado em diferentes situações para análise, Tsou et al. (2015) por exemplo empregaram como parte do PEC-PCD para avaliar a qualidade dos dados VGI em áreas urbanas, enquanto Yang et al. (2018) utilizaram para comparar dados VGI com dados oficiais em uma área rural e Liu et al. (2020) integraram o RMSE na avaliação da precisão dos dados VGI usando o Método PEC-PCD em um estudo de área metropolitana. Todos eles mostraram como o RMSE foi uma ferramenta eficaz para identificar e quantificar erros espaciais nos dados colaborativos.

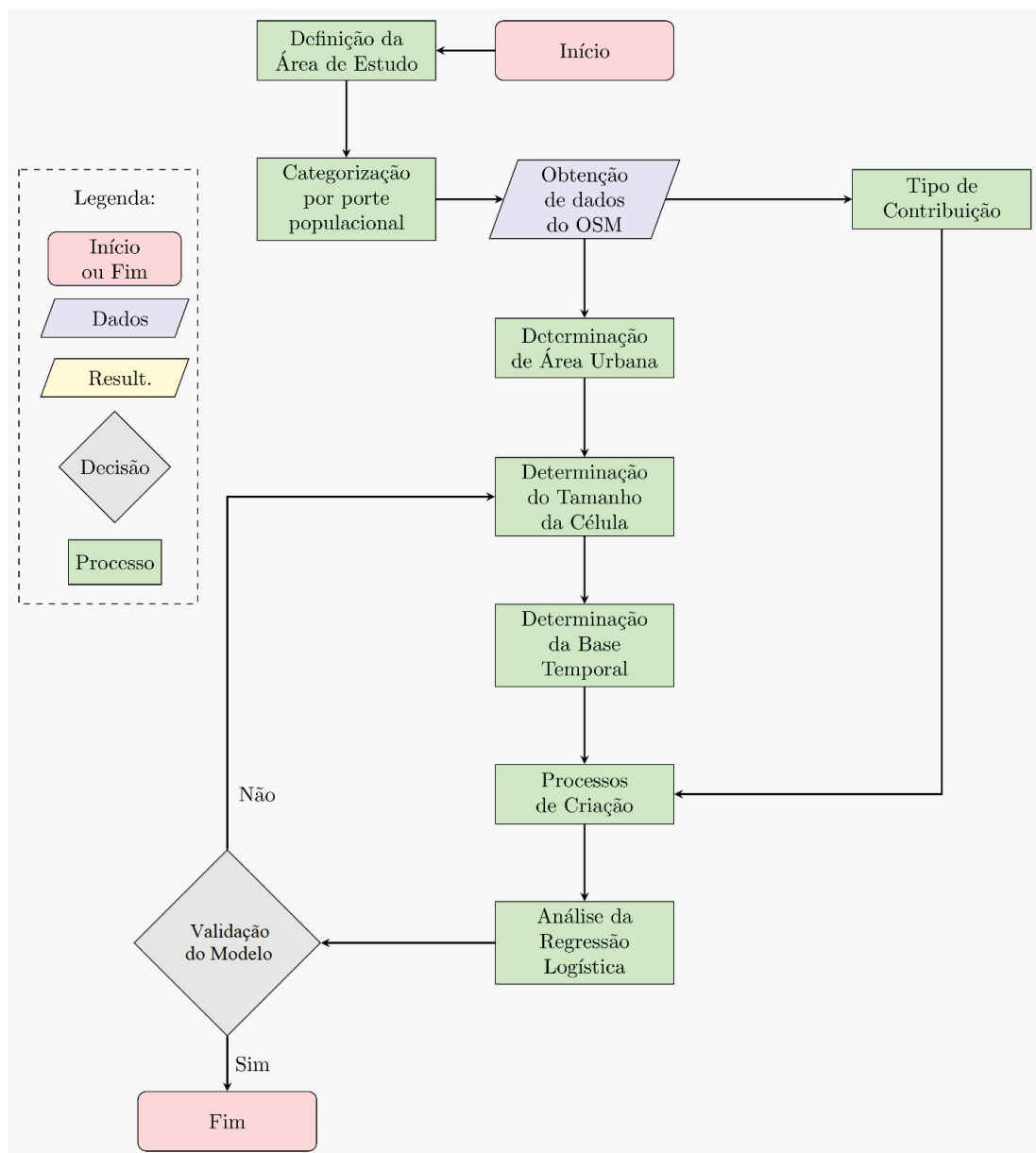
Como vimos, diversos autores têm explorado a aplicação do RMSE para garantir que os dados cartográficos digitais atendam aos requisitos de exatidão estabelecidos. Kraak (2004) destacou a importância do RMSE na avaliação da qualidade dos dados geográficos em SIG, enquanto Di Gregorio (2005) aplicou o RMSE na avaliação da precisão de mapas digitais, enfatizando a necessidade de atender aos critérios do PEC-PCD.

Outros estudos também reforçam a relevância do RMSE no contexto do PEC-PCD como Voženílek (2009) que analisou a exatidão de mapas digitais e outros produtos cartográficos digitais, empregando o RMSE como uma medida-chave para avaliar a conformidade com os requisitos de exatidão do PEC-PCD. Todos esses estudos ilustram como o RMSE é uma ferramenta essencial na garantia da qualidade e confiabilidade dos dados cartográficos digitais, permitindo melhorias contínuas e com produtos que atendam aos altos padrões de precisão exigidos na cartografia digital.

3 MATERIAL E MÉTODOS

Este capítulo apresenta a metodologia adotada para explorar as condições das ferramentas de análise de dados voluntários, buscando quantificar a confiabilidade das análises geoespaciais considerando a variação das informações contidas na base OSM em diferentes contextos de municípios brasileiros, sob os critérios da análise do tamanho da área das células, da variação temporal, classe de contribuição e dos parâmetros do modelo determinado em cada situação. O fluxograma da Figura 2 apresenta uma esquematização da execução da metodologia.

Figura 2 - Fluxograma de execução da metodologia



Este fluxograma delinea as etapas chave da pesquisa, começando com a "Definição da Área de Estudo" e avançando através de etapas cruciais. A "Categorização por Porte Populacional" destaca a seleção de quinze municípios brasileiros, representativos da diversidade urbana, baseada nos critérios populacionais do IBGE. A "Obtenção dos Dados do OSM" é detalhada, incluindo métodos de extração direta, uso do QuickOSM no QGIS, além do planejamento de utilizar a API da plataforma OH SOME. Com os dados do OSM, analisa-se o "Tipo de Contribuição" realizada, classificando-as em individuais ou em massa, para avaliar como esses valores afetam o processo de criação e diversificam a regressão logística. A "Determinação do Área Urbana" foi realizada com o objetivo de selecionar áreas com maior número de contribuições, evitando que os resultados sejam distorcidos pela inclusão de áreas sem dados. A "Determinação do Tamanho da Célula" enfatiza a importância de variar esse parâmetro na análise de qualidade de dados geoespaciais em plataformas VGI, destacando a influência de fatores como histórico de edições e padrões de contribuições. A "Determinação da Base Temporal" foi fundamentada nos períodos selecionados de 1, 5 e 10 anos para cada município. Reúne-se dados e definições, preparando para a "Análise da Regressão Logística", crucial para entender a relação entre a qualidade dos dados no OSM e padrões de contribuição ao longo do tempo, para atender aos objetivos desta pesquisa. A "Validação do Modelo" é apresentada como um pilar essencial, abrangendo análises comparativas em diferentes contextos urbanos, tamanhos de células e tipos de contribuições, contribuindo para diretrizes metodológicas robustas. O fluxograma conclui com a verificação da correspondência dos resultados; se validados, encerra o processo, se não, retorna à "Determinação do Tamanho da Célula" para ajustes necessários. Para isso, essa análise foi baseada nos trabalhos de parâmetros de modelos propostos por Brückner et al. (2021) e por Elias (2022) com as características das contribuições dos usuários ao mapeamento colaborativo, permitindo averiguar a hipótese de que o método em questão pode ser utilizado em qualquer período, tamanho de malha ou local de análise e suas melhorias.

3.1 DEFINIÇÃO DA ÁREA DE ESTUDO

Considerando o contexto de informações VGI, em especial no Brasil, foram escolhidas municípios que sintetizam a diversidade das distintas realidades urbanas brasileiras. Por isso, foram selecionadas quinze municípios, entre capitais, de médio e pequeno porte, distribuídas nas cinco regiões brasileiras. Segundo Maia (2005), o IBGE classifica os municípios em pequenos, médios e grandes, adotando como critério seu número de habitantes. Municípios pequenos são os que têm população menor que 20 mil habitantes; médios têm entre 20 mil e 500 mil e grandes, acima de 500 mil habitantes.

Dentro do escopo desta pesquisa, considerando a vastidão geográfica e a diversidade demográfica do Brasil, buscou-se selecionar municípios que representem com precisão as diferentes realidades urbanas nas cinco regiões brasileiras. Conforme já dito, segundo o IBGE, os municípios podem ser categorizados por seu porte em termos populacionais. A seleção dos municípios, abrangendo desde pequenos até grandes metrópoles (Quadro 1), foi estrategicamente definida para refletir a heterogeneidade urbana do país. Com esta abordagem pretende-se ver diferentes realidades geográficas e características de cada município e região.

Quadro 1 - Municípios selecionadas para área de estudo

Região	Porte	Município	Habitantes (H)	Área (km²)	Densidade Populacional (H/km²)
Norte	Pequeno	Taguatinga/TO	14 mil	2.441,01	5,98
	Médio	Santarém/PA	300 mil	22.887,08	18,55
	Grande	Belém/PA	1.3 milhões	1.059,46	1.230,25
Nordeste	Pequeno	Joaquim Gomes/AL	17 mil	400,02	57,52
	Médio	Juazeiro do Norte/CE	270 mil	248,56	1.105,62
	Grande	Fortaleza/CE	2.4 milhões	314,93	7.775,52
Centro- Oeste	Pequeno	Matupá/MT	20 mil	5.253,25	3,84
	Médio	Anápolis/GO	380 mil	918,38	426,29

Região	Porte	Município	Habitantes (H)	Área (km²)	Densidade Populacional (H/km²)
	Grande	Goiânia/GO	1.4milhões	739,49	1.970,90
Sudeste	Pequeno	Ponto Belo/ES	6.5mil	353,21	18,04
	Médio	Juiz de Fora/MG	560 mil	1.429,01	376,64
	Grande	São Paulo/SP	12.3milhões	1.521,11	7.528,20
Sul	Pequeno	Cacequi/RS	11 mil	2.441,01	4,70
	Médio	Chapecó/SC	220 mil	626,06	407,76
	Grande	Curitiba/PR	1.8milhões	435,04	4.078,53

Fonte: IBGE (2022), elaboração: Autora (2024)

Classificadas como grandes municípios, Belém/PA, representando a região Norte, com uma população de aproximadamente 1,3 milhões, e um IDH de Educação de 0,730, Fortaleza/CE, com cerca de 2,4 milhões de habitantes e um IDH de Educação de 0,736, representando o Nordeste, no Centro-Oeste, Goiânia/GO, com uma população próxima de 1,4 milhões e um IDH de Educação de 0,773, São Paulo/SP, com cerca de 12,3 milhões de habitantes, e um IDH de Educação de 0,810, da região Sudeste e Curitiba/PR, que abriga cerca de 1,8 milhão de pessoas, na região Sul do país e um IDH de Educação de 0,823 (IBGE, 2022).

Os municípios de porte médio selecionados foram Santarém/PA, com cerca de 300 mil habitantes e um IDH de Educação de 0,684. Juazeiro do Norte/CE, com cerca de 270 mil habitantes e um IDH de Educação de 0,719, no Nordeste. Anápolis/GO tem aproximadamente 380 mil habitantes e um IDH de Educação de 0,760 no Centro-Oeste; Juiz de Fora/MG, com uma população de 560 mil habitantes e um IDH de Educação de 0,773, no Sudeste, e Chapecó/SC, no Sul, que possui cerca de 220 mil habitantes e um IDH de Educação de 0,793 (IBGE, 2022).

Para municípios de porte pequeno, buscou-se diversidade nos estados ainda não citados. Taguatinga/TO, com uma população de cerca de 14 mil habitantes e um IDH de Educação de 0,654. Joaquim Gomes/AL, com aproximadamente 17 mil habitantes e um IDH de Educação de 0,603, Matupá/MT, com cerca de 20 mil

habitantes e um IDH de Educação de 0,642, Ponto Belo/ES, com cerca de 6,5 mil habitantes e um IDH de Educação de 0,591 e Cacequi/RS, com aproximadamente 11 mil habitantes e um IDH de Educação de 0,640 (IBGE, 2022).

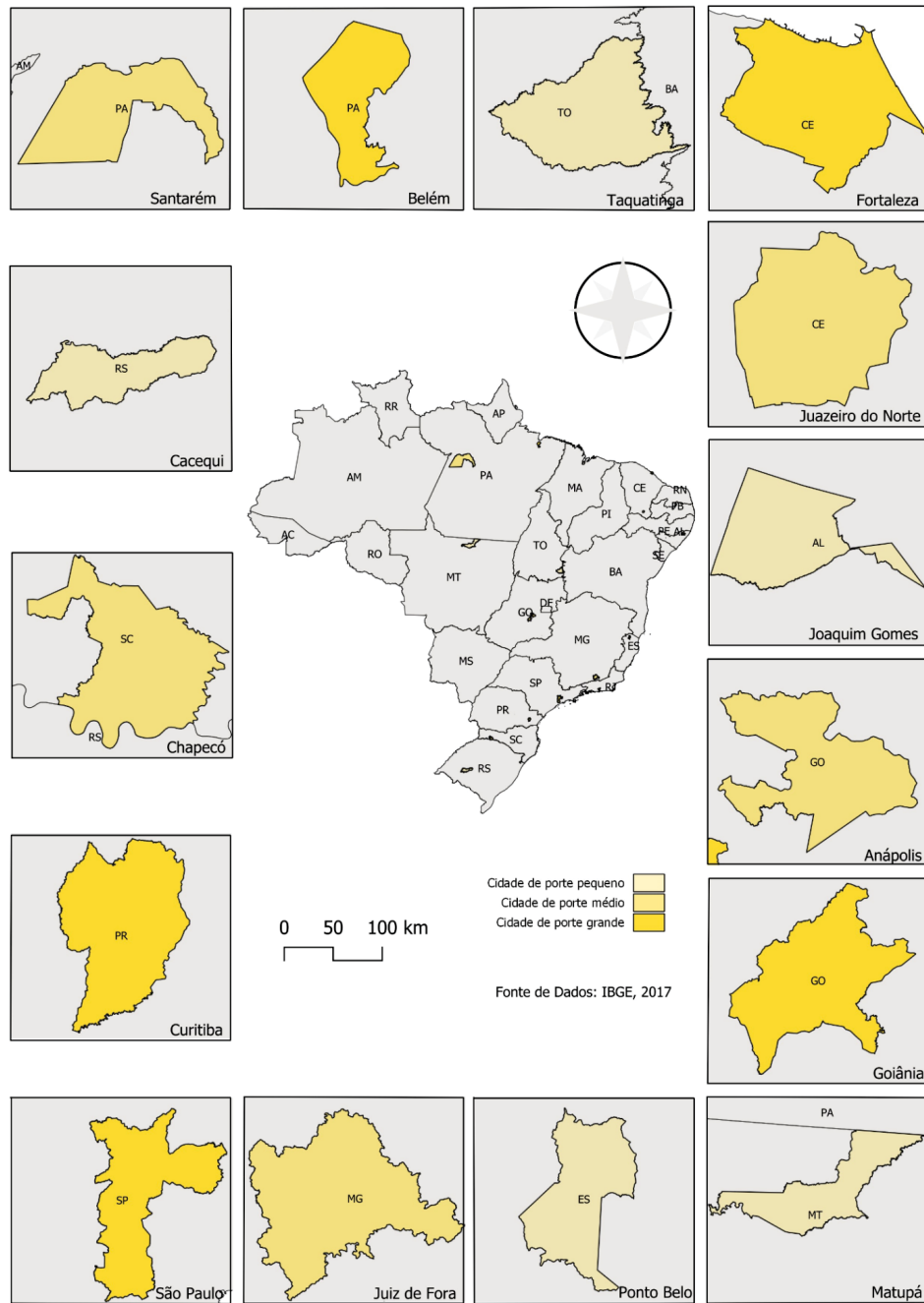
Segundo o IBGE (2022), os municípios mencionados apresentam uma diversidade econômica em suas respectivas regiões do Brasil. Belém, situada na Região Norte, é um importante centro econômico, impulsionado principalmente pela indústria, comércio e o setor de serviços. No entanto, seu IDH de Educação reflete a necessidade de maiores investimentos em infraestrutura educacional, especialmente em áreas mais periféricas, para acompanhar o desenvolvimento econômico e uma população mais qualificada. No Nordeste, Fortaleza desempenha um papel crucial na economia, com destaque para o turismo, a indústria têxtil e a pesca. O município apresenta um bom desempenho econômico, mas o IDH de Educação revela a importância de continuar investindo em formação, visto que, apesar do crescimento, há desafios no acesso à educação de qualidade para todas as camadas da população. Goiânia, localizada no Centro-Oeste, é um polo agroindustrial em crescimento, com uma forte presença na produção de alimentos e agricultura. Seu IDH de Educação mais elevado em comparação com outros municípios da região indica um progresso na área educacional, fundamental para sustentar o desenvolvimento agroindustrial com uma mão de obra mais qualificada. São Paulo, no Sudeste, é o principal centro econômico do país, contribuindo de forma significativa para o PIB nacional, com uma economia diversificada que inclui finanças, tecnologia, manufatura e serviços. O município, com um IDH de Educação elevado, destaca-se por ser um polo educacional e intelectual, o que alimenta sua economia robusta e tecnológica. Curitiba, na Região Sul, é reconhecida por sua economia diversificada, abrangendo setores como o automobilístico, tecnologia e educação. Com um IDH de Educação também elevado, o município se consolida como um centro de inovação e qualificação, favorecendo o desenvolvimento de sua economia tecnológica (IBGE, 2022).

Nos municípios de porte médio, como pode ser visto na Figura 3, Santarém é um ponto central na economia da Região Norte, destacando-se pela agricultura, pesca e comércio. No entanto, o IDH de Educação do município revela desafios em garantir acesso à educação de qualidade em áreas mais isoladas, o que impacta diretamente na capacidade de desenvolver economicamente. Juazeiro do Norte é

importante no Nordeste, com ênfase no comércio, turismo religioso e indústria. Contudo, o IDH de Educação aponta para uma necessidade de melhorias educacionais, especialmente para sustentar o turismo e o comércio local com uma mão de obra mais preparada. Anápolis, no Centro-Oeste, é um polo industrial em crescimento, com foco em manufatura e logística. O município se beneficia de um IDH de Educação que reflete sua infraestrutura educacional relativamente boa, favorecendo o desenvolvimento industrial com trabalhadores mais qualificados. Juiz de Fora, no Sudeste, possui uma economia diversificada, abrangendo manufatura, tecnologia e serviços. O IDH de Educação do município sugere uma população bem-educada, o que contribui para o sucesso de sua diversificação econômica. Chapecó, na Região Sul, destaca-se pela agroindústria e o setor agropecuário. Apesar de seu desenvolvimento econômico, o IDH de Educação do município indica que investimentos em educação podem ser essenciais para manter sua competitividade no setor agrícola e industrial (IBGE, 2022).

Nos municípios de pequeno porte, representados na Figura 3, Taguatinga é um município agrícola em Tocantins, contribuindo para a produção de alimentos na região. Contudo, o IDH de Educação do município revela desafios em garantir acesso à educação de qualidade, o que pode limitar seu desenvolvimento a longo prazo. Joaquim Gomes, em Alagoas, tem uma economia centrada na agricultura e pecuária, mas o IDH de Educação evidencia que ainda há muito a ser feito para melhorar o acesso à educação, fator fundamental para o progresso econômico local. Matupá, em Mato Grosso, é um município agroindustrial, com foco na produção de grãos e agricultura familiar. O IDH de Educação relativamente baixo mostra que a qualificação da mão de obra ainda é um obstáculo para o crescimento sustentável do município. Ponto Belo, no Espírito Santo, destaca-se pela pesca e agricultura, mas seu IDH de Educação indica que investimentos educacionais são necessários para diversificar ainda mais a economia e promover o desenvolvimento local. Cacequi, no Rio Grande do Sul, tem uma economia diversificada, incluindo agricultura, pecuária e comércio local. No entanto, o IDH de Educação relativamente baixo sugere que o desenvolvimento educacional deve ser uma prioridade para que o município continue a crescer de forma equilibrada e sustentável (IBGE, 2022).

Figura 3 - Figura do mapa de localização dos municípios da área de estudo



Fonte: Autora (2024)

A tentativa de generalizar a classificação de municípios esbarra na diversidade de critérios adotados por países, ausência de uma regra universal que contemple as particularidades de cada nação. Diante desse panorama, surge a necessidade de abrir espaço para a discussão sobre a estratificação proposta. As três classes tradicionais — pequeno, médio e grande porte — adotada no Brasil pelo IBGE, podem não ser suficientes para abranger a complexidade de contribuições quando se considera as complexidades urbanas presentes na base do OSM.

Cada país, permeado por sua própria identidade demográfica e estrutura urbana, estabelece parâmetros específicos para categorizar o porte de suas cidades. Segundo Costa (2002) e Dos Ramos (2016), na Alemanha, por exemplo, a classificação minuciosa em categorias como "Landstadt" (espaço rural: até 5.000 hab.), "Kleinstadt" (cidade pequena: entre 5.000–19.999 hab.), "Mittelstadt" (cidade média: entre 20.000–99.999 hab.) e "Großstadt" (cidade grande: mais de 100.000 hab.) proporciona uma visão detalhada, levando em consideração o número de habitantes. Já na Espanha, segundo o Instituto Nacional de Estatística (INE, 2023) orienta a classificação dos municípios em categorias como "Município" (municípios pequenos, com uma população menor), "Ciudad" (cidades maiores, muitas vezes com uma população acima de 10.000 habitantes) e "Gran Ciudad" (grandes cidades, frequentemente com população superior a 100.000 habitantes). Na França, segundo o Instituto Nacional de Estatística e Estudos Econômicos da França (INSEE, 2023) categorizam-se em "Commune" (pequenas cidades ou vilarejos, que pode ter uma população relativamente pequena, variando geralmente de algumas centenas a alguns milhares de habitantes), "Ville" (cidades de médio porte, com populações que variam de dezenas de milhares a algumas centenas de milhares de habitantes e "Ville Capitale" ou "Métropole" (engloba as grandes cidades, muitas vezes capitais de regiões ou a capital nacional) refletindo a diversidade populacional.

Ao alinhar essas considerações com a abordagem adotada na análise comparativa dos resultados, torna-se evidente que a diversidade intrínseca nas dimensões urbanas é um fator crucial a ser considerado. A proposta de categorização flexível por habitantes, em contraste com as classificações tradicionais, sugere uma abordagem mais dinâmica, adaptada às realidades demográficas diversas.

Explorando alternativas, a proposta de delimitação por faixas específicas de habitantes surge como uma possibilidade. Considerando a diversidade de realidades demográficas e potenciais contribuidores, sugere-se que categorias como 20 mil a 250 mil habitantes, 250 mil a 500 mil habitantes, 500 mil a 750 mil habitantes e acima de 750 mil habitantes poderia oferecer uma abordagem mais refinada e ajustada às diferentes escalas urbanas. Essa subdivisão auxiliou refletindo as variações nas dimensões urbanas, adaptando-se tanto a realidades densamente povoadas quanto a regiões menos habitadas.

3.2 OBTENÇÃO DOS DADOS PARA DESENVOLVIMENTO DAS ANÁLISES

Para a realização deste trabalho, foram utilizados dados provenientes do *OpenStreetMap*. A aquisição das últimas versões das feições editadas dos dados pode ser realizada de algumas formas, dentre elas, destaca-se a extração direta na plataforma do OSM, permitindo a seleção de uma área ou retângulo envolvente para a extração dos arquivos vetoriais. Outra alternativa possível consiste na utilização do complemento do QGIS denominado QuickOSM, que possibilita a extração de dados de maneira semelhante à plataforma do OSM, através da escolha de uma área específica, que foi a forma utilizada. Por fim, é possível ainda o download dos dados pela plataforma Geofabrik que segue um formato diferente, uma vez que as regiões para download já estão predefinidas, abrangendo áreas mais extensas, como continentes e países. Essas abordagens oferecem flexibilidade na obtenção de dados do OSM, podendo ser ajustadas às necessidades do usuário e o tamanho da célula desejado.

Para esta pesquisa, utilizou-se ainda a API da plataforma OHSOME, desenvolvida pela Universidade de Heidelberg, na Alemanha. Essa ferramenta oferece recursos avançados para a extração de dados históricos das feições do (OSM), além de ser usada para obter informações sobre o número de funcionalidades adicionadas ao longo do tempo, o número de contribuintes em uma área específica ou conjunto de funcionalidades, bem como a obtenção de categorias específicas. A documentação da API enfatiza a necessidade de realizar consultas específicas, delimitando as coordenadas geográficas da região de interesse, para a extração de dados, que neste caso foram os municípios de pequeno, médio e grande porte, o tamanho mínimo do retângulo circundante e o intervalo de tempo pretendido para as análises, proporcionando uma visão temporária precisa do ambiente geoespacial em estudo.

Um diferencial significativo da API da plataforma OHSOME é a capacidade de fornecer informações sobre a quantidade de usuários que contribuíram em um determinado local ao longo do tempo. Esse aspecto adiciona uma camada de compreensão social ao contexto geoespacial, permitindo uma análise mais abrangente das dinâmicas colaborativas na construção e atualização do OSM. A

integração dessa API amplia substancialmente as possibilidades de investigação, oferecendo uma visão mais completa e contextualizada das mudanças e contribuições no cenário mapeado, permitindo analisar como as contribuições dos usuários do OSM evoluíram nesse cenário ao longo do tempo, por exemplo, adicionando uma dimensão social à análise geoespacial e possibilitando uma compreensão das dinâmicas colaborativas.

3.2.1 Software Utilizado

O software utilizado como ferramenta para relacionar o banco de dados à qualidade dos dados, foi o QGIS. Segundo a OSGeo (2015), o QGIS é um Sistema de Informações Geográficas de código aberto, licenciado sob a *General Public License* (GNU). Desenvolvido em linguagens C++ e Python, é compatível com diversos sistemas operacionais. O software está acessível através do site oficial do QGIS, disponibilizando uma plataforma gratuita para os usuários, além disso permite a elaboração de extensões que implementam rotinas não incluídas nas versões básicas do software, atendendo assim às necessidades específicas dos usuários.

Dada sua natureza como um aplicativo baseado em uma biblioteca de código aberto, os usuários têm a oportunidade de contribuir para o desenvolvimento do programa, escrevendo novas rotinas destinadas a diversas aplicações relacionadas.

3.2.2 Desenvolvimento de aplicações

Com base nas implementações realizadas por Elias et al. (2023) e Elias (2022), foi elaborado um código que permite a obtenção de informações específicas sobre a quantidade de usuários ativos e suas contribuições no OSM dentro de intervalos de tempo e áreas geográficas definidas. Utilizou-se o Python como linguagem de programação, aproveitando a API da plataforma OHSOME para acessar e extrair os dados necessários. O código começa com a definição de parâmetros de entrada, como as datas de início e término do período analisado e as coordenadas do *bounding box* que delimita o município de interesse. Com esses parâmetros, o código faz uma requisição à API, que retorna dados sobre as contribuições dos usuários no OSM para a área e período especificados. Essas

informações são cruciais para entender a dinâmica de atualizações do mapa e a atividade da comunidade de mapeadores.

Além de realizar as requisições à API, o código também trata e verifica os dados recebidos, garantindo que erros de comunicação ou dados incompletos sejam identificados e gerenciados apropriadamente. Isso é essencial para manter a integridade dos resultados e a confiabilidade das análises subsequentes.

3.3 TAMANHO DAS CÉLULAS

A determinação do tamanho da célula para análise de qualidade de dados geoespaciais em plataformas VGI desempenha um papel importante justificado pela discrepância e heterogeneidade existente entre os espaços urbanos, sendo necessário variar sua definição na tentativa de se identificar um valor que seja apropriado e representativo das características das localidades. Pesquisadores como Martini, Kuper e Breunig (2019), Brovelli e Zamboni (2018), Zhang e Malczewski (2018), Brovelli, Minghini e Moliari (2016), Zhou et al. (2022), têm realizado a avaliação da qualidade dos dados nas plataformas VGI com base no princípio de subdividir a área de estudo em células utilizando processos de avaliação de qualidade extrínseca com base em indicadores ISO:19157 ou parâmetros intrínsecos associados às características das contribuições e contribuidores. A qualidade dos dados em plataformas VGI, como o *OpenStreetMap*, é influenciada por diversos fatores, incluindo histórico de edições, perfil dos contribuintes e padrões de contribuições e relacioná-los com os métodos de qualidade previstos pela ISO 19.157 (ISO, 2013).

Camboim et al. (2022) desenvolveram uma abordagem interativa para determinar o cálculo do tamanho das grades, utilizando inicialmente pontos de acurácia posicional conhecida. A hipótese é que quando a subárea analisada é de tamanho ideal, é possível interpolar o erro dentro da célula por meio de uma interpolação ponderada pelo inverso da distância ponderada (IDW) e gerar resíduos mínimos nos pontos de controle. Além disso, ao subdividir consecutivamente a grade, a curva de erro médio quadrático em relação ao tamanho da célula se

aproxima da estabilidade, revelando que há um valor no qual a maioria dos municípios analisados convergem em relação às discrepâncias geométricas.

Camboim et al. (2022) realizou ainda uma análise preliminar em um município brasileiro, Curitiba, com cerca de 28 mil pontos de acurácia conhecida, foram separados 2.800 pontos de controle e o município foi dividido em células de 250 m a 8 km. A partir deste estudo preliminar, foi observado que, à medida que o tamanho da célula diminui, os resultados tendem a ser mais constantes, o que corrobora a hipótese dessa relação com a qualidade dos dados.

No contexto da análise de padrões temporais de contribuições para o *OpenStreetMap* em municípios brasileiros, Peterlini et al. (2023) abordaram a relevância do tamanho da célula ou grade na identificação desses padrões. O estudo focou em cinco capitais brasileiras: Curitiba, São Paulo, Goiânia, Fortaleza e Belém, analisando as contribuições de 2008 a 2022, utilizando um script Python, considerando as tags "*highway*" e "*buildings*", e empregando células concêntricas variando de 1 a 8 km. Os resultados indicaram um comportamento consistente das curvas em cada município com o aumento do tamanho da célula, sendo possível observar que células de cerca de 3km apresentariam um desempenho consistente para comparar os padrões dos municípios. Além disso, observou-se que municípios como Goiânia e Belém, que não possuem uma Infraestrutura de Dados Espaciais (IDE), apresentaram uma curva de crescimento gradual, enquanto municípios com IDEs próprias, que buscam facilitar a disseminação de informações geoespaciais, com conjunto integrado de tecnologias, mecanismos e procedimentos de coordenação e monitoramento, necessários para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento e o uso dos dados geoespaciais, como em Fortaleza e São Paulo, mostraram padrões distintos de contribuições ao longo do tempo. Este estudo fornece uma base sólida para a escolha do tamanho da célula na análise de padrões temporais de contribuições.

Além da análise das contribuições, o estudo realizado por Peterlini et al. (2023) também se aprofundou na qualidade do ajuste da curva, representada pelo RMSE de cada tamanho de célula. Observou-se que o aumento no tamanho da célula tende a apresentar um melhor ajuste da curva, estabilizando ou até mesmo melhorando após um determinado ponto. Não apenas destacou a importância do

tamanho da célula na modelagem da curva logística, mas também ajudou com análises sobre o comportamento das contribuições em diferentes municípios, com base em sua infraestrutura de dados espaciais e histórico de contribuições.

A análise do tamanho da célula, conforme explorado em Elias (2022), Elias et al. (2023) e Peterlini et al. (2023), revelou que a medida utilizada poderia influenciar na interpretação dos padrões temporais de contribuições para o *OpenStreetMap*. No entanto, a profundidade e amplitude dessa influência ainda não foram totalmente esgotadas. O tamanho da célula pode afetar diretamente a granularidade e, conseqüentemente, a precisão dos dados analisados. Células muito grandes podem mascarar variações locais, enquanto células muito pequenas podem introduzir ruído ou variabilidade excessiva nos resultados. Com base nas análises realizadas, é evidente que a escolha do tamanho da célula desempenha um papel crucial na qualidade e na interpretação dos dados geoespaciais, particularmente em ambientes colaborativos como o OSM.

Dada a importância de obter detalhes a partir dos dados, é crucial entender e otimizar essa variável. Ao seguir as diretrizes estabelecidas pelos estudos de Camboim et al. (2022), Elias (2022) e Peterlini et al. (2023), que abordaram a importância do dimensionamento da célula na análise de padrões temporais de contribuições, foi possível delinear um método confiável para determinar o tamanho ideal da célula. Continuar a investigar a influência do tamanho da célula permitiu uma compreensão mais aprofundada de como podem afetar os resultados, agregando robustez à metodologia adotada nesta pesquisa de dissertação. Além disso, ao entender melhor essa dinâmica, pode-se estabelecer diretrizes mais claras para futuras pesquisas e aplicações práticas, sendo necessário entender qual tamanho da célula é suficientemente representativo do padrão de contribuições dos municípios.

3.4 ERRO MÉDIO QUADRÁTICO

Nesta pesquisa, o RMSE foi analisado em dados VGI ao longo de períodos de 1, 5 e 10 anos para cada uma dos 15 municípios de pequeno, médio e grande porte,

divididas nas 5 regiões brasileiras, entre a observação e a curva resultante. Essa análise abrangente compreende a variação da qualidade dos dados ao longo do tempo, aprimorando a análise da qualidade de dados geoespaciais colaborativos. A análise em diferentes períodos de tempo e em municípios de diferentes portes é essencial para capturar a dinâmica da evolução dos dados VGI, identificando possíveis tendências, padrões e desafios específicos em diferentes contextos urbanos e temporais.

Além da análise das contribuições, é essencial considerar a qualidade do ajuste da curva, representada pelo RMSE de cada tamanho de célula (Peterlini et al., 2023). O RMSE, como uma medida de dispersão dos erros entre os valores previstos e observados, desempenha importante papel na avaliação da precisão dos modelos de curva logística utilizados para representar as contribuições em diferentes municípios, analisando o comportamento da curva, incluindo uma avaliação das tendências observadas.

Para chegar aos valores de análise apresentados, foi necessário um processo prévio de manipulação dos dados. Inicialmente, os dados foram processados no algoritmo desenvolvido por Elias et al. (2023) e devidamente alteradas as informações de interesse, período de análise e tamanho da célula, pela autora deste trabalho. Esse programa processa os dados e fornece um arquivo que contém o número de células e o valor de RMSE de cada uma das células analisadas. No entanto, é importante destacar que muitas células possuíam valores nulos, enquanto outras não continham informações. Para lidar com essa heterogeneidade nos dados, foi calculada a média entre os valores presentes nas células que continham informações, a fim de obter valores mais precisos. Após realizar esses cálculos e análises, os resultados foram apresentados nas tabelas fornecidas no capítulo dos resultados e discussões. Vale ressaltar que esse mesmo processo foi realizado para todos os municípios e para os três períodos de tempo em questão (1, 5 e 10 anos).

3.5 PARÂMETROS DA CURVA

A regressão logística é uma técnica que, no contexto do mapeamento colaborativo, pode ser usada como uma ferramenta para, por exemplo, entender como a qualidade dos dados geoespaciais, no OSM está relacionada aos padrões

de contribuição ao longo do tempo e entender os fatores que a afetam, temática desta pesquisa. A relação entre a qualidade dos dados e essas contribuições pode ser modelada pela função logística, com a forma de uma curva em "S", como pode ser visto na Figura 4, este modelo é descrito a partir de uma trajetória semelhante a uma função sigmoidal e pode estar associada ao padrão de contribuições em uma determinada área, que começa com poucas contribuições, apresenta crescimento e tende a se estabilizar ao longo do tempo (Elias et al., 2023).

É preciso determinar os parâmetros da regressão logística para determinar o comportamento destes parâmetros considerando as variações devidas ao tamanho das células e classes das contribuições (individuais ou massa). Esses coeficientes descrevem como as variáveis independentes influenciam a variável dependente. A modelagem matemática envolve o ajuste desta curva logística aos dados, permitindo que sejam feitas análises de como a qualidade dos dados varia com base no comportamento das contribuições. Por exemplo, valores baixos no coeficiente θ , associado também ao valor de D , pode indicar áreas com crescimento abrupto no mapeamento, o que pode afetar a qualidade dos dados geoespaciais. A partir do histórico de contribuições acumulado ao longo do tempo, a modelagem dos dados foi realizada por meio de uma regressão logística de quatro parâmetros, conforme mostra a Equação 1.

$$(1) \quad y = A + ((B - A) \div (1 + e^{(C-x)/D}))$$

Onde:

A: assíntota superior

B: assíntota inferior

C: ponto médio da curva logística (no eixo x)

D: inclinação da curva logística curva

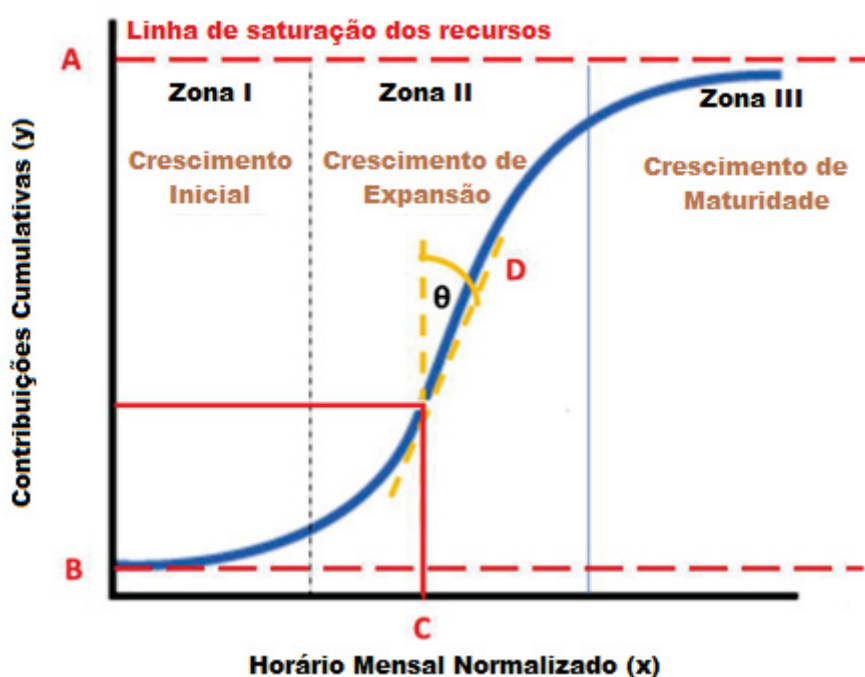
x: valor mensal normalizado

y: número acumulado de funcionalidades inseridas no OSM

Conforme mostrado na Figura 4 e Equação 1, os parâmetros que descrevem a curvatura são as assíntotas inferior e superior, a inclinação e ponto médio da curvatura. A assíntota inferior representa o valor mínimo de feições presentes na célula. Da mesma forma, a assíntota superior representa o máximo de contribuições

na célula. Os demais parâmetros derivam da relação das assíntotas com a variação temporal representada ao longo do eixo X. A consideração do parâmetro D está associado à inclinação da curva, valores reduzidos na regressão podem sinalizar características específicas do padrão de crescimento acumulado. O ângulo teta é oriundo da relação dos parâmetros C e D e representa a taxa de crescimento de contribuições acumuladas na célula ao longo do tempo. Por exemplo, em regressões que mostram mudanças bruscas, o parâmetro C tende a se assemelhar ao mês (valor x) em que ocorreu a maior incorporação de elementos, podendo assim observar o comportamento das curvas nas células de análise, nas diversos municípios, incorporando avaliação das tendências que serão observadas.

Figura 4 - Parâmetros e zonas da curva logística



Fonte: Adaptado Elias et al. (2023)

Conforme descrito por Elias et al. (2023), a Figura 4 nos mostra a relação entre o comportamento da curva e os diferentes estágios de contribuições, subdivididos em:

- Zona I - Início (etapa inicial das atividades de mapeamento);
- Zona II - Crescimento (representa o crescimento do número de contribuições) e;

- Zona III - Saturação ou Maturidade (estágio de saturação em que uma determinada área atinge o limite máximo de feições que podem ser inseridas).

A primeira área representa aquelas células com os estágios iniciais de contribuições com valor mínimo definido pela assíntota inferior (B). Nessa fase, as células recebem tipicamente um pequeno número de contribuições ou existem pouca atividade de mapeamento voluntário.

Na segunda etapa, o aumento da inclinação da curva representa que a acumulação das contribuições se torna mais significativa. Isso é representado com o crescimento da curva em função da sua inclinação representada pelo C, no ponto de flexão, gerando θ . Células com θ próximo a 0° , ou seja, com baixa inclinação da curva em relação à linha vertical passante pelo ponto médio (C), representam um crescimento vegetativo frente às contribuições individuais ou muito espaçadas. Da mesma forma, uma curva íngreme, ou seja, com forte angulação, representa contribuições em massa. É crucial destacar que o gráfico utilizado para representar essas zonas não está estritamente referenciado a uma localização geográfica específica, mas sim à trajetória temporal das contribuições. Em determinados estágios, a curva logística associada a uma célula pode apresentar um ângulo θ próximo a 90 graus ou mesmo 0 graus, dependendo do referencial escolhido. Essas contribuições são aquelas realizadas por empresas ou pessoas que realizam a contribuição de um grande número de feições temporalmente localizadas em horas ou dias, como no caso de mapatonas.

A figura 6 ainda apresenta a assíntota superior (A) como limite da curva. Contudo, na fase de maturação ressalta-se que a área da célula contém informações que são passíveis de um volume maior ou menor de contribuições na medida que a região representada tem feições que sejam atualizáveis ao longo do tempo. Por exemplo, células que contenham áreas de vegetação ou corpos hídricos terão uma assíntota mais baixa em relação ao número máximo de feições existentes frente aquelas células que sejam mais urbanizadas, com destaque a feições lineares de vias e feições pontuais.

Portanto, além da análise das fases de contribuição, explora-se a relação entre as células ao longo do tempo, considerando a variação no número de feições. A dinâmica temporal pode fornecer informações sobre o estado de maturidade das

áreas em estudo. Se, em dois momentos distintos, o número de feições em uma célula apresenta diferenças significativas, isso sugere um processo de crescimento em andamento. Por outro lado, a estabilidade no número de feições indica uma maturação naquela região. Essa abordagem temporal oferece uma perspectiva de entender as mudanças no ambiente geoespacial, destacando áreas em transição e aquelas que atingiram um estado de maturidade.

Ao dividir o processo de análise em duas etapas distintas, a primeira centrada na viabilidade de alteração da célula e a segunda na avaliação do potencial de mudança, pode-se identificar quais parâmetros da regressão logística são mais suscetíveis a variações. A ausência de alterações em áreas consideradas saturadas sugere uma estabilidade que pode ser quantificada, enquanto a presença de mudanças permite avaliar a porcentagem da célula que está sujeita a alterações.

Explorar a maturidade das células em análise não se restringe apenas à contagem absoluta de feições, mas também à interpretação do ângulo θ em momentos distintos, considerando a presença ou ausência de contribuições em massa. Esta análise oferece um indicativo valioso da dinâmica temporal das células, servindo como um potencial marcador da maturidade sem depender de dados externos. Ao considerar a relação intrínseca entre células geoespaciais em momentos distintos, é possível discernir nuances significativas no desenvolvimento do mapeamento colaborativo representada pela alteração do ângulo θ . A disparidade no número de feições presentes em uma célula pode sugerir um estado de crescimento das feições mapeadas. Em contraste, a estabilidade no valor de θ pode denotar que a maturidade da célula pode ter sido já alcançada independente do valor da célula. Esta análise temporal proporciona uma janela singular para a compreensão da dinâmica evolutiva das áreas em estudo. O enfoque na análise de mudanças de área, portanto, direciona-se principalmente às regiões em crescimento ou transição. A determinação do potencial de alteração em uma célula ao longo do tempo é crucial para compreender a evolução do mapeamento colaborativo.

A complexidade da análise é ampliada ao considerar que as zonas I, II e III não estão rigidamente vinculadas a um único local geográfico. Diferentemente de modelos que pressupõem uma distribuição espacial uniforme, a natureza dinâmica do mapeamento colaborativo permite que essas zonas se manifestem de maneira

não uniforme no espaço. Em outras palavras, uma área geográfica pode transitar por diferentes estágios de contribuição, representados pelas zonas I, II e III.

Partindo dos resultados apresentados por Elias et al. (2023), de que se a utilização da regressão logística na análise de dados VGI não apenas permite modelar o comportamento das contribuições ao longo do tempo, mas também oferece uma estrutura para compreender a dinâmica de maturidade das células, que existem relações dos padrões espaço temporais de contribuições no OSM com a qualidade dos dados geoespaciais, foi possível realizar a sua modelagem e extrair os seus parâmetros para análise em diferentes municípios brasileiros, possibilitando a validação do modelo utilizado.

As considerações sobre a possibilidade de alterações em uma célula de análise, o processo analítico desdobra-se em fases distintas. Uma avaliação para determinar a aptidão da célula para modificações, estabelecendo critérios para o seu potencial de transformação ao longo de um período.

A análise explora a classe de feições presentes na célula, delimitando, assim, o espectro de possibilidades para a área em estudo. Este aprofundamento conduz a uma interpretação dos parâmetros na equação logística, elucidando o papel crucial que desempenham na representação do crescimento e da maturidade das contribuições. Nesse contexto, a identificação de quais parâmetros de regressão possuem flexibilidade para alterações, esclarecendo a dinâmica local do mapeamento colaborativo ao constatar a ausência de alterações em áreas saturadas e, por contraste, a presença de modificações, surgindo uma oportunidade única de examinar a porcentagem da célula suscetível a transformações.

Com a subdivisão dos municípios avaliados em células homogêneas, extração e espacialização dos parâmetros da Regressão Logística, a estimativa dos comportamentos atípicos nos padrões de contribuição no OSM, a obtenção e espacialização dos parâmetros da regressão logística, bem como, da completude dos dados, foi possível realizar análises estatísticas classificando os padrões de contribuição e obter características espaciais da distribuição de tais padrões e suas relações com outras variáveis de descrição do território e do mapeamento colaborativo.

A combinação de análise temporal e espacial proporciona uma visão abrangente, permitindo a identificação de áreas em crescimento, estáveis ou maduras, e a avaliação do potencial de alteração em função das características específicas das células e das contribuições associadas. Com a necessidade de explorar métricas de qualidade com base em parâmetros intrínsecos, foi realizada análise de métricas de qualidade na regressão, de forma semelhante à metodologia apresentada por Grinberger et al. (2021) calculando o erro quadrático. Essas métricas ajudaram a avaliar a qualidade dos dados geoespaciais de uma maneira quantitativa.

A partir das comparações, esta etapa corresponde à validação do modelo utilizado, no intuito de verificar se as mesmas poderão ser utilizadas em qualquer contexto dos dados VGI. Esta etapa buscou a confiabilidade e generalização dos resultados obtidos. A validação verificou a competência e a aplicabilidade das ferramentas em diferentes contextos urbanos, considerando variáveis como tamanho da célula, período e bases de referência.

A validação do modelo por meio de análises comparativas em diferentes localidades, verificando a concordância e a estabilidade dos resultados independentemente das dimensões geográficas e características específicas de cada município, bem como a análise de diferentes tamanhos de células, períodos temporais e bases de referência serão considerados, permitiu uma avaliação abrangente da capacidade das ferramentas em lidar com a heterogeneidade inerente aos dados do *OpenStreetMap*.

Ao considerar a diversidade de municípios, tamanhos de células, tipo de contribuições (individuais e em massa) e bases temporais, a validação do modelo torna-se uma etapa crucial para orientar a tomada de decisões informadas quanto à escolha e aplicação das ferramentas em distintos contextos. A transparência e consistência alcançadas nesse processo de validação contribuíram para o desenvolvimento de diretrizes metodológicas, permitindo a utilização das mesmas.

4 RESULTADOS E DISCUSSÕES

Neste capítulo, detalha-se os resultados alcançados, seguida de uma discussão que visa elucidar suas implicações no contexto dos objetivos propostos por esta pesquisa. A metodologia empregada permitiu a execução de análises, cujos resultados não somente corroboram com as hipóteses levantadas, mas também oferecem novas perspectivas. O processo analítico, fundamentado em abordagens quantitativas e qualitativas, possibilitou a obtenção de dados críticos para a compreensão das questões em investigação.

4.1 DEFINIÇÃO DO TAMANHO DA CÉLULA

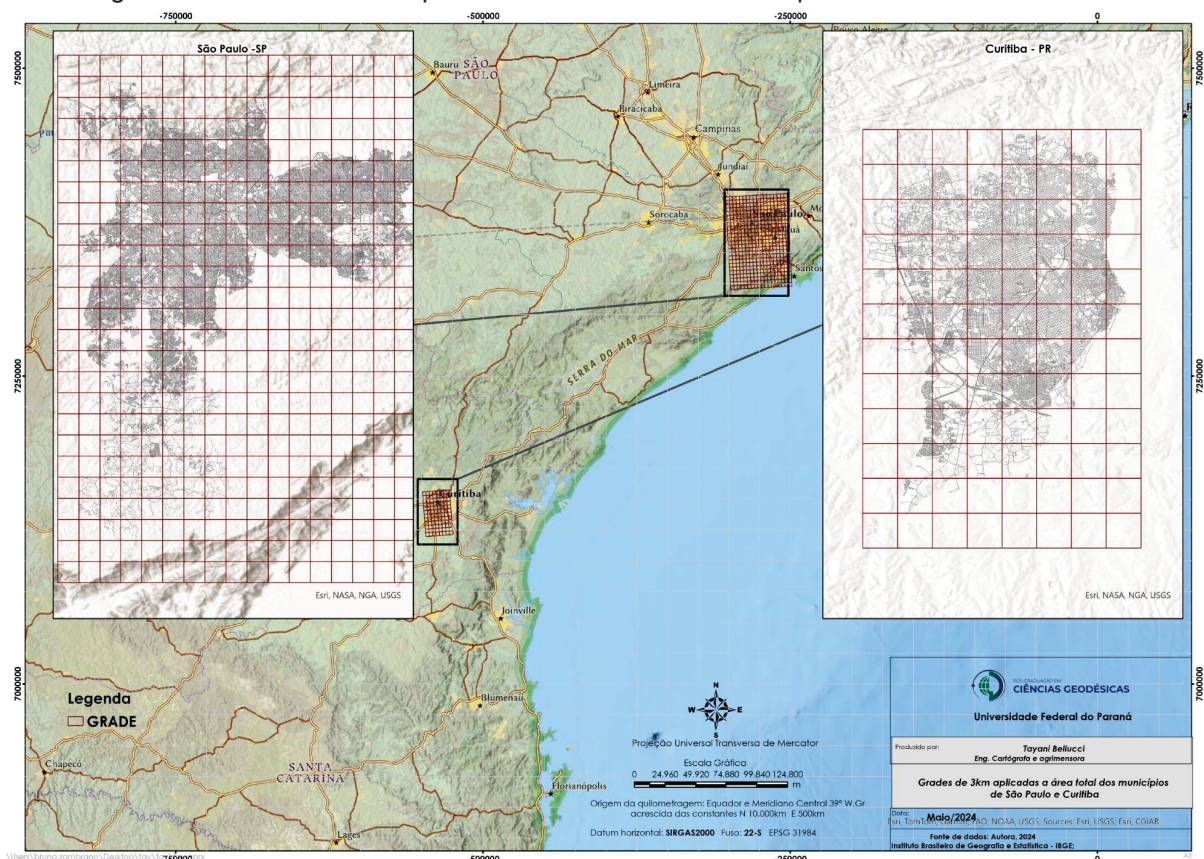
A constatação de que a grade de 3 km foi suficiente para as análises propostas neste trabalho reflete não apenas a consistência dos resultados obtidos em vários municípios, mas também a consonância com a literatura preexistente. Conforme destacado por Camboim et al. (2022), a redução do tamanho da célula tende a gerar resultados mais estáveis, enquanto Peterlini et al. (2023) ressaltaram a significância do ajuste da curva em relação ao tamanho da célula, evidenciando melhorias substanciais à medida que o tamanho aumenta.

Essa escolha de adotar uma célula de 3 km como padrão para as análises subsequentes não foi tomada de forma arbitrária, mas sim fundamentada em uma compreensão aprofundada das implicações da variação do tamanho da célula. Como discutido por Elias (2022) e Peterlini et al. (2023), a escolha do tamanho da célula pode comprometer a granularidade dos dados se for muito grande ou introduzir ruído excessivo se for muito pequeno. Portanto, a seleção de um tamanho específico da célula não apenas aprimora a precisão dos resultados, mas também auxilia na análise de relevância e a robustez das conclusões extraídas.

Ao estabelecer a grade de 3 km como padrão nesta pesquisa, buscou-se não apenas a consistência metodológica, mas também promover uma abordagem empiricamente embasada na análise de padrões temporais de contribuições no OSM. Essa escolha estratégica ajuda no processo analítico e permite uma investigação mais incisiva dos resultados em questão.

Inicialmente, as grades de 3 km foram aplicadas à área total do município de acordo com a delimitação geográfica do IBGE, conforme pode ser visto na Figura 5, de Curitiba e São Paulo, usadas como exemplo. Essa abordagem proporcionou uma visão geral dos padrões de contribuições no OSM em cada município, permitindo uma comparação inicial entre diferentes localidades. No entanto, para uma análise mais aprofundada e direcionada aos objetivos específicos deste trabalho, foi realizada uma segunda etapa de análise, na qual as grades de 3 km foram geradas novamente, mas dessa vez apenas na área urbana delimitada manualmente de cada município, como pode ser visto na Figura 6.

Figura 5 - Grades de 3 km aplicadas à área total dos municípios de São Paulo e Curitiba



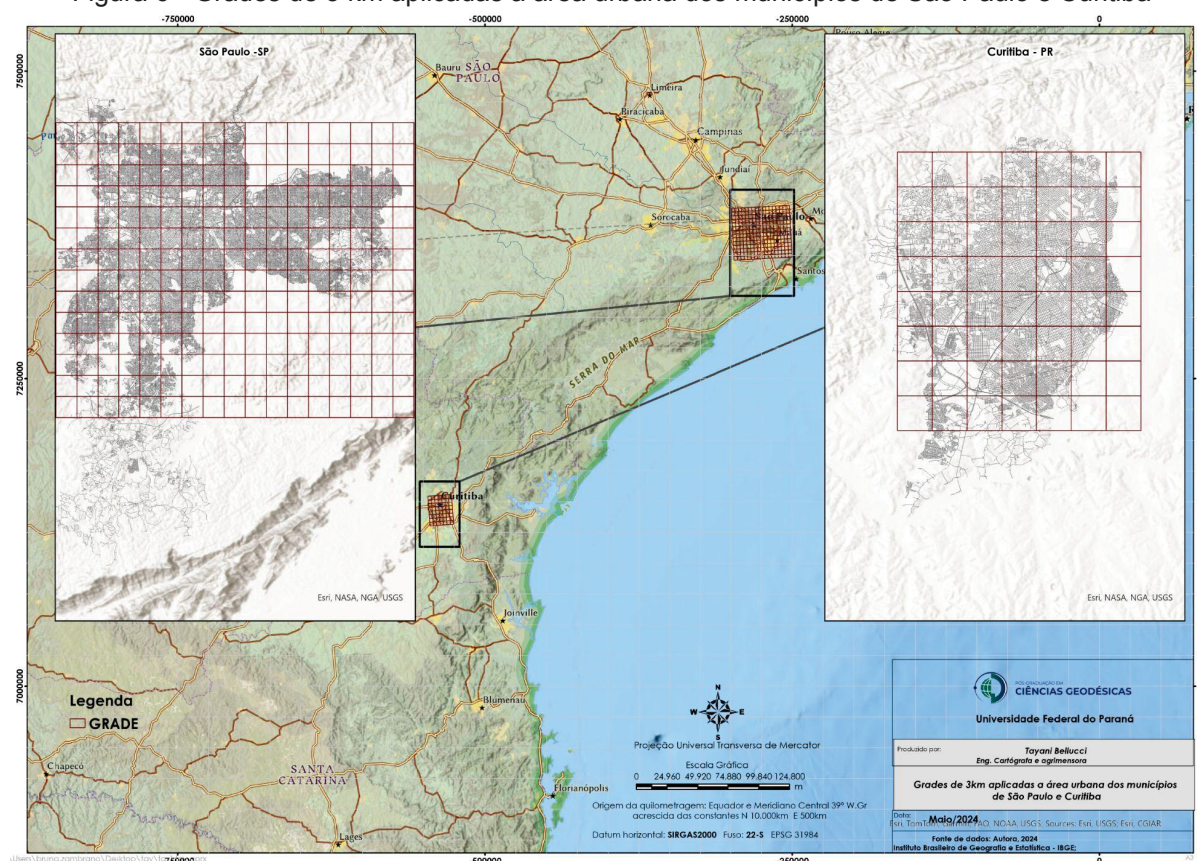
Fonte: Autor (2024)

É importante destacar que a área considerada nas análises não corresponde, estritamente, à definição oficial de área urbana conforme critérios legais ou cartográficos, mas sim a uma delimitação manual realizada para focar em regiões de interesse dentro do município. Esse recorte manual foi adotado para adequar a análise às áreas com maior densidade de dados inseridos no OSM, refletindo a

atividade dos voluntários nas zonas urbanizadas ou de maior relevância. Apesar de o termo 'área urbana' ser utilizado ao longo do estudo, ele deve ser entendido como uma delimitação operacional realizada de maneira personalizada para capturar os aspectos mais significativos da contribuição de dados em cada município analisado.

Essa diferenciação entre a área total do município e a área urbana permitiu uma análise mais detalhada e focada nos aspectos relacionados à qualidade dos dados de VGI dentro das zonas urbanas. Ao restringir as análises à área urbana, foi possível observar diferenças significativas nos padrões de contribuições, na distribuição dos dados e na densidade de contribuidores. Como exemplo, a Figura 6 das grades de 3 km nos municípios de São Paulo e Curitiba destacam visualmente essas diferenças e permitem uma análise comparativa entre duas localidades relevantes em termos de tamanho urbano, atividade no OSM.

Figura 6 - Grades de 3 km aplicadas à área urbana dos municípios de São Paulo e Curitiba



Fonte: Autor (2024)

Para os municípios de Joaquim Gomes, Cacequi, Ponto Belo, Matupá e Taguatinga, todas de pequeno porte, a área urbana é muito menor que os demais

municípios analisados, conforme dados do IBGE; entre estes municípios citados a área urbana média é de aproximadamente 10 km².

Devido às pequenas dimensões das áreas urbanas desses municípios, foi inviável aplicar a grade de 3 km exclusivamente nessas regiões. O próprio programa utilizado não aceitava a definição de uma célula tão grande para áreas tão reduzidas. Esse cenário evidencia a necessidade de ajustar a metodologia de análise para municípios de pequeno porte ou com áreas urbanas restritas.

Os resultados obtidos reforçam a importância de adaptar o tamanho da célula de análise conforme a dimensão das áreas urbanas. Em municípios de pequeno porte ou com áreas urbanas limitadas, como observado em Joaquim Gomes, Cacequi, Ponto Belo, Matupá e Taguatinga, é essencial utilizar grades menores para garantir uma análise mais precisa e relevante dos dados de VGI no contexto do OSM. Essa abordagem permite capturar nuances e padrões específicos dessas localidades, contribuindo para uma compreensão mais aprofundada do panorama do mapeamento colaborativo nessas áreas.

Ao adotar uma abordagem que prioriza a análise na área urbana, em oposição à delimitação municipal geral, identificamos a necessidade de ajustar o tamanho da célula conforme a dimensão das áreas urbanas. Isso se deve ao fato de que abranger a área conforme a delimitação municipal acaba incluindo áreas verdes, reservas, áreas hidrográficas, entre outras, que podem interferir nos dados, causando falsas impressões de valor e resultando em conclusões não realistas. A variação do tamanho da célula demonstrou ser um elemento-chave para evitar distorções nos dados, oferecendo uma visão mais precisa e confiável do panorama do mapeamento colaborativo nessas áreas.

4.2 CÁLCULO DO RMSE

Conforme pode ser visto abaixo na Tabela 1, a variação na quantidade de grades por município pode influenciar significativamente os valores do RMSE. Este cálculo é feito para avaliar a qualidade das regressões, pois afeta a distribuição espacial das observações e a representatividade das amostras em relação à área total do município. Em municípios com maior quantidade de grades, a densidade de

amostragem é maior, o que pode resultar em estimativas mais precisas do RMSE devido à maior densidade de amostragem, enquanto municípios com menor quantidade de grades podem ter valores de RMSE menos precisos devido à menor densidade de amostragem, calculada tomando-se a raiz quadrada da média dos quadrados dos erros, onde o erro bruto é a diferença entre o valor previsto pelo modelo e o valor real, levando em conta o número de amostras, o valor observado para a amostra e o valor previsto pelo modelo, ou seja, quanto menor o RMSE, melhor o modelo está se saindo, tudo isso calculado automaticamente pelo algoritmo desenvolvido por Elias et al. (2023).

Tabela 1 - Valor médios de RMSE por município

Município	N° Células	1 ano	5 anos	10 anos
Taguatinga	532	0.63%	3.12%	11.35%
Joaquim Gomes	84	0.06%	3.85%	19.35%
Matupá	1922	0.14%	1.70%	6.22%
Ponto Belo	40	0.58%	7.65%	31.38%
Cacequi	640	4.15%	3.88%	11.11%
Santarém	3870	1.18%	3.03%	10.74%
Juazeiro do Norte	49	0.19%	3.75%	11.09%
Anápolis	238	0.70%	4.87%	19.57%
Juiz de Fora	378	0.14%	4.22%	24.61%
Chapecó	182	1.28%	8.58%	25.65%
Belém	247	1.05%	7.64%	14.92%
Fortaleza	72	5.83%	63.85%	69.05%
Goiânia	210	0.55%	4.27%	13.48%
São Paulo	450	0.45%	4.16%	9.18%
Curitiba	96	0.49%	6.90%	15.11%

Fonte: Autor (2024)

Observa-se que, das cinco regiões analisadas nos três portes de classificação, os municípios de pequeno porte tendem a apresentar valores de RMSE mais baixos nos três períodos analisados, nas regiões norte e centro-oeste, em comparação com os municípios de médio e grande porte, além disso, é o menor valor também para região nordeste no período de um ano, e na região sul nos períodos de 5 e 10 anos, ou seja, 60% dos menores erros médio quadrático estão nos municípios de pequeno porte.

Ao comparar os valores de RMSE entre municípios de diferentes portes, observa-se que apresentam valores de RMSE consistentemente baixos em todos os períodos analisados (1, 5 e 10 anos), possivelmente devido a um menor volume de informações registradas. Nos municípios de médio porte, mostram uma variação nos valores de RMSE, com alguns municípios apresentando valores mais baixos em comparação com outras. Já nas de grande porte, tendem a apresentar valores de RMSE mais elevados em comparação com municípios de pequeno e médio porte.

Após esta análise inicial, que indicou uma variação significativa nos valores de RMSE em função do porte dos municípios e da densidade de grades de amostragem, decidiu-se realizar uma avaliação adicional focada exclusivamente nas áreas urbanas de cada município, como podem ser vistos os resultados na Tabela 2.

Esta etapa visou verificar se a extensão total da área urbana, comparativamente à área total do município, exerce influência determinante sobre a precisão das estimativas do RMSE. Para isso, foram reprocessados os dados, isolando as observações pertinentes às zonas urbanizadas, a fim de obter um novo conjunto de valores de RMSE.

Tabela 2 - Valor médios de RMSE da área urbana por município

Município	Nº Células	1 ano	5 anos	10 anos
Taguatinga	16	1.45%	10.47%	26.48%
Joaquim Gomes	4	0.36%	5.58%	10.76%
Matupá	42	0.02%	1.60%	7.30%
Ponto Belo	4	0.00%	6.48%	58.25%

Município	Nº Células	1 ano	5 anos	10 anos
Cacequi	6	3.65%	13.85%	18.53%
Santarém	209	0.52%	5.45%	11.46%
Juazeiro do Norte	9	0.13%	3.78%	7.40%
Anápolis	54	0.07%	3.10%	13.05%
Juiz de Fora	42	0.12%	2.38%	9.06%
Chapecó	16	0.91%	9.73%	23.85%
Belém	40	0.53%	11.96%	16.73%
Fortaleza	36	11.08%	9.69%	10.54%
Goiânia	100	0.52%	4.14%	11.46%
São Paulo	240	0.61%	5.08%	10.07%
Curitiba	70	0.26%	6.60%	16.42%

Fonte: Autor (2024)

Este procedimento visou elucidar se as características das áreas urbanas, como a maior concentração de pontos de dados e a possível homogeneidade na distribuição espacial das observações, poderiam resultar em variações nos erros de estimativa quando comparados aos dados anteriores que contemplavam o município como um todo, visto que esta é uma métrica de avaliação utilizada para medir o desempenho de modelos de regressão.

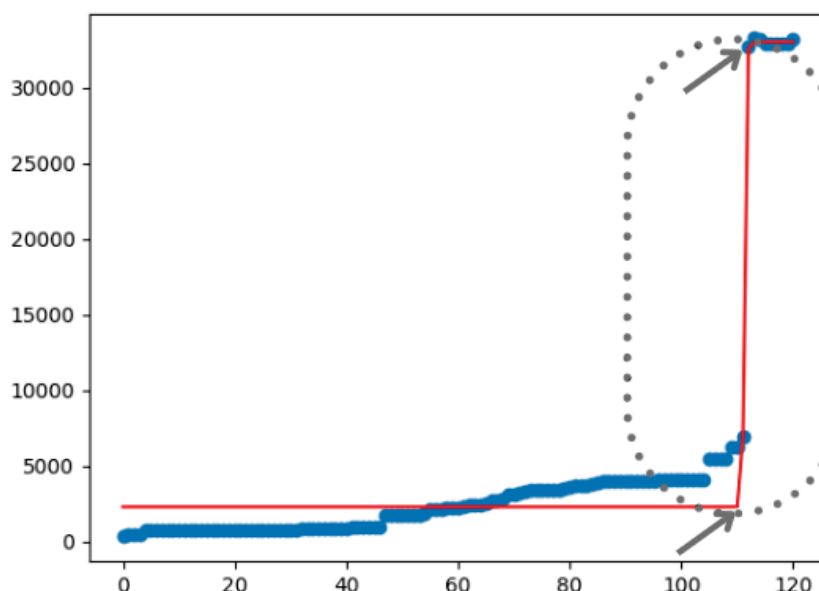
Ao concentrar a análise nas áreas urbanas, observou-se que das cinco municípios de pequeno porte analisadas nos três períodos, 60% apresentaram um aumento nos valores de RMSE. Já nos municípios de médio e grande porte, apenas 26,6% registraram um aumento no RMSE, porém foi notável também no período de 1 ano para municípios de médio porte e de 10 anos para os municípios de grande porte, 100% dos valores diminuíram. Em uma análise geral, de todos os períodos e portes, 62,2% dos valores de RMSE obtidos na análise das áreas urbanas foram inferiores aos registrados na análise da área total anteriormente realizada. Esses resultados indicam uma tendência de maior precisão nas estimativas de RMSE ao focar na análise das áreas urbanas, sugerindo que as características específicas

dessas áreas, como a densidade de dados e a distribuição espacial homogênea das observações, contribuem para uma maior acurácia nas estimativas de RMSE em comparação com a abordagem anterior que considerava a área total de cada município.

4.3 DADOS EM MASSA

Na análise da regressão logística processada em cada uma das grades geradas para os 15 municípios, exclusivamente nas áreas urbanas, em algumas delas o valor do ângulo θ foi de 0° (parâmetro d). No processamento, após a geração resultante dos dados em forma de tabelas, a identificação dessas células foi realizada a partir de uma mensagem de erro, que representa uma solução não possível, limitada a condição do modelo utilizado no código, diretamente relacionado ao alto percentual de inserções de recursos sobre o total acumulado, conforme pode ser visto no exemplo da Figura 7, com a cidade de Fortaleza/CE analisada no período de 10 anos (120 meses).

Figura 7 - Exemplo de grade limitada a condição do modelo utilizado na cidade de Fortaleza



Fonte: Autor (2024)

Na Tabela 3, apresentamos uma visão detalhada do número de grades afetadas por uma solução não possível (*overflow*) nos respectivos municípios e nos períodos de análise em que tais eventos ocorreram.

Tabela 3 - Relação de nº de grades atingidas por dados em massa por município

Município	Nº de Grades	1 ano	5 anos	10 anos
Taguatinga	532	106	215	214
Joaquim Gomes	84	6	36	20
Matupá	1922	201	612	893
Ponto Belo	40	5	2	2
Cacequi	640	43	208	299
Santarém	3870	1403	952	1817
Juazeiro do Norte	49	3	10	4
Anápolis	238	2	13	4
Juiz de Fora	378	3	32	52
Chapecó	182	7	20	12
Belém	247	19	60	71
Fortaleza	72	5	15	11
Goiânia	210	3	8	20
São Paulo	450	22	41	30
Curitiba	96	4	4	4

Fonte: Autor (2024)

Inicialmente, as regressões logísticas foram processadas para as grades que cobriam a área total dos 15 municípios, onde já foram identificados fenômenos limitados a condição do modelo utilizado, indicando um alto percentual de inserções de dados em uma única grade sobre o total acumulado. Posteriormente, uma análise mais focada foi realizada exclusivamente nas áreas urbanas delimitadas manualmente de cada município. Esta segunda fase de análise também revelou

limites à condição do modelo utilizado, como pode ser visto na Tabela 4 permitindo uma comparação direta entre os comportamentos de contribuição nas áreas totais dos municípios e nas suas respectivas áreas urbanas.

Tabela 4 - Relação de nº de grades área urbana atingidas por dados em massa por município

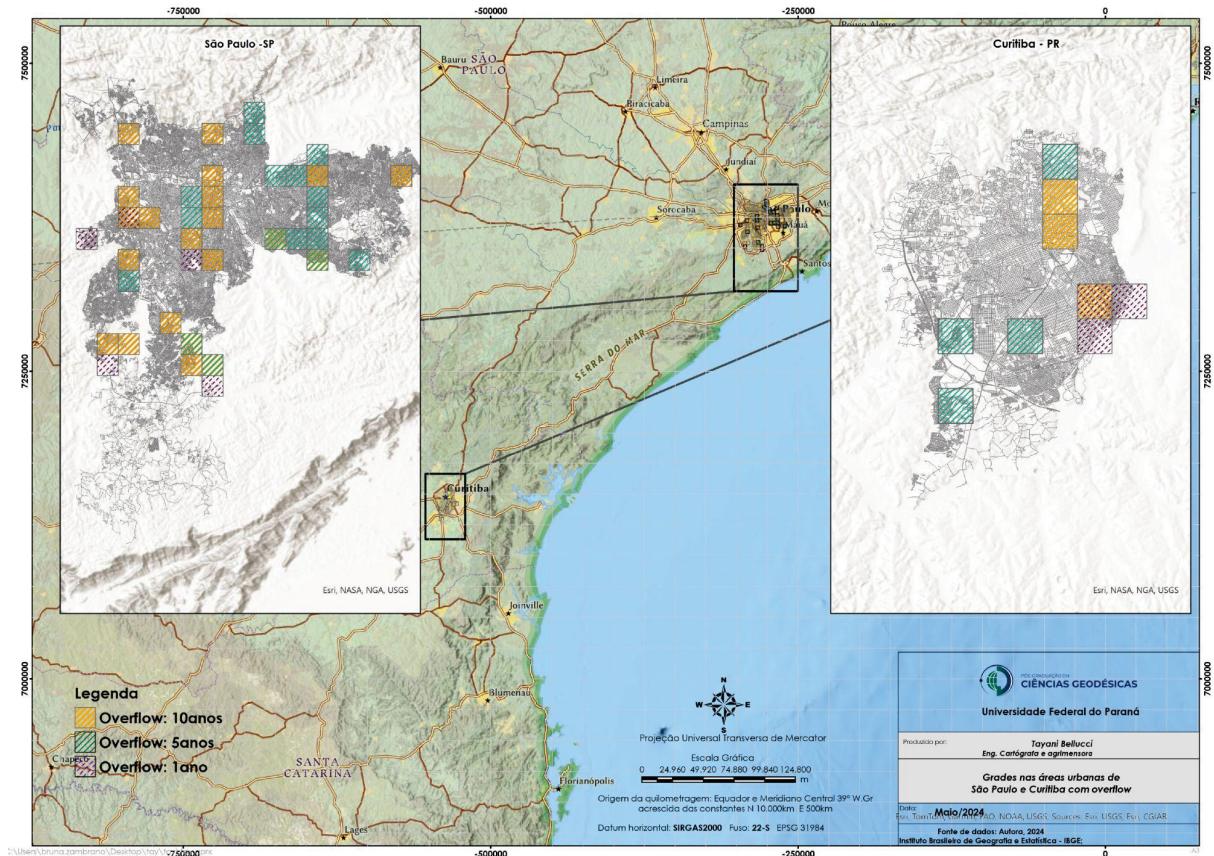
Município	Nº de Grades	1 ano	5 anos	10 anos
Taguatinga	16	0	4	1
Joaquim Gomes	4	0	0	0
Matupá	42	0	15	14
Ponto Belo	4	1	0	0
Cacequi	6	1	1	1
Santarém	209	22	43	32
Juazeiro do Norte	9	0	0	0
Anápolis	54	3	2	1
Juiz de Fora	42	2	1	0
Chapecó	16	1	2	1
Belém	40	1	1	4
Fortaleza	36	3	4	3
Goiânia	100	12	2	1
São Paulo	238	5	15	21
Curitiba	56	3	4	3

Fonte: Autor (2024)

Posteriormente, a análise se concentrou nas áreas urbanas de São Paulo e Curitiba, que se destacam pelo alto número de contribuidores e contribuições no OSM e pela complexidade de seus padrões urbanos. A análise detalhada das grades onde foram detectados limites a condição do modelo utilizado nessas áreas urbanas permite, conforme pode ser visto na Figura 8, não apenas visualizar os desafios específicos enfrentados, mas também oferece uma plataforma para discutir

como as contribuições em massa podem influenciar a qualidade dos dados geoespaciais coletados.

Figura 8 - Grades nas áreas de São Paulo e Curitiba com limites a condição do modelo utilizado

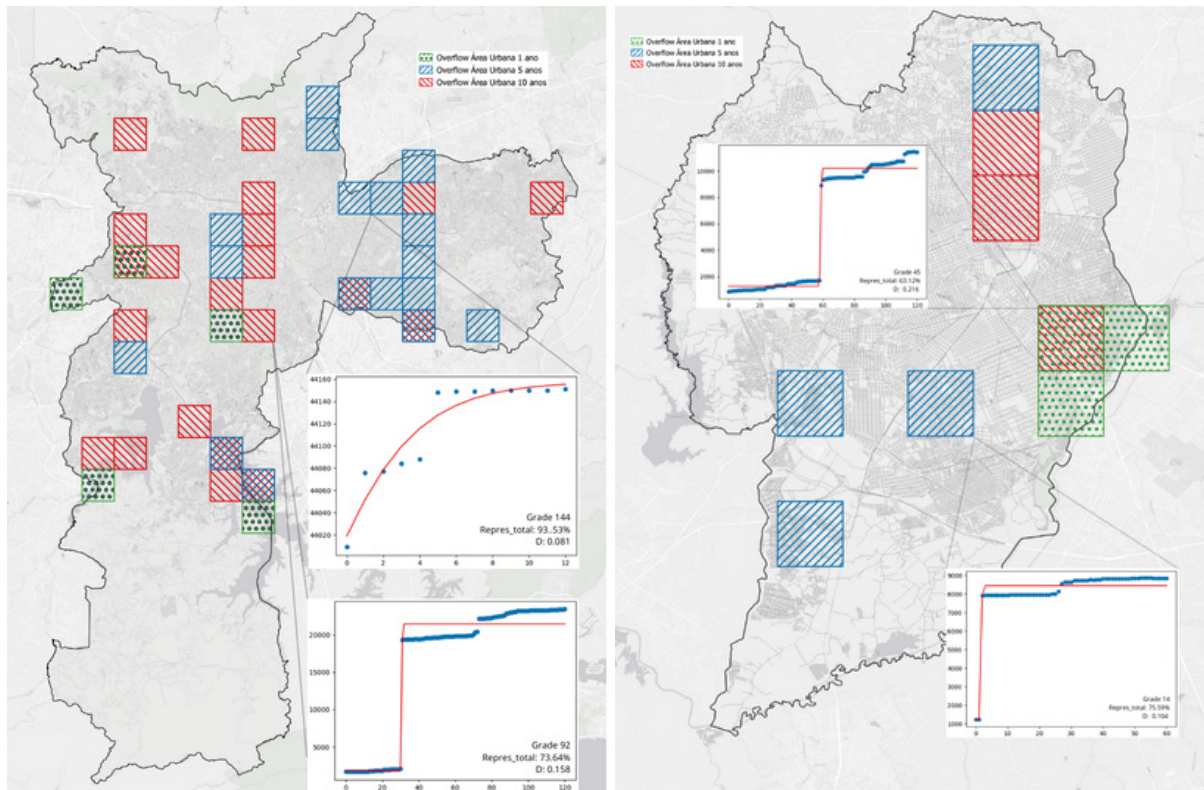


Fonte: Autor (2024)

A análise dos limites da condição do modelo utilizado nas áreas urbanas destacou a importância da delimitação das áreas para uma análise realista desses eventos. Observou-se uma significativa redução no número de ocorrências ao examinar diretamente as áreas urbanas, em comparação com a análise das áreas totais dos municípios. Essa diferença evidencia a influência da delimitação geográfica na interpretação correta dos dados de contribuição no OSM, destacando a necessidade de direcionar a análise para as áreas de interesse específicas, como exemplificado na Figura 9 a seguir. Além disso, vale ressaltar que a função senoidal utilizada na regressão logística pode interferir, especialmente devido aos valores máximos e mínimos dos parâmetros A e D, os quais tendem a se aproximar de 90° ou 180° . Esses valores extremos podem resultar em erros na função de regressão

logística, influenciando diretamente a ocorrência de erros e a interpretação dos dados de contribuição.

Figura 9 - Grades limitadas a condição do modelo utilizado nos municípios de São Paulo e Curitiba



Fonte: Autor (2024)

Considerando os 15 municípios abrangidos pelas áreas de estudo, os quais englobam localidades de pequeno, médio e grande porte, ao longo dos períodos de 5, 10 e 15 anos, foi observado que no período de 1 ano apenas Cacequi apresentou uma coincidência de limitações do modelo utilizado na grade de maior colaboração em relação à representação total, representando 6,66% dos municípios. No período de 5 anos, Taguatinga, Matupá e Curitiba, registraram uma coincidência representando 20% dos municípios, enquanto no período de 10 anos, Taguatinga, Anápolis e Curitiba também apresentaram uma incidência também representando 20%.

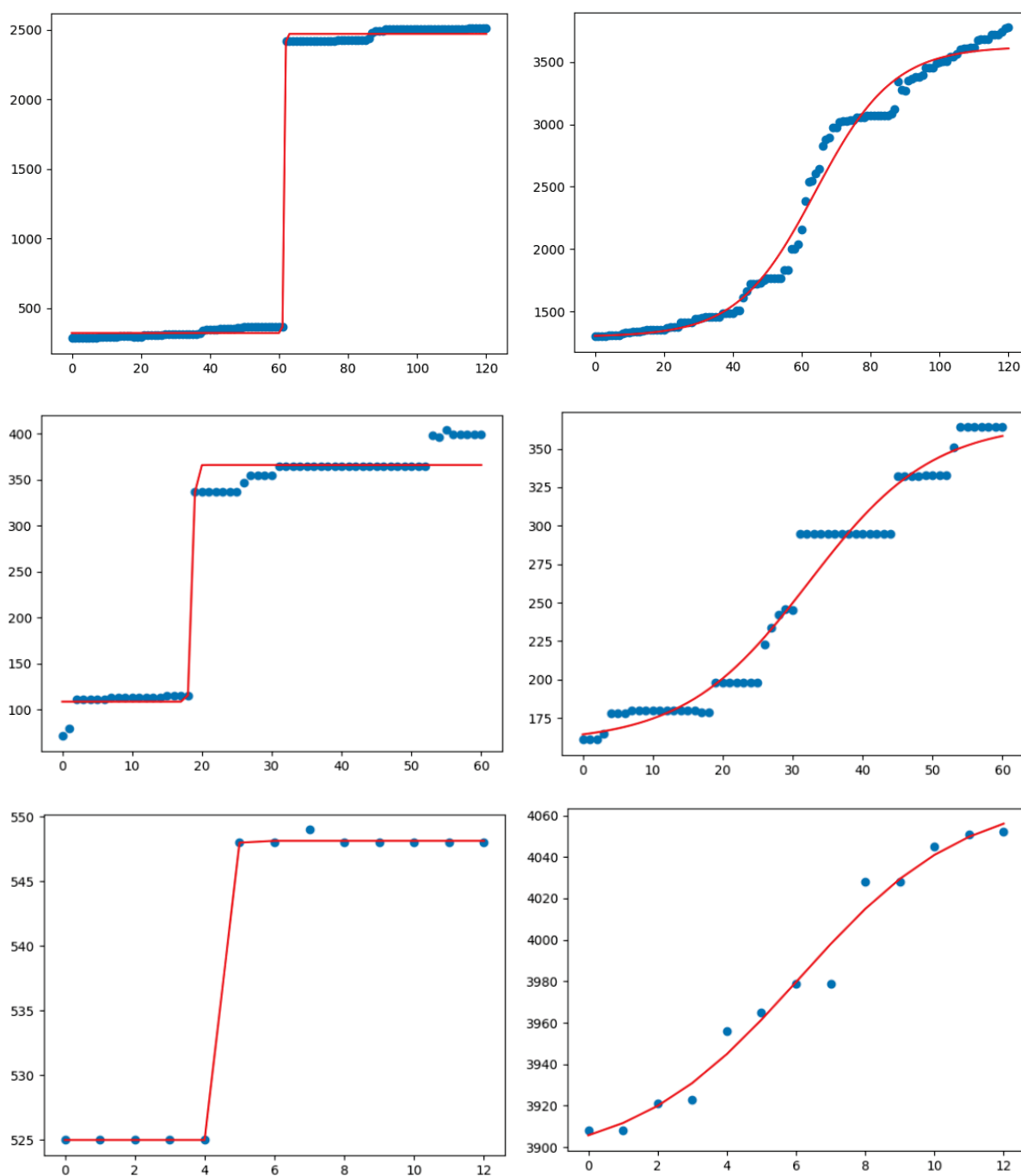
Dessa forma, a hipótese levantada por Elias (2022) não se mostrou aplicável nesse contexto específico. Elias (2022) conduziu testes na região metropolitana em torno de Curitiba e no município de Curitiba, encontrando células afetadas por limitações do uso do modelo, devido a contribuições em massa, ao analisar grades

de 1x1km, conforme analisado por ele. Contudo, ao adotar células de 3x3km e considerar a análise dos dias com maior contribuição divididos pelo número total de contribuições, não foi observada a mesma tendência de ocorrência de soluções não possíveis, neste caso, demonstrando uma divergência em relação aos resultados encontrados por Elias (2022). Essa divergência indica a importância de considerar diferentes fatores e abordagens metodológicas ao analisar a ocorrência de soluções não possíveis e a contribuição em massa no contexto do mapeamento colaborativo no OSM.

Nos gráficos das Figuras 10 e 11, de Curitiba e São Paulo, o azul representa o crescimento acumulado de contribuições, e o vermelho representa a curva de regressão logística, no período de 10 anos, com grades 3x3km. Vemos uma figura com salto no número de dados inseridos e outras com crescimento vegetativo, com quantidades proporcionais, o que gerou uma curva mais suave. Isto é possível descrever a suavidade da curva em termos do parâmetro de inclinação D (Equação 1), em que quanto menor o valor obtido, maior será a sua inclinação.

Portanto, conforme explorado por Elias (2022), Elias et al. (2023) e Peterlini et al. (2023), confirmou-se que a medida utilizada pode influenciar na interpretação dos padrões temporais de contribuições para o *OpenStreetMap*. Células grandes podem mascarar variações locais, enquanto células pequenas podem introduzir ruído excessivo, ou seja, a escolha do tamanho da célula é essencial para a qualidade e interpretação dos dados geoespaciais em plataformas colaborativas como o OSM.

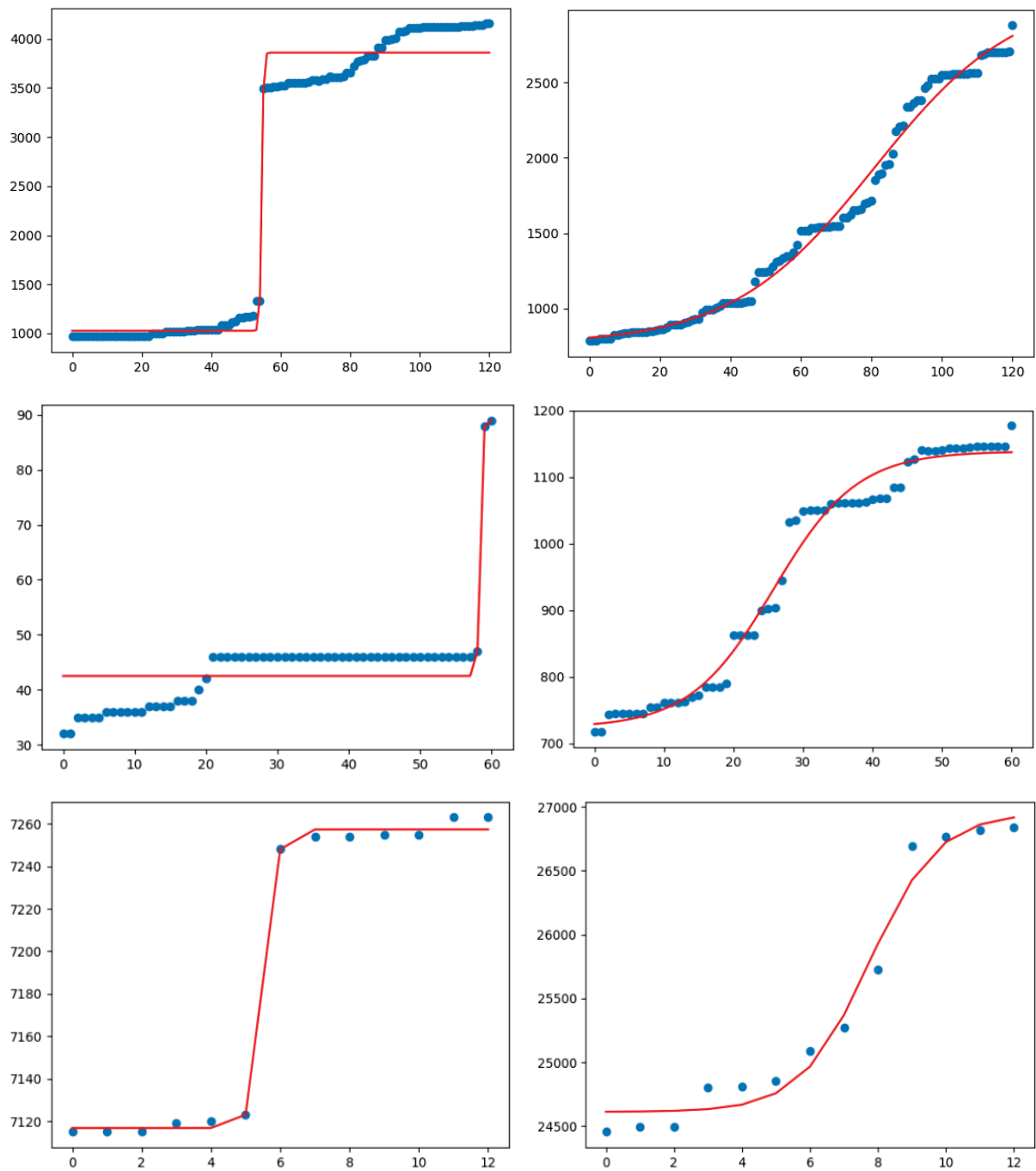
Figura 10 - Grades 3x3 da cidade de Curitiba no período de 10, 5 e 1 ano, com e sem soluções do modelo utilizado



Fonte: Autor (2024)

Desenvolver gráficos da quantidade acumulada no período de 1, 5 e 10 anos possibilitou identificar e avaliar a contribuição padrão, além do comportamento de saltos abruptos. Figura 11 apresenta exemplos dos resultados obtidos onde em uma grade exemplificada a regressão logística apresentou limitações no uso do modelo, indicando o salto abrupto de São Paulo e na outra um crescimento vegetativo.

Figura 11 - Grades 3x3 da cidade de São Paulo no período de 10, 5 e 1 ano, com e sem soluções do modelo utilizado



Fonte: Autor (2024)

Como já dito anteriormente, ao medir a taxa de crescimento dos dados, podemos determinar quais áreas estão recebendo mais atenção dos mapeadores voluntários e em quais períodos. Isso é fundamental para compreender como e quando as comunidades estão se mobilizando para contribuir com dados, o que

pode estar relacionado a eventos específicos, campanhas de mapeamento, ou mudanças nas necessidades de dados geográficos.

A inclinação da curva logística é essencialmente a taxa de variação da função logística em pontos específicos. Para calcular a inclinação da curva, foi implementada uma função que deriva a expressão da curva logística ajustada, permitindo determinar o ângulo de inclinação em graus, conforme apresentado no exemplo da Figura 12. Este cálculo analisa a intensidade de crescimento dos dados VGI ao longo do tempo em diferentes regiões, refletindo a intensidade e a qualidade das contribuições voluntárias.

Figura 12 - Exemplo de valores de inclinação da curva de regressão por grade

Grade	2	: Inclinação = -89.8960845434814 graus
Grade	3	: Inclinação = 88.98676979065624 graus
Grade	4	: Inclinação = 89.51657833805282 graus
Grade	5	: Inclinação = 89.57373566650884 graus

Fonte: Autor (2024)

Isso pode acontecer por diferentes motivos, com valores extremamente grandes ou D sendo muito pequeno, faz com que o denominador da exponencial se aproxime de zero. Essas situações podem levar a resultados com erro de limitação do modelo utilizado.

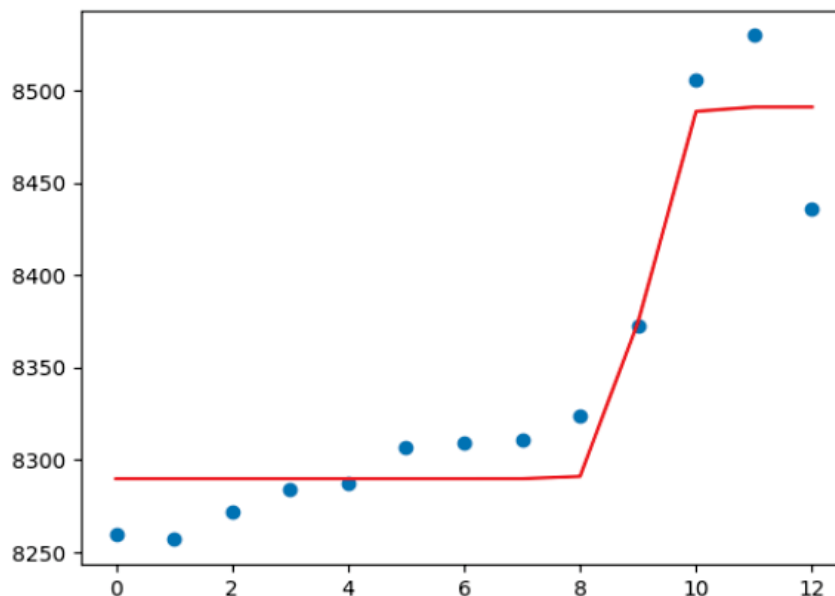
Em São Paulo, no período de 10 anos, as áreas com erro apresentam uma representação total elevada. Isso indica uma contribuição massiva, semelhante ao observado em Curitiba, conforme já apresentado por Elias (2023). Esses erros são significativos e refletem uma participação em massa dos dados.

Para análises anuais, a sensibilidade do modelo de regressão logística é reduzida. Embora exista uma contribuição significativa em determinadas grades, o modelo não se mostra eficaz, mesmo com as análises sendo realizadas mês a mês para diferentes regiões e intervalos de tempo.

A análise mensal a longo prazo demonstra maior efetividade. No período temporal de um ano, os dados não apresentam uma adequação satisfatória, como evidenciado na Figura 13 abaixo. Nesta figura, a exclusão de dados é evidente, conforme indicado pela bolinha azul, resultando em uma queda, porém não ocorre a

modelagem correta da curva.

Figura 13 - Exemplo de modelagem inconsistente



Fonte: Autor (2024)

Na análise de Elias (2022), quando aplicada a modelagem para Curitiba foi possível identificar áreas afetadas por endereçamentos. No entanto, em nossa análise de 1 ano, não foi possível identificar a ocorrência desses eventos no período da regressão. A relação entre as ocorrências e a inclusão em massa dos dados não foi encontrada, indicando a necessidade de utilizar outros modelos matemáticos de regressão para uma análise eficaz.

Considerando um recorte temporal de apenas um ano, não é possível identificar variações significativas. Portanto, para realizar uma análise consistente, é necessário um período maior, como 5 ou 10 anos, para a identificação de limites do uso do modelo. Isso pode ser visualmente notado anteriormente nas Figuras 9 e 10 de São Paulo e Curitiba. As comparações das regressões em diferentes períodos, com ou sem soluções do modelo utilizado, demonstram que o modelo não é adequado para poucos dados e intervalos curtos de tempo.

Portanto o programa demonstrou-se eficaz e sensível para análises mensais em diferentes cidades nos períodos de 5 e 10 anos, permitindo a elaboração precisa dos gráficos de regressão logística e a especificação das áreas afetadas por dados

em massa, porém para análises anuais, a eficácia do modelo foi reduzida, não sendo capaz de capturar variações significativas, o que sugere a necessidade de empregar outros métodos de análise para períodos mais curtos.

4.4 USUÁRIOS OSM

Ao aplicar o código desenvolvido aos diferentes municípios analisados, observamos uma variabilidade significativa no número de usuários ativos e no volume de contribuições no OSM. Por exemplo, em municípios como São Paulo e Curitiba, como pode ser visto no Tabela 5, as contribuições foram substancialmente maiores, refletindo uma comunidade de mapeamento mais ativa e possivelmente uma maior consciência sobre a importância do mapeamento colaborativo. Em contrapartida, municípios menores como Matupá e Cacequi apresentaram um número menor de contribuições, o que pode indicar uma oportunidade de crescimento e de engajamento da comunidade na manutenção dos dados geográficos.

Tabela 5 - Número de usuários por intervalo de tempo em cada município

Município	1 ano	5 anos	10 anos
São Paulo	2806917	1906867	166957
Curitiba	452764	232250	35387
Belém	28062	11694	3913
Fortaleza	497766	55862	12017
Chapecó	31277	13397	6213
Cacequi	1323	412	229
Matupá	988	557	368
Juiz de Fora	11464	7823	3934
Taguatinga	12393	7681	3907
Joaquim Gomes	489	324	236
Santarém	17244	3012	1712
Goiânia	37134	21050	11065

Município	1 ano	5 anos	10 anos
Anápolis	2664	1399	251
Juazeiro do Norte	5029	2787	1547
Ponto Belo	624	36	20

Fonte: Autor (2024)

Os resultados obtidos realçam a disparidade entre áreas urbanas densamente povoadas e regiões menos densas em termos de dados geográficos disponíveis e atualizados. Esta análise permite discutir sobre a relevância do engajamento comunitário no mapeamento colaborativo, que é crucial para a manutenção da precisão e da atualidade dos mapas no OSM.

Visto que foi feita uma implementação específica no código utilizado para este trabalho para determinar o número de contribuintes por município, surgiu a oportunidade de analisar a relação entre o número de contribuintes e o IDH. Embora essa análise não seja o objetivo principal, os dados fornecidos pela ferramenta permitem explorar como o desenvolvimento humano pode influenciar a participação de dados geográficos voluntários. O IDH é um indicador compreensivo que reflete não apenas a capacidade econômica, mas também o acesso à educação e à saúde, proporcionando uma visão ampla sobre as condições de vida nas respectivas localidades. Os dados coletados revelam variações significativas no desenvolvimento humano entre os municípios estudados, conforme apresentado na Tabela 6. Tais discrepâncias são indicativas de desigualdades no acesso a recursos essenciais e oportunidades.

A discussão a seguir busca explorar a possível relação entre o número de contribuintes e o valor de IDH, considerando como variações no IDH podem refletir nas contribuições para a OSM. Essa análise visou identificar se municípios com IDH mais elevado apresentam maior número de contribuições, sugerindo uma correlação entre desenvolvimento humano e participação da população na produção de dados geográficos voluntários.

Tabela 6 - IDH dos municípios analisadas

Município	IDH
Curitiba	0.823
São Paulo	0.805
Goiânia	0.799
Fortaleza	0.754
Belém	0.746
Chapecó	0.727
Juiz de Fora	0.711
Anápolis	0.66
Santarém	0.648
Juazeiro do Norte	0.642
Matupá	0.631
Cacequi	0.626
Ponto Belo	0.605
Taguatinga	0.497
Joaquim Gomes	0.386

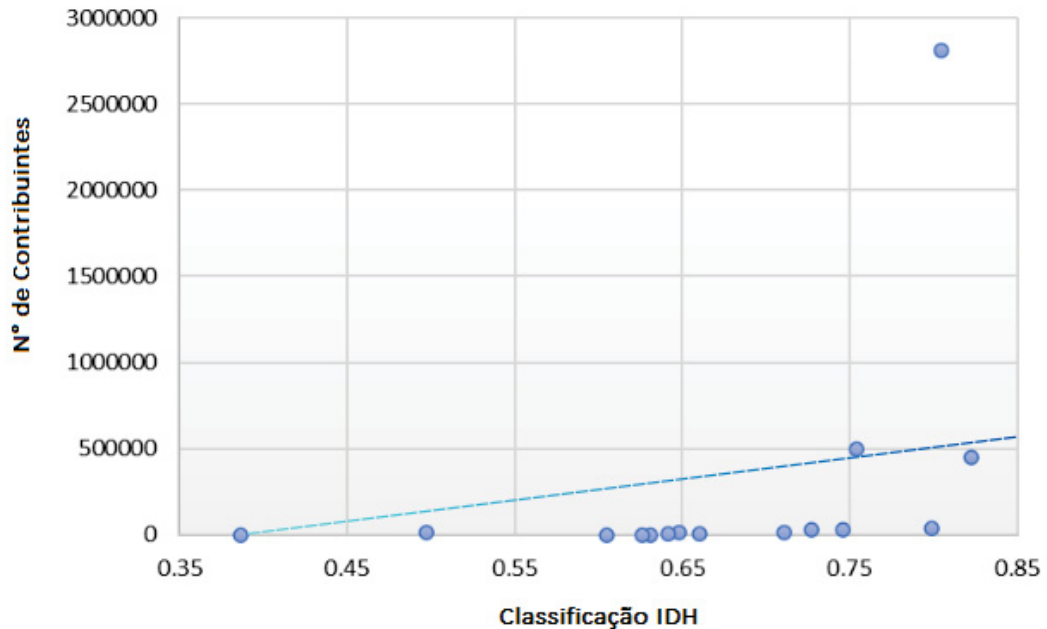
Fonte: AtlasBR (2024)

Municípios como São Paulo (0.805) e Curitiba (0.823) apresentam índices superiores, refletindo melhor infraestrutura, maior investimento público e privado e um mercado de trabalho mais diversificado. Em contraste, municípios como Joaquim Gomes (0.386) e Taguatinga (0.497) mostram valores consideravelmente mais baixos.

Ao analisar o número de usuários do OSM nos municípios ao longo de três períodos distintos (1, 5 e 10 anos), observa-se uma notável estabilidade na classificação por número de usuários entre os municípios. Em todos os períodos analisados, a ordem dos municípios, com base no volume de usuários, manteve-se

consistentemente similar, apresentando variações mínimas, que não excederam duas posições em apenas alguns casos isolados.

Figura 14 - Dispersão do n° de contribuintes x IDH



Fonte: Autora (2024)

Como informado anteriormente através de uma implementação específica no código utilizado para este trabalho, foi possível determinar o número de contribuintes por município e visualizar a relação com o IDH através de scatterplots com linhas ajustadas. Os resultados indicam que municípios com IDH mais elevado, como São Paulo e Curitiba, tendem a ter um maior número de usuários contribuindo para o OSM, em comparação com municípios com IDH mais baixo, como Joaquim Gomes e Taguatinga. Essa correlação sugere que o desenvolvimento humano, refletido pelo IDH, está relacionado à capacidade e ao interesse da população em participar de iniciativas de VGI, ou seja, municípios com maior número de contribuidores tendem a ter contribuição em massa com mais frequência, tendo um crescimento vegetativo mais alto do que municípios menores, com menor número de contribuidores.

Esta análise comparativa entre a estabilidade do número de usuários e os indicadores de IDH nos municípios sugere que, em municípios com maior IDH, as infraestruturas e condições socioeconômicas favorecem a continuidade e o crescimento do uso de VGI, como OSM, enquanto municípios com IDH menor podem enfrentar barreiras que limitam tal expansão.

5 CONCLUSÃO

Considerando os resultados obtidos e as análises realizadas ao longo deste estudo, é possível inferir que a análise da qualidade dos dados de VGI no contexto do OSM é um processo complexo que requer não apenas rigor metodológico, mas também consideração cuidadosa dos diversos fatores que influenciam a precisão e relevância das conclusões. A análise dos municípios de diferentes portes, juntamente com a variação do tamanho da célula de acordo com a área urbana estudada, revelou a influência dos dados na ocorrência de erros devido a limitação do uso do modelo escolhido e na interpretação. A identificação da interferência da função senoidal nos resultados destaca a importância de considerar não apenas os modelos estatísticos utilizados, mas também as limitações e nuances inerentes ao processo de análise.

Do ponto de vista técnico, a observação de que a função senoidal pode sofrer com as características dos dados presentes em uma região geográfica, assim os programas devem incluir funções descritivas para análise. O erro causado por valores abruptos de contribuição, ressalta a necessidade contínua de refinamento e aprimoramento das técnicas de análise para melhoria de precisão dos resultados. Além disso, a constatação de que o tamanho das células também desempenha um papel crucial na qualidade dos resultados sublinha a importância de adaptar o tamanho da célula de análise de acordo com a dimensão das áreas urbanas estudadas.

Por outro lado, do ponto de vista científico, a constatação da alta heterogeneidade das áreas municipais no Brasil e a necessidade de variação da grade de estudo conforme o tamanho da área analisada apontam para a importância de uma abordagem adaptativa e contextualizada na análise da qualidade dos dados de VGI, ou seja, conforme visto é necessário adaptação do tamanho da grade especificamente baseado na área urbana de análise para que não haja homogeneização de dados, além de um período maior que um ano conforme analisado para que os dados possam ser analisados corretamente.

O programa demonstrou-se válido para análises mensais nas diferentes cidades nos períodos de 5 e 10 anos, sendo eficaz e sensível na elaboração dos

gráficos de regressão logística e na especificação das áreas e grades afetadas por dados em massa. Para esses períodos mais longos, foi possível identificar padrões e variações relevantes, conforme evidenciado nos resultados apresentados. Entretanto, para análises realizadas no período de um ano, a eficácia e a sensibilidade do modelo de regressão logística foram significativamente reduzidas. Os dados não apresentaram variações adequadas, e a modelagem da curva tornou-se imprecisa, indicando a necessidade de empregar outros métodos de análise e modelos matemáticos alternativos para períodos mais curtos. Assim, a presente pesquisa contribui não apenas para o avanço do conhecimento científico no campo da análise de dados geoespaciais, mas também para a aplicabilidade prática desses conhecimentos no contexto do mapeamento colaborativo e na tomada de decisões informadas em diferentes áreas de atuação.

Para analisar e avaliar a aplicabilidade e a eficácia das ferramentas, é fundamental destacar que a mera utilização de um *plugin* ou *software* não garante resultados precisos ou interpretáveis. A análise realizada revelou que a relação entre o tamanho da área urbana e o tamanho da célula, bem como a inclinação da curva e o período de análise, desempenham papéis cruciais na precisão dos resultados. A ferramenta de análise deve ser constantemente avaliada e adaptada conforme as necessidades específicas do estudo. Não se trata apenas de aplicar dados e gerar gráficos, mas sim de entender que os resultados são fruto de uma combinação de variáveis, nesse caso, incluindo o tamanho da célula, o período de contribuição e as informações fornecidas pelo usuário. Essas descobertas sublinham a necessidade de uma abordagem crítica e meticulosa ao utilizar ferramentas de análise, fazendo com que as conclusões sejam não apenas baseadas em dados, mas também contextualizadas e interpretadas corretamente segundo as particularidades de cada estudo.

6 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Com base nas análises e conclusões apresentadas nos capítulos anteriores, as recomendações para trabalhos futuros visam expandir e aprofundar nossa compreensão sobre os temas abordados. Este capítulo oferece diretrizes e sugestões para pesquisas subsequentes, explorando áreas de interesse que surgiram durante este estudo

Recomenda-se a continuidade das pesquisas nos diferentes fatores relacionados com a inserção de feições no VGI ao longo do tempo, incluindo a exploração de outras formas de modelagem matemática. Além disso, considerar a inclusão de outras variáveis que possam influenciar a qualidade e a precisão dos dados obtidos através do VGI. Recomenda-se a utilização de grades estatísticas ou setores censitários para a delimitação das áreas de estudo, o que traria maior consistência e alinhamento com padrões demográficos amplamente utilizados.

Uma sugestão adicional é a incorporação de índices de desenvolvimento, como o Índice de Desenvolvimento da Educação Básica (IDEB), como variável contextual para análises de qualidade dos dados. Esse índice poderia fornecer uma base comparativa interessante para avaliar a relação entre desenvolvimento educacional e a precisão das contribuições voluntárias, além de permitir uma análise mais segmentada por áreas urbanas e rurais dentro dos municípios. A análise por setor censitário também se apresenta como uma abordagem promissora, permitindo uma compreensão mais granular e detalhada das dinâmicas de inserção de dados em diferentes partes dos municípios.

Outra sugestão é a automatização da análise das grades utilizadas, de modo que, com base no tamanho da área analisada, seja apresentado automaticamente o tamanho de grade ideal a ser utilizado. Esta automação não apenas agilizaria o processo de análise, mas também auxiliaria com uma abordagem mais precisa e eficiente na determinação das grades para diferentes tipos de áreas.

Além disso, é recomendável considerar a automatização da inserção das listas de coordenadas no programa utilizado, ou até mesmo criar uma forma de identificação direta no mapa da área de análise, nas regiões urbanizadas. Isso

facilitaria significativamente o manuseio e a organização dos dados, tornando o processo mais intuitivo e menos suscetível a erros humanos.

Essas recomendações são fundamentais para o avanço contínuo da pesquisa em VGI e na avaliação da qualidade dos dados obtidos, proporcionando uma base sólida para futuros estudos e aplicações práticas.

REFERÊNCIAS

- ALMEIDA, et al. Controle de Qualidade Posicional do Sistema Rodoviário do OpenStreetMap na Região Central de Viçosa, MG. **Boletim de Ciências Geodésicas**, 2021.
- AL-BAKRI. **Developing Tools and Models for Evaluating Geospatial Data Integration of Official and VGI Data Sources**. Doctoral Thesis, School of Civil Engineering and Geosciences, Newcastle University, 2012.
- ATKINSON, Anthony B. **Desigualdade: O que pode ser feito?** . Imprensa da Universidade de Harvard, 2015.
- ARDILA, Alfredo et al. The influence of the parents' educational level on the development of executive functions. **Developmental neuropsychology**, v. 28, n. 1, p. 539-560, 2005.
- ARSANJANI, Jamal Jokar et al. An exploration of future patterns of the contributions to OpenStreetMap and development of a Contribution Index. **Transactions in GIS**, v. 19, n. 6, p. 896-914, 2015.
- ANTONIOU, Vyron; SKOPELITI, Andriani. O Impacto da Contribuição do Microambiente na Qualidade dos Dados: O Caso do OSM. In: Foody, G et al (eds.), **Maapeamento e Sensor Cidadão**. Londres: Ubiquity Press, 2017.
- ANTONIOU, V.; MORLEY, J.; HAKLAY, M. P. The role of experience, knowledge and mapping practices in contributing to VGI quality. **Future Internet**, v. 8, n. 2, p. 14-29, 2016.
- BARRON, Christopher; NEIS, Pascal; ZIPF, Alexander. A comprehensive framework for intrinsic OpenStreetMap quality analysis. **Transactions in GIS**, v. 18, n. 6, p. 877-895, 2014..
- BIELECKA, Elżbieta. GEOGRAPHICAL DATA SETS FITNESS OF USE EVALUATION. **Geodetski vestnik**, v. 59, n. 2, 2015.
- BILDIRICI, I. O.; ULUGTEKIN, N. N. web mapping with Google Maps Mashups: overlaying geodata. In: **A Special Joint Symposium of ISPRS Technical Commission IV & AutoCarto in Conjunction with ASPRS/CaGIS**. 2010. p. 15-19.
- BEARDEN, M. J. The National Map Corps: The USGS'Volunteer Geographic Information Program. In: **VGI Workshop (NCGIA)**. Santa Barbara: NCGIA, 2007.
- BÉGIN, D.; DEVILLERS, R.; ROCHE, S. Assessing Volunteered Geographic Information (vgi) Quality Based on CONTRIBUTORS'Mapping Behaviours. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 40, p. 149-154, 2013.

BORGES, Karla AV; DAVIS JR, Clodoveu A.; LAENDER, Alberto HF. Modelagem conceitual de dados geográficos. **CASANOVA, et. al. Banco de Dados Geográfico. MundoGEO: Curitiba**, p. 83-136, 2005.

BRASIL. Ministério do Planejamento, Orçamento e Gestão –Comissão Nacional de Cartografia (CONCAR). **Plano de Ação para Implantação da Infraestrutura Nacional de Dados Espaciais –INDE**, 2010.

BRAVO, J. V. M.; SLUTER, C. R. The Geospatial Data Quality subject in the Volunteered Geographic Information era. **BCG - Boletim de Ciências Geodésicas**, v. 21, n. 1, p. 56–73, 2015.

BROVELLI, Maria Antonia; MINGHINI, MARco; MOLINARI, Monia Elisa. An automated GRASS-based procedure to assess the geometrical accuracy of the OpenStreetMap Paris road network. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 41, p. 919-925, 2016.

BROVELLI, Maria Antonia; ZAMBONI, Giorgio. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. **ISPRS International Journal of Geo-Information**, v. 7, n. 8, p. 289, 2018.

BRÜCKNER, Josephine et al. Assessing shop completeness in OpenStreetMap for two federal states in Germany. **AGILE: GIScience Series**, v. 2, p. 20, 2021.

BUDHATHOKI, Nama R.; HAYTHORNTHWAITE, Caroline. Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. **American Behavioral Scientist**, v. 57, n. 5, p. 548-575, 2013.

CAMBOIM, Silvana et al. The cell size issue in OpenStreetMap data quality parameter analyses: an interpolation-based approach. **Editors**, p. 42, 2022.

CAMBOIM, Silvana Philippi; BRAVO, João Vitor Meza; SLUTER, Claudia Robbi. An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. **ISPRS International Journal of Geo-Information**, v. 4, n. 3, p. 1366-1388, 2015.

CÂMARA, Gilberto et al. Introdução à ciência da geoinformação. **São José dos Campos: INPE**, v. 345, 2001.

CASTILLO, Gualda R. et al. Associations between parental educational-occupational levels and cognitive performance in Spanish adolescents: the AVENA study. **Psicothema**, 2011.

CHEN, J.; PAVLIDIS, T. Urban information systems: From theory to applications. Boca Raton: **CRC Press**, 2013.

CIEPŁUCH, Błażej et al. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In: **Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences 20-23rd July 2010**. University of Leicester, 2010. p. 337.

CORMODE, Graham; KRISHNAMURTHY, Balachander. Key differences between web 1.0 and web 2.0. **First Monday**, 2008.

COSTA, Eduarda Marques. Cidades médias-Contributos para a sua definição. **Finisterra**, v. 37, n. 74, 2002.

CRUZ, D. T.; SANTOS, A. P. Controle de qualidade posicional do sistema rodoviário do Openstreetmap na região central De Viçosa-MG. **VI Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**, 2016.

DAVIS JR, Clodoveu A.; FONSECA, Frederico T.; CÂMARA, Gilberto. Infraestruturas de dados espaciais na integração entre ciência e comunidades para promover a sustentabilidade ambiental. In: **Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais**. 2009.

DEVILLERS, R.; JEANSOULIN, R. Spatial Data Quality: Concepts. In R. Devillers and R. Jeansoulin, eds. **Fundamentals of Spatial Data Quality**. London: **ISTE Ltd**. Ch. 2, p.31-42, 2006.

DE OLIVEIRA, Daniel Ferreira et al. Avaliação intrínseca dos dados do Openstreetmap e os aspectos espaço-temporais no estudo de caso do rompimento da barragem de Brumadinho: Intrinsic evaluation of Openstreetmap data and space-time aspects in the case study of Brumadinho's Dam breaking. **Geosciences= Geociências**, v. 41, n. 3, p. 797-810, 2022.

DE SOUZA, T. M.; SCHMIDT, M. A. R. • Avaliação da acurácia posicional de dados colaborativos do OpenStreetMap: Eixos viários de bairro no município de Uberlândia (MG): Evaluation of the positional accuracy of OpenStreetMap collaborative data: neighborhood road axes in the municipality of Uberlândia (MG). **Geosciences= Geociências**, v. 42, n. 2, p. 281-290, 2023.

DI GREGORIO, A. **Land cover classification system: classification concepts and user manual: LCCS**. Food & Agriculture Org., 2005.

DOS RAMOS, Mariana Filipa Castilho Furtado et al. Ökosiedlung-A importância dos espaços comuns. 2016.

DROR, Talia; DOYTHER, Yerach; DALYOT, Sagi. Investigating the use of historical node location data as a source to improve OpenStreetMap position quality. **Open Source Geospatial Science for Urban Studies: The Value of Open Geospatial Data**, p. 55-73, 2021.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-CQDG - Especificação Técnica para Controle de Qualidade de Dados Geoespaciais**, v. 1, 2016.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-EDGV F-Ter. – Especificação Técnica para a Estruturação dos Dados Geoespaciais Vetoriais - 1ª parte**, 2015.

EGENHOFER, Max J.; FRANZOSA, Robert D. Point-set topological spatial relations. **International Journal of Geographical Information System**, v. 5, n. 2, p. 161-174, 1991.

ELIAS, E. N. N.; DE OLIVEIRA FERNANDES, V. Qualidade dos dados geoespaciais do openstreetmap para os indicadores de acurácia posicional, acurácia temática e completude. **Geografia (Londrina)**, v. 30, n. 2, p. 255-275, 2021.

ELIAS, E. N. N.; SANTOS, D. L.; CAMBOIM, S. P.; DELAZARI, L. S.; SCHMIDT, M. A. R. Avaliação e representação da qualidade de dados geoespaciais a partir do desenvolvimento de ferramentas em ambiente SIG. In: **XI Colóquio Brasileiro de Ciências Geodésicas**, 2020.

ELIAS, E. N. N. Qualidade de dados geoespaciais em plataforma de mapeamento colaborativo. 2021.

ELIAS, E. N. N. et al. Discovering spatiotemporal patterns on data quality assessment in collaborative mapping: A preliminary study in an area of Brazil. **Abstracts of the ICA**, v. 3, p. 73, 2021.

ELIAS, E. N. N. et al. Exploring spatio-temporal patterns of OpenStreetMap (OSM) contributions in heterogeneous urban areas. **Boletim de Ciências Geodésicas**, v. 29, p. e2023005, 2023.

ELIAS, E. N. N. et al. The quality of OpenStreetMap in a large metropolis in northeast Brazil: Preliminary assessment of geospatial data for road axes. **Boletim de Ciências Geodésicas**, v. 26, n. 03, p. e2020012, 2020.

ELIAS, E. N. N. et al. QPEC: QGIS Toolkit for Evaluating Geospatial Data Positional Accuracy according to the Brazilian Cartographic Accuracy Standard. **Anuário do Instituto de Geociências**, v. 46, 2023.

ELIAS, E. N. N. **Determinação e avaliação de indicadores espaço-temporais da qualidade de dados no mapeamento colaborativo**. Curitiba, 2022. Tese (Doutorado em Ciências Geodésicas) – Universidade Federal do Paraná, Paraná.

ELWOOD, Sarah; LESZCZYNSKI, Agnieszka. Privacy, reconsidered: New representations, data practices, and the geoweb. **Geoforum**, v. 42, n. 1, p. 6-15, 2011.

ELWOOD, S.; GOODCHILD, M. F.; SUI, D. Z. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. **Annals of the Association of American geographers**, 102, (3), pp. 571-590, 2012.

ESTES, John E.; MOONEYHAN, D. Wayne. Of maps and myths. **Photogrammetric engineering and remote sensing**, v. 60, n. 5, p. 517-524, 1994.

FERSTER, C.; FISCHER, J.; MANAUGH, K.; NELSON, T.; WINTERS, M. Using OpenStreetMap to inventory bicycle infrastructure: a comparison with open data from cities. **International Journal of Sustainable Transportation**, 14(1), 64-73, 2020.

FLANAGIN, A. J.; METZGER, M. J. The credibility of volunteered geographic information. **GeoJournal**, 137-148, 2008.

FONTE, C. C.; BAPTISTA, J. V.; BARRA, A.; MARTINS, F. The impact of population density and access to the internet on the quality of volunteered geographic information. **International Journal of Geographical Information Science**, v. 31, n. 1, p. 98-123, 2017.

GANAPATI, S. Uses of Public Participation Geographic Information Systems Applications in E-Government. **Public Administration Review**, v. 71, n. 3, p. 425-434, 2011.

GARCIA, A. V.; HILLESHEIM, J. Pobreza e desigualdades educacionais: uma análise com base nos Planos Nacionais de Educação e nos Planos Plurianuais Federais. **Educar em revista**, p. 131-147, 2017.

GIRRES, J. F.; TOUYA, G. Quality assessment of the French OpenStreetMap dataset. **Transactions in GIS**, 14(4), 435-459, 2010.

GOODCHILD, Michael F. Citizens as sensors: the world of volunteered geography. **GeoJournal**, v. 69, p. 211-221, 2007.

GOODCHILD, Michael F.; GLENNON, J. Alan. Crowdsourcing geographic information for disaster response: a research frontier. **International Journal of Digital Earth**, v. 3, n. 3, p. 231-241, 2010.

GOODCHILD, Michael F.; LI, Linna. Assuring the quality of volunteered geographic information. **Spatial statistics**, v. 1, p. 110-120, 2012.

GOETZ, Marcus. Using crowdsourced indoor geodata for the creation of a three-dimensional indoor routing web application. **Future Internet**, v. 4, n. 2, p. 575-591, 2012.

GOLDANI, Ana Maria. Família, gênero e políticas: famílias brasileiras nos anos 90 e seus desafios como fator de proteção. **Revista Brasileira de Estudos de População**, v. 19, n. 1, p. 29-48, 2002.

GRIFFIN, A. L.; WHITE, T.; FISH, C. TOMIO, B.; HUANG, H.; SLUTER, C. R.; BRAVO, J. V. M.; FABRIKANT, S.; BLEISCH, S.; YAMADA, M. PICANÇO JR, P. Designing across map use contexts: A research agenda. **International Journal of Cartography**, Special Issue, vol. 3, 2017. pp. 1-25

GRINBERGER, A. Yair et al. An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. **Transactions in GIS**, v. 25, n. 2, p. 622-641, 2021.

GRÖCHENIG, Simon; BRUNAUER, Richard; REHRL, Karl. Digging into the history of VGI data-sets: results from a worldwide study on OpenStreetMap mapping activity. **Journal of Location Based Services**, v. 8, n. 3, p. 198-210, 2014.

HAKLAY, Mordechai. How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. **Environment & Planning** (under review), 2008.

HAKLAY, Mordechai. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. **Environment and planning B: Planning and design**, v. 37, n. 4, p. 682-703, 2010.

HAKLAY, M.; WEBER, P. OpenStreetMap: User-generated street maps. **IEEE Pervasive Computing**, v. 7, n. 4, p. 12-18, 2008.

HECHT, R.; KUNZE, C.; HAHMANN, S. Measuring completeness of building footprints in OpenStreetMap over space and time. **ISPRS International Journal of GeoInformation**, 2(4), 1066-1091, 2013.

HOWE, J. The rise of crowdsourcing. **Wired magazine**, 14(6), 1-4, 2006.

INE - Instituto Nacional de Estadística. Acesso: 11 de outubro de 2023. Disponível em: <https://www.ine.es/>

INSEE - Institut national de la statistique et des études économiques. Acesso: 11 de outubro de 2023. Disponível em: <https://www.insee.fr/fr/accueil>

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Avaliação da qualidade de dados geoespaciais**. Rio de Janeiro: IBGE, Manuais Técnicos em Geociências n. 13, 1ª ed. 2017.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Avaliação da qualidade de dados geoespaciais**. Rio de Janeiro: IBGE, Manuais Técnicos em Geociências n. 13, 2ª ed. 2019.

ISO 19113. **Geographic information – Quality principles**. International Organization for Standardization, 2002.

ISO 19114. Geographic information – Quality **evaluation procedures**. International Organization for Standardization, 2003.

ISO:19157. **Geographic Information - Data Quality**. International Organization for Standardization, 2013.

JACOBSEN, Geise M. et al. Qual é a participação de fatores socioeconômicos na inteligência de crianças?. **Revista Neuropsicologia Latinoamericana**, v. 5, n. 4, p. 32-39, 2013.

JAMIESON, J. Many (to platform) to many: web 2.0 application infrastructures. **First Monday**, vol. 21, n. 6, 2016.

JARRET, K. Interactivity is evil! A critical investigation of web 2.0. **First Monday**, vol. 13, n. 3, 2008.

JOHNSON, Peter A.; SIEBER, Renee E. Motivations driving government adoption of the Geoweb. **GeoJournal**, v. 77, p. 667-680, 2012.

JOHNSON, Peter A.; SIEBER, Renee E. Situating the adoption of VGI by government. In: **Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice**. Dordrecht: Springer Netherlands, 2012. p. 65-81.

KRAAK, M. J. (2004). The role of accuracy in operational GIS. **International Society for Photogrammetry and Remote Sensing**. In Accuracy, p. 127-138, 2004

LE GUILCHER, Arnaud; OLTEANU-RAIMOND, A.-M.; BALDE, Mamadou Bailo. Analysis of Massive Imports of Open Data in Openstreetmap Database: a Study Case for France. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 4, p. 99-106, 2022.

LIU, X.; ZHANG, Y.; WANG, L. Evaluating the accuracy of VGI data in metropolitan areas using RMSE and PECPCD. **Journal of Urban and Regional Information Systems**, v. 45, n. 2, p. 121-140, 2020.

LONGLEY, Paul A. et al. **Sistemas e ciência da informação geográfica**. Bookman Editora, 2009.

LÓPEZ-DE-IPÍÑA, Diego; LORIDO, Tania; LÓPEZ, Unai. Indoor navigation and product recognition for blind people assisted shopping. In: **Ambient Assisted Living: Third International Workshop, IWAAL 2011, Held at IWANN 2011, Torremolinos-Málaga, Spain, June 8-10, 2011. Proceedings 3**. Springer Berlin Heidelberg, 2011. p. 33-40.

LOSHIN, D. Monitoring Data Quality Performance: Using Data Quality Metrics. White Paper, Informatica. 2010.

MACHADO, A. A. **Compatibilização semântica entre o modelo de dados do OpenStreetMap e a Especificação Técnica Para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV)**. Tese (Doutorado em Ciências Geodésicas). Universidade Federal do Paraná, Curitiba, 2020.

MACHADO, Adriana Alexandria; CAMBOIM, Silvana Philippi. Mapeamento colaborativo como fonte de dados para o planejamento urbano: desafios e potencialidades. **urbe. Revista Brasileira de Gestão Urbana**, v. 11, 2019.

MAIA, D. S. Pequenas cidades: como defini-las. **IX Simpósio Nacional de Geografia Urbana-SIMPURB**, p. 1-20, 2005.

MARMOT, Michael. A lacuna na saúde: o desafio de um mundo desigual. **The Lancet**, v. 10011, pág. 2442-2444, 2015.

MARTINI, A.; KUPER, Paul Vincent; BREUNIG, Martin. Database-supported change analysis and quality evaluation of OpenStreetMap Data. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 4, p. 535-541, 2019.

MILLER, Harvey J. Beyond sharing: cultivating cooperative transportation systems through geographic information science. **Journal of Transport Geography**, v. 31, p. 296-308, 2013.

MONDZECH, Juliane; SESTER, Monika. Quality analysis of OpenStreetMap data based on application needs. **Cartographica: The International Journal for Geographic Information and Geovisualization**, v. 46, n. 2, p. 115-125, 2011.

MOONEY, Peter; CORCORAN, Pdraig. Characteristics of heavily edited objects in OpenStreetMap. **Future Internet**, v. 4, n. 1, p. 285-305, 2012.

MOONEY, Peter; CORCORAN, Pdraig; WINSTANLEY, Adam. A study of data representation of natural features in openstreetmap. In: **Proceedings of GIScience**. 2010. p. 150-156.

MOONEY, Peter; CORCORAN, Pdraig. Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. **Transactions in GIS**, v. 18, n. 5, p. 633-659, 2014.

MUTTAQIEN, Bani Idham; OSTERMANN, Frank O.; LEMMENS, Rob LG. Modeling aggregated expertise of user contributions to assess the credibility of OpenStreetMap features. **Transactions in GIS**, v. 22, n. 3, p. 823-841, 2018.

NASIRI, Afsaneh et al. Improving the quality of citizen contributed geodata through their historical contributions: The case of the road network in OpenStreetMap. **ISPRS International Journal of Geo-Information**, v. 7, n. 7, p. 253, 2018.

NEIS, Pascal; ZIELSTRA, Dennis. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. **Future internet**, v. 6, n. 1, p. 76-106, 2014.

NEIS, Pascal; ZIELSTRA, Dennis; ZIPF, Alexander. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. **Future Internet**, v. 4, n. 1, p. 1-21, 2011.

NEIS, Pascal; ZIPF, Alexander. Analyzing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. **ISPRS International Journal of Geo-Information**, v. 1, n. 2, p. 146-165, 2012.

OLIVEIRA, D. F. et al. Intrinsic evaluation of Openstreetmap data and space-time aspects in the case study of Brumadinho's Dam breaking. **Geosciences**, v. 41, n. 3, p. 797-810, 2022.

OSM. OpenStreetMap Foundation. Disponível em: <http://wiki.osmfoundation.org/wiki/>. Acesso em: 20/10/2023.

OSM. OpenStreetMap stats. Disponível em: https://www.openstreetmap.org/stats/data_stats.html. Acesso em: 20/10/2023.

OXLEY, A. web 2.0 Applications of Geographic and Geospatial Information. **Bulletin of the American Society for Information Science and Technology**, vol.35, n. 4, 2009.

O'REILLY, T. What is web 2.0: Design patterns and business models for the next generation of software. **Communications & Strategies**, vol.65, n.1, 2007.

PACHECO, M.; OLIVEIRA, R. Acurácia e Completude dos Dados do OpenStreetMap para Sistemas de Transporte Urbano. **Anais do Congresso Brasileiro de Cartografia**, 2020.

PAIVA, C. A. **Inferência da qualidade de dados VGI a partir de métricas intrínsecas**. Curitiba, 2021. 159 f. Tese (Doutorado em Ciências Geodésicas) – Universidade Federal do Paraná, Paraná.

PAIVA, Caio dos Anjos; CAMBOIM, Silvana Philippi. Inference of positional accuracy of collaborative data from intrinsic parameters. **Transactions in GIS**, v. 26, n. 4, p. 1898-1913, 2022.

PAIVA, Caio dos Anjos; CAMBOIM, Silvana Philippi. A Dinâmica de Colaborações OpenStreetMap e sua Relação com as Atividades de Uso e Ocupação do Solo: um Estudo Segundo Zoneamento de Curitiba. **Rev. Bras. Cartogr**, v. 73, n. 1, 2021.

PAPAPESIOS, Nikolaos et al. Exploring the use of crowdsourced geographic information in defence: challenges and opportunities. **Journal of Geographical Systems**, v. 21, p. 133-160, 2019.

PERKINS, C. Cultures of Map Use. **The Cartographic Journal**, vol. 45, n. 2, 2008. pp. 150-158.

PETERLINI, Tayani Bellucci et al. Identifying Temporal Patterns of Contributions to OpenStreetMap in Brazilian Cities: The Role of Cell Size. **Abstracts of the ICA**, v. 6, p. 197, 2023.

PITOMBO, Cira Souza et al. **Estudos de relações entre variáveis socioeconômicas, de uso do solo, participação em atividades e padrões de viagens encadeadas urbanas**. 2007. Tese de Doutorado. Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

PROFERES, N. web 2.0 user knowledge and the limits of individual and collective power. **First Monday**, vol. 21, n. 6, 2016.

RAYMOND, E. S. **The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary**. O'Reilly Media, 1997.

REDMAN, Thomas C. **Data quality for the information age**. Artech House, Inc., 1997.

ROCHA, Marisa PC. Desenvolvimento de referencial teórico para um sistema de informações gerenciais (SIG) para parlamentares e assessores na Câmara Legislativa do Distrito Federal: em busca de um modelo conceitual. **Ciência da Informação**, v. 32, p. 80-88, 2003.

ROUSE, L. J.; BERGERON, S. J.; HARRIS, T. M. Participating in the Geospatial web: Collaborative Mapping, Social Networks and Participatory GIS. IN: SCHARI, A. (ed.); TOCHTERMANN, K. (ed.). **The Geospatial web: How Geobrowsers, Social Software and the web 2.0 are shaping the Network Society**. Advanced Information and Knowledge Processing, Springer, 2007.

ROSA, Roberto. Geotecnologias na geografia aplicada. **Revista do Departamento de Geografia**, v. 16, p. 81-90, 2005.

SANTOS, A., BERNARDES, A. Avaliação da Qualidade dos Dados do OpenStreetMap no Contexto de Planejamento Urbano em Campinas, SP. **Revista Brasileira de Cartografia**, 2020.

SANTOS, A., et al. Qualidade Posicional dos Dados do OpenStreetMap na Região de Viçosa, MG. **Geografia: Revista da Universidade Estadual de Londrina**, 2017.

SEE, L.; MOONEY, P.; FOODY, G.; BASTIN, L.; COMBER, A.; ESTIMA, J.; FRITZ, S.; KERLE, N.; JIANG, B.; LAAKSO, M.; LIU, H.; MILCINSKI, G.; NIKSIC, M.; PAINHO, M.; PODOR, A.; OLTEANU-RAIMOND, A.; RUTZINGER, M. Crowdsourcing, Citizen Science or Volunteered Geographic Information? The current state of Crowdsourced Geographic Information. **ISPRS International Journal of Geo-Information**, vol. 5, n. 55, 2016.

SCHARL, A., Tochtermann, K., Eds.; **The Geospatial web: How Geobrowsers, Social Software and the web 2.0 are Shaping the Network Society**; Springer: London, UK, 2007.

SEHRA, Sukhjit Singh; SINGH, Jaiteg; RAI, Hardeep Singh. A systematic study of OpenStreetMap data quality assessment. In: 2014 11th **International Conference on information technology: new generations**. IEEE, 2014. p. 377-381.

SEHRA, Sukhjit Singh; SINGH, Jaiteg; RAI, Hardeep Singh. Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. **Future Internet**, v. 9, n. 2, p. 15, 2017.

SENARATNE, Hansi et al. A review of volunteered geographic information quality assessment methods. **International Journal of Geographical Information Science**, v. 31, n. 1, p. 139-167, 2017.

SENARATNE, H.; MOBASHERI, A.; ZIPF, A. A geographic approach to combining social media and authoritative data towards identifying useful information for disaster management. **International Journal of Geographical Information Science**, v. 31, n. 2, p. 219-242, 2017.

SETO, Toshikazu; KANASUGI, Hiroshi; NISHIMURA, Yuichiro. Quality verification of volunteered geographic information using osm notes data in a global context. **ISPRS International Journal of Geo-Information**, v. 9, n. 6, p. 372, 2020.

SILVA, L.S.L.; CAMBOIM, S.P. Authoritative cartography in Brazil and collaborative mapping platforms: challenges and proposals for data integration. **Bulletin of Geodetic Sciences**. 27(spe): e2021003, 2021.

SKARLATIDOU, Artemis; HAKLAY, Muki. Public web mapping: preliminary usability evaluation, 2006.

SLUTER, Claudia Robbi; CAMBOIM, Silvana Philippi. The National Topographic Mapping as an Indispensable Database for a Brazilian National Spatial Data Infrastructure (NSDI). In: Proceedings of the **XXIV International Cartographic Conference, Santiago de Chile**. Chile: NSDI, 2009.

SLUTER, Claudia Robbi et al. Uma proposta de simbologia cartográfica topográfica para mapas de grande escala de áreas urbanas no Brasil. **A Revista Cartográfica**, v. 55, n. 4, pág. 362-377, 2018.

SOARES, HERICLIS. Informação geográfica voluntária aplicada na infraestrutura urbana do município de Raul Soares-MG. 2018.

STEINMANN, Renate et al. Gender dimensions in UGC and VGI: A desk-based study. 2013.

TARDIN, et al. Interoperabilidade dos Dados do OpenStreetMap em Projetos de Planejamento Urbano. **Revista Geociências**, 2019.

TEIXEIRA, M. DE S.; SCHMIDT, M. A. R. Evaluation of the positional accuracy of OpenStreetMap collaborative data: neighborhood road axes in the municipality of Uberlândia (MG). **Geosciences**, v. 42, n. 2, p. 281-290, 2023.

TIAN, Y.; ZHOU, Q.; FU, X. An Analysis of the Evolution, Completeness and Spatial Patterns of OpenStreetMap Building Data in China. **ISPRS International Journal of GeoInformation**, 8(1), 35, 2019.

TORVALDS, Linus. **Just for fun: The story of an accidental revolutionary**. Harper Audio, 2001.

TOUYA, G.; ANTONIOU, V.; OLTEANU-RAIMOND, A. M.; VAN DAMME, M. D. Assessing crowdsourced POI quality: Combining methods based on reference data, history, and spatial relations. **ISPRS International Journal of GeoInformation**, 6(3), 80. (2017a).

TOUYA, G.; ANTONIOU, V.; CHRISTOPHE, S.; SKOPELITI, A. Production of Topographic Maps with VGI: Quality Management and Automation. In Mapping and the Citizen Sensor; FOODY, G.; SEE, L.; FRITZ, S.; MOONEY, P.; OLTEANU-

TSOU, M. H.; LEITNER, M. Understanding and mitigating the impacts of low data quality on urban analyses. **Annals of GIS**, v. 21, n. 3, p. 209-218, 2015.

RAIMOND, A.-M.; FONTE, C.C.; ANTONIOU, V., Eds.; **Ubiquity Press**: London, UK, 2017; pp. 137–164, 2017.

VIANA, Cláudia M.; ENCALADA, Luis; ROCHA, Jorge. The value of OpenStreetMap historical contributions as a source of sampling data for multi-temporal land use/cover maps. **ISPRS International Journal of Geo-Information**, v. 8, n. 3, p. 116, 2019.

VOZENILEK, Vit. Artificial intelligence and GIS: mutual meeting and passing. In: 2009 **International conference on intelligent networking and collaborative systems**. IEEE, 2009. p. 279-284.

WAGLE, Nimisha; ACHARYA, Tri Dev. Past and present practices of topographic base map database update in Nepal. **ISPRS International Journal of Geo-Information**, v. 9, n. 6, p. 397, 2020.

WEST, Andrew G. et al. Trust in collaborative web applications. **Future Generation Computer Systems**, v. 28, n. 8, p. 1238-1251, 2012.

WESTERHOLT, Rene et al. Abundant topological outliers in social media data and their effect on spatial analysis. **PloS one**, v. 11, n. 9, p. e0162360, 2016.

WITT, Raphael; LOOS, Lukas; ZIPF, Alexander. Analysing the impact of large data imports in OpenStreetMap. **ISPRS International Journal of Geo-Information**, v. 10, n. 8, p. 528, 2021.

YAN, Y.; FENG, C. C.; HUANG, W.; FAN, H.; WANG, Y. C.; ZIPF, A. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. **International Journal of Geographical Information Science**, 1-27, 2020.

YANG, C.; LIU, K.; CHEN, J. Analyzing the quality of volunteered geographic information in rural areas using RMSE and PECPCD. **Geographical Information Science**, v. 34, n. 6, p. 561-579, 2018.

ZHANG, Hongyu; MALCZEWSKI, Jacek. Accuracy evaluation of the Canadian OpenStreetMap road networks. **International Journal of Geospatial and Environmental Research**, v. 5, n. 2, 2017.

ZHOU, Qi et al. Assessing OSM building completeness for almost 13,000 cities globally. **International Journal of Digital Earth**, v. 15, n. 1, p. 2400-2421, 2022.

ZHOU, Q. Exploring the relationship between density and completeness of urban building data in OpenStreetMap for quality estimation. **International Journal of Geographical Information Science**, 32(2), pp. 257-281, 2018.

ZIELSTRA, Dennis; ZIPF, Alexander. Quantitative studies on the data quality of OpenStreetMap in Germany. In: **Proceedings of GIScience**. 2010.

ZIELSTRA, D.; HOCHMAIR, H. H. Comparative study of pedestrian accessibility to transit stations using free and proprietary network data. Transportation Research Record: **Journal of the Transportation Research Board**, v. 2217, p. 145-152, 2011.