

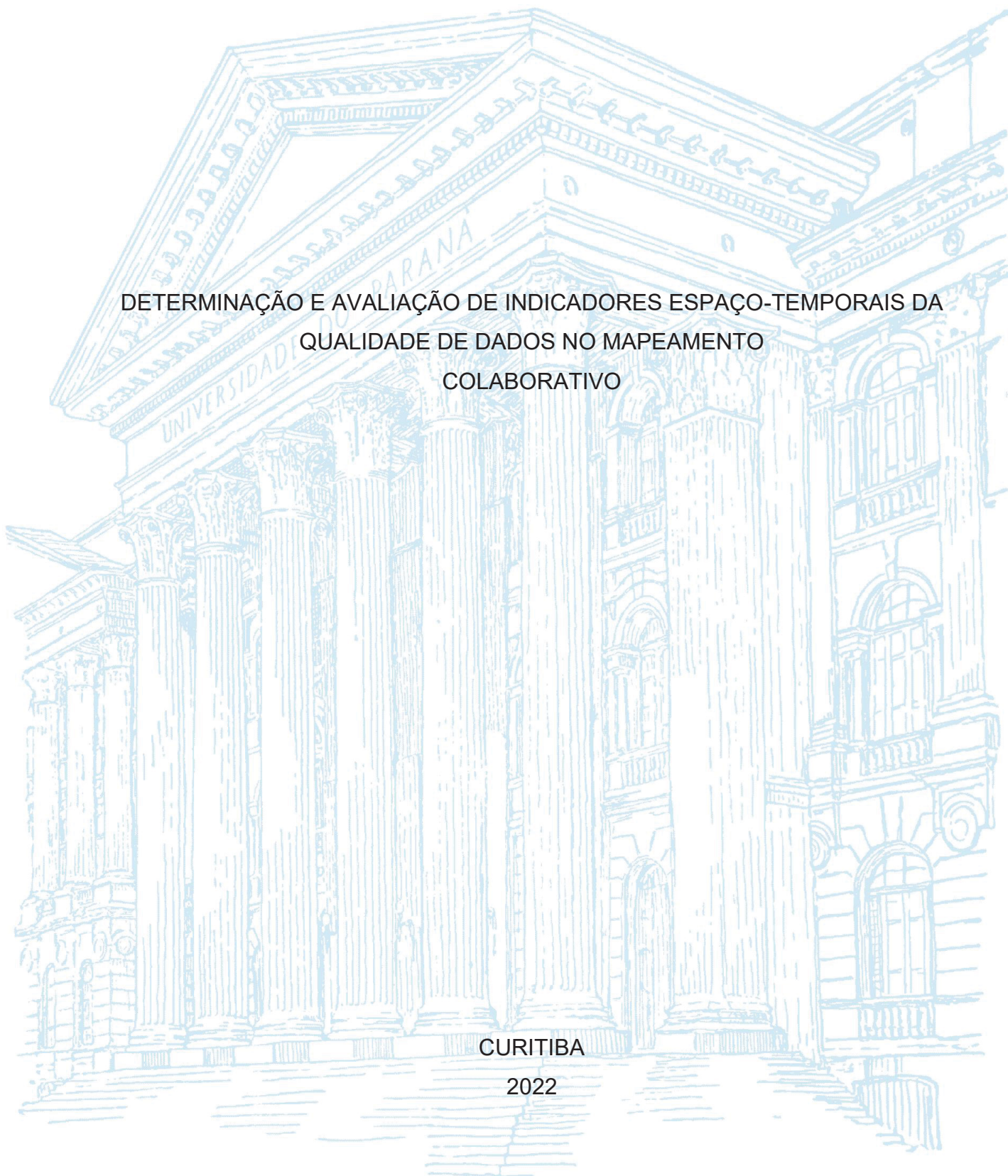
UNIVERSIDADE FEDERAL DO PARANÁ

ELIAS NASR NAIM ELIAS

DETERMINAÇÃO E AVALIAÇÃO DE INDICADORES ESPAÇO-TEMPORAIS DA
QUALIDADE DE DADOS NO MAPEAMENTO
COLABORATIVO

CURITIBA

2022



ELIAS NASR NAIM ELIAS

DETERMINAÇÃO E AVALIAÇÃO DE INDICADORES ESPAÇO-TEMPORAIS DA
QUALIDADE DE DADOS NO MAPEAMENTO
COLABORATIVO

Tese apresentada como requisito parcial à obtenção do título de Doutor em Ciências Geodésicas no curso de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra da Universidade Federal do Paraná – UFPR.

Orientador: Prof. Dr. Marcio Augusto Reolon Schmidt.

Co-Orientadora: Prof^a. Dr^a. Silvana Philippi Camboim.

CURITIBA

2022

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

Elias, Elias Nasr Naim

Determinação e avaliação de indicadores espaço-temporais da qualidade de dados no mapeamento colaborativo / Elias Nasr Naim Elias. – Curitiba, 2022.

1 recurso on-line : PDF.

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

Orientador: Marcio Augusto Reolon Schmidt

Coorientador: Silvana Philippi Camboim

1. Dados Geospaciais. 2. Análise de regressão logística. 3. Banco de dados temporais. 4. OpenStreetMap. I. Universidade Federal do Paraná. II. Programa de Pós-Graduação em Ciências Geodésicas. III. Schmidt, Marcio Augusto Reolon. IV. Camboim, Silvana Philippi. V. Título.



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 tese de Doutorado de **ELIAS NASR NAIM ELIAS** intitulada: **DETERMINAÇÃO E AVALIAÇÃO DE INDICADORES ESPAÇO-TEMPORAIS DA QUALIDADE DE DADOS NO MAPEAMENTO COLABORATIVO**, sob orientação do Prof. Dr. MARCIO AUGUSTO REOLON SCHMIDT, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor 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, 28 de Novembro de 2022.

Assinatura Eletrônica
28/11/2022 16:38:44.0

MARCIO AUGUSTO REOLON SCHMIDT
Presidente da Banca Examinadora

Assinatura Eletrônica
28/11/2022 16:41:34.0

VIVIAN DE OLIVEIRA FERNANDES
Avaliador Externo (UNIVERSIDADE FEDERAL DA BAHIA)

Assinatura Eletrônica
28/11/2022 16:39:19.0

ANDREA LOPES IESCHECK
Avaliador Externo (UNIVERSIDADE FEDERAL DO RIO GRANDE DO
SUL)

Assinatura Eletrônica
28/11/2022 16:38:40.0

CAIO DOS ANJOS PAIVA
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Dedico este trabalho a todos que acreditam que a educação transforma a vida das pessoas e aos que lutam em prol da democracia, políticas públicas e pelo ensino público de qualidade para todos.

AGRADECIMENTOS

Alcançar essa etapa não representa só o fim do ciclo do doutorado, mas a realização de um sonho que se iniciou desde antes da graduação. Então primeiramente agradeço a todas as políticas de ações afirmativas que me deram subsídios para viver a universidade pública ao longo dos últimos 12 anos, desde os auxílios, recursos, restaurante universitário, bolsas de iniciação científica, extensão, mestrado até o doutorado. Sem todo o investimento não conseguiria chegar até aqui.

Aos meus pais Nasr e Marlene por todo suporte e apoio durante toda a minha vida. Em especial a minha mãe que transformou os meus sonhos nos seus próprios sonhos e vibrou em todos os momentos. Você faz parte disso, mãe. Essa vitória é nossa!

À minha tia Conceição pelas palavras, conselhos, orações e por sempre acreditar em mim e estar presente. À minha madrinha Janete e minhas tias Rebecca e Marwa que sempre estiveram presentes na minha vida e à prima Fabiana pela aproximação ao longo desses anos e por sempre me receber de braços abertos.

À Adjaci pela amizade de sempre, preocupação e me incentivar a chegar mais longe. Um dia um amigo que não está mais entre nós, por razões adversas, nos apresentou e isso foi um melhores presentes que ganhei.

À minha amiga irmã de uma vida inteira Luana por mais esse reencontro em Curitiba e por me fazer sentir parte da sua família. Obrigado por essa conexão que já transcende mais de 20 anos.

À UFPR e a todos os professores e colegas do PPGCG por todos os ensinamentos, auxílios e suporte desde a minha chegada em Curitiba e ingresso no programa. Ao Erickson e à Mônica por toda ajuda para resolver questões burocráticas.

À Universidade Federal da Bahia (UFBA) por todo suporte na minha formação pessoal e profissional entre a graduação e o mestrado. Levo comigo as melhores lembranças.

À CAPES pelo fomento que me permitiu cursar o doutorado.

Ao meu orientador Prof. Dr. Marcio Schmidt pela amizade, incentivo, paciência, confiança e por sempre acreditar no meu potencial. À minha coorientadora Prof. Dra. Silvana Camboim por me receber de braços abertos na UFPR, empolgação, incentivo e por acreditar na minha capacidade de realizar análises tão desafiadoras. Faltam palavras para agradecer toda a ajuda de ambos.

Aos membros da banca examinadora de qualificação e tese Prof. Dra. Andrea Iescheck, Prof. Dr. Caio Paiva, Prof. Dra. Julia Strauch, Prof. Dra. Luciene Delazari e Prof. Dra. Vivian Fernandes, pelas valiosas contribuições nas diferentes etapas desse trabalho.

À Prof. Dra. Vivian Fernandes por sempre acreditar, muitas vezes mais do que eu, que esse momento chegaria e por todo incentivo desde a IC.

Aos amigos que fiz no PPGCG: Everton, Kauê, Niédja, Rubens, Samo e Sergin pelos conselhos, paciência, conversas e por sempre estarem presentes e tornarem minha caminhada mais leve. Vocês já estão fazendo muita falta! Ao Samir e à Andrea, essenciais nesse último ano de doutorado para que tudo se concretizasse. Ao Fabricio pela parceria nas publicações e ao Raphael e Cynthia pela disponibilidade pra tirar minhas dúvidas em programação. Ao Fabiano, Felipe, Livia, Tulio e Vinicius por toda ajuda e bons momentos, principalmente na época das disciplinas.

Aos colegas do LabGeoLivre, LabCarto, LFSR e LARAS pelos bons momentos, cafés, saídas e conversas.

Ao grupo dos Mapeadores Livres da UFPR.

Aos amigos de Salvador: Ediana, Elaine, Gabi, Lore, Marcos, Nicole e Ricardo que mesmo distantes fisicamente parte desses anos, sempre se mantiveram presentes e foram fundamentais em vários momentos.

Por fim e mais importante, agradeço à Deus por ter sido o meu sustento nos momentos mais difíceis e por me dar forças e coragem para alcançar os meus objetivos.

“The world isn’t split into good people and Death Eaters. We’ve all got both light and dark inside us. What matters is the part we choose to act on.” – Sirius Black.

Harry Potter and the Order of the Phoenix
David Yates (2007), baseado em Rowling (2003)

RESUMO

Questões relacionadas com a desatualização do mapeamento oficial são recorrentes em diferentes localidades do Brasil e do mundo, principalmente devido ao custo-benefício para a sua produção e manutenção. Neste contexto, pesquisas têm direcionado esforços em avaliar o potencial de informações oriundas de plataformas de mapeamento colaborativo, no âmbito de estabelecer o seu potencial de integração e a determinação da qualidade. Abordagens tradicionais se baseiam em comparações em relação a bases oficiais existentes, todavia, esta nem sempre é a realidade para as cidades brasileiras, uma vez que, o mapeamento oficial pode ser inexistente ou desatualizado. Este aspecto tem impulsionado pesquisas a relacionar a qualidade dos dados colaborativos em relação a seus parâmetros intrínsecos, caracterizados por históricos de edições, quantidade de contribuições e contribuidores. Com base em tais questões, questiona-se nesta pesquisa se é possível modelar os padrões espaço-temporais de qualidade intrínseca que influenciam na completude dos dados da plataforma OpenStreetMap (OSM) para obter a sua qualidade. Além disso, questiona-se também se é possível desenvolver ferramentas para a avaliação da qualidade extrínseca na qual seja possível identificar e discutir questões acerca da heterogeneidade dos dados. Dessa forma, este trabalho objetivou o desenvolvimento de uma metodologia para modelar e avaliar padrões espaço-temporais dos indicadores de qualidade intrínseca dos dados plataforma OSM e sua relação com os indicadores de qualidade tradicionais. Foi desenvolvido um procedimento metodológico para modelar as contribuições ao longo do tempo e, a partir dos parâmetros obtidos, identificar de que maneira os padrões se comportam e quais os fatores que influenciam na sua heterogeneidade. Foram desenvolvidos complementos para avaliar e visualizar a acurácia posicional e completude no OSM, de modo que auxiliem na tomada de decisões. A modelagem matemática deu-se a partir da Regressão Logística, em células de 1x1 km. Como resultados, notou-se que o parâmetro de inclinação da curva permitiu diferenciar regiões com grandes contribuições e de crescimento gradativo ao longo do tempo, e até mesmo, a sinergia entre o OSM e os dados oficiais, a partir da importação de feições. Além disso, notou-se que existe uma relação direta da completude dos dados e a saturação da curva (estabilidade das contribuições nos últimos anos), principalmente nas análises que envolvem os eixos viários. Conclui-se que é possível utilizar os padrões espaço-temporais de contribuição como medida de qualidade intrínseca, diante das questões relacionadas com a qualidade dos dados e recomenda-se a continuidade das análises, utilizando diferentes regiões de estudo, categoriais e tamanhos de células.

Palavras-chave: OpenStreetMap, Qualidade dos dados Geoespaciais, Heterogeneidade dos dados Geoespaciais, Avaliação da qualidade intrínseca, Regressão Logística.

ABSTRACT

Issues related to the outdated status of official mapping are recurrent in different locations in Brazil and around the world, especially due to the cost-effectiveness of its production and maintenance. In this context, research is focusing its efforts on evaluating the potential of information coming from collaborative mapping, in the context of establishing their integrating and determining quality potential. Traditional approaches are based on comparisons with existing official databases, however, this is not always the reality for Brazilian cities, since the official mapping may be non-existent or outdated. This aspect has driven research to relate the quality of collaborative data in relation to its intrinsic parameters, characterized by historical editions, number of contributions and contributors. Based on these questions, this research asks whether it is possible to model the spatiotemporal patterns of intrinsic quality that influence the completeness of the OpenStreetMap (OSM) platform data to obtain its quality. In addition, it is also questioned whether it is possible to develop tools to evaluate the extrinsic quality in which it is possible to identify and discuss questions over the heterogeneity of the data. Thus, this research aimed to develop a methodology to model and evaluate spatiotemporal patterns of the intrinsic quality indicators of data from OSM and its relationship with the traditional quality indicators. A methodological procedure was developed to model the contributions over time and based on the obtained parameters, identify how the patterns behave and which factors influence their heterogeneity. Complements were developed to evaluate and visualize the OSM positional accuracy and completeness, in order to help in decision making. The Logistic Regression based the mathematic modeling, in cells of 1x1 km. As a result, it was possible to notice that the curve slope parameter allowed to differentiate regions with large contributions and gradual growth over time, and even the synergy between OSM and official data, from the import of features. Besides that, it was noticed that there is a direct relationship between the completeness of data and the curve saturation (stability of the contributions in the last years), especially in the analysis involving road axles. It is concluded that it is possible to use spatiotemporal patterns of contributions as a measure of intrinsic quality, regarding the issues related to data quality. It is recommended to continue the analysis using different study regions, categories and cell sizes.

Keywords: OpenStreetMap, Geospatial Data Quality, Heterogeneity of Geospatial Data, Intrinsic Quality Assessment, Logistic Regression.

LISTA DE FIGURAS

FIGURA 1 – FLUXOGRAMA DE ETAPAS DESENVOLVIDAS NA TESE	31
FIGURA 2 – HIERARQUIA ASSOCIADA AOS TERMOS <i>CROWDSOURCING</i> , UGC, VGI E INFORMAÇÃO GEOGRÁFICA INVOLUNTÁRIA	38
FIGURA 3 – MÉTODOS PARA AVALIAR A QUALIDADE DE DADOS NO VGI.....	52
FIGURE 4 – DOUBLE BUFFER METHOD	84
FIGURE 5 – EXAMPLE OF INSERTING A CONDITION IN THE QGIS FIELD CALCULATOR USING THE MODEL BUILDER TO CONSULT INFORMATION REGARDING A GIVEN SAMPLE OF POINTS.	86
FIGURE 6 – LOCALISATION MAPS.....	87
FIGURE 7 – OPERATIONAL FLOWCHART FOR ASSESSING POSITIONAL ACCURACY IN POINT FEATURES USING THE METHODS OF MERCHANT (1982); AND GALO & CAMARGO (1994).	88
FIGURE 8 – OPERATIONAL FLOWCHART OF THE TOOLS FOR THE ASSESSMENT OF POSITIONAL ACCURACY IN LINEAR FEATURES ACCORDING TO THE SCALES ESTABLISHED IN THE ET-ADGV...	91
FIGURE 9 – FUNCTIONING OF THE TOOL FOR EVALUATING THE POSITIONAL ACCURACY OF LINEAR FEATURES FROM INSERTING THE BUFFER SIZE.....	92
FIGURE 10 – INTERFACE FOR ASSESSING POSITIONAL ACCURACY OF LINEAR FEATURES USING THE ET-CQDG METHOD.....	94
FIGURE 11 – EXAMPLE WITH THE RESULTS OBTAINED IN THE TABLE OF ATTRIBUTES FROM THE ET-CQDG METHOD.	95
FIGURE 12 – MAP OF PLANIMETRIC DISCREPANCIES OBTAINED IN IN SALVADOR-BA: A. PUNCTUAL FEATURES; B. LINEAR FEATURES.....	98
FIGURE 13 – BEHAVIOR OF THE POINT FEATURE DISCREPANCIES SAMPLE	99
FIGURE 14 – LOCALISATION MAP OF THE STUDY AREA.	113
FIGURE 15 – THE LOGISTIC CURVE PARAMETERS AND ZONES	115
FIGURE 16 – ARRANGEMENT OF OSM FEATURES IN CURITIBA (A) AND SÃO JOSÉ DOS PINHAIS (B) CELLS.....	116
FIGURE 17 – SEQUENCE FOR EXTRACTING THE INFORMATION	118

FIGURE 18 – SOFTWARE FOR COMPUTING THE REGRESSION	118
FIGURE 19 – FEATURES AVAILABLE IN OSM (A) AND THE PERCENTAGE.....	119
FIGURE 20 – CELLS IN WITH OVERFLOW (A) AND THE % OF THE LARGEST CONTRIBUTION MONTH (B).....	121
FIGURE 21 – BEHAVIOR OF LOGISTIC REGRESSION IN DIFFERENT CELLS EVALUATED.....	122
FIGURE 22 – BEHAVIOR EXAMPLES OF THE ABRUPT JUMPS IN THE CONTRIBUTION PATTERNS IN DIFFERENT CURITIBA CELLS ..	122
FIGURE 23 – LOGISTIC REGRESSIONS IN DIFFERENT AREAS BEFORE AND AFTER FILTERING THE ADDRESS POINTS	123
FIGURE 24 – CELL IN WHICH THE COLLECTIVE MAPPING ACTIVITY TOOK PLACE.	124
FIGURE 25 – CONTRIBUTIONS IN THE HIGHEST MONTH OVER THE TOTAL (AFTER FILTERING)	125
FIGURE 26 – CELL IN WHICH REPRESENTATIVE EXCLUSION OF FEATURES WAS IDENTIFIED	126
FIGURE 27 – NRMS OBTAINED BEFORE (A) AND AFTER FILTERING THE ADDRESSES (B).	127
FIGURA 28 – EVOLUTION OF OSM CONTRIBUTIONS FROM 2008 TO 2020 IN TWO AREAS IN BRAZIL.....	141
FIGURE 29 – DISCREPANCIES BEHAVIOUR IN SALVADOR-BA.....	142
FIGURE 30 – REGRESSIONS IN DIFFERENT REGIONS OF CURITIBA-PR.....	148
FIGURA 31 – COMPLETUDE DOS EIXOS VIÁRIOS (HIGHWAYS) DO OSM EM CURITIBA	155
FIGURA 32 – COMPLETUDE DAS EDIFICAÇÕES (BUILDINGS) DO OSM EM CURITIBA	156
FIGURA 33 – FLUXOGRAMA DE EXECUÇÃO DO NOVO COMPLEMENTO DESENVOLVIDO.....	159
FIGURA 34 – EXEMPLO DE IDENTIFICAÇÃO DE FEIÇÃO HOMOLOGA.....	160
FIGURA 35 – APLICAÇÃO DE CONDICIONANTE PARA A CLASSE A DO PEC PCD	161
FIGURA 36 – INTERFACE DO COMPLEMENTO DESENVOLVIDO.....	161
FIGURA 37 – RESULTADOS OBTIDOS NA TABELA DE ATRIBUTOS DO QGIS	162

FIGURA 38 – MAPA DA ACURÁCIA POSICIONAL DOS EIXOS VIÁRIOS A PARTIR DA APLICAÇÃO DO COMPLEMENTO	163
---	-----

LISTA DE QUADROS

QUADRO 1 – MATRIZ DE AMARRAÇÃO COM OS CAPÍTULOS DA TESE	30
QUADRO 2 – INDICADORES DE QUALIDADE PARA AVALIAÇÃO DA QUALIDADE DE DADOS GEOESPACIAIS.....	47
QUADRO 3 – PRINCIPAIS PRODUTOS E CONCLUSÕES OBTIDAS EM CADA CAPITULO	164

LISTA DE TABELAS

TABELA 1 – PADRÃO DE EXATIDÃO CARTOGRÁFICA PARA PRODUTOS CARTOGRÁFICOS.....	69
TABLE 2 – CARTOGRAPHIC ACCURACY STANDARD FOR CARTOGRAPHIC PRODUCTS.....	80
TABLE 3 – RESULTS OBTAINED FOR ASSESSING POSITIONAL ACCURACY IN POINT FEATURES.....	96
TABLE 4 – RESULTS OBTAINED FOR ASSESSING POSITIONAL ACCURACY IN LINEAR FEATURES.....	96

SUMÁRIO

1.	INTRODUÇÃO	18
1.1	OBJETIVOS	25
1.1.1	Objetivo Geral.....	25
1.1.2	Objetivos Específicos	25
1.2	JUSTIFICATIVA.....	25
1.3	ESTRUTURAÇÃO DA TESE.....	28
2.	FUNDAMENTAÇÃO TEÓRICA	32
2.1	O MAPEAMENTO COLABORATIVO	32
2.1.1	A <i>web 2.0</i>	32
2.1.2	A definição do Mapeamento Colaborativo	34
2.1.3	Os dados geoespaciais e o VGI no contexto da <i>web 2.0</i>	40
2.1.4	O VGI e a questão da Qualidade dos Dados Geoespaciais	44
2.1.5	O projeto <i>OpenStreetMap</i>	53
2.2	O MAPEAMENTO DE REFERÊNCIA E A QUALIDADE DE DADOS GEOESPACIAIS	56
2.2.1	O VGI e a questão da Qualidade dos Dados Geoespaciais	56
2.2.1.1	Mapeamento de Referência no Brasil	59
2.2.2	A Qualidade de Dados Geoespaciais	62
2.2.2.1	Avaliação da Qualidade de Dados Geoespaciais no Brasil.....	68
2.2.2.2	Indicadores de qualidade para avaliação da qualidade de dados geoespaciais.....	70
3.	QPEC: QGIS TOOLKIT FOR EVALUATING GEOSPATIAL DATA POSITIONAL ACCURACY ACCORDING TO THE BRAZILIAN CARTOGRAPHIC ACCURACY STANDARD	74
3.1	INTRODUCTION	75
3.2	METHODS FOR THE ASSESSMENT OF PLANIMETRIC POSITIONAL ACCURACY OF GEOSPATIAL DATA IN BRAZIL	79
3.3	METHODOLOGY AND DATA	85
3.3.1	Planimetric Positional Accuracy of Point Features	87
3.3.2	Planimetric Positional Accuracy of Linear Features.....	90
3.4	RESULTS AND DISCUSSION	93

3.5	CONCLUSIONS AND RECOMENDATIONS.....	99
3.6	ACKNOWLEDGMENTS	101
3.7	REFERENCES	101
4.	EXPLORING SPATIO-TEMPORAL PATTERNS OF OPENSTREETMAP (OSM) CONTRIBUTIONS IN HETEROGENEOUS URBAN AREAS.....	108
4.1	INTRODUCTION	109
4.1.1	A current panorama of geospatial big data integration and analysis	111
4.2	METHODOLOGY	112
4.2.1	Obtaining the data and adjustment of the logistic curve	114
4.2.2	Applications development.....	117
4.3	RESULTS AND DISCUSSION	119
4.4	CONCLUSIONS AND RECOMMENDATIONS.....	128
4.5	REFERENCES	129
5.	OS PADRÕES ESPAÇO-TEMPORAIS DE CONTRIBUIÇÕES E A QUALIDADE DOS DADOS GEOESPACIAIS NO OSM.....	132
5.1	AVALIAÇÃO E REPRESENTAÇÃO DA QUALIDADE DE DADOS GEOESPACIAIS A PARTIR DO DESENVOLVIMENTO DE FERRAMENTAS EM AMBIENTE SIG.....	134
5.1.1	Referências	136
5.2	DATA QUALITY ANALYSIS TOOLS AND THEIR USE FOR THE DETECTION OF SPATIO-TEMPORAL HETEROGENEITIES IN VGI CONTRIBUTIONS	137
5.2.1	Introduction.....	138
5.2.2	Methodology	140
5.2.3	Results.....	140
5.2.4	Final Considerations.....	142
5.2.5	References	142
5.3	DISCOVERING SPATIOTEMPORAL PATTERNS ON DATA QUALITY ASSESSMENT IN COLLABORATIVE MAPPING: A PRELIMINARY STUDY IN AN AREA OF BRAZIL.....	144
5.3.1	References	149
5.4	O POTENCIAL DOS PADRÕES ESPAÇO-TEMPORAIS PARA A OBTENÇÃO DA COMPLETUDE NO MAPEAMENTO COLABORATIVO..	150

5.4.1	Referências	153
5.5	CONSIDERAÇÕES QUANTO A COMPLETUDE DOS DADOS DO OSM.	155
5.6	CONSIDERAÇÕES QUANTO AO DESENVOLVIMENTO DE COMPLEMENTO QGIS 3.X PARA A AVALIAÇÃO DA ACURÁCIA POSICIONAL.....	158
6.	CONCLUSÕES E RECOMENDAÇÕES	163
6.1	RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	168
7.	REFERÊNCIAS.....	169
8.	APÊNDICES.....	187
8.1	O MAPEAMENTO DE REFERÊNCIA E A QUALIDADE DE DADOS GEOESPACIAIS SINERGIAS ENTRE A PRODUÇÃO DOS DADOS GEOESPACIAIS DE REFERÊNCIA OFICIAIS E COLABORATIVOS: UMA PROPOSIÇÃO DE EIXOS POTENCIAIS	187
8.2	O A PROPOSAL FOR A QGIS PLUGIN FOR SPATIO-TEMPORAL ANALYSIS OF OSM DATA QUALITY: THE CASE STUDY FOR THE CITY OF SALVADOR, BRAZIL.....	190
8.2.1	References	193
8.3	THE CELL SIZE ISSUE IN OPENSTREETMAP DATA QUALITY PARAMETER ANALYSES: NA INTERPOLATION-BASED APPROACH ..	195
8.3.1	References	198

1. INTRODUÇÃO

A atualização cartográfica é um aspecto importante a ser considerado na produção de dados geoespaciais, visto que a dinâmica de alterações que ocorrem no espaço geográfico requer que existam metodologias que acompanhem essas mudanças. O tempo para a atualização do mapeamento de topográfico, em escala regional, pode variar internacionalmente de 10 a 15 anos e, neste sentido, órgãos oficiais de mapeamento devem estabelecer medidas alternativas para a identificação de alterações em feições territoriais (CAROTI; PIEMONTE; NESPOLI, 2017). No que diz respeito ao Brasil, Silva e Camboim (2020) revelam que determinadas regiões sequer possuem um mapeamento topográfico.

Os mapeamentos topográficos são desenvolvidos por Agências Nacionais de Mapeamento (ANM) para a representação de feições com um propósito geral e que sirvam para uma variedade de usos e usuários, por meio da representação simultânea de diferentes categorias de informação (KENT, 2009). De acordo com Maulia (2018), a produção tradicional do mapeamento topográfico requer que existam requisitos que atendam a padrões estabelecidos em normatizações quanto aos métodos de obtenção de dados, ferramentas aplicadas e profissionais com conhecimento específico acerca do processo e, portanto, possui custos onerosos. Além disso, visto que a estrutura tradicional de execução do mapeamento é caracterizada por cobertura aerofotogramétrica e atualmente cobertura laser e implantação de rede de apoio através de técnicas de levantamentos geodésicos, estruturados em metodologias com especificações técnicas detalhadas, o processo de produção cartográfica requer muito tempo para a sua finalização, o que causa impacto direto na regularidade de mapas atualizados (SANTOS *et al.*, 2011).

A abordagem de que mapeamentos de referência são produzidos por agências oficiais e atendem a uma série de padrões estabelecidos foi, por longo período, a única caracterização associada com a produção de dados geoespaciais (ELWOOD; GOODCHILD; SUI, 2012). Nos últimos anos novas fontes de dados e metodologia de coletas foram desenvolvidas, e o acesso à rede mundial de computadores, disponibilização de dados abertos, além do crescente uso de *smartphones*, tem permitido a produção de um grande volume de dados diariamente por toda a população mundial conectada de forma *online*.

A ampliação do volume, variedade e velocidade de atualização de informações caracteriza o *big data*, onde fluxo de dados em meio digital pode ocorrer a partir de diferentes fontes, como a modelagem numérica, uso de *smartphones*, acesso à *internet* e uso de redes sociais (YANG *et al.*, 2017). É válido ainda ressaltar que, atualmente, os avanços tecnológicos viabilizam a qualquer indivíduo que tenha computador ou smartphone com acesso à internet a geração dados geoespaciais (GANAPATI, 2011). É a chamada era da *web 2.0*, termo criado por O'Reilly em 2005, definindo um fenômeno no qual os usuários tornam-se agentes fundamentais na produção e gerenciamento de dados, deixando de ser apenas visualizadores e consumidores de dados (CORMODE e KRISHNAMURTHY, 2008).

Neste contexto, Goodchild (2007) apresentou o conceito de informações Geográficas Voluntárias, do inglês *Volunteered Geographic Information* (VGI), em que os dados geoespaciais, que até então eram desenvolvidos apenas por especialistas, passam a ser produzidos também, por indivíduos que tenham acesso à *internet* e receptores GNSS (*Global Navigation Satellite System*) nos *smartphones*, o que permite com que estes publiquem informações geográficas sobre determinado local e até mesmo, validem as informações postadas por outros usuários em plataformas online. Além disso, é importante citar outros aspectos que estão ligados diretamente ao VGI, como a disseminação e disponibilização de imagens de alta resolução e motivações associadas ao *Open Science*, *Citizen Science* e de Participação Cidadã.

Conforme abordado por Yan *et al.* (2020), mesmo após mais de 15 anos que o termo VGI foi apresentado, atualmente é possível notar que a temática se tornou um tópico importante no que tange a obtenção de dados geoespaciais. Os autores ressaltam que a importância no VGI foi dada, não só por pesquisadores, mas também usuários de dados geoespaciais, visto que estas fontes de dados são indispensáveis para compreender a dinâmica espacial da Terra.

Atualmente existe uma série de técnicas para a obtenção de dados geoespaciais que partem desde levantamentos topográficos e aerofotogramétricos até o uso do VGI. Em relação ao VGI, é importante que exista um eficiente e eficaz manuseio e integração no volume de dados, no âmbito de detectar questões relacionadas ao crescimento urbano e, além disso, torna oportuno o estabelecimento de metodologias para avaliar a qualidade, extrair e integrar as informações relevantes desta crescente quantidade de dados geográficos (BROVELLI, *et al.*, 2019). O potencial do VGI para complementar produtos cartográficos de referência está

associado ao fato de que as mudanças podem ser detectadas de forma mais rápida, o que viabiliza a atualização de dados geoespaciais (DU *et al.*, 2017; FAIRBAIRN; AL-BAKRI, 2013). Du *et al.* (2017), ainda enfatiza o custo reduzido para obtenção de feições oriundas de VGI em relação ao mapeamento topográfico. Além disso, a obtenção de dados geoespaciais que acompanhem as rápidas mudanças que ocorrem no espaço geográfico, são desafios inerentes às agências oficiais de mapeamento (MAULIA, 2018).

É importante considerar que, no Brasil, pesquisas mais recentes têm direcionado esforços para estabelecer a viabilidade do uso de feições oriundas de VGI em relação ao mapeamento topográfico. Machado (2020) desenvolveu a compatibilização semântica entre o modelo conceitual do mapeamento colaborativo da plataforma *OpenStreetMap* (OSM) e o modelo conceitual do mapeamento topográfico oficial brasileiro. Paiva (2021) desenvolveu uma metodologia de avaliação da qualidade intrínseca do OSM e Silva (2022) desenvolveu uma metodologia para a integração do mapeamento colaborativo do OSM, ampliando as análises de Machado (2020) para o mapeamento topográfico brasileiro para escalas pequenas.

Em uma análise quanto às características do VGI realizada por Yan *et al.* (2020), baseada em definições descritas por Coote e Rackham (2008), são apresentados aspectos relacionados ao objetivo, custo, gestão, fonte, coleção, qualidade, licenciamento e acesso. Dentre estas características, a qualidade das feições oriundas de VGI é um dos aspectos mais importantes a ser considerado, visto que, conforme é apresentado pelos autores, desde as primeiras pesquisas até as mais recentes, evidenciam-se a falta de controle de qualidade nos processos de coleta de dados e a carência de metadados padrão sobre a qualidade dos dados.

Os indicadores para a avaliação da qualidade de dados geoespaciais são atribuídos pela ISO (*International Organization for Standardization*) 19157 (ISO, 2013) e são definidos por Acurácia Posicional, Acurácia Temática, Consistência Lógica, Completude e Usabilidade. No Brasil, a Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG), foi desenvolvida em 2015 pela Diretoria do Serviço Geográfico (DSG) e aprovada pelo exército. Nestas são definidos os critérios para a avaliação da qualidade de dados geoespaciais. Estes critérios são derivados da ISO 19157 (ISO, 2013) e são apresentados uma série de métodos avançados de amostragem para que as análises sejam desencadeadas,

complementados pelo Manual Técnico para Avaliação da Qualidade de Dados Geoespaciais do Instituto Brasileiro de Geografia e Estatística (IBGE, 2019).

Visto que no mapeamento topográfico os dados são produzidos por profissionais e de forma homogênea, que não está associada com as premissas dos dados produzidos nas plataformas VGI, uma série de pesquisas tem sido desenvolvidas ao redor do mundo no intuito de realizar análises dos parâmetros de qualidade destes tipos de dados visando a determinação da potencialidade de integração de dados VGI e bases de referências oficiais. Dentre estas pesquisas, é possível citar os trabalhos de Silva (2021), Paiva (2021), Jasin e Hamandani (2020), Joaquim (2020), Machado (2020), Fernandes, Elias e Zipf (2020), Ferster *et al.* (2019), Ibrahim, Ramadan e Hefny (2019), Zhang e Malczewski (2018), Brovelli e Zamboni (2018), Camboim, Bravo e Sluter (2015), Arsanjani *et al.* (2015), Forghani e Delavar (2014), Hecht, Kunze e Hahmann (2013), Giurres e Touya (2010) e Haklay (2010).

É importante considerar que para avaliar a qualidade dos dados no VGI com base nos trabalhos citados foram utilizados produtos cartográficos de referência para proceder com as análises. Contudo, existem casos que este produto cartográfico não está disponível ou não existe e, neste contexto, pesquisas sobre a qualidade do VGI têm sido desenvolvidas a partir da avaliação de parâmetros intrínsecos de qualidade, como histórico de edições, quantidade de contribuintes, quantidade de contribuições, etc. Dessa forma, podem ser citados os trabalhos de Tian, Zhou e Fu (2019), Martini, Kuper e Breunig (2019) e Minghini e Frassinelli (2019).

A partir das abordagens estabelecidas em relação aos dados provenientes do mapeamento colaborativo, bem como, diante dos trabalhos mencionados, notou-se que um dos principais desafios relacionados com a qualidade deste tipo de informação está relacionado diretamente com a heterogeneidade dos dados, visto que, as características das informações representadas podem variar de acordo com o local, perfil dos contribuintes e dinâmica das contribuições. Tal aspecto implica diretamente na qualidade de feições oriundas do VGI e faz com que existam variações em relação a área de estudo ou até no mesmo local.

Conforme é abordado por Touya *et al.* (2017), dentre os indicadores de qualidade que podem ser heterogêneos no VGI, os mais significativos correspondem a Acurácia Posicional, Acurácia Temática, Completude e Consistência Lógica. Na Acurácia Posicional, os autores destacam que a heterogeneidade pode levar os especialistas a usarem características incompatíveis e em uma mesma escala.

Conforme destacam Brovelli e Zamboni (2018), para determinar correspondências entre diferentes tipos de dados geoespaciais, é necessário que estes tenham escalas semelhantes ou que estejam representados em níveis de detalhes similares e no OSM, as diferentes formas de coleta, faz com que os dados geoespaciais dispostos tenham diferentes tipos de características quanto a sua precisão.

No que diz respeito à heterogeneidade na Acurácia Temática e na Consistência Lógica, a cartografia automatizada requer a dependência de informações temáticas para classificar produtos cartográficos, bem como, topologias de redes geográficas para identificar características importantes e técnicas de simbolização (TOUYA *et al.*, 2017). A plataforma OSM, por exemplo, possui um modelo semântico global que pode ser entendido com diferentes interpretações pelos colaboradores.

Na Acurácia Temática, Touya *et al.* (2017) afirmam que a heterogeneidade dos dados faz com que os processos dependam da geometria das feições e aspectos semânticos são utilizados de acordo com a sua disponibilidade. Segundo os autores, a heterogeneidade na Completude levanta o problema dos “espaços vazios” nos mapas e no que se refere ao VGI, pode significar regiões com vazios ou incompletas. É válido ressaltar que o principal fator abordado na literatura, quanto ao aspecto caracterizado, diz respeito à própria concentração da população em diferentes locais.

Visto as constatações e desafios relacionados com a heterogeneidade dos dados no VGI, pesquisas mais recentes têm direcionado aplicações específicas para feições oriundas de VGI em mapeamentos topográficos a partir de resultados obtidos ao avaliar a sua qualidade. Tais aplicações são descritas nos próximos parágrafos.

Maulia (2018) e Olteanu-Raimond *et al.* (2017), por exemplo, especificam que de acordo com a qualidade obtida, feições oriundas de VGI podem ser utilizadas para integrar, detectar mudanças ou reportar erros no mapeamento topográfico. Neste contexto, uma pesquisa realizada por Ivanovic *et al.* (2019) avaliou o potencial de rastros obtidos a partir de VGI para detectar atualizações no mapeamento topográfico.

Além destes, conforme é abordado por Yan *et al.* (2020), muitas pesquisas têm direcionado esforços para avaliar a qualidade no VGI a partir da combinação de parâmetros intrínsecos e extrínsecos, no intuito de compreender as características do VGI e se a combinação de diferentes parâmetros melhora a qualidade. Neste contexto, é possível citar os trabalhos de Dror, Doytsher e Daltoy (2020), Nasiri *et al.* (2018) e Zhou (2018) que a partir de uma avaliação da qualidade extrínseca,

utilizaram em paralelo características intrínsecas associadas ao VGI. Os trabalhos de Dror, Doytsher e Daltoy (2020) e Nasiri *et al.* (2018) avaliaram a acurácia posicional de feições a partir da combinação do histórico de edições de feições contribuídas no VGI e Zhou (2018) relacionou a completude com a densidade de feições dispostas.

Além das pesquisas citadas, também é importante compreender como é dada a dinâmica de contribuições no VGI. Na pesquisa desenvolvida por Rajaram e Manjula (2019), por exemplo, foi proposto um método para avaliar a proficiência dos contribuintes no VGI e determinar a credibilidade dos dados contribuídos, Viana, Encalada e Rocha (2019) estudaram o valor de contribuições históricas como fonte de dados de amostragem para mapas multitemporais e de uso da terra e Ivanovic *et al.* (2019), que avaliaram o potencial do VGI para detectar atualizações no mapeamento de referência a partir de parâmetros associados ao período em que as feições foram editadas ou atualizadas.

Diante dos aspectos descritos e das características abordadas, é notável que a comunidade científica tem direcionado esforços para compreender as características do VGI quanto à qualidade dos dados, no intuito de estabelecer metodologias para que seja possível utilizá-las na complementação de mapeamentos topográficos. 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.

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). Para isso, entender a qualidade dos dados nessas regiões é essencial.

Pesquisas nesta temática têm buscado entender como a dinâmica dos padrões espaço-temporais na história das contribuições podem ser medidas e modeladas para extrair informações relevantes. Grinberger *et al.* (2021) abordou esse aspecto no trabalho, que mediu quantidades de contribuições ao longo do tempo para detectar eventos no OSM. Além disso, Brückner *et al.* (2021) estimou a completude

das lojas de varejo (que possuíam em suas tags as categorias de shop e amenity) no OSM. Ambos os trabalhos foram baseados na quantidade acumulada de contribuições ao longo do tempo e utilizaram o modelo de regressão logística em seus dados. Este modelo é descrito a partir de uma trajetória semelhante a uma função sigmoide 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. Brückner *et al.* (2021) ainda utilizou outros modelos de regressão em suas análises, caracterizados pela regressão logística de três parâmetros e a hipérbole retangular. Ao apresentar a função de logística para estimar o comportamento dos dados no OSM, Grinberger *et al.* (2021) relaciona as características da curva com as premissas do trabalho desenvolvido por Gröchenig, Brunauer and Rehr (2014) que propuseram uma abordagem para identificar desenvolvimentos regionais e temporalmente diferentes associados à evolução do mapeamento. O modelo proposto permitiu classificar os diferentes estágios de atividade nas contribuições realizadas no OSM como Start, Growth and Saturation. Neste contexto, Arsanjani *et al.* (2015) propuseram um Índice de Contribuição (IC) baseado na análise dos padrões espaço-temporais de contribuições no OSM. O IC foi construído a partir do número de contribuições, média do número de versões das contribuições, média da quantidade de atributos e número de usuários. Além das pesquisas citadas, Witt, Loos and Zipf (2021) analisaram o impacto de grandes importações de dados no OSM para Holanda e a Índia e Guilcher, Olteanu-Raimond e Balde (2022) estudaram a evolução de importações massivas no OSM na região da França.

Visto os aspectos apresentados e as constatações de que feições dispostas em plataformas VGI são heterogêneas, esta pesquisa buscou entender como a evolução desses parâmetros ao longo do tempo pode ser caracterizada como uma medida de qualidade adicional, além dos parâmetros intrínsecos usuais. Questiona-se, então, como problema de pesquisa se é possível modelar os padrões espaço-temporais da qualidade intrínseca temporal que influenciam na completude dos dados da plataforma OpenStreetMap (OSM) e se por meio da sua identificação é possível obter a qualidade dos dados geoespaciais em uma determinada região. Neste contexto, questiona-se também se é possível desenvolver ferramentas para a avaliação da qualidade extrínseca na qual seja possível identificar e discutir questões acerca da heterogeneidade dos dados.

A partir dos questionamentos abordados, parte-se da hipótese de que se existirem relações dos padrões espaço temporais de contribuições no OSM com a qualidade dos dados geoespaciais, então é possível realizar a sua modelagem e extrair os seus parâmetros. Além disso, caso existam ferramentas que viabilizem o cálculo desses parâmetros e espacialização dos resultados, bem como, a avaliação da qualidade extrínseca de dados geoespaciais, podem ser obtidos aspectos que acarretam na heterogeneidade dos dados e a proposição de um novo indicador de qualidade intrínseca do OSM.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Desenvolver uma metodologia para modelar e avaliar padrões espaço-temporais dos indicadores de qualidade intrínseca dos dados plataforma OSM e sua relação com os indicadores de qualidade tradicionais.

1.1.2 Objetivos Específicos

- a) Desenvolver ferramentas para avaliação da Acurácia Posicional de feições pontuais e lineares;
- b) Modelar os padrões temporais de contribuições do OSM;
- c) Analisar os principais fatores que influenciam nos padrões temporais de contribuições e explorar a interação entre dados oficiais e colaborativos neste contexto;
- d) Relacionar os parâmetros temporais e a completude dos dados do OSM.

1.2 JUSTIFICATIVA

O desenvolvimento de metodologias alternativas para obtenção de dados geoespaciais de forma mais rápida e que acompanhem a dinâmica de alterações do espaço tem sido alvo de pesquisas, visto as limitações de tempo e custo, já abordadas, em relação ao mapeamento de referência. Neste contexto, Martins Jr. *et*

al. (2016), afirmam que é desafiador estabelecer uma forma para a aquisição de um dado geoespacial de modo econômico e confiável.

Apesar de questões relacionadas com a atualização cartográfica serem inerentes ao desenvolvimento de novas técnicas, com abordagens tecnológicas, Guimarães *et al.* (1998) já caracterizavam o tempo por ser um aspecto importante para perda de elementos fiéis de feições representadas em produtos cartográficos, visto que as mudanças, naturais ou antrópicas, podem ocasionar a alteração da realidade representada.

No Brasil, as produções de dados geoespaciais de referência em nível nacional, com escala inferior a 1:25.000, estão a cargo do Instituto Brasileiro de Geografia e Estatística (IBGE) e da DSG. Contudo, para mapeamentos na escala de 1:25.000, a cobertura total é de 5% (SLUTER *et al.*, 2019; Silva e Camboim, 2021). No que diz respeito ao mapeamento em escala grande, esta é de competência para os gestores de municípios e estados.

Conforme apontam Klein *et al.* (2017), apesar de países como a Alemanha, Áustria e Suíça possuírem redes de referência altamente densificadas para a realização de levantamentos, essa não é uma realidade no Brasil, visto que alguns municípios ainda carecem de redes de referência para manter mapeamentos atualizados. Camboim, Bravo e Sluter (2015) afirmam que a vasta extensão territorial do Brasil torna custoso o investimento de mapeamentos no Brasil, principalmente os financiados por órgãos governamentais. É válido ressaltar que, apesar das disparidades associadas com a produção de dados geoespaciais de demais países em relação ao Brasil, nos trabalhos que envolvem atualização cartográfica já citados na introdução, questões relacionadas com o custo e tempo para atualização do mapeamento de referência são problemas a nível mundial, em que se enfatiza a necessidade de medidas alternativas para amenizar tal problema.

Muitas pesquisas têm direcionado esforços para avaliar a viabilidade da participação cidadã para representar as mudanças, de forma mais rápida, que ocorrem no espaço geográfico. A integração de informações de conteúdos gerados por usuários na internet em dados governamentais representa uma mudança de paradigma no manuseio do Sistema de Informação Geográfica (SIG), visto que esta é dominante por especialistas, e o cidadão passa a interagir em dados geoespaciais de referência por meio da sua modificação e contestação (JOHNSON e SIEBER, 2012).

No contexto do uso do VGI, conforme já abordado, visto que os métodos para coleta e disposição dos dados ocorrem de forma diferente do mapeamento de referência, muitos trabalhos direcionaram esforços para avaliação da qualidade no VGI no intuito de identificar e mensurar o seu potencial. Maulia (2018), por exemplo, por meio do uso dos indicadores de qualidade, avaliou o potencial de dados do OSM para fins de atualização cartográfica e identificou que, mesmo não sendo possível realizar a integração destas feições em relação ao mapeamento de referência, é possível utilizá-las para detectar regiões onde houve mudanças.

As características da qualidade no VGI permitem obter variações nas discrepâncias obtidas, completude das feições contribuídas, atributos preenchidos corretamente, modelagem geométrica das feições e relações semânticas a um dado de referência. Análises realizadas considerando tais características permitem detectar a desatualização do mapeamento de referência em categorias específicas de feições que estejam representadas corretamente representadas no VGI e, além disso, estudos associados com os indicadores citados permitem obter incertezas associadas com a escala de mapeamento e identificar a viabilidade para fins de atualização cartográfica

Olteanu-Raimond *et al.* (2017) apontam que com a adoção da política de dados abertos, Agências Oficiais de Mapeamento de vários países tem disponibilizado seus dados de seus dados para o público e cita a Finlândia, Holanda, Grã-Bretanha e a França, que divulgaram parte de seus dados para acesso gratuito. A maior parte desta disponibilização ocorreu no conjunto de dados da plataforma OSM, que resultou em aprimoramentos da referida plataforma. Neste contexto, voluntários podem contribuir com a identificação de alterações decorridas no mapeamento de referência e destacar para as agências de mapeamento áreas que precisam ser vistoriados e atualizados (OLTEANU-RAIMOND *et al.* 2017). Finlândia, França, Países Baixos, Espanha e Suíça utilizam dados VGI em mapeamentos oficiais para fins de detecção de alterações.

Visto que nos procedimentos de avaliação da qualidade a heterogeneidade dos dados é um aspecto recorrente no VGI, nas pesquisas mais recentes ao redor do mundo que visam avaliar o potencial destas informações, notou-se que tem crescido as verificações que tendem a determinar as suas características intrínsecas relacionadas ao histórico de edições, perfil das contribuintes e padrões de contribuições e relacioná-los com os métodos de qualidade previstos pela ISO 19.157

(ISO, 2013). Contudo, no Brasil ainda são poucas as pesquisas que estudam as influências dos padrões de contribuições e heterogeneidade na qualidade dos dados geoespaciais do VGI.

As pesquisas mais recentes nesta temática no Brasil são 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), nos quais realizou-se a compatibilização semântica de estruturação dos dados colaborativos da plataforma OSM em relação a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV), Silva *et al.* (2018) que avaliaram a viabilidade do uso de dados VGI para o mapeamento de municípios de pequeno porte, Martins Junior *et al.* (2016) apresentaram uma metodologia para a elaboração de uma base de dados voluntária e um estudo de caso relacionado a postos de combustíveis e Camboim, Bravo e Sluter (2015) investigaram a completude dos dados do OSM em áreas heterogêneas no Estado do Paraná. No contexto apresentado é possível notar a viabilidade no uso do VGI na atualização de dados geoespaciais, ou até mesmo, para validar a detecção de alterações desencadeadas por outros tipos de análises é necessário compreender as características que estão relacionadas a estas feições.

1.3 ESTRUTURAÇÃO DA TESE

A tese apresentada está estruturada em seis capítulos. O primeiro capítulo descreve o problema, hipótese, objetivos e justificativa em torno de toda abordagem do trabalho. O segundo capítulo apresenta o estado da arte de pesquisas que envolvem a qualidade de dados no mapeamento colaborativo, bem como, referencial teórico sobre as definições de VGI e de Mapeamento Topográfico. Além disso, descreve as legislações para a produção do Mapeamento Topográfico no Brasil e as principais normatizações que definem os indicadores para avaliação da qualidade em diferentes países.

O terceiro capítulo corresponde a um artigo científico que trata do desenvolvimento de uma aplicação, a partir de complemento no QGIS para a avaliação da acurácia posicional de dados geoespaciais, com base na legislação brasileira e métodos estatísticos previstos na literatura. Para validação dos dados foram utilizadas amostras oriundas do OSM, o que permitiu apresentar e discutir questões relacionadas com a heterogeneidade dos dados e as diferenças na

magnitude dos dados em uma mesma região. O capítulo 4 deu-se pelo desenvolvimento de um artigo científico onde é proposta uma metodologia para obtenção dos padrões de contribuição no OSM com base no acumulado de contribuições ao longo do tempo. Neste contexto, extraiu-se todo 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.

Com base nas constatações e aspectos obtidos nos capítulos 3 e 4, o capítulo 5 apresenta a concatenação de quatro resumos expandidos científicos publicados que relacionam os padrões de contribuições com a acurácia posicional e completude no OSM. No primeiro artigo realizou-se a complementação da ferramenta desenvolvida para avaliação da acurácia posicional do capítulo 3 e foi desenvolvido um novo plugin para avaliação da completude de eixos viários do OSM. Tal medida é baseada na diferença entre os comprimentos de vias nas feições avaliadas e de referência. O segundo artigo discute a heterogeneidade dos dados com base nas respostas obtidas na modelagem do histórico de edições e descreve os diferentes padrões obtidos, com base na dinâmica de contribuições e seu comportamento. A aplicação foi realizada na região central de Curitiba-PR, entre 01-01-2008 e 01-08-2021. Assumiu-se que por se tratar de uma região com uma quantidade mais expressiva de contribuições, a interação dos usuários seria mais significativa, permitindo a modelagem dos dados. O terceiro artigo tratou da discussão da heterogeneidade dos dados com base na sua acurácia posicional e do comportamento das contribuições ao longo do tempo. As aplicações desenvolvidas deram subsídio para compreender os aspectos das contribuições em diferentes regiões do Brasil e estimar o comportamento das discrepâncias em uma dada região. O quarto artigo trata do estudo da relação entre os parâmetros obtidos na modelagem matemática de feições e a completude dos dados no OSM. Foram selecionadas as categorias de vias e edificações e com base dos resultados obtidos para os cálculos de comprimento e área, foram calculados os percentuais. Avaliou-se também, a inclinação da curva nos últimos anos de análise para extrair respostas.

A partir de todas as análises realizadas, o capítulo 6 aborda as conclusões baseadas nos objetivos descritos para esta tese e também apresenta conclusões para trabalhos futuros. O Quadro 1 apresenta a matriz metodológica dos capítulos 3,4 e 5 e

o fluxograma da Figura 1 apresenta uma esquematização de como os capítulos desenvolvidos se relacionam.

QUADRO 1 – MATRIZ DE AMARRAÇÃO COM OS CAPÍTULOS DA TESE

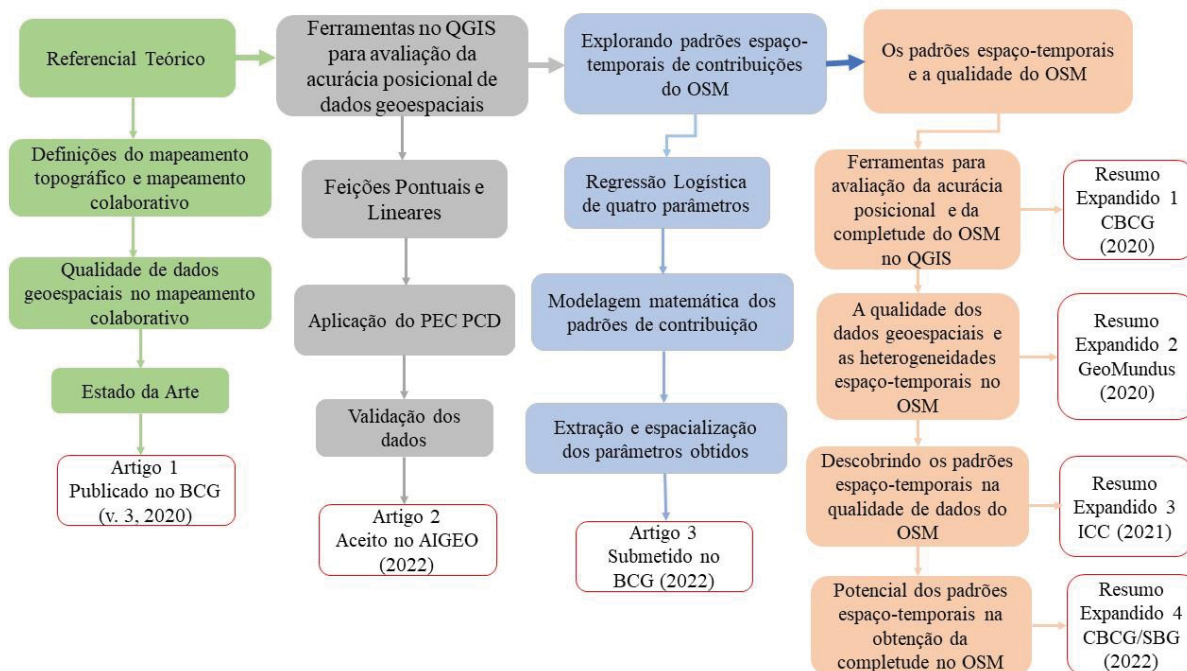
Título	Problema	Objetivo	Justificativa	Método
<p>QPEC: QGIS toolkit for evaluating geospatial data positional accuracy according to the brazilian cartographic accuracy standard</p>	<p>Como desenvolver ferramentas para avaliação da acurácia posicional em ambiente SIG na qual seja possível analisar as discrepâncias com base na distribuição espacial dos erros?</p>	<p>Desenvolver ferramentas no QGIS para avaliação da acurácia posicional de feições pontuais e lineares.</p>	<p>Necessidade do desenvolvimento de aplicações para avaliar a qualidade dos dados geoespaciais de forma factível.</p> <p>As diferentes fontes de dados geoespaciais disponíveis atualmente demandam a existência de ferramentas em que seja possível analisar a heterogeneidade dos dados.</p>	<p>Implementação de fluxos automatizados no QGIS e estrutura SQL-LIKE para aplicação das condicionantes estabelecidas pelo PEC PCD.</p>
<p>Exploring spatio-temporal patterns of OpenStreetMap (OSM) contributions in heterogeneous urban areas</p>	<p>É possível desenvolver uma medida de adicional de qualidade intrínseca com base na modelagem matemática do histórico de edições no OSM?</p>	<p>Desenvolver uma metodologia para obter padrões de contribuições no OSM a partir da Regressão Logística.</p>	<p>Necessidade de propor novas medidas de qualidade no OSM baseadas em aspectos intrínsecos.</p> <p>Aplicações voltadas para estimar o estágio de mapeamento no VGI.</p>	<p>Subdivisão da região avaliada em células homogêneas. Extração e espacialização dos parâmetros da Regressão Logística.</p> <p>Estimativa dos comportamentos atípicos nos padrões de contribuição no OSM.</p>

<p>Os padrões espaço-temporais de contribuições e a qualidade dos dados geoespaciais no OSM*</p>	<p>É possível estabelecer relações entre a modelagem matemática dos padrões de contribuição e a completude dos dados no OSM?</p> <p>Como melhorar o QPEC para permitir a identificação de feições homologas?</p>	<p>Avaliar a existência de relações entre a completude das <i>highways</i> e <i>buildings</i> do OSM e os parâmetros obtidos na Regressão Logística.</p> <p>Propor e implementar melhorias no QPEC para permitir a identificação de feições homologas e visualizar a heterogeneidade dos dados.</p>	<p>Necessidade explorar métricas de qualidade com base em parâmetros intrínsecos.</p> <p>Possibilidade de ampliar discussões acerca da heterogeneidade dos dados do OSM.</p>	<p>Subdivisão da região avaliada em células homogêneas.</p> <p>Obtenção e espacialização dos parâmetros da regressão logística, bem como, da completude dos dados.</p> <p>Complementação do QPEC a partir de Implementação de fluxos automatizados no QGIS que permitiram identificar feições homologas.</p>
--	--	---	--	--

*Capítulo formado pelo desenvolvido de quatro resumos expandidos.

FONTE: O autor (2022).

FIGURA 1 – FLUXOGRAMA DE ETAPAS DESENVOLVIDAS NA TESE



FONTE: O autor (2022).

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são abordados aspectos conceituais que subsidiam o desenvolvimento desta pesquisa. Dessa forma, são apresentadas questões relacionadas com a definição e evolução do mapeamento colaborativo e seu estado da arte, características da qualidade de dados geoespaciais e vertentes para relacionar estas temáticas, a partir das peculiaridades associadas a este tipo de dado geoespacial.

Parte da literatura desenvolvida neste capítulo está apresentada no artigo “*The quality of OpenStreetMap in a large metropolis in northeast Brazil: Preliminary assessment of geospatial data for road axes*”, publicada no periódico Boletim de Ciências Geodésicas (vol. 3, 2020), podendo ser acessada a partir do endereço online: <https://doi.org/10.1590/s1982-21702020000300012>.

2.1 O MAPEAMENTO COLABORATIVO

2.1.1 A web 2.0

Nas pesquisas que abordam o uso do mapeamento colaborativo em contextos relacionados a identificação e análises do potencial destes dados geoespaciais, a principal motivação está associada aos avanços tecnológicos computacionais das últimas décadas e do conceito de VGI trazido por Goodchild (2007), e da mudança no formato da *internet*, de *web 1.0* para *web 2.0* (CORMODE e KRISHNAMURTHY, 2008). Esta mudança está associada ao fato de que o usuário que acessa plataformas web, passa, também, a interagir e contribuir com o fornecimento de informações, além de ser consumidor de informações. Tal contextualização é descrita em trabalhos como Zielstra e Zipf (2010), Giurres e Touya (2010), Forghani e Delavar (2014), Senaratne *et al.* (2017), Brovelli e Zamboni (2018), Martini, Kuper e Breunig (2019), Basiri, Haklay e Mooney (2019), Jasim e Al-Hamadani (2020) e Seto, Kanasugi e Nishimura (2020).

É importante considerar que, independentemente do tipo de análise e o período em que esta foi desencadeada, a viabilidade em avaliar dados geoespaciais oriundos de contribuições em plataformas online, quase sempre está relacionada com a grande quantidade de dados fornecidos diariamente. Além disso, pode-se

destacar a facilidade na coleta e posterior obtenção com a qual estes são caracterizados. Tais aspectos divergem das fontes de dados geoespaciais que formam o mapeamento de referência, onde o rigor técnico para a coleta das informações inviabiliza com que alterações no espaço geográfico sejam representadas em curto período ou em tempo real.

A mudança no formato da internet de *web 1.0* para *web 2.0* contribuiu diretamente para tornar viável estudos quanto ao potencial de dados geoespaciais fornecidos por usuários. A *web 2.0* caracteriza a combinação de inovações da web dos últimos anos e trata de uma plataforma que está associada ao uso de recursos tecnológicos para a sua construção, onde os usuários são objetos destacados em primeiro plano (CORMODE e KRISHNAMURTHY, 2008).

Cormode e Krishnamurthy (2008) ainda destacam que na *web 1.0*, eram poucos os produtores de conteúdo, visto que os usuários atuavam como consumidores. Em outras palavras, a *web 1.0* permitiu com que informações fossem encontradas e lidas. Não haviam muitas interações e contribuições de conteúdos por parte dos usuários (NAIK e SHIVALINGAIAH, 2008). Por sua vez, a *web 2.0* trata de uma coleção de aplicativos online de código aberto, que permite a interação e controle do usuário e, dessa forma, viabiliza expandir as suas experiências e conhecimentos, como participantes em diferentes processos. (CONSTANTINIDES e FONTAIN, 2008).

Constantinides e Fontain (2008) afirmam que o termo *web 2.0* é amplamente utilizado, apesar de ambiguidades associadas com o seu significado concreto. Contudo, em um consenso descrito pelos autores, o termo auxilia a estabelecer uma visão comum e fornece uma plataforma para que sejam desenvolvidos aplicativos associados a serviços online e controlados por usuários. O aspecto democrático da *web 2.0* pode ser exemplificado através da existência de diferentes grupos de usuários na internet que podem realizar trocas de conteúdo de texto, áudio ou vídeo entre si e realizar interações em outros grupos. Além disso, torna possível combinar e renderizar conteúdo de diferentes formas (CORMODE e KRISHNAMURTHY, 2008).

No que diz respeito a sistemas que caracterizam a *web 2.0*, baseados nas definições e aspectos descritos, podem exemplificados as seguintes plataformas *online*:

- *Flickr*: Plataforma que permite com que usuários realizem o compartilhamento de fotografias tomadas em qualquer região do mundo;
- *Youtube*: Permite com que usuários submetam e compartilhem gravações de vídeo na plataforma, de forma que demais pessoas possam ter acesso e realizar interações (comentários) sobre estas;
- *Facebook*: Plataforma de rede social que permite com que usuários ao redor do mundo compartilhem textos, áudios e vídeos, integram sobre elas e se comuniquem entre si;
- *Wikipedia*: Enciclopédia *online*, editada e mantida por usuários, em que estes podem contribuir com a inserção de textos com as mais diversas definições.

No que diz respeito à Wikipedia, a sua abordagem é interessante para descrever aspectos da *web* 2.0, visto que, usuários atuam de forma conjunta para produzir e atualizar conhecimentos. Este aspecto difere de enciclopédias tradicionais, pois a descrição das informações é determinada previamente e permanece estática (PAROUTIS e SALEH, 2009).

Diante da contextualização abordada quanto aos avanços tecnológicos, mudanças no formato da internet, em que esta passa a ter uma abordagem voltada ao contexto da inserção e interação de dados, de forma constante, por usuários e em diferentes plataformas, o foco deste trabalho, é realizar estudos voltados para o mapeamento colaborativo, a partir de abordagens associadas com plataformas de informação geográfica voluntária. Neste contexto, as próximas subseções abordam conceitos e definições associadas ao mapeamento colaborativo, bem como, descrevem algumas características associadas ao potencial do uso destes dados geoespaciais.

2.1.2 A definição do Mapeamento Colaborativo

O termo Mapeamento Colaborativo, do inglês *Collaborative Mapping*, foi apresentado por MacGillavry (2003) e se refere a uma iniciativa de desenvolver modelos de diferentes locais do mundo, nos quais qualquer indivíduo possa acessar e analisar estes modelos de forma automática. O autor ressalta que à medida que as informações geográficas são criadas de forma colaborativa, os usuários têm a

autonomia de armazená-las em seus computadores, celulares ou realizar o compartilhamento em meio digital.

Nas bibliografias que abordam as contribuições de usuários em plataformas *online* e que estão relacionadas com a *web 2.0* (subseção 2.1.1), a expressão mais difundida associada ao mapeamento colaborativo é o VGI (GOODCHILD, 2007). Goodchild (2007) quando descreve o fornecimento de informações geográficas de forma voluntária por indivíduos, ressalta que plataformas online têm permitido com que cidadãos possam fornecer conteúdo por meio de uma “colcha de retalhos global”. Já no título do trabalho, o autor remete o termo *citizens as sensors*, em português “cidadãos como sensores”, e a partir de um contexto histórico e de uma abordagem que mostra que conjuntos de ações coletivas e voluntárias podem representar impactos em SIGs, Goodchild (2007) define o que vem a ser VGI, já descrito anteriormente.

A partir do contexto apresentado, uma das definições que preconizou o VGI foi o SIG Público Participativo, do inglês *Public Participatory GIS* (PPGIS). O PPGIS surgiu no ano de 1996, nas reuniões organizadas pelo National Center for Geographic Information and Analysis, realizada nos Estados Unidos (SEE *et al.*, 2016; BROWN; KYTTÄ, 2014, KOUKOLETOS, 2012). Em uma pesquisa realizada por Brown e Kyttä (2014), baseada em definições de Tulloch (2008), apresentam o PPGIS como um campo dentro da ciência da informação geográfica que destaca as maneiras com a qual os usuários utilizam as diferentes tecnologias associadas aos dados geoespaciais no intuito de participar de processos públicos, como mapeamento e tomada de decisão. Uma definição que é utilizada em paralelo com o PPGIS é o SIG Participativo, do inglês *Participatory GIS* (PGIS). Estudos de Brown e Kyttä (2014) apontam que o PGIS surgiu com abordagens participativas em áreas rurais de países em desenvolvimento e corresponde ao resultado de uma fusão espontânea de métodos de Aprendizagem e Ação Participativa com tecnologias de informação geográfica.

Pesquisas realizadas por Yan *et al.* (2020), explicam que embora algumas literaturas apontem diferenças quanto ao PPGIS e PGIS, quanto aos métodos e tecnologias para coleta de dados, os dois termos são frequentemente utilizados de forma intercambiável em trabalhos relacionados ao VGI. Neste contexto, Yan *et al.* (2020) e Brown e Kyttä (2014) explicam que o entendimento quanto ao PGIS e

PPGIS abriram caminhos para compreender o fenômeno VGI, visto que os três conceitos englobam a perspectiva compartilhada do mapeamento colaborativo.

Conforme aponta Koukoletsos (2012), notando-se que através de um SIG é possível estabelecer ferramentas que permitam explorar e marginalizar as comunidades públicas, a ideia do PPGIS é utilizar o SIG para capacitar comunidades excluídas por meio do acesso a dados espaciais, de forma que estes possam ser participativos nos processos de tomada de tomadas de decisões. See *et al.* (2016) apontam que a relevância do PPGIS foi identificada nos processos de planejamento urbano, conservação da natureza e desenvolvimento rural.

A partir de análises desencadeadas por Yan *et al.* (2020) o PPGIS e o PGIS visam envolver populações marginalizadas e sub representadas no planejamento de políticas associadas aos resultados sociais, o que facilita trocas de conhecimento e o VGI enfatiza o aproveitamento de ferramentas por indivíduos para que estes possam contribuir voluntariamente e disseminar o Conteúdo Gerado pelo Usuário, do inglês User-Generated Content (UGC).

No que diz respeito ao VGI, Goodchild (2007) destaca que este é um caso especial de UGC, visto que está associado com conteúdos obtidos em relação à superfície da Terra. O UGC, conforme aborda a OCDE (2007), é uma das principais características associadas com a web participativa. Neste contexto, estão incluídas várias formas de mídia e trabalhos desenvolvidos por usuários a partir da internet e de ferramentas tecnológicas, o que inclui textos, áudios, imagens e combinados. Os autores ainda afirmam que um dos aspectos centrais é que os conteúdos são recomendados e classificados pelos usuários.

Longueville, Ostländer e Keskitalo (2009) exemplificam esse tipo de conteúdo a partir de trilhas GPS de carros e pontos de interesse, como restaurantes e cafeterias. Ainda baseado nesta expressão, um termo apresentado por Antoniou, Morley e Haklay (2010) é o de Conteúdo Espacial Gerado pelo Usuário, do inglês User Generated Spatial Content. Os autores utilizaram tal expressão devido ao fato de que em suas pesquisas, identificou-se que o termo “voluntário” pode ser enganoso devido às particularidades dos dados gerados e as reais intenções dos seus provedores e neste contexto reconhecem a expressão de Conteúdo Espacial Gerado pelo Usuário em uma abordagem mais geral, mas precisa. Mesmo estando adequada, esta definição não é tão difundida quanto o VGI (KOUKOLETOS, 2012).

Uma expressão bastante difundida quanto a geração de conteúdo por usuários é o *Crowdsourcing*. O *Crowdsourcing* é um termo definido por Howe (2006) e diz respeito a uma prática comercial em que uma determinada atividade é terceirizada para os usuários. See *et al.* (2016) afirmam que a expressão também está relacionada com uma solução de baixo custo, o envolvimento de uma grande quantidade de usuários e de ter valor associado a um modelo de negócios. Papapesios *et al.* (2018) ao descreverem o *Crowdsourcing* afirmam ser caracterizada por ser uma atividade participativa aberta onde é permitido que qualquer pessoa possa propor uma tarefa específica de coleta de dados para várias pessoas. Em contrapartida, See *et al.* (2016) afirmam que nem todo tipo de *Crowdsourcing* precisa ser aberto para qualquer tipo de usuário, mas pode se restringir geograficamente ou a determinados grupos com experiência sobre uma determinada abordagem. Além disso, para Koukolestos (2012) a sua definição segue uma abordagem diferente da expressão VGI, visto a natureza comercial dada para a contribuição de indivíduos.

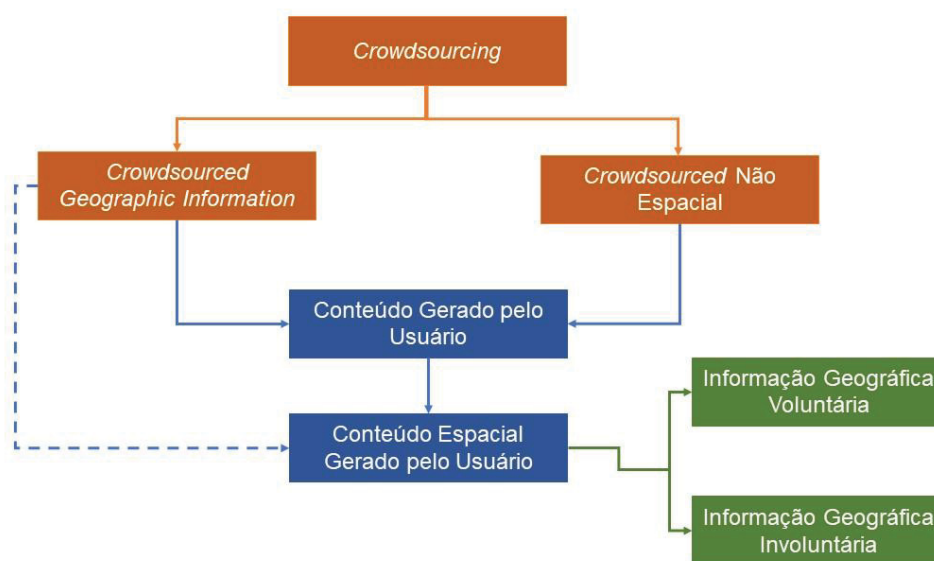
Em uma revisão bibliométrica realizada por See *et al.* (2016) quanto às diferentes expressões associadas ao VGI, é descrito que suas buscas foram baseadas a partir da expressão *Crowdsourced Geographic Information* que, conforme os autores apontam, foi utilizada como “guarda-chuva” para inclusão dos diferentes termos disponíveis. Neste contexto, tal expressão está relacionada a qualquer dado contribuído por usuários com uma referência geográfica que possa ser potencialmente mapeada (SEE *et al.*, 2016). A partir desses aspectos, Papapesios *et al.* (2018) relacionam esta abordagem com o termo VGI, visto que tal expressão foi apresentada por Goodchild (2007) como abrangente para as definições de informações geográficas que são criadas por usuários e, em contrapartida muitos dados podem ser colaborados em plataformas online com geolocalização e fornecer informações atualizadas sobre uma determinada área, sem que necessariamente exista uma ação voluntária direta. Twitter e Flickr são exemplos que caracterizam este tipo de plataforma (PAPAPESIOS *et al.*, 2018; SEE *et al.*, 2016).

A partir do contexto apresentado e no que tange às informações geográficas que são compartilhadas na web sem que necessariamente exista uma contribuição voluntária direta, Fischer (2012) apresentou a expressão Informação Geográfica Involuntária que se define a partir de dados que possuem informação quanto a

sua localização geoespacial, mas que não foram inseridos com caráter voluntário pelos usuários e que podem ser utilizados para uma série de aplicações. See *et al.* (2016) comentam que estes tipos de dados normalmente são obtidos em tempo real por meio de vários tipos de plataformas de mídia social.

Por meio das abordagens descritas quanto ao *Crowdsourcing* e do *Crowdsourced Geographic Information*, foi possível notar a existência de uma hierarquia quanto a estes termos em relação aos que estão associados com os dados geoespaciais colaborativos na *web*, especialmente quando a abordagem está relacionada com o VGI. Dessa forma, o esquema da Figura 2 apresenta os aspectos hierárquicos de tais expressões.

FIGURA 2 – HIERARQUIA ASSOCIADA AOS TERMOS *CROWDSOURCING*, UGC, VGI E INFORMAÇÃO GEOGRÁFICA INVOLUNTÁRIA



FONTE: Adaptado de Papapesios *et al.* (2018).

Conforme apresentado na Figura 2, o *Crowdsourced Geographic Information* engloba dados georreferenciados que são inseridos por usuários na *web*, seja estes por ações voluntárias diretas ou por Informação Geográfica Involuntária. Papapesios *et al.* (2018), See *et al.* (2016) e Craglia, Ostermann e Spinsanti (2012) ainda diferenciam as contribuições de dados geoespaciais a partir de usuários ativos no caso do VGI e usuários passivos para a Informação Geográfica Involuntária, o que permite classificar os aspectos de uma determinada contribuição.

Dentre alguns outros conceitos abordados em Goodchild (2007), alguns que chamaram a atenção quanto às suas definições, são os de *humans as sensors*, em português “humanos como sensores”, e *citizen science*, do português ciência cidadã. No que diz respeito aos humanos como sensores, o autor cita que cada cidadão é um sintetizador inteligente e capaz de interpretar informações sobre um determinado local. Já na ciência cidadã, é utilizada para descrever comunidades ou redes de cidadãos que atuam em algum domínio específico, o que torna necessário um certo grau de conhecimento quanto a temática em que o usuário atua.

Uma série de termos são utilizados no contexto das pesquisas que abordam as colaborações em plataformas *online*. See *et al.* (2016) afirmam que alguns destes termos estão relacionados com a natureza espacial dos dados, como é o caso do VGI e da Neogeografia, e outros possuem uma aplicação muito mais ampla, como por exemplo o *Crowdsourcing*, Ciência do Cidadão e UGC. Apesar de possuírem diferenças nas definições é comum que estes termos sejam utilizados de forma conjunta, visto a ideia básica relacionada com o envolvimento do cidadão no desenvolvimento de atividades ligadas com a ciência da informação geográfica.

Nas pesquisas de Martins Jr. e Silva (2018), See *et al.* (2016) e Koukoletsos (2012) foram apresentadas as definições para cada termo na temática das colaborações na web. É importante considerar que nos trabalhos de Martins Jr. e Silva (2018) e See *et al.* (2016), foram realizadas pesquisas bibliográficas e apresentadas as definições a partir de um contexto histórico. Martins Jr. e Silva (2018) ainda propõem uma hierarquização dos termos baseada nas suas definições características. Baseadas nas constatações e aspectos trazidos por estes autores, os próximos parágrafos apresentam as principais definições e referências acerca do assunto abordado.

O termo Neogeografia, do inglês *Neogeography*, foi trazido por Turner (2006) e diz respeito a criação e compartilhamento de mapas a partir do uso de ferramentas e aplicações disponíveis gratuitamente. A abordagem associada com a Neogeografia é caracterizada pelas diferenças relacionadas com as técnicas utilizadas para a geração do mapeamento de referência. De acordo com Byrne e Pickard (2016) a descrição da Neogeografia está relacionada à redução de barreiras de custo e acesso para obtenção de dados geoespaciais, o que permitiu que usuários não especialistas também pudessem elaborar estes tipos de dados. É importante considerar que, em uma abordagem trazida por Koukoletsos (2012),

embora o conceito clássico de Neogeografia esteja relacionado com usuários não especialistas, existe a possibilidade de estes deterem algum conhecimento a respeito de definições que englobam os dados geoespaciais previamente.

2.1.3 Os dados geoespaciais e o VGI no contexto da *web 2.0*

Conforme descrito, os avanços computacionais e a mudança no formato da *internet* têm viabilizado com que usuários compartilhem um grande número de dados em plataformas online diariamente. Neste contexto, Haklay, Singleton e Parker (2008), apontavam a crescente quantidade de usuários que interagem em plataformas de informação geográfica. Apesar da defasagem temporal da pesquisa, é importante descrever que a reflexão na mudança no formato da internet em relação aos dados geoespaciais, visto que, os autores apresentam que no final de 2007 plataformas, como o *Google Maps* e o *Google Earth* possuíam 71,5 e 22,5 milhões de usuários cadastrados, respectivamente. Haklay, Singleton e Parker (2008) descrevem que as características associadas ao uso da *internet*, permitiram a mudança na natureza dos aplicativos que, em sua totalidade, são denominados de *Geographic World Wide Web* ou *geoweb*.

Define-se como *geoweb* ou *web geospatial* a fusão de informações geográficas em conteúdo *online*. Estes conteúdos, na maioria das vezes, estão implicitamente associados em estruturas da *web 2.0* especialmente no que se refere a processos que abordam a geração de conteúdo a partir de usuários (HAKLAY, SINGLETON; PARKER 2008; ELWOOD; LESZCZYNSKI, 2011; ROCHE; PROPECK-ZIMMERMANN; MERICKSKAY 2013).

Zhang (2019) descreve a *geoweb* a partir de uma visão geográfica da *web 2.0* onde, as suas tecnologias online viabilizam o desenvolvimento e gerenciamento das informações geoespaciais. Dessa forma, em contraste com os SIGs tradicionais existentes, a *geoweb* fornece maior acessibilidade através de interfaces gráficas para o usuário em modelos em que os serviços são predominantemente gratuitos o que promove a participação do usuário neste processo, no intuito de abordar problemas sociais, afirmar o conhecimento local, melhorar comunidades e mudar políticas públicas (SIEBER, *et al.* 2016).

A aplicação prática central que está associada com a *geoweb* diz respeito à caracterização da informação geográfica de um determinado conteúdo online ou a atribuição de nome e informações sobre a sua localização em conteúdos que caracterizam a *web 2.0* (ELWOOD e LESZCZYNSKI, 2011).

A partir das definições apresentadas, a *geoweb* está diretamente relacionada com as características da *web 2.0*, principalmente nos avanços tecnológicos e a interação dos usuários na disposição das informações. Tal aspecto está associado ao fato de que, a maior parte das análises e constatações da *geoweb*, são possíveis graças as tecnologias que estão vinculadas com a *web 2.0* (CROOKS *et al.*, 2014).

Contudo, é importante mencionar que, em uma abordagem de Crooks *et al.* (2014) baseada em aspectos apresentados por Maguire (2006), são descritas diferenças caracterizadas por uma evolução da *geoweb 1.0* e *geoweb 2.0*. Em linhas gerais, as diferenças são descritas baseadas nos mesmos aspectos da *web 1.0* e da *web 2.0*, mas numa vertente associada aos dados geoespaciais. Dessa forma, é apresentada a evolução da disposição de mapas na web, visto que deixam de ter aspectos estáticos e passam a ser dinâmicos, os serviços da web de uso direto (*Google Maps*, *Big Maps*, etc.) e a existência dos portais de catálogo para acesso aos dados geoespaciais (*geodata.gov*, *INSPIRE*, *USGS*, etc).

No contexto da *web 2.0*, houve um aumento significativo na quantidade de dados compartilhados voluntariamente (SETO; KANASUGI; NISHIMURA, 2020). A partir das abordagens que caracterizam uma Infraestrutura Nacional de Dados Espaciais (INDE), o autor destaca a sua relação com o VGI, visto que grupos de indivíduos podem representar informações de forma independente de acordo com necessidades e aspectos de comunidades locais. Dessa forma, os conjuntos de informações fornecidas viabilizam criar uma “cobertura de retalhos”. É importante considerar que, a partir de uma plataforma com ferramentas apropriadas, em que as informações fornecidas possam ser carregadas, torna-se viável realizar ajustes contínuos e a periodicidade na atualização dos dados podem estar relacionadas com necessidades locais (GOODCHILD, 2007).

Ao comparar aspectos do VGI com os métodos convencionais utilizados para a produção de dados geoespaciais, Elwood, Goodchild e Sui (2012) apresentam diferenças em determinados eixos. Estas diferenças estão associadas com o conteúdo da informação, tecnologias utilizadas para a aquisição dos dados,

qualidade dos dados, métodos e técnicas utilizados para manusear os dados, processos sociais que desencadeiam a criação do VGI e os impactos gerados.

Apesar das características e definições do VGI estarem diretamente relacionadas aos avanços tecnológicos e *web 2.0*, Elwood, Goodchild e Sui (2012) afirmam que os esforços para a coleta de informações oriundas por usuários não são novos e que uma série de ações permitiu transformar o VGI em um fenômeno. Dentre as pesquisas apresentadas por estes autores, pode-se destacar os levantamentos de uso da terra na Grã-Bretanha (1930-1940) desenvolvidos por Stamp (1931), em que as atividades foram desenvolvidas, principalmente, por professores e crianças, o Projeto *Domesday* da BBC (OPENSHAW; RHIND; GODDARD, 1986) que em foi montada um arquivo digital com dados espaciais do país, com grande parte dos dados oriundos de contribuições voluntárias e Bunge (1971), em que foram envolvidos moradores urbanos em um processo denominado de “Expedições Geográficas”.

Em uma abordagem prévia, é possível notar que os aspectos relacionados com a coleta de informações voluntárias estão associados com a detecção de aspectos locais sobre uma determinada área, a possibilidade de concatenar e corrigir informações obtidas a partir de diferentes fontes e a viabilidade em detectar mudanças de forma mais rápida de acordo com as necessidades dos usuários. Além disso, é importante considerar a importância dada para este tipo de informação, visto o surgimento de metodologias alternativas para a obtenção de dados geoespaciais.

Os avanços tecnológicos e a *web 2.0* permitiram dispor uma série de ferramentas e plataformas online em que fosse possível compartilhar dados geoespaciais. Estas plataformas, estejam elas associadas a mídias sociais, privadas ou que caracterizem projetos que são desenvolvidos por cidadãos, permitem com que sejam estabelecidas novas possibilidades de instituições governamentais com os cidadãos (KHAN e JOHNSON, 2020).

Mais de uma década se passou desde a abordagem pioneira do VGI e destaca-se que este 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). Conforme descreve Capineri (2016) o VGI pode ser considerado uma inovação onde são combinados a tecnologia, as práticas sociais e o relacionamento de poder. A tecnologia está associada com os dispositivos de localização baseados na *web 2.0* e

o fenômeno do UGC está relacionado a uma mudança cultural que permitiu adotar a colaboração e compartilhamento da informação (CAPINERI, 2016).

Na maioria das vezes a informação voluntária contribuída a partir de uma estrutura de banco de dados e disponibilizado gratuitamente para que demais usuários da internet interessados possam obtê-los e manuseá-los (NEIS e ZIELSTRA, 2014). Neis e Zielstra (2014) descrevem que para contribuir nestas plataformas é necessário que o usuário tenha acesso à internet e hardware adicional, como smarthphone ou computador. Os autores comentam que apesar destes requisitos parecerem triviais atualmente, existem alguns aspectos limitantes do VGI, que recaem na abordagem trazida por Goodchild (2007) no que diz respeito aos países em desenvolvimento. Este aspecto está relacionado com o fato de que muitas vezes algumas plataformas VGI suportam apenas a língua inglesa como idioma e as taxas de analfabetismo que podem dificultar com que determinadas áreas sejam mapeadas. Apesar dos aspectos apresentados Neis e Zielstra (2014) apontam a popularidade e disseminação mundial da plataforma OSM.

Em uma pesquisa realizada por Senaratne *et al.* (2016) foram descritos diferentes tipos de VGI de acordo com as suas características e do contexto de uso. Dessa forma, os autores descrevem três diferentes tipos de VGI, que podem ser em forma de mapas, imagens ou texto. O VGI baseado em mapa é caracterizado por todas as fontes VGI onde estão inclusas feições relacionadas com as primitivas geométricas de ponto, linha e polígono, em que estas caracterizam os elementos básicos para construção de um mapa e que podem ser exemplificados a partir das plataformas OSM, *Wikimapia*, *Google Map Maker* (descontinuado em 2017), *Map Insight*, dentre outros (SENARATNE *et al.*, 2016).

O VGI baseado em imagem é produzido implicitamente a partir de plataformas online como *Flickr*, *Panoramio* e *Instagram*, em que os contribuidores fazem a tomada de fotografias de objetos e anexam a referência espacial aos mesmos (SENARATNE *et al.*, 2016). Neste contexto, Koukolestsos (2012) aborda que uma categoria em que uma gama de informações é fornecida por multidões de usuários é referente a plataformas de compartilhamento de fotografias. Os autores ainda destacam que quando as fotografias contribuídas contém as coordenadas geográficas como parte dos seus metadados, estas podem ser consideradas VGI. Nas descrições abordadas e do caráter implícito abordado por Senaratne *et al.* (2016) fica evidente que, apesar da inserção de informações por usuários em

plataformas online já caracterizar o VGI, em alguns casos as informações com referência espacial não caracterizam uma contribuição voluntária ativa. O VGI baseado em texto é realizado de forma implícita a partir de plataformas *online*, como os *Blogs* em geral, *Twitter* ou *Reddit* e neste caso os usuários contribuem por meio de informações geográficas na forma textual a partir dos seus dispositivos com acesso à *internet* (SENARATNE *et al.*, 2016).

2.1.4 O VGI e a questão da Qualidade dos Dados Geoespaciais

Durante longo período, a produção, disseminação e uso de dados geoespaciais eram realizados, em sua totalidade, por usuários que possuíam conhecimentos específicos em cartografia, contudo, esta realidade tem se alterado com os avanços tecnológicos e o desenvolvimento de metodologias alternativas para a geração de dados geoespaciais. Nesta vertente Koukoletsos (2012) aponta novos desafios enfrentados por especialistas em cartografia no que diz respeito a geração de dados geoespaciais de referência e podem ser citados:

- O desenvolvimento de ferramentas para fins de levantamento e produção de mapas, pode ocasionar na exigência de treinamento mudança significativa na cadeia de produção de dados;
- A rapidez no crescimento de redes viárias, expansão de cidades e outras construções humanas, torna os dados geoespaciais desatualizados mais cedo, o que ocasiona em discrepâncias significativas no planejamento na frequência de atualizações cartográficas;
- A crescente demanda por categorias de feições, diferentes das clássicas, para alimentar novas aplicações que utilizam dados geoespaciais (mapas de ciclismo, mapas que auxiliem pessoas com deficiência a se locomover em cidades, aplicativos para monitoramento de fenômenos sociais e desastres naturais);
- A dificuldade em atender a demanda da produção de novos tipos de dados geoespaciais e ao mesmo tempo, garantir a qualidade e periodicidade na obtenção desses dados;
- A diminuição no financiamento para a produção de dados geoespaciais, visto que o fornecedor pode ser uma Agência Nacional de Mapeamento e depender de fundos governamentais ou um fornecedor comercial e depender de vendas e

demanda do mercado. Além disso, o custo da produção de dados pode aumentar, visto a exigência de informações cada vez mais detalhadas.

Existe uma série de métodos para a obtenção de dados geoespaciais e estes incluem feições oriundas de VGI (BROVELLI, *et al.*, 2019). Neste contexto, uma pesquisa realizada por Wagle e Archarya (2020), no intuito de comparar metodologias passadas e futuras de atualização de Mapas Topográficos de Base do Nepal, apresentaram o VGI como tendência mundial para fins de atualização cartográfica, visto a periodicidade regular na inserção de feições e as vantagens econômicas.

Conforme apontam Goodchild e Li (2012), no ano de 2004 ficou evidente que indivíduos são capazes de fornecer dados geoespaciais em plataformas online, com alta qualidade e sem custos. Os mesmos autores ainda ressaltaram efetividade do voluntário para fornecer informações sobre um determinado local, em relação a usuários especialistas de agências governamentais, visto a proximidade com o objeto de estudo e apontaram a plataforma OSM devido aos esforços, na época, de fornecer uma alternativa aos dados geoespaciais de referência. Neste contexto, já em 2008, Flanagan e Metzger (2008) abordaram sobre a credibilidade associada ao VGI e apontaram os avanços e melhorias para as informações geográficas e destacaram os benefícios associados com a disponibilidade de dados geoespaciais, visto a conhecer a auxiliar no conhecimento sobre aspectos específicos de um determinado local.

Uma pesquisa realizada por Budhathoki (2010) revelou que o termo VGI atraiu, cada vez mais, a atenção, na academia, nos negócios e no governo. O autor trouxe alguns exemplos que demonstram tais características e apontou que a Google em 2009 abriu a sua plataforma online para que usuários de 100 diferentes países pudessem realizar edições. Um dos fatores que contribui para a crescente demanda do VGI está relacionado com a representação de informações alternativas para fins de mapeamento. Estas informações podem estar relacionadas a sistemas de navegação em veículos e sites de viagem ou imobiliários (ELWOOD; GOODCHILD; SUI, 2012). No que diz respeito a feições que caracterizam o mapeamento de referência, Elwood, Goodchild e Sui (2012) abordam que nem todos os tipos de dados podem ser realizados por meio de colaborações, como é o caso de Sistema de Manutenção de Controle Geodésico e aquisição de ortofotos, contudo

os autores citam exemplos de atividades VGI para a geração de informações de transporte e hidrografia.

Dada a importância do VGI em diferentes aplicações para a obtenção de dados geoespaciais, uma série de estudos têm sido realizados no intuito de verificar a viabilidade em utilizar conteúdos gerados por usuários para obter informações em diferentes vertentes. Neste contexto podem ser citados os trabalhos de Papapesios *et al.* (2018), onde foi realizado um estudo quanto aos desafios e oportunidades do uso de VGI na defesa (manutenção da paz, ajuda humanitária e ajuda a desastres e guerras), Giuffrida *et al.* (2019) que realizaram um estudo para identificar o potencial do VGI na tomada de decisões sobre transportes e o trabalho de Rota, Albert e Giné (2019) realizaram o monitoramento do uso de áreas públicas protegidas utilizando o VGI.

Na pesquisa de Elwood, Goodchild e Sui (2012), estes diferenciaram as iniciativas VGI para produção de conjuntos de dados de estrutura (ligados ao mapeamento de referência) e dados que não são de estrutura. Os autores apontaram uma série de pesquisas que partem dos esforços VGI, como por exemplo o trabalho de Tulloch (2008), em que foi identificada a presença e localização de piscinas vernais no Estado de Nova Jersey a partir de pesquisas de campo realizadas por voluntários. No que diz respeito a resposta a desastres, um estudo realizado por Goodchild e Glennon (2010) mostrou a aplicação do uso do VGI nos incêndios florestais em Santa Barbara – Califórnia, nos anos de 2007 até 2009.

É válido destacar que pesquisas mais recentes têm relacionado dados VGI com outras fontes de obtenção de dados geoespaciais, como imagens de satélites. Goldblatt, Jones e Mannix (2020) por exemplo, propuseram uma metodologia que se baseava em medições detectadas de forma remota (a partir de imagens de satélite) para prever a completude de edificações no OSM, Olteanu-Raimond *et al.* (2020), ao identificar mudanças na cobertura do solo, também utilizando de imagens de satélite, realizaram *mapatons* em usuários com diferentes perfis para a validação das alterações identificadas, o que caracteriza o uso do VGI.

Nas abordagens referentes ao uso de feições oriundas do VGI, a principal temática discutida está diretamente relacionada com a qualidade dos dados. As pesquisas, em geral, consideram importante a avaliação da integridade do VGI visto que, conforme descrito, este não possui o mesmo rigor técnico associado ao mapeamento de referência. Esta característica é abordada em trabalhos como

Flanagin e Metzger (2008), Goodchild e Glennon (2010), Al-Bakri (2012), Goodchild e Li (2012), Neis e Zipf (2012), Elwood, Goodchild e Sui (2012), Bégin, Devillers e Roche (2013), Senaratne *et al.*, (2016), Seto, Kanasugi e Nishimura (2020).

Conforme descrito, visto que as feições oriundas de VGI não possuem os mesmos critérios de produção associados aos dados de referência, uma série de pesquisas tem sido desenvolvidas ao redor do mundo no intuito de que a qualidade destes dados seja avaliada e possam ser definidos os parâmetros que viabilizem com que estas possam complementar mapeamentos oriundos de outras fontes de dados.

Esta avaliação da qualidade deu-se inicialmente a partir de indicadores clássicos definidos por normas e especificações técnicas para a produção do mapeamento de referência. No Quadro 2 é apresentado, resumidamente, as definições associadas a estes indicadores, baseadas nas descrições da ISO 19.157 (ISO, 2013).

QUADRO 2 – INDICADORES DE QUALIDADE PARA AVALIAÇÃO DA QUALIDADE DE DADOS GEOESPACIAIS

Indicador de Qualidade	Definição
Acurácia Posicional	Acurácia referente a posição de feições em sistema de referência espacial caracterizado
Acurácia Temática	Refere-se à acurácia dos dados quantitativos, o quão corretas estão, quantitativamente, as classificações de feições e seus respectivos relacionamentos
Acurácia Temporal	Qualidade dos relacionamentos temporais entre determinadas feições de acordo com a caracterização dos seus respectivos atributos
Compleitude	Existência ou ausência de feições, seus respectivos atributos e relacionamentos
Consistência Lógica	Nível de aderência relacionada com os aspectos lógicos para a estrutura de dados, relacionamento e atribuição dos elementos
Usabilidade	Baseia-se nos requisitos dos usuários e na compatibilidade das informações contidas nos elementos e sua real necessidade de uso

FONTE: Adaptado de ISO 19157 (ISO, 2013).

É importante considerar que na subseção 2.2, as definições do Quadro 2 são apresentadas com maiores detalhes e em uma abordagem histórica. No que diz respeito às pesquisas que envolvem a avaliação da qualidade no VGI uma das primeiras análises associadas ao dado colaborativo foram desenvolvidas por Haklay (2010), em que foi avaliada a acurácia posicional e compleitude de dados da plataforma VGI do *OpenStreetMap* (OSM) em relação ao mapeamento de referência

do Reino Unido (*Ordnance Survey*) e enxergou-se o potencial no uso destas informações. Na avaliação da acurácia posicional, Haklay (2010) utilizou o método de comparação de buffer e a avaliação da completude foi dada através da comparação baseada em grades a partir do comprimento das estradas. Neste contexto, uma série de análises e métodos foram desenvolvidos em outras pesquisas com base nos critérios avaliados por Haklay (2010). Jasin e Hamandani (2020) realizaram pesquisas acerca da acurácia posicional de vias em diferentes regiões do Iraque, Ferster *et al.* (2019) avaliaram a completude de ciclovias no Canadá, Ibrahim, Ramadan e Hefny (2019), avaliaram a acurácia posicional de feições no Egito, Zhang e Malczewski (2018) realizaram a avaliação da qualidade de vias no Canadá e Brovelli e Zamboni (2018), realizaram a avaliação da acurácia posicional e completude de edificações na Itália.

Para avaliar a qualidade das informações com base nos indicadores dos trabalhos citados é necessário a existência de um produto cartográfico de referência para que as comparações possam ser desencadeadas. Entretanto, algumas vezes os dados de referência são indisponíveis devido aos altos custos ou restrições no seu licenciamento (ZHOU, 2018), então algumas pesquisas têm direcionado esforços para avaliar a qualidade dados VGI sem considerar parâmetros extrínsecos, ou seja, de forma intrínseca. Zouh (2018), por exemplo, propôs a utilização de um indicador de densidade de edificações e realizou experimentos a partir da completude em áreas urbanas da Nova Zelândia e Estados Unidos. Martini, Kuper e Breunig (2019) e Minghini e Frassinelli (2019) fizeram uma análise do histórico de edições e características representadas na Alemanha e em diferentes países da Europa.

No que diz respeito aos parâmetros intrínsecos para a avaliação da qualidade VGI, informações quanto a quantidade de feições representadas, períodos em que as últimas edições foram desencadeadas, quantidade de edições e editores, podem indicar aspectos relevantes quanto ao comportamento espacial das atualizações que são desencadeadas e relacioná-las com outros fatores. Tal característica é evidenciada nas pesquisas desenvolvidas por Tian, Zhou e Fu (2019) e Hecht, Kunze e Hahmann (2013). No primeiro identificou-se a variação da quantidade de edificações do OSM entre os anos de 2013 e 2017 em diferentes regiões da China e no segundo avaliou-se as diferenças de completude de edificações em regiões da Alemanha nos anos de 2011 e 2012. Além disso,

pesquisas mais recentes mostram que é possível combinar diferentes tipos de indicadores no intuito de obter características associadas ao VGI. Camboim, Bravo e Sluter (2015), por exemplo, avaliaram uma série de aspectos em municípios do Paraná, relacionados com: Comprimento de estradas rurais em OSM em relação ao comprimento de estradas rurais no banco de dados topográfico de referência na escala de 1: 250.000; Densidade de estradas em áreas urbanas, Quantidade de edificações em área urbana; Porcentagem de estradas classificadas; Realização da última edição e; Quantidade de editores. Estes aspectos foram correlacionados com uma série de parâmetros socioeconômicos das regiões de estudo, podendo citar as distâncias até a capital, renda e quantidade de habitantes. Em Zhou (2018), foi realizada uma correlação entre a completude de edificações do OSM em relação a sua densidade.

É válido ressaltar que foi identificado, nestes trabalhos, que os principais desafios dos dados VGI estão associados com a heterogeneidade das informações, visto que, estas diferem de acordo com a área de estudo e os elementos representados. Neste contexto, nas metodologias abordadas por Ibrahim, Ramadan e Hefny (2019) e Brovelli e Zamboni (2018), as aplicações desenvolvidas levaram em consideração as características da heterogeneidade.

Apesar das questões associadas com a heterogeneidade dos dados VGI serem desafios apresentados em pesquisas recentes, nas primeiras análises envolvendo estas plataformas, tal característica já era abordada ao realizar avaliações da qualidade, principalmente no que diz respeito a acurácia posicional e a completude. Giurres e Touya (2010) realizaram análises da qualidade do OSM em uma região da França e constataram a heterogeneidade por meio dos indicadores utilizados, onde, ao avaliar a acurácia posicional de segmentos lineares, a partir do método da distância de Hausdorff, encontrou-se que a média das distâncias obtidas foi de 13,57 m e o desvio padrão foi de 8,32 m, onde esta dispersão dos resultados representa a heterogeneidade dos dados do OSM nesta área.

Forghani e Delavar (2014), ao avaliarem a qualidade das vias do OSM em uma região do Irã, por meio de métodos que englobaram verificações geométricas a partir do comprimento das vias, geometria mínima dos limites, centro meridiano e distribuição direcional, com a subdivisão da área de estudo por meio de uma grade com células, notaram que a qualidade dos dados do OSM varia consideravelmente

na área de estudo e apontam a heterogeneidade como principal problema neste conjunto de dados.

No que diz respeito a completude em dados VGI, é válido ressaltar as diferenças na quantidade de feições representadas em áreas urbanas e rurais e que esta se torna problemática em áreas rurais, visto a pouca quantidade de representações nestas áreas (BROVELLI e ZAMBONI, 2018; GUIRRES; TOUYA, 2010; ZIELSTRA; ZIPF, 2010). Hecht, Kunze e Hahmann (2013) ainda afirmam que a completude das feições diminui a medida a que aumentam as distâncias dos locais em relação aos centros urbanos. Tian, Zhou e Fu (2019) ao avaliarem a completude de edificações do OSM na China, a partir da subdivisão por prefeituras, atribuíram a heterogeneidade com a distribuição da quantidade de feições nesta região, visto que em 2017, a quantidade de edificações em Jiangsu era 175.215 e em Guizhou era 1.463.

Zhang e Malczewski (2018) ao avaliarem a quantidade de vias do OSM no Canadá, evidenciaram que o produto cartográfico utilizado como referência para as análises apresentou maior quantidade de feições representadas e ao comparar estes resultados com verificações realizadas em uma região de Londres, as diferenças evidenciaram a heterogeneidade espacial destes dados no Canadá. Os autores também identificaram esta heterogeneidade ao comparar a quantidade de vias com o nome preenchido.

Além das características que associam a heterogeneidade na completude com a quantidade de feições representadas em plataformas VGI, Hecht, Kunze e Hahmann (2013) ao avaliarem a completude de edificações em regiões da Alemanha do OSM, por meio de análises temporais, revelou que aspectos relacionados com a forma como estas feições são modeladas geometricamente, também indicaram a heterogeneidade. Brovelli e Zamboni (2018) descrevem a heterogeneidade da plataforma VGI do OSM a partir dos diferentes métodos de coleta que são realizados pelos usuários, neste sentido, são apontados os seguintes processos para a entrada de feições: Atividades de campo em que o operador grava pontos ou trilhas de GPS, geralmente com receptores de baixo custo; Interpretação de imagens de satélite disponíveis na plataforma OSM e a edição de feições de forma online e; Importação de conjuntos de conjuntos de dados geoespaciais gratuito e abertos, onde é a realizada a mesclagem com as feições já existentes.

No que diz respeito a classificação de feições oriundas de VGI, bem como, a precisão dos atributos preenchidos, estas estão associadas com a acurácia temática, visto que este indicador tem o propósito de verificar o quão corretas estão, quantitativamente e qualitativamente, as categorias avaliadas. Contudo, antes de proceder às análises, é necessário que uma avaliação semântica seja realizada, no intuito de compatibilizar o significado das feições oriundas de plataformas VGI em relação aos dados de referência.

Arsanjani *et al.* (2015), para realizar a avaliação da qualidade do uso do solo em regiões da Alemanha no OSM, precisou proceder este estudo semântico e atribuir o significado das categorias do OSM em relação ao dado de referência. Neste estudo, foram encontradas dificuldades quanto ao entendimento dos idiomas, onde termos com erro de digitação em alemão, precisaram ser traduzidos para o inglês, bem como, feições representadas de forma incompleta ou desconhecidas, representadas pelos contribuidores.

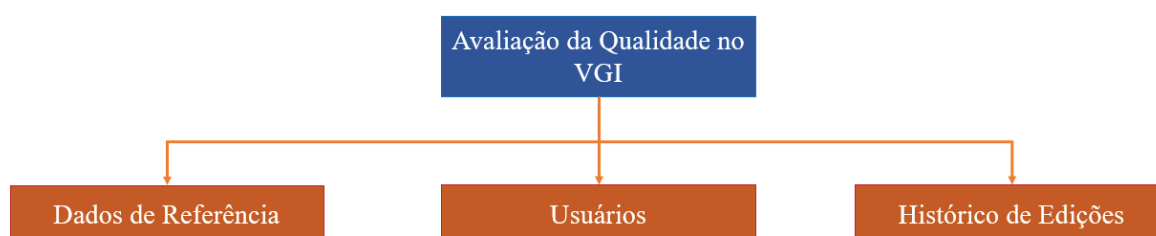
Giurres e Touya (2010) ao avaliar aspectos semânticos da classificação de vias do OSM notaram que as vias principais (“auto estrada” e “primário”) estavam semanticamente corretas, visto que quase todas as feições avaliadas possuíam homólogos nas feições de referência. Contudo, ocorreu uma subestimação, por parte dos contribuidores, da importância das vias secundárias de referência, visto que estas no OSM estavam classificadas como residencial e terciária no OSM. Estes autores destacam essa característica das especificações que abordem de forma clara a classificação semântica dos elementos, como é o caso das vias principais, além disso, a baixa precisão em algumas categorias pode estar relacionada ao fato de que as especificações, apesar do alcance global, podem não ser comumente aceitas, o que ocasiona variações semânticas de acordo com o perfil dos colaboradores.

Na abordagem que trata do uso parâmetros intrínsecos em VGI, pesquisas têm direcionado esforços no intuito de compreender o comportamento das contribuições baseados em perfis de usuários, históricos de edições, tempo de vida útil do contribuinte, quantidade de contribuintes por feição e por área, padrões de contribuições, etc. No trabalho realizado por Sehra, Singh e Rai (2017), por exemplo, foi desenvolvida uma caixa de ferramentas no *software QGIS* para a avaliação da qualidade intrínseca de feições do OSM, a partir de arquivos do histórico de edições na Índia, Tian, Zhou e Fu (2019) avaliaram padrões de crescimento na China a partir

do histórico de edições no OSM, Jacobs (2018) utilizou o histórico de edições e dados dos contribuintes do OSM ao realizar a avaliação da Acurácia Temporal no Canadá e Viana, Encalada e Rocha (2019) realizaram um estudo com o histórico de edições do OSM em Portugal para avaliar a viabilidade em usá-los como fonte de dados de amostragem para mapas multitemporais da cobertura e uso da Terra.

A partir das contextualizações abordadas quanto a avaliação da qualidade no VGI, Nasiri *et al.* (2018) apontam que esta pode ser realizada a partir de três vieses conforme é apresentado na Figura 3.

FIGURA 3 – MÉTODOS PARA AVALIAR A QUALIDADE DE DADOS NO VGI



FONTE: Adaptado de Nasiri *et al.* (2018).

No viés da avaliação da qualidade do VGI relacionado com os dados de referência, estes são comparados com conjuntos de dados que formam o mapeamento de referência, como os exemplos descritos anteriormente (NASIRI *et al.*, 2018). No que corresponde a avaliação da qualidade a partir dos usuários, Nasiri *et al.* (2018) comentam que pesquisas têm sido realizadas para investigar a qualidade dos dados no VGI a partir do comportamento dos usuários e tem sido explorado o impacto do número de colaboradores na qualidade dos dados. No que diz respeito ao histórico de edições, os autores comentam que este viés não emprega dados de referência para a avaliação da qualidade.

Pesquisas mais recentes têm contemplado a avaliação da qualidade no VGI a partir do uso de dados de referência em conjunto com aspectos relacionados com as características dos usuários e com o histórico de edições. Nasiri *et al.* (2018), por exemplo, avaliou a Acurácia Posicional e a Completude das edições mais recentes no OSM do Irã e em seguida realizaram a mesma as mesmas verificações com o conjunto de histórico de edições e verificaram que a qualidade melhorou em 14%.

Em uma pesquisa realizada por Touya *et al.* (2017a) foi avaliada a qualidade do OSM ao longo do tempo a partir da combinação de dados de referência, histórico de edições e relações espaciais.

2.1.5 O projeto *OpenStreetMap*

Visto as características apresentadas para os diferentes tipos de VGI, o foco deste trabalho está relacionado com procedimentos a partir do VGI baseado em mapas, mais especificamente a plataforma OSM, devido a aspectos que serão descritos no decorrer dos próximos parágrafos. O OSM corresponde a uma plataforma VGI *online* e com conteúdo gratuito, em que os usuários podem editar e atualizar informações (TIAN; ZHOU; FU, 2019).

O OSM foi criado em 2004 por Steve Coast com a proposta de um sistema *web* de mapas livres que viabilizassem ao usuário a criação, visualização e distribuição de dados geográficos livremente, composta por estradas ruas e caminhos (BRAVO, 2017; NEIS; ZIPF, 2012). A principal motivação para desenvolver o OSM ocorreu devido a identificação de restrições no uso e obtenção de dados do mapeamento de referência europeia do *Ordnance Survey* (CHILTON, 2011). Neste contexto, no desenvolvimento do projeto OSM a abordagem central está relacionada ao acesso gratuito de informações geográficas atualizadas, visto que o elevado custo de dados geoespaciais precisos inviabiliza a disponibilização destes para todos os indivíduos (HAKLAY; WEBER, 2008). Dessa forma, de acordo com Perkins (2011) e Haklay e Weber (2008) o objetivo do OSM é compartilhar informações sem restrições privadas ou a partir de direitos autorais, o que vem permitindo o crescimento da plataforma quanto a abrangência de informações. Haklay e Weber (2008) ainda descrevem que o OSM segue o modelo de produção entre pares característico ao *Wikipedia*.

O OSM é atualmente a maior coleção colaborativa de dados geoespaciais e é amplamente utilizada em projetos para integração em dados de referência (Brovelli e Zamboni, 2018). A popularidade do OSM fez com que outros projetos, como o Here Map Creator, Map Share e Waze, fossem estabelecidos (NEIS; ZIELSTRA, 2014). Neis e Zielstra (2014) afirmam que no passado tais plataformas provedoras de dados proprietários tinham uma funcionalidade limitada, visto com que só era

permitido que os usuários relatassem erros nos mapas através de notas e atualmente é possível que estes realizem criem e façam correções nas feições dispostas. Além disso, ao comparar o OSM com estas plataformas, os autores destacam que o OSM possui licença de banco de dados *Open Data Commons Open Database License* (ODbL) que permite com que os usuários possam copiar, distribuir, transmitir e adaptar seus dados, desde que a fonte seja devidamente referenciada.

Em uma abordagem histórica, é importante comentar que nos primeiros anos de criação do OSM a maior parte informações geográficas foram coletadas a partir de computadores de mão com GPS (NEIS; ZIELSTRA, 2014), contudo no final do ano de 2006 a *Yahoo* permitiu com que o OSM utilizasse seu serviço *web* de imagens de satélite (até 2011) e em 2010 a *Microsoft Aerial Bing* também permitiu o uso das suas imagens, o que permitiu um aumento na inserção de feições no OSM (NEIS; ZIELSTRA, 2014, NEIS; ZIPF, 2012; HAKLAY; WEBER, 2008). Além disso, países como a Holanda, Áustria e Estados Unidos, Espanha e França importaram informações cadastrais sobre as suas edificações para o banco de dados do OSM (NEIS; ZIELSTRA, 2014).

Conforme as conceituações apresentadas é válido mencionar que no OSM todos os servidores e interfaces utilizados são criados e compartilhadas por voluntários. Neste contexto, informações extraídas da *Wiki OSM Fountation* (OSM, 2020) apresentam a Fundação OSM que corresponde a uma organização internacional sem fins lucrativos e que o seu desenvolvimento tem o intuito de incentivar o crescimento e distribuição de dados geoespaciais de forma gratuita para que usuários possam fazer uso e compartilhá-los. A partir destes aspectos é importante considerar que o OSM visa a construção de um mapa editável para o mundo inteiro, a partir do seu potencial ativo em relação dados geoespaciais oriundos do mapeamento de referência (RAJARAM; MANJULA, 2019; SENARATNE *et al.*, 2016).

Conforme já destacado, os fatores que impulsionaram o crescimento do OSM estão ligados aos avanços tecnológicos e a demanda para a obtenção de dados geoespaciais de forma mais rápida e acessível. Neste contexto, Haklay e Weber (2008) já apontavam as características do OSM a partir da quantidade de contribuintes cadastrados e feições mapeadas na época. Os autores destacam que o OSM já tinha mais de 33.000 usuários registrados e 3.500 usuários ativos em

2008. É importante ressaltar que atualmente existem mais de 9,4 milhões de usuários cadastrados e mais de 23,4 bilhões de feições armazenadas no OSM (OSM, 2022).

A partir de uma revisão bibliográfica realizada por Senaratne *et al.* (2016) a respeito de métodos para avaliação da qualidade em VGI, os autores explicam que o OSM é o projeto mais importante pois, além de fornecer um mapa gratuito do mundo, possui milhões de colaboradores cadastrados, comunidades de mapeadores ativos em diferentes regiões do mundo e são fornecidos mecanismos para que a contribuição ocorra de forma gratuita e flexível (fornecimento de mapa, roteamento, busca por localização de pontos de interesse, etc.).

A dimensão espacial dos dados do OSM é dada a partir das primitivas geométricas de ponto, linha e polígono (SENARATNE *et al.*, 2016) que podem ser classificados a partir de uma ou várias etiquetas (tags) que são atribuídas por uma chave (*key*) e um valor (*value*) (OSM, 2020b; BROVELLI; ZAMBONI, 2018; BRAVO, 2017; SENARATNE *et al.*, 2016, NEIS; ZIELSTRA, 2014, NEIS; ZIPF, 2012). Na *wiki* OSM (OSM, 2020b) é descrito que uma etiqueta não é um elemento, mas uma propriedade aplicada em feições no OSM. Neste contexto, cada feição geográfica que pode atribuída a partir de diferentes combinações de chaves e valores que seguem padrões associados ao OSM (SENARATNE *et al.*, 2016). Apesar da inserção de feições no OSM ser caracterizada pelas primitivas geométricas de ponto, linha e polígono, conforme já descrito, as informações geográficas no banco de dados do OSM são armazenadas a partir de três tipos diferentes de objetos, sendo estes: Nós (*Nodes*), Caminhos (*Ways*) e Relações (*Relations*) (OSM, 2020c; NEIS; ZIPF, 2012). O “Nó” corresponde a uma feição pontual do OSM, associada a um par de coordenadas geográficas de latitude e longitude, Os “Caminhos” são caracterizados pelas feições lineares e poligonais mapeadas no OSM e a “Relação” define as relações lógicas ou geográficas entre os objetos representados no OSM (OSM, 2020c; NEIS; ZIPF, 2012). É importante considerar que o OSM possui uma *wiki*, onde são descritas todas as características associadas ao projeto, bem como, e as definições de todas as etiquetas a partir das combinações citadas. Haklay e Weber (2008) e Neis e Zielstra (2014) destacam que na *wiki* constam as informações sobre o projeto e oferece orientações a respeito das práticas realizadas pelos colaboradores por meio de uma extensa documentação da sua infraestrutura técnica.

Para que possam ser realizadas edições no OSM os colaboradores podem utilizar diferentes editores. Conforme dados da *Wiki OSM* (OSM, 2020d) os três principais editores de dados no OSM são o iD, Potlatch 2 e JOSM. O *iD* é um editor OSM programado em *JavaScript* com o objetivo de apresentar uma interface amigável e simples para que os usuários possam realizar as edições (OSM, 2020e), por sua vez, o *Potlatch 2* é o precursor do iD e que ainda é utilizado por alguns usuários (OSM, 2020d). O JOSM (*Java OpenStreetMap*) é caracterizado por ser uma ferramenta desktop que permite com que os usuários realizem as edições em modo offline, para posteriormente serem incorporadas ao OSM (BRAVO, 2017).

2.2 O MAPEAMENTO DE REFERÊNCIA E A QUALIDADE DE DADOS GEOESPACIAIS

Nesta subseção são abordados conceitos associados ao mapeamento de referência e a qualidade de dados geoespaciais, bem como, abordagens quanto às normatizações estabelecidas no Brasil. Na subseção 2.2.1 estão descritos os principais aspectos que definem o mapeamento de referência.

2.2.1 O VGI e a questão da Qualidade dos Dados Geoespaciais

No que diz respeito ao Mapeamento de Referência, antes de trazer suas definições propriamente ditas, é importante destacar seu conceito em diferentes bibliografias, é comumente descrito pelo termo Mapeamento Topográfico. Conforme é descrito por Kent (2009), o Mapeamento de Referência é realizado por Agências Nacionais de Mapeamento (ANM) para a representação de feições com um propósito geral e que sirvam para uma variedade de usos e usuários, por meio da representação simultânea de diferentes temas.

Em um contexto histórico, Kent e Hopfstock (2018) explicam que se um determinado produto cartográfico tem o intuito de estabelecer noções humanas de lugar e espaço, os mapas desenvolvidos para fornecer observações da forma e caráter da superfície da Terra estavam entre os primeiros a serem desenvolvidos. Neste sentido, no que baseados em definições que descrevem o surgimento do Mapeamento de Referência, os autores explicam que tais mapas já representavam

feições mesmo antes dos procedimentos para mapear elementos em escala fixa que aderem ao Mapeamento de Referência, advindo no século XVI. Conforme é descrito por Machado e Camboim (2019) e Perret, Gribaudi e Barthelemy (2015) o primeiro Mapeamento de Referência realizado com as premissas similares aos levantamentos com técnicas de precisão conhecidos atualmente, foi desenvolvido na França em 1747 por César-François Cassini de Thury e seus engenheiros, a partir de um serviço contratado pelo rei Luís XV, para mapear os detalhes mais sutis de todo o reino francês a partir da divisão do território em uma grade de 180 retângulos (80 km x 50 km). O produto final tinha escala de 1:86.000 e devido a dificuldades financeiras e mudanças no cenário governamental, a finalização do produto final foi dada em 1815.

Atualmente os Mapas de Referência mantém a sua hierarquia entre os produtos cartográficos, além de serem produzidos pelas ANM, devido a precisão científica com a qual a suas representações correspondem às características do mundo real (KENT; HOPFSTOCK, 2018). Kent e Hopfstock (2018) ressaltam que as ANM geralmente são responsáveis por manter o Mapeamento de Referência completo e detalhado do território nacional, este aspecto inclui garantir que exista precisão na representação de características naturais e culturais, como relevo, água, vegetação, transporte, áreas urbanas e toponímias.

O Mapeamento Topográfico de Referência consiste no resultado de diferentes análises que objetivam representar o espaço territorial por meio de uma série de mapas em escalas padrão (MACHADO e CAMBOIM, 2019). As séries de Mapeamentos de Referência nacionais oferecem informações topográficas em escalas pequenas e grandes e fornecem uma estrutura de referência espacial em relação a informações sobre a Terra (KENT; HOPFSTOCK, 2018). Kent (2009) destaca que embora o objetivo geral para a série de Mapas de Referência de cada país seja o mesmo, existem diferenças na maneira como as feições são representadas de acordo com as escolhas sugeridas a partir da escala.

No que diz respeito ao Mapeamento de Referência em escala grande, foco desta tese, de acordo com pesquisas realizadas por Machado e Camboim (2019), Silveira (2019) e Robinson *et al.* (1995) estes são atribuídos pelas escalas de 1:10.000, 1:5.000, 1:2.000, 1:1000 e 1:500. Neste contexto, é importante destacar que um estudo realizado por Machado e Camboim (2019) a partir de diferentes conceitos quanto ao Mapeamento Topográfico em escalas grandes, definem o

mesmo como mapas de uso geral aplicáveis para diferentes usos e usuários e que servem de referência para a produção de outros tipos de mapeamento. As autoras ainda ressaltam que o Mapeamento de Referência em escala grande mostra com precisão e acurácia a localização de objetos na superfície terrestre.

Conforme já mencionado, a realização do Mapeamento de Referência ocorre geralmente por ANMs e todo o processo de produção e avaliação da qualidade é resguardado por especificações e normas técnicas, que permitem a obtenção de mapas com informações condizentes com a realidade (em escala) e que possam disponibilizados e utilizados para realização de análises. Neste contexto, em uma pesquisa realizada por Olteanu-Raimond *et al.* (2017), os autores citam uma série de organizações nacionais de países da Europa que atuam na produção de Mapeamento de Referência, como por exemplo: *Federal Agency for Cartography and Geodesy* (Alemanha), *National Geographic Institute of Spain* (Espanha), *Ordnance Survey Ireland* (Irlanda), *Hellenic Military Geographical Service* (Grécia), *National Land Survey of Iceland* (Islândia), *Ordnance Survey of Great Britain* (Grã Bretanha) e *National Agency for Cadastre and Land Registration* (Romênia).

Diante das abordagens apresentadas, Silveira (2019) realizou um estudo em diferentes bibliografias quanto às escalas de produtos cartográficos que são desenvolvidos por órgãos de mapeamento de diferentes países. Dentre os descritos pela autora, podem ser citados a *Swiss Society of Cartography* (Suíça) onde são desenvolvidas cartas topográficas nas escalas de 1:5.000 e 1:10.000 e que servem de referência para a realização de produtos cartográficos em escalas menores (1:25.000, 1:50.000, 1:100.000, 1:200.000, 1:500.000 e 1:1.000.000), o Instituto Nacional de Geografia (IGN) da França que produz produtos cartográficos em escala pequena do país que variam de 1:20.000 até 1:5.000.000), o mapeamento nas escalas de 1:1.000 e 15.000 (escala grande) na França que são realizados por instituições de cadastro urbano e o *Institut Cartographic de Catalunya* (Catalunha) que desenvolve cartas topográficas nas escalas de 1:500, 1:2.000 e 1:5.000 onde, a partir destas, são produzidos produtos cartográficos em escalas menores.

2.1.1.1 Mapeamento de Referência no Brasil

Em uma abordagem histórica quanto ao desenvolvimento de produtos cartográficos no Brasil, no trabalho desenvolvido por Archela e Archela (2008) é apresentada uma ordem cronológica com fatos que foram desencadeados quanto a sua evolução e, dentre estes, é possível citar alguns que estão relacionados com a produção do Mapeamento de Referência no Brasil, como a Fundação da Comissão Geográfica de São Paulo (1886) onde se iniciaram os levantamentos com operações de triangulação, a criação do Serviço Geográfico Militar (1890) para execução de trabalhos geodésicos e geográficos, a organização do Serviço Geográfico do Exército e a divulgação da Carta do Brasil ao Milionésimo (1922) e o primeiro projeto realizado pelo IBGE (1938) referente a determinação das Coordenadas das Cidades e Vilas.

É importante destacar que apesar da criação do Serviço Geográfico Militar ter ocorrido em 1890, a uniformização da Cartografia no Brasil foi dada em 1946 a partir do Decreto-Lei 9.210 (BRASIL, 1946) (SILVEIRA, 2019; ARCHELA; ARCHELA, 2008). No ano de 1967, a partir do Decreto-Lei 243 (BRASIL, 1967), foram fixadas as Diretrizes e Bases da Cartografia no Brasil e a criação da Comissão de Cartografia (COCAR), que em 1994 passou a ser denominada de Comissão Nacional de Cartografia (CONCAR). Neste contexto, a produção do mapeamento no Brasil ocorre a partir do Sistema Cartográfico Nacional (SCN), que é formado por instituições públicas e privadas, que atuam no intuito de executar a produção cartográfica, onde pode-se citar o IBGE, a DSG e a CONCAR (MACHADO, 2020).

Machado (2020) destaca que a CONCAR é constituída por agentes da defesa do Brasil, Associação de Empresas de Aerofotogrametria e do secretário-geral Conselho Nacional de Geografia do IBGE e que uma das suas atribuições está ligada na elaboração das Instruções Reguladoras das Normas Técnicas da Cartografia Terrestre Nacional, que foram estabelecidas pelas Decreto-Lei 89.817 em 1984 (BRASIL, 1984). Desta forma, o Plano de Ação da INDE (BRASIL, 2010) explica que as definições do Decreto-Lei 243 de 1967 (BRASIL, 1967) e do Decreto-Lei 89.817 (BRASIL, 1984) caracterizam a hierarquia e os diferentes tipos de produtos cartográficos do SCN do mapeamento sistemático básico terrestre a partir das escalas de 1:1.000.000 e 1:25.000 e obedecem a um padrão internacional para a sua nomenclatura.

A partir das abordagens apresentadas nos parágrafos anteriores é possível notar que a produção do Mapeamento de Referência no Brasil em escala pequena está a cargo de instituições governamentais referentes ao IBGE e da DSG. No que diz respeito ao IBGE, conforme apontado por Brasil (2010), a produção da Cartografia Terrestre é atribuída nas escalas menores que 1:250.000 nas estruturas vetoriais e matriciais. No que diz respeito às estruturas vetoriais, a Especificação Técnica do IBGE define um modelo de dados vetoriais no intuito de garantir a sua consistência lógica (BRASIL, 2010).

Na DSG, a produção da Cartografia Terrestre é atribuída para as escalas de 1:250.000 e maiores com estruturas vetoriais e matriciais a partir de diferentes Especificações Técnicas em que são apresentadas as suas finalidades e atribuições. Nos próximos parágrafos são apresentadas as características descritas para cada uma delas. Além disso, compete a DSG o fornecimento de normas e especificações para o Mapeamento de Referência em escala grande (maiores que 1:25.000). É importante destacar que todas as normatizações e Especificações Técnicas homologadas pela CONCAR podem ser usadas tanto pelo IBGE, quanto pela DSG.

A Especificação Técnica de Produtos de Conjuntos de Dados Geoespaciais (ET-PCDG) define os padrões dos produtos de conjuntos de dados geoespaciais a serem observados na produção de produtos cartográficos vetoriais e matriciais (BRASIL, 2010). A ET-PCDG teve a sua primeira edição disponibilizada no ano de 2014 (DSG, 2014) e a segunda edição disponibilizada no ano de 2016 (DSG, 2016), onde ambas estão aprovadas pelo exército.

A Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV) define o modelo conceitual para dados vetoriais a partir da consistência que associadas aos seus conceitos e aderências de diferentes categorias de dados geoespaciais (BRASIL, 2010). A ET-EDGV tem o intuito de padronizar estruturas de dados, no intuito de viabilizar o compartilhamento, interoperabilidade e a racionalização de recursos entre produtores e usuários que fazem uso dos dados geoespaciais. A versão atual é a 3.0 (DSG, 2017). A Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV) define regras para a aquisição da geometria dos dados no intuito de garantir a consistência lógica do atributo, geometria e consistência topológica (BRASIL, 2010). Neste contexto, é importante destacar que na ET-ADGV são definidas as regras para aquisição das

classes presentes na ET-EDGV, caracterizadas por geometrias e atributos. A versão atual é a ET-ADGV 3.0 (DSG, 2017)

A partir das caracterizações apresentadas para a ET-EDGV e ET-ADGV é importante comentar que, apesar de estar previsto o Mapeamento Topográfico em escala pequena pela DSG, a ET-EDGV 3.0 (DSG, 2017) divide as categorias de informação em Mapeamento Topográfico em Escala Pequena e Mapeamento Topográfico em Escala Grande. O modelo conceitual para o Mapeamento Topográfico em Escala Grande passou a ser definido a partir da ET-EDGV Força Terrestre 1ª edição (DSG, 2015a) visto as necessidades de ações de segurança no Brasil a partir da Copa da Confederações (2013), Copa do Mundo (2014) e Jogos Olímpicos (2016).

Machado (2020) destaca que estes eventos requerem o uso de dados geoespaciais temáticos no âmbito da defesa e segurança, além de bases cartográficas em escalas grandes elaboradas por prefeituras municipais. Contudo, a autora afirma que, os mapeamentos em escala grande produzidos pelos municípios não atendiam aos padrões esperados pela DSG, visto que não possuíam um padrão único, o que dificultava a interoperabilidade e integração dos bancos de dados. Dessa forma, a partir da primeira edição da ET-EDGV Força Terrestre (DSG, 2015a) passaram a ser modeladas as estruturas lógicas dos dados geoespaciais a partir da escala de 1:1.000.

É importante mencionar que, apesar das abordagens trazidas, o Mapeamento de Referência em escala grande é, normalmente, de responsabilidade de órgãos federais, estaduais e municipais por meio da contratação de iniciativa privada (MACHADO; CAMBOIM, 2019; BRASIL, 2010).

A Especificação Técnica para a Representação de Dados Geoespaciais (ET-RDG) tem o intuito de garantir a consistência na representação de classes de objetos, sendo um componente necessário para a representação gráfica de dados geoespaciais. Neste contexto, para cada produto atribuído a ET-PCDG são definidos os estilos e representação gráfica para a classe associada ao mesmo. No que tange a situação atual desta especificação, esta ainda está em elaboração e conforme é apresentado na plataforma online da DSG, as normas utilizadas para a apresentação de dados geoespaciais correspondem a 2ª edição do Manual Técnico T 34-700 - Convenções Cartográficas. O MT 34-700 é dividido em duas partes: Normas para emprego dos símbolos (DSG, 1998) e Catálogos símbolos (DSG,

2000). Dessa forma, a primeira parte foi substituída pela ET-EDGV e a segunda parte será substituída pela ET-RDG.

A Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG) define os procedimentos para a realização do controle de qualidade de dados geoespaciais vetoriais (BRASIL, 2010). A ET-CQDG (DSG, 2015b) foi desenvolvida em 2015 e a sua primeira edição foi aprovada pelo exército. Visto que nesta tese os procedimentos metodológicos englobam processos associados com a avaliação da qualidade, os próximos subitens apresentam abordagens quanto à qualidade de dados geoespaciais, desde aspectos históricos até os indicadores utilizados atualmente.

2.2.2 A Qualidade de Dados Geoespaciais

O projeto e desenvolvimento do SIG ocorreu de forma pioneira na década de 1960 e, desde então, uma série de avanços e desenvolvimento de ferramentas tem sido desencadeados para que aplicações pudessem ser realizadas a partir do manuseio de informações geográficas em meio digital (GOODCHILD, 2002). Neste contexto, Goodchild (2002) explica que a importância dada na literatura quanto a acurácia e incerteza no SIG são mais recentes quando comparadas com a época da sua criação. Embora nas décadas de 1960 e 1970 algumas pesquisas em cartografia analítica tenham sido realizadas em um viés estatístico, a importância destas abordagens foram trazidas a partir da década de 1980 (GOODCHILD, 2002). É importante considerar que entre os anos de 1970 e 1980, ao utilizarem o SIG, os usuários passaram a notar que, na maioria das vezes, não se tinha conhecimento acerca da qualidade das feições de entrada o que dificultava direcionar as aplicações que seriam viáveis para as mesmas (KOUKOLESTSOS, 2012).

A partir das abordagens trazidas, pode ser citado como exemplo o volume desenvolvido em 1988 por Goodchild e Copal (1989), que se refere a uma compilação de documentos criados a partir de discussões realizadas por grupos de pesquisadores que atuavam diretamente com o manuseio de dados geoespaciais, onde foram debatidos problemas associados com a acurácia das feições. Esta reunião aconteceu em Montecino-California e foi convocada pelo novo Centro

Nacional de Informação e Análise Geográfica no intuito de estabelecer uma agenda de pesquisa nessa área pelos próximos dois anos (GOODCHILD; GOPAL, 1989).

Desta forma, Devillers e Jeansoulin (2006) citam duas conferências realizadas pela comunidade científica na década de 1990 onde a temática estava relacionada com a qualidade de dados geoespaciais, sendo elas: Simpósio Internacional de Avaliação de Precisão Espacial em Recursos Naturais e Ciências Ambientais, onde foram abordadas a incerteza em relação, principalmente, ao campo dos recursos naturais e do meio ambiente e o Simpósio Internacional de Qualidade de Dados Espaciais que aborda a qualidade dos dados geoespaciais de forma geral.

Visto os aspectos iniciais abordados é possível notar que, mesmo em épocas mais remotas, a compreensão quanto aos aspectos que envolvem a qualidade de dados geoespaciais já era levada em consideração. Neste contexto, 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.

Nos contextos apresentados é possível notar que as conceituações iniciais que envolviam a qualidade de um dado geoespacial eram dadas a partir da acurácia espacial dos dados coletados, ou seja, o quão corretas estão as coordenadas geográficas de determinadas feições, contudo o conceito de qualidade está associado com uma gama muito maior de parâmetros, que partem desde o processo de aquisição, gerenciamento, comunicação até o uso dos dados geoespaciais (DEVILLERS; JEANSOULIN, 2006).

Devillers *et al.* (2010) publicaram um trabalho em que foram descritas as conquistas, falhas e oportunidades dos últimos 30 anos de pesquisas acadêmicas no âmbito da qualidade de dados geoespaciais.

Dentre as conquistas estabelecidas é importante mencionar que os autores citam as diferentes metodologias criadas para avaliar a qualidade espacial de elementos que possuem características específicas, como dados vetoriais, matriciais, Modelo Digital de Elevação (MDE), além de abordagens no SIG para a medição da acurácia posicional de feições pontuais, lineares e poligonais, atributo de modelagem e incertezas temporais e propagação de incertezas. Outra conquista

descrita por Devillers *et al.* (2010) está relacionada com as questões semânticas na avaliação da qualidade de dados geoespaciais, onde esta vertente foi desencadeada a partir do maior compartilhamento de dados geoespaciais em plataformas online.

Os autores comentam que para entender questões complexas e responder aos desafios globais, é necessário que exista a combinação de diferentes conjuntos de dados que utilizam padrões diferentes de acordo com o local e, dessa forma, a qualidade da semântica tem impacto direto na existência de feições, na sua categorização, caracterização dos atributos, aspectos qualitativos e acurácia temporal.

No que diz respeito às falhas, um aspecto importante abordado por Devillers *et al.* (2010) está relacionado com a questão dos metadados, ou seja, a forma de comunicação das informações a respeito da qualidade dos dados. Os autores descrevem que a publicação dos metadados foi a principal abordagem dos produtores, nas décadas que antecederam a publicação do artigo, para comunicar a qualidade dos dados geoespaciais para os usuários, sejam eles profissionais ou leigos. A geração dos metadados é atividade demorada e revelou limitações na capacidade de compreensão do usuário dos possíveis usos de dados.

A partir de todos os aspectos abordados é possível notar a importância dada a qualidade de dados geoespaciais no âmbito acadêmico, principalmente quanto atreladas ao SIG e ao desenvolvimento computacional que permite o compartilhamento de informações. Além dos interesses isolados quanto a estudos envolvendo esta temática, o crescimento dessa abordagem viabilizou o desenvolvimento de grupos de pesquisa ao redor do mundo como um tópico reconhecido nas pesquisas em geomática e as diretrizes para a avaliação da qualidade de dados geoespaciais se tornasse presente em normas e padrões internacionais (DEVILLERS *et al.*, 2010). Nos próximos parágrafos são descritas algumas normas e especificações que abordam a avaliação da qualidade de dados geoespaciais, bem como, os indicadores utilizados.

Conforme já mencionado e de acordo com Chrisman (2006) antes da caracterização e disseminação do termo relacionado com a “qualidade”, o termo mais operativo era o de “acurácia” e, neste contexto, são mencionadas as Normas de Acurácia dos Mapas, do inglês *National Map Accuracy Standards* (NMAS), desenvolvidas pelos Estados Unidos na década de 1940. O autor afirma que nas NMAS eram abordados aspectos relacionados com a acurácia posicional de

produtos cartográficos impressos em que eram definidos limites admissíveis de erros, onde 90% dos pontos testados deverão estar dentro destes limites. É importante ressaltar que as considerações abordadas pelas NMAS serviram de subsídio para diferentes aplicações ao redor do mundo quanto a avaliação da acurácia posicional em produtos cartográficos.

No que diz respeito à mudança de cenário da qualidade de dados geoespaciais para a avaliação a partir da consideração de outros elementos, além da acurácia posicional, Chrisman (2006) considerou as ramificações de três mudanças no contexto dos dados geoespaciais, sendo estas caracterizadas pelo reconhecimento da acurácia dos atributos a partir da premissa do sensoriamento remoto, o movimento topológico e as questões práticas relacionadas com a consistência lógica dos elementos e o surgimento do conceito de adequação ao uso.

No que diz respeito à adequação ao uso, o autor explica que as novas formas de gerenciamento multiplicaram as oportunidades de que determinadas informações sejam manuseadas por diferentes usuários, o que pode alterar as pretensões iniciais estabelecidas. Neste contexto a abordagem de qualidade evoluiu para uma definição baseada em “adequação ao uso” visto que, um mesmo produto pode estar adequado para determinados grupos de usuários em detrimento de outros.

A partir das abordagens trazidas é importante considerar que os procedimentos para a avaliação da qualidade são previstos nas etapas que compreendem a produção de um produto cartográfico de referência, a partir de indicadores que são atribuídos por padrões adotados por diferentes países. Uma pesquisa realizada por Al-Bakri (2012) apresenta diferentes padrões internacionais com diretrizes para a obtenção da qualidade de dados geoespaciais. Conforme é apresentado pelo referido autor a partir de meados da década de 1980, foram realizados esforços por instituições que atuavam diretamente com dados geoespaciais para o desenvolvimento de padrões que viabilizasse a avaliação da qualidade em produtos cartográficos a partir da definição de indicadores.

Al-Bakri (2012) afirma que desde 1992 o *Spatial Data Transfer Standard* (SDTS) foi implementado por diferentes instituições governamentais e privadas dos Estados Unidos a partir da definição dos seguintes indicadores de qualidade para avaliação da qualidade de dados geoespaciais: Linhagem, Acurácia Posicional, Acurácia do Atributo, Consistência Lógica e Completude. É importante considerar

que os indicadores da SDTS foram aceitos por profissionais que atuam na geração de dados geoespaciais e são utilizados pela maioria dos padrões que abordam esta temática (AL-BAKRI, 2012).

A partir dos aspectos apresentados quanto ao SDTS nos Estados Unidos, Al-Bakri (2012) comenta que a demanda pela disseminação de dados geoespaciais viabilizou com que os EUA estabelecessem uma National Spatial Data Infrastructure (NSDI) e, neste contexto, uma parte da NSDI era estabelecer um conteúdo de metadados a partir do *Federal Geographic Data Committee* (FGDC) por meio do padrão *Content Standards for Digital Geospatial*, aceito em 1994 pelo FGDC. No que diz respeito a qualidade, o padrão seguiu os mesmos componentes estabelecidos pela SDTS.

Na Austrália e Nova Zelândia a agência nacional que atua em relação aos dados geoespaciais é a *Australian and New Zealand Land Information Council* (ANZLIC) em que são fornecidos os padrões de dados e viabiliza com que usuários acessem informações espaciais a partir de uma plataforma online (AL-BAKRI, 2012). De acordo com Al-Bakri (2012) o primeiro padrão da ANZLIC foi desenvolvido em 1996 e uma segunda versão foi realizada em 2001, em que os indicadores de qualidade utilizados pela ANZLIC são semelhantes ao da SDTS.

Outro exemplo apresentado por Al-Bakri (2012) é referente ao *European Committee Standardization* (CEN) em que a primeira tentativa de definir um projeto europeu associado com normas de informação geográfica foi estabelecido em 1991 pelo CEN. O objetivo desta norma trata da descrição, definição e estruturação de dados geoespaciais de forma padronizada, além do desenvolvimento de um sistema padrão que que seja possível atualizar e transferir informações geográficas (AL-BAKRI, 2012). No que diz respeito aos componentes para a avaliação da qualidade, além dos descritos pelo SDTS, estão inclusas as análises de Acurácia Semântica e Acurácia Temporal dos dados geoespaciais (AL-BAKRI, 2012).

Visto as normatizações citadas, é possível notar o aumento da necessidade pela unificação de padrões para permitir o manuseio de dados geoespaciais e, neste contexto, a ISO formou um Comitê Técnico voltado para as informações geográficas que viabilizou o desenvolvimento da série ISO 19.100, que diz respeito a um grupo de metadados espaciais (AL-BAKRI, 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).

Na ISO 19113 (ISO, 2002) é definido um esquema de qualidade aplicável aos dados geográficos. Conforme aponta Koukoletsos (2012) na ISO 19113 (ISO, 2002) são estabelecidos cinco elementos para a avaliação da qualidade de dados geoespaciais, denominados por Completude, Consistência Lógica, Acurácia Posicional, Acurácia Temporal e Acurácia Temática. Além disso, o autor ainda destaca que três elementos podem ser utilizados com indicadores de qualidade não quantitativa, atribuídos pela finalidade dos dados, formas de usar os dados e a linhagem. A linhagem está relacionada com a descrição da fonte dos dados, métodos de coleta utilizados e transformações desencadeadas.

A ISO 19114 (ISO, 2003) trata do desenvolvimento de um guia de metodologias de avaliação da qualidade de dados geoespaciais e a ISO 19157 (ISO, 2013) define os componentes, procedimentos e princípios para a qualidade dos dados geoespaciais. Além disso, na ISO 19157 (ISO, 2013) é definido um conjunto de medidas de qualidade de dados relacionados com a avaliação e fornecimento de relatórios de qualidade.

É importante considerar que a ISO 19157 (ISO, 2013) trata de uma atualização das definições estabelecidas na ISO 19113 (ISO, 2002) e na ISO 19114 (ISO, 2003) em que, além dos indicadores apresentados na ISO 19113 (ISO, 2002), também é acrescida a componente de usabilidade.

No Brasil, os padrões que tratam da qualidade de dados geoespaciais estão definidos pela ET-CQDG (DSG, 2015) (conforme já mencionado) e no Manual Técnico em Geociências do IBGE referente a Avaliação da Qualidade de Dados Geoespaciais, com sua primeira versão desenvolvida em 2017 e segunda versão em 2019. Ambos os padrões citados se baseiam fortemente nas definições dos indicadores estabelecidos pela ISO 19157 (ISO, 2013) para a definição dos procedimentos a serem adotados. A seguir são abordadas, com maiores detalhes, as normatizações utilizadas no Brasil para a avaliação da qualidade de dados geoespaciais.

2.2.2.1 Avaliação da Qualidade de Dados Geoespaciais no Brasil

No Brasil, a primeira documentação que normatizou as questões referentes à qualidade de dados em produtos cartográficos referiu-se ao Decreto nº 89.817 - Normas Técnicas da Cartografia Nacional (BRASIL, 1984). Nesta foram descritas questões referentes a acurácia posicional e definiu-se critérios referentes a exatidão e distribuição de erros, com base em um indicador estatístico da qualidade posicional, denominado de Padrão de Exatidão Cartográfica (PEC). O PEC é dividido em três classes, denominadas de A, B e C, onde na classe A existe um maior rigor na avaliação do produto cartográfico e na classe C um menor rigor. Para cada classe está estabelecido um valor de PEC e, cada valor está associado a um Erro Padrão (EP). Este EP corresponde a 60,8% do valor do PEC, que está associado com a probabilidade de 90% da distribuição normal. Os valores são estabelecidos tanto para a avaliação planimétrica quanto altimétrica.

Em 2011, foi criada pela DSG a Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais - ET-ADGV 2.1.3 (DSG, 2011). Nesta especificação foram atualizados os padrões de acurácia posicional previstos no Decreto Lei nº 89.817 (Brasil, 1984), uma vez que, a mesma já não atendia as necessidades atuais aplicadas em meio digital. Foi estabelecido na ET-ADGV (DSG, 2011) um novo indicador estatístico, o Padrão de Exatidão Cartográfica para Produtos Cartográficos Digitais (PEC-PCD).

O PEC-PCD possui as mesmas premissas do PEC, contudo adicionou-se uma nova classe e os produtos cartográficos passaram a ser classificados em A, B, C e D. O PEC-PCD está associado com a escala de um determinado produto cartográfico e permite com que este seja classificado a partir do erro máximo obtido em discrepâncias que compõem uma amostra de pontos. Trata-se de um indicador estatístico onde, a partir do gráfico probabilístico de distribuição normal, considera-se 90% de confiança. A Tabela 1 apresenta os valores estabelecidos no PEC e no PEC-PCD planimétrico, assim como, os EPs associados aos mesmos.

TABELA 1 – PADRÃO DE EXATIDÃO CARTOGRÁFICA PARA PRODUTOS CARTOGRÁFICOS.

PEC (1984)	PEC-PCD (2011)	PEC Planimétrico	Erro Padrão Planimétrico
	A	0,28 mm x Denominador de Escala	0,17 mm x Denominador de Escala
A	B	0,5 mm x Denominador de Escala	0,3 mm x Denominador de Escala
B	C	0,8 mm x Denominador de Escala	0,5 mm x Denominador de Escala
C	D	1,0 mm x Denominador de Escala	0,6 mm x Denominador de Escala

FONTE: Adaptada da ET-ADGV 2.1.3 (DSG, 2011).

Conforme mostra a Tabela 1, as classes A, B e C do PEC passaram a ser denominadas de B, C e D do PEC-PCD, respectivamente, sendo acrescida uma nova classe A ao mesmo. Tal modificação está associada com o maior rigor estabelecido na geração de produtos cartográficos digitais. É possível notar também, que os valores estão associados com a escala pretendida, ou seja, para encontrar o valor do PEC e o EP a uma dada escala, multiplicam-se os valores da tabela 1 pelo seu denominador e prossegue-se com as análises.

No ano de 2015, foi criada pela DSG a Especificação Técnica para Controle de Qualidade de Dados Geoespaciais - ET-CQDG (DSG, 2015) que abrange metodologias para a avaliação da qualidade em produtos cartográficos. No que diz respeito a acurácia posicional, a ET-CQDG (DSG, 2015) prevê que sejam calculadas as Distâncias Euclidianas (DE) de uma amostra de pontos homólogos entre o produto cartográfico referência e o avaliado. A partir da obtenção das DE, para encontrar a classificação do produto cartográfico em uma dada escala, são avaliadas duas condições: se 90% das DE são menores ou iguais ao valor do PEC-PCD; e se o Erro Médio Quadrático (EMQ) é menor ou igual ao EP. Além da Acurácia Posicional, a ET-CQDG também prevê os indicadores de Acurácia Temática, Completude e Consistência Lógica para a avaliação da qualidade em dados geoespaciais.

Outro documento que trata sobre a qualidade de dados geoespaciais no Brasil refere-se ao Manual Técnico em Geociências do IBGE, onde sua primeira versão foi publicada no ano de 2017 e a segunda versão no ano de 2019 (IBGE, 2019). IBGE (2019) aborda que a concepção dos conceitos associados com a

qualidade de dados geoespaciais, além de considerar os elementos necessários para a sua apresentação, é necessário considerar conceitos associados com as perspectivas de produtores e usuários de dados geoespaciais e a adoção de normas internacionais em nível nacional.

No contexto apresentado é importante ressaltar, mais uma vez, que tanto a ET-CQDG (DSG, 2015) quanto o Manual Técnico em Geociências (IBGE, 2019) apresentam as premissas dos indicadores de qualidade pautadas na ISO 19.157 (ISO, 2013) e possuem critérios semelhantes para a amostragem de dados para proceder com as análises estatísticas referentes a qualidade. IBGE (2019) ainda apresenta a usabilidade como um dos indicadores para a avaliação da qualidade de dados geoespaciais. Na subseção a seguir são apresentadas as definições dos indicadores de qualidade citados.

2.2.2.2 Indicadores de qualidade para avaliação da qualidade de dados geoespaciais

Para a conceituação dos indicadores desta subseção foram apresentadas definições apresentadas na ISO 19.157 (ISO, 2013), na ET-CQDG (DSG, 2015) e no Manual Técnico em Geociências (IBGE, 2019). Contudo, antes de proceder com as definições propriamente ditas, é importante contextualizar a subdivisão abordada para os conceitos de qualidade com base em algumas abordagens trazidas na literatura.

Uma pesquisa realizada por Maranhão (2013) baseada nas definições de Devillers e Jeansoulin (2006) e de outros autores, apresenta a qualidade no âmbito dos dados geoespaciais a partir de dois grupos denominados qualidade interna e qualidade externa. A qualidade interna está associada a produtos que estão isentos de erros e, por sua vez, a qualidade externa está relacionada com produtos cartográficos que atendem à necessidade do usuário.

No que diz respeito a qualidade interna, Maranhão (2013) explica que esta é baseada em estimar a ausência de erros nos dados geoespaciais a partir das descrições incluídas em seus metadados. A autora ainda comenta que as descrições incluem as escalas, resoluções espaciais e temporais, data de coleta, métodos de aquisição, entre outros. É importante considerar que a avaliação da

qualidade interna é realizada a partir de indicadores definidos por normas e especificações e que, apesar de diferentes países adotarem diferentes métodos para a realização das análises, existe um consenso quanto aos parâmetros que devem ser considerados na avaliação da qualidade de dados geoespaciais. Nesta subseção são descritos os indicadores utilizados pela ISO 19157 (ISO, 2013).

A qualidade externa está diretamente relacionada com a aderência de um determinado produto cartográfico às necessidades de um usuário ou grupos de usuários. Quanto aos elementos utilizados para a avaliação da qualidade externa de dados geoespaciais, uma pesquisa realizada por Maranhão (2013), a partir de conceitos trazidos por Bedard e Vallière (1995), apresenta os parâmetros de definição, cobertura, linhagem, precisão, legitimidade e acessibilidade. É importante comentar que, neste caso, a precisão representa o quanto um determinado dado geoespacial é aceitável para uma determinada necessidade ou aplicação.

No que diz respeito aos indicadores para a avaliação da qualidade interna de dados geoespaciais, conforme já apresentado, estes são definidos pela ISO 19157 (ISO, 2013) e também utilizados nas especificações brasileiras quanto à temática da qualidade de dados geoespaciais (IBGE, 2019; DSG, 2015). Nos próximos parágrafos estes indicadores são definidos.

A completude dos dados indica excesso ou omissão de feições existentes em uma determinada base de dados (IBGE, 2019; DSG, 2015). Na ET-CQDG (DSG, 2015) a avaliação da completude é referenciada a partir de duas análises, a primeira é relacionada com a presença de objetos em uma determinada base de dados em relação às especificações responsáveis pela aquisição e a segunda análise é dada a partir da ausência dos elementos. A documentação quanto a completude em um produto cartográfico geralmente é dada por meio da apresentação percentual de feições em excesso ou omissas a partir de uma amostra de feições.

A consistência lógica é caracterizada pelo cumprimento de regras lógicas de uma determinada estrutura de dados, atributos e a suas relações (DSG, 2015). Conforme é apontado pelo IBGE (2019), este indicador tem o intuito de avaliar a integridade estrutural de um conjunto de dados, a partir de análises da sua aderência em relação aos modelos de dados geoespaciais utilizados. O IBGE (2019) e a ISO 19157 (ISO, 2013) subdivide a consistência lógica em elementos:

- **Consistência conceitual:** Refere-se à aderência de um determinado conjunto de dados geoespaciais em relação ao universo de discussão, ou seja, as regras que são definidas pelo modelo conceitual;
- **Consistência de domínio:** Trata da avaliação da conformidade em relação a valores de domínios estabelecidos previamente;
- **Consistência de formato:** Nesta análise é considerado o grau em que os dados são armazenados conforme a estrutura física do conjunto de dados. É examinado o modelo de armazenamento para uma determinada estrutura de dados em hardware ou software;
- **Consistência topológica:** É caracterizada pelos aspectos geométricos e topológicos dos dados geoespaciais, como situações de adjacência e pertinência.

A Acurácia Posicional refere-se à obtenção das discrepâncias, em coordenadas planimétricas e altimétricas, das feições de um produto cartográfico em relação a um dado de referência, onde a sua qualidade é conhecida. Conforme já foi discutido no início da subseção 2.2.2, a avaliação da acurácia corresponde ao primeiro indicador utilizado por especialistas em cartografia para a validação de projetos de mapeamento. O IBGE (2019) e a ISO 19157 (ISO, 2013) definem a acurácia posicional a partir de três elementos:

- **Acurácia Posicional absoluta ou externa:** refere-se às discrepâncias de coordenadas observadas em um determinado produto cartográfico em relação as suas coordenadas homólogas, ditas verdadeiras;
- **Acurácia Posicional relativa ou interna:** refere-se a discrepâncias de coordenadas observadas em um determinado produto cartográfico em relação a outras feições no conjunto de dados, tidas como verdadeiras;
- **Acurácia Posicional dos dados em grade:** É definida pela discrepância de dados (estrutura matricial) com o valor tido como verdadeiro.

A Acurácia Temática é caracterizada pela interpretação correta de feições e atributos, bem como, o correto enquadramento nas classes que descrevem o seu modelo conceitual (DSG, 2015). Este indicador indica a acurácia dos atributos de um conjunto de dados geoespaciais em relação ao seu significado real no terreno (IBGE, 2019). O IBGE (2019) e a ISO 19.157 (ISO, 2013) atribuem três elementos para a avaliação da acurácia temática:

- Correção da classificação: Comparação da representação da classe de feições ou seus atributos com a realidade no terreno;
- Correções dos atributos não quantitativos: É obtida a partir de verificações de atributos que não podem ser contados ao realizar as análises comparativas em relação a atributos das mesmas feições na fonte de maior precisão (DSG, 2015). O IBGE (2019) exemplifica avaliação da acurácia temática de dados não quantitativos a partir da ocorrência do nome de um rio preenchido incorretamente;
- Correção dos atributos quantitativos: É obtida a partir da verificação de atributos que podem ser contados ao se realizar as análises comparativas em relação a atributos das mesmas feições na fonte de maior precisão (DSG, 2015). O IBGE (2019) exemplifica a avaliação da avaliação da acurácia temática de dados qualitativos a partir da descrição do número de pistas em um trecho rodoviário e a altitude de uma pista ou ponto de pouso (metros).

A Acurácia Temporal está relacionada com a referência à exatidão encontrada no componente temporal dos dados geoespaciais (DSG, 2015). Conforme é abordado pelo IBGE (2019) baseado em Ariza-López (2002) ao considerar o tempo para gerenciar uma determinada informação é necessário considerar o período em que ocorreram as mudanças no mundo real, o tempo da observação da evidência e o tempo em que as mudanças são descritas no banco de dados. O IBGE (2019) e a ISO 19.157 (ISO, 2013) apresentam a acurácia temporal dividida em três elementos:

- Acurácia de uma medida temporal: Refere-se a correta referência temporal do item avaliado. Dessa forma, mesmo que um dado geoespacial esteja disponível em um determinado período, é importante que conste a época em que o mesmo foi produzido;
- Consistência temporal: Refere-se a correta ordem de alterações a partir de uma sequência de fatos, quando relatado;
- Validade temporal: É caracterizada pela validade dos dados em relação a um determinado período, ou seja, um atributo de uma feição pode estar vinculado a uma determinada data ou periodicidade.

A Usabilidade é caracterizada pelo aspecto da qualidade que é baseado em requisitos específicos de usuários em relação a um determinado dado (IBGE, 2019).

Conforme é apontado pelo IBGE (2019) e na ISO 19.157 (ISO, 2013) todos os demais indicadores para a avaliação da qualidade de dados geoespaciais podem ser utilizados na usabilidade, visto que este descreve aspectos específicos da qualidade relacionados com a adequação de um determinado conjunto de dados em relação a sua aplicação ou conformidade de acordo com requisitos estabelecidos.

3. QPEC: QGIS TOOLKIT FOR EVALUATING GEOSPATIAL DATA POSITIONAL ACCURACY ACCORDING TO THE BRAZILIAN CARTOGRAPHIC ACCURACY STANDARD

QPEC: CONJUNTO DE FERRAMENTAS NO QGIS PARA AVALIAÇÃO DA ACURÁCIA POSICIONAL DE DADOS GEOESPACIAIS DE ACORDO COM O PADRÃO DE EXATIDÃO CARTOGRÁFICA BRASILEIRO

Autores: Elias Nasr Naim Elias, Samoel Gihel, Fabricio Rosa Amorim, Marcio Augusto Reolon Schmidt, Silvana Phillippi Camboim, Vivian de Oliveira Fernandes.

Periódico: Artigo aceito para publicação no periódico Anuário de Geociências (2022).

Abstract:

This paper presents the development of a QGIS plugin to support evaluating the planimetric positional quality for point and linear features based on the metrics established by Brazilian legislation. For this purpose, we used the QGIS environment Graphical Modeler, which consists of an interface to concatenate a series of processes into a single algorithm. The set of tools, called QPEC, allows for performing the statistical tests from the automatic identification of the sample size and discrepancies. In order to demonstrate the implemented functionalities, a case study was carried out. In this illustrative example, the vector files from the Cartographic and Cadastral System of the Municipality of Salvador - BA (SICAD) were the reference data, and their homologous OpenStreetMap (OSM) features were the analysed database. The results obtained are presented in the attributes table. In addition, the spatial distribution of the discrepancies is visualised through the visual variable colour value in a quartile classification. The creation of this toolset corroborates the feasibility of developing more visual, automated and complete

interfaces to support users of geospatial data in analysing the quality of the information available, especially when it involves free applications with open-source code.

Keywords: Planimetric Positional Accuracy; QGIS Plugin; Python

Resumo:

Este documento apresenta o desenvolvimento de um plugin de QGIS para apoiar a avaliação da acurácia posicional planimétrica para feições pontuais e lineares, com base nas métricas estabelecidas pela legislação brasileira. Para este fim, utilizamos o Modelador Gráfico no ambiente QGIS, que consiste numa interface para concatenar uma série de processos num único algoritmo. O conjunto de ferramentas, denominado QPEC, permite realizar os testes estatísticos a partir da identificação automática do tamanho da amostra e das discrepâncias. A fim de demonstrar as funcionalidades implementadas, foi realizado um estudo de caso. Neste exemplo ilustrativo, os arquivos vetoriais do Sistema Cartográfico e Cadastral do Município de Salvador - BA (SICAD) foram os dados de referência, e as suas feições homólogas no OpenStreetMap (OSM) foram a base de dados analisada. Os resultados obtidos são apresentados na tabela de atributos. Além disso, a distribuição espacial das discrepâncias é visualizada através da variável visual valor de cor variável em uma classificação em quartis. A criação deste conjunto de ferramentas corrobora a viabilidade do desenvolvimento de interfaces mais visuais, automatizadas e completas para apoiar os usuários de dados geoespaciais na análise da qualidade da informação disponível, especialmente quando envolve aplicações gratuitas com código-fonte aberto.

Palavras-chave: Acurácia posicional planimétrica; *Plugin* para *QGIS*; *Python*.

3.1 INTRODUCTION

The union, states, and municipalities need geospatial information to represent the territory. Unfortunately, many cartographic products are delivered with an unclear quality due to non-adherence to standards and legislation and the immediate use perspective, thus opposing the reuse of geospatial data. The lack of experience and knowledge on the quality issue reinforces these. Most of the time,

quality assurance occurs when professionals use data in their activities and compare it with the geospatial reality (Silva Júnior, França & Moraes 2022; Pedreira, Oliveira & Santos 2020; Luz & Antunes 2015).

In territorial planning, the execution or management of any public or private project, it is essential to work with consistent, organised, and current information. In public institutions, Cartography provides information for public policies governed by legislation, standards, and technical specifications that guarantee the reliability of cartographic products. In Brazil, for example, the Brazilian standard for assessing the positional accuracy of geospatial data is given by Decree-Law nº 89,817 of 1984 (Brasil 1984). This standard was called the Cartographic Accuracy Standard (PEC). It was updated to Cartographic Accuracy Standard for Digital Cartographic Products (PECPCD) in one of the topics of the Technical Specification for the Acquisition of Vector Geospatial Data - ET-ADGV (DSG 2011), created by the Brazilian Army Geographic Service Directorate (DSG) in 2011. In 2016, DSG developed the Technical Specification for Quality Control of Geospatial Data - ET-CQDG (DSG2016), which establishes methodologies for assessing the quality of geospatial data based on the International Organization for Standardization (ISO) 19157 (ISO 2013). Further details regarding the PECPCD are presented in section 2.

ISO 19157 (ISO 2013) assigns the indicators for evaluating geospatial data. These present the following indicators: Positional Accuracy, Thematic Accuracy; Completeness; Logical Consistency, Temporal Quality, and Usability. Among these indicators, positional accuracy is the main characteristic to be considered as it refers to the ability of a device, or a method, to estimate the position (latitude and longitude) of a feature on the earth's surface. (Roberto 2013). Furthermore, the presence of random errors (precision) and systematic errors (trends) in the measurements is taken into account (Monico et al. 2009).

It is necessary to recognise the impossibility of making any free-of-error cartographic product. User satisfaction is also a challenging and laborious dimension to measure. With the popularisation of the Global Navigation Satellite System (GNSS) and the access to maps through digital globes such as Google Earth, among others, cartographic information has been widely used in the development of applications and incorporated into a series of tools and systems. In recent years the need to know the quality of the products has increased. The automation of processes expands the range of uses of geoinformation and, in some ways, helps to minimise

some types of errors, but on the other hand, it may assume previous knowledge about the data sources and the various processes applied to them that are not so common among all users.

Geospatial data quality is essential in map production, whether for general purposes or to represent thematic aspects. In the classical view, official map agencies generate reference maps to represent all visible features in the landscape (Keates 1973) and general-purpose territorial features in ways that serve a variety of uses and users through the simultaneous presentation of various themes (Kent 2009). This method was considered the primary way to produce maps until recently (Elwood, Goodchild & Sui 2012). However, computational technological advances have allowed anyone with a computer or smartphone with internet access to generate geoinformation (Ganapati 2011).

The increase in geospatial information volume, variety and velocity is associated with big data. This currently data-intensive landscape, known as Big Data, covers the flow of data in digital media from different sources, such as numerical modelling, smartphones, internet access, and social networks (Yang et al. 2017). In this aspect, Robinson et al. (2017) point out challenges regarding representation, user context, and updates that constantly occur rapidly. This aspect can be evidenced in Volunteered Geographic Information (VGI) platforms because the user's participation allows the data to be updated and made available in real time. This method is a supplementary alternative to the traditional availability of cartographic products, usually coming from the agencies responsible for national cartography.

Currently, there are several methods for obtaining geospatial data, ranging from topographic and aerial photogrammetric surveys to VGI. Thus, efficient treatment and integration in the volume of data become essential. In this scenario, methodologies arise to extract, integrate and evaluate the quality of the relevant information from this ever-increasing amount of geographic data (Brovelli et al., 2019). In the context of geospatial data quality, although methodologies with efficient statistical and theoretical approaches exist, the speed with which information is obtained in real-time and the availability of open data requires free software tools that assist users in these assessments. Furthermore, in the methods used to assess quality, criteria must be established to spatially represent the results and consider the different sources and heterogeneity of the data. Assessing the quality of geospatial

data is to investigate the veracity of the information through processes to measure and document divergences, including them, when possible, in cartographic products.

It is essential to point out that there are studies focused on positional quality that start from the establishment of different statistical criteria for carrying out the tests, the implementation of computational tools aimed at automating the process to validations performed on geospatial data obtained from different sources. Among these studies, the works by Lisboa et al. (2019); Cunha et al. (2019); Santos et al. (2016); Cintra & Nero (2015), Galo & Camargo (1994) and Merchant (1982). The theme of these works concerns the establishment of different statistical methods for determining positional accuracy based on sample sets. Furthermore, concerning implementing tools for the automatic assessment of positional quality, it was noted that there are efforts to verify data from VGI. This characteristic can be evidenced in works such as Mozas-Calvache & Ariza-López (2019); Martini, Kuper & Breunig (2019); Brovelli & Zamboni (2018); Zhou (2018) and Brovelli, Minghini & Moliari (2016). Within the scope of geospatial data services available on the web, there is the research by Silva Júnior, França, and Moraes (2022) that evaluated the planimetric positional accuracy of Google Earth images. In research by Pedreira, Oliveira, and Santos (2020) and Pessoa et al. (2020), for example, the altimetric positional accuracy of a Digital Elevation Model (DEM) generated from a UAV was evaluated respectively. The positional error of Digital Surface Models (DSM) obtained in the post-processing of data acquired with UAV concerning the number and distribution of control points on the ground.

Regarding developing applications in geospatial data quality, extensive research has focused on creating open-source tools to compare cartographic products. The objective is to evaluate the positional discrepancies and data behaviour through statistical tests and reported errors. Garcia-Balboa, Ureña-Câmera & Xavier (2021), for example, developed a web application, ICPos, in which it is possible to evaluate the positional accuracy based on the insertion of control points using different quality standards. As highlighted by the authors, in ICPos, it is possible to calculate and automatically create a complete report for the input data according to the user's specifications. Besides ICPos, researchers such as Zanetti et al. (2016), Soares et al. (2018) and Barbosa, Meirelles & Santos (2021) evaluated the positional accuracy of features from the GeoPEC software. GeoPEC, developed by researchers at the Federal University of Viçosa, Brazil, allows the evaluation of

the positional accuracy of cartographic products based on the PECPCD from the insertion of coordinates of point features. It also allows for statistical tests associated with sampling, normality and identification of data error trends.

Santos (2015) and Santos et al. (2016) developed models in the ArcGIS Model Builder tool to obtain discrepancies between linear features from different methods. Teixeira & Santos (2019) developed similar applications in QGIS and compared the magnitude of the discrepancies obtained. Foz et al. (2021) also implemented quality control processes in cartographic series updating from GIS and Computer Aided Design (CAD) data combination.

It is possible to notice that the presented tools emphasise the discrepancies' calculation and statistical analysis of a specific sample. Unfortunately, complete GIS applications in this context are still scarce. It would be helpful in these solutions to enable the entry of layers and simultaneously display quality spatial behaviour maps and the attribute table classification values, according to the statistical analysis performed. In this way, this work aimed at developing QGIS 3.X software tools to evaluate the positional accuracy of punctual and linear features. For the point features assessment, it implements the methods of Merchant (1982) and Galo & Camargo (1994), combined with the conditions established by the ET-CQDG (DSG 2016). For the linear features, it applied the double buffer method developed by Santos (2015). Section 2 details the application of the explained methods.

3.2 METHODS FOR THE ASSESSMENT OF PLANIMETRIC POSITIONAL ACCURACY OF GEOSPATIAL DATA IN BRAZIL

In Brazil, the first documentation that addressed positional accuracy in cartographic products referred to Decree-Law nº 89,817 - Technical Norms of National Cartography (Brasil, 1984). Criteria for accuracy and distribution of errors were established based on a statistical indicator of positional quality called the PEC. The PEC classifies the products into three groups, called A, B, and C, where in class A, there is greater rigour in the evaluation of the cartographic product, and in class C, there is less rigour. The norm establishes the PEC value for each class, and each value is associated with a Standard Error (EP). The EP of each class is 60.8% of the PEC ($EP = 0.608 \cdot PEC$), which corresponds to a 90% probability in the normal distribution. There are values for both planimetric and altimetric evaluation.

In 2011, the DSG created the ET-ADGV specification (DSG, 2011). The purpose of this norm is to update the standard in Decree-Law nº 89,817 (Brasil, 1984) since it no longer meets the current needs of digital mapping. In addition, the ET-ADGV (DSG 2011) established a new statistical indicator, the PEC-PCD. This indicator has the same premises as the PEC. In addition, however, a new class was added, and the cartographic products were classified into A, B, C, and D.

PEC-PCD is associated with the scale of cartographic products and allows them to be classified by the discrepancies found in a sample of points. It is a statistical indicator where, from the probabilistic graph of normal distribution, 90% confidence is considered. It is worth noting that the DSG has already developed two more recent versions of the ET-ADGV in 2015 and 2016, called the Technical Specification for the Acquisition of Geospatial Data Vector for Defense of the Land Force - ET-ADGV (DSG 2015a; DSG 2016) in which aspects associated with the PEC-PCD are also addressed. Table 2 presents the values established in the PEC, the planimetric PEC-PCD, and the Planimetric Standard Errors (EP) associated with them.

TABLE 2 – CARTOGRAPHIC ACCURACY STANDARD FOR CARTOGRAPHIC PRODUCTS.

PEC (1984)	PEC-PCD (2011)	Planimetric PEC	Planimetric Standard Error (E.P.)
	A	0,28 mm x Scale Denominator	0,17 mm x Scale Denominator
A	B	0,5 mm x Scale Denominator	0,3 mm x Scale Denominator
B	C	0,8 mm x Scale Denominator	0,5 mm x Scale Denominator
C	D	1,0 mm x Scale Denominator	0,6 mm x Scale Denominator

FONTE: Adapted from ET-ADGV (DSG 2011).

As shown in Table 2, classes A, B, and C of the PEC started to be called B, C, and D of the PEC-PCD, respectively, and a new class A was added. It is also possible to note that values are associated with the desired scale, that is, to find the value of the PEC and EP at a given scale, the values in Table 2 are multiplied by their denominator, and the analysis continues.

In 2015, the DSG created the ET-CQDG (DSG 2016), which covers methodologies for quality assessment in cartographic products considering similar indicators and definitions to those addressed by ISO 19157 (ISO 2013). Regarding positional accuracy, the ET-CQDG (DSG 2016) provides for calculating Euclidean Distances (DE) from a sample of homologous points between the reference cartographic product and the evaluated one. After obtaining the DE, to find the classification of the cartographic product on a given scale, two conditions are evaluated: whether 90% of the DE is less than or equal to the PEC PCD value and if the Root Mean Square Error (RMSE) error is less than or equal to the EP.

It is essential to highlight that, in addition to the procedures established for the evaluation of planimetric positional accuracy established in the ET-CQDG (DSG 2016), in the ET-ADGV (DSG 2011), and Brazil (1984), there are statistical procedures provided for in the literature, such for example, the method proposed by Merchant (1982). This method was adapted by Galo & Camargo (1994) so that there was a statistical connection with the PEC, allowing cartographic product classification based on trend and precision analysis. The Student t-test gives the trend analysis. It allows for identifying error trends in some of the cartographic products evaluated. Furthermore, it verifies whether the mean of a set of discrepancies can be considered statistically equal to zero (null hypothesis). In this context, the sample t-value based on the discrepancies of each sample feature is calculated by Equation 1 as follows:

$$(1) \quad t_x = \frac{\overline{\Delta X}}{S_x} \sqrt{n}$$

In equation 1, $\overline{\Delta X}$ and S_x , and n are equivalent to the discrepancies' mean, standard deviation and sample size. The analysis compares the calculated t-value to a tabulated one in t-student distribution. The tabulated t-value is based on the sample size and on its significance level "a/2", which has $(n - 1)$ degrees of freedom, as shown in equation 2:

$$(2) \quad |t_x| \leq t_{(n-1; \frac{a}{2})}$$

Thus, if the module calculated for the sample "t" (Equation 1) is less than or equal to the tabulated "t" value, the null hypothesis is accepted, i.e., the cartographic product can be considered free of significant trends. On the contrary, the null hypothesis is rejected if the inequality is not satisfied. Since the PEC PCD corresponds to a statistical indicator with a 90% confidence probability, the "t" value is calculated considering this confidence level.

The chi-square statistical test gives the precision analysis. It allows for comparing the variance obtained in a set concerning the EP of the PEC-PCD. The null hypothesis is that the variance is statistically smaller than the EP. This comparison allows for classifying the cartographic product according to Table 2. The EP is assumed as the resultant of the two planimetric components. Considering that the trend analysis is obtained for each component of the X and Y coordinate system, this value can be calculated from Equation 3.

$$(3) \quad \sigma_x = EP/\sqrt{2}$$

With the value obtained in Equation 3, the sample chi-square test is applied to verify if the discrepancies' standard deviation can be considered statistically equal or smaller than the assumed EP for a given PCD PEC class. Equation 4 gives the calculation of the sample chi-square:

$$(4) \quad X_x = (n - 1)S_x^2/\sigma_x^2$$

The output from the operation of Equation 4 is compared to the tabulated chi-square value to test whether the test's null hypothesis is satisfied. This comparison is made based on Equation 5.

$$(5) \quad X_x^2 \leq X_{(n-1;\alpha)}^2$$

The tabulated chi-square has (n - 1) degrees of freedom and confidence interval α . It is verified if the expression of Equation 5 is true within a dispersion indicator of 90% (Nogueira Jr. 2003; Galo & Camargo 1994). This procedure is

performed based on the expected EP for a given class of the PEC PCD. If the condition is not satisfied, the lower class is tested until its acceptance, and the sample can be classified.

Despite the time lag of the analysis established by Merchant (1982) to the criteria established by the ET-CQDG (DSG 2016), several studies to evaluate positional quality in geospatial data in Brazil still use its statistical tests for trend and precision. These tests show the trend of errors, their directions, and the variance around the sample set mean. An example is the work of Oliveira & Brito (2019). In addition, Santos (2015) states that several countries' cartographic standards base positional quality assessment on methodologies comparing homologous points. Nogueira Jr. (2003) and Pereira & Nero (2012) also identified this aspect.

According to Tong et al. (2013), studies evaluating the positional accuracy of linear features have increased in interest as an object of study by the scientific community. Studies carried out by Santos (2015) show that homologous lines between the reference cartographic product and the product to be evaluated are analysed to evaluate the positional quality of linear features. Also called Buffer Overlay Statistics, the double buffer method was proposed by Tveite&Langaas (1999). It concerns the application of the Error Band model to the reference vector elements and their counterparts to be tested. This model creates a range in the linear reference feature, characterised as the most likely position (Santos 2015). The model generated for the uncertainty range is the application of a buffer around the evaluated line segments to establish their positional uncertainty.

Santos (2015) establishes a sequence of methodological steps for evaluating the positional accuracy of linear features from the double buffer method in association with the parameters established by Decree-Law nº 89,817 (Brasil 1984). The steps for applying the double buffer method are as follows:

- a) Select a sample with n homologous test (LT) and reference (LR) lines;
- b) From the selected linear features, apply a buffer of size x in each line i of the LT and LR sample. The buffer size is used as the PEC PCD value for the scale and class used in the evaluation process, following Decree-Law nº 89,817 (Brasil, 1984);
- c) Calculate the buffer area in each LT;

d) Overlap the LT and LR buffers and calculate the value of A_F , which is the LR buffer area that does not intersect the LT buffer;

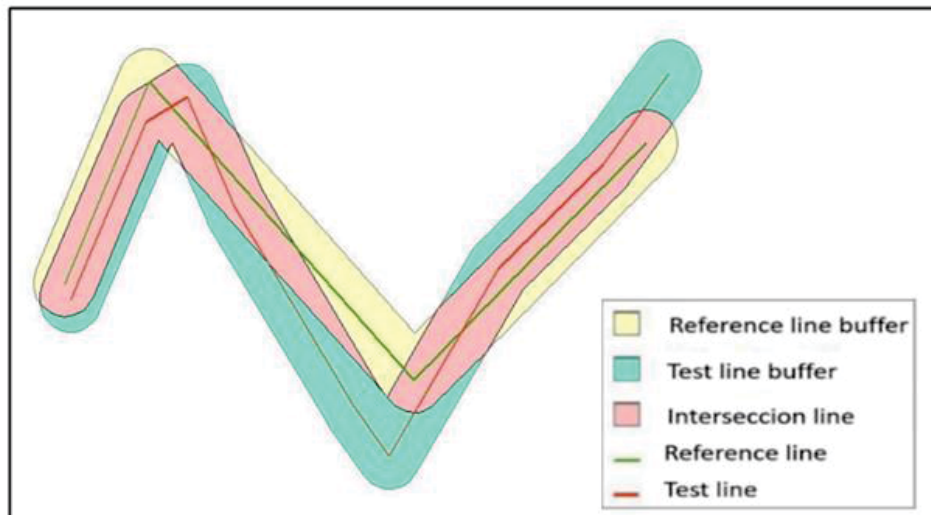
e) Calculate the Mean Discrepancy (DM) for each test line i by applying Equation 6;

$$(6) \quad DM_i = \pi \cdot x \cdot \frac{(\sum A_F)_i}{A_T}_i$$

f) For alignment with the parameters established by Decree-Law nº (Brasil 1984), it is evaluated if 90% of the tested lines have DM smaller or equal to the PEC PCD value (of the class and scale used to generate the buffer x), and also, if the RMSE of the DMs is smaller or equal to the EP value (also to the PEC PCD). The sample is classified according to the class and scale evaluated if both conditions are satisfied.

Figure 4 shows an example of the generation of polygons resulting from homologous linear features from applying the Double Buffer method.

FIGURE 4 – DOUBLE BUFFER METHOD



SOURCE: Double buffer method. Source Cruz and Santos (2016).

The conditions used are similar to those described in ET-CQDG (DSG 2016). Instead of the DE of each point feature, it calculates a Mean Discrepancy (DM). A mathematical operation divides the area of the reference buffer that does not intersect the evaluated buffer for the total area of the evaluated buffer. The result of

this division is multiplied by the size of the applied buffer and by π . The buffer size must be related to the value of the PEC-PCD on a given scale, allowing the analysis to proceed.

3.3 METHODOLOGY AND DATA

The present study developed two tools in QGIS 3.X software, one for assessing the positional accuracy of point features and the other for linear features. The tools were developed from pre-existing processes in QGIS through the Graphical Modeler, where it is possible to concatenate a series of operations in a single analysis and develop an interface. The model created is stored and can be exported in the python programming language. QGIS field calculation processes form the bulk of the procedures used. The attribute table of the analysed layer is populated with the result of area and discrepancy calculations and the conditional classification based on the available quality assessment methods. This conditional sorting was performed by scripts using a SQL-like language. The initial calculations and statistical analysis elements were derived from the entry features. The initial values are the X and Y coordinates for point features, used to calculate the main parameters: count, mean and stdev (Standard Deviation) and sqrt (Square Root).

Besides the mathematical operations, the buffer of each pair of features was obtained in the case of linear features. The intersections between their areas were sufficient for the statistical procedures. Figure 5 shows an example of a condition script used in constructing one of the tools.

FIGURE 5 – EXAMPLE OF INSERTING A CONDITION IN THE QGIS FIELD CALCULATOR USING THE MODEL BUILDER TO CONSULT INFORMATION REGARDING A GIVEN SAMPLE OF POINTS.

```

CASE WHEN count("ID")>30
THEN 'Nao Aplicavel'
WHEN (count("ID")-1)*((stdev("dif_E")^2)/((1.70/sqrt(2))^2)) <="qui_tabelado"
AND (count("ID")-1)*((stdev("dif_N")^2)/((1.70/sqrt(2))^2)) <="qui_tabelado"
THEN 'Classe A'
WHEN (count("ID")-1)*(((stdev("dif_E"))^2)/((3.0/sqrt(2))^2)) <="qui_tabelado"
AND (count("ID")-1)*(((stdev("dif_N"))^2)/((3.0/sqrt(2))^2)) <="qui_tabelado"
THEN 'Classe B'
WHEN (count("ID")-1)*(((stdev("dif_E"))^2)/((5.0/sqrt(2))^2)) <="qui_tabelado"
AND (count("ID")-1)*(((stdev("dif_N"))^2)/((5.0/sqrt(2))^2)) <="qui_tabelado"
THEN 'Classe C'
WHEN (count("ID")-1)*(((stdev("dif_E"))^2)/((6.0/sqrt(2))^2)) <="qui_tabelado"
AND (count("ID")-1)*(((stdev("dif_N"))^2)/((6.0/sqrt(2))^2)) <="qui_tabelado"
THEN 'Classe D' ELSE 'Rejeitado'
END

```

SOURCE: The Authors (2022).

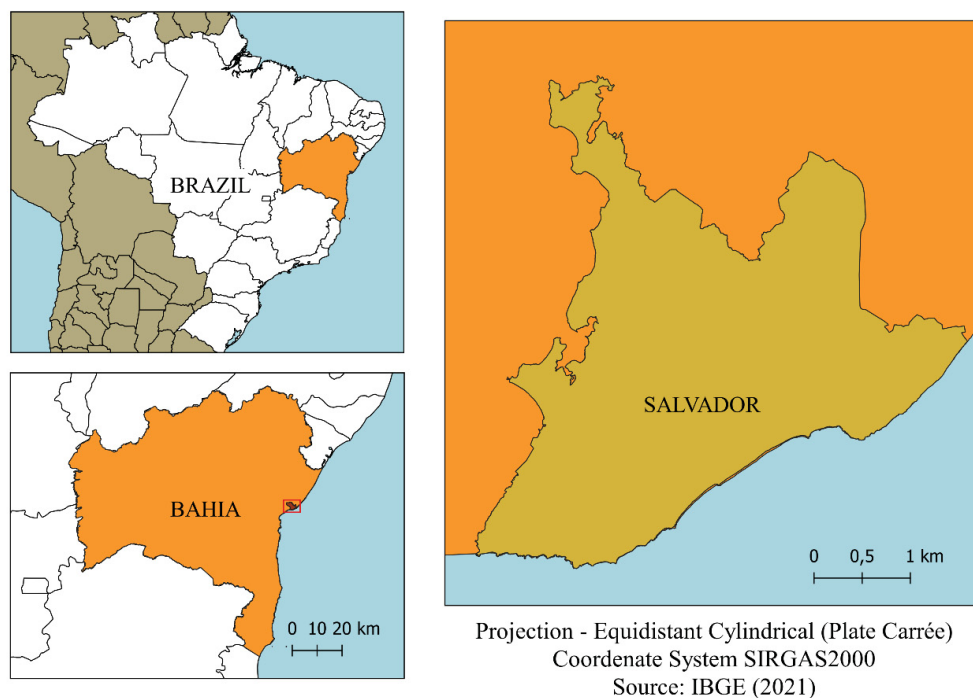
The example described in Figure 5 refers to applying the chi-square test for a sample of point features at a scale of 1:10,000. First, the result is calculated based on Equation 3, and the value found is compared with tabulated information of the referred test. Then, using conditionals "CASE", "WHEN", and "THEN", the PEC PCD classes were tested until all of them were accepted or rejected at the evaluated scale.

The methodological steps used in this work correspond to the statistical tests of Merchant (1982) and Galo & Camargo (1994) for the punctual features, as well as the application of the methodology proposed by the ET-CQDG (DSG 2016) to apply the PEC PCD. In addition, the double buffer method was applied to the linear features, as Santos (2015) and Santos et al. (2016) described. Both approaches were presented in section 3.2.

Sets of point and linear feature samples from Salvador-BA were used, with 20 features for each. Reference data were extracted from reference vector files from the 2006 Cartographic and Cadastral System of the Municipality of Salvador (SICAD). The homologous features evaluated corresponded to data obtained from the OSM collaborative mapping platform. The datasets were obtained circa April 2018 and integrated a series of analyses carried out by Elias & Fernandes (2021) for the OSM geospatial data quality assessment in Salvador-BA. According to the authors, the number of features available in the period exceeded 27,500. The choice

of sample size was based on criteria adopted by Merchant (1982), where 20 features in a given area are sufficient to assess cartographic product quality. It is worth mentioning that the ET-CQDG (DSG 2016) already establishes a series of advanced sampling criteria, which are based on ISO 19.157 (ISO 2013), where procedures are described as a function of the size of the area (lots). However, since this is not the focus of this work, the statistic considered by Merchant (1982) was adopted. Figure 6 displays the study area location map.

FIGURE 6 – LOCALISATION MAPS

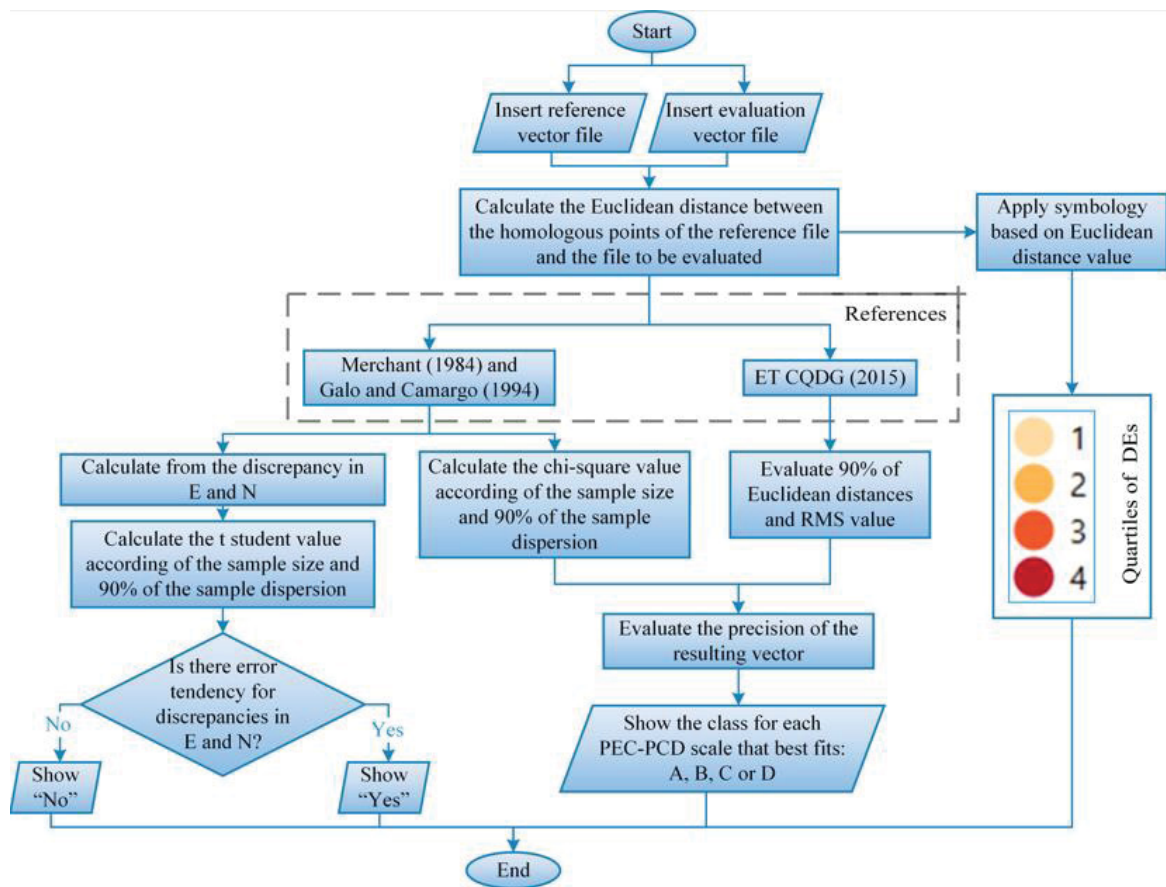


SOURCE: IBGE (2021).

3.3.1 Planimetric Positional Accuracy of Point Features

The methods used to assess the positional accuracy of point features corresponded to the statistical tests established by Merchant (1982) and Galo & Camargo (1994) and the constraints established in the ET-CQDG (DSG 2016). In addition, each of these methods had a specific flow developed. Also, before entering the data to perform the computations, an initial step that must be performed is the selection of feature samples — the tool analyses each planimetric component of the coordinates pair evaluated. Figure 7 shows the flowchart for assessing positional accuracy in point features.

FIGURE 7 – OPERATIONAL FLOWCHART FOR ASSESSING POSITIONAL ACCURACY IN POINT FEATURES USING THE METHODS OF MERCHANT (1982); AND GALO & CAMARGO (1994).



FONTE: The Authors (2022).

As shown in Figure 7, the vector layer of the reference point features and their respective counterparts to be evaluated must be previously inserted in the QGIS software in metric coordinates. Since these are point features, this configuration allows us to obtain the discrepancies, calculate the DEs in meters, and apply the PEC PCD definitions. Reference files can be, for example, points obtained from mapping techniques or indirectly from a cartographic product whose positional quality is known. In addition, reference and evaluated files must be associated with the same "ID" value in the Attributes Table to perform operations on homologous features.

The method of Merchant (1982) and Galo & Camargo (1994) evaluates the existence of a trend in the North and East directions. The sample size of each input file of 30 features follows the valid recommendations for the Student t-test, which,

according to Barbetta (2005), this test is ideal for small samples. Oliveira & Brito (2019) also considered this aspect in the analysis. For samples bigger than 30, the message "Not Evaluated" will be displayed, and the test will not be performed.

This threshold used for the sample size was established according to the premises of the mentioned statistical method. In samples where the number of elements is greater than 30, for instance, the z test (Normal Distribution) must be applied, which will be implemented in the future to complement the tools developed. Besides, it is vital to comment that the minimum number of 20 elements refers to the quantity established by Merchant (1982) for evaluating the positional accuracy of cartographic products. This value was established in the sample selection to validate the developed tools that could be higher or lower based on the user's needs. As discussed above, if the t Student test is chosen and the sample size exceeds 30, the statistical analysis will not be performed.

After meeting the sampling criterion, the tools calculate the discrepancies between the coordinates in the East and North directions, the DE, and the values referring to the t Student and Chi-Square. These values are necessary for testing the conditions, verifying trends in the features, and classifying the cartographic product according to each PEC PCD scale.

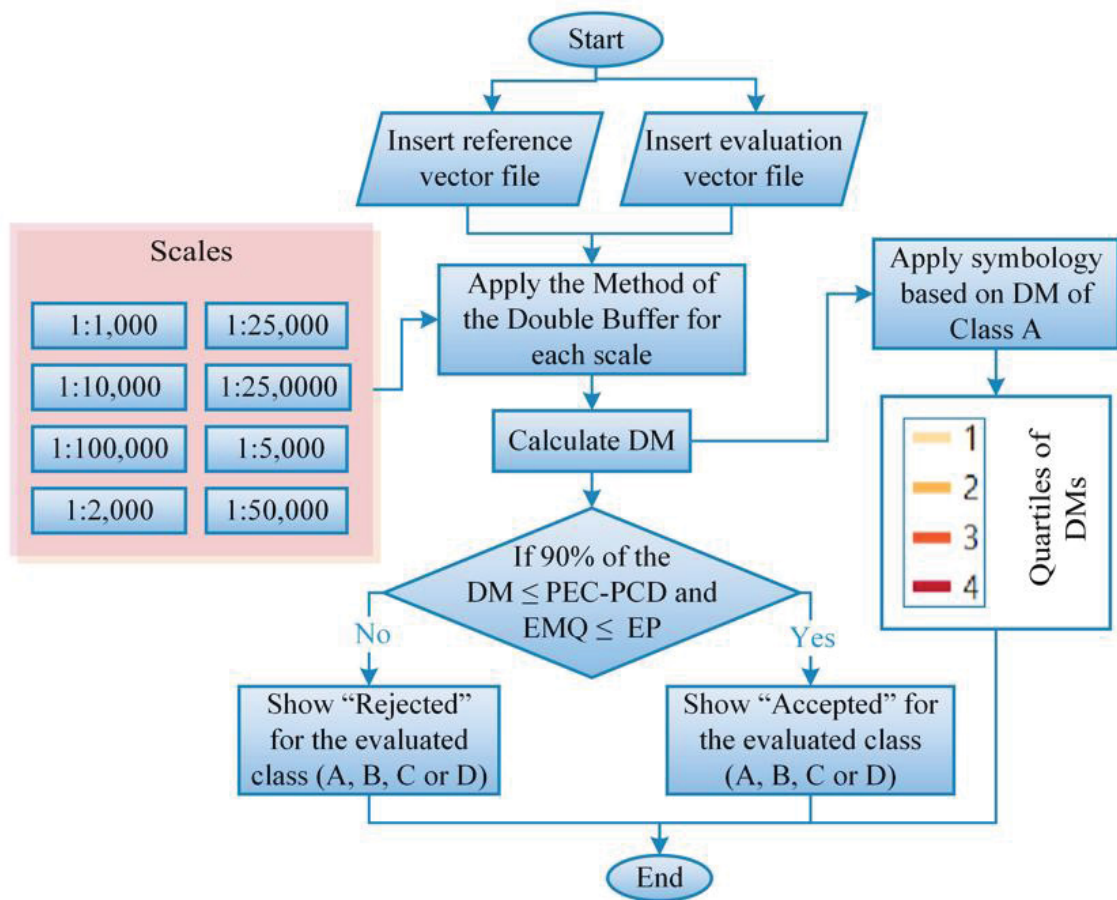
In the ET-CQDG (DSG 2016), the sampling premises are based on the total number of features available and the breakdown of the study area into regularly spaced grid squares. In addition, the standard provides a maximum threshold of unaccepted features for a given sample to be accepted. However, the statistical approach of the PEC PCD, different from what is described for the Student t-test, is not restricted to a minimum sample size. For this reason, the developed application does not establish a threshold for the number of features inserted when choosing this method. Finally, complying with described requirements, the script calculates the SD for each feature and uses the results to compute the samples' RMSE.

In the next step, the algorithm tests two conditions for each PECPCD class and scale: if 90% of the Euclidean distances are smaller than the PEC and if the RMSE value is smaller than the EP. If these conditions are met, the routine displays the class in which the samples conform to the desired scale. Otherwise, the value displayed is: "rejected".

3.3.2 Planimetric Positional Accuracy of Linear Features

The methodology for evaluating positional accuracy in linear features was based on the double buffer method, presented in item 1. The basic principle consists of creating buffers in linear features with a size equal to the values associated with the PEC PCD in the scales established by the ET-ADGV (DSG 2011) (1:1,000, 1:2,000, 1:5,000, 1:10,000, 1:25,000, 1:50,000, 1:100,000 and 1:250,000). The user can then choose the scale with which to determine the PEC PCD, and when running the process, each class displays the result of the conditionals that are part of the algorithm. It is relevant to mention that the scales used here are the standard ones of the Brazilian systematic mapping, addressed in the ET-CQDG (DSG 2016) and ET-ADGV (DSG 2011). Figure 8 shows the tools' functions for assessing linear features' positional accuracy.

FIGURE 8 – OPERATIONAL FLOWCHART OF THE TOOLS FOR THE ASSESSMENT OF POSITIONAL ACCURACY IN LINEAR FEATURES ACCORDING TO THE SCALES ESTABLISHED IN THE ET-ADGV.

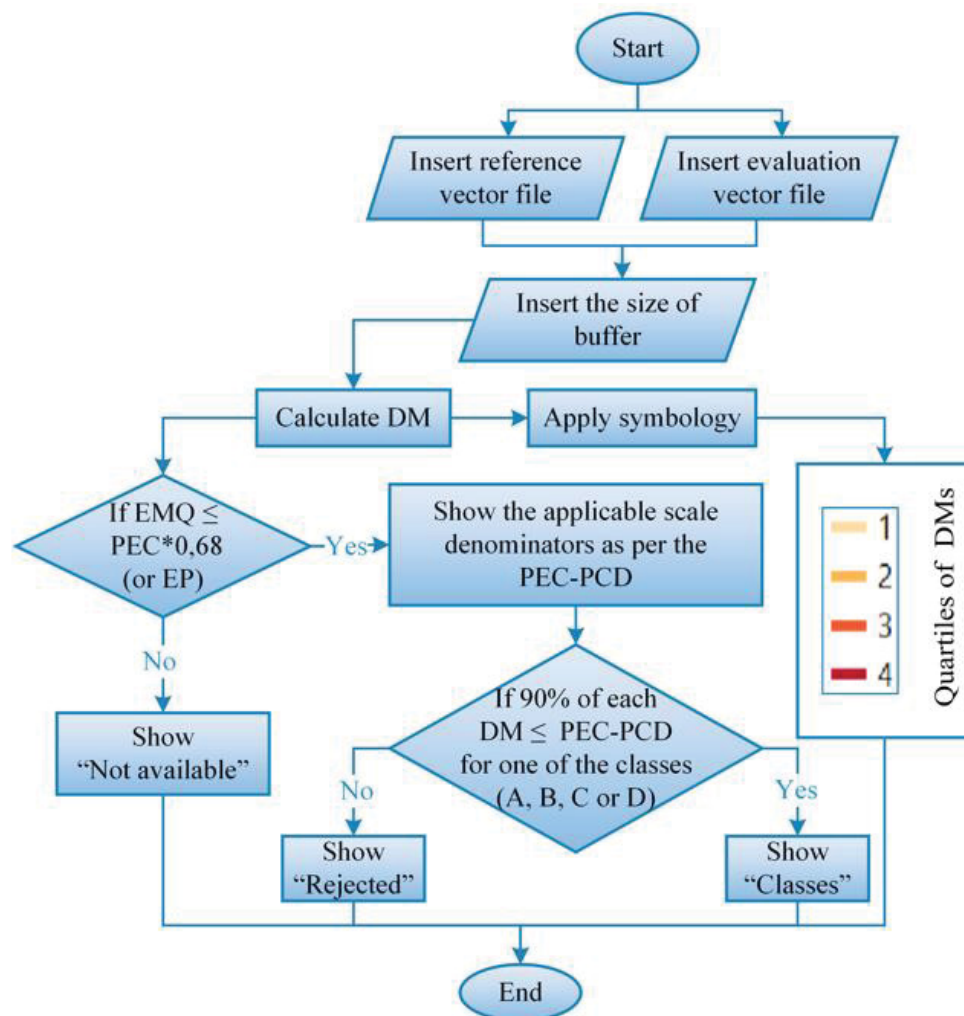


SOURCE: The Authors (2022).

As shown in the flowchart in Figure 8, the user must indicate the desired scale and the vector layers with the linear reference features and those under evaluation. Such layers must be in projected coordinates, and homologous line pairs must have corresponding ID values. Such layers must be in projected coordinates, and homologous line pairs must have corresponding ID values. The input data conditions are the same as those established for point features. The choice of scale allows us to apply the buffer to each linear feature and calculate the DM. The DM is calculated for each PEC-PCD class on the chosen scale. For each class, the conditions set for the method are tested. If 90% of the DM is less than the PCD PEC, and if the RMSE value is less than the EP, the value "Accepted" will be registered. Otherwise, it will be "Rejected".

In addition to evaluating the positional linear feature quality from the choice of one of the PEC-PCD scales, we have implemented a model using the double buffer method. In this application, the user supplies the value of the buffer to be evaluated, and then the PEC-PCD scales and classes are obtained relative to the supplied value. The flowchart in Figure 9 shows how this tool works.

FIGURE 9 – FUNCTIONING OF THE TOOL FOR EVALUATING THE POSITIONAL ACCURACY OF LINEAR FEATURES FROM INSERTING THE BUFFER SIZE.



SOURCE: The Authors (2022).

In this tool, in addition to inserting reference and evaluated vector files, the desired buffer size is also inserted for performing the analyses, as presented in section 2 above. In this case, it is essential to note that the buffer distance represents the user's desired tolerable error for the verification. This value enables the calculation of the range of standard scales of the PEC-PCD will be considered. For example, if a value of 20 meters is entered, the possible scales are 1: 25,000 and 1:

50,000. Once the scales are selected, whether the calculated RMSE is in the EP range of each scale is evaluated. If it does, it is possible to proceed with the remaining steps.

After verifying if the buffer entered by the user conforms to any test scale, the next step is to check if there is any PEC PCD class (A, B, C or D) in which the sample is accepted. Otherwise, the message "Not applicable" is displayed. Then, the DM of each feature is calculated and tested. This step verifies if 90% of DM's are less or equal to each class' value in the PEC PCD. This analysis is performed simultaneously with the RMSE verification; it is also necessary to be less than or equal to the EP of the evaluated class. As previously presented, this metric is associated with the PEC PCD EP corresponding to this percentage. If this condition is met, the class found will be shown. Otherwise, the message "Rejected" will be displayed.

Another significant characteristic of these tools is the ability to generate cartographic representations to visualise the results of the analysis of the quality parameters. In this aspect, the flowcharts of Figures 7, 8 and 9 show that the representation of the positional accuracy applied to the result uses a sequential colour scheme based on the values of DE or DM (in the case of linear features), classified into quartiles. The tools with the developed codes are stored on GitHub (Elias 2022; Elias 2019).

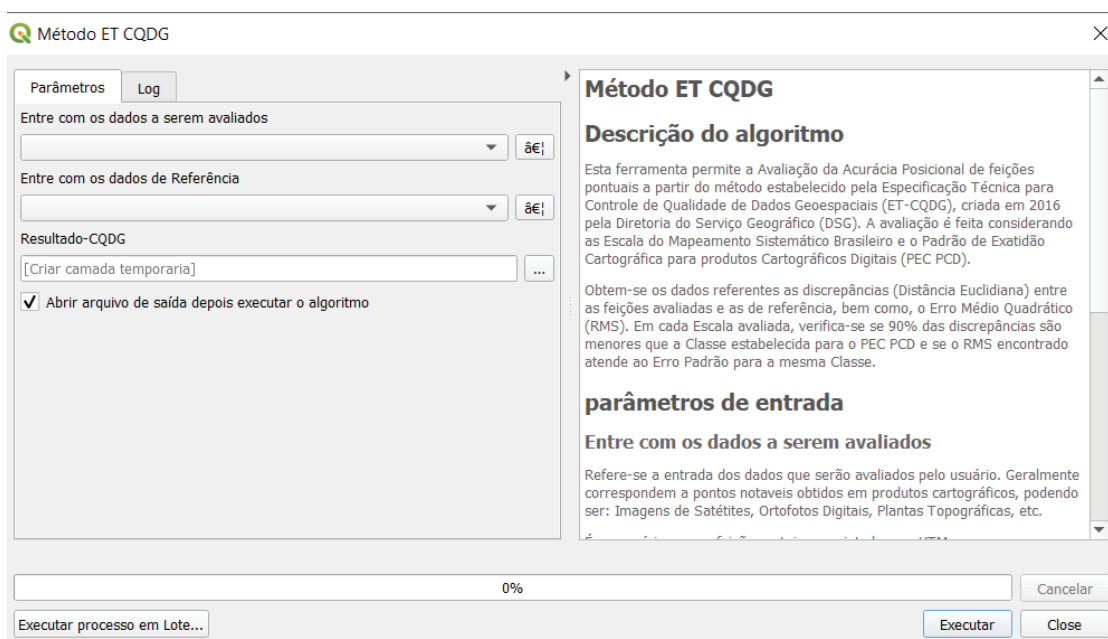
3.4 RESULTS AND DISCUSSION

From the methodology procedures applied, the tools developed show to be promising since the positional accuracy of punctual and linear features in QGIS can be obtained, selecting the desired statistical test or analysis and visualising the behaviour of discrepancies in a given region. Therefore, in addition to the results obtained for the evaluated samples, application-related aspects are provided in this item. Figure 10 shows one of the interfaces running on QGIS.

This interface evaluates the ET-CQDG (DSG 2016) method, so the user can insert the reference and evaluated vector files and save the resulting vector file. In addition, a brief description of the algorithm was also inserted to understand the process better. It is essential to highlight that the interfaces built for the other methods presented a structure similar to that of Figure 10.

The classifications obtained according to the PEC-PCD are presented in QGIS Attributes Table, both for punctual and linear features, according to specifics of each tool, presented with flowcharts in item 2. Figure 11 shows an example from the Table of attributes obtained with the results in assessing the positional accuracy of point features using the ET-CQDG (DSG 2016) method. Although the results in Figure 11 presented rows of each sample feature, these are equivalent to the final global analysis performed from the combination of the whole set. Such an approach justifies that all values in each column, except for DE, present the same answer. In this context, the sample was rejected on 1:1,000, 1:2,000 and 1: 5,000. On scales 1:10,000 and 1:25,000, classes D and B were achieved, respectively, and from 1:50,000, class A was obtained.

FIGURE 10 – INTERFACE FOR ASSESSING POSITIONAL ACCURACY OF LINEAR FEATURES USING THE ET-CQDG METHOD.



SOURCE: The Authors (2022).

In the example shown in Figure 11, it is possible to notice that the Attributes Table shows the results referring to the DE of each feature, the RMSE, and the corresponding class for each PEC PCD scale. It is worth mentioning that in the other tests, the results are presented similarly, each with particularities. For example, in the evaluation of linear features by the PEC PCD scale (Figure 8), the classes (A, B, C,

and D) are displayed as results and for each one, "Accepted" or "Rejected" is displayed.

FIGURE 11 – EXAMPLE WITH THE RESULTS OBTAINED IN THE TABLE OF ATTRIBUTES FROM THE ET-CQDG METHOD.

ID ▲	Dist. Euclidiana	RMS	1:1000	1:2000	1:5000	1:10000	1:25000	1:50000	1:100000	1:250000
1	6,298646	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
2	4,471328	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
3	3,453445	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
4	1,683770	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
5	7,766315	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
6	1,597060	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
7	3,681186	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
8	2,946831	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
9	7,066316	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
10	1,708404	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
11	1,036062	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
12	7,046330	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
13	2,949449	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
14	10,581139	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
15	0,564979	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
16	1,961830	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
17	6,048125	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
18	0,617573	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
19	1,444939	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A
20	1,136351	4,627	Rejeitado	Rejeitado	Rejeitado	Classe C	Classe B	Classe A	Classe A	Classe A

SOURCE: The Authors (2022).

The input data for evaluating the point and linear features are the reference data from the SICAD database and the evaluated data from OSM. Vertices of pavements and curbs were the punctual features, and the linear features, in their entirety, were formed by road axes. With a significance level of 90% in the punctual features, the PEC PCD value for class A corresponded to 14m and EP of 8.5m. In classes B and C, these values were 7m and 4.25m and 2.8 and 1.7m. With the same

significance level of 90% for the linear features, class A was reached at the scale of 1:25,000, corresponding to a PEC PCD of 14m (buffer size) and EP of 8.5m.

Such values are valuable because they guide the decision-making by managers and users for specific applications based on the quality of the available data. Moreover, they make it possible to determine how statistically accurate they are when evaluated together. Tables 3 and 4 present the samples' point and linear features results.

TABLE 3 – RESULTS OBTAINED FOR ASSESSING POSITIONAL ACCURACY IN POINT FEATURES.

Method	Class A	Class B	Class C	Class D	RMSE (m)
ET-CQDG (DSG, 2016)	1:50.000	1:25.000	-	1:10.000	4.627
Merchant (1982) and Galo & Camargo (1994)	1:50.000	1:25.000	1:10.000	-	-

SOURCE: The Authors (2022).

TABLE 4 – RESULTS OBTAINED FOR ASSESSING POSITIONAL ACCURACY IN LINEAR FEATURES.

Method	Class A	Class B	Class C	Class D
<i>Double Buffer</i> (Santos 2015)	1:25.000	1:25.000	1:10.000	1:10.000

SOURCE: The Authors (2022).

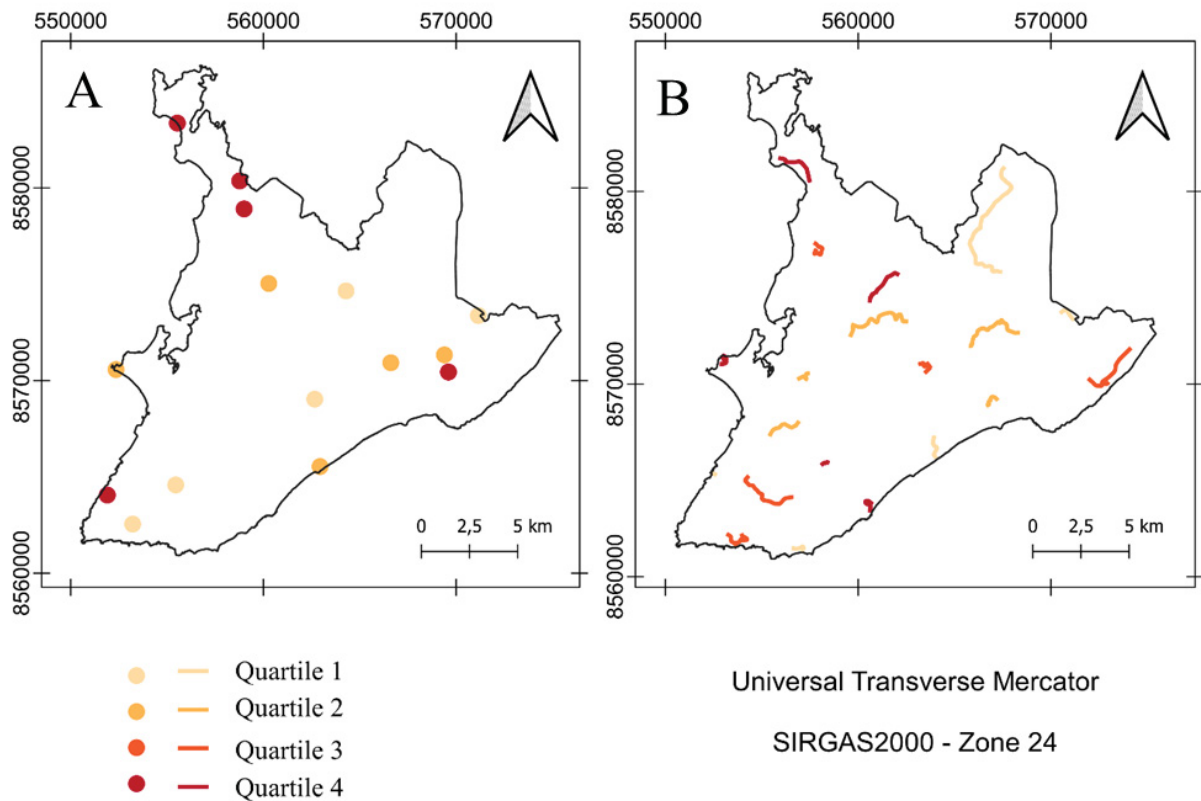
As shown in Table 3, different methodologies to assess positional quality in point features allowed us to highlight aspects related to the similarities and differences in obtaining the results. For example, when performing analyses from the ET-CQDG (DSG 2016) and the Decree-Law nº 89,817 (Brasil 1984), on a scale of 1:10,000, the sample was classified as PEC PCD D class. On the other hand, this same scale is classified as C when using Merchant (1982) and Galo & Camargo (1994). Furthermore, it was considered as a weighting factor that, with this method, it is possible to assess the existence of errors, which allows for obtaining more conclusive results from the evaluated sample. Although it is not the focus of this article to study the different characteristics associated with the exposed methods, this aspect is essential since there are vast works in which such methodologies are used

to evaluate the PEC PCD. However, there are few analyses in which it is possible to compare them.

By evaluating the linear features, Table 4 shows the results for this sample as class PEC PCD A and B on a 1:25,000 scale. This characteristic is related to the fact that it was accepted for referred classes since the PEC PCD scale triggers the analysis. The same aspect was presented at a scale of 1:10,000, accepted in classes C and D.

In order to make the visualisation of results standardised and interoperable, the DEs and DMs were classified into quartiles. Quartiles 1 and 4 represent, respectively, the smallest and largest discrepancy range, which gradually increases in the intermediate quartiles. This aspect makes it possible to identify regions with the same characteristics and allows obtaining new parameters for modelling the data based on the characteristics of a particular region. Furthermore, given the described issues, the quartiles were differentiated by the colour value, in which the lowest quartiles are lighter and get darker as they increase. Figure 12 presents the map with the visualisation of the discrepancies for the sets of point samples and linear features.

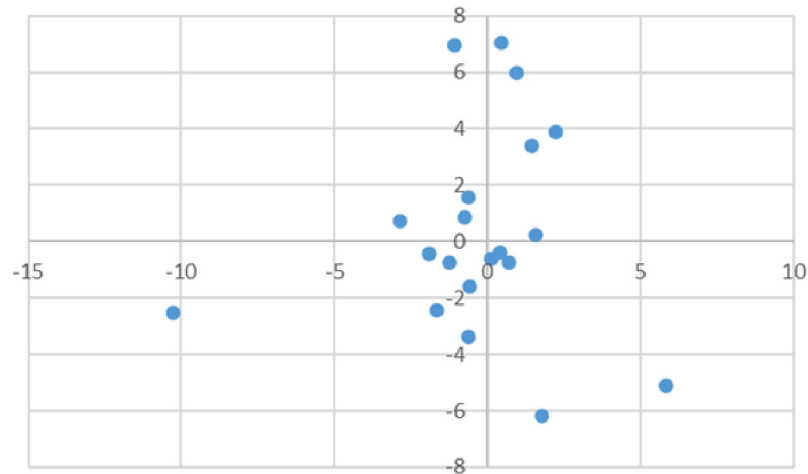
FIGURE 12 – MAP OF PLANIMETRIC DISCREPANCIES OBTAINED IN IN SALVADOR-BA:
A. PUNCTUAL FEATURES; B. LINEAR FEATURES.



SOURCE: The Authors (2022).

When analysing the behaviour of the results, it was possible to observe the differences in the patterns of discrepancies along the study region. This aspect can be evidenced by observing the DE values of the punctual features in Figure 11 since the values varied from 0.56m (quartile 1) to 10.58m (quartile 4). In the category in question, 5 features were in quartile 1 ($DE \leq 1.521\text{m}$), 5 in quartile 2 ($1.521\text{m} < DE \leq 2.948\text{m}$), 5 in quartile 3 ($2.948\text{m} < DE \leq 6.173\text{m}$) and 5 in quartile 4 ($DE > 6.173$). Furthermore, no error tendencies were found in the sample of punctual features, and even though it was classified as class A of the PEC PCD on a scale of 1:50,000, it does not reveal homogeneous behaviour of the discrepancies. In order to evidence the approached issues, the graphic of Figure 13 presents the discrepancies behaviour of the punctual features in the north (Y coordinate) and east (X coordinate) components.

FIGURE 13 – BEHAVIOR OF THE POINT FEATURE DISCREPANCIES SAMPLE



SOURCE: The Authors (2022).

As presented in Figure 13, it is possible to note that the sample behaviour presents a scatter of the calculated discrepancies. Furthermore, these results show the heterogeneity of collaborative data. That is, the quality of a given feature may be related to the time when the contribution was made, the platform, the study region, and the satellite image which feeds it, among other aspects.

However, despite this characteristic, the discrepancies in representation automatization (Figure 13) highlighted an important aspect related to the distribution of the sample results, evidenced in the analysis of point features. This characteristic may be associated with aspects of the study area or the sample distribution, which presented different magnitudes of discrepancies, as previously mentioned and presented in the analysis. In the tool development, the main difficulties were associated with statistical tests of tendency concatenation and precision since it was necessary to establish stopping criteria from the Student's t and Chi-Square Tables and evaluate linear features. In addition, there was difficulty in obtaining the best method to develop since each buffer generates a different DM for each feature.

3.5 CONCLUSIONS AND RECOMENDATIONS

Based on the products obtained, it is initially highlighted the importance of automated methodologies used for the assessment of positional accuracy in order that the verifications can be carried out more quickly, given the large volume of geospatial data currently available. Furthermore, in the processes related to

cartographic production, especially on a large scale, it is fundamental to incorporate quality analysis methods. A set of models created in a widely used open-source tool such as QGIS makes it possible to quantify, visualise and document in an automated way the spatial distribution of the quality parameters defined by national and international standards. An example of the importance of quality geovisualization could be evidenced in the results obtained for the study area. In these results, we can observe that the discrepancies tended to increase as they approached the peripheral regions of the study area (mainly in point features).

The increase in the number of geospatial data sources and the participation of non-specialised users in this process has resulted in heterogeneity in the information quality. In this context, developing quality assessment tools, where the results can be presented in GIS software, can increase the user's security in decision-making related to using such information. For example, it can help identify characteristics for integrating data from collaborative mapping with official data. As a case study in this paper, we compared OpenStreetMap and authoritative data. In a scenario of using OSM data for cartographic updating, the knowledge regarding the quality allows the evaluation of the suitability for use and documentation in the metadata of the quality reports obtained.

In addition, it is worth highlighting the importance of tools developed in open-source software that can be reused and easily accessed and manipulated by any user. Also, interfaces were developed in QGIS for data entry and visualisation of the results. Furthermore, the results displayed the essential parameters to assess the features of the evaluated database, such as discrepancies; the existence of error trends; class by the scale of the PEC-PCD and the visualisation of the results. Finally, it is essential to highlight that evaluating quality aspects in a GIS environment enables future analyses associating characteristic parameters of a specific area, such as socioeconomic aspects and the number of inhabitants.

As recommendations for future works, it is suggested to investigate methods to evaluate positional accuracy from polygonal features and to develop tools to evaluate thematic accuracy, completeness and logical consistency. Furthermore, in analysing the linear features, it is advisable to explore other metrics for evaluating the positional quality, mainly those aimed at vertex verification, such as the Hausdorff Distance method. Moreover, also on linear features, it is recommended to explore new methodologies for evaluating the positional accuracy, considering the

planimetric components separately. Such improvement would allow for exploring the data heterogeneity and performing analyses similar to Figure 13. Beyond that, as the tool also covered the representation of the results obtained for the positional quality, it is recommended to perform tests to improve the results visualisation based on the user requirements.

3.6 ACKNOWLEDGMENTS

The authors would like to thank the Graduate Program in Geodetic Sciences at the Federal University of Paraná – Brazil. This study was carried out with the support of the Coordination for the Improvement of Higher Education Personnel (CAPES) - Funding Code 001.

3.7 REFERENCES

Barbetta 2005, *Estatística aplicada às Ciências Sociais*, vol. 7, Universidade Federal de Santa Catarina, Florianópolis.

Barbosa, L.S., Silva, S.K.M. & Santos, A.P. 2021, 'Avaliação da acurácia posicional planimétrica de uma ortofoto gerada através de uma plataforma VANT, considerando a influência do tamanho amostral na normativa do INCRA', *Revista Brasileira de Geomática*, v. 9, n. 2, pp. 158-1 78. <https://doi.org/10.3895/rbgeo.v9n2.13698>.

Brasil, Decreto nº 89.817, de 20 de junho de 1984. *Normas Técnicas da Cartografia Nacional*. Diário Oficial da União, Brasília, DF, 22 jul. 1984. Seção 1, p. 8884.

Brovelli, M.A., Boccardo, P., Bordogna, G., Pepe, A., Crespi, M., Munafò, M. & Pirotti, F. 2019, 'Urban Geo Big Data', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4-W14-23-2019, pp. 23-30. <https://doi.org/10.5194/isprs-archives-XLII-4-W14-23-2019>.

Brovelli, M. A., Minghini, M. & Molinari, E. 2016, 'Database-supported change analysis and quality evaluation of OpenStreetMap Data', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-

B7-919-2016, pp. 919-925. <https://doi.org/10.5194/isprsarchives-XLI-B7-919-2016>.

Brovelli, M.A. & Zamboni, G. 2018, 'A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints', *ISPRS International Journal of Geo-Information*, vol. 7, n°. 8, pp. 289. <https://doi.org/10.3390/ijgi7080289>.

Carvalho, J.A.B. & Silva, D.C. 2018, 'Métodos para avaliação da acurácia posicional altimétrica no Brasil'. *Revista Brasileira de Cartografia*, vol. 70, n°. 2, pp. 725-744. <https://doi.org/10.14393/rbcv70n2-45404>.

Cintra, J.P. & Nero, M.A. 2015, 'New Method for Positional Cartographic Quality Control in Digital Mapping'. *Journal of Surveying Engineering*, vol. 141, no. 3, pp. 04015001-1-04015001-10. [https://doi.org/10.1061/\(asce\)su.1943-5428.0000141](https://doi.org/10.1061/(asce)su.1943-5428.0000141).

Cruz, D.T., & Santos, A.P. 2016, 'Controle de qualidade posicional do sistema rodoviário do Openstreetmap na região central De Viçosa-MG', *Anais do VI Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, Universidade Federal de Pernambuco*.

Cunha, M.M., Secatto, G.Z., Galindo, J.R.F. & Santos, A.P. 2019, 'Proposta de um método de avaliação da acurácia posicional baseado na modificação do Buffer Simples', *Revista Brasileira de Cartografia*, vol. 71, no. 4, pp. 1193-1218. <https://doi.org/10.14393/rbcv71n4-49301>.

Diretoria do Serviço Geográfico (DSG). 2011, *Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV)*, vol. 2, Brasília: DSG.

Diretoria do Serviço Geográfico (DSG). 2016, *Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG)*, vol. 1, DSG, Brasília.

Diretoria do Serviço Geográfico (DSG). 2015a, *Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais de Defesa da Força Terrestre (ET-ADGV-DefesaFT)*, vol. 1, DSG, Brasília.

Diretoria do Serviço Geográfico (DSG). 2016, Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais de Defesa da Força Terrestre (ET-ADGV-DefesaFT). vol. 2. DSG, Brasília.

Elias, E.N.N. GitHub. Plugin_PECPCD, 8 dez. 2019. Disponível em: (https://github.com/eliasnaim/AcuraciaPosicional_PEC-PCD). Acessado em: 17 dez. 2022.

Elias, E.N.N. GitHub. Plugin_PECPCD, 19 out. 2022. Disponível em: (https://github.com/eliasnaim/Plugin_PECPCD). Acessado em: 17 dez. 2022.

Elias, E.N.N., & Fernandes, V. de O. 2021, 'Qualidade dos Dados Geoespaciais do OpenStreetMap para os indicadores de Acurácia Posicional, Acurácia Temática e Completude' *Geografia (Londrina)*, vol. 30 no. 2, pp. 255-275.
<https://doi.org/10.5433/2447-1747.2021v30n2p255>.

Elwood, S., Goodchild, M.F. & Sui, D.Z. 2012, 'Researching volunteered geographic information: Spatial data, geographic research, and new social practice', *Annals of the Association of American geographers*, vol. 102, no. 3, pp. 571-590.
<https://doi.org/10.1080 / 00045608.2011.595657>.

Fonte, C.C., Antoniou, V., Bastin, L., Estima, J., Arsanjani, J.J., Bayas, J-C., L., See, L. & Vatsava, R. 2017, 'Assessing VGI Data Quality', in: G. Foody, L. See, S. Fritz, P. Mooney, A-M. Olteanu-Raimond, C. Fonte & V. Antoniou (eds), *Mapping and the Citizen Sensor*, Ubiquity Press, London, pp.137–163. <https://doi.org/10.5334/bbf>.

Foz, E.T., Muñoz-Nieto, Á. L., González-Aguilera, D. & González, P. R. 2021, 'Implementación de procesos de control de calidad na actualización de series cartográficas urbanas mediante CAD y SIG'. *Revista Cartográfica*, n. 103, pp. 47-69.

Galo, M. & Camargo, P. 1994, 'Utilização do GPS no controle de qualidade de cartas'. *Anais do Congresso Brasileiro de Cadastro Técnico Multifinalitário*, Universidade Federal de Santa Catarina, Florianópolis, pp. 41-48.

Ganapati, S. 2011, 'Uses of Public Participation Geographic Information Systems Applications in E-Government', *Public Administration Review*, vol. 71, no. 3, pp. 425-434. <https://doi.org/10.1111/j.1540-6210.2011.02226.x>.

García-Balboa, J.L., Ureña-Cámara, M. A., & Xavier, E. M. A. 2021, 'iCPos: una herramienta web para la elaboración de Informes de Calidad Posicional', *Revista Cartográfica*, n. 103, pp. 183-199. <https://doi.org/10.35424/rcarto.i103.997>.

Goodchild, M. F. 2010, 'Foreword', in: W. Shi (ed), *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses*. CRC Press, Boca Raton, pp. XIX.

IBGE. Malha Municipal 2021. Rio de Janeiro: IBGE, 2021. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=acesso-ao-produto>. Acessado em: 14 de Novembro de 2022.

https://geofpt.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2019/UFs/BA/. Acesso em: 15 jun. 2020.

International Organisation for Standardisation. 2013, *ISO 19157: Geographic Information – Data Quality*, Switzerland.

Keates, J.S. 1973, *Cartographic Design and Production*, Longman, London.

Kent, A. J. 2009, 'Topographic Maps: Methodological Approaches for Analysing Cartographic Style', *Journal of Map & Geography Libraries*, vol. 5, no. 2, pp. 131–156. <https://doi.org/10.1080/15420350903001187>.

Lisboa, M H., Santos, A.P., Medeiros, N. Das G.& Abreu, M.V.S. 2019, 'Avaliação da Acurácia Posicional Tridimensional de Produtos Cartográficos utilizando um Elipsoide de Incertezas', *Revista Brasileira de Cartografia*, vol. 71, nº. 4, pp.1040-1063. <https://doi.org/10.14393/rbcv71n4-49488>.

Luz, C.C., & Antunes, A.F.B. 2015. 'Validação da tecnologia VANT na atualização de

bases de dados cartográficos geológicos. Estudo de caso: Sistema cárstico do rio João Rodrigues', *Revista Brasileira de Cartografia*, nº. 67/7.

Martini, A., Kuper, P.V., Breunig, M. 2019, 'Database-supported change analysis and quality evaluation of OpenStreetMap Data'. *Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2-W5-535-2019, pp. 535-541. <https://doi.org/10.5194/isprs-annals-IV-2-W5-535-2019>.

Merchant, D.C. 1982, 'Spatial accuracy standards for large scale line maps', *American Congress on Surveying and Mapping Proceedings*, ACSM, Falls Church, pp 222-231.

Mozas-Calvache, A.T. & Ariza-López, F.J. 2019, 'Analysing the positional accuracy of GNSS multi-tracks obtained from VGI sources to generate improved 3D mean axes', *International Journal of Geographical Information Science*, vol. 33, nº. 11, pp. 2170-2187. <https://doi.org/10.1080/13658816.2019.1645335.4>

Mônico, J.F.G., Póz, A.P.D., Galo, M., Santos, M.C. & Oliveira, L.C. 2009, 'Acurácia e precisão: revendo os conceitos de forma acurada', *Boletim de Ciências Geodésicas*, vol. 15, nº. 3, pp. 469-483.

Nogueira Jr., J.B. 2003, 'Controle de qualidade de produtos cartográficos: uma proposta metodológica', Master Thesis, Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista Júlio de Mesquita Silva.

Oliveira, D.V. & Brito, J.L.S. 2019. 'Avaliação da Acurácia Posicional de Dados Gerados por Aeronave Remotamente Pilotada', *Revista Brasileira de Cartografia*, vol. 71, no. 4, pp. 934-959. <https://doi.org/10.14393/rbcv71n4-45080>.

Pedreira, W.J.P., de Andrade Oliveira, J. & Santos, P.S. 2020, 'Avaliação da Acurácia Altimétrica usando a Tecnologia VANT', *Revista Caminhos da Geografia*. v. 21, n. 3 pp. 209–222. <https://doi.org/10.14393/RCG217348071>.

Pereira, T.A.J. & Nero, M.A. 2012, 'Análise de normas de controle de qualidade

posicional em cartografia ao redor do mundo: exemplos atuais de alguns países'. *Anais do IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação*, Universidade Federal de Pernambuco, Recife.

Pessoa, G.G.; Carrilho, A.C., Miyoshi, G.T., Amorim, A. & Galo, M. 2021, 'Assessment of UAV-based digital surface model and the effects of quantity and distribution of ground control points', *Int. J. Remote Sens*, v. 42, n. 1, pp. 65–83. <https://doi.org/10.1080/01431161.2020.1800122>.

Roberto, A.J. 2013, 'Extração de Informação Geográfica a partir de Fotografias Aéreas obtidas com VANTs para apoio a um SIG Municipal', Master Thesis, Departamento de Geociências, Ambiente e Ordenamento do Território, Universidade do Porto.

Robinson, A.C., Demšar, U., Moore, A.B., Buckley, A., Jiang, B., Field, K., Kraak, M., Camboim, S. P. & Sluter, C.L. 2017, 'Geospatial big data and cartography: research challenges and opportunities for making maps that matter', *International Journal of Cartography*, vol. 3, nº. sup1, pp. 32-60. Mar. <https://doi.org/10.1080/23729333.2016.1278151>.

Santos, A.P. 2015, 'Controle de Qualidade Cartográfica: Metodologias para Avaliação da Acurácia Posicional em Dados Espaciais', PhD Thesis, Departamento de Engenharia Civil, Universidade Federal de Viçosa.

Santos, A.P., Medeiros, N. G., Santos, G.R. & Rodrigues, D.D. 2016, 'Avaliação da acurácia posicional planimétrica em modelos digitais de superfície com o uso de feições lineares', *Boletim de Ciências Geodésicas*, vol. 22, nº. 1, pp. 157-174. <https://doi.org/10.1590/S1982-21702016000100009>.

Silva Júnior, J.A.D., França, T.M.S., & Moraes, V.S.D. 2022, 'Avaliação da acurácia planimétrica de imagens Google Earth numa zona urbana e rural o estado de Pernambuco', Brasil. *Revista cartográfica*, n.105, pp. 117-133. <https://doi.org/10.35424/rcarto.i105.1386>.

Soares, G., Inocencio, L. C., Veronez, M.R., da Silveira, L.G., Bordin, F. & Marson, F. P. 2018, 'Analysis of positional and geometric accuracy of objects in survey with unmanned aerial vehicle (UAV)', *In IEEE International Geoscience and Remote Sensing Symposium*, Valencia – Spain, pp. 5693-5696. <https://doi.org/10.1109/IGARSS.2018.8517315>.

Teixeira, P.S. & Santos, A.D.P. 2019, 'Implementação de ferramenta para Controle de Qualidade Cartográfica no software QGIS'. Proceedings XX GEOINFO, São José dos Campos, SP, Brazil, pp. 212-217.

Tong, X., Sun, T., Fana, J., Goodchild, M. F. & Shi, W. 2013, 'A statistical simulation model for positional error of line features in Geographic Information Systems (GIS)', *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, pp. 136-148. <https://doi.org/10.1016/j.jag.2012.08.004>.

Tveite, H. & Langaas, S. 1999, 'An accuracy assessment method for geographical line data sets based on buffering', *International Journal of Geographical Information Science*, vol. 13, no. 1, pp. 27-47.

Yang, C., Yu, M., Hu, F., Jiang, Y. & Li, Y. 2017, 'Utilising cloud computing to address big geospatial data challenges', *Computers, Environment and Urban Systems*, vol. 61, pp. 120-128. <https://doi.org/10.1016/j.compenvurbsys.2016.10.010>.

Zanetti, J., Paula, R. M., Santos, A. P., & Medeiros, N. G. 2016, 'Avaliação da acurácia posicional planimétrica de ortoimagens disponibilizadas nos sistemas de informações geográficas'. *Revista Brasileira de Cartografia*, v. 68 n. 7, pp. 1341-1352.

Zhou, Q. 2018, 'Exploring the relationship between density and completeness of urban building data in OpenStreetMap for quality estimation', *International Journal of Geographical Information Science*, vol. 32, no. 2, pp. 257-281. <https://10.1080/13658816.2017.1395883>.

4. EXPLORING SPATIO-TEMPORAL PATTERNS OF OPENSTREETMAP (OSM) CONTRIBUTIONS IN HETEROGENEOUS URBAN AREAS

Autores: Elias Nasr Naim Elias, Fabricio Rosa Amorim, Marcio Augusto Reolon Schmidt, Silvana Philippi Camboim.

Periódico: Artigo Submetido ao Boletim de Ciências Geodésicas (2022).

Abstract:

The potential of intrinsic parameters to estimate geospatial data quality in Voluntary Geographic Information (VGI) platforms is a recurrent theme in Cartography research. The spatial-temporal distribution in these platforms is very heterogeneous, depending on several factors such as input availability, number, and motivation of the volunteers, especially in developing countries. The most recent approaches have been aiming to detail temporal patterns as an additional measure of quality in VGI. This research proposes a methodology to identify and analyse the behaviour of the contribution parameters over time (2007-2022) of the OSM platform and differentiate the influences that affect its growth. A part of the Metropolitan region of Curitiba was the study area, subdivided into 1 x 1 km cells. The cumulative growth of the contributions was calculated and later adjusted via a Logistic Regression. The parameters obtained made it possible to identify abruptly growing cells caused by external data import, mass contributions, or collective mapping activities. Also, the heterogeneity in the growth of the data available in OSM over time was evidenced. Such intrinsic temporal parameters have the potential to be used to estimate the spatial distribution of data quality, even capturing local heterogeneities given the granularity of the proposed grid. Furthermore, the proposed methodology promoted the investigation of a new indicator of intrinsic quality based on modelling the spatio-temporal evolution of OSM feature insertions. The results made it possible to relate the contribution dynamics in different mapping stages and differentiate each evaluated cell behaviour.

Keywords: Geospatial Data Quality; Collaborative Mapping; Intrinsic Parameters; Logistic Regression.

4.1 INTRODUCTION

OpenStreetMap (OSM) is one of the world's biggest and most popular platforms for collaborative collection (Teimoory, Abbaspour and Chehreghan 2021; Brovelli and Zamboni 2018), with more than 7.8 million registered users, overtaking 16.7 billion edited features. Seeing the promising aspects of VGI, focusing on OSM, different researchers worldwide have focused efforts to comprehend its characteristics, especially in data quality and viability in integrating processes (Brovelli and Zamboni 2018; Nasiri et al. 2018; Zhang, Malczewski 2017; Sehra, Singh and Rai 2017; Haklay 2010;).

Geospatial data quality is essential to topographic mapping and is part of specific standards for its production (ISO 19157 2013; DSG 2015). In VGI, this measurement is even more indispensable since the contributed data is heterogeneous and can be affected by several factors related to the area, the profile of the contributors, the quantity, and the dynamic of the contributions. Extrinsic approaches from comparisons with topographic mapping (Brovelli and Zamboni 2018; Zhang and Malczewski 2017; Haklay 2010) and intrinsic (Sehra, Singh and Rai 2017) characterised by analysis of edition history, the number of collaborators and contributions describe the procedures for evaluating VGI's quality.

The evaluation of intrinsic parameters is even more relevant in developing countries, where the chronic lack of resources for cartography many times results in the absence of updated data for comparison. Besides that, this situation makes the data from VGI even more necessary to complement the existing topographic mapping (Camboim, Bravo and Sluter 2015). In Brazil, for example, research by Silva e Camboim (2020) reveals that in the 1:25.000 scale, little more than 5% of the country's extension has available topographic mapping. Thus, understanding the data's quality in these regions is essential.

The classic intrinsic parameters, characterised by the number of contributions, contributors, and edition history, were based on the Linus Law. This law applied to the VGI context, as addressed by Haklay et al. (2010), states that as the quantity of contributors in a given spatial unit increases, the greater and better known its quality. In the early days of intrinsic parameter research, the only temporal variable was the date of the last data edition. However, Paiva and Camboim (2021) evidenced that this approach alone would not explain the entire behaviour of quality

in VGI. Such aspect converges with research recently conducted aimed at detailing temporal patterns and thereby improving the understanding of the spatial distribution of its quality (Le Guilcher, Olteanu-Raimond and Balde, 2022; Grinberger et al., 2021; Brückner et al., 2021; Witt, Loos and Zipf 2021; Arsanjani et al. 2015; Gröching, Brunauer and Rehr 2014).

Concerning mathematical modelling, Grinberger et al. (2021) measured the number of contributions over time to detect events in the OSM platform worldwide. The authors proposed procedures to identify these large-scale data production events in the OSM history and analyse their patterns. Brückner et al. (2021) estimated the completeness of OSM retail stores in Germany. Both pieces of research are based on accumulated contributions over time and use the four-parameter logistic regression model in their data. This model is described from an "S" shaped curve (sigmoid function). This curve format can be associated with the pattern of contributions in a given area, which starts with few contributions, increases and gradually stabilises over time. Brückner et al. (2021) used other regression models in their analysis. When presenting the logistic function to estimate the behaviour of the OSM data, Grinberger et al. (2021) relate the characteristics of the curve with the premises of the work developed by Gröchenig, Brunauer, and Rehr (2014), who proposed an approach to identify regional and temporally different developments associated with the mapping evolution. The proposed model allowed us to classify the different stages of activity in the contributions accomplished in OSM, such as Start, Growth, and Saturation. In this context, Arsanjani et al. (2015) proposed a Contribution Index (CI) based on analysing the Spatio-temporal patterns of OSM contributions. The CI encompasses the number of contributions, the average number of versions, the average number of attributes, and the number of users. In addition to the research mentioned above, Witt, Loos and Zipf (2021) analysed the impact of significant imports in OSM for the Netherlands and India, and Guilcher, Olteanu-Raimond and Balde (2022) studied the evolution of massive imports in OSM in France.

Considering the questions presented, as there is a finite number of features in a region, this research aims to study the evolution of contributions in a homogeneous cell, looking for patterns that vary according to the location of this cell, working explicitly with data in large urban centres of Brazil. The hypothesis is that if the existent patterns are known, it will be possible to determine in the future which

stage a particular cell is and, therefore, how close it is to having its complete mapping. This information is beneficial to use more robust data and encourage mapping in regions that are still poorly mapped, reducing heterogeneities, especially in the poorest and peripheral areas.

This research proposes a methodology to identify and analyse the behaviour of Spatio-temporal parameters of the OSM contributions and differentiate the influences that affect the accumulated growth of the insertion of features over time, either by mass contributions or mapathons. Besides that, the interaction between official and collaborative data and the synergy obtained by importing data into OSM were analysed. Finally, the adequacy of the logistical model was verified for the accumulated count of point, linear and polygonal features mapped in a specific area and the measure of the smallest possible size of the surrounding rectangle. The study was conducted in an area of the metropolitan region of Curitiba (Brazil). In this context, it was possible to identify and compare different influences that affect a big metropolis and validate them to its adjacent regions.

4.1.1 A current panorama of geospatial big data integration and analysis

The current paradigm of the technological environment constitutes the era of Big Data, with all the unique opportunities it presents for the use and production of spatial data (Robinson et al., 2017). The territory portraying, formerly strongly restricted to official agents' and traditional methods, costly and limited to specialists, nowadays happens dynamically by several agents, including the citizens themselves, through several IT tools. Globally, proposals for alternative sources to traditional ones and the reuse of open geospatial or statistical data are discussed. The United Nations, through the Statistics Division (UNSD) and the Committee of Experts on Global Geospatial Information Management (UN-GGIM), emphasize the importance of this vision in documents such as the United Nations Integrated Geospatial Information Framework - UN-IGIF (UN-GGIM, 2022). Countries like Australia are at the forefront of this integration (Kitchin, 2015). The trend of general use of Big Data, including diverse sources such as social networks, sensors, and device monitoring, among others, to improve official databases (Tam and Van Halderen, 2020). In Brazil, large government databases, such as RAIS - Annual Report of Social

Information (Ministry of Labour and Previdence, 2022) and CNEFE - National Register of Addresses for Statistical Purposes (IBGE,2022), and cadasters in various instances, have the potential to be leveraged. The advantage is to portray specific scenarios with comprehensiveness and dynamism, as in the case of RAIS, created to manage labour data and which can be important information about companies, providing addresses of Points of Interest that can be geocoded and integrated with existing databases. For instance, Paiva and Camboim (2021) compared data on economic activity permits available at the Open Data Portal of Curitiba's City Hall with data quality parameters from OpenStreetMap.

There is also research on recent applications in urban areas using open geospatial data, whether they are mined from proprietary applications based on crowdsourcing, government open data and SDis, or involve collaborative mapping. For example, Cerqueira and Diniz (2022) explored the distribution of urban equipment to find primary and secondary centralities in the Metropolitan Region of Belo Horizonte - Minas Gerais. The authors used Google Places of Interest data. Also, with data from the capital city of Minas Gerais, Paiva and Camboim (2022) worked with official and collaborative data to search for intrinsic geospatial data quality models. Finally, Elias et al. (2020) explored the extrinsic quality of road axes from OSM's VGI platform in Salvador – Bahia, compared with the authoritative municipal dataset. The parameters analyzed were positional accuracy, thematic accuracy and completeness, exposing the data heterogeneity in the region.

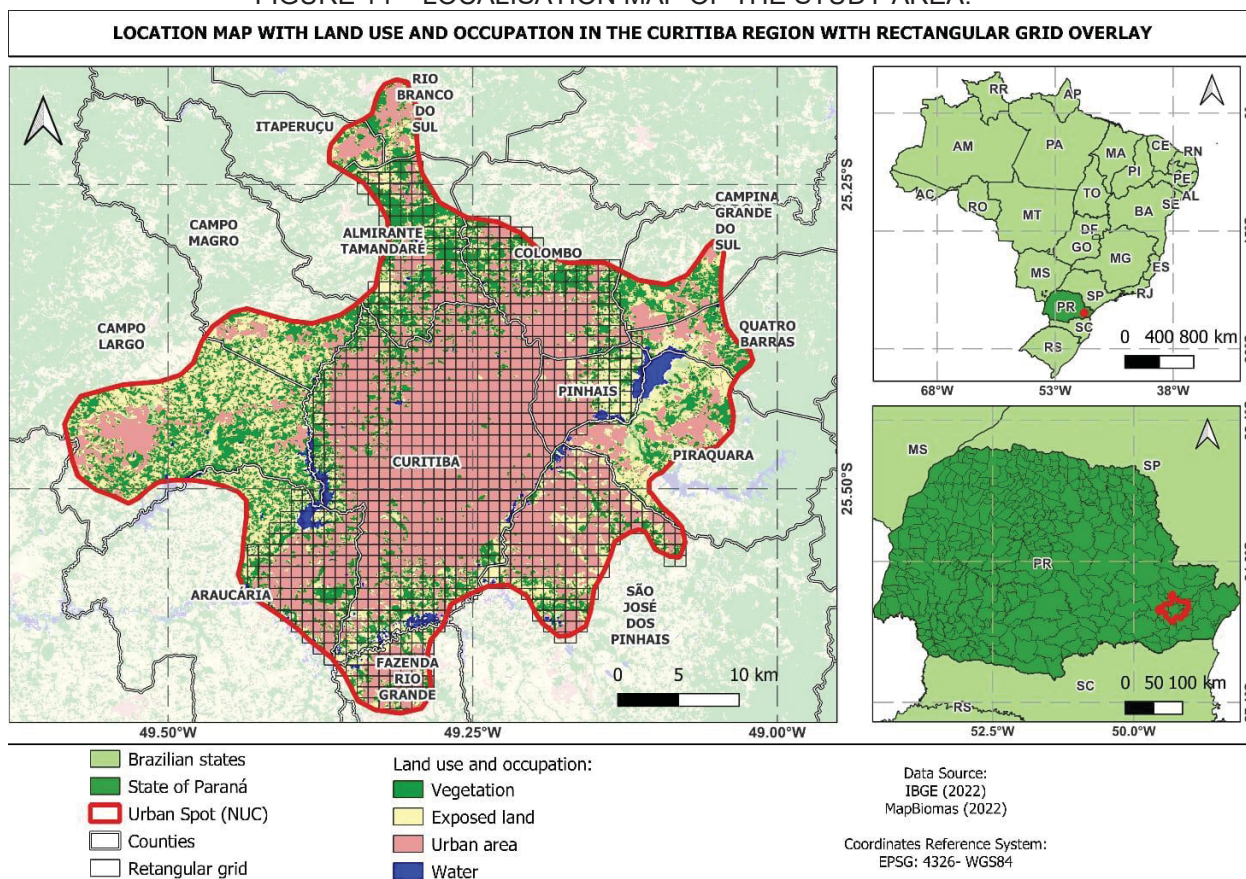
4.2 METHODOLOGY

The first step to obtain and analyse the Spatio-temporal patterns of OSM contributions was to subdivide the study region into a grade with cells of 1x1 km (Considering approximately, the equivalence along the Equator line). This procedure was essential to identify micro patterns of OSM collaboration. The second step corresponds to obtaining the monthly and daily accumulated amount of punctual, linear and polygonal features inserted into the platform, in the periods from 11/2007 to 10/2022, for each evaluated cell. The third step comprehended the data adjustment from logistic regression and the extraction of its parameters for each cell. Finally, the fourth step was creating visual representations of the parameters to observe the spatial aspects influencing the contribution patterns.

This research's study area comprehends a part of NUC of the Metropolitan Region of Curitiba (Região Metropolitana de Curitiba - RMC), located in southern Brazil. As described by COMEC (2022), the RMC has 3.223.836 inhabitants, Brazil's eighth most populous metropolitan region. Besides that, it is the second largest metropolitan region in the country, in its extension, with 16.581,21km². Figure 14 presents the location map of the study area.

An important aspect worth mentioning is concerning the heterogeneous aspects of the municipalities. For example, in the land use map (Figure 14), Curitiba concentrates most of its extension as built-up areas, while the others present some parts with vegetation and exposed soil. This characteristic influenced the cell delimitation for the analyses, prioritising urbanised areas. The selection is necessary because there is a direct relationship between the amount of existing information, the local population, and the number of contributors and contributions. However, although the cells are primarily distributed in built-up areas, the contribution dynamics are different between the metropolis core and peripheral areas.

FIGURE 14 – LOCALISATION MAP OF THE STUDY AREA.



SOURCE: The Authors (2022).

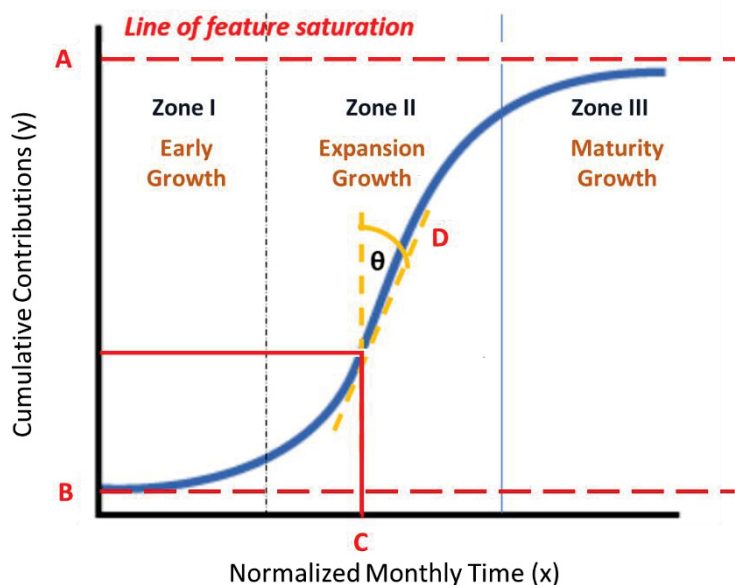
4.2.1 Obtaining the data and adjustment of the logistic curve

The OHSOME Application Programming Interface (API), at <https://api.ohsome.org>, was used to obtain the number of features inserted over time into OSM. This API allows extracting the history of feature editions in OSM, the number of contributors in a specific area or set of features, adding up the information, and even obtaining the set from specific feature categories. The programming language was Python, and the software QGIS 3.x was used for the data gathering and manipulation. The API documentation makes explicit that the data extraction must be accomplished through specific queries, from delimiting the geographic coordinates in the region of interest. Then, an interactive process that takes the cell's bounding box as input calculates the contribution data and performs mathematical modelling for each. From the contribution history accumulated over time, the data modelling was accomplished through the four-parameter logistic regression, as shown in Equation 1.

$$(1) \quad y = a + \left(\frac{b-a}{1+e^{(c-x)/d}} \right)$$

In Equation 1, the parameters a , b , c , and d model the curve and represent the upper asymptote, the lower asymptote, the midpoint of the logistic curve (on the x -axis) and the steepness of the logistic. The x value refers to the normalised monthly value, and the y represents the accumulated amount of features inserted in OSM. The graph in Figure 15 shows the logistic curve parameters and zones.

FIGURE 15 – THE LOGISTIC CURVE PARAMETERS AND ZONES



SOURCE: The Authors (2022).

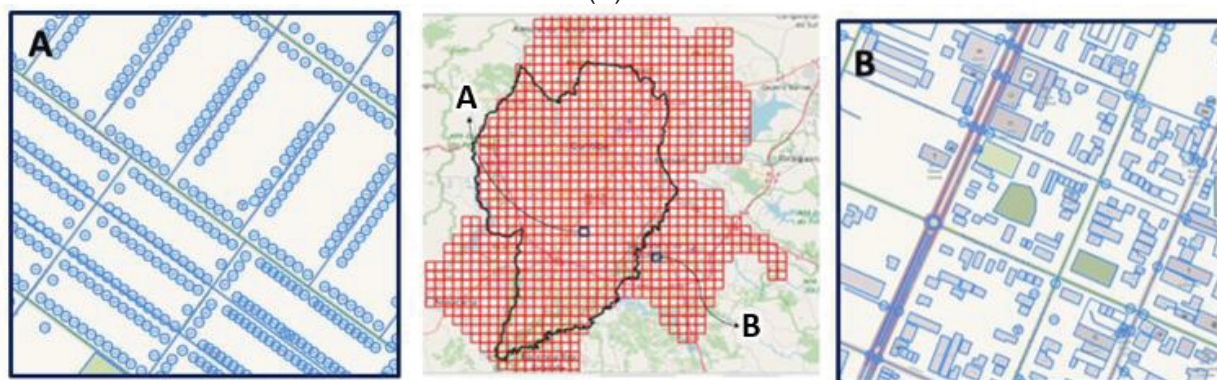
As depicted in Figure 15, the curve's parameters describe its characteristics, from which conclusions can be derived about the contribution patterns over time. For example, in Brückner et al. (2021), one of the criteria used by the authors was the value of the upper asymptote in relation to the most recent number of features. Furthermore, considering that the parameter d is related to the slope of the curve, low values in the regression may indicate specific aspects of the accumulated growth pattern. For instance, in regressions that present abrupt leaps, the parameter c tends to be similar to the month (value x) in which the largest insertion of features occurred.

Figure 15 presents a relation between the behaviour of the curve and the different stages of contributions from zones I, II, and III. Zone I characterises the initial stage of mapping activities. Zone II represents the growth of the number of contributions, and zone III characterises a saturation stage in which a given area reaches the maximum limit of features that can be inserted. An important variable analysed in this research was the presence of high slopes in zone II of the curves. When this slope is too high, the θ angle tends to zero, and it is possible to detect by the model areas of significant contributions or data import. This research used the term "cell with abrupt growth" in this case.

Besides the regressions graphs, it was possible to obtain the number of monthly inserts in this step. This procedure calculates the number of available features in a specific month, subtracted from the monthly value that precedes it. This analysis shows the relationship between the number of features in the highest insertion month and the most recent date. Finally, the regressions were analysed similarly to the methodology presented by Grinberger et al. (2021) by calculating the Normalized Root Mean Squared Error (NRSME) of the median value of each cell's feature.

In obtaining and modelling data, it was noticed that Curitiba presented an atypical behaviour in the contributions pattern compared to the other municipalities that are a part of the NUC. The analysis of the monthly growth and the obtained values evidenced this aspect. Furthermore, abrupt jumps can be observed in the number of features inserted over time, besides the organic growth (characterised by gradual contributions and different collaborators). Considering that Curitiba is the largest municipality of the evaluated region, it is worth highlighting that the Institute of Research and Urban Planning of Curitiba (Instituto de Pesquisa e Planejamento Urbano de Curitiba - IPPUC) maintains an open geospatial data platform. This website provides points with parcel addresses, urban equipment, and street axes, which do not occur in smaller municipalities. It was then identified that the insertions of the data made available by IPPUC into OSM appear as abrupt leaps in the curve. To exemplify, Figure 16 presents the arrangement of the features in OSM considering the presence (A) and absence (B) of IPPUC data.

FIGURE 16 – ARRANGEMENT OF OSM FEATURES IN CURITIBA (A) AND SÃO JOSÉ DOS PINHAIS (B) CELLS.



SOURCE: The Authors (2022).

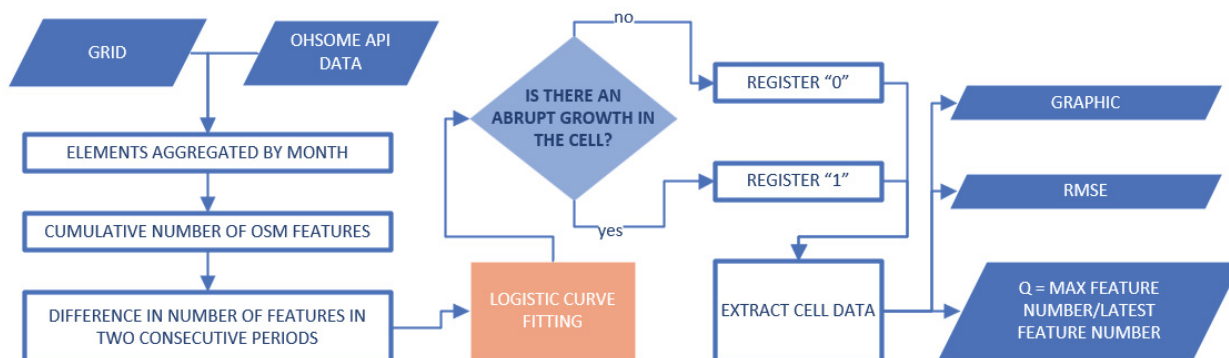
As presented in region A of Figure 16, it is possible to notice the set of point features around the blocks. It was identified that such elements correspond to the address information of each parcel available in the region. IPPUC (2022) provides, in shapefile format, the more than 300.000 parcels in the municipality, with their respective addresses, available. In turn, it is worth noting that Brazilian municipalities do not commonly make such data available, including those belonging to NUC. For example, in region B of Figure 16, there are no address points.

All point features with the tag key "addr:street" were filtered out. The analysis produced graphs with the model and parameters of the logistic regression, the graph of the monthly contributions, a list of NRMSE values, the quotient between the number of features in the highest contribution month and the most recent date, as well as the relations between the amount of punctual, linear and polygonal features. Besides the described information, cells that presented the $\theta = 0^\circ$ at some point of the curve generation were also computed. These cells are essential once they allow identifying different contribution patterns over time.

4.2.2 Applications development

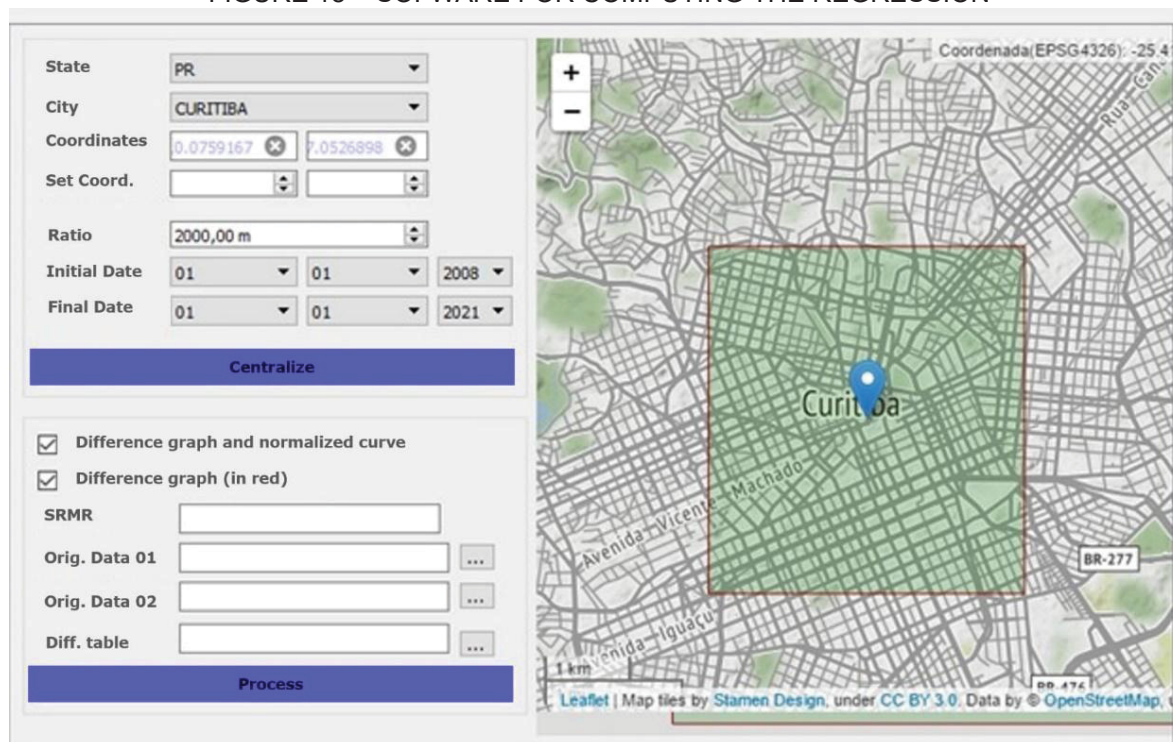
The main Python libraries used in the application were `spicy` to optimise and `matplotlib` to calculate the logistic regression and plot the data, and `NumPy` for mathematical operations. The algorithm takes as inputs both the contribution files and the cells' bounding boxes, generating the output lists after the computation. Figure 17 presents the flowchart of the procedure. Developing an interface enables the user to select the target region, insert data, and calculate the regression, thus supporting data extraction from different Brazilian regions. Figure 18 presents that interface. Since the contribution patterns tend to present heterogeneous characteristics according to the region, this software allows the selection of the target area and the time interval.

FIGURE 17 – SEQUENCE FOR EXTRACTING THE INFORMATION



SOURCE: The Authors (2022).

FIGURE 18 – SOFTWARE FOR COMPUTING THE REGRESSION



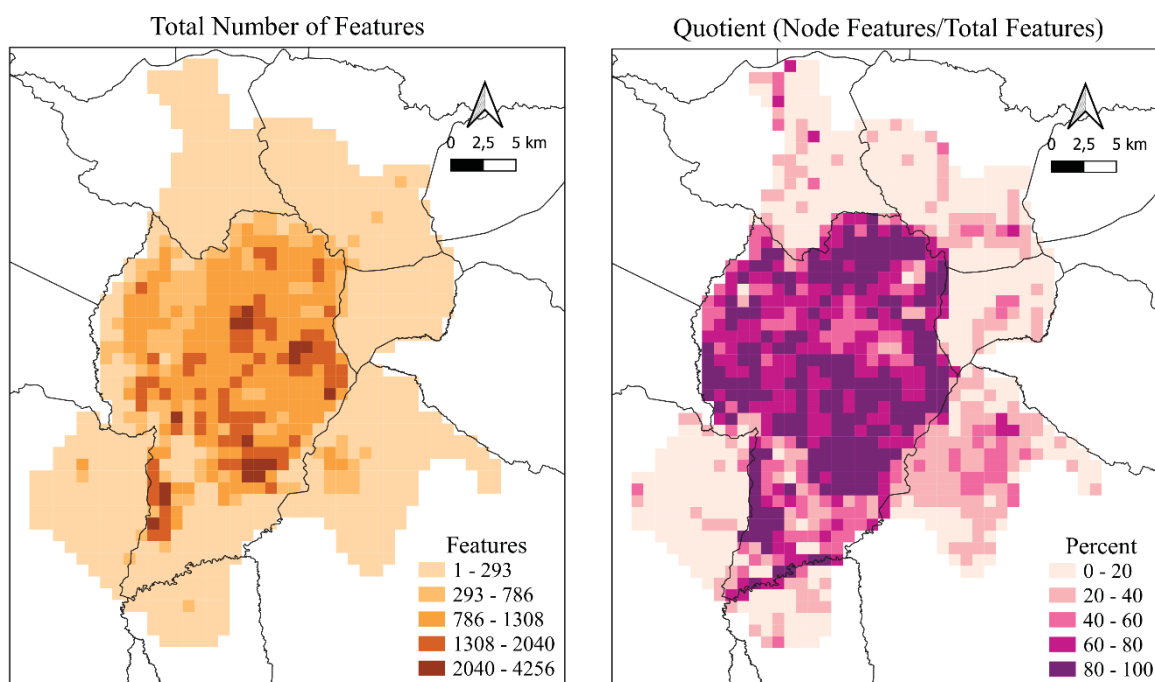
SOURCE: The Authors (2022).

Considering the interface in Figure 18, the user can select the target municipality and insert the geographic coordinates of the region of interest and the length of the enclosing rectangle from which the data will be obtained. Finally, the target interval and the path to store the lists with the results are selected. The interface icons were derived from PyQT and were developed using qt designer, besides folium, for visualising the base map. The developed scripts are available for download in the GitHub online repository.

4.3 RESULTS AND DISCUSSION

Based on the described methodology, 1074 cells with 1x1km had their behaviour evaluated. Figures 19 (A) and 19 (B) show the representation of the existing number of features in OSM and the quotient between the number of points features and the total number of features.

FIGURE 19 – FEATURES AVAILABLE IN OSM (A) AND THE PERCENTAGE OF POINT FEATURES (B).



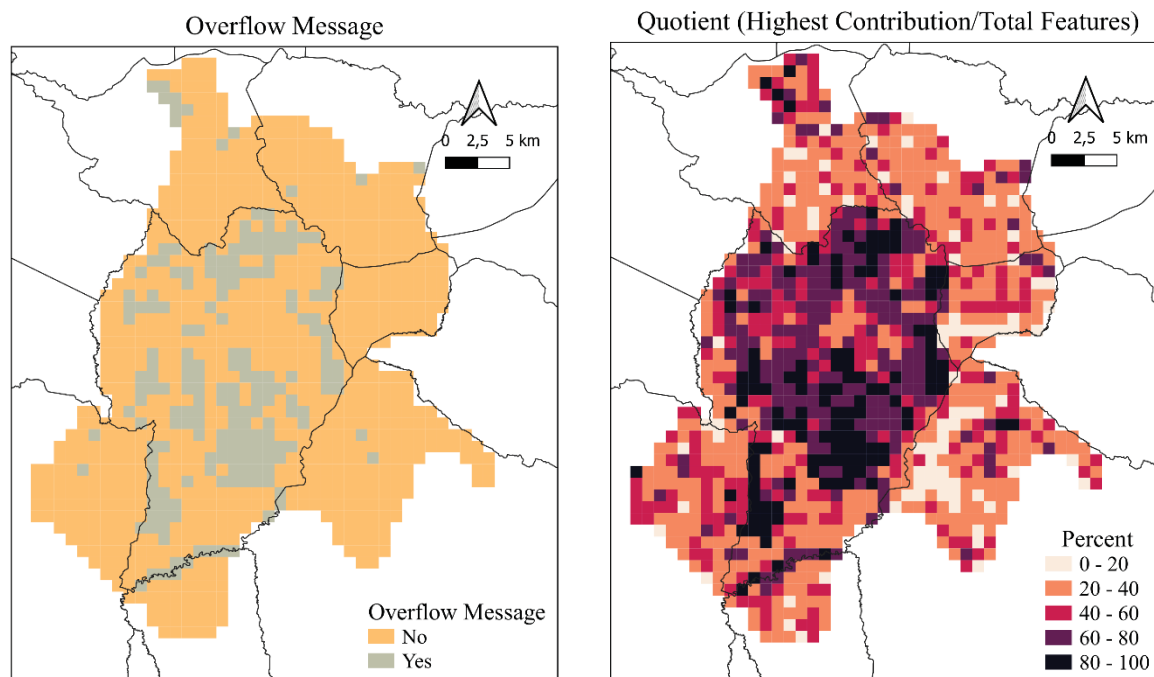
SOURCE: The Authors (2022).

Concerning the data availability, the areas close to the central region of Curitiba had more significant contributions. The mean and median number of features in Curitiba were 804.59 and 834. In the other regions, these values were 92.63 and 65 features. Besides, it was possible to notice a convergence with the population data, considering that Curitiba is equivalent to approximately 2/3 of the NUC region, in addition to the higher rate of urbanisation of the capital. It was also noticed that the central region of the municipality has a higher cell with a predominance of points, more than 60% of the total (Figure 19 B). In the regression analysis procedures, in some cells, the value of the θ angle was 0° (d parameter), predominantly in regions with abrupt leaps in the accumulated contributions. In data processing, identifying these cells was accomplished from an overflow message

directly related to the high percentage of feature inserts in a month about the total accumulated. The insert of points with Curitiba's addresses justifies this aspect. The maps in Figures 20(A) and 20(B) present the spatialisation of cells with abrupt growth and the calculated percentage.

When analysing Figures 20(A) and 20(B), an initial statement is that the imports of addresses influence the whole pattern of contributions of Curitiba since a single month was equivalent to more than 80% of the total features. The other municipalities of the NUC did not evidence this aspect, and the percentages varied predominantly between 0% and 40%. Of the 527 cells that intercepted Curitiba, in 233, the percentage (Figure 20B) was higher or equal to 70%, and the abrupt jumps accounted for more than 80% of them (Figure 20A). The overflow message was captured in all cells in which the percentage was higher than 80%. The abrupt growth was not predominant in the cells of other municipalities of the NUC, considering that regions with low contribution characterised those that occurred. An exception to this behaviour occurred in a cell in Araucária, where abrupt growth indicated a significant contribution in a residential area. In this context, despite being predominantly urban, as presented in the land use map in Figure 14, the neighbouring municipalities of the NUC include agricultural and forestry areas, which naturally have a different pattern of features than urban areas. Figure 21 presents examples of the behaviour of the curve obtained in the logistic regression, of evaluated cells, in different aspects.

FIGURE 20 – CELLS IN WITH OVERFLOW (A) AND THE % OF THE LARGEST CONTRIBUTION MONTH (B).

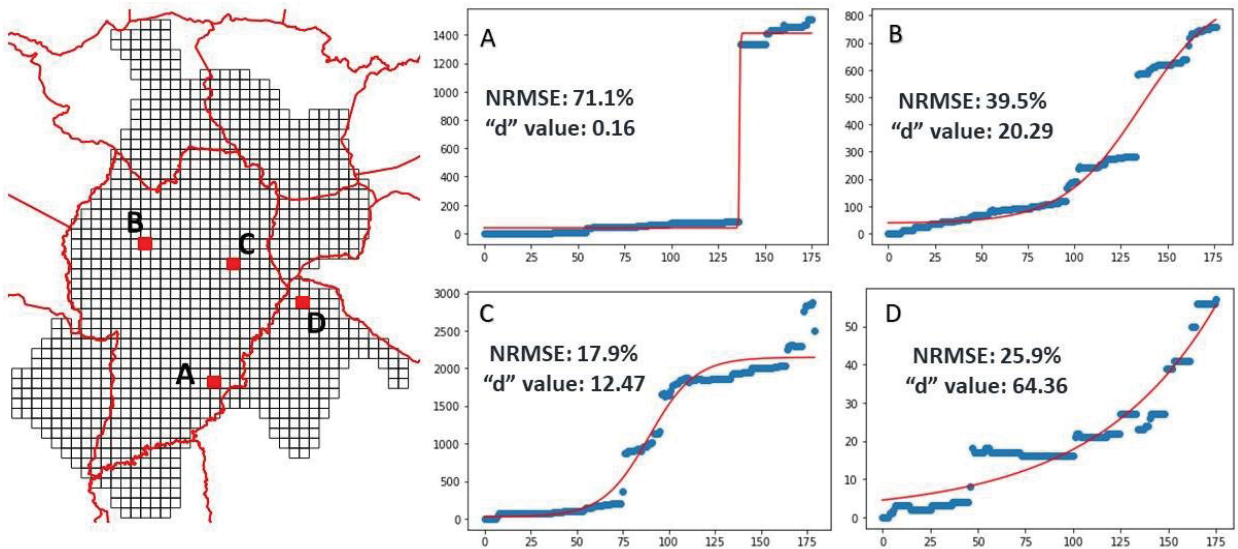


SOURCE: The Authors (2022).

In the graphs of Figure 21, blue represents the accumulated growth of contributions, and red represents the logistic regression curve. The jump in the number of inserted features in graph A is due to the insertion of IPPUC addresses points. There were 1248 contributions in a single day, corresponding to 80% of the total. In graph B, the inserts are evident, as, in a single month, 299 features were inserted, corresponding to 39% of the total. In turn, this importation occurred in a mix of organic features, with proportional amounts, which generated a smoother curve. It is possible to describe the smoothness of the curve in terms of the slope d parameter (Equation 1), in which the lower the value obtained, the steeper its slope. The accumulated contributions in graphs C and D did not present any leaps. However, in graph C, it is possible to notice a mix of significant contributions in short periods, beyond the fact that the Zone III of the curve is close to reaching a saturation level, considering last year's stability. In this region, most of the building polygons were mapped. In these features, the points with the addresses possibly have not even been inserted. Instead, the edition may have included inserting the address tags into an existing building. This aspect justifies the non-existence of abrupt jumps in this cell.

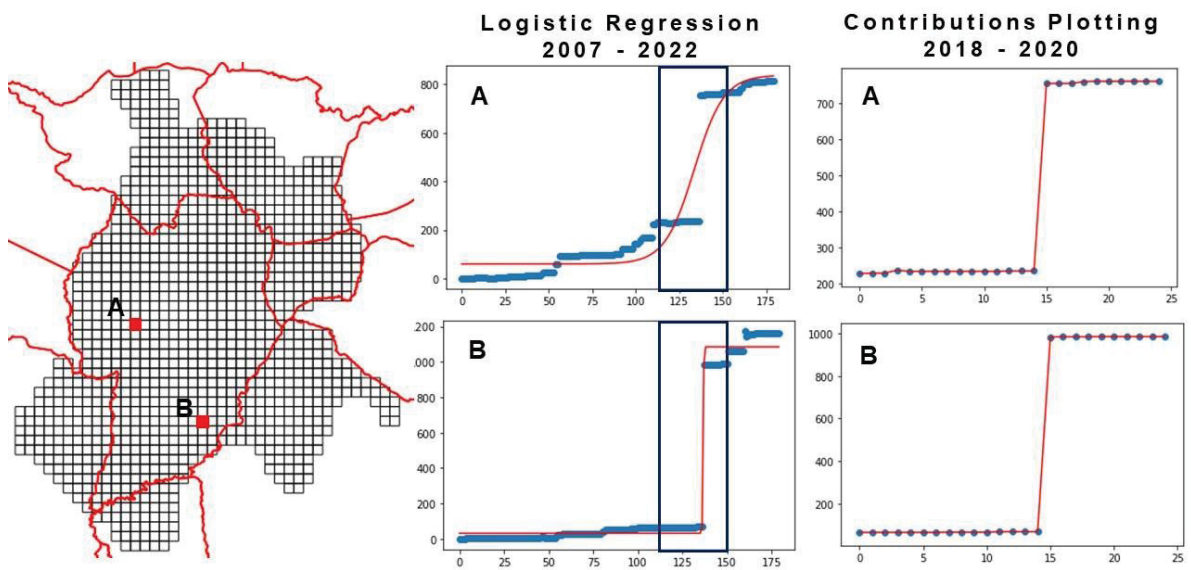
The import of the addresses in Curitiba occurred between 2018 and 2019. Therefore, creating graphs of the accumulated monthly quantity of features from 01/01/2018 to 01/01/2020 enabled identifying and evaluating the contribution patterns and their magnitude, in addition to the abrupt jumps behaviour. Figure 22 presents examples of the results obtained.

FIGURE 21 – BEHAVIOR OF LOGISTIC REGRESSION IN DIFFERENT CELLS EVALUATED



SOURCE: The Authors (2022).

FIGURE 22 – BEHAVIOR EXAMPLES OF THE ABRUPT JUMPS IN THE CONTRIBUTION PATTERNS IN DIFFERENT CURITIBA CELLS

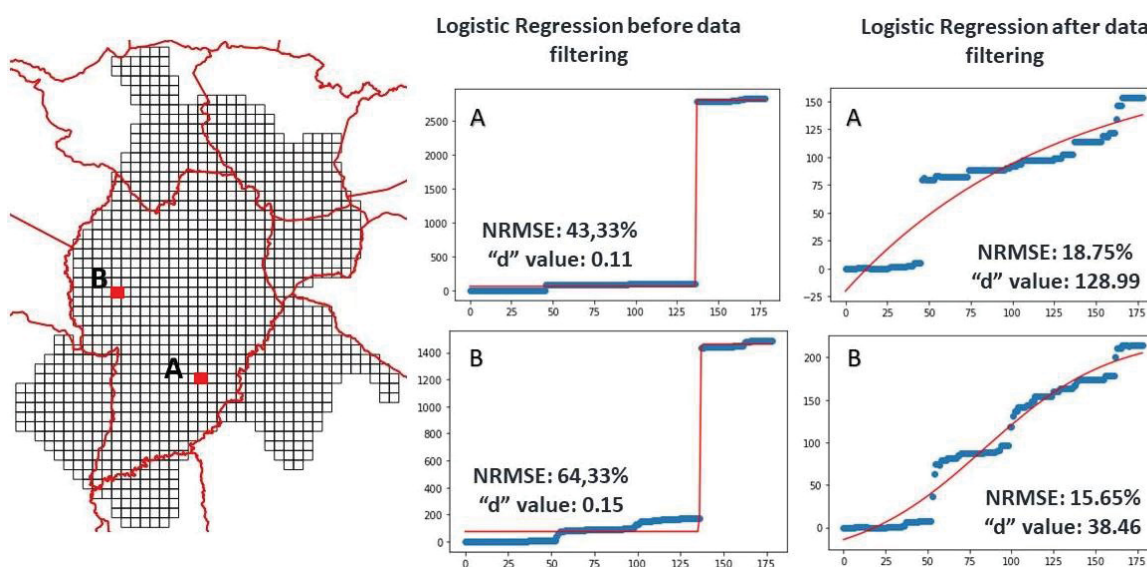


SOURCE: The Authors (2022).

With the graphs shown in Figure 22, it is possible to note that address imports influenced the total accumulated contributions. In region A, 519 features were added in a single month during the period (2018-2020). This amount corresponds to 63.4% of the total accumulated number of features by 2022. In region B, the number of contributions in a single month was 916 features, corresponding to more than 78% of the available features. In addition, the logistic regression presented the overflow message, indicating the abrupt jump.

Figure 23 presents the behaviour of the logistic curve in different cells before and after the filtering of the OSM tag key "addr:street".

FIGURE 23 – LOGISTIC REGRESSIONS IN DIFFERENT AREAS BEFORE AND AFTER FILTERING THE ADDRESS POINTS



SOURCE: The Authors (2022).

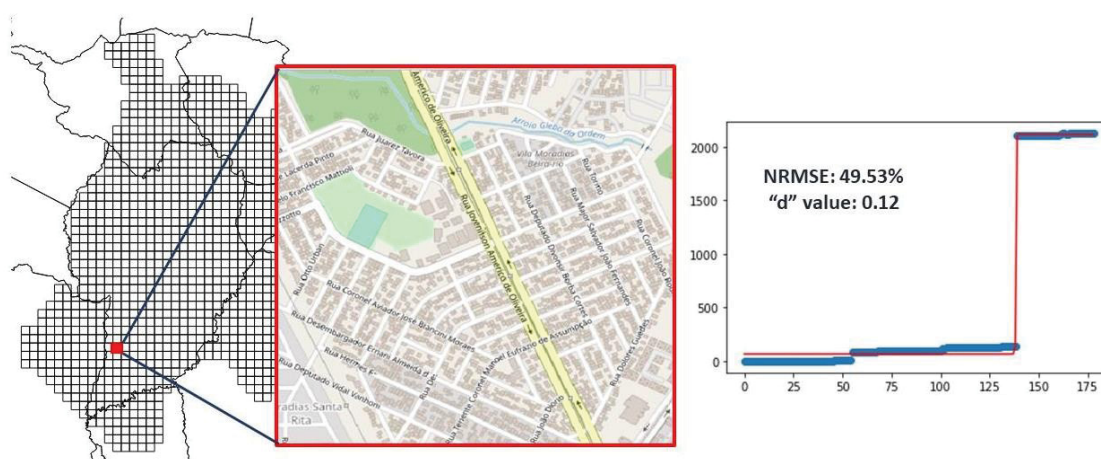
With the graphs shown in Figure 23, it is possible to notice that, from the data filtering, it was possible to obtain more smooth curves in the regression calculation. An important aspect is that in graph A, even though a significant contribution represents a jump, it was not modelled in the regression after it occurred with gradual contributions. Besides that, in the mathematical modelling that contemplates all the contributions, the high number of inserts in a single month made the regression indicate a possible data saturation, both in graphs A and B. This aspect was not evidenced after the filtering, which indicates that the regions are still growing (Zone II). Another question was about the values obtained for parameter "d." After filtering the data, it was possible to notice a significant increase in its value. In this context, in

steeper curves, the slope tends to zero (Figure 15), which allows for identifying areas of high growth in a short period or even measuring the behaviour of the contributions. It was also possible to analyse regions influenced by significant contributions or events besides inserting address points. In this context were identified, in Curitiba, cell clusters in which the contribution percentage in a single month was equivalent to more than 80% of the total amount.

In the region comprising Figure 24, it was observed that the described behaviour was buildings inserted simultaneously in a collective mapping activity in 2019, in which the set of changes was called "Mapeo Colectivo realizado en el Marco del Master en Desarrollo Urbano y Territorial (MDUT) Universidad Politecnica de Catalunya (UPC) #MapPyOSM." This characteristic evidenced the abrupt growth observed in the logistic regression, besides having 1962 features inserted in a single month, equivalent to more than 90% of the accumulated total.

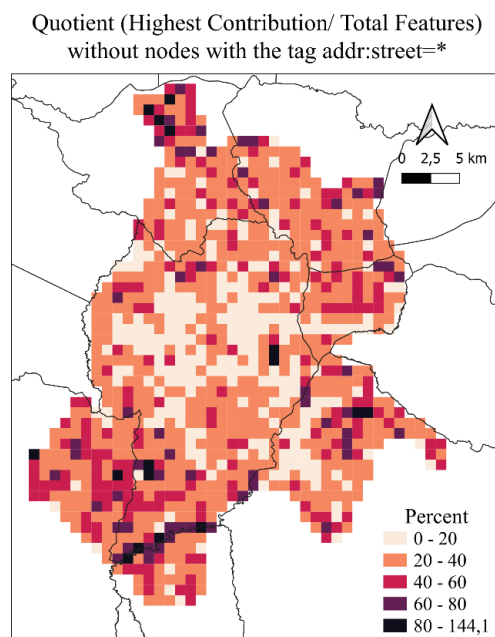
Therefore, when observing the distribution of the contributions accumulated over time in heterogeneous regions, it is possible to observe patterns and behaviours that may help understand the dynamic of OSM contributions in a specific time and space. For example, the map in Figure 25 shows the spatial distribution of the month's percentage with the most contributions concerning the accumulated total after filtering the OSM address tags.

FIGURE 24 – CELL IN WHICH THE COLLECTIVE MAPPING ACTIVITY TOOK PLACE.



SOURCE: The Authors (2022).

FIGURE 25 – CONTRIBUTIONS IN THE HIGHEST MONTH OVER THE TOTAL
(AFTER FILTERING)



SOURCE: The authors (2022).

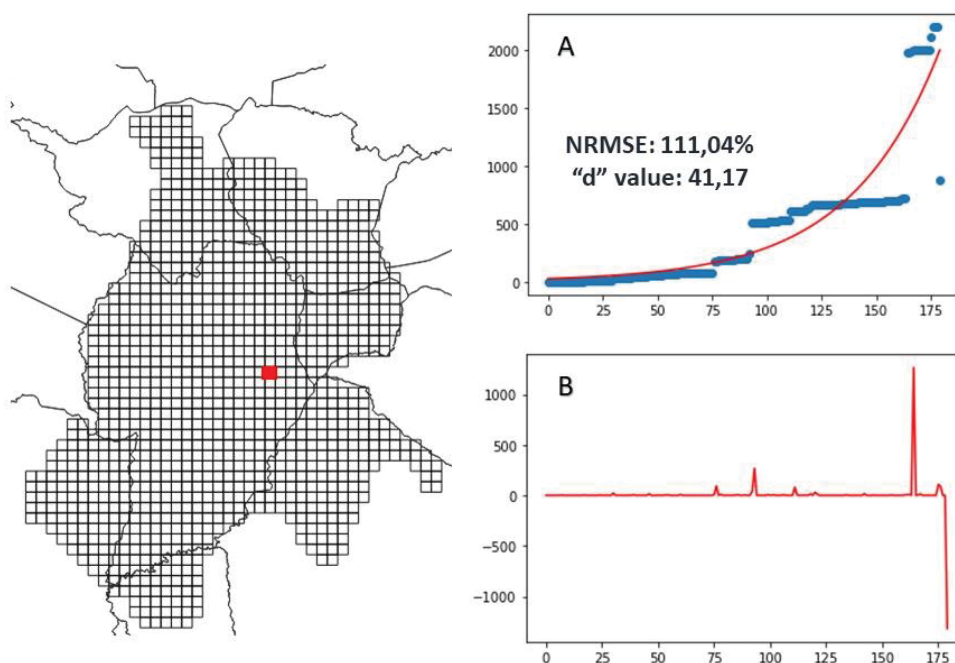
When comparing the answers obtained in Figure 25 about Figure 21B, it is possible to notice that there was a decrease of cells with peaks in the insertion of features in Curitiba, demonstrating quantitatively that the address points have a significant weight in the municipality analysis. This decrease made the results of Curitiba's cells equivalent to the total of the NUC. The predominant percentage was between 0% and 40%, obtained in more than 75% of the total cells.

A critical issue observed in Figure 26 was the presence of two cells in a specific region of Curitiba where the calculated percentage exceeded 100%. In these regions, it was noticed that there was a representative exclusion of features in the most recent month evaluated. Figure 26 shows the data pattern in this region. Graph A of Figure 26 presents the adjustment of the data from the logistic regression, and graph B corresponds to the number of features inserted or excluded by month. In the most recent month, 1319 features were excluded, and this value converges to the most significant number added (1265 features). This fact is most apparent in graph B, considering that there are two extreme peaks of addition and exclusion. However, the curve in graph A indicates that this is a typical growing cell (Zone II). The modelling of the curve has not yet adapted to this abrupt decrease, probably because the exclusion was in the later epochs of the study period and of the short relative time gap between the contribution spike and its exclusion. Nevertheless, this

characteristic corroborates the initial premise of data heterogeneity and that the Spatio-temporal patterns can vary according to the evaluated period and region. All analysis results, with graphs for the 1074 cells of the region, can be interactively consulted at https://bit.ly/NUC_notebook.

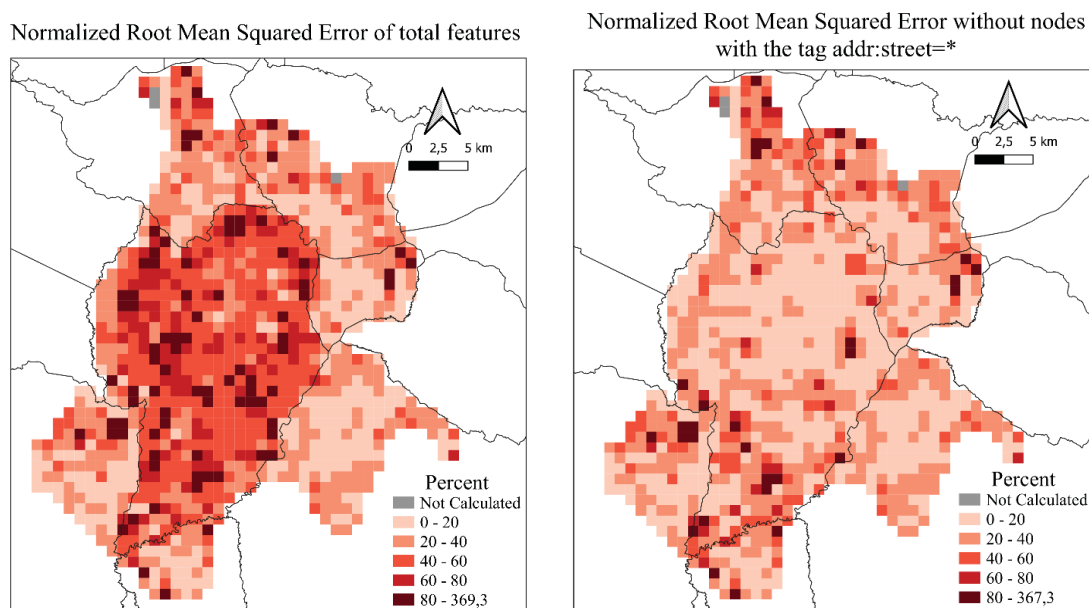
The maps shown in Figures 27A and 27B depict the spatialisation of the NRMS obtained in each cell before and after filtering the "addr:street" tag key.

FIGURE 26 – CELL IN WHICH REPRESENTATIVE EXCLUSION OF FEATURES WAS IDENTIFIED



SOURCE: The authors (2022).

FIGURE 27 – NRMS OBTAINED BEFORE (A) AND AFTER FILTERING THE ADDRESSES (B).



SOURCE: The authors (2022).

When evaluating the regression quality through each cell NRMS, the first finding is that the addresses imports impacted the results obtained in Curitiba. In the analysis with all inserted features (Figure 27A), the mean and median NRMS were 41.247% and 35.7%, respectively. After filtering the addresses (Figure 27B), these values decreased to 27.536% and 20.6%, respectively. Therefore, there was an improvement in the regressions obtained, mainly due to the mass data importation to Curitiba. While in Curitiba, we observed that the NRMS was reduced after address filtering, there were no significant variations in the other NUC regions. The mean and median NRMS values of the total number of features found in the cells intersecting Curitiba were 53.325% and 49.2%, while in the other NUC regions, 29.546% and 22.2% were found. After address filtering, the results reached 25.455% and 19.1% in Curitiba and 29.552% and 22.2% in the other locations.

In addition, three cells (represented in grey in Figure 27) failed in the NRMS calculation, as the median was zero. The curve failed to model the contributions in these regions due to the low amount of data input and the accumulated growth without a typical pattern. As explained in the previous analysis of Figure 20, even if it is not an abrupt jump, regions with few features or unusual contribution behaviour over time can also present this behaviour. After address filtering, the overflow no longer appeared in Curitiba, comparable to the pattern observed in other NUC

regions (Figures 21 and 23). This last analysis also reinforces the use limitation of the presented method for areas where the number of contributions and their behaviour are compatible with the logistic curve.

4.4 CONCLUSIONS AND RECOMMENDATIONS

This work sought to develop an open-source methodology and tools to obtain the Spatio-temporal patterns of OSM and its mathematical modelling using logistic regression. The results point to spatial data heterogeneity since the evolution of the contributions varies according to the region's characteristics. Among the causes of this variation is the existence of official open data for import in OSM, the initiatives to encourage collaborative mapping, and the characteristics of the population and urbanisation of each area. The abrupt jumps detected are not negative aspects, as they show areas where data growth is growing faster, either by the availability of open data or by focused actions of mapping communities in specific areas. However, such contributions differ from the organic growth in areas where such phenomena do not occur. Furthermore, by verifying the parameters of the steepness of the logistic regression, it was possible to identify these aspects and the areas affected by significant accumulated contributions in a single period over time. Besides that, it was possible to notice that the contributions' dynamic can vary in the same region, which corroborates the prerogative that Spatio-temporal aspects in VGI can be directly related to the quality of geospatial data, especially regarding completeness and temporality.

Regarding the contribution modelling, it was possible to notice that the temporal aspects can be described through mathematical functions and by identifying the dynamics related to the evaluated region. In this context, for future works, it is recommended to continue research on the different factors related to the insertion of features in VGI over time, including other forms of mathematical modelling and other variables. In addition, it is recommended to explore the parameters found when obtaining the regressions. Also, there is a recommendation to perform spatial statistical analyses to classify the contribution patterns, which are the spatial characteristics of the distribution of such patterns and their relationships with other variables of the description of the territory and the collaborative mapping.

4.5 REFERENCES

Arsanjani, J.J. et al. 2015. An exploration of future patterns of the contributions to OpenStreetMap and development of a Contribution Index. *Transactions in GIS*, 19(6), pp. 896-914.

Brovelli, M.A. and Zamboni, G. 2018. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of Geo-Information*, 7(8), pp. 1-25.

Brückner, J., Schott, M., Zipf, A., and Lautenbach, S. 2021. Assessing shop completeness in OpenStreetMap for two federal states in Germany. *AGILE GIScience Series*, 2, p. 20.

Camboim, S.P.; Bravo, J.V.M.; and Sluter, C.R. 2015. An investigation into the completeness of, and updates to, the Open Street Map data in a heterogeneous area in Brazil. *ISPRS International Journal of Geo-Information*, 4(3), pp. 1366-1388.

COMEC – Coordenação da Região Metropolitana de Curitiba. 2022. A Região Metropolitana de Curitiba. Available at: <<https://www.comec.pr.gov.br/Pagina/Regiao-Metropolitana-de-Curitiba>> [Accessed 13 October 2022].

Diretoria do Serviço Geográfico (DSG), 2015. Especificação Técnica para Controle de Qualidade de Dados Geoespaciais Vetoriais (ET-CQDG). Brasil.

Grinberger, A.Y. et al. 2021. An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS*, 25(2), pp. 622-641.

Gröching, S., Brunauer, R., and Rehl, K. 2014. Digging into the history of VGI datasets: Results from a worldwide study on OpenStreetMap mapping activity. *Journal of Location Based Services*, 8(3), pp. 198–210.

Haklay, M. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4), pp. 682-703.

Haklay, M., Basiouka, S., Antoniou, V., and Ather, A. 2010. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal*, 47(4), pp. 315–322.

IBGE – Instituto Brasileiro de Geografia e Estatística. 2021. Estimativas de População. Available at: <https://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2021/estimativa_dou_2021.pdf> [Accessed 13 October 2022].

IBGE – Instituto Brasileiro de Geografia e Estatística. 2022. Censo Demográfico 2022 - IBGE. Available at: <<https://censo2022.ibge.gov.br/sobre/geografia-censitaria/enderecamento.html>> [Accessed 29 December 2022].

IPPUC – Instituto de Pesquisa e Planejamento Urbano de Curitiba: Dados Geográficos. Available at: <<http://www.ippuc.org.br/geodownloads/geo.html>> [Accessed 13 October 2022].

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

Kitchin, R. 2015. The opportunities, challenges and risks of big data for official statistics, *Statistical Journal of the IAOS*, 31(3), pp. 471–481.

Le Guilcher, A., Olteanu-Raimond, A.-M. and Balde, M.B. 2022. Analysis of Massive Imports of Open Data in OpenStreetMap Database: A Study Case for France, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, V-4-2022, pp.99–106.

Nasiri, A. et al. 2018. 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*, 7(7), p. 253.

Paiva, C. dos A. and Camboim, S.P. 2021. A Dinâmica de Colaborações OpenStreetMap e sua Relação com as Atividades de Uso e Ocupação do Solo: um Estudo Segundo Zoneamento de Curitiba. *Revista Brasileira de Cartografia*, 73(1), pp. 73-87.

Paiva, C. dos A. and Camboim, S.P. 2022. Inference of positional accuracy of collaborative data from intrinsic parameters. *Transactions in GIS*, 26(4), pp. 1898–1913.

RAIS - Annual Report of Social Information. 2021. Available at: <http://www.rais.gov.br/sitio/sobre.jsf>. [Accessed 29 December 2022].

Robinson, A.C. et al. 2017. Geospatial big data and cartography: Research challenges and opportunities for making maps that matter. *International Journal of Cartography*, 3(sup1), pp. 32–60.

Sehra, S.S., Singh, J. and Rai, H.S. 2017. Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. *Future Internet*, 9(2), p. 15.

Silva, L.S.L. and Camboim, S.P. 2020 Brazilian NSDI ten years later: current overview, new challenges and propositions for national topographic mapping. *Bulletin of Geodetic Sciences*. 26(4) p. e2020018.

Teimoory, N., Abbaspour R.A. and Chehreghan A. 2021 Reliability extracted from the history file as an intrinsic indicator for assessing the quality of OpenStreetMap. *Earth Science Informatics*, 14 (3), pp. 1413–1432.

UN-GGIM. 2022. United Nations Integrated Geospatial Information Framework (UN-IGIF) Available at: <<https://ggim.un.org/IGIF/part1.cshtml>> [Accessed 29 December 2022].

Witt, R., Loos, L., and Zipf, A. 2021. Analysing the Impact of Large Data Imports in OpenStreetMap. *ISPRS International Journal of Geo-Information*, 10(8), p. 528.

Tam, S.-M. and Van Halderen, G. 2020. The five V's, seven virtues and ten rules of big data engagement for official statistics. *Statistical Journal of the IAOS*, 36(2), pp. 423–433.

Zhang, H. and Malczewski, J. 2018. Accuracy Evaluation of the Canadian OpenStreetMap Road Networks. *Internacional Journal Geospatial and Environmental Research*, 5, pp. 1-14.

5. OS PADRÕES ESPAÇO-TEMPORAIS DE CONTRIBUIÇÕES E A QUALIDADE DOS DADOS GEOESPACIAIS NO OSM.

Conforme abordado na seção 1, este capítulo apresenta quatro resumos expandidos, publicados em eventos, na qual são discutidos resultados que relacionam os padrões espaço temporais de contribuições do OSM e a qualidade dos dados geoespaciais. Nos referidos trabalhos, todas as aplicações foram derivadas dos produtos obtidos nos artigos dos capítulos 3 e 4, além de destacar a heterogeneidade dos dados e a sua relação com a disposição final dos resultados.

Nas subseções 5.1 e 5.2, são apresentadas complementações das ferramentas desenvolvidas no capítulo 3. A subseção 5.1 consiste no desenvolvimento de uma nova ferramenta para a avaliação da completude de eixos viários do OSM de forma automática. As análises são realizadas tanto a partir da sua geometria, quanto no percentual de atributos preenchidos. Os resultados são computados a partir de uma grade com células na região de estudo e o usuário pode selecionar o melhor tamanho que se adequa ao seu propósito. Na subseção 5.2, é disposta uma nova forma de visualização dos resultados da acurácia posicional da amostra de feições, por meio da interpolação das discrepâncias. Em paralelo, o complemento QPEC (capítulo 3) foi atualizado e, na análise de feições pontuais, permite-se com que o usuário escolha uma escala específica para obtenção dos resultados ou exibi-las para todas as escalas do Mapeamento Sistemático Brasileiro. Além disso, foram computados os comportamentos das contribuições ao longo do tempo em células de 5x5 km de todas as capitais dos Estados brasileiros. Verificou-se que as diferentes áreas são influenciadas por aspectos específicos, que podem estar relacionados com grades contribuições ou importações de dados para o OSM.

A identificação destes aspectos permitiu estabelecer parâmetros para a modelagem dos dados.

Os subitens 5.3 e 5.4 são caracterizados pela modelagem das contribuições a partir da Regressão Logística, apresentada na seção 4. Dessa forma, por meio dos parâmetros encontrados, permitiu-se obter relações com a qualidade intrínseca no OSM e a completude dos dados. Na subseção 5.3 identificou-se que determinadas regiões Curitiba-PR possuem mesclas de contribuições gradativas ao longo do tempo e outras caracterizadas apenas por grandes contribuições, realizadas em um único período. Notou-se que o padrão encontrado foi semelhante nas regiões centrais do município e ia se diferenciando à medida que eram analisadas áreas mais afastadas. Por fim desenvolveu-se uma interface para visualizar os padrões de contribuição em diferentes regiões do Brasil, a partir da seleção do retângulo envolvente do local desejado (também apresentada na seção 4). Na subseção 5.4 foram utilizados os parâmetros obtidos na regressão logística para obter a completude dos dados do OSM nas primitivas geométricas de eixos viários e edificações. Foram avaliados elementos relacionados com a saturação da curva nos últimos anos de contribuição, inclinação da curva e comportamento da assíntota superior. Avaliou-se o potencial desses parâmetros, nas categorias descritas, que se mostraram promissores nos eixos viários.

Além dos resumos expandidos, também são apresentadas as subseções 5.5 e 5.6 que descrevem aplicações relacionadas com a avaliação da completude e da acurácia posicional. A subseção 5.5 aborda um novo processamento dos dados descritos na subseção 5.4 e uma nova avaliação da completude. Além disso, é realizada a espacialização dos resultados das feições classificadas como edificações e vias (tags *buildings* e *highways*) no OSM em Curitiba. A subseção 5.6 descreve o desenvolvimento de uma nova ferramenta para a avaliação da acurácia posicional de dados geoespaciais de eixos viários com base no QPEC (seção 3). Além dos procedimentos já descritos, a ferramenta baseia-se na identificação de feições homologas referentes aos cruzamentos de vias, obtenção das discrepâncias e aplicação do PEC PCD.

5.1 AVALIAÇÃO E REPRESENTAÇÃO DA QUALIDADE DE DADOS GEOESPACIAIS A PARTIR DO DESENVOLVIMENTO DE FERRAMENTAS EM AMBIENTE SIG

Autores: Elias Nasr Naim Elias, Denis Leonardo Santos, Silvana Philippi Camboim, Luciene Stamato Delazari, Marcio Augusto Reolon Schmidt.

Evento: XI Colóquio Brasileiro de Ciências Geodésicas Curitiba, 06 a 08 de maio de 2020.

Abstract:

A qualidade de dados geoespaciais é um aspecto importante a ser considerado em mapeamentos, sejam estes produzidos para fins gerais (referência) ou para representar aspectos temáticos. Os mapeamentos de referência são gerados por agências oficiais de mapeamento com o propósito de representar todas as feições visíveis na paisagem [1] e feições territoriais de uso geral de forma que sirvam para uma variedade de usos [2]. Este tipo de informação foi por muito tempo, a única abordagem referente a produção cartográfica [3]. Contudo, os avanços tecnológicos computacionais têm permitido que qualquer indivíduo que tenha computador ou smartphone com acesso à internet possa gerar geoinformação [4]. Tal processo tornou-se possível a partir do estabelecimento do conceito de web 2.0, criado por [5]. De acordo com [6], define-se como *web 2.0* um fenômeno cibernético onde os usuários tornam-se elementos fundamentais na geração e gerenciamentos de determinadas informações, além de serem consumidores destas. Com o aumento das fontes de dados geoespaciais, incluindo plataformas de informação geográfica voluntária, e a participação do usuário neste processo torna-se importante que existam ferramentas para avaliar a qualidade destas informações e que estas estejam à disposição do usuário. Além disso, é viável que estas ferramentas se adequem com as necessidades dos produtores e consumidores de dados geoespaciais. Tratando-se da qualidade de dados espaciais, os indicadores para a avaliação de produtos cartográficos foram atribuídos pela [7] e revalidados pela [8]. Estes indicadores são caracterizados por: Acurácia Posicional; Acurácia Temática; Completude; Consistência Lógica; Qualidade Temporal; e Usabilidade. Na literatura são abrangentes os trabalhos associados com a qualidade de dados geoespaciais que utilizam os indicadores descritos e quando associados com a implementação de ferramentas para a avaliação automática, notou-se que, muitos trabalhos têm

concentrado esforços em verificar dados oriundos a partir de informações geográficas voluntárias. Embora existam diversas aplicações para a avaliação da qualidade de dados geoespaciais, são escassos os trabalhos em que o usuário, por meio de uma ferramenta em ambiente de Sistema de Informações Geográficas (SIG), possa visualizar, a partir da inserção de amostras de feições, tanto os resultados dos testes utilizados nas análises, quanto a distribuição espacial dos erros obtidos. Neste contexto, este trabalho objetiva a avaliação da acurácia posicional e da completude de feições de dados vetoriais a partir do desenvolvimento de uma aplicação geoespacial de código aberto, no *software QGIS 3.10*. Foram desenvolvidos três conjuntos de ferramentas no *software QGIS 3.10* para avaliar: Acurácia Posicional de Feições Pontuais; Acurácia Posicional de Feições Lineares; e Completude. A aplicação consistiu na combinação de processos pré-existentes no *software* em questão, a partir do *Model Builder*. Este corresponde a um modelador gráfico que permite desenvolver projetos a partir de uma interface simples. A sequência metodológica para a realização dos procedimentos compreendeu as seguintes etapas: Escolha dos indicadores a serem avaliados; análise das condicionantes estatísticas a serem utilizadas; utilização do *Model Builder* do *QGIS* para o desenvolvimento da aplicação; estabelecimento da visualização da qualidade geoespacial das feições e armazenamento dos complementos. Para a avaliação da qualidade posicional de feições pontuais, foram criadas duas funcionalidades. A primeira correspondeu ao método caracterizado pelos testes estatísticos estabelecidos por [9] e adaptados por [10] para avaliar tendência e precisão de conjuntos amostrais e a segunda correspondeu às condicionantes estabelecidas pela [11], aplicando o Padrão de Exatidão Cartográfica para Produtos Cartográficos Digitais (PEC-PCD). No que diz respeito às feições lineares, utilizou-se o método do buffer duplo, criado por [12] e adaptado para os padrões do PEC-PCD em 2015 por [13]. Foram criadas 9 funcionalidades, permitindo ao usuário selecionar uma das oito escalas do PEC-PCD ou entrar com o tamanho de buffer (erro) para identificar a escala que se adequa. A visualização da qualidade do conjunto amostral deu-se a partir das distâncias euclidianas obtidas entre os dados de referência e os avaliados, utilizando a subdivisão destas em quartis, com variação da luminosidade do tom da cor para representar os valores de discrepância. Para a completude, foi criada uma ferramenta de avaliação da qualidade física e de atributos para o *software* utilizado, sendo assim, utilizou-se o

método estabelecido por [14] e adaptado por [15]. Desta forma, a citada ferramenta apresenta duas funcionalidades, sendo a primeira um provedor da análise da qualidade física, sendo esta, a completude geométrica entre uma base de referência e dados provenientes da base *Open Street Map* (OSM), representado por uma grade regular, na qual a resolução é escolhida pelo usuário, com valores variando entre o intervalo de -1 a 1 expressando a desigualdade entre os dados, tomando como base os dados de referência. A segunda funcionalidade traz uma avaliação semântica, fazendo uma análise percentual de preenchimento das feições com o campo de nomenclatura das vias, sendo representada a partir de uma grade regular, com graduação a partir de 0% até 100%, onde, ambas funcionalidades têm variação na matiz da cor para representação visual final da avaliação. É importante ressaltar a otimização do processo de análise do dado avaliado, visto que, a informação resultante é armazenada em arquivos vetoriais, permitindo realizar análises espaciais com outros parâmetros associados a características da área de estudo utilizada. Recomenda-se para trabalhos futuros a continuidade da criação de ferramentas com os demais indicadores de qualidade, bem como, a realização de testes com usuários para que as interfaces de inserção de dados e a disposição dos resultados esteja de acordo com as suas necessidades.

Palavras-chaves: Aplicação Geoespacial; Qualidade de Dados; Acurácia Posicional; Completude.

5.1.1 Referências

[1] KEATES, J. S. *Cartographic Design and Production*. UK: Longman , 1973. 240p.

[2] KENT, A. J. *Topographic Maps: Methodological Approaches for Analyzing Cartographic Style*. *Journal of Map & Geography Libraries* , vol. 5, n. 2, 2009. pp. 131– 156.

[3] 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*, v. 102, n. 3, p. 571-590, 2012.

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

[5] 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.

[6] CORMODE, G.; KRISHNAMURTHY, B. Key differences between Web 1.0 and Web 2.0. *First Monday* , vol.13, n.6, 2008.

[7] ISO 19114: Geographic Information - Quality Evaluation Procedures . International Organization for Standardization, 2001.

[8] ISO 19157: Geographic Information - Data Quality . International Organization for Standardization, 2013.

[9] MERCHANT, D. C. Spatial Accuracy Standards for Large Scale Line Maps. In: *Technical Congress on Surveying and Mapping . Proceedings*. v. 1, 1982.

[10] GALO, M., DAL POZ, A. P.; FERREIRA, F. M. Uso de Feições no Controle de Qualidade em Cartografia. In: *CBC - Congresso Brasileiro de Cartografia* , 19. Porto Alegre. Anais... Porto Alegre, Brasil, 2001.

5.2 DATA QUALITY ANALYSIS TOOLS AND THEIR USE FOR THE DETECTION OF SPATIO-TEMPORAL HETEROGENEITIES IN VGI CONTRIBUTIONS

Autores: Elias Nasr Naim Elias, Fabricio Rosa Amorim, Marcio Augusto Reolon Schmidt and Silvana Philippi Camboim.

Periódico ou Evento: **GeoMundus Conferece (2020)**

Abstract:

The quality of data from VGI platforms is a recurring theme in recent research. The spatiotemporal distribution of contributions on such platforms is quite heterogeneous, depending on several factors such as the availability of inputs, the

number and motivation of volunteers, among other causes that are still little explored in the literature, especially in developing countries. The present work aims the development of tools to evaluate and visualize quality heterogeneity of the available features on the OpenStreetMap's (OSM) Volunteered Geographic Information (VGI) platform in Brazil. The analyzes were developed using two tools addressing different parameters of data quality. The first one referred to the identification of the history of editions evolution between 2008 and 2020 in different regions from Brazil. The second one focused on the spatialization of the positional discrepancies of the OSM features, in the city of Salvador. In terms of temporal distribution, it was noted that there are not homogeneous patterns, since some regions presented linear growth behaviour. In others, there was a large number of mass contributions. In discrepancy's analysis, it varied from 0.27m to 11.27m, which indicates the heterogeneity of positional data quality. These are the initial results on the development of a toolset to provide users visualization and analysis of these parameters on open platforms.

Keywords: Geospatial Data Quality, Data Heterogeneity, OpenStreetMap.

5.2.1 Introduction

On the steps that permeate the generation of geospatial data, one of the most critical and discussed topics by the scientific community concerns the development of methodologies for the evaluation and representation of the quality of this type of data. It is essential to consider that through time, GIS users began to notice that in most cases, they did not know the quality of the mapped features, which made it difficult to use them in some applications (KOUKOLESTSOS, 2012). In this context, Devillers and Jeansoulin (2006) pointed out the existence of problems related to the evaluation of the quality in geospatial data. They highlighted its intensification from the technological advances and the availability of spatial information in web platforms, beyond the different applications related to the use of spatial data in digital platforms.

About the parameters for evaluating the quality of geospatial data, a series of standards around the world provide the indicators to evaluate the quality of a cartographic product. One of the first standards that approach the quality of

geospatial data was the Spatial Data Transfer Standard (SDTS), developed in 1992 and implemented by several governmental and private institutions in The United States (AL-BAKRI, 2012). According to Al-Bakri (2012), the mapping standards around the world that were developed based on the standards established by STDS to define quality aspects, such as Australian and New Zealand Land Information Council (ANZLIC) and the European Committee Standardization (CEN). The most recent definition is provided by the International Organization for Standardization (ISO) 19.157 (ISO, 2013) in which are approached the indications of positional accuracy, thematic accuracy, completeness, logical consistency and usability.

With the development of technology, it is now possible to obtain spatial data from different sources. The creation of such information, previously restricted to specialists, is now possible through contributions from lay users through collaborative platforms Volunteered Geographic Information (VGI), such as defined by Goodchild (2007). In this context, it is vital to incorporate procedures for quality evaluation, as the diverse sources that contemplate these methodologies (Brovelli et al., 2019). A bibliometric review about the quality of the geospatial data performed by Bielcka and Burek (2019), emphasized that, since 2004, the geospatial data provided by volunteers has become the object of extensive research and these studies analyze, not only the different types of errors but also the behaviour of the volunteers in a particular area.

Data quality on VGI platforms has been the subject of extensive research, being a recurring objective in this field to integrate such data into authoritative geospatial databases. Among these research, it is possible to quote the research of Jasin and Hamandani (2020), Ibrahim, Ramadan and Hefny (2019), Tian, Zhou and Fu (2019), Minghini and Frassinelli (2019), Brovelli and Zamboni (2018). These works use traditional quality parameters described by the ISO 19.157 (ISO, 2013) as well as additionally the use of parameters intrinsic to VGI data, such as the history of editions, the number of users and collaboration patterns.

Some recent research that involve the VGI quality parameters have been trying to comprehend how much the characteristics and the dynamics of the contributions interfere in its quality (DROR, DOYTHER and DALTOY, 2020, NASIRI et al., 2018). One aspect identified in these works is the heterogeneity of the data, where the quality parameters can vary according to the study area. In this context, for example, in the same area, the features from VGI can be fully mapped in a portion

and also have considerable voids in another portion, or the positional precision may vary from 0.10cm to 10m. In developing countries, the number of researches is smaller but, studies like those from Camboim, Bravo and Sluter(2015) indicate an even more explicit heterogeneity since even the official mapping is quite fragmented and dispersed over time. Since these heterogeneities are so present in the VGI data, we propose that if we can analyze and visualize these discrepancies using open source tools, it will be possible to understand their causes better. In this way, we can identify the elements that affect the quality of the OSM data and enable better decision making regarding the use of this information for a given purpose.

5.2.2 Methodology

To work with positional and temporal quality parameters, we divided the analysis into two phases. The first step corresponded to a preliminary evaluation of the behaviour of the history of editions of the OSM's VGI platform in Brazil. The data for the two selected Brazilian urban regions (5x5 km rectangles), from the years 2008 to 2020, come from the University of Heidelberg's OHSOME Application Programming Interface (API)(<https://heigit.org/big-spatial-data-analytics-en/ohsome/>). We created an application using the python language to compute and generate graphs showing the evolution of updates over time.

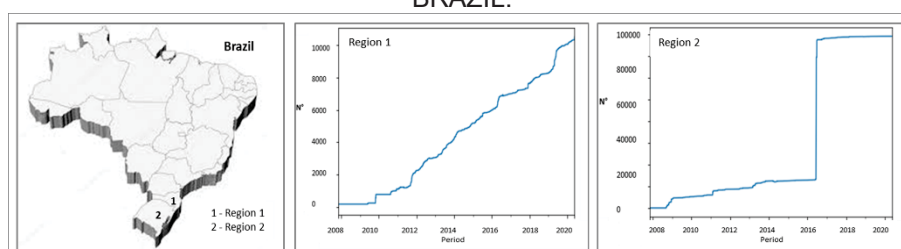
The second tool developed allows the comparison of data from a reference base with the OSM data. It was developed as a plugin for the QGIS software also using the python language. To demonstrate its application, we used 20 points in the city of Salvador, capital of the state of Bahia. In this analysis, the Euclidean distance between homologous points in the two bases is calculated. These differences were later interpolated on a surface to demonstrate the variation in the positional quality of OSM data in the city.

5.2.3 Results

The temporal behaviour of the editions in the study areas analyzed with the first tool was quite distinct. Despite having equal areas and being all urban, one could have more than 100,000 contributions, while the others did not exceed 2,000.

Additionally, it was noted that the pattern of contributions over time could vary from area to area. Figure 28, for example, shows two OSM regions that presented different growth patterns. While one of them remained linear throughout time, another presented a mass contribution in 2016. It is essential to consider that these patterns can help to understand better the way in which data were obtained over time in a given location and help to infer the impact of this behaviour on the quality of the available data.

FIGURA 28 – EVOLUTION OF OSM CONTRIBUTIONS FROM 2008 TO 2020 IN TWO AREAS IN BRAZIL.

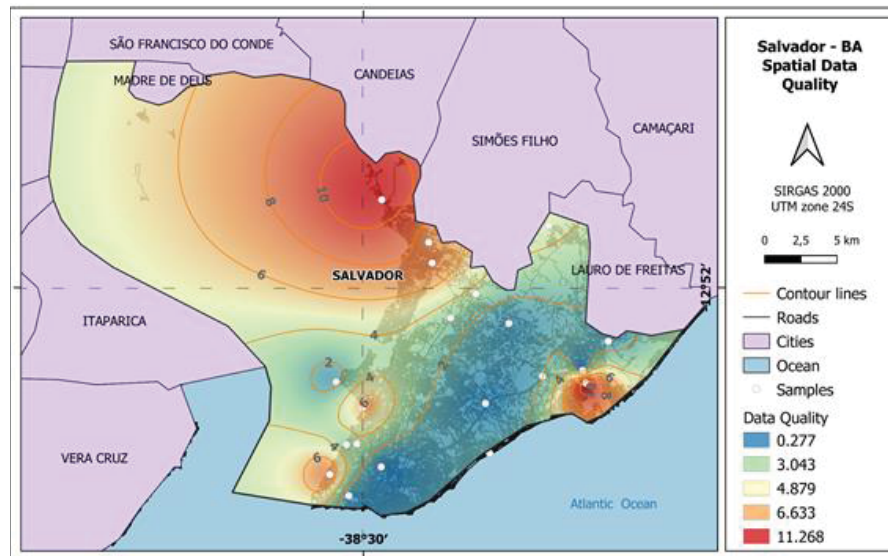


FONTE: The Authors (2020).

The graphic differences between regions 1 and 2 reveal discrepancies in terms of the behaviour of the editions. On region 1, the linear growth reflects that the OSM input features were made by gradual contributions of users and regions, while on region 2, the existence of mass contributions caused stability in the number of contributions from 2016.

However, even within these analyzed regions, the data obtained can be heterogeneous. The second tool allows the comparison with official reference bases in order to allow the visualization of positional quality in a region. Figure 29 presents the interpolation of the euclidean distances obtained on the evaluation of the positional accuracy of the OSM's point and linear features of the county of Salvador-BA. We can see that the discrepancies are not evenly distributed throughout the city, but we find regions where the results of positional accuracy are consistently worse. The absolute values in the samples varied from 0.27m up to 11.27m. Tools like this plugin can help users understand the differences in data quality parameters and help advance the understanding of what causes this local inequality.

FIGURE 29 – DISCREPANCIES BEHAVIOUR IN SALVADOR-BA



SOURCE: The Authors (2022).

5.2.4 Final Considerations

The results of these analyzes corroborate the need for open tools that provide users with knowledge about the spatio-temporal quality of collaborative mapping data. Future studies on the distribution of these parameters can help to model the causes of such heterogeneities. Thus, it is possible to outline actions that contribute to improving the mapping in underrepresented areas. Additionally, by knowing more about the nature of the distribution of contributions, it is possible to plan the integration of these data with official maps and contribute to mapping in Brazil that, due to its extension and lack of investment in cartography, require updated geospatial data.

5.2.5 References

Al-Bakri. Developing Tools and Models for Evaluating Geospatial Data Integration of Official and VGI Data Sources. 2012. PhD. School of Civil Engineering and Geosciences, Newcastle University.

Bielecka, E., Burek, E. (2019). Spatial data quality and uncertainty publication patterns and trends by bibliometric analysis. *Open Geosciences*, 11(1), pp. 219-235.

Brovelli, M. A.; Boccardo, P.; Bordogna, G.; Pepe, A.; Crespi, M.; Munafò, M.; Pirotti, F. Urban Geo Big Data. 2019. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, FOSS4G 2019 - Academic Track*. Bucharest: Romania.

Brovelli, M. A., Zamboni, G. 2018. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of GeoInformation*, 7(8), pp. 1-25.

Camboim, S.P.; Bravo, J.V.M.; Sluter, C.R. 2015. An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. *ISPRS International Journal of GeoInformation*, 4, pp 1366-1388.

Devillers, R., Jeansoulin, R. 2006. Spatial Data Quality: Concepts. In *Fundamentals of Spatial Data Quality* R. Devillers and R. Jeansoulin, (EDS). Ch. 2, pp.31-42. London: ISTE Ltd.

Dror T., Doytsher Y., Dalyot S. 2020. Investigating the Use of Historical Node Location Data as a Source to Improve OpenStreetMap Position Quality. In *Open Source Geospatial Science for Urban Studies* A. Mobasher (EDS). *Lecture Notes in Intelligent Transportation and Infrastructure*. Springer, Cham.

Goodchild, M. F. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), pp. 211-221.

Ibrahim, M. H., Darwish, N. R., Hefny, H. A. 2019. An Approach to Control the Positional Accuracy of Point Features in Volunteered Geographic Information Systems. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 10(6), pp. 169-175.

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

Jasim, S., Al-Hamadani, O. 2020. Positional Accuracy Assessment for Updating Authoritative Geospatial Datasets Based on Open Source Data and Remotely Sensed Images. *Journal of Engineering*, 26(2) pp. 70-84.

Koukoletsos, T. 2012. A Framework for Quality Evaluation of VGI linear datasets. PhD., University College London.

Minghini, M., Frassinelli, F. 2019. OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data, Software and Standards*, 4(1), pp. 9.

Nasiri, A.; Ali Abbaspour, R.; Chehregan, A. Jokar Arsanjani, J. 2018. Improving the quality of citizen contributed geodata through their historical contributions: The case of the road network in OpenStreetMap. *ISPRS International Journal of GeoInformation*, 7(7), pp. 253.

Neis, P., Zipf, A. 2012. Analyzing the contributor activity of a volunteered geographic information Project —The case of OpenStreetMap. *ISPRS International Journal of GeoInformation*, 1(2), pp. 146-165.

Tian, Y., Zhou, Q., Fu, X. 2019. An Analysis of the Evolution, Completeness and Spatial Patterns of OpenStreetMap Building Data in China. *ISPRS International Journal of GeoInformation*, 8(1), pp. 35.

5.3 DISCOVERING SPATIOTEMPORAL PATTERNS ON DATA QUALITY ASSESSMENT IN COLLABORATIVE MAPPING: A PRELIMINARY STUDY IN AN AREA OF BRAZIL

Autores: Elias Nasr Naim Elias, Fabricio Rosa Amorim, Marcio Augusto Reolon Schmidt, Silvana Philippi Camboim.

Periódico/Evento: Abstracts of the International Cartographic Association, 3, 2021. 30th International Cartographic Conference (ICC 2021), 14–18 December 2021, Florence, Italy.

Keywords: VGI, OpenStreetMap, Spatiotemporal Patterns, Geospatial Data Quality.

Abstract:

Technological and computational advances have enabled geospatial data to be obtained and updated daily in recent decades. This aspect characterizes big data, where data flow in digital media comes from different sources, such as numerical modelling, use of smartphones and internet access (YANG *et al.*, 2017). Given the aspects above, it is essential to consider the emergence of new data sources and methodologies for obtaining geospatial data, different from the classical approach associated with topographic mapping. As discussed by Brovelli, *et al.*, 2019 there are currently different methods for obtaining geospatial data, ranging from topographic and aerial surveys to the use of Volunteered Geographic Information (VGI). The authors emphasize that it is essential to establish methodologies to assess quality, extract and integrate relevant information from different data sources.

Quality is an important aspect to be considered in geospatial data, as it allows determining its suitability for use for specific purposes. In VGI, this issue becomes even more mitigating because the data have heterogeneous aspects, as they can vary according to the location, characteristics of collaborations and collaborators profile. Therefore, quality assessment procedures in the VGI are described from extrinsic approaches, through parameters established by the International Organization for Standardization 19.157 (ISO, 2013) (BROVELLI and ZAMBONI, 2018; ZHANG and MALCZEWSKI, 2017; HAKLAY, 2010) intrinsic (SEHRA, SINGH and RAI, 2017), characterized by the history of editions, the number of contributors and contributions; or even by combining extrinsic and intrinsic parameters (NASIRI *et al.*, 2018).

The assessment of intrinsic parameters is even more relevant in developing countries, where the chronic lack of resources for cartography often results in a lack of up-to-date data to provide a comparison. In addition, this situation makes data coming from VGI even more necessary to complement existing reference mappings (CAMBOIM, BRAVO and SLUTER, 2015). For this, understanding the quality of data in these regions is essential. Therefore, this research aims to go beyond the usual intrinsic parameters, such as the number of contributions and contributors, but to understand how the evolution of these parameters over time can be characterized as an additional quality measure.

Research on this topic has sought to understand how the dynamics of spatiotemporal patterns in the history of contributions can be measured and modelled to extract relevant information. Grinberger *et al.* (2021) addressed this aspect in work, who measured amounts of contributions over time to detect events on the OpenStreetMap (OSM) platform. Furthermore, Brückner *et al.* (2021) estimated the completeness of OSM retail stores. Both works were based on the accumulated number of contributions over time and used the logistic regression model in their data. This model is described from an "S"-shaped trajectory and may be associated with the pattern of contributions in a given area, which starts with few contributions, shows growth and tends to stabilize over time.

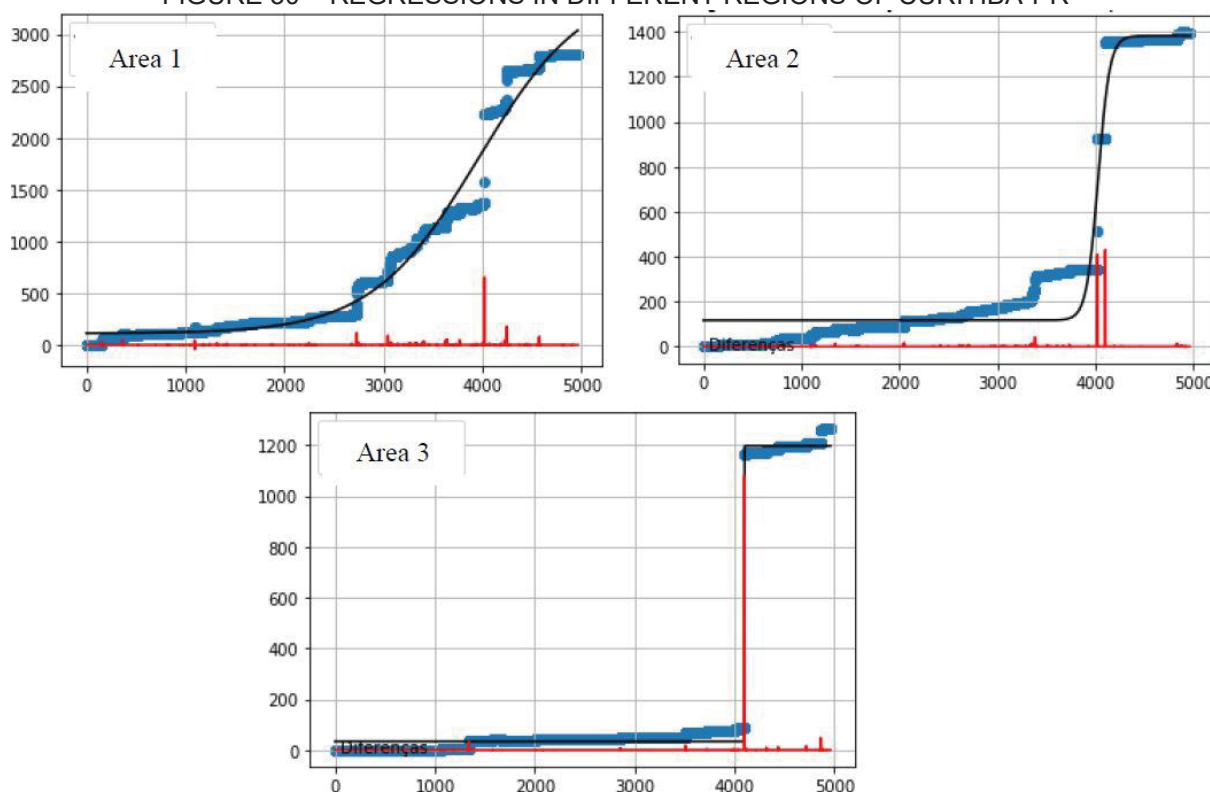
As there is a finite number of features in a region, we proposed to study the evolution of the contributions in a homogeneous cell, looking for patterns that vary according to the location of this cell, explicitly working with data in large urban centres in Brazil. The hypothesis is that if we know the existing patterns, we will be able in the future to determine at what stage a given cell is and, therefore, how close it is to have its complete mapping. This information is beneficial for using data that are already more robust and encouraging mapping in regions that are still poorly mapped, working to reduce heterogeneities, especially in the poorest and most peripheral areas.

Given the questions presented, it is questioned whether the answers obtained regarding the spatiotemporal patterns can be related to the quality of the data in the VGI and whether the correlations regarding the regression models can be measured in the same study area and their variations can be attributed to spatial patterns. Furthermore, it starts from the hypothesis that obtaining quality in VGI platforms in developing countries, such as Brazil, can be a guide to finding parameters for their integration concerning topographic mapping. This work aimed to obtain the spatiotemporal patterns of the history of editions of the OSM platform in different parts of the municipality of Curitiba-Parana-Brazil. The adequacy of the logistic model was verified for the accumulated amount of punctual, linear and polygonal features mapped in a given area and the measurement of the smallest possible size of the surrounding rectangle. Furthermore, the feasibility of identifying the saturation of features in an area from the particularities and outliers of the data was identified.

The methodological procedures included obtaining historical data from the OSM from the OHSOME Application Programming Interface (API) (<https://heigit.org/bigspatial-data-analytics-en/ohsome/>) and the adaptation of scripts in Python language provided by researchers connected to OHSOME to obtain logistic regressions and their parameters. A Python interface was implemented to obtain the results. The input data are the geographic coordinates of the place to be evaluated, the surrounding rectangle's minimum size, and the intended time interval for the analyses. The output data are the tables with the number of accumulated features and the successive differences, the residuals for the logistic regression and the graphs containing the accumulated contributions and the performed regression.

The test was carried out in the central region of Curitiba-PR, between 01-01-2008 and 01-08-2021. It was assumed that since it is a region with a more significant amount of contributions, the interaction of users would be more significant, representing the intended logistical model. Furthermore, from the regression analysis in the region, it was noted that a 1x1 km enveloping rectangle would be sufficient to verify the behaviour of the growth of features in an area. To illustrate the aspects found for the evaluated municipality, Figure 30 presents the regression graphs for three different areas of Curitiba-PR. The first (area 1) is characterized by the central region and the other two (area 2 and area 3) by regions gradually distant from the first region.

FIGURE 30 – REGRESSIONS IN DIFFERENT REGIONS OF CURITIBA-PR



SOURCE: The Authors (2021).

As shown in Figure 30, in the blue and black colours are the accumulated growth of features in the OSM and the logistic regression and the number of daily contributions in red. The X axis represents the period, represented in days and in Y the number of features inserted or removed. A critical aspect observed is a more significant iteration of contributions in the central region, exhibiting gradual growth. In the other regions, in addition to less contributions, the interaction was reduced to the point that many contributions described growth in a single period in almost its entirety. Such aspects may be related to imports of data or unusual activities dated in a single period that interfere with the behaviour of the data. This aspect is evidenced in the graph of region 3, in which the contributions were concentrated on a single date.

The approach allowed preliminary estimates to verify the growth of features inserted in the OSM in the evaluated region. The study was applied from the accumulated amount of punctual, linear and polygonal features over time, and the behaviour of logistic regressions showed that the dynamics of contributions influence the quality of the model. In addition, it was noted that it is possible to estimate the saturation of the number of features inserted in the OSM by comparing the levels of

completeness in subareas of the same region. This study is the first part of a larger project, including the specialization of residues and responses of regression models and their comparison with responses obtained concerning data quality evaluation procedures to obtain a tool for visualizing the framing of cells according to the spatiotemporal pattern of contributions.

5.3.1 References

Camboim, S. P., Bravo, J. V. M., Sluter, C. R. (2015). An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. *ISPRS International Journal of Geo-Information*, 4(3), 1366-1388.

Brovelli, M. A., Boccardo, P., Bordogna, G., Pepe, A., Crespi, M., Munafò, M., & Pirotti, F. (2019). Urban Geo Big Data. In *2019 Free and Open Source Software for Geospatial, FOSS4G 2019* (Vol. 42, No. 4, pp. 23-30). International Society for Photogrammetry and Remote Sensing.

Brovelli, M. A., Zamboni, G. (2018). A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of Geo-Information*, 7(8), 289.

Brückner, J., Schott, M., Zipf, A., Lautenbach, S. (2021). Assessing shop completeness in OpenStreetMap for two federal states in Germany, *AGILE GIScience Ser.*, 2, 20.

Grinberger, A. Y., Schott, M., Raifer, M., Zipf, A. (2021). An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS*, 25(2), 622-641.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4), 682-703.

ISO 19.157. (2013). Geographic Information - Data Quality. International Organization for Standardization.

Nasiri, A., Ali Abbaspour, R., Chehreghan, A., Jokar Arsanjani, J. (2018). 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, 7(7), 253.

Sehra, S. S., Singh, J., Rai, H. S. (2017). Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. Future Internet, 9(2), 15.

Yang, C., Yu, M., Hu, F., Jiang, Y., Li, Y. (2017). Utilizing cloud computing to address big geospatial data challenges. Computers, environment and urban systems, 61, 120-128.

Zhang, H., Malczewski, J. (2017). Accuracy evaluation of the Canadian OpenStreetMap road networks. International Journal of Geospatial and Environmental Research, 5(2).

5.4 O POTENCIAL DOS PADRÕES ESPAÇO-TEMPORAIS PARA A OBTENÇÃO DA COMPLETUDE NO MAPEAMENTO COLABORATIVO

Autores: Elias Nasr Naim Elias, Marcio Augusto Reolon Schmidt, Silvana Philippi Camboim.

Evento: XII Colóquio Brasileiro de Ciências Geodésicas e V Simpósio Brasileiro de Geomática, Curitiba 08-11 de Novembro de 2022.

Resumo:

A atualização cartográfica é um aspecto importante a ser considerado na produção de dados geoespaciais, visto que a dinâmica de alterações que ocorrem no espaço geográfico requer que existam metodologias que acompanhem essas mudanças. No Brasil tais questões ainda são mais relevantes, visto que falta de investimentos para a realização do Mapeamento Topográfico corroboram em regiões

desatualizadas ou que sequer tenham sido mapeadas [1]. Atrélicas ao Mapeamento Topográfico, produzidos por Agências Nacionais, e em face aos avanços tecnológicos, atualmente os dados geoespaciais podem ser obtidos por diferentes fontes e que podem se complementar. Nesse contexto, podem ser citadas as Informações Geográficas Voluntárias, do inglês *Volunteered Geographic Information* (VGI) [2]. No VGI, os dados geoespaciais passam a ser produzidos também, por indivíduos que tenham acesso à *internet* e receptores GNSS (*Global Navigation Satellite System*) nos *smartphones*, o que permite com que estes publiquem informações geográficas sobre determinado local e até mesmo, validem as informações postadas por outros usuários em plataformas online. Diante dessas questões, uma série de pesquisas têm buscado compreender a qualidade de feições oriundas de VGI, inclusive para estimar o seu potencial de integração com dados oficiais. A mensuração da qualidade no VGI é importante, visto que os dados contribuídos são heterogêneos e podem ser afetados por diferentes fatores, relacionados com a área, perfil dos contribuintes, quantidade e dinâmica de contribuições. Os procedimentos de avaliação da qualidade no VGI são divididos em duas abordagens. As avaliações extrínsecas partem da comparação com dados externos, usualmente de bases de dados de referência [3, 4, 5]. Já as avaliações intrínsecas [6] são caracterizadas por análises de fatores internos aos dados colaborativos, como histórico de edições, número de colaboradores e contribuições. Há ainda aplicações que se caracterizam pela combinação de parâmetros extrínsecos e intrínsecos [7]. Nas análises que contemplam os parâmetros intrínsecos, abordagens mais recentes têm buscado relacionar o histórico de edições acumulado a modelos matemáticos para estimar padrões a partir da colaboração ao longo do tempo [8, 9]. A modelagem matemática apresentada nas pesquisas de [8] e [9] foi dada partir do ajuste dos dados a uma Regressão Logística, que é descrita por meio de uma trajetória em forma de "S" (função sigmoide). O comportamento da função de logística, quando associado as edições em plataformas de VGI, pode ser descrito em três estágios principais: Início, Crescimento e Saturação. Baseado nesses estágios, a inserção de feições e elementos em uma determinada área se inicia com poucas contribuições, apresenta crescimento e tende a se estabilizar ao longo do tempo. Em virtude das questões apresentadas, parte-se da hipótese de que se for possível modelar os padrões espaço-temporais de contribuições no VGI é possível relacionar as respostas

obtidas com a sua qualidade, em áreas urbanas e heterogêneas do Brasil. Desta forma, essa pesquisa propõe o desenvolvimento de uma metodologia para estimar a completude de feições oriundas da plataforma de VGI do OpenStreetMap (OSM), por meio dos parâmetros obtidos na modelagem matemática das contribuições acumuladas ao longo do tempo. Os procedimentos metodológicos são descritos a partir das seguintes etapas: Subdivisão da região de estudo em uma grade com quadriculas de 1x1 Km; Extração e plotagem da quantidade de contribuições mensais acumulada em um dado intervalo temporal, de cada quadricula; Ajuste dos dados a partir da Regressão Logística de quatro parâmetros; Obtenção dos parâmetros da curva de Logística e análise da sua adequação por meio do cálculo do seu Erro Médio Quadrático, normalizado pela mediana das contribuições; Obtenção da inclinação da curva nos três últimos anos de contribuições para calcular a sua saturação; Espacialização dos resultados encontrados; e Validação das análises. Um aspecto importante a ser mencionado é que a subdivisão da região de estudo permite avaliar a heterogeneidade dos dados de forma mais concisa e viabiliza estimar correlações espaciais, a partir de diferentes variáveis. Para a validação dos procedimentos propostos foi selecionada a região do município de Curitiba-Paraná, Brasil. As categoriais de feições utilizadas corresponderam aos comprimentos dos eixos viários e as áreas das edificações, denominadas, respectivamente, pelas *tags* de *Highway* e *Building* no OSM. Além disso, foram utilizados os eixos viários do Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC) (atualizados em maio de 2022) e as edificações disponibilizadas pela *Microsoft* [10], extraídas a partir de imagens. Esses dados foram utilizados para estabelecer uma estimativa de área construída por quadricula, que indica o alcance máximo da curva. Para a obtenção do histórico de contribuições foi utilizada a *Application Programming Interface* (API) do *OHSOME* [11] e foram extraídas no período que corresponde a novembro/2007 e julho/2022. Por fim, os parâmetros da Regressão e os resultados, para cada quadricula, foram obtidos a partir de scripts na linguagem de programação *Python* e dispostos no *software QGIS 3.24* para a espacialização e realização das análises. Dentre as principais constatações obtidas, destaca-se a diferença no comportamento das contribuições ao longo do tempo das duas categorias avaliadas. Enquanto as vias (*highways*), em sua maior parte, foram inseridas a partir de contribuições gradativas, as edificações (*buildings*) apresentaram contribuições pontuais em uma única data. Tal constatação foi

observada a partir de dois aspectos: O parâmetro de inclinação da curva Logística “d” e a presença de *overflows* no ajuste da Regressão. Valores muito pequenos de “d” indicam uma curva mais acentuada (perpendicular ao eixo Y). Dessa forma, nas vias (*highways*) enquanto cerca de 16,9% das quadriculas o valor de “d” foi menor que 1, esse percentual superou 45% nas edificações (*buildings*). Além disso, saltos abruptos nas contribuições, normalmente foram seguidos de uma mensagem de overflow ao proceder com a regressão. Visto que tal mensagem foi mais comum nas edificações, evidencia-se as questões citadas. No que diz respeito às análises de saturação das vias, em mais de 99% das quadriculas o comprimento no OSM superou os do IPPUC. Tal resultado coincide com o percentual de saturação obtido, visto que em mais de 88% das quadriculas esse valor foi maior que 80%. Por sua vez, nas edificações, apesar de mais de 78% das quadriculas apresentarem percentual de saturação maior que 80%, a área total das edificações, em todas as quadriculas, a quantidade de feições foi menor que as da *Microsoft*. Ao analisar o comportamento das regressões, notou-se que os resultados obtidos estão diretamente relacionados com a dinâmica de contribuições ao longo do tempo e com regiões que não possuem feições. Nas vertentes apresentadas, conclui-se que os aspectos espaço-temporais se mostram promissores como novos parâmetros intrínsecos para a estimativa da completude, desde que as análises considerem a heterogeneidade dos dados e das diferentes categorias de feições. Recomenda-se a continuidade dos experimentos, a partir de novos modelos estatísticos que englobem a correlação espacial dos dados, bem como o refinamento dos parâmetros do modelo matemático obtido.

Palavras-chaves: *OpenStreetMap*; Atualização; VGI; Qualidade de dados; Modelagem Matemática; Regressão Logística.

5.4.1 Referências

1. SILVA, L. S. L; CAMBOIM, S. P. Brazilian NSDI ten years later: current overview, new challenges and propositions for national topographic mapping. 2020. Boletim de Ciências Geodésicas [online]. 2020, v. 26, n. 04, e2020018. Available from: <<https://doi.org/10.1590/s1982-21702020000400018>>.

2. GOODCHILD, M. F. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), pp. 211-221, 2007.
3. BROVELLI, M. A.; ZAMBONI, G. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of GeoInformation*, 7(8), pp. 1-25, 2018
4. ZHANG, H.; MALCZEWSKI, J. Accuracy Evaluation of the Canadian OpenStreetMap Road Networks. *International Journal Geospatial and Environmental Research*, 5, pp. 1-14, 2018.
5. HAKLAY, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4), 682-703, 2010.
6. SEHRA, S. S.; SINGH, J.; RAI, H. S. Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. *Future Internet*, 9(2), 15, 2017.
7. NASIRI, A.; ALI ABBASPOUR, R.; CHEHREGHAN, A. JOKAR ARSANJANI, J. Improving the quality of citizen contributed geodata through their historical contributions: The case of the road network in OpenStreetMap. *ISPRS International Journal of GeoInformation*, 7(7), 253, 2018.
8. Brückner, J., Schott, M., Zipf, A., Lautenbach, S. Assessing shop completeness in OpenStreetMap for two federal states in Germany, *AGILE GIScience Ser.*, 2, 20, 2021.
9. Grinberger, A. Y., Schott, M., Raifer, M., Zipf, A. An analysis of the spatial and temporal distribution of large- scale data production events in OpenStreetMap. *Transactions in GIS*. 25(2), 622-641, 2021.

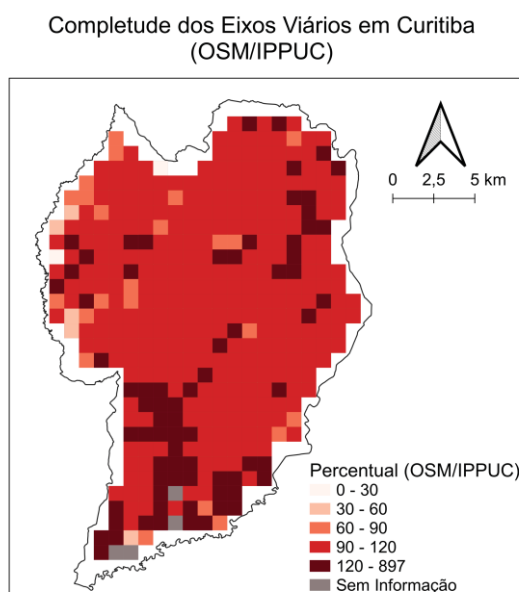
10. Microsoft - GlobalML Building Footprints. Disponível em: <<https://github.com/microsoft/GlobalMLBuildingFootprints>>. Acessado em: 08 de agosto de 2022.

11. OHSOME API. Disponível em: <<https://docs.ohsome.org/ohsome-api/v1/>>. Acessado em: 08 de agosto de 2022.

5.5 CONSIDERAÇÕES QUANTO A COMPLETUDE DOS DADOS DO OSM

A partir dos experimentos abordados na subseção 5.4, foi realizado um novo processamento dos dados considerando a filtragem dos eixos viários do OSM, de tal forma que se tornassem equivalentes aos dados do IPPUC. Dessa forma, foram desconsideradas *tags* que não estão representadas nas feições de referência, podendo ser exemplificadas pela tag “*highway=footway*” e “*highway=service*”. Além disso, aumentou-se a variação temporal para obtenção das regressões e o período considerado foi de Novembro de 2007 até Outubro de 2022. É importante comentar que a subdivisão das células ocorreu em quadriculas de 1x1 km (considerando aproximadamente, a equivalência ao longo da linha do Equador). O mapa da Figura 31 apresenta a completude obtida para os eixos viários.

FIGURA 31 – COMPLETUDE DOS EIXOS VIÁRIOS (HIGHWAYS) DO OSM EM CURITIBA

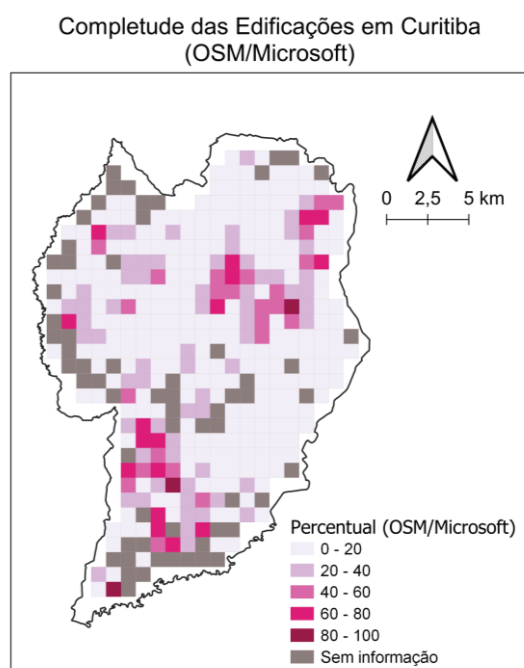


FONTE: O autor (2022).

Conforme apresentado no mapa da Figura 31, a realização das análises ocorreu a partir da extração das células que estavam contidas no município de Curitiba e foram desconsideradas as quais não ocorreu nenhum tipo de adição ou exclusão de dados no OSM ao longo do tempo na categoria avaliada e foram analisadas 409 células. Notou-se que o percentual predominante de completude estava entre 90% e 120% ao longo da região de estudo, presente em 74,57% do total de células. Além disso, em mais de 97% das células a estabilidade da Regressão Logística nos últimos três anos superou 80%, o que indica que as vias estão próximas de um estágio de saturação. Vale-se considerar que o cálculo da saturação foi dado a partir do quociente de um determinado mês no ano de 2019 em relação ao mês mais recente.

No que diz respeito a avaliação da completude das edificações do OSM em relação aos dados da Microsoft, a Figura 32 apresenta o mapa com os resultados obtidos em Curitiba.

FIGURA 32 – COMPLETUE DAS EDIFICAÇÕES (BUILDINGS) DO OSM EM CURITIBA



FONTE: O autor (2022).

Os resultados obtidos no mapa da Figura 32 também foram oriundos do novo processamento dos dados, considerando a variação temporal de Novembro de 2007 até Outubro de 2022. As verificações ocorreram nas células em que ocorreram adição ou exclusão de feições ao longo tempo para a categoria avaliada. É válido

ressaltar que tal filtragem das células (nas vias e nas edificações do OSM) ocorreu para que pudessem ser avaliadas regiões onde existissem dados para a realização da regressão logística. Foram analisadas 350 células das quais o percentual de completude predominante variou entre 0% e 20%, presente em 68,86% das células. No que diz respeito a saturação das contribuições, 74,57% das células apresentaram percentual maior que 80%. Tal aspecto reforça as questões descritas na subseção 5.4, visto que a saturação dos dados convergiu com a completude das vias e estava diretamente relacionada com os padrões de contribuições das edificações.

Nas respostas obtidas quanto ao parâmetro “d”, relacionado com a inclinação da curva, notou-se que nas edificações 52,86% das células apresentaram valores menores do que 1, sendo que a sua média e a mediana foram, respectivamente, 12,127 e 0,53. Por sua vez, nas vias, 27,14% das células apresentaram valores menores do que 1 e a média e mediana obtidas foram, respectivamente, 15,3 e 4,547. Os resultados encontrados reforçam as questões apresentadas quanto a predominância de saltos abruptos ou contribuições acumuladas nas edificações e o crescimento gradativo ao longo do tempo nas vias.

No cálculo do EMQ para avaliar a qualidade das regressões, inicialmente tentou-se normalizar os valores obtidos pela mediana das contribuições (conforme proposto no capítulo 4 e na subseção 5.4), contudo algumas células apresentaram o valor de mediana igual a zero. O referido valor foi predominante nas edificações, apresentada em 52,29% das células. Tal característica ocorreu devido ao fato das contribuições se iniciarem após o mês central do conjunto de dados em determinadas regiões. No intuito de suprir tal questão o EMQ foi normalizado pelo valor mais recente disponível. Nas edificações os valores do EMQ foram normalizados em relação a área mais recente do OSM e nas vias pelos comprimentos disponíveis em cada célula. Um aspecto importante a ser comentado é que devido a comportamentos atípicos nas contribuições das edificações, principalmente relacionados com saltos abruptos, 50 células obtiveram valores de EMQ que tendiam a zero, por sua vez, nas vias tal aspecto só foi apresentado em uma única célula. Ao normalizar os valores obtidos para cada célula das edificações, a média e mediana obtidos foram, respectivamente 9,52% e 6,02%. Nas vias, tais valores foram, respectivamente, 6,14% e 5,6%.

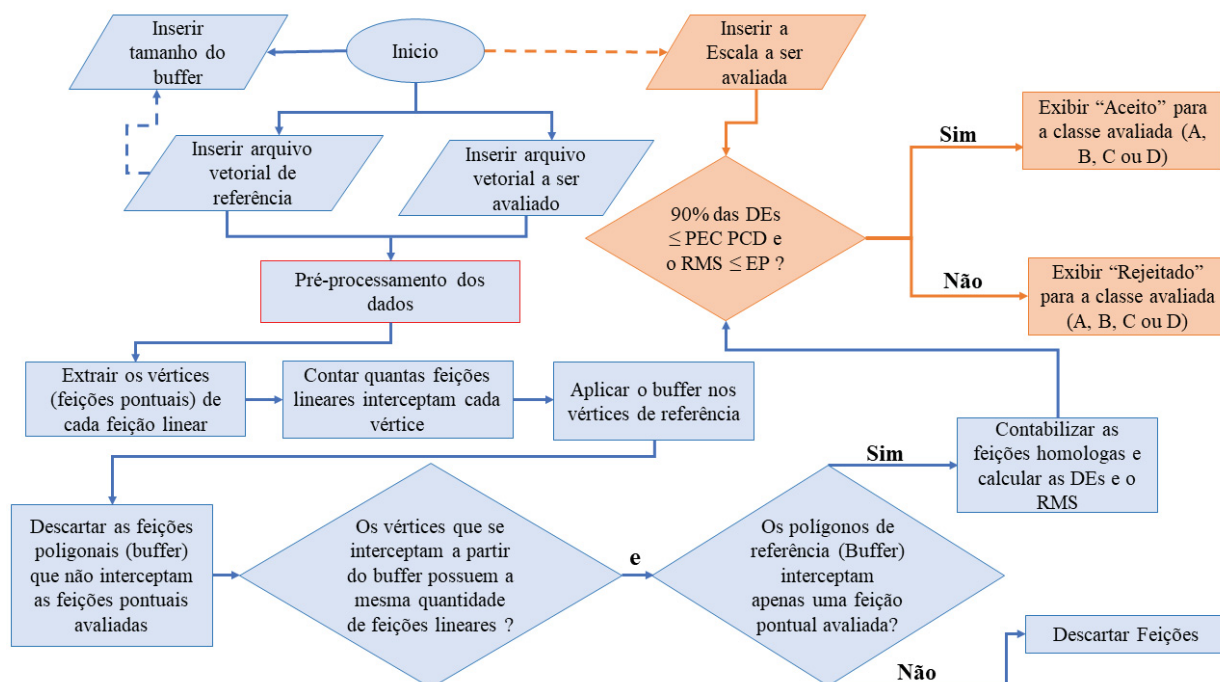
5.6 CONSIDERAÇÕES QUANTO AO DESENVOLVIMENTO DE COMPLEMENTO QGIS 3.X PARA A AVALIAÇÃO DA ACURÁCIA POSICIONAL

Visto as questões descritas no capítulo 3 para o desenvolvimento do complemento QPEC e com as novas aplicações oriundas deste, explanadas nas subseções 5.1 e 5.2, esta subseção apresenta uma nova versão da referida ferramenta, com a qual incrementou-se uma etapa preliminar de identificação de pontos homólogos ao processo de avaliação da acurácia posicional. O desenvolvimento do complemento considerou feições lineares que correspondem a eixos viários e a sua estrutura foi construída de forma semelhante ao QPEC, utilizando o modelador gráfico do QGIS 3.X. Além disso, a aplicação foi baseada em procedimentos estabelecidos por Paiva e Camboim (2022), na qual são determinadas etapas para identificar feições homologas entre dados de referência e avaliados.

A partir dos aspectos apresentados a execução do complemento é dada por meio da extração de feições pontuais que correspondem aos cruzamentos de eixos viários, tanto nas feições lineares de referência quanto nas avaliadas, e contabiliza-se quantas feições lineares interceptam tais cruzamentos. Em seguida é gerada uma área de abrangência (*buffer*) nas feições pontuais de referência e verifica-se se os cruzamentos possuem a mesma quantidade de interseções em relação as feições avaliadas que estão contidas no *buffer*. Além disso, avalia-se se as feições possuem um único candidato homólogo. Caso tais condições sejam atendidas, são extraídas as Distâncias Euclidianas (DE), procedendo com o cálculo do Erro Médio Quadrático (EMQ) e a aplicação do PEC PCD para uma dada escala.

A Figura 33 apresenta um fluxograma de execução do complemento desenvolvido.

FIGURA 33 – FLUXOGRAMA DE EXECUÇÃO DO NOVO COMPLEMENTO DESENVOLVIDO



FONTE: O autor (2022).

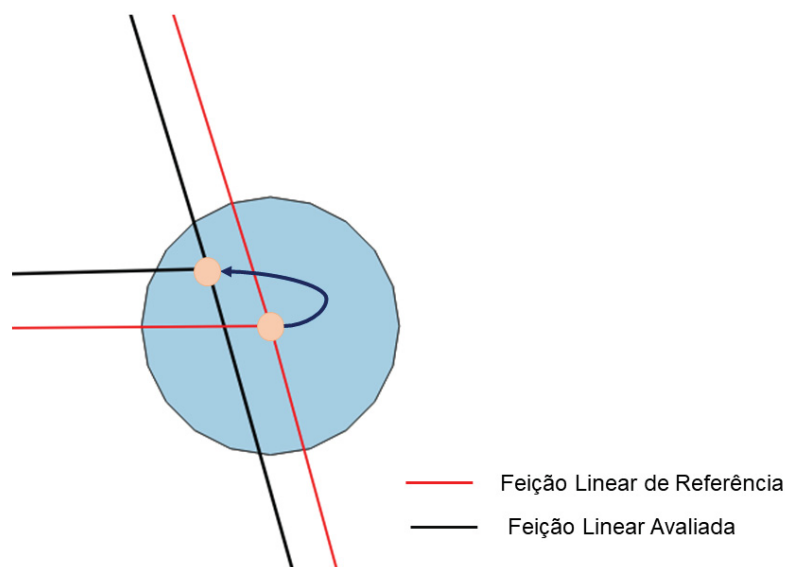
Conforme apresentado na Figura 33, além da inserção dos arquivos vetoriais com as feições lineares para a realização das etapas, o usuário também deve inserir o tamanho do buffer para detecção dos pontos homologos e a escala desejada para a realização do cálculo do PEC PCD. Antes da extração dos vértices que correspondem aos cruzamentos, é realizado um pré-processamento no intuito de viabilizar com que os conjuntos de dados sejam equivalentes e proceder com as análises. Tal etapa é necessária visto que as feições podem não estar estruturadas topologicamente da mesma maneira, o que pode ocasionar a detecção de vértices com quantidade de cruzamentos divergentes e fornecer resultados errôneos.

Diante das questões apresentadas quanto ao pré-processamento dos dados e visto que o enfoque desta tese está relacionado com a determinação da qualidade de dados geospaciais no VGI, tais aspectos tornam ainda mais oportuno o estabelecimento de métricas para tornar as feições oriundas de diferentes fontes e estruturas comparáveis. Dessa forma, nesta etapa foram utilizados diferentes algoritmos já existentes no QGIS, sendo estes: Dissolver, Linhas com Quebra e Excluir Geometrias Duplicadas. A captura e equivalência da quantidade de cruzamentos em cada feição pontual deu-se pelos algoritmos de Associar Atributos

por Localização, Unir Atributos pela Localização (Sumário) e Selecionar por uma Expressão.

A Figura 34 apresenta um exemplo de cruzamento na qual os homólogos foram identificados a partir da aplicação do buffer na feição pontual de referência.

FIGURA 34 – EXEMPLO DE IDENTIFICAÇÃO DE FEIÇÃO HOMOLOGA



FONTE: O autor (2022).

Conforme apresentado na Figura 34, a partir do *buffer* realizado no cruzamento das feições lineares de referência foi detectado um cruzamento homologo nas feições avaliadas. Além disso, identificou-se que ambos eram interceptados pela mesma quantidade de linhas (3 feições lineares), de tal forma que foram atendidas as condicionantes estabelecidas previamente e a DE entre as feições pontuais foi computada.

Por fim, as condicionantes para a aplicação do PEC PCD são dadas de forma semelhante ao abordado ao longo de todo o capítulo 3, nas quais os scripts foram desenvolvidos no QGIS utilizando a estrutura SQL-LIKE. Dessa forma, a partir da inserção do denominador de escala desejado, aplicam-se os testes estabelecidos na Tabela 2 para as Classes A, B, C e D e exibe-se para cada uma delas a mensagem de “Aceito” ou “Rejeitado” de acordo com a resposta obtida. Os resultados das DE, EMQ e as classificações finais são apresentados na tabela de atributos do QGIS.

A Figura 35 apresenta as condicionantes estabelecidas para a Classe A do PEC PCD na estrutura descrita.

FIGURA 35 – APLICAÇÃO DE CONDICIONANTE PARA A CLASSE A DO PEC PCD

```
CASE WHEN sum("Distancia"<=0.28*"va"/1000) >= 0.9*count("ID_REF_2")
AND sqrt(sum("Distancia"^2)/count("ID_REF_2"))<=0.17*"va"/1000
THEN 'Aceito'
ELSE 'Rejeitado'
END
```

FONTE: O autor (2022).

A Figura 36 apresenta a interface para entrada os dados descritos na Figura 33.

FIGURA 36 – INTERFACE DO COMPLEMENTO DESENVOLVIDO

FONTE: O autor (2022).

Para a validação da ferramenta desenvolvida, realizou-se um experimento na região que compreende o município de Curitiba-Paraná, descrita no capítulo 4.

Foram utilizados como dados de referência os eixos viários do IPPUC e os dados avaliados corresponderam as feições do OSM, classificados com a *key* “*highway*”. É importante considerar que foi realizada uma filtragem das *tags*, no intuito de tornar os dados compatíveis entre si e foram consideradas as vias principais do OSM. Por exemplo, foram excluídas as *tags* “*highway=service*” e “*highway=footway*”, por compreender que não existiriam homólogos nas feições do IPPUC. Tal operação viabilizou com que pontos homólogos fossem identificados de forma errônea.

Para realização dos experimentos utilizou-se um *buffer* de 6m e o PEC PCD foi aplicado para uma escala de 1:10.000 e foram identificados 15.024 pontos homólogos. O tamanho do *buffer* foi baseado nas premissas de Haklay (2010). A Figura 37 apresenta a tabela de atributos com os resultados obtidos.

FIGURA 37 – RESULTADOS OBTIDOS NA TABELA DE ATRIBUTOS DO QGIS

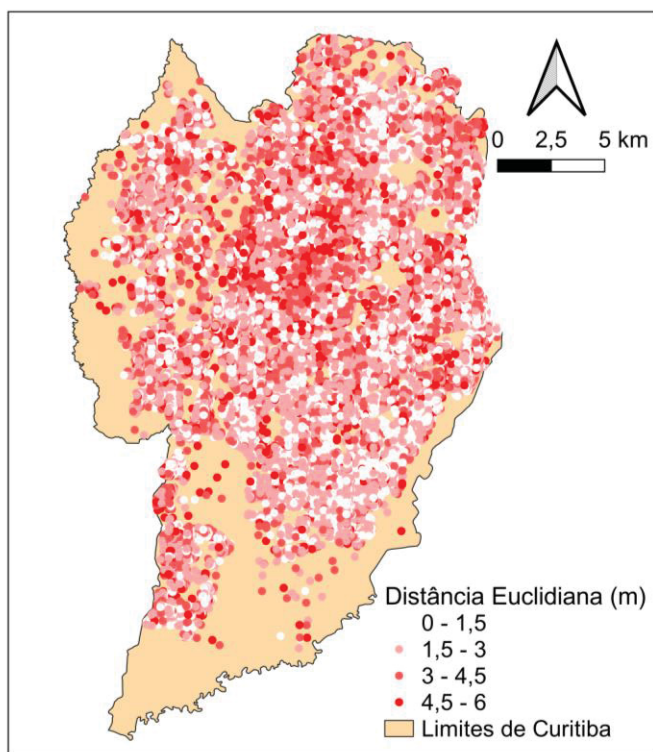
ID_REF_2	ID_REF_2_count	ID_AVAL_2	ID_AVAL_2_count	Distancia	RMS	va	Classe_A	Classe_B	Classe_C	Classe_D
3005	6	5072	6	1,013	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
5915	6	14811	6	4,220	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
5979	6	14288	6	0,617	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
6104	6	12463	6	1,798	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
9526	6	12149	6	1,498	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
11230	6	21415	6	1,164	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
11890	6	20246	6	0,722	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
14621	6	28188	6	0,694	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
16691	6	14164	6	2,809	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito
18022	6	19817	6	3,433	2,738	10000,00...	Rejeitado	Aceito	Aceito	Aceito

FONTE: O autor (2022).

A partir dos resultados apresentados na Figura 37, notou-se que a escala avaliada para o conjunto de feições foi aceita a partir da Classe B do PEC PCD. Dessa forma, os valores médio e mediano das DE foram, respectivamente, de 2,411m e 2,250m. Além disso, o Desvio Padrão obtido correspondeu a 1,297m. O mapa da Figura 38 apresenta a distribuição e magnitude das discrepâncias obtidas ao longo da região de estudo.

FIGURA 38 – MAPA DA ACURÁCIA POSICIONAL DOS EIXOS VIÁRIOS A PARTIR DA APLICAÇÃO DO COMPLEMENTO

Acurácia Posicional de Eixos Viários em Curitiba



FONTE: O autor (2022).

Visto a aplicação do complemento desenvolvido nas feições do OSM, bem como a análise dos resultados e espacialização das discrepâncias obtidas, é possível notar que estes auxiliam na identificação de padrões quanto a qualidade dos dados e verificações da sua heterogeneidade. Tal aspecto pode ser evidenciado ao observar o comportamento das DE ao longo de Curitiba, no mapa da Figura 35, onde ocorreram variações de até 5m. A ferramenta desenvolvida está disponível no seguinte repositório *online*: https://github.com/eliasnaim/QPEC_vias_automatico.

6. CONCLUSÕES E RECOMENDAÇÕES

Neste capítulo são apresentados os resultados referentes a cada uma das etapas descritas na metodologia, bem como, as discussões a respeito desses resultados. No quadro 3 constam os principais produtos e conclusões obtidas em cada capítulo.

QUADRO 3 – PRINCIPAIS PRODUTOS E CONCLUSÕES OBTIDAS EM CADA CAPÍTULO

Título	Síntese dos Resultados	Contribuições	Limitações	Proposta de Estudos Futuros
QPEC: QGIS toolkit for evaluating geospatial data positional accuracy according to the brazilian cartographic accuracy standard	Plugin para QGIS que permite avaliar e visualizar a acurácia posicional de feições pontuais e lineares, de forma automática, com base no PEC PCD. Os resultados são apresentados na tabela de atributos e espacialmente.	Concatenar análises estatísticas e visualização dos erros de forma simultânea. Em fontes de dados heterogêneos (como VGI) é possível avaliar o comportamento das discrepâncias em diferentes áreas.	O usuário necessita selecionar as feições homologas de forma manual. A classificação final obtida ainda se restringe a normatização brasileira.	Aplicar funções que permitam a identificação automática de feições homologas. Estudar novas variáveis para a visualização os resultados.
Exploring spatio-temporal patterns of OpenStreetMap (OSM) contributions in heterogeneous urban areas	Modelagem da evolução da inserção de feições no OSM ao longo do tempo. Análise da interação de dados oficiais e colaborativos. Avaliação da heterogeneidade dos dados a partir dos percentuais de similaridade.	Metodologia para obter padrões de contribuições no OSM a partir da Regressão Logística. Proposta de novo indicador de qualidade intrínseca no mapeamento colaborativo.	A validação foi realizada apenas Núcleo Urbano Central de Curitiba em células de 1x1 km.	Avaliar novos modelos matemáticos para obtenção dos padrões de contribuição. Ampliar os estudos para outras regiões.
Avaliação e representação da qualidade de dados geospaciais a partir do desenvolvimento de ferramentas em ambiente SIG *	Ampliação do QPEC e desenvolvimento de plugin para a avaliação e visualização automática da completude geométrica e do atributo de eixos viários do OSM.	Visualização da heterogeneidade da qualidade dos dados no OSM. Auxilia na tomada de decisões quanto ao seu comportamento em uma determinada região	A aplicação é baseada na diferença entre o comprimento de vias de referência e do OSM. Não são aplicadas funções para a identificação automática de feições homologas.	Propor funções para identificação automática de feições homologas. Desenvolver ferramentas para os demais indicadores de qualidade.

Data quality analysis tools and their use for the detection of spatio-temporal heterogeneities in VGI contributions *	Análise da evolução das contribuições ao longo do tempo em células de 5x5 km nas capitais brasileiras. Visualização da heterogeneidade da acurácia posicional do OSM.	Visão geral do comportamento dos dados do OSM em diferentes regiões do Brasil. Descrição da heterogeneidade dos dados com base nos padrões de contribuição. Melhorias no plugin QPEC.	As células testadas foram apenas nas regiões que contemplam as capitais dos estados brasileiros.	Ampliar as regiões de estudo para realização das análises.
Discovering spatiotemporal patterns on data quality assessment in collaborative mapping: a preliminary study in an area of Brazil *	Avaliação das influências que afetam na qualidade dos dados em relação aos padrões de contribuição no OSM. Estimativas da interação dos usuários no OSM como base na região e na modelagem das contribuições.	Desenvolvimento de interface para a obtenção dos padrões de contribuições a partir da delimitação do local e Descrição das diferenças entre os padrões de contribuição.	A avaliação foi realizada apenas no município de Curitiba-PR, em células de 1x1 km.	Ampliar as verificações para demais regiões de estudo, bem como, utilizar diferentes tamanhos de células.
O potencial dos padrões espaço-temporais para a obtenção da completude no mapeamento colaborativo *	Percentual de completude dos eixos viários do OSM foi maior que as edificações. Enquanto que os eixos viários, em sua maior parte, foram inseridos a partir de contribuições gradativas, as edificações apresentaram contribuições pontuais em uma única data.	Estimativas do Potencial dos parâmetros da Regressão Logística e de saturação dos dados para a obtenção da completude do OSM.	As análises foram realizadas em células de 1x1 Km no município de Curitiba-PR.	Continuidade dos experimentos, a partir de novos modelos estatísticos que englobem a correlação espacial dos dados, bem como o refinamento dos parâmetros do modelo matemático obtido.

* Resumos Expandidos publicados em eventos.

FONTE: O autor (2022).

Conforme destacado ao longo deste trabalho os avanços tecnológicos e a *web 2.0* vem permitindo com que diferentes usuários, não necessariamente especialistas, sejam produtores de dados geoespaciais, com enfoque nas plataformas de mapeamento colaborativo. Uma questão discutida pela comunidade científica, neste contexto, trata da qualidade desses tipos de dados e o seu potencial de integração em mapeamentos topográficos. Em países em desenvolvimento, como o Brasil, tal questão ainda é mais relevante visto a falta de investimentos e recursos para produção cartográfica atualizada.

Com base nas ferramentas desenvolvidas, descritas no objetivo 1, permitiu-se concatenar análises estatísticas e a visualização dos erros de forma simultânea. Em fontes de dados heterogêneos (como VGI) é possível avaliar o comportamento das discrepâncias em diferentes áreas. Tal aspecto converge com as questões de heterogeneidade e as discussões dessa característica no OSM. Dessa forma, as considerações realizadas permitiram estabelecer diretrizes para a modelagem dos dados, associada ao objetivo 2.

No objetivo 2, os parâmetros extraídos na modelagem dos dados permitiram avaliar as influências que afetam na qualidade dos dados em relação aos padrões de contribuição no OSM. Os resultados obtidos no NUC permitiram evidenciar como os padrões convergem nas regiões centrais do município de Curitiba, em relação as áreas periféricas, permitindo relacionar mesclas de contribuições orgânicas e importação de dados ou grandes contribuições.

No que concerne à gestão pública, é fundamental que existam ferramentas e processos que permitam mensurar e armazenar a qualidade de dados geoespaciais de forma mais rápida, eficiente e, especialmente, que estejam disponíveis em *softwares* livres para manipulação. Tal aspecto viabiliza a tomada de decisões e permite estabelecer para quais aplicações um determinado tipo de dado se adequa.

Diante das questões apresentadas, esta pesquisa teve como intuito identificar e suprir lacunas na literatura associadas com a qualidade de dados geoespaciais no mapeamento colaborativo, com enfoque na mensuração da heterogeneidade dos dados. Tal aspecto é importante, visto que, a qualidade deste tipo de informação pode variar de acordo com uma série de fatores, associadas ao local, perfil das contribuições e dos contribuintes. Explorou-se então, uma nova medida de qualidade intrínseca relacionada com o acumulado de contribuições ao longo do tempo e os parâmetros obtidos a partir da sua modelagem matemática.

Essa modelagem permitiu obter e discretizar os diferentes fatores que influenciaram nos padrões de contribuições, caracterizadas por importação de dados, mapatonas ou até mesmo grandes contribuições.

Notou-se que a dinâmica de contribuições corrobora a prerrogativa de que aspectos espaço temporais no VGI podem estar diretamente relacionados com a qualidade dos dados, principalmente na sua completude e temporalidade. Além disso, foi possível avaliar como a sinergia entre dados de referência e colaborativos influenciam na obtenção de padrões temporais. Por meio das questões descritas, a sequência de procedimentos metodológicos dos capítulos 4 e 5 e a posterior espacialização dos dados evidencia o cunho inovador desta pesquisa, visto que traz novos parâmetros para avaliar a heterogeneidade no mapeamento colaborativo e estabelecer como estes se apresentam em regiões com diferentes características.

Visto os avanços na tecnologia da geoinformação, as ferramentas e algoritmos obtidos contribuem para o desenvolvimento de novos sistemas de avaliação e visualização da qualidade dos dados, seja de forma extrínseca ou intrínseca. Torna-se possível realizar análises específicas para visualizar fenômenos e padrões de uma determinada região, na qual tal aspecto foi comprovado ao comparar o comportamento das contribuições, dos padrões e da qualidade dos dados a partir da subdivisão das regiões de estudo utilizadas. Além disso, o fato de terem sido produzidas a partir de aplicações *open source* e a entrada dos dados se dar por meio de uma interface simples, torna viável a reprodutibilidade do método.

No que diz respeito a atualização cartográfica e a integração de dados de referência e colaborativos, no capítulo 4 constatou-se que para atingir a zona III do estágio de mapeamento, que se caracteriza a estabilidade e saturação do modelo logístico foi necessário que, em sua totalidade, ocorrem grandes inserções de feições em um período ou importação de dados oficiais. Tal característica foi modelada a partir do parâmetro “d”, que define a inclinação da curva e o quanto ela está acentuada. No capítulo 5 esta característica também é apresentada ao calcular a inclinação nos três últimos anos e observar a relação direta entre a estabilidade da curva com a completude dos dados, principalmente na categoria de eixos viários. Tais aspectos reforçam a contribuição apresentada nesta tese, na apresentação de parâmetros para mensurar a qualidade intrínseca no VGI.

Considerando o grande volume de dados gerados diariamente na *web* e os esforços da comunidade científica em automatizar processos e apresentar os

resultados de forma homogênea, a sequência metodológica descrita em cada um dos capítulos e a forma como as aplicações foram desenvolvidas, permitiu com que os objetivos pretendidos fossem alcançados a partir do agrupamento dos diferentes grandes conjuntos de dados e a sua posterior visualização. Por fim, este aspecto permite com que usuários consigam visualizar a confiabilidade dos dados e comportamento das contribuições, além da continuidade dos experimentos e avaliação por parte de gestores de órgãos públicos quanto à usabilidade dos dados.

6.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Diante dos resultados obtidos e dos procedimentos metodológicos apresentados, no que diz respeito ao capítulo 3, recomenda-se explorar as ferramentas desenvolvidas e propor novas formas de visualização das discrepâncias posicionais. Além disso, um aspecto importante diz respeito à identificação automática de feições homólogas para a avaliação da qualidade. Dessa forma, para trabalhos futuros é recomendado explorar os diferentes aspectos das categorias de feições, filtragem de dados para que feições que representam o mesmo elemento possa ser identificada e incrementada ao código. Por fim, recomenda-se o desenvolvimento de complementos para avaliação dos indicadores de Acurácia Temática e Consistência Lógica.

No que diz respeito aos capítulos 4 e 5, para trabalhos futuros 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, e outras variáveis. Além disso, recomenda-se explorar os parâmetros encontrados ao obter as regressões e realizar análises estatísticas espaciais no intuito de se classificar 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. É viável, também, a continuidade dos experimentos, a partir de novos modelos estatísticos que englobam a correlação espacial dos dados, bem como o refinamento dos parâmetros do modelo matemático obtido. Além disso, sugere-se a extrapolação das análises para outros municípios e localidades no intuito de validar a metodologia proposta.

7. REFERÊNCIAS

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.

ANTONIOU, V.; MORLEY, J.; HAKLAY, M. Web 2.0 geotagged photos: Assessing the spatial dimension of the phenomenon. **Geomatica**, 64(1), 99-110, 2010.

ARCHELA, R. S.; ARCHELA, E. Síntese cronológica da cartografia no Brasil. **Portal de Cartografia das Geociências**, 1(1), 93-110, 2008.

ARSANJANI, J. J.; MOONEY, P.; ZIPF, A.; SCHAUSS, A. Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets In: Jokar Arsanjani, J., Zipf, A., Mooney, P., Helbich, M., OpenStreetMap in GIScience: experiences, research, applications, **Springer Press**, pp. 37-58, 2015.

BARBETTA, P. A. Estatística aplicada às ciências sociais. 5. ed. rev. Florianópolis: **Ed. UFSC**, 2005.

BASIRI, A.; HAKLAY, M.; FOODY, G.; MOONEY, P. Crowdsourced geospatial data quality: Challenges and future directions, **International Journal of Geographical Information Science**, 33:8, 1588-1593, 2019.

BÉDARD, Y.; VALLIÈRE, D. Qualité des données à référence spatiale dans un contexte gouvernemental. **Université Laval**, Québec, 53, 1995.

BÉGIN, D.; DEVILLERS, R.; ROCHE, S. **Assessing Volunteered Geographic Information (VGI) quality based on contributors' mapping behaviours**. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 8th International Symposium on Spatial Data Quality. 30 May - 1 June 2013, Hong Kong, 2013.

BORTOLINI, E.; SILVA, L. S. L.; MACHADO, A. A.; PAIVA, C. P.; CAMBOIM, S. P. **Potenciais categorias de informações geográficas do mapeamento colaborativo para o Oficial**. In: X Colóquio Brasileiro de Ciências Geodésicas, Curitiba. Anais do X CBCG, 2018.

BRASIL. Decreto 243, de 28 de fevereiro de 1967. **Fixa as Diretrizes e Bases da Cartografia Brasileira**, 1967.

BRASIL. Decreto 89.817, de 20 de junho de 1984. **Estabelece as Instruções Reguladoras da Normas Técnicas da Cartografia Nacional**, 1984.

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. **Identificação e caracterização de tarefas de uso e geração de geoinformação no mapeamento colaborativo**. Tese (Doutorado em Ciências Geodésicas). Universidade Federal do Paraná, Curitiba, 2017.

BROVELLI, M. A.; BOCCARDO, P.; BORDOGNA, G.; PEPE, A.; CRESPI, M.; MUNAFÒ, M.; PIROTTI, F. **Urban Geo Big Data**. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, FOSS4G 2019 - Academic Track. 26-30 August 2019, Bucharest: Romania, 2019.

BROVELLI, M. A.; ZAMBONI, G. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. **ISPRS International Journal of GeoInformation**, 7(8), pp. 1-25, 2018.

BROWN, G.; KYTTÄ, M. Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research. **Applied geography**, 46, 122-136, 2014.

BUDHATHOKI, N. **Participants' Motivations to Contribute to Geographic Information in an Online Community**. Doctoral Thesis, University of Illinois, Urbana-Champaign, IL, USA, 2010.

BUNGE, W. Fitzgerald: Geography of a revolution, **Cambridge, MA: Schenkman**, 1971.

BYRNE, D.; PICKARD, A. J. Neogeography and the democratization of GIS: a metasynthesis of qualitative research. **Information, Communication & Society**, 19(11), 1505-1522, 2016.

CAMBOIM, S. P.; BRAVO, J. V. M.; and Sluter, C. R. An investigation into the completeness of, and updates to, the OpenStreetMap data in a heterogeneous area in Brazil. **ISPRS International Journal of GeoInformation**, 4(3), pp. 1366-1388, 2015.

CAPINERI, C. The nature of volunteered geographic information, In: European Handbook of Crowdsourced Geographic Information, ed. by CAPINERI, C.; HAKLAY, M.; HUANG, H.; ANTONIOU, V.; KETTUNEN, J.; OSTERMANN, F.; PURVES, R. pp. 15–33, **Ubiquity Press**, London, 2016.

CAROTI, G.; PIEMONTE, A.; NESPOLI, R. **UAV-Borne photogrammetry: a low cost 3D surveying methodology for cartographic update**. In: MATEC Web of Conferences, vol. 120, pp. 09005. EDP Sciences, 2017.

CHILTON, S. OS and OpenStreetMap. **Sheetlines**, vol. 91, pp. 20-27, 2011.

COOTE, A.; RACKHAM, L. **Neogeographic data quality—is it an issue**. In: Proceedings of the Annual Conference of the Association for Geographic Information (AGI), Stratford-upon-Avon, UK: AGI, 1–17, 2008.

CONSTANTINIDES, E.; FOUNTAIN, S. J. Web 2.0: Conceptual foundations and marketing issues. **Journal of direct, data and digital marketing practice**, 9(3), 231-244, 2008.

CORMODE, G.; KRISHNAMURTHY, B. Key Differences between web 1.0 and web 2.0. **First Monday**, 13(6), 2008.

CRAGLIA, M.; OSTERMANN, F.; SPINSANTI, L. Digital Earth from vision to practice: making sense of citizen-generated content. **International Journal of Digital Earth**, 5(5), 398-416, 2012.

CHRISMAN, N. Development in the Treatment of Spatial Data Quality. In: Devillers, R., Jeansoulin, R. (eds.) *Fundamentals of Spatial Data Quality*, pp. 21–30. **Iste**, London, 2006.

CROOKS, A.T.; HUDSON SMITH, A.; CROITORU, A.; STEFANIDIS, A. The Evolving GeoWeb, in ABRAHART R. J.; SEE, L. M. (eds.), *Geocomputation (Second Edition)*, **CRC Press**, Boca Raton, FL, pp. 69-96, 2014.

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

DE LONGUEVILLE, B.; OSTLÄNDER, N.; KESKITALO, C. **Addressing vagueness in Volunteered Geographic Information (VGI) – A case study**, GSDI-11 Conference Proceedings, 21p., 2009.

DEVILLERS, R.; STEIN, A.; BÉDARD, Y.; CHRISMAN, N.; FISHER, P.; SHI, W. Thirty years of research on spatial data quality: achievements, failures, and opportunities. **Transactions in GIS**, 14(4), 387-400, 2010.

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.

DROR T.; DOYTSHER Y.; DALYOT S. Investigating the Use of Historical Node Location Data as a Source to Improve OpenStreetMap Position Quality. In:

Mobasheri A. (eds) **Open Source Geospatial Science for Urban Studies**. 2020. Lecture Notes in Intelligent Transportation and Infrastructure. Springer, Cham.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **Manual Técnico T 34-700 Convenções Cartográficas (1ª Parte) Normas para o Emprego dos Símbolos**, 2ª Edição, 1998.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **Manual Técnico T 34-700 Convenções Cartográficas (2ª Parte) Catálogo de Símbolos**, 2ª Edição, 2000.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-CQDG – Especificação Técnica para Controle de Qualidade de Dados Geoespaciais**, 2015b.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-ADGV – Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais - 1ª parte**, 2015a.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-ADGV – Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais - 2ª parte**, 2016b.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-ADGV 3.0 – Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais**, 2017a.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-EDGV 2.1.3. – Especificação Técnica para a Estruturação dos Dados Geoespaciais Vetoriais**, 2010.

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

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-EDGV F-Ter. – Especificação Técnica para a Estruturação dos Dados Geoespaciais Vetoriais** - 2ª parte, 2016a.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-PCDG - Especificação Técnica para Produtos de Conjuntos de Dados Geoespaciais**, 1 ed, 2014.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-PCDG - Especificação Técnica para Produtos de Conjuntos de Dados Geoespaciais**, 2 ed, 2016.

DU, H.; ALECHINA, N.; JACKSON, M.; HART, G. A method for matching crowd-sourced and authoritative geospatial data. **Transactions in GIS**, 21(2), 406-427, 2017.

FERNANDES, V. O.; ELIAS, E. N.; ZIPF, A. Integration of authoritative and volunteered geographic information for updating urban mapping: challenges and potentials. **International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 43, 2020.

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.

ELWOOD, S.; LESZCZYNSKI, A. Privacy, reconsidered: New representations, data practices, and the geoweb. **Geoforum**, 42(1), 6-15, 2011.

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.

FAIRBAIRN, D.; AL-BAKRI, M. Using geometric properties to evaluate possible integration of authoritative and volunteered geographic information. **ISPRS International Journal of Geoinformation**, 2(2), 349-370, 2013.

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**, 72(3-4), 137-148, 2008.

FORGHANI, M.; DELAVAR, M. R. A quality study of the OpenStreetMap dataset for Tehran. **ISPRS International Journal of Geoinformation**, 3(2), 750-763, 2014.

GALO, M., CAMARGO, P. O. **Utilização do GNSS no controle de qualidade de cartas**. In: Congresso Brasileiro De Cadastro Técnico Multifinalitário, 1. – COBRAC, Anais... v. 2, Florianópolis, 1994

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

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

GIUFFRIDA, N.; LE PIRA, M.; INTURRI, G.; IGNACCOLO, M. Mapping with stakeholders: An overview of public participatory GIS and VGI in transport decision-making. **ISPRS International Journal of Geoinformation**, 8(4), 198, 2019.

GOODCHILD, M. F. Citizens as sensors: the world of volunteered geography. **GeoJournal**, 69(4), pp. 211-221, 2007.

GOODCHILD, M. F. Theoretical Models for Uncertain GIS. In: W. Shi, P.F. Fisher and M.F. Goodchild, eds. Spatial Data Quality. **London: Taylor & Francis**. Introduction to Part I & ch. 1, p.1-17, 2002.

GOODCHILD, M. F.; GLENNON, J. A. Crowdsourcing geographic information for disaster response: a research frontier. **International Journal of Digital Earth**, 3(3), 231-241, 2010.

GOODCHILD, M. F.; GOPAL, S. (Eds.). The accuracy of spatial databases. **CRC Press**, 1989.

GOODCHILD, M. F.; LI, L. Assuring the quality of volunteered geographic information. **Spatial Statistics**, 1, 110-120, 2012.

GOLDBLATT, R.; JONES, N.; MANNIX, J. Assessing OpenStreetMap Completeness for Management of Natural Disaster by Means of Remote Sensing: A Case Study of Three Small Island States (Haiti, Dominica and St. Lucia). **Remote Sensing**, 12(1), 118, 2020.

GUIMARAES, R. G.; IMAI, N. N.; DA SILVA, E. A.; ELIAS, A. R. **Sistema de atualização cartográfica apoiado por computador**. In: Simpósio Brasileiro de Sensoriamento Remoto. Santos (SP). Cdroom, 1998.

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

HAKLAY, M.; SINGLETON, A.; PARKER, C. Web mapping 2.0: The neogeography of the GeoWeb. **Geography Compass**, 2(6), 2011-2039, 2008.

HAKLAY, M.; WEBER, P. Openstreetmap: User-generated street maps. **IEEE Pervasive Computing**, 7(4), 12-18, 2008.

IBRAHIM, M. H.; DARWISH, N. R.; HEFNY, H. A. An Approach to Control the Positional Accuracy of Point Features in Volunteered Geographic Information Systems. **International Journal of Advanced Computer Science and Applications (IJACSA)**, 10(6), pp. 169-175, 2019.

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 2859 – 1. **Sampling procedures for inspection by attributes: part 1: sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection**. International Organization for Standardization, 1999.

ISO 2859 – 2. **Sampling procedures for inspection by attributes: part 2: sampling plans indexed by limiting quality (LQ) for isolated lot inspection**. International Organization for Standardization, 1985.

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.

IVANOVIC S. S.; OLTEANU-RAIMOND A. M.; MUSTIÈRE S.; DEVOGELE T. Potential of Crowdsourced Traces for Detecting Updates in Authoritative Geographic Data. In: KYRIAKIDIS P.; HADJIMITSIS D.; SKARLATOS D.; MANSOURIAN A.

(eds) **Geospatial Technologies for Local and Regional Development**. AGILE 2019. Lecture Notes in Geoinformation and Cartography. Springer, Cham.

JACOBS, K. T. **Quality Assessment of Volunteered Geographic Information: An Investigation into the Ottawa-Gatineau OpenStreetMap Database**. Master Thesis. Carleton University, Ottawa, Ontario, 2018.

JASIM, S.; AL-HAMADANI, O. Positional Accuracy Assessment for Updating Authoritative Geospatial Datasets Based on Open Source Data and Remotely Sensed Images. **Journal of Engineering**, 26(2) pp. 70-84, 2020.

JOHNSON, P. A.; SIEBER, R. E. Motivations driving government adoption of the Geoweb. **GeoJournal**, vol. 77, 2012.

KEATES, J. S. Cartographic design and production. **UK: Longman**, 1973.

KENT, A. Topographic maps: methodological approaches for analyzing cartographic style. **Journal of Map & Geography Libraries**, 5(2), pp. 131-156, 2009.

KENT, A. J.; HOPFSTOCK, A. Topographic Mapping: Past, Present and Future. **The Cartographic Journal**, v. 55, n. 4, p.305-308, 2018.

KHAN, Z. T.; JOHNSON, P. A. Citizen and government co-production of data: Analyzing the challenges to government adoption of VGI. **The Canadian Geographer/Le Géographe canadien**, 2020.

KLEIN, I.; GUZATTO, M. P.; HASENACK, M.; CABRAL, C. R.; LIMA, A. P. B.; FRITSCHÉ, S.; REGINA JR., L. A. M.; MOMO, G. F. Rede de Referência Municipal para Estações Livres: Uma proposta de baixo custo e grande abrangência. **Revista Brasileira de Cartografia**, vol. 69, 2017.

KOUKOLETSOS, T. **A Framework for Quality Evaluation of VGI linear datasets**. Doctoral Thesis, University College London, 2012.

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, A. A.; CAMBOIM, S. P. Desambiguação dos Termos Mapeamento Topográfico em Grandes Escalas e Mapeamento Cadastral no Brasil. **Revista Brasileira de Cartografia**, 71(2), 295-327, 2019.

MACHADO, A. A.; CAMBOIM, S. P. Mapeamento colaborativo como fonte de dados para o planejamento urbano: desafios e potencialidades. **urbe. Revista Brasileira de Gestão Urbana**, 11, 2019.

MACGILLAVRY, E. Collaborative Mapping: By the people, for the people. **Society of Cartographers Bulletin**, 37(2): 43-45, 2003.

MAGUIRE, D. **GeoWeb 2.0: Implications for ESDI**, In: Proceedings of the 12th EC-GI&GIS Workshop, Innsbruck, Austria, 2006.

MARANHÃO, V. C. **Modelagem e controle de qualidade de uma infraestrutura de dados espaciais para o Estado de Pernambuco**. Dissertação (Mestrado em Ciências Geodésicas Tecnologias das Geoinformação). Universidade Federal de Pernambuco (UFPE). Recife, 2013.

MARTINI, A.; KUPER, P.V.; BREUNIG, M. **Database-supported change analysis and quality evaluation of OpenStreetMap Data**. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences ISPRS, GEOSPATIAL WEEK. 535-541 June 2019, Enschede: The Netherlands, 2019.

MARTINS JUNIOR, O. G.; STRAUCH, J. C. M.; SANTOS, C. J. B. D.; BORBA, R. L. R.; SOUZA, J. M. D. Informação geográfica voluntária no processo de reambulação. **Boletim de Ciências Geodésicas**, 22(4), 613-629, 2016.

MARTINS JUNIOR, O. G.; SILVA, L. F. C. F. Proposta de Hierarquia para Conceitos de Cartografia Colaborativa. **Anuário do Instituto de Geociências**, 41(3), 560-567, 2019.

MAULIA, M. **Development of an update procedure for authoritative spatial data by the combination with crowdsourced information**. Master Thesis. Technische Universität Dresden, 2018.

MEDEIROS, G. F. B. **OpenStreetMap: Uma Análise Sobre a Evolução de Dados Geográficos Colaborativos no Brasil**. Trabalho de Conclusão de Curso (Graduação em Engenharia da Computação). Universidade de Brasília, 2017.

MERCHANT, D. C. **Spatial Accuracy Standards for Large Scalle Line Maps**. In: Technical Congress on Surveing and Mapping. Proceedings. v. 1, 1982.

MINGHINI, M.; FRASSINELLI, F. OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date?. **Open Geospatial Data, Software and Standards**, 4(1), 9, 2019.

NAIK, U.; SHIVALINGAIAH, D. Comparative Study of Web 1.0, Web 2.0 and Web 3.0, **Allahhabad: CALIBER**, 2008.

NASIRI, A.; ALI ABBASPOUR, R.; CHEHREGHAN, A. JOKAR ARSANJANI, J. Improving the quality of citizen contributed geodata through their historical contributions: The case of the **Road** network in OpenStreetMap. **ISPRS International Journal of GeoInformation**, 7(7), 253, 2018.

NEIS, P.; ZIELSTRA, D. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. **Future Internet**, 6(1), 76-106, 2014.

NEIS, P.; ZIPF, A. Analyzing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. **ISPRS International Journal of GeoInformation**, 1(2), 146-165, 2012.

OCDE – Organização para Cooperação e Desenvolvimento Econômico.

Participative Web: User-Created Content. Techn. Ber. Directorate for Science, Technology, Industry. Committee for Information, Computer and Communications Policy. 74p, 2007.

OLTEANU-RAIMOND, A. M.; HART, G.; FOODY, G. M.; TOUYA, G.; KELLENBERGER, T.; DEMETRIOU, D. The scale of VGI in map production: A perspective on European National Mapping Agencies. **Transactions in GIS**, 21(1), 74-90, 2017.

OLTEANU-RAIMOND, A.-M.; SEE, L.; SCHULTZ, M.; FOODY, G.; RIFFLER, M.; GASBER, T.; JOLIVET, L.; LE BRIS, A.; MENEROUX, Y.; LIU, L.; POUPÉE, M.; GOMBERT, M. Use of Automated Change Detection and VGI Sources for Identifying and Validating Urban Land Use Change. **Remote Sens**, 12, 1186, 2020.

OPENSHAW, S.; RHIND, D.; GODDARD, J. Geography, geographers, and the BBC Domesday Project. **Area**, 18(1): 9–13, 1986.

OSM. **OpenStreetMap Foundation.** Disponível em: <http://wiki.osmfoundation.org/wiki/>. Acesso em: 20/07/2020.

OSM. **OpenStreetMap stats.** Disponível em: https://www.openstreetmap.org/stats/data_stats.html. Acesso em: 20/07/2020a.

OSM. **Tags.** Disponível em: https://wiki.openstreetmap.org/wiki/Tags#Keys_and_value. Acesso em 20/07/2020b.

OSM. **Elements.** Disponível em: <https://wiki.openstreetmap.org/wiki/Elements>. Acesso em: 20/07/2020c.

OSM. **Editors.** Disponível em: <https://wiki.openstreetmap.org/wiki/Editors>. Acesso em: 20/07/2020c.

OSM. **Pt:iD**. Disponível em: <https://wiki.openstreetmap.org/wiki/Pt:iD>. Acesso em: 20/07/2020d.

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, C. A.; Camboim, S. P. Inference of positional accuracy of collaborative data from intrinsic parameters. **Transactions in GIS**, 26(4), 1898-1913, 2022.

PAPAPESIOS, N.; ELLUL, C.; SHAKIR, A.; HART, G. Exploring the use of crowdsourced geographic information in defence: challenges and opportunities. **Journal of Geographical Systems**, 21(1), 133-160, 2019.

PAROUTIS, S.; AL SALEH, A. Determinants of knowledge sharing using Web 2.0 technologies. **Journal of knowledge management**, 2009.

PERRET, J.; GRIBAUDI, M.; BARTHELEMY, M. Roads and cities of 18th century France. **Scientific data**, 2(1), p. 1-7, 2015.

PERKINS, C. Researching mapping: methods, modes and moments in the (im) mutability of OpenStreetMap. **Global Media Journal: Australian Edition**, 5(2), p. 1-12, 2011.

RAJARAM, G.; MANJULA, K. R. Exploiting the Potential of VGI Metadata to Develop A Data-Driven Framework for Predicting User's Proficiency in OpenStreetMap Context. **ISPRS International Journal of GeoInformation**, 8(11), 492, 2019.

ROBINSON, A.H.; MORRISON, J.L.; MUEHRCKE, J.L.; KIMERLING A.J.; GUPTILL, S.C. **Elements of Cartography**, 6ª Edição. Nova Iork, 1995.

ROCHE, S.; PROPECK-ZIMMERMANN, E.; MERICKSKAY, B. GeoWeb e gestão de crise: questões e perspectivas da informação geográfica voluntária. **GeoJournal**, 78, p. 21–40, 2013.

ROTA, J.; ALBERT, P.; GINÉ, S. Visitor monitoring in protected areas: an approach to Natura 2000 sites using Volunteered Geographic Information (VGI). **Geografisk Tidsskrift-Danish Journal of Geography**, 119(1), p. 69-83, 2019.

SANTOS, T.; FREIRE, S.; FONSECA, A.; TENEDÓRIO, J. A. Producing a building change map for urban management. **EARSel EProceedings**, vol. 10, no. 1, p. 56, 2011.

SEE, L.; MOONEY, P.; FOODY, G.; BASTIN, L.; COMBER, A.; ESTIMA, J.; FRITZ, S.; KERLE, N.; JIANG, B.; LAAKSO, M.; LIU, H.-Y.; MILČINSKI, G.; NIKŠIČ, M.; PAINHO, M.; PÖDÖR, A.; OLTEANU-RAIMOND, A.-M.; RUTZINGER, M. Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. **ISPRS International Journal of Geographic Information**, 5(55): 23p, 2016.

SEHRA, S. S.; SINGH, J.; RAI, H. S. Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. **Future Internet**, 9(2), 15, 2017.

SENARATNE, H.; MOBASHERI, A.; ALI, A. L.; CAPINERI, C.; HAKLAY, M. A review of volunteered geographic information quality assessment methods. **International Journal of Geographical Information Science**, 31(1), 139-167, .2017.

SETO, T.; KANASUGI, H.; NISHIMURA, Y. Quality Verification of Volunteered Geographic Information Using OSM Notes Data in a Global Context. **ISPRS International Journal of GeoInformation**, 9(6), 372, 2020.

SIEBER, R. E.; ROBINSON, P. J.; JOHNSON, P. A.; CORBETT, J. M. Doing public participation on the geospatial web. **Annals of the American Association of Geographers**, 106(5), 1030-1046, 2016.

Silva, L. S. L. **Integração de dados provenientes de mapeamento colaborativo na cartografia de referência do Brasil**. Curitiba, 2022. 153 f. Tese (Doutorado em Ciências Geodésicas) – Universidade Federal do Paraná, Paraná.

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. DOI: 10.1590/s1982-21702021000100003

SILVA, P. L.; SPERANDIO, V. G.; RODRIGUES, L. F.; DIAS, V. E. C.; LOTI, L. B.; LISBOA, M. H.; OLIVEIRA, G. D.; MIRANDA, G. H.; MARTINS, G. S.; MONTEIRO, C. R.; MENDES, V. F.; LISBOA FILHO, J. Bases Cartográficas para Municípios de Pequeno Porte geradas por Informação Geográfica Voluntária. **Revista Brasileira de Cartografia**, v. 70, p. 1664-1685, 2018.

SILVEIRA, F. **Proposição de símbolos pontuais para o mapeamento topográfico em escala grande com base na percepção de usuários**. Dissertação (Mestrado em Ciências Geodésicas). Universidade Federal do Paraná, Curitiba, 2019.

SLUTER, C. R.; CAMBOIM, S. P.; IESCHECK, A. L.; PEREIRA, L. B.; CASTRO, M. C.; YAMADA, M. M.; ARAÚJO, V. S. A Proposal for topographic map symbols for large-scale maps of urban areas in Brazil. **The Cartographic Journal**, 55(4), 362-377, 2018.

STAMP, D. The land utilization survey of Britain. **The Geographical Journal**, 78(1): 40–47, 1931.

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.

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-

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

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).

TULLOCH, D. Public participation GIS (PPGIS), K. Kemp (Ed.), Encyclopedia of geographic information science, **SAGE Publications**, Inc, Thousand Oaks, CA, pp. 352-355, 2008.

TURNER, A. Introduction to neogeography. **O'Reilly Media**, Inc, 2006.

VIANA, C. M.; ENCALADA, L.; ROCHA, J. The value of OpenStreetMap historical contributions as a source of sampling data for multi-temporal land use/cover maps. **ISPRS International Journal of GeoInformation**, 8(3), 116, 2019.

WAGLE, N.; ACHARYA, T. D. Past and Present Practices of Topographic Base Map Database Update in Nepal. **ISPRS International Journal of Geo-Information**, 9(6), 397, 2020.

YANG, C.; YU, M.; HU, F.; JIANG, Y.; LI, Y. Utilizing cloud computing to address big geospatial data challenges. **Computers, Environment and Urban Systems**, 61, 120-128, 2017.

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.

ZHANG, H.; MALCZEWSKI, J. Accuracy Evaluation of the Canadian OpenStreetMap Road Networks. **International Journal Geospatial and Environmental Research**, 5, pp. 1-14, 2018.

ZHANG, S. Public participation in the Geoweb era: Defining a typology for geo-participation in local governments. **Cities**, 85, 38-50, 2019.

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, D.; ZIPF, A. Quantitative studies on the data quality of OpenStreetMap in Germany. **Proceedings of GIScience**, 2010.

8. APÊNDICES

8.1 O MAPEAMENTO DE REFERÊNCIA E A QUALIDADE DE DADOS GEOESPACIAIS SINERGIAS ENTRE A PRODUÇÃO DOS DADOS GEOESPACIAIS DE REFERÊNCIA OFICIAIS E COLABORATIVOS: UMA PROPOSIÇÃO DE EIXOS POTENCIAIS

Autores: Everton Bortolini, Leonardo Scharth Loureiro Silva, Elias Nasr Naim Elias, Silvana Philippi Camboim, Marcio Augusto Reolon Schmidt.

Evento: 2º Simpósio Brasileiro de Infraestrutura de Dados Espaciais – SBIDE (2020).

Resumo:

A crescente demanda por informação geoespacial tem se evidenciado, uma vez que as sociedades vêm aumentando o espectro de tarefas e serviços que dependem de mapeamento completo e atualizado, sobretudo nas maiores escalas. Uma das novidades que a *web 2.0* proporcionou, tanto em termos de novas ferramentas quanto de novas técnicas, foi o mapeamento colaborativo, que permite que qualquer pessoa não especialista em cartografia possa se tornar um mapeador. Diante das dificuldades e das limitações para prover mapeamento de referência que atenda plenamente às demandas atuais, as organizações governamentais, responsáveis pela sua execução, compostas por cartógrafos com capacitação formal e experiência profissional, tem estudado incorporar informações provenientes de mapeamento colaborativo. Assim, para alcançar o enlace entre ambos os tipos de mapeamento é necessário compreender o sentido que as informações geográficas trafegam entre eles. Apesar de já existirem pesquisas que exploram esta demanda, ainda há necessidade de estudar possibilidades e restrições quanto a utilização desse tipo de mapeamento no Brasil. Portanto, este trabalho tem o intuito de aprofundar como o estudo das questões legais, semânticas e de disponibilidade tornam possível a sinergia entre a comunidade que mantém o mapeamento colaborativo, *OpenStreetMap* e as instituições responsáveis pelos dados de referência oficiais brasileiros em diversas escalas. Assim, partimos das premissas que as interações entre os mapeamentos de referência e colaborativo não são unidirecionais e de que existem tipos de categorias e contextos das informações geográficas em que as mesmas transitam entre estes mapeamentos. A partir disso,

temos por objetivo neste trabalho descrever quais e como se dá às relações entre os dois tipos de mapeamento. Este resumo traz uma compilação de resultados dos trabalhos relacionados ao nosso grupo de pesquisa. Assim, consultamos fontes de dados que passaram pela: 1) Análise documental de legislações que definem a localização de feições geográficas; 2) Análise das informações geográficas existentes na base de dados do *OpenStreetMap* quanto à completude de algumas áreas de estudos e tendências relacionadas com o aumento na quantidade de feições mapeadas; 3) Análise das classes de informações geográficas do modelo de dados usado no Brasil, a Estrutura de Dados Geoespaciais Vetoriais (ET-EDGV), em pareamento ao modelo do *OpenStreetMap*. A partir das análises que realizamos, obtivemos os resultados que apresentamos na sequência. As informações geográficas cuja definição espacial é dada por dispositivos legais, como limites administrativos ou de áreas de preservação sociocultural e ambiental, sempre sairão do mapeamento oficial para o mapeamento colaborativo. Contudo, esses tipos de informações, que visam representar a delimitação de uma entidade, têm muitas vezes sua definição dada por feições que são de outras categorias como rios ou vias. Pontos de referência, ou seja, marcos geodésicos e afins, também são feições que vão do oficial para colaborativo por serem definições legais, na mesma lógica dos limites. Feições relativas ao relevo, no caso as curvas de nível, por uma limitação tecnológica das ferramentas de mapeamento colaborativo ainda devem vir do mapeamento oficial. No que se refere à categoria de vegetação, os mapeamentos de referência costumam possuir mais detalhamento em relação às classificações das feições. Por outro lado, embora também possuam bom nível de detalhamento, as informações colaborativas possuem maior capacidade de delimitação e identificação de áreas cobertas por vegetação ou aquelas que vêm sofrendo degradação da sua cobertura. Nas análises de completude e no histórico de feições mapeadas no *OpenStreetMap*, em diferentes regiões, notamos a existência de contribuições em massa em um determinado período, além disso, identificamos que este aspecto ocorreu em diferentes épocas para uma mesma área de estudo. Desta forma, ao partir do pressuposto que este fato indica importações realizadas por um único contribuinte e que estas podem ter sido desencadeadas a partir de dados oficiais, notamos que este aspecto influencia com que novas contribuições sejam realizadas de forma gradativa por voluntários, contudo esta influência ocorre de maneira distinta em cada área de estudo avaliada. Portanto,

para os grupos feições ora não citados podemos considerar a possibilidade de que elas possam contribuir duplamente nos dois tipos de mapeamento, pois quanto mais informações geográficas vão do oficial para o colaborativo, mais facilitada é a produção de novas informações no colaborativo, que por sua vez podem voltar ao oficial. Neste contexto vale pontuar que este tipo de ganho também é possível quando organizações, não só governamentais e focadas em cartografia, distribuem seus dados de forma aberta, e que o mapeamento oficial pode se beneficiar de forma direta dessas fontes, ou indiretamente quando elas serviram para alimentar as bases de dados das plataformas de mapeamento colaborativo. Por fim, este trabalho representa um passo nas discussões sobre a integração entre o mapeamento oficial e o colaborativo. Conforme apresentamos, concluímos para as várias categorias citadas, quanto ao fluxo de informações geográficas entre os dois tipos de mapeamento, que existe uma tendência que muitas destas informações sejam importadas do oficial para o colaborativo. Contudo, ainda existe uma contribuição potencial do colaborativo no em enriquecer os dados existentes do mapeamento de referência. A partir da metodologia abordada, notamos a necessidade de estabelecer uma abordagem explicativa do que observamos nos resultados. As análises dos itens 1 e 2 da metodologia são exploratórias e a análise do item 3 no caso é descritiva. Em todos os casos é preciso destacar a quantidade de categorias a serem compatibilizadas entre os modelos, o que limita uma descrição completa. Para trabalhos futuros é interessante avançar em detalhes sobre as informações geográficas contidas nas bases do mapeamento oficial, em bases colaborativas como o *OpenStreetMap*. Além das bases nomeadas, também podemos considerar a prospecção dos bancos de dados com informações geográficas que funcionam sob a política de dados abertos. Recomendamos, ainda, investigar mais a fundo os padrões de crescimento na quantidade de feições mapeadas, a partir das diferentes categorias que são importadas e contribuídas gradativamente. Estas duas últimas considerações podem ajudar a responder alguns apontamentos apresentados neste resumo, pois as bases abertas, independente do tipo de organização ou de indivíduo a elas relacionados, podem também interagir com o mapeamento oficial e colaborativo, formando assim um terceiro ente nas análises.

8.2 O A PROPOSAL FOR A QGIS PLUGIN FOR SPATIO-TEMPORAL ANALYSIS OF OSM DATA QUALITY: THE CASE STUDY FOR THE CITY OF SALVADOR, BRAZIL

Authors: Elias Elias, Fabricio Amorim, Leonardo Silva, Marcio Schmidt, Silvana Camboim and Vivian Fernandes.

Event: Proceedings of the Academic Track, State of the Map 2021 July 9-11, 2021 - Online Conference.

Abstract:

The development of methodologies to evaluate geospatial data quality is one of the most important aspects to be considered while obtaining this data. For developing countries, such as Brazil, the lack of investment for the maintenance of topographic mapping, especially on a big scale, is a recurrent challenge for the National Mapping Agencies [1]. For example, studies reveal areas in Brazil that have never been mapped and that the topographic mapping in the 1:25,000 scale is nearly 5% of its extension [2].

Technological advances have enabled a series of methodologies for obtaining geospatial data [3]. One example is presented as Volunteered Geographic Information (VGI) [4]. In this case, the update of information may occur faster and with a reduced cost in detriment to the traditional structures of topographic mapping [5]. A successful case of VGI is the OpenStreetMap (OSM) project, which presents the growth in the number of contributors and contributions or mapped features. To understand the quality of OSM features and their integration potential in topographic mapping, different surveys worldwide have put efforts to evaluate its quality, whether by its extrinsic [6, 7] or intrinsic [8] aspects. In this regard, some studies have evaluated the quality of OSM features by combining extrinsic and intrinsic aspects, like [9], which evaluated the positional accuracy of OSM based on the combination of edit history. Besides that, recent works have focused on comprehending spatial and temporal aspects of events in OSM contributions [10], as well developing add-ons for evaluating data quality, as presented by [11], where the authors developed a QGIS toolbox to evaluate parameters of the intrinsic quality of OSM features.

The literature identifies the heterogeneity of the data as one of the main challenges for the integration processes. The value of the quality parameters may

vary according to the study area, the indicator used or even the temporal variations in the dynamics of the geographical space itself. In this context, to understand the integration of OSM data to the topographic mapping, it is crucial to connect aspects related to the quality and heterogeneity of data. Research work like [1] argues that, based on the obtained quality, the resources resulting from VGI may be used to integrate, detect changes or report errors. Therefore, classifying resources from OSM according to their usability in a certain region becomes essential, especially in developing countries like Brazil. Besides that, research that explores issues of quality, heterogeneity, and contributions patterns of OSM is still not widespread in developing countries [12].

Given the importance of classifying OSM features according to their usability for a given region, especially in developing countries, few researchers have explored quality, heterogeneity, and contribution pattern issues in OSM in Brazil. Based on the issues addressed, we propose a hypothesis that understanding aspects of the extrinsic and intrinsic quality of the quality of OSM features will help decision making regarding the integration of such data in topographic mapping. The main focus is on the spatio-temporal aspects of contributions in developing countries. Thus, this research has the objective to evaluate the extrinsic quality of OSM features for the county of Salvador-Bahia-Brazil (the northeast region of the country). Therefore, we investigated indicators of positional accuracy, thematic accuracy and completeness, the visualisation of heterogeneity of data, and the analysis of the edition history. To accomplish the evaluation of extrinsic quality, the OSM features were compared to the topographic mapping of the country from the Cartographic and Cadastral System of the County of Salvador (Sistema Cartográfico e Cadastral do Município do Salvador - SICAD, 2006) and features from the Urban Development Company of the State of Bahia (Companhia de Desenvolvimento Urbano do Estado da Bahia - CONDER).

The analysis of positional and thematic accuracy was made through procedures of feature sampling. The analysis of completeness occurred from comparing the total number of available features. The verified categories were features from the road system, religious, educational, and health buildings. We divided the municipality of Salvador into sub-regions to identify different local patterns of quality in the analysis of thematic accuracy and completeness. Visualisation allows obtaining the data's heterogeneity through a plugin developed in the software QGIS,

making the planimetric positional evaluation for point and line features. The statistical procedures for developing the plugins were realised based on the Brazilian law to evaluate geospatial data quality analysis [13] and based on the method of double buffer proposed by [14]. The plugin is available, and it is possible to be accessed in the online repository [15]. Even though the final results comprehend aspects of Brazilian law, they can be replicated to obtain the discrepancies and posterior adjustments. We used the Ohsome Application Programming Interface (API) [16] to identify the patterns concerning the OSM editing history. Thus, from the adaptations performed in scripts given by researchers linked to Ohsome, it was possible to identify the aspects of OSM contributions between 2008 and 2020. We also tested the generation of regression curves and calculated the number of daily contributions to identify these patterns. These verifications were occasionally created through the generation of an evolving rectangle of 5x5 km in the study area. The disposition of the rectangle was given through a visual analysis with a larger quantity of OSM features.

The evaluation of extrinsic evaluation highlighted the variability of the results obtained in [17]. In analysing the positional accuracy, the scale found varied from 1:20,000 to 1:30,000, while the discrepancies between the mapped coordinates and the reference one varied between 0.12m and 10.27m. In analysing completeness, it was observed that features that corresponded to the road system presented better results concerning the other categories. The road system presented a completeness percentage of 82%, while in the other features, the variation was from 29% to 46%. When analysing thematic accuracy, it turns out that the primary source of errors is related to the absence of names in editing. In the analysis of the OSM contribution history growth of represented features, it was possible to notice a near-linear function, with an R^2 value of 0.94. As there is a finite amount of mappable elements at a given area, we can use this function to model the contributions patterns over time until the region is saturated. Besides that, it was possible to observe that the patterns of collaboration can be affected by different variables because it was noticed that in 2016, more than 800 features were added in a short period. These aspects can be related to events such as data importation or mapathons.

The development of add-ons for evaluating OSM data quality that departs from the making of statistical procedures up to visualising the heterogeneity of data will assist in the decision-making as to data quality. The development of QGIS

plugins for OSM data quality assessment that execute from statistical procedures to visualisation of data heterogeneity will assist in decision making regarding data quality.

From the add-on developed, it was possible to notice that the magnitude of discrepancies did not present patterns and that this may vary according to the period of editing and the database used for the contributions. Based on the obtained results, we noticed the relevance in identifying the aspects of quality and heterogeneity in OSM contributions. For Brazil, identifying these characteristics may numerally indicate the integration potential of these data to authoritative mapping. Besides that, it will estimate the influence of unusual agents, like it is the case of data import in the contributions. The continuity of the studies is recommended to identify the causes of different patterns of growth and the continuity of studies to automatise the quality procedures.

8.2.1 References

- [1] Maulia, N. (2018). Development of an update procedure for authoritative spatial data by the combination with crowdsourced information. Master's thesis, Technische Universität Dresden.
- [2] Silva, L. S., & Camboim, S. P. (2020). Brazilian Nsdi Ten Years Later: Current Overview, New Challenges And Propositions For National Topographic Mapping. *Boletim De Ciências Geodésicas*, 26(4).
- [3] Brovelli, M. A., Boccardo, P., Bordogna, G., Pepe, A., Crespi, M., Munafò, M., & Pirotti, F. (2019). Urban Geo Big Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W14, 23-30.
- [4] Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221.

- [5] Du, H., Alechina, N., Jackson, M., & Hart, G. (2017). A method for matching crowd-sourced and authoritative geospatial data. *Transactions in GIS*, 21(2), 406-427.
- [6] Brovelli, M. A., & Zamboni, G. (2018). A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of Geo-Information*, 7(8), 289.
- [7] Zhang, H., & Malczewski, J. (2017). Accuracy evaluation of the Canadian OpenStreetMap Road networks. *International Journal of Geospatial and Environmental Research*, 5(2).
- [8] Minghini, M., & Frassinelli, F. (2019). OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data, Software and Standards*, 4(1), 9.
- [9] Nasiri, A., Ali Abbaspour, R., Chehreghan, A., & Jokar Arsanjani, J. (2018). 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*, 7(7), 253.
- [10] Grinberger, A. Y., Schott, M., Raifer, M., & Zipf, A. (2021). An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS*, 25(2), 622- 641.
- [11] Sehra, S. S., Singh, J., & Rai, H. S. (2017). Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. *Future Internet*, 9(2), 15.
- [12] Camboim, S. P., Bravo, J. V. M., & Sluter, C. R. (2015). An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. *ISPRS International Journal of Geo-Information*, 4(3), 1366-1388.

[13] DSG. (2016). Norma da Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG). Diretoria do Serviço Geográfico, 1, 94, Brasília-DF.

[14] Santos, A. P. (2015). Controle de qualidade cartográfica: Metodologias para avaliação da acurácia posicional em dados espaciais. PhD thesis, Universidade Federal de Viçosa.

[15] Elias, E. (2019). AcuraciaPosicional_PEC-PCD. Retrieved from https://github.com/eliasnaim/AcuraciaPosicional_PEC-PCD.

[16] Heidelberg Institute for Geoinformation Technology (2021). Ohsome. Retrieved from <https://heigit.org/big-spatial-data-analytics-en/ohsome>.

[17] Elias, E. N. N. (2019). Qualidade de Dados Geoespaciais em Plataforma de Mapeamento Colaborativo. Master thesis, Universidade Federal da Bahia.

8.3 THE CELL SIZE ISSUE IN OPENSTREETMAP DATA QUALITY PARAMETER ANALYSES: NA INTERPOLATION-BASED APPROACH

Authors: Silvana Camboim, Caio Paiva, Marcio Schmidt, Elias Elias, Everton Bortolini, Guilherme Neivas.

Event: Proceedings of the Academic Track, State of the Map 2022 August 19 - 21, 2022 | Florence, Italy.

Abstract:

Knowing the quality of a given geospatial data allows measuring how much its use can be viable in specific applications and assist in decision making. ISO 19157 [1] established that the geospatial data quality indicators are positional accuracy, temporal accuracy, thematic accuracy, logical consistency, and completeness. These measures are represented by values that summarize the condition of a product as a whole. These values tend to be homogeneous throughout the evaluated area in traditional mapping. In contrast, in VGI, data quality can be affected by several conditions related to editing history, contribution period, and

contributor profiles [2,3]. Given the mentioned aspects, data quality in VGI platforms tends to be heterogeneous, i.e., the results may show significant discrepancies according to the area assessed or even within the same region.

Given the heterogeneity issues described, several researchers around the world have performed the quality assessment of these types of information based on the principle of subdividing the study area into cells [4–8]. Such a procedure has been used in extrinsic quality assessment processes based on ISO 19157 indicators or intrinsic parameters associated with the characteristics of the contributions and contributors. Given the results obtained, the representation of the quality of the data from sub-areas makes it possible to obtain analyses regarding the existence of patterns and establish relationships with other agents and their predominance. The discretization of space into rectangular or hexagonal grids is central to this type of analysis.

The subdivision can occur using grids or based on different sizes of areas. The units with non-equal area cells allow us to perform analyses accepting other features or spatial phenomena that define these dimensions (e.g., neighbourhood border, a river or a railway track, areas with different population densities, and the dichotomy between rural and urban areas). However, these methods make operations difficult because they demand that the area value weigh the values; and the spatial analysis considering the neighbourhood is more complex. Units with equal area cells (e.g. grid with rectangular or hexagonal cells) solve these two limitations. However, the problem of the grid of cells not conforming to spatial phenomena or features reappears. In order to conform to them, it is necessary to determine the optimal size of the cells.

However, one issue remains little discussed: how to determine the size of such cells. Using too large a cell would treat unequal areas equally. On the other hand, using too small a cell and the increased computational cost of the process, ultimately, the ability to generalize the results is lost. Therefore, in this work, we seek to develop an interactive approach for determining the grid cell size calculation, initially using points of known positional accuracy. The hypothesis here is that when the analyzed subarea is of optimal size, one can interpolate the error within the cell via an IDW and generate minimal residuals at the control points. Furthermore, by consecutively subdividing the grid, the mean squared error versus cell size curve will approach stability, thus revealing the optimal size for a given region.

IDW interpolation calculates cell values using sample point sets. This method considers that the higher weights in the interpolation should be due to the proximity of the unknown value point. Thus, if we had a homogeneous behaviour of the quality parameter in an individual area, by interpolation, we could estimate the quality of the points where this value was unknown.

The methodological procedures developed using python in the QGIS environment are: For the study area, points of known positional accuracy are chosen (in our case, intersections of the road system), from which a random subset of 10% is separated as a control set.

- Definition of a first grid;
- The points are used for interpolation within each cell by the IDW method.

The Root-mean-square deviation (RMSE) is calculated using the control points for each cell and the average of the RMSEs for the entire area;

- Definition of a second grid with half the resolution of the first grid;
- Repeat the process described in item 3 for the second grid;
- Calculate the differences between the average error values of the second grid and the first grid and check their significance;
- Repetition of the process described in case there are still values indicated as significant.

In a first analysis, we did a preliminary study for a Brazilian city, Curitiba, with about 28 thousand points of known accuracy. We separated 2.8 thousand control points, and the city was divided into 8 km to 250 m cells. From the preliminary study performed, it was noted that the method show promise in obtaining the necessary analyses to identify the aspects proposed in this work. Furthermore, it was noticed that, as the cell size decreased, the results tended to be more constant, which corroborates the hypothesis of this relationship with data quality. The next steps are to continue the analyses, starting with the verifications and the representation of the magnitude of the differences between different cell sizes.

Although it is a method that still has a relatively high computational cost to be realized, the results are exciting and can be optimized. It is assumed that if it is possible to identify the minimum cell size in which it is possible to estimate the quality of the features, this will help in decision making regarding the incorporation of procedures in different areas. This method may need even smaller clippings in

regions with very heterogeneous characteristics concerning their surroundings (e.g., slums). It is an initial approach to resolve with data a fundamental issue arising from the lack of knowledge of the granularity of discrepancies for each study area.

8.3.1 References

- [1] International Organization for Standardization (2013). ISO 19157:2013 Geographic Information – Data quality. Retrieved from <https://www.iso.org/standard/32575.html>.
- [2] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), 877–895.
- [3] Minghini, M., & Frassinelli, F. (2019). OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date?. *Open Geospatial Data, Software and Standards*, 4(1), 1–17.
- [4] Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, 47(4), 315–322.
- [5] Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28(4), 700–719.
- [6] Martini, A., Kuper, P.V., & Breunig, M. (2019). Database-supported change analysis and quality evaluation of OpenStreetMap Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 535–541.
- [7] Brovelli, M.A., & Zamboni, G. (2018). A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of Geo-Information*, 7(8), 289.

[8] Zhang, H., & Malczewski, J. (2017). Accuracy evaluation of the Canadian OpenStreetMap Road networks. *International Journal of Geospatial and Environmental Research*, 5(2).