Elias, Elias Nasr Naim; Fernandes, Vivian de Oliveira;
Alixandrini Junior, Mauro José; Schmidt, Marcio Augusto Reolon
The quality of OpenStreetMap in a large metropolis in northeast
Brazil: Preliminary assessment of geospatial data for road axes

The quality of OpenStreetMap in a large metropolis in northeast Brazil: Preliminary assessment of geospatial data for road axes

Elias Nasr Naim Elias[1] - ORCID: 0000-0003-2289-5055

Vivian de Oliveira Fernandes[2] - ORCID: 0000-0002-2851-9141

Mauro José Alixandrini Junior [2] - ORCID: 0000-0002-5376-7171

Marcio Augusto Reolon Schmidt [1] - ORCID: 0000-0003-2716-2360

[1]Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas, Curitiba - Paraná, Brazil.
E-mail: elias.naim@ufpr.br, marcio.schmidt@ufu.br

[2]Universidade Federal da Bahia, Programa de Pós-Graduação em Engenharia Civil, Salvador - Bahia, Brazil.
E-mail: vivian.fernandes@ufba.br, mauro.alixandrini@ufba.br

Abstract:

This paper evaluates the data quality of road axes using the OpenStreetMap (OSM) collaborative mapping platform. OSM was chosen owing to the abundance of data and registered contributors (~ 6 million). We assumed the OSM collaborative data could complement the reference mappings by its quality parameters. We used the cartographic quality indicators of positional accuracy, thematic accuracy, and completeness to validate vector files from OSM. We analyzed the positional accuracy of linear features and we developed the automation of the positional accuracy process. The tool verified the completeness of road axes and thematic accuracy. The positional accuracy of linear features was also used, performed to obtain a range of scales, which reflected the characteristics of mapped areas and varied from 1:22,500 to 1:25,000. The completeness of road axes was 82% of the checked areas. By evaluating the thematic accuracy, we found that the absence of road axes toponymy in editions caused errors in the OSM features (i.e., 58% of road axes without information). As such, we concluded that collaborative data complements the reference cartography by measuring the heterogeneity of information in various regions and filtering the OSM data, despite its being useful for certain analyses.

Keywords: Data quality; OpenStreetMap; Geospatial data; Road axis; Completeness; Positional Accuracy; Thematic Accuracy.

1. Introduction

Measuring quality in geospatial data is essential for reference mappings with higher accuracy or thematic purposes. The data quality process is beyond the assessment of geometric accuracy and relates to the completeness, thematic accuracy and, logical consistency. According to Keates (1973), reference mappings are developed by official mapping agencies to represent all features visible in a landscape for various uses (Kent, 2009). This concept has been the only approach associated with the production of geospatial data (Elwood, Goodchild and Sui 2012). However, technological development has changed the collection, processing, analysis, and dissemination of geospatial data. Development of new data sources, establishment of digital process methodologies, and the Internet have allowed the daily production of enormous data volumes. The increase in the volume of geospatial information available in virtual databases is associated with big data. The term "big data" refers to a flow of digital data in different sources, such as numerical modeling, smartphones, and access to the Internet and social networks (Yang et al. 2017).

Technological advances have allowed any individual with a computer or smartphone and internet access to generate geospatial data (Ganapati 2011). This characteristic is associated with web 2.0, which is defined by a phenomenon in which users become fundamental agents in the production and management of data, ceasing to be just viewers and consumers (Cormode and Krishnamurthy 2008). According to Cormode and Krishnamurthy (2008), the term web 2.0 was coined around 2004, and many of the first truly web 2.0 sites began emerging in late 2003 and early 2004.

In this context, Goodchild (2007) introduced the concept of Voluntary Geographic Information (VGI), where individuals with access to the Internet can publish geographical information on a specific location, validate information posted by other users, and produce geospatial data, functions that were previously available only to the specialists. According to Johnson and Sieber (2012), as VGI gained popularity, the governmental bodies of some countries, responsible for the generation of reference mapping, used VGI for the maintenance and update of otherwise costly and time-consuming cartographic bases. In addition, the official mapping agencies generating geospatial data are to embrace rapid technological advances in the geographical space (Maulia 2018).

VGI platforms are used to generate and update geospatial data. For example, since May 2020, over 6.5 million users and 14.7 billion features have been recorded in the OpenStreetMap platform (OSM). OSM was created in 2004 by Steve Coast, who established a web-based map data under an open license to add, visualize, and distribute geographic data. Steve Coast developed OSM in response to many restrictions for obtaining certain data from the European-based Ordnance Survey (Chilton 2011). According to Neis and Zipf (2012) and Bravo (2017), OSM enabled the establishment of a universal database of highways, streets, and paths. As a result, one can share map data without restrictions or copyrights, which substantially increases the platform's information coverage (Perkins 2011).

Several techniques are used to obtain geospatial data, ranging from the field-based topographic and aerial photogrammetric surveys to the web-based collaborative VGI platforms. Efficient data handling and integration help detect issues related to urban growth, establish methodological procedures to assess the data quality, and extract and integrate the growing amounts of geographic data (Brovelli et al. 2019). The instructions on how to assess the quality of geographic metadata are provided by the International Organization for Standardization in ISO 19157 (ISO 2013):

- Positional accuracy refers to the accuracy of the feature position (i.e., points, lines, areas) within a spatial reference system. It is usually assessed by comparing the position of features with their counterparts in reference data, which represent the "true" position;

- Thematic accuracy refers to the accuracy of classes or thematic tags associated with specific locations or objects placed in geographic space. Classes can be assigned to pixels in a land cover map, whereas tags can be assigned to a vector-encoded entity, such as a highway, river, building, or green area;

- Logical consistency refers to the adherence to logical rules for data structure, attribution, and relationships

as per the product's specifications;

- Completeness refers to the presence or absence of features, their attributes, and relationships compared to the product's specification. It is divided into a) commission, which explains the presence of excess data in a dataset, and b) omission, which explains the absence of data from a dataset.

- Usability (or fitness-for-use) refers to the external quality of a dataset and user needs. All mentioned data quality elements may be combined to describe the overall usability of a specific dataset for a particular use, i.e., its fitness-for-purpose.

The VGI data lacks the standard quality validation process for its products in terms of reference mapping. Therefore, a series of studies have analyzed the quality of these data types, aiming at their integration. The first analysis (Haklay 2010) compared the positional accuracy and completeness of data between the OSM platform and the United Kingdom reference mapping system. The possibility of using the VGI data was verified. The latest studies on the evaluation of the VGI data quality are summarized in Table 1.

Table 1: The latest research outcomes involving evaluation of VGI data quality in the world

Item	Authors	Country of study area	Analysis
1	Jasin and Hamandani (2020)	Iraq	Positional accuracy of pathways
2	Ferster et al. (2019)	Canada	Completeness of Bike Paths
3	Martini, Kuper and Breunig (2019)	Germany	Intrinsic quality assessment (History of editions and characteristics of contributors)
4	Mozas-Calvache and Ariza-López (2019)	Spain	Positional Accuracy of GNSS trails
5	Ibrahim, Darwish e Hefny (2019)	Egypt	Positional accuracy of features
6	Minghini and Frassinelli (2019)	Different european countries	Intrinsic quality assessment (Edit history)
7	Küçük and Anbaroğlu (2019)	Turkey	Positional Accuracy of Buildings
8	Zhang and Malczewski (2018)	Canada	Assessment of road quality
9	Brovelli and Zamboni (2018)	Italy	Positional Accuracy and Completeness of Buildings
10	Zhou (2018)	United States and New Zealand	Completeness of Buildings
11	Brovelli, Minghini and Moliari (2016)	Paris	Positional accuracy of pathways

Table 1 permits the observation of the characteristics of developed methods, defines knowledge gaps, and establishes a hypothesis for this research. We noticed that collaborative mapping lacks quick data updates on the changes in geographic space and the heterogeneity of information. The integration of reference mapping and VGI data requires frequent checks on the data quality related to the study location, number and characteristics of the contributors, form of data entry, and time of editing.

The topic of data heterogeneity has been covered by Jasim and Al-Hamadani (2020), Minghini and Frassinelli (2019), and Zhou (2018). Because their analyses were carried out in different study areas, it is still unclear whether heterogeneity serves as a quality indicator per ISO 19157 (ISO 2013). Regardless, heterogeneity should be evaluated and presented for the user. So far, Minghini and Frassinelli (2019) suggested the most suitable approach, where they were able to detect the issues associated with VGI data updating. Jasim and Al-Hamadani (2020), Küçük and Anbaroğlu (2019) analyzed geospatial data in the commercial software ArcGIS, which limits the user access by its high license cost. It also complicates the reusing of new tools developed by other users with a focus on application

development, the feature that is possible under an open OpenStreetMap license.

Quality control is important when working with geospatial data, especially with VGI, where no specifications for data creation exist. In geoinformation, the ISO principles and guidelines are considered for quality assessment. Recently, the ISO 19113 (ISO 2002) and ISO 19114 (ISO 2003) standards have been replaced by ISO 19157 (ISO 2013). The updated standard defines the following data quality elements: completeness, logical consistency, positional accuracy, temporal quality, thematic accuracy, and usability.

In Brazil, the production of reference cartographic data at a national level (a scale of less than 1:25,000) is coordinated by the Brazilian Institute of Geography and Statistics (IBGE) and the Directorate of the Geographic Service (DSG). It is important to point out that we present the acronyms in portuguese language. DSG provides matrix and vector topographic maps, ortho-images, and digital surface models at scales ranging from 1:25,000 to 1:250,000. IBGE supplies the sheets in matrix and raster formats at various scales and provides access to continuous bases in Brazil at the 1:1,000,000 and 1:250,000 scales. Mapping in Brazil at a scale of 1:25,000 provides a total area coverage of 5% (Sluter et al. 2019). Germany, Austria, and Switzerland have dense reference networks for the surveys, while some municipalities in Brazil still lack them for updated mapping (Klein et al. 2017). Camboim et al. (2015) pointed out that mapping Brazil, with its vast territory, is costly and is predominantly financed by governmental agencies. As such, we suggest integrating alternative methods for the quicker and cheaper production of geospatial data in Brazil by incorporating the quality-associated VGI data into reference cartographic bases.

The first evaluation of the geospatial data quality of cartographic products in Brazil dates back to the Decree-Law no. 89.817 "Technical Norms of National Cartography" (Brazil 1984). It established a set of diagnostic criteria for the accuracy and distribution of errors based on a statistical indicator of positional quality, the so-called Cartographic Accuracy Standard (PEC). PEC is divided into three classes (i.e., A, B, and C), where class A is characterized by the highest rigor of evaluation of the cartographic product, and class C has the lowest rigor. A PEC value is established for each class and is associated with a standard error. This standard error corresponds to 60.8% of the PEC value and a 90% probability of the normal distribution. The PEC values are used for both planimetric and altimetric assessments.

In 2011, the Technical Specification for the Acquisition of Vector Geospatial Data -ET-ADGV (DSG 2011) was presented by the Directorate of the Geographic Service (DSG). This specification updated the standards of the Decree-Law no. 89.817 (Brazil 1984), which failed to meet the needs of digital media. A new statistical indicator for ET-ADGV (DSG 2011) was established to revise the Cartographic Accuracy Standard for Digital Cartographic Products (PEC-PCD). This indicator is similar to PEC with an extra class (D). Thus, the cartographic products are classified as A, B, C, and D. PEC-PCD is associated with the scale of certain cartographic products and allows their classification by the maximum error obtained by the discrepancies that draw up a sample of points. It serves as a statistical indicator with a 90% confidence level based on a probabilistic graph of normal distribution. More recent versions of ET-ADGV developed by DSG in 2015 and 2016, the so-called Technical Specification for the Acquisition of Vector Geospatial Data for Defense of the Land Force -ET-ADGV-DefenseFT (DSG 2015; DSG 2016), addressed several aspects of PEC-PCD. In 2015, DSG created the Technical Specification for Quality Control of Geospatial Data - ET-CQDG (DSG 2015), which covers methodologies for assessing the quality of cartographic products similar to ISO 19157 (ISO 2013). ET-CQDG (DSG 2015) allows calculating the Euclidean distances of a sample of homologous points between the reference and evaluated cartographic products. The calculated Euclidean distance permits the classification of the cartographic product on a scale and evaluates it based on two conditions: 1) whether 90% of the Euclidean distance is less than or equal to the PEC-PCD value, and 2) whether the square mean error is less than or equal to the standard error.

Although several studies have verified the quality of OSM, gaps regarding its use still exist, especially given the lack of a standard in the data provision and VGI heterogeneity. Moreover, some VGI methodologies have already been incorporated into the reference mapping. As such, we aim to assess the quality of the OSM data from different sectors of the same study area using various sample spaces by measuring the OSM quality aspects and obtaining the characteristics of data heterogeneity. The study results show that we may successfully incorporate the existing geospatial information from other parts of the world and improve the parameters for integrating this data into the reference mapping.

2. Methodology

The study area is located in Salvador (BA), Brazil, with a population of approximately 2.9 million people (2017). Salvador (BA) is the largest city in the Northeast Region, with a total area of 693 km2 and a population density of 4,187/km². The reference map used for this study forms part of the cartographic and cadastral system of the municipality (SICAD). Figure 1 shows the subdivision of the municipality of Salvador (BA) by prefectures and neighborhoods. Besides the subdivision by prefectures and neighborhoods, the physical completeness of the road axes was determined by their overlapping with the Salvador (BA) development regions (e.g., central area, coastal expansion, *miolo,* and *Subúrbio*).
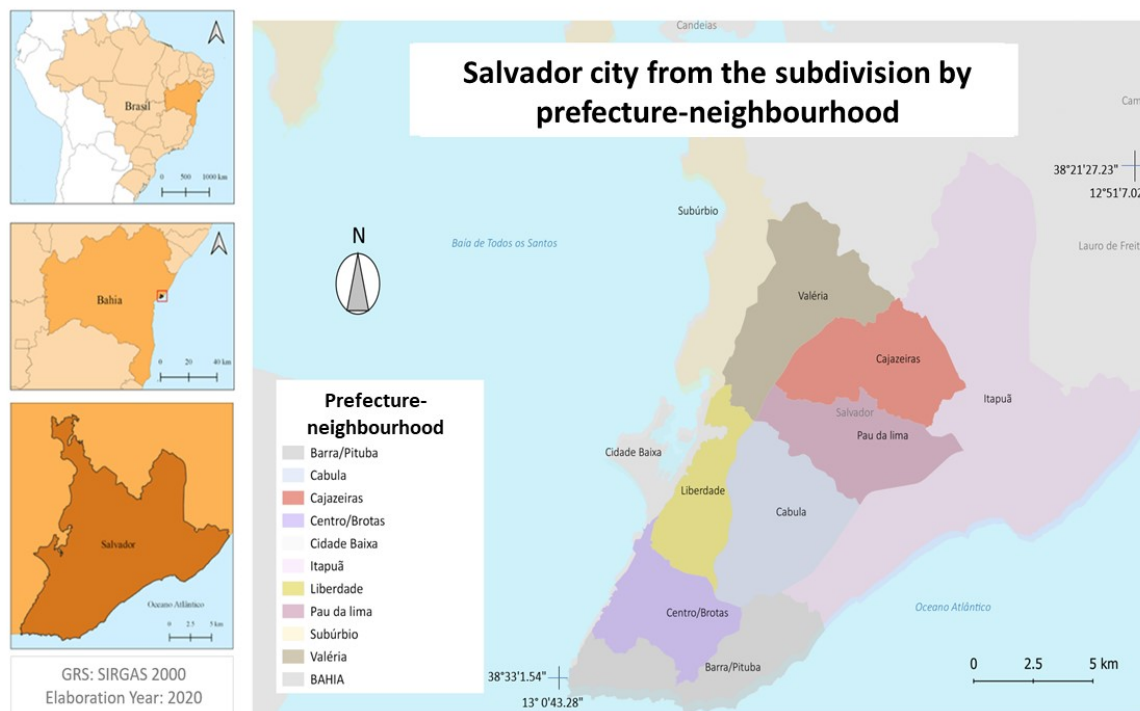

Figure 1: Localization map of Salvador city from the subdivision by prefecture-neighbourhood

First, we evaluated the positional accuracy of the road axes using the double buffer method by Tveite and Langaas (1999) and adapted by Santos (2015). We followed the Cartographic Accuracy Standard for digital cartographic products in our statistical tests. Merchant (1982) presented a sampling method to select punctual features, where at least 20 control points are needed to evaluate the positional quality. We applied the same methodology using two samples with 20 linear features.

The behavior of point features tends to differ from that of linear features, and more studies are needed on the sampling method for linear features. However, this issue was ignored herein as we established the positional accuracy following Merchant (1982). Other studies have shown that the positional accuracy is the main indicator of the geospatial data quality (Ibrahim, Ramadan and Hefny 2019; Fonte et al. 2017; Roberto 2013; ISO, 2013). Theoretical analyses on the positional accuracy led to the development of QGIS 3.10 using the Python scripts. This new tool automated the assessment of the positional accuracy of linear features according to the PEC-PCD scales and permitted the spatial visualization of the discrepancies from quartile subdivision. In this way, the user can select the scale and data to evaluate and view the reference data in the graphical interface. The results are presented as an attributes table in the primary screen of QGIS 3.10. This script is available at https://github.com/eliasnaim/AcuraciaPosicional_PEC-PCD.

As stated previously, VGI heterogeneity is one of the main issues to be resolved. It is associated with the number of features represented, different ways to obtain data, and the time of the edits. In Brazil, a preliminary study by Elias et al. (2018) assessed the positional accuracy in different sample sets of point features and revealed a scale variation from 1:20,000 to 1:30,000 in the OSM database. Based on this evidence, the positional accuracy was analyzed using two sample sets in the region of the checks.

Therefore, more than one sample should be used to evaluate positional quality more accurately in a particular region. Santos (2015) found that 90% of the tested lines had an average discrepancy less than or equal to the PEC value (class and scale used to generate the buffer distance), whereas the root mean square of the average discrepancy was less than the PEC-PCD standard error value. If both conditions are satisfied, the cartographic product is classified according to its class and scale. Figure 2 shows an example of the application of the double buffer method.
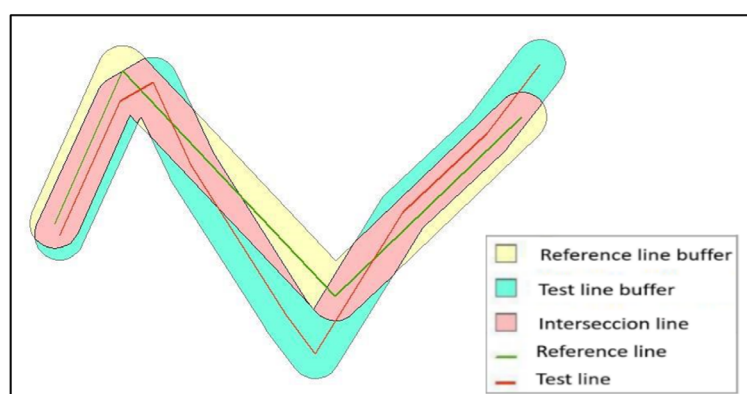
Search: Cruz and Santos (2016)

Figure 2: Double buffer method

Second, we evaluated the completeness of road axes using the method by Haklay (2010), where the dimensions of the road axes were compared to the reference data, and the study area was subdivided into the grids with a spatial resolution of 1 x 1 km. The study by Haklay (2010) pioneered the use of grids to assess the quality of the OSM data, and more studies followed Brovelli et al. (2016), Zhang and Malczewski (2018), Brovelli and Zamboni (2018) and Martini, Kuper and Breunig (2019). Herein, we also subdivided the study area and linear features corresponding to the road axes in OSM and SICAD into the grids with a spatial resolution of 1 x 1 km. The evaluation of the magnitude of discrepancies verified the completeness in the lengths between these two databases in relation to each grid.

We performed the completeness analysis by filtering the vector files missing in the reference data and present on the platform (e.g., bicycle paths, pedestrian paths, bus lanes, tracks, race tracks). Thus, we obtained the results that resembled the related attributes.

Third, we evaluated the thematic accuracy by confirming the tags of the road axes in OSM. The verifications were completed using the name and hierarchy of the road axes (Figure 3).

Figure 3: Compatibility between SICAD and OSM

Other attributes of the road axes in OSM, such as the number of traffic lanes, types of tracks (e.g., single, double, triple), and maximum speeds were not evaluated because they lacked in the SICAD database, making the comparative analysis impossible.

The sampling method used to select the evaluated items for the thematic accuracy corresponded to the parameters established by ISO 2859-1 (ISO 1999) and ET-CQDG (DSG 2015). Although the sampling procedure outlined in ISO 2859-1 (ISO 1999) foresees the sample rejection to a number of incorrect elements, these elements were computed to verify the types of inconsistencies inherent to routes vector files in OSM.

To verify the hierarchy of routes, we harmonized the categories assigned to OSM (Figure 3) and SICAD following the methodology by Zhang and Malczewski (2018). They assessed the quality of cartography of road axes of Canada in OSM and adapted the official classification of Canadian routes of Digital Mapping Technologies Inc. to OSM. Herein, we aligned the hierarchy of road axes in OSM and SICAD based on the classification provided in Chapter III of the Brazilian Traffic Code, Law 9503/1997. Table 2 shows the compatibility of the OSM and Brazilian CTB road hierarchy classifications.

Table 2: Road hierarchy of OSM and CTB

Road hierarchy	
Brazilian Traffic Code (CTB)	OpenStreetMap (OSM)
Fast Transit Track	Motorway, Trunk
Arterial route (I and II)	Primary, primary link secondary, secondary link
Collector	Tertiary, tertiary link
Local	Residential, Service, Living Street, Footway, Track
--	Unclassified, unknown

Source: Adapted from chapter III of the Brazilian Traffic Code (CTB)

We evaluated the thematic accuracy relative to the quantity (percentage) of correct items in the OSM and SICAD verified road hierarchies. Bicycle paths, pedestrian paths, bus lanes, and race tracks in OSM were left unharmonized and untested in the sample space being incompatible with the road axes in SICAD. We also verified that the names of road axes were compatible between OSM and SICAD. Being mandatory for filling by a collaborating user, all attributes of the road hierarchy in OSM were present. The filling of the names in OMS is optional. As such, empty fields were observed for some categories. It allowed the evaluation of the completeness of this attribute of road axes in said sample space.

The completeness and thematic accuracy analyses were performed for Salvador (BA) using the subdivision of prefectures and neighborhoods. Figure 4 presents the flowchart of the methodological procedures used.
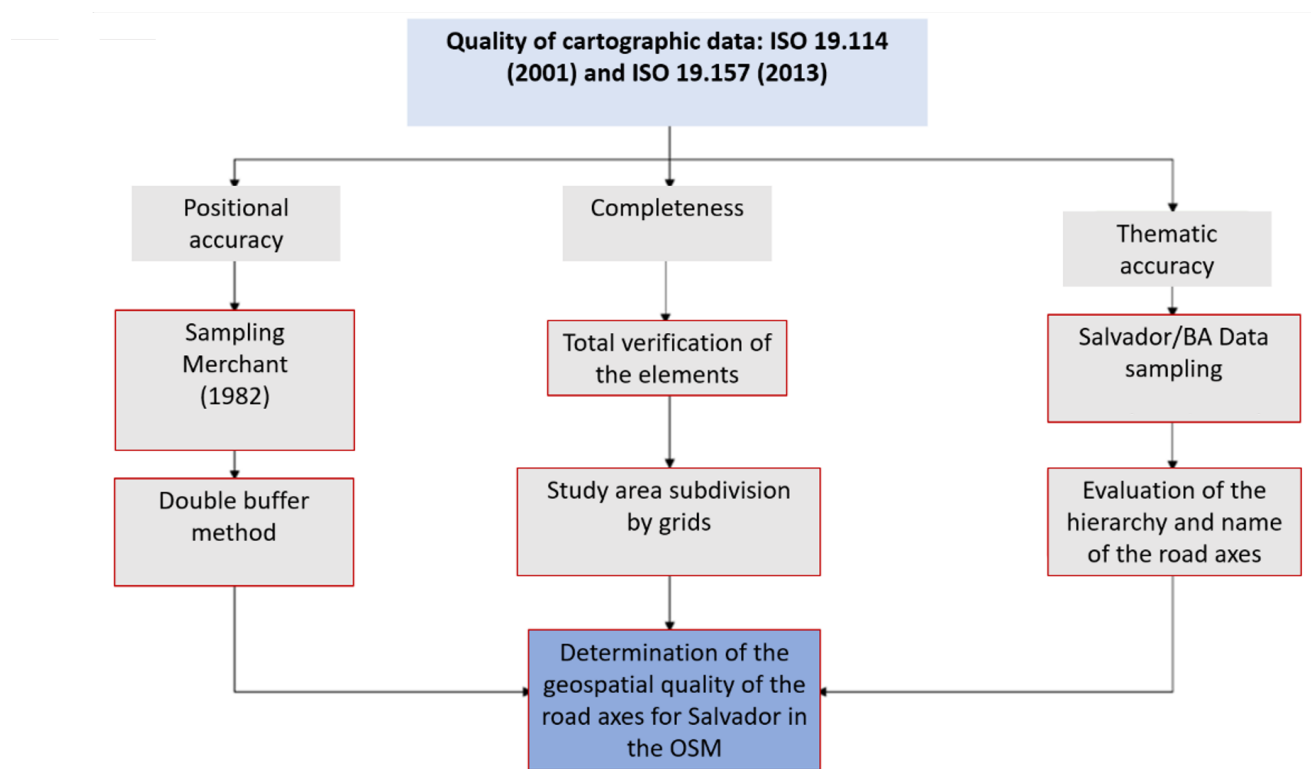
Figure 4: Flowchart of methodological procedures.

3. Results

3.1 Positional Accuracy

To automatically evaluate the positional accuracy, we used QGIS 3.10; and its user interface design is shown in Figure 5.
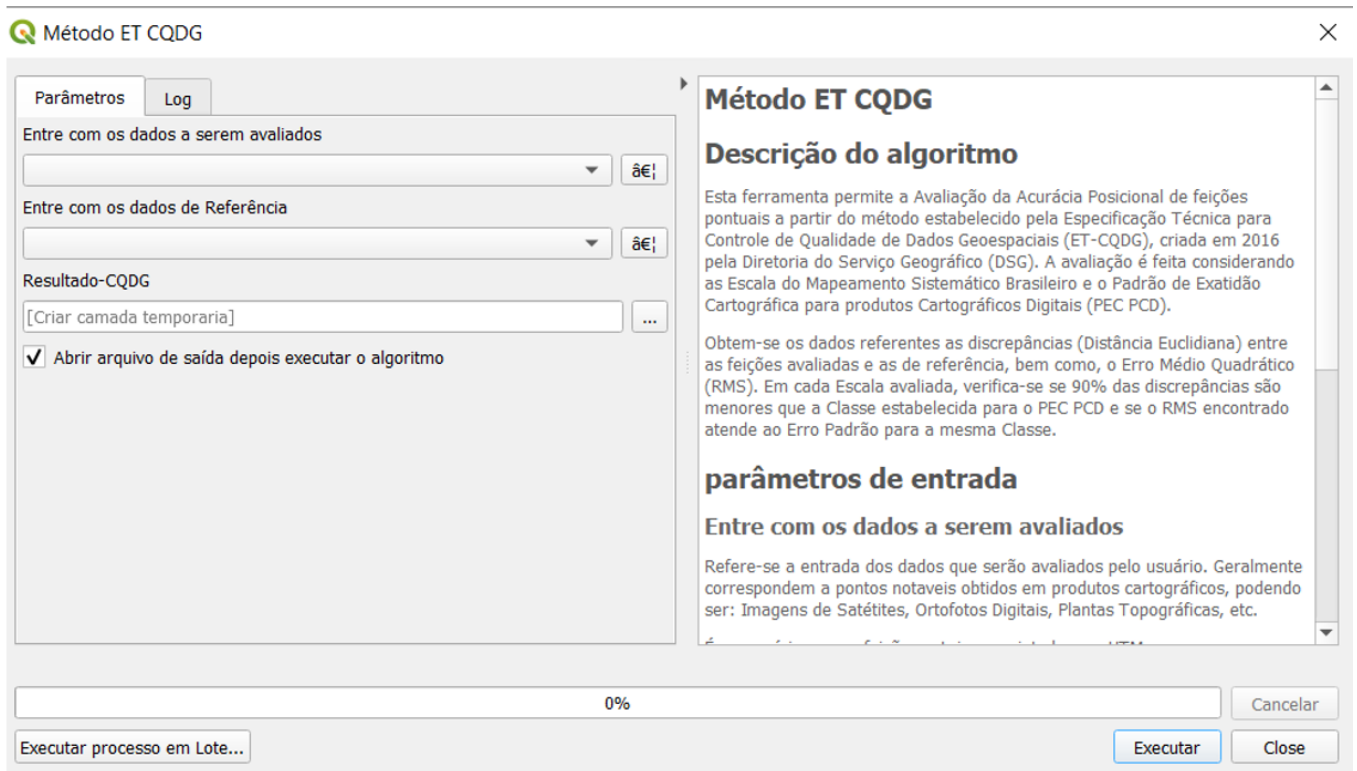

Figure 5: Example of the tool interface developed for QGIS 3.10.

The average area discrepancies and RMS are outlined in Table 3 on scales of 1:22,500 and 1:25,000, respectively. Figure 6 shows the map with the distribution of linear features in the study area.

Table 3: Results of Positional Accuracy obtained for samples 1 and 2

Results	Sample 1	Sample 2
Scale	1:22,500	1:25,000
RMS	3.81	4.2
Standard error (SE)	3.82	4.26
Threshold value	6.3	7
Buffer (%)	100%	100%
Average deviation (m)	1.77	1.35
Maximum deviation (m)	5.99	6.5

Distribution of samples 1 and 2 of the linear features

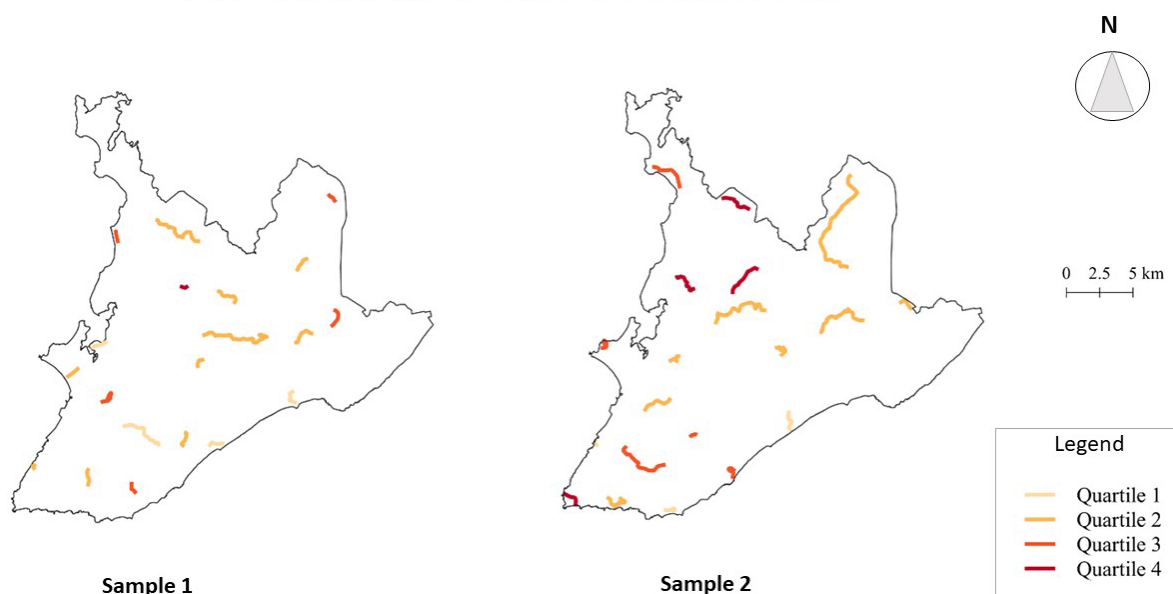Sample 1   Sample 2

Figure 6: Distribution of linear features in the study area.

Using the calculated average discrepancies, we validated the analysis on the scales applying the established verifications for 90% of the obtained values. The variations in average discrepancies were related to the threshold values in the buffer, where the evaluated scale was characterized. Different results were obtained for the two analyzed samples for class A.

To scrutinize the discrepancies, we analyzed the arrangement of road axes in OSM and the Salvador geodatabase. The heterogeneity of the discrepancies may be related to the period in which the editions were initiated or to the BING image characteristics. Figure 7 demonstrates apparent displacements in the linear features. Figure 8 reveals almost identical linear features of OSM and the Salvador geodatabase, providing spatial evidence of their variations. In addition to the features related to the editing period of the element associated with the parameters of the image used as support for this, however, these errors can also be attributed to the level of detail, with which the linear features were edited in OSM.

Figure 7: Linear Example 1 OSM linear features evaluated in relation to the Salvador Geodatabase.

Figure 8: Linear Example 2 OSM linear features evaluated in relation to the Salvador Geodatabase.

3.2 Completeness

We evaluated the physical completeness of vector files by characterizing linear features and calculating the discrepancies in the length of linear tracks using the 1 × 1 km grid system.

The results of the completeness analysis were superimposed on the vector files of the development regions of Salvador (BA) (i.e., middle, suburb, coastal expansion, and central) The city subdivisions originated from the IBGE census tracts, which generated traffic zones in the source-destination survey (1995), and the Sustainable Mobility Plan of Salvador (BA) (PLANMOB, 2017). These are shown in figure 9.

The length of track axes after the exclusion of these data was determined to be 3,451,706.92 m. A previous evaluation demonstrated the potential to estimate the number of taxpayers in this municipality. The cartographic base was used as a reference with a temporal variation of 12 years in relation to OSM. The discrepancy of 1,003,516.90 m revealed the completeness of features in this area (Table 4).

Table 4: Results obtained for completeness evaluation

Analyzed Feature	Number of grids (units)	Percentage (%)
Number of empty cells	1	0.31
SICAD cartographic base with more features than OSM	55	17.14
SICAD cartographic base with fewer features than OSM	265	82.55

The obtained results showed the growth in the representation of road axes in Salvador (BA). Considering the discrepancies and temporal variations in the SICAD database in the study area, the OSM vector files and the completeness methodology are deemed to have the potential to verify expansion aspects in certain localities, especially in the regions with significant changes in a certain period.

The *Miolo* region presented greater discrepancies compared to the other regions (Figure 9), indicating urban growth associated with the construction of a new subway line to the municipality and the expansion of traffic routes.
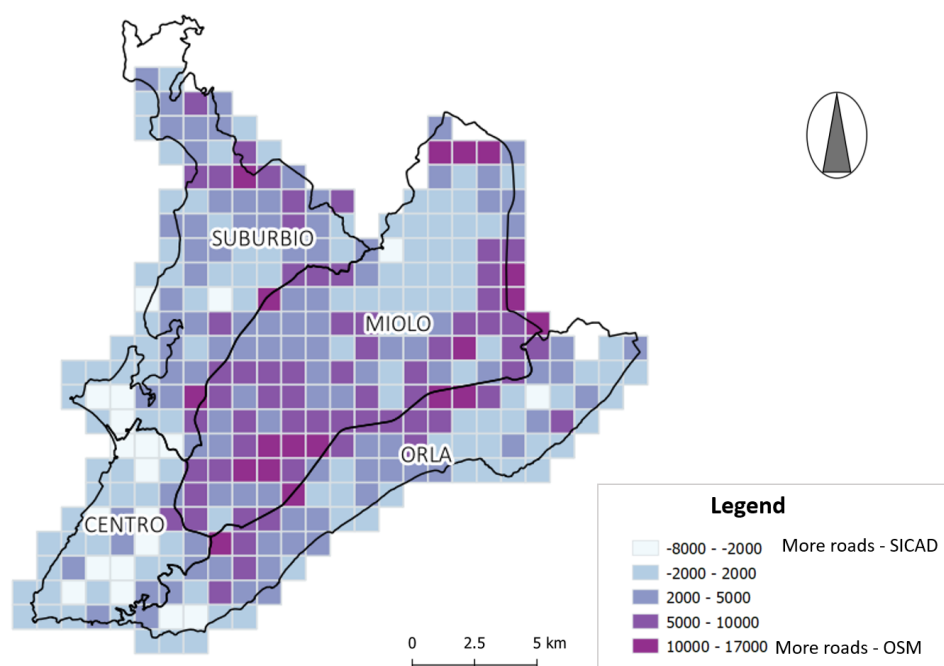
Completness of road axes from grid mesh

Figure 9: Evaluation of the completeness of the road axes from the grid mesh.

Despite the outdated mapping of Salvador (BA), the authoritative-type mapping (2006) allowed the identification of different dynamics associated with the contributions, mainly in the completeness assessment. The temporal variation of OSM in relation to the geospatial data of SICAD permitted the recognition of smaller domains within the study area with higher contributions in recent years. Therefore, we were able to compare the initiatives for expansion, growth, and construction.

3.3 Thematic Accuracy

The population sample and the number of evaluated elements were calculated according to the ET-CQDG (DSG 2015). Although this specification presents an acceptable limit of errors for a given sample, it was neglected herein. First, we checked the number of items to be evaluated from the sample population. The total number of edited routes in OSM reached 19,082 elements, and the total number of evaluated elements amounted to 315. Table 5 shows the verification of the thematic accuracy of the linear features classification (type) in OSM, using vector files in SICAD for reference. The sample consisted of approximately 32 features (points) in each polygon. Table 5 confirms the high percentage of correctly classified road axes in OSM, where the correct values ranged from 65% (*Pau da Lima*) to 90% (*Cajazeiras*). Out of the 315 items checked, 242 elements were classified correctly (77%).

Table 5. Thematic Accuracy in the classification according to the hierarchy of the road axes in Salvador-BA.

Thematic Accuracy (Classification of Road Axes)	
City Hall - Neighborhood	Percentage (correct)
Barra-Pituba	81%
Centro-Brotas	72%
Cabula-Tancredo Neves	81%
Cajazeiras	90%
Itapuã-Ipitanga	69%
Cidade Baixa	84%
Pau da Lima	65%
Liberdade-São Caetano	74%
Valéria	74%
Subúrbio	78%
Total	77%

Considering a 90% confidence level for a population of 19,082 and a sample of 315 elements, the differences greater than the sampling error (4.61%) were regarded representative. Correctness it is around 77% by the thematic accuracy. As such, OSM is a reliable tool that can be used in the future. It is possible indicate high variability due 25% correctness amplitude in the neighborhood comparison greater than 4.61% sampling error.

As mentioned earlier, the insertion of the classification (type) in the OSM road axes by a contributing user is mandatory, making such high confidence levels possible.

Figure 10 displays the spatialization of the percentage of correctly sorted items from each subdivision (prefecture and neighborhood).
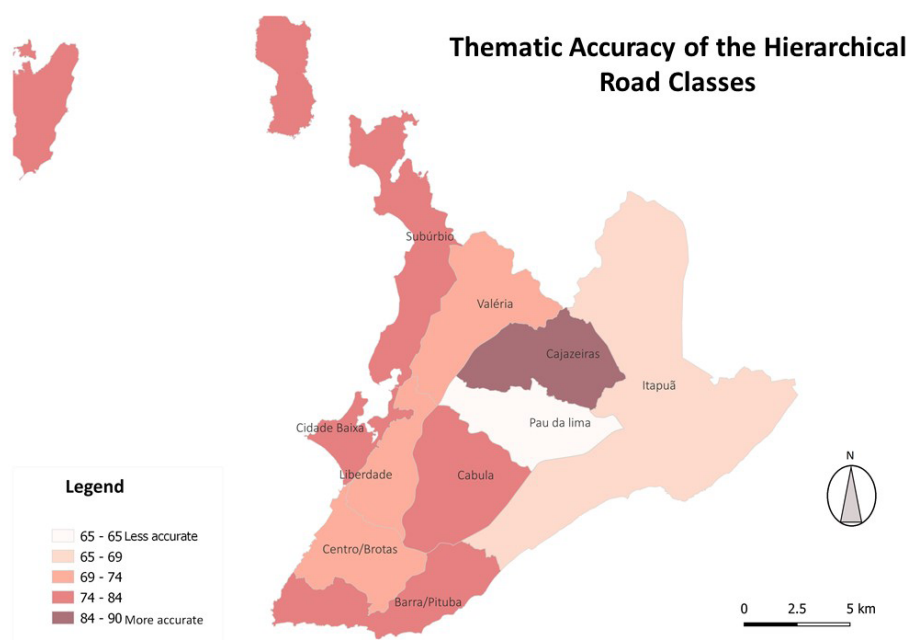

Figure 10: Thematic Accuracy in the classification according to the hierarchy of the road axes in Salvador-BA.

Table 6. Results obtained in verifying the road axes name of the OSM in relation to the Salvador Geodatabase.

City Hall -Neighborhood	Thematic accuracy (Names of road axis)	
	Percentage (Empty Items)	Percentage (correct)
Barra-Pituba	13%	84%
Centro-Brotas	22%	69%
Cabula-Tancredo Neves	47%	44%
Cajazeiras	61%	32%
Itapuã-Ipitanga	44%	47%
Cidade Baixa	71%	23%
Pau da Lima	77%	23%
Liberdade-São Caetano	77%	16%
Valéria	84%	16%
Subúrbio	84%	13%
Total	58%	37%

Out of the 315 evaluated linear features, 116 (37%) were classified as correct in relation to the Salvador geodatabase (Table 6), 17 linear features had incorrect names, and 182 features were nameless. The pathways with similar spellings of the same name were treated as correct. The main sources of errors for the road axes names in OSM were related to the items without information. The absence of names in the evaluated features was encountered in 58% of the analyzed samples. The analysis performed by prefecture and neighborhood suggested certain variability in the obtained values, which ranged from 13% (Subúrbio) to 84% (Barra-Pituba). For example, the prefectures of *Barra-Pituba* and *Centro-Brotas* had the highest percentages of correct names equaling 84% and 69%, respectively. In contrast, the prefectures and neighborhoods of *Liberdade-São Caetano, Valéria,* and *Subúrbio* showed the lowest percentages of correct names with 16%, 16%, and 13%, respectively. Regarding the incorrect names, we made queries to the vector files to investigate the origin of the errors. We verified that only one feature located in the prefecture and neighborhood of *Liberdade-São Caetano*, was incorrectly named. In the other 16 road axes with incorrect names, the errors were related to their topological consistency. In general, the majority of the road axes in OSM were named correctly.

Certain breaks in the vector files of the reference axes in the Salvador geodatabase characterized changes in their names and other attributes and differed from the homologous OSM features (Figure 10). This indicated the discrepancies related to the logical consistency of linear OSM features. The inconsistencies presented in Figure 11 demonstrate the flaws in the topological structure of reference axes. This error was verified for 16 linear features out of the 17 incorrectly named ones.
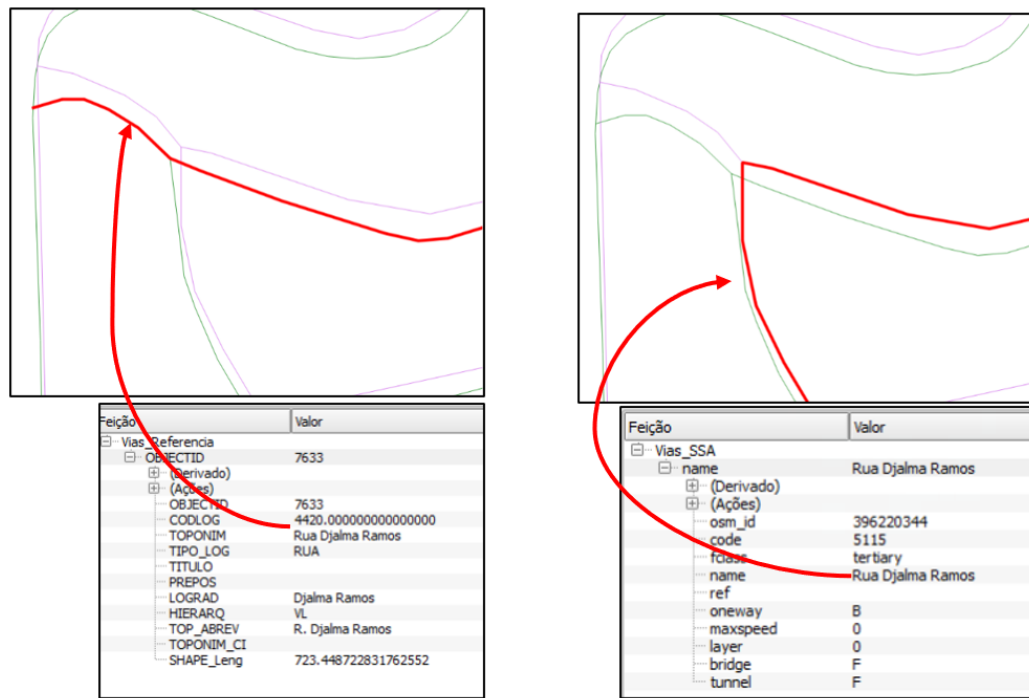
Figure 11: Topological road axis at SICAD and OSM base.

As a result, the primary source of errors associated with the OSM road axes names was related to the absence of this element. In the prefecture and neighborhoods of *Liberdade-São Caetano, Valéria*, and *Subúrbio*, road axes with absent names reached 77%, 84%, and 84%, respectively.

4. Conclusion

Through multiple analyses performed in the study area, we examined the viability of using the OSM road axes features for mapping purposes by checking the veracity of geospatial data compared to authoritative mapping. This characteristic was observed in the completeness assessment, where the regions with the largest contributions in the past 12 years represented urban growth areas.

By evaluating the positional accuracy at a scale of 1:22,500, we were able to observe the limitations related to the precise identification of pathway locations and traces in the vertices. The layout of the OSM roads revealed a heterogeneous pattern, likely due to the period in which the editing was performed, the level of detail applied by the users, and the image quality.

We also evaluated the representativeness and found that 80% of the samples were representative of the dataset.

We expect that OSM will be adopted as the future VGI data standard, considering its potential as associative support in the studies related to urban growth in certain regions with outdated project areas. However, more studies are needed to analyze the data heterogeneity.

Some European countries have been incorporating official cartographic data into their spatial data infrastructure. Despite OSM being a universal tool for spatial data infrastructure, in Brazil, this practice has not been incorporated yet.

As for future work recommendations, we suggest automating the method, establishing a conceptual framework for the OSM data quality assessment, and linking the analyses to different graphical primitives.

ACKNOWLEDGEMENT

AUTHOR´S CONTRIBUTION

Elias Nasr Naim Elias and Vivian de Oliveira Fernandes conceived of the presented idea. Elias Nasr Naim Elias developed the study, experiments and writing. All the authors contributed with preparation of draft manuscript, review of the proposed method, discussion of the results, data analysis, paper revision and refinement

REFERENCES

Brasil. 1984. Decreto nº 89.817, de 20 de junho de 1984. Normas Técnicas da Cartografia Nacional. Brasil.

Brasil. 1997. Lei nº 9.503, de 23 de setembro de 1997. Código de Trânsito Brasileiro. Brasil.

Bravo, J. V. M. Identificação e caracterização de tarefas de uso e geração de geoinformação no mapeamento colaborativo. 2017. PhD. Universidade Federal do Paraná.

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.

Brovelli, M. A. et al. 2019. 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.

Brovelli, M. A. Minghini, M. and Molinari, E. 2016. Database-supported change analysis and quality evaluation of OpenStreetMap Data. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXIII ISPRS CONGRESS. 919-925 July 2016, Prague: Czech Republic.

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.

Chilton, S. 2011. OS and OpenStreetMap. Sheetlines, 91, pp. 20-27.

Cormode, G. and Krishnamurthy, B. 2008. Key Differences between web 1.0 and web 2.0. First Monday, 13(6).

Cruz, D. T and Santos, A.F.P. 2016. Controle de qualidade posicional do sistema rodoviário do Openstreetmap na região central De Viçosa-MG. In: VI Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, VI SIMGEO. August 2016. Pernambuco: Brasil.

Diretoria de Serviço Geográfico (DSG), 2011. Especificação Técnicas para a Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV). Brasil.

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

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

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

Elias, E. N. N., Jesus, E. G. V. and Fernandes, V. O. 2018. Avaliação da dispersão e heterogeneidade dos dados colaborativos do Openstreetmap. In: VII Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, VII SIMGEO. 154-163 October 2018. Pernambuco: Brasil.

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

Ferster C. et al. 2019 Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. International Journal of Sustainable Transportation, 14(1), pp. 64-73.

Fonte, C. C. et al. 2017. Assessing VGI Data Quality. In: G. Foody et al., eds. Mapping and the Citizen Sensor. London: Ubiquity Press, pp.137–163.

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

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

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.

Ibrahim, M. H. Darwish, N. R. and 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 19113, 2002. Geographic information - Quality principles. International Organization for Standarization.

ISO 19114, 2003. Geographic Information - Quality Evaluation Procedures. International Organization for Standarization.

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

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

Jasim, S. and 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.

Johnson, P. A. and Sieber, R. E. 2012. Motivations driving government adoption of the Geoweb. GeoJournal, 77(5), pp. 667-680.

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

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

Klein, I. et al. 2017. Rede de referência municipal para estações livres: uma proposta de baixo custo e grande abrangência. Revista Brasileira de Cartografia, 69(3), pp. 519-532.

Küçük, K. and Anbaroğlu, B. 2019. OpenStreetMap Binalarının Mekansal Doğruluğunun Analiz Edilmesi. Türkiye Coğrafi Bilgi Sistemleri Dergisi, 1 (1), pp. 5-13.

Martini, A. Kuper, P.V. and Breunig, M. 2019. 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.

Maulia, M. Development of an update procedure for authoritative spatial data by the combination with crowdsourced information. 2018. Master. Technische Universitat Dresden.

Merchant, D. C. 1982. Spatial accuracy standards for large scale line maps. In: Proceedings of the Technical Congress on Surveying and Mapping, 1, pp. 222-231.

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

Mozas-Calvache, A. T. and 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, 33(11), pp. 2170-2187.

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

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

Roberto, A. J. 2013. Extração de Informação Geográfica a partir de Fotografias Aéreas obtidas com VANTs para apoio a um SIG Municipal. Master. Universidade do Porto.

Santos, A. D. P. 2015. Controle de qualidade cartográfica: metodologias para avaliação da acurácia posicional em dados espaciais. PhD, Universidade Federal de Viçosa.

Secretaria Municipal De Mobilidade De Salvador (SEMOB), 2018. Relatório Técnico RT14: Plano de Mobilidade Urbana Sustentável de Salvador. [pdf] Salvador: SEMOB. Available at: <http://www.mobilidade.salvador.ba.gov.br/documentos/RT_14-PlanMob_SSA TOMO_I.pdf.> [Accessed 9 July 2020].

Sluter, C. R. et al. 2019. Proposal for Topographic Map Symbols for Large-Scale Maps of Urban Areas in Brazil. The Cartographic Journal, 55, pp. 362-377.

Tveite, H. and Langaas, S. 1999. An accuracy assessment method for geographical line data sets based on buffering. International journal of geographical information science, 13(1), pp. 27-47.

Yang, C. et al. 2017. Utilizing cloud computing to address big geospatial data challenges. Computers, Environment and Urban Systems, 61, pp. 120-128.

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

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, 32(2), pp. 257-281.