COMPARISON BETWEEN THE DOUBLE BUFFER METHOD AND THE EQUIVALENT RECTANGLE METHOD FOR THE QUANTIFICATION OF DISCREPANCIES BETWEEN LINEAR FEATURES

Comparação entre o método do buffer duplo e o método do retângulo equivalente para a quantificação de discrepâncias entre feições lineares

Leandro Luiz Silva de França[1] - ORCID: 0000-0003-0863-1926

Luiz Felipe Coutinho Ferreira da Silva[2] - ORCID: 0000-0003-1744-0515

[1]Diretoria de Serviço Geográfico, 3º Centro de Geoinformação, Olinda - PE, Brasil.
E-mail: franca.leandro@eb.mil.br

[2]Instituto Militar de Engenharia, Engenharia Cartográfica, Rio de Janeiro - RJ, Brasil.
E-mail: felipe@ime.eb.br

Abstract:

Currently, in Brazil, for the assessment of the Positional Accuracy of non-point features (lines and polygons), there is no standard norm of execution. This work aims to compare the results of two methodologies that allow determining the average value of the discrepancies between linear features. The first, Equivalent Rectangle Method, aims to determine the discrepancy by considering an equivalent rectangle for the polygon obtained from the two homologous lines. The second, Double Buffer Method applies a buffer on both lines and obtains the average discrepancy value based on the relation of the areas of the generated polygons. These methods were compared in two steps. Initially, an experiment was performed with features of known measurements, where the displacement of the homologous lines was controlled in azimuth and distance. In this step, it was verified that the shape of the feature and the direction of the displacement interfere in the results of both methods when compared to the traditional procedure of measurement of discrepancies by homologous points. In the second stage, we evaluated the vector data of the OpenStreetMap (class of roads), with reference to a more accurate vector dataset produced for the Mapping of the State of Bahia. As a result, for the 1:25,000, 1:50,000, 1:100,000 and 1:250,000 scales, it was obtained, respectively, the PEC-PCD for the Equivalent Rectangle Method "C", "B", "A" and "A" and the PEC-PCD for the Double Buffer Method "R", "C", "B" and "A", where "R" means that it has not achieved the minimum PEC-PCD classification.

Keywords: Equivalent Rectangle Method; Double Buffer Method; Positional Accuracy; OSM.

Resumo:

Atualmente, no Brasil, para a avaliação da Acurácia Posicional de feições não pontuais (linhas e polígonos), não existe uma norma padrão de execução. Este trabalho tem como objetivo comparar os resultados de dois métodos que permitem determinar o valor médio das discrepâncias entre as feições lineares. O primeiro, Método do Retângulo Equivalente, visa determinar a discrepância considerando um retângulo equivalente para o polígono obtido a partir de duas linhas homólogas. O segundo, Método do Buffer Duplo, aplica um buffer nas duas linhas e obtém o valor de discrepância média baseado na relação das áreas dos polígonos gerados. A comparação dos métodos foi realizada em duas etapas. Inicialmente, foi realizada uma experimentação com feições de medidas conhecidas, onde o deslocamento das linhas homólogas foi controlado em azimute e distância. Nessa etapa verificou-se que a forma da feição e a direção do deslocamento interferem nos resultados de ambos os métodos se comparado ao procedimento tradicional de medição de discrepâncias por pontos homólogos. Na segunda etapa, foram avaliados os dados vetoriais do *OpenStreetMap* (classe de rodovias), tendo por referência um conjunto de dados vetoriais (classe trecho rodoviário e arruamento) produzido para o Mapeamento do Estado da Bahia. Como resultado, para as escalas 1:25.000, 1:50.000, 1:100.000 e 1:250.000, foram obtidos, respectivamente, pelo Método do Retângulo Equivalente PEC-PCD "C", "B", "A" e "A" e pelo Método do Buffer Duplo PEC-PCD "R", "C", "B" e "A", onde "R" significa que não obteve a classificação mínima do PEC-PCD.

Palavras-chave: Método do Retângulo Equivalente; Método do Buffer Duplo; Acurácia Posicional; OSM.

1. Introduction

In the last years, the cartography has received the voluntary and non-commercial contribution of several institutions for global mapping, a phenomenon known as Volunteered Geographic Information (VGI) (Goodchild, 2007). In this context, it stands out the OpenStreetMap (OSM) (Sehra et al., 2014). Concomitantly, new discussions about the quality of these data have increased, considering the following quality elements: completeness, logical consistency, positional accuracy, temporal accuracy and thematic accuracy (Ming et al., 2013).

For paper topographic charts (known as analogical products) produced officially in Brazil for systematic mapping, the positional accuracy follows the Cartographic Accuracy Standard (*Padrão de Exatidão Catográfica* - PEC) which parameter values were established in Decree No. 89.817 of June 20, 1984 (Brazil, 1984), transcribed through the Table 1.

Table 1: Mean Error (*Erro Médio* - EM) and Standard Error (*Erro-Padrão* - EP) values for PEC classification of printed products (topographic map).

Type	PEC	1:25,000		1:50,000		1:100,000		1:250,000	
		EM	EP	EM	EP	EM	EP	EM	EP
Planimetric measurements	A	12,5	7,5	25	15	50	30	125	75
	B	20	12,5	40	25	80	50	200	125
	C	25	15	50	30	100	60	250	150

Source: DSG, 2016a.

With the recent evolution of digital cartography and the geotechnology's resources, new products require other quality parameters, including a refinement of the positional accuracy assessment (Ariza-López et. al, 2007) and computational programs to perform it (Nero et. al, 2017).

The Brazilian Geographic Service Bureau (*Diretoria de Serviço Geográfico* - DSG, 2016) established the quality assessment parameters described in the Technical Specifications for Geospatial Set Products (ET-PCDG) and Geospatial Data Quality Control (ET-CQDG).

The new positional accuracy standards, known as the Cartographic Accuracy Standard for Digital Cartographic Products (PEC-PCD), consider small and large scales (Table 2 and Table 3). It should be noted that the "B", "C" and "D" from PEC-PCD correspond respectively to "A", "B" and "C" classes of the Decree No. 89.817 (DSG, 2015) and a new class A was created.

Table 2: Mean Error (*Erro Médio* - EM) and Standard Error (*Erro-Padrão* - EP) values for the classification of PEC-PCD for Vector Geospatial Data Set (*Conjunto de Dados Geoespaciais Vetoriais* - CDGV) on small scales.

Type	PEC-PCD	1:25,000		1:50,000		1:100,000		1:250,000	
		EM	EP	EM	EP	EM	EP	EM	EP
Planimetric measurements	A	7,00	4,25	14,00	8,50	28,00	17,00	70,00	42,50
	B	12,50	7,50	25,00	15,00	50,00	30,00	125,00	75,00
	C	20,00	12,50	40,00	25,00	80,00	50,00	200,00	125,00
	D	25,00	15,00	50,00	30,00	100,00	60,00	250,00	150,00

Source: DSG, 2016a.

Table 3: Mean Error (*Erro Médio* - EM) and Standard Error (*Erro-Padrão* - EP) values for the classification of PEC-PCD for Vector Geospatial Data Set (*Conjunto de Dados Geoespaciais Vetoriais* - CDGV) on large scales.

Type	PEC-PCD	1:1,000		1:2,000		1:5,000		1:10,000	
		EM	EP	EM	EP	EM	EP	EM	EP
Planimetric measurements	A	0,28	0,17	0,56	0,34	1,40	0,85	2,80	1,70
	B	0,50	0,30	1,00	0,60	2,50	1,50	5,00	3,00
	C	0,80	0,50	1,60	1,00	4,00	2,50	8,00	5,00
	D	1,00	0,60	2,00	1,20	5,00	3,00	10,00	6,00

Source: DSG, 2016a.

## 2. Positional Accuracy of Linear Features

Even in the most recent standards, such as ET-CQDG, the quality assessment procedures regarding positional accuracy leave some uncertainties regarding the performance of the evaluation for line and polygon feature types (Santos et al., 2015). The adopted procedure remains the determination of quantitative percentages in relation to the maximum permissible error as well as the comparison of the mean square error with the standard error between the observations related to a point feature obtained in the field and its cartographic representation (DSG, 2016a).

According to Ferreira and Cintra (1999), the choice of the points to be evaluated is a subjective task that, when conducted by different professionals, can produce disparate results, distorting the real results. One of the reasons is the difficulty in determining many homologous points (Figure 1) and even if points are taken at regular intervals, the results may be different depending on the number of intervals and the initial marking point, illustrated by means of Figure 2.
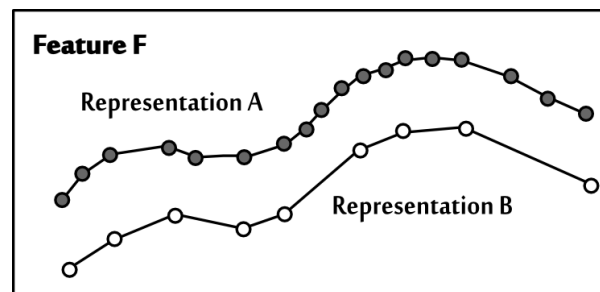
Figure 1: Different representations of the same feature.
Source: Ferreira da Silva and Cintra (1999)
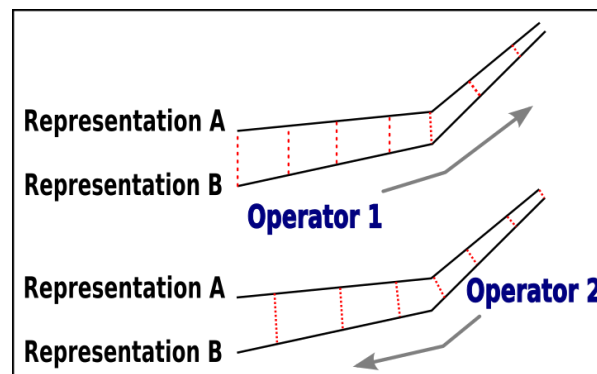
Figure 2: Measurements of discrepancies at regular intervals.
Source: Ferreira da Silva and Cintra (1999)

The calculation of discrepancies through distance measurements between homologous points is the traditional method and is still widely used. Reference is made, for example, to Helbich et al. (2012), which use the junctions (crossings between highways of a German city) to make the statistical evaluation through a two-dimensional regression model, besides studying the pattern of spatial association through the distribution function G (Getis and Ord, 1992).

Similarly, Brovelli et al. (2016) evaluated the positional accuracy of OSM buildings in the city of Milan (Italy) through a semi-automatic method of detecting homologous points in relation to a set of reference data.

To automate the evaluation and standardization, other methods of positional accuracy have been developed. Santos et al. (2015) presented the following methods for evaluating planimetric positional accuracy: Epsilon Band (Area Method), Single Buffer, Double Buffer, Hausdorff Distance and Vertex Influence. Among the studied methods, it was concluded that Double Buffer would be the most recommended because this method makes it possible to verify outliers and trends in the data, besides of reaching results closer to the traditional method, both in the classification of the positional accuracy of the features and in the descriptive statistics of errors.

In Brazil, Fonseca Neto et al. (2017) used the Double Buffer Method to determine the planimetric positional accuracy of linear features in the evaluation of the quality of orthoimages generated from a sensor embedded in a drone platform. Santos et al. (2016) used the same method to evaluate the planimetric positional accuracy of digital surface models (DSM) from linear features extracted from the models (ridges and drainage lines).

Cruz and Santos (2016) also applied the Double Buffer Method to evaluate the OSM road system of the central Brazilian region of Viçosa-MG, obtaining for that situation PEC-PCD class C for the 1: 10,000 scale.

The measurement of the mean discrepancy of each linear feature in relation to its homologous reference feature (which can be measured on field or obtained by more accurate dataset) allows the determination of the root mean square error of the sample and the percentage of the absolute value of discrepancies lower than a specific tolerance and, thus, the PEC-PCD classification.

In this context, this work compares the results of the mean discrepancies and the PEC-PCD classification obtained by the Equivalent Rectangle Method (ERM) with the results calculated using the Double Buffer Method (DBM). The research was performed in two stages:

1) Experimentation with features of distinct figures with controlled azimuth and distance for the displacements between pairs of homologous lines.

2) A case study with the roads from the OSM's vector data in relation to the equivalent from the mapping of the Brazilian State of Bahia's CDGV.

3. Equivalent Rectangle Method (ERM)

This method comes as an alternative to the traditional method of measuring discrepancies from homologous points. It proposes to perform this task by measuring the area between two representations of the same feature, replacing the resulting geometry with an equivalent rectangle with the same area and the same perimeter (Ferreira da Silva and Cintra, 1999).

Figure 3 shows the representations A and B for a Feature F and its respective equivalent rectangle of x_1 and x_2 sides, whose area (S) and perimeter ($2p$) are obtained by Equations 1 and 2, respectively.
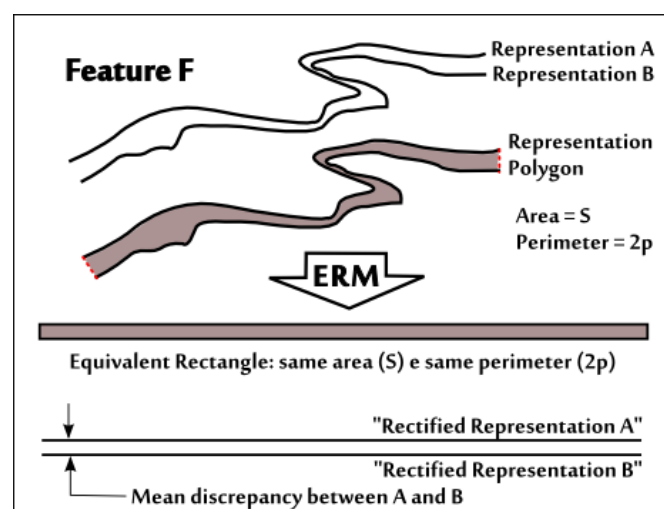

Figure 3: Equivalent Rectangle.
Source: Ferreira da Silva and Cintra (1999)

$$S = x_1 . x_2 \tag{1}$$

$$2p = 2x_1 + 2x_2 \tag{2}$$

From (1) and (2) is obtained the quadratic equation (Equation 3):

$$
\begin{aligned}
x_2 &= p - x_1 \\
S &= x_1 . (p - x_1) \\
x_1^2 &- px_1 + S = 0
\end{aligned}
\tag{3}
$$

The discrepancy value x_1 (smaller side of the rectangle) can be calculated from Equation 3, through the Bhaskara's Formula:

$$x_1 = \frac{p - \sqrt{p^2 - 4S}}{2} \tag{4}$$

According to Ferreira and Cintra (1999), due to the fact that the rectangle is adopted as a geometric model, lines with many salient angles tend to generate underestimated values of x_1, that is, the difference for the real value is proportional to the quantity and size of salient angles in the polygon, which usually does not occur in practice.

Some precautions should be taken to apply the method:

- If there is a cross between the feature's representations, these lines must be broken at the points of intersection and the original polygon must be fractioned into smaller polygons;
- Both representations must be in the same Coordinate Reference System (CRS).

The main advantages of this method are the great speed in obtaining the results, the reduction of bias and the uniformity of the evaluation.

4. Double Buffer Method (DBM)

The DBM has been suggested by Santos et al. (2016), Fonseca Neto et al. (2017) and Cruz and Santos (2016) for the evaluation of the positional accuracy of linear features due to the best results presented in studies developed by Santos et al. (2015).

This method was originally proposed by Tveite and Langaas (1999) under the name Buffer Overlay Statistics. It is based on the application of an uncertainty value in the two lines: Reference Line (RL) and Test Line (TL), that is, a buffer is created for each line. Then, the difference of the polygon generated by the RL in relation to the polygon generated by the TL is determined. The area of this difference allows the calculation of the mean discrepancy between the features, which is necessary for the evaluation of the positional accuracy.

The value of the mean discrepancy md is obtained by Equation 5, where x is the buffer distance, ΣA_{diff} is the sum of the areas of the difference between the polygons, i.e., the areas of the RL buffer that does not intersect the TL buffer and AB_{TL} is the area of the TL's buffer.

$$md_i = \pi . x . \frac{\sum A_{diff}}{AB_{TL}} \qquad (5)$$

Figure 4 illustrates a practical example of the Method for an *x* buffer distance, highlighting the hatched area as the difference of the RL buffer over the TL buffer.

The value *x* of the buffer used by Santos et al. (2015) corresponds to the EM of the PEC-PCD according to the scale and class to be evaluated.
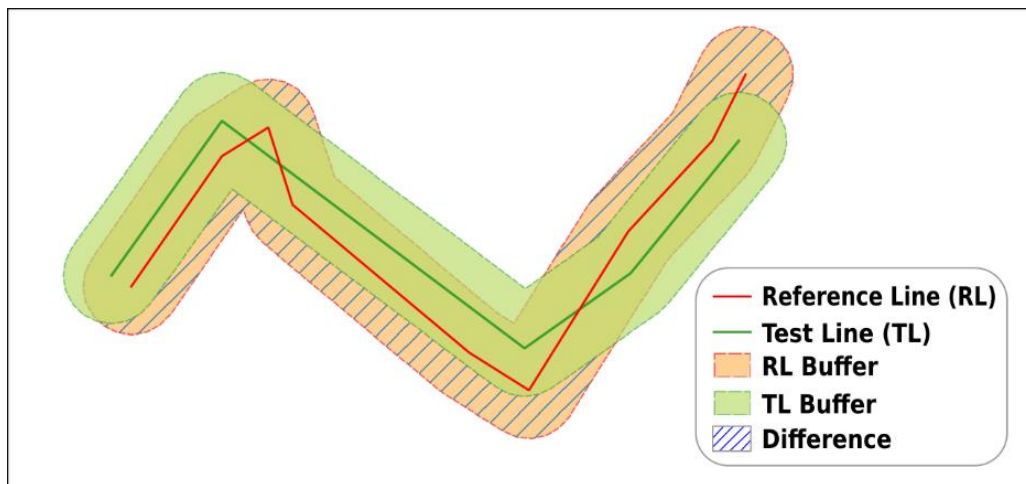
Figure 4: Double Buffer Method
**Source:** Adapted from Cruz and Santos (2016)

5. Materials and Methods

5.1 Controlled Experiment

The objective of this procedure is to investigate the results of the ERM and DBM applied to features with different shapes and different types of displacements, given a certain distance, for homologous pairs of lines.

In this experiment, three features of different configurations were built using the QGIS Advanced Digitizing Tools and the CADDigitize plugin. These features and the respective measurements are represented in Figure 5, where the first is a straight line, the second a 90° inflection line in the middle point and the latter a semi-circumference.

New identical lines were also created with position changed considering a certain azimuth. In all situations the translation distance of 20 meters was considered, that is, in the case of the traditional method, the discrepancy between all its homologous points is exactly 20 meters corresponding to the distance of the translation.
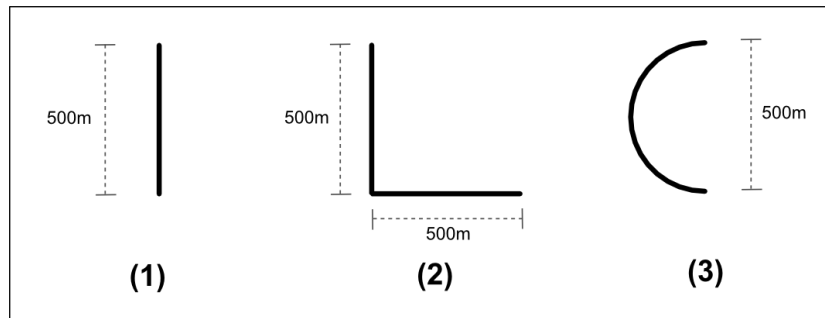
Figure 5: Shape and size of features for the controlled experiment.

In order to obtain different configurations of the pairs of homologous lines and eliminating some symmetrical relative positions, the azimuths shown in Figure 6 were considered, where the test lines are dashed in red and the reference lines are in black color.
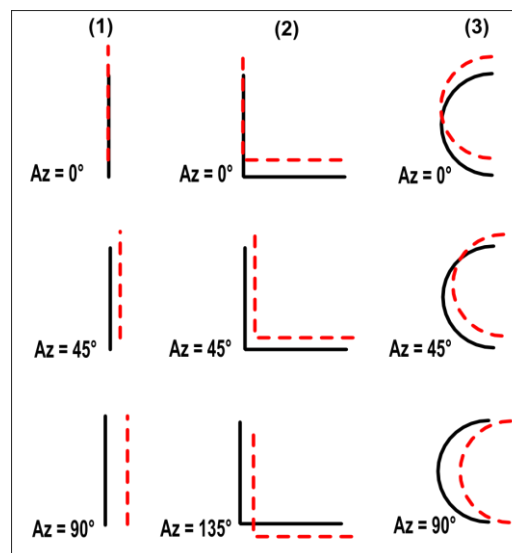
Figure 6: Homologous pairs used in the controlled experiment.

For each of the nine situations presented, the ERM and DBM were applied. For the ERM, when the lines intersect themselves, those lines were segmented at the points of intersection, creating two polygons, and then were calculated the discrepancy based upon the sums of the areas and semi-perimeters of those polygons.

5.2 Case Study

The area of study corresponds to Map Index (MI) 1721-4-NO with nomenclature SC-23-Z-B-V-4-NO referring to the Brazilian Systematic Mapping's frame for the 1:25,000 scale.

The evaluated data consider the linear features of the Roads class of the OSM (Ramm, 2017), corresponding to all types of highways, paths and accesses. This data can be obtained free of charge through the Geofabrik server (2016). In this work, the highways of the OSM were denominated Test Lines (TL).

The data adopted as reference correspond to the classes *Trecho Rodoviário* and *Arruamento,* belonging to the Transportation Systems category of ET-EDGV 2.1.3 (DSG, 2010). These data are available in the Geographic Database of the Brazilian Army (*Banco de Dados Geográfico do Exército* - BDGEx) through the following website: http://www.geoportal.eb.mil.br/mediador/. In this work, the combination of these two classes was performed in a single class called Reference Lines (RL).

The reference features were obtained by heads-up vectorization on orthoimage whose planimetric positional accuracy was evaluated for the Bahia Mapping Project (Penha et al., 2012). The acquisition scale was 1: 2,000 on orthoimage with a spatial resolution of 60 cm.

The work area, as well the TL and RL are shown in Figure 7. The mapped region covers the city of Xique-xique in the state of Bahia.
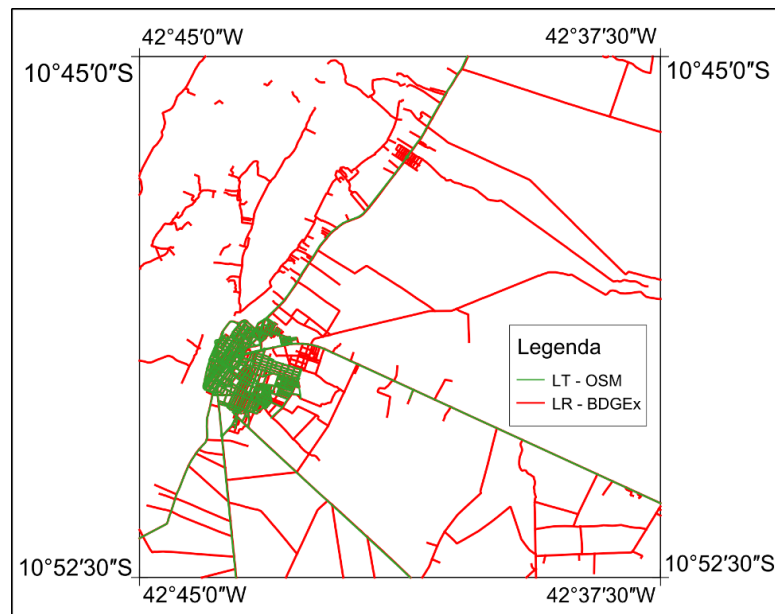

Figure 7: Work area.

For the execution of this work, the open source software QGIS 2.18 and tools created in Python were used. The developed tools are available in the following Github repository: www.github.com/LEOXINGU/linear_pos_accuracy.

Before applying the QGIS tools for the ERM and DBM, it is necessary to perform the geoprocessing that are described in the workflow (Figure 8).

The objective of obtaining the LT and LR with the same Coordinate Reference System (CRS) SIRGAS 2000/UTM 23S is to have all the measurements in metric units, making possible the classification of the PEC-PCD.
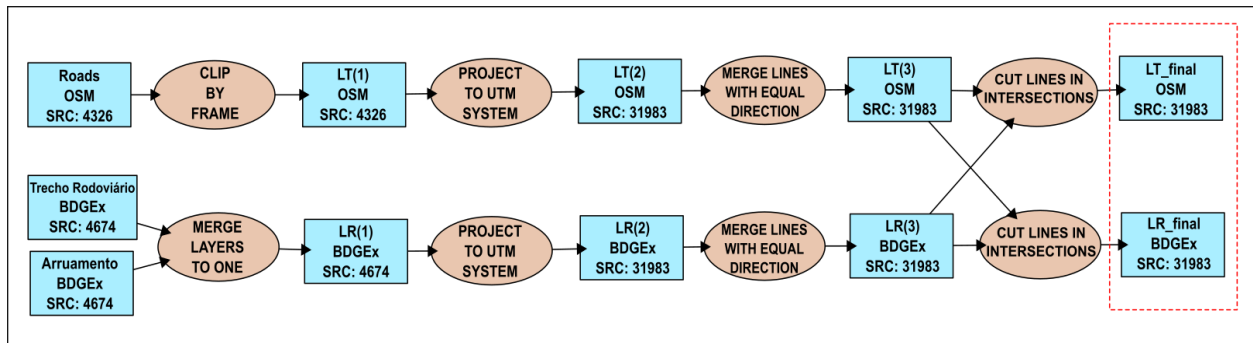
Figure 8: Workflow for data preparation.

The tools that "merge lines with equal direction" and "cut lines in intersections" allow automatic identification of a greater number of homologous features. Besides, the ERM requires intersecting lines be segmented at the intersections.

Figure 9a shows an example of the "merge lines with equal direction" tool application, and Figure 9b illustrates the result of the "cut lines in intersections" tool.
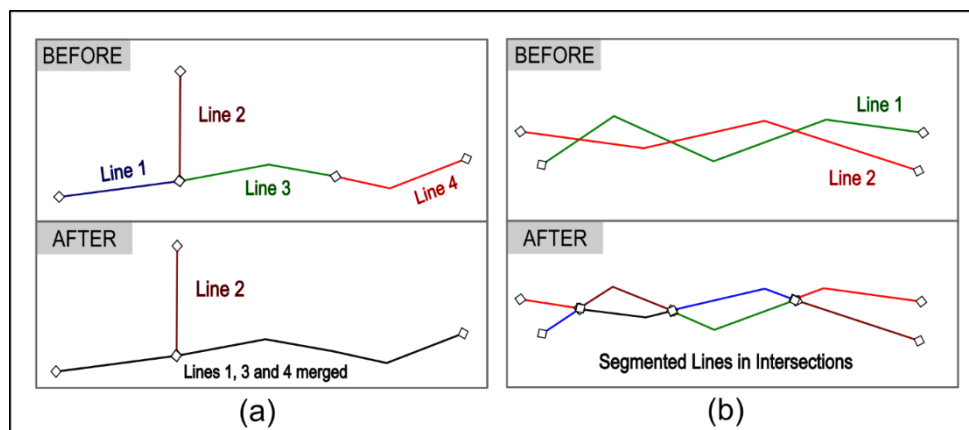

Figure 9: a) merge lines with equal direction; b) cut lines in intersections.

After pre-processing, the data is ready to be evaluated. The evaluation was performed using the following QGIS tools: "*MBD – Método do Buffer Duplo*" and "*MRE – Método dos Retângulos Equivalentes*".

The identification of homologous features is performed automatically following the criteria: "Two lines are homologous when one is completely inside the relationship buffer of the other and vice versa". If there are more than one line within the relationship buffer of the other, then the nearest line is considered.

It is worth noting that this artifice was implemented with the purpose of automatically identifying homologous lines, not being suggested by any of the studied methods. However, it systematically enables the evaluation of several homologous pairs in a set of lines.

In order to classify the PEC-PCD of linear features, it was considered the value of the discrepancies weighted by the respective lengths through an analogy in what ET-CQDG recommends for pairs of homologous points. For that, the two criteria were followed:

1st *criterion*: A sample whose sum of the features' lengths with discrepancies lower than the EM shall fit into a cartographic class (PEC-PCD) when that sum is greater than 90% of the total length. Table 4 presents an example of evaluation applying this criterion.

Table 4: Example of PEC-PCD evaluation considering the EM.

Mean discrepancy	Length (m)
7,9	40
8,7	10
4,1	30
13,6	20

In this case, considering the scale 1:25,000 and class B, the EM is 12.5 meters (Table 2). The sum of lengths where the mean discrepancy is less than the EM corresponds to 80.0 meters, or 80% of the total length. Therefore, this sample is not classified in class B for the 1:25,000 scale.

The length value may refer to the length of the RL or TL, as well as the sum of both. In this work, the length of the RL was used.

2nd *Criterion*: The value of the Weighted Root Mean Square Error (RMSE$_W$) is compared to the EP of the PEC-PCD table. If the value is lower, then the sample belongs to the evaluated class. If it is larger, it walks on the PEC-PCD table until it finds a case where RMSE$_W$ is lower or, if it is not found, the sample is considered "not compliant" for any of the classes, called as "R" class.

The RMSE$_W$ is calculated by Equation 6, where *dm* and *l* correspond, respectively, to the value of the mean discrepancy and length for each feature of the sample.

$$RMSE_W = \sqrt{\frac{\sum_{i=1}^{N} dm_i^2 . l_i}{\sum_{i=1}^{N} l_i}} \tag{6}$$

6. Results and Discussion

6.1 Controlled Experiment's Results

Initially, it was found that the discrepancy values applying the DBM may have different results considering the same pair of homologous features. This is proven in the simplest case, where we have two parallel lines of 500 meters of length separated at 20 meters with 90 ° azimuth (Figure 6), several values of discrepancies were obtained, presented in Table 5.

Similarly to Santos et al. (2015), the EM was used as buffer distance. Table 5 reveals that the discrepancy is influenced by the buffer distance which is different for each scale and class of the PEC-PCD (Tables 2 and 3).

Table 5: Different values (in meters) of discrepancies for the same situation applying the DBM.

Discrepancies of Great Scales					
PEC-PCD	1:1.000	1:2.000	1:5.000	1:10.000	Summary	
A	0,88	1,76	4,40	8,80	Mean	9,12
B	1,57	3,14	7,85	15,71	Std deviation	8,93
C	2,51	5,03	12,57	25,13	Maximum	31,42
D	3,14	6,28	15,71	31,42	Minimum	0,88
Discrepancies of Small Scales					
PEC-PCD	1:25.000	1:50.000	1:100.000	1:250.000	Summary	
A	21,99	31,63	32,06	32,98	Mean	32,15
B	31,56	31,98	32,57	33,92	Std deviation	2,93
C	31,83	32,34	33,17	34,89	Maximum	35,39
D	31,98	32,57	33,52	35,39	Minimum	21,99

For the larger scales a greater variation of the calculated discrepancy was presented, on the other hand, for the small scales, there was a smaller variation. In the latter case, the average obtained was 32.15 meters.

The main reason for the large variation in the calculated discrepancies is the lack of intersection between the polygons when a small buffer value was applied (in the cases where EM is less than 10 meters). As there is no intersection, the discrepancy of the DBM becomes directly proportional to the value *x* of the buffer (Equation 5).

Thus, the DBM makes sense in cases where there is an intersection between the polygons generated from the lines' buffer. In the situation of two parallel lines separated at 20 meters, this only happens for small scales, except for the case of PEC-PCD "A" and 1: 25,000 scale, where the EM is 7.0 meters (Table 2).

Therefore, it can be observed in Table 5 that, for small scales, the DBM overestimates the value of the discrepancies, considering that a displacement of 20 meters was forced and, using this method, reached the value of 32.15 meters for one of the situations and values greater than 20 in most cases (Table 5).

Applying the ERM for the two parallel lines, the exact value of 20 meters of discrepancy was obtained, which is expected value for the situation. In this method, the calculated value is independent of the scale, that is, the discrepancy will be the same for all situations.

The shape of the linear features and the displacement azimuth between the homologous pairs present different results for the different situations presented in Figure 6 for both the ERM and the DBM. Table 6 presents the results of the discrepancies calculated for the nine situations presented in Figure 6.

Table 6: Discrepancies in meters for the 3 features of the controlled experiment.

Feature 1			Feature 2			Feature 3		
Az	ERM	DBM	Az	ERM	DBM	Az	ERM	DBM
0°	0,00	6,27	0°	9,90	17,73	0°	12,82	21,89
45°	13,97	24,9	45°	14,06	23,92	45°	12,58	21,89
90°	20,00	32,15	135°	14,06	22,98	90°	12,63	22,03

In Table 6, the discrepancy value of the DBM corresponds to the mean of the calculated values for small scales, so the discrepancies for large scales were not considered.

Considering the shape of the features, Feature 1 (straight segment) was the one that had greater variation of the values of discrepancy in both methods. On the other hand, Feature 3 (half-circumference) obtained a smaller variation of the discrepancies.

It is evident that in both methods, in situations where a part of the homologous pairs is coincident (Feature 1 and Feature 2 with 0 ° azimuth), the values of the discrepancies were smaller. In the case where the discrepancy was zero, it means that the lines were parallel and, then, there is no area between them.

However, for the 500 meters of the features, only 40 meters was not coincident. In this situation, to assert that the discrepancy is 20 meters (calculated by the traditional method) can bring a distorted idea from the point of view that the features coincided in most of its extensions. Nevertheless, to consider that the discrepancy is null, as is the case of the ERM, is unfear because it disregards that there is a discrepancy.

Discrepancies by DBM generally overestimated the 20 meters discrepancy imposed on feature points. On the other hand, in none of the cases did the ERM extrapolate the amount imposed.

6.2 Case Study's Results

Within the study area, initially, the amount of RL from BDGEx was 1552 features. The number of TL related to OSM was 285. After the pre-processing described in Figure 8, the amounts were changed to 3568 and 3219 for the BDGEx and OSM layers, respectively.

The criterion for identifying pairs of homologous lines was the same for both methods, applying a relationship buffer of 30 meters. With this parameter, the number of pairs of homologous lines automatically identified was 2122.

Figure 10 shows pairs of homologous lines (the result of the DBM tool) and Figure 11 allows visualization of the polygons generated by these lines (the result of the ERM tool).

Both QGIS tools, developed by the authors, generate a shapefile type layer. The ERM tool creates a layer of polygons and its attributes include the average discrepancy for each feature. Similarly, the DBM tool creates a multiline layer for pairs of homologous lines. In this case, their attributes include the discrepancy for each scale situation and category of PEC-PCD, because the buffer value is different for each of these occasions.

Due to the DBM presents different values of discrepancies, the average of the discrepancies calculated for small scales was considered for each homologous pair, allowing a comparison with the ERM.
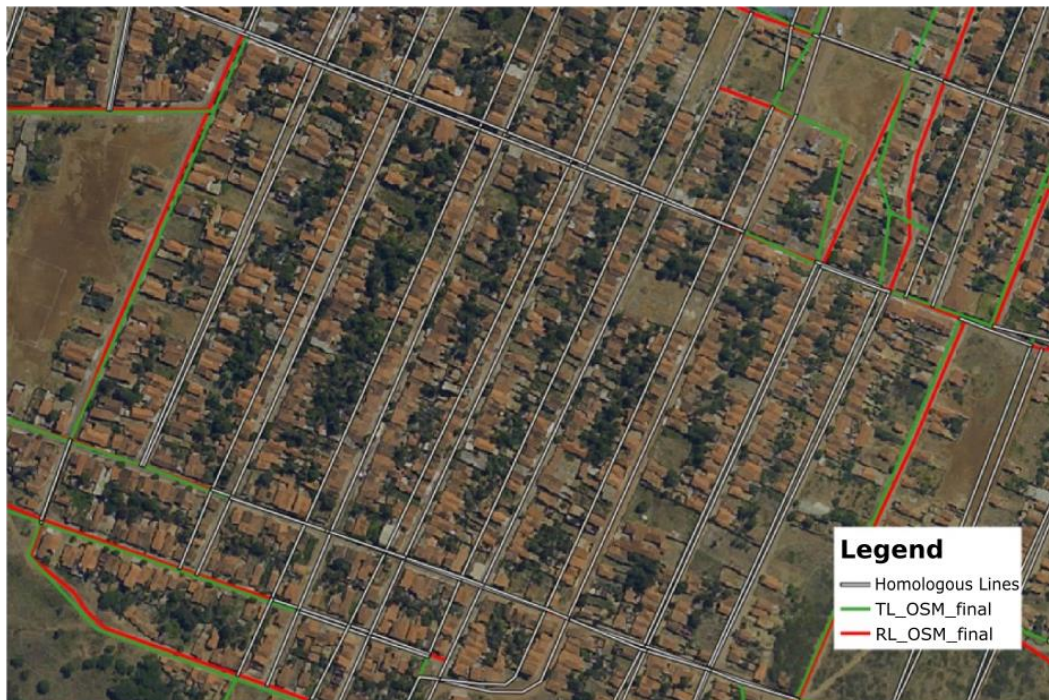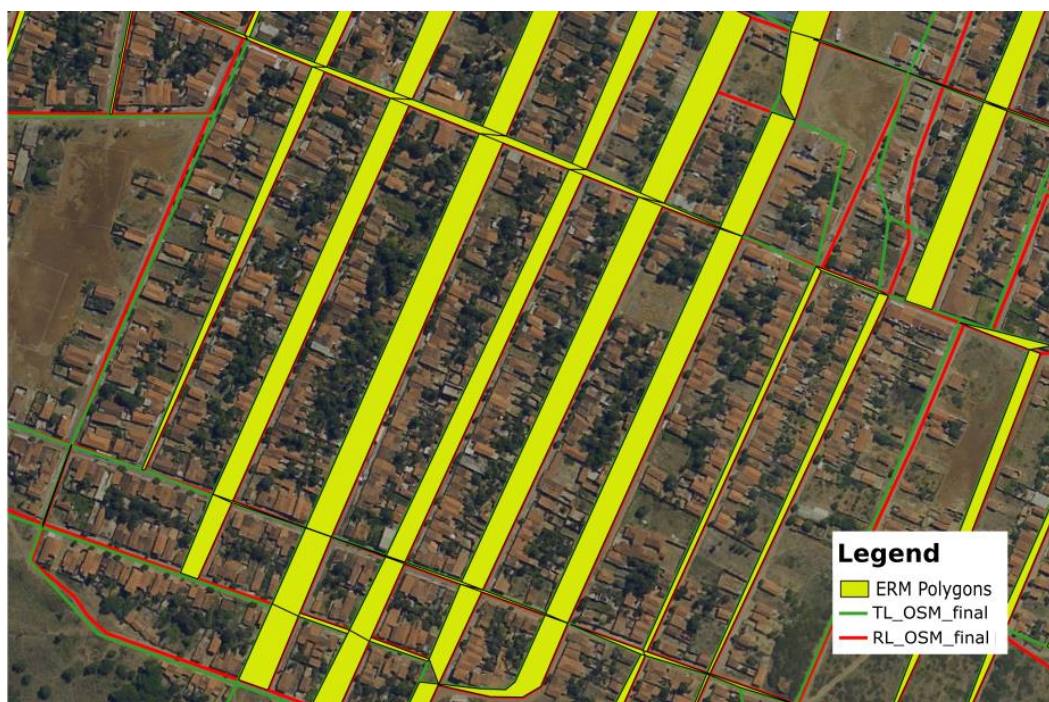
Figure 10: Homologous lines.

Figure 11: Polygons generated from homologous lines.

Table 7 presents the mean, weighted mean, RMSE, weighted RMSE, minimum and maximum of the discrepancies obtained in each method for the total number of homologous lines.

In both methods, the weighted calculus by the feature-length was higher than those calculated without weighting, being this difference more evident for the ERM.

Table 7: Descriptive results of discrepancies calculated for the two methods.

	ERM	DBM
Mean	1,619	8,343
Weighted Mean	4,993	12,138
RMSE	4,122	15,531
Weighted RMSE	7,616	16,982
Minimum	0	0
Maximum	24,727	69,704

In this study case, the pairs of features were obtained through the application of a 30-meter relationship buffer. That is, for any of the pairs, the maximum discrepancy would be 30 meters, which was not verified in DBM, which had discrepancies up to 69.7 meters (Table 7).

Figure 12 corresponds to the scatter diagram, where each point represents the discrepancies obtained by method for each pair of homologous features. The Pearson correlation coefficient calculated was 0.432, which highlights the low correlation between the results. Although, by inspecting the graph, a tendency of greater correlation of the results is observed as the values of discrepancy by ERM are higher.
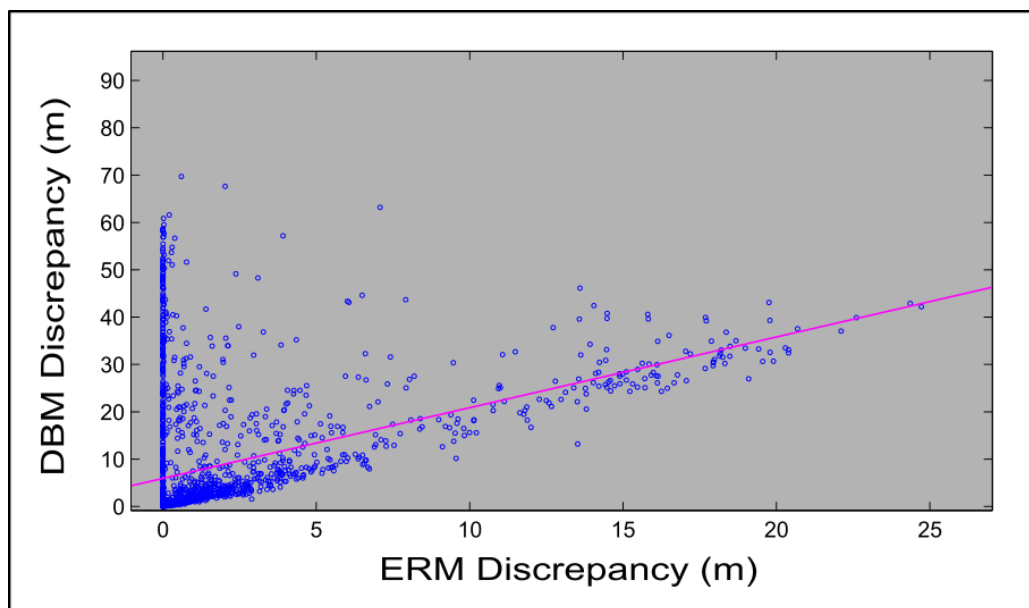

Figure 12: Scatter diagram between the values obtained for each method.

Table 8 presents the results of the PEC-PCD for OSM roads data, qualifying the results of the mean discrepancies of homologous lines for each method.

This assessment was made based on the two weighted calculation criteria for comparison with EM and EP values of Table 2.

Table 8: Result of PEC-PCD for each method.

	1:25,000	1:50,000	1:100,000	1:250,000
DBM	R	C	B	A
ERM	C	B	A	A

7. Conclusion

The studied methods allowed the determination of the mean discrepancy between pairs of homologous lines. The results of these methods were analyzed through controlled experimentation with distinct features varying the azimuth at a fixed distance. A case study with the OSM roads data was also carried out in relation to those available in the BDGEx.

In the DBM, for the same pair of homologous lines, different values of discrepancy were obtained (Table 5). This variation happened due to the adoption of the EM values as buffer distance which is different for each scale and PEC-PCD class. In the case of the ERM, there is only one discrepancy value.

Three cases were presented to prove that the feature configuration and the displacement azimuth interfere in the determination of the discrepancy in both methods. In cases where the lines are parallel with coincident parts, the ERM underestimates the value of the discrepancy by assigning the value 0 (zero).

Both the controlled experiment (for small scales) and the case study showed that the DBM overestimates the value of the discrepancy between linear features (Tables 6 and 7).

By visual inspection, it was verified that the ERM discrepancies reached results very close to the values calculated with the distance measuring tool of the QGIS. On the other hand, DBM discrepancies have almost double the result obtained by the other method in most pairs of homologous lines. This is proven through the spreading diagram (Figure 12).

Fonseca Neto et al. (2017), Cruz and Santos (2016), Santos et al. (2016) and Santos et al. (2015) arbitrated for the *x* value of the buffer the EM that in their works they denominate simply of PEC.

It is considered that the use of buffer distance requires detailed investigation. A fixed value in millimeters (mm) should be arbitrated for the buffer instead of EM. The fixed value would be transformed into meters according to the scale for calculating the discrepancy.

However, it is clear the DBM is more rigorous than the ERM in the assessment of Positional Accuracy. Any product approved in DBM will inevitably be approved in the ERM and, most likely, in the Traditional Method by homologous points.

The Traditional Method differs of both studied methods due to the advantage of detecting systematic errors, nevertheless the ERM and DBM detect automatically and efficiently the precision for linear features.

In terms of innovation, it is worth noting the use of the "adapted" criteria for the evaluation of the PEC-PCD for linear features that consider the weighting based on the length of the features. That is, larger features have average discrepancy values with greater weight than smaller features. Although this innovation is not standardized, the proposal is interesting to be studied and applied.

In the classification criteria of the PEC-PCD for linear features, the non-use of weighting by feature-length may fake the results, indiscriminately assigning the same weight to lines with different lengths.

As the calculated discrepancy value in both methods is a mean estimate, then a probability distribution based on the linear characteristic "length" as weight can be considered to evaluate the sample (Montgomery and Runger, 2002).

From the point of view of the results of the evaluation of the quality of the highways of the OSM, the results were worse than those found by Cruz and Santos (2016), being a possible reason, the region is a rural locality, pointed out by Sehra et al. (2014) and Helbich et al. (2012) as an area of low contribution and data validation.

The variety of open data resources unquestionably collaborate for economy in mapping works. However, it is worth emphasizing the necessary cares in their use, regarding the quality of the products that will be developed, in a way that meets its purpose, reflected mainly in the scale to which the product will be applied.

REFERENCES

Ariza-López, F. J., Atkinson-Gordo, A. D. J., Nero, M. A. and Cintra, J. P. 2007. La Componente Posicional de los Datos Geográficos: Análisis crítico de los Métodos de Evaluación y Reporte en Hispanoamérica. *Revista Cartográfica*, 83, pp.95–152.

Brasil 1984. Decreto nº 89.817 de 20 de Junho de 1984. *Normas Técnicas da Cartografia Nacional.*

Brovelli, M. A., Minghini, M., Molinari, M. E. and Zamboni, G. 2016. Positional Accuracy assessment of the OpenstreetMap buildings layer through automatic homologous pairs detection: The method and a case study. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, v. XLI-B2.

Cruz, D. T. and Santos, A. P. 2016. *Controle de Qualidade Posicional do Sistema Rodoviário do OpenStreetMap na região central de Viçosa-MG. VI Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação*.

DSG. Diretoria do Serviço Geográfico. 2015. *Norma da Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais de Defesa da Força Terrestre – ET-ADGV*. 1st Ed.

DSG. Diretoria do Serviço Geográfico. 2010. *Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais – ET-EDGV*. Edição 2.1.3.

DSG. Diretoria do Serviço Geográfico. 2016a. *Norma da Especificação Técnica para Controle de Qualidade de Dados Geoespaciais – ET-CQDG*. 1st Ed.

DSG. Diretoria do Serviço Geográfico. 2016b. *Norma da Especificação Técnica para Produtos de Conjunto de Dados Geoespaciais – ET-PCDG*. 2nd Ed..

Ferreira da Silva, L. F. C., & Cintra, J. P. 1999. Quantificação de Discrepâncias entre Feições Lineares por Retângulos Equivalentes. *Revista Brasileira de Cartografia*, 51, pp. 1–8.

Fonseca Neto, F. D., et al. 2017. Avaliação da Qualidade Posicional de dados espaciais gerados por VANT utilizando feições pontuais e lineares para aplicações cadastrais. *Boletim de Ciências Geodésicas*, 23(1), pp.134–149. doi: 10.1590/s1982-21702017000100009

Geofabrik. 2016. *Open Street Map Data Extracts*. Avaliable at: <http://download.geofabrik.de/>. [Acessed september 2017].

Getis, A. and Ord, J. K. 2012. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), pp.189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x

Goodchild, M. F. 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 4(4), pp.211–222. doi: 10.1007/s10708-007-9111-y

Helbich, M., Ameluxen, C., Neis, P. and Zipf, A. 2012. *Comparative Spatial Analysis of Positional Accuracy of Open Street Map and Proprietary Geodata. Angewandte Geoinformatik 2012*. Herbert Wichmann Verlag (pp. 24–33). Berlin, Offenbach: VDE Verlag GmbH.

Ming, W., Quingquan, L., Qui Qwu, H. and Meng, Z. 2013. Quality analysis of Open Street Map Data. *8th International Symposium on Spatial Data Quality*, Hong Kong. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-2*(W1),

Montgomery, D. C. and Runger, G. C. 2002. *Applied Statistics and Probability for Engineers* (3rd ed.). USA: John Wiley & Sons.

Nero, M., Cintra, J., Ferreira, G., Pereira, T. and Faria, T. 2017. A computational tool to evaluate the sample size in map positional accuracy. *Boletim de Ciências Geodésicas*, 23(3), pp.445–460. doi: 10.1590/s1982-21702017000300030

Penha, A. L. T., Morita, C. Y. and Cerqueira, R. W. 2012. Geração de Base Cartográfica Digital a partir de produtos fotogramétricos para a geração de ortofotocarta, carta topográfica e banco de dados geográficos – o caso do projeto de mapeamento do Estado da Bahia. *IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação*. SIMGEO.

Santos, A. P., Medeiros, N. G., Santos, G. R. and Rodrigues, D. D. 2015. Controle de Qualidade Posicional em dados espaciais utilizando feições lineares. *Boletim de Ciências Geodésicas*, 21(2), pp.233–250. doi: 10.1590/S1982-21702015000200013

Santos, A. P., Medeiros, N. G., Santos, G. R. and Rodrigues, D. D. 2016. Avaliação da Acurácia Poscional Planimétrica em Modelos Digitais de Superfície com o Uso de Feições Lineares. *Boletim de Ciências Geodésicas*, 22(1), pp.157–174. doi: 10.1590/S1982-21702016000100009

Sehra, S. S., Singh, J., Rai, H. S. 2014. A Systematic Study of OpenStreetMap Data Quality Assessment. *11th International Conference on Information Technology: New Generations.* Ramm, F. 2017. Open Street Map Data in Layered GIS Format. Disponível: <https://download.geofabrik.de/osm-data-in-gis-formats-free.pdf>. [Acessed september 2017].

Tveite, H. and Langaas, S. 1999. Na Accuracy Assessment Method for Geographical Line Data Sets Based on Buffering. *International Journal of Geographical Information Science,* 13(1), pp.27–47. doi: 10.1080/136588199241445