A comprehensive quality assessment framework for linear features from Volunteered Geographic Information

Article *in* International Journal of Geographical Information Science · October 2020

DOI: 10.1080/13658816.2020.1832228

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

A comprehensive quality assessment framework for linear features from Volunteered Geographic Information

Hao Wu , Anqi Lin , Keith C. Clarke , Wenzhong Shi , Abraham Cardenas-Tristan & Zhenfa Tu

To cite this article: Hao Wu , Anqi Lin , Keith C. Clarke , Wenzhong Shi , Abraham Cardenas-Tristan & Zhenfa Tu (2020): A comprehensive quality assessment framework for linear features from Volunteered Geographic Information, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2020.1832228

To link to this article: https://doi.org/10.1080/13658816.2020.1832228

Published online: 12 Oct 2020.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2020.1832228

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

A comprehensive quality assessment framework for linear features from Volunteered Geographic Information

Hao Wu[a,b], Anqi Lin[a,b], Keith C. Clarke[c], Wenzhong Shi[d], Abraham Cardenas-Tristan[e] and Zhenfa Tu[a,b]

[a]Hubei Provincial Key Laboratory for Geographical Process Analysis and Simulation, Central China Normal University, Wuhan, PR China; [b]College of Urban and Environmental Sciences, Central China Normal University, Wuhan, PR China; [c]Department of Geography, University of California, Santa Barbara, CA, USA; [d]Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Hong Kong, China; [e]Faculty of Engineering-Geomatics Engineering, Autonomus University of San Luis Potosí, San Luis Potosi, Mexico

ABSTRACT
The majority of spatial data provided as Volunteered Geographic Information (VGI) are roads and other linear map features. Such data have been widely used in routing and navigation, road network update, emergency response, urban planning and more. Due to the lack of cartographic standards and issues with volunteer credibility, the quality of VGI linear features remains a concern and could seriously hinder the broad application of VGI data. This research proposes a comprehensive quality assessment framework for VGI linear features which adopts factor analysis to integrate two novel quality metrics with six other commonly used metrics, and further examines the spatial autocorrelation and semantic correlation of VGI linear feature quality. The OpenStreetMap road network of Allegheny County, Pennsylvania (USA) was selected as an example to test the proposed framework. Our results suggest that the proposed metrics, Box-counting dimension difference and Link accuracy are feasible for detecting quality issues and are important supplements to the common quality metrics. The findings also show that significant spatial autocorrelation exists in spatial completeness, positional accuracy, and logical consistency. Road type such as Tertiary, Residential, Service and Link has been proven to be a typical indicator of the different quality elements for VGI linear features.

ARTICLE HISTORY
Received 7 February 2020
Accepted 30 September 2020

KEYWORDS
Volunteered geographic information; linear features; quality assessment; factor analysis; OpenStreetMap

1. Introduction

Volunteered Geographic Information (VGI) has proven highly successful in enabling volunteers to contribute spatial data at very low cost (Goodchild 2007), particularly roads and other linear map features (Barrington-Leigh and Millard-Ball 2017). The availability of large amounts of VGI linear features as well as others has greatly facilitated many applications in routing and navigation (Bergman and Oksanen 2016), road network updates (Manandhar *et al*. 2019), emergency response (Lin *et al*. 2020), urban planning (Moeinaddini *et al*. 2014), and other fields (Fonte *et al*. 2015, Hong and Yao 2019, Yang *et*

CONTACT Hao Wu haowu@mail.ccnu.edu.cn

al. 2020). Like any other user-generated content, the uncertain quality of VGI linear features has impeded their further applications because they are contributed by non-professional, amateur volunteers without rigorous data specifications (Goodchild and Glennon 2010, Goodchild and Li 2012, Severinsen *et al*. 2019). The topological and geometric characteristics of linear features also do not allow a standardized evaluation method consistent with points and polygons (Xie and Levinson 2007). Therefore, a robust understanding of VGI linear feature quality is necessary if we would like to further exploit where problematic data are clustered and what potential factors are related to the data quality.

The history of assessing spatial data quality dates back to the 1980s, with attention to error measurement in cartography (Goodchild and Gopal 1989). Nevertheless, it was not until 2002 that the five quality elements, completeness, logical consistency, positional accuracy, temporal accuracy, and thematic accuracy were formally standardized by ISO 19113 (ISO 2002). For VGI linear features, it was reported that completeness, positional accuracy, and logical consistency should be assigned more importance than the others because they are main elements ensuring the OSM primary application in routing and navigation (Neis *et al*. 2012, Barron *et al*. 2014, Mobasheri *et al*. 2018b). Specifically, the road network for navigation service should be complete and accurate enough so that the position of vehicles could be displayed correctly on the road where they are actually located (Seo *et al*. 2009). Meanwhile, navigability heavily depends on the correct topological relation of the road segments (Sehra *et al*. 2016, 2019). Among these three quality elements, logical consistency is inherent in the data structure and so can be easily identified by error detection tools such as Keep Right (keepright.at) and Osmose (osmose.openstreetmap.org). However, assessing the completeness and positional accuracy usually requires comparison with an external dataset that has no standardised tool or method, and therefore inhibits use.

At present, OpenStreetMap (OSM), as one of the leading VGI projects, boosts significant achievements in mapping road networks worldwide, but suffers from the credibility concern (Yan *et al*. 2020). The completeness and accuracy assessments of VGI linear features in the case of OSM remain a hot topic and have led to numerous studies. Regarding the completeness of VGI linear features, Haklay (2010) launched the fundamental quality study of OSM road network by comparing the road length between OSM and Ordnance Survey datasets in England, with this approach used in Germany (Zielstra and Zipf 2010) and France (Girres and Touya 2010). Forghani and Delavar (2014) proposed four heuristic metrics to measure the uncertainty of OSM road networks. Mobasheri *et al*. (2018a) applied a polygon-based method to match and compare sidewalk between OSM and GPS trace. Chehreghan and Abbaspour (2018) put forward geometric metrics to assess the completeness through the distance and orientation of VGI linear features. With respect to positional accuracy, Goodchild and Hunter (1997) proposed the classic buffer method that has been widely applied for measuring the positional error of VGI linear features. Girres and Touya (2010) calculated the Euclidean distance for homologous nodes and the Hausdorff distance for corresponding links between OSM and authoritative data in France, while Stark (2010) evaluated the error distance between OSM and reference data. Kalantari and La (2015) measured positional accuracy with an index of shape similarity. Despite these significant contributions to the quality assessment of VGI linear features, little substantial research has yet been attempted to investigate whether these existing metrics can fully detect quality issues under the reality of the rapid growth in crowdsourcing data. After all, not only does complexity of road networks increase in metropolitan areas, there is also an emergence of data sourced from uploaded driven GNSS tracks, out-of-copyright maps, street view imagery, and overhead photography. These new and varied sources have extensively challenged effective accuracy and quality assessment for VGI linear features.

More importantly, researchers have made use of the aforementioned metrics to identify the quality bias of VGI linear features and have drawn meaningful insights (Senaratne *et al*. 2017). It has been reported that OSM road networks in well-governed countries with accessible Internet services tend to be more complete (Barrington-Leigh and Millard-Ball 2017). Also, cities usually received more data contributions with higher quality than rural areas in many regions and countries such as Ireland (Ciepłuch *et al*. 2010), England (Fairbairn and Al-Bakri 2013), and China (Zhang *et al*. 2015, Zhou and Tian 2018). This has inspired the development of proxy indicators that can reveal the data quality without comparison with external data sources. Zielstra and Zipf (2010) suggested that low population density areas, especially rural villages, were characterized by poor VGI data coverage. Exel *et al*. (2010) pointed out that three components of volunteer background, namely local knowledge, experience and recognition, had significant impact on the VGI data quality. Keßler and Groot (2013) proved that frequent revisions had a negative influence on OSM data quality. Camboim *et al*. (2015) examined the correlation of VGI data quality with economic and developmental variables, such as GDP and average income. Mobasheri *et al*. (2017) developed the data indicators, such as total number of highways, buildings, residential facilities, and OMS users, that were correlated with the completeness of sidewalks. Truong *et al*. (2019) identified the useful behaviours of mapping contributors that indicate reliable data. Notably, the effective application of these quality indicators usually depends on the availability of richer data such as metadata, demographic information, socioeconomic factors, and contributor background (Antoniou and Skopeliti 2015, Haklay 2016, Gardner and Mooney 2018). Actually, such data, especially at a fine scale, are usually either unrecorded or inaccessible due to commercial interests or licensing restrictions (Basiri *et al*. 2019). To advance our capability to rapidly determine the data quality, it is critically important to develop new indicators by exploring rules or patterns from VGI data itself. For example, quantifying the spatial autocorrelation of VGI linear feature quality will help to identify spatial clusters of high/low data quality and further predict quality issues at specific places. Meanwhile, road type, a key attribute of road networks, also has great potential to indicate the VGI data quality as it reflects the functional accessibility of the transportation infrastructure to human activities (Lee *et al*. 2017). Therefore, there is still room to explore the spatial correlation and semantic correlation of VGI linear feature quality.

Here we propose a comprehensive quality assessment framework for VGI linear features. It integrates two novel quality metrics with six other commonly used metrics through factor analysis, and further examines the spatial autocorrelation and semantic correlation of VGI linear feature quality. The thorough understanding of VGI linear features, from direct quality measurements to potential quality indicators, should facilitate deeper insight into this collaborative spatial data product and thereby allow a targeted quality control policy.

2. Methodology

2.1. Typical quality metrics for VGI linear features

2.1.1. Geometry-based quality metrics

The spatial quality of VGI linear features can be investigated by comparing the features with an authoritative dataset from a geometric perspective, a common idea in previous studies (Goodchild *et al*. 1992, Hunter 1999). In this study, we took the road data from the OSM project as a case study to measure the quality of VGI linear features.

Road length difference (RLD) is the most common metric used to compare the completeness of the OSM road network. It can be evaluated by calculating the total length of the OSM roads within a certain area, and then comparing it with that of a reference map within the same extent. None of any certain road type is filtered out when making the comparison. There are two types of completeness errors, including omission (missing data) and commission (excessive data). Both of them lead to the length difference between these two linear datasets. RLD can be calculated by Equation (1):

$$RLD = \left| \sum RL_{OSM} - \sum RL_{REF} \right| \tag{1}$$

where RLD is the road length difference; $\sum RL_{OSM}$ is the total length of OSM road network, $\sum RL_{REF}$ is the total length of reference road network. The difference in the overall length indicates the completeness of VGI data as it assumes that the reference data can better represent the ground truth.

Median centre has been widely used in measuring the central tendency of a set of features (Kuhn and Kuenne 1962). For a road network, the median centre indicates the location that minimizes the overall Euclidean distance to all road segments in the study area, which gravitates towards an area with the most features. In detail, an iterative algorithm is adopted and a candidate point is refined at each step in the iterative process until it represents the location that minimizes the Euclidean distance to all road segments. The Median centre distance (MCD) between OSM dataset and reference dataset has been used previously for assessing OSM data quality in Tehran (Forghani and Delavar 2014). Smaller MCD values indicate similar spatial distributions of the two datasets, and thus better OSM data quality. It can be calculated using Equation (2):

$$MCD = \sqrt{(X_{OSM\_MC} - X_{REF\_MC})^2 + (Y_{OSM\_MC} - Y_{REF\_MC})^2} \tag{2}$$

where MCD is the distance between the median centre position of OSM and reference road network; (X_{OSM_MC}, Y_{OSM_MC}) is the median centre coordinate of VGI dataset, and (X_{REF_MC}, Y_{REF_MC}) is the median centre coordinate of the reference road network.

Node accuracy (NA) in this study refers to the positional errors from the VGI node to the corresponding node in the reference dataset using automatic point matching (Seo *et al*. 2009). In particular, only nodes at crossroads and the end of any road segments will be tested, and the pseudo nodes that connect two edges of a single road will be filtered out when creating the road network topology. The reference nodes are buffered with a given radius, and then matched with the nearest VGI node from those falling in the buffer zones. If there is a match, the Euclidean distance of corresponding nodes will be derived by differencing the coordinates of the matching nodes. The average distance for all pairs is the measure of NA using Equation (3):

$$NA = \sum_{i=1}^{n} \sqrt{(X_{OSM\_ND} - X_{REF\_ND})^2 + (Y_{OSM\_ND} - Y_{REF\_ND})^2} / n \quad (3)$$

where (X_{OSM_ND}, Y_{OSM_ND}) is the node coordinate of VGI dataset, $(X_{REF_ND}, Y_{REF_ND})$is the node coordinate of the reference road network.

2.1.2. Topology–based quality metrics

Topological consistency can be measured by examining errors against certain rules that define the spatial relationship within the feature class. A large number of topological errors remain unsolved in OSM dataset, so OSM encourages users to apply quality assurance tools such as Keep Right and Osmose to detect and correct those bugs (Hashemi and Abbaspour 2015, OSM 2020b). Therefore, topological consistency is important and necessary in assessing the quality of VGI linear features. In this study, the values of the three most common topological errors were chosen as quality metrics, namely Floating island (FI), Almost junction (AJ) and Intersection without junction (IWJ) as shown in Figure 1.

FI (Figure 1(a)) refers to roads that are not connected to the rest of the road network. Theoretically, any road segment on the ground should be accessible from at least one other road except for one tiny island in the middle of the sea far away from the mainland. AJ (Figure 1(b)) signals that the nodes at the end of the road segment are very close, but not connected to, other segments. Roads that are closer than a minimum distance, often 10 m, to another road are considered to be AJ. As for IWJ (Figure 1(c)), it is a metric indicating multiple roads crossing on the same layer but without a common node as the junction. For navigation proposes, the node splitting the road segments at the intersection is necessary to make the turn possible.

2.2. Novel quality metrics for VGI linear features

Although the metrics covered in Section 2.1 have been successfully used to assess the quality of VGI linear features, the question of whether they are sufficient for detecting all

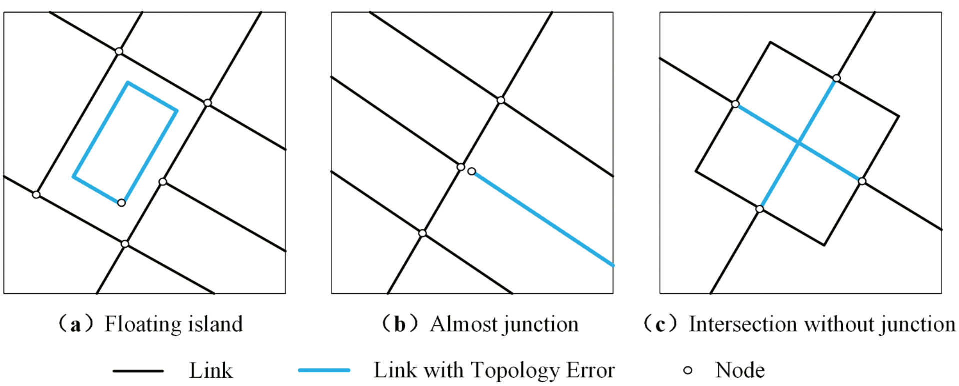


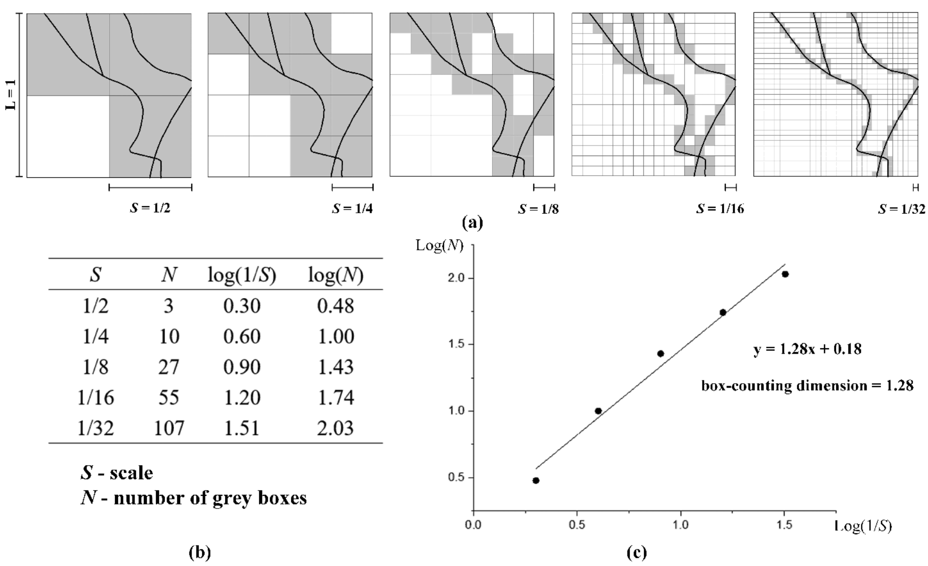
Figure 1. Graph illustration of three common topological errors, (a) Floating island, (b) Almost junction, and (c) Intersection without junction.

quality issues remains largely unanswered. For instance, a small road length difference between OSM and reference road network does not indicate a high degree of completeness when there is a mix of commission errors and omission errors with similar length in the OSM dataset. Due to the complexity of road network structures, it is necessary to introduce new metrics to reduce the bias in measuring quality in OSM datasets. In this context, we propose two novel quality metrics, the Box-counting dimension difference (BDD) and the Link accuracy (LA).

2.2.1. Box-counting dimension difference metric

Fractal geometry is an efficient branch of mathematics that can describe a variety of complex structures, including road networks. The description of the structural fractality of road networks allows a better understanding of the complexity and dynamics of a road system. Given this, the fractal dimensions of complex networks can be calculated by the simple box-counting method. In particular, the box-counting dimension is a statistical magnitude measuring the space-filling efficiency, which ranges from 1 to 2, and the value increases with the higher space-filling efficiency of the structure (Sreelekha *et al*. 2017). The method adopts a set of successive square boxes to cover the study area and count the boxes that overlap any part of the network (Figure 2). Then, the above step is repeated several times with different box sizes. The log of the power–law relationship between box size and box count forms a straight line and the slope gives the box-counting dimension by the following equation:

$$D = \lim_{s \to 0} - \frac{\log N(s)}{\log s} \tag{4}$$



S	N	log(1/S)	log(N)
1/2	3	0.30	0.48
1/4	10	0.60	1.00
1/8	27	0.90	1.43
1/16	55	1.20	1.74
1/32	107	1.51	2.03

S - scale
N - number of grey boxes

(b)

Figure 2. Illustration of estimating Box-counting dimension for road network, including (a) Overlapping boxes at different scales on road network, (b) Counting the number of colored boxes with respect to different scales, and (c) Plotting Log-log graph to calculate the box-counting dimension.

where D denotes box-counting fractal dimension, s is the size of the box, and $N(s)$ is the number of boxes overlapping the road network at the box scale s. The absolute value of the difference between the box-counting fractal dimension of the OSM and reference datasets was applied as BDD:

$$BDD = |D_{OSM} - D_{REF}| \tag{5}$$

where D_{OSM} is the box-counting fractal dimension of the OSM road network and D_{REF} is the box-counting fractal dimension of the reference road network. BDD measures the difference of structural complexity between the two road networks. Smaller BDD values represent more similar fractal structures, in the sense that spatial distribution of OSM geometries is more consistent with the reference geometries.

2.2.2. Link accuracy metric

The Buffer method index (BMI) is commonly used for evaluating the positional accuracy of linear features (Goodchild and Hunter 1997), which computes the percentage of the test linear feature's length inside the buffer around the reference feature. Considering the length ratio fails to detect the geometrical similarity between those two datasets, the proposed LA measures the overlap ratio of two buffer bands derived from both test and reference datasets (Figure 3). Taking advantage of the metric Surface distance that calculates the distance between corresponding polygons by the rate of the intersection and union area of the two polygons (Vauglin 1997), LA can be computed using the following equation:

$$\mathrm{LA} = 1 - \frac{S(OSM \cap REF)}{S(OSM \cup REF)} \tag{6}$$

where $S(OSM \cap REF)$ is the intersection area of the buffers of OSM and the reference road network, and $S(OSM \cup REF)$is the union area of the buffers of the OSM and reference road networks. LA is defined in the interval [0, 1]. If the two buffers have a 100% match, the intersection area would be the same as the union area, and LA would be equal to 0. If the two buffers were completely disjoint, the intersection area would be 0, and LA would be

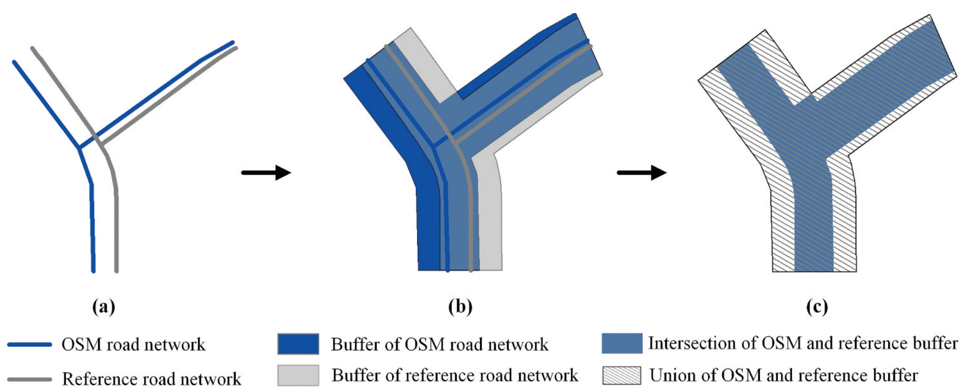


Figure 3. Procedure to measure the Link accuracy of road network: (a) Match reference road network and the corresponding OSM road network to be tested, (b) Create buffers around the reference road network and the OSM road network, and (c) Determine the intersection and union of their buffers.

equal to 1. In other words, the lower the positional accuracy of linear features, the higher the LA value.

2.3. A comprehensive quality assessment framework for VGI linear features

To provide a thorough insight into the quality of VGI linear features, we developed a comprehensive quality assessment framework consisting of four steps (Figure 4): data processing; metrics computation; quality assessment; and result interpolation. Data processing involves map projection transformation, data clipping, topology creation, and grid cell generation for both the VGI and reference datasets. Considering the spatial heterogeneity of VGI data quality, the same 1 km grid was overlaid over both the VGI and reference linear features in the study area as a means of calculating the metrics for each cell. It was reported that there were slight variances between the results of quality measures under different cell sizes at 0.5 km, 1 km, and 2 km in a study of OSM buildings (Zhou 2018). Given that, the regular cell size, 1 km, was used in our study. After that, five geometry-based metrics (RLD, MCD, BBD, NA, LA) and three topology-based metrics (IWJ,

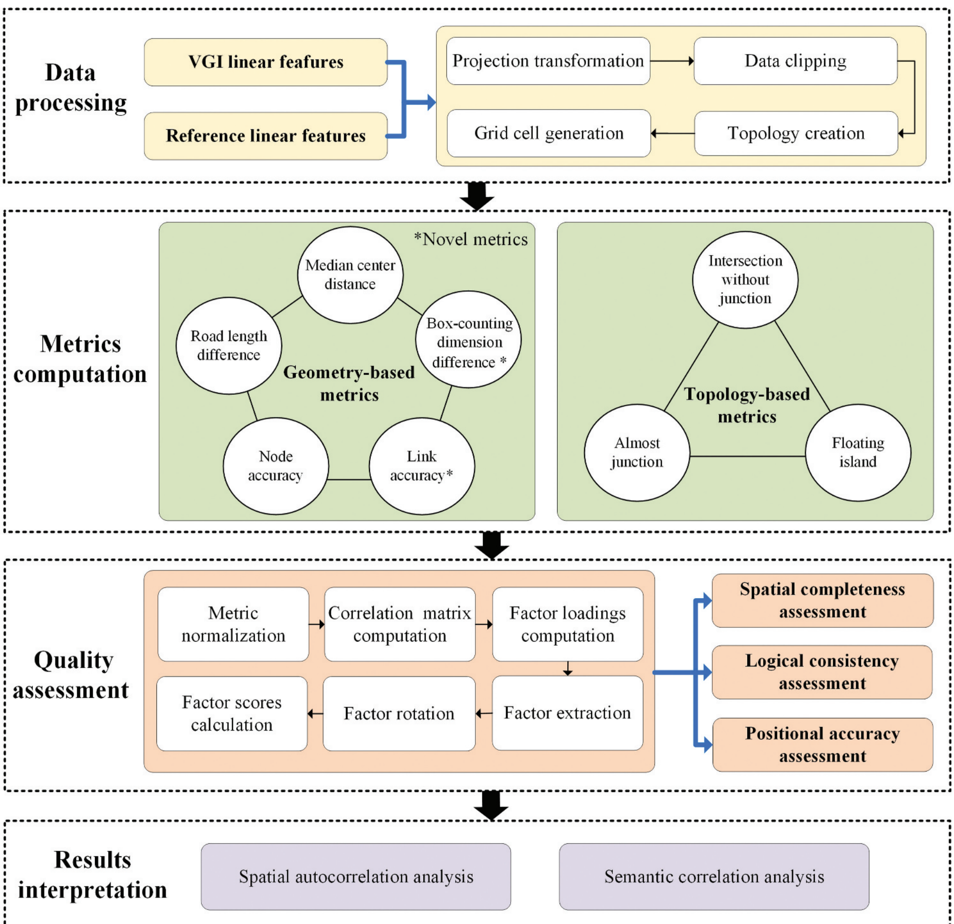


Figure 4. The comprehensive quality assessment framework for VGI linear features.

AJ and FI) were computed for each grid cell. The five geometry-based metrics were measured by the comparison to extrinsic data-based grounding, and the three topology-based metrics were measured by the comparison to the intrinsic grounding in rules (Mocnik *et al.* 2018).

To reveal the underlying dimensions in the multiple metrics, factor analysis was performed to extract the main factors that can provide a straightforward and comprehensive quality assessment. All eight quality metrics were normalized to the range (0, 1) and used to compute the correlation matrix and the factor loadings. According to Kaiser's criteria (eigenvalue greater than 1), the factors are extracted by principal component analysis. Then, factor rotation was applied to get a simplified structure that loads each metric strongly onto one specific factor. The factor scores of spatial completeness, positional accuracy and logical consistency for each cell were calculated. Eventually, the results are analysed and interpreted to detect spatial autocorrelation as well as semantic correlation.

3. Study area and data

3.1. Study area

The OSM road network for Allegheny County, with the City of Pittsburgh in the centre, was selected for the quality assessment of VGI linear features (Figure 5). Allegheny County is the second-most populous county in Pennsylvania with a population of 1,223,048 and an area of 1,930 km^2. More than one-quarter of the people live in the City of Pittsburgh, which is the most urbanized and developed area in Allegheny County. Pittsburgh is one of the pioneering cities developing autonomous vehicle (AV) technology, and was designated an AV proving site by the United States Department of Transportation in 2017. Aptiv, a global auto parts company testing driverless car in Pittsburgh, has explored OSM capability for road perception in autonomous driving (Zheng and Izzat 2018, Zheng *et al.*

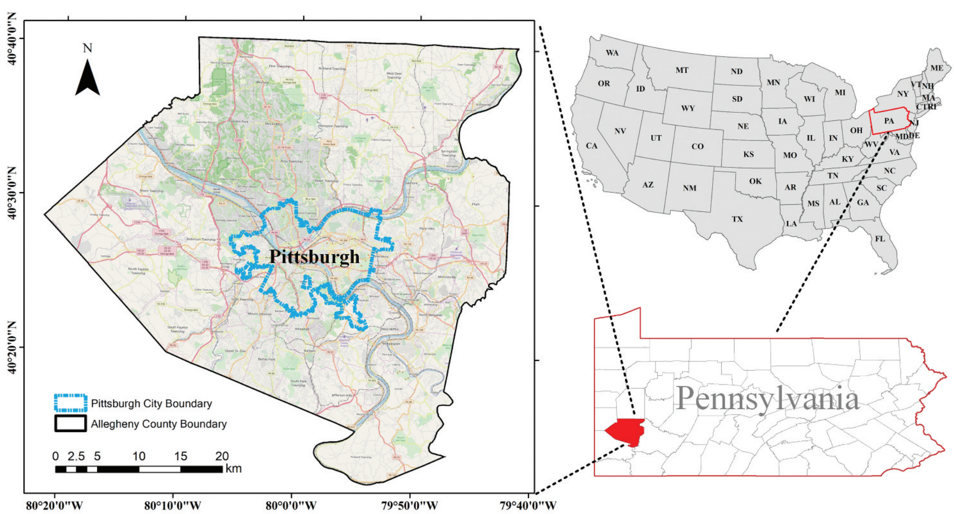


Figure 5. Study area, Allegheny County, Pennsylvania, USA.

2019). It is worthwhile to examine the OSM road network quality in this area as it is a potential date source for AV development. For context, OSM data have been applied in routing, navigation, and other services by many companies around the world, such as Skobbler, Mapbox, and Mapla (OSM 2020a).

3.2. Data and data processing

The OSM road dataset within the boundary of Allegheny County was downloaded from download.geofabrik.de in July 2018. Reference road data and the county boundary in shapefile format for the same area came from the Western Pennsylvania Regional Data Center (WPRDC, data.wprdc.org). The OSM road dataset was reprojected to NAD83 Pennsylvania South (EPSG: 2272), the same as the authoritative dataset from WPRDC. Both OSM and the reference dataset were clipped to the shape of Allegheny County, and the topology for both datasets was generated. Due to the quality heterogeneity, both datasets were gridded at a resolution of 1 km as the basic quality assessment unit, which generated 1,849 cells excluding those with only minor or no data.

4. Results

4.1. Computation results of the eight quality metrics

The five geometry-based metrics (RLD, MCD, BBD, NA, LA) and three topology-based metrics (IWJ, AJ and FI) discussed above were calculated separately for all 1,849 cells through batch processing in ArcGIS 10.2 and python scripts. The quantitative results for the eight metrics were normalized to the interval [0, 1] (Figure 6). In general, the quality of the OSM road network is good, as the mean of most metrics are less than 0.2. The distributions of RLD, MCD and BDD are positively skewed. However, the range of RLD is less than that of MCD and BDD, which implies that the values of RLD are less dispersed, and OSM and reference data are more consistent in terms of feature length. It can be seen that NA and LA are approximately symmetric distributions. This is reasonable because the positional errors for both nodes and links are random errors which should follow the normal distribution. Also, the IWJ, AJ and FI values are less dispersed and the data quality from topological perspective is much better than other metrics.

The thematic maps for the eight metrics (Figure 7) illustrate the relatively high spatial heterogeneity of the eight data quality metrics across the study area. The three topological quality metrics (IWJ, AJ and FI) are substantially different from the other metrics as there are clear sets of high values near the centre where Pittsburgh is located. Meanwhile, the similar pattern such as the high values in the north west corner for both RLD and BBD is to be expected, as the metrics extracted from the comparison between two spatial datasets might contain duplicate information.

4.2. Spatial data quality assessment results

The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity both indicated a factor analysis was appropriate using these data. The KMO index was 0.657, which exceeds 0.5, while Bartlett's Test was significant ($p < 0.05$). The first

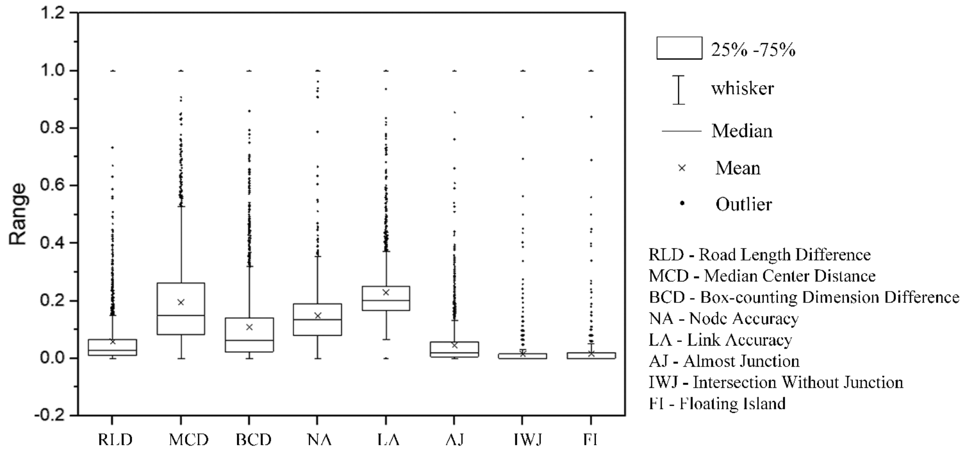


Figure 6. Box plots of the eight quality metrics of the OSM road network.

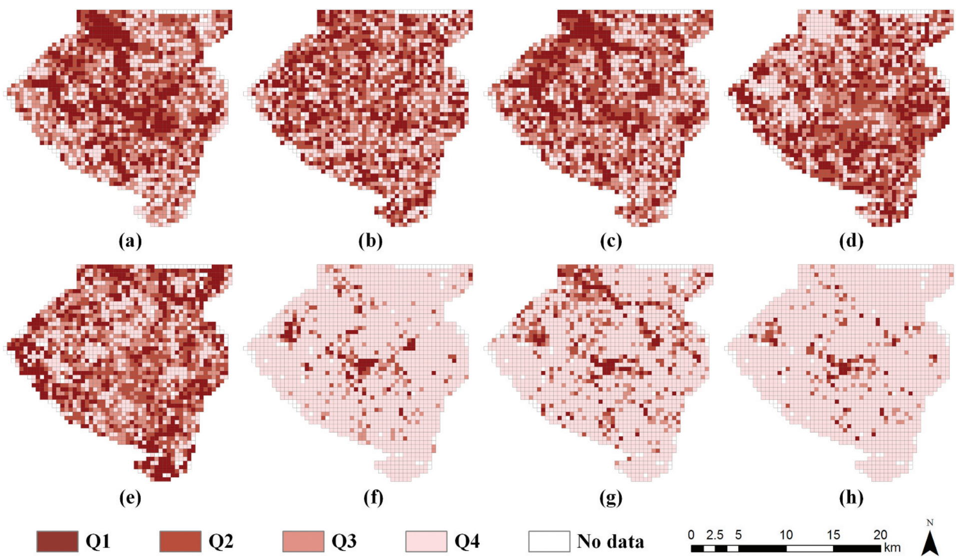


Figure 7. Thematic maps of the eight quality metrics for the OSM road network: (a) Road length difference, (b) Median centre distance, (c) Box-counting dimension difference, (d) Node accuracy, (e) Link accuracy, (f) Almost junction, (g) Intersection without junction, and (h) Floating island. Q1 is the upper quartile.

three factor analysis components (Table 1) have eigenvalues greater than 1 (Kaiser's criteria), and account for a cumulative 71.8% of the total variance explained. Therefore, three factors were determined to capture sufficient information from the original eight quality metrics. The Varimax rotation was then utilized to polarize the variable loadings (Table 2). Across the three factors, spatial completeness is the most important, explaining 37.2% of the total variance, and corresponds with the quality metrics RLD, MCD and BBD; logical consistency accounts for 15.0% of the total variance and corresponds with IWJ, AJ

Table 1. Eigenvalues from principal components analysis of the eight quality metrics.

Component	Eigenvalue	Proportion(%)	Cumulative(%)
1	2.973*	37.2	37.2
2	1.572*	19.7	56.8
3	1.196*	15.0	71.8
4	0.730	9.1	80.9
5	0.612	7.7	88.6
6	0.580	7.2	95.8
7	0.182	2.3	98.1
8	0.155	1.9	100

*Indicates that eigenvalues are larger than 1, and corresponding factors are retained.

Table 2. Rotated component matrix of eight quality metrics.

Metrics	Factor1	Factor 2	Factor 3
RLD	0.785*	0.291	−0.230
MCD	0.750*	−0.084	0.109
BDD	0.867*	0.245	−0.135
NA	−0.375	−0.029	0.692*
LA	0.128	−0.086	0.866*
IWJ	0.055	0.917*	0.024
AJ	0.121	0.658*	−0.221
FI	0.135	0.907*	0.016

* Indicates the largest loading of each metric on a certain factor among the three.

and FI; positional accuracy accounts for 19.7% of the total variance and corresponds with NA and LA.

Factor scores were calculated for the three main factors, spatial completeness, logical consistency, and positional accuracy, are given in Figure 8. It is much easier to identify the spatial patterns for the three factors in Figure 8 than in Figure 7. In particular, highly incomplete data were more common in rural areas than urban areas (Figure 8(a)). For logical consistency, high values were mainly in the middle of the study area where Pittsburgh is located and the values gradually decreased from the centre to the

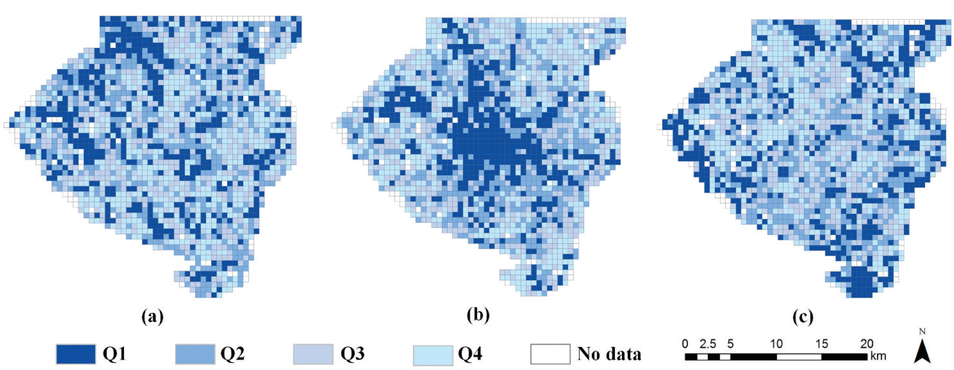


Figure 8. Thematic maps of the three quality factors for the OSM road network: (a) Spatial completeness, (b) Logical consistency, and (c) Positional accuracy. Q1 is the upper quartile.

surrounding areas (Figure 8(b)). Large positional errors were mostly around the boundary of Allegheny County (Figure 8(c)).

5. Discussion

5.1. Complexity of quality assessment for VGI linear features

As the most typical VGI linear feature, the road network consists of interconnected lines and points, which possess many different structural and geometrical properties (Xie and Levinson 2007, Wu *et al.* 2014). Due to the complex properties of road networks, it is necessary but difficult for us to thoroughly examine spatial data quality especially for important elements such as spatial completeness and positional accuracy.

To demonstrate the necessity of the supplemental metrics and the combination of multiple metrics for examining spatial completeness, we compared both values and percentile ranks of RLD and MCD with proposed BDD across four typical examples (Figure 9). For OSM road networks that are very consistent with the reference dataset, all three metrics give the same results with a high degree of completeness (Figure 9(a)). However, for those that are not, these three metrics have clear differences (Figure 9(b-d)). When omission errors approximate commission errors (Figure 9(b)) RLD (top 6th percentile) overestimates the completeness of the OSM road network, while both MCD (88th percentile) and BDD (57th percentile) provide more reasonable assessments. Although both MCD and BDD can make up for the deficiency of RLD, MCD always functions worse than BDD. For example, a minor outlier in the OSM dataset shown in Figure 9(c) causes a

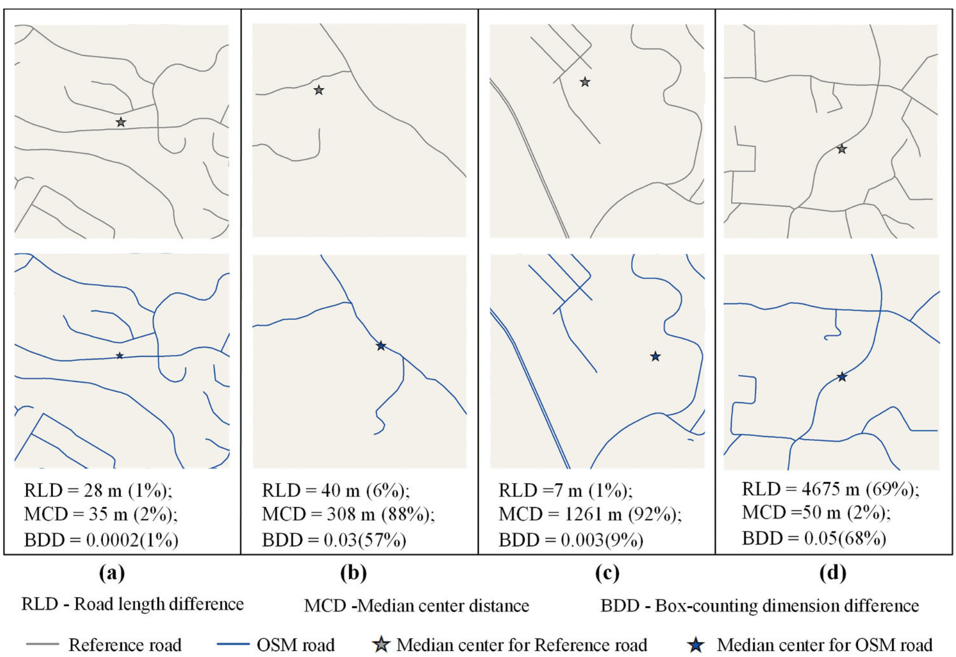


Figure 9. Comparative analysis of the spatial completeness of the OSM road network derived from RLD, MCD, and BDD metrics for four typical examples.

clear shift of the median centre, underestimating the completeness. More importantly, there is an obvious difference between the OSM and reference datasets where the MCD (top 2nd percentile) indicates high consistency (Figure 9(d)). This further suggests that MCD cannot thoroughly measure the completeness because it easily leads to bias. From the results discussed above, the proposed BDD is a necessary supplementary metric to RLD and MCD for complex quality assessment on VGI linear features so as to achieve a better understanding of the spatial completeness.

The positional accuracy with respect to nodes and links of the road network should also be taken into consideration (Goodchild and Hunter 1997). Positional error of nodes has been fully discussed in previous studies (Safra *et al*. 2010, Jackson *et al*. 2013), whereas there is no standard for positional uncertainty of links because the distance between linear features is much more complicated to measure (Zhang *et al*. 2019). The commonly used BMI, computing the percentage of length from the test linear feature inside the buffer around reference one, fails to detect the geometrical similarity between the two datasets. Instead of calculating the length ratio by BMI, our proposed LA measures the overlapping ratio of two buffer bands derived from both test and reference datasets. Figure 10 shows that the overlapping area of buffer bands (blue) gradually dominates, while the union area (red) shrinks. This indicates that the OSM road network is increasingly approaching the reference road network. Both BMI and LA were used to assess the positional accuracy for the four examples. BMI varied from 0.95 to 1.01 (a narrow range of 0.06), while LA changed from 0.64 to 0.92 (a much wider range of 0.28). This suggests that LA has a better identification capability for assessing the detailed difference between the test and reference datasets than BMI. In essence, LA not only keeps the relative positional relationship between the corresponding linear features, but also takes the shape and curvature of linear features into account. This indicates that our proposed LA can better expose the shortcomings of the OSM linear features, allowing us to make a more insightful assessment in terms of positional accuracy.

5.2. Validation of spatial autocorrelation in VGI linear feature quality

One of the main purposes of VGI quality assessment is to identify where the reliable data is located and where the poor data lie. Many studies have generally concluded that VGI

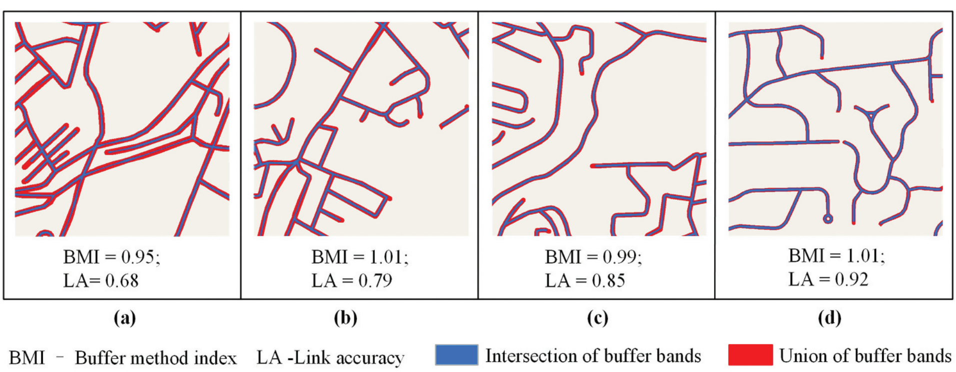


Figure 10. Comparative analysis between the Buffer method index and the proposed Link accuracy for the OSM road network using four typical examples.

has better quality in urban than rural areas (Neis and Zielstra 2014, Ma *et al*. 2015, Wu *et al*. 2019).

In this context, both Global and Local Moran's I were adopted to examine the spatial autocorrelation of the spatial completeness, logical consistency, and positional accuracy (Figure 11). As is well known, the global Moran's I is suitable to measure the overall spatial autocorrelation at the macro level, and the local Moran's I can identify local clusters and outliers at the local level. We used the Queen's contiguity to define the spatial weights for the local and global Moran's I calculations. Significant positive spatial autocorrelation (Global Moran's I > 0) was found for all three quality elements across the study area. It implies that these three quality elements are characterized of spatial clustering to some extent. However, local Moran's I results indicate there is spatial non-stationarity in the distribution characteristics. For example, in terms of spatial completeness, high-high clusters are mainly located in rural areas while low-low clusters are in Pittsburgh, which means that the spatial completeness of the road network in the urban area is much higher than in the rural areas. This trend is similar to that identified in previous studies (Neis *et al*. 2012, Koukoletsos *et al*. 2012, Barrington-Leigh and Millard-Ball 2017). More surprisingly, we found interesting results for logical consistency and positional accuracy (Figure 11(b, c)). For logical consistency, high-high spatial clusters were mostly in the urban area while low-low spatial clusters were in the rural areas, which suggests that topological errors happen more frequently in the urban areas. It reveals that the large amount of volunteers in an urban area does not ensure a better topological consistency of VGI linear features,

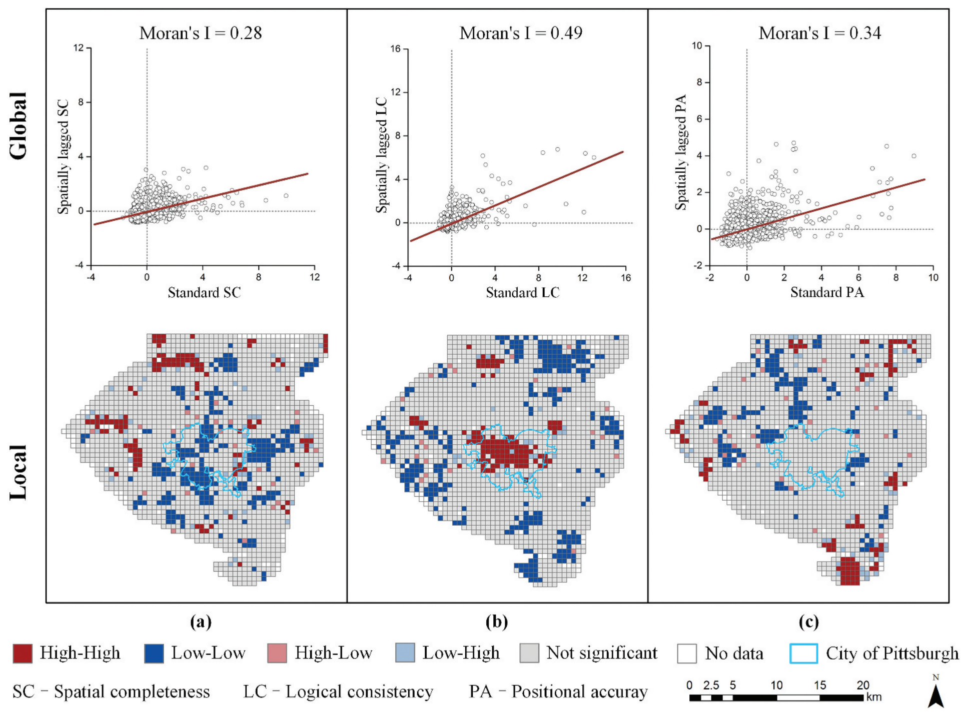


Figure 11. Spatial autocorrelation analysis for three factors using global and local Moran's I, including (a) Spatial completeness, (b) Logical consistency, and (c) Positional accuracy.

even if it can contribute to data completeness as discussed above. Particularly, with the intense update frequency from high VGI user activity, the side effect of loose cartography standards and nonprofessional data editing on logical consistency become more prominent in urban areas. As for positional accuracy, there are fewer and smaller clusters than the other two quality elements, which corresponds to the fact that the distribution of positional accuracy approximates a normal distribution (Figure 6). Meanwhile, high-high spatial clusters are mostly located at the boundaries of Allegheny County, mostly covered with forest where the positional accuracy might be affected by the intermittent loss of GPS signals due to tree obstructions.

Given the above, one should consider different distribution characteristics of spatial clusters as well as the potential reasons for the three quality elements if the intent is to develop targeted quality control strategies facing the different quality issues of VGI linear features. For example, OSM should promote their project or design incentive mechanism especially in rural areas in order to increase the awareness of the need for more data in those areas and motivate users to contribute more data. This is inspired by a sidewalk enrichment strategy proposed by Mobasheri *et al*. (2018b). The topology should also be checked more frequently in the urban area with complicated road networks. GPS trajectories, especially from forest areas, should follow stricter data entry protocols regarding the GPS devices, weather conditions, user's professional level and other factors, to assure better positional accuracy from such data.

5.3. Correlation between road type and VGI linear feature quality

Given that VGI data is the result of public participation, many external quality indicators, such as socioeconomic factors, population density, and contributors' background, have been applied to assess VGI data quality rather than the inherent attribution (Neis *et al*. 2013, Fogliaroni *et al*. 2018). Road type as the essential attribution of road network has considerable potential to indicate VGI data quality because it refers to the accessibility of the transportation infrastructure for various human activities (Lee *et al*. 2017). The OSM road system classifies the road into Motorway, Trunk, Primary, Secondary, Tertiary, Unclassified, Residential, Service as well as Link for all kinds of roads (OSM 2019). The Motorway, Trunk, Primary and Secondary road types have very low Pearson's correlations with any of the three quality elements (Table 3). This is perhaps because that these fundamental roads are mostly imported from TIGER data in 2007 and 2008, so they have uniform quality standards, and have largely remain unchanged since then (Zielstra *et al*. 2013). The other types, Tertiary, Residential, Service and Link roads, are significant. For example, Tertiary roads are positively correlated with positional accuracy (0.140**), which means that they generally have lower positional accuracy (the lower the factor scores, the better the data quality). This is because Tertiary roads often connect small towns and villages where fewer people provide GPS traces, and many linear features are digitized based on satellite images. Residential roads show significantly negative correlation with spatial completeness (−0.426**), suggesting that Residential roads are more complete than other road types. This is mainly due to the higher population density and more frequent human activities in residential areas, which improves road spatial completeness. Moreover, Residential roads are negatively correlated with logical consistency (−0.187**) because the roads in residential areas are

Table 3. Pearson correlation coefficients between road types and quality elements.

	Spatial completeness	Logical consistency	Positional accuracy
Motorway	−0.071**	0.087**	−0.011
Trunk	−0.043	0.055*	−0.019
Primary	−0.041	0.076*	0.032
Secondary	−0.026	−0.055*	−0.016
Tertiary	0.052*	−0.105*	0.140**
Residential	−0.426**	−0.143**	−0.051*
Unclassified	0.155*	−0.021	0.010
Service	0.588**	0.175*	−0.187**
Link	−0.062**	0.288**	−0.067**

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

generally simple and well organized so that there are fewer topological errors. In contrast, Link roads are positively correlated with logical consistency (0.288**) as they refer to slip roads or ramps in complicated network structures where topological errors are more likely to occur. As for Service roads located in industrial estates, camp sites, and parking lots where numerous data contributors are expected, the negative correlation (−0.187**) indicates a higher positional accuracy. This corresponds to the common assertion that frequent user activities improve positional accuracy (Mullen *et al*. 2014). It is particularly interesting that the volunteer contribution of Service roads has a negative effect on spatial completeness (R = 0.588**). Indeed, this indicates a low consistency between OSM and reference data, rather than the incompleteness of Service roads in the OSM dataset because Service roads are mostly beyond the scope of authoritative cartography.

The varying data qualities of different road types suggest that road type can be used as a quality indicator independent from extrinsic data. Specifically, it could be classified as a comparison to an intrinsic grounding in a rule gained from OSM data itself according to the grounding-based ontology (Mocnik *et al*. 2018). The findings can also provide detailed quality control instructions for what kind of road should be given more attention in terms of spatial completeness, logical consistency, and positional accuracy.

6. Conclusions

A perception of lack of reliability has seriously impeded the adoption of VGI linear features, so a thorough quality study can facilitate the applicable mechanisms for correcting problematic data and quality assurance. To address this issue, this study proposed a comprehensive quality assessment framework for VGI linear features, which adopts factor analysis to integrate two novel quality metrics with six commonly used metrics, and further examines the spatial autocorrelation and semantic correlation of VGI linear feature quality. The main conclusions can be summarized as follows:

(1) The proposed metrics, Box-counting dimension difference and Link accuracy are efficient in detecting the quality as supplements to the commonly used quality metrics. This can reduce the uncertainty of a single metric so as to more precisely quantify the degree of data incompatibility for the reference-based approach.

(2) Spatial autocorrelation was shown to exist in all three quality elements, namely spatial completeness, positional accuracy, and logical consistency. Highly complete VGI linear features are clustered in urban areas and there are fewer in rural areas, while the opposite is true for logical consistency. As for positional accuracy, there are fewer and smaller clusters than the other two quality elements as these are approximately random errors.

(3) Road type has been proven to be a typical quality indicator of VGI linear features. We demonstrated that Tertiary, Residential, Service and Link roads are correlated with different quality elements, but the sign of the correlation coefficient varies.

Although part of the quality metrics are reference-based, the spatial clustering characteristic and the influence of road type indicate the data quality without the use of authoritative data. The findings could help with designing detailed quality control strategies facing different quality issues. In particular, the grid-based quality assessment results can help us to quickly filter out cells with no or few quality issues and select typical areas where problematic data are clustered. Accordingly, data correction can be arranged and performed focusing on those typical cells rather than the entire study area. This can greatly reduce the cost of checking each road individually, and therefore more efficiently improve the data quality.

Despite the achievements in this study, there are several aspects that deserve further investigation. Besides road networks, our proposed comprehensive quality assessment framework can also be applied to sidewalks, buildings, points of interest, and other features from VGI datasets, although new quality metrics may need to be developed. VGI projects such as Wikimapia, Google MyMaps, Map Insight and Flickr also provide considerable and valuable data in different formats and deserve research on their quality issues in the future. In addition, the source code can be implemented in an open environment, such as QGIS, that allows more people to use and possibly extend the proposed quality assessment framework.

Acknowledgments

The authors are grateful to the Western Pennsylvania Regional Data Center (data.wprdc.org) for providing reference road data of Allegheny County. The authors would like to express special thanks to all the anonymous reviewers and the editor for their comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China[42071358, 41671406];Scientific Research Project of Department of Natural Resources of Hubei Province [ZRZY2020KJ01];Fundamental Research Funds for the Central Universities[CCNU20TS035].

Notes on contributors

Hao Wu is currently a professor at Central China Normal University. His research interests include GIScience, remote sensing and GNSS, focusing on land use and cover change, spatio-temporal data analysis and mining, and volunteered geographic information.

Anqi Lin is a Ph. D. candidate at Central China Normal University and her research interests focus on spatio-temporal data analysis and mining, and volunteered geographic information.

Keith C. Clarke is currently a professor at University of California Santa Barbara. His research interests are environmental simulation modeling, urban growth using cellular automata, terrain mapping and analysis, and real-time visualization.

Wenzhong Shi is the Head and Chair Professor in the Department of Land Surveying and Geoinformatics at the Hong Kong Polytechnic University. His research interests include GIScience and remote sensing, focusing on uncertainties and quality control of spatial data, satellite images and LiDAR data, 3D modeling, and human dynamics.

Abraham Cardenas-Tristan is currently a professor at Autonomus University of San Luis Potosí, San Luis, Mexico. His current research interests include point cloud processing and deformation monitoring of mine slope.

Zhenfa Tu is currently an associate professor at Central China Normal University. His research interests include GIScience and software development, focusing on spatial data organization and management.

ORCID

Hao Wu http://orcid.org/0000-0001-5751-7885
Keith C. Clarke http://orcid.org/0000-0001-5805-6056
Wenzhong Shi http://orcid.org/0000-0002-3886-7027

Data and codes availability statement

The data and codes that support the findings of this study are available in 'figshare.com' with the identifier at the permanent link: https://doi.org/10.6084/m9.figshare.11815533.v1.

References

Antoniou, V. and Skopeliti, A., 2015. Measures and indicators of vgi quality: an overview. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W5, 345–351. doi:10.5194/isprsannals-II-3-W5-345-2015

Barrington-Leigh, C. and Millard-Ball, A., 2017. The world's user-generated road map is more than 80% complete. *Plos One*, 14 (10), e0224742. doi:10.1371/journal.pone.0224742

Barron, C., Neis, P., and Zipf, A., 2014. A comprehensive framework for intrinsic openstreetmap quality analysis. *Transactions in Gis*, 18 (6), 877–895. doi:10.1111/tgis.12073

Basiri, A., *et al.*, 2019. Crowdsourced geospatial data quality: challenges and future directions. *International Journal of Geographical Information Science*, 33 (8), 1588–1593. doi:10.1080/13658816.2019.1593422

Bergman, C. and Oksanen, J., 2016. Conflation of OpenStreetMap and mobile sports tracking data for automatic bicycle routing. *Transactions in Gis*, 20 (6), 848–868.

Camboim, S., Bravo, J., and Sluter, C., 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. doi:10.3390/ijgi4031366

Chehreghan, A. and Abbaspour, R.A., 2018. An evaluation of data completeness of VGI through geometric similarity assessment. *International Journal of Image and Data Fusion*, 9 (4), 319–337. doi:10.1080/19479832.2018.1504825

Ciepłuch, B., *et al*., 2010. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences*. Leicester, England.

Exel, M., Dias, E., and Fruijtier, S., 2010. The impact of crowdsourcing on spatial data quality indicators. *In Proceedings of GiScience 2011*, 14–17. *Zurich*, Switzerland.

Fairbairn, D. and Al-Bakri, M., 2013. Using geometric properties to evaluate possible integration of authoritative and volunteered geographic information. *ISPRS International Journal of Geo-Information*, 2 (2), 349–370. doi:10.3390/ijgi2020349

Fogliaroni, P., D'Antonio, F., and Clementini, E., 2018. Data trustworthiness and user reputation as indicators of VGI quality. *Geo-Spatial Information Science*, 21 (3), 213–233. doi:10.1080/10095020.2018.1496556

Fonte, C.C., *et al*., 2015. Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science*, 29 (7), 1269–1291. doi:10.1080/13658816.2015.1018266

Forghani, M. and Delavar, M.R., 2014. A quality study of the OpenStreetMap dataset for Tehran. *ISPRS International Journal of Geo-Information*, 3 (2), 750–763. doi:10.3390/ijgi3020750

Gardner, Z. and Mooney, P., 2018. Investigating gender differences in OpenStreetMap activities in Malawi: a small case-study. *AGILE Conference*. Lund, Sweden.

Girres, J.F. and Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in Gis*, 14 (4), 435–459. doi:10.1111/j.1467-9671.2010.01203.x

Goodchild, M. and Gopal, S., 1989. The accuracy of spatial database. London: Taylor and Francis

Goodchild, M., Haining, R., and Wise, S., 1992. Integrating GIS and spatial data analysis: problems and possibilities. *International Journal of Geographical Information Systems*, 6 (5), 407–423. doi:10.1080/02693799208901923

Goodchild, M.F., 2007. Citizens as Sensors: the world of volunteered geography. *Geojournal*, 69 (4), 211–221. doi:10.1007/s10708-007-9111-y

Goodchild, M.F. and Glennon, J.A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3 (3), 231–241. doi:10.1080/17538941003759255

Goodchild, M.F. and Hunter, G.J., 1997. A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science*, 11 (3), 299–306. doi:10.1080/136588197242419

Goodchild, M.F. and Li, L.N., 2012. Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110–120. doi:10.1016/j.spasta.2012.03.002

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning. B, Planning & Design*, 37, 682–703. doi:10.1068/b35097

Haklay, M., 2016. *Why is participation inequality important?* London: Ubiquity Press.

Hashemi, P. and Abbaspour, R.A., 2015. *Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept*. Switzerland: Springer International Publishing Switzerland.

Hong, Y. and Yao, Y., 2019. Hierarchical community detection and functional area identification with OSM roads and complex graph theory. *International Journal of Geographical Information Science*, 33 (8), 1569–1587. doi:10.1080/13658816.2019.1584806

Hunter, G.J. 1999. New tools for handling spatial data quality: moving from academic concepts to practical reality. *URISA 1999 Conference*. Chicago, 167–180.

ISO, 2002. ISO 19113: 2002. Geographic information — quality principles. Geneva: ISO

Jackson, S., *et al*., 2013. Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS International Journal of Geo-Information*, 2 (2), 507. doi:10.3390/ijgi2020507

Kalantari, M., and La, V. 2015. Assessing OpenStreetMap as an open property map. In: J. Jokar Arsanjani, A. Zipf, P. Mooney, and M. Helbich , eds. OpenStreetMap in GIScience, Lecture Notes in Geoinformation and Cartography. Cham: Springer, 255–272.

Keßler, C. and Groot, R.T.A.D., 2013. Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In: D. Vandenbroucke, B. Bucher, and J. Crompvoets, eds.. Geographic Information Science at the Heart of Europe. Lecture Notes in Geoinformation and Cartography. Cham: Springer, 21–37

Koukoletsos, T., Haklay, M., and Ellul, C., 2012. Assessing data completeness of VGI through an automated matching procedure for linear data. *Transactions in Gis*, 16 (4), 477–498. doi:10.1111/j.1467-9671.2012.01304.x

Kuhn, H.W. and Kuenne, K.E., 1962. An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics. *Journal of Regional Science*, 4 (2), 21–34.

Lee, M., *et al*., 2017. Morphology of travel routes and the organization of cities. *Nature Communications*, 8 (1), 2229. doi:10.1038/s41467-017-02374-7

Lin, A., *et al*., 2020. A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster. *International Journal of Disaster Risk Reduction*, 49, 101682.

Ma, D., Sandberg, M., and Jiang, B., 2015. Characterizing the heterogeneity of the OpenStreetMap data and community. *ISPRS International Journal of Geo-Information*, 4 (2), 535–550. doi:10.3390/ijgi4020535

Manandhar, P., *et al*., 2019. Towards automatic extraction and updating of VGI-based road networks using deep learning. *Remote Sensing*, 11 (9), 1012–1122. doi:10.3390/rs11091012

Mobasheri, A., *et al*., 2017. Are crowdsourced datasets suitable for specialized routing services? Case study of openstreetmap for routing of people with limited mobility. *Sustainability*, 9 (6), 997. doi:10.3390/su9060997

Mobasheri, A., *et al*., 2018a. Enrichment of OpenStreetMap data completeness with sidewalk geometries using data mining techniques. *Sensors*, 18 (2), 509. doi:10.3390/s18020509

Mobasheri, A., Zipf, A., and Francis, L., 2018b. OpenStreetMap data quality enrichment through awareness raising and collective action tools-experiences from a European project. *Geo-Spatial Information Science*, 21 (3), 234–246. doi:10.1080/10095020.2018.1493817

Mocnik, F.-B., *et al*., 2018. A grounding-based ontology of data quality measures. *Journal Of Spatial Information Science*, 16, 1–25.

Moeinaddini, M., Asadi-Shekari, Z., and Shah, M.Z., 2014. The relationship between urban street networks and the number of transport fatalities at the city level. *Safety Science*, 62, 114–120. doi:10.1016/j.ssci.2013.08.015

Mullen, W.F., *et al*., 2014. Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. *Geojournal*, 80, 587–605. doi:10.1007/s10708-014-9564-8

Neis, P. and Zielstra, D., 2014. Recent developments and future trends in volunteered geographic information research: the case of OpenStreetMap. *Future Internet*, 6 (1), 76–106. doi:10.3390/fi6010076

Neis, P., Zielstra, D., and Zipf, A., 2012. The street network evolution of crowdsourced maps: openStreetMap in Germany 2007–2011. *Future Internet*, 4 (1), 1–21. doi:10.3390/fi4010001

Neis, P., Zielstra, D., and Zipf, A., 2013. Comparison of volunteered geographic information data contributions and community development for selected world regions. *Future Internet*, 5 (2), 282–300. doi:10.3390/fi5020282

OSM, 2020a. *Commercial OSM Software and Services* [online]. Available from: https://wiki.openstreetmap.org/wiki/Commercial_OSM_Software_and_Services [Accessed 20 May 2020].

OSM, 2020b. Quality assurance [online]. Available from: https://wiki.openstreetmap.org/wiki/Quality_assurance [Accessed 20 May 2020]

Safra, E., *et al*., 2010. Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets. *International Journal of Geographical Information Science*, 24 (1), 69–106. doi:10.1080/13658810802275560

Sehra, S.S., *et al*., 2019. Extending processing toolbox for assessing the logical consistency of OpenStreetMap data. *Transactions in Gis*, 24 (1), 44–71. doi:10.1111/tgis.12587

Sehra, S.S., Singh, J., and Rai, H.S., 2016. Analysing OpenStreetMap data for topological errors. *International Journal of Spatial, Temporal and Multimedia Information Systems*, 1 (1), 87–101. doi:10.1504/IJSTMIS.2016.076800

Senaratne, H., *et al*., 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31 (1), 139–167. doi:10.1080/13658816.2016.1189556
Seo, S., O'Hara, and O'Hara, C.G., 2009. Quality assessment of linear data. *International Journal of Geographical Information Science*, 23 (12), 1503–1525. doi:10.1080/13658810802231456
Severinsen, J., *et al*., 2019. VGTrust: measuring trust for volunteered geographic information. *International Journal of Geographical Information Science*, 33 (8), 1683–1701. doi:10.1080/13658816.2019.1572893
Sreelekha, M.G., Krishnamurthy, K., and Anjaneyulu, M.V.L.R., 2017. Fractal Assessment of Road Transport System. *European Transport*, 65, 5.
Stark, H.J., 2010. Quality Assessment of Volunteered Geographic Information (VGI) Based on Open Web Map Services and ISO/TC 211 19100-Family Standards. FOSS4G. Barcelona, Spain.
Truong, Q.T., de Runz, C., and Touya, G., 2019. Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. *International Journal of Geographical Information Science*, 33 (8), 1651–1682. doi:10.1080/13658816.2018.1556395
Vauglin, F., 1997. *Modèles statistiques des imprécisions géométriques des objets géographiques linéaires*. PhD. Université de Marne-la-Vallée.
Wu, H., *et al*., 2014. Assessing the effects of land use spatial structure on urban heat islands using HJ-1B remote sensing imagery in Wuhan, China. *International Journal of Applied Earth Observation and Geoinformation*, 32, 67–78. doi:10.1016/j.jag.2014.03.019
Wu, H., *et al*., 2019. Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. *International Journal of Geographical Information Science*, 33 (5), 1040–1061. doi:10.1080/13658816.2019.1568441
Xie, F. and Levinson, D., 2007. Measuring the structure of road networks. *Geographical Analysis*, 39 (3), 336–356. doi:10.1111/j.1538-4632.2007.00707.x
Yan, Y.W., *et al*., 2020. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science*, 34 (9), 1765–1791. doi:10.1080/13658816.2020.1730848
Yang, C., *et al*., 2020. Big spatiotemporal data analytics: a research and innovation frontier. *International Journal of Geographical Information Science*, 34 (6), 1075–1088. doi:10.1080/13658816.2019.1698743
Zhang, W.-B., Leung, Y., and Ma, J.-H., 2019. Analysis of positional uncertainty of road networks in volunteered geographic information with a statistically defined buffer-zone method. *International Journal of Geographical Information Science*, 33 (9), 1807–1828. doi:10.1080/13658816.2019.1606430
Zhang, Y., *et al*., 2015. Density and diversity of OpenStreetMap road networks in China. *Journal of Urban Management*, 4 (2), 135–146. doi:10.1016/j.jum.2015.10.001
Zheng, Y. and Izzat, I.H., 2018. Exploring OpenStreetMap capability for road perception. *2018 IEEE Intelligent Vehicles Symposium*. Changshu, Suzhou, China, 1438–1443.
Zheng, Y., Izzat, I.H., and Hansen, J.H.L., 2019. Exploring OpenStreetMap availability for driving environment understanding. *preprint*, arXiv:1903.04084.
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), 257–281. doi:10.1080/13658816.2017.1395883
Zhou, Q. and Tian, Y., 2018. The use of geometric indicators to estimate the quantitative completeness of street blocks in OpenStreetMap. *Transactions in GIS*, 22 (6), 1550–1572. doi:10.1111/tgis.12486
Zielstra, D., Hochmair, H.H., and Neis, P., 2013. Assessing the effect of data imports on the completeness of OpenStreetMap – A United States case study. *Transactions in GIS*, 17 (3), 315–334. doi:10.1111/tgis.12037
Zielstra, D. and Zipf, A., A comparative study of proprietary geodata and volunteered geographic information for Germany. ed. *Proceedings of the Thirteenth AGILE International Conference on Geographic Information Science*, 2010 Guimarães, Portugal.