A project report on

Assessing the quality of OpenStreetMap building data in Singapore

Submitted by

CHEN WAI HOONG

In partial fulfilment of the requirements for the award of the degree of

MASTER OF SCIENCE IN APPLIED GEOGRAPHIC INFORMATION SYSTEMS

Under the supervision of

DR. FILIP BILJECKI

GE6226 GIS Research Project

Department of Geography

Faculty of Arts and Social Sciences

National University of Singapore

July 2020

Acknowledgement

First and foremost, I would like to express my gratitude and thank my project advisor, Dr. Filip Biljecki, for his willingness to supervise my work. This project would not have been possible without his incredible guidance and advice.

I would also like to thank Dr. Feng Chen-Chieh, the module facilitator of GE6226, for assisting with all my queries and being such a supportive professor. I am also grateful to all my university professors, lecturers, and staff for their tremendous support and help during the past two semesters.

In addition, I am thankful to all my wonderful coursemates, groupmates, and friends, who have made my university life a pleasant and unforgettable experience. Also, I would like to convey my sincere appreciation to my good friends, Noah, Li Ming, and Hui En, for always being there for me throughout my time in Singapore.

Most of all, I would like to take this opportunity to thank my lovely parents and family members for being my pillars of support. I would not be able to complete this project without their endless and unconditional support.

Last but not least, special thanks to Daphne for her invaluable insights and encouragement, besides taking the time to proofread this report.

Table of Contents

List of tables by pages

Main report

Appendices

List of figures by pages

Executive summary

As OpenStreetMap is getting increasingly popular due to its open-license nature and collaborative aspect, its data quality is increasingly under scrutiny from many geospatial enthusiasts and scientists. Given that many web services and scientific researchers are relying on OpenStreetMap data as the primary data source, data inaccuracy would cause unforeseen problems. Therefore, it is imperative to assess the quality of OpenStreetMap data in order to identify areas of improvement and to improve the reliability of OpenStreetMap data. While the assessment of OpenStreetMap data quality is an ongoing task in many countries, there is a lack of such assessments in Singapore. Therefore, this study was conducted to address this research gap. Five quality metrics of Housing & Development Board (HDB) buildings were studied and analysed as part of the assessment of OpenStreetMap building data quality in Singapore: completeness, positional accuracy, shape accuracy, orientation accuracy, and attribute accuracy. The results of this study suggest that the completeness of HDB building data in Singapore is close to perfect, with 97.67% of the HDB blocks being mapped in OpenStreetMap. Taking all quality metrics into account, it was concluded from this study that the overall quality of HDB buildings in Singapore is fairly good, with some room for improvement. With regard to improving the overall quality of OpenStreetMap data, this study recommends that the OpenStreetMap community explores building a data quality warning system for its users. In addition, correlation analyses revealed that both the median age of planning areas and the mean age of HDB buildings have weak relationships with the data quality of HDB buildings in Singapore. Furthermore, this study has also found that it is currently not feasible to use attributes of HDB buildings in OpenStreetMap to build semantically rich 3D building models, as these attributes are mostly unfilled.

1. Introduction

1.1 Project concept

Volunteered Geographic Information (VGI) has gained popularity since the emergence of Web 2.0 technology. The VGI concept is associated with crowdsourcing (Goodchild & Glennon, 2010), in which volunteers generate and create geospatial data. This concept has captured the attention of many private businesses and researchers, as most VGI data are freely available on the Internet, while authoritative datasets are expensive and restrictive (Antoniou & Skopeliti, 2015).

While OpenStreetMap is not entirely equivalent to VGI, it is often considered as a representation of VGI. Goodchild (2007) listed OpenStreetMap as one of the many VGI projects to have emerged recently, and it has grown into one of the most extensive open-license mapping databases in the world. As of July 2020, there are 6 million members in OpenStreetMap (https://wiki.openstreetmap.org/wiki/Stats). Besides being a free geospatial data repository, OpenStreetMap has also contributed to education and knowledge (Arsanjani, Mooney, Zipf, & Schauss, 2015a).

Due to OpenStreetMap's enormous database and open nature, numerous research projects rely on OpenStreetMap as the primary data source. For instance, Hentschel and Wagner (2010) integrated OpenStreetMap geodata in their study of autonomous robot navigation. On the other hand, the OpenStreetMap database was used in the research of environmental information delivery systems (Ciepluch, Mooney, Jacob, & Winstanley, 2009). Essentially, the versatility and usefulness of OpenStreetMap data in academia cannot be understated.

Since OpenStreetMap is freely accessible to anyone with internet access, the collection and creation of geodata on this platform do not depend on specialists. This characteristic means that thousands of volunteers may contribute to the OpenStreetMap project at any time (Arsanjani, Zipf, Mooney, & Helbich, 2015b), effectively decreasing the costs and time required to produce and update geospatial data in OpenStreetMap.

Moreover, the vast OpenStreetMap database has also led to the creation of hundreds of web services and mobile applications (Amirian, Basiri, Gales, Winstanley, & McDonald, 2015), many of which are crucial in disaster and humanitarian efforts. For example, Rahman, Alam, and Chowdhury (2012) succeeded in designing an affordable disaster evacuation system for Bangladesh. Another example is the usage of OpenStreetMap during the Haitian earthquake in 2010 (Soden & Palen, 2014), in which remote areas mapped by OpenStreetMap volunteers aided the authorities in searching for survivors through crisis mapping.

Nonetheless, OpenStreetMap is not without shortcomings. Due to its crowdsourced nature, OpenStreetMap data may not always be accurate, as data quality depends on user input. A study by Zheng and Zheng (2014) showed that only 66% of OpenStreetMap's data in China are accurate, while a whopping 71% of the data are less detailed when compared to Baidu's datasets. Besides, Zielstra and Zipf (2010) found that the completeness of OpenStreetMap data is lower in rural regions when compared with urban regions in Germany. These demonstrate that OpenStreetMap data is highly heterogeneous.

As the popularity and amount of geospatial data in OpenStreetMap continue to gain traction, many researchers, such as Flanagin and Metzger (2008), have reiterated the importance of establishing methods to assess VGI data quality. Goodchild (2008) suggested that the accuracy of crowdsourced geospatial data needs to be evaluated to address the data quality issues caused by untrained VGI volunteers.

In recent times, OpenStreetMap has become a focus of VGI data quality assessments due to its massive user base and popularity. More so, there are many disaster management services that rely on the database. The quality of data is crucial as it may affect the efficiency of rescue missions. Furthermore, OpenStreetMap's building footprint data are often used to build 3D building models (Fan, Zipf, Fu, & Neis, 2014). Building footprint data that are low in quality and accuracy fail to reflect the actual shape of the buildings, propagating to inaccurate 3D building models. These erroneous models may be used in scientific analyses (energy demand estimation, floor space calculation, etc.) and will lead to unreliable results. These problems highlight the importance of data quality assessment, as well as the need to improve data accuracy in OpenStreetMap.

Given that there is a lack of assessment on OpenStreetMap data quality in Singapore, this project aims to provide a general overview of the quality of OpenStreetMap building data in Singapore.

1.2 Term of reference

1.2.1 Scope of work

The main objective of this project is to assess the quality of OpenStreetMap building data in Singapore using several metrics and quality indicators, through various geospatial analysis methods. As a regional technology hub, Singapore's OpenStreetMap data are highly relevant and widely used. Therefore, it is imperative to assess the quality of OpenStreetMap data in Singapore, as well as to propose recommendations to improve its quality and reliability.

Secondary objectives of this project include understanding whether the quality of OpenStreetMap data is related to socio-economic and other factors, as well as determining the feasibility of using attributes of Singapore's OpenStreetMap buildings in creating semantically rich 3D building models.

1.2.2 Study area

The study area of this project is Singapore. As of now, there are no known studies and researches which focus on the quality of Singapore's OpenStreetMap data.

At first look, most features such as buildings, roads, railways, bus stops, etc. in Singapore are well-mapped in OpenStreetMap. The high level of detail is likely due to Singapore's small landmass and very high population density. Despite this, the degree of quality and accuracy of OpenStreetMap data in Singapore remains unknown.

Many OpenStreetMap buildings in Singapore contain vertical information required to generate basic 3D models, which is taken advantage of by services such as OSM Buildings. OSM Buildings is a library and web map which allows users to visualise OpenStreetMap buildings on 2D and 3D maps. The 3D building models in OSM Buildings were generated using the attributes (floor count, colour of the building, etc.) of buildings in OpenStreetMap.

Due to constraints concerning data availability, budget, and time frame, the focus of this project is on all Housing & Development Board (HDB) buildings in Singapore, which account for most of the residential buildings in the nation. However, the hypothesis is that assessing the quality of this subset of residential buildings in Singapore provides a strong hint at the quality of other buildings as well. Furthermore, this study sets the scene for the quality assessment of buildings nation-wide for future work.

2. Approaches and data availability

2.1 Evaluation of existing methods in assessing the quality of OpenStreetMap data

According to Antoniou and Skopeliti (2015), researchers follow the guidelines set by the International Organization for Standardization in ISO 19157:2013 when it comes to assessing geospatial data quality of VGI using authoritative data. There are six categories of data quality metrics in ISO 19157:2013: completeness, positional accuracy, thematic accuracy, logical consistency, temporal quality, and usability element (ISO, 2013). Nonetheless, in instances where authoritative data was not available, numerous researches have began analysing new quality metrics as proxies to study the quality of VGI data.

Data completeness is arguably the most important quality metric of OpenStreetMap. Hecht, Kunze, and Hahmann (2013) used unit-based elements such as the area and number of buildings to study the completeness of OpenStreetMap data. In another study of data completeness, Mashhadi, Quattrone, and Capra (2015) compared the number of points of interest between OpenStreetMap and ground truth data.

In terms of positional accuracy, the Euclidean distance between the points of OpenStreetMap and reference data were used as a proxy by researchers such as Mashhadi et al. (2015), and Girres and Touya (2010). On the other hand, Koukoletsos, Haklay, and Ellul (2011) calculated the percentage of OpenStreetMap features within the buffer zone of reference data to estimate positional accuracy.

The shape accuracy of OpenStreetMap buildings can be analysed by comparing them with buildings in authoritative data. A high shape accuracy would mean that the buildings in

OpenStreetMap were mapped comprehensively. Fan et al. (2014), and Mooney, Corcoran, and Winstanley (2010) are a few notable researchers who resorted to turning function to analyse the shape accuracy between OpenStreetMap and reference data.

As for semantic accuracy assessment, Fan et al. (2014) analysed the correspondences between buildings in the authoritative reference data and OpenStreetMap data. Contrarily, some researchers looked into the number of elements with specific values as a proxy for semantic accuracy (Arsanjani, Barron, Bakillah, & Helbich, 2013).

Generally, there is no standardised approach in assessing the quality of OpenStreetMap data. The methods mentioned above are non-exhaustive, considering that this is one of the most discussed topics among VGI enthusiasts. Some of the methods outlined above were used in this project and will be described in Section 2.3.

2.2 Data sources

The data for this project were retrieved from a few public sources. The following section outlines the details of all datasets in this project:

(i) OpenStreetMap dataset

The OpenStreetMap dataset of Singapore was downloaded via Overpass Turbo on 26 May 2020. Filters were applied to download polygons representing most buildings located in Singapore, without including any point and line features.

All OpenStreetMap polygons that represent HDB buildings constructed before 2018 were extracted from this data, to align with the reference dataset, which contains only the HDB buildings constructed before 2018. The extraction process will be explained in Section 2.3. The resulting dataset contains 11,806 buildings and would be referred to as “OSM dataset”.

(ii) Reference dataset (hereafter referred to as “HDB dataset”)

The reference dataset, which acts as the ground truth of HDB buildings in Singapore, was extracted from ArcGIS Online. The authoritative source of this dataset is from the Housing and Development Board, and it contains the polygons of all HDB buildings built before 2018. These buildings are located in 31 planning areas in Singapore (Figure 1). The reference dataset contains 11,924 HDB buildings, and the attributes of this dataset include postal codes.

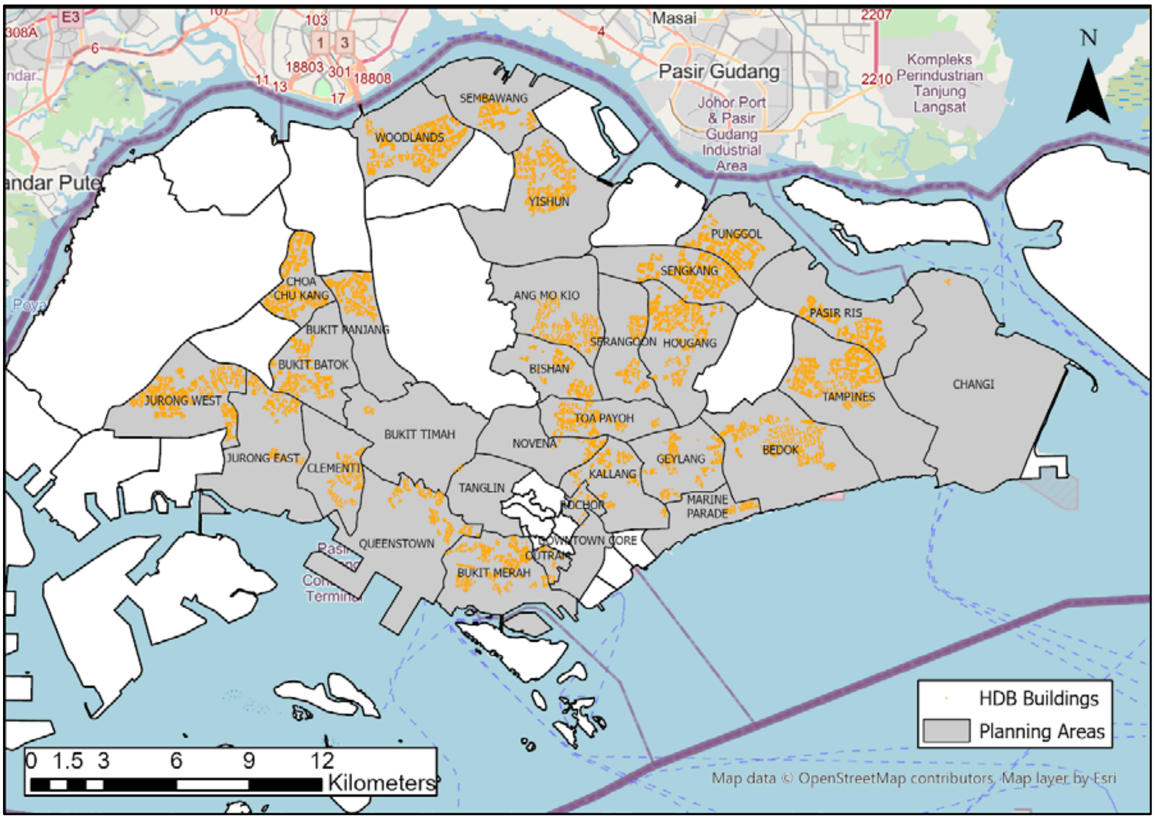

Figure 1. The location of HDB buildings constructed before 2018.

(iii) HDB Property Information

The HDB Property Information spreadsheet was downloaded from Singapore's open data portal (data.gov.sg), and it contains all relevant information about HDB buildings in Singapore, e.g., block number, street name, number of floors, year of completion, etc. However, it does not include postal codes.

(iv) Planning Areas of Singapore

The Planning Areas of Singapore shapefile was downloaded from data.gov.sg. This shapefile was used to display analysis results and to visualise the disparity of OpenStreetMap data quality across planning areas in Singapore.

Furthermore, a 1km by 1km grid map was created. The 1km squared grid map enabled us to identify the spatial patterns of various OpenStreetMap data quality metrics in a more detailed manner.

(v) Population Trends, 2019

The Population Trends 2019 PDF booklet was obtained from the Department of Statistics Singapore website (singstat.gov.sg). It contains the June 2019 population census.

2.3 Project timeline and workflow

Table 1 shows the overall timeline of this project.

Table 1. Overall project timeline.

Date **Task**	**12 March**	**13 March - 30 April**	**1 May - 19 May**	**20 May**	**21 May - 18 June**	**19 June**	**20 June - 19 July**	**20 July**
• First meeting with project advisor - discussion on potential topics and scope of the project								
• Conduct preliminary research and literature review • Assess data availability								
• Finalise project topic and objectives • Conduct literature review • Collect and preprocess data								
• Second meeting with project advisor - discussion on project workflow and methodology								
• Collect and preprocess data • Analyse data								
• Third meeting with project advisor - discussion on analysis results								
• Write the report								
• Project submission deadline								

This project comprises of three major phases, which will be explained in the following section:

(i) Literature review and data collection

Upon deciding on the project topic and study area, papers on quality assessment of OpenStreetMap data were reviewed for a more in-depth understanding of the concept and existing methods. The most suitable quality assessment methods for this project were then determined, and the scope of the project was established with realistic targets while considering critical aspects such as time frame and data availability.

Multiple OpenStreetMap datasets were extracted from various data extraction mediums, for instance, Geofabrik, BBBike, and Overpass Turbo. All downloaded datasets were reviewed to determine the best set of data for this project. The data extracted via Overpass Turbo was chosen as it contains all building attributes from OpenStreetMap. Besides this, HDB dataset and other supporting data were also obtained.

(ii) Data preprocessing

To avoid map projection errors, both OpenStreetMap and HDB datasets were projected to the SVY21 Singapore TM coordinate system (WKID: 3414), to ensure that the buildings from both datasets were positioned correctly on the same map in ArcGIS Pro.

As the OpenStreetMap dataset contains most buildings in Singapore, ArcGIS Pro’s “Select By Location” tool was used to select buildings in this dataset that overlap and correspond to the buildings in HDB dataset (Figure 2). A new feature class containing these selected buildings was then created (hereafter referred to as “OSM dataset”).

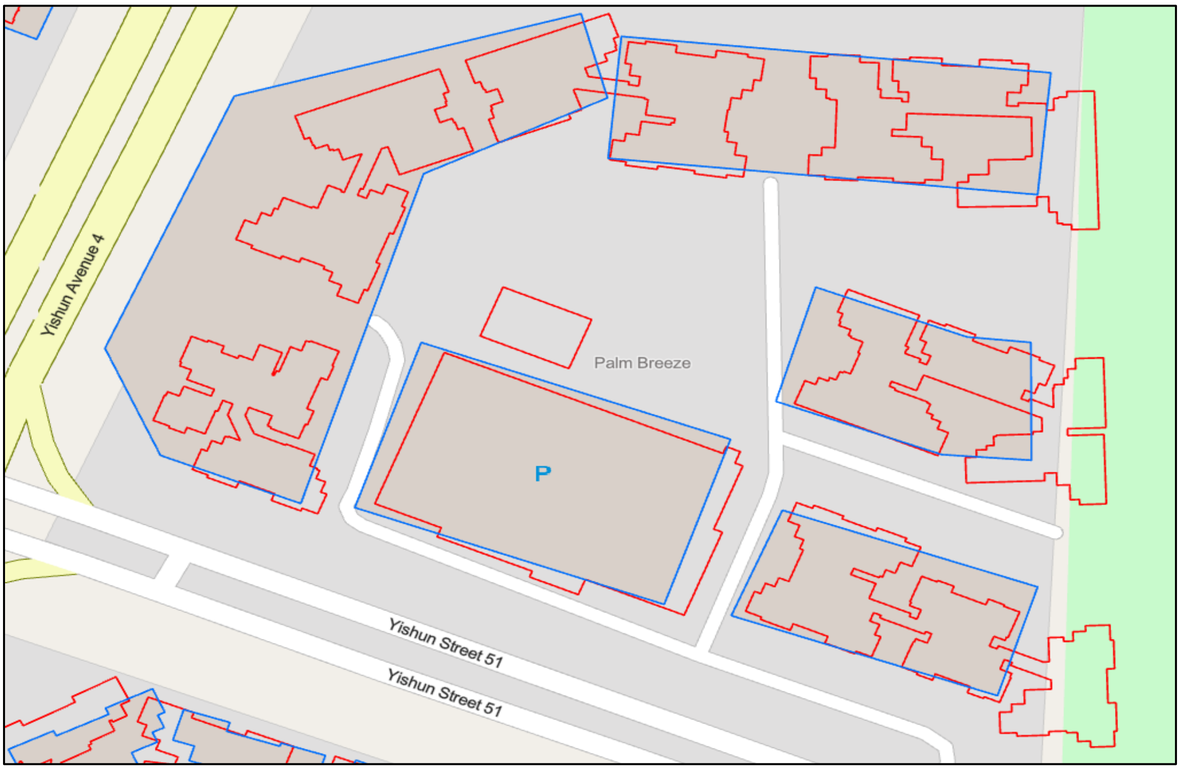

Figure 2. OpenStreetMap buildings (blue) that overlap buildings in the HDB dataset (red).

The OSM dataset was then screened to ensure that only buildings corresponding to the HDB buildings in HDB dataset remain in this dataset. The screening was performed by comparing the shape, orientation, and attributes of both corresponding buildings. It was a straightforward process as the HDB buildings in Singapore are typically constructed in clusters within a designated area and are well-spaced apart.

Postal codes were generated for every street address in the HDB Property Information spreadsheet so that the spreadsheet data can be joined to the HDB dataset before analysing attribute accuracy.

This was done using a Python code published by NUS Urban Analytics Lab in GitHub, which is based on OneMap API provided by OneMap[1].

(iii) Data analysis

In this study, five selected OpenStreetMap quality metrics were assessed, namely completeness, positional accuracy, shape accuracy, orientation accuracy, and attribute accuracy. Apart from these, Getis-Ord Gi* hotspot analyses were performed on positional accuracy, shape accuracy, and orientation accuracy to identify the locations where there are statistically significant clusters of low (coldspots) and high (hotspots) values. The following describes all analyses performed in this study:

(a) Completeness

Before calculating the completeness of HDB buildings in OpenStreetMap, the semantic accuracy was determined, by categorising corresponding buildings in HDB and OSM datasets into five types of semantic relations based on building-to-building cardinalities (Xu, Chen, Xie, & Wu, 2017). Each type of relation (Figure 3) denotes the different degrees of semantic accuracy. Table 2 describes them in detail:

[1] The code used to generate postal codes is provided by NUS Urban Analytics Lab at https://github.com/ualsg/hdb3d-code/blob/master/gc.py

Table 2. Interpretation of each category of relation.

Relation	Interpretation
1:1	• A building in HDB dataset corresponds to a building in OSM dataset • Semantically accurate
1:0	• A building in HDB dataset does not correspond to any building in OSM dataset • Semantically inaccurate
1:n	• A building in HDB dataset corresponds to several (n) buildings in OSM dataset • Semantically partially accurate
n:m	• Several (n) buildings in HDB dataset corresponds to several (m) buildings in OSM dataset • Semantically inaccurate
n:1	• Several (n) buildings in HDB dataset correspond to a building in OSM dataset • Semantically partially accurate

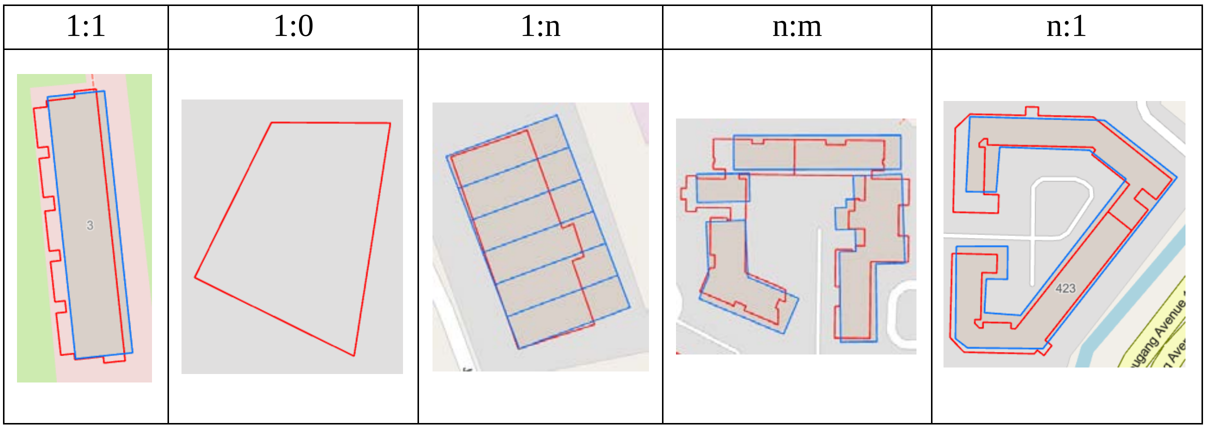

Figure 3. Buildings in HDB (red) and OSM (blue) datasets shown in each category of relation.

Buildings that exist in OpenStreetMap but do not correspond to any building in the reference dataset are usually categorised into 0:1 relation category. However, this relation does not apply to this study. A unique ID was assigned to every corresponding building pairs with 1:1 relation. This is to ensure that no buildings have multiple 1:1 relations; hence each building in OSM dataset corresponds to one unique building in HDB dataset.

The completeness of HDB buildings in Singapore was then calculated using the number of buildings in every semantic relation category; the formula is shown below (Hecht et al., 2013):

$$\text{Completeness} = \frac{Total\ number\ of\ buildings\ in\ OSM\ dataset}{Total\ number\ of\ buildings\ in\ HDB\ dataset}\ x\ 100$$

The breakdown of the calculation of completeness is shown in Appendix A.

(b) Positional accuracy

For every corresponding building pairs in HDB and OSM datasets with 1:1 relation, a pair of centroids were generated. The centroids mark the middlemost point of each building. Using the unique ID assigned to every corresponding HDB and OSM building pairs, a straight polyline was created to connect the centroids between every corresponding building pairs. The length of the polyline represents the distance between each corresponding HDB and OSM buildings, which would be used as the indicator of positional accuracy in OpenStreetMap (Fan et al., 2014).

(c) Shape accuracy

The Hausdorff distance is used as the data quality indicator to assess the shape accuracy between corresponding buildings in HDB and OSM datasets. The Hausdorff distance measures the greatest distance from a point to another closest point between two polylines, and it is commonly proposed as a method to measure shape and line similarity. For example, Girres and Touya (2010) used Hausdorff distance as a proxy to measure the accuracy of roads and coastlines in OpenStreetMap by analysing the maximum deviation between polyline features.

In this assessment, a Hausdorff distance was generated for every building pairs with 1:1 relation using a custom processing tool in QGIS. The unique ID was used by the tool to recognise corresponding buildings in HDB and OSM datasets. The source Python code used to generate Hausdorff distance in the custom processing tool was obtained from a public GitHub repository[2]. Since the code would only work on polylines, all corresponding buildings in HDB and OSM datasets were converted into polylines. In this quality assessment, only the outer perimeter/shape of the buildings was considered. Therefore, all polylines which represent the interior courtyards of buildings were removed.

(d) Orientation accuracy

A minimum bounding rectangle was generated for all buildings in HDB and OSM datasets to estimate the orientation of the buildings. For every minimum bounding rectangle created, the orientation of the longer side of the rectangle was measured in decimal degrees, clockwise from north. The absolute orientation difference between every corresponding HDB and OSM building pairs with 1:1 relation was calculated as an indicator of orientation accuracy.

(e) Attribute accuracy

As explained in Section 2.3, the data from HDB Property Information spreadsheet was joined to HDB dataset to allow the comparison of attributes between the OpenStreetMap and reference datasets. The percentage of attributes that were filled correctly in OpenStreetMap was determined to evaluate attribute accuracy (Girres & Touya, 2010).

---

[2] The code used to generate Hausdorff distance is provided by Anita Graser at https://github.com/anitagraser/QGIS-Processing-tools/blob/master/1.1/scripts/hausdorff_distance_pairwise.py

Since attributes of OpenStreetMap buildings are increasingly used to build 3D building models, the number of buildings that contain attributes useful for building semantically rich 3D models was also analysed.

(f) Correlation analyses

To understand whether the quality of OpenStreetMap data is influenced by different socio-economic indicators such as the median population age across planning areas in Singapore, correlation analyses were performed. Using the data from Population Trends, 2019 PDF booklet, the median age of every planning area was calculated using the following formula:

$$Median = L + w\ \frac{(\frac{n}{2} - c)}{f}$$

where L = lower limit of the age group containing the median

w = width of age group

n = total population

c = cumulative count/frequency up to L

f = count/frequency in the median age group

Similar correlation analyses were also performed against the mean age of HDB buildings of planning areas in Singapore. The mean age of HDB buildings in every planning area was calculated using the “Year Completed” attribute from the HDB Property Information spreadsheet. The full table displaying the median age of population and the mean age of HDB buildings in every planning area can be found in Appendix B.

3. Results

3.1 Discussion

3.1.1 Completeness

The completeness of HDB buildings in OpenStreetMap is 97.67%. There are 275 buildings that exist in the reference dataset but were not mapped in OpenStreetMap (1:0 relation).

Concerning semantic accuracy, the overall results are shown in Table 3. Close to 95% of the buildings in HDB and OSM datasets fall under the 1:1 relation category. The results suggest that the boundaries of most individual HDB buildings can be distinguished from the satellite images in OpenStreetMap, which is likely given that most HDB buildings in Singapore are well-spaced apart.

It was observed that many buildings with 1:0 relation in the reference dataset are rather small HDB buildings such as community pavilions; therefore, they may have been missed by OpenStreetMap contributors or not deemed as buildings.

Besides, the total percentage of buildings in HDB dataset categorised into 1:n, n:m, and n:1 relation categories do not exceed 3%. These HDB buildings are inter-connected; hence the boundary of the buildings may be difficult to distinguish in the satellite images used in OpenStreetMap.

Table 3. Overall results of semantic accuracy.

Relation	HDB dataset		OSM dataset	
	Number of buildings	**Percentage (%)**	**Number of buildings**	**Percentage (%)**
1:1	11312	94.87	11312	95.82
1:0	275	2.31	N/A	N/A
1:n	97	0.81	317	2.69
n:m	135	1.13	132	1.12
n:1	105	0.88	45	0.38
Total	**11924**	**100**	**11806**	**100**

The proportion of buildings in each category of relation across planning areas in Singapore are shown in pie charts in Figure 4 (HDB dataset) and Figure 5 (OSM dataset). In both datasets, Outram, Novena, Punggol, and Rochor have a lower proportion of building with 1:1 relation (<90%) compared to other planning areas. The results of Outram and Rochor may be skewed due to the low number of HDB buildings located in these relatively small planning areas. However, in the case of Novena, it is mostly caused by the semantic inaccuracy of a few rows of landed HDB terrace houses at Jalan Ma’mor. A single row of multiple terrace houses here is considered as a single HDB building in HDB dataset; however, these terrace houses were mapped as individual buildings in OpenStreetMap (Figure 6).

The full results of semantic accuracy broken down into every planning area for both HDB and OSM datasets are displayed in Appendix C and Appendix D.

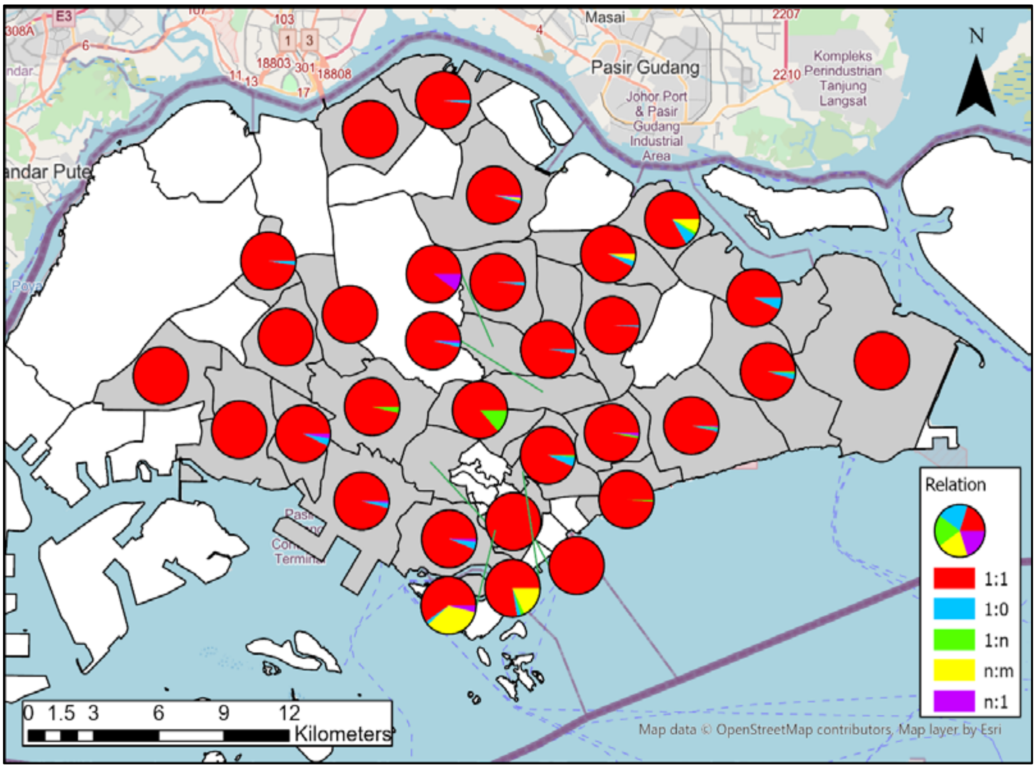

Figure 4. Proportion of buildings (HDB dataset) in each category of relation across planning areas in Singapore.

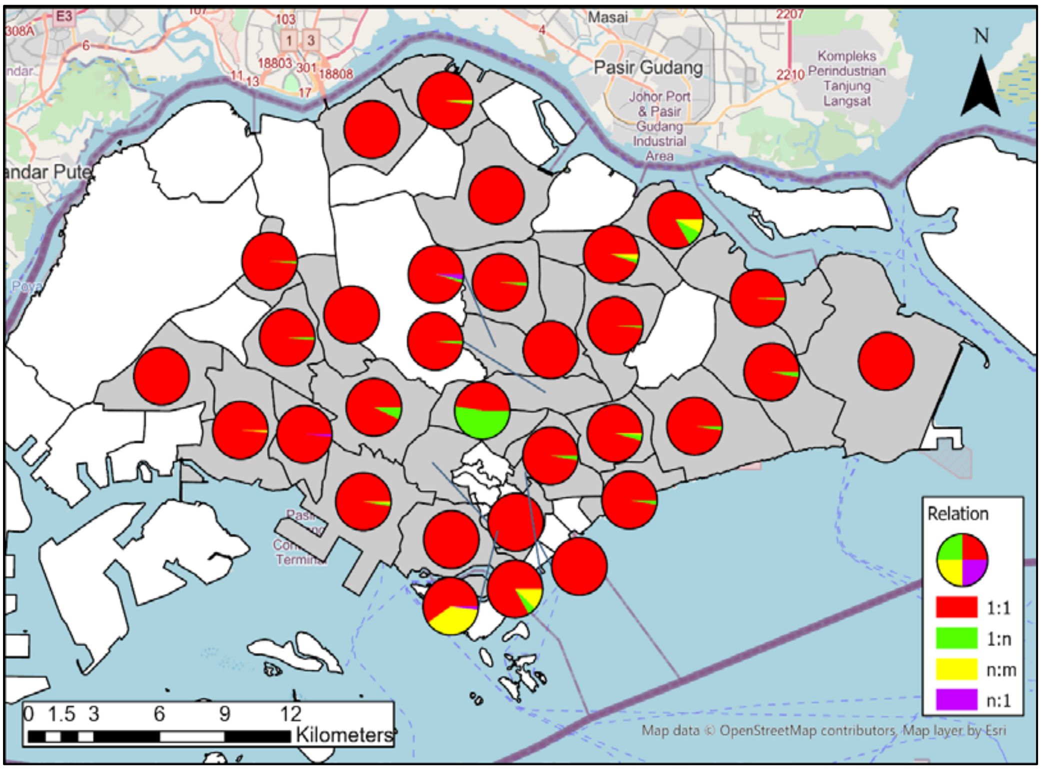

Figure 5. Proportion of buildings (OSM dataset) in each category of relation across planning areas in Singapore.

Figure 6. Landed HDB terrace houses at Jalan Ma'mor.

3.1.2 Positional accuracy

As shown in Table 4, the mean offset distance between corresponding buildings in HDB and OSM datasets in Singapore is 4.06m, while the highest and lowest offset distance is 31.95m and 0.04m, respectively. The results for every planning area in Singapore are shown in Appendix E.

Table 4. Overall results of positional accuracy.

Mean offset distance	4.06m
Standard deviation	2.96m
Median	3.39m
Highest offset distance	31.95m
Lowest offset distance	0.04m

Figure 7 and Figure 8 display the results across planning areas and grid maps in Singapore. Clementi, Bukit Batok, and Bukit Timah are the planning areas with the three highest mean offset distances (>5m). In contrast, Marine Parade and Downtown Core have the two lowest mean offset distances (<2m). The map in Figure 9 shows that statistically significant hotspots are scattered all across the country. Interestingly, in a few planning areas, namely Pasir Ris, Bukit Merah, Clementi, Bukit Batok, and Bukit Timah, the majority of the buildings are considered hotspots.

Given that OpenStreetMap buildings are mainly mapped using Bing satellite images, both HDB and OSM datasets were inspected visually, with Bing satellite image as basemap. It was determined that the offset is likely due to Bing map's low-resolution satellite images, in line with research results discovered by Fan et al. (2014) and Hecht et al. (2013). Furthermore, Brovelli and Zambroni (2018) mentioned that the spatial accuracy of OpenStreetMap data mapped from satellite images is usually similar to the accuracy of the satellite images; hence the accuracy of satellite images may affect the positional accuracy of OpenStreetMap buildings.

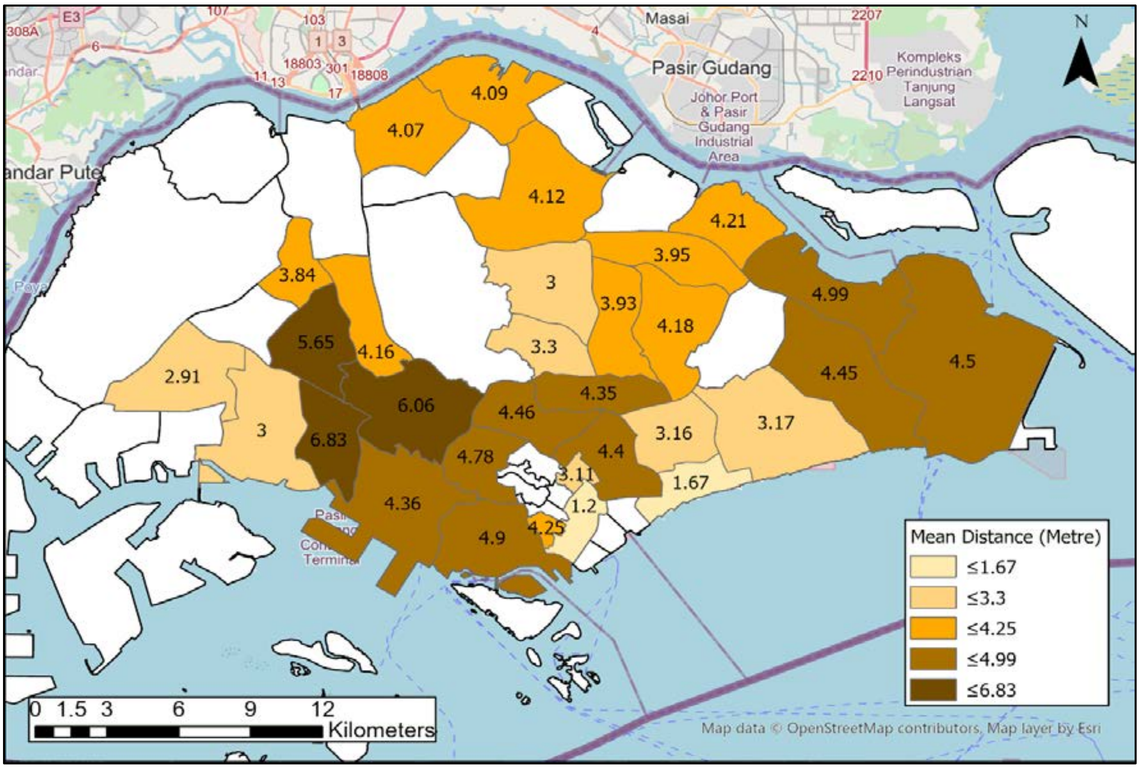

Figure 7. Mean offset distance across planning areas in Singapore.

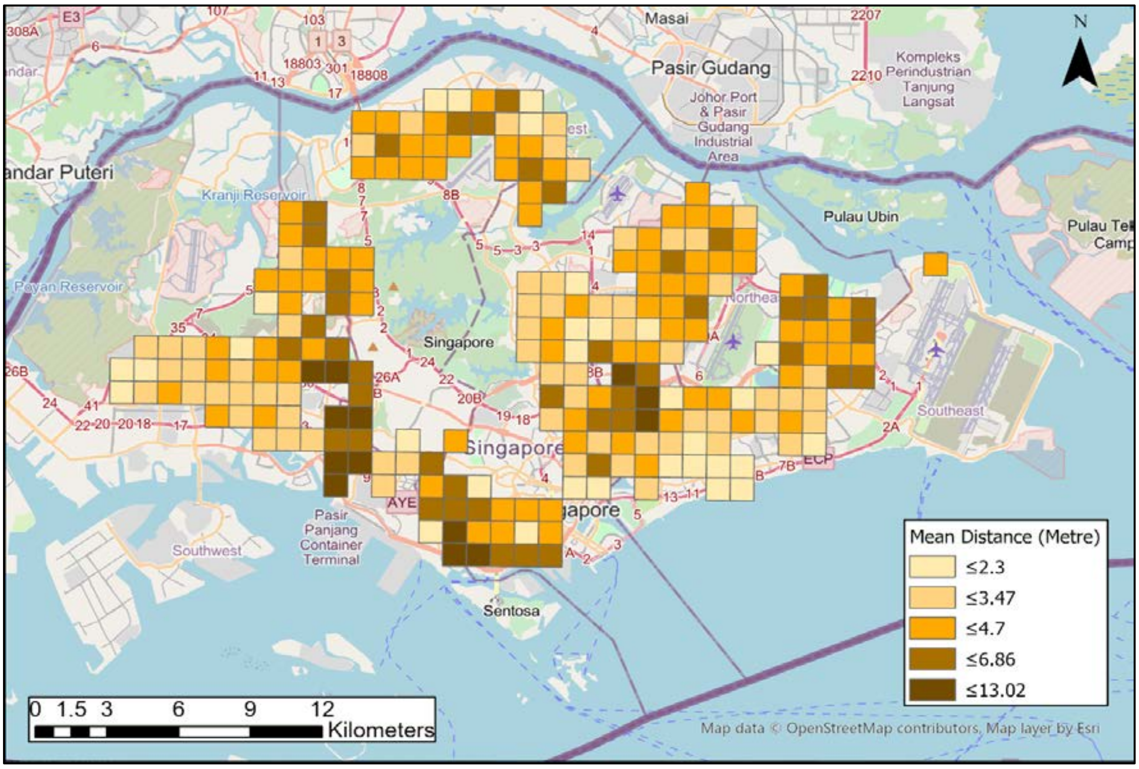

Figure 8. Mean offset distance across 1km squared grids in Singapore.

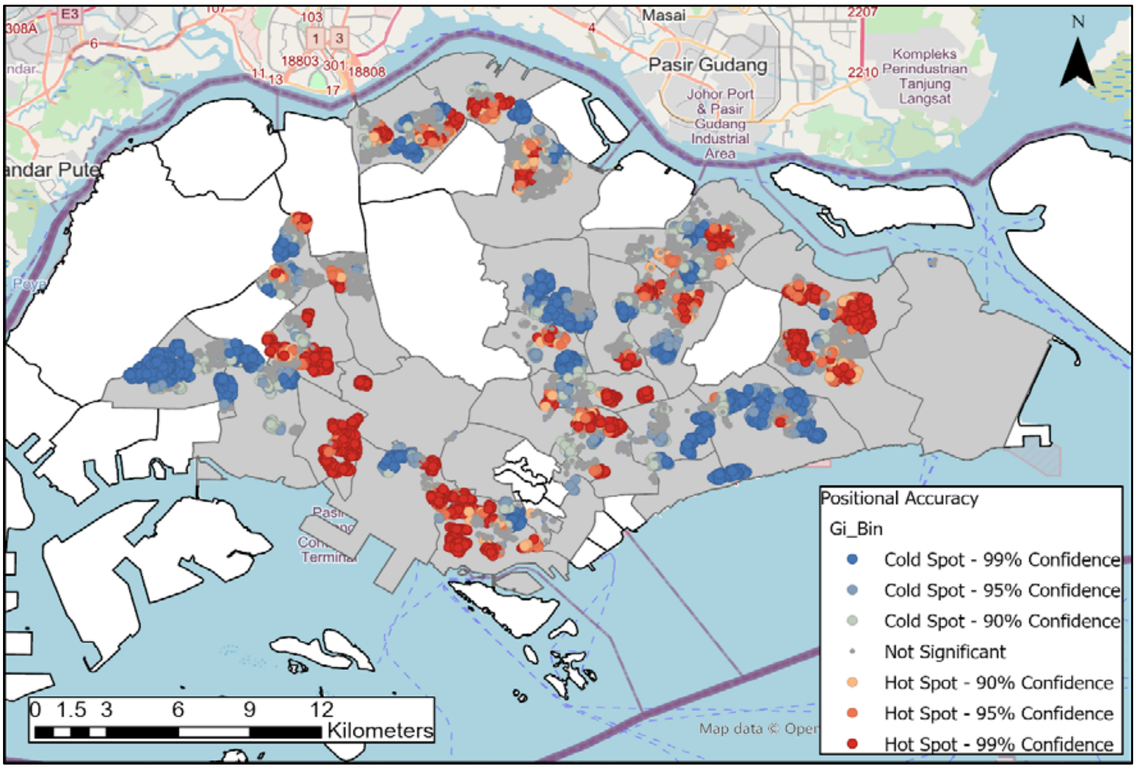

Figure 9. Hotspots and coldspots of offset distance in Singapore.

3.1.3 Shape accuracy

Table 5 shows the overall results of shape accuracy, while the results for every planning area are shown in Appendix F. In terms of shape accuracy of OpenStreetMap buildings, a low Hausdorff distance would mean that the shape accuracy is high and closely resemble the shape of buildings in the reference dataset. On average, the Hausdorff distance for HDB buildings in Singapore is 19.26m.

Table 5. Overall results of shape accuracy.

Mean Hausdorff distance	19.26m
Standard deviation	13.18m
Median	15.43m
Highest Hausdorff distance	146.44m
Lowest Hausdorff distance	0.62m

Figure 10 and Figure 11 depict the mean Hausdorff distance in various regions of Singapore. The maps show that the planning areas surrounding Central Water Catchment and eastern parts of Singapore have a higher mean Hausdorff distance compared to other parts of the country, with Ang Mo Kio and Tanglin coming on top (>30m). Similar to the observation made above, Figure 12 shows that most of the hotspot clusters are located in planning areas where the mean Hausdorff distance is generally higher. Conversely, coldspots are mostly concentrated towards western and north-eastern parts of the country.

It was ascertained that most of the HDB buildings with high Hausdorff distance have complicated building shapes in the reference dataset but were digitised with simplified polygons in OpenStreetMap. Plausible explanation is that some OpenStreetMap contributors may not have the patience to digitise the buildings accurately as it is more time-consuming. Some contributors may digitise the buildings carefully according to the shape of the building, while others may simply create simplified polygons (Husen, Idris, & Ishak, 2018). Additionally, the low-resolution Bing satellite images made digitising complex-shaped buildings accurately a difficult task; moreover, the outline of roofs may be difficult to discern from bird's-eye view (Fan at al., 2014).

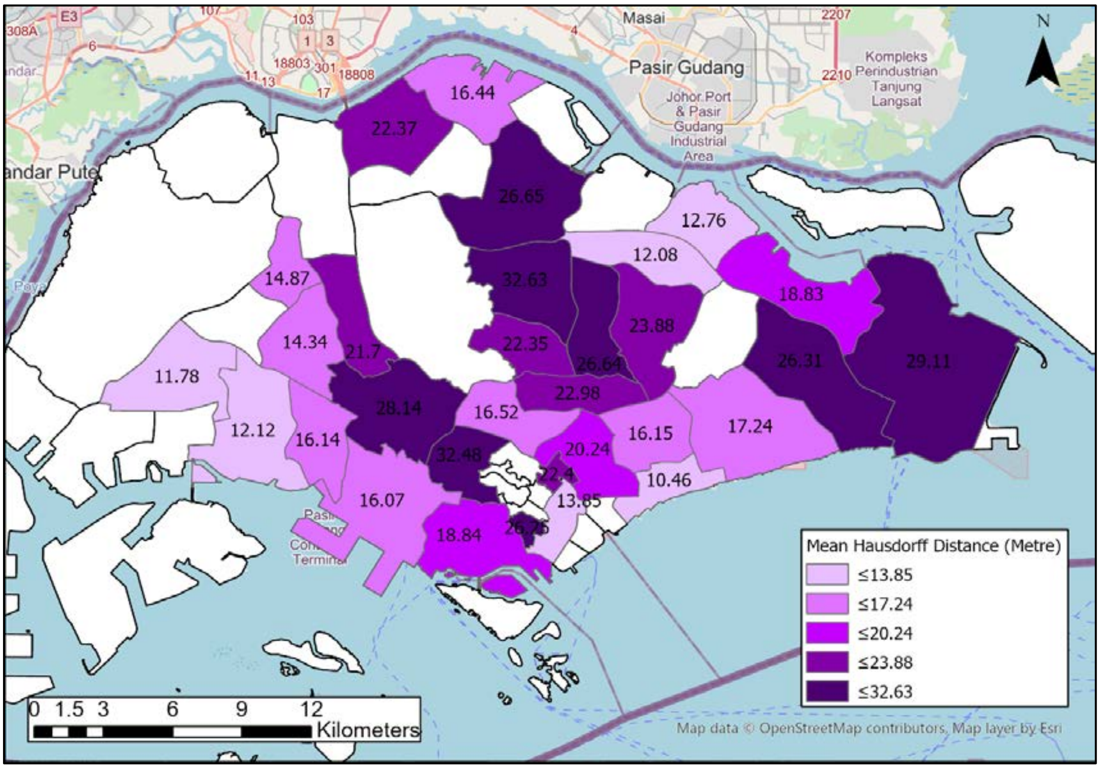

Figure 10. Mean Hausdorff distance across planning areas in Singapore.

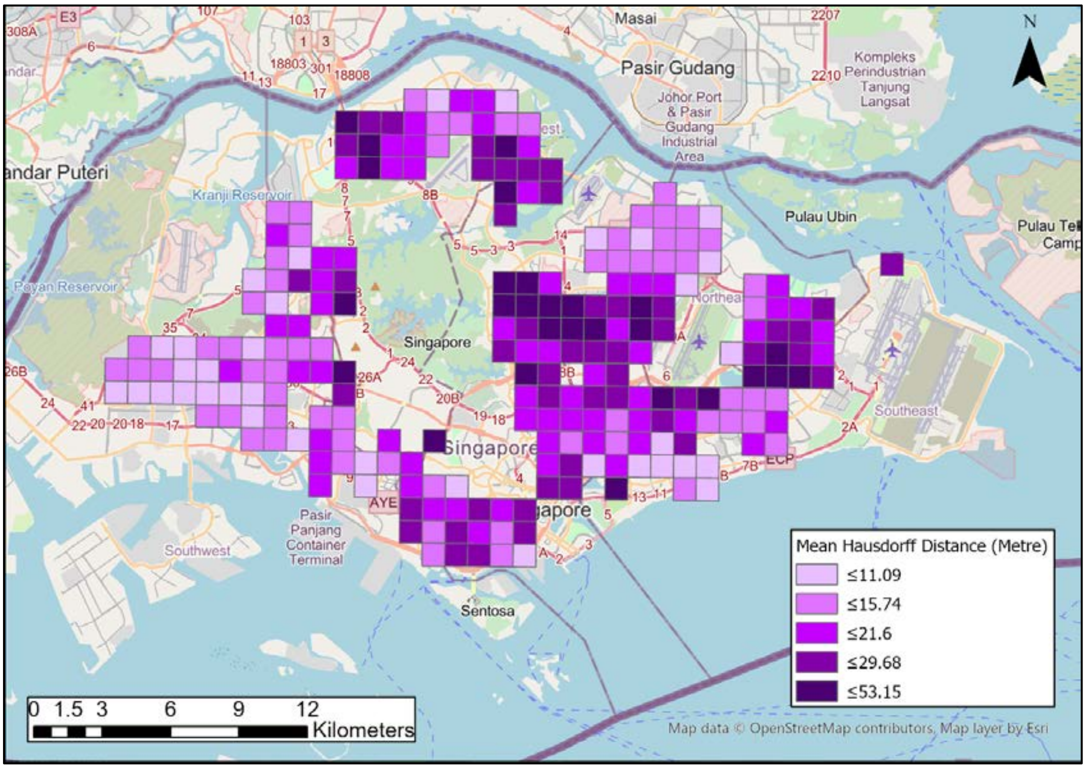

Figure 11. Mean Hausdorff distance across 1km squared grids in Singapore.

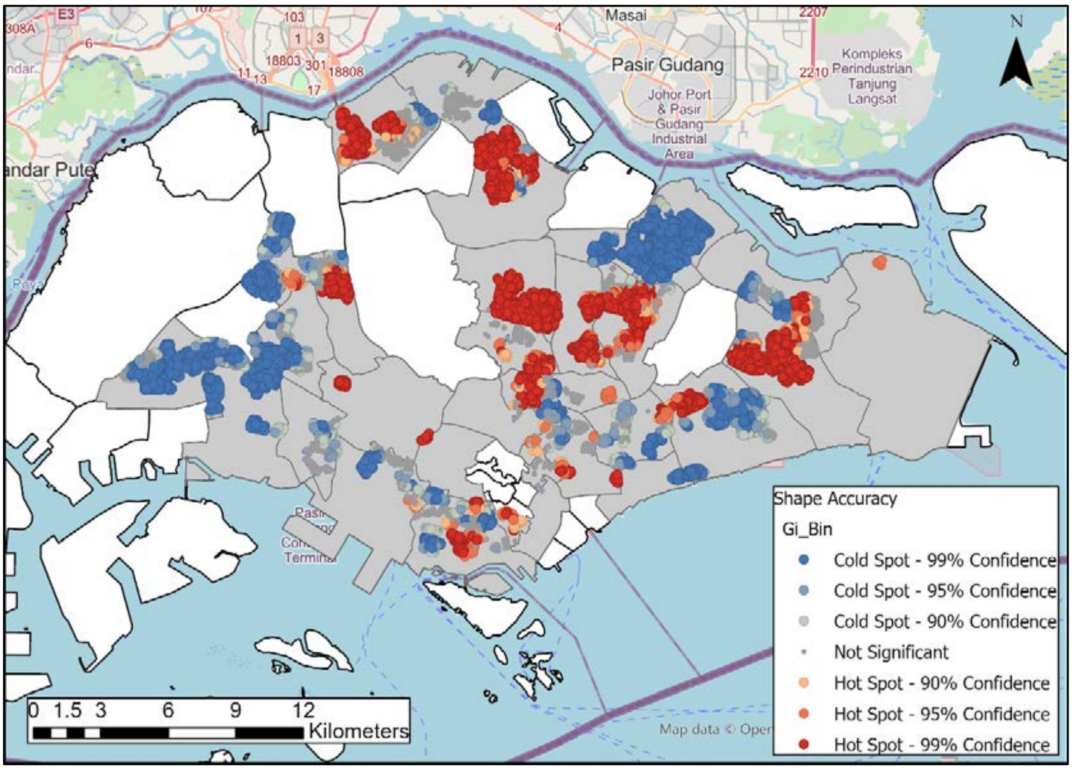

Figure 12. Hotspots and coldspots of Hausdorff distance in Singapore.

3.1.4 Orientation accuracy

Table 6 shows that the mean absolute orientation difference between corresponding buildings is 3.2°, while the difference between the highest and lowest value is 90°. A thorough inspection revealed that about 84% of the corresponding buildings in HDB and OSM datasets have an absolute orientation difference of less than 3°. The mean absolute orientation difference of every planning area in Singapore are displayed in Appendix G.

Table 6. Overall results of orientation accuracy.

Mean absolute difference	3.2°
Standard deviation	9.9°
Median	0.7°
Highest absolute difference	90°
Lowest absolute difference	0°

Figure 13 and Figure 14 display the mean absolute orientation difference across different areas in Singapore. In general, HDB buildings in OSM dataset that are located at the north and north-eastern parts of Singapore have a higher absolute difference in building orientation (>4°) when compared with the HDB dataset. Notably, planning areas located in these regions of Singapore contains some of the newest HDB towns, namely Punggol, Sengkang, and Sembawang. Changi, however, is considered as an outlier in the overall results as there are only seven HDB buildings within the planning area. Overall, statistically significant hotspots can be found in the same regions mentioned above (Figure 15), while coldspots are mostly found in the west of Singapore.

It is not surprising that the highest absolute orientation difference is 90°. Since the orientation of a building was measured using a minimum bounding rectangle, the shape accuracy was also an influential factor. While most HDB buildings in Singapore typically have a long rectangular shape, some exceptions have a shape that resembles a square. It was discovered that some corresponding square-shaped buildings in HDB and OSM datasets have a huge absolute difference in orientation (>80°). The massive difference in orientation angle is likely caused by the shape inaccuracy of buildings mapped by OpenStreetMap contributors. According to Lokhat and Touya (2016), many building footprints with perfectly square angles have been mapped with different angles in a VGI dataset by unskilled contributors. Therefore, a square-shaped building in OpenStreetMap that was digitised with a slight variation in angles or length may be interpreted by the software as a rectangular building that has a perpendicular dominant angle to the dominant angle of the corresponding building from the reference dataset.

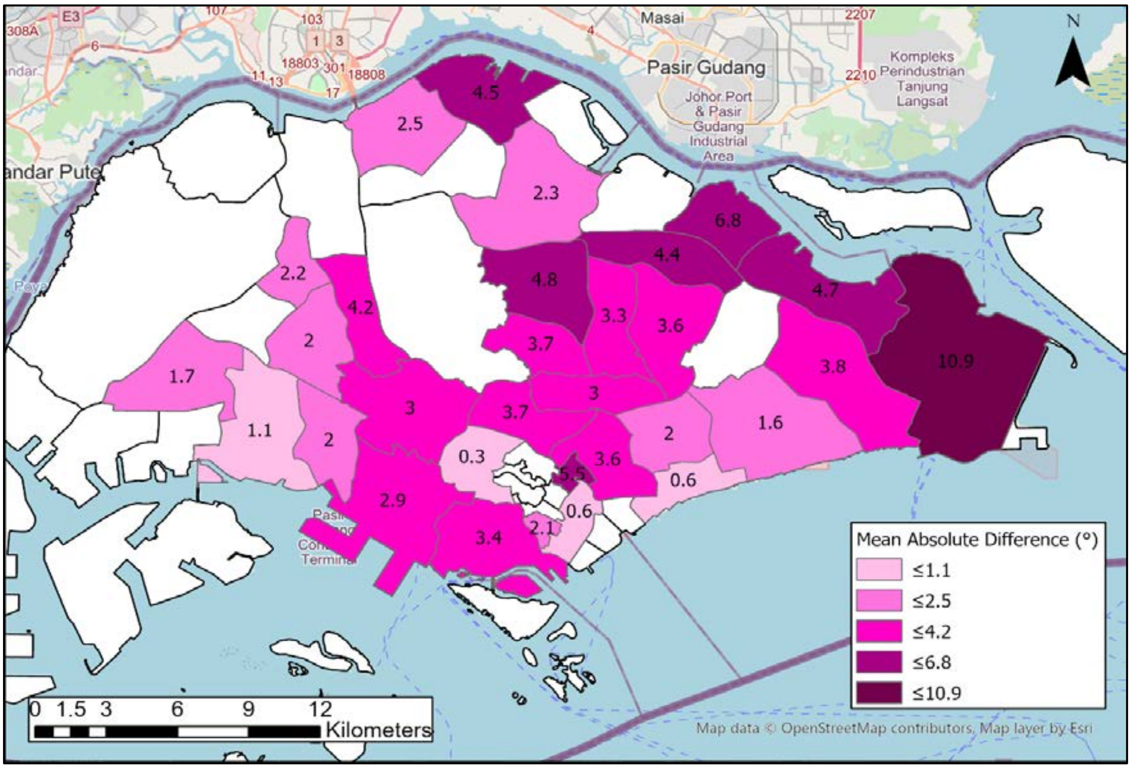

Figure 13. Mean absolute difference across planning areas in Singapore.

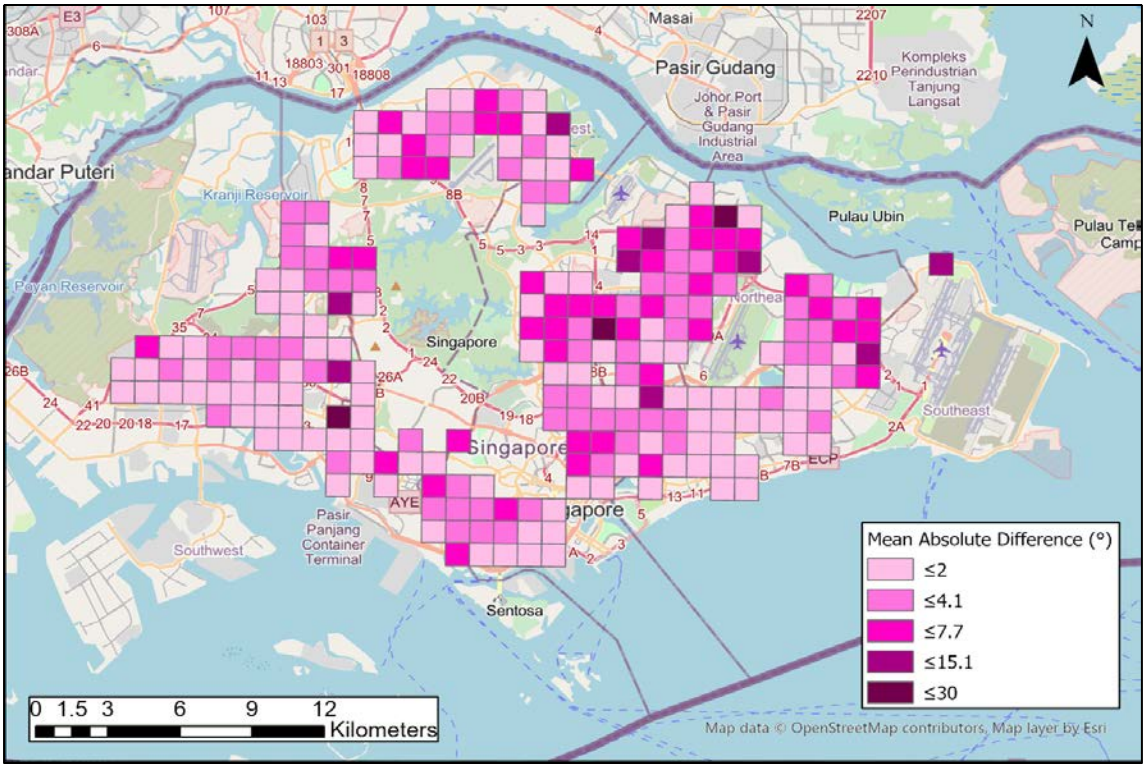

Figure 14. Mean absolute difference across 1km squared grids in Singapore.

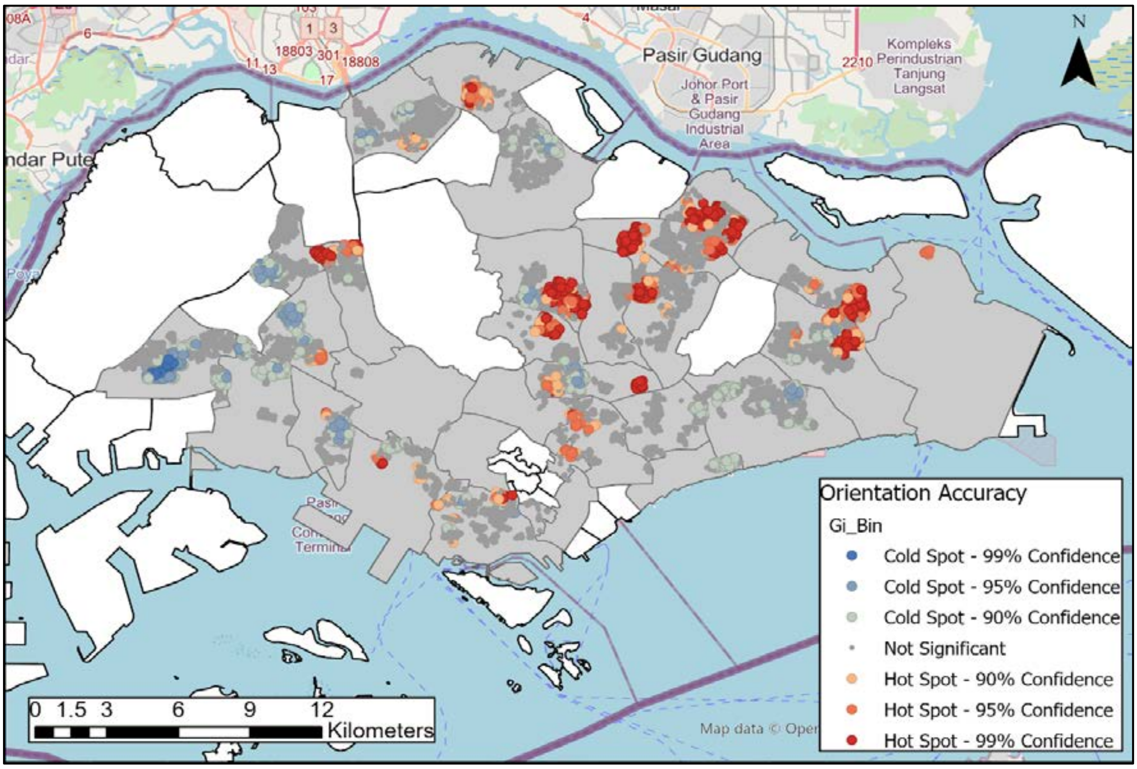

Figure 15. Hotspots and coldspots of absolute orientation difference in Singapore.

3.1.5 Attribute accuracy

The overview of attribute accuracy in OSM dataset is shown in Table 7. The percentage of correct attribute matches of important attribute fields such as postal code, block number, street name, and the number of floors hovers between 45% to 60%, except for the year of completion.

Table 7. Overall results of attribute accuracy. (Total buildings: 11312)

Attribute field	Number of correct attribute matches	Percentage (%)
Postal Code	5810	51.36
Block Number	5402	47.75
Street	6452	57.04
Year Completed	230	2.03
Number of Floors	5247	46.38

The attribute accuracy of every planning areas in Singapore are shown in Appendix H. The bar charts in Figure 16 illustrate the difference in the number of correct attribute matches between the five attribute fields across planning areas in Singapore. While planning areas in eastern Singapore such as Pasir Ris, Tampines, and Bedok have a higher percentage of correct attribute matches, the spatial trend in terms of attribute accuracy appears random for most of Singapore.

According to Camboim, Bravo, and Sluter (2015), attribute completeness in OpenStreetMap relies on individual efforts of contributors to provide attribute details. We can infer from the results table that not all OpenStreetMap contributors were inspired to achieve high attribute accuracy in OpenStreetMap. On the other hand, Arsanjani et al. (2013) stated that “expert mappers” would pay attention to data quality in OpenStreetMaps, and provide accurate semantic information based on their expertise of the area. In this case, we can assume that the HDB buildings with many correct attributes in OpenStreetMap were most likely mapped by contributors who may have excellent local knowledge, and were keen to improve the OpenStreetMap data quality in those areas.

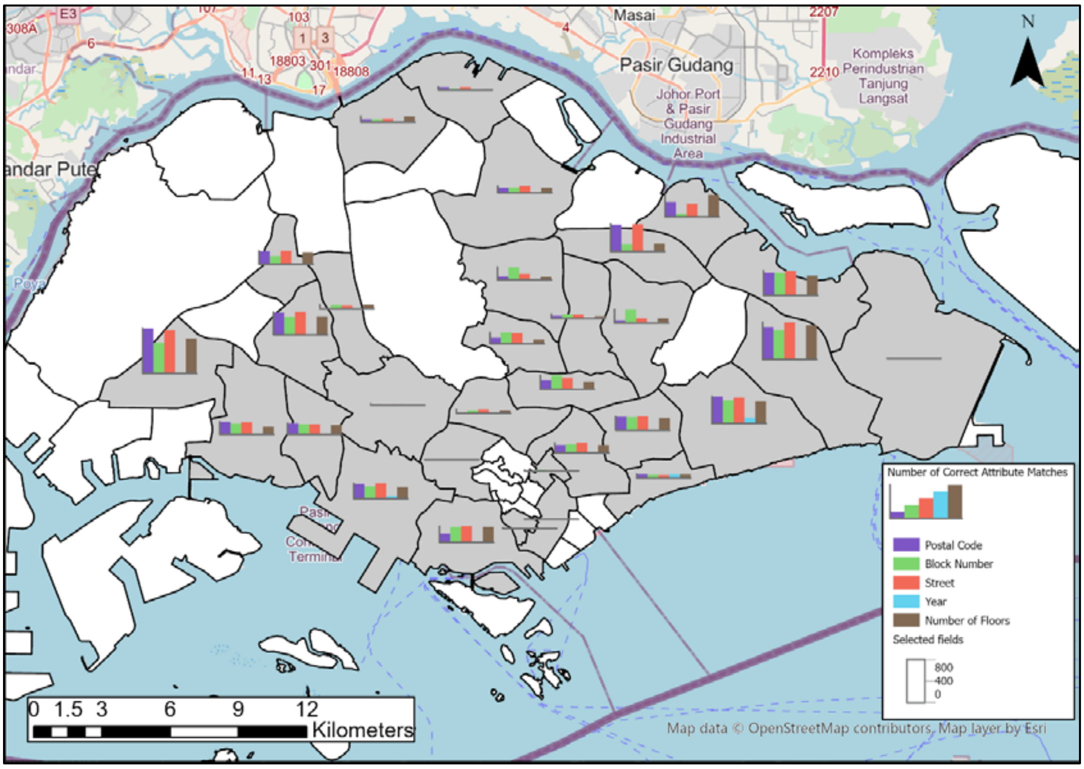

Figure 16. The number of correct attribute matches between five attribute fields across planning areas in Singapore.

In addition, the number of correct specific tags in OpenStreetMap buildings were also analysed as a proxy to assess attribute accuracy, and the results are shown in Table 8. Given that only 38% of the "residential" tag is correct, it reveals that many OpenStreetMap contributors were not aware of the type of residential buildings they mapped in Singapore.

Table 8. The number of correct specific tags. (Total buildings: 11312)

Tag	Number of buildings in OpenStreetMap	Percentage (%)
addr:city=Singapore	7310	64.62
addr:country=SG	5974	52.81
residential=HDB/hdb	4393	38.83
residential=yes	0	0

Table 9 shows the number and percentage of buildings that contain attributes useful for creating semantically rich 3D building models. It can be deduced that the majority of HDB buildings in

Singapore lack these attributes; therefore, it is not feasible to build semantically rich 3D building models using these attributes at the current time.

Table 9. Attributes useful for creating semantically rich 3D building models. (Total buildings: 11312)

Attribute field	Number of buildings in OpenStreetMap	Percentage (%)
roof:shape	248	2.19
roof:colour	264	2.33
roof:material	342	3.02
roof:levels	140	1.24
roof:orientation	57	0.5
roof:direction	8	0.07
height	721	6.37
surface	14	0.12
building:levels:underground	65	0.57
building:colour	411	3.63
building:levels	6103	53.95
levels	3	0.03

3.1.6 Correlation analyses

Table 10 shows the Pearson's correlation coefficient, r, calculated from correlation analyses conducted using the mean values of three quality metrics in every planning area against the planning areas' (1) median age of population, and (2) mean age of HDB buildings. The scatter plots used in these analyses are shown in Figure 17.

Table 10. Summary of correlation analyses.

	Median age of population (2019)	Mean age of HDB buildings
	Correlation coefficient, r	**Correlation coefficient, r**
Positional accuracy	-0.153132	-0.172797
Shape accuracy	0.180288	0.301116
Orientation accuracy	-0.39982	-0.26776

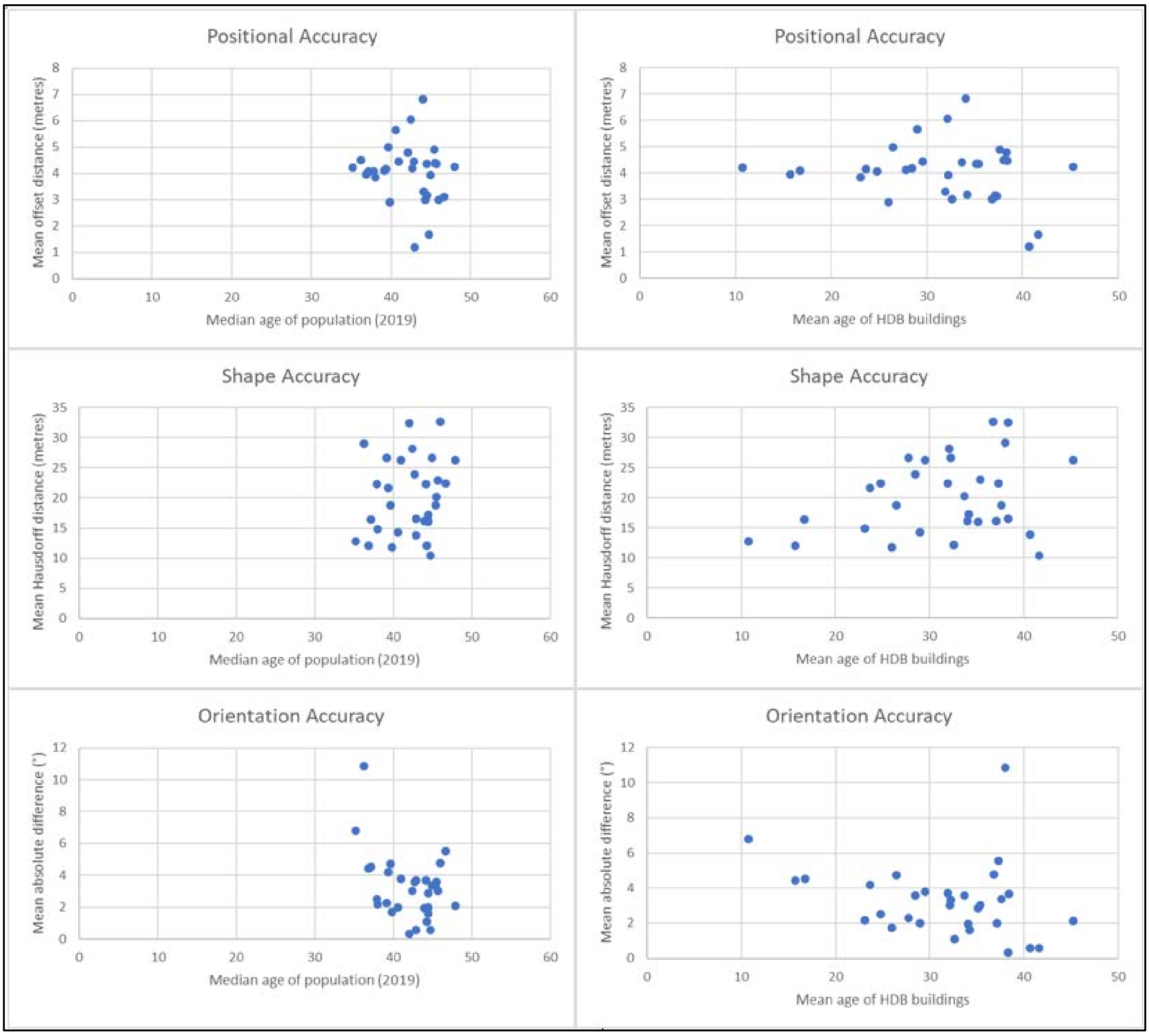

Figure 17. Scatter plots used in correlation analyses.

From the results table, it can be determined that both median age of population and mean age of HDB buildings have weak relationships with positional accuracy and shape accuracy of OpenStreetMap buildings in Singapore. On the other hand, orientation accuracy has a moderately weak negative relationship with median age of population, but a weak negative relationship with mean age of HDB buildings.

In other words, the median age of population and the mean age of HDB buildings in Singapore's planning areas are not sufficiently reliable to predict the quality of OpenStreetMap data in terms of positional accuracy, shape accuracy, and orientation accuracy.

3.2 Evaluation

Overall, the completeness of HDB buildings in OpenStreetMap is 97.67%. The near-perfect completeness level of HDB buildings gives a strong clue that a high percentage of other buildings in Singapore are likely mapped in OpenStreetMap.

Besides, more than 95% of HDB buildings in Singapore have 1:1 semantic relation, suggesting that the users were able to distinguish most of the individual HDB buildings from the satellite images in OpenStreetMap. Furthermore, the mean offset distance between buildings in OpenStreetMap and reference datasets is 4.06m, which may be caused by technical limitations such as the low-resolution of Bing satellite images. Moreover, the huge disparity between the highest and lowest Hausdorff distance indicates that there is large heterogeneity in terms of shape accuracy of HDB buildings in OpenStreetMap.

On the other hand, the mean absolute orientation difference of 3.2° suggests that the mapped HDB buildings generally have the same dominant direction as the corresponding buildings in the reference dataset, but with varying orientation accuracies. Apart from this, the attribute accuracy of HDB buildings in OpenStreetMap is far from perfect, with only 50% to 60% of the important attributes being correctly filled.

Considering all aspects of quality metrics discussed above, we can objectively deduce that the overall quality of OpenStreetMap HDB building data in Singapore is fairly good and could be valuable for different purposes, such as research.

3.3 Limitations

Since this study does not cover all buildings in Singapore, the overall results do not reflect the full picture of OpenStreetMap building data quality in Singapore.

Furthermore, since there are no other reliable methods to extract HDB buildings from all OpenStreetMap buildings in Singapore accurately, only the buildings in OpenStreetMap dataset which are overlapping the buildings in the reference dataset were extracted. As a result, HDB buildings that may be missing from the reference dataset but exist in OpenStreetMap could not be analysed.

4. Conclusion and recommendation

In this study, five quality metrics were analysed using various geospatial analysis methods to determine the quality of OpenStreetMap building data in Singapore. This study suggests that the OpenStreetMap HDB building data in Singapore reaches near-perfect completeness, with 97.67% of HDB buildings in Singapore being mapped in OpenStreetMap.

As discussed in Section 3.2, the quality of HDB buildings in OpenStreetMap was concluded objectively as fairly good. The analysis results were displayed in grid map and planning area map of Singapore, which allowed us to determine the spatial patterns of each quality metric. Besides, hotspot analyses were performed to identify the spatial clusters of statistically significant high and low values. These provided an overview of the locations in Singapore where OpenStreetMap data quality can be improved.

Through correlation analyses, socio-economic factors such as the median age of population and the mean age of HDB buildings in planning areas were found to have weak relationships with the quality of OpenStreetMap data in Singapore. Furthermore, we have also determined that it is currently not feasible to utilise the OpenStreetMap attributes of HDB buildings in Singapore to build semantically rich 3D building models.

While existing error detection tools such as Keep Right and Osmose are useful in detecting potential errors in OpenStreetMap, they are unable to detect errors in real-time. It is therefore recommended that the OpenStreetMap community build upon these existing technologies by exploring the possibility of building a real-time warning system, which has the ability to warn users instantly if the mapped feature is below a certain threshold of acceptable quality.

Future extension of this project may include other HDB facilities such as courtyards, sports facilities, gardens, playgrounds, etc. as part of the assessment of OpenStreetMap HDB building quality in Singapore. Ultimately, the data quality assessment can be extended to all buildings nation-wide.

References

Amirian, P., Basiri, A., Gales, G., Winstanley, A., & McDonald, J. (2015). The next generation of navigational services using OpenStreetMap data: The integration of augmented reality and graph databases. In *OpenStreetMap in GIScience* (pp. 211-228). Cham, Switzerland: Springer

Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of VGI quality: An overview. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *1*, 345-351.

Arsanjani, J. J., Barron, C., Bakillah, M., & Helbich, M. (2013). *Assessing the quality of OpenStreetMap contributors together with their contributions*. AGILE 2013, Leuven, Belgium.

Arsanjani, J. J., Mooney, P., Zipf, A., & Schauss, A. (2015a). Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In *OpenStreetMap in GIScience* (pp. 37-58). Cham, Switzerland: Springer

Arsanjani, J. J., Zipf, A., Mooney, P., & Helbich, M. (2015b). An introduction to OpenStreetMap in Geographic Information Science: Experiences, research, and applications. In *OpenStreetMap in GIScience* (1-15). Cham, Switzerland: Springer

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

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

Ciepluch, B., Mooney, P., Jacob, R., & Winstanley, A. C. (2009). Using OpenStreetMap to deliver location-based environmental information in Ireland. *SIGSPATIAL Special*, *1*(3), 17-22.

Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, *28*(4), 700-719.

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, *72*(3-4), 137-148.

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

Goodchild, M. F. (2008). Spatial accuracy 2.0. *Proceedings of the Eighth International Symposium on Spatial Data Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, China (pp 1-7).

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

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

Hecht, R., Kunze, C., & Hahmann, S. (2013). Measuring completeness of building footprints in OpenStreetMap over space and time. *ISPRS International Journal of Geo-Information, 2*(4), 1066-1091.

Hentschel, M., & Wagner, B. (2010). Autonomous robot navigation based on OpenStreetMap geodata. *13th International IEEE Conference on Intelligent Transportation Systems*, Funchal, Portugal (pp 1645-1650).

Husen, S., Idris, N. H., & Ishak, M. H. I. (2018). The quality of OpenStreetMap in Malaysia: A preliminary assessment. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4249*, 291-298.

International Organisation for Standardisation (2013). *Geographic information - Data quality* (ISO Standard No. 19157:2013). Retrieved from https://www.iso.org/standard/32575.html

Koukoletsos, T., Haklay, M., & Ellul, C. (2011). *An automated method to assess data completeness and positional accuracy of OpenStreetMap*. Proceedings GISRUK 2011, University of Portsmouth, Portsmouth, UK.

Lokhat, I., & Touya, G. (2016). Enhancing building footprints with squaring operations. *Journal of Spatial Information Science, 13*, 33-60.

Mashhadi, A., Quattrone, G., & Capra, L. (2015). The impact of society on volunteered geographic information: The case of OpenStreetMap. *OpenStreetMap in GIScience* (pp. 125-141). Cham, Switzerland: Springer International Publishing

Mooney, P., Corcoran, P., & Winstanley, A. C. (2010). Towards quality metrics for OpenStreetMap. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (514-517). New York, USA: Association for Computing Machinery

Rahman, K. M., Alam, T., & Chowdhury, M. (2012). Location based early disaster warning and evacuation system on mobile phones using OpenStreetMap. *2012 IEEE Conference on Open Systems*, Kuala Lumpur, Malaysia (pp 1-6).

Soden, R., & Palen, L. (2014). From crowdsourced mapping to community mapping: The post-earthquake work of OpenStreetMap Haiti. *Proceedings of the 11th International Conference on the Design of Cooperative Systems*, Nice, France (pp 311-326).

Xu, Y., Chen, Z., Xie, Z., & Wu, L. (2017). Quality assessment of building footprint data using a deep autoencoder network. *International Journal of Geographical Information Science, 31*(10), 1929-1951.

Zheng, S., & Zheng, J. (2014). Assessing the completeness and positional accuracy of OpenStreetMap in China. In *Thematic Cartography for the Society* (pp. 171-189). Cham, Switzerland: Springer.

Zielstra, D., & Zipf, A. (2010). Quantitative studies on the data quality of OpenStreetMap in Germany. *Proceedings of the Sixth International Conference on Geographic Information Science, GIScience 2010*, Zurich, Switzerland (pp 20–26).

Appendices

Appendix A. The breakdown of the calculation of completeness.

	Number of buildings (For calculation of completeness)	
Relation	**HDB Dataset**	**OSM Dataset**
1:1	11312	11312
1:0	97	-
1:n	97	97*
n:m	135	135*
n:1	105	105*
Total	**11924**	**11649**

$$\text{Completeness} = \frac{Total\ number\ of\ buildings\ in\ OSM\ dataset}{Total\ number\ of\ buildings\ in\ HDB\ dataset}\ x\ 100$$

$$\text{Completeness} = \frac{11312 + 97 + 135 + 105}{11312 + 275 + 97 + 135 + 105}\ x\ 100$$

$$\text{Completeness} = 97.67\%$$

*All semantic relations were incorporated into the calculation of completeness, refer to Table 3 for the overall results of all semantic relations. The number of buildings in 1:n, n:m, and n:1 relations of both HDB and OSM datasets are different due to semantic inaccuracy. In this case, the number of buildings in 1:n, n:m, and n:1 relations of HDB dataset was used for both HDB and OSM datasets in the calculation, as the completeness of HDB buildings in these semantic relations are considered 100%, regardless of the number of buildings involved.

Appendix B. Median age of population and the mean age of HDB buildings of every planning area in Singapore.

Planning Area	Median age of population (2019)	Mean age of HDB buildings
Ang Mo Kio	46.02	36.78
Bedok	44.44	34.2
Bishan	44.12	31.9
Bukit Batok	40.59	28.95
Bukit Merah	45.46	37.62
Bukit Panjang	39.33	23.65
Bukit Timah	42.45	32.11
Changi	36.25	38
Choa Chu Kang	38.05	23.09
Clementi	43	34.05
Downtown Core	42.93	40.67
Geylang	44.44	37.09
Hougang	42.7	28.46
Jurong East	44.28	32.59
Jurong West	39.83	25.97
Kallang	45.55	33.68
Marine Parade	44.71	41.61
Novena	42.91	38.38
Outram	47.96	45.23
Pasir Ris	39.68	26.46
Punggol	35.21	10.76
Queenstown	44.43	35.1
Rochor	46.68	37.31
Sembawang	37.15	16.72
Sengkang	36.86	15.69
Serangoon	44.91	32.22
Tampines	40.98	29.5
Tanglin	42.07	38.33
Toa Payoh	45.67	35.39
Woodlands	37.88	24.81
Yishun	39.2	27.78

Appendix C. Semantic accuracy (HDB dataset) of every planning area in Singapore.

	Number of buildings in HDB dataset					
Planning Areas	**1:1**	**1:0**	**1:n**	**n:m**	**n:1**	**Percentage of buildings with 1:1 relation (%)**
Ang Mo Kio	430	8	4	-	4	96.41
Bedok	574	9	6	2	6	96.15
Bishan	247	1	2	-	27	89.17
Bukit Batok	439	-	3	2	-	98.87
Bukit Merah	536	24	2	3	11	93.06
Bukit Panjang	415	1	1	-	-	99.52
Bukit Timah	27	-	1	-	-	96.43
Changi	7	-	-	-	-	100
Choa Chu Kang	614	15	3	2	4	96.24
Clementi	224	10	-	-	7	92.95
Downtown Core	3	-	-	-	-	100
Geylang	309	3	5	3	6	94.79
Hougang	629	7	3	-	2	98.13
Jurong East	258	1	1	2	-	98.47
Jurong West	869	4	4	-	2	98.86
Kallang	246	14	3	-	2	92.83
Marine Parade	77	-	1	-	-	98.72
Novena	97	1	15	-	-	85.84
Outram	31	1	-	18	2	59.62
Pasir Ris	483	35	2	2	-	92.53
Punggol	507	43	11	46	3	83.11
Queenstown	306	9	2	3	4	94.44
Rochor	35	1	1	8	-	77.78
Sembawang	285	5	2	2	2	96.28
Sengkang	736	28	8	31	2	91.43
Serangoon	249	6	1	-	-	97.27
Tampines	813	28	10	2	4	94.87
Tanglin	3	-	-	-	-	100
Toa Payoh	355	9	2	-	5	95.69
Woodlands	845	-	2	-	-	99.76
Yishun	663	12	2	9	12	94.99
Total	**11312**	**275**	**97**	**135**	**105**	**94.87**

Appendix D. Semantic accuracy (OSM dataset) of every planning area in Singapore.

	Number of buildings in OSM dataset				
Planning Areas	**1:1**	**1:n**	**n:m**	**n:1**	**Percentage of buildings with 1:1 relation (%)**
Ang Mo Kio	430	8	-	2	97.73
Bedok	574	13	3	3	96.8
Bishan	247	4	-	8	95.37
Bukit Batok	439	6	3	-	97.99
Bukit Merah	536	5	2	5	97.81
Bukit Panjang	415	2	-	-	99.52
Bukit Timah	27	2	-	-	93.1
Changi	7	-	-	-	100
Choa Chu Kang	614	7	3	2	98.08
Clementi	224	-	-	3	98.68
Downtown Core	3	-	-	-	100
Geylang	309	10	4	3	94.79
Hougang	629	8	-	1	98.59
Jurong East	258	2	3	-	98.1
Jurong West	869	8	-	1	98.97
Kallang	246	6	-	1	97.23
Marine Parade	77	2	-	-	97.47
Novena	97	107	-	-	47.55
Outram	31	-	20	1	59.62
Pasir Ris	483	5	3	-	98.37
Punggol	507	55	45	1	83.39
Queenstown	306	4	4	2	96.84
Rochor	35	2	5	-	83.33
Sembawang	285	4	3	1	97.27
Sengkang	736	16	24	1	94.72
Serangoon	249	2	-	-	99.2
Tampines	813	25	4	2	96.33
Tanglin	3	-	-	-	100
Toa Payoh	355	6	-	2	97.8
Woodlands	845	4	-	-	99.53
Yishun	663	4	6	6	97.64
Total	**11312**	**317**	**132**	**45**	**95.82**

Appendix E. Positional accuracy of HDB buildings of every planning area in Singapore.

Planning Area	Mean offset distance (metres)
Ang Mo Kio	3
Bedok	3.17
Bishan	3.3
Bukit Batok	5.65
Bukit Merah	4.9
Bukit Panjang	4.16
Bukit Timah	6.06
Changi	4.5
Choa Chu Kang	3.84
Clementi	6.83
Downtown Core	1.2
Geylang	3.16
Hougang	4.18
Jurong East	3
Jurong West	2.91
Kallang	4.4
Marine Parade	1.67
Novena	4.46
Outram	4.25
Pasir Ris	4.99
Punggol	4.21
Queenstown	4.36
Rochor	3.11
Sembawang	4.09
Sengkang	3.95
Serangoon	3.93
Tampines	4.45
Tanglin	4.78
Toa Payoh	4.35
Woodlands	4.07
Yishun	4.12

Appendix F. Shape accuracy of HDB buildings of every planning area in Singapore.

Planning Area	Mean Hausdorff distance (metres)
Ang Mo Kio	32.63
Bedok	17.24
Bishan	22.35
Bukit Batok	14.34
Bukit Merah	18.84
Bukit Panjang	21.76
Bukit Timah	28.14
Changi	29.11
Choa Chu Kang	14.87
Clementi	16.14
Downtown Core	13.85
Geylang	16.15
Hougang	23.88
Jurong East	12.12
Jurong West	11.78
Kallang	20.24
Marine Parade	10.46
Novena	16.52
Outram	26.25
Pasir Ris	18.83
Punggol	12.76
Queenstown	16.07
Rochor	22.4
Sembawang	16.44
Sengkang	12.08
Serangoon	26.64
Tampines	26.31
Tanglin	32.48
Toa Payoh	22.98
Woodlands	22.37
Yishun	26.65

Appendix G. Orientation accuracy of HDB buildings of every planning area in Singapore.

Planning Areas	Mean absolute difference (°)
Ang Mo Kio	4.8
Bedok	1.6
Bishan	3.7
Bukit Batok	2
Bukit Merah	3.4
Bukit Panjang	4.2
Bukit Timah	3
Changi	10.9
Choa Chu Kang	2.2
Clementi	2
Downtown Core	0.6
Geylang	2
Hougang	3.6
Jurong East	1.1
Jurong West	1.7
Kallang	3.6
Marine Parade	0.6
Novena	3.7
Outram	2.1
Pasir Ris	4.7
Punggol	6.8
Queenstown	2.9
Rochor	5.5
Sembawang	4.5
Sengkang	4.4
Serangoon	3.3
Tampines	3.8
Tanglin	0.3
Toa Payoh	3
Woodlands	2.5
Yishun	2.3

Appendix H. Attribute accuracy of HDB buildings of every planning area in Singapore.

	Percentage of buildings in OpenStreetMap with correct attribute matches (%)				
Planning Area	**Postal Code**	**Block Number**	**Street**	**Year Completed**	**Number of Floors**
Ang Mo Kio	0.54	2.08	0.96	0	0.45
Bedok	4.4	3.78	4.23	0.95	3.61
Bishan	1.03	1.86	1.79	0	0.71
Bukit Batok	3.57	2.99	3.7	0	3.03
Bukit Merah	1.45	2.52	2.66	0	2.55
Bukit Panjang	0.08	0.58	0.54	0	0.69
Bukit Timah	0.17	0	0.17	0	0
Changi	0	0.01	0.01	0	0.04
Choa Chu Kang	2.14	1.33	2.18	0	1.98
Clementi	1.86	1.71	1.71	0	1.56
Downtown Core	0.01	0.03	0.01	0	0
Geylang	2.23	2.16	2.26	0	2.05
Hougang	0.45	2.21	0.77	0	0.77
Jurong East	2.07	1.84	2.02	0	1.26
Jurong West	7.36	4.99	7.03	0	5.7
Kallang	1.31	1.46	1.71	0	1.29
Marine Parade	0.65	0.57	0.52	0.62	0.62
Novena	0.13	0.55	0.65	0	0.55
Outram	0.04	0.14	0.12	0	0.08
Pasir Ris	3.71	3.63	4.04	0	3.24
Punggol	2.41	0.53	2.14	0	3.64
Queenstown	2.41	2.03	2.51	0.47	1.9
Rochor	0.01	0.19	0.09	0	0.17
Sembawang	0.46	0.24	0.5	0	0.16
Sengkang	4.4	1.24	4.47	0	1.32
Serangoon	0.45	0.67	0.65	0	0.37
Tampines	5.25	4.75	6.05	0	5.59
Tanglin	0	0	0	0	0.01
Toa Payoh	1.44	2.3	1.87	0	1.22
Woodlands	0.48	0.46	0.54	0	0.98
Yishun	0.86	0.9	1.14	0	0.86
Total	**51.36**	**47.75**	**57.04**	**2.03**	**46.38**