Assessing the Quality of OpenStreetMap Data in South Africa in Reference to National Mapping Standards

Lindy-Anne Siebritz[1], George Sithole[2]

[1]Chief Directorate: National Geo-Spatial Information. Department of Rural Development and Land Reform, Cape Town, South Africa, lasiebritz@ruraldevelopment.gov.za

[2]Geomatics Division, University of Cape Town, Cape Town, South Africa, George.sithole@uct.ac.za

This paper has been through a process of double-blind peer review

Abstract

The introduction and success of Volunteered Geographic Information (VGI) has gained the interest of National Mapping Agencies (NMAs). VGI is geographic information that is freely generated by non-experts and shared using VGI initiatives available on the Internet. NMAs are looking to this volunteer information to maintain their topographic databases; however, the main concern is the quality of the data. The purpose of this study is to assess whether it is feasible to use VGI to update authoritative databases, specifically in South Africa. The data from OpenStreetMap (OSM), which is one of the most successful VGI initiatives, was compared to a reference data set provided by the NMA of South Africa i.e. the Chief Directorate: National Geo-Spatial Information (CD: NGI). The results demonstrate that the OSM quality is heterogeneous across South Africa. Some of the quality measures are higher, while others are lower than that stipulated by the CD: NGI. OSM data is therefore not within CD: NGI's national mapping requirements.

1. Introduction

Globally, there is an increasing demand for spatial data. NMAs are therefore constantly looking for ways to improve their mapping process. While official mapping is declining worldwide, the demand for spatial data is rising (Mcdougall, 2009). Instead of competing with open initiatives, NMAs are realising that open initiatives present a great opportunity for collaboration (McLaren, 2012). This study is focused on the integration of VGI into authoritative data in South Africa. OpenStreetMap (OSM) was chosen as the source of VGI as it is easy to obtain the vector data without any costs. The CD: NGI was chosen as the source of authoritative geo-spatial data. The feasibility of integrating the two data sets was determined by assessing the quality of OSM data with reference to the CD: NGI geo-spatial data standards.

For this investigation, the internal quality was measured by assessing the positional and semantic accuracy of roads, the geometric accuracy of amenity buildings and the completeness of roads. The internal quality refers to what the data is really like and is not dependent on the perception of the user (Mostafavi et al., 2004).

This paper is divided into seven main sections. Section two highlights previous investigations into the quality of OSM data. Section three presents some of the problems that hinder the integration of VGI into authoritative data. The methodology used to assess the quality of OSM road and building data is given in section four. In section five the results are presented. Sections six and seven contain the implications and conclusions of the results, respectively.

2. Previous Quality Assessments of OpenStreetMap Data

In 2009, one researcher investigated the positional accuracy of the OSM road network by comparing it to different authoritative data sets (Kounadi 2009). Thereafter, other researchers performed various quality assessments, including the completeness, geometric accuracy and attribute accuracy (Haklay & Elull 2010, Girres & Touya 2010). The varied results obtained in these studies show that the geographic location and extent; the reference data set; the method and execution of testing and the date of the OSM data extraction influence the results considerably. However, generally it was concluded that:

- Contributions are heterogeneous across study areas (Girres & Touya, 2010, Siebritz et al., 2011).
- In some areas OSM roads compare well in positional accuracy (Kounadi 2009).
- Many features are being added without the necessary attribute information (Haklay & Ellul, 2010).

Two quality assessments on the South African OSM data was done in 2012 (du Plooy 2012, Hankel 2012). However, the motivation for this study was to compare the OSM data specifically to the CD: NGI data and to increase the size and number of test areas.

3. Integrating VGI and Authoritative Data

Many researchers believe that any level of VGI ingestion will result in reduced map production costs (Johnson & Sieber, 2013). Other researchers have a different perspective, they believe that there are costs involved in re-aligning the workflow to incorporate VGI and other costs, like Internet access necessary to obtain the latest updates and training staff in VGI usage (Johnson & Sieber, 2013). Another challenge is government mapping agencies' acceptance of spatial data produced by non-experts (Johnson & Sieber, 2013). It is a known fact that the public deem authoritative spatial information to be without error, even if this is not necessarily the case. The opposite holds true for most, not all, NMAs - they do not deem volunteer information to be of high quality. It has been said that the reason for government mapping agencies' reluctance to accept volunteer information is the legal implications associated with incorrect information disseminated to the public (Johnson & Sieber, 2013).

Other technical issues to consider when integrating VGI data into NMA data include different reference systems, different representation of features, how to handle duplication of features and different file formats. There is also the physical structure of spatial databases and naming conventions to consider. This discussion will be limited to data structure issues.

3.1 Data Structures

There are distinct differences between the data models of authoritative and volunteer data in the way the data is acquired, stored and the policies and standards governing the data. The differences between the CD: NGI and OSM data models are listed in table 1.

Table 1. Comparison of OSM and CD: NGI Data Models

Element	CD: NGI	OSM
Data Elements & Storage	Comprised of points, lines, polygons and attributes and stored in an object-relational Oracle 10g database.	Comprised of nodes, ways (open, closed, closed-filled), tags and relations and stored in a SQL database (Ramm et al., 2011).
Acquisition/ Compilation	Field and office annotation data, map deletion sheets, ortho-rectified images and compilation corrections are used during heads-up digitising from aerial imagery.	Contributions made via GPS track uploads, organisational mass uploads and heads-up digitising from aerial/satellite imagery (Ramm et al., 2011).
Policies/Licensing	CD: NGI distributes spatial data under the Copyrights Act No 98 of 1978. CD: NGI owns the data. Users may use and modify the data, but they may not sell the data. The laws of the Copyrights Act do make allowance for exceptions if the application for copyrights is found to be valid (e.g. data vendors).	OSM operates under the Open Database License. Volunteers own their contributions. Users may use and modify the data, but they may not sell the data (*OpenStreetMap Main Page*, 2013).
Quality Assurance	The topographic compiler performs two checks on their own work and then the final quality control is done by a senior employee who ensures compliance to the standard for capture of topographic data.	Various tools are available for reporting bugs and detecting errors, automatically or manually (*OpenStreetMap Foundation*, 2013).

3.2 Examples of Integrating VGI and Authoritative Spatial Data

3.2.1 Victoria Department of Sustainability and Environment - Notification for Edit Service (NES)

The State of Victoria in Australia initiated a program where government users of Victoria's Corporate Spatial Data Library (CSDL) may annotate updates or corrections to the base spatial data set via an on-line editor (Coleman et al., 2010, Thomas et al., 2009) These changes and updates are sent to the data custodians for verification (Coleman et al., 2010). The program is still running and the results have shown a considerable decrease in map update time (Coleman et al., 2010).

3.2.2 Ordnance Survey Open Data

Instead of enriching the authoritative data set, Pourabdollah et al. (2013) focused their study on enriching the OSM data set using authoritative data from Ordnance Survey's Open Data. The enriched OSM data was served over open Web services (Pourabdollah et al., 2013). The authors state that they encountered various conflation issues.

4. Methodology for Determining the Quality of OSM Data in South Africa

The aim of this investigation is to determine how well the OSM data compares with the CD: NGI data quality requirements. The CD: NGI standard for the Capture of Topographic Data states: i) features captured by photogrammetric methods must have a positional accuracy not exceeding 10 metres at the 95% confidence level and ii) features shall be correctly classified at the 90% confidence interval. The standard does not include accuracy statements for the other quality aspects. The OSM data was thus assessed with respect to the positional accuracy and semantic accuracy. In addition, the completeness of roads was assessed. Due to time and technical constraints, other quality elements (e.g. thematic accuracy) were not investigated. This section presents the method for determining the positional accuracy and semantic accuracy of roads, the geometric accuracy of amenity buildings and the completeness of roads.

4.1 Data Sources and Data Preparation

The OSM data sets were obtained in shapefile format and covered the period from October 2006 to April 2012. The reference vector (in shapefile format) and raster data were obtained from the CD: NGI. In order to overlay the OSM data onto the CD: NGI data, the OSM data was projected onto the Transverse Mercator projection type. Twenty-seven test sites were chosen throughout South Africa, three test areas per province (see figure 1). Each test area represented high urban density or low urban density or commercial. Before quality tests could be done, the OSM data had to be preprocessed. Essentially, the quality assessment will be used to determine feasibility of integrating the CD: NGI and OSM data. Thus, the OSM data was transformed to resemble the CD: NGI data. The sections below describe the preprocessing of the OSM data.

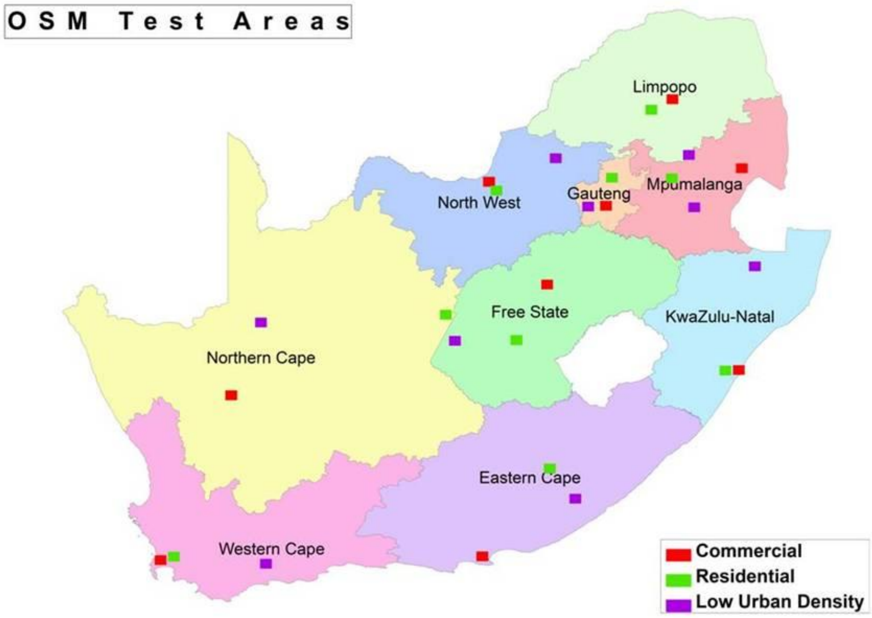

Figure 1. Map showing the twenty-seven test areas (not to scale)

4.1.1 Filtering

Attributed road features were extracted from the OSM line data sets. The same process was applied to the polygon data sets. Attributed buildings were extracted for comparison to the CD: NGI building (area) features. The OSM hole polygons consists of an external and internal boundary. These polygons were generalised into simple polygons in order to match the CD: NGI polygons, by removing the internal boundaries.

4.1.2 Generating Centrelines

The CD: NGI generalises all multi-lane (e.g., dual lane) roads into a single centreline, whereas OSM allows for multiple centrelines. Single centrelines were generated automatically using the appropriate tool in the Esri ArcMap software in order to remove any bias. Generalisation compromises the accuracy. However, the methodology for assessing the positional accuracy is based on matching roads between the two data sets. In figure 2, the dual-way road is represented by two centrelines in the OSM data set, while the CD: NGI uses a single centreline. Neither of the OSM centrelines is within 10m of the corresponding CD: NGI road. However, the generalised centreline in figure 2 falls within the buffer. In the case where the generated centreline was outside of the buffer (that is, no match), the road was not included in the assessment.

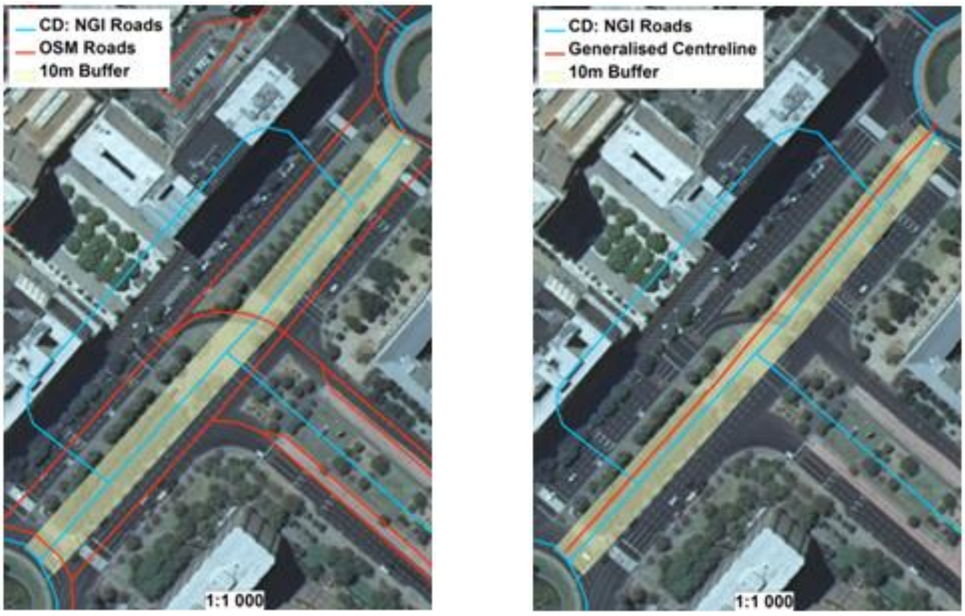

Figure 2: Example of generalised OSM centreline within 10m buffer

4.1.3 Feature Matching

Corresponding roads between the CD: NGI and OSM data were identified in order to eliminate any bias. A buffer size was chosen based on the CD: NGI standard for capturing topographical features. A 10 m buffer was generated for each CD: NGI road feature. However, because of road intersections, some of the buffers included parts of other roads. A semi-automatic method was developed to remove unwanted road sections (see figure 3).

Corresponding amenity buildings between the OSM and CD: NGI data sets were identified by comparing their centroids. If the centroid of an OSM polygon was found to be within a CD: NGI polygon, the two polygons most likely represented the same feature on the ground.

4.2 Computing the Positional Accuracy of Roads

The method by Goodchild & Hunter (1997) was chosen to compute the positional accuracy of OSM road features because: i) no assumption is made about the accuracy of the test data and ii) it is insensitive to outliers. According to this method, the positional accuracy may be determined by the percentage of OSM road that is within the buffer of the corresponding CD: NGI road feature (Goodchild & Hunter, 1997). For this investigation, the buffer width was set to the CD: NGI stated positional accuracy of 10m (see figure 3).

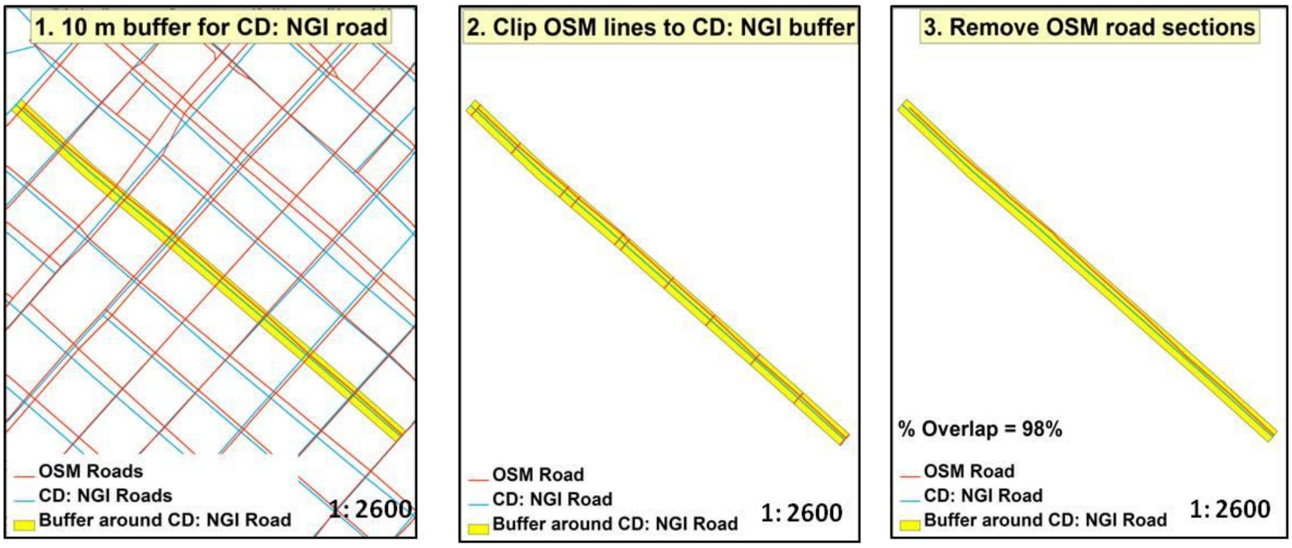

Figure 3. Computing the positional accuracy of roads using the Goodchild & Hunter (1997) method

4.3 Computing the Geometric Accuracy of Polygons

The Hausdorff distance was computed to find the positional accuracy of polygons. The Hausdorff distance gives the "maximum distance of a set to the nearest point in the other set" (Gregoire & Bouillot, 1998). In addition, the ratio of the areas and compactness were computed for shape comparisons:

$$Area\ Ratio = {}^{Area\ of\ OSM\ polygon}\!/\!_{Area\ of\ CD{:}\ NGI\ polygon} \quad [1]$$

Compactness measures how regular a polygon is, in other words, how much a polygon deviates from a predefined shape (e.g. a square) (Lee et al., 2004):

$$C = {}^{Area}\!/\!_{0.282\ \times\ perimeter^2} \quad [2]$$

Incorrect polygon matches were identified by a large Hausdorff distance. Each outlier was inspected manually in order to confirm an incorrect match.

4.4 Computing the Semantic Accuracy of Roads

Semantic accuracy refers to how well volunteers are able to name or identify the features they digitise. In terms of the CD: NGI and OSM road classes, the semantic accuracy is determined by how often a feature is classified as the same type of feature in both data sets. The class names differed; therefore, it was necessary to determine which road classes defined the same type of features between the two data sets. The biggest data set, the Western Cape commercial data set was chosen to determine how many times an OSM road class matched a certain CD: NGI road class for corresponding road features. Based on these results as well as the road class definitions, three road matches were selected (see Table 2). The positive road class matches were then used as a standard to assess the semantic accuracy of the other test areas.

Table 2. Road class matches between the CD: NGI and OSM (Chief Directorate: National Geo-

Spatial Information, 2013, Ramm et al., 2011)

CD: NGI Road	CD: NGI Definition	OSM Road	OSM Definition
Main Road	Main roads link the large towns, which are not on national or arterial routes, to the nearest major centre or city.	Primary	Major long-distance (inter-city) road
National Freeway	A national freeway is a dual carriageway (double road, each having two or more lanes) that is free of obstructions. No robots or intersections slow down the traffic and a minimum speed limit keeps slow vehicles off the freeway. Access is limited i.e. traffic must join and leave a freeway via an on or off ramp only.	Motorway	Large grade-separated, limited access freeway (or motorway)
Street	Streets make it possible to get access to the buildings in a town. They divide urban areas into blocks of houses or other buildings. Three basic patterns of streets can be described as grid, radial or irregular.	Residential	Residential street. Most inner-city roads use this type unless they are freeways.

4.5 Completeness

The completeness of a data set can be defined as the omission of data and the presence of excess data (Girres & Touya 2010). For this investigation, the completeness refers to how much of the CD: NGI roads exist within the OSM data set for a given area. The completeness was computed by dividing the total length of the April 2012 OSM road data set by the total length of the CD: NGI road data as at April 2012 (Haklay & Ellul, 2010).

5. Results

5.1 Positional Accuracy of Roads

The CD: NGI roads were buffered and the percentages of matching OSM roads within these buffers were computed. The percentage match for each test area was computed. These were then used to compute the weighted averages per province first and then for each settlement category. The count of features was used as the weight and was computed in order to remove any bias. The weighted percentage overlap for each province is presented in table 3. The percentages for the nine provinces are in the range of 64.8-94.3%. Five provinces have a percentage greater than 80% and four out of the five have percentages within 10% of the CD: NGI requirement. Considering the 5% error within the CD: NGI data, these four provinces may have an absolute accuracy of 95% and greater. The bigger and more developed cities tend to have the highest percentages. The location, which is linked to the number of features, also influences the average percentage overlap. The remaining non-overlapping percentages are due to incorrect digitising in the OSM data set or omission of feature detail in the CD: NGI data set or missing road sections in the OSM data set. The weighted percentage overlap per settlement category is shown in table 4. The percentages for commercial and residential categories are very similar, while the percentage for low urban density is about 10% lower. The best way to describe the OSM road positional accuracy is the percentage overlap per province, e.g. 93% of the OSM roads in Gauteng are within 10m of the CD: NGI roads.

Table 3. Comparing the positional accuracies of roads per province

Province	Weighted Mean	Province	Weighted Mean	Province	Weighted Mean
North West	64.8%	Eastern Cape	78.7%	KZN	90.6%
Mpumalanga	71.1%	Free State	80.9%	Gauteng	92.9%
Northern Cape	77.1%	Western Cape	89.4%	Limpopo	94.3%

Table 4. Comparing the positional accuracies of roads per settlement category

Settlement Type	Weighted Mean	Settlement Type	Weighted Mean	Settlement Type	Weighted Mean
Commercial	84.9%	Residential	85.7%	Low urban density	74.1%

5.2 Geometric Accuracy of Polygons

Only seven out of the twenty-seven test areas had OSM polygon data representing amenities. Figure 4 shows the average Hausdorff distances for each test area. The average Hausdorff distances for the commercial test areas range from 9.90m to 22.03m with a weighted average of 11.29m. The averages for the residential areas range from 11.34m to 17.36m and a weighted average of 12.54m. Considering the size of buildings, these deviations from the 10m requirement is not significant.

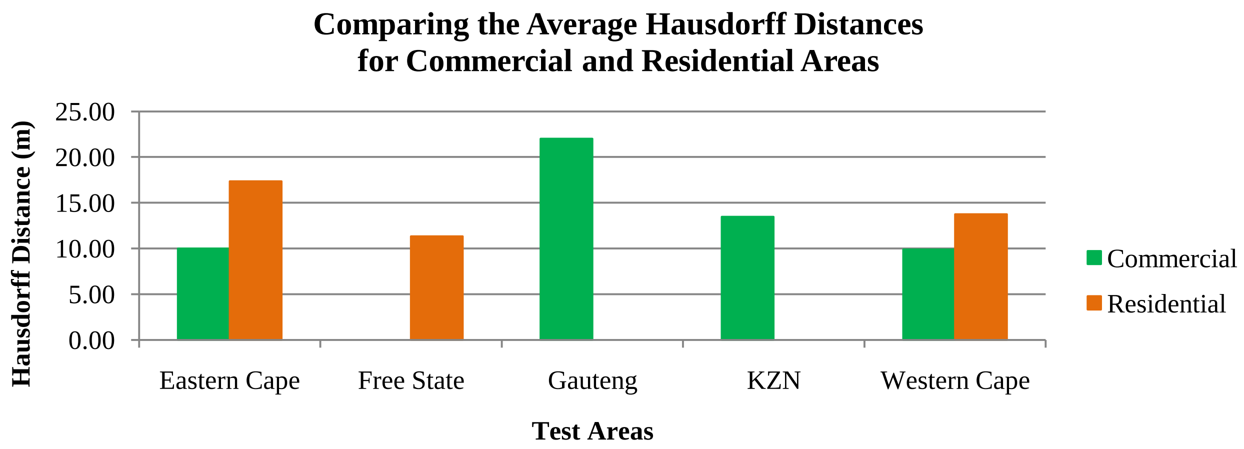

Figure 4. Graph comparing the average Hausdorff distances for commercial and residential areas

The compactness values were computed for every pair of corresponding polygons. The averages per province are presented in table 5. The commercial compactness averages for both the CD: NGI and OSM polygons are mostly consistent between provinces. This is reflected by the low standard deviations of 0.02 and 0.03 for CD: NGI and OSM, respectively. There is a difference of 0.01 between the CD: NGI and OSM weighted averages. The polygons therefore have a similar compactness. The residential averages for the CD: NGI is varying between provinces, compared to the commercial compactness averages. The standard deviation is 0.12, slightly greater. The OSM residential averages are somewhat more consistent with a small standard deviation of 0.04. The weighted average difference is higher with a value of 0.18. The CD: NGI polygons have a greater average compactness, thus the OSM polygons are less compact in residential areas. The CD: NGI weighted average compactness for commercial and residential polygons differ by 0.01, which means the CD: NGI compactness is generally consistent.

Table 5 also compares the area ratios for the commercial and residential areas. The commercial area ratios range from 0.97 to 1.37 and from 0.97 to 1.34 for the residential category. The similarity in range is seen by the equal standard deviations of 0.18. The weighted average ratios are slightly different for the 2 categories with 1.06 for commercial and 1.27 for residential areas. For the commercial weighted average, which is close to one, it can be said that the CD: NGI and OSM polygons occupy similar areas. Polygons in residential areas are slightly less similar in area.

Table 5. Comparing the compactness values and area ratios between CD: NGI and OSM polygons

Province	Compactness Difference						Area Ratio Differences			
	Commercial			Residential			Commercial		Residential	
	CD: NGI Average Compactness	OSM Average Compactness	Compactness Difference	CD: NGI Average Compactness	OSM Average Compactness	Compactness Difference	Area Ratio	Std Deviation	Area Ratio	Std Deviation
Eastern Cape	0.60	0.54	0.06	0.53	0.50	0.18	1.37	0.55	1.07	0.48
Free State	-	-	-	0.65	0.41	0.18	-	-	1.34	0.38
Gauteng	0.57	0.55	0.02	-	-	-	1.34	0.19	-	-
Kwazulu Natal	0.62	0.59	0.03	-	-	-	1.20	0.32	-	-
Western Cape	0.60	0.60	-0.01	0.41	0.44	0.03	0.97	0.38	0.97	0.30
Weighted Mean	0.60	0.59		0.61	0.43		1.06		1.27	
Std Dev	0.02	0.03		0.12	0.04		0.18		0.18	

Table 6. Comparing the weighted average percentages for CD: NGI and OSM road class matches

Province	Commercial (%)			Residential (%)			Low Urban Density (%)		
	National Freeway	Main Road	Street	National Freeway	Main Road	Street	National Freeway	Main Road	Street
Eastern Cape	100.0	-	34.6	-	-	90.4	-	-	100
Free State	-	0	100.0	-	-	81.6	-	-	96.2
Gauteng	-	16.7	54.8	-	0	96.5	-	-	-
Kwazulu Natal	50.0	18.4	8.5	-	-	0	-	-	94.6
Limpopo	0	100	79.7	-	0	100.0	-	0	86.7
Mpumalanga	0	-	16.2	0	0	100.0	-	0	100.0
North West	-	0	100.0	-	-	100.0	-	-	100.0
Northern Cape	0	0	58.9	-	12.4	91.1	0	-	100.0
Western Cape	52.4	38.0	22.8	100	-	85.5	-	0	98.1
Weighted Mean	39.1	28.9	32.3	66.7	7.7	91.4	0	0	97.9
Std Dev	41.0	36.0	34.8	70.7	6.3	31.8	0	0	4.6
Std Error	6.1	3.1	0.9	28.9	1.73	1.04	0	0	0.3

5.3 Semantic Accuracy of Roads

Semantic accuracy is expressed here as the number of matches between CD: NGI and OSM road classes, this is presented in table 6. The weighted means were computed for each road class within the three settlement categories. As before, the weight was used to remove bias of different counts. The percentage matches for commercial areas were found to be the lowest. In both the residential and low urban density categories the averages for the "street" (or the residential class in OSM) category is very high. As would be expected, the count for this road class is high for residential and low urban density areas. This does not necessarily mean that the percentage match should be high. One explanation for the high percentages for the “street” category is that there is a default naming with this road class. It appears that most of the time people will classify a road as a residential road. For residential and low urban density areas, it just happens to be the correct classification. The low percentages for the commercial category provides a better indication of how often people classify roads correctly.

5.4 Completeness of OSM Data

The completeness is based on the initial assumption that the CD: NGI data set is complete. Thus, in cases where the completeness percentage exceeds 100% completeness, the assumption no longer holds true. Because no ground truth data was used, it cannot be said with certainty that the percentages above 100% imply omission in the CD: NGI data set. Instead, it implies commission in the OSM data set. The three graphs in figures 5, 6 and 7 show the completeness of OSM data from October 2006 to April 2012. In all three graphs, the OSM data reaches a peak and then evens out. The commercial category had the most sites reaching their peaks during the periods 2009 to 2010 and 2011 to 2012. The residential category had the most sites reaching their peaks during the period 2011 to 2012 and for the low urban density category, during the 2010 to 2011 period. This shows that OSM South Africa received the most contributions from 2010 to 2012. Perhaps more people were exposed to volunteer mapping during this period. Specific events in time, like the 2010 FIFA Soccer World Cup, may have motivated citizens to participate in volunteers mapping in SA (Siebritz et al., 2011). For commercial areas, three of the test areas did not reach a completeness of 100%, although two of the three areas had a maximum in the 93-96% range. The residential category had five sites with a maximum completeness below 100%. In the low urban density category, only two sites reached a completeness value of 100% and above. The results show that less developed areas have a lower level of completeness. The completeness graphs also show that commercial areas received contributions much quicker than the other two categories and therefore even out sooner.

Figure 5. Comparison of OSM road completeness for commercial areas (2006-2012)

Figure 6. Comparison of OSM road completeness for residential areas (2006-2012)

Figure 7. Comparison of OSM road completeness for low urban density areas

6. Analysis and Discussion

The OSM roads do not meet the CD: NGI accuracy requirements. However, the overlap percentages are high for most provinces considering; i) that OSM does not enforce accuracy, ii) that data is being generated by non-professionals and iii) the methods of data collection. In terms of the settlement categories, the percentages for the residential and commercial areas are close to the CD: NGI quality requirements, while for the low urban density areas they are not. The OSM semantic accuracy of roads is only high for one of the road categories, which gives a bad impression of the non-expert's understanding of feature classification. Overall, the semantic accuracy of OSM data is not high. This confirms a previous statement that freedom in OSM tagging results in semantic interoperability problems (Baglatzi et al., 2012). By enforcing pre-set attribute ranges; the users' tagging options may be limited leading to stable classification.

The average Hausdorff distances for the OSM amenity buildings compare well with the CD: NGI's stated positional accuracy of 10m. The commercial areas have a higher positional accuracy. The shape comparisons show that polygons in commercial areas compare well with the CD: NGI polygons, whereas those in residential areas are less consistent.

Commercial and residential areas generally have a high level of completeness and in many cases exceed the CD: NGI data set. Low urban density areas are in most cases not complete. The completeness of low urban density areas may increase as more people become involved with volunteer mapping.

7. Conclusion

The investigation has demonstrated that for the purpose of integration, pre-processing of OSM data cannot be avoided, and it may be time consuming, especially for larger areas. Moreover, the quality of OSM data was found to be heterogeneous across South Africa (Siebritz et al., 2011). More developed areas receive more contributions than low urban density areas.

The data only meets some of the CD: NGI quality requirements. What this means is that the integration of OSM data into the CD: NGI data may be complicated. The degree to which these complications can be overcome will determine the level of integration. An integration workflow has been proposed, but was not presented here (Siebritz, 2014). Another option is to use the OSM data for detecting changes to the landscape only. Future work will investigate the modalities of integrating VGI data into CD: NGI data.

8. References

OpenStreetMap Foundation, 2013, viewed January 17, 2014, http://wiki.osmfoundation.org/ .

OpenStreetMap Main Page, 2013 viewed January 21, 2013, http://wiki.openstreetmap.org.

Baglatzi, A, Kokla, M & Kavouras, M 2012, 'Semantifying OpenStreetMap', in *The 11th International Semantic Web Conference*. Boston, pp. 39–48.

Chief Directorate: National Geo-Spatial Information 2013, *Chief Directorate : National Geo – spatial Information Standard for the Capture of Topographical Data*, Cape Town.

Coleman, DJ, Nkhwanana, NJ & Sabone, B 2010, Volunteering Geographic Information to authoritative databases: Linking contributor motivations to program characteristics', *Geomatica*, 64(1), pp.383–396.

Girres, J & Touya, G 2010, 'Quality assessment of the French OpenStreetMap dataset', *Transactions in GIS*, 14(4), pp.435–459, available at: http://doi.wiley.com/10.1111/j.1467-9671.2010.01203.x .

Goodchild, MF & Hunter, GJ 1997, 'A simple positional accuracy for linear features', *International Journal of Geographical Information Science*, 11(3), pp.299–306.

Gregoire, N & Bouillot, M 1998, 'Hausdorff distance between convex polygons', *Web Project for the Course CS 507 Computational Geometry*, available at: http://cgm.cs.mcgill.ca/~godfried/teaching/cg-projects/98/normand/main.html.

Haklay, M & Ellul, C 2010, 'Completeness in volunteered geographical information - the evolution of OpenStreetMap coverage in England (2008-2009)', *Journal of Spatial Information Science*, 2, p.18.

Hankel, L 2012, *Assessment of the quality of OpenStreetMap data in Woodhill , Kameeldrift and Mamelodi for the use of location based services*, University of Pretoria.

Johnson, PA & Sieber, RE 2013, 'Situating the Adoption of VGI by Government', In Sui, Elwood, & Goodchild, eds. *Crowdsourcing Geographic Knowledge: Volunteered Georgaphic Information (VGI) in Theory and Practice*. New York London: Springer Dordrecht Heidelberg, pp. 65–82.

Kounadi, O 2009, *Assessing the quality of OpenStreetMap data*, University College of London.

Lee, SC, Wang, Y & Lee, ET 2004, 'Compactness measure of digital shapes', in *2004 IEEE Region 5 Conference: Annual Technical & Leadership Workshop*. Oklahoma, pp. 103–105, available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1300173.

Mcdougall, K 2009, 'The Potential of Citizen Volunteered Spatial Information for Building SDI', in *GSDI-11 Conference*. Rotterdam, p. 10.

McLaren, R 2012, 'Between a rock and a hard location', *Geospatial World*, 2(12), pp.38–40.

Mostafavi, MA, Edwards, G & Jeansoulin, R 2004, 'An ontology-based method for quality assessment of spatial data bases', in *Third International Symposium on Spatial Data Quality*. Austria, pp. 49–66.

du Plooy, N 2012, *Assessment of OpenStreetMap Data Quality Across Different Urban Hierarchic Levels: A Comparative Study*, University of Pretoria.

Pourabdollah, A, Morley, J, Feldman, S & Jackson, M 2013, 'Towards an Authoritative OpenStreetMap: Conflating OSM and OS OpenData National Maps' Road Network', *ISPRS International Journal of Geo-Information*, 2(3), pp.704–728.

Ramm, F, Topf, J & Chilton, S 2011, *OpenStreetMap*, England: UIT Cambridge Ltd.

Siebritz, L 2014. *Assessing the accuracy of OpenStreetMap data in South Africa for the purpose of integrating it with authoritative data*, University of Cape Town.

Siebritz, L, Sithole, G & Zlatanova, S 2011, 'Assessment of the homogeneity of Volunteered Geographic Information in South Africa', in *XXII International Society for Photogrammetry & Remote Sensing Congress*. Melbourne, p. 6.

Thomas, E, Hedberg, O, Thompson, B, Rajabifard, A & Victorian Spatial Council 2009, 'A Strategy Framework to Facilitate Spatially Enabled Victoria', in *GSDI 11 World Conference*. Rotterdam, Netherlands, p. 19. Available at: http://www.gsdi.org/gsdi11/papers/pdf/167.pdf.