2017

Accuracy Evaluation of the Canadian OpenStreetMap Road Networks

Hongyu Zhang
The University of Western Ontario, hzhan338@uwo.ca

Jacek Malczewski
The University of Western Ontario, jmalczew@uwo.ca

Accuracy Evaluation of the Canadian OpenStreetMap Road Networks

Abstract
Volunteered geographic information (VGI) has been applied in many fields such as participatory planning, humanitarian relief and crisis management. One of the reasons for popularity of VGI is its cost-effectiveness. However, the coverage and accuracy of VGI cannot be guaranteed. The issue of geospatial data quality in the OpenStreetMap (OSM) project has become a trending research topic because of the large size of the dataset and the multiple channels of data access. This paper focuses on a national study of the Canadian OSM road network data for the assessment of completeness, positional accuracy, attribute accuracy, semantic accuracy and lineage. The OSM road networks in Canada have generally reliable quality compared to Digital Mapping Technologies Inc. Urban areas and footways received more contributions than rural areas and motorways, and imported road segments from GeoBase have slightly better quality than the national OSM dataset. The findings of the map quality can potentially guide cartographic product selection for interested parties and offer a better understanding of future improvement of OSM quality. In addition, the study presents the OSM contributions influenced by data import and remote mapping.

Keywords
Canada, quality, OpenStreetMap, volunteered geographic information

Acknowledgements
Thanks to the server infrastructure and staff members at the Shared Hierarchical Academic Research Computing Network (SHARCNET), especially Mr. Ge Baolai from the University of Western Ontario for his support in this project. The authors would also like to thank the anonymous reviewers for their valuable comments and suggestions.

1 INTRODUCTION

The advancement and availability of technology such as Web 2.0, the Global Positioning System (GPS) and high-speed internet has resulted in the proliferation of geospatial data in the 21st century. Users are no longer limited to browsing but also creating content online, and geographers are particularly interested in user-generated content (UGC) with spatial information (Coleman, Georgiadou, and Labonte 2009). Different concepts have been defined to describe this worldwide phenomenon, namely volunteered geographic information (VGI) (Goodchild 2007), coerced geographic information (McKenzie and Janowicz 2014), neogeography (Turner 2006), crowd-sourced geodata (Barron, Neis, and Zipf 2014), public participation GIS (Sieber 2006), collaborative GIS (Balram and Dragicevic 2008), participatory GIS (Elwood 2006), and community integrated GIS (Elmes et al. 2005). Compared to other concepts, VGI targets end-users, who are usually laypeople with their own needs and motivations (Flanagin and Metzger 2008). In this paper, only information collected through an opt-in provision is considered as VGI (Harvey 2013).

OpenStreetMap (OSM) is one of the early and long-lasting VGI mapping projects that aims to develop a free and accessible world map. Established in 2004, OSM has grown quickly in recent years, with the total number of registered users passing 3.5 million on March 15, 2017 according to the OSMstats website maintained by Pascal Neis. The project utilizes the Open Database License (ODbL) from Open Data Commons, which allows data to be freely accessed from multiple servers (e.g., Planet OSM, Geofabrik or Mapzen) in different formats (e.g., OSM Extensible Markup Language (XML), Protocol Buffer Binary Format (PBF) or shapefile). Tags are stored as "key:value" pairs, which are displayed as attributes associated with map features. Nodes, ways and relations construct the OSM project together, where ways are made of multiple nodes (points), and relations (e.g., bus routes) consist of at least one tag with an ordered list of nodes, ways and/or relations. Applications based on OSM are very diverse, and include but are not limited to navigation (e.g., for driving, biking or walking[1]), cartography for specific purposes (e.g., for wheelchair users[2], humanitarian relief[3] and land use/land cover mapping (e.g., Jokar Arsanjani et al. 2015)) and 3D city models (e.g., Over et al. 2010).

Users want to know how accurate the OSM data are, but an applicable quality evaluation approach for a large country like Canada has not been fully explored. Since road networks have received with the most attention from OSM contributors (Gröchenig, Brunauer, and Rehrl 2014), the objective of this study is to assess the extrinsic quality of OSM street networks in Canada. Results of this research can be used as a guide of OSM quality improvement and decision-making in cartographic product selection. Broadly speaking, this paper provides deeper insights into the uncertainty of VGI and geocomputation in data quality.

The remainder of the paper is organized as follows. Section 2 presents related studies of OSM quality. Section 3 describes the process of data preparation and the how-to methods of selected quality elements. Section 4 shows and discusses an assessment of the Canadian OSM quality. Finally, Section 5 provides a summary of findings and future research directions.

---

[1] http://www.openrouteservice.org/
[2] https://wheelmap.org/
[3] https://www.hotosm.org/

2 RELATED WORK

Zhang and Malczewski (2017) performed an extensive review of quality evaluation on OSM and found 60 relevant articles as of July 2016. OSM data used in those articles were accessed starting from 2007, which matches the founding year of the notion of VGI (Goodchild 2007). Both extrinsic and intrinsic metrics can evaluate spatial data quality. While extrinsic assessment compares OSM data to a reliable (and usually authoritative) reference dataset using quality measures derived from the ISO standards, intrinsic assessment measures OSM quality through proxies that are known as quality indicators (Antoniou and Skopeliti 2015). Examples of quality measures include completeness, positional accuracy, attribute accuracy and semantic accuracy (Van Oort 2006). Most reviewed articles used quality measures to compare OSM data with governmental or commercial datasets in selected European regions (Zhang and Malczewski 2017). For instance, Haklay (2010) first examined the completeness and positional accuracy of OSM streets in London and other parts of England in 2007.

Only seven of the 60 studies were implemented nationally, indicating potential difficulties of small scale OSM quality analysis (Zhang and Malczewski 2017). Few studies have focused on North America, primarily because of the less comprehensive data compared to European countries in the first number of years of the OSM project. Three trends, including improved quality over time, better urban coverage, and more comprehensive walkways, can be summarized from the national studies. Zielstra and Zipf (2010) found that although the total road length of OSM did not catch up with the data from Tele Atlas and Multinet, the number of roads increased very quickly in Germany (Zielstra and Zipf 2010). The topological and completeness errors decreased over the years from 2007 to 2011 as well (Neis, Zielstra, and Zipf 2011). Moreover, city centers received more contributions than rural areas, and spatial heterogeneity was observed in terms of completeness (Zielstra and Zipf 2010). Similar to what Zielstra and Zipf discovered in 2010, populated regions in Germany were found to have better attribute accuracy and completess than uninhabited regions (Ludwig, Voss, and Krause-traudes 2011). In the U.K., while dense areas had the best attribute accuracy, the middle to large sized cities had the worst quality, leaving less populated areas in the middle (Pourabdollah et al. 2013). Furthermore, both Ludwig, Voss, and Krause-traudes (2011) and Neis, Zielstra, and Zipf (2011) found that walkways had better completeness than motorways. In rural Germany, OSM could produce better routes of pedestrian navigation than TomTom, while TomTom generally outperformed OSM in car navigation because of reasons such as the lack of turn restrictions in OSM (Neis, Zielstra, and Zipf 2011).

Two phenomena, namely remote mapping and data import, potentially influence the quality of OSM. With the availability of satellite images, more and more contributors have become “armchair mappers” who only trace objects from aerial photos without local knowledge or without making measurements with GPS devices (Neis, Zielstra, and Zipf 2013). Although more detailed studies are needed, “armchair mapping” (or remote mapping) may cause various quality issues because of language barriers, limited image resolutions, lack of cartographic skills and loosely enforced specifications. At the same time, contributors tried to improve the regional maps through importing data from available authoritative data sources. Zielstra, Hochmair, and Neis (2013) compared the OSM streets in the United States between 2006 and 2012 to TIGER/Line data, which was fully imported to OSM in 2007 and 2008. Although the import action dramatically increased the completeness of street networks in OSM, especially in sparsely populated areas, the import also resulted in systematic errors (Zielstra, Hochmair, and Neis 2013). For example, OSM does not share the same road classification system with other databases such as TIGER/Line, which led to either incorrectly classified or unclassified roads in the U.S. Previously linked walkways and motorways may have been disconnected due to the import as well (Zielstra, Hochmair, and Neis 2013). Similar to the U.S., attention should be paid in Canada to the impacts of data import (from GeoBase starting at 2008) and associated systematic error propagation (Tenney 2014). Figure 1 shows the relationship between CanVec, GeoBase and the National Road Network (NRN), which essentially refers to the same road network dataset. All of them are produced by Natural Resources Canada and can be accessed from Open Government of Canada. For the consistency with OpenStreetMap Wiki, the three sources were aggregated as GeoBase for the remaining of the paper.

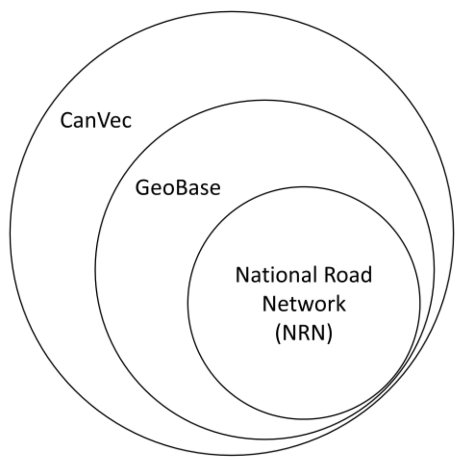

Figure 1. The relationship between selected geospatial data products from Natural Resources Canada

3 DATA AND METHODS

This research focuses on the quality of OSM in Canada. To the best of the authors' knowledge, previous studies have not covered the Canadian OSM quality in detail. Two databases were compared to evaluate the extrinsic OSM quality. The reference data were the national road networks from Digital Mapping Technologies Inc. (DMTI) published on Sept. 1, 2015, of which the positional accuracy ranges from 0.6 (urban) to 30 m (rural) (DMTI Spatial Inc. 2015). NRN was not chosen as the reference dataset, because NRN is a source of data import for the Canadian OSM and has an uncertain positional accuracy ("in meters" as of April 23, 2016 on its website). The OSM data were extracted from the full history dump and then processed using open-sourced packages on a Linux server (see Figure 2). Using the Osmium Tool, time filter was first applied to retrieve the latest versions of the global OSM data on the last modified date of the reference data. The Canadian data were then clipped using the Canada.poly file from Geofabrik and the OSM History Splitter. Next, tag filter was executed in Osmosis to keep all motorways with a "highway" tag (all but footway, steps, path, track, raceway and cycleway, because the number of trails in OSM is much more than that in DMTI). Finally, street networks in Canada were loaded from PBF to the PostgreSQL database combining exports from Osmosis (using PostGIS simple schema for separate tag tables) and Imposm 3.

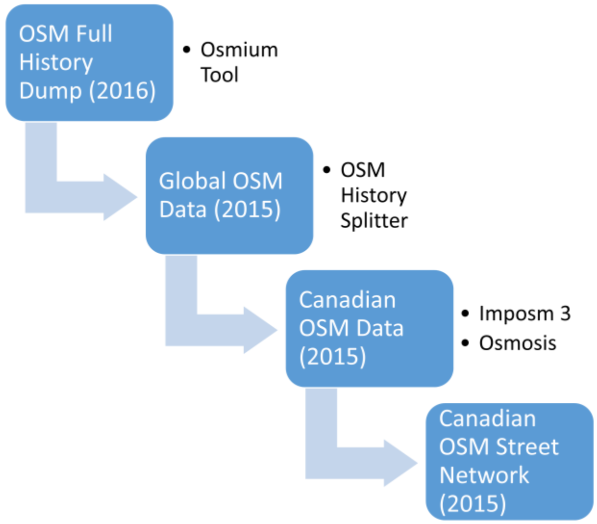

Figure 2. OSM data extraction

Quality measures were analyzed by the following methods (see Table 1) after projecting both datasets into NAD 1983 Statistics Canada Lambert. Road classification (e.g., primary highways) (see Table 2) and corresponding attributes (e.g., street name) were first pre-processed based on Zhang and Malczewski (2017). All OSM street name components (prefix and suffix directions and street types) except core street names were cleaned, capitalized, and transformed to abbreviated forms (e.g., BLVD for Boulevard) to match the format in DMTI using a self-developed python script. Completeness was then evaluated by total road length and road density. This unit-based approach was widely used in previous studies (e.g., Haklay 2010; Zielstra and Zipf 2010). Next, geometric and attribute feature changes were detected using ArcGIS. Only matched roads were kept for the following steps. Buffer analysis was employed to measure the positional accuracy (Goodchild and Hunter 1997). Buffers of widths from 5 to 30 m, with a 5-m interval, were generated around the reference street networks, and the percentages of OSM roads that fell within the buffers were computed. In terms of attribute accuracy, tag presence reflects the completeness of road attributes through counting the number of non-empty OSM tags when reference tags are available, and Levenshtein distance (Levenshtein 1966) represents the steps required to transform one string to another. The absolute differences of numeric attributes, such as road classification, were calculated to evaluate semantic accuracy. Source information was filtered individually to figure out the impacts of data import on Canadian OSM quality. It is worth noting that some steps mentioned above (e.g., geometric feature matching and buffer analysis) were completed using the “divide-and-conquer” approach (Cormen et al. 2009), because the entire dataset was too big to be processed within a reasonable amount of time and sometimes caused software crashes. National street networks were split into census tracts or dissemination areas for computation, then merged back for data report and visualization.

Table 1. Overview of the methods

Measures	Methods
Completeness	Road length (e.g., Haklay 2010; Zielstra and Zipf 2010)
	Road density (Hochmair, Zielstra, and Neis 2015)
Positional accuracy	Buffer analysis (Goodchild and Hunter 1997)
Attribute accuracy	Tag presence (Ludwig, Voss, and Krause-traudes 2011)
	Levenshtein distance (Levenshtein 1966)
Semantic accuracy	Number difference (Ludwig, Voss, and Krause-traudes 2011)
Lineage	Source
	Version
	Last modified date

Table 2. Matches of road classification (adapted from Zhang and Malczewski 2017)

Rank	Road Classification of DMTI	Road Classification of OSM
0	Not available	Unclassified
1	Expressways	Motorway, motorway_link
2	Primary highways	Trunk, trunk_link
3	Secondary highways	Primary, primary_link
4	Major roads	Secondary, secondary_link
5	Local roads	Tertiary, tertiary_link, residential, service
6	Trails	Footway, steps, path, track, raceway, cycleway
7	Proposed roads	Not available
8	Proposed highways	Not available

4 RESULTS AND DISCUSSION

4.1 Completeness

Figure 3 illustrates the total road length differences between the source and reference datasets. A positive number means DMTI has a longer road length, and a negative number represents OSM has better completeness. Results are aggregated based on DMTI's road classification (Table 2). This situation was also identified in the U.S. (Zielstra and Hochmair 2012) and Germany (Neis, Zielstra, and Zipf 2011). A significantly higher number of total road lengths in DMTI can be observed, which is consistent with previous findings in Germany (Neis, Zielstra, and Zipf 2011; Zielstra and Zipf 2010) and the U.S. (Zielstra, Hochmair, and Neis 2013). However, Zhang and Malczewski (2017) found that in London, Ontario, OSM had a longer total road length, which indicates the spatial heterogeneity of OSM quality in Canada. In the same study, the unclassified OSM roads were discovered to be mainly local roads through manual examination (Zhang and Malczewski 2017), which may be the case nationally as well. It is worth noting that although granularity affects road length, the differences between the two datasets are large enough to reach the conclusion.

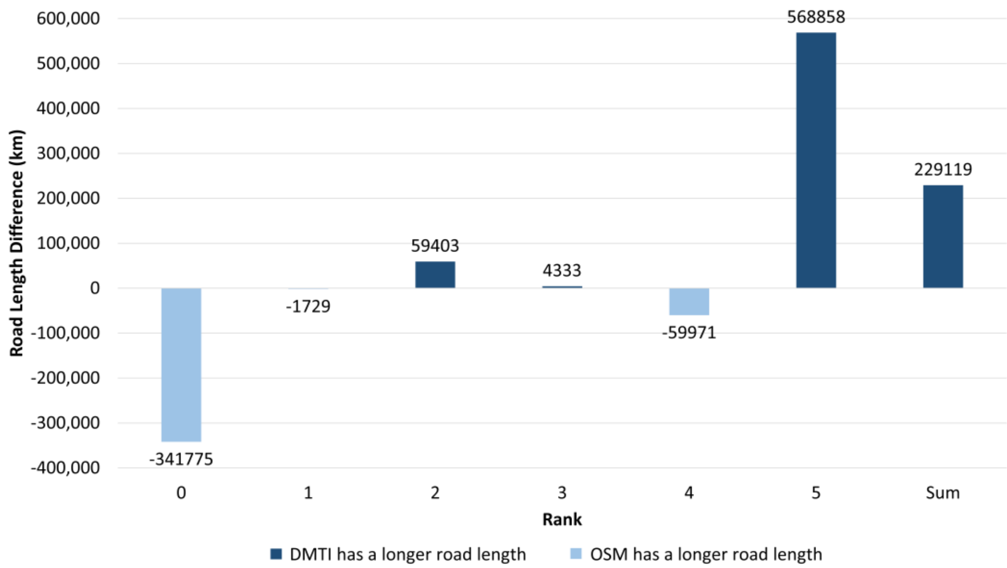

Figure 3. Classified road length differences

Using a cell size of 250 m, Figure 4 shows the spatial distribution of the differences in road density. The cell size is identical to the case study of London, Ontario (Zhang and Malczewski 2017), which provides enough details and can be computed within a reasonable amount of time. The top and bottom 0.5% of the data have been removed to reduce the effects of outliers. Here, green pixels represent a higher road density of DMTI; pink pixels represent a higher road density of OSM; and yellow pixels represent no difference. The maximum absolute value of green pixels (4.12) is significantly larger than that of pink pixels (0.52), indicating that the overall road density of DMTI is higher than that of OSM. In many cases, urban regions such as the Great Toronto Area (GTA), Ottawa and Quebec City have higher road densities in OSM, while remote regions such as the northern territories have higher road densities in DMTI. This pattern is similar to the OSM street networks of Germany in 2009, where completeness ranged from 97% in densely populated zones to 18% in uninhabited areas (Ludwig, Voss, and Krause-traudes 2011). The significant longer length of local roads in DMTI (see Figure 3) can probably be explained by Figure 4. Compared to the relatively small areas of urban centres, the vast rural regions in Canada received less attention from the OSM contributors, which resulted in the shorter length of local roads in OSM. Saskatchewan (SK) and Newfoundland and Labrador (NL) have an “anomalous” spatial pattern though, where OSM generally outperforms DMTI in road density. Data import, which started near the end of 2008 in Canada, plays an important role here. Manual comparisons in the provinces found that many unpaved roads were present in OSM but not DMTI. These roads were sourced from GeoBase in urban Saskatchewan and NLRoads in rural Newfoundland. The overall urban-focused pattern is not impacted because of the relative small number of road segments in Newfoundland.

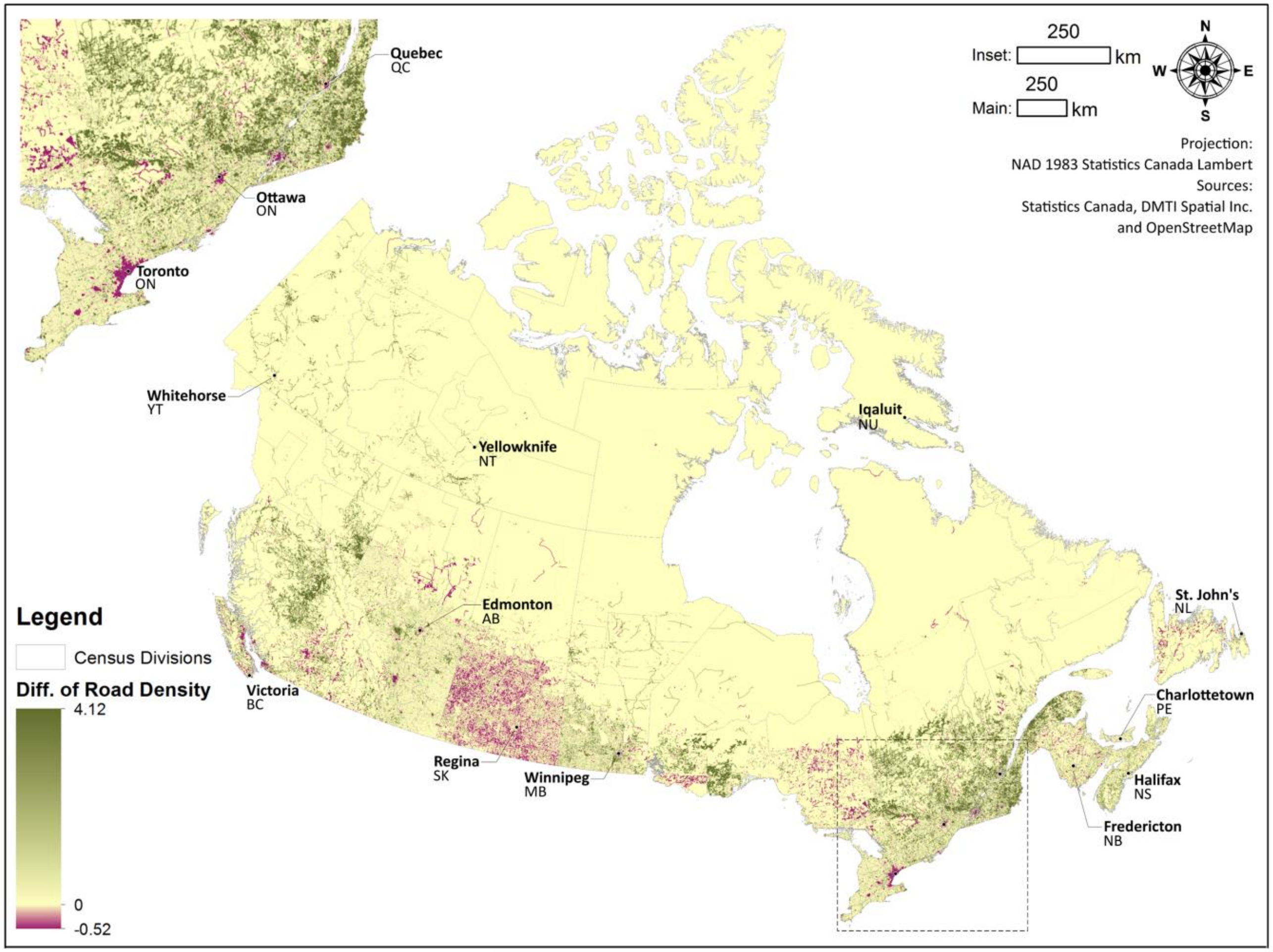

Figure 4. Differences of road density

4.2 Positional Accuracy

Figure 5 shows the results of the buffer analysis. Approximately 60% of roads of DMTI have a 25-m or better positional accuracy, while the rest have a guarantee of 30-m accuracy. Overall, 91.5% of roads of OSM fall within the 30-m buffer, in which 77.5% are within 5 m, 8.3% between 5 and 10 m, and 5.7% between 10 and 30 m away from the reference dataset. It is worth noting that an OSM road segment would be less likely to fall within a buffer if the corresponding positional accuracy of the reference dataset is too low. Compared to Germany in 2009, the Canadian OSM streets have a 4.5% increase in positional accuracy within the 5-m buffer, but a 8.5% decrease in total – all German OSM streets were within the 30-m buffer of Navteq data (Ludwig, Voss, and Krause-traudes 2011). In terms of road classification, primary and secondary highways have relatively low positional accuracy, whereas local roads are the most accurate ones at the 5-m buffer. This phenomenon can probably be explained by Linus' Law, which was found to generally apply to positional accuracy in London, England where local roads received more attentions (Haklay et al. 2010). Moreover, primary and secondary highways are usually wider than local roads, so errors are more likely to occur if the highways were traced by road lanes instead of centre lines. Similarly, the expressways have a relatively low positional accuracy at the 5-m buffer but jumped to the top at the 30-m buffer potentially because of data import. Finally, since rank 2 and 3 only account for 12% of the road segments in the reference dataset while local roads take over 83%, extreme cases have a higher impact on the accuracy of the primary and secondary highways as well.

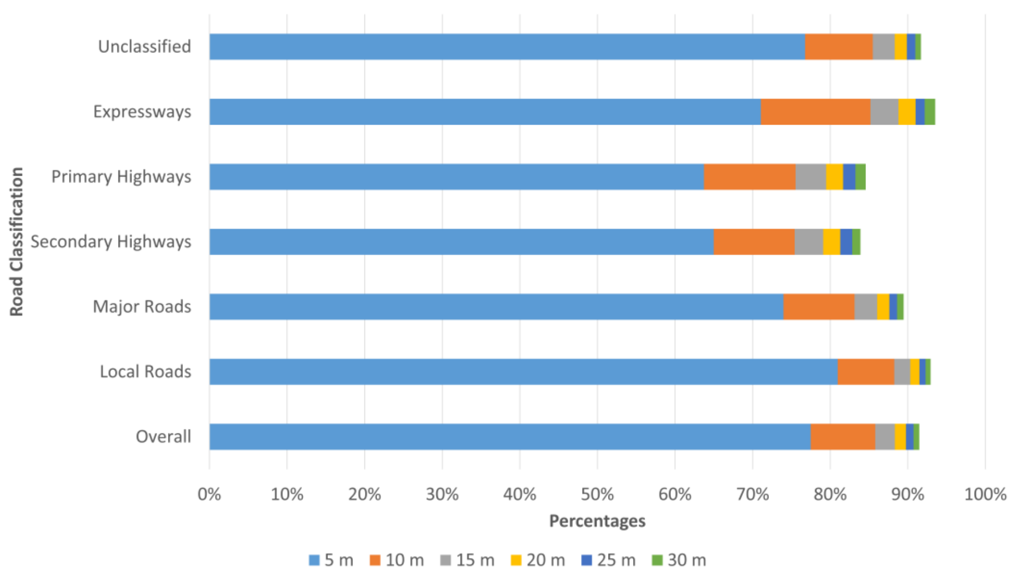

Figure 5. Results of the buffer analysis

4.3 Attribute Accuracy

Figure 6 shows the tag presence rates of Canadian OSM street names, which have been divided into five components to match the attributes in DMTI. For the most part, French road names have prefix street types (e.g., Rue Sherbrooke Ouest), and English road names have suffix street types (e.g., Wonderland Road North). In comparison to London, Ontario (Zhang and Malczewski 2017), the national tag presence rates dropped from mostly 90% and above to a minimum of 52%, which once again indicates the spatial heterogeneity of the Canadian OSM quality. Suffix directions have close tag presence rates both locally and nationally. This may suggest OSM contributors either do not know or do not care about most of the suffix directions of Canadian streets. Core street names and suffix street types have the highest presence rates, which is understandable since a common street name consists of the two components. Like the results in Section 3.1.2, Linus' Law plays a role in attribute accuracy as well. Primary and secondary highways usually have lower tag presence rates, while major and local roads have higher percentages of presence, except for core street names which are potentially influenced by the data import from GeoBase. Neis, Zielstra, and Zipf (2011) discovered that unclassified roads had the highest ratio (61%) of missing names or route numbers in Germany in 2011, which is not the case in Canada. Overall, the tag presence rates of Canadian OSM street names are comparable with those in Germany (82.5% to 94.4%) in 2009 (Ludwig, Voss, and Krause-traudes 2011).

Figure 7 shows the Levenshtein distance of the Canadian OSM street names. The larger the Levenshtein distance, the greater the variations between strings. A Levenshtein distance of 0 represents that two strings are the same, and a value of 1 to 3 usually means a typo. Prefix and suffix directions, with a maximum text length of 1, have almost perfect spelling accuracy. The percentages of completely matched prefix and suffix street types and core street names are about 87%, 71% and 57% respectively. Some extreme Levenshtein distances with a maximum value of 79 were identified in core street names; however, this component also has the largest maximum text length. The average Levenshtein distance of core street names is 3.09, which is higher than that of core street names (0.80) in London, Ontario (Zhang and Malczewski 2017).

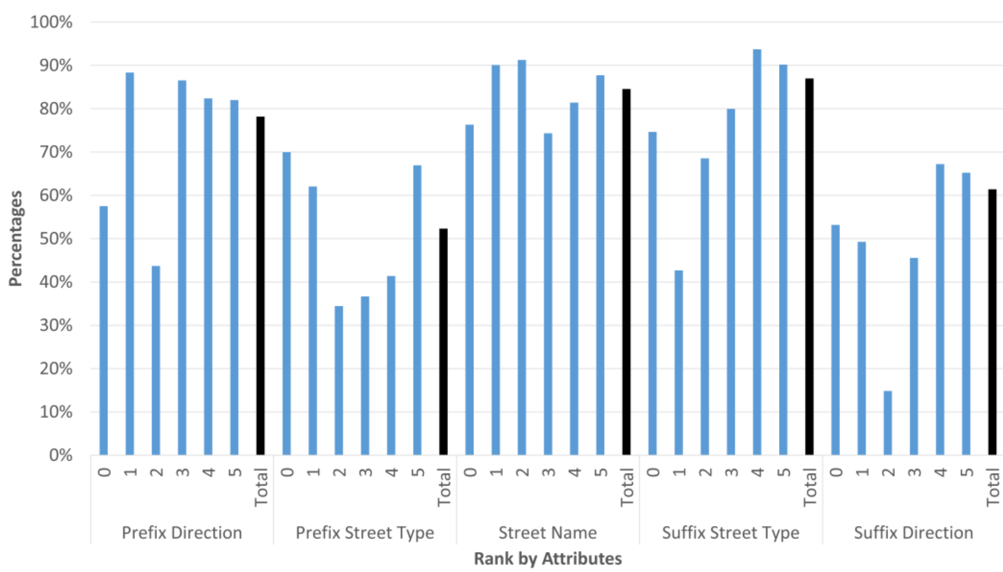

Figure 6. Tag presence rates of street names

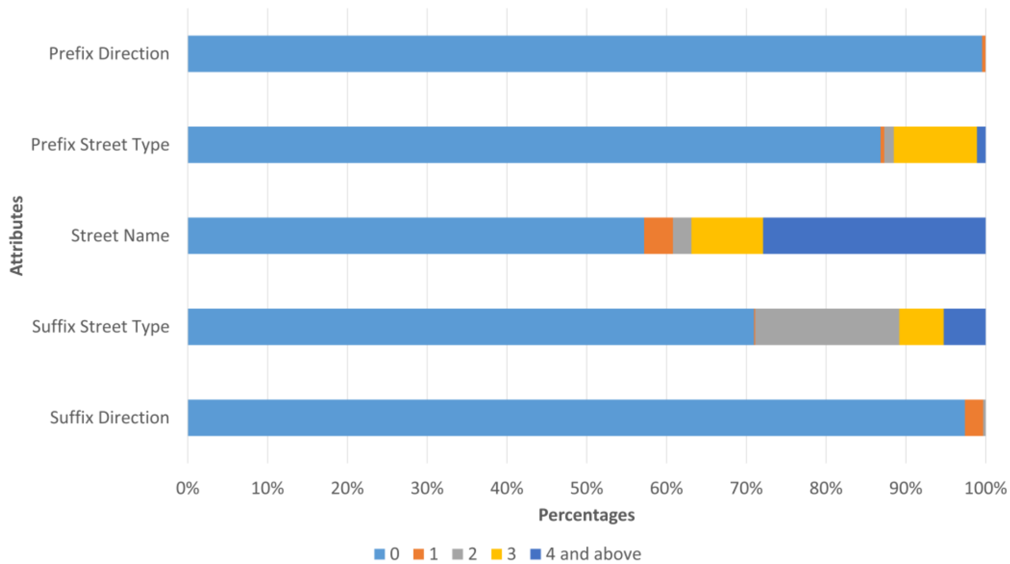

Figure 7. Levenshtein distance of street names

4.4 Semantic Accuracy

Table 3 shows the absolute differences of numeric attributes in the two datasets. More than 40% of the road networks were misclassified in OSM. Among those, more than half were local roads but remained unclassified. One possibility behind the misclassification is the import from GeoBase, which was the case with the TIGER/Line import in the U.S. (Zielstra, Hochmair, and Neis 2013). The majority of the rest of the misclassified roads had a rank difference of 1 (e.g., a local road is classified as a major road or vice versa), which is understandable because of the incompatible classification schema and classification ambiguity and conceptual plausibility (Ali et al. 2014). In fact, because of its worldwide coverage, OSM only has a comparison-based road classification system ("from the most to least important"). For example, in OpenStreetMap Wiki, primary roads are defined as "the next most important roads in a country's system. (Often link larger towns.)". Without industrial specifications, this definition can be extremely confusing without a context. Some OSM road classifications (e.g., motorway and trunk) have an international equivalence section under their dedicated wiki pages, but Canada is not listed in any of these incomplete sections.

In terms of number of lanes, lots of data were missing in DMTI (value equals to 0). Thus, the semantic accuracy of presented number of lanes in OSM could not be fully evaluated, and the actual accuracy rate was probably higher than 39.5%. In contrast, tunnel, bridge and one-way flags have nearly perfect accuracy. It is worth noting that the total numbers of positive flags (value equals to 1) in both datasets are very small, which leads to this high accuracy.

Table 3. Number differences of attributes

Differences	Percentages	Differences	Percentages
Rank		Tunnel	
0	59.1%	0	100.0%
1	12.3%	1	0.0%
2	1.5%	Bridge	
3	0.4%	0	99.2%
4	0.1%	1	0.8%
5	26.6%	-	-
Number of Lanes		One-way	
0	39.5%	0	99.2%
1	3.0%	1	0.8%
2	56.9%	2	0.0%
3 to 9	0.7%	-	-

4.5 Lineage

Using the associated metadata and a self-developed python script, a summary was produced for the existing sources in Canadian OSM. The major source is GeoBase, which consisted of around 77% of the entire Canadian OSM road networks. About 17% of the roads had unknown sources, and the remaining 6% were attributed to Bing, Yahoo, Newfoundland and Labrador Roads and others.

Figure 8 presents the percentage differences of selected quality metrics between GeoBase-sourced road segments and the entire OSM dataset, which shows the impacts of data import on attribute and semantic accuracy. Completeness and positional accuracy were not included because of their aggregated results. Most quality measures have slightly improved accuracy percentage-wise, which is probably due to the removal of outliers from vandalism. Four quality measures have decreased accuracy and require further exploration for logical explanations. One possibility is that contributions from remote mappers without local knowledge substantially increased the complexity of OSM quality. As remote mappers, ground-truth is not an option, so they solely rely on the (potentially) out-of-date top-view satellite images with limited resolution. Without local knowledge, remote mappers also cannot fully contribute and validate attribute information. In the case of Canada where about 77% of the road networks are sourced from one governmental database, data import can potentially be a more essential factor. Bots are able to create various types of systematic errors, which can be difficult when tracking and understanding from the perspective of human behaviors.

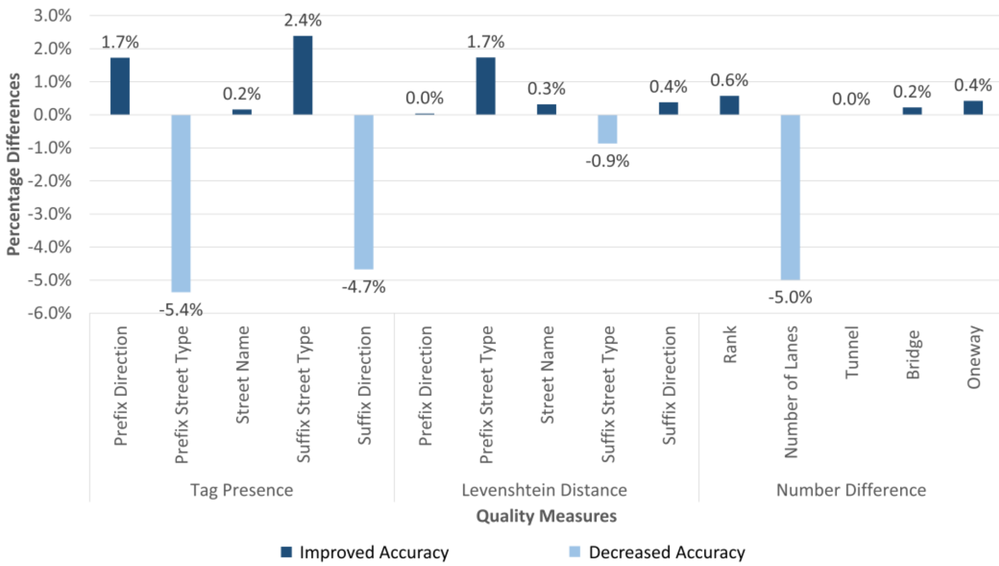

Figure 8. Percentage differences between GeoBase-sourced road segments and the whole dataset

5 SUMMARY AND OUTLOOK

This study evaluated the extrinsic quality of the Canadian OSM street networks in terms of completeness, positional accuracy, attribute accuracy, semantic accuracy and lineage. The overall OSM quality in Canada is comparable with DMTI, although spatial heterogeneity is a common theme across all quality measures. Urban areas received more contributions than rural areas, and footways were favoured over motorways by contributors in general. GeoBase-sourced road segments have slightly and commonly improved quality. For future work, other features, such as buildings and points of interest, can be evaluated. Measures such as temporal quality and logical consistency can be examined in addition.

Results of this study have some implications on OSM quality improvement in the future. For instance, do the activities of remote mappers decrease the overall quality of the project? Is local knowledge necessary to create accurate maps? How can the uniformity of OSM quality be increased? Are strict specifications better or worse for the project, and should contributors have their current degree of freedom? While data import boosts up the number of map features dramatically in a short period, does this action impair the motivations of OSM contributors and the sustainability of the project in the long term? These questions are worth discussing and can potentially contribute to quality improvements in VGI.

REFERENCES

Ali, A.L., Schmid, F., Al-Salman, R. and Kauppinen, T. (2014) Ambiguity and Plausibility: Managing Classification Quality in Volunteered Geographic Information. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL '14* (pp. 143–152).

Antoniou, V. and Skopeliti, A. (2015) Measures and Indicators of VGI Quality: An Overview. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 345–351.

Balram, S. and Dragicevic, S. (2008) Collaborative Geographic Information Systems: Origins, Boundaries, and Structures. In *E-Learning for Geographers: Online Materials, Resources, and Repositories: Online Materials, Resources, and Repositories* (pp. 1-23). Hershey: IGI Global.

Barron, C., Neis, P. and Zipf, A. (2014) A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6), 877–895.

Coleman, D.J., Georgiadou, Y. and Labonte, J. (2009) Volunteered Geographic Information: The Nature and Motivation of Produsers. *International Journal of Spatial Data Infrastructures Research*, 4(1), 332–358.

Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2009) *Introduction to Algorithms*. 3rd ed. Cambridge: The MIT Press.

DMTI Spatial Inc. (2015) *CanMap® Content Suite Support Manual*. Markham: DMTI Spatial Inc.

Elmes, G., Dougherty, M., Challig, H., Karigomba, W., McCusker, B. and Weiner, D. (2005) Local Knowledge Doesn't Grow on Trees: Community-Integrated Geographic Information Systems and Rural Community Self-Definition. In P. F. Fisher (Ed.). *Developments in Spatial Data Handling* (pp. 29–39). Berlin: Springer.

Elwood, S. (2006) Critical Issues in Participatory GIS: Deconstructions, Reconstructions, and New Research Directions. *Transactions in GIS*, 10(5), 693–708.

Flanagin, A.J. and Metzger, M.J. (2008) The Credibility of Volunteered Geographic Information. *GeoJournal*, 72(3–4), 137–148.

Goodchild, M. (2007) Citizens as Sensors: The World of Volunteered Geography. *GeoJournal*, 69(4), 211-221.

Goodchild, M. and Hunter, G.J. (1997) A Simple Positional Accuracy Measure for Linear Features. *International Journal of Geographical Information Science*, 11 (3), 299–306.

Gröchenig, S., Brunauer, R. and Rehrl, K. (2014) Estimating Completeness of VGI Datasets by Analyzing Community Activity over Time Periods. In J. Huerta, S. Schade and C. Granell (Eds.). *Connecting a Digital Europe Through Location and Place* (pp. 3–18). Cham: Springer.

Haklay, M. (2010) How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design*, 37(4), 682–703.

Haklay, M., Basiouka, S., Antoniou, V. and Ather, A. (2010) How Many Volunteers Does It Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal*, 47(4), 315–322.

Harvey, F. (2013) To Volunteer or to Contribute Locational Information? Towards Truth in Labeling for Crowdsourced Geographic Information. In D. Sui, S. Elwood and M. Goodchild (Eds.). *Crowdsourcing Geographic Knowledge* (pp. 31–42). Dordrecht: Springer.

Hochmair, H.H., Zielstra, D. and Neis, P. (2015) Assessing the Completeness of Bicycle Trail and Designated Lane Features in OpenStreetMap for the United States and Europe. *Transactions in GIS*, 19(1), 63–81.

Jokar Arsanjani, J. and Bakillah, M. (2015) Understanding the Potential Relationship between the Socio-Economic Variables and Contributions to OpenStreetMap. *International Journal of Digital Earth*, 8(10), 1–16.

Levenshtein, V.I. (1966) Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10(8), 707-710.

Ludwig, I., Voss, A. and Krause-traudes, M. (2011) A Comparison of the Street Networks of Navteq and OSM in Germany. In S. Geertman, W. Reinhardt and F. Toppen (Eds.). *Advancing Geoinformation Science for a Changing World* (pp. 65–84). Berlin: Springer.

McKenzie, G. and Janowicz, K. (2014) Coerced Geographic Information: The Not-so-Voluntary Side of User-Generated Geo-Content. In *Eighth International Conference on Geographic Information Science*.

Neis, P., Zielstra, D. and Zipf, A. (2011) The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, 4(4), 1–21.

Neis, P., Zielstra, D. and Zipf, A. (2013) Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet*, 5(2), 282–300.

Over, M., Schilling, A., Neubauer, S. and Zipf, A. (2010) Generating Web-Based 3D City Models from OpenStreetMap: The Current Situation in Germany. *Computers, Environment and Urban Systems*, 34(6), 496–507.

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

Sieber, R. (2006) Public Participation Geographic Information Systems: A Literature Review and Framework. *Annals of the Association of American Geographers*, 96(3), 491–507.

Tenney, M. (2014) Quality Evaluations on Canadian OpenStreetMap Data. http://rose.geog.mcgill.ca/ski/system/files/fm/2014/tenney.pdf.

Turner, A. J. (2006) *Introduction to Neogeography*. Sebastopol: O'Reilly Media.

Van Oort, P. (2006) *Spatial Data Quality: From Description to Application*. Wageningen: Wageningen University.

Zhang, H. and Malczewski, J. (2017) Quality Evaluation of Volunteered Geographic Information: The Case of OpenStreetMap. In C. Campelo, M. Bertolotto and P. Corcoran (Eds.). *Volunteered Geographic Information and the Future of Geospatial Data* (pp. 19–46). Hershey: IGI Global.

Zielstra, D. and Hochmair, H.H. (2012) Using Free and Proprietary Data to Compare Shortest-Path Lengths for Effective Pedestrian Routing in Street Networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2299, 41–47.

Zielstra, D., Hochmair, H.H. and Neis, P. (2013) Assessing the Effect of Data Imports on the Completeness of Openstreetmap - A United States Case Study. *Transactions in GIS*, 17(3), 315–334.

Zielstra, D. and Zipf, A. (2010) A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. In *13th AGILE International Conference on Geographic Information Science* (Vol. 2010).