

## AN AUTOMATED GRASS-BASED PROCEDURE TO ASSESS THE GEOMETRICAL ACCURACY OF THE OPENSTREETMAP PARIS ROAD NETWORK

M. A. Brovelli, M. Minghini \*, M. E. Molinari

Department of Civil and Environmental Engineering, Politecnico di Milano, Como Campus, Via Valleggio 11, 22100 Como, Italy –  
(maria.brovelli, marco.minghini, moniaelisa.molinari)@polimi.it

Commission VII, SpS 10 - FOSS4G: FOSS4G Session (coorganized with OSGeo)

**KEY WORDS:** Accuracy, FOSS4G, GRASS, Open data, OpenStreetMap, Road network, Volunteered Geographic Information

### ABSTRACT:

OpenStreetMap (OSM) is the largest spatial database of the world. One of the most frequently occurring geospatial elements within this database is the road network, whose quality is crucial for applications such as routing and navigation. Several methods have been proposed for the assessment of OSM road network quality, however they are often tightly coupled to the characteristics of the authoritative dataset involved in the comparison. This makes it hard to replicate and extend these methods. This study relies on an automated procedure which was recently developed for comparing OSM with any road network dataset. It is based on three Python modules for the open source GRASS GIS software and provides measures of OSM road network spatial accuracy and completeness. Provided that the user is familiar with the authoritative dataset used, he can adjust the values of the parameters involved thanks to the flexibility of the procedure. The method is applied to assess the quality of the Paris OSM road network dataset through a comparison against the French official dataset provided by the French National Institute of Geographic and Forest Information (IGN). The results show that the Paris OSM road network has both a high completeness and spatial accuracy. It has a greater length than the IGN road network, and is found to be suitable for applications requiring spatial accuracies up to 5-6 m. Also, the results confirm the flexibility of the procedure for supporting users in carrying out their own comparisons between OSM and reference road datasets.

### 1. INTRODUCTION

Volunteered Geographic Information (VGI), which refers to the provision of geospatial contents by ordinary people (Goodchild, 2007), has now become a well-established practice in the GIS and geography domains. Started in 2004 OpenStreetMap (OSM, <http://www.openstreetmap.org>) is today the most popular VGI project. Until now it has attracted about two and half million contributors (<http://wiki.openstreetmap.org/wiki/Stats>) and it has produced the largest and most diverse geospatial database of the world. The increasing success of OSM is largely due to the open license of data (<http://www.openstreetmap.org/copyright>) which allows everyone to copy, distribute, transmit and adapt data provided that credit is made to OSM and its contributors. As a consequence, interest has recently grown in both the usage of OSM data – e.g. by companies, institutions and humanitarian organizations – and the multi-disciplinary research – e.g. from the analytical, social, political and ethical perspectives – on the OSM project by the academic community (Jokar Arsanjani et al., 2015a).

The aspect of VGI that literature has most addressed is probably that of quality, fitness-for-use and fitness-for-purpose (Ali and Schmid, 2014). Thanks to its popularity OSM is exploited as a primary source for the assessment of VGI quality. The multiple research studies in this field can be classified into three groups. The first includes studies where OSM datasets are compared to reference datasets produced by authoritative bodies and thus considered as ground truth. Notable examples are provided by Haklay (2010), Girres and Touya (2010) and Neis et al. (2011). The second group of studies on OSM quality is focused on user activities (see e.g. Jokar Arsanjani et al., 2015b) while the third concentrates on the history of OSM data, which is stored in the OSM database together with the data. Examples include the works by Keßler et al. (2011) and Keßler and de Groot (2013).

The greatest number of available studies on OSM data quality (and similarly the study presented in this paper) belong to the first group. Comparing OSM datasets with the corresponding authoritative datasets, such as those released by Commercial Mapping Companies (CMC) and National Mapping Agencies (NMA), is the most immediate and rational operation one can think to. Availability of reference data for comparison with OSM is increasingly favoured by the open licenses of geospatial datasets released by NMA and CMC and, in some cases, by the permission of exploiting proprietary datasets for pure research purposes (e.g. Ludwig et al., 2010). However, the fact that these datasets are accessible does not imply that OSM comparisons against them can be easily carried out. As observed by Brovelli et al. (2016), VGI and authoritative datasets exist for different reasons and are created, managed, curated, edited and updated under very different workflows. For this reason, the procedures proposed so far are typically very specific and tightly coupled to the data model and characteristics of the reference dataset under consideration. Also, very few studies can be found in literature where some software or automated procedures are developed to facilitate these comparisons. The result is that the methods and procedures available in the literature, despite being rigorous and scientifically valid, are generally hard to be reused or extended when a different reference dataset is involved.

This paper fits in this context by focusing the quality assessment on OSM road network datasets. These datasets are one of the most typical elements occurring within the OSM database (the OSM project itself was originally born for mapping roads) and their quality is crucial for many applications such as routing and navigation. The paper describes the quality assessment of the Paris OSM road network dataset through a comparison against the French official dataset provided by the French National Institute of Geographic and Forest Information (IGN). The work is performed through an automated procedure developed by the

---

\* Corresponding author

authors and released with open source license (Brovelli et al., 2016). In contrast to the typical issues of existing procedures described above, this is a general tool providing users with a high degree of flexibility in order to facilitate the comparison of OSM against any reference road network dataset. Provided that users are familiar with the reference road dataset used they can adjust the values of the parameters involved, thus customizing the whole procedure and obtaining their own OSM data quality assessment.

The remainder of the paper has four parts. First, it reviews the extant literature relevant to the assessment of OSM road quality. Then the procedure for comparing OSM against reference road network datasets is briefly recalled and described. Next, the application of the procedure on the Paris case study is presented in detail. The paper concludes with a discussion on the main outcomes of the study, lessons learnt and directions for future research.

## 2. RELATED WORK

The parameters which are typically used to measure VGI quality include some traditional assessment parameters for geospatial data such as completeness, up-to-dateness, positional accuracy, attribute accuracy, temporal accuracy, semantic accuracy and logical consistency (see e.g. Guptill and Morrison, 1995). More recently the concepts of fitness-for-use and fitness-for-purpose were introduced to recognise that quality has different degrees of suitability according to specific purposes and users' demands (Devillers and Jeansoulin, 2006). Many studies investigating the quality of OSM data are available in literature. Most of them are specifically focused on the assessment of road network quality. In the following, a review of the most notable works focused on the comparison between OSM and authoritative road datasets – the same approach applied in this study – is presented.

The earliest studies on the quality of OSM road networks were proposed by Haklay (2010) and Ather (2009) who measured the completeness and positional accuracy of London and England OSM road datasets compared with reference data provided by the Ordnance Survey, the UK NMA. Girres and Touya (2010) extended this work to France and provided more different statistics (geometric accuracy, attribute accuracy, semantic accuracy, completeness, logical consistency, temporal accuracy, lineage, and usage) about OSM roads by means of a comparison with the data provided by the French IGN. Zielstra and Zipf (2010) investigated the completeness of German OSM roads by comparing them with data from TeleAtlas and Helbich et al. (2012) measured the positional accuracy of OSM and TomTom data using official survey data as the reference dataset. All these works highlighted the high potential of VGI and OSM as accurate data sources but also emphasized some limitations due to their heterogeneity in terms of completeness, with coverage typically lacking in rural areas. Other comparisons between OSM and reference road network datasets from commercial providers were proposed by Ludwig et al. (2011) and Neis et al. (2011) who used Navteq and Tom Tom, respectively. In particular, Ludwig et al. (2011) analyzed OSM road network datasets for populated roads in Germany showing a satisfactory completeness of object/names and accuracy of OSM data for map generation purpose. Neis et al. (2011) studied the evolution of the German OSM road dataset from 2007 to 2011. The work showed a significant increase in street network data (exceeding Tom Tom dataset) but still highlighted some deficiencies in route network for car navigation. A similar study was conducted for Florida by Zielstra and Hochmair (2011) who compared the OSM road dataset with commercial datasets from Navteq and TeleAtlas. Heterogeneity was shown in terms of completeness. However, in contrast with European results, OSM coverage in rural areas was found to be generally higher than in urban areas.

Many other works investigating OSM road network quality were performed in the same years. Kounadi (2011) compared the OSM road dataset in Athens with the data provided by the Hellenic Military Geographical Service (HMGS) highlighting an overall good data quality, especially for length completeness, name accuracy and positional accuracy. Ciepluch et al. (2011) analyzed the completeness of Ireland OSM roads with respect to the Ordnance Survey Ireland dataset and concluded that OSM in city areas could be an accurate alternative to commercial data to develop Location-based Services (LBS). Jokar Arsanjani et al. (2013) compared the OSM roads of Heidelberg (Germany) with an authoritative dataset from the Federal Agency for Cartography and Geodesy. OSM data reliability was assessed in terms of completeness, positional accuracy and semantic accuracy, according to different categories of contributors.

More recently, Graser et al. (2014) proposed a novel toolbox for street network comparison. The toolbox was applied in the Vienna region to evaluate OSM road data quality compared to the official Austrian dataset, the “Graph Integration Platform” (GIP), and showed good results for positional accuracy and network length with still some deficiencies in terms of attribute completeness. Forghani and Delevar (2014) compared the road datasets from OSM and the Municipality of Tehran computing new quality metrics indexes (Minimum Bounding Geometry, Directional Distribution, etc.) which highlighted a medium quality of OSM data. Siebritz and Sithole (2014) analyzed the quality of OSM roads and buildings in South Africa using as a reference the dataset provided by the South African NMA (CD-NGI). Authors showed the heterogeneity of OSM quality across South Africa and stated that, despite some OSM quality indexes are good, the dataset did not meet the accuracy requirements for the integration with the CD-NGI database. With the same purpose of integrating VGI and authoritative data, Fairbairn and Al-Bakri (2013) proposed a method to assess the positional and shape quality of OSM road datasets. The procedure was tested in areas of UK and Iraq through a comparison with official datasets from the Ordnance Survey and the General Directorate for Survey, respectively, and results showed that integration of OSM data for large scale mapping applications was not viable.

Finally the procedure used in this study, developed by Brovelli et al. (2016), was applied by Antunes et al. (2015) to assess the positional differences between the OSM road network available for some regions of Coimbra Municipality (Portugal) and the reference data provided by the Coimbra City Hall.

## 3. METHODOLOGY

As mentioned above, the comparison between the OSM and the IGN Paris road network datasets, described later in Section 4, is performed using a novel procedure that was recently developed by the authors. A detailed explanation is provided by Brovelli et al. (2016). In the current state the procedure returns quantitative measures of completeness and positional accuracy of the OSM road network dataset. In the following a brief summary of this procedure (including some new features which were recently added) is presented to highlight its main purpose and sequence of operations. Section 4 will then show in detail how each step of the procedure is implemented on the Paris case study and the corresponding results in terms of OSM quality assessment.

From the technical perspective, the procedure consists of a set of three independent modules for the open source GRASS GIS software (<http://grass.osgeo.org>), that are written in Python and available with a graphical user interface (GUI). The modules are in turn released as open source. Their most up-to-date version is maintained at <https://github.com/MoniaMolinari/OSM-roads-comparison/tree/master/GRASS-scripts>. Using the abbreviations OSM for the OpenStreetMap road network dataset and REF for

the reference road network dataset exploited in the comparison, the main role of each module is described in the following:

1. Preparation and preliminary comparison of OSM and REF, computation of global statistics on their geometrical similarity (providing measure of OSM completeness), and sensitivity analysis on the value of buffer parameter which is key in Step 2.
2. Geometric preprocessing of OSM to only extract its subset of roads having a correspondence in REF.
3. Evaluation of OSM positional accuracy using a grid-based approach. Two separate analysis are available. The first computes the maximum deviation of OSM from REF; the second evaluates OSM positional accuracy against one or more user-specified thresholds.

Each of these three steps is fully driven by the user. This means that the user can adjust the numerical values of the parameters involved according to both the customization of the analysis he wants to achieve and the specific road network reference dataset he is using. Indeed it is assumed that the user is highly familiar with the characteristics of this dataset such as its data structure, nominal scale and accuracy. Also some operations are optional, i.e. the user can decide whether to execute them or not.

#### 4. RESULTS

This section describes the application of the presented GRASS-based procedure to compare the OSM and authoritative road network datasets for Paris city. The latter was made available by the French IGN in the frame of the European COST Action IC1203 'European Network Exploring Research into Geospatial Information Crowdsourcing: Software and methodologies for harnessing geographic information from the crowd (ENERGIC)' in which the authors are involved. The three steps are separately presented in the following together with the related results.

##### 4.1 Step 1: Preliminary comparison of the datasets

In this first step the two datasets are imported in GRASS and undergo an initial comparison. A preliminary requirement for the whole procedure is that the datasets are actually comparable, i.e.: a) the scale and level of detail of the reference dataset are similar to those of OSM; and b) the two datasets represent exactly the same categories of roads. Once verified that the IGN and OSM datasets are fully comparable, the input required from the user is a list of buffer values to be computed around the IGN dataset. For each buffer, the algorithm computes the length of the OSM roads included in the buffer (both in map units and as a percentage of the total length). A plot is then generated (see Figure 1) which shows how the length of OSM roads included in the buffer increases when increasing the buffer. This is an addition compared to the original work by Brovelli et al. (2016) and acts as a sensitivity analysis on the buffer parameter that the user has to choose in Step 2 (see Subsection 4.2).

A text file is also returned which, in addition to the length of OSM roads falling inside and outside each buffer, provides the total length of the two datasets. The total length of OSM dataset is 3124 km and the total length of IGN dataset is 2686 km. Thus OSM dataset has 438 km of roads more than the IGN dataset, a quantity corresponding to about 16% of the total length of IGN road dataset. This is a first measure of the high completeness of the Paris OSM road network, which confirms the outcomes of previous literature studies (e.g. Zielstra and Zipf, 2010) stating that, the more urban is an area, the higher is the number of contributors and the higher the completeness of OSM data.

Having initially verified that the two datasets include the same categories of roads, the reasons why the OSM dataset is much

longer than the IGN dataset can be mainly explained as follows. In the Paris OSM database many more pedestrian paths are mapped compared to the IGN dataset. Also, due to a difference between the OSM and IGN data models, cycleways as well as separate carriageways belonging to the same road are mapped in OSM as separate entities, while they converge in a single road in the IGN dataset. Figure 2 shows examples of these cases.

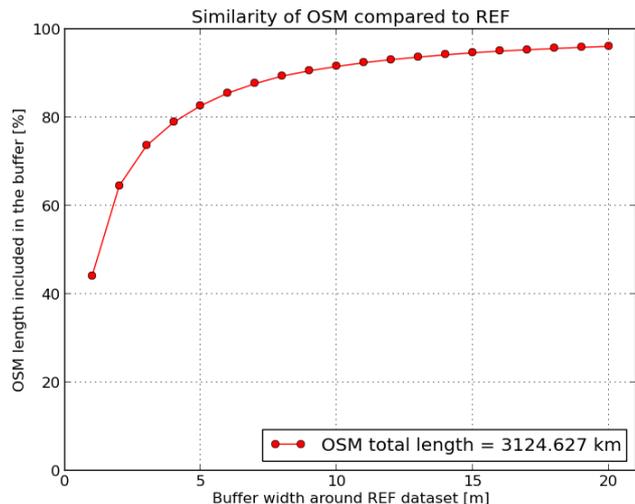


Figure 1. Length of Paris OSM roads (as a percentage of their total length) included in the buffers from 1 to 20 m around the Paris IGN road dataset

##### 4.2 Step 2: Geometric pre-processing of OSM dataset

After getting an idea of the similarity between the two datasets, the purpose of Step 2 is to perform a geometric pre-processing of the OSM dataset in order to make it fully comparable with the reference dataset. The comparison and quality assessment will be then performed in Step 3. The pre-processing is aimed at removing from the OSM dataset all the portions of roads which have no correspondence in the IGN dataset. This is performed through a well-defined series of processing operations, which are described in the following by justifying the choices made in terms of parameter values.

The first operation is proposed as optional (i.e. the user can skip it) and consists of a generalization of the IGN dataset with the only purpose of reducing the processing time required by the following operations. The procedure makes use of the Douglas-Peucker generalization algorithm. The choice of the value for the threshold parameter of the algorithm (which corresponds to the maximum error introduced by the generalization) is left to the user. Ideally, this value should be less than or equal to the nominal accuracy of the reference dataset used. Knowing from the French IGN that the Paris road dataset has an accuracy of 1 m, the generalization is run three times with thresholds of 1 m, 0.5 m and 0 m (i.e. with no generalization). This allows to run a sensitivity analysis on this parameter to assess the variation of results on OSM positional accuracy after the application of Step 3. This analysis will be addressed in Subsection 4.3.3.

At this point the OSM dataset is pre-processed by removing all its road portions falling outside a buffer around the IGN dataset. The choice of the buffer value is again left to the user. The idea is to apply a buffer around the IGN roads such that, among all the OSM roads, only the ones corresponding to the IGN roads are included. The sensitivity analysis run in Step 1 can facilitate this choice. Referring to Figure 1, the best choice should fall on a buffer value in correspondence of which the increase of the percentage of OSM roads included in the buffer starts to

decrease (or, in other words, the curve starts to become horizontal). According to this method, a value of 11 m is chosen for the buffer.

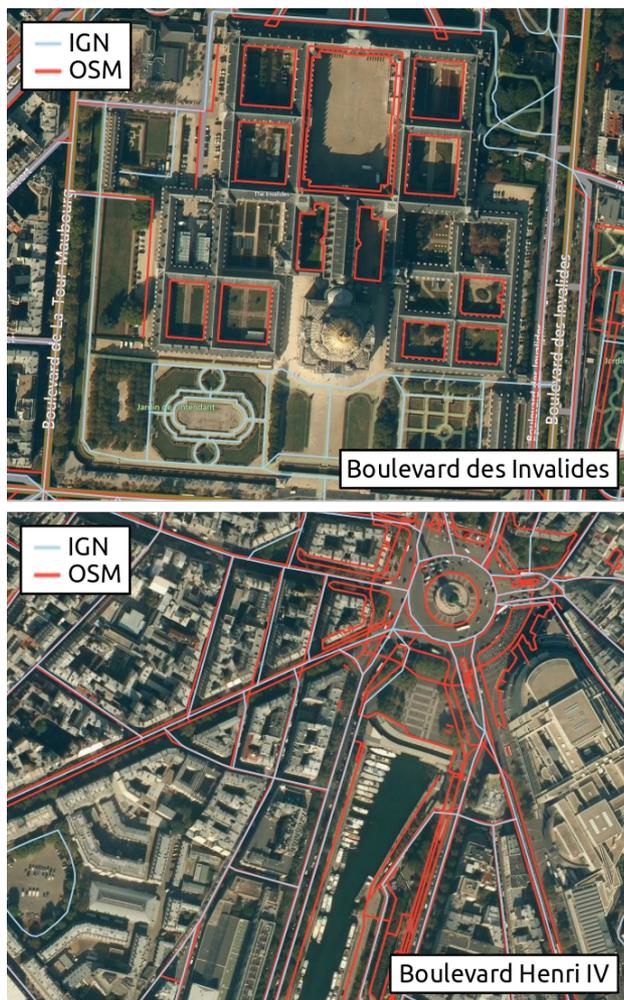


Figure 2. Examples of areas where OSM dataset shows more pedestrian pathways than IGN dataset (top) and multiple line entities for the same IGN road (bottom) (basemap: © Bing Maps; data: © IGN and © OpenStreetMap contributors)

Therefore, the OSM dataset is cleaned by removing all the road portions falling outside a buffer of 11 m computed around the IGN dataset. A subsequent refinement of this OSM “cleaning”, which further deletes the individual OSM line segments having no correspondence in the IGN dataset, is performed. Interested readers are invited to check the details in Brovelli et al. (2016). The total length of the pre-processed OSM dataset (included in a text file returned from Step 2) is 2880 km. This means that 244 km of OSM roads have been removed but the pre-processed OSM dataset (which now includes only the same roads as IGN) has still 194 km of roads more compared to the IGN dataset.

### 4.3 Step 3: Evaluation of OSM positional accuracy

The purpose of Step 3 is to evaluate the positional accuracy of the OSM dataset, which has been pre-processed in Step 2 and is now fully comparable to the IGN dataset. To take into account the typically heterogeneous nature of OSM, whose geometrical accuracy can vary from area to area, a grid-based approach is used. The procedure allows either to import a vector layer to be used as the grid or to manually create the grid. In this case, a grid with square cells of size 1 km x 1 km is created (see Figure

3 and Figure 5). Two separate analysis to assess the positional accuracy of the OSM road dataset are then available. They are discussed in Subsections 4.3.1 and 4.3.2. Subsection 4.3.3 evaluates instead how the results of OSM geometrical accuracy vary when varying the value of the most critical parameter of the whole procedure, i.e. the generalization threshold described in Step 2.

**4.3.1 Evaluation of OSM maximum deviation:** The first analysis quantifies the maximum deviation of the OSM dataset from the IGN dataset for each grid cell. The user is required to enter the values of two parameters. The first is the upper bound value for the deviation and should be ideally equal to the buffer value used to pre-process OSM dataset in Step 2. The second parameter specifies the percentage of the OSM road length to be considered. While this value can be 100%, literature suggests that a slightly lower percentage allows to neglect the outliers that have been inevitably introduced by the procedure itself (e.g. by an imperfect cleaning of the OSM dataset in Step 2) as well as the non-optimal choices of parameters done by the user (Koukoletsos 2012). Choosing parameter values of 11 m for the maximum deviation and 95% for the percentage of OSM road length to be considered, an iterative algorithm based on a bisection search is run. For each grid cell the algorithm returns the maximum deviation of the most accurate 95% of the OSM dataset. This is computed as the minimum width of the buffer around the IGN dataset within which the 95% of the OSM road length is included (see Figure 3).

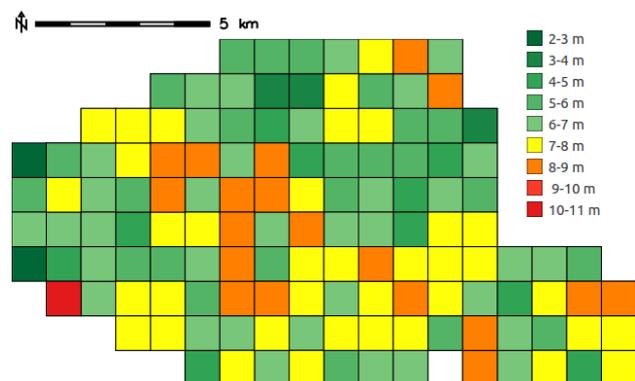


Figure 3. Maximum deviation of the Paris OSM dataset from the IGN dataset for each grid cell

The results clearly demonstrate the high positional accuracy of the Paris OSM road network. Besides a local variability, which produces maximum deviations ranging from 2 to more than 10 m, the average maximum deviation is found to be 6.6 m. Also, taking a look at the data it can be seen that the main reason for the highest values of maximum deviations is the fact that the “cleaning” of OSM dataset performed in Step 2 is not optimal – or, better to say, can never be optimal. Figure 4 shows an example where the maximum deviation of the OSM dataset from the IGN dataset is caused by the presence of portions of OSM roads which were not deleted by Step 2.

**4.3.2 Evaluation of OSM geometrical accuracy against user-specified thresholds:** The second analysis is intended to be run when users have to evaluate the positional accuracy of OSM against one or more specific targets, e.g. in the case they need to know whether the OSM road dataset can be used for a specific application requiring a minimum positional accuracy. In this second case, the only parameter value the user has to enter is the accuracy target (or the list of accuracy targets) to be considered in the evaluation. For each accuracy target a map is computed that, for each grid cell, returns the percentage of the OSM road length whose deviation from the IGN dataset is less

than or equal to that target. Figure 5 shows the map computed for an accuracy target equal to 8 m.

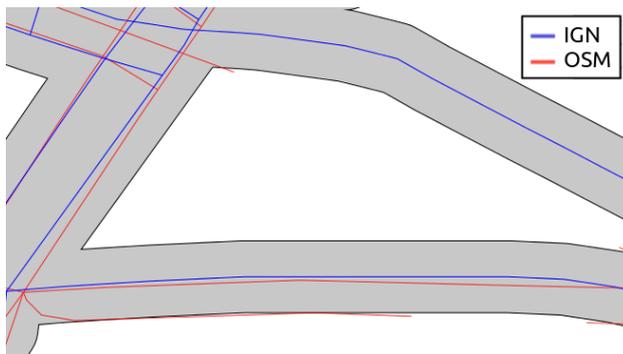


Figure 4. Detail of an area where the maximum deviation of the OSM dataset from the IGN dataset is caused by portions of OSM roads having no correspondence in the IGN dataset (data: © IGN and © OpenStreetMap contributors)

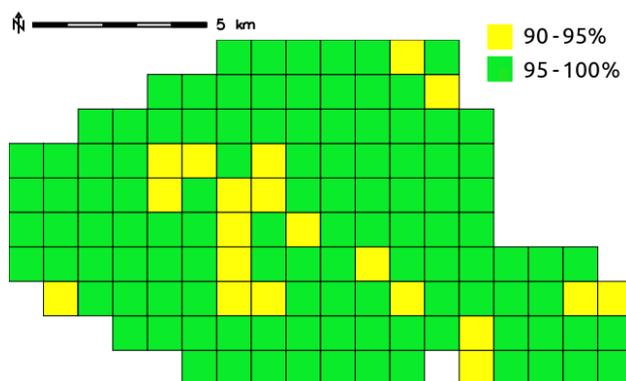


Figure 5. Percentage of OSM road length having a positional accuracy higher than or equal to 8 m for each cell

Figure 5 shows that, with very few exceptions, the percentage of OSM road length compliant with an accuracy target of 8 m is close to 100% for all the grid cells. However users should bear in mind that, due to the outliers introduced by the procedure as well as the non-optimal choice of parameters, percentage values around 90% correspond as well to high positional accuracies. Thus, iterating this analysis for lower accuracy targets, it can be concluded that the Paris OSM road dataset is suitable for applications requiring positional accuracies of 5-6 m.

#### 4.3.3 Sensitivity analysis on the generalization threshold:

The OSM accuracy results found in Step 3 are the consequence of the choice of parameter values performed by the user during Step 2. A crucial operation is therefore a sensitivity analysis on all these parameters, in order to assess how sensitive the results are to the parameter variation. Here we present a sensitivity analysis on the threshold of the Douglas-Peucker generalization algorithm used in Step 2. As mentioned in Subsection 4.2, the pre-processing of the OSM dataset in Step 2 is run three times with thresholds of 1 m, 0.5 m and 0 m (i.e. with no generalization). It should be noted that Step 2, which is the most computationally intensive, takes approximately 1/3 of the time if run with a threshold of 1 m compared to a threshold of 0.

The second analysis of Step 3 (see Subsection 4.3.2) is run on the three versions of the pre-processed OSM dataset (i.e. generalized with thresholds of 1 m, 0.5 m and 0 m), and using accuracy targets of 6 m, 8 m and 10 m. For each of these targets and for each of the three OSM datasets, the number of grid cells

is counted where the percentage of OSM road length compliant with the target is higher than or equal to 95%. Surprisingly, the number of grid cells is the same in all the cases. To further investigate whether the generalization of the OSM dataset does not really influence the results, the two extreme cases (i.e. no generalization and generalization with a threshold of 1 m) are considered. Table 1 shows the statistics (average, standard deviation, minimum and maximum absolute value) on the differences between the percentages of OSM road length compliant with each accuracy target (6 m, 8 m and 10 m) in the two cases. It is clear that the differences are almost neglectable. This means that a generalization introducing an error limited to the nominal accuracy of the IGN dataset has no influence on the OSM accuracy results. In addition, it brings also the great advantage of reducing the computational time of Step 2 of about 2/3.

	average	st. dev.	min(abs)	max(abs)
6 m	0.001	0.104	0	0.683
8 m	0.012	0.137	0	1.307
10 m	0.008	0.090	0	0.536

Table 1. Statistics on the differences between the percentages of OSM road length compliant with accuracy targets of 6 m, 8 m, and 10 m, in the cases of no generalization and generalization with threshold of 1 m

## 5. CONCLUSIONS

The increasing popularity of VGI and OSM has raised interest on the quality assessment of these user-generated data. This is favoured by the frequent release of official and authoritative datasets from NMA and CMC with open licenses. Focusing on road network datasets, a number of procedures have been proposed over the last years to evaluate OSM quality through a comparison with road datasets from NMA and CSC. However, the majority of these procedures were specifically designed for the particular reference dataset exploited in the study and are not easily replicable or extendable to other reference dataset. On the contrary, the work presented in this paper has made use of an automated procedure which was specially developed to allow comparison of OSM road network datasets with any reference dataset. The procedure returns measures of OSM completeness and positional accuracy. It is designed to be fully driven by the user, who can customize the analysis by adapting the parameter values according to the characteristics of the reference dataset OSM is compared with.

An application of the procedure is presented which assesses the quality of the Paris OSM road network. Results certify the high completeness and positional accuracy of this dataset. In terms of completeness, the analysis shows that the total length of OSM roads is higher than the total length of the IGN road dataset both before and after the OSM geometric pre-processing which deletes the OSM portions of roads having no correspondence in the IGN dataset. The analysis on positional accuracy returns an average maximum deviation of OSM from the IGN dataset of about 6-7 m and provides evidence for OSM suitability for applications requiring accuracies of 5-6 m. The sensitivity analysis performed on the generalization threshold parameter shows an almost null sensitivity of results when increasing the extent of the generalization. Though interesting, this result is specific to this particular application and cannot be generalized. Further tests on different datasets and with higher generalization thresholds are needed to evaluate whether the accuracy results are really insensitive to the variation of this parameter.

This study, along with others already completed (e.g. Antunes et al., 2015), demonstrates the high flexibility and adaptability of

the procedure. From the authors' perspective, it represents a valuable means for supporting a number of users, researchers and practitioners in carrying out their own quality assessments between OSM and reference road datasets. The open source nature allows the procedure to be easily and continuously improved and extended. Authors are currently working in a number of directions. First, a parallel processing approach is under development in order to reduce the computational time of Step 2. This would allow the exposure of the whole procedure on the Internet as a Web Processing Service (WPS), which is currently available only for Step 1. An extension of the comparison to the data attributes (in addition to the simple geometries) is also planned for the near future.

#### ACKNOWLEDGEMENTS

The authors acknowledge support from COST Actions TD1202 'Mapping and the Citizen Sensor' and IC1203 'European Network Exploring Research into Geospatial Information Crowdsourcing: software and methodologies for harnessing geographic information from the crowd (ENERGIC)'. Finally the authors would like to thank the French IGN for providing the Paris road network dataset exploited in the study.

#### REFERENCES

- Ali, A. L., Schmid, F., 2014. Data quality assurance for volunteered geographic information. In: *Geographic Information Science*. Springer International Publishing, pp. 126-141.
- Antunes, F., Fonte, C. C., Brovelli, M. A., Minghini, M., Molinari, M., Mooney, P., 2015. Assessing OSM Road Positional Quality with Authoritative Data. In: *Proceedings of the VIII Conferência Nacional de Cartografia e Geodesia (VIII CNCG)*, Lisbon, Portugal, 29-30 October 2015.
- Ather, A. A., 2009. *Quality Analysis of OpenStreetMap Data*. MSc Thesis, University College London, London.
- Brovelli, M. A., Minghini, M., Molinari, M. E., Mooney, P., 2016. Towards an automated comparison of OpenStreetMap with authoritative road datasets. *Transactions in GIS*.
- Ciepluch, B., Mooney, P., Jacob, R., Zheng, J., Winstanely, A. C., 2011. Assessing the quality of open spatial data for mobile location-based services: Research and applications. *Archives of Photogrammetry, Cartography and Remote Sensing*, 22, pp. 105-116.
- Devillers, R., Jeansoulin, R., 2006. *Fundamentals of spatial data quality*. ISTE, London.
- Fairbairn, D., Al-Bakri, M., 2013. Using Geometric Properties to Evaluate Possible Integration of Authoritative and Volunteered Geographic Information. *ISPRS International Journal Geo-Information*, 2(2), pp. 349-370.
- Forghani, M., Delevar, M. R., 2014. A Quality Study of the OpenStreetMap Dataset for Tehran. *ISPRS International Journal Geo-Information*, 3(2), pp. 750-763.
- Girres, J. F., Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), pp. 435-459.
- Goodchild, M. F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), pp. 211-221.
- Graser, A., Straub, M., Dragaschnig, M., 2014. Towards an Open Source Analysis Toolbox for Street Network Comparison: Indicators, Tools and Results of a Comparison of OSM and the Official Austrian Reference Graph. *Transactions in GIS*, 18(4), pp. 510-526.
- Guptill, S. C., Morrison, J. L., 1995. *Elements of spatial data quality*. Elsevier, Oxford.
- 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), pp. 682-703.
- Helbich, M., Amelunxen, C., Neis, P., Zipf, A., 2012. Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. In: *Angewandte Geoinformatik 2012*. Herbert Wichmann Verlag, VDE Verlag GmbH, Berlin/Offenbach, pp. 24-33.
- Jokar Arsanjani, J., Barron, C., Bakillah, M., Helbich, M., 2013. Assessing the Quality of OpenStreetMap Contributors together with their Contributions. In: *Proceedings of the 16th AGILE conference on Geographic Information Science*, Leuven, Belgium, 14-17 May 2013.
- Jokar Arsanjani, J., Zipf, A., Mooney, P., Helbich, M., 2015a. An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications. In: *OpenStreetMap in GIScience*. Springer International Publishing, pp. 1-15.
- Jokar Arsanjani, J., Mooney, P., Helbich, M., Zipf, A., 2015b. An exploration of future patterns of the contributions to OpenStreetMap and development of a contribution index. *Transactions in GIS*, 19(6), pp. 896-914.
- Keßler, C., Trame, J., Kauppinen, T., 2011. Tracking editing processes in volunteered geographic information: the case of OpenStreetMap. In: *Identifying objects, processes and events in spatio-temporally distributed data (IOPE)*, workshop at Conference on Spatial Information Theory, Belfast, USA, 12 September 2011.
- Keßler, C., de Groot, R. T. A., 2013. Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In: *Geographic information science at the heart of Europe*. Springer International Publishing, pp. 21-37.
- Koukoletsos, T., 2012. *A Framework for Quality Evaluation of VGI Linear Datasets*. PhD Thesis, University College London, London.
- Kounadi, O., 2011. *Assessing the quality of OpenStreetMap data*. MSc Thesis, University College of London, London.
- Ludwig, I., Voss, A., Krause-Traudes, M., 2011. A comparison of the street networks of Navteq and OSM in Germany. In: Geertman, S., Reinhardt, W., Toppen, F. (eds), *Advancing Geoinformation Science for a Changing World*. Springer International Publishing, pp. 65-84.
- Neis, P., Zielstra, D., Zipf, A., 2011. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007-2011. *Future Internet*, 4(1), pp. 1-21.
- Siebritz, L. A., Sithole, G., 2014. Assessing the Quality of OpenStreetMap Data in South Africa in Reference to National Mapping Standards. In: *Proceedings of the Second AfricaGEO Conference*, Cape Town, South Africa, 1-3 July 2014.

Zielstra, D., Zipf, A., 2010. A comparative study of proprietary geodata and volunteered geographic information for Germany. In: *Proceedings of the 13th AGILE international conference on geographic information science*, Guimarães, Portugal, 11-14 May 2010.

Zielstra, D., Hochmair, H. H., 2011. Digital street data: Free versus proprietary. *GIM International*, 25, pp. 29-33.