

# A proposal for a QGIS plugin for spatio-temporal analysis of OSM data quality: the case study for the city of Salvador, Brazil

Elias Elias<sup>1\*</sup>, Fabricio Amorim<sup>1</sup>, Leonardo Silva<sup>1</sup>, Marcio Schmidt<sup>1</sup>, Silvana Camboim<sup>1</sup> and Vivian Fernandes<sup>2</sup>

<sup>1</sup> Programa de Pós Graduação em Ciências Geodésicas, Federal University of Parana, Curitiba, Brazil; [elias\\_naim2008@hotmail.com](mailto:elias_naim2008@hotmail.com), [fabricioamorimeac@hotmail.com](mailto:fabricioamorimeac@hotmail.com), [scharth.leo@gmail.com](mailto:scharth.leo@gmail.com), [marcio.schmidt@ufu.br](mailto:marcio.schmidt@ufu.br), [silvanacamboim@gmail.com](mailto:silvanacamboim@gmail.com)

<sup>2</sup> Programa de Pós Graduação em Engenharia Civil, Federal University of Bahia, Salvador, Brazil; [vivian.fernandes@ufba.br](mailto:vivian.fernandes@ufba.br)

\* Author to whom correspondence should be addressed.

This abstract was accepted to the Academic Track of the State of the Map 2021 Conference after peer-review.

The development of methodologies to evaluate geospatial data quality is one of the most important aspects to be considered while obtaining this data. For developing countries, such as Brazil, the lack of investment for the maintenance of topographic mapping, especially on a big scale, is a recurrent challenge for the National Mapping Agencies [1]. For example, studies reveal areas in Brazil that have never been mapped and that the topographic mapping in the 1:25,000 scale is nearly 5% of its extension [2].

Technological advances have enabled a series of methodologies for obtaining geospatial data [3]. One example is presented as Volunteered Geographic Information (VGI) [4]. In this case, the update of information may occur faster and with a reduced cost in detriment to the traditional structures of topographic mapping [5]. A successful case of VGI is the OpenStreetMap (OSM) project, which presents the growth in the number of contributors and contributions or mapped features. To understand the quality of OSM features and their integration potential in topographic mapping, different surveys worldwide have put efforts to evaluate its quality, whether by its extrinsic [6, 7] or intrinsic [8] aspects. In this regard, some studies have evaluated the quality of OSM features by combining extrinsic and intrinsic aspects, like [9], which evaluated the positional accuracy of OSM based on the combination of edit history. Besides that, recent works have focused on comprehending spatial and temporal aspects of events in OSM contributions [10], as well developing add-ons for evaluating data quality, as presented by [11], where the authors developed a QGIS toolbox to evaluate parameters of the intrinsic quality of OSM features.

The literature identifies the heterogeneity of the data as one of the main challenges for the integration processes. The value of the quality parameters may vary according to the study area, the indicator used or even the temporal variations in the dynamics of the geographical space itself. In this context, to understand the integration of OSM data to the

Elias, E.N.N., Amorim, F.R., Silva, L.S., Schmidt, M.A., Camboim, S.P. & Fernandes, V.O. (2021). A proposal for a QGIS plugin for spatio-temporal analysis of OSM data quality: the case study for the city of Salvador, Brazil

In: Minghini, M., Ludwing, C., Anderson, J., Mooney, P., Grinberger, A.Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2021 Online Conference, July 09-11 2021, 27-30. Available at <https://zenodo.org/communities/sotm-2021>

DOI: [10.5281/zenodo.5112244](https://doi.org/10.5281/zenodo.5112244)



topographic mapping, it is crucial to connect aspects related to the quality and heterogeneity of data. Research work like [1] argues that, based on the obtained quality, the resources resulting from VGI may be used to integrate, detect changes or report errors. Therefore, classifying resources from OSM according to their usability in a certain region becomes essential, especially in developing countries like Brazil. Besides that, research that explores issues of quality, heterogeneity, and contributions patterns of OSM is still not widespread in developing countries [12].

Given the importance of classifying OSM features according to their usability for a given region, especially in developing countries, few researchers have explored quality, heterogeneity, and contribution pattern issues in OSM in Brazil. Based on the issues addressed, we propose a hypothesis that understanding aspects of the extrinsic and intrinsic quality of the quality of OSM features will help decision making regarding the integration of such data in topographic mapping. The main focus is on the spatio-temporal aspects of contributions in developing countries. Thus, this research has the objective to evaluate the extrinsic quality of OSM features for the county of Salvador-Bahia-Brazil (the northeast region of the country). Therefore, we investigated indicators of positional accuracy, thematic accuracy and completeness, the visualisation of heterogeneity of data, and the analysis of the edition history. To accomplish the evaluation of extrinsic quality, the OSM features were compared to the topographic mapping of the country from the Cartographic and Cadastral System of the County of Salvador (Sistema Cartográfico e Cadastral do Município do Salvador - SICAD, 2006) and features from the Urban Development Company of the State of Bahia (Companhia de Desenvolvimento Urbano do Estado da Bahia - CONDER).

The analysis of positional and thematic accuracy was made through procedures of feature sampling. The analysis of completeness occurred from comparing the total number of available features. The verified categories were features from the road system, religious, educational, and health buildings. We divided the municipality of Salvador into sub-regions to identify different local patterns of quality in the analysis of thematic accuracy and completeness. Visualisation allows obtaining the data's heterogeneity through a plugin developed in the software QGIS, making the planimetric positional evaluation for point and line features. The statistical procedures for developing the plugins were realised based on the Brazilian law to evaluate geospatial data quality analysis [13] and based on the method of double buffer proposed by [14]. The plugin is available, and it is possible to be accessed in the online repository [15]. Even though the final results comprehend aspects of Brazilian law, they can be replicated to obtain the discrepancies and posterior adjustments. We used the Oshome Application Programming Interface (API) [16] to identify the patterns concerning the OSM editing history. Thus, from the adaptations performed in scripts given by researchers linked to Oshome, it was possible to identify the aspects of OSM contributions between 2008 and 2020. We also tested the generation of regression curves and calculated the number of daily contributions to identify these patterns. These verifications were occasionally created through the generation of an evolving rectangle of 5x5 km in the study area. The disposition of the rectangle was given through a visual analysis with a larger quantity of OSM features.

The evaluation of extrinsic evaluation highlighted the variability of the results obtained in [17]. In analysing the positional accuracy, the scale found varied from 1:20,000 to 1:30,000, while the discrepancies between the mapped coordinates and the reference one

varied between 0.12m and 10.27m. In analysing completeness, it was observed that features that corresponded to the road system presented better results concerning the other categories. The road system presented a completeness percentage of 82%, while in the other features, the variation was from 29% to 46%. When analysing thematic accuracy, it turns out that the primary source of errors is related to the absence of names in editing. In the analysis of the OSM contribution history growth of represented features, it was possible to notice a near-linear function, with an R2 value of 0.94. As there is a finite amount of mappable elements at a given area, we can use this function to model the contributions patterns over time until the region is saturated. Besides that, it was possible to observe that the patterns of collaboration can be affected by different variables because it was noticed that in 2016, more than 800 features were added in a short period. These aspects can be related to events such as data importation or mapathons.

The development of add-ons for evaluating OSM data quality that departs from the making of statistical procedures up to visualising the heterogeneity of data will assist in the decision-making as to data quality. The development of QGIS plugins for OSM data quality assessment that execute from statistical procedures to visualisation of data heterogeneity will assist in decision making regarding data quality

From the add-on developed, it was possible to notice that the magnitude of discrepancies did not present patterns and that this may vary according to the period of editing and the database used for the contributions. Based on the obtained results, we noticed the relevance in identifying the aspects of quality and heterogeneity in OSM contributions. For Brazil, identifying these characteristics may numerally indicate the integration potential of these data to authoritative mapping. Besides that, it will estimate the influence of unusual agents, like it is the case of data import in the contributions. The continuity of the studies is recommended to identify the causes of different patterns of growth and the continuity of studies to automatise the quality procedures.

## References

- [1] Maulia, N. (2018). Development of an update procedure for authoritative spatial data by the combination with crowdsourced information. Master's thesis, Technische Universitat Dresden.
- [2] Silva, L. S., & Camboim, S. P. (2020). Brazilian Nsdi Ten Years Later: Current Overview, New Challenges And Propositions For National Topographic Mapping. *Boletim De Ciências Geodésicas*, 26(4).
- [3] Brovelli, M. A., Boccardo, P., Bordogna, G., Pepe, A., Crespi, M., Munafò, M., & Pirotti, F. (2019). Urban Geo Big Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W14, 23-30.
- [4] Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221.
- [5] Du, H., Alechina, N., Jackson, M., & Hart, G. (2017). A method for matching crowd-sourced and authoritative geospatial data. *Transactions in GIS*, 21(2), 406-427.
- [6] Brovelli, M. A., & Zamboni, G. (2018). A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of Geo-Information*, 7(8), 289.
- [7] Zhang, H., & Malczewski, J. (2017). Accuracy evaluation of the Canadian OpenStreetMap road networks. *International Journal of Geospatial and Environmental Research*, 5(2).
- [8] Minghini, M., & Frassinelli, F. (2019). OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data, Software and Standards*, 4(1), 9.

- [9] Nasiri, A., Ali Abbaspour, R., Chehreghan, A., & Jokar Arsanjani, J. (2018). Improving the quality of citizen contributed geodata through their historical contributions: The case of the road network in OpenStreetMap. *ISPRS International Journal of Geo-Information*, 7(7), 253.
- [10] Grinberger, A. Y., Schott, M., Raifer, M., & Zipf, A. (2021). An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS*, 25(2), 622-641.
- [11] Sehra, S. S., Singh, J., & Rai, H. S. (2017). Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. *Future Internet*, 9(2), 15.
- [12] Camboim, S. P., Bravo, J. V. M., & Sluter, C. R. (2015). An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. *ISPRS International Journal of Geo-Information*, 4(3), 1366-1388.
- [13] DSG. (2016). Norma da Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG). *Diretoria do Serviço Geográfico*, 1, 94, Brasília-DF.
- [14] Santos, A. P. (2015). Controle de qualidade cartográfica: Metodologias para avaliação da acurácia posicional em dados espaciais. PhD thesis, Universidade Federal de Viçosa.
- [15] Elias, E. (2019). AcuraciaPosicional\_PEC-PCD. Retrieved from [https://github.com/eliasnaim/AcuraciaPosicional\\_PEC-PCD](https://github.com/eliasnaim/AcuraciaPosicional_PEC-PCD)
- [16] Heidelberg Institute for Geoinformation Technology (2021). Ohsome. Retrieved from <https://heigit.org/big-spatial-data-analytics-en/ohsome>
- [17] Elias, E. N. N. (2019). Qualidade de Dados Geoespaciais em Plataforma de Mapeamento Colaborativo. Master thesis, Universidade Federal da Bahia.