International Cartographic Association
Association Cartographique Internationale

Identifying Temporal Patterns of Contributions to OpenStreetMap in Brazilian Cities: The Role of Cell Size.

Tayani Bellucci Peterlini [a], Marcio Augusto Reolon Schmidt [a,b], Elias Nasr Naim Elias [a], Silvana Phillip Camboim [a], Caio dos Anjos Paiva [a].

[a] *Universidade Federal do Paraná - UFPR, Tayani Bellucci Peterlini – tayanibellucci@ufpr.br, Marcio Augusto Reolon Schmidt - marcio.schmidt@ufpr.br; Elias Nasr Naim Elias - elias.naim@ufpr.br; Silvana Phillip Camboim -silvanacamboim@ufpr.br; Caio dos Anjos Paiva -anjospaiva@ufpr.br.*
[b] *Universidade federal de Uberlândia - UFU - marcio.schmidt@ufu.br*

Keywords: VGI intrinsic parameters, data quality, temporal contribution analysis

Abstract:

Data quality from collaborative mapping platforms like OpenStreetMap is critical in determining their potential use. When official data is available, extrinsic assessment can provide quality parameters such as positional accuracy and completeness. However, there is often a lack of such reference datasets in developing countries, despite the need for collaborative data for cartographic updating. For example, mapping coverage at the 1:25,000 scale in Brazil is only 5% (SILVA and CAMBOIM, 2021). At larger scales, the existence and updating of spatial data are even more concerning in the jurisdictions of more than 5,000 municipal governments. In these cases, using intrinsic parameters is a viable alternative when dealing with areas with few or outdated official data. This topic has been the subject of several studies (BROVELLI and ZAMBONI, 2018; HECHT et al., 2013; HAKLAY et al., 2010; MOONEY, CORCORAN and WINSTANLEY, 2010 and ATHER, 2009). These intrinsic parameters are typically based on a region's number of contributors and editions. Recently, intrinsic temporal parameters have also been proposed. The dynamics of contributions over time can provide valuable information. Elias et al. (2023a) used the sigmoidal curve regression model proposed by Brückner et al. (2021) to describe contributions over time and obtain curve parameters mathematically. In this research, the authors applied a regular 1x1 km grid in a metropolitan area in southern Brazil. The resulting sigmoidal curves allow for the identification of parameters in the respective squares, enabling the identification of local patterns of contribution history in the same city. In this type of analysis, the size of the chosen cell is crucial because it needs to be small enough to detect significant variations but not so small as to unnecessarily increase computational costs without providing modelling gains. Camboim et al. (2022) explored this issue in their search for an optimal solution for observing local heterogeneities.

In addition to studying within the same city, it can be helpful to compare cities, as Zhou et al. (2022) did in their global analysis of 13,000 areas. In this case, cell size is again a critical variable, and it is necessary to understand which cell size is sufficiently representative of the city's contributions pattern. Therefore, this study aims to provide a preliminary analysis of the effect of cell size on the quality of logistic curve modelling in five Brazilian state capitals. Figure 1 shows the cities: Curitiba (South) with 1.9 million inhabitants, São Paulo (Southeast) with 12.4 million, Goiânia (Midwest) with 1.6 million, Fortaleza (Northeast) with 2.7 million, and Belém (North) with 1.5 million. For all locations, a Python script consulting the Ohsome API (URL1) analysed the contributions from 2008 to 2022 aggregated by month. The analysis considered the tags "highway" and "buildings" and used concentric cells with sides ranging from 1 to 8 km, as shown in the Curitiba example.

Figure 1. Location of the five cities and cell example for Curitiba.

The graphs in Figure 2 demonstrate a very consistent behaviour of the curves in each city with increasing cell size. Even with the varying total number of features on 31-12-2022 (from a minimum of 16,000 features in 8km² in Belém to about

https://doi.org/10.5194/ica-abs-6-197-2023

160,000 in São Paulo), we can see a pattern of local contributions. Goiânia and Belém (the cities with fewer mapped features) have a gradual growth curve, very organic. Both cities do not have Spatial Data Infrastructure, which would explain the lack of imported data to the OSM base. On the other hand, Fortaleza (ide.sefin.fortaleza.ce.gov.br) and São Paulo (geosampa.prefeitura.sp.gov.br) are two cities that have developed their own SDIs. As a result, the incorporation of mass data recently in Fortaleza and approximately halfway through the period studied in São Paulo is noticeable. Another interesting observation in São Paulo is how the number of collaborations continues to rise even after imports, showing a positive synergy between the official and collaborative maps. Curitiba has an intermediate situation, as there are imports, but not so significant. The city has a spatial data portal (ippuc.org.br/geodownloads/geo.htm) but not an SDI, with limited data availability.

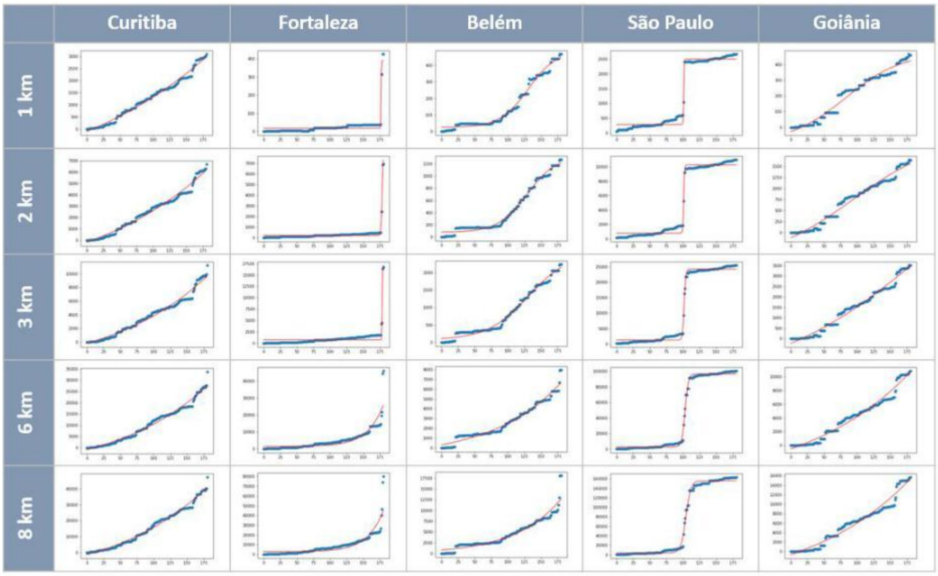

Figure 2. Cumulative contributions and logistic curve regression by city and cell size

The RMSE of each cell size represents the quality of the curve fit. The number of features normalised the value at the end of the period, whose results are in the graph of Figure 3. We can observe that the increase in the cell size tends to present a better curve fit, sometimes going up or stabilising after a certain point and in this preliminary experiment with only five cities, observing that cells of about 3km would present a consistent performance to compare the cities' patterns.

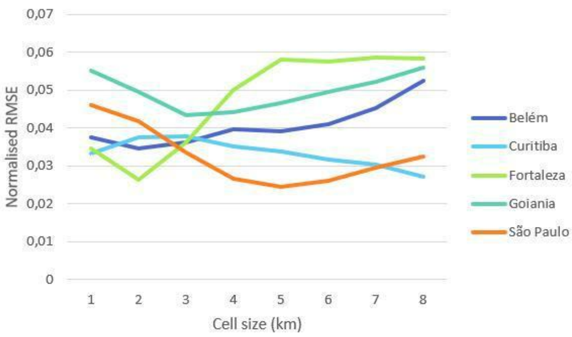

Figure 3: Normalised RMSE behaviour for the five cities

The spatial data situation heterogeneity between Brazilian municipalities has multiple reasons, mainly the economic and historical inequalities between regions. For example, Neis and Zielstra (2014) describe that to contribute to these platforms, users need to have access to the internet and additional hardware, such as smartphones or computers. The authors comment that although these requirements may seem trivial nowadays, some limiting aspects of VGI align with Goodchild's approach (2007), particularly concerning developing countries. This aspect is related to the fact that some VGI platforms often only support the English language, and the illiteracy rates may hinder mapping specific areas. However, there is also an impact of public policies in this respect. The National Spatial Data Infrastructure, when enacted in 2008, was mandatory only for federal institutions. Thus, each municipality must identify the importance of investing in open geospatial data and follow it independently. In this scenario, we must have the mechanisms to observe the consequence of local public policies' discrepancies in data availability. When replicating the experiment for other municipalities, it is expected that it would be possible to find patterns and necessary actions locally, besides demonstrating how much the investment in creating and disseminating open geospatial data can positively contribute to promoting collaborative mapping actions.

https://doi.org/10.5194/ica-abs-6-197-2023 |

References

Ather, A. (2009). 'A quality analysis of OpenStreetMap data', unpublished MEng dissertation, UCL, London, UK brovelli, M. A.; Zamboni, G. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. ISPRS International Journal of GeoInformation, 7(8), pp. 1-25, 2018.

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

Haklay, M. 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, 2010.

Mooney, Peter, Padraig Corcoran, and Adam C. Winstanley. "Towards quality metrics for OpenStreetMap." Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. 2010.

Neis, P.; Zielstra, D. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. Future Internet, 6(1), 76-106, 2014

Zhou, Q., Zhang, Y., Chang, K., & Brovelli, M. A., 2022. Assessing OSM building completeness for almost 13,000 cities globally. International Journal of Digital Earth, 15(1), 2400–2421. https://doi.org/10.1080/17538947.2022.2159550

Abstracts of the International Cartographic Association, 6, 197, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa.
https://doi.org/10.5194/ica-abs-6-197-2023