The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B4-2026
XXV ISPRS Congress 2026 "From Imagery to Understanding", Commission IV, 4–11 July 2026, Toronto, Canada

Integrating Microsoft Building Footprints and OpenStreetMap to Improve Building Representation

Cidália C. Fonte[1,2], Joaquim Patriarca[2,3], Diogo Duarte[2,4]

[1] University of Coimbra, Department of Mathematics, Largo D. Dinis, 3000 – 143 Coimbra, Portugal, cfonte@uc.pt
[2] Institute for Systems Engineering and Computers at Coimbra- INESCC, Rua Sílvio Lima, 3030-790 Coimbra, Portugal
[3] University of Coimbra, CISUC, Department of Informatics Engineering, Rua Sílvio Lima, 3030-290 Coimbra, Portugal
[4] University of Coimbra, Department of Electrotechnical Engineering, Rua Sílvio Lima, 3030-790 Coimbra, Portugal

Keywords: Buildings, Microsoft Building Footprints, OpenStreetMap, Quality

Abstract

This paper investigates whether integrating the Microsoft Building Footprints (MBF) dataset with building footprints contributed by the OpenStreetMap (OSM) community can improve the spatial quality of building data. Specifically, we assess whether the resulting hybrid dataset enhances completeness and positional accuracy relative to the original MBF and OSM datasets. The evaluation was conducted in a study area encompassing both urban and rural environments, using 1:5,000 topographic cartography as the reference dataset. The merged MBF+OSM dataset successfully captured 87% of the buildings represented in the reference cartography, outperforming the standalone MBF and OSM datasets, which captured 81% and 70%, respectively. These results demonstrate that combining MBF and OSM footprints provides a more comprehensive representation of buildings and can offer a valuable alternative for applications requiring detailed, up-to-date building information.

1. Introduction

Accurate and up-to-date information on buildings in both urban and rural environments is essential for many applications. These include planning and coordinating emergency response to natural disasters—such as floods or forest fires—and the management of urban settlements, where the location of key facilities (e.g., commercial, industrial, educational, or health services) must be known (Hong et al., 2024; Oostwegel et al., 2025; So, 2023). Large-scale topographic maps typically provide this information; however, their production is costly and time-consuming. As a result, updates may occur only at multi-year intervals, and in some parts of the world such maps are not available at all.

Land Use Land Cover Maps, such as Urban Atlas, CLC+ Backbone, or Corine Land Cover (Copernicus, 2016; EEA, 2022; European Commission, 2020) offer valuable insight into urbanized areas, but they do not delineate individual buildings or provide their precise locations. In contrast, datasets such as the Microsoft Building Footprints (MBF) and the volunteered geographic information produced by the OpenStreetMap (OSM) project map individual buildings and are available worldwide. Despite this advantage, these two sources differ substantially in methodology and exhibit complementary strengths and weaknesses(Fonte & Patriarca, 2025). For example, OSM typically exhibits high positional accuracy where data exist, but coverage remains incomplete in many regions. MBF, on the other hand, offers greater coverage but lower positional accuracy for many buildings.

To address these limitations, this paper proposes a methodology to merge MBF and OSM building footprints in order to create a more complete and spatially accurate dataset. The approach leverages the higher positional quality of OSM where available, while relying on MBF in areas where OSM data are missing. The quality of the original datasets and the merged dataset is evaluated in terms of completeness and positional accuracy using a 1:5,000 topographic map as reference.

Section 2 describes the study area and the datasets used. Section 3 evaluates the completeness and positional accuracy of the original datasets in comparison with the cartography. Section 4 presents the methodology for merging the datasets, and Section 5 reports the results. Finally, Section 6 provides conclusions and outlines directions for future work.

2. Study area and datasets

2.1 Study area

The study area used considered in the tests made in this paper is presented in Figure 1. It includes part of an urban area (the city of Coimbra, in Portugal) that includes residential, commercial and industrial areas, and a surrounding rural area.

2.2 Microsoft Building Footprints (MBF)

Bing Maps released a global building footprint dataset in 2022, known as the Microsoft Building Footprints (MBF), containing approximately 1.4 billion buildings (Microsoft, 2022). The dataset has been updated regularly with new releases and is distributed under the Open Data Commons Open Database License (ODbL), which allows free use provided that appropriate attribution is given and that any derived products are shared under the same license.

The MBF dataset was generated from Bing Maps imagery acquired between 2014 and 2024 using an automated two-step workflow: (1) semantic segmentation and (2) polygonization. In the first step, a deep neural network trained to detect buildings assigns each pixel a confidence score representing the likelihood that it corresponds to a building. Pixels with a confidence score above 0.5 are classified as building pixels, producing a raster mask of probable building locations. In the second step, these building-classified pixels are converted into vector polygons. Figure 2 illustrates the MBF building footprints available for the study area.

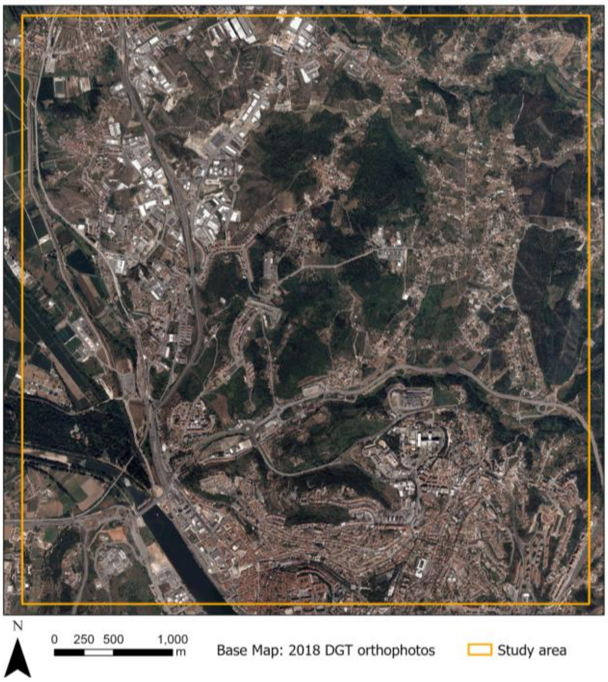

Figure 1. Study area.

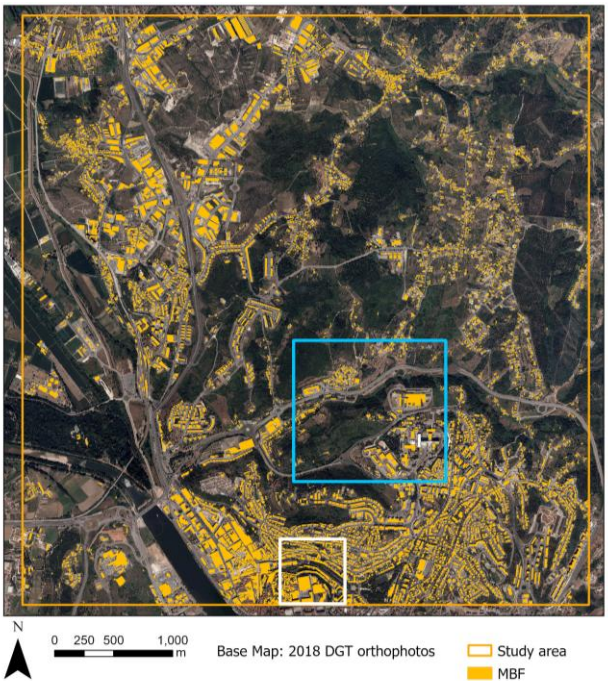

Figure 2. MBF in the study area. The white square delimits subregion A and the blue square subregion B.

2.3 OpenStreetMap (OSM)

The collaborative project OSM was created in 2004 by Steve Coast (Fox, 2012) with the goal of producing and maintaining a freely accessible, crowdsourced map of the world. Anyone can contribute to OSM by adding new features or updating and deleting existing ones. OSM data is stored in a vector format and consists of points, lines, and relations. Each feature has associated tags, which are composed of a key and a value. One of the used keys is “building,” which has dozens of possible values characterizing building types, such as apartments, hotels, schools, residential, industrial, or commercial (OSM map features, 2025). OSM is managed by the OSM Foundation, and the data is also distributed under the Open Database License (ODbL). Figure 3 illustrates the OSM building footprints available for the study area.

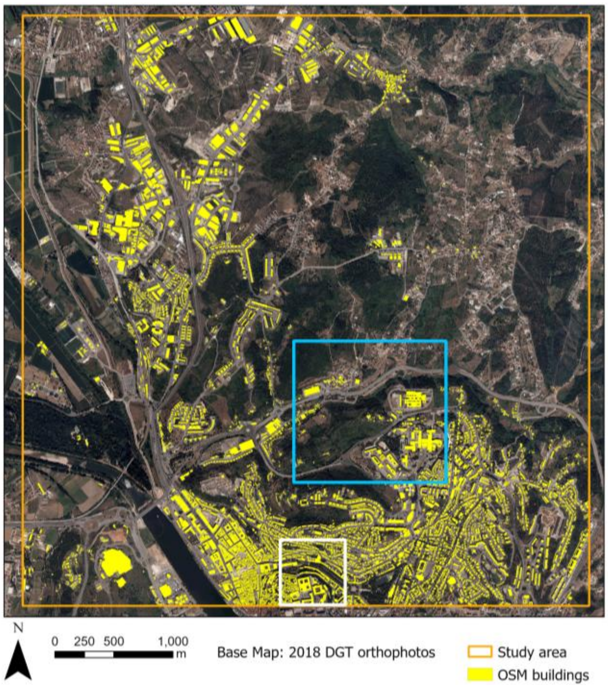

Figure 3. OSM buildings in the study area. The white square delimits subregion A and the blue square subregion B.

2.4 Topographic cartography

Figure 4 shows the buildings mapped in a 1:5,000 topographic map of 2007, kindly provided by the Coimbra Municipality.

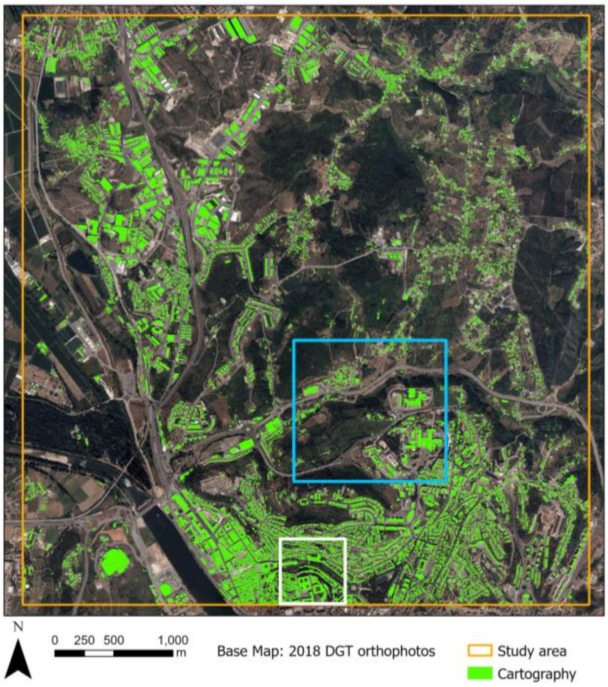

Figure 4. Buildings of the study area available in the cartography.

This dataset was used as a reference to compare the MBF, OSM building footprints, and the dataset generated by merging these two sources. Although this cartographic product is more than 18 years older than the MBF and OSM data, the region has experienced minimal changes over this period, making it suitable for evaluating the quality of the different building datasets.

3. Comparative analysis of the datasets

3.1 MBF and OSM quality

The MBF dataset includes quality metrics calculated for each region using sample buildings. For instance, precision ranges from 92% (for Caribbean) to 97% (for Central Asia), while recall ranges from 71% (for Africa) to 86% (for Europe).

Regarding OSM, even though data imports are occasionally made into OSM, most of data is edited manually by contributors through interpretation of high-resolution imagery. As OSM is a crowdsource project, the quality of OSM data has the characteristics typically associated with this type of data, such as heterogeneous quality (regarding, for example, completeness, positional quality and thematic quality) over space and type of feature (Antoniou & Skopeliti, 2015; Fonte et al., 2015; Minghini & Frassinelli, 2019). Several studies have been made regarding the quality of OSM data, and in particular OSM buildings (Biljecki et al., 2023; Borkowska & Pokonieczny, 2022; Herfort et al., 2023; Minghini & Frassinelli, 2019; Zhang et al., 2022). Such studies confirm that in some areas OSM data is very accurate and complete, while in other areas it has low levels of completeness and low quality.

Fonte & Patriarca (2025) compared MBF with the buildings extracted from OSM for a region in Portugal, and concluded that MBF showed higher levels of completeness for the analyzed area, particularly in rural areas, while OSM buildings showed better positional accuracy. Therefore, both datasets have complementary strengths and weaknesses.

3.2 Datasets comparison for the study area

Table 1 shows the buildings area in the three datasets and Table 2 the area differences between the datasets. The results are in accordance with the expectations that MBF and the cartography are more complete than OSM. MBF shows 56,287m² more built-up area than the topographic map. This difference has two explanations. First, the topographic map dates from 2007, so new buildings constructed since then would appear only in MBF. Second, MBF may include some areas incorrectly identified as buildings.

Data set	Area (m^2)
Cartography	2 358 287
MBF	2 414 574
OSM	1 966 100

Table 1. Area of the buildings in the study area in the datasets.

Data set	Area (m^2)
Cartography - MBF	-56 287
Cartography - OSM	392 187
MBF - OSM	448 474

Table 2. Area of the buildings in the study area in the datasets.

Figure 5 and Figure 7 show the buildings and buildings parts that are both in MBF and the cartography and that only belong to one of the datasets, for the subregions A and B within the study area (corresponding, respectively, to the white and blue boxes in Figures 1 to 3). The first region is located within the urban fabric, and the second one is in the area surrounding the city. Figure 6 and Figure 8 show the same areas as Figures 5 and 7, respectively, but for OSM data and the cartography.

In Figure 5 we can see that MBF (shown in orange) includes regions that should not be mapped as building (MBF commission errors), namely open areas within large buildings and smaller areas resulting from inaccurate polygon delineation.

Figure 5. Buildings of a subregion A of the study area available only in MBF, only in the cartography, and in both datasets.

Figure 6. Buildings of a subregion A of the study area available only in OSM, only in the cartography, and in both datasets.

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-41-2026

However, there are also other areas that are mapped in the cartography and absent from MBF (shown in green), some of which are in fact omission errors, also visible in Figure 7, where a large building is absent from MBF. Figure 6 shows the same area as Figure 5 but compares the buildings extracted from OSM with the cartography. We can see that in this subregion A (urban area) there are less commission errors in OSM than in MBF. It is also clear that the geometry of the buildings in OSM is much closer to their geometry in the cartography than the ones in MBF.

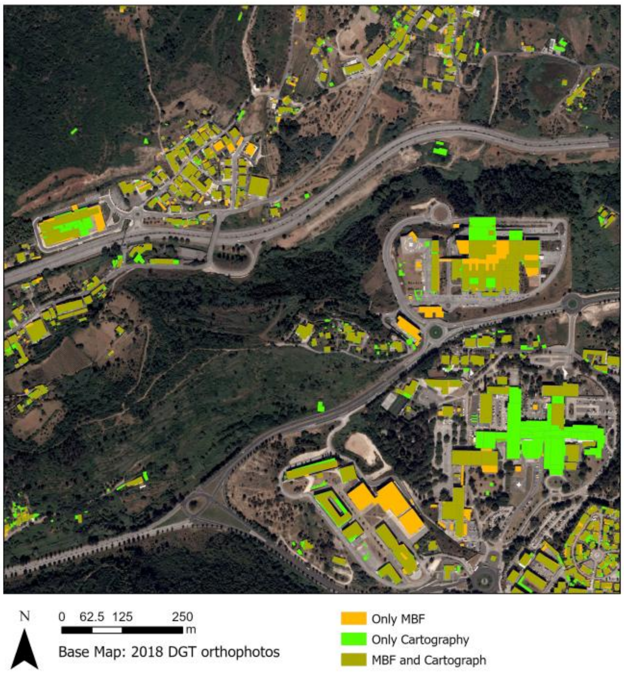

Figure 7. Buildings of a subregion B of the study area available only in MBF, only in the cartography, and in both datasets.

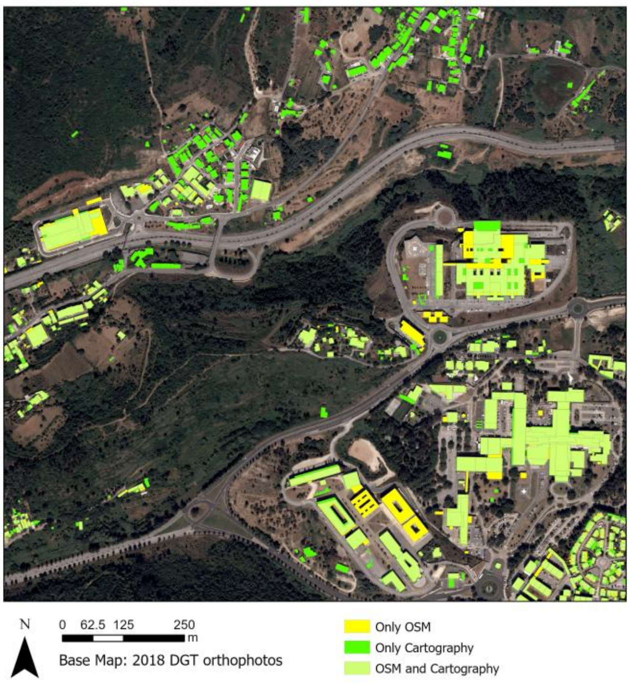

Figure 8. Buildings of a subregion B of the study area available only in OSM, only in the cartography, and in both datasets.

In Figure 7 (subregion B) we can see some new buildings mapped in MBF that were not in the cartography, but, in most cases, they are also mapped in OSM (Figure 8). However, the small buildings on top of the area shown in Figures 7 and 8 are absent from OSM (OSM omission errors) but are both in MBF and the cartography. These examples show that there are omission and commission errors in both datasets. In the cartography, a few buildings could also be identified that were demolished after 2007; therefore, they no longer represent a building and are in fact not included either in MBF nor in OSM.

4. Methodology

Because MBF has fewer omissions than OSM, and OSM buildings are likely to have higher positional accuracy in populated areas, we developed a methodology to merge the two datasets. The resulting dataset aims to achieve completeness similar to MBF while leveraging OSM geometries in regions where they are available to improve positional accuracy.

4.1 Merging MBF and OSM buildings

The methodology used to create a new dataset by merging MBF and OSM buildings, tested in this paper, consists of: 1) identifying the MBF that intersect OSM buildings; 2) remove such buildings from the MBF dataset, and 3) merge the dataset resulting from the previous step with the OSM buildings (see Figure 9). The aim is to replace MBF that are also mapped in OSM with the OSM geometries, which are likely to have higher levels of detail and positional accuracy.

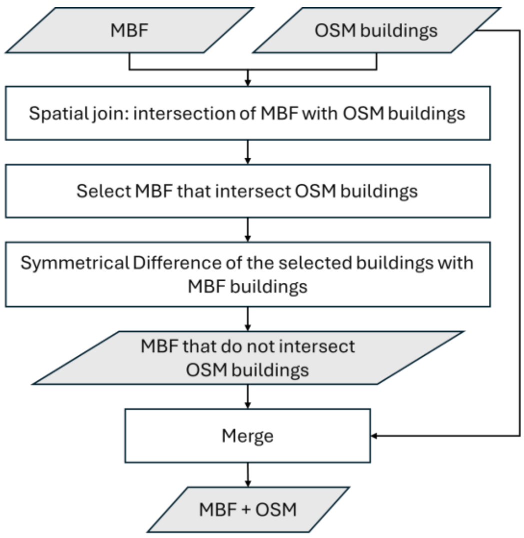

Figure 9. Diagram of the methodology used to merge MBF and OSM buildings in the study area.

4.2 Quality assessment

The resulting dataset, referred to as MBF+OSM, is then compared with the cartography. The results are also compared with the characteristics of the original datasets presented in Section 3. The metric Intersection over Union (IoU) is computed to compare the original MBF and OSM datasets, as well as the MBF+OSM dataset, with the cartography. Then, the Overlap Area Ratio (OAR) is computed for the MBF and cartography relative to both datasets, the OSM and the cartography also relative to both datasets, and the cartography and the MBF+OSM, relative to the three original datasets.

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-41-2026

5. Results

5.1 Quality assessment

Figure 10 shows the dataset obtained by merging MBF and OSM buildings with the methodology described in the previous section. Figure 11 shows the areas of the buildings in the four datasets within the study area, as well as the difference relative to the area of the buildings in the cartography.

The results show that the dataset with larger area is indeed the MBF+OSM, which has 171 968 m^2 of buildings more than the cartography, and 100 000 m^2 more than MBF.

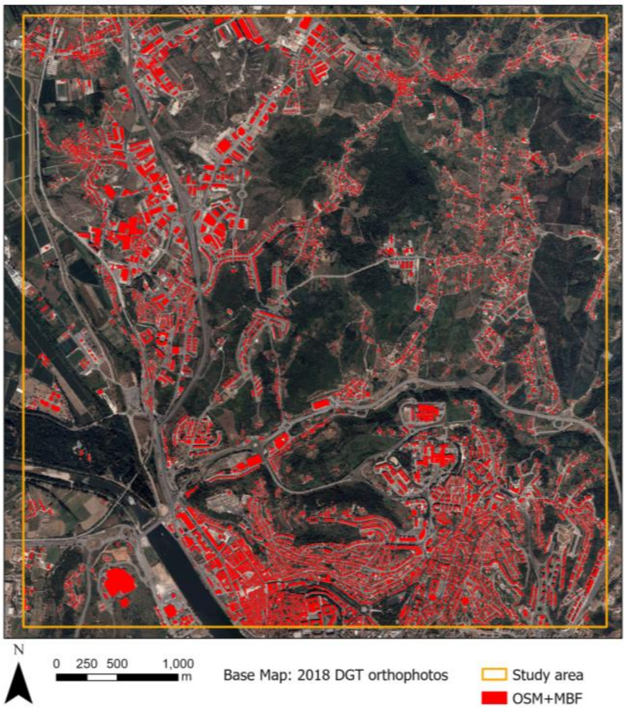

Figure 10. Buildings dataset for the study area obtained by applying the methodology to merge MBF and OSM buildings.

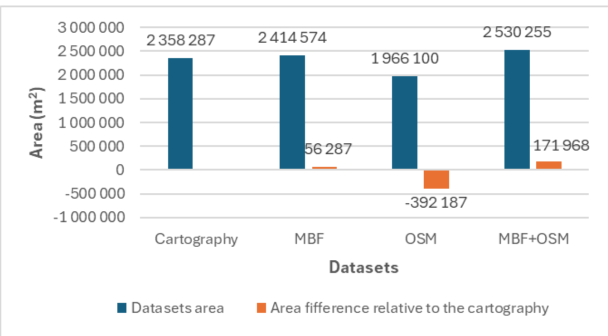

Figure 11. Buildings area in the four datasets (in blue) and the area differences between MBF, OSM and MBF+OSM and the cartography (in orange).

Table 3 shows the IoU computed for the cartography buildings and, respectively, MBF, OSM and MBF+OSM. A value of 72% is obtained for the IoU of the cartography and MBF+OSM, which shows that the dataset closer to the cartography is indeed the MBF+OSM, while the one with the lowest similarity is OSM (62%), even though only 5% lower than MBF (67%).

Datasets	IoU (%)
Cartography and MBF	67
Cartography and OSM	62
Cartography and MBF+OSM	72

Table 3. Intersection over union (IoU) of the datasets.

Table 4 shows the OAR of the intersection of each dataset under analysis and the cartography, relative to the area of the original datasets. The results confirm that the dataset that presents higher similarity with the cartography is MBF+OSM, with an OAR(C∩MO)/C) equal to 87%, while the values obtained for MBF ((C∩M) / C) and OSM (C∩O) / C) are 81% and 70%, respectively. The high value obtained for MBF+OSM is a good result, taking into consideration that there are also omissions in the cartography, due to the datasets temporal differences, that are mapped in the newer datasets.

Datasets		OAR
(Cartography ∩ MBF)/Cartography	(C∩M) / C	81
(Cartography ∩ MBF)/MBF	(C∩M) / M	79
(Cartography ∩ OSM)/Cartography	(C∩O) / C	70
(Cartography ∩ OSM)/OSM	(C∩O) / O	84
(Cartography∩MBF+OSM)/Cartography	(C∩MO) /C	87
(Cartography ∩ MBF+OSM)/MBF	(C∩MO) /M	85
(Cartography ∩ MBF+OSM)/OSM	(C∩MO) /O	104

Table 4. Overlap Area Ratio (OAR) between the intersection of MBF (M), OSM buildings (O) and MBF+OSM (MO) with the cartography, in relation to the three original datasets.

The OAR between the intersection of the original MBF and the cartography relative to MBF ((C∩M) / M) is only 79%, while the OAR of the intersection of the original OSM and the cartography relative to OSM ((C∩O) / O) is 84%. These results show that a higher percentage of OSM data is overlapped with the cartography when compared to the overlapping of MBF with the cartography.

When the MBF+OSM dataset is considered, the OAR of its intersection with the cartography is computed not only in relation to the cartography, but also in relation to MBF ((C∩MO)/M) and OSM ((C∩MO)/O), with resulting values of 85% and 104%, respectively. The first value is 2% lower than the one obtained in relation to the MBF, due to the fact that MBF includes a larger area mapped as buildings than the cartography. The 104% obtained when OAR is computed in relation to OSM, confirms that MBF was useful to add missing building in OSM.

Figures 12 and 13 show the merged building dataset for, respectively, the subregion A and B, highlighting the parts that are only mapped in the MBF+OSM (in red) and the ones that are common to the MBF+OSM and the cartography (in brown). The green areas show the regions that are only mapped in the cartography.

6. Conclusions and future work

A methodology to merge MBF and OSM buildings is tested, to assess if the new buildings dataset is closer to the buildings shown in a topographic map. This study was tested in a small area, which included an urban setting and some nearby rural areas. High quality data exists in OSM for the urban area, but such data was missing in the rural region, which was an indicator that the merge with the MBF data would be useful.

The results show that the new dataset is indeed closer to the buildings represented in a 5:000 topographic cartography, than either one of the original datasets. This shows that the proposed approach may be useful to obtain a dataset of the buildings with higher quality than either MBF or OSM, especially in regions where no topographic cartography is available. It may also help in the identification of possible new buildings present in the study area, which may indicate that the cartography, if available, needs to be updated.

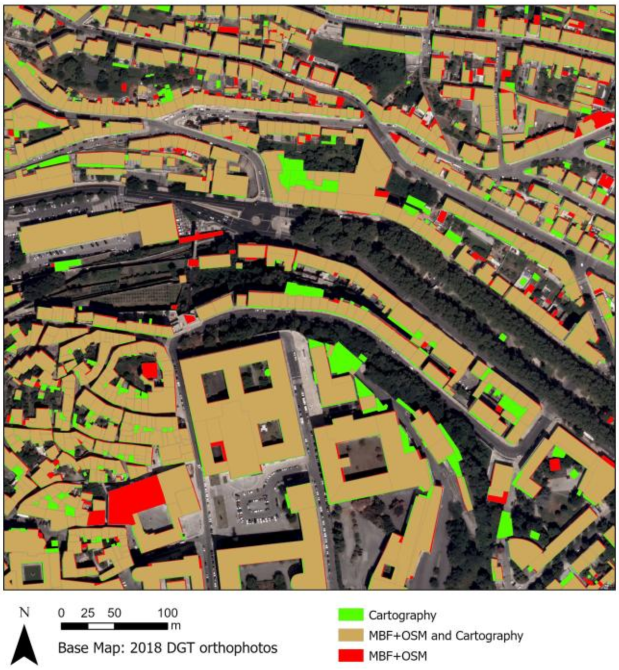

Figure 12. Buildings dataset resulting from the merge of MBF and OSM buildings for the subregion A of study area.

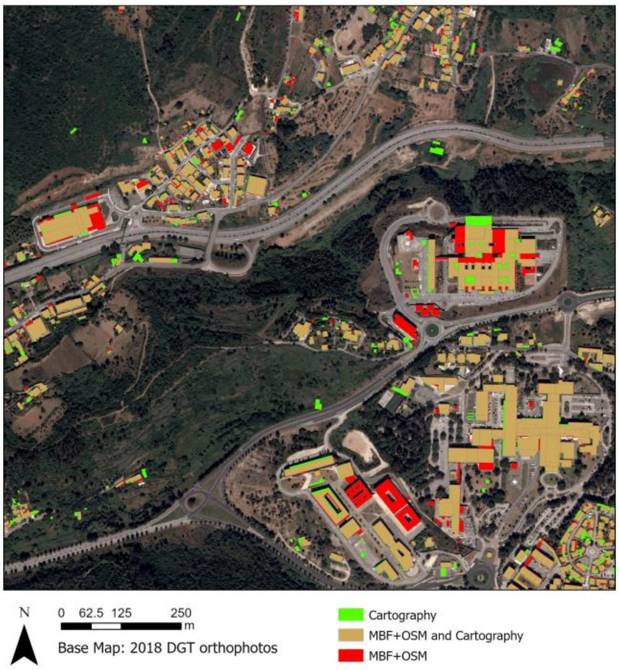

Figure 13. Buildings dataset resulting from the merge of MBF and OSM buildings for the subregion B of study area.

The value added by the proposed methodology when applied to other areas, in relation to the use of the original MBF or OSM datasets, depends upon the data available in OSM and its quality. For regions where almost no data is available in OSM, or the available buildings data have low quality, the generation of a new dataset by merging MBF and OSM may not be worth the effort. However, as OSM has data for most of the world, and data are constantly being added and edited by the OSM contributors (so that it is kept updated); and also as the quality of the building data in OSM is in general very good, the methodology is likely to be useful to obtain a dataset of the buildings with improved quality when compared with the use of either MBF or OSM data alone.

Future work will test the application of the proposed methodology to other areas of the world, with different levels of data completeness and positional quality in OSM, to assess if the results obtained for this study area are reproduced for other parts of the world. Improvements to the merging methodology will also be tested, so that not only the intersection of MBF and OSM buildings are used to identify which MBF buildings are replaced by OSM buildings, but to identify polygons corresponding to the same buildings.

Acknowledgements

The authors thank the Coimbra Municipality for providing access to the 2007 topographic cartography of the municipality with a nominal scale of 1:5000.
This work was supported by the Portuguese Foundation for Science and Technology under the project grants UID/00308/2025 with DOI UID/00308/2025.

References

Antoniou, V., & Skopeliti, A. (2015). MEASURES AND INDICATORS OF VGI QUALITY: AN OVERVIEW. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-3-W5*, 345–351. https://doi.org/10.5194/isprsannals-II-3-W5-345-2015

Biljecki, F., Chow, Y. S., & Lee, K. (2023). Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. *Building and Environment*, *237*, 110295. https://doi.org/10.1016/j.buildenv.2023.110295

Borkowska, S., & Pokonieczny, K. (2022). Analysis of OpenStreetMap Data Quality for Selected Counties in Poland in Terms of Sustainable Development. *Sustainability*, *14*(7), 7. https://doi.org/10.3390/su14073728

Copernicus. (2016). *Mapping Guide v4.7 fora European Urban Atlas*.

EEA. (2022). *CLC+ BACKBONE Product Specification and User Manual* (3.0; p. 131). https://land.copernicus.eu/user-corner/technical-library/clc-bb_user_manual

European Commission. (2020). *CORINE Land Cover 2018 (vector), Europe, 6-yearly—Version 2020_20u1, May 2020*. EEA geospatial data catalogue. https://sdi.eea.europa.eu/catalogue/copernicus/api/records/71c95a07-e296-44fc-b22b-415f42acfdf0

Fonte, C. C., Bastin, L., Foody, G., Kellenberger, T., Kerle, N., Mooney, P., Olteanu-Raimond, A.-M., & See, L. (2015). VGI

QUALITY CONTROL. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-3/W5*, 317–324. https://doi.org/10.5194/isprsannals-II-3-W5-317-2015

Fonte, C. C., & Patriarca, J. (2025). Comparison of Buildings Extracted from the Collaborative Project OpenStreetMap with the Microsoft Building Footprints dataset. *2025 7th Experiment@ International Conference (exp.at’25)*, 388–393. https://doi.org/10.1109/exp.at2565440.2025.11348602

Fox, K. (2012). History of OpenStreetMap. Em *OpenStreetMap: «It’s the Wikipedia of maps»*. https://www.theguardian.com/theobserver/2012/feb/18/openstreetmap-world-map-radicals

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2023). A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. *Nature Communications*, *14*(1), 3985. https://doi.org/10.1038/s41467-023-39698-6

Hong, Z., Zhang, H., Tong, X., Liu, S., Zhou, R., Pan, H., Zhang, Y., Han, Y., Wang, J., & Yang, S. (2024). Rapid Fine-Grained Damage Assessment of Buildings on a Large Scale: A Case Study of the February 2023 Earthquake in Turkey. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 5204–5220. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. https://doi.org/10.1109/JSTARS.2024.3362809

Microsoft. (2022). *Microsoft Building Footprints*. https://github.com/microsoft/GlobalMLBuildingFootprints

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. https://doi.org/10.1186/s40965-019-0067-x

Oostwegel, L. J. N., Schorlemmer, D., & Guéguen, P. (2025). From Footprints to Functions: A Comprehensive Global and Semantic Building Footprint Dataset. *Scientific Data*, *12*(1), 1699. https://doi.org/10.1038/s41597-025-06132-z

OSM map features. (2025). *Map features—OpenStreetMap Wiki*. https://wiki.openstreetmap.org/wiki/Map_features

So, E. (2023). Data and its role in reducing the risk of disasters in the built environment. *Natural Hazards*, *119*(2), 1127–1130. https://doi.org/10.1007/s11069-022-05590-7

Zhang, Y., Zhou, Q., Brovelli, M. A., & Li, W. (2022). Assessing OSM building completeness using population data. *International Journal of Geographical Information Science*, *36*(7), 1443–1466. https://doi.org/10.1080/13658816.2021.2023158