International Cartographic Association
Association Cartographique Internationale

Geometric comparison of open Microsoft Building Footprints and official data - a case study from Poland.

Wojciech Dawid [a,*]

[a] *Faculty of Civil Engineering and Geodesy, Military University of Technology, 00908 Warsaw, Poland, wojciech.dawid@wat.edu.pl*

* Corresponding author

Keywords: geometric quality, building footprints, displacement, positional error

Abstract:

The rapid development of open-access geospatial datasets, most notably the Microsoft Building Footprints (MSBF), has transformed the landscape of urban modelling and spatial analysis. However, the geometric fidelity of these machine-learning-derived datasets, when compared to authoritative sources such as the Polish Land and Building Register (EGiB), remains a critical area of investigation. This study presents a detailed geometrical comparison of MSBF and EGiB data, focusing on the development and evaluation of a unified framework for assessing spatial agreement through high-dimensional metric analysis.

While existing literature often relies on isolated metrics like area difference or centroid distance, this research addresses a significant methodological gap: the lack of a holistic, integrated representation of geometric congruence. This research proposes a robust evaluation framework comprising 19 distinct geometric metrics categorized into dimensions of positional accuracy, orientation, overlap, shape similarity, local spatial context, and geometric detail. The employment of such varied geometric descriptors is increasingly prevalent in studies assessing the quality of open datasets, where building polygons frequently exhibit simplified geometry relative to authoritative registries (Basaraner, 2021; Fan et al., 2014). By evaluating 19 metrics simultaneously, this approach moves beyond simplistic comparisons, allowing for a comprehensive analysis of building geometry that captures subtle structural discrepancies often missed by single-indicator studies. Examples of geometry discrepancies are presented in Figure 1.

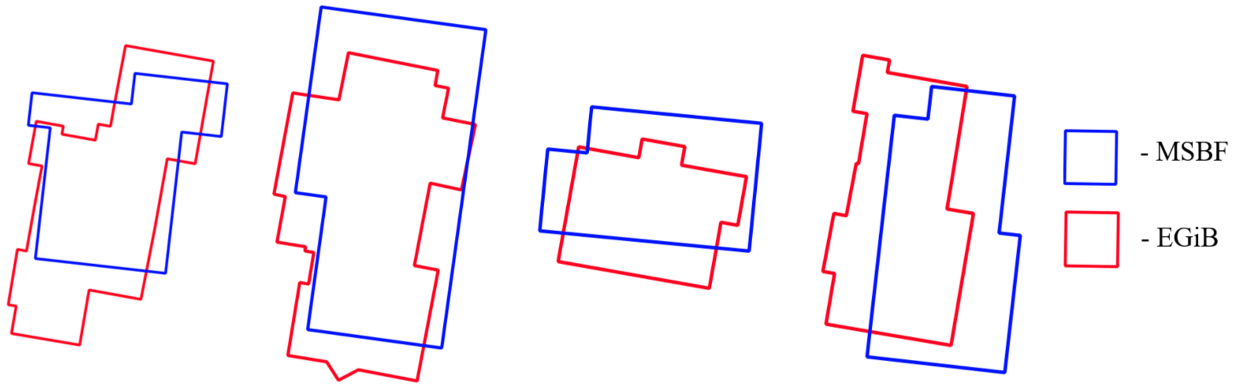

Figure 1. Discrepancies in geometry of EGiB and MSBF buildings.

The methodology is applied to the TOR Housing Estate in Lublin, Poland, a 13.5-hectare area comprising 220 buildings with a well-preserved interwar architectural style. This specific site provides a unique testing ground due to its high spatial stability over time, allowing for a precise comparison without the interference of recent construction. The results demonstrate that MSBF footprints exhibit a systematic positional shift and significant geometric simplification compared to the official registry. Specifically, the analysis identifies a mean positional error of approximately 4.0 meters and a substantial reduction in geometric complexity, with MSBF polygons containing an average of 4.5 fewer wall segments than their EGiB counterparts. Ultimately, this study demonstrates that while AI-generated footprints provide valuable large-scale data, their geometric precision remains insufficient for high-accuracy cadastral applications in historical urban layouts.

The findings from the thorough analysis are summarized below:

https://doi.org/10.5194/ica-abs-12-29-2026

1. Systematic positional displacement

The data reveals a consistent directional bias in the MSBF dataset. The mean centroid distance is 3.84 m, but this displacement is not random. The buildings are shifted an average of 2.01 meters to the East and 2.96 m to the North. The average azimuth of displacement 43.46° confirms a dominant North-East shift. This suggests a systematic error in the satellite imagery or in an automatic footprints generation procedure. In a cadastral context, a nearly 4-meter North-East shift is significant, as it frequently places building footprints across parcel boundaries or onto neighbouring properties.

2. Geometric simplification and wall detail

One of the most striking findings is the walls difference, which averages -4.57. This negative value indicates that MSBF polygons consistently use fewer vertices and segments to represent the same buildings. In the TOR estate, where semi-detached houses often feature complex architectural outlines, the AI model significantly generalizes these shapes, reducing them to simpler rectangular or simplified polygonal forms. This simplification is further reflected in the Hausdorff distance, which averages 5.56 meters, indicating that at certain points, the MSBF boundary deviates from the EGiB boundary by more than five meters.

3. Area and Volume Overestimation

On average, MSBF footprints are larger than the official EGiB records, with a mean area difference of 26.97 m^2. While the median difference is smaller (25.82 m^2), some buildings show an overestimation of up to 264 m^2. This "bloating" effect, combined with an overlap ratio of 0.38, suggests that while the footprints are generally placed in the correct location, their oversized and simplified boundaries do not align precisely with the actual building perimeters.

4. Orientation and Angular Accuracy

Despite the positional and area errors, the AI models perform relatively well in capturing building alignment. The median orientation difference is only 4.9°, although the mean is higher (20.04°) due to a few extreme outliers where buildings were rotated nearly 90°. For the vast majority of the TOR estate, the MSBF footprints correctly reflect the primary axis of the interwar street layout.

Summary of key differences are presented in Table 1.

Metric	Mean value	Interpretation
Positional error	4.01 m	Significant shift for cadastral standards.
Centroid distance	3.84 m	Systematic displacement (mostly North-East).
Area difference	26.97 m^2	Tendency for MSBF to overestimate building footprint size.
Walls difference	-4.57	Strong geometric simplification (loss of architectural detail).
Hausdorff distance	5.56 m	High maximum boundary deviation.
F1 score	0.54	Moderate overall geometric correspondence.

Table 1. Key differences of MSBF and EGiB

Acknowledgements

This work was supported by Military University of Technology in Warsaw, Faculty of Civil Engineering and Geodesy, Institute of Geospatial Engineering and Geodesy: [Grant Number 531-000130-W400-22]

References

Basaraner, M., 2021. Geometric and semantic quality assessments of building features in OpenStreetMap for some areas of Istanbul. PCR 52, 94–107. https://doi.org/10.2478/pcr-2020-0010

Fan, H., Zipf, A., Fu, Q., Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. International Journal of Geographical Information Science 28, 700–719. https://doi.org/10.1080/13658816.2013.867495

https://doi.org/10.5194/ica-abs-12-29-2026