ISSN 1831-9424

European Commission

DBSM R2025: EU Digital Building Stock Model update including satellite-based attributes

A new version of the EU Digital Building Stock Model including authoritative footprints and per-building estimates for building height, compactness, construction epoch and use

Martínez, A. M., Kakoulaki, G., Florio, P., Politis, P., Anselmo, S., Freire, S., Krasnodębska, K., Gounari, O.

2025

Joint Research Centre

EUR 40347

Contact information
Name: Ana M. Martínez
Address: Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
Email: ana.martinez@ec.europa.eu
Tel.: +39 0332 78 5273

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC142133

EUR 40347

PDF ISBN 978-92-68-28465-0 ISSN 1831-9424 doi:10.2760/0629989 KJ-01-25-323-EN-N

Luxembourg: Publications Office of the European Union, 2025

How to cite this report: Martinez, A. M., Kakoulaki, G., Florio, P., Politis, P., Anselmo, S. et al., DBSM R2025: EU Digital Building Stock Model update including satellite-based attributes - A new version of the EU Digital Building Stock Model including authoritative footprints and per-building estimates for building height, compactness, construction epoch and use, Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/0629989, JRC142133.

Content

Abstract

The EU Digital Building Stock Model (DBSM) is designed to provide a comprehensive and detailed geospatial database of individual buildings within the European Union, focusing primarily on their energy-related characteristics, although DBSM has a wide range of applications. The main goal of this initiative is to support key energy policies, such as the Energy Performance of Buildings Directive, by facilitating more targeted and informed investment decisions in the context of the building renovation wave. DBSM also aims to support the recent European Affordable Housing initiative, for example by enabling more precise "what-if" analyses and assessments of higher granularity. The current report outlines the latest developments in the Digital Building Stock Model dataset, specifically the updated version known as DBSM R2025. This new iteration offers significant enhancements over its predecessor, including more precise and detailed building polygons, as well as an expanded collection of building attributes and a complete EU-27 geographic coverage (including Azores, Madeira and Canary islands). These attributes encompass a range of factors such as building height, compactness, epoch of construction, and use type. The report provides an in-depth examination of the methodological approach employed in the development of these enhancements and discusses the validation processes undertaken to ensure the accuracy and reliability of the data.

Acknowledgements

The authors would like to thank the work of Cristiano Giovando, Fabio Lana and Manuel Fernandez in relation to DBSM.

We are also grateful to our colleagues in DG EUROSTAT for providing the authoritative datasets from several member states for comparison with DBSM R2023.

We would also like to thank Bagdagul Tan for her design on the cover page of this report.

The authors would also like to acknowledge the work of the EUBUCCO team from Mercator Research Institute of Global Commons and Climate Change and the Technical University of Berlin for their efforts homogenizing authoritative sources of building footprint data into their EUBUCCO database and making them openly available.

Authors

Ana M. Martínez[1], Georgia Kakoulaki[1], Pietro Florio[1], Panagiotis Politis[2], Sebastiano Anselmo[1], Sergio Freire[3], Katarzyna Krasnodębska[4], Olympia Gounari[5].

[1]European Commission, Joint Research Centre (JRC), Ispra, Italy
[2] European Dynamics Belgium S.A., Brussels, Belgium
[3] Engineering Ingegneria Informatica SpA, Rome, Italy
[4] Institute of Geography and Spatial Organization Polish Academy of Sciences, Warsaw, Poland
[5] TRASYS International, Brussels, Belgium

1 Definitions

- Building: any roofed structure erected above ground for any human use, including structures in slums, informal settlements, and refugees/IDP camps[1] (Pesaresi et al., 2024). Some agricultural facilities, such as large greenhouses have been excluded from DBSM R2025 when identified. Also large football stadiums have been removed when identified.
- Building footprint: a polygon delineating the gross surface and the boundaries of the planar, top-view of a building, as it could be drawn overlaid on an aerial image. As such, a building footprint is sometimes referred to as building "roofprint", to distinguish it from the building intersection with the ground. It necessarily comprises outdoor-facing walls.
- Authoritative source: a data product issued from any national or local authority with mapping competences, acknowledged by administrative institutions (e.g. National Mapping and Cadastral Agencies). This definition excludes maps involving human digitisation unsupervised from such authorities (e.g. community-based – like OpenStreetMap) and maps generated with computer-vision algorithms (like Microsoft Global Machine Learning Building Footprints).

[1] UNHCR, 2023. Geoservices. https://data.unhcr.org/en/geoservices/ Accessed: 12 May 2025.

2 Introduction

The European Commission has set as a goal for Europe to be the first continent to reach climate neutrality by 2050. Buildings represent one of the largest energy-intensive sectors, responsible for 36% of energy-related greenhouse gas emissions and approximately 40% of the energy consumption (European Parliament and Council of the European Union, 2024, 2023).

The European Commission has adopted several initiatives to reach climate neutrality and decarbonize the current and future building stock, as part of the European Green Deal (European Commission, 2019). The optimal implementation and monitoring of these initiatives require the use of detailed building datasets at EU scale, as highlighted by the needs of the following policies:

— An open access high-resolution digital model of the building stock is essential to plan and optimize building renovation strategies, help channelling investments in line with the Renovation Wave initiative (European Commission, 2020), as well as to guide the development of regulations and guidelines related to the integration of solar photovoltaic systems in building construction and renovation. The Energy Performance of Buildings Directive or EPDB (European Parliament and Council of the European Union, 2024) contributes to the objective of reducing GHG emissions by at least 60% in the building sector by 2030 compared to 2015, and achieving a decarbonised, zero-emission building stock by 2050.

— The European Union's Solar Rooftop initiative, which was launched as part of the EU solar energy strategy (European Commission, 2022a), aims to promote the widespread adoption of rooftop solar photovoltaic (PV) systems across the EU member states, setting targets for different types of buildings. A high-resolution Digital Building Stock Model (DBSM) will allow for a more accurate estimation of the solar rooftop potential at individual building level and the possibility to aggregate data in several ways (e.g. by building size, floor-area, building use, etc.), while it is also possible to look at customized areas, such as neighbourhoods or entire cities. Potential users of this dataset will not only be European policy makers, but also local and national governments and organizations in countries that are currently lagging in terms of data availability.

— Additionally, the Affordable Housing Plan is a major recent initiative launched by the European Commission aiming to address the pressing issue of housing affordability in the EU, and to promote more inclusive, sustainable, and equitable housing markets (European Commission and European Parliament, 2025; European Parliament, 2025, 2024). For the first time, a dedicated Commissioner for Housing, Dan Jørgensen, has been appointed to oversee and address the complex housing issues affecting the EU. To bolster Commissioner Jørgensen's efforts, a specialized Task Force for Housing has been established, bringing together expertise and resources to tackle the pressing housing needs and concerns of European citizens. By treating housing as a European issue, the EU is poised to develop more effective and coordinated solutions to address the housing challenges facing its member states and citizens. In this context, a detailed European Digital Building Stock Model can be crucial to guide and inform decisions based on granular data beyond aggregated regional statistics.

Furthermore, the United Nations (UN) Sustainable Development Goal (SDG)[2] 7 aims to "ensure access to affordable, reliable, sustainable and modern energy for all."

Currently, the building and associated energy-relevant characteristics necessary for effective policy design are often inaccessible or heterogeneous. Information about buildings at the EU level is available as aggregated statistics or coarse raster maps like the Global Human Settlement Layer[3]. It is now essential to move from aggregated data at administrative level to individual buildings, to which relevant attributes are attached. Their relevance is confirmed by the European Union legislation (European Commission, 2022b), which mandates free access to *High-Value Datasets*, including geospatial building data from Member States, in the broader framework of *Common European Data Spaces* (European Parliament, 2022). In this context, the Buildings datasets in particular, are expected to have a granularity up to the scale of 1:5000, a geographical coverage of the entire Member State (in a single or multiple datasets) and the following key attributes: unique identifier, footprint of the building, number of floors and type of use. This will help in the quest to build a high-resolution Digital Building Stock Model at European scale, which can provide information with different configurations and aggregation levels. This is essential to plan and optimize renovation interventions in buildings and monitor development and status. Side benefits of the DBSM also include topics related to the area of urban planning, disaster risk management such as fires, earthquakes, floods, etc.

Detailed information on building characteristics remains scattered across different member states, regions and even municipalities. While some of this data is openly available, other is not. Several initiatives are underway to unify and homogenise these different building datasets from authoritative sources, including the Geographic Information System of the Commission (GISCO)[4], in EUROSTAT or the EuroGeographics and their OME2 Project[5]. Over the past two years, numerous initiatives have leveraged these sources and combined with community-based sources such as Open Street Map to create open maps of the building stock under a common scheme. This scheme aims to identify all individual buildings at EU or global scales:

— EUBUCCO v0.1 (Milojevic-Dupont et al., 2023): is a comprehensive database featuring individual building footprints for approximately 202 million buildings spanning across all 27 European Union countries, as well as Switzerland. The database includes three key attributes for a significant portion of the buildings: building height (73%), construction year (24%), and building type (46%). EUBUCCO is developed by the Mercator Research Institute of Global Commons and Climate Change and the Technical University of Berlin.

[2] Link to the SDG: https://www.sciencedirect.com/topics/earth-and-planetary-sciences/sustainable-development-goals, last accessed on 12 May 2025.

[3] Link to the Global Human Settlement Layer site: https://human-settlement.emergency.copernicus.eu/, last accessed on 12 May 2025.

[4] Link to GISCO site: https://ec.europa.eu/eurostat/web/gisco, last accessed on 12 May 2025.

[5] Link to OME2 Project, Eurographics: https://eurogeographics.org/open-maps-for-europe/ome2-progress/, last accessed on 12 May 2025.

— Overture (Overture Maps Foundation, 2023): is an open global map including six data “themes”: addresses, buildings, base, division, places and transportation, produced by the combination of many data sources including OpenStreetMap, Meta, Microsoft, Esri, OpenAddresses and others. Overture Maps Foundation is a Joint Development Foundation Project, an affiliate of the Linux Foundation, driven by the companies Amazon, Meta, Microsoft and TomTom.

— JRC’s DBSM R2023 (Florio et al., 2023): the first version of the DBSM geospatial dataset was publicly released in the JRC Data Catalogue[6]. DBSM R2023 contains building footprints as the results of a conflation process of three open access datasets: OpenStreetMap, Microsoft Global Machine Learning Building Footprints and the European Settlement Map. The description of the procedure for this first version, as well as the quality verification of the map, has been included in (Florio et al., 2023). The objective is to release in an agile way, “often”, with incrementally more accurate and richer (in terms of building attributes) versions of the model. DBSM is developed and supported by the Joint Research Centre of the European Commission.

The DBSM is a comprehensive and homogeneous pan-European building stock model, offering a wide range of applications and benefits to various stakeholders. It is openly accessible, allowing it to facilitate bottom-up solutions that span from individual housing to neighbourhoods and entire countries. By leveraging the DBSM, public administrators, policymakers, and other stakeholders can effectively achieve their goals in various energy-related areas, including: land use planning and management, infrastructure development, renewable energy planning in cities and rural areas, demographic management, disaster risk management, emergency preparedness and many more.

Applications

- Support the affordable housing initiatives and energy-related policies
- Energy efficiency assessments
- Energy consumption of buildings
- Grid optimization and smart cities
- Sustainable development and resource management
- Assessing the potential for PV solar installations and BIPV
- Disaster risk management (earthquake, flood risk vulnerability)
- Urban planning
- Investment analysis and property valuation

[6] Link in the JRC Data Catalogue to download DBSM R2023: https://data.jrc.ec.europa.eu/collection/id-00382, last accessed on 12 May 2025.

- Behavioral studies and social equity
- Outdoor and indoor air quality

This report describes in a concise manner the content of DBSM R2025 and the methodology followed in this updated version. DBSM R2025 is the result of the conflation of geospatial data with buildings from three main open sources: cadastral data from administrative authorities included in EUBUCCO (EUB), building polygons from Open Street Map (OSM) and Microsoft Global Machine Learning Building Footprints (MSB). Several new key building attributes were introduced in DBSM R2025, including a) height, b) compactness, c) construction epoch, and d) use (residential/non-residential), which are derived from satellite sources (Global Human Settlement Layer).

2.1 Structure of the document

The rest of the document is structured as follows:

- Section 3 includes an overview of the dataset with a brief description of the attributes available per building.
- Section 4 includes a description of the methodology followed for the update of the buildings footprints (conflation process and post-processing), including some validation.
- Section 5 presents a description of the main attributes included in this version and a comparison with authoritative sources in the city of Turin, Italy.
- Section 6 provides information about the data and code availability.
- Section 7 and 8 focus on limitations and conclusions respectively.

3 Overview of DBSM R2025

In addition to the building footprint polygon, DBSM R2025 includes a comprehensive set of attributes for each feature, aimed at enriching the information available for each building. Table 1 shows the list of attributes included.

The attributes 'compactness' and 'epoch' (building age grouped in 10-year bins) are derived from satellite-based sources from the Global Human Settlement Layer (GHSL) following the methodology described in GHS-OBAT (Florio et al., 2025). The integration of the height attribute follows a similar approach to the one described in (Florio et al., 2025), with the additional integration of data from the Urban Atlas Building Height as described in Section 5.1.

Table 1 List and description of building attributes per building polygon in DBSM R2025.

Attribute name	Description
unique_id	The format of the unique identifier is as follows: `NUTS3_GISCOgridID_LON_LAT` - `NUTS3` represents the country and the region id, version 2024 of the Nomenclature of territorial Unites for Statistics at 100k resolution. - `GISCOgridID` represents the cell identifier according to the INSPIRE specification, X coordinate of the lower left corner. Y coordinate of the lower left corner. - `LON` and `LAT` are the longitude and latitude of the building centroid in degrees (WGS84). Data type: string
source	Data source of the building footprint polygon. Categorical values: -eub: EUBUCCO, only from authoritative sources -osm: OpenStreetMap -msb: Microsoft Global Machine Learning Building Footprints Data type: string
height	Height of the building polygon, extracted from Urban Atlas Building Height when available, or otherwise from GHS-BUILT-H. Numeric value. Unit: meters. Data type: float
shapefactor	Defined as the surface to volume ratio of the building polygon. It can be considered as an inverse proxy for building compactness. Unit: m^2/m^3. Data type: float
epoch	Estimated age of construction from GHS-AGE grouped in 10-year bins from 1980. Categorical values: 0: no data 1: before 1980 2: 1980-1990 3: 1990-2000 4: 2000-2010 5: 2010-2020

	Data type: integer
use	Main use of the building derived from GHS-BUILT-C and the BUTYPE building typology derived from GHS-BUILT-S. Categorical values: 0: no use assigned 1: Residential 2: Non-residential Data type: integer
area	The area of the building footprint. Unit: square meters (m^2). Data type: float
eub_json	Inherited attributes from EUBUCCO's dataset for a given polygon, which include "height", "age" and "type". Data type: string
osm_json	Inherited attributes from OSM's dataset for a given polygon[7]: building, levels, roof-shape, height, date. Data type: string
msb_json	Inherited attributes from MSB's dataset for a given polygon: height. Data type: string

Source: JRC analysis

More information on these attributes is provided in Section 4.1.3.3 for the unique_id, Section 5.1 for height and Section 5.2 for shapefactor, epoch and use and also in (Florio et al., 2025).

[7] The criteria used to determine whether two polygons from EUBUCCO and OSM refer to the same buildings is as follows. If any of the ratios of the intersection of the areas for the two polygons over the area for each building is higher or equal than 50%, then the polygon extracted from EUBUCCO will inherit some of the selected attributes from the OSM layer.

4 Update of building footprints

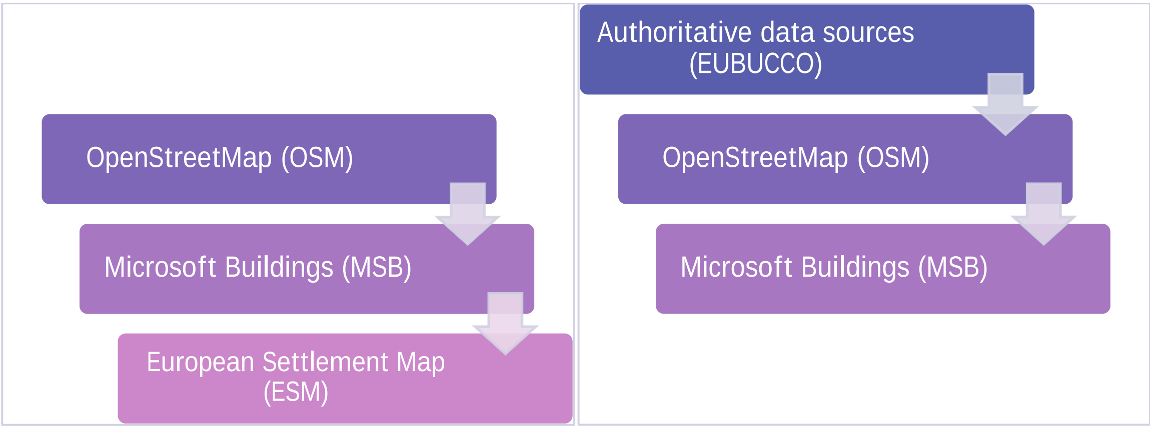
DBSM R2025 includes a new, more accurate collection of building footprints compared to its predecessor, thanks to the consideration of building polygons extracted from authoritative sources when readily available.

4.1 Methodology description

4.1.1 DBSM R2025 conflation method

In broad terms, the hierarchical conflation process in DBSM R2025 has additionally considered data from authoritative sources compared to DBSM R2023. In DBSM R2023, the conflation process included building footprints from OSM, MSB and a vectorised version of the European Settlement Map (ESM) (Florio et al., 2023), see diagram in Figure 1 left. Whereas DBSM R2025 now includes building footprints from authoritative data sources collected in the EUBUCCO (EUB) dataset (Milojevic-Dupont et al., 2023) and downloaded in January 2023. Priority is given to the EUB data whenever the source of the data is an authoritative source. In cases where the EUB data source is OSM, we instead incorporated more recent OSM data, extracted in April 2024. The last step of the conflation process is the incorporation of MSB buildings (accessed on November 2024), as shown in Figure 1 right.

Figure 1 Left: list of data sources in hierarchical priority for DBSM R2023; right: list of data sources in hierarchical priority for DBSM R2025, where authoritative cadastral data sources have been included and ESM has been excluded.

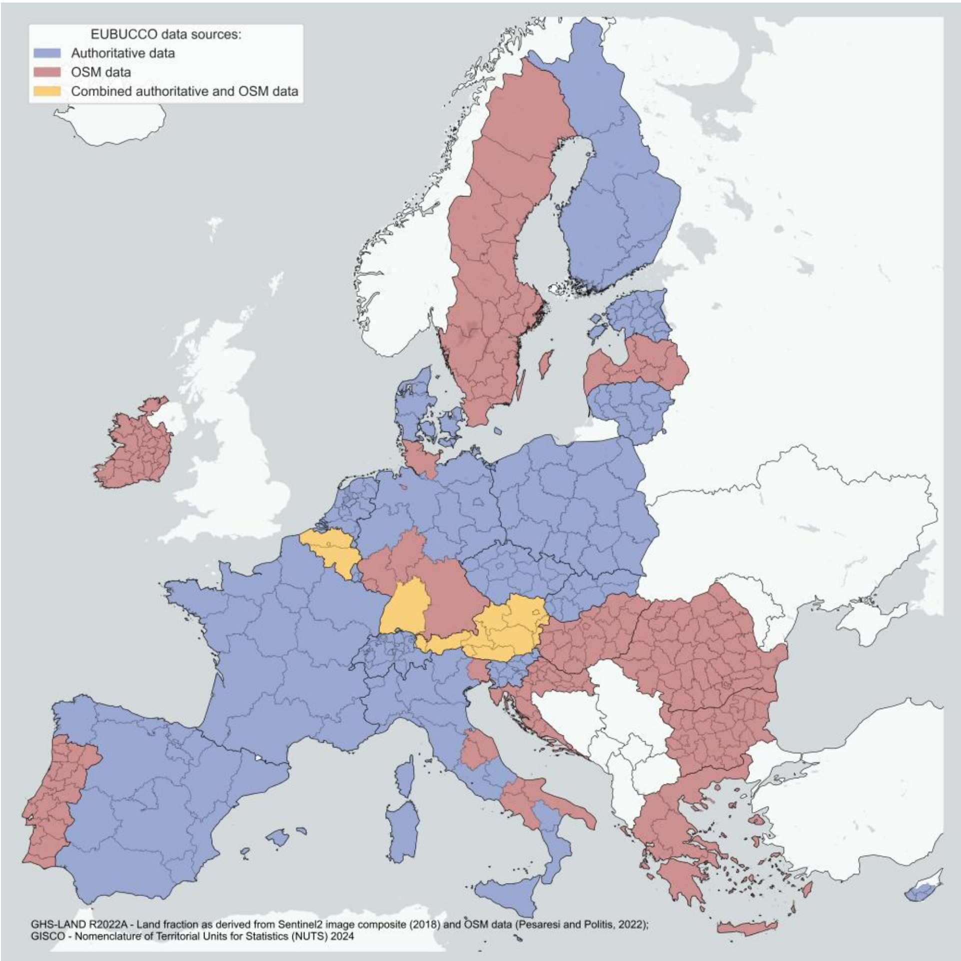


Source: JRC analysis.

More precise information of the data sources:

1. Authoritative data sources – EUBUCCO v0.1 (EUB) has been incorporated as highest priority when a polygon is available. Only data from authoritative sources in EUBUCCO have been considered (blue areas in Figure 2), meaning that building polygons from OSM in EUBUCCO v0.1 have not been used in DBSM R2025. In countries where both OSM and authoritative sources were combined in EUBUCCO v0.1, only polygons extracted from authoritative sources have been considered in the conflation process for DBSM R2025. Information on the height, age and type of the building is also retained. The data is distributed under ODbL-1.0 license, and the authoritative sources included retain their original licenses as listed in the metadata[8].

Figure 2 EUBUCCO v0.1 data sources in administrative units at level 1 from the database of Global Administrative Areas (GADM L1 units), version 4.1[9].



Source: JRC analysis.

[8] EUBUCCO's table information on input dataset https://api.eubucco.com/v0.1/files/cc5b1b47-e1fc-4887-8a26-d8878648c3df/download, last accessed on 15 May 2025.

[9] Database of Global Administrative Areas: https://gadm.org/download_world.html, last accessed on 15 May 2025.

2. OpenStreetMap (OSM): it has been accessed on April 2024 using the OGR2OGR GDAL tool. The building tags kept from the numerous available tags are: building, levels, roof-shape, height, and construction date. The equivalence with the original OSM tags is displayed in Table 2. The data is distributed under ODbL-1.0 license.

Table 2 List of OSM tags considered.

building	'building', 'building:levels', 'building:use', 'building:colour', 'building:material', 'abandoned:building', 'was:building:use', 'building:prefabricated', 'disused:building', 'building:flats', 'building:levels:underground', 'building:min_level', 'demolished:building', 'building:style', 'was:building', 'building:type', 'building:cladding', 'building:part', 'building:roof:levels', 'building:material:concrete', 'building:facade:colour', 'building:architecture', 'building:condition', 'building:year_built', 'disused:building:use', 'building:facade:material', 'historic:building', 'ruined:building', 'ruins:building', 'source:building', 'building:walls', 'building:1910', 'maybe:building', 'building_type', 'planned:building', 'building:architect', 'not:building', 'was:building:levels', 'building:units', 'building:use:residential', 'building:structure', 'seamark:building:function', 'building:name', 'former:building', 'building:source', 'building:shape', 'building:parts', 'building:roof'
levels	'building:levels', 'roof:levels', 'building:levels:underground', 'building:roof:levels', 'was:building:levels'
roof-shape	'roof:orientation', 'roof:shape', 'roof:levels', 'roof:colour', 'roof:material', 'was:roof:material', 'roof:height', 'roof:level', 'roof:material:1839', 'source:roof:material', 'roof', 'roof:direction', 'building:roof:levels', 'roof:angle', 'roof:material:note', 'roof:material:1749-1940', 'roof:material:1941-1997', 'building:roof'
height	min_height', 'height', 'roof:height', 'seamark:light:height', 'maxheight', 'seamark:light:1:height', 'seamark:light:2:height', 'seamark:light:3:height', 'seamark:landmark:height', 'maxheight:physical', 'height_source', 'source:height'
date	'check_date', 'start_date', 'check_date:opening_hours', 'end_date', 'survey:date', 'construction_date', 'opening_date', 'source:start_date', 'thatch_date', 'ruins:start_date', 'ref:IE:smr:date', 'thatch:start_date', 'source:date', 'amenity:end_date', 'thatch:end_date', 'end_date:railway', 'date', 'check_date:compressed_air', 'check_date:currency:XBT', 'check_date:wheelchair'

Source: JRC analysis

3. Microsoft Global Machine Learning Building Footprints (MSB) dataset downloaded in May 2024. Buildings from MSB are considered as a third priority and included when they do not overlap with the EUB-OSM conflated dataset. The MSB height was only included as an attribute when the confidence level (provided by the data source) is higher than 90%. The height information is also retained. The data is distributed under ODbL-1.0 license.

4.1.2 Methodological differences between DBSM R2023 and DBSM R2025

There are a few differences regarding the conflation process followed in DBSM R2025 with respect to the one followed in DBSM R2023.

One of the first differences between the two versions is that in DBSM R2023, the entire conflation process was implemented using the QGIS Model Designer. In contrast, the conflation and post-processing for DBSM R2025 have been developed using Python 3.10. This transition allows for a more flexible and transparent methodology. The in-house Python code enables easier parallelization and optimization of the conflation process, reducing the processing time to just two days for all EU-27 countries for over 600 million buildings (cumulatively across the original data sources) in vector format. This number is reduced to about 271 million buildings in the final dataset (after conflation and post-processing steps). The source code will be soon published in code.europa.eu.

In DBSM R2023, as shown in (Florio et al., 2023), a vectorised version of the European Settlement Map (ESM) was used as a third source in the hierarchical conflation process. In DBSM R2025, the ESM vectorised is not considered, as most of the areas where ESM was filling the gaps in OSM and MSB in DBSM R2023, are now covered by authoritative sources of better quality.

The advantage of introducing building footprints from authoritative sources, is that these tend to be more reliable (although this is not always the case, see examples in Section 7). However, this also means that the rich collection of OSM building attributes is potentially lost when selecting the footprint polygons from authoritative sources over OSM. To avoid losing this information, which can be particularly interesting for creating training data collections for machine learning algorithms, we aimed to find a rule to identify when two footprint polygons from EUB and OSM represent the same building, allowing the EUB footprint polygon to inherit the OSM attributes. This rule is based on the overlapping area, and it is defined as follows:

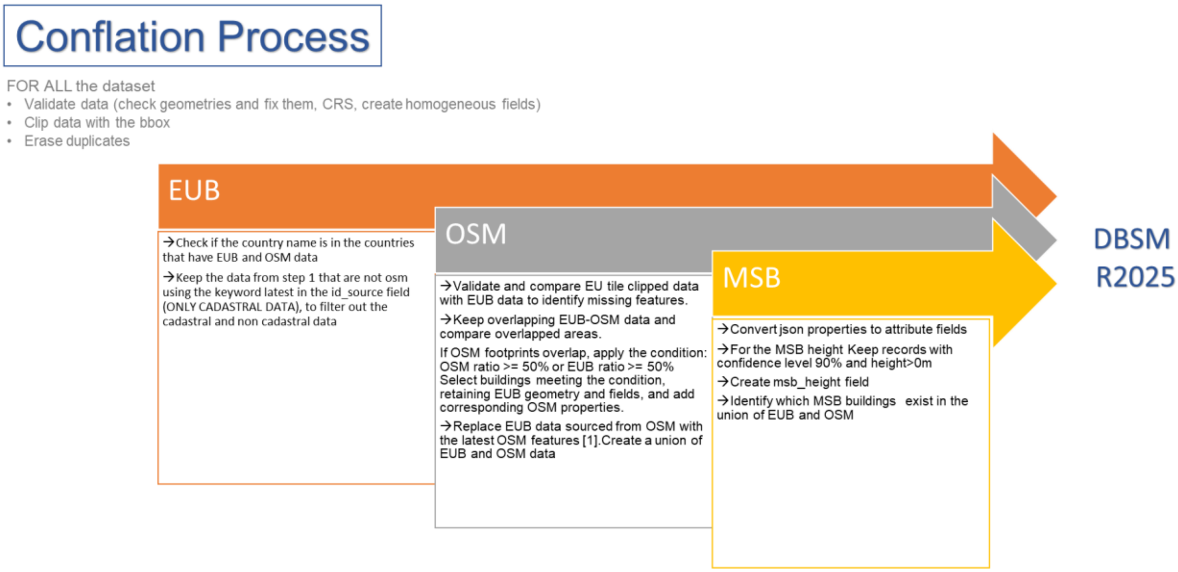
$$ratio_{OSM} = \frac{intersection(area_{EUB}, area_{OSM})}{area_{OSM}}$$

$$ratio_{EUB} = \frac{intersection(area_{EUB}, area_{OSM})}{area_{EUB}}$$

- If the area $ratio_{OSM}$ or $ratio_{EUB}$ is larger or equal than 50%, it is assumed that the two building polygons represent the same building, and the EUB polygon inherits the OSM attributes.
- If the area $ratio_{OSM}$ or $ratio_{EUB}$ is lower than 50%, no building attributes are inherited.

Another difference between the two versions is the geometric difference analysis. In DBSM R2023, if two building footprint polygons from different sources overlapped by 20%, then a geometric difference analysis would be performed to extract all parts of the MSB not covered by the OSM one. After thorough visual checks, it was noticed that in many cases, small overlaps were due to misplacement of the same building, which led to the presence of duplicated buildings or the addition of extra geometries leading to larger building footprint areas. For this reason, in DBSM R2025, the geometric difference is not included. More details about the conflation process in DBSM R2025 can be found in Figure 3.

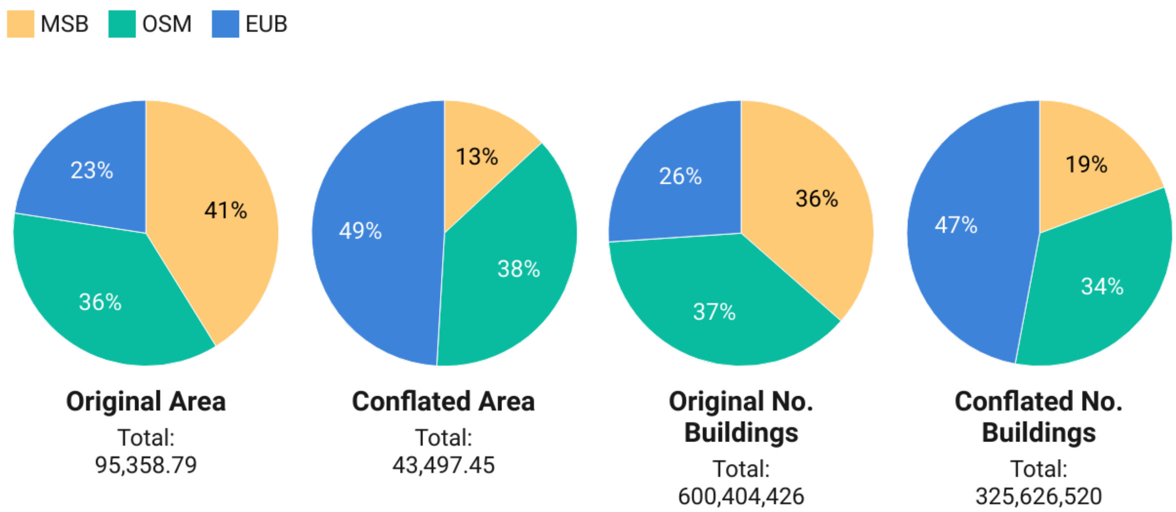
Figure 3 Methodology of the conflation process followed in DBSM R2025



Source: JRC analysis.

The total building footprint number and the corresponding area in m^2 from the original datasets and the results of the conflation process per source, are presented in Figure 4. It can be observed that the original MSB dataset corresponds to 41% of the total area from all the sources, but after the conflation process decreased to 13%, as the EUB and OSM datasets completed most of the country surfaces. The conflation resulted in 49% of the area representation from EUB, corresponding to 47% of the total number of buildings. OSM follows with 38% of area coverage, corresponding to 34% of the total number of buildings.

Figure 4 Percentage of available area in the original datasets (Original Area) and after conflation (Conflated Area); number of buildings in the source original datasets (Original No. Buildings) and after conflation (Conflated No. Buildings)



Source: JRC analysis.

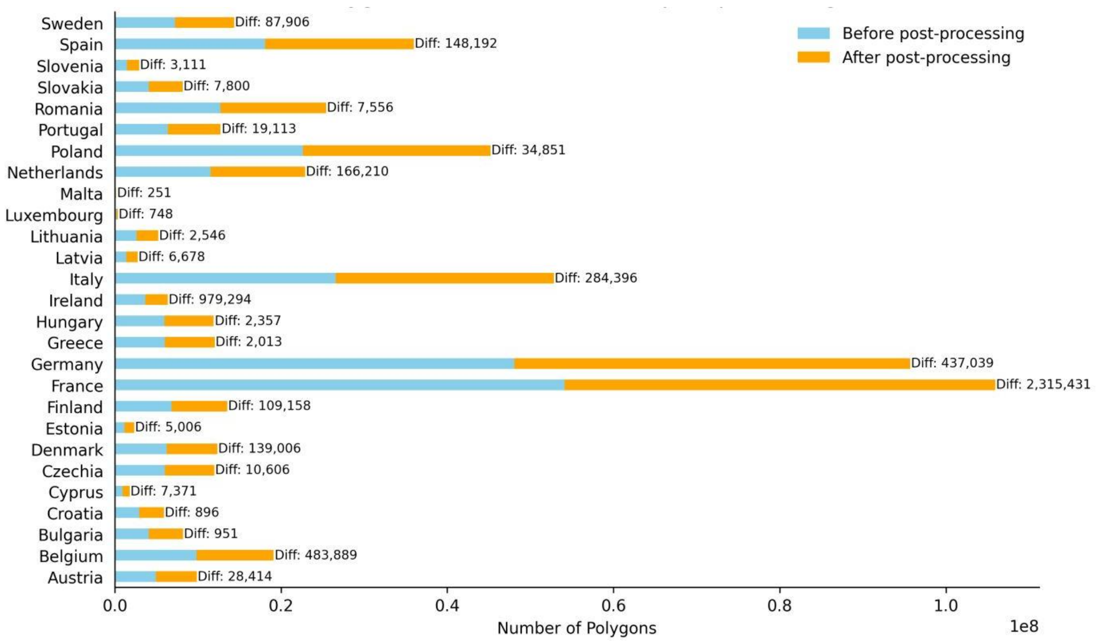
4.1.3 Post-processing steps

There are three main post-processing steps to the above-mentioned conflation process:

4.1.3.1 Removing duplicate and overlapping buildings

— Why do these duplicates exist? Some of the data sources contain duplicated and overlapping buildings. In some cases, especially in OSM data where users may have digitised the footprint of the whole parcel and the actual building footprint. In other cases, there are chimneys or house extensions digitized over the main building. The duplicate buildings were eliminated using build-in functions in Python by examining identical geometries of buildings. These values range from 251 buildings removed in Malta to 2,315,431 buildings removed in France. In total we removed 1.2% of total conflated buildings. Figure 5 shows the number of buildings removed per country.

Figure 5 Number of polygons before and after the removal of duplicates and number of removed polygons (diff.) per country.



Source: JRC analysis.

— There are cases where several smaller buildings are contained in a single larger polygon: in this case, we have noticed that most of the time (although not always), this corresponds to smaller fractions of a building being individualised. We have decided to remove these smaller overlapping buildings, although we acknowledge there will be cases where maintaining this would have been preferred.

4.1.3.2 Identification of large false positives and very small polygons

— After visual inspection, and as documented in GitHub by the EUBUCCO team[10], we identified several large polygons, particularly in Spain originating from authoritative data sources in EUBUCCO. These polygons often represent lakes, water reservoirs, and photovoltaic installations, or groups of buildings plus other areas combined in a single polygon. Additionally, we have identified large polygons covering agricultural facilities (such as large greenhouses), ruins, mines, port areas and other non-built-up surfaces, which were subsequently deleted. For a more automated review, we visually inspected all building footprint polygons larger than 2 ha that overlapped with a water mask using the GHS-LAND (Pesaresi and Politis, 2022) and forestry areas using the Corine Land Cover[11] Polygons that didn't correspond to individual buildings were removed after detailed visual inspection; the total number of polygons removed per country is shown in Table 3. Figure 6 provides examples of large polygons that did not correspond to actual buildings in various areas of Spain.

Table 3 Number of large (>2 ha) building polygons overlapping the water mask removed per country.

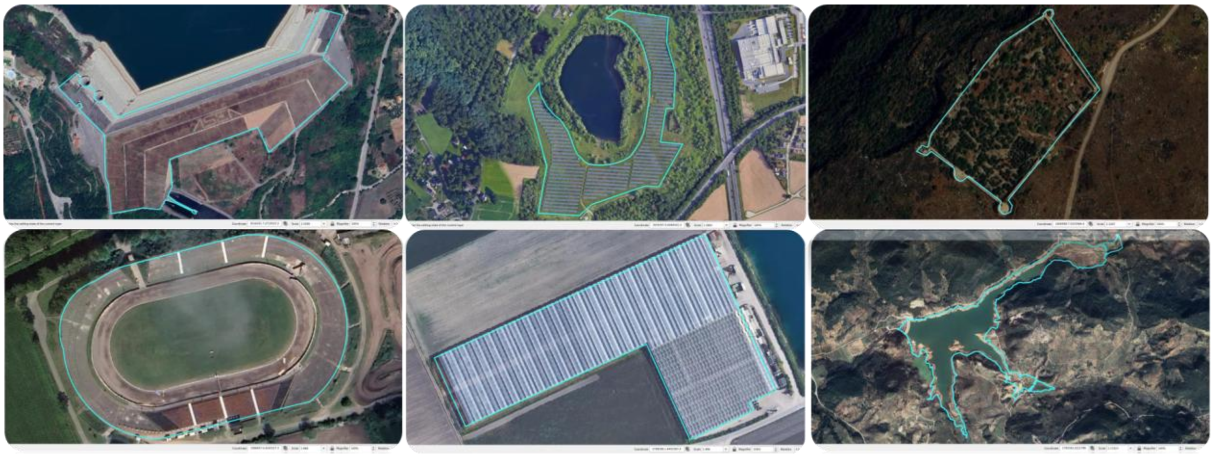
Country	Total number of false positive buildings > 2 ha	Country	Total number of false positive buildings > 2 ha
Austria	1	Italy	16
Belgium	1	Lithuania	4
Cyprus	1	Netherlands	19
Denmark	1	Poland	26
Finland	25	Portugal	15
France	40	Romania	5
Germany	17	Slovakia	16
Greece	7	Slovenia	2
Hungary	4	Spain	447
Ireland	1	Sweden	7
		Total	**665**

Source: JRC analysis

[10] eubucco-analysis/factsheets/spain_analysis.md at main · ai4up/eubucco-analysis · GitHub. Accessed on 2 May 2025.

[11] https://land.copernicus.eu/en/products/corine-land-cover/clc2018. Accessed on 2 May 2025.

Figure 6 Visual examples of large polygons not corresponding to real buildings in Spain.



Source: JRC analysis.

— In addition, during thorough visual inspection, we identified building footprints that did not correspond to real buildings, but represented olive trees, windmills, boats parked at a marine dock, etc. These buildings had an area that does not exceed 5 m^2 or even in some cases tiny geometries from the original sources < 0.1 m^2 existed. Figure 7 shows a screenshot with examples of these observations. Visual inspection has highlighted that most of these polygons did not correspond to real buildings. Thus, we decided to remove all building footprint polygons with area smaller than 5 m^2.

Figure 7 Examples of polygons of small size features (olive trees, wind farms) identified as false positive buildings.

Source: JRC analysis.

4.1.3.3 Assignment of the building identifier

The post-processing scripts also assign a unique identifier to each of the building polygons, which has the following format:

`NUTS3_GISCOgridID_LON_LAT`

For example, MT002_N2400E4500_X34845_Y68358.

For the creation of the unique identifier, we utilized the Nomenclature of Territorial Units for Statistics (NUTS), a geographical classification system used by the European Union to divide member states into regional levels for statistical purposes (European Commission and Council of the European Union, 2003). Specifically, we chose the smallest regional division, NUTS3 version of 2024 and 100k resolution. By including the NUTS ID in the unique identifier, we enable users to aggregate data from the smallest region (NUTS3) to the country level (NUTS0).

The next component of our unique identifier represents the GISCO grid cell. The GRID_id is the grid cell identifier according to the INSPIRE specification, with X coordinate of the lower left corner and Y coordinate of the lower left corner[12]. This feature allows users to aggregate data at different grid scales, ranging from 1 km to 100 km.

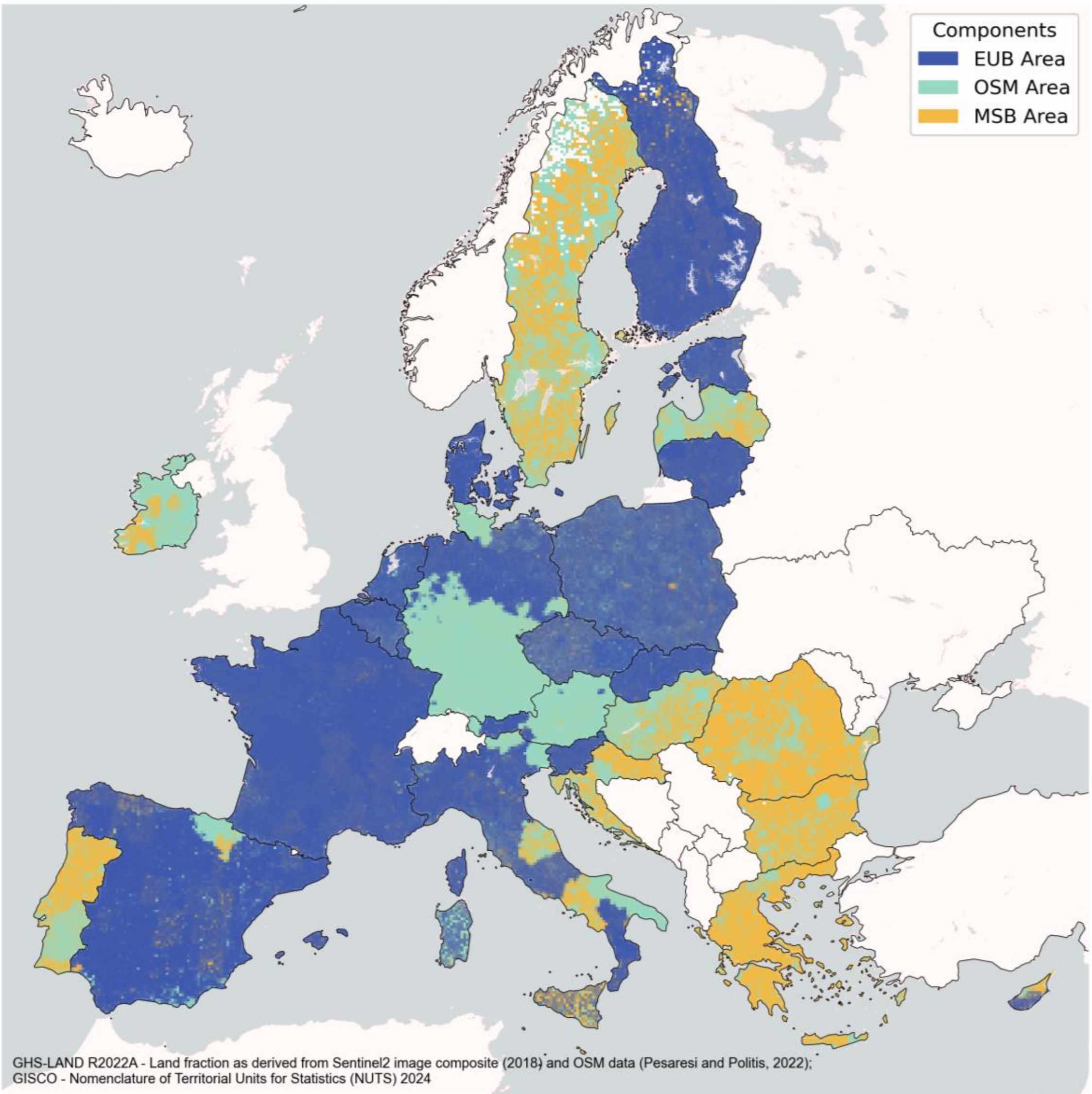
Finally, the last part of the unique identifier encodes the latitude and longitude of the building's footprint centroid, providing precise spatial referencing in degrees, in WGS84.

4.1.3.4 Statistics of building polygons in DBSM R2025

The above-mentioned post-processing steps resulted in the removal of 16.7% of total conflated buildings, corresponding to 54,410,061 building footprints in total. The total final number of building footprints is equal to 271,216,459, with a corresponding total footprint area of 37,370 km^2. An overview of the areas covered per data source is presented in Figure 8. The distribution of the EUB area here follows closely the distribution of the area for authoritative data in EUBUCCO in Figure 2 with some notable exceptions in Spain (Basque Country and Navarra) where no authoritative data was available in the EUBUCCO version used; and in Belgium, parts of Germany and Austria, where both authoritative and OSM data is present in EUBUCCO. MSB area takes a predominant role in countries like Greece, Bulgaria, Romania, Croatia, Portugal and Sweden to some extent.

[12] GISCO Grids, metadata descriptor: https://gisco-services.ec.europa.eu/grid/GISCO_grid_metadata.pdf, last accessed on 26 May 2025.

Figure 8 Geographical distribution of the original data sources of the buildings in DBSM R2025. The colour of each pixel is proportional to the mix of EUB area (blue), OSM area (green) and MSB area (yellow).



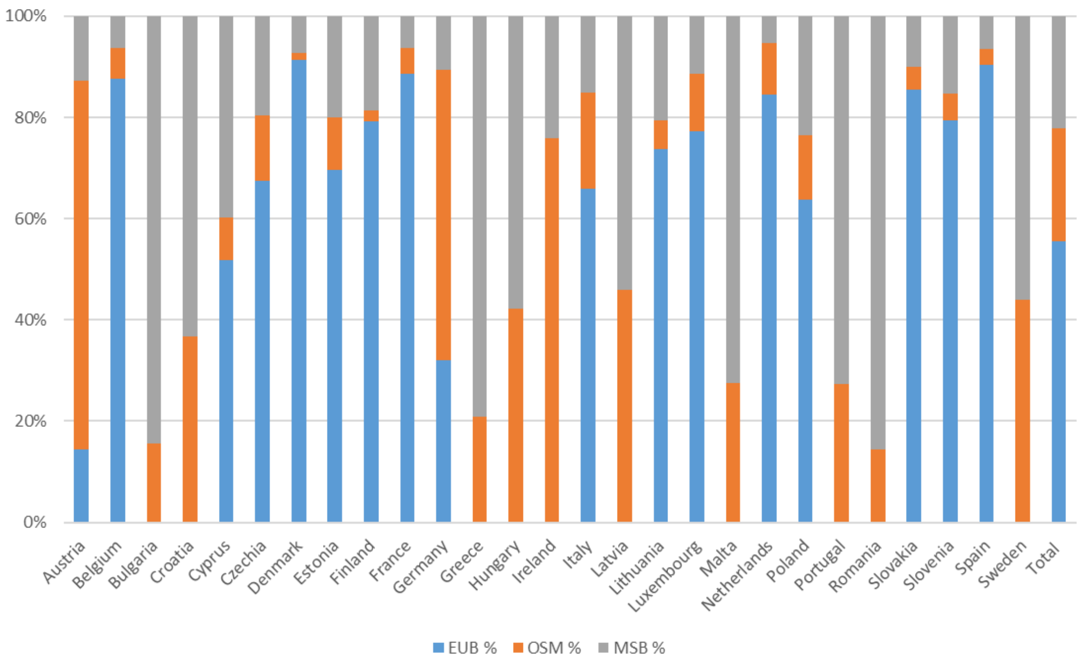
Source: JRC analysis.

Figure 9 shows the percentage of buildings for each of the input data sources: EUB, OSM and MSB and the overall for EU (with the exact number presented in Table 4.

There are 15 countries where the data source coming from authoritative sources is higher than 50% (Belgium, Cyprus, Czechia, Denmark, Estonia, Finland, France, Italy, Lithuania, Luxembourg, Netherlands, Poland, Slovakia, Slovenia and Spain). In 3 other countries (Austria, Germany and Ireland) the most popular input source is OSM in terms of number of buildings. There are 9 countries where more than 50% of the buildings come from MSB (Bulgaria, Croatia, Greece, Hungary, Latvia, Malta, Portugal, Romania and Sweden). However, if we are looking at the percentage of footprint area covered, then Hungary, Latvia and Sweden are dominated by OSM (see Figure 10).

Overall, in terms of footprint ground area covered, the majority of the surface comes from authoritative sources (57%), 28% of the area from OSM and only 15% from MSB. The final representation in terms of number of buildings corresponds to 55% from EUB, 23% from OSM and 22% from MSB, which would indicate that the buildings incorporated from MSB tend to be smaller ones.

Figure 9 Percentage of number of buildings per input data source per country in DBSM R2025



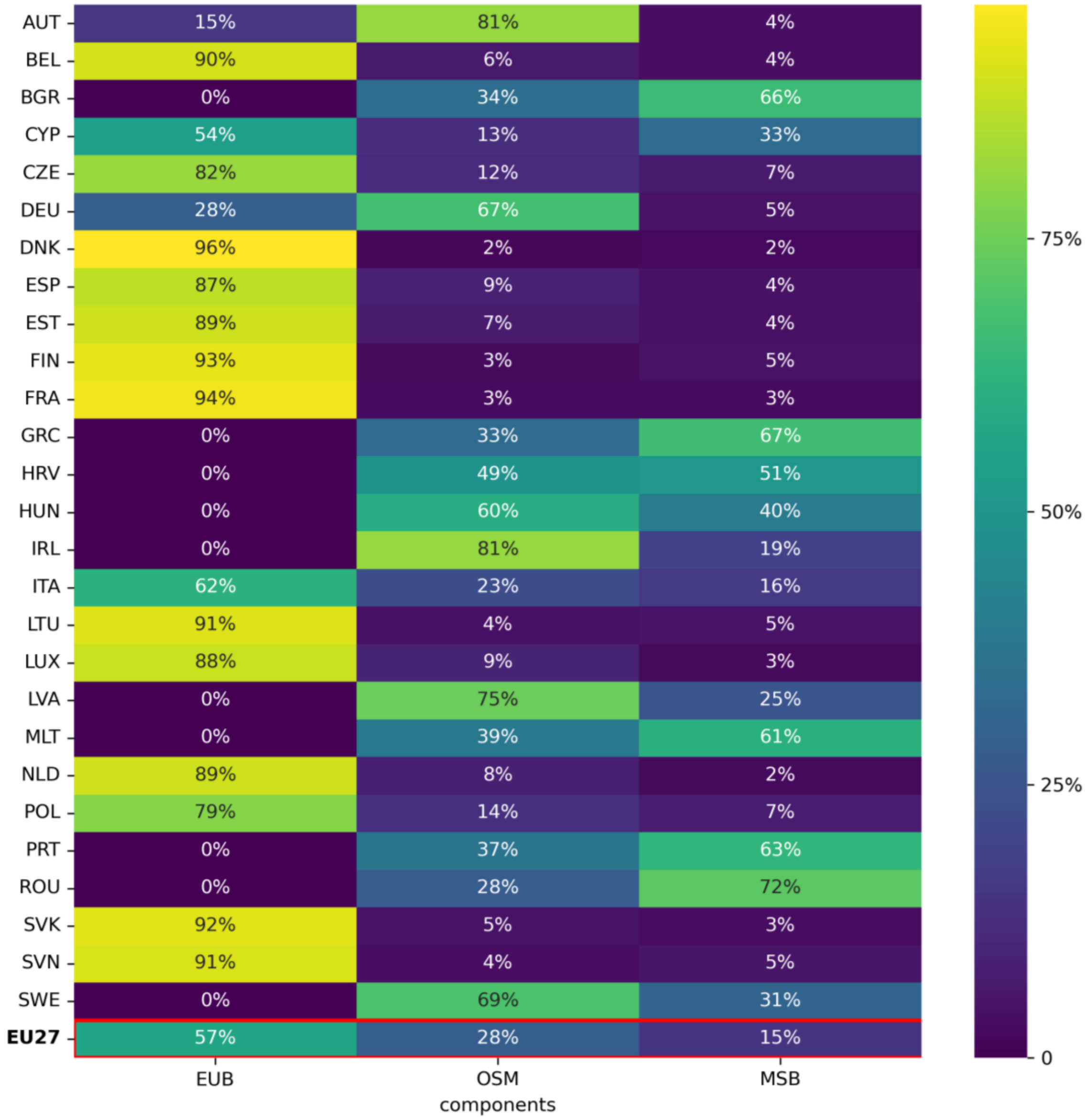
Source: JRC analysis.

Table 4 Final number of buildings per data source per country and overall number for EU27.

Country	EUB	OSM	MSB	Total
Austria	702,540.00	3,581,023.00	629,295.00	4,912,858.00
Belgium	8,155,661.00	572,446.00	583,425.00	9,311,532.00
Bulgaria	0	641,797.00	3,457,191.00	4,098,988.00
Croatia	0	1,080,610.00	1,869,624.00	2,950,234.00
Cyprus	460,299.00	75,245.00	353,997.00	889,541.00
Czechia	4,037,043.00	771,050.00	1,176,004.00	5,984,097.00
Denmark	5,563,236.00	85,166.00	445,187.00	6,093,589.00
Estonia	803,198.00	121,121.00	230,717.00	1,155,036.00
Finland	5,315,504.00	144,554.00	1,242,481.00	6,702,539.00
France	45,921,571.00	2,633,642.00	3,237,061.00	51,792,274.00
Germany	15,221,811.00	27,366,792.00	5,024,828.00	47,613,431.00
Greece	0	1,254,548.00	4,759,651.00	6,014,199.00
Hungary	0	2,515,614.00	3,432,104.00	5,947,718.00
Ireland	0	2,776,878.00	880,442.00	3,657,320.00
Italy	17,310,027.00	4,976,259.00	3,976,414.00	26,262,700.00
Latvia	0	631,456.00	741,908.00	1,373,364.00
Lithuania	1,924,350.00	145,613.00	538,825.00	2,608,788.00
Luxembourg	143,523.00	21,469.00	20,998.00	185,990.00
Malta	0	20,776.00	54,604.00	75,380.00
Netherlands	9,589,734.00	1,159,446.00	605,500.00	11,354,680.00
Poland	14,404,030.00	2,887,115.00	5,289,769.00	22,580,914.00
Portugal	0	1,737,114.00	4,618,243.00	6,355,357.00
Romania	0	1,822,410.00	10,884,419.00	12,706,829.00
Slovakia	3,486,360.00	182,805.00	405,858.00	4,075,023.00
Slovenia	1,160,268.00	77,398.00	223,413.00	1,461,079.00
Spain	16,195,070.00	567,347.00	1,149,013.00	17,911,430.00
Sweden	0	3,133,449.00	4,008,120.00	7,141,569.00
Total	150,394,225.00	60,983,143.00	59,839,091.00	271,216,459.00

Source: JRC analysis

Figure 10 Percentage comparison of footprint ground area covered by each of the 3 data sources: EUB, OSM, and MSB, overall and per country.



Source: JRC analysis.

4.2 Validation of building footprints

We believe that no data can be considered as ground truth at the moment of writing (Minghini et al., 2024). For this reason, comparisons of DBSM R2025 have been made against GHS Built-up surface data (GHS-BUILT-S), the Overture buildings map and authoritative sources collected by GISCO EUROSTAT at aggregate 10-km tile level per country, and more into detail against cadastral data issued from the municipality authority in the city of Turin, Italy.

The GHS-BUILT-S spatial raster dataset (Martino Pesaresi and Politis, 2023) depicts the global distribution of the built-up (BU) surface in 10 m tiles for the year 2018, supported by a Sentinel-2 (S2) image composite (GHS-composite-S2 R2020A). The built-up surface is the gross surface (including the thickness of the walls) bounded by the building wall perimeter with a spatial generalization matching the 1:10K topographic map specifications, compatible with the concept of “building footprint”.

The Overture maps Foundation is producing a dataset with six data themes: addresses, buildings, base, division, places and transportation. Overture buildings, used for comparison here, is a collection of conflated building features from different open datasets[13]. The first release date was in April 2023. The data format is GeoParquet files stored on both AWS and Azure.

Since DBSM R2023 did not include data from authoritative sources, part of the validation was made comparing it with authoritative data sources. Now that data from EUBUCCO is considered as first option in the hierarchical conflation process, this type of comparison is no longer valid. However, in the framework of the High Value Data Regulation, Eurostat is collecting building footprints from National Mapping and Cadastral Agencies, complementing the efforts made by the Eurogeographics Consortium. The dataset includes exclusively data from authoritative sources. In spite of its realisation being still in progress[14], a preliminary comparison in a few European countries could be carried out here to offer a validation perspective against authoritative data collected by GISCO EUROSTAT.

One could argue that adding data from OSM and MSB to building footprints from authoritative sources could lead to an overestimation of the building stock, but we also need to consider that some of the cadastral databases were collected some years ago (since 2007 for the data from Abruzzo in Italy to a few datasets from 2021, being the average for the available countries the year 2017 roughly)[15], and likely new buildings have been constructed in the meantime.

4.2.1 Comparison with GHS-BUILT-S

The building polygons in the DBSM R2025 dataset are compared to the European Commission's GHSL Built-up surface layer for epoch 2018 at 10 m resolution (GHS-BUILT-S R2023A) (Martino Pesaresi and Politis, 2023): the aim is to understand the coverage of built-up areas at a pan-

[13] The conflation priority and other details of the Overture buildings dataset are here: https://docs.overturemaps.org/guides/buildings/. Last accessed on 17 May 2025.

[14] The Eurostat dataset can be visualised here: https://observablehq.com/@eurostat-ws/building-demography

The data source is available here: https://gisco-services.ec.europa.eu/pub/Inspire/ANNEX-1/Buildings/GPKG/. Last accessed on 17 May 2025.

[15] EUBUCCO’s input dataset metastable: https://api.eubucco.com/v0.1/files/cc5b1b47-e1fc-4887-8a26-d8878648c3df/download. Last accessed on 29 May 2025.

European level. GHS-BUILT-S is a spatial raster dataset that estimates the distribution of built-up surfaces from 1975 to 2030 at 5-year intervals, with a 100 m resolution. As an independent data source, GHS-BUILT-S relies on Sentinel-2 imagery from 2018, eliminating discrepancies between country authorities and temporal misalignments typical of community-based data like OSM.

Previous assessments of GHSL built-up surface estimates have shown a tendency to overestimate built-up surfaces (Uhl and Leyk, 2022). To address this, an adjustment factor is calculated as the ratio of the total built-up area estimated using DBSM R2025 data over all Europe, to the total built-up area derived from the GHS layer. This “cumulative” adjustment factor enables the evaluation of built-up areas derived from GHS-BUILT-S layers, considering the temporal accuracy and coverage of the reference data.

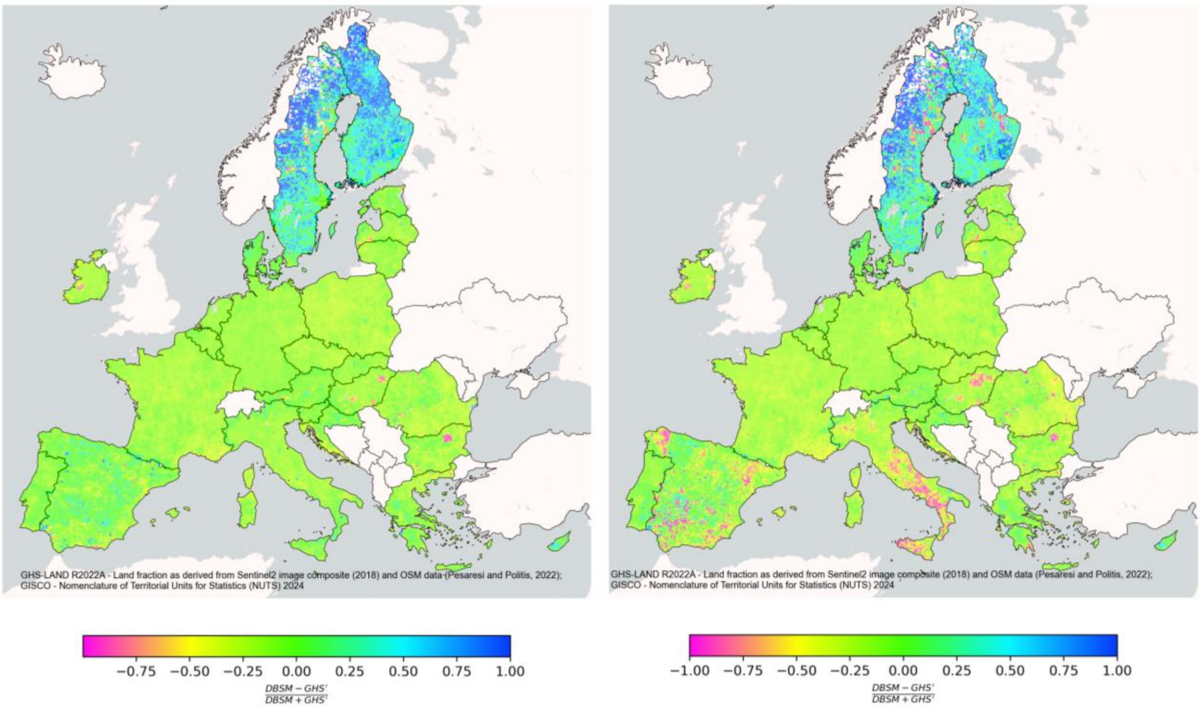
The adjustment factor is estimated at a pan-European level and is equal to 0.71, with respect to DBSM R2025. This factor was set to 0.68 in the comparison with DBSM R2023 (Florio et al., 2023). For the spatially explicit pan-European level analysis, the GHS layer for 2018 is multiplied by the computed adjustment factor to obtain an adjusted layer (GHS'), which mitigates the built-up surface overestimation. DBSM R2025 is then compared to the adjusted GHS' layer at a pan-European level, using aggregates per 10 km-side grid cells.

To assess the completeness of DBSM R2025 data against the adjusted GHS' layer, the ratio of the difference between areas derived from both layers to their sum is calculated (completeness check as Normalised Difference Index - NDI). This results in an indicator with values ranging from -1 to +1, where:

- -1 indicates that only GHS' data is available for a given grid cell
- +1 indicates that only DBSM data is available for a given grid cell
- 0 indicates convergence between the two datasets in terms of built-up area per grid cell

This completeness check provides a quantitative measure of the agreement between DBSM R2025 and the adjusted GHS' layer, allowing for the identification of areas with discrepancies in built-up area coverage. The results and comparison of GHS’ with DBSM R2023 are also shown in Figure 11 left, where we can see how the new version of DBSM features more uniform results (Figure 11 right), indicating that some of the gaps existing in particular in the South European areas are now being covered by the authoritative data sources from EUBUCCO. We can still observe differences in Sweden and Finland, which are probably due to the sensitivity of the relative comparison metric to the small amount of buildings in such areas; further discrepancies are due to commission errors by certain DBSM sources, especially in forest areas, and to omission gaps in GHS-Built-S, caused by limitations of the satellite data in those areas.

Figure 11 Left: Comparison (completeness check as Normalised Difference Index - NDI metric) of DBSM R2025 with a calibrated version of GHS Built-up surface 2023R for 2018 (adjustment factor = 0.71), in 10-km tiles over EU27; and right: DBSM R2023 with another calibrated version of GHS Build-up surface (adjustment factor = 0.67).

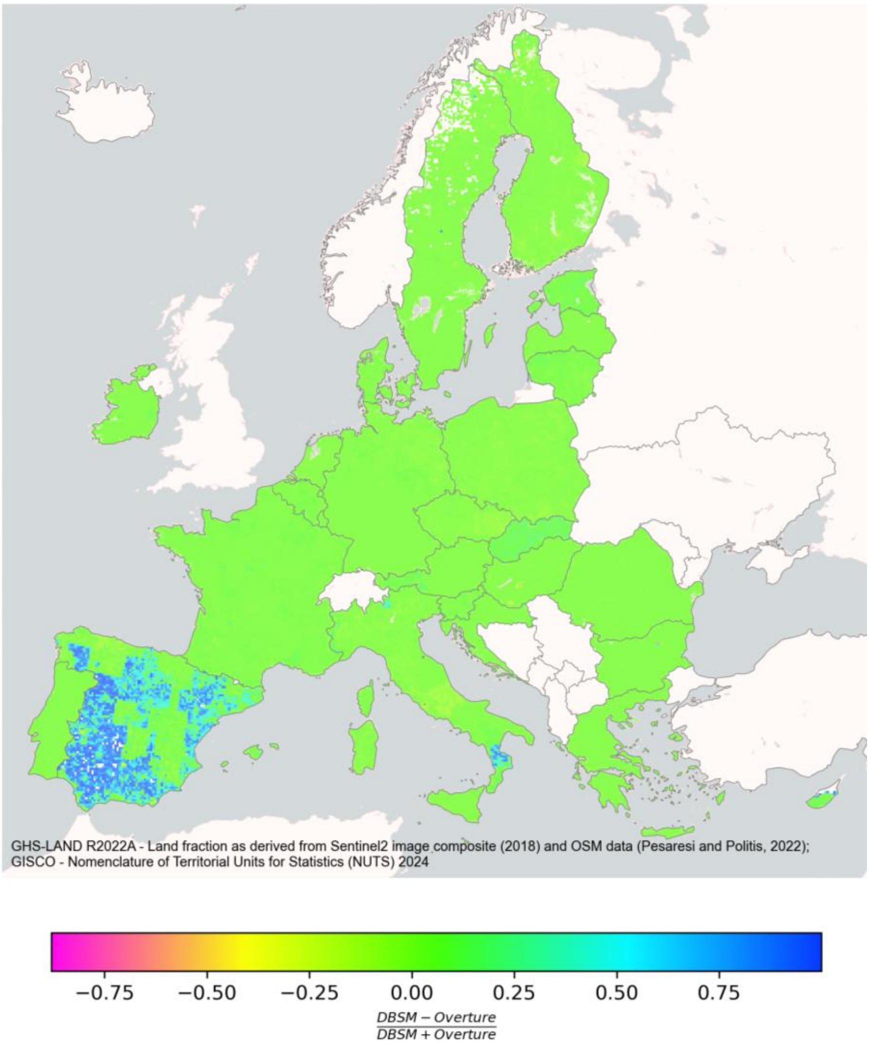


Source: JRC analysis.

4.2.2 Comparison with Overture

We have carried out a comparison of DBSM R2025 with the Overture release v.2024-07-22.0, using the same Normalised Difference Index - NDI metric as before for completeness check. The result is shown in Figure 12, where we can see how the two datasets show a significant alignment, except for large areas in Spain, and other smaller zones in Southern Italy, Cyprus, and Åland in Finland. Many of these gaps in Overture have been addressed in most recent releases, making the dataset more complete over Europe.

Figure 12 Comparison (Normalised Difference Index - NDI metric) of built-up surface in DBSM R2025 vs Overture Buildings v.2024-07-22.0, in 10-km tiles over EU27.

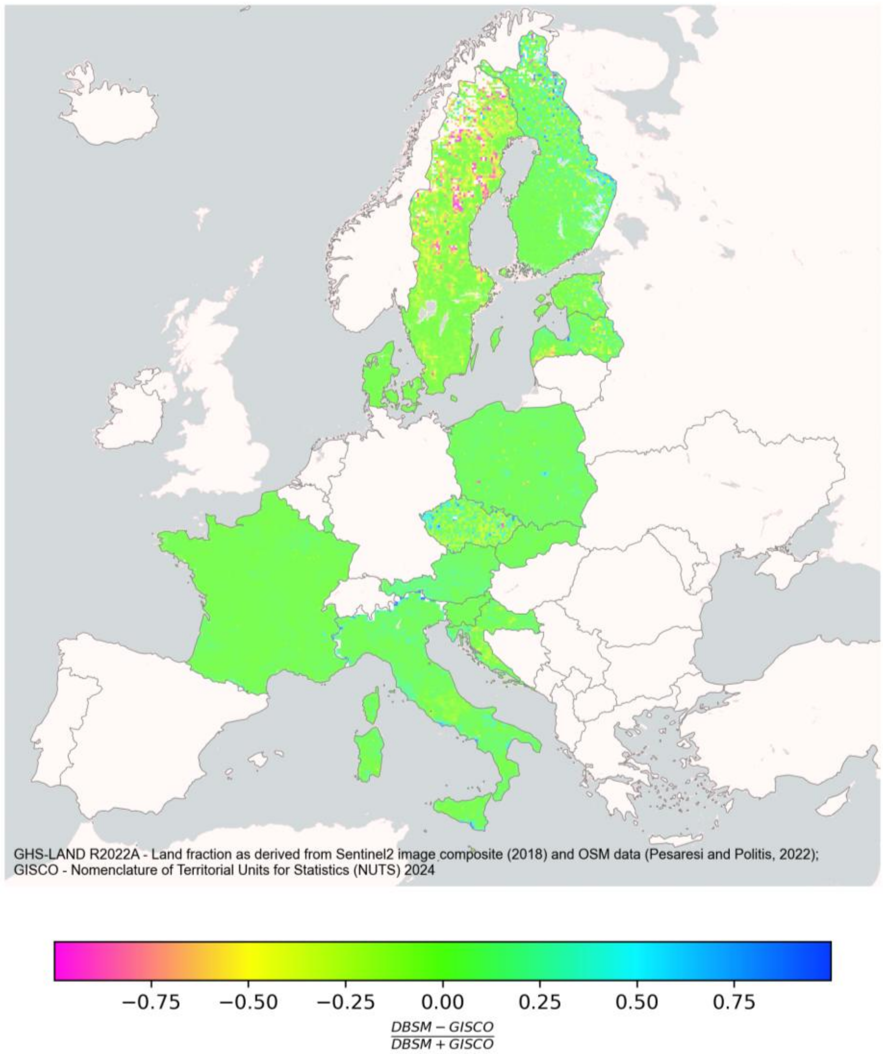


Source: JRC analysis.

4.2.3 Comparison with Eurostat GISCO

The Geographic Information System of the Commission (GISCO) at Eurostat is collecting building footprints from national authoritative sources. Being the Eurostat building datasets still incomplete, a completeness check index (Normalised Difference Index - NDI) could be computed only for 15 EU countries, and needs to be expanded further. However, as shown in Figure 13, such preliminary analysis shows a sound agreement between cumulative built-up surfaces assessed at 10-km tile level (as the values are close to 0). Few omission anomalies have to be flagged in scarcely built areas of Sweden. As soon as such authoritative data is conveyed to Eurostat and becomes increasingly available to partner institutions and the public, its incorporation in DBSM will grant improved accuracy and usability.

Figure 13 Comparison (Normalised Difference Index - NDI metric) of built-up surface in DBSM R2025 vs GISCO Buildings v.2025-05-07, in 10-km tiles over EU27.



Source: JRC analysis.

4.2.4 Comparison with authoritative sources in the city of Turin

As a next step, a close-up validation of DBSM R2025 has been performed in a smaller area in the city of Turin in Italy, where accurate data is readily available at city level, for the building footprints and the different building attributes included in DBSM R2025.

The Municipal Technical Map of the City of Turin (CTC)[16], updated in 2024, has been adopted as reference dataset. In case of data gaps in the Municipal Technical Map, the last version of the BDTRE (*Base Dati Territoriale di Riferimento degli Enti piemontesi*, reference geodataset for Piedmont Region local administrations), updated in 2024[17], was queried. In both cases, information is divided between three layers, which corresponds to volumetric units, buildings and minor buildings (garages, service buildings). The two validation datasets differ in terms of scale of acquisition, with the CTC being returned on a 1:1000 scale and the BDTRE at 1:10000.

The total area considered in this case study is 130.058 km^2. The building data in DBSM R2025 for Turin consists of 94.4% from authoritative data, in particular the BDTRE, updated in 2021. Other footprints are derived from OSM (3.42%) and MSB (2.14%).

The building geometries in DBSM R2025 are often similar to the BDTRE one in terms of detail, as they mainly come from this source, including wider surfaces than the actual ones. The total area covered by building footprints is summarized in Table 5.

Table 5 Total area in km^2 covered by building footprints in the area considered for the city of Turin, Italy.

DBSM	23.69 km^2
CTC	25.51 km^2
BDTRE	24.35 km^2

Source: JRC analysis

As a result, when validating building footprints against the CTC, it results in an overestimation of the area equal to 15.88%, while compared to the area covered by BDTRE, this value is reduced to 2.87%. This is due to the differences in terms of detail of the building footprints in the validation datasets, as CTC building outlines are more detailed.

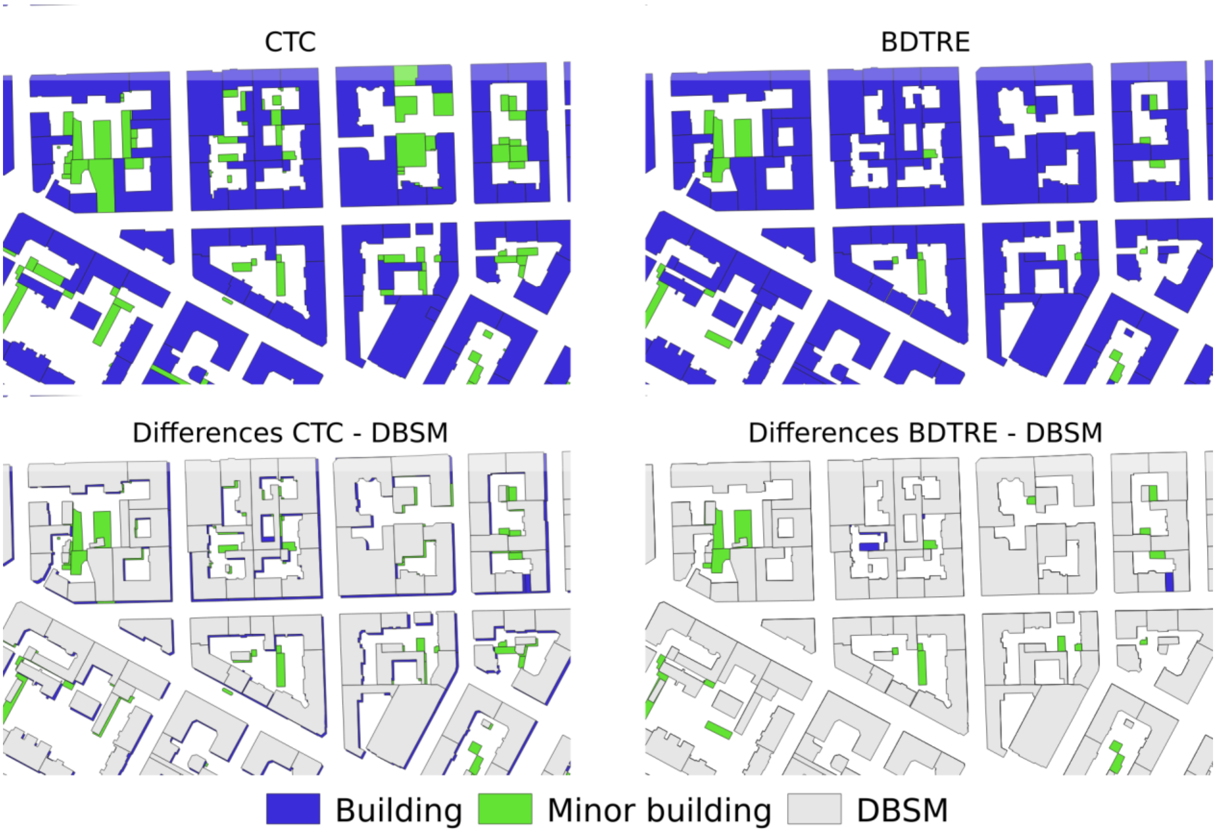
Higher differences between the two validation datasets can be observed for the buildings omitted in DBSM R2025. A surface area equal to 4.81% of the total DBSM R2025 building surface in the area analysed (4.54 km^2 in total) is absent when compared to the BDTRE. Comparing CTC with DBSM R2025, resulted in 13% of missing area for DBSM R2025. However, when comparing to the BDTRE, the majority of buildings not included in DBSM R2025 are minor buildings (e.g. garages or deposits), which account for 93.9% of the missed area from BDTRE in DBSM R2025, and 41.9% of the missed area when compared to the CTC. Figure 14 shows examples of minor buildings (in green) that are not included in the DBSM_v2 when compared to CTC (left) and BDTRE (right). These differences could in part be attributed to the decision to delete polygons smaller than 5 m^2 in DBSM R2025. On

[16] CTC dataset is accessible here: http://geoportale.comune.torino.it/viscotoga/?printEnabled=true&ricercaTopoEnabled=true&lang=it&topic=CARTOGRAFIA&bgLayer=2&layers=Viario_Viario_Corsi20180507120726829,Carta_tecnica___fogli_1_100020191211104305393. Last accessed on 2 May 2025.

[17] BDTRE dataset is accessible here: https://geoportale.igr.piemonte.it/cms/bdtre/bdtre-2. Last accessed on 2 May 2025.

the other hand, while according to the CTC there are differences also in the buildings (in blue), these differences cannot be observed when comparing the DBSM_v2 to the BDTRE.

Figure 14 For a particular area in the city of Turin, screenshots corresponding to: Top left: Original CTC data, Top right: original BDTRE data, Bottom left: DBSM R2025 in grey, differences between CTC and DBSM R2025 for minor (green) and larger buildings (blue), Bottom right: DBSM R2025 in grey, differences between BDTRE and DBSM R2025 for minor (green) and larger buildings (blue).



Source: JRC analysis.

In conclusion, in Turin DBSM R2025, in spite of minor differences, footprints can be considered reliable with an accuracy corresponding to the 1:10000 scale dataset.

5 Additional attributes

Most attributes computed at building level rely on several components of the GHS-BUILT (Built-Up) layer of the Global Human Settlement Layer (GHSL)[18] release R2023A. In particular, attributes rely on the products described herewith. For detailed descriptions on such data, see the distribution report (European Commission. Joint Research Centre., 2023) or the dedicated scientific literature (Pesaresi et al., 2024).

The GHS-BUILT-H spatial raster dataset (M. Pesaresi and Politis, 2023a) depicts the distribution of the building heights generalized at the resolution of 100 m, and referred to the year 2018. The input data used to predict the building heights are the ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30) (Japan Aerospace Exploration Agency, 2016), the NASA Shuttle Radar Topographic Mission data - 30m (SRTM30) (NASA Shuttle Radar Topography Mission, 2013), and the Sentinel-2 global pixel based image composite from L1C data for the period 2017-2018 (Corbane et al., 2020).

The GHS-BUILT-C spatial raster datasets (M. Pesaresi and Politis, 2023b) describe the built environment's morphology and main functional use, classifying areas into residential and non-residential at 10 m resolution, through analysis of Sentinel-2 satellite imagery features such as radiometry, texture, and morphology. The definition of built-up surfaces in GHSL R2023 aligns with the INSPIRE designation[19], including aboveground buildings and temporary structures for sheltering humans or goods, while excluding underground structures. Residential areas are primarily for housing, potentially mixed with non-conflicting uses, whereas non-residential areas are designated for industrial, commercial, or infrastructural purposes.

The GHS-AGE (Uhl et al., 2025) identifies the earliest 10-year period between 1980 and 2020 when at least 50% of the current built-up area existed in a 100 m grid cell. This attribute represents the estimated construction year of the majority of buildings in that cell, serving as a proxy for the construction year of the predominant buildings within the area. The limits of the epochs are designed based on the availability of satellite imagery.

5.1 Building heights (Urban Atlas Building Heights)

Building height is an essential information to many different use cases, including the estimation of useful (heated or cooled) area of buildings, cast shadow on building-integrated solar modules, outdoor climate and environment (sky view factor) (Demuzere et al., 2019) or the accurate disaggregation of the population from enumeration areas of censuses to buildings (Palacios-Lopez et al., 2022).

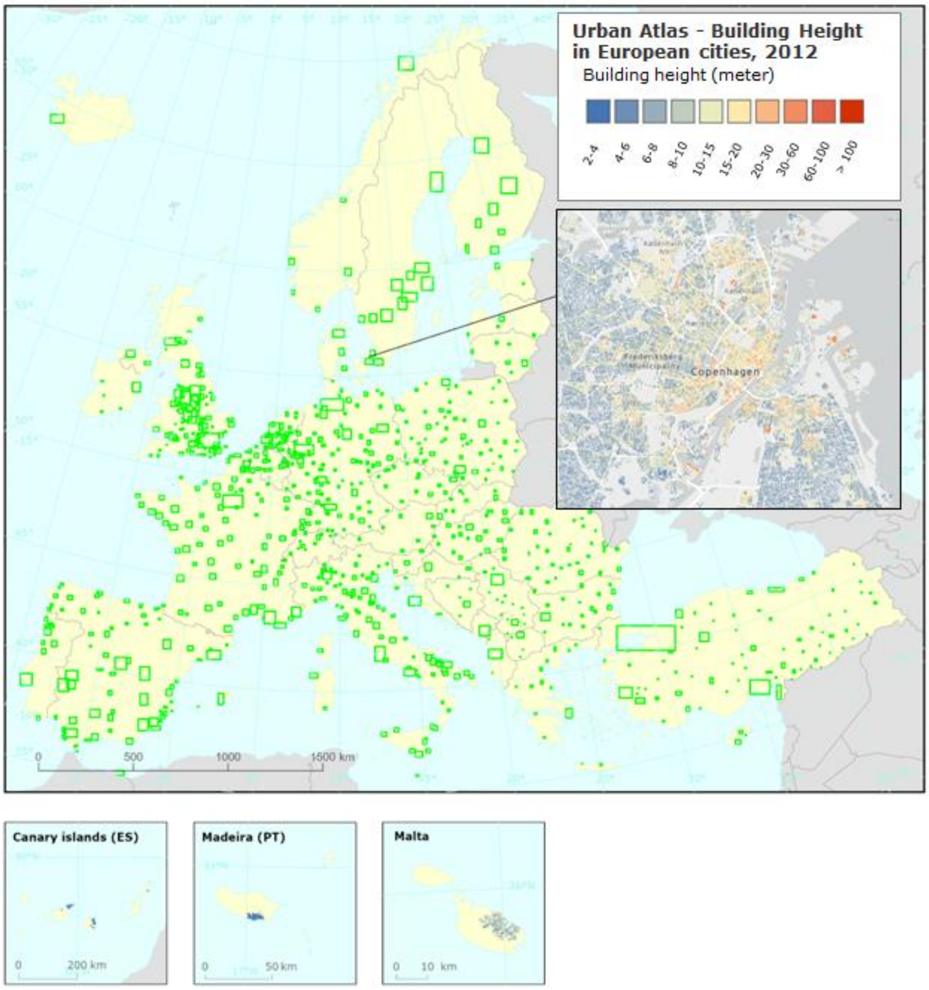
The height attribute has been assigned based on the availability of the Urban Atlas Building Height 2012 version 3 (European Environment Agency and European Environment Agency, 2022). The Urban Atlas Building Height 2012, from now on referred to here as UA, is a 10 m raster layer containing height information for about 870 core urban areas of selected cities in Europe (within

18 Link to the Global Human Settlement Layer site: https://human-settlement.emergency.copernicus.eu/. Last accessed on 12 May 2025.

19 Link to INSPIRE Data Specification on Buildings – Technical Guidelines: https://inspire.ec.europa.eu/id/document/tg/bu. . Last accessed on 12 May 2025.

the 38 members of the European Environment Agency, excluding Vaduz, Liechtenstein), and the United Kingdom. See Figure 15.

Figure 15 Core cities covered by Urban Atlas Building Heights 2012.



Source: EEA geospatial data catalogue.

As the UA has higher horizontal resolution compared to GHS-BUILT-H R2023A (100 m) (M. Pesaresi and Politis, 2023a), it is expected that the height values assigned to individual buildings are also more accurate. However, the fact that the temporal extent of the data from Urban Atlas Building Height dates from 01-01-2012 to 31-12-2014, could imply that the height of newer buildings is assessed incorrectly.

For this reason, we carried out some comparisons with height information available from authoritative sources when considering only GHS-BUILT-H as source, compared to using height information from UA when available. The results are presented in Table 6. The metrics reported correspond to the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Median Absolute Error (MedAE), the Pearson Correlation Coefficient r and the total number of building polygons considered within areas covered by UA in selected countries. Only countries with a comparable definition of building height, compatible with satellite-based estimates were used to this end; these include heights estimated as distance from ground to the lowest point of the roof, to the median/mean roof height, to eaves or to the 70th percentile of the point cloud generated during scanning. It can be observed that in all the cases, the error values obtained in building footprint polygons where heights are extracted from UA, are smaller than those obtained with GHS-BUILT-H. Based on this, we could say that in those areas covered by UA, we could expect error values a bit smaller than one floor height (between 1.65 – 2.89 m), whereas in those areas not covered by UA these errors are expected to be higher than one floor (between 3.19 – 4.83 m). The percentage of buildings covered by UA in DBSM is 14%, including all major cities in EU.

Table 6 Results of comparing building heights from authoritative sources (eub-heights), with heights extracted from Urban Atlas Building Heights 2012 (UA-heights) and heights extracted from GHS- BUILT-H (GHSL-heights).

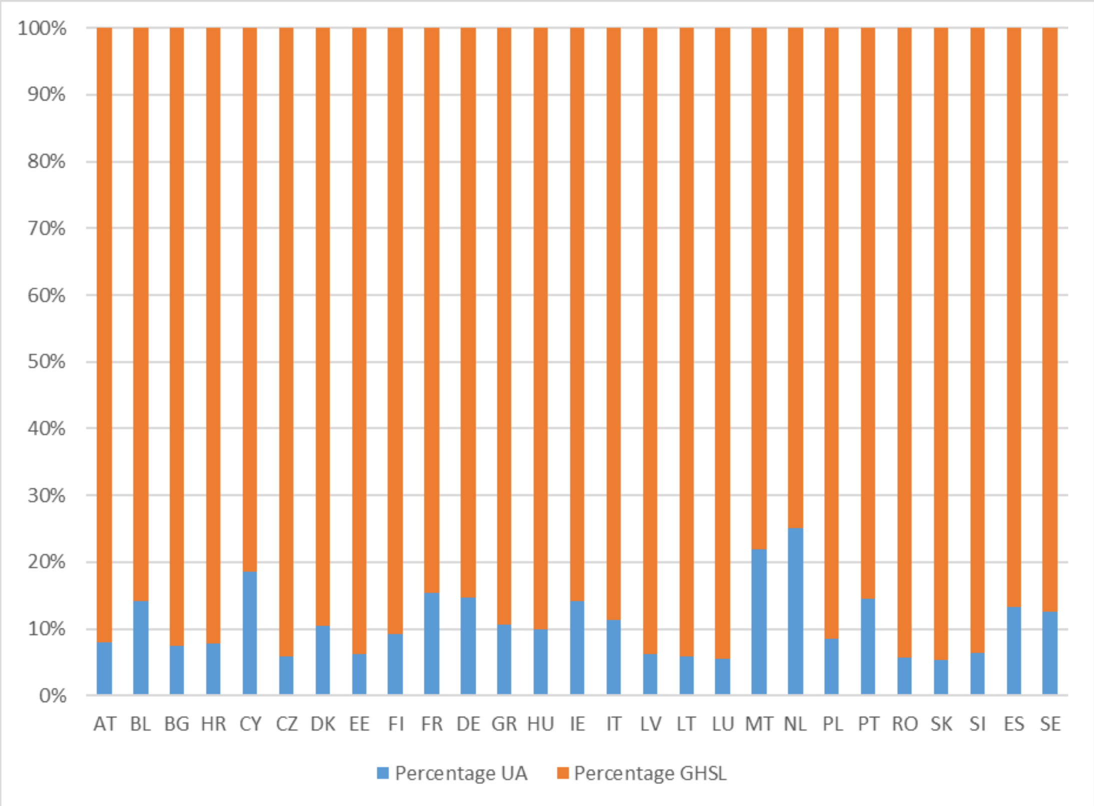
	UA-heights vs eub-heights	GHSL-heights vs eub-heights
MAE	1.96	3.61
MAPE	43.44	89.87
RMSE	2.89	4.83
MedAE	1.65	3.19
r	0.73	0.36
Buildings count	14 728 256	14 723 734

Source: JRC analysis

Based on this results, the height attribute for each building in DBSM R2025 is assigned by applying a simple decision rule. If a value from the Urban Atlas Building Height dataset (UA) is available for the building, it is selected due to its higher spatial resolution and accuracy. Otherwise, the value is taken from the GHSL-derived height layer GHS-BUILT-H. This logic ensures that the final height field reflects the most reliable data source available per building footprint with fall-back mechanisms to maximize completeness across the dataset.

Figure 16 shows the percentage of buildings per country where the height is extracted from UA and GHSL. There are two countries, Malta and The Netherlands, where the percentage of buildings coming from UA is higher than 20%, around 10 countries with more than 10% coverage, and the rest below this percentage.

Figure 16 Percentage of buildings where height is extracted from UA and GHSL per country

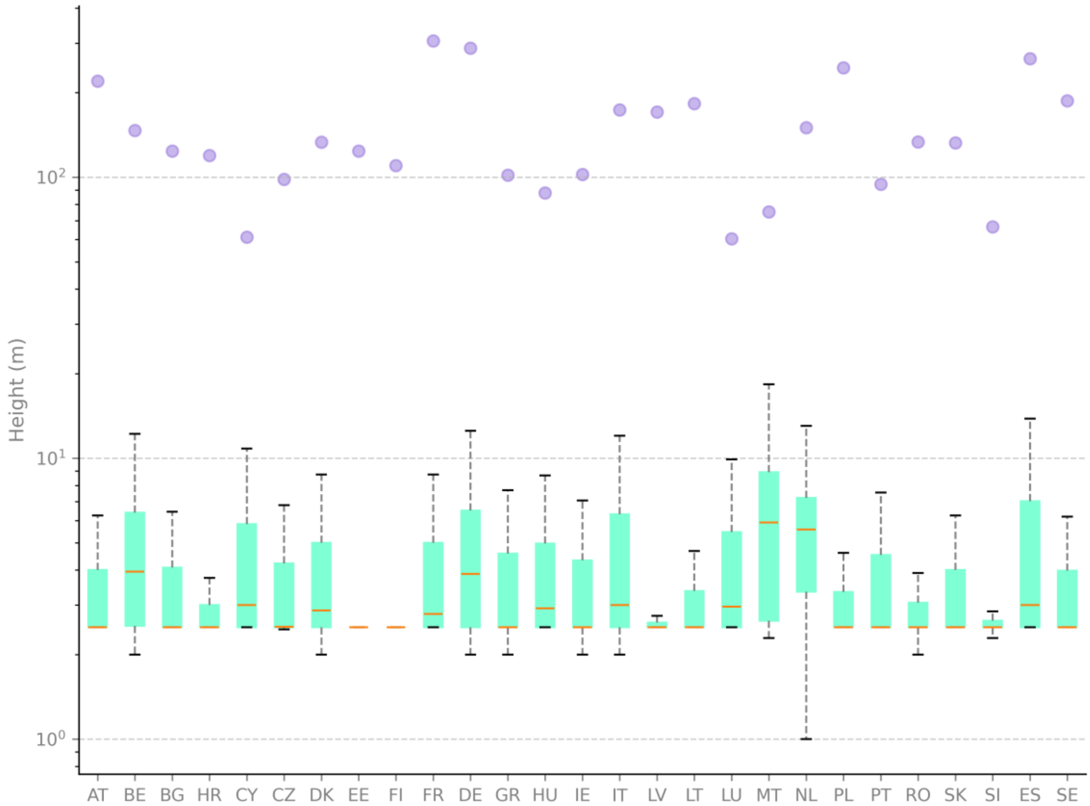


Source: JRC analysis.

Figure 17 shows the distribution of the heights for the different Member States. Here we can notice that Malta and Netherlands are showing the largest amount of higher buildings. The lowest height for most countries is around 2.5 m, as this is the lowest value considered in most cases in GHS-BUILT-H. Overall, the smallest buildings in terms of height are found in Hungary, Estonia, Finland, Latvia, Lithuania, Poland, Romania and Slovenia.

Future analysis could assess whether considering information from GHS-BUILT-H instead of the Urban Atlas for buildings on the 2010-2020 epoch could lead to more accurate estimates, thanks to more up-to-date observations, since most of the UA reference data was obtained in the time period between 2012 and 2014, although at a higher resolution.

Figure 17 Distribution of the building height per country depicted as whisker plots in logarithmic scale. The boxes represent the first and third quartile (with median orange line). The purple circles represent the maximum values.



Source: JRC analysis.

5.2 Building shape factor (compactness), construction epoch (age) and use

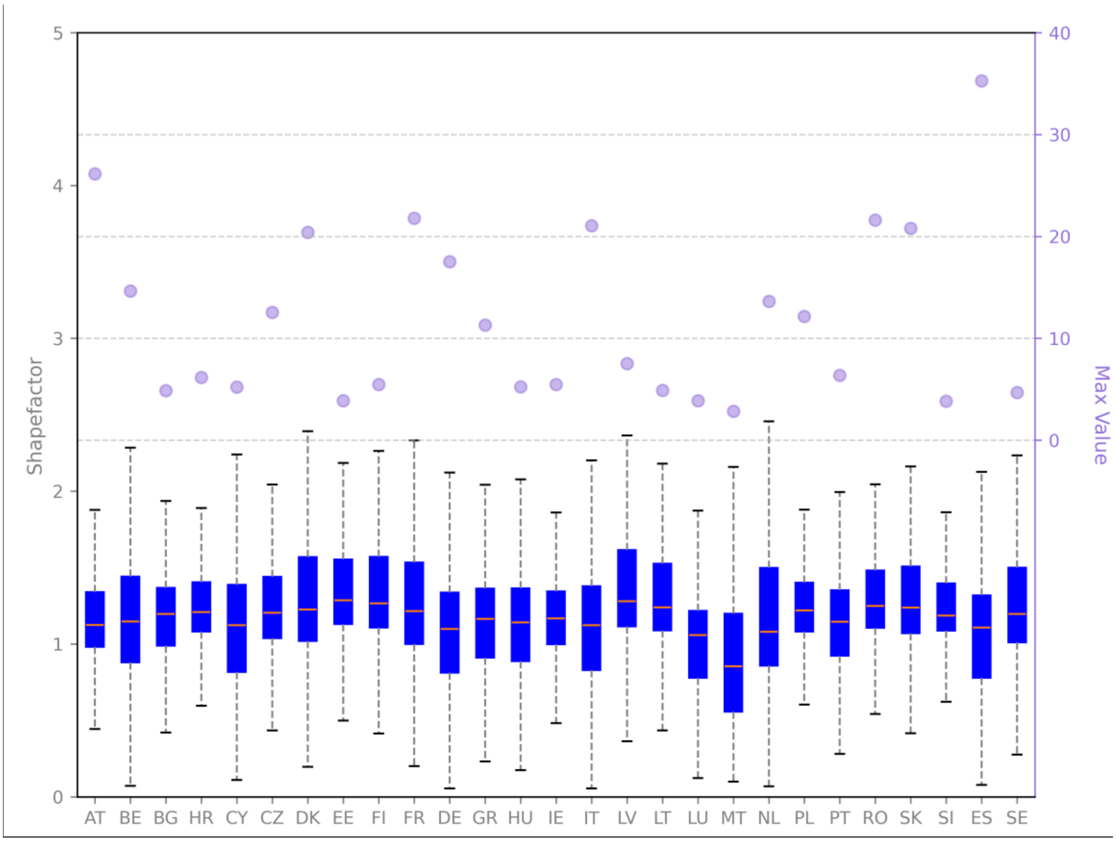
The methodology for assigning these three attributes: shape factor, construction epoch and use, is outlined in GHS-OBAT (Florio et al., 2025), where it is applied to the global Overture building dataset. For a more in-depth explanation of these attributes, refer to (Florio et al., 2025). Notably, the validation process followed in this study is also relevant to DBSM R2025, given that the majority of the validation data originates from EU27 countries. The following section provides a brief overview of each attribute, along with overall statistics for each of them in DBSM R2025.

Building shape factor (or surface to volume ratio, as an inverse proxy of the building compactness) can be used as a proxy for estimating energy demands in buildings. It is defined as the ratio between the outdoor exposed building envelope surface and the heated volume. In general, higher values of the shape factor make it harder for buildings to be energy efficient and vice versa: in cold climates, lower values of shape factor are preferred to minimize heat loss, whereas in hot, dry climates is recommended to reduce heat gain (Institut Catalá d'Energia et al., 2004). Additionally, buildings in the city centres tend to have lower values of shape factor than those in the city outskirts or in rural areas. .

Figure 18 shows the distribution of the shape factor, or surface to volume ratio, of buildings per country. Malta stands out as the member state with lower values of shape factor. Countries like

Belgium, Cyprus, Germany, Greece, Hungary, Italy, Luxembourg, The Netherlands, Portugal and Spain have their first quartile below 1, although often show high dispersion. On the other end, countries like Latvia, Estonia, or Finland show slightly higher overall shape factor values, followed closely by many other countries.

Figure 18 Distribution of the building shape factor (inverse proxy for compactness) per country depicted as whisker plots. The boxes represent the first and third quartile (with median orange line). The purple circles represent the maximum values.

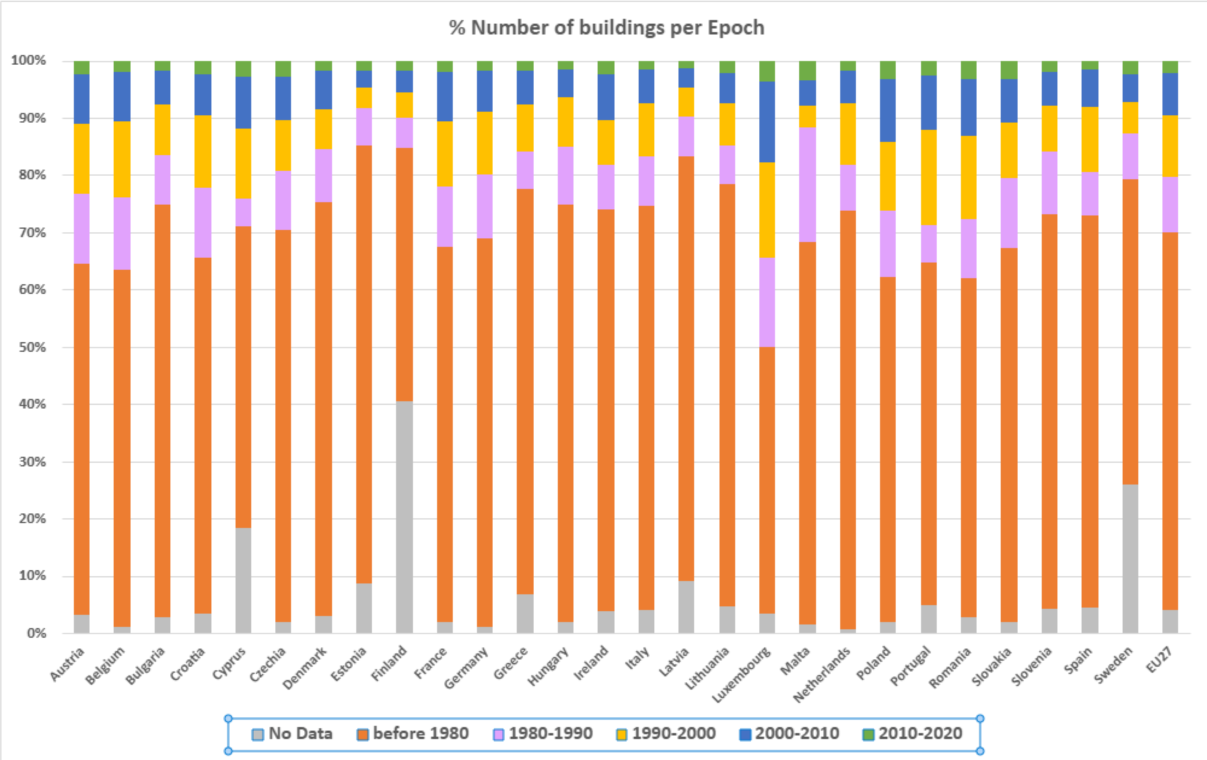


Source: JRC analysis.

Building age (epoch) is another essential factor to estimate energy demand in buildings, as older buildings tend to have poorer energy performance. Authoritative information on the age of construction is available for just a few countries in Europe and scattered in local heterogeneous data sources. In order to provide a first estimate for all member states in the EU, we have used the information derived from satellite data in GHS-OBAT (Florio et al., 2025). The building construction year is estimated using the GHS-BUILT-S dataset, which measures built-up surfaces globally from 1975 to 2020. The dataset is queried to find the year when 50% or more of the built-up surface was present in each 100 m grid cell, representing the likely construction epoch as described in the gridded GHS-AGE data product (Uhl et al., 2025). The estimated construction year is then categorized into 10-year time steps, called epochs, ranging from before 1980 to 2010-2020, and expressed as an integer from 1 to 5 (0 is used to identify “no data” for areas outside of the GHS domain).

Figure 19 shows the percentage of buildings estimated to belong to these different categories, where we see that the epoch of buildings estimated to be built before 1980 is largely dominating the rest.

Figure 19 Percentage of buildings per country per construction epoch category.



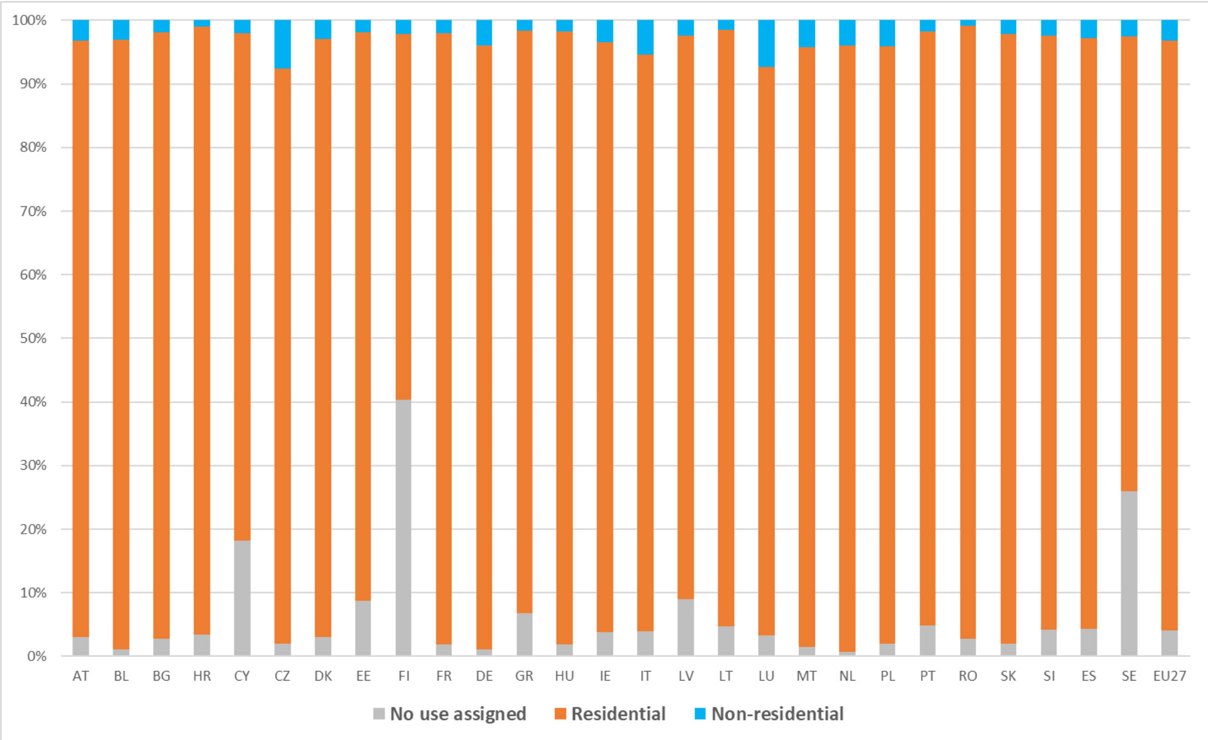
Source: JRC analysis.

Building use refers to the main function of the building divided into residential and non-residential. The building use attribute is determined by combining the GHS-BUILT-C dataset and the BUTYPE raster grid, a 100-meter resolution dataset derived from the GHSL R2023A dataset. The classification is expressed as an integer: 0 for areas outside the built-up domain, 1 for residential, and 2 for non-residential. Residential areas are those used predominantly for housing, while non-residential areas are used exclusively for other purposes, identified using Sentinel-2 satellite imagery (Florio et al., 2025).

In future versions of DBSM, we plan to assess the combination of the building function as defined in OSM when available, with the satellite-derived information on the building use.

Figure 20 shows the percentage of residential and non-residential buildings per Member State. The percentage of residential buildings overall in EU27 is about 93%, while non-residential buildings represent about 3% of all buildings (4% are unassigned). According to DBSM R2025, countries like Czechia, Luxembourg, Italy or Germany show the higher percentage of non-residential buildings according to DBSM R2025. On the other end, Netherlands, Belgium, Germany, Malta, and Czechia, show the lowest values. There are a few case, 4% overall as mentioned before, where the use of the building is unknown, because the identified polygon does not overlap a GHSL built-up area.

Figure 20 Percentage of residential and non-residential buildings per country.



Source: JRC analysis.

5.3 Validation of height, use and construction epoch in the city of Turin

This section considers the same data analysed in Section 4.2.4 with the aim to quantify the differences encountered for the building height, construction epoch and use between alternative and more up-to-date authoritative sources and DBSM R2025 for a small area.

While both use and construction epoch of the buildings in DBSM R2025 are derived from GHSL, for the height information, the delineation of height from Urban Atlas is generally preferred as demonstrated in Section 5.1. Therefore, in 49058 cases (87%) this source was selected for the area of Turin. In the remaining 13% of buildings, GHSL data was adopted to estimate heights in DBSM R2025. As the volumetric unit – defined as part of a building with homogeneous height, for which the volume can be quantified as the product of footprint area and height - represents the minimum unit of analysis, all elaborations were performed at this scale, after having spatially joined information contained in other layers. Differences have been calculated for every volumetric unit as follows:

- Height: the height difference has been calculated by subtracting the validation height to the DBSM height, with a result measured in metres

$$\Delta_h = h_{DBSM} - h_{CTC,BDTRE}\ [m]$$

- Use: as the DBSM R2025 distinguishes only between residential and non-residential buildings, information contained in both CTC and BDTRE (local authority cadastral maps) is simplified to residential and non-residential. DBSM R2025 data are validated considering that in the residential category also mixed uses are included. There are also non applicable cases when there is no validation data (CTC and BDTRE mark the use as “unknown” or “other”) or when the information is not contained in DBSM R2025, as shown in Figure 20.

- Construction epoch: DBSM R2025 classifies buildings in classes of epochs of ten years from 1980. CTC differentiates buildings built before 1918, from 1919 to 1945, from 1946 to 1960, then in ten years classes until 2000. Years after 2000 are divided as follows: until 2005, until 2012, then it is disaggregated by construction year until 2023. With the exception of the years 2010-2012, the classes or epochs correspond in both datasets: for this reason, buildings are flagged as being in *partial agreement* or *disagreement*. Buildings from authoritative sources with construction years after 2021 are marked as *disagreement*.

The resulting information is then aggregated, from volumetric unit to building level, according to the unique id attribute of DBSM R2025 and joined with the DBSM itself based on the unique id. When information on the building attributes from CTC is present, buildings in DBSM R2025 are compared with these; otherwise, it is necessary to resort to BDTRE values.

Results have been assessed based on the number of units and their footprint area, thus making it possible to understand not only the absolute values, but also the relevance that different classes have on the building stock.

5.3.1 Height

Our results showed that height differences range from an underestimation of 163.29 m to an overestimation of 38.58 m. The highest underestimation can be observed for the skyscraper of the Piedmont Region, in the Southern part of Turin (opened in late 2022), which is 191 m tall. It is possible that the error derives from the year of the last update of the input height data considered in DBSM R2025. The opposite situation occurs in a small building located in the courtyard of another newly-constructed building. The coarse resolution of the input data could also make it difficult to distinguish it from the neighbouring buildings, resulting in an overestimated height.

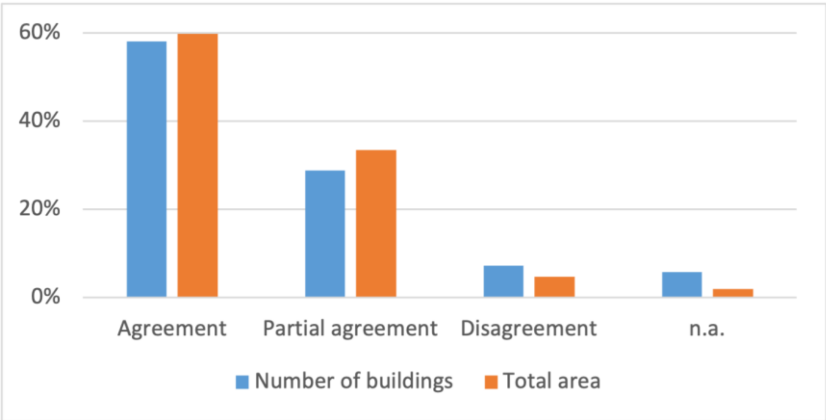
The RMSE is equal to 2.32 m, whereas the MAE is equal to 3.66 m. The error in terms of RMSE is lower than the error values obtained overall for heights extracted from UA in Section 5.2. The error in terms of MAE is similar to that obtained overall for heights coming from GHSL. This could indicate that there are no large differences in terms of height in this area, as the RMSE metric tends to penalise these higher errors more.

In order to better interpret the results, we are grouping them in the following categories. In particular:

- For discrepancies within 2.5 m from the validation value, therefore within a ±1 floor discrepancy, agreement between the DBSM R2025 and the validation dataset was acknowledged
- For discrepancies until 10 m, buildings have been flagged with partial agreement
- In case of higher discrepancies, disagreement has been acknowledged.

It is observed that 58.13% of buildings, accounting for 59.86% of the total surface, show agreement in height values. On the other hand, 28.14% (representing 33.4% of the total surface) return disagreement.

Figure 21 Percentage of buildings (number and area) whose height in DBSM R2025 is in agreement, partial agreement and disagreement with authoritative sources.



Source: JRC analysis.

As for the spatial distribution of the errors, it is possible to observe a concentration of overestimations comprised between 2.5 m and 10 m in the historic centre. All other classes are equally distributed across the city.

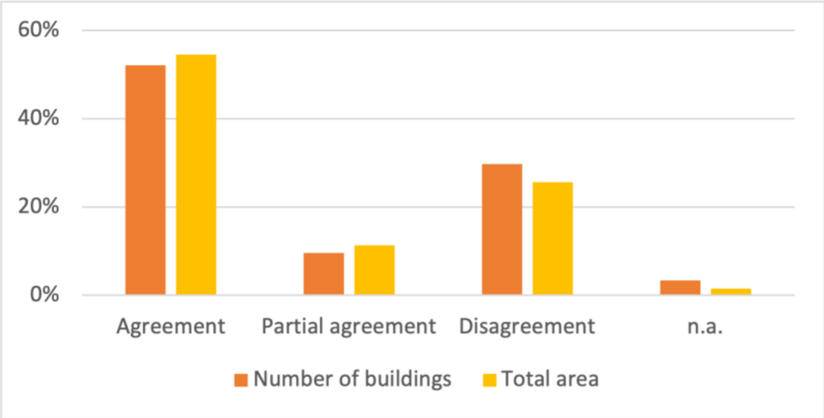
5.3.2 Use (residential/non-residential)

The general results of the validation of the attribute for main use differ from the ones about height. In this case, the 3 categories considered are defined as follows:

- Agreement: the use for all volumetric units from the authoritative sources correspond to that of the overlapping building in DBSM R2025.
- Disagreement: the use for all volumetric units from the authoritative sources does not match the use of the overlapping building in DBSM R2025.
- Partial agreement: some volumetric units in the authoritative sources match the use value of the overlapping building in DBSM R2025.
- n.a.: at least one of the two sources has missing values.

As shown in Figure 22, while the share of agreements is similar (52.19% of buildings) when compared to the height validation described above, partial agreement is significantly lower, due to a higher incidence of disagreeing values (29.83% of buildings). Partial agreement is negligible, accounting for 3.44% of the buildings.

Figure 22 Percentage of buildings (number and area) for which building use (residential / non-residential) in DBSM R2025 is in agreement, partial agreement and disagreement with authoritative sources



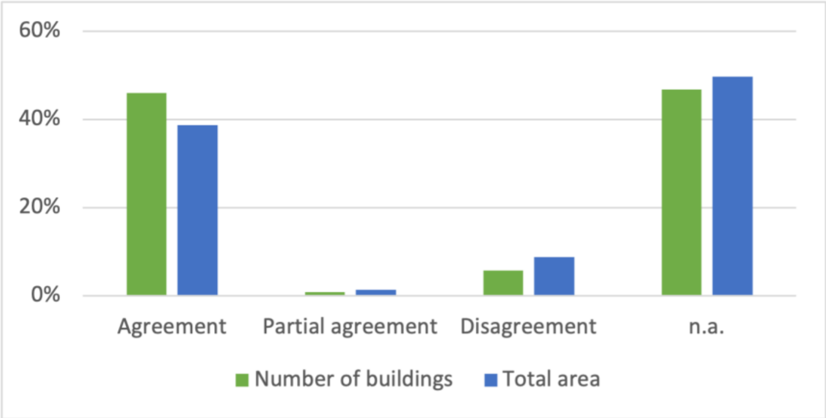
Source: JRC analysis.

As for the spatial patterns, it is possible to observe that three of the most relevant industrial complexes of the city (FIAT factory in Mirafiori, Iveco factory in the extreme North and the former ThyssenKrupp foundry) are misclassified, possibly due to their proximity to residential areas, and the heterogeneity of roof materials and shapes that influence the classification.

5.3.3 Construction epoch

Finally, the last attribute to be validated is the epoch (date) of construction. This is also the most challenging, as a comprehensive dataset on the thematic layer is unavailable and CTC information is incomplete. As expected given the datasets limitations, 47.16% of buildings, accounting for 50.42% of the total surface, have no information on the age of construction in CTC. This highlights the challenges to validate correctly this attribute of the DBSM R2025. Nevertheless, and following the same definition included in the previous section for agreement, disagreement and partial agreement; agreeing values account for 46.3% of the buildings (88% of the buildings for which data are available). Finally, a negligible share of buildings (0.78%) has been flagged as partially agreeing: as for some of the volumetric units the validation returned results agreement, in others disagreement. Figure 23 shows a summary of the resultsSource: JRC analysis.

Figure 23 Percentage of buildings (number and area) whose building construction epoch in DBSM R2025 is in agreement, partial agreement and disagreement with authoritative sources.



Source: JRC analysis.

To conclude, we can see that despite some differences, the comparisons with the building estimates for height, use and construction epoch from authoritative sources in the city of Turin show a high degree of agreement compared to the satellite-based approached followed in DBSM R2025.

6 Code repositories and data download

6.1 Data

All releases of DBSM are publicly made available in the Joint Research Centre Data Catalogue:

https://data.jrc.ec.europa.eu/collection/id-00382

DBSM R2025 is distributed under the Open Data Commons Open Database License[20] (ODbL). The various licenses of the authoritative datasets incorporated from EUBUCCO are specified in https://api.eubucco.com/v0.1/files/cc5b1b47-e1fc-4887-8a26-d8878648c3df/download.

6.2 Code

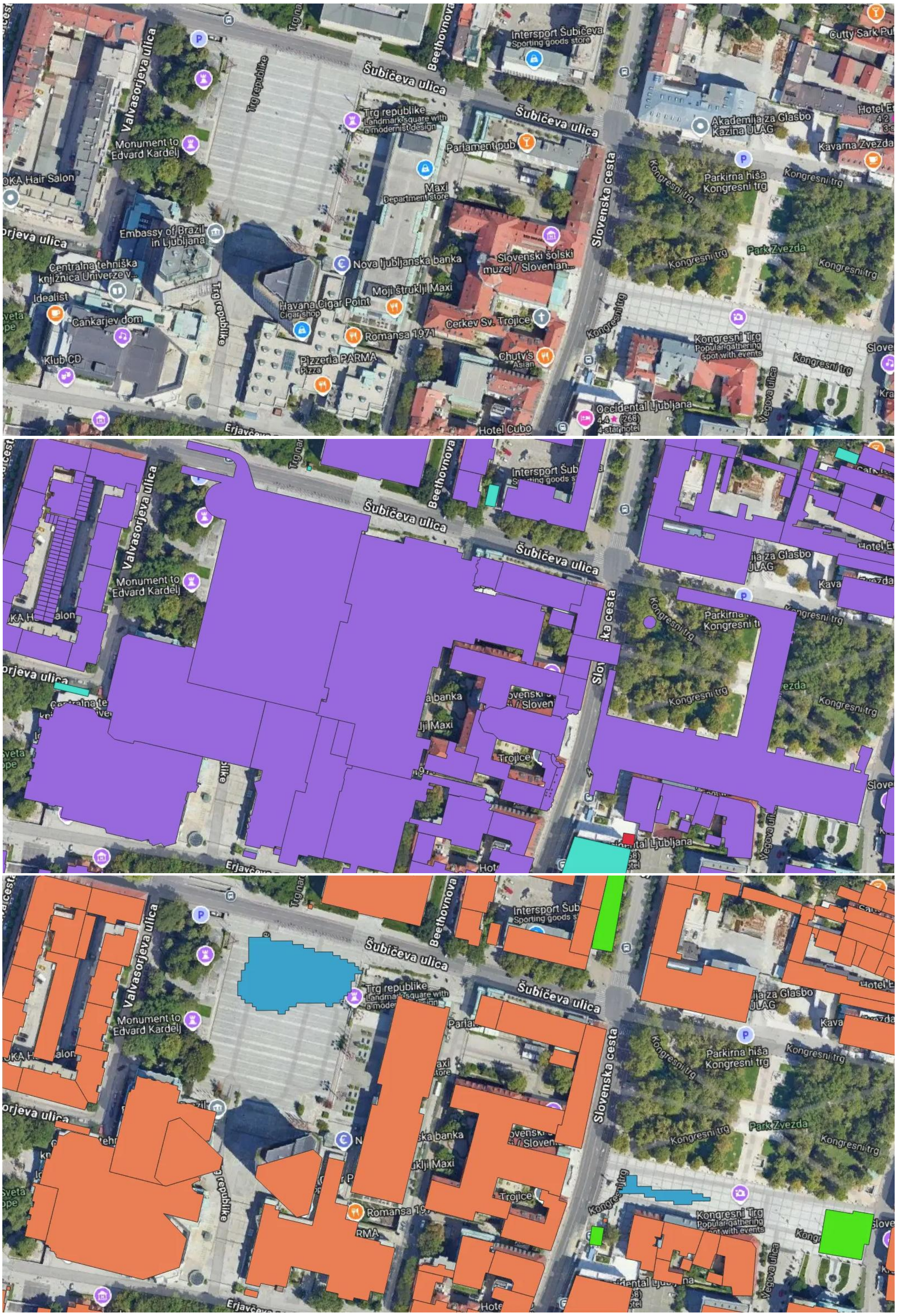
The python code corresponding to the conflation process and the attribute assignment will be made available in code.europa.eu.

[20] ODbL license: https://opendatacommons.org/licenses/odbl/, last accessed on 27 May 2025.

7 Limitations

- Authoritative sources over Open Street Map: in DBSM R2025, the decision has been made to assign higher priority to building polygons coming from authoritative sources, that is, from EUBUCCO over Open Street Map, but we acknowledge that they may not always be the most accurate, updated or reliable option. For example, Figure 24 illustrates a case in Slovenia where the buildings sourced from authoritative sources provide less comprehensive coverage compared to those selected from OSM in DBSM R2023. However, in future updates, we plan to integrate data from GISCO EUROSTAT and evaluate whether, for certain countries, OSM data surpasses the quality of authoritative sources.
- Courtyard areas and open spaces: additionally, our analysis has revealed instances where courtyard areas or other open spaces are incorrectly included within building polygons, as also evident in Figure 24 (middle figure), which depicts the Trg republike square in Ljubljana.
- Varying level of detail (footprints): while footprints typically represent individual buildings, this is not always the case; given the different sources and methods used for mapping buildings, oftentimes features include building blocks (i.e. clusters of adjacent buildings contained within a city block), or even as a whole city block in urban areas; in rare cases, an irregular cluster of buildings (or small settlement) comprising in-between streets can be represented by a single feature (i.e. as a built-up patch), typically in rural areas.
- Low temporal coherence: reference date for building mapping is unknown or it spans a relatively large window, in particular in the case of building polygons extracted from Microsoft Global Machine Learning Building Footprints, but not only.

Figure 24 Example of an area in Slovenia (46.0503383, 14.5016667). Top: Google satellite image, Middle: DBSM R2025 where EUBBUCO is selected (purple); Bottom: DBSM R2023 where OSM is selected (orange).



Source: JRC analysis.

8 Conclusions and future work

The introduction of DBSM R2025 marks a qualitative improvement in the accuracy of building footprint data available across Europe, thanks to the incorporation of authoritative data sources extracted from EUBUCCO and conflated with Open Street Map and Microsoft Global Machine Learning Building Footprints using a simple methodology developed in Python. A first validation process has been undertaken, comparing the accuracy of DBSM R2025 against existing alternatives, like GHSL-BUILT-S R2023A, Overture, and authoritative sources. The results indicate a high level of agreement in terms of footprint area with these sources.

This updated version provides a comprehensive dataset where nearly every individual building is accompanied by an initial satellite-derived evaluation of its key attributes, including height, compactness, epoch of construction, and use. While there is still scope for refining the precision of these attributes in the future, the results also indicate a reasonable margin of error in the current version, with variations for building heights ranging from one floor in most European cities (thanks to the incorporation of data from the Urban Atlas Building Height) to two floors in other areas. The validity of the satellite-derived values and estimated accuracy for use and construction epoch is reported in (Florio et al., 2025). Overall, the level of accuracy for these attributes is deemed sufficient for various applications, offering a reliable foundation for informed decision-making, although the needs of each case should always be assessed individually.

The DBSM R2025 dataset has the potential to become a valuable tool in the implementation and evaluation of numerous energy-related policies, including the renovation wave initiative, the promotion of renewable energy, and the affordable housing plan. By leveraging this robust and accurate data, policymakers and stakeholders can develop more effective strategies, make data-driven decisions, and monitor progress towards their goals, ultimately contributing to a more sustainable and energy-efficient built environment.

There are several ongoing developments related to the DBSM project. These include:

- API publication: The DBSM dataset is being made available through an Application Programming Interface (API), which will facilitate access and usage of the data for various stakeholders.
- Upcoming minor release: A minor coming release of the DBSM is planned to include an initial assessment of solar photovoltaic rooftop potential.
- Incorporation of EUROSTAT GISCO footprints: The project is evaluating the potential incorporation of building footprints from EUROSTAT's GISCO (Geographical Information System of the Commission) database. This could further enhance the accuracy and completeness of the DBSM.
- Parallel research is being conducted to improve the accuracy of the following aspects:
    - Age of construction: Refining the estimation of building ages with machine learning, potentially also to reduce the size of the building epochs.
    - Use: Identifying a more granular primary purpose of buildings (e.g., single-family houses, apartments, commercial, industrial, etc.).
    - Roof type: Assessing the types of roofs used in buildings (e.g., flat, pitched, green roofs).

- Energy demand: Estimating the energy requirements of buildings, which could be useful for energy efficiency and sustainability analyses.

The aim of the DBSM project is to actively evolve to provide more comprehensive and accurate data on buildings, which can support various applications, such as urban planning, energy management, and environmental sustainability.

References

Corbane, C., Politis, P., Kempeneers, P., Simonetti, D., Soille, P., Burger, A., Pesaresi, M., Sabo, F., Syrris, V., Kemper, T., 2020. A global cloud free pixel- based image composite from Sentinel-2 data. Data Brief 31, 105737. https://doi.org/10.1016/j.dib.2020.105737

Demuzere, M., Bechtel, B., Middel, A., Mills, G., 2019. Mapping Europe into local climate zones. PLOS ONE 14, e0214474. https://doi.org/10.1371/journal.pone.0214474

European Commission, 2022a. COM/2022/221 EU Solar Energy Strategy., eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2022:221:FIN.

European Commission, 2022b. Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use, https://eur-lex.europa.eu/eli/reg_impl/2023/13.

European Commission, 2020. COM/2020/662 final. A Renovation Wave for Europe - greening our buildings, creating jobs, improving lives, eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0662.

European Commission, 2019. The European Green Deal, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2019%3A640%3AFIN.

European Commission, Council of the European Union, 2003. REGULATION (EC) No 1059/2003 on the establishment of a common classification of territorial units for statistics (NUTS), https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32003R1059.

European Commission, European Parliament, 2025. Conference on affordable housing – addressing the European housing crisis., https://energy.ec.europa.eu/events/conference-affordable-housing-addressing-european-housing-crisis-2025-03-24_en.

European Commission. Joint Research Centre., 2023. GHSL data package 2023., https://data.europa.eu/doi/10.2760/098587. Publications Office, LU.

European Environment Agency, European Environment Agency, 2022. Urban Atlas Building Height 2012 (raster 10 m), Europe - version 3, Oct. 2022. https://doi.org/10.2909/42690E05-EDF4-43FC-8020-33E130F62023

European Parliament, 2025. Rising housing costs in the EU: the facts (infographics). Article 20241014STO24542, . Directorate General for Communication, European Parliament. https://www.europarl.europa.eu/topics/en/article/20241014STO24542/rising-housing-costs-in-the-eu-the-facts-infographics.

European Parliament, 2024. P10_TA(2024)0066 – Setting up a special committee on the Housing Crisis in the European Union, and defining its responsibilities, numerical strength and term of office, https://eur-lex.europa.eu/eli/C/2025/1982/oj.

European Parliament, 2022. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance), http://data.europa.eu/eli/reg/2022/868/oj/eng.

European Parliament, Council of the European Union, 2024. PE/102/2023/REV/1 Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the energy performance of buildings (recast), https://eur-lex.europa.eu/eli/dir/2024/1275/oj.

European Parliament, Council of the European Union, 2023. DIRECTIVE (EU) 2023/1791 on energy efficiency and amending Regulation (EU) 2023/955 (recast). https://eur-lex.europa.eu/eli/dir/2023/1791/.

Florio, P., Giovando, C., Goch, K., Pesaresi, M., Politis, P., Martinez, A., 2023. Towards a pan-EU building footprint map based on the hierarchical conflation of open datasets: the Digital Building Stock Model - DBSM. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLVIII-4/W7-2023, 47–52. https://doi.org/10.5194/isprs-archives-XLVIII-4-W7-2023-47-2023

Florio, P., Politis, P., Krasnodębska, K., Uhl, J.H., Melchiorri, M., Martinez, A., Kakoulaki, G., Pesaresi, M., Kemper, T., 2025. GHS-OBAT: Global, Open Building Attributes data at footprint level including age, function, height and compactness. Data in Brief.

Institut Catalá d'Energia, Energy and Resources Institute, Asia Urbs Programme, 2004. Sustainable Building Design Manual: Sustainable building design practices. Energy and Resources Institute.

Japan Aerospace Exploration Agency, 2016. ALOS World 3D 30 meter DEM. V3.2, Jan 2021. Distributed by OpenTopography. https://doi.org/10.5069/G94M92HB

Milojevic-Dupont, N., Wagner, F., Nachtigall, F., Hu, J., Brüser, G.B., Zumwald, M., Biljecki, F., Heeren, N., Kaack, L.H., Pichler, P.-P., Creutzig, F., 2023. EUBUCCO v0.1: European building stock characteristics in a common and open database for 200+ million individual buildings. Sci. Data 10, 147. https://doi.org/10.1038/s41597-023-02040-2

Minghini, M., Thabit Gonzalez, S., Gabrielli, L., 2024. Pan-European open building footprints: analysis and comparison in selected countries. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLVIII-4/W12-2024, 97–103. https://doi.org/10.5194/isprs-archives-XLVIII-4-W12-2024-97-2024

NASA Shuttle Radar Topography Mission, 2013. Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF

Overture Maps Foundation, 2023. Free and open building data normalized to one schema, overturemaps.org.

Palacios-Lopez, D., Esch, T., MacManus, K., Marconcini, M., Sorichetta, A., Yetman, G., Zeidler, J., Dech, S., Tatem, A.J., Reinartz, P., 2022. Towards an Improved Large-Scale Gridded Population Dataset: A Pan-European Study on the Integration of 3D Settlement Data into Population Modelling. Remote Sens. 14, 325. https://doi.org/10.3390/rs14020325

Pesaresi, Martino, Politis, P., 2023. GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030). https://doi.org/10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA.

Pesaresi, M., Politis, P., 2023a. GHS-BUILT-H R2023A - GHS building height, derived from AW3D30, SRTM30, and Sentinel2 composite (2018) European Commission, Joint Research Centre (JRC) [Dataset]. https://doi.org/10.2905/85005901-3A49-48DD-9D19-6261354F56FE PID:

Pesaresi, M., Politis, P., 2023b. GHS-BUILT-C R2023A - GHS Settlement Characteristics, derived from Sentinel2 composite (2018) and other GHS R2023A data. European Commission, Joint Research Centre (JRC) [Dataset]. https://doi.org/10.2905/3C60DDF6-0586-4190-854B-F6AA0EDC2A30

Pesaresi, M., Politis, P., 2022. GHS-LAND R2022A - Land fraction as derived from Sentinel2 image composite (2018) and OSM data. . PID: http://data.europa.eu/89h/ab7ad451-5ed5-44a6-a4d0-9f7a4e848ceeEuropean Commission, Joint Research Centre (JRC). https://doi.org/10.2905/AB7AD451-5ED5-44A6-A4D0-9F7A4E848CEE

Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J.H., Carioli, A., Corbane, C., Dijkstra, L., Florio, P., Friedrich, H.K., Gao, J., Leyk, S., Lu, L., Maffenini, L., Mari-Rivero, I., Melchiorri, M., Syrris, V., Van Den Hoek, J., Kemper, T., 2024. Advances on the Global Human

Settlement Layer by joint assessment of Earth Observation and population survey data. Int. J. Digit. Earth 17, 2390454. https://doi.org/10.1080/17538947.2024.2390454

Uhl, J.H., Politis, P., Pesaresi, M., 2025. GHS-AGE R2025A – Global gridded estimates of the dominant age of the built stock (1975-2020). . European Commission, Joint Research Centre (JRC) [Dataset]. https://doi.org/10.2905/d503bb56-9884-4e4d-bb8f-d86711d9f749

List of abbreviations and definitions

Abbreviations	Definitions
DBSM	Digital Building Stock Model
EU	European Union
EUB	EUBUCCO
GHSL	Global Human Settlement Layer
GHS-AGE	Global Human Settlement Age, data released under GHSL (GHSL R2024A)
GHS-BUILT-S	Global Human Settlement Built-up Surface, data released under GHSL (GHSL R2023A)
GHS-OBAT	Global Human Settlement – Open Buildings Attribute Table
GISCO	Geographic Information System of the Commission
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MedAE	Median Absolute Error
MSB	Microsoft Global Machine Learning Building Footprints
NDI	Normalised Difference Index
NUTS	Nomenclature of Territorial Units for Statistics
OSM	Open Street Map
r	Pearson's Correlation Coefficient
RMSE	Root Mean Square Error

List of figures

List of tables

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
Joint-research-centre.ec.europa.eu

Publications Office of the European Union