GFZ
Helmholtz Centre
POTSDAM

HELMHOLTZ CENTRE POTSDAM
GFZ GERMAN RESEARCH CENTRE
FOR GEOSCIENCES

**Cecilia I. Nievas, Marius Kriegerowski,
Felix Delattre, Nicolás García-Ospina,
Karsten Prehn, Fabrice Cotton**

The European High-Resolution Exposure (EHRE) Model

Scientific Technical Report STR23/05

www.gfz-potsdam.de

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Recommended citation of the publication

Nievas, C.I., Kriegerowski, M., Delattre, F., Garcia Ospina, N., Prehn, K., Cotton, F. (2023): The European High-Resolution Exposure (EHRE) Model, Scientific Technical Report STR; 23/05, Potsdam: GFZ German Research Centre for Geosciences. https://doi.org/10.48440/gfz.b103-23055

Imprint

Helmholtz Centre Potsdam
GFZ German Research Centre for Geosciences
Telegrafenberg
D-14473 Potsdam

Published in Potsdam, Germany
2023

DOI: https://doi.org/10.48440/gfz.b103-23055
URN: urn:nbn:de:kobv:b103-23055

This work is published in the GFZ series Scientific Technical Report STR and electronically available via our repository GFZpublic: https://tinyurl.com/GFZ-ST-Reports

HELMHOLTZ CENTRE POTSDAM
GFZ GERMAN RESEARCH CENTRE
FOR GEOSCIENCES

The European High-Resolution Exposure (EHRE) Model

Cecilia I. Nievas, Marius Kriegerowski, Felix Delattre,
Nicolás García-Ospina, Karsten Prehn, Fabrice Cotton

October 2023

Author's ORCIDs:

Cecilia I. Nievas: https://orcid.org/0000-0001-9260-9590

Marius Kriegerowski: https://orcid.org/0000-0001-5673-8432

Felix Delattre: https://orcid.org/0009-0003-6267-983X

Nicolás García-Ospina: https://orcid.org/0009-0000-0069-4485

Karsten Prehn: https://orcid.org/0000-0001-9745-2350

Fabrice Cotton: https://orcid.org/0000-0002-9242-3996

Contents

Abstract

This report presents the European High-Resolution Exposure (EHRE) model, which we created by combining the state-of-the art exposure model of the European Seismic Risk Model 2020 (ESRM20) with building footprints and data from OpenStreetMap previously processed by OpenBuildingMap. We present the method used to generate the model and the software we developed for this purpose, as well as a number of outputs that give an overview of what the model consists of and enable discussion.

1. Introduction

The general lack of systematic approaches for the collection of building data relevant to characterise the seismic fragility of structures in large regions (such as countries or continents) has led, in the past two decades, to the development of different strategies and methods for the creation of building exposure models based on sources of information with different degrees of completeness and detail. Some of the most common sources are national census data, official statistics from international bodies such as Eurostat, expert opinion gathered through workshops and other participatory processes, remote sensing techniques by themselves or combined with machine learning algorithms, datasets created by large research institutions such as the European Commission's Joint Research Centre (JRC) or the US National Aeronautics and Space Administration (NASA), and volunteered geographic information systems (e.g., Crowley et al., 2020; Figueiredo and Martina, 2016; Geiß et al., 2015, 2017; Liuzzi et al., 2019; Paprotny et al., 2020; Pelizari et al., 2021; Sousa et al., 2017; Wieland et al., 2015; Yepes-Estrada et al., 2017).

At the European level, the state-of-the art exposure model for the European Seismic Risk Model 2020 (Crowley et al., 2021) has been recently published, building upon previous efforts and concluding with a transparent, publicly-available model that covers 44 countries, as described in Crowley et al. (2020). In the ESRM20 model, the exposure is defined at different resolutions for different occupancy cases and countries, the determining factor being the resolution of the input data available to define the model. Residential and commercial exposure use administrative units while industrial exposure uses administrative units in some countries and 30-arcsec cells in others. A higher-resolution version of ESRM20 has also been developed by distributing the administrative-level exposure onto 30-arcsec cells proportionally to population counts from the WorldPop dataset[1].

Volunteered geographic information (VGI) systems have become increasingly relevant in recent years, with OpenStretMap (OSM) being one of the most prominent initiatives both in terms of coverage and growth rate. OSM has not been created with the purpose of developing building exposure models, but the extent to which the volunteer-based mapping community is committed to mapping as many features as possible all around the globe makes OSM one of the largest existing geographic databases and a very relevant source of information for risk modellers. In OSM, features such as buildings are mapped as geo-located geometric objects to which so-called *tags* are assigned in order to indicate properties, such as the number of storeys or intended use. Such data has already been successfully used in the development of exposure models (e.g., Geiß et al., 2017), and even within ESRM20 itself, as it incorporates information from the industrial model of Sousa et al. (2017). Recognising the potential of OSM to become the most complete worldwide building database, OpenBuildingMap (OBM; Schorlemmer et al., 2020, 2023) harvests OSM to retrieve building footprints and process them, calculating or inferring additional parameters of relevance along the way.

In the present work, we combine the ESRM20 exposure model with building footprints and data from OpenStreetMap (via OpenBuildingMap) to create the European High-Resolution Exposure (EHRE) model, taking advantage of the state-of-the-art knowledge on structural properties, occupancy and replacement cost of buildings stemming from ESRM20, as well as of data on the location, number of storeys and use of individual buildings mapped in OBM. The method followed has been designed as a general large-scale approach that works with all 44 countries covered by ESRM20 and is serving as a preliminary step towards the development of a worldwide Global Dynamic Exposure (GDE) model (Schorlemmer et al., 2020, 2023) that automatically retrieves and incorporates any changes in OSM.

---

[1] https://www.worldpop.org/

The work presented herein is the outcome of this continental-level approach. The resulting exposure model contains a large volume of data, whose full dimension is challenging to explore. As it will become apparent in the upcoming chapters, a full investigation of the impact of modelling choices as well as data quality in different countries is a step still to come, but some key aspects are presented herein with the purpose of enabling such analyses and guiding future efforts on refinements and improvements.

Chapter 2 of this report starts with an overview of the method used to create the EHRE model, and rapidly moves on to the details on the most important steps. It concludes with summary statistics and associated discussions to be picked up in the conclusions and outlook for the future of Chapter 3. Systematically organised details on software and data accessibility can be found in Chapter 4.

2. The model

2.1 Overview

The European High-Resolution Exposure (EHRE) model is a high-resolution exposure model that combines:

- the aggregated exposure model of the European Seismic Risk Model 2020 (ESRM20; Crowley et al., 2020, 2021);
- data on individual buildings from OpenStreetMap (OSM), processed as OpenBuildingMap (Schorlemmer et al., 2020, 2023);
- remote sensing-derived built-up areas from the Global Human Settlement Layer (GHSL; Corbane et al., 2018), processed with a series of dedicated software (Oostwegel et al., 2023).

Aggregated exposure models usually define the number of buildings in a geographic area according to defined building classes that are relevant to characterize the behaviour of the buildings when subject to specific natural hazards. It is not uncommon for the building classes to be separated into occupancy categories, such as residential, commercial and industrial. Moreover, they also tend to provide information on the number of occupants and the replacement cost of each building, as these allow risk modellers to carry out economic and human loss assessments. As a very simplified fictitious example, an aggregated exposure model may say that the administrative district of Sample City has 300,000 residential buildings, 200,000 of which are of class A and 100,000 of which are of class B, and that each building of class A usually has an average of 15 occupants and costs 5 million EUR to rebuild, while each building of class B has an average of 4 occupants and costs 300,000 EUR to rebuild. It is *aggregated* because it does not specify *exactly* where these buildings are located *within* Sample City.

As shown in Figure 1, the process of generating the EHRE model starts from three fronts. On the one hand, aggregated exposure models are distributed onto zoom-level 18 quadtiles (defined in EPSG:3857)[2] proportionally to the built-up areas expected in each tile. The building classes are taken directly from the aggregated exposure model as a function of location (all Sample City tiles get assigned Sample City classes). At the same time, polygons from OSM (so-called *ways*) that are tagged as buildings are retrieved and processed to create OpenBuildingMap (OBM; Schorlemmer et al., 2020, 2023). The four most important pieces of information associated with an OBM building are: its unique OSM identifier (OSM ID), its footprint (geometry, location, orientation), its number of storeys and its occupancy. Knowing the building's location

---

[2] These quadtiles are obtained by subdividing the world (longitude in range [-180.0, +180.0], latitude in range [-85.0511, +85.0511]) in EPSG:3857 projection into four tiles of level 1, subsequently dividing these into four tiles each (level 2) and so on, up to the desired zoom level (18 in our case).

and occupancy allows us to assign building classes (as defined in the aggregated exposure model) to the OBM buildings. Two attempts are made to narrow down the building classes assigned to a building, one based on its number of storeys and the other based on occupancy details of commercial buildings.

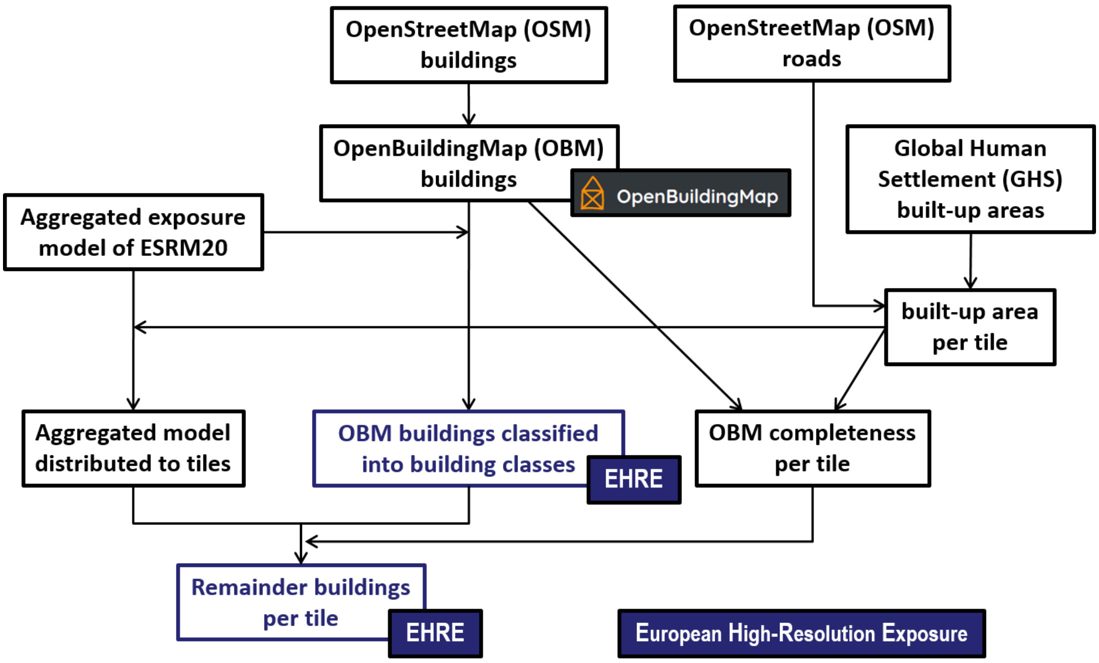


Figure 1 General overview of the components of the European High-Resolution Exposure model.

One of the challenges associated with the use of building data from OpenStreetMap is the fact that OSM is not complete everywhere, that is, not all existing buildings are represented in OSM. The third front thus focuses on estimating which tiles are complete and which are incomplete. This is done, as described by Oostwegel et al. (2023), by comparing the remote sensing-derived estimates of built-up areas from the Global Human Settlement Layer (GHSL; Corbane et al., 2018), adjusted to take into account the width of roads, against the built-up area resulting from the summation of OSM building footprints. If the ratio between the two built-up area values for a tile is above a selected threshold, the tile is considered complete; otherwise, it is considered incomplete.

This completeness classification is the key ingredient to move on to the final stage, which consists in combining the number of buildings stemming from the aggregated exposure model and OBM. If a tile is complete, all its exposure is then defined by the OBM buildings located in the tile. If a tile is incomplete then the number of buildings expected to exist in the tile from the distribution of the aggregated exposure model is compared against the number of OBM buildings: if the latter is larger, then all the exposure is defined by the OBM buildings but, if the latter is smaller, the number of so-called *remainder* buildings is calculated as the difference between the two. These *remainder* buildings represent the buildings expected to exist in reality but not yet represented in OpenStreetMap and are assigned the centroid (in EPSG:4326) of the tile as their location (i.e., they are “lumped” at the centre of the tile). This comparison and calculation is carried out separately for each occupancy case. The EHRE model is thus a combination of individual OBM buildings and remainder buildings on tiles, together with their associated numbers of occupants and replacement costs.

Taking the European Seismic Risk Model 2020 (ESRM20; Crowley et al., 2020, 2021) as a starting point, the EHRE model covers the same three occupancy cases (residential, commercial and industrial) and 44 European countries as ESRM20: Albania, Andorra, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Gibraltar, Greece, Hungary, Ireland, Iceland, Isle of Man, Italy, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Moldova, Monaco, Montenegro, Netherlands, Norway, North Macedonia, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Türkiye, and the United Kingdom. In the ESRM20 model, the exposure is defined at different resolutions for different occupancy cases and countries. Residential and commercial exposure use administrative units while industrial exposure uses administrative units in some countries and 30-arcsec cells in others. Building classes are defined in terms of the GEM Building Taxonomy v3.0 (Silva et al., 2022), also known as GED4ALL.

A version of the ESRM20 model distributed onto 30-arcsec cells for all occupancy cases has become available in recent months, after the coding and running of the EHRE software was finalised. This distribution onto 30-arcsec cells is proportional to population counts from the WorldPop dataset[3], which is conceptually similar to the distribution of the admin-level ESRM20 into zoom-level 18 tiles using the Global Human Settlement Layer (GHSL; Corbane et al., 2018) done for EHRE. For this reason, we do not expect fundamental differences to arise from taking one or the other resolution level of ESRM20 as the starting point for EHRE, particularly due to the use of the concept of *data units* (instead of administrative units) and *filler data units*, as explained in Section *2.4 Processing the ESRM20 aggregated exposure model*.

2.2 OpenBuildingMap

The processing of OSM buildings within OpenBuildingMap (OBM; Schorlemmer et al., 2020, 2023) is carried out using rabotnik-obm[4], which, in turn, uses rabotnik[5]. In order to run, rabotnik-obm requires a list of OSM building IDs. For the version of OBM used for EHRE, the lists of OSM IDs to process was defined by means of polygons drawn to enclose European countries or regions of interest, which were intersected with the geometries of OSM buildings in a local replication of the OSM database. This local replication is constantly updated by running the well-known OSM importer Imposom[6] with the configuration file whose contents can be found in Appendix *8.4 Configuration file for Imposom (replication of the OSM database)*. Once the list of OSM IDs has been defined, the IDs are passed to rabotnik-obm, which distributes the work across all available workers in the server. This manual definition of the polygons to process means that the OSM buildings included in EHRE are a static version of OSM at the time of processing the polygon, as opposed to the automatic updating of the Global Dynamic Exposure (GDE) model (Schorlemmer et al., 2020, 2023).

Buildings are identified in OSM by means of the tag “building:yes”. Tags are the technical way in which OSM contributors assign attributes to features of the map and consist of a *key* (“building”) and a *value* (“yes”). OBM retrieves from OSM all polygons (OSM *ways*) tagged with the keys “building” or “building:part”, or with the key “aeroway” and keys “terminal” or “hangar”. Polygons that comply with these criteria but are also tagged with (i) the key “building” and values "no", or "none", or "No", or "bridge", or "pier", or "road", (ii) the key “building:part” and values "no", or "none", or "No", and/or (iii) the key “man_made” with values "bridge", or "pier", are skipped.

---

[3] https://www.worldpop.org/

[4] https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/rabotnik-obm

[5] https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/rabotnik

[6] https://imposm.org/

When processing building polygons from OSM, rabotnik-obm retrieves/determines the following attributes:

- OSM ID;
- geometry of the footprint;
- number of storeys (above ground), if available;
- floor space (i.e. footprint area times the number of storeys), if the number of storeys is available;
- relation ID (if the building polygon is part of an OSM relation);
- quadkey of the zoom-level 18 quadtile that contains the centroid of the building footprint;
- occupancy type.

In the version of OBM used for EHRE, buildings that are represented in OSM through a series of polygons linked together by means of a so-called *relation* are treated as multiple separate buildings. These building “parts” may have different numbers of storeys and may be assigned different occupancy types. However, we later bring these different parts together within the EHRE code when assigning building classes to the OBM buildings, by means of the relation ID.

The processing of buildings by rabotnik-obm in order to obtain/define the aforementioned parameters is controlled by rules that are run sequentially for any individual OSM ID:

1. *GetBuilding*: This rule retrieves and stores the building footprint geometry. If the building is part of a relation, it also retrieves and stores the associated relation ID. If the OSM ID already exists in the database, it updates its fields; if the OSM ID does not yet exist in the database, it creates a new entry.
2. *GetQuadkey*: This rule retrieves and stores the quadkey at zoom level 18 of the tile that contains the centroid of the building’s footprint.
3. *GetStoreys*: This rule retrieves and stores the number of storeys, if the information is available, and calculates the estimated floor space. When the number of storeys indicated in OSM (by means of the tag “building:levels”) is not an integer, rabotnik-obm rounds the number up to the closest integer.
4. *GetGEMBuildingOccupancy*: This rule retrieves all tags associated with a building polygon, maps them to occupancy types as per the GEM Building Taxonomy v3.0, and decides on a final occupancy type through a series of rules, whose order of verification is relevant to the final outcome, as the verification of occupancy rules ends when one rule is able to return a value.

In OBM, occupancy is classified as per the GEM Building Taxonomy v3.0 (Silva et al., 2022), in which occupancy strings consist of three letters, which point at the general class, and can be followed by a number and/or a combination of numbers and letters, which point at a more detailed class. For example, RES1 refers to a single dwelling residential building (e.g. a detached house) while RES2 refers to a multi-unit residential building (e.g. a semi-detached house, an apartment block), and RES alone points at a residential building of unknown subtype. A list of all occupancy strings as per the GEM Building Taxonomy can be found in Appendix *8.1 Occupancy types of the GEM Building Taxonomy v3.0*. Explanations regarding the meaning of the occupancy strings can be found in the online glossary of the OpenQuake platform[7].

The steps to assign occupancy types are the following:

1. Tags associated with the building are retrieved from OSM.
2. OSM tags are mapped into occupancy strings as per the GEM Building Taxonomy.
3. All resulting occupancy strings associated with a building are evaluated as a group according to rules that decide on the final occupancy to be assigned to the building.

---

[7] https://taxonomy.openquake.org/

As mentioned earlier, OSM contributors assign attributes to features of the map by means of so-called *tags*. There are no limits to the number or content of the tags that can be assigned to any feature, but the OSM community regularly agrees on best practice so as to encourage a degree of homogeneity.

For any particular building, the process to assign occupancy types starts by retrieving three kinds of tags:

- Building tags: OSM tags attached to the building polygon.
- Points of Interest (PoI) tags: OSM tags associated with points of interest (usually defined using *nodes*) that are contained within the building polygon. The attributes of the points of interest are often used to indicate the presence of a business, office, etc.
- Land use tags: OSM tags associated with polygons tagged as land use polygons (and not buildings) that are intersected by the building polygon.

The number of tags that can be retrieved for any individual building is unlimited and all tags associated with a building (as defined above) are considered for the final classification. It does happen that a building may only have one associated tag that is “building:yes” and that no further tags are available to classify its occupancy, in which case the occupancy remains unknown (labelled as UNK).

Once tags have been retrieved, they are mapped into GEM Building Taxonomy occupancy strings by means of a mapping scheme, which is based on the description of common OSM tags[8] and can be found in Section *8.2 Mapping of OSM tags into occupancy types*. The output of this process is one occupancy string per tag associated with the building (from any of the three kinds of tags).

The final step runs a series of rules (in sequence) to assign a unique occupancy type to each building. The first rule is called *RuleOverridingOccupancy* and consists of searching for each occupancy string of a pre-selected hierarchical subset within the list of occupancy strings associated with the building and, if found, adopting it as the final occupancy. The pre-selected subset is called the set of "overriding occupancies" and can be found in Section *8.3 Overriding occupancies*. The order in which the occupancy strings are named in the subset is hierarchical: the first string takes precedence over all other strings, the second takes precedence over all other strings except for the first one, and so on. An airport building is a simple example to demonstrate the logic behind this rule, as it can be associated with tags of all kinds, such as restaurants, cafes, retail shops, offices, police, customs, etc., but the building is still, above all, an airport. As a consequence, if COM10 (=airport) is present in the list of occupancy strings associated with a building, the final occupancy of the building should be COM10 (=airport).

If the first rule (*RuleOverridingOccupancy*) returns no final occupancy, the rabotnik-obm software moves on to the second rule, and so on. Many rules from the second one onward use additional tools that interpret the different levels of detail that a certain occupancy string has, determine whether there are repeated strings in the list of all strings associated with a building, and count numbers of repetitions. "Level of detail" refers, for example, to the difference between having the string RES, which is a residential building of unknown type, or RES2, which is a residential multi-dwelling building, or RES2A, which is a 2-unit residential building (duplex): RES2A (detail level 2) is a sub-type of RES2 (detail level 1), which is in turn a subtype of RES (detail level 0).

The second and subsequent rules are:

- *RuleOneUniqueTag*: This rule addresses the trivial case in which there is only one unique occupancy string associated with the building and results in this unique occupancy string being adopted. The unique string may be repeated. For example, two COM1 strings lead to COM1.

---

[8] https://taginfo.openstreetmap.org/

- *RuleTwoTagsSameLevel0*: If there are two unique occupancy strings (irrespective of whether they are repeated or not), and one is a sub-type of the other, the more detailed sub-type is adopted. For example, if the strings are RES, RES1 and RES, the final occupancy is RES1. One exception to this is the case in which the unique strings are COM (commercial and/or public, unknown subtype) and COM7 (covered parking garage), in which case the final occupancy is COM (this could be a commercial building that should not be reduced to *just* a covered parking garage).
- *ResidentialWithGarage*: If the individual occupancy strings are a combination of COM7 (covered parking garage) and residential occupancies of the kind RES1 (residential, single dwelling) and/or RES2 (residential, multi-unit, unknown type) and/or RES4 (institutional housing), the final occupancy is RES, as it is interpreted to represent a residential building with a garage.
- *RuleShoppingMall*: This rule is intended to catch potential combinations of occupancy strings that are likely to indicate the presence of a shopping mall or arcade. Two sub-cases lead to the final occupancy being COM1 (i.e., retail trade):
    1. the individual occupancy strings are COM1 (retail trade) and/or COM5 (restaurants, bars, cafes) irrespective of how many times they appear, or
    2. there are at least two cases of COM1 and all other strings are COM and/or COM2 (wholesale trade and storage) and/or COM3 (professional/technical offices) and/or COM5 (restaurants, bars, cafes) and/or COM7 (covered parking garage) and/or COM11 (recreation and leisure) and/or ASS3 (cinema or concert hall).
- *RuleManyTagsSameLevel0*: If there are more than two unique occupancy strings (irrespective of whether they are repeated or not), and they are all subtypes of the same main type (of detail level 0), the main type is returned as the final occupancy. For example, if the unique strings are IND (industrial, unknown subtype), IND1 (heavy industrial), IND2 (light industrial), the final occupancy is IND (no decision is made between “heavy” and “light”).
- *RuleResidentialAndCommercial*: If the individual occupancy strings are a combination of residential occupancies of the kind RES1 (residential, single dwelling) and/or RES2 (residential, multi-unit, unknown type) and/or RES4 (institutional housing), and commercial occupancies of the kind COM1 (retail trade) and/or COM3 (professional/technical offices) and/or COM5 (restaurants, bars, cafes), the final occupancy is MIX1 (i.e., a mix of mostly residential and commercial).
- *RuleCommercialAndIndustrial*: If the individual occupancy strings are a combination of any industrial (IND) sub-type (IND, IND1, IND2) with commercial occupancies of the kind COM (commercial and/or public, unknown subtype) and/or COM1 (retail trade) and/or COM2 (wholesale trade and storage) and/or COM3 (professional/technical offices) and/or COM5 (restaurants, bars, cafes) and/or COM7 (covered parking garage) and/or COM11 (recreation and leisure), the final occupancy is MIX5 (i.e., a mix of mostly industrial and commercial).
- *RuleResidentialAndIndustrial*: If the individual occupancy strings are a combination of residential occupancies of the kind RES1 (residential, single dwelling) and/or RES2 (residential, multi-unit, unknown type) and/or RES4 (institutional housing) with any industrial (IND) sub-type (IND, IND1, IND2), the final occupancy is MIX4 (i.e., a mix of mostly residential and industrial).
- *RuleUnknown*: If none of the previous rules apply, the final occupancy is UNK (unknown).

Figure 2 shows a schematic representation of the algorithm used by rabotnik-obm to assign occupancy types to each building, together with an application example.

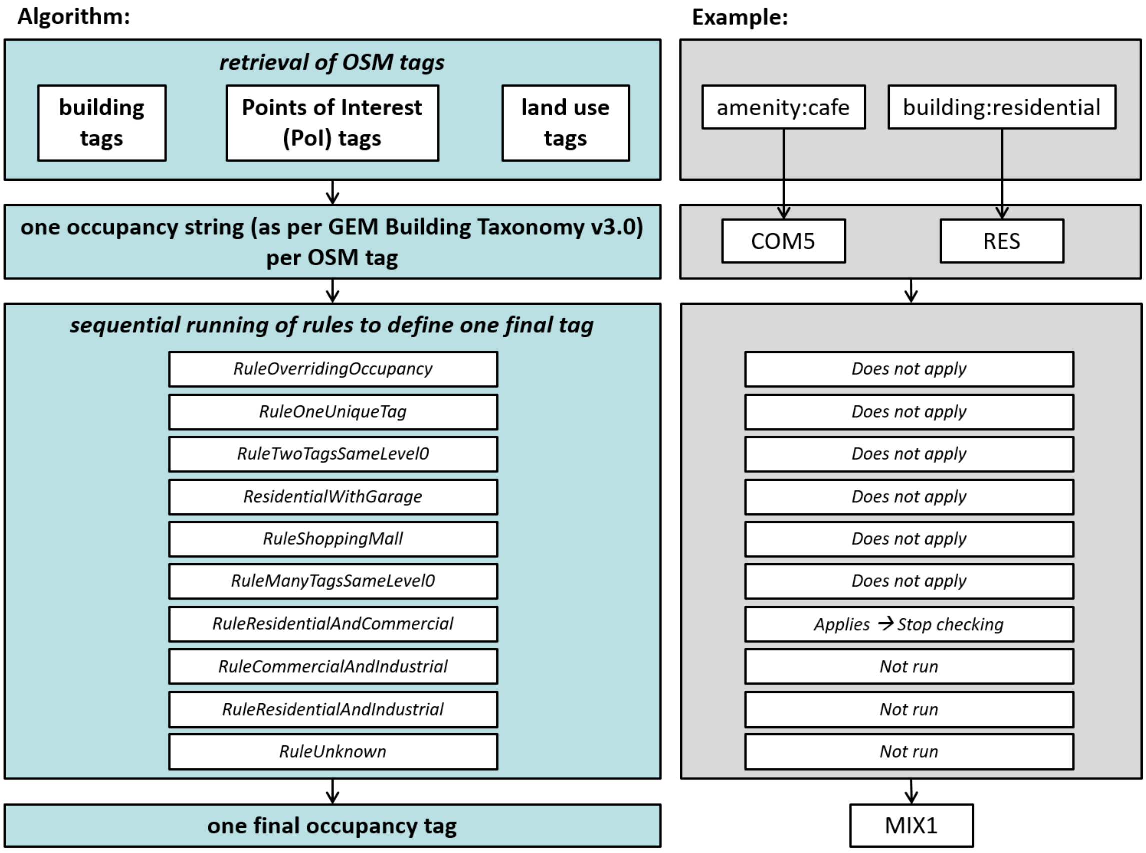


Figure 2 Schematic representation of the OBM algorithm to assign one occupancy tag per building (left) and example (right).

The OpenBuildingMap database used to create the EHRE model was populated with OSM data retrieved between June and September 2022 and is not automatically updated like the main OpenBuildingMap database[9]. It consists of 192,358,699 OSM buildings, most of which correspond to the 44 European countries covered by the ESRM20 model, but a few of which likely lie in neighbouring countries such as Belarus or Georgia (due to using hand-drawn polygons to define areas to query OSM). Of these 192,358,699 OSM buildings, 13,924,577 (7.24%) have an assigned number of storeys, 122,110 (0.06%) have a relation ID, and 159,554,743 (82.95%) have an assigned occupancy type other than “UNK” (unknown).

2.3 Built-Up areas and completeness

Being able to estimate the completeness of OSM, that is, whether all buildings that exist in reality are mapped in OSM, is fundamental to be able to incorporate OSM building data into any sort of analysis (Oostwegel et al., 2023). Within EHRE, we use information on completeness to decide whether the final number of buildings in a tile comes fully from OSM or from a combination of OSM and ESRM20, while estimates of built-up areas per tile are used to distribute ESRM20 onto the tiles. These two pieces of information are determined at the tile level by the obmgapanalysis[10] software, using the multi-temporal information layer of built-up surfaces of the Global Human Settlement Layer (GHSL, Corbane et al., 2018), which is derived from global Landsat satellite data collected from 1975 to 2014 and has a spatial resolution of 30 meters (GHSL code

---

[9] https://www.openbuildingmap.org

[10] https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/obmgapanalysis

GHS_BUILT_LDSMT_GLOBE_R2018A), as well as information on roads from OpenStreetMap. As a 30-m resolution is not enough to resolve streets and other roads, an 8-m width is assumed for each *way* used to represent an OSM road (i.e., a 4-m buffer to each side) and this is subtracted from the GHSL built-up pixels before calculating the resulting built-up area. It is noted that many wide avenues/roads are represented in OSM with different lines in the two opposite traffic directions, which means that the resulting width is larger than 8 meters under this assumption. The obmgapanalysis software carries out these tasks by processing lists of keys of the zoom-level 18 tiles associated with a selected country (represented by means of the GADM 3.6[11] country boundaries), which are generated in advance by the auxiliary code quadtreegrid[12].

In order to have an estimate of the degree of completeness of OSM in the tile, the summation of the area of the footprints of OSM buildings and the ratio of OSM built-up area to GHSL built-up area are calculated as well. A tile is assumed to be complete if the ratio is equal to or larger than 0.25, and it is assumed incomplete otherwise, as shown by Equation (1).

$$r = \frac{A_{OSM}}{A_{GHSL-roads}} \begin{cases} \geq 0.25 \Rightarrow \text{complete} \\ < 0.25 \Rightarrow \text{incomplete} \end{cases} \tag{1}$$

The calculation of the built-up ratio (of OSM to GHSL built-up areas) as well as the automatic completeness assessment (ratio ≥ 0.25 considered as complete) are implemented as functions in the PSQL database to which obmgapanalysis writes its outputs. These functions are manually run after obmgapanalysis has been ran for both the GHSL and the OSM built-up areas (for a geographic area of interest).

This threshold of 0.25 was selected by means of a study focused on the region of Attica, Greece, whose completeness was assessed manually by visually comparing OSM footprints against Bing Aerial imagery. As shown in Figure 3, manually-assessed tiles were classified into three groups, namely (i) complete tiles, (ii) incomplete + almost complete tiles, and (iii) unknown (due, for example, to clouds not allowing to see the buildings in the aerial imagery), and the relative cumulative rates of exceedance of the different values of built-up ratio were plotted. For example, Figure 3 tells us the trivial observation that 100% of tiles manually classified as complete had a built-up ratio equal to or larger than zero, and the more interesting fact that around 60% of the tiles in this group had a built-up ratio of ~0.3 or larger. Selecting a threshold of 0.25 (as marked vertically in the figure) implies that 65% of the complete tiles of Attica get correctly classified as complete, while 35% of them get incorrectly classified as incomplete, and that around 85% of the incomplete (or almost complete) tiles of Attica get correctly classified as incomplete, while around 15% of them get incorrectly classified as complete.

As the completeness status of a tile is used to decide whether remainder buildings are calculated or not, it is preferable to misclassify a complete tile as incomplete, than to do the opposite. When a tile is misclassified as incomplete we move on to the step of calculating the difference between the number of buildings coming from the aggregated model and the number of OBM buildings. If the latter is larger than the former, then no remainder buildings will be added. However, when an incomplete tile is misclassified as complete, we eliminate the opportunity of making this comparison and take the number of OBM buildings as the final number of buildings in the tile. The number of OBM buildings is taken into account in any of the two cases.

---

[11] https://gadm.org/download_country36.html

[12] https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/quadtreegrid

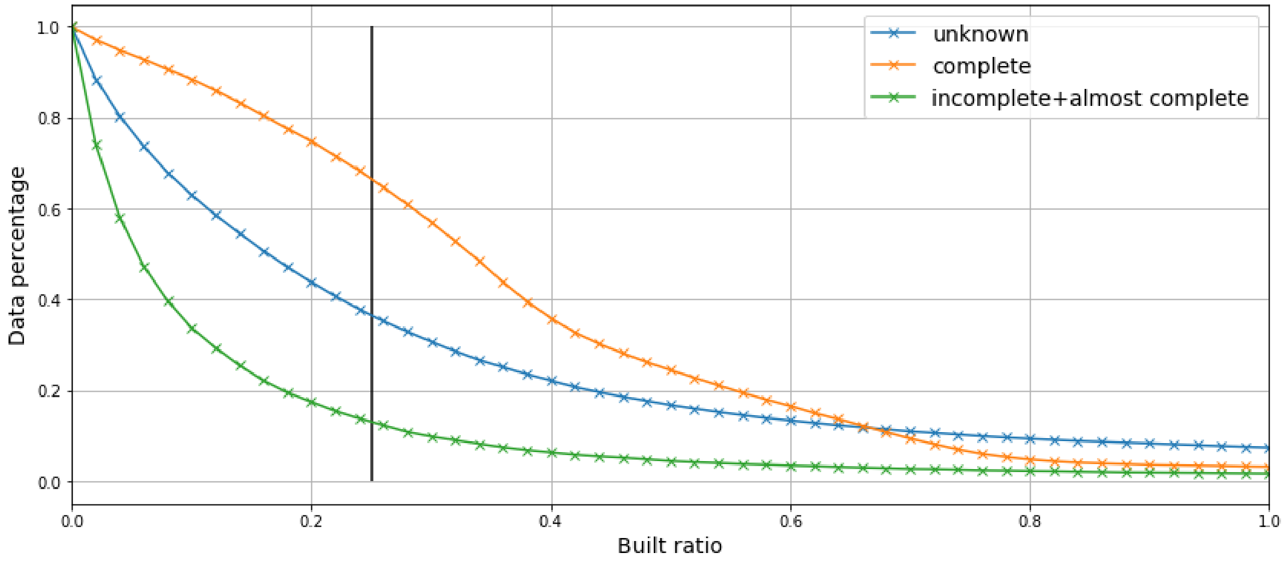


Figure 3 Proportion (0-1) of unknown (blue), complete (orange) and incomplete or almost complete (green) tiles of Attica whose built-up ratio is equal to or larger than the values in the horizontal axis. The black vertical line indicates the selected threshold of 0.25.

It is noted that the built-up area and completeness estimations are carried out considering all occupancy cases together, as it is not possible to distinguish cases in the GHSL data. As the GHSL built-up values are used to distribute aggregated exposure models onto the tiles, this means that a tile that is full of, say, educational buildings, will get assigned residential buildings in the same proportion. In other words, distributing aggregated models using remote sensing-derived built-up areas assumes that all occupancy cases are evenly distributed within the administrative unit, which is generally not true (think of industrial zones, or purely residential neighbourhoods). This also applies in terms of the completeness: the completeness status affects all occupancy cases, but it could be, for example, that the tile is complete in terms of commercial buildings but not residential ones. This is an inherent limitation of the method that could only be solved by classifying built-up areas into occupancy cases at a high-enough resolution level as, for example, marking a whole neighbourhood as residential could lead to lots of commercial or mixed residential-commercial use buildings being ignored.

The final outcome of this process thus consists, firstly, on GHSL built-up areas (with consideration of OSM roads) and, secondly, on an automatic assessment of completeness.

2.4 Processing the ESRM20 aggregated exposure model

We have created the ehre-importer[13] to carry out the processing (i.e., import) of aggregated exposure models so that they can be used to create the EHRE model. In a nutshell, the ehre-importer distributes the aggregated exposure models onto zoom-level 18 quadtiles (defined in EPSG:3857), both in terms of numbers and classes of buildings. In more detail, the processing of the aggregated exposure model of the European Seismic Risk Model 2020 (ESRM20; Crowley et al., 2020, 2021) for the creation of the EHRE model comprises three main aspects:

---

[13] https://git.gfz-potsdam.de/ehre/ehre-software/ehre-importer

1. Assignment of numbers of buildings to each tile (so-called *aggregated buildings*, as they stem from the aggregated model).
2. Assignment of building classes to each tile.
3. Definition of attributes (replacement cost, number of occupants) of building classes.

In order to be able to tackle the first two, it is relevant to introduce the concept of *data-unit tiles*, which is the geographic unit in which all calculations of the EHRE model are carried out. As the name indicates, data-unit tiles stem from the intersection of zoom-level 18 tiles and what we call *data units*, which are the smallest geographic areas in which the aggregated exposure model is defined. In other words, data units define the resolution of the aggregated exposure model, as it is at this level that the model indicates the number of buildings of each class that exist, without any further detail on their spatial distribution. Data units are often administrative units, but they do not need to be. In the case of the ESRM20 model (Crowley et al., 2020) used for EHRE, residential and commercial exposure use administrative units (with associated geodata files provided by the model[14]) while industrial exposure uses administrative units in some countries and 30-arcsec cells in others (as a consequence of the method used to create the model, see Sousa et al., 2017). Other models may use other aggregation strategies, such as Voronoi cells (e.g., Pittore et al., 2020; Gomez-Zapata et al., 2021). In our model, *data units* refer equally to administrative units, 30-arcsec cells, Voronoi cells, etc.

While data units contain the information on numbers of buildings of each class, there are certain properties of the buildings, their costs or their number of occupants that may be defined at a larger scale. In the case of ESRM20 this larger scale is that of individual countries, which in EHRE we name *exposure entities*. An exposure entity may thus be subdivided into one or more data units.

Because aggregated exposure models can be defined at different resolution levels for different occupancy cases, we treat each occupancy case (residential, commercial, industrial) virtually as a separate model. When intersecting zoom-level 18 tiles with data units to create the data-unit tiles, one particular tile may be fully contained in a data unit of commercial exposure (e.g., administrative level 2) and be split in two by data units of the residential exposure (e.g., administrative level 4). In our model, such a tile would thus comprise just one individual data-unit tile for commercial exposure but two data-unit tiles for residential exposure. As a consequence, when we speak of a data unit (or a data-unit tile) we imply a data unit (or a data-unit tile) of a particular occupancy case. Even if, say, residential and commercial exposure are defined at the same resolution, independent data-unit tiles are created for each occupancy case.

Once the data-unit tiles of a data unit are created, the ehre-importer assigns them weights proportionally to their built-up area (determined as described in Section *2.3 Built-Up areas and completeness*). The weight of a data-unit tile of tile *i* and data unit *j* is calculated as per Equation (2):

$$w_{ij} = A_{ij} \Big/ \sum_{k=1}^{k=n} A_{kj} \tag{2}$$

where A_{ij} is the built-up area (determined from GHSL and roads) of the data-unit tile of tile *i* and data unit *j*, and ΣA_{kj} is the summation across all data-unit tiles *k* that belong to data unit *j*. If the summation of built-up areas is zero, the weights are calculated proportionally to the surface area instead. The number of buildings assigned to this data-unit tile of tile *i* and data unit *j* is calculated by multiplying the weight w_{ij} and the total number of buildings in data unit j (of all building classes), $N_{b\,j}$, as per Equation (3):

---

[14] https://git.gfz-potsdam.de/ehre/datasources/-/tree/master/ESRM20_boundaries

$$N_{b\,ij} = N_{b\,j} \cdot w_{ij} = N_{b\,j} \cdot A_{ij} \Big/ \sum_{k=1}^{k=n} A_{kj} \tag{3}$$

The assignment of building classes to each data-unit tile is much simpler as it only depends on the data unit the data-unit tile belongs to. Conceptually, we treat the assignment of building classes in terms of proportions (that sum up to 1), so that the total number of buildings in a data-unit tile multiplied by these proportions yields the number of buildings of each class.

The main challenge that arises during the assignment of building classes is the case in which the exposure entity is not fully covered by data units in the aggregated exposure model, which could make it impossible to assign building classes to OBM buildings potentially located in these areas later on in the process. This is particularly relevant for ESRM20 industrial exposure defined in terms of 30-arcsec cells (as there is no definition of exposure outside of the cells), though not exclusive. Our processing thus identifies gaps in the geographic coverage of the exposure entities and assigns them the weighted average of building classes of the whole exposure entity, by means of creating so-called *filler data units*. Filler data units are partitioned (according to parameters and limits set by the user) so as to avoid the creation of very large data units that would make later processing too long. Once created and stored, these *filler data units* are treated like all other data units and are subdivided into data-unit tiles accordingly. Data-unit tiles of such cases are assigned zero aggregated buildings. Figure 4 shows, as an example, the industrial data units around the area of Rome, Italy.

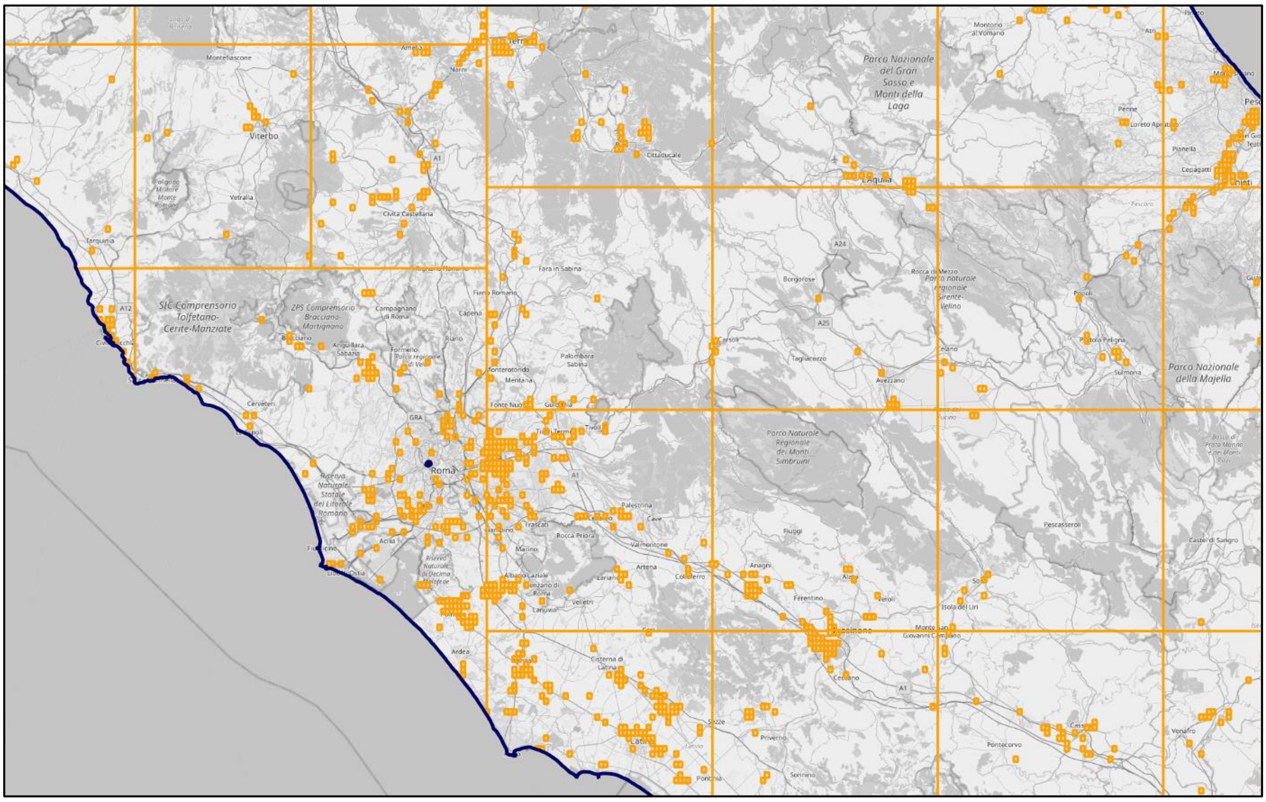
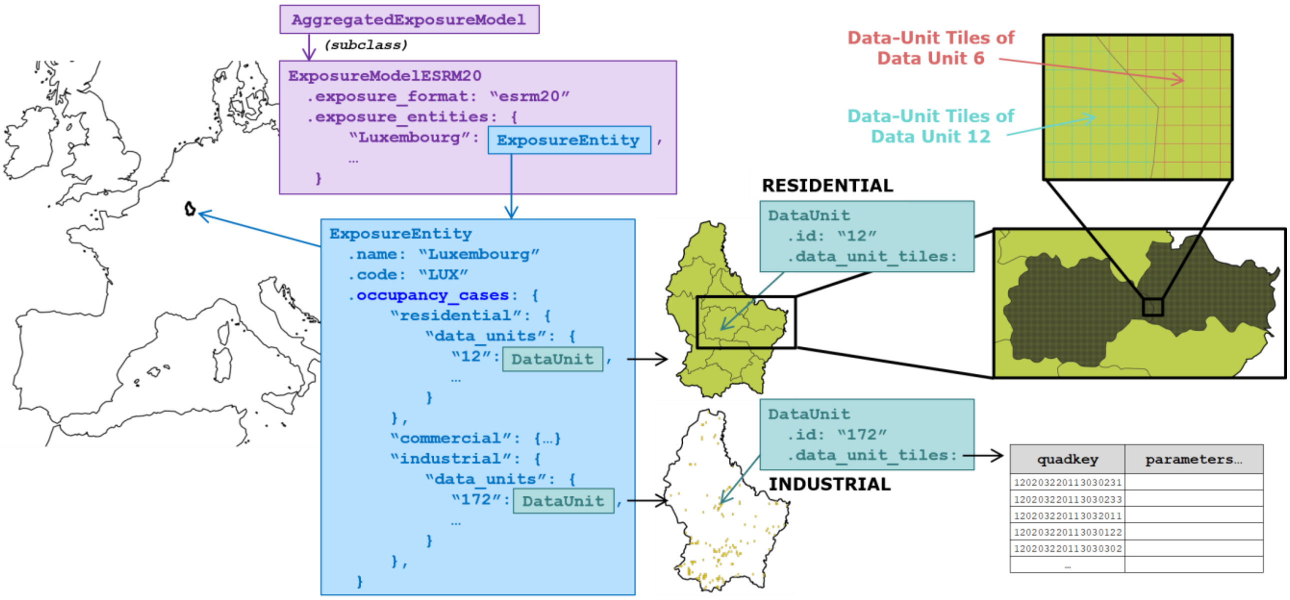


Figure 4 Industrial data units around the area of Rome, Italy. Smaller rectangles are the ESRM20 30-arcsec cells, while larger units are the fillers.

The overall organisation of the geographic space by the ehre-importer in terms of the five concepts of (i) aggregated exposure model, (ii) exposure entity, (iii) occupancy cases, (iv) data unit and (v) data-unit tile is schematically shown in Figure 5 using Luxembourg as an example.



```
AggregatedExposureModel
  (subclass)
ExposureModelESRM20
  .exposure_format: "esrm20"
  .exposure_entities: {
      "Luxembourg": ExposureEntity ,
      …
   }
```

```
ExposureEntity
  .name: "Luxembourg"
  .code: "LUX"
  .occupancy_cases: {
      "residential": {
          "data_units": {
              "12": DataUnit ,
              …
          }
      },
      "commercial": {…}
      "industrial": {
          "data_units": {
              "172": DataUnit ,
              …
          }
      },
  }
```

```
DataUnit
  .id: "12"
  .data_unit_tiles:
```

```
DataUnit
  .id: "172"
  .data_unit_tiles:
```

quadkey	parameters…
120203220113030231	
120203220113030233	
120203220113032011	
120203220113030122	
120203220113030302	
…	

Figure 5 The five core concepts used by the ehre-importer to organise the geographic space, exemplified for Luxembourg.

The final aspect of the processing of aggregated exposure models is the definition of their replacement cost and number of occupants of building classes. The ESRM20 model provides the replacement cost of structural components, non-structural components and contents of each building, and the number of occupants during the day, night and transit times, as well as a value independent of the time of the day. The disaggregation of total replacement costs is a function of occupancy case only, which means that, just like in ESRM20, we treat the total replacement cost as an attribute of the building that can then be multiplied by the corresponding factors to obtain the cost of structural components, non-structural components and contents. The distribution of occupants in different times of the day is also done by means of factors that depend not only on the occupancy case but also on the country, and that can be applied to the value that is independent of time. The latter is not given a name in ESRM20, but we call them "census occupants" as they stem from simply considering the whole population of the country distributed across residential buildings and commercial/industrial activities. In the case of residential buildings, the sum of census occupants of all buildings is equal to the population of the country as indicated in the census, while in the case of commercial and industrial buildings the sums correspond to the total population assumed to work in the commercial and industrial sectors, which is taken in ESRM20 as 40% of the work force (Crowley et al., 2021). The day, night and transit factors allow us to consider the movement of people around the city and its impact on the number of occupants of each individual building.

The census occupants and total replacement cost of a building class of a particular occupancy case can vary from data unit to data unit, even within the same exposure entity. Within each data unit (and, implicitly, an occupancy case), values of census occupants and total replacement costs can be unequivocally assigned to a building class defined by three attributes that we call "building_class_name", "settlement_type" and "occupancy_subtype". These have a one-to-one correspondence with the "taxonomy", "settlement_type" and “occupancy_type” fields of the ESRM20 model. The building class name corresponds to the structural description of the building using the GEM Building Taxonomy v3.0 (Silva et al., 2022). The settlement type can be “urban”, “rural”, “big city” or “all”, and is used in the ESRM20 residential exposure to distinguish across different degrees of urbanisation, as this affects the replacement costs. The occupancy subtype is used in the ESRM20 commercial exposure to provide further details on the occupancy type, and can be “hotels”, “offices” or “trade”; it is also used in the residential exposure model for Cyprus, to distinguish between “apartment” and “single” housing. The category “all” is used when the field is not relevant for the particular occupancy case (e.g., the occupancy subtype of all residential and industrial building classes is “all”).

In all but two countries covered by ESRM20 is the combination of “building_class_name”, “settlement_type” and “occupancy_subtype” (and data unit ID and occupancy case) sufficient to arrive at a unique value of census occupants and of total replacement cost per entry of the ESRM20 exposure CSV file. The two exceptions are Italy and Portugal, as the processing of their input census data was carried out by local experts who used different criteria for the allocation of these parameters. In these cases, the same combination of building class name, settlement type and occupancy subtype can be named more than once for the same data unit ID and occupancy case and have different census occupants and replacement costs. Here, we add up the total number of census occupants and the total replacement cost of all assets associated with that combination (for one data unit and occupancy case) and divide them by the total number of assets to obtain an average value that we finally adopt.

As mentioned earlier, several European countries have their industrial exposure defined in terms of 30-arcsec cells, following the method of Sousa et al. (2017). These are: Albania, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Moldova, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. As the ESRM20 model does not include a geodata file with the geometries of the 30-arcsec cells used to define industrial exposure and only provides their centroids, we create the geometries ourselves around these points as a preliminary step before running the ehre-importer. The current version of the ehre-importer relies on running in advance a Python script that is part of the former prototype version of the EHRE code. The script is called *SERA_creating_industrial_cells.py* and can be found in the GitLab repository of the EHRE prototype code[15]. Instructions on the steps to follow can be found in the README file of ehre-importer, and details on the algorithm works can be found in the documentation of the EHRE prototype code, under “Industrial Cells”[16].

The version of ESRM20 used for EHRE corresponds to the master branch of the ERM20 exposure GitLab repository[17] cloned onto our server on 6 October 2022. This includes updates to ESRM20 v1.0, which was released on 23 November 2021.

2.5 Assigning building classes to OBM buildings

In EHRE, OBM Buildings are assigned ESRM20 building classes based on their location (data unit to which their centroid belongs) and their occupancy case. This task is carried out by our ehre-core[18] software, which is more broadly in charge of merging together the OBM buildings with the ESRM20 model, once the latter has been

---

[15] https://git.gfz-potsdam.de/ehre/ehre-software/prototype/ehre-prototype

[16] https://git.gfz-potsdam.de/ehre/ehre-software/prototype/ehre-prototype/-/blob/master/docs/08_Industrial_Cells.md

[17] https://gitlab.seismo.ethz.ch/efehr/esrm20_exposure

[18] https://git.gfz-potsdam.de/ehre/ehre-software/ehre-core

processed by the ehre-importer. The assignment of building classes to OBM buildings is one of the two fundamental tasks performed by ehre-core, the other being the calculation of remainder buildings in the tiles (see Section *2.6 Calculating remainder buildings*). While retrieving OBM buildings of different occupancy cases, the ehre-core groups building parts that belong to the same *relation ID* as well and treats them as one building thereafter; if the parts have different number of storeys, it takes the largest value.

As mentioned earlier, the ESRM20 model covers three occupancy cases:

- Residential: buildings used mainly for residential purposes.
- Commercial: it includes offices, wholesale and retail trade, hotels and restaurants.
- Industrial: it includes the manufacturing, mining, quarrying, and construction sectors.

Based on the description of the ESRM20 exposure model (Crowley et al., 2020), we have adopted the following grouping of occupancy types per occupancy case:

- Residential: RES (residential, unknown type), RES1 (residential, single dwelling), RES2 (residential, multi-unit, unknown type), RES2A (idem, 2 units), RES2B (idem, 3-4 units), RES2C (idem, 5-9 units), RES2D (idem, 10-19 units), RES2E (idem, 20-49 units), RES2F (idem, 50+ units), RES4 (institutional housing, including military, college, prisons and nursing homes), RES6 (informal housing), MIX1 (mostly residential, and commercial), MIX2 (mostly commercial, and residential), MIX4 (mostly residential, and industrial).
- Commercial: COM (commercial and public, unknown type), COM1 (retail trade), COM2 (wholesale trade and storage), COM3 (offices, professional/technical services), COM5 (restaurants, bars, cafes), RES3 (temporary lodging; e.g. hotels).
- Industrial: IND (industrial, unknown type), IND1 (heavy industrial), IND2 (light industrial), MIX3 (mostly commercial, and industrial), MIX5 (mostly industrial, and commercial), MIX6 (mostly industrial, and residential).

Crowley et al. (2020) describes the industrial exposure as covering light industries, while naming the manufacturing, mining, quarrying, and construction sectors. However, the choice to use the term “light industries” stems from the desire to explicitly exclude activities such as energy production, which are not part of the model (Crowley and Despotaki, pers. comm. 2019).

Only OBM buildings with any of the occupancy types listed above can be assigned ESRM20 building classes. All the rest remain unknown and are not ready to be included in a damage/loss assessment. However, the interested user of the ehre-core software can make decisions and assign building classes and their associated fragility/vulnerability models to some of these buildings, under their own judgement.

It is noted that including COM and IND (in which the specific commercial/industrial activity is unknown) within the commercial and industrial cases may lead to a larger number of buildings being considered than the strict equivalent of ESRM20. We opted for this over the alternative of keeping them out, firstly, because they *may* belong to the cases covered by ESRM20 and, secondly, because excluding them results in OBM buildings for which no building classes are available, which means that they cannot be included in a damage/loss assessment. Moreover, none of the attribution of building classes to individual OBM buildings can be deemed to be the exact truth, as the process of retrieving data from OSM and assigning occupancy types in itself is uncertain by nature, as is the definition of building classes within ESRM20. OBM buildings with assigned building classes are always a model and the assignment may not match reality.

As a starting point, we assign an OBM building all possible ESRM20 building classes associated with their data unit and occupancy case, together with their associated probabilities (equal to their proportion within the data unit). If the number of storeys of the OBM building is available from OSM, we only retain building classes compatible with that number of storeys. If the building is a commercial building, we use its specific occupancy type to narrow down the building classes making use of the "occupancy subtype" field (see Section *2.4 Processing the ESRM20 aggregated exposure model*):

- If the building is RES3 or COM5, we assign building classes with occupancy subtype "hotels" (which is used to gather both hotels and restaurants in ESRM20).
- If the building is COM3, we assign building classes with occupancy subtype "offices".
- If the building is COM1 or COM2, we assign building classes with occupancy subtype "trade".

If after narrowing down the possible building classes there are no classes left, this means that there is an inconsistency between the potential building classes (as a function of data unit ID and occupancy case) and the attributes of the building. In this case, we re-assign all possible building classes and log a warning.

2.6 Calculating remainder buildings

Once all OBM buildings of each occupancy case associated with a specific data-unit tile are identified, our ehre-core software calculates the number of remainder buildings taking into account the completeness of the tile. If a tile is complete, all its exposure is then defined by the OBM buildings located in the tile, and no remainder buildings are added. If a tile *i* is incomplete then the number of remainder buildings in tile *i* and data unit *j* (i.e., data-unit tile *ij*) of a specific occupancy case is calculated as per Equation (4):

$$N_{b\ remainder\ ij} = \max\left[0;\ N_{b\ aggregated\ ij} - N_{b\ OBM\ ij}\right] \tag{4}$$

These remainder buildings are assigned all possible building classes of their data unit (and, implicitly, occupancy case), with the same proportions as in the aggregated model (ESRM20). This means that we do not take into account the building classes assigned to individual OBM buildings to re-calculate the proportions of the remainder buildings. This decision stems from the desire to separate the processing of ESRM20 (distribution onto tiles, etc.) from the processing of OSM/OBM buildings, so as to facilitate re-computation when a change occurs in the future. A consequence of this decision is that the distribution of building classes in the EHRE model may become different from that of the input aggregated exposure models, and that the summation of census people in residential buildings may start to diverge from the census data. However, if the number of OBM buildings for which assigned building classes can be narrowed down is small, the overall proportions are easily maintained.

2.7 Data format and exporting EHRE

All the software used for the creation of OBM and EHRE mentioned in previous sections write their outputs to PostgreSQL (PSQL) databases:

- rabotnik-obm writes its output to the OBM Buildings database, whose structure is described here: https://git.gfz-potsdam.de/ehre/databases/database-obmbuildings.
- obmgapanalysis writes its output to the OBM Tiles database, whose structure is described here: https://git.gfz-potsdam.de/ehre/databases/database-obmtiles.
- ehre-importer and ehre-core write their output to the EHRE Tiles database, whose structure is described in Appendix *8.5 EHRE Tiles database structure* as well as in the corresponding GitLab repository: https://git.gfz-potsdam.de/ehre/databases/database-ehretiles.

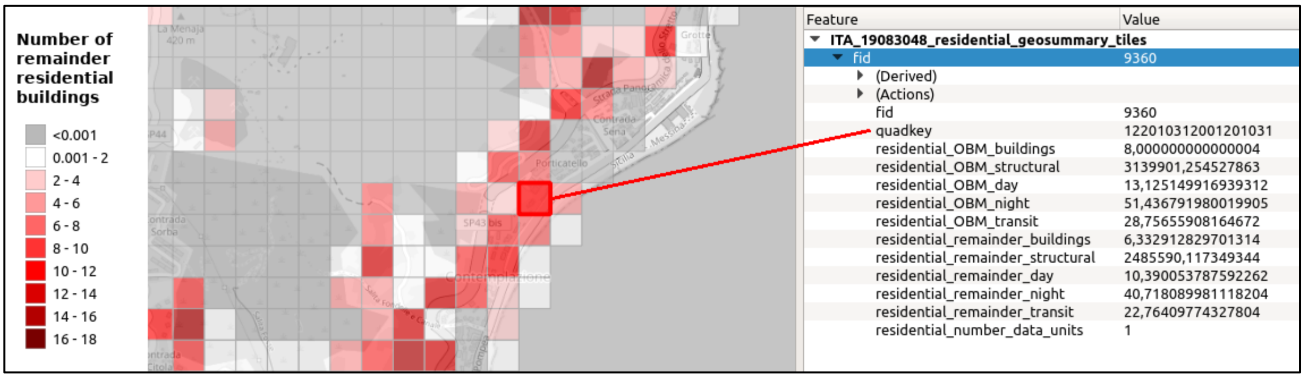
We have created the ehre-exporter[19] software to be able to convert these data into other desired output formats. In its current version, the ehre-exporter can export EHRE into the OpenQuake-engine[20] (Pagani et al., 2014; Silva et al., 2014) exposure CSV file format as well as create GeoPackage (GPKG) files with summary values per tile (e.g., numbers of OBM/remainder buildings, associated replacement costs and number of occupants), which we are naming GeoSummary tiles (Figure 6). When creating OpenQuake exposure CSV files, additional GeoPackage (GPKG) files with the geometries of the tiles and the OBM buildings are generated as well. These allow a user to post-process damage and loss results calculated by running the OpenQuake engine to plot these outputs on a map.

The geographic area to be exported by ehre-exporter is selected firstly by exposure entity (i.e., country) and then by a series of refined options:

- by exposure entity: the quadkeys of the tiles to output are determined by the software as all those associated in the database with the listed exposure entities;
- by data unit ID: the quadkeys of the tiles to output are determined by the software as all those associated in the database with the listed data unit IDs;
- by quadkeys: the user provides a file that lists quadkeys directly;
- by bounding box: the software determines the quadkeys associated with a user-defined bounding box.

All of the options above are constrained by the exposure entities listed by the user, as any quadtile (identified by a quadkey) can be associated with more than one exposure entity, but a user might want to export the data associated only with a particular exposure entity (e.g., a tile in the boundary of Italy and Switzerland, but the user is interested only in Italian exposure).

The user of ehre-exporter is also able to indicate whether they wish to export OBM buildings, remainder buildings and/or aggregated buildings (buildings on the tile as per ESRM20), as well as which types of building replacement costs (structural, non-structural, contents, total) and number of building occupants (at day, night and/or transit times) to include in the output files.



Feature	Value
ITA_19083048_residential_geosummary_tiles	
fid	9360
(Derived)	
(Actions)	
fid	9360
quadkey	122010312001201031
residential_OBM_buildings	8,000000000000004
residential_OBM_structural	3139901,254527863
residential_OBM_day	13,125149916939312
residential_OBM_night	51,436791980019905
residential_OBM_transit	28,75655908164672
residential_remainder_buildings	6,332912829701314
residential_remainder_structural	2485590,117349344
residential_remainder_day	10,390053787592262
residential_remainder_night	40,718089981118204
residential_remainder_transit	22,76409774327804
residential_number_data_units	1

Figure 6 Example of contents of a GeoSummary file created with ehre-exporter (and visualised on QGIS) for which the user selected to include only residential OBM and remainder buildings.

---

[19] https://git.gfz-potsdam.de/ehre/ehre-software/ehre-exporter

[20] https://github.com/gem/oq-engine

2.8 Summary outputs and discussion

While the amount of data involved in the EHRE model is quite large, this section aims at presenting a few summary values and observations stemming from it.

Number of tiles per country

All of the EHRE model covers around 700,000,000 zoom-level 18 tiles.

Table 1 shows the number of tiles associated with each European country covered by ESRM20 and, consequently, EHRE. These result from the intersection of zoom-level 18 tiles and the geodata files of administrative boundaries from the ESRM20 model.

Table 1 Number of zoom-level 18 tiles associated with each country covered by the EHRE model.

Country	ISO3 Code	Number
Sweden	SWE	93,511,050
Finland	FIN	78,478,925
Norway	NOR	77,643,842
Türkiye	TUR	55,595,995
France	FRA	49,953,566
Germany	DEU	38,970,954
Spain	ESP	37,342,322
Poland	POL	35,561,000
United Kingdom	GBR	30,875,010
Iceland	ISL	24,535,968
Italy	ITA	24,119,845
Romania	ROU	21,059,470
Greece	GRC	9,515,038
Latvia	LVA	9,252,027
Bulgaria	BGR	8,876,402
Hungary	HUN	8,624,513
Lithuania	LTU	8,596,008
Ireland	IRL	8,422,232
Czechia	CZE	8,093,897
Austria	AUT	7,900,307
Estonia	EST	7,213,812
Portugal	PRT	6,659,680
Serbia	SRB	6,535,948
Denmark	DNK	5,928,779
Croatia	HRV	4,934,189
Slovakia	SVK	4,829,858
Netherlands	NLD	4,310,179
Bosnia and Herzegovina	BIH	4,258,268
Switzerland	CHE	3,792,436
Belgium	BEL	3,271,711
Republic of Moldova	MDA	3,153,676
Albania	ALB	2,177,270
North Macedonia	MKD	1,904,965
Slovenia	SVN	1,824,879
Montenegro	MNE	1,064,660
Kosovo	XKX	865,651
Cyprus	CYP	382,703
Luxembourg	LUX	267,637
Isle of Man	IMN	73,496
Andorra	AND	36,262
Malta	MLT	22,101
Liechtenstein	LIE	15,423
Gibraltar	GIB	582
Monaco	MCO	247

Number of buildings per country

The whole of the EHRE model contains 219,000,466 buildings, of which 207,103,640 (94.6%) are residential, 6,406,816 (2.9%) are commercial, and 5,490,010 (2.5%) are industrial. The distribution of these numbers of buildings across different countries can be observed in the maps in Figure 7 through Figure 10. These and all subsequent maps use country boundaries as provided by ESRM20.

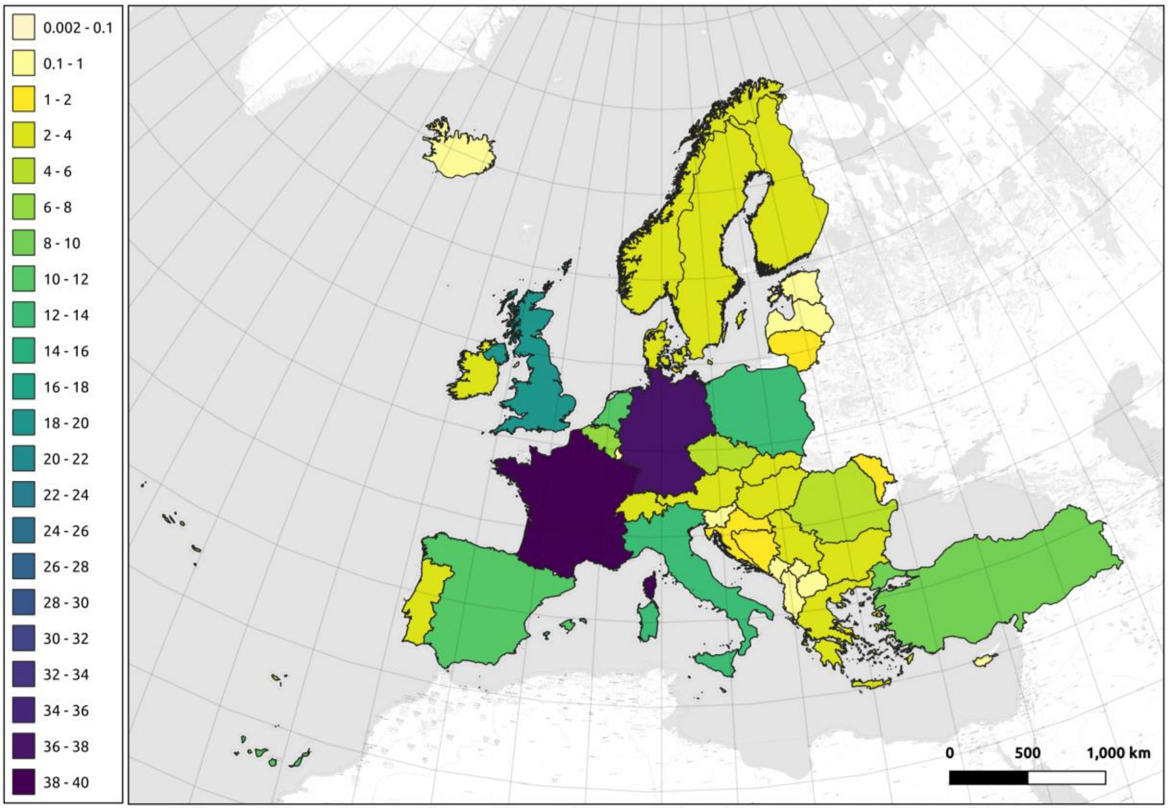


Figure 7 Total number of buildings (residential, commercial and industrial) per country in the EHRE model. Colour scale in millions of buildings.

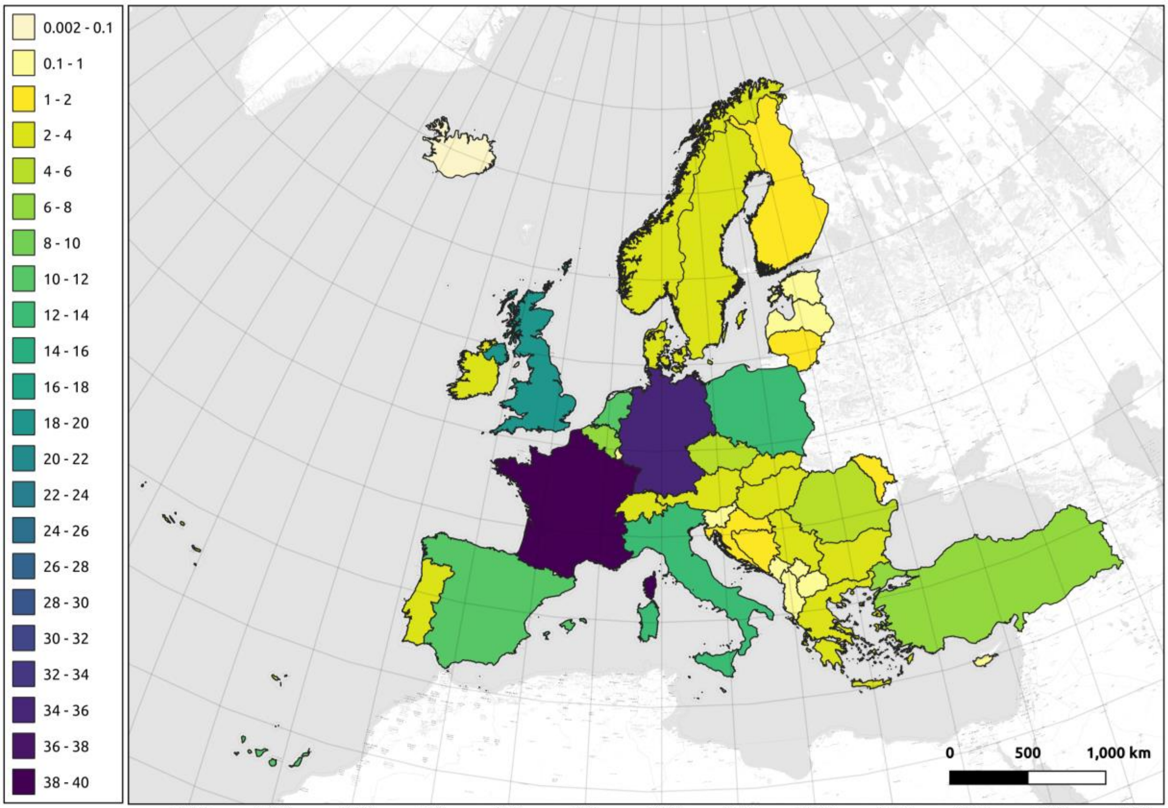


Figure 8 Total number of residential buildings per country in the EHRE model. Colour scale in millions of buildings.

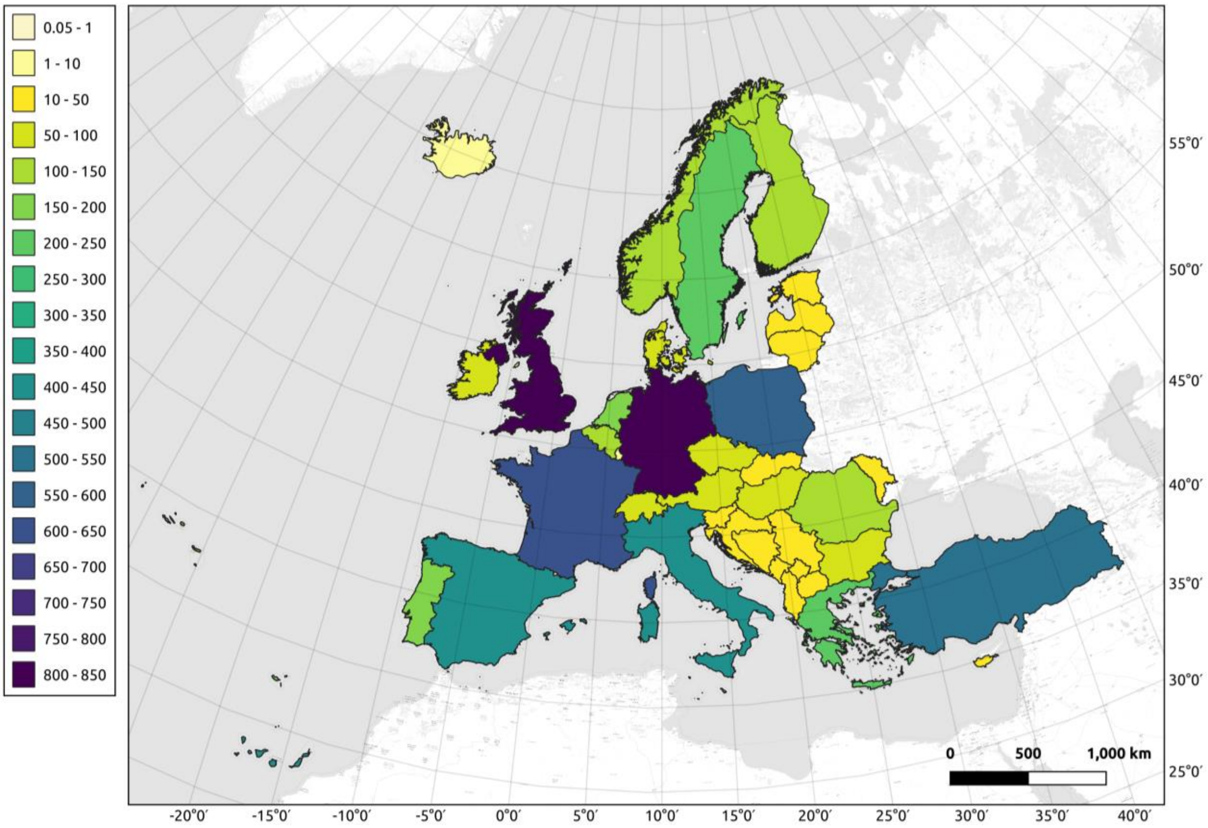


Figure 9 Total number of commercial buildings per country in the EHRE model. Colour scale in thousands of buildings.

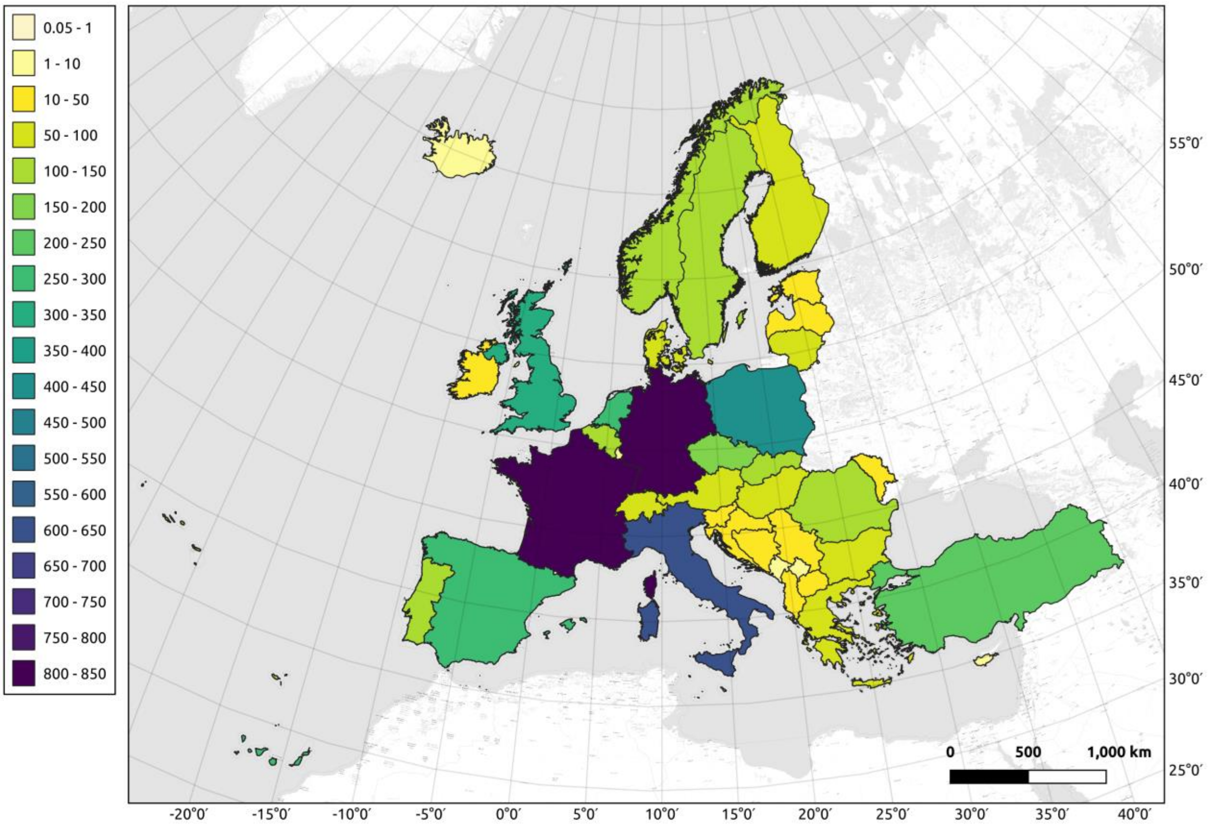


Figure 10 Total number of industrial buildings per country in the EHRE model. Colour scale in thousands of buildings.

The 219,000,466 total (residential + commercial + industrial) buildings in EHRE represents a 53.6% increase with respect to the total 142,569,366 of ESRM20. This relative difference in the number of buildings between EHRE and ESRM20 varies significantly for different countries and occupancy cases.

Table 2 lists the number of buildings in the EHRE model per country and occupancy case (residential, commercial, industrial, all three) as well as the ratios of the EHRE model to the ESRM20 model in each case. The ratios are plotted in the maps from Figure 11 through Figure 14.

The reasons for the discrepancies in the final number of buildings between the EHRE and the ESRM20 model can be multiple and may even be different for different countries. Understanding which (if any) of these reasons apply to each country and/or smaller administrative areas requires an in-depth investigation not undertaken yet. Some of the sources of the discrepancies (discussed one by one further down) may be:

- the building of new structures in the years between the gathering of data used by ESRM20 and the time of retrieval of OSM data,
- the assignment of occupancy cases to OSM buildings in OBM,
- the equivalence of occupancy cases in OBM and ESRM20,
- the way in which individual buildings are represented in OSM,
- the processing of parts of buildings (*relations*) in OBM,
- the recognition of built-up areas from remote sensing in GHSL,
- the completeness of OBM associated with all occupancy cases,
- the time difference in between calculating the completeness of OBM and retrieving OBM buildings and data for EHRE,
- EHRE’s distribution of ESRM20 first and then comparison against OBM,
- the resolution of the geodata files of administrative boundaries,
- the determination of the number of buildings in the ESRM20 model in itself.

Impact of the passage of time

The ESRM20 exposure model combines input data from different points in time, as this ultimately depends on the dates at which censuses and statistical activities are conducted in different countries or regions. According to Crowley et al. (2020), the oldest residential census data used in ESRM20 is that of Italy, Lithuania, Luxembourg, Slovakia and the United Kingdom (all from 2001), and the newest is that of the Netherlands and Norway (from 2017). Data for commercial and industrial buildings also stem from different years. As the OpenBuildingMap database used to create the EHRE model was populated with OSM data retrieved between June and September 2022, some increases in the number of buildings from ESRM20 to EHRE might be due to more recently-built structures.

Table 2 Number of buildings in the EHRE model and ratio with respect to the ESRM20 model (EHRE/ESRM20) per country.

Country	Residential		Commercial		Industrial		All	
	Number	Ratio	Number	Ratio	Number	Ratio	Number	Ratio
Albania	638,802.3	1.068	27,159.5	0.923	15,848.0	0.995	681,809.7	1.059
Andorra	10,560.2	0.837	2,604.7	0.613	739.7	0.613	13,904.6	0.769
Austria	3,497,012.4	1.596	83,893.0	0.890	69,666.1	2.314	3,650,571.4	1.576
Belgium	6,127,399.7	1.664	104,504.6	0.989	111,292.1	1.880	6,343,196.4	1.649
Bosnia and Herzegovina	1,070,965.6	0.993	13,360.4	0.940	12,350.0	1.875	1,096,676.0	0.998
Bulgaria	2,095,839.5	1.017	73,477.2	0.910	69,095.7	1.342	2,238,412.4	1.021
Croatia	1,738,490.3	1.080	36,880.4	1.086	28,771.9	1.066	1,804,142.6	1.080
Cyprus	270,210.3	1.022	15,943.9	1.031	6,614.1	0.969	292,768.3	1.021
Czechia	4,571,061.5	2.093	91,358.5	0.839	184,197.2	2.302	4,846,617.2	2.042
Denmark	3,253,833.2	2.118	62,980.3	0.768	99,132.5	1.448	3,415,945.9	2.025
Estonia	470,458.2	2.218	12,073.3	0.937	19,987.4	2.019	502,518.8	2.139
Finland	1,920,515.6	1.553	114,859.8	1.864	72,986.8	1.825	2,108,362.1	1.575
France	38,213,657.0	2.645	614,266.0	1.030	801,215.6	2.666	39,629,138.6	2.582
Germany	35,219,795.0	1.861	805,714.9	1.292	840,705.0	2.162	36,866,214.9	1.849
Gibraltar	2,109.5	0.521	738.1	0.527	132.3	2.242	2,979.9	0.541
Greece	3,160,489.1	1.036	234,373.6	0.939	57,801.3	1.128	3,452,664.0	1.030
Hungary	3,240,645.7	1.209	71,242.3	0.883	90,916.8	1.464	3,402,804.9	1.205
Iceland	92,291.5	1.536	5,475.0	0.645	5,820.7	1.062	103,587.3	1.399
Ireland	2,307,425.0	1.244	64,897.0	1.399	39,046.3	1.390	2,411,368.3	1.250
Isle of Man	23,485.8	1.503	3,197.4	0.947	1,053.4	1.160	27,736.6	1.393
Italy	12,128,298.3	1.068	415,573.3	0.761	641,755.2	2.236	13,185,626.9	1.082
Kosovo	288,732.2	1.166	31,012.8	0.828	7,778.4	1.088	327,523.5	1.121
Latvia	473,773.6	1.355	19,401.3	0.955	27,138.1	2.056	520,313.0	1.358
Liechtenstein	15,716.7	1.506	1,451.6	0.462	876.6	1.478	18,044.9	1.273
Lithuania	1,459,505.1	2.876	48,882.8	0.699	51,917.1	2.382	1,560,305.0	2.604
Luxembourg	189,318.2	1.583	9,838.3	1.071	4,161.7	1.982	203,318.3	1.553
Malta	80,569.0	0.854	11,028.0	0.846	8,336.0	0.844	99,933.0	0.852
Moldova	1,081,660.6	1.370	20,122.8	1.048	11,203.7	2.811	1,112,987.0	1.370
Monaco	3,315.8	0.362	301.2	0.381	50.2	0.301	3,667.2	0.362
Montenegro	171,842.1	1.137	10,084.4	0.894	2,546.5	1.015	184,473.0	1.118
Netherlands	11,061,498.0	2.158	161,224.3	0.650	264,082.7	2.113	11,486,805.1	2.089
North Macedonia	427,085.6	0.957	30,819.8	0.867	10,542.0	0.875	468,447.4	0.949
Norway	3,612,354.4	2.354	137,424.0	1.134	104,978.0	0.998	3,854,756.3	2.189
Poland	12,578,890.4	1.992	584,456.6	1.089	417,526.4	2.422	13,580,873.4	1.933
Portugal	3,369,440.0	1.005	161,472.1	0.930	105,997.3	1.041	3,636,909.4	1.002
Romania	5,640,644.2	1.072	146,124.1	0.842	100,004.6	1.359	5,886,772.9	1.069
Serbia	2,485,130.8	1.093	39,219.2	0.894	30,156.6	1.254	2,554,506.6	1.091
Slovakia	2,165,880.2	2.512	48,505.3	0.829	106,729.8	1.412	2,321,115.4	2.329
Slovenia	607,187.5	1.311	23,114.3	0.842	20,497.4	0.949	650,799.2	1.271
Spain	10,158,112.0	1.092	447,249.3	0.877	288,431.6	1.465	10,893,792.9	1.088
Sweden	2,965,040.3	1.467	210,334.7	0.815	132,663.8	1.335	3,308,038.8	1.391
Switzerland	2,248,555.2	1.299	84,666.1	0.717	65,649.4	2.627	2,398,870.8	1.280
Türkiye	7,705,676.5	0.921	508,569.6	0.851	217,244.7	1.103	8,431,490.9	0.920
United Kingdom	18,260,366.0	1.245	816,940.1	1.349	342,369.6	1.702	19,419,675.7	1.255
Total	207,103,639.8	1.556	6,406,816.0	0.987	5,490,010.3	1.822	219,000,466.1	1.536

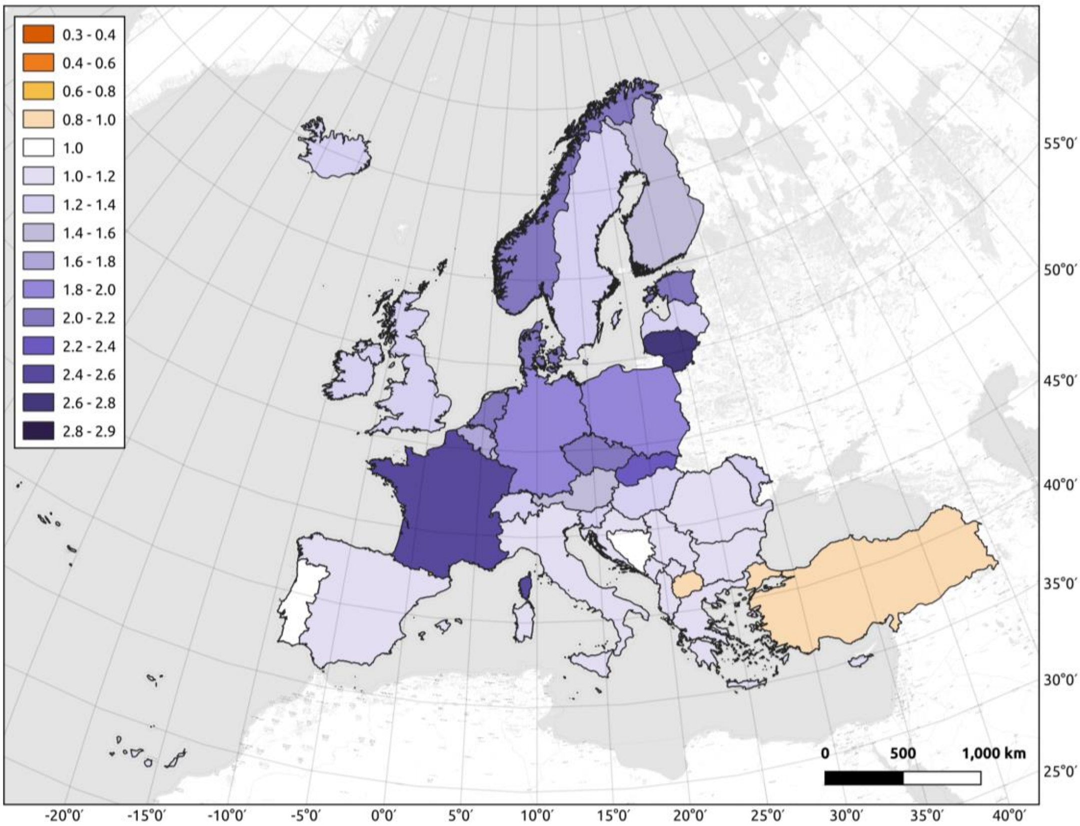


Figure 11 Ratio of EHRE to ESRM20 total (residential, commercial and industrial) buildings per country (1 defined as 0.99-1.01).

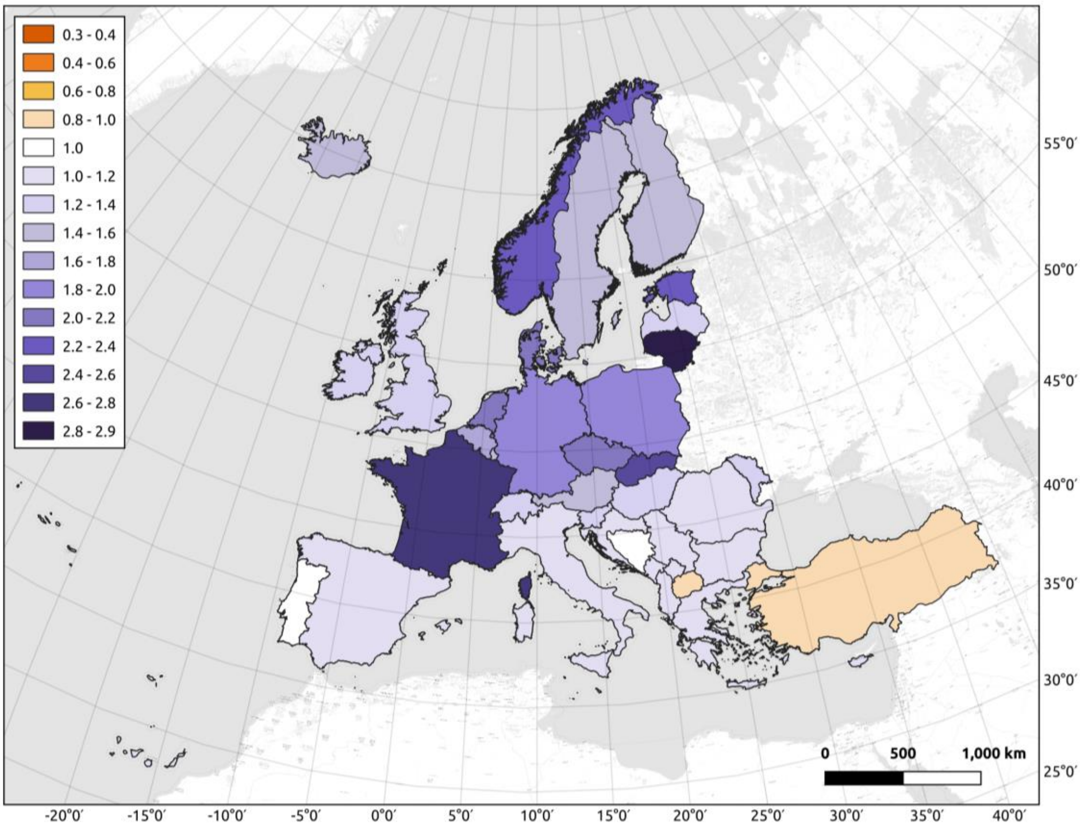


Figure 12 Ratio of EHRE to ESRM20 residential buildings per country (1 defined as 0.99-1.01).

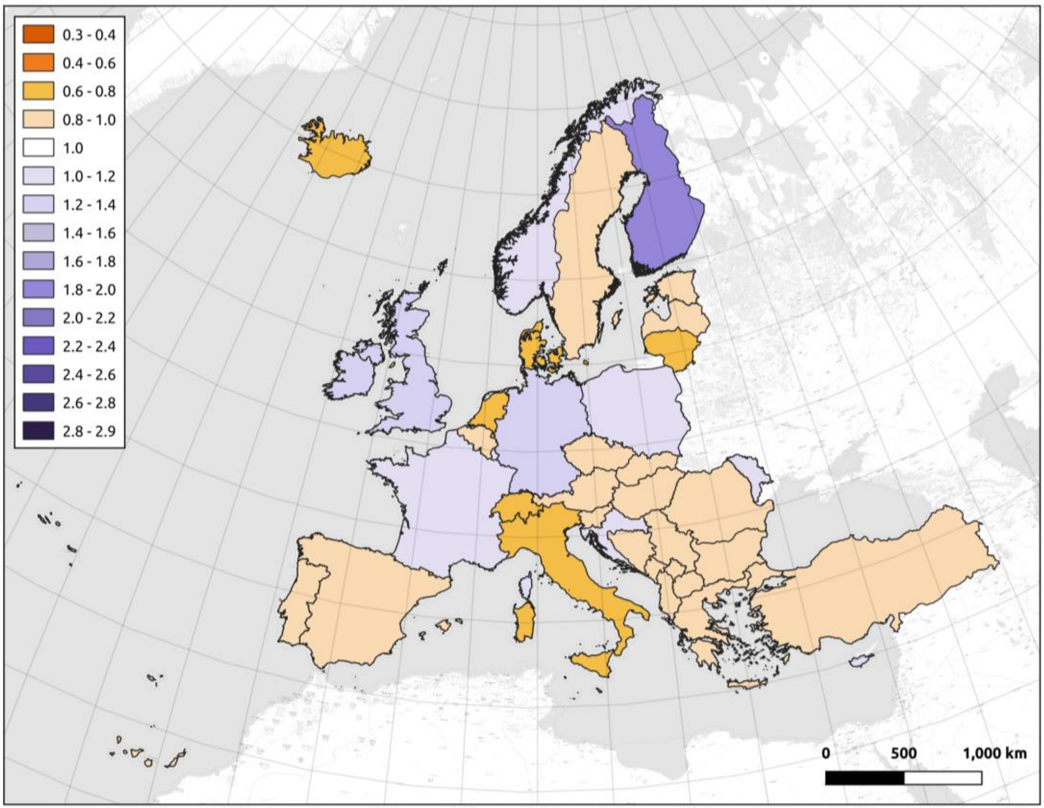


Figure 13 Ratio of EHRE to ESRM20 commercial buildings per country (1 defined as 0.99-1.01).

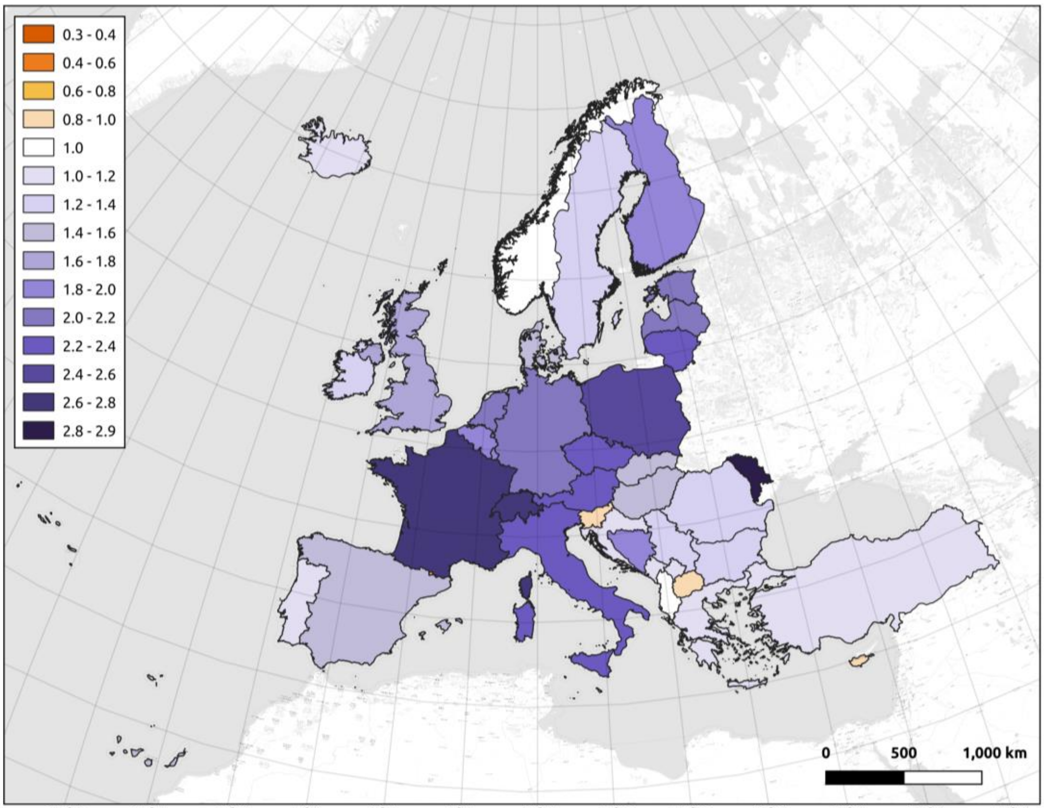


Figure 14 Ratio of EHRE to ESRM20 industrial buildings per country (1 defined as 0.99-1.01).

Uncertainty in the assignment of occupancy cases in OBM

The assignment of occupancy cases to OSM buildings by rabotnik-obm is a very complex problem for which the procedure described in Section *2.5 Assigning building classes to OBM buildings* has a series of limitations. Firstly, there are OSM buildings for which there is no information in OpenStreetMap other than the observation that they are buildings, or for which the available information has not been enough for the current algorithm to assign an occupancy type (17.05% of the buildings in the OBM database used for EHRE correspond to such a situation in which unknown, i.e. “UNK”, occupancy has been assigned). Such buildings cannot be classified as residential, commercial or industrial and are thus not part of the EHRE model; in tiles classified as complete (where the total number of buildings is taken directly from OMB), this can result in an underestimation of residential, commercial or industrial buildings (the tile might actually be complete but the OBM buildings cannot be assigned to an occupancy case and get left out). Secondly, the logical rules defined in order to be able to select a final occupancy type for a building from a series of occupancy strings that do not match (i.e., the case in which several tags apply to the same building) are limited in their capacity to capture all possible combinations and outcomes. This limitation can lead to misclassifications that place OBM buildings in the wrong category and, ultimately, wrong occupancy case. Such misclassifications can lead equally to an over- or underestimation of the number of buildings.

Equivalence of occupancy cases in OBM and ESRM20

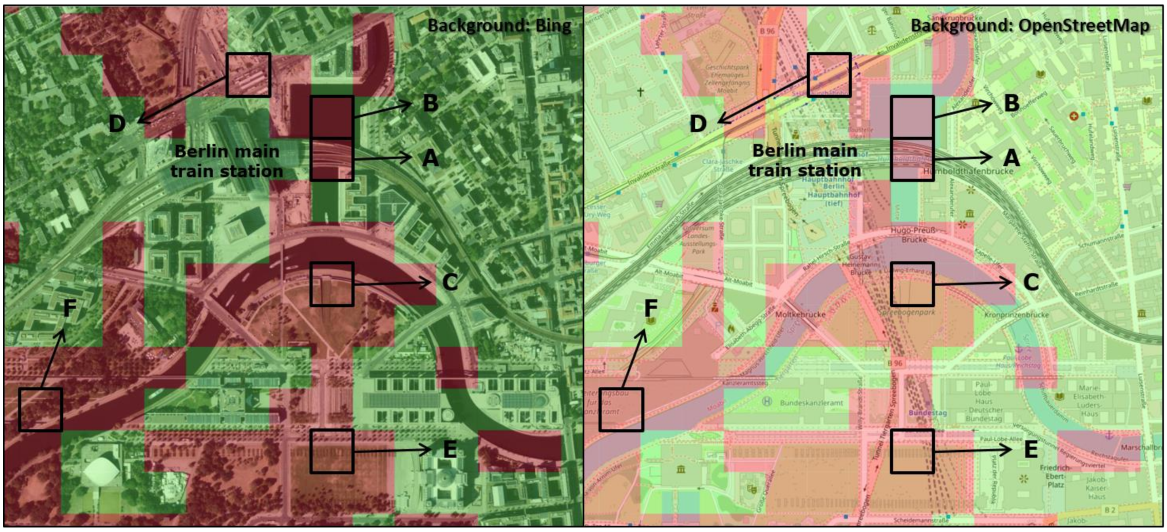
The grouping of occupancy types (e.g., “RES”, “RES1”, etc.) into occupancy cases (e.g., residential) carried out by ehre-core may lead to discrepancies in the number of buildings in each occupancy case as well. As explained in Section *2.5 Assigning building classes to OBM buildings*, OBM buildings that resulted in “COM” or “IND”, that is, commercial/industrial building of unknown sub-type, are included in the EHRE model, even if their actual (unknown) sub-type may be of a kind not included in the ESRM20 model. This may lead to the comparison between the two models not being done in the same terms, with EHRE potentially including sub-types that are not included in ESRM20.

Representation of individual buildings in OSM and processing of relations in OBM

Individual buildings can be represented in OSM in different ways, and these might not be consistent with the way buildings are defined in building census data (in the few countries where they are conducted). In a study focused on the city of Cologne, Germany, Nievas et al. (2022) found buildings seemingly representing sometimes a structure and sometimes an address, as well as buildings with vertical irregularities being represented in OSM as different polygons with different number of storeys. Paprotny et al. (2020) noted that sometimes neighbouring buildings within an urban block are represented in OSM as one artificially large building. The latter is not a problem if both polygons are marked in OSM as belonging to the same so-called *relation*, as rabotnik-obm collects relation IDs associated with the buildings, and ehre-core treats members of the same relation as one building. The case in which the building parts are marked as belonging to more than one relation has not been addressed in rabotnik-obm yet, and the software only keeps the first relation ID that it finds for each building, which may lead to parts of a building not being joined together by ehre-core (particularly when a system of nested relations has been used in OSM to define a building) and a subsequent overestimation of the total number of buildings. This last point is not expected to be having a large impact on EHRE, as only 122,110 (0.06%) of the 192,358,699 buildings in the OBM database used for EHRE have a relation ID.

Recognition of built-up areas from remote sensing

As the distribution of the ESRM20 onto tiles and the classification of tiles as complete/incomplete are carried out based on remote sensing-derived estimates of built-up area from GHSL, any overestimation of built-up area in the latter has the double effect of increasing both the chance of tiles being classified as incomplete and the number of ESRM20 buildings allocated to them. Figure 15 shows the consequences of built-up area being overestimated in tiles occupied with railway tracks, water and green spaces in the area around the main train station in Berlin, Germany. As can be observed in the tiles marked as an example, remainder residential and commercial buildings are allocated to these tiles in EHRE when, in fact, there are no residential and commercial buildings there.



A quadkey: 120210233222003203
0 OBM buildings
7.35 remainder buildings (res + com)

C quadkey: 120210233222021001
0 OBM buildings
3.99 remainder buildings (res + com)

E quadkey: 120210233222021201
0 OBM buildings
5.15 remainder buildings (res + com)

B quadkey: 120210233222003201
0 OBM buildings
3.73 remainder buildings (res + com)

D quadkey: 120210233222002133
0 OBM buildings
5.06 remainder buildings (res + com)

F quadkey: 120210233222020032
0 OBM buildings
4.64 remainder buildings (res + com)

Figure 15 Example of remainder buildings added to EHRE around the area of the main train station in Berlin due to inaccuracies in the estimation of the built-up area. Green and red tiles are classified as complete and incomplete, respectively.

Impossibility to distinguish occupancy cases to determine OBM completeness

Linked to this is the fact that completeness can only be determined for all occupancy cases together, and also includes buildings that do not belong to residential, commercial or industrial occupancies, and buildings whose occupancy cannot be determined in OBM. A tile may be classified as complete because the sum of the areas of OBM buildings in the tile leads to that conclusion but, if many of these buildings do not have tags that allow them to have an occupancy type assigned to them, the final number of buildings of residential, commercial and industrial will be smaller than in reality. If this happens with a tile classified as incomplete, the effect can be the opposite, as buildings from ESRM20 are taken into account for the calculation of remainder buildings, but the OBM buildings available belong to neither residential nor commercial nor industrial occupancies and thus no (or very few) OBM buildings are subtracted.

Time elapsed between processing completeness and retrieving OBM buildings for EHRE

Some discrepancies might stem from the fact that the OBM building data directly incorporated into the EHRE model (via ehre-core) and the OBM building data fed into the obmgapanalysis software to determine the completeness of OBM do not correspond to the exact same point in time. The built-up area and completeness analysis of different countries was carried out in between May and November 2022, while the retrieval of OSM data with rabotnik-obm took place between June and September 2022; within these time ranges, countries may have been processed in different orders for these different purposes. The most likely outcome of this time shift is the potential misclassification of tiles as incomplete if buildings were added to OSM and retrieved by rabotnik-obm after the completeness had been estimated with these buildings missing in OSM (more likely than the opposite case of buildings being deleted). However, given that the maximum time shift would be less than half a year, we do not expect this to have a large impact in the resulting estimates of completeness, particularly when considering all other uncertainties associated with the completeness estimation itself (i.e., the accuracy of the built-up areas, the use of a fixed ratio of 0.25 as threshold, the fact that different occupancy cases cannot be separated).

Distribution of ESRM20 onto tiles

Another way in which potential inaccuracies in the identification of built-up areas may propagate to an overall divergence between the number of buildings in EHRE and ESRM20 comes with the fact that, to create EHRE, ESRM20 is first distributed onto the zoom-level 18 tiles (proportionally to the estimated built-up areas), and then this number is compared against the number of OBM buildings (in incomplete tiles) or simply discarded, if the tile is complete. This merging of ESRM20 and OBM at the tile level, instead of at the level of the resolution of ESRM20 (e.g., administrative units), means that discrepancies can arise from these assumptions of spatial distribution. Put in simple words, the same number of buildings can be distributed across a region in different ways, and comparing the two at the tile level once the distribution is done can lead to different conclusions than comparing the two at the level of the coarser resolution. As a conceptual example, Figure 16 shows how a simple swap of the number of ESRM20 buildings assigned to two incomplete tiles leads to a different total number of buildings in EHRE and, consequently, a different ratio with respect to ESRM20.

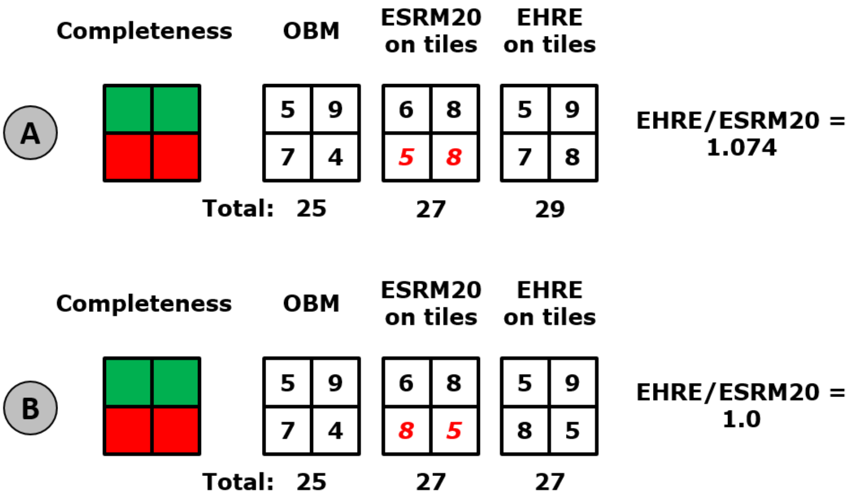


Figure 16 Fictitious examples of calculations of the total number of buildings (of a certain occupancy case) in EHRE. Green and red tiles are classified as complete and incomplete, respectively. The only difference in the input data between A and B is the number of buildings in the two incomplete tiles (5 and 8), which are swapped. "EHRE on tiles" refers to the final number of buildings (OBM plus remainder) in the EHRE model in each case.

Resolution of ESRM20 administrative boundaries

The resolution of the administrative boundaries used in ESRM20 (and, consequently, in EHRE) might contribute to the discrepancies in the final number of buildings. It has been observed (e.g., in Athens, Greece) that some of these boundaries are quite coarsely defined when confronted against higher-resolution boundaries, with within-city boundary lines often crossing through buildings and blocks instead of following streets or rivers, which are usually used to subdivide urban spaces. As the number of ESRM20 buildings assigned to each data unit tile depends on the intersection between data units (represented by the ESRM20 geodata files) and tiles, a coarse resolution boundary may lead to building numbers being erroneously allocated to a different data unit, a difference that will later propagate to the calculation of remainder buildings when confronted against the OBM buildings in the tile. This issue can result in either the over- or underestimation of numbers of buildings.

Uncertainties in the number of buildings in the ESRM20 model

The number of buildings in ESRM20 itself includes uncertainties for some countries, as this information is not readily available everywhere. Of most interest for the present discussion, though, are the cases in which the total number of buildings in ESRM20 were obtained or calibrated against national statistics of numbers of buildings. Table 3 shows the number of residential buildings in the EHRE model (in terms of the total as well as per individual OBM and remainder buildings components), in the ESRM20 model, and as per national census/statistical data for an arbitrary selection of countries. Within this selection, the number of residential buildings in the EHRE model seems the most consistent with the ESRM20 model and census data for Albania, Greece, Italy, Portugal, Romania and Spain. It is interesting that these six countries have the smallest ratios of number of OBM buildings to total EHRE buildings of the eleven, though Italy (51.7%) is quite close to Finland (58.5%) in this regard; the meaning of this observation (if there is any) remains to be investigated. It is noted that Portugal and Spain were expected to have more residential buildings in EHRE than in ESRM20 due to the fact that ESRM20 does not include exposure data for the Canary Islands (Spain), or the islands of Madeira and Azores (Portugal), while EHRE and the census data do. As the numbers of buildings in ESRM20 for Austria, Finland, Germany, Norway and Switzerland are very close to those from the census/statistical data, and the ratios of EHRE buildings to ESRM20 and census data obtained are too large to be due only to an increase in the number of buildings in the years in between the census and the present time, a more in-depth investigation is needed for these countries (and others with large discrepancies in Table 2). The examples in Table 3 highlight the need for a systematic revision of the resulting exposure models at the country level, to be carried out in the future.

The impact of large potential over- or underestimations in the total number of buildings is not limited to the buildings per se but propagates as well to the total replacement cost and, most importantly, the total number of exposed people. This is relevant because ESRM20 makes sure that the total sum of census population in the residential buildings matches the total number of inhabitants reported in national census data by adjusting the average number of occupants per building if necessary. The current version of the EHRE model does not adjust the average number of people per building to maintain this overall statistic. As a consequence, a significant increase in the number of buildings from ESRM20 to EHRE leads to a significant discrepancy in the overall number of inhabitants of the country (as is the case for Germany or Austria, for example). Making such an adjustment is not trivial, and thus potential strategies to tackle this issue need to be investigated.

Table 3 Number of residential buildings in the EHRE model (total and individual OBM and remainder components), in ESRM20, and in census/statistical data, as well as EHRE/ESRM20 and EHRE/statistics ratios for a selection of countries.

Country	OBM	EHRE remainder	EHRE total	ESRM20	ESRM20 Year	EHRE/ ESRM20	Census or Statistics	Stats Year	EHRE/ Stats	Stats Source
Albania	169,503	469,299.3	638,802.3	598,267.0	2011	1.068	598,267	2011	1.068	a
Austria	2,931,494	565,518.4	3,497,012.4	2,191,280.0	2011	1.596	1,978,794	2011	1.767	b
Finland	1,123,208	797,307.6	1,920,516	1,236,900.0	2011	1.553	1,328,682	2022	1.445	c
							1,294,426	2017	1.484	d
Germany	31,312,996	3,906,799.0	35,219,795.0	18,922,618.0	2011	1.861	18,922,614	2011	1.861	e
Greece	607,409	2,553,080.1	3,160,489.1	3,051,157.9	2011	1.036	3,246,008	2011	0.974	f
Italy	6,276,273	5,852,025.3	12,128,298.3	11,354,381.3	2001	1.068	12,187,698	2011	0.995	g
Norway	2,867,247	745,107.4	3,612,354	1,534,879.6	2017	2.354	1,592,339	2023	2.269	h
							1,534,929	2017	2.353	i
Portugal	664,351	2,705,089.0	3,369,440.0	3,353,762.0	2011(*)	1.005	3,573,416	2021	0.943	j
							3,544,389	2011	0.951	k
Romania	1,207,976	4,432,668.2	5,640,644.2	5,260,264.0	2011	1.072	5,117,940	2011	1.102	l
Spain	3,224,707	6,933,405.0	10,158,112.0	9,305,735.0	2011(**)	1.092	9,730,999	2011	1.044	m
Switzerland	1,799,537	449,018.2	2,248,555.2	1,730,823.0	2016	1.299	1,730,415	2016	1.299	n
							1,765,551	2020	1.274	o
							1,664,581	2021(***)	1.351	p

Notes

(*) ESRM20 does not include buildings in the Azores and Madeira islands, while the census data and EHRE do.
(**) ESRM20 does not include buildings in the Canary islands, while the census data and EHRE do.
(***) The Swiss Earthquake Risk Model (ERM_CH23) discards buidings with volume smaller than 200 m^3.

Sources

a Institute of Statistics of Albania: https://www.instat.gov.al/en/themes/censuses/census-of-population-and-housing/#tab2
b Statistics Austria: https://www.statistik.at/en/statistics/population-and-society/housing/stock-of-buildings
c Statistics Finland: https://pxdata.stat.fi/PxWeb/pxweb/en/StatFin/StatFin__rakke/statfin_rakke_pxt_116g.px/
d Statistics Finland: https://statfin.stat.fi/PxWeb/pxweb/en/StatFin_Passiivi/StatFin_Passiivi__rakke/statfinpas_rakke_pxt_001_201700.px/
e Federal Statistical Office of Germany: https://ergebnisse.zensus2011.de
f Hellenic Statistical Authority: https://www.statistics.gr/documents/20181/1204358/A1601_SKT01_DT_DC_00_2011_01_F_EN.pdf
g Italian National Institute of Statistics: https://www.istat.it
h Statistics Norway: https://www.ssb.no/en/bygg-bolig-og-eiendom/bygg-og-anlegg/statistikk/bygningsmassen
i Statistics Norway: https://www.ssb.no/en/statbank/table/03158/tableViewLayout1/
j Statistics Portugal: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0011786&contexto=bd&selTab=tab2
k Statistics Portugal: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0005967&contexto=bd&selTab=tab2
l Romanian Central Population and Housing Census Commission: https://insse.ro/cms/files/statistici/comunicate/alte/2012/Comunicat%20DATE%20PROVIZORII%20RPL%202011e.pdf
m Spanish National Statistics Institute: https://www.ine.es/censos2011_datos/cen11_datos_resultados.htm#
n Swiss Federal Statistical Office: https://www.bfs.admin.ch/bfs/de/home/statistiken/bau-wohnungswesen.assetdetail.3982429.html
o Swiss Federal Statistical Office: https://www.bfs.admin.ch/bfs/en/home/statistics/construction-housing.assetdetail.17924951.html
p Wiemer et al. (2023) Earthquake Risk Model of Switzerland (ERM-CH23), Swiss Seismological Service, ETH Zurich, https://doi.org/10.12686/a20

Number of OBM buildings with narrowed-down building classes

As explained in Section *2.5 Assigning building classes to OBM buildings*, all possible ESRM20 building classes associated with a specific location and occupancy case are first assigned to each OBM building, and classes that are incompatible with known information about the building are subsequently discarded. The current version of EHRE uses two pieces of data for this purpose:

- the number of storeys, which is used to narrow down building classes for residential, commercial and industrial buildings;
- the occupancy subtype of commercial buildings.

As building classes for residential and industrial buildings are only narrowed down based on the reported number of storeys, the overall percentage of residential and industrial buildings for which it was possible to narrow down the number of assigned building classes is quite low, given that only 7.24% of the buildings (of all occupancy types) in the OpenBuildingMap database used to create the EHRE model have an assigned number of storeys. When focusing just on residential and industrial occupancy cases, the percentage of buildings with known number of storeys are 7.61% and 5.95%, respectively. The final percentage of buildings for which it was possible to assign a reduced number of building classes is smaller than these values because (1) all possible building classes are assigned to a building (and a warning is logged) when all ESRM20 building classes are incompatible with the number of storeys of the OBM building, and (2) OBM statistics refer to individual building parts, while in the EHRE statistics building parts are already grouped if they are part of a relation.

The percentage of commercial buildings for which narrowing down was possible is larger, as it depends not only on the number of storeys but also, and more importantly, on the occupancy subtype. As the commercial occupancy case gathers buildings with occupancy types COM (no known subtype), COM1 (retail trade), COM2 (wholesale trade and storage), COM3 (offices, professional/technical services), COM5 (restaurants, bars, cafes), and RES3 (temporary lodging; e.g. hotels), all commercial OBM buildings except those classified only as COM can have their assigned building classes narrowed down. There are 2,263,793 building parts that fall under the commercial occupancy case in the OBM database used for EHRE, of which 732,935 (32.38%) are COM, 865,210 (38.22%) are COM1, 68,668 (3.03%) are COM2, 101,256 (4.47%) are COM3, 158,997 (7.02%) are COM5, and 336,727 (14.87%) are RES3, which means that around 67.61% of OBM commercial buildings can have their building classes narrowed down as a function of their subtype (hotels/restaurants, offices, trade). For each occupancy type within the commercial case, the number and percentage (with respect to its own type) of buildings with available number of storeys in the OBM database are 60,144 (8.21%) COM, 128,677 (14.87%) COM1, 15,346 (22.35%) COM2, 30,798 (30.42%) COM3, 15,262 (9.60%) COM5, and 158,634 (47.11%), which means that, overall, 18.06% of building parts that fall under the commercial occupancy case in the OBM database used for EHRE have number of storeys that can be used to narrow down the building classes.

Table 4 summarises the number of OBM buildings in the EHRE model with narrowed down and non-narrowed down number of assigned ESRM20 building classes, for all occupancy cases and countries, as well as the percentage of buildings with narrowed down classes. Though there is a clear variability across countries, the trends match those mentioned in the previous paragraphs. A large proportion of buildings for which the number of possible building classes has not been narrowed down implies the possibility of aggregating buildings into larger spatial units without loss of accuracy in the resulting damage and losses due to earthquake action, particularly since it is rare to have site properties/amplification models at the building level and the distance between different building centroids of the same block is too small to cause a large difference in the resulting ground motions. Table 4 helps distinguish countries and occupancy cases for which such aggregation is possible.

Table 4 Number of OBM buildings in the EHRE model with narrowed down (“yes”) and non-narrowed down (“no”) number of assigned ESRM20 building classes, and percentage buildings with narrowed down classes, per occupancy case and country.

Country	Residential			Commercial			Industrial		
	Yes	No	Percent.	Yes	No	Percent.	Yes	No	Percent.
Albania	6,982	162,521	4.12%	1,553	267	85.33%	270	1,654	14.03%
Andorra	68	3,559	1.87%	160	12	93.02%	0	49	0.00%
Austria	88,919	2,842,418	3.03%	26,164	21,356	55.06%	1,132	57,359	1.94%
Belgium	53,420	4,638,996	1.14%	22,461	9,538	70.19%	850	78,757	1.07%
Bosnia and Herzegovina	4,771	462,970	1.02%	2,570	1,901	57.48%	66	9,040	0.72%
Bulgaria	54,345	314,803	14.72%	6,397	3,797	62.75%	1,939	29,308	6.21%
Croatia	11,398	706,798	1.59%	6,940	4,164	62.50%	124	10,020	1.22%
Cyprus	7,588	62,287	10.86%	2,674	414	86.59%	19	1,985	0.95%
Czechia	2,005,028	1,859,360	51.88%	24,488	4,133	85.56%	15,642	130,974	10.67%
Denmark	57,584	2,681,084	2.10%	14,387	5,077	73.92%	855	64,499	1.31%
Estonia	10,288	346,661	2.88%	2,779	1,950	58.77%	278	14,938	1.83%
Finland	159,739	963,464	14.22%	67,791	6,316	91.48%	4,107	46,750	8.08%
France	348,689	33,716,309	1.02%	198,366	51,655	79.34%	4,745	646,764	0.73%
Germany	1,965,877	29,333,774	6.28%	256,498	369,855	40.95%	26,636	689,205	3.72%
Gibraltar	1	331	0.30%	73	48	60.33%	0	109	0.00%
Greece	36,866	570,257	6.07%	16,930	4,880	77.62%	663	15,285	4.16%
Hungary	51,316	1,481,999	3.35%	7,633	1,713	81.67%	858	50,250	1.68%
Iceland	4,736	73,166	6.08%	1,926	1,068	64.33%	168	3,883	4.15%
Ireland	229,106	1,013,198	18.44%	23,706	8,645	73.28%	897	23,022	3.75%
Isle of Man	109	14,141	0.76%	343	264	56.51%	0	374	0.00%
Italy	167,425	6,107,857	2.67%	96,302	21,310	81.88%	5,253	513,105	1.01%
Kosovo	3,565	103,993	3.31%	562	298	65.35%	22	1,852	1.17%
Latvia	22,163	253,119	8.05%	4,944	2,281	68.43%	1,782	18,422	8.82%
Liechtenstein	624	13,682	4.36%	252	154	62.07%	17	689	2.41%
Lithuania	18,190	1,249,701	1.43%	3,775	7,387	33.82%	388	41,409	0.93%
Luxembourg	5,988	145,024	3.97%	2,332	1,927	54.75%	34	3,026	1.11%
Malta	232	5,345	4.16%	358	35	91.09%	6	1,003	0.59%
Moldova	18,303	475,995	3.70%	3,624	1,711	67.93%	182	8,430	2.11%
Monaco	291	355	45.05%	81	15	84.38%	1	3	25.00%
Montenegro	850	74,668	1.13%	1,070	97	91.69%	14	804	1.71%
Netherlands	174,361	9,384,881	1.82%	25,001	10,987	69.47%	1,594	215,851	0.73%
North Macedonia	2,555	45,899	5.27%	1,004	152	86.85%	86	1,210	6.64%
Norway	181,759	2,685,329	6.34%	50,623	10,510	82.81%	540	45,592	1.17%
Poland	2,093,575	8,048,714	20.64%	218,090	42,241	83.77%	79,907	241,821	24.84%
Portugal	41,092	623,258	6.19%	23,969	4,419	84.43%	1,183	32,652	3.50%
Romania	72,324	1,135,647	5.99%	4,309	1,382	75.72%	0	42,845	0.00%
Serbia	16,240	723,011	2.20%	4,782	3,194	59.95%	665	12,972	4.88%
Slovakia	92,771	1,806,732	4.88%	8,820	3,621	70.89%	0	68,577	0.00%
Slovenia	9,802	366,318	2.61%	3,647	1,176	75.62%	66	7,969	0.82%
Spain	1,304,895	1,919,720	40.47%	56,717	11,578	83.05%	41,763	130,787	24.20%
Sweden	79,774	1,642,347	4.63%	22,835	7,180	76.08%	1,394	64,773	2.11%
Switzerland	91,750	1,706,496	5.10%	17,127	9,956	63.24%	1,346	55,018	2.39%
Türkiye	46,563	1,162,029	3.85%	523	31,055	1.66%	1,713	66,769	2.50%
United Kingdom	1,070,906	8,302,380	11.43%	269,923	63,294	81.01%	0	219,253	0.00%
Total	10,612,828	129,230,596	7.59%	1,504,509	733,013	67.24%	197,205	3,669,057	5.10%

Note: All ESRM20 industrial building classes for Andorra, Isle of Man, Romania, Slovakia and the United Kingdom have one storey, which results in the method used to calculate this table yielding 0% of industrial OBM buildings with narrowed-down building classes for these cases.

Example of EHRE for a city

In this section we show, as an example, the resulting EHRE model around the Italian town of Syracuse, on the south-eastern coast of the island of Sicily, as shown in Figure 17.

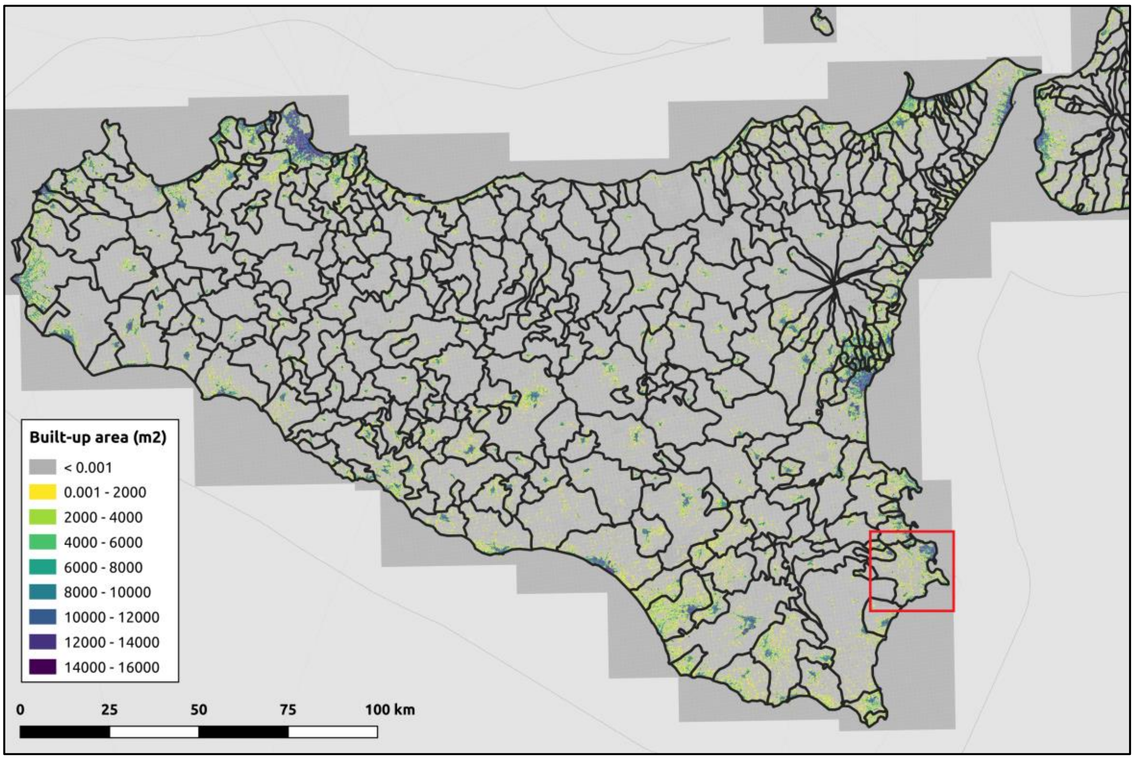


Figure 17 Built-up area estimated with the obmgapanalysis software for the island of Sicily, based on GHS. Town of Syracuse and surrounding areas marked in red rectangle.

While Figure 17 depicts the built-up area per tile as determined with the obmgapanalysis software for the whole island of Sicily, Figure 18 zooms in to the area around the town of Syracuse. The resulting completeness estimates of the tiles (calculated as per the procedure described in Section *2.3 Built-Up areas and completeness*) is shown in Figure 19.

The OBM buildings of the area, classified into the three occupancy cases of ESRM20 (residential, commercial, industrial) and “other” are depicted in Figure 20, while Figure 21 through Figure 23 show the buildings for each occupancy case separately, superimposed on the EHRE tiles, whose colour scales show the resulting number of remainder buildings, calculated as per the procedure described in Section *2.6 Calculating remainder buildings*. In Figure 21 and Figure 22, the solid black lines indicate the administrative level 3 boundaries for which ESRM20 defines residential and commercial exposure in Italy. In Figure 23, the solid black lines depict the 30-arcsec cells used by ESRM20 to define industrial exposure, and the filler industrial data units defined by ehre-importer to cover all the territory. As can be observed, several industrial OBM buildings fall within the filler units.

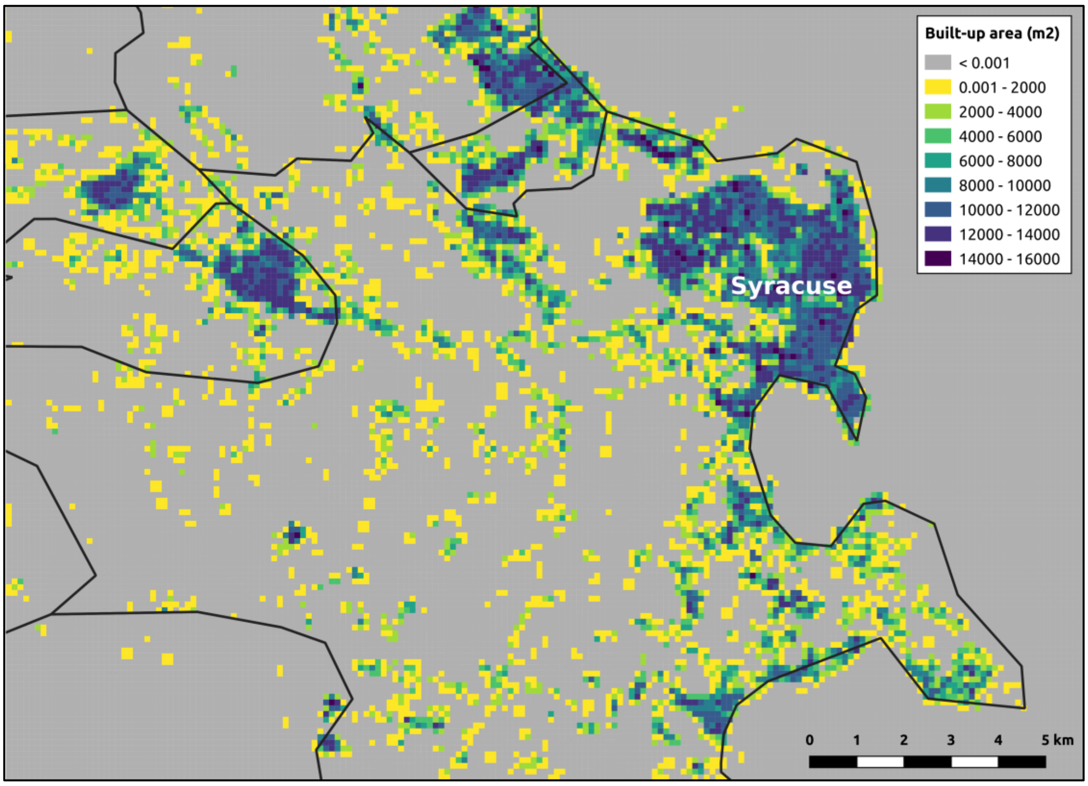


Figure 18 Built-up area estimated with the obmgapanalysis software for area around the town of Syracuse.

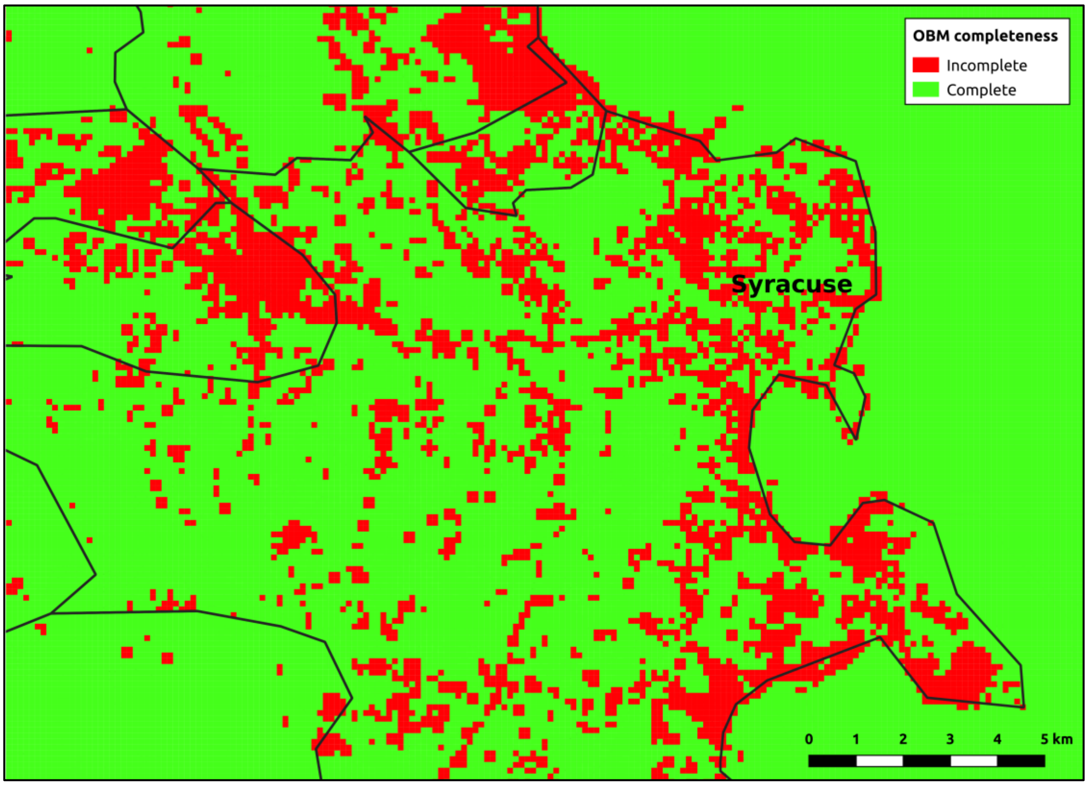


Figure 19 Completeness of OBM estimated with the obmgapanalysis software for area around the town of Syracuse.

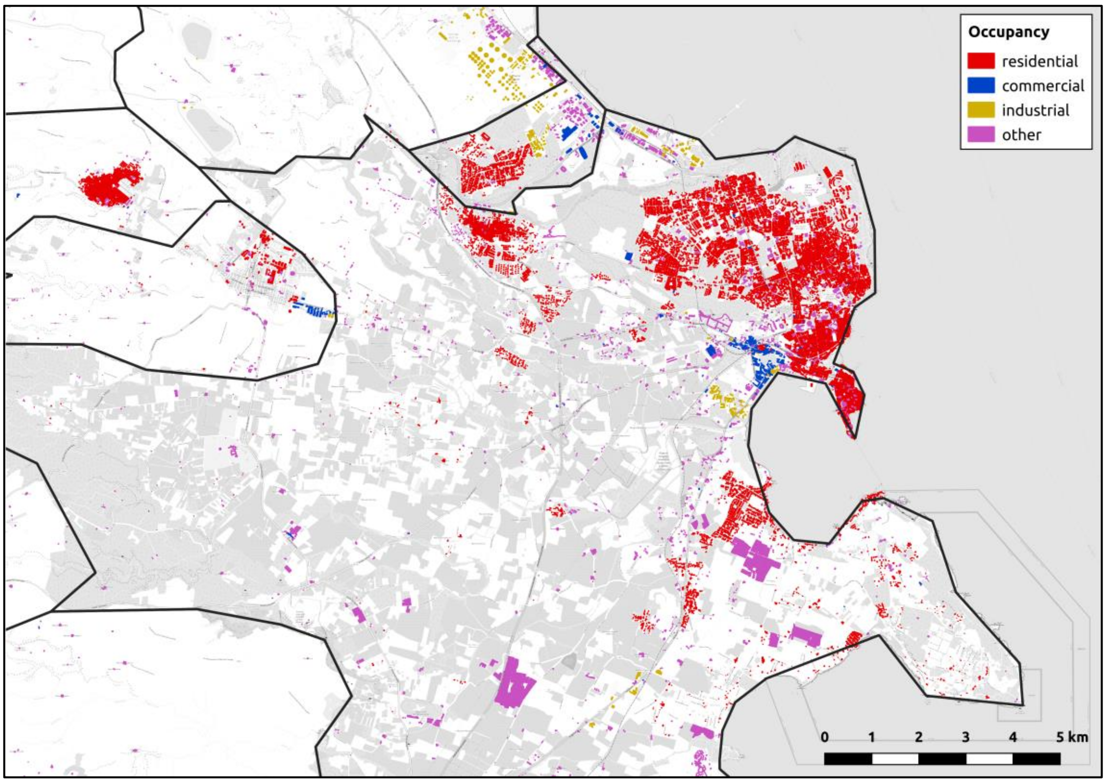


Figure 20 OBM buildings classified into residential, commercial and industrial occupancy cases compatible with ESRM20 in the area around the town of Syracuse. Background: OpenStreetMap.

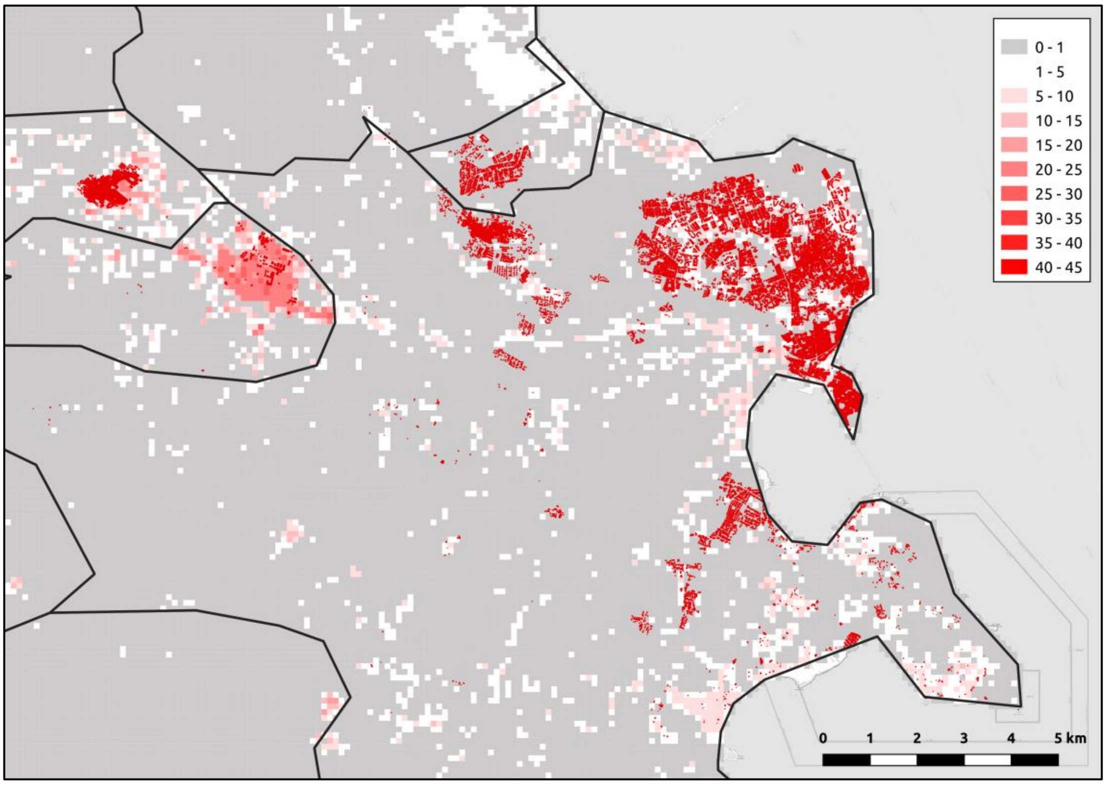


Figure 21 OBM residential building footprints over EHRE tiles with colour scale indicating number of remainder residential buildings in the area around the town of Syracuse. Thick black lines are ESRM20 administrative level 3 boundaries.

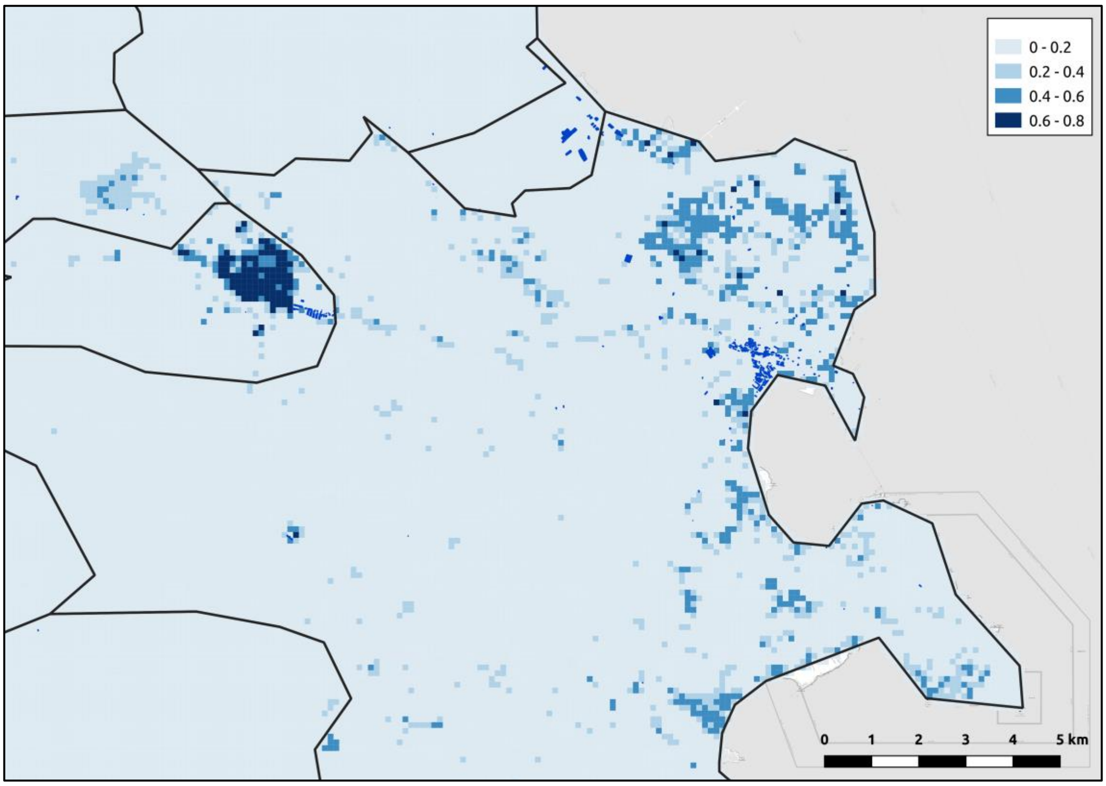


Figure 22 OBM commercial building footprints over EHRE tiles with colour scale indicating number of remainder commercial buildings in the area around the town of Syracuse. Thick black lines are ESRM20 administrative level 3 boundaries.

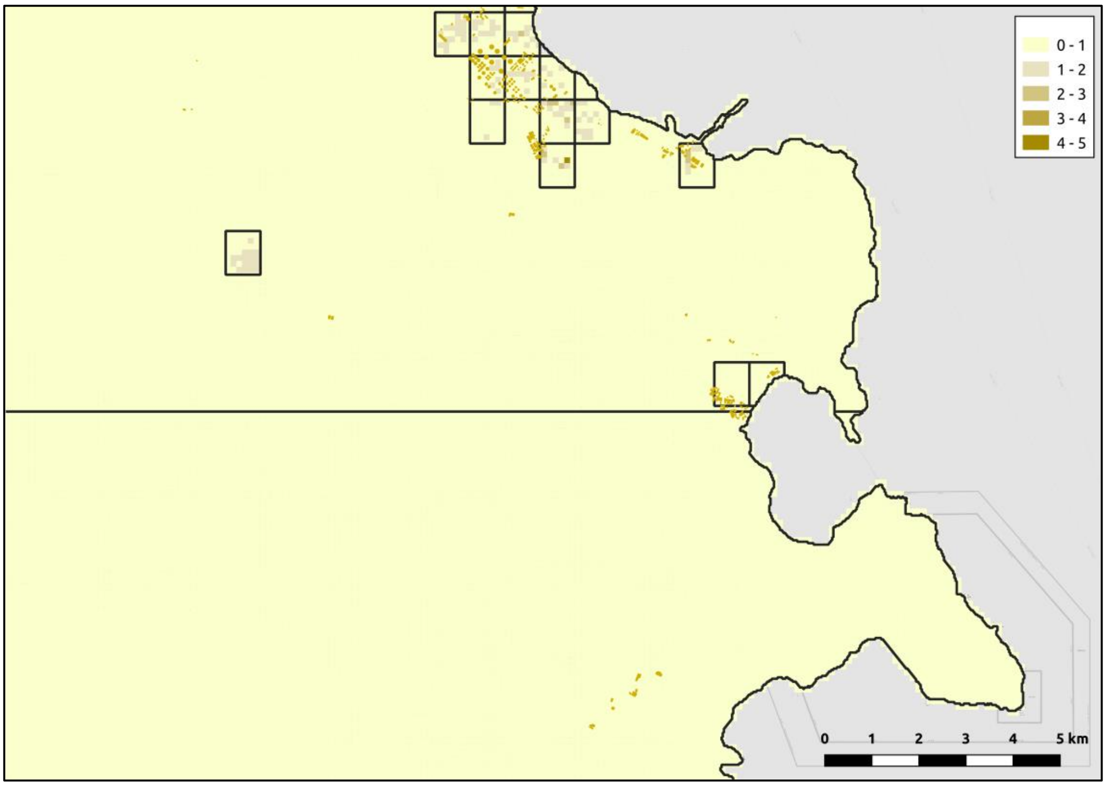
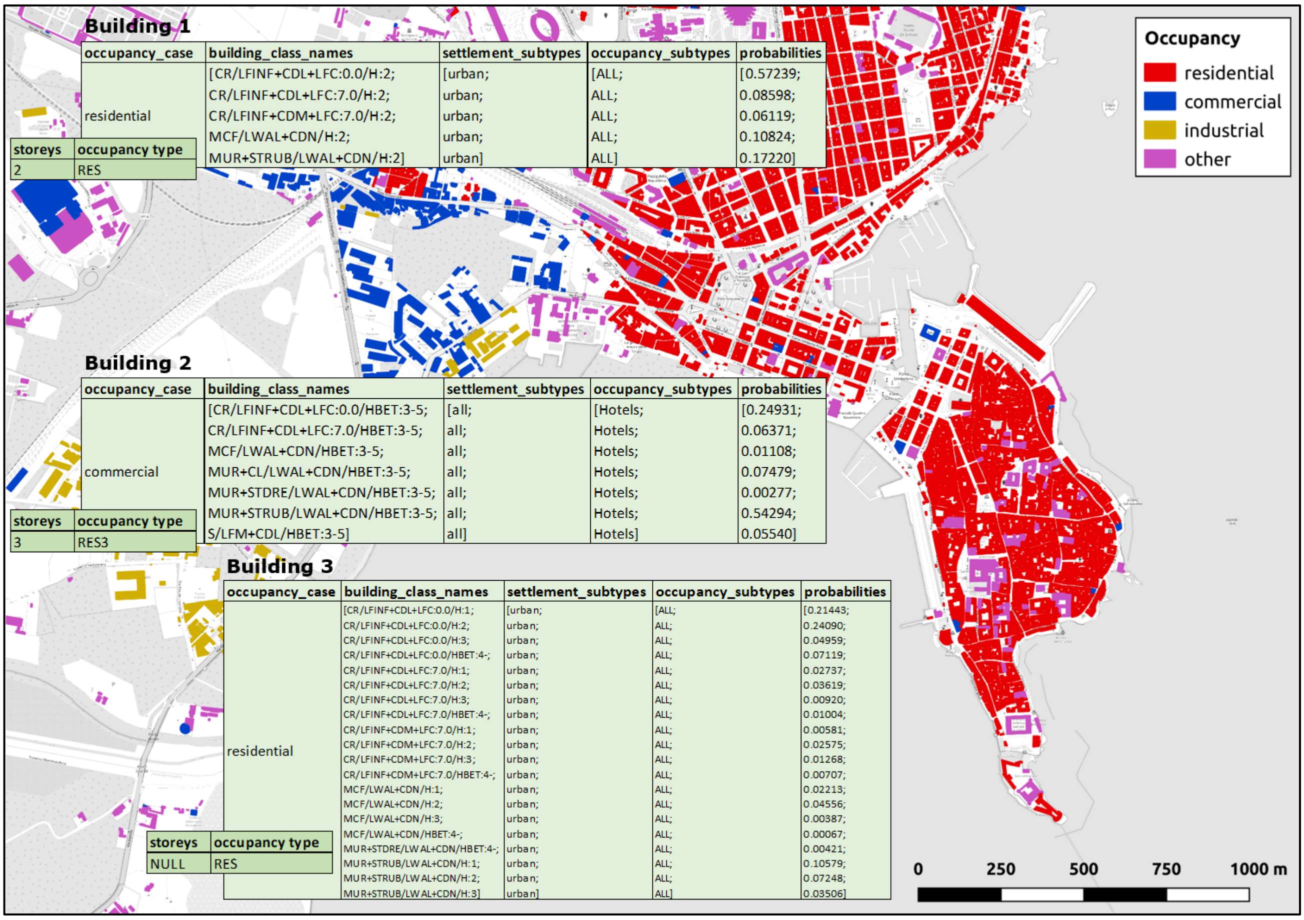


Figure 23 OBM industrial building footprints over EHRE tiles with colour scale indicating number of remainder industrial buildings in the area around the town of Syracuse. Thick black lines are ESRM20 industrial 30-arcsec cells and filler industrial data units defined by ehre-importer.

The map in Figure 24 zooms in the southern area of the city and shows the most relevant data associated with individual building footprints in OBM and EHRE, for three example buildings. Their exact location is not shown but all of them are within the area covered by the figure. Building 1 shows a case of a residential building with 2 storeys, for which this information has been used to narrow down the 20 possible building classes to a final set of five possibilities, while building 3 gets assigned all 20 possible building classes associated with this residential data unit because its number of storeys is not indicated in OSM (and, consequently, OBM). Building 2 is an example of a 3-storey commercial building whose occupancy type is RES3 (temporary lodging; e.g. hotels), to which seven building classes get assigned, narrowed down both by occupancy subtype ("Hotels") and number of storeys.



Building 1

occupancy_case	building_class_names	settlement_subtypes	occupancy_subtypes	probabilities
residential	[CR/LFINF+CDL+LFC:0.0/H:2;	[urban;	[ALL;	[0.57239;
	CR/LFINF+CDL+LFC:7.0/H:2;	urban;	ALL;	0.08598;
	CR/LFINF+CDM+LFC:7.0/H:2;	urban;	ALL;	0.06119;
	MCF/LWAL+CDN/H:2;	urban;	ALL;	0.10824;
	MUR+STRUB/LWAL+CDN/H:2]	urban]	ALL]	0.17220]

storeys	occupancy type
2	RES

Building 2

occupancy_case	building_class_names	settlement_subtypes	occupancy_subtypes	probabilities
commercial	[CR/LFINF+CDL+LFC:0.0/HBET:3-5;	[all;	[Hotels;	[0.24931;
	CR/LFINF+CDL+LFC:7.0/HBET:3-5;	all;	Hotels;	0.06371;
	MCF/LWAL+CDN/HBET:3-5;	all;	Hotels;	0.01108;
	MUR+CL/LWAL+CDN/HBET:3-5;	all;	Hotels;	0.07479;
	MUR+STDRE/LWAL+CDN/HBET:3-5;	all;	Hotels;	0.00277;
	MUR+STRUB/LWAL+CDN/HBET:3-5;	all;	Hotels;	0.54294;
	S/LFM+CDL/HBET:3-5]	all]	Hotels]	0.05540]

storeys	occupancy type
3	RES3

Building 3

occupancy_case	building_class_names	settlement_subtypes	occupancy_subtypes	probabilities
residential	[CR/LFINF+CDL+LFC:0.0/H:1;	[urban;	[ALL;	[0.21443;
	CR/LFINF+CDL+LFC:0.0/H:2;	urban;	ALL;	0.24090;
	CR/LFINF+CDL+LFC:0.0/H:3;	urban;	ALL;	0.04959;
	CR/LFINF+CDL+LFC:0.0/HBET:4-;	urban;	ALL;	0.07119;
	CR/LFINF+CDL+LFC:7.0/H:1;	urban;	ALL;	0.02737;
	CR/LFINF+CDL+LFC:7.0/H:2;	urban;	ALL;	0.03619;
	CR/LFINF+CDL+LFC:7.0/H:3;	urban;	ALL;	0.00920;
	CR/LFINF+CDL+LFC:7.0/HBET:4-;	urban;	ALL;	0.01004;
	CR/LFINF+CDM+LFC:7.0/H:1;	urban;	ALL;	0.00581;
	CR/LFINF+CDM+LFC:7.0/H:2;	urban;	ALL;	0.02575;
	CR/LFINF+CDM+LFC:7.0/H:3;	urban;	ALL;	0.01268;
	CR/LFINF+CDM+LFC:7.0/HBET:4-;	urban;	ALL;	0.00707;
	MCF/LWAL+CDN/H:1;	urban;	ALL;	0.02213;
	MCF/LWAL+CDN/H:2;	urban;	ALL;	0.04556;
	MCF/LWAL+CDN/H:3;	urban;	ALL;	0.00387;
	MCF/LWAL+CDN/HBET:4-;	urban;	ALL;	0.00067;
	MUR+STDRE/LWAL+CDN/HBET:4-;	urban;	ALL;	0.00421;
	MUR+STRUB/LWAL+CDN/H:1;	urban;	ALL;	0.10579;
	MUR+STRUB/LWAL+CDN/H:2;	urban;	ALL;	0.07248;
	MUR+STRUB/LWAL+CDN/H:3]	urban]	ALL]	0.03506]

storeys	occupancy type
NULL	RES

Figure 24 Example of most relevant information available at the building level from OpenBuildingMap (smaller tables) and EHRE (larger tables). These three buildings are located in the area shown in the map. Background: OpenStreetMap.

3. Conclusions and outlook for the future

The creation of the EHRE model has relied on the development of dedicated software as well as on a detailed understanding of the contents of the ESRM20 model it is based upon. The software is open-source and can be used by third parties to generate the model locally or as an inspiration for other endeavours that require similar architectures and tasks. Moreover, it can be built upon to further develop or improve different aspects of the model, as will be discussed in the following paragraphs.

The relatively small percentage of residential and industrial OBM buildings for which it was possible to narrow down the number of ESRM20 building classes assigned to them highlights the need to develop additional criteria that enable such reduction. A possibility would be to implement in ehre-core the mapping schemes of ESRM20, which connect variables available from census data, such as number of storeys, number and/or type of dwellings (e.g., single houses, detached/semidetached houses, row houses, apartment blocks), material of the structure, or year of construction with distributions of ESRM20 building classes. The availability of relevant parameters in OpenStreetMap and their incorporation to OpenBuildingMap (via the rabotnik-obm code) becomes of uttermost importance here, as the current use of the number of storeys within ehre-core shows that only a small percentage of buildings in OSM have such information, and most building attributes are not widely used (this observation is shared with Paprotny et al., 2020). Potential lines of communication with the OSM community for the definition and more systematic mapping of attributes that are useful for exposure modelling for risk assessment might be worth exploring.

When all possible ESRM20 building classes are assigned to an OBM building, the gain from increasing the resolution of the model by knowing the specific location of the footprint will likely be relatively small, as other components of the risk calculation, such as the site response model, are likely defined at a coarser resolution (e.g., ESRM20's at 30-arcsec, Weatherill et al., 2023), and estimates of ground motion do not necessarily vary largely within a few hundred metres. The impact of using the EHRE model in its full resolution or a coarser counterpart should be explored in the future by means of damage and loss calculations, similar to what has been done by Dabbeek et al. (2021) for ESRM20.

The final numbers of buildings per country and occupancy case reported in Section *2.8 Summary outputs and discussion* have highlighted the need for a detailed analysis of the output for each country, so as to be able to understand the sources of discrepancies between the EHRE model and other models/data (ESRM20 or original national census data) and potentially design strategies to tackle cases that might point at a negative impact of modelling decisions and/or data quality on the final product. As discussed, small percentage differences between models can be easily explained due to a variety of reasons (e.g., the passage of time), but large discrepancies, particularly with respect to official government data, require more in-depth investigations. While a general approach that can be applied at the continental level allows for the creation of large-scale models such as EHRE, local differences need to be explored when interpreting diverging results. Some aspects to consider are potential differences in the definition of what is considered a building (in OSM, in census data, in ESMR20), the accuracy of the automatic completeness calculation, including the appropriateness of the 0.25 built-up ratio as a threshold applicable to all countries (when derived only based on the region of Attica, Greece) and the determination of built-up areas in the first place, and the treatment of buildings that are part of a *relation* and the assignment of occupancy cases to OSM buildings in OpenBuildingMap. Regarding the latter, apart from benchmarking the current accuracy of rabotnik-obm in the classification of buildings into occupancy types, more sophisticated approaches such as that of Kunze and Hecht (2015) could be considered in the future.

In the meantime, potential uses of the EHRE model need to keep in mind that large discrepancies with respect to ESRM20 in the total number of buildings, particularly residential ones, results in the total number of inhabitants of a country not matching the official census data, which might be problematic when calculating human injuries and losses. Simple adjustments, such as a proportional reduction/amplification of the number of people per building to match the census total, could be considered.

Irrespective of the existence of discrepancies with respect to ESRM20 or not, the possibility of re-calculating the number of occupants and replacement cost per building based on the more fundamental parameters used in ESRM20 instead of the final ESRM20 values per building class could be evaluated. In the case of costs, for example, ESRM20 provides reconstruction costs per m^2 per country, occupancy case, occupancy sub-case in the case of commercial buildings (offices, trade, hotels/restaurants) and settlement type in the case of residential buildings (rural, urban, big city), which could be directly used in EHRE combined with the area of the building footprint from OpenBuildingMap, multiplied by the number of storeys, to obtain the expected replacement cost of OBM buildings. If not available in OSM, the number of storeys would come from the definition of the assigned building classes. Something similar can be done considering the average area (m^2) of residential construction per inhabitant. EHRE currently takes the number of occupants and replacement cost per building directly from ESRM20.

Being first full versions, the ehre-importer, ehre-core and ehre-exporter software have not been fully investigated for running efficiency, except during the development of some of their components. If these codes were to be used in the future for large-scale implementations of the model, they might benefit from benchmarking and optimisation. The running times reported in Appendix *8.6 Software running times* might be of use for this purpose. One aspect that might be particularly worth investigating is the trade-off in computational and storage demand between the case in which a large number of variables are stored in a compact way and final parameters are computed by the ehre-exporter once a desired output format is selected, as is the case now, and an alternative case in which several output formats might be pre-computed. The current version of the ehre-exporter code is particularly slow and, if it were to be impossible to optimise further, the latter alternative might be a better way forward.

4. Software and data accessibility

The whole software stack used to create EHRE is based on free and open-source software. All software that has been developed within the project is likewise published under AGPLv3+ license.

All software developed specifically for EHRE can be found under https://git.gfz-potsdam.de/ehre/ehre-software:

- **ehre-importer**: https://git.gfz-potsdam.de/ehre/ehre-software/ehre-importer.
- **ehre-core**: https://git.gfz-potsdam.de/ehre/ehre-software/ehre-core.
- **ehre-exporter**: https://git.gfz-potsdam.de/ehre/ehre-software/ehre-exporter.
- **EHRE** (deprecated) **prototype code** (still needed in part by ehre-importer to create the geometries of the 30-arcsec industrial cells of ESRM20): https://git.gfz-potsdam.de/ehre/ehre-software/prototype/ehre-prototype.

Further details on how to use the software, how to set up the configuration files, as well as how the calculations are being carried out, can be found within the *README* files and *docs* folders in each of the repositories enumerated above. Appendix *8.6 Software running times* contains the running times for ehre-importer and ehre-core for each country.

The EHRE model uses the following software developed for OpenBuildingMap (Schorlemmer et al., 2020, 2023):

- **rabotnik**: The version used for EHRE can be found in https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/rabotnik. The most up-to-date version can be obtained from the original repository: https://git.gfz-potsdam.de/globaldynamicexposure/rabotnik/rabotnik.
- **rcom**: The version used for EHRE can be found in https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/rcom. The most up-to-date version can be obtained from the original repository: https://git.gfz-potsdam.de/globaldynamicexposure/rabotnik/rcom.

- **rabotnik-obm**: The version used for EHRE can be found in https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/rabotnik-obm. The most up-to-date version can be obtained from the original repository: https://git.gfz-potsdam.de/globaldynamicexposure/rabotnik/rabotnik-obm.
- **obmgapanalysis**: https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/obmgapanalysis. The original repository has been archived under https://git.gfz-potsdam.de/globaldynamicexposure/openbuildingmap/obmgapanalysis.
- **quadtreegrid**: https://git.gfz-potsdam.de/ehre/obm-used-by-ehre/quadtreegrid. The original repository has been archived under https://git.gfz-potsdam.de/globaldynamicexposure/openbuildingmap/quadtreegrid.

Details on all database structures used by these software can be found here: https://git.gfz-potsdam.de/ehre/databases.

Details on external data sources used in EHRE can be found here: https://git.gfz-potsdam.de/ehre/datasources.

The whole database of EHRE and its associated static OBM database are large, and so are the files produced by ehre-exporter. The current version of EHRE does not include an API or service to access these data, but segments associated with small geographic areas can be made available upon demand.

5. Acknowledgements

Initial steps and key decisions of the European High-Resolution Exposure (EHRE) model benefited greatly from the contribution of Danijel Schorlemmer, for which we are thankful. We would also like to thank Laurens Oostwegel for running part of the code to process GHS built-up areas and determine OBM completeness, and Lars Lingner for his IT technical support.

We would like to express our gratitude to Helen Crowley, Jamal Dabbeek and Venetia Despotaki for providing access to early versions of the exposure model of the European Seismic Risk Model 2020 (ESRM20; Crowley et al., 2020, 2021) and for their support in understanding all its components.

This work was partially funded by:

- the Real-time Earthquake Risk Reduction for a Resilient Europe (RISE) project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 821115;
- the Large-scale EXecution for Industry and Society (LEXIS) project, which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 825532;
- the Airborne Observation of Critical Infrastructures (Luftgestützte Observation Kritischer Infrastrukturen, LOKI in German) project, which has received funding from the German Federal Ministry for Education and Research (BMBF) under funding code (FKZ) 03G0890D.

Maps created using QGIS. Background maps: OpenStreetMap.

6. Disclaimer

The geographical boundaries used for EHRE are adopted without modification from original source data on which it is based, and are intended for the purpose of scientific research. Neither GFZ nor the developers of the EHRE model endorse or take any specific view on territorial jurisdictions presented therein.

7. References

Corbane C, Florczyk A, Pesaresi M, Politis P, Syrris V (2018) GHS-BUILT R2018A - GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014) - OBSOLETE RELEASE. European Commission, Joint Research Centre (JRC) [Dataset] doi: 10.2905/jrc-ghsl-10007 PID: http://data.europa.eu/89h/jrc-ghsl-10007

Crowley H, Despotaki V, Rodrigues D, Silva V, Toma-Danila D, Riga E, Karatzetzou A, Fotopoulou S, Zugic Z, Sousa L, Ozcebe S, Gamba P (2020) Exposure model for European seismic risk assessment. Earthquake Spectra 36(1_suppl):252-273. https://doi.org/10.1177/8755293020919429.

Crowley H, Dabbeek J, Despotaki V, Rodrigues D, Martins L, Silva V, Romão, X, Pereira N, Weatherill G, Danciu L (2021) European Seismic Risk Model (ESRM20), EFEHR Technical Report 002, V1.0.1, 84 pp, https://doi.org/10.7414/EUC-EFEHR-TR002-ESRM20.

Dabbeek J, Crowley H, Silva V, Weatherill G, Paul N, Nievas C (2021) Impact of exposure spatial resolution on seismic loss estimates in regional portfolios. Bulletin of Earthquake Engineering 19, 5819–5841. https://doi.org/10.1007/s10518-021-01194-x.

Figueiredo R, Martina M (2016) Using open building data in the development of exposure data sets for catastrophe risk modelling. Natural Hazards and Earth System Sciences 16(2):417-429. https://doi.org/10.5194/nhess-16-417-2016

Geiß C, Pelizari PA, Marconcini M, Sengara W, Edwards M, Lakes T, Taubenböck H (2015) Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. ISPRS J Photogramm Remote Sens 104:175–188. https://doi.org/10.1016/j.isprsjprs.2014.07.016.

Geiß C, Schauß A, Riedlinger T, Dech S, Zelaya C, Guzmán N, Hube MA, Arsanjani JJ, Taubenböck H (2017) Joint use of remote sensing data and volunteered geographic information for exposure estimation: evidence from Valparaíso, Chile. Nat Hazard 86(1):81–105. https://doi.org/10.1007/s11069-016-2663-8.

Gomez-Zapata JC, Brinckmann N, Harig S, Zafrir R, Pittore M, Cotton F, Babeyko A (2021) Variable-resolution building exposure modelling for earthquake and tsunami scenario-based risk assessment: an application case in Lima, Peru. Natural Hazards and Earth System Sciences 21(11):3599-3628. https://doi.org/10.5194/nhess-21-3599-2021

Jaiswal KS, Wald DJ (2010) Development of a semi-empirical loss model within the USGS Prompt Assessment of Global Earthquakes for Response (PAGER) System. Proceedings of the 9th US and 10th Canadian Conference on Earthquake Engineering: Reaching Beyond Borders, July 25-29, Toronto, Canada.

Kunze C, Hecht R (2015) Semantic enrichment of building data with volunteered geographic information to improve mappings of dwelling units and population. Computers, Environment and Urban Systems, 53, 4-18, https://doi.org/10.1016/j.compenvurbsys.2015.04.002.

Liuzzi M, Pelizari PA, Geiß C, Masi A, Tramutoli V, Taubenböck H (2019) A transferable remote sensing approach to classify building structural types for seismic risk analyses: the case of Val d'Agri area (Italy). Bull Earthq Eng 17(9):4825–4853. https://doi.org/10.1007/s10518-019-00648-7.

Nievas, CI, Pilz M, Prehn K, Schorlemmer D, Weatherill G, Cotton F (2022) Calculating earthquake damage building by building: the case of the city of Cologne, Germany. Bulletin of Earthquake Engineering 20(3):1519-1565. https://doi.org/10.1007/s10518-021-01303-w.

Oostwegel LJN, Garcia Ospina N, Evaz Zadeh T, Shinde S, Schorlemmer D (2023) Automatic global building completeness assessment of OpenStreetMap using remote sensing data. EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-13160, https://doi.org/10.5194/egusphere-egu23-13160.

OpenStreetMap (OSM) https://www.openstreetmap.org

Pagani M, Monelli D, Weatherill G, Danciu L, Crowley H, Silva V, Henshaw P, Butler L, Nastasi M, Panzeri L, Simionato M, Vigano D (2014) OpenQuake engine: An open hazard (and risk) software for the global earthquake model. Seismological Research Letters 85(3):692-702. https://doi.org/10.1785/0220130087.

Paprotny D, Kreibich H, Morales-Nápoles O, Terefenko P, Schröter K (2020) Estimating exposure of residential assets to natural hazards in Europe using open data. Natural Hazards and Earth System Sciences 20(1):323-343. https://doi.org/10.5194/nhess-20-323-2020

Pelizari PA, Geiß C, Aguirre P, Santa María H, Peña YM, Taubenböck H (2021) Automated building characterization for seismic risk assessment using street-level imagery and deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 180, 370-386. https://doi.org/10.1016/j.isprsjprs.2021.07.004.

Pittore M, Haas M, Silva V (2020) Variable resolution probabilistic modeling of residential exposure and vulnerability for risk applications. Earthquake Spectra 36(1_suppl):321-344. https://doi.org/10.1177/8755293020951582

Schorlemmer D, Beutin T, Cotton F, Garcia Ospina N, Hirata N, Ma KF, Nievas C, Prehn K, Wyss M (2020) Global Dynamic Exposure and the OpenBuildingMap – A big-data and crowd-sourcing approach to exposure modeling. EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-18920. https://doi.org/10.5194/egusphere-egu2020-18920

Schorlemmer D, Cotton F, Delattre F, Evaz Zadeh T, Kriegerowski M, Lingner L, Oostwegel L, Prehn K, Shinde S (2023) An open, dynamic, high-resolution dynamic exposure model for Europe. RISE Project Deliverable 2.13. Available online: http://rise-eu.org/export/sites/rise/.galleries/Deliverables/Deliverable_2.13.pdf

Silva V, Crowley H, Pagani M, Monelli D, Pinho R (2014) Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment. Natural Hazards 72:1409-1427. https://doi.org/10.1007/s11069-013-0618-x

Silva V, Brzev S, Scawthorn C, Yepes C, Dabbeek J, Crowley H (2022) A building classification system for multi-hazard risk assessment. International Journal of Disaster Risk Science 13:161–177. https://www.doi.org/10.1007/s13753-022-00400-x.

Sousa L, Silva V, Bazzurro P (2017) Using open-access data in the development of exposure data sets of industrial buildings for earthquake risk modeling. Earthquake spectra 33(1):63-84. https://doi.org/10.1193/020316eqs027m.

Weatherill G, Crowley H, Roullé A, Tourlière B, Lemoine A, Hidalgo CG, Kotha SR, Cotton F (2023) Modelling site response at regional scale for the 2020 European Seismic Risk Model (ESRM20). Bulletin of Earthquake Engineering 21, 665–714. https://doi.org/10.1007/s10518-022-01526-5.

Wieland M, Pittore M, Parolai S, Begaliev U, Yasunov P, Tyagunov S, Moldobekov B, Saidiy S, Ilyasov I,Abakanov T (2015) A multiscale exposure model for seismic risk assessment in Central Asia. Seismol Res Lett 86(1):210–222. https://doi.org/10.1785/0220140130.

Yepes-Estrada C, Silva V, Valcárcel J, Acevedo AB, Tarque N, Hube MA, Coronel G, Santa María H (2017) Modeling the residential building inventory in South America for seismic risk assessment. Earthq Spectra 33(1):299–322. https://doi.org/10.1193/101915eqs155dp.

8. Appendices

8.1 Occupancy types of the GEM Building Taxonomy v3.0

For further details, please refer to the online glossary of the OpenQuake platform[21].

Table 5 Occupancy types defined in the GEM Building Taxonomy v3.0.

Occupancy	Description
COM	Commercial and public, unknown type
COM1	Retail trade
COM2	Wholesale trade and storage (warehouse)
COM3	Offices, professional/technical services
COM4	Hospital/medical clinic
COM5	Restaurants, bars, cafes
COM6	Public building (e.g. gallery, museum, monument building, library)
COM7	Covered parking garage
COM8	Bus station
COM9	Railway station
COM10	Airport
COM11	Recreation and leisure (sports and recreational facilities with fewer than 1000 spectators)
RES	Residential, unknown type
RES1	Residential, single dwelling
RES2	Residential, multi-unit, unknown type
RES2A	Residential, multi-unit, 2 Units (duplex)
RES2B	Residential, multi-unit, 3-4 Units
RES2C	Residential, multi-unit, 5-9 Units
RES2D	Residential, multi-unit, 10-19 Units
RES2E	Residential, multi-unit, 20-49 Units
RES2F	Residential, multi-unit, 50+ Units
RES3	Temporary lodging (e.g. hotels)
RES4	Institutional housing (group housing, including military, college, prisons, and nursing homes)
RES5	Mobile home
RES6	Informal housing
IND	Industrial, unknown type
IND1	Heavy industrial
IND2	Light industrial
MIX	Mixed, unknown type
MIX1	Mostly residential and commercial
MIX2	Mostly commercial and residential
MIX3	Mostly commercial and industrial
MIX4	Mostly residential and industrial
MIX5	Mostly industrial and commercial
MIX6	Mostly industrial and residential
AGR	Agriculture, unknown type
AGR1	Produce storage

---

[21] https://taxonomy.openquake.org/

Occupancy	Description
AGR2	Animal shelter
AGR3	Agricultural processing
ASS	Assembly, unknown type
ASS1	Religious gathering
ASS2	Arena
ASS3	Cinema or concert hall
ASS4	Other gatherings
GOV	Government, unknown type
GOV1	Government, general services
GOV2	Government, emergency response
EDU	Education, unknown type
EDU1	Pre-school facility
EDU2	School
EDU3	College/university, offices and/or classrooms
EDU4	College/university, research facilities and/or labs
OCO	Other occupancy type

8.2 Mapping of OSM tags into occupancy types in OBM

Table 6 shows the mapping of OSM tags of buildings and points of interest into GEM Building Taxonomy v3.0 occupancy types used by rabotnik-obm.

Table 6 OpenBuildingMap's mapping of tags from buildings and points of interest into occupancy types as per the GEM Building Taxonomy 3.0.

key	value	GEM_occupancy
aerialway	station	UNK
aeroway	hangar	COM10
aeroway	terminal	COM10
amenity	arts_centre	COM6
amenity	bank	COM3
amenity	bar	COM5
amenity	bbq	COM11
amenity	bicycle_rental	UNDECIDABLE
amenity	biergarten	COM5
amenity	bureau_de_change	COM1
amenity	bus_station	COM8
amenity	cafe	COM5
amenity	car_rental	COM1
amenity	childcare	EDU1
amenity	cinema	ASS3
amenity	clinic	COM4
amenity	college	EDU3
amenity	community_centre	ASS4
amenity	courthouse	GOV1
amenity	dentist	COM3
amenity	doctors	COM3
amenity	driving_school	COM1
amenity	embassy	GOV1
amenity	events_venue	ASS4
amenity	fast_food	COM5
amenity	ferry_terminal	UNK
amenity	fire_station	GOV2
amenity	food_court	COM5
amenity	fuel	UNDECIDABLE
amenity	grave_yard	ASS1
amenity	hospital	COM4
amenity	ice_cream	COM5
amenity	internet_cafe	COM1
amenity	kindergarten	EDU1
amenity	library	COM6
amenity	marketplace	COM1
amenity	mobile_money_agent	COM1
amenity	nightclub	COM11
amenity	nursing_home	RES4
amenity	parking	UNDECIDABLE
amenity	pharmacy	COM1
amenity	place_of_worship	ASS1
amenity	police	GOV2
amenity	post_office	COM3
amenity	prison	RES4
amenity	pub	COM5
amenity	public_building	GOV
amenity	recycling	IND

key	value	GEM_occupancy
amenity	restaurant	COM5
amenity	school	EDU2
amenity	shelter	UNDECIDABLE
amenity	social_facility	ASS4
amenity	studio	COM3
amenity	swimming_pool	COM11
amenity	taxi	UNDECIDABLE
amenity	telephone	UNDECIDABLE
amenity	theatre	ASS3
amenity	toilets	UNDECIDABLE
amenity	townhall	GOV1
amenity	university	EDU3
amenity	veterinary	COM3
building	allotment_house	RES1
building	apartments	RES2
building	barn	AGR1
building	bungalow	RES3
building	bunker	UNDECIDABLE
building	cabin	RES6
building	chapel	ASS1
building	church	ASS1
building	civic	GOV1|COM6|COM11
building	college	EDU3
building	commercial	COM
building	commercial;residential	MIX2
building	construction	UNDECIDABLE
building	cowshed	AGR2
building	detached	RES1
building	dormitory	RES4
building	farm	AGR
building	farm_auxiliary	AGR
building	garage	RES
building	garages	RES
building	government	GOV1
building	grandstand	COM11|ASS2
building	greenhouse	AGR3
building	hangar	COM10
building	hospital	COM4
building	hotel	RES3
building	house	RES1
building	hut	RES6
building	industrial	IND
building	kindergarten	EDU1
building	kiosk	COM1
building	manufacture	IND2
building	mosque	ASS1
building	office	COM3
building	parking	COM7
building	public	GOV
building	residential	RES
building	retail	COM1
building	school	EDU2
building	semi	RES2A
building	semidetached_house	RES2A
building	service	IND
building	shed	UNDECIDABLE
building	shop	COM
building	silo	AGR1
building	slurry_tank	AGR3
building	stable	AGR2
building	storage_tank	IND
building	sty	AGR2
building	supermarket	COM1
building	temple	ASS1
building	terrace	RES2
building	toilets	UNDECIDABLE
building	train_station	COM9
building	transformer_tower	UNDECIDABLE
building	transportation	COM8|COM9
building	trullo	RES1
building	university	EDU3|EDU4
building	warehouse	COM2
building	yes	UNDECIDABLE
building:type	apartment_building	RES2
building:type	apartments	RES2
building:type	barn	AGR1
building:type	dwelling_house	RES1
building:type	garage	RES
building:type	greenhouse	AGR3
building:type	house	RES1
building:type	residential	RES
building:type	semidetached_house	RES2A
building:use	commercial	COM
building:use	residential	RES
building:use	residential;industrial	MIX4
building:use	terrace	RES2
historic	monument	COM6
historic	castle	COM6
historic	ruins	COM6
landuse	allotments	RES
landuse	brownfield	UNDECIDABLE
landuse	cemetery	UNDECIDABLE
landuse	commercial	COM
landuse	farm	AGR
landuse	farmland	AGR
landuse	farmyard	AGR
landuse	forest	UNDECIDABLE
landuse	garages	RES
landuse	grass	UNDECIDABLE
landuse	grassland	UNDECIDABLE
landuse	landfill	IND
landuse	industrial	IND
landuse	meadow	AGR
landuse	mine	IND1
landuse	military	GOV
landuse	orchard	AGR

key	value	GEM_occupancy
landuse	park	UNDECIDABLE
landuse	plant_nursery	AGR3
landuse	quarry	IND1
landuse	railway	UNDECIDABLE
landuse	recreation_ground	COM11
landuse	reservoir	UNDECIDABLE
landuse	residential	RES
landuse	retail	COM1
landuse	village_green	UNDECIDABLE
landuse	vineyard	AGR
landuse	wood	UNDECIDABLE
landuse	winter_sports	UNDECIDABLE
leisure	common	ASS4
leisure	dog_park	UNDECIDABLE
leisure	escape_game	COM11
leisure	garden	UNDECIDABLE
leisure	golf_course	UNDECIDABLE
leisure	ice_rink	COM11|ASS2
leisure	hackerspace	COM11
leisure	marina	UNDECIDABLE
leisure	miniature_golf	UNDECIDABLE
leisure	nature_reserve	UNDECIDABLE
leisure	park	UNDECIDABLE
leisure	pitch	COM11
leisure	playground	UNDECIDABLE
leisure	sports_centre	COM11
leisure	stadium	ASS2
leisure	swimming_area	COM11
leisure	swimming_pool	COM11
leisure	water_park	COM11
man_made	storage_tank	IND
office	government	GOV1
office	company	COM3
office	yes	COM3
office	estate_agent	COM3
office	insurance	COM3
office	lawyer	COM3
office	educational_institution	EDU
office	telecommunication	COM3
office	association	COM3
office	ngo	COM3
office	diplomatic	GOV1
office	it	COM3
office	administrative	COM3
office	employment_agency	COM3
office	accountant	COM3
office	research	EDU4
office	religion	COM3|ASS1
office	architect	COM3
office	financial	COM3
office	tax_advisor	COM3
office	newspaper	COM3
office	advertising_agency	COM3
office	notary	COM3
office	political_party	COM3
office	logistics	COM3

key	value	GEM_occupancy
office	travel_agent	COM3
office	energy_supplier	COM3
office	therapist	COM3
office	foundation	COM3
office	physician	COM3
office	financial_advisor	COM3
office	consulting	COM3
office	water_utility	COM3
office	coworking	COM3
office	forestry	COM3
office	property_management	COM3
office	charity	COM3
shop	alcohol	COM1
shop	bakery	COM1
shop	beauty	COM1
shop	butcher	COM1
shop	car	COM1
shop	car_parts	COM1
shop	car_repair	COM1
shop	clothes	COM1
shop	convenience	COM1
shop	doityourself	COM1
shop	electronics	COM1
shop	florist	COM1
shop	furniture	COM1
shop	hairdresser	COM1
shop	hardware	COM1
shop	kiosk	COM1
shop	mall	COM1
shop	mobile_phone	COM1
shop	shoes	COM1
shop	supermarket	COM1
shop	yes	COM1
shop	variety_store	COM1
shop	optician	COM1
shop	jewelry	COM1
shop	gift	COM1
shop	greengrocer	COM1
shop	department_store	COM1
shop	books	COM1
shop	bicycle	COM1
shop	travel_agency	COM1
shop	chemist	COM1
shop	sports	COM1
shop	laundry	COM1
shop	confectionery	COM1
shop	stationery	COM1
shop	pet	COM1
shop	computer	COM1
shop	vacant	COM1
shop	tyres	COM1
shop	beverages	COM1
shop	newsagent	COM1
shop	dry_cleaning	COM1
shop	cosmetics	COM1
shop	motorcycle	COM1

key	value	GEM_occupancy
shop	garden_centre	COM1
shop	funeral_directors	COM1
shop	copyshop	COM1
shop	tailor	COM1
shop	tobacco	COM1
shop	toys	COM1
shop	farm	COM1
shop	deli	COM1
shop	interior_decoration	COM1
shop	seafood	COM1
shop	massage	COM1
shop	ticket	UNDECIDABLE
shop	storage_rental	COM2
shop	trade	COM1
shop	houseware	COM1
shop	photo	COM1
shop	pastry	COM1

key	value	GEM_occupancy
shop	wine	COM1
shop	outdoor	COM1
shop	paint	COM1
shop	general	COM1
shop	art	COM1
shop	bookmaker	COM1
shop	boutique	COM1
shop	charity	COM1
shop	pawnbroker	COM1
shop	second_hand	COM1
shop	fabric	COM1
shop	kitchen	COM1
shop	medical_supply	COM1
shop	tattoo	COM1
tourism	hotel	RES3
tourism	museum	COM6

Table 7 shows the mapping of OSM tags of land use polygons into GEM Building Taxonomy v3.0 occupancy types used by rabotnik-obm.

Table 7 OpenBuildingMap's mapping of tags from land use polygons into occupancy types as per the GEM Building Taxonomy v3.

key	value	GEM_occupancy
amenity	university	EDU3
amenity	school	EDU2
amenity	college	EDU3
amenity	library	COM6
amenity	fuel	UNDECIDABLE
amenity	parking	UNDECIDABLE
amenity	cinema	ASS3
amenity	theatre	ASS3
amenity	place_of_worship	ASS1
amenity	hospital	COM4
landuse	allotments	RES
landuse	animal_keeping	AGR2
landuse	brownfield	UNDECIDABLE
landuse	cemetery	UNDECIDABLE
landuse	commercial	COM
landuse	education	EDU
landuse	farm	AGR
landuse	farmland	AGR
landuse	farmyard	AGR
landuse	forest	UNDECIDABLE
landuse	garages	RES
landuse	government	GOV
landuse	grass	UNDECIDABLE
landuse	grassland	UNDECIDABLE
landuse	greenhouse	AGR3
landuse	greenhouse_horticulture	AGR3
landuse	industrial	IND
landuse	landfill	IND
landuse	leisure	COM11
landuse	meadow	AGR
landuse	mine	IND1
landuse	military	GOV
landuse	orchard	AGR
landuse	park	UNDECIDABLE
landuse	plant_nursery	AGR3
landuse	quarry	IND1
landuse	railway	UNDECIDABLE
landuse	recreation_ground	COM11
landuse	religious	ASS1
landuse	reservoir	UNDECIDABLE
landuse	residential	RES
landuse	retail	COM1
landuse	school	EDU2
landuse	village_green	UNDECIDABLE
landuse	vineyard	AGR
landuse	wood	UNDECIDABLE
landuse	winter_sports	UNDECIDABLE
leisure	common	ASS4
leisure	dog_park	UNDECIDABLE
leisure	escape_game	COM11
leisure	garden	UNDECIDABLE
leisure	golf_course	UNDECIDABLE
leisure	ice_rink	COM11
leisure	hackerspace	COM11
leisure	marina	UNDECIDABLE
leisure	miniature_golf	UNDECIDABLE
leisure	nature_reserve	UNDECIDABLE
leisure	park	UNDECIDABLE
leisure	pitch	COM11
leisure	playground	UNDECIDABLE
leisure	sports_centre	COM11
leisure	stadium	ASS2
leisure	swimming_area	COM11
leisure	swimming_pool	COM11
leisure	water_park	COM11
natural	beach	UNDECIDABLE
natural	bare_rock	UNDECIDABLE
natural	dune	UNDECIDABLE
natural	fell	UNDECIDABLE
natural	glacier	UNDECIDABLE
natural	grassland	UNDECIDABLE
natural	heath	UNDECIDABLE
natural	land	UNDECIDABLE
natural	mud	UNDECIDABLE
natural	scrub	UNDECIDABLE
natural	scree	UNDECIDABLE
natural	sand	UNDECIDABLE
natural	tundra	UNDECIDABLE
natural	wetland	UNDECIDABLE
natural	wood	UNDECIDABLE
place	neighbourhood	UNDECIDABLE
place	quarter	UNDECIDABLE
place	suburb	UNDECIDABLE
tourism	theme_park	COM11
tourism	zoo	AGR2|COM5|COM11
wetland	bog	UNDECIDABLE
wetland	swamp	UNDECIDABLE
wetland	wet_meadow	UNDECIDABLE
wetland	marsh	UNDECIDABLE
wetland	reedbed	UNDECIDABLE
wetland	saltern	UNDECIDABLE
wetland	tidalflat	UNDECIDABLE
wetland	saltmarsh	UNDECIDABLE
wetland	mangrove	UNDECIDABLE

8.3 Overriding occupancies

The first rule run by rabotnik-obm to decide on a final occupancy type for an OSM building based on all its tags is called *RuleOverridingOccupancy* and consists in searching for each occupancy string of a pre-selected hierarchical subset within the list of occupancy strings associated with the building and, if found, adopting it as the final occupancy. The pre-selected subset is called the set of "overriding occupancies" and is shown below in Table 8. The order in which the occupancy strings are named in the subset is hierarchical: the first string takes precedence over all other strings, the second takes precedence over all other strings except for the first one, etc.

Table 8 OpenBuildingMap's hierarchical list of overriding occupancies.

Occupancy	Description
COM10	airport
COM9	railway station
COM8	bus station
COM4	hospital/medical clinic
GOV2	government emergency response
GOV1	government general services
COM6	public building (gallery;museum;monument building;library)
ASS2	arena
EDU2	school
EDU3	offices and/or classrooms of college/university
EDU4	research facilities and/or labs of college/university
RES3	temporary lodging (hotels;motels;guest lodges;cabins)
ASS1	religious gathering
AGR1	produce storage
AGR2	animal shelter
AGR3	agricultural processing

8.4 Configuration file for Imposom (replication of the OSM database)

The rabotnik-obm software reads OSM from a local replication of the OSM database that is constantly updated by running the well-known OSM importer Imposom[22] with the configuration file whose contents are shown below:

```
areas:
  area_tags: [buildings, landuse, leisure, natural, aeroway]
  linear_tags: [highway, barrier]
tags:
  load_all: true
  exclude: [created_by, source, "tiger:*"]
tables:
  building_polygons:
    columns:
      - name: osm_id
        type: id
      - name: geometry
        type: validated_geometry
      - name: building
        key: building
```

---

[22] https://imposm.org/

```
      type: string
    - name: tags
      type: hstore_tags
  mapping:
    building:part:
      - __any__
    building:
      - __any__
    aeroway:
      - terminal
      - hangar
  filters:
    reject:
      building: ["no", "none", "No", "bridge", "pier", "road"]
      building:part: ["no", "none", "No"]
      man_made: ["bridge", "pier"]
  type: polygon
building_relations:
  columns:
    - name: osm_id
      type: id
    - name: geometry
      type: validated_geometry
    - name: building
      key: building
      type: string
      from_member: true
    - name: member
      type: member_id
    - name: index
      type: member_index
    - name: role
      type: member_role
      from_member: true
    - name: type
      type: member_type
    - name: tags
      type: hstore_tags
  mapping:
    type: [building]
  type: relation_member
lands:
  columns:
    - name: osm_id
      type: id
    - name: geometry
      type: validated_geometry
    - key: name
      name: name
      type: string
    - name: type
      type: mapping_value
    - name: key
      type: mapping_key
  filters:
    reject:
      building: [__any__]
  mapping:
    amenity:
      - university
      - school
      - college
      - library
      - fuel
      - parking
      - cinema
      - theatre
```

```
  - place_of_worship
  - hospital
landuse:
  - allotments
  - animal_keeping
  - brownfield
  - cemetery
  - commercial
  - education
  - farm
  - farmland
  - farmyard
  - forest
  - garages
  - government
  - grass
  - grassland
  - greenhouse
  - greenhouse_horticulture
  - landfill
  - leisure
  - industrial
  - meadow
  - mine
  - military
  - orchard
  - park
  - plant_nursery
  - quarry
  - railway
  - recreation_ground
  - religious
  - reservoir
  - residential
  - retail
  - school
  - village_green
  - vineyard
  - wood
  - winter_sports
leisure:
  - common
  - dog_park
  - escape_game
  - garden
  - golf_course
  - ice_rink
  - hackerspace
  - marina
  - miniature_golf
  - nature_reserve
  - park
  - pitch
  - playground
  - sports_centre
  - stadium
  - swimming_area
  - swimming_pool
  - water_park
natural:
  - beach
  - bare_rock
  - dune
  - fell
  - glacier
  - grassland
  - heath
```

```
      - land
      - mud
      - scrub
      - scree
      - sand
      - tundra
      - wetland
      - wood
    place:
      - neighbourhood
      - quarter
      - suburb
    tourism:
      - theme_park
      - zoo
    wetland:
      - bog
      - swamp
      - wet_meadow
      - marsh
      - reedbed
      - saltern
      - tidalflat
      - saltmarsh
      - mangrove
  type: polygon
roads:
  columns:
    - name: osm_id
      type: id
    - name: geometry
      type: geometry
    - name: type
      type: mapping_value
    - name: class
      type: mapping_key
  filters:
    reject:
      area: ["yes"]
  mappings:
    roads:
      mapping:
        highway:
          - motorway
          - motorway_link
          - trunk
          - trunk_link
          - primary
          - primary_link
          - secondary
          - secondary_link
          - tertiary
          - tertiary_link
          - road
          - path
          - track
          - service
          - footway
          - bridleway
          - cycleway
          - steps
          - pedestrian
          - living_street
          - unclassified
          - residential
          - raceway
          - bridleway
```

```
          - corridor
          - raceway
          - construction
        man_made:
          - pier
          - groyne
  type: linestring
spots:
  columns:
    - name: osm_id
      type: id
    - name: geometry
      type: geometry
    - key: name
      name: name
      type: string
    - name: type
      type: mapping_value
    - name: key
      type: mapping_key
  mapping:
    aerialway:
      - station
    amenity:
      - arts_centre
      - bank
      - bar
      - bbq
      - bicycle_rental
      - biergarten
      - bureau_de_change
      - bus_station
      - cafe
      - car_rental
      - childcare
      - cinema
      - clinic
      - college
      - community_centre
      - courthouse
      - dentist
      - doctors
      - driving_school
      - embassy
      - events_venue
      - fast_food
      - ferry_terminal
      - fire_station
      - food_court
      - fuel
      - grave_yard
      - hospital
      - ice_cream
      - internet_cafe
      - kindergarten
      - library
      - marketplace
      - mobile_money_agent
      - nightclub
      - nursing_home
      - parking
      - pharmacy
      - place_of_worship
      - police
      - post_office
      - prison
      - pub
```

```
    - public_building
    - recycling
    - restaurant
    - school
    - shelter
    - social_facility
    - studio
    - swimming_pool
    - taxi
    - telephone
    - theatre
    - toilets
    - townhall
    - university
    - veterinary
  building:
    - dormitory
  historic:
    - monument
    - castle
    - ruins
  landuse:
    - allotments
    - brownfield
    - cemetery
    - commercial
    - farm
    - farmland
    - farmyard
    - forest
    - garages
    - grass
    - grassland
    - landfill
    - industrial
    - meadow
    - mine
    - military
    - orchard
    - park
    - plant_nursery
    - quarry
    - railway
    - recreation_ground
    - reservoir
    - residential
    - retail
    - village_green
    - vineyard
    - wood
    - winter_sports
  leisure:
    - common
    - dog_park
    - escape_game
    - garden
    - golf_course
    - ice_rink
    - hackerspace
    - marina
    - miniature_golf
    - nature_reserve
    - park
    - pitch
    - playground
    - sports_centre
    - stadium
```

```
        - swimming_area
        - swimming_pool
        - water_park
      office:
        - __any__
      shop:
        - __any__
    type: point
```

8.5 EHRE Tiles database structure

The database that contains the EHRE model consists of the following tables:

- **aggregated_sources:** ID, name and format of input aggregated exposure models (ESRM20 herein), as shown in Table 9.

Table 9 Details of table aggregated_sources of the database.

Field	Data Type	Description
aggregated_source_id	smallint	Serial primary key.
name	varchar	Name of the source.
format	varchar	Format of the source files.

- **data_units:** Data-units are the smallest geographical unit where an aggregated exposure model is defined, i.e. where data is available for a particular occupancy case. This table basically contains data from ESRM20, as processed and interpreted by ehre-importer. Fields and details are presented in Table 10.

Table 10 Details of table data_units of the database.

Field	Data Type	Description
data_unit_id	varchar	Identifier of the data unit.
occupancy_case	occupancycase	Residential, commercial or industrial.
aggregated_source_id	smallint	Identifier of the source of the aggregated model.
exposure_entity	char(3)	Identifier of the exposure entity. If a country, ISO3 code.
buildings_total	float	Total number of buildings as per the aggregated exposure model.
dwellings_total	float	Total number of dwellings as per the aggregated exposure model.
people_census	float	Total number of census people as per the aggregated exposure model.
cost_total	float	Total replacement cost of buildings as per the aggregated exposure model.
geometry	geometry	Geometry of the data unit.

- **data_unit_tiles:** Data-unit tiles result from the intersection of data units with zoom level 18 quadtiles; they are the geographic unit in which the comparison between number of buildings resulting from distributing ESRM20 buildings onto tiles and OBM buildings take place. The contents of this table are shown in Table 11. The number of aggregated and remainder buildings is the total in the data unit tile (for a specific occupancy case and data unit ID); when multiplied by the proportions of building classes associated with the corresponding occupancy case and data unit ID (which are stored in the **data_units_buildings** table described below), the numbers of aggregated and remainder buildings per building class are obtained.

Table 11 Details of table data_unit_tiles of the database.

Field	Data Type	Description
quadkey	char(18)	Zoom-level 18 tile identifier.
aggregated_source_id	smallint	Identifier of the source of the aggregated model.
occupancy_case	occupancycase	Residential, commercial or industrial.
exposure_entity	char(3)	Identifier of the exposure entity. If a country, ISO3 code.
data_unit_id	varchar	Identifier of the data unit.
size_data_unit_tile_area	float	Area of the data-unit tile, in m^2.
size_data_unit_tile_built_up_area	float	Built-up area within the data-unit tile, in m^2.
fraction_data_unit_area	float	size_data_unit_tile_area / area of the data unit.
fraction_data_unit_built_up_area	float	size_data_unit_tile_built_up_area / built-up area of the data unit.
aggregated_buildings	float	Number of buildings from (aggregated_source_id, occupancy_case, data_unit_id) in the data-unit tile.
obm_buildings	smallint	Number of OBM buildings of occupancy_case in the data-unit tile.
remainder_buildings	float	Number of remainder buildings of occupancy_case in the data-unit tile.

- **data_units_buildings:** This table stores parameters associated with building classes in a particular data unit: the proportion in which the building class is present in the data unit, the number of census-derived people per building (i.e., not accounting for time of the day) and the total replacement cost per building (including structural and non-structural components, as well as contents), all as per ESRM20, processed and interpreted by ehre-importer. Each combination of the **building_class_name**, **settlement_type**, **occupancy_subtype**, **occupancy_case** and **data_unit_id** fields is a building class. Even though the number of storeys is included within the **building_class_name** (as per the GEM Building Taxonomy), the fields **storeys_min** and **storeys_max** are included to facilitate comparison against the number of storeys from individual OBM buildings. Table 12 contains a detailed explanation of all the fields of this table.

Table 12 Details of table data_units_buildings of the database.

Field	Data Type	Description
building_class_name	varchar	Building class as per the GEM Building Taxonomy v3.0.
settlement_type	settlement	Rural, urban, big city or all.
occupancy_subtype	varchar	Details on the occupancy, if relevant to characterise the building classes.
aggregated_source_id	smallint	Identifier of the source of the aggregated model.
occupancy_case	occupancycase	Residential, commercial or industrial.
exposure_entity	char(3)	Identifier of the exposure entity. If a country, ISO3 code.
data_unit_id	varchar	Identifier of the data unit.
proportions	float	Proportions in which this building class is present in the data unit.
census_people_per_building	float	Number of census-derived people per building.
total_cost_per_building	float	Total replacement cost per building.
storeys_min	smallint	Minimum number of storeys of the building class.
storeys_max	smallint	Maximum number of storeys of the building class.

- **exposure_entities_costs_assumptions:** This table stores the currency in which the replacement costs of buildings of a particular occupancy case in a particular exposure entity (country) of an aggregated exposure model (i.e., ESRM20) are given, as well as the factors by which to multiply total replacements costs from the table **data_units_buildings** (see above) to obtain the costs of structural components, non-structural components and building contents. In the current version of ESRM20 these factors are only dependent on the occupancy case and independent of the country, but this table envisages the possibility of a distinction by country in the future. All fields and details are presented in Table 13.

Table 13 Details of table exposure_entities_costs_assumptions of the database.

Field	Data Type	Description
exposure_entity	char(3)	Identifier of the exposure entity. If a country, ISO3 code.
occupancy_case	occupancycase	Residential, commercial or industrial.
aggregated_source_id	smallint	Identifier of the source of the aggregated model.
structural	float	Factor to obtain the cost of the structural components from the total cost.
non_structural	float	Factor to obtain the cost of the non-structural components from the total cost.
contents	float	Factor to obtain the cost of the building contents from the total cost.
currency	varchar	Currency in which the replacement costs are expressed.

- **exposure_entities_population_time_distribution:** This table stores the factors by which the census number of occupants per building from the table **data_units_buildings** (see above) can be multiplied to obtain an estimate of the people in the buildings at a certain time of the day. All fields and details are presented in Table 14. The factors used in EHRE are those of ESRM20, which are a modified version of the PAGER population distribution model (Jaiswal and Wald, 2010).

Table 14 Details of table exposure_entities_population_time_distribution of the database.

Field	Data Type	Description
exposure_entity	char(3)	Identifier of the exposure entity. If a country, ISO3 code.
occupancy_case	occupancycase	Residential, commercial or industrial.
aggregated_source_id	smallint	Identifier of the source of the aggregated model.
day	float	Factor to obtain the number of occupants during day time (approx. 10 am to 6 pm).
night	float	Factor to obtain the number of occupants during night time (approx. 10 pm to 6 am).
transit	float	Factor to obtain the number of occupants during transit time (approx. 6 am to 10 am and 6 pm to 10 pm).

- **ehre_buildings:** EHRE buildings result from the assignment of ESRM20 building classes to OpenBuildingMap buildings. For each OBM building (i.e., each entry in this table), **building_class_names**, **settlement_types**, **occupancy_subtypes** and **probabilities** are arrays of the same length, and the elements of one array correspond to those of the other. As in the **data_units_buildings** table described above, a building class is defined by the combination of the parameters **building_class_names**, **settlement_types**, **occupancy_subtypes**, **occupancy_case** and **data_unit_id**. The expected replacement cost and number of occupants of each individual OBM building listed herein can be calculated by retrieving the fields **total_cost_per_building** and **census_people_per_building** for the corresponding building classes from the table **data_units_buildings**, and combining them with the factors for different times of the day and for different cost categories stored in the tables **exposure_entities_population_time_distribution** and **exposure_entities_costs_assumptions** described above. Table 15 presents all fields and details for this table.

Table 15 Details of table ehre_buildings of the database.

Field	Data Type	Description
osm_id	integer	OpenStreetMap ID of the building (ID of the relation if building is presented by one).
aggregated_source_id	smallint	Identifier of the source of the aggregated model.
occupancy_case	occupancycase	Residential, commercial or industrial.
data_unit_id	varchar	Identifier of the data unit the OBM building belongs to.
quadkey	char(18)	Quadkey of the zoom-level 18 tile to which the centroid of the building belongs.
building_class_names	array of varchar	Array of building classes as per the GEM Building Taxonomy v3.0.
settlement_types	array of settlement	Array of rural, urban, big city or all.
occupancy_subtypes	array of varchar	Array of details on the occupancy, if relevant to characterise the building classes.
probabilities	array of float	Array of probabilities of the building belonging to each building class.
geometry	geometry	Footprint of the building.

The tables above make use of two personalised enumerated types defined within the database:

- **occupancycase**, which can be residential, commercial or industrial;
- **settlement**, which can be rural, urban, big city or all, and stems from ESRM20.

8.6 Software running times

Table 16 presents the time it took to run ehre-importer and ehre-core for each country and occupancy case using a server with 64 cores (3.2 GHz AMD Opteron 6386 SE, with a maximum of 48 cores allocated to running the code) and 1 TB RAM. It is noted that the server was not exclusively running ehre-importer and/or ehre-core the whole time, and so part of its capacity was used for different tasks at several points in time. Table 16 reports the observed running times, some of which may have been affected by the running of other tasks in the same server at the same time.

While in the case of ehre-importer the controlling factors are the overall size of the country in the EPSG:3857 projection (which defines the number of tiles it covers, see Table 1) and the resolution of the ESRM20 exposure model (number of data units and their respective sizes), in the case of ehre-core the number of OBM buildings also comes into play.

Table 16 Running times (hours) of the ehre-importer and ehre-core software per country and occupancy case.

Country	ehre-importer				ehre-core				Total
	Res	Com	Ind	Total	Res	Com	Ind	Total	
Albania	0.550	1.917	1.133	3.600	1.550	1.650	1.567	4.767	8.367
Andorra	0.050	0.050	0.050	0.150	0.117	0.092	0.088	0.296	0.446
Austria	10.600	10.600	12.783	33.983	14.100	9.933	19.150	43.183	77.167
Belgium	0.600	2.683	9.717	13.000	4.767	2.750	16.500	24.017	37.017
Bosnia and Herzegovina	1.000	1.017	2.617	4.633	1.467	1.200	3.683	6.350	10.983
Bulgaria	1.300	3.917	14.167	19.383	2.650	3.650	31.967	38.267	57.650
Croatia	1.233	1.333	5.667	8.233	2.600	2.217	6.400	11.217	19.450
Cyprus	0.183	0.683	1.217	2.083	1.500	0.650	0.983	3.133	5.217
Czechia	1.600	6.450	17.717	25.767	5.517	6.467	58.600	70.583	96.350
Denmark	1.917	1.700	6.567	10.183	5.600	3.233	20.083	28.917	39.100
Estonia	10.417	42.417	18.067	70.900	24.217	25.550	32.383	82.150	153.050
Finland	139.733	132.183	78.667	350.583	120.000	115.383	88.883	324.267	674.850
France	28.717	10.833	53.650	93.200	189.867	11.317	95.350	296.533	389.733
Germany	8.700	12.433	37.933	59.067	41.283	8.667	56.200	106.150	165.217
Gibraltar	0.003	0.003	0.003	0.008	0.139	0.139	0.139	0.417	0.425
Greece	N/A	N/A	N/A	N/A	1.883	1.800	6.117	9.800	N/A
Hungary	4.600	4.450	12.300	21.350	5.267	4.533	32.033	41.833	63.183
Iceland	13.217	34.867	12.500	60.583	11.633	41.450	11.517	64.600	125.183
Ireland	6.717	6.717	8.000	21.433	5.767	4.950	10.667	21.383	42.817
Isle of Man	0.267	0.267	0.050	0.583	0.317	0.283	0.217	0.817	1.400
Italy	5.000	5.750	38.217	48.967	16.500	13.833	49.117	79.450	128.417
Kosovo	0.467	0.467	0.467	1.400	0.783	0.617	0.633	2.033	3.433
Latvia	14.000	13.567	8.167	35.733	12.750	20.733	16.300	49.783	85.517
Liechtenstein	0.017	0.017	0.017	0.050	0.100	0.067	0.067	0.233	0.283
Lithuania	2.050	7.517	10.533	20.100	2.950	7.267	23.850	34.067	54.167
Luxembourg	0.250	0.183	0.833	1.267	0.400	0.283	1.283	1.967	3.233
Malta	0.033	0.050	0.067	0.150	0.108	0.079	0.288	0.475	0.625
Moldova	2.000	2.000	3.567	7.567	2.183	1.733	4.450	8.367	15.933
Monaco	0.002	0.002	0.002	0.005	0.083	0.050	0.050	0.183	0.188
Montenegro	0.833	0.833	0.867	2.533	1.000	0.883	1.167	3.050	5.583
Netherlands	2.600	3.950	12.633	19.183	5.367	3.417	27.933	36.717	55.900
North Macedonia	0.600	0.600	1.767	2.967	0.917	0.800	2.617	4.333	7.300
Norway	196.233	170.717	56.967	423.917	331.333	83.700	70.717	485.750	909.667
Poland	7.200	32.467	48.833	88.500	18.400	21.917	143.267	183.583	272.083
Portugal	1.933	2.567	10.417	14.917	14.850	15.250	13.350	43.450	58.367
Romania	4.283	5.300	18.767	28.350	13.083	4.867	34.367	52.317	80.667
Serbia	1.200	1.200	5.317	7.717	2.083	1.733	7.267	11.083	18.800
Slovakia	1.250	1.767	10.183	13.200	2.467	1.900	32.783	37.150	50.350
Slovenia	0.500	0.533	3.067	4.100	1.000	0.867	4.133	6.000	10.100
Spain	10.117	55.717	47.767	113.600	12.133	43.250	88.783	144.167	257.767
Sweden	40.100	40.450	96.200	176.750	46.550	43.950	145.417	235.917	412.667
Switzerland	1.917	4.567	6.450	12.933	6.683	3.583	15.217	25.483	38.417
Türkiye	21.283	21.217	20.500	63.000	17.633	16.250	17.700	51.583	114.583
United Kingdom	37.483	283.817	50.150	371.450	20.883	79.833	89.217	189.933	561.383
Total (hours)	582.8	929.8	744.6	2257.1	970.5	612.8	1282.5	2865.8	5113.0
Total (days)	24.3	38.7	31.0	94.0	40.4	25.5	53.4	119.4	213.0
Total (weeks)	3.5	5.5	4.4	13.4	5.8	3.6	7.6	17.1	30.4

ISSN 2190-7110