European Commission

JRC TECHNICAL REPORT

Comparative analysis of different methodologies and datasets for Energy Performance Labelling of buildings

ELISE Energy and Location Applications
Final Report

Martirano, G., Pignatelli, F., Vinci, F. (JRC)

Struck, C. (SAXION)

Coors, V., Fitzky, M. (HFT)

Hernández Moral, G., Serna-González, V., Ramos-Díez, I., Valmaseda, C. (CARTIF)

2022

ELISE

Enabling digital government through geospatial & location intelligence

Joint Research Centre

EUR 30963 EN

Contact information
Name: Francesco Pignatelli
Address: Via E. Fermi, 2749 – 21027 Ispra (VA), Italy
Email: Francesco.PIGNATELLI@ec.europa.eu
Tel.: +39. 033278-6319

EU Science Hub
https://ec.europa.eu/jrc

JRC124885

EUR 30963 EN

PDF ISBN 978-92-76-46608-6 ISSN 1831-9424 doi:10.2760/746342

Luxembourg: Publications Office of the European Union, 2022

How to cite this report: Martirano, G., Pignatelli, F., Vinci, F., Struck, C., Coors, V., Fitzky, M., Hernández Moral, G., Serna-González, V., Ramos-Díez, I., Valmaseda, C., *ELISE Energy and Location Applications*: *Use case "Comparative analysis of different methodologies and datasets for Energy Performance Labelling of buildings" – Final Report*, EUR 30963 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-46608-6, doi:10.2760/746342, JRC124885.

Contents

Acknowledgements

Special thanks to Marco Minghini and Lorena Hernandez Quiros of JRC/B.6 for their thorough review and valuable comments and suggestions.

Authors

Giacomo Martirano, Francesco Pignatelli, Fabio Vinci (JRC B.6, Ispra, IT)

Gema Hernández Moral, Victor Iván Serna-González, Iván Ramos-Díez, César Valmaseda (CARTIF Foundation, Valladolid, ES)

Volker Coors, Matthias Fitzky (Stuttgart University of Applied Sciences, Stuttgart, DE)

Christian Struck (Saxion University of Applied Sciences, Enschede, NL)

Abstract

According to studies carried out by European Commission Directorate-General for Energy (DG ENER), buildings are responsible for approximately 40% of the primary energy consumption in Europe. Therefore, there is a vital need to take actions to improve the energy efficiency of the building stock. Predictions of the heat demand at the building level, for an entire district or city, could provide valuable support to different stakeholders involved in the energy efficiency policy cycle.

However, these predictions are hampered by the lack of standardised calculation methodologies and interoperable building data to perform energy simulations. Another drawback is the low degree of comparability of the predictions. The latter has different causes: different calculation methodologies, diverse accuracy of building data, heterogeneous encoding of data and different ways of representing and visualising data.

Predictions of energy heat demand using the simulation software SimStadt have been produced, analysed and compared in four different case studies in three different Member States. The simulations were done with 3D building data of different accuracy and from different sources, which made it possible to identify significant causes of mismatch between simulations and real consumption scenarios. Several mapping exercises between the CityGML standard and the INSPIRE Directive data models have been documented to improve the interoperability of input and output datasets used in the simulations.

The conclusions drawn can support stakeholders involved in energy policy cycle aiming to assess the energy performance of their building stock in different geographical areas. A preliminary costs and benefits analysis of the assessment can be done re-using the methodology described in the report.

Five recommendations have been also formulated, suggested by the potential implications that the conclusions of the report may have on several policy-related discussions regarding the improvement of the energy efficiency of the building stock.

The reported activities have been executed in the frame of the Energy & Location Applications of the ELISE (European Location Interoperability Solutions for e-Government) action of the ISA2 (Interoperability solutions for public administrations, businesses and citizens) Programme.

Keywords

Energy efficiency, Location interoperability, energy performance of buildings, energy labels, energy heat demand, SimStadt, 3D building data, buildings, CityGML, ELISE action, Interoperability, Energy simulations

Executive Summary

According to studies carried out by European Commission Directorate-General for Energy (DG ENER)[1], buildings are responsible for approximately 40% of the primary energy consumption in Europe. Therefore, the need to improve the energy efficiency of the building stock is vital to support the European Green Deal[2] objectives while leveraging data-driven innovations and the opportunities that Digital Government Transformation can bring.

This publication addresses the energy efficiency challenge in the form of a use case named "Comparative analysis of different methodologies and datasets for Energy Performance Labelling of buildings ". This use case is part of the Energy & Location Applications activity carried out by the *European Location Interoperability Solutions for e-Government (ELISE)*, Action 10 of the ISA2 (Interoperability Solutions for Public Administrations, Business and Citizens) Programme, which aimed at making:

- a comparative analysis of different methodologies for Energy Performance Labelling of buildings applied to sample datasets of buildings of Germany (DE), the Netherlands (NL) and Spain (ES);
- the results of the comparative analysis reusable in other geographical areas by organisations aiming to assess the energy performance of the building stock and interested to preliminary assess costs & benefits of applying the same (or similar) methodologies based on the availability of datasets similar to those used in the comparative analysis.

The results presented in this report could be relevant to the nowadays EU policy context, because energy efficiency of buildings is one of the pillars of the European Green Deal[3] and, in particular, of the Renovation Wave strategy[4] and one of the seven flagship areas for investments and reforms[5] foreseen by the Recovery and Resilience Facility. The use case might also be relevant to the ongoing revision of the Energy Performance of Buildings Directive (EPBD)[6] part of the European Commission's "Fit for 55 package"[7].

In this regard, the use case could support the national long-term renovation strategies by assessing the energy performance of the current building stock. Besides, it could be useful to provide different future renovation scenarios through local predictions of the heat demand at building level for an entire district or city. These methodologies could provide valuable support to three different types of stakeholders involved in the energy efficiency policy cycle: Public Administrations involved in energy policymaking at regional/local level (i), businesses working in the sector of energy renovation of buildings, utility companies, Energy Service Companies (ESCOs) (ii), citizens acting as building/building unit owners/tenants and/or willing to sell/buy/rent/rent out a building/building unit (iii). All of the three types of stakeholders, aiming to assess, for different purposes, the energy performance of buildings in a specific area, can use the results of the analyses to preliminary estimate costs and benefits of similar predictions to be made in their regions/countries.

However, these predictions are affected by the lack of standardised calculation methodologies and harmonised and interoperable building data needed to perform energy simulations. The ultimate drawback is represented by the poor comparability of the predictions, caused by different calculation methodologies, input building data of different accuracy, heterogeneous encoding of input/output data and different ways of representing/visualising output data.

Predictions of energy heat demand using the simulation software SimStadt have been produced, analysed and compared across four different case studies in 3 different Member States, using 3D building data of different accuracy and provided by different sources, which made it also possible to identify the main sources of mismatch between simulations and real consumption scenarios. Moreoer, several mapping exercises between CityGML and INSPIRE data models have been documented, to improve the interoperability of input and/or output datasets used in the simulations.

A comparative analysis of the simulation results has been done, aiming at providing insight into the following aspects:

- identify the main obstacles to find and pre-process the input data required by the simulations, including the need to adapt the building physical library used by the simulation software to local contexts,

1 https://ec.europa.eu/jrc/en/energy-efficiency/buildings
2 https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en
3 https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en
4 https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/renovation-wave_en
5 https://ec.europa.eu/info/business-economy-euro/recovery-coronavirus/recovery-and-resilience-facility_en
6 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2018.156.01.0075.01.ENG
7 https://www.europarl.europa.eu/legislative-train/theme-a-european-green-deal/package-fit-for-55

- identify the main factors influencing the accuracy of the simulation results,
- estimate the influence of the accuracy of the CityGML LoD (Level of Detail) of the input data on the accuracy of the simulations results,
- identify the main sources of mismatch to be considered when comparing the simulation results with real energy consumption data.

For each of the above-listed aspects, the following main conclusions were drawn:

- Despite the availability of 3D city models as open data is increasing, information required by the energy simulations, such as building age, is often available only under restricted conditions.
- In the case of the simulations for the test area of Enschede (NL), the building physic library natively present in SimStadt and related to Germany has been successfully adapted to the Dutch building typologies, proving the viability of the adaptation.
- The preparation of the 3D building data as input data for the energy simulations requires software tools that require specific skills.
- A verification methodology to guide the interpretation of the results and their differences has been introduced.
- The improved accuracy of the simulation results depending on the better accuracy of the 3D building input data has not been demonstrated. Several comparisons between results obtained with LOD1 and LOD2 CityGML datasets have shown that some aspects of the building fabric are better considered using LOD1 datasets, e.g. the reduced over-estimation of the floor area.
- When comparing energy simulations with real energy consumption data, it is important to highlight that energy simulations do not consider user behaviours or possible energy efficiency interventions made on (parts of) the simulated buildings, which strongly impact energy consumption.
- When comparing the energy performance of buildings in different Member States, it is much better to compare absolute values expressed in $KWh/m^2/y$ rather than comparing the labels because the interval values the latter refers to are fixed by country-dependant national laws.
- Although all the simulations in this report have been made with the SimStadt software, in the Spanish case, the simulations have also been done using another software (ENERGIS). However, assessing the dependency of the simulation results on the simulation software would require additional investigations which are out of the scope of the work undertaken.

Finally, the following recommendations were formulated:

- **Recommendation 1**: 3D city models at different levels of detail, including information required by the energy simulations such as building age, should be made available as High-Value Datasets[8] and shared according to FAIR[9] principles, possibly within Energy Data Spaces[10].
- **Recommendation 2**: An EU common methodology to assess and document the quality, expressed in terms of different quality components (e.g. accuracy, completeness, up-to-date), of the input/output data used for the simulations of energy heat demand for building should be developed.
- **Recommendation 3**: Building physic libraries modelling the different building typologies in the different Member States should be developed adopting common semantics and shared under FAIR conditions.
- **Recommendation 4:** An EU common methodology to validate the results of the simulations of energy heat demand for buildings, obtained with different simulation software, should be developed.
- **Recommendation 5**: Adequate digital skills needed for an accurate assessment of the energy performance of the building stock should be formalised at the EU level, and the set-up of adequate education and training initiatives should be encouraged to fill in the related skill gaps.

[8] High Value Datasets defined in the DIRECTIVE (EU) 2019/1024 on open data and the re-use of public sector information (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024).

[9] Findability, Accessibility, Interoperability, Reusability. The FAIR Guiding Principles for scientific data management and stewardship (https://doi.org/10.1038/sdata.2016.18)

[10] Common European Data Spaces, as defined in the European Strategy for data (https://digital-strategy.ec.europa.eu/en/policies/strategy-data)

1 Introduction

The ELISE action Energy & Location Applications consist of a series of use cases aimed to show how location data can support different types of stakeholders engaged in energy policies' cycle at different geographical scales, from local up to EU level.

In particular, one of the use cases, named "Comparative analysis of different methodologies and datasets for Energy Performance Labelling of buildings", aimed:

- to make a comparative analysis of different methodologies for Energy Performance Labelling of buildings, applied to sample datasets of buildings of Germany (DE), the Netherlands (NL) and Spain (ES);
- to make the results of the comparative analysis re-usable in other geographical areas (Member States) by parties aiming to assess the energy performance labels of their building stock and interested to preliminary assess costs & benefits of applying the same (or similar) methodologies based on the availability of similar datasets, with respect to those used in the comparative analysis.

The problem addressed by the use case is that, according to DG ENER studies[11], buildings are responsible for approximately 40% of the primary energy consumption in Europe and there is a vital need to take actions to improve the energy efficiency of the building stock.

The use case is therefore relevant in the wider current EU policy context, because energy efficiency of buildings is one of the pillars of the European Green Deal[12] and, in particular, of the Renovation Wave strategy[13] and one of the seven flagship areas for investments and reforms[14] foreseen by the Recovery and Resilience Facility. In a more specific EU legislative context, the use case is relevant to the on-going revision of the Energy Performance of Buildings Directive (EPBD)[15], as part of the "Fit for 55 package"[16] of the European Commission, because it could support the national long-term renovation strategies, providing assessments of the energy performance of the current building stock as well as of different future renovation scenarios.

At local level, predictions of the heat demand at building level for an entire district or city could provide valuable support to different types of stakeholders involved in the energy efficiency policy cycle. These predictions are however affected by the lack of standardized calculation methodologies and of harmonized and interoperable building data needed to perform energy simulations.

The ultimate drawback is represented by the poor comparability of the predictions, caused by different calculation methodologies, input building data of different accuracy, heterogeneous encoding of input/output data and different ways to represent/visualize output data. One approach to tackle this issue is the use of building archetypes in building energy models. However, in their review paper, Reinhart et al (2016) [1] point out that building archetypes inherit a high source of uncertainty regarding how well they represent the building stock. On the other hand, the application of 3D building models to modelling urban energy systems and to simulate heating demand on city scale has made substantial progress in recent years. [2] - [5].

In this use case, predictions of energy heat demand using the simulation software SimStadt [6] have been produced, analysed and compared in 4 different case studies in 3 different Member States and with 3D building data of different accuracy and provided by different sources, allowing also to identify the main sources of mismatch between simulations and real consumption scenarios.

Different types of stakeholders, all aiming to assess, for different purposes, the energy performance of buildings in a specific area, can use the results of the analyses to preliminary estimate costs and benefits of similar predictions to be made in their regions/countries:

- Public Administrations involved in energy policy making at regional/local level,
- Businesses working in the sector of energy renovation of buildings, utility companies, Energy Service Companies (ESCOs),
- Citizens acting as building/building unit owners/tenants and/or willing to sell/buy/rent/rent out a building/building unit.

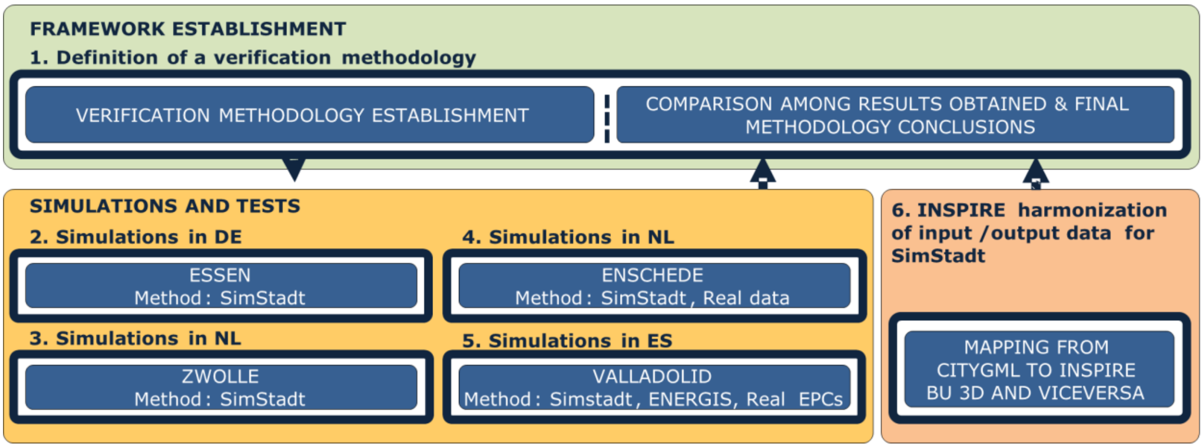
11 https://ec.europa.eu/jrc/en/energy-efficiency/buildings
12 https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en
13 https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/renovation-wave_en
14 https://ec.europa.eu/info/business-economy-euro/recovery-coronavirus/recovery-and-resilience-facility_en
15 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2018.156.01.0075.01.ENG
16 https://www.europarl.europa.eu/legislative-train/theme-a-european-green-deal/package-fit-for-55

The use case has been executed in 6 steps, shortly described below, represented in Figure 1 and separately addressed in the following sections.

1. Step 1: definition of a verification methodology, used to assess the results of the simulations made in the 4 different test areas (Section 2)
2. Step 2: simulations in DE, in the test area of Essen (Section 3)
3. Step 3: simulations in NL, in the test area 1 of Zwolle (Section 4)
4. Step 4: simulations in NL, in the test area 2 of Enschede (Section 5)
5. Step 5: simulations in ES, in the test area of Valladolid (Section 6)
6. Step 6: INSPIRE harmonisation of input/output data used in the simulations (Section 7).

Final conclusions, summarising the results achieved in each test area and describing the main achievements and lessons learnt, are elaborated in Section 8.

Figure 1. Use case approach



Source: own elaboration, CARTIF, 2020

2 Verification methodology

A verification approach, shown in Figure 2, has been developed to evaluate the prediction accuracy of simulations obtained from available geospatial data and using different simulation tools.

The framework consists of three main areas:

- Area 1 – detail of geometrical model representation;
- Area 2 – inter-model prediction accuracy;
- Area 3 – absolute prediction accuracy.

Area 1 addresses the impact of the geometrical level of detail on the relative prediction accuracy of the annual energy demand for heating. Knowledge about the impact of the chosen Level of Detail (LOD) on the prediction accuracy could reduce the costs for input data pre-processing before running the simulation. The use of input data with different level of detail and from different data sources is documented in the following sections, whilst more details on the level of detail concept, inherited from CityGML standard, are provided in the following sub-section 2.1.

Area 2 assesses the relative prediction accuracy among different simulation tools [7]. The SimStadt simulation tool has been used for all the test areas documented in this report and, for the only test area in Spain, the SimStadt predictions performance are compared with predictions obtained with another simulation tool (ENERGIS [8]). To the authors' best knowledge, publications about similar inter-model comparisons at district scale, are not yet available in the literature.

Area 3 is related to the comparison of SimStadt predictions performance with measured energy use data for the considered neighbourhood. Details of the comparisons made for the different test areas are provided in the related sections. For the test area in Spain an additional comparison has been made with data obtained from EPCs (Energy Performance Certificates).

Whilst the approach indicates the possibility of including the performance of individual buildings into verification exercises, no such verification has been made in the activities documented in this report.

2.1 CityGML and its levels of detail

CityGML is an open data model and XML-based format for the storage and exchange of virtual 3D city models. It is an application schema for the Geography Markup Language version 3.1.1 (GML3), the extendible international standard for spatial data exchange issued by the Open Geospatial Consortium (OGC) and the ISO TC211. The aim of the development of CityGML is to reach a common definition of the basic entities, attributes, and relations of a 3D city model.

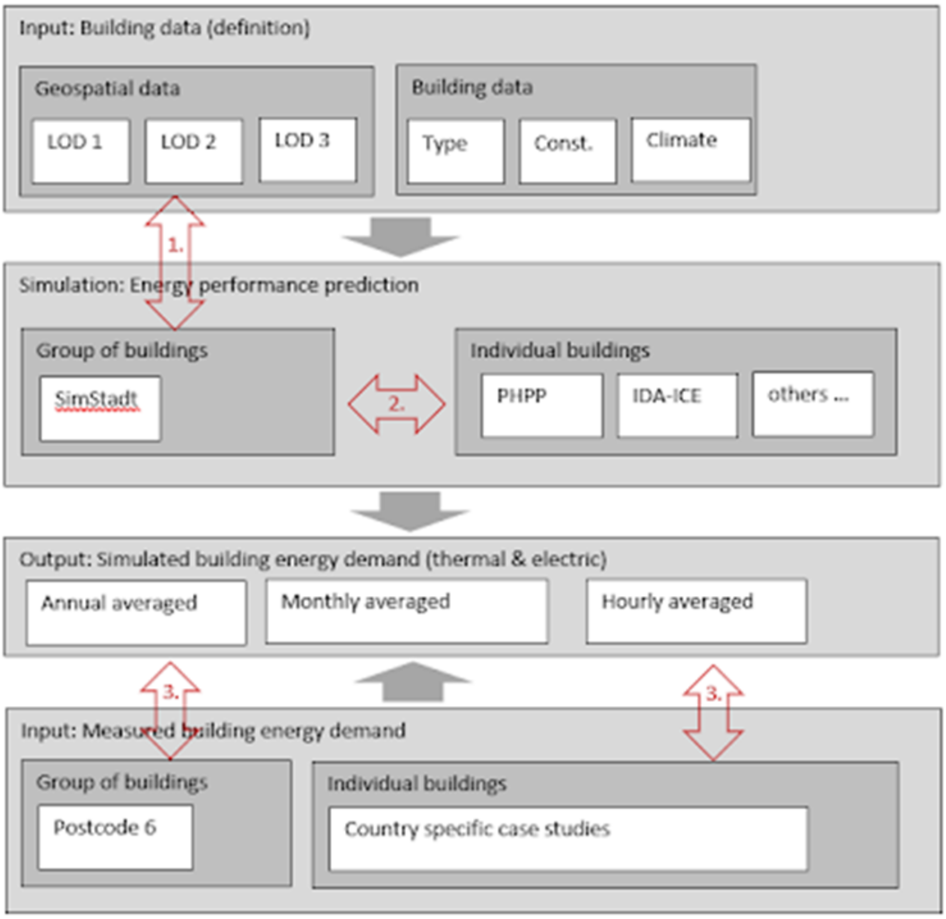
CityGML defines the classes and relations for the most relevant topographic objects in cities and regional models with respect to their geometrical, topological, semantical, and appearance properties[17].

- The building model is one of the most detailed thematic concepts of CityGML. It allows for the representation of thematic and spatial aspects of buildings and building parts in five levels of detail (LoD), from LOD0 to LOD4, shown in Figure 3: LoD 0 that offers a 2D model for buildings has been included in the latest version of City GML (v2.0).
- LoD 1 with block models (flat roofs)
- LoD 2 with the shape of roofs
- LoD 3 with accurate description of exterior (including openings: doors and windows)
- LoD 4: interior model

In this use case only the LoD1 and LoD2 datasets have been considered, because they are the two formats used by the SimStadt.

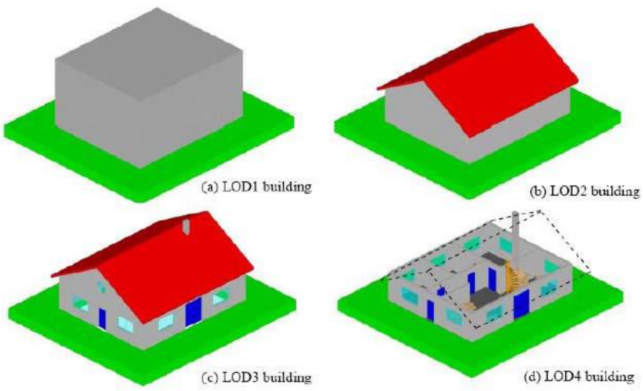
[17] https://www.citygmlwiki.org/index.php?title=Citygml_Wiki

Figure 2. Verification approach



Source: own elaboration, Saxion, 2020

Figure 3. The 4 CityGML level of details



Source: D2.8.III.2 INSPIRE Data Specification on Buildings – Technical Guidelines.

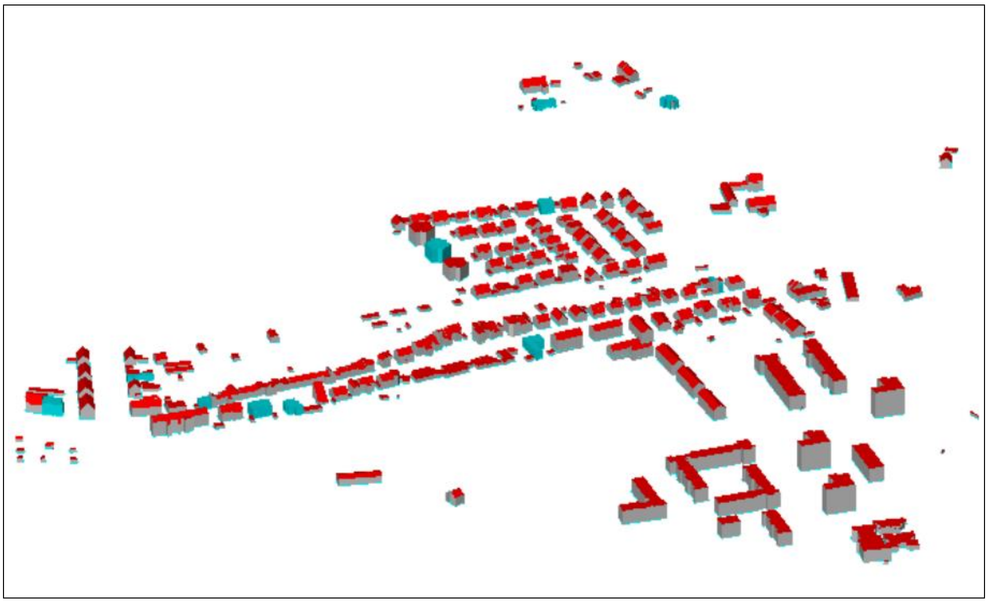
3 Simulations in DE (test area of Essen)

In this section the simulations made in the test area of Essen in Germany are described. Two different datasets, corresponding to LOD1 and LOD2 CityGML level of details, have been used as input data for energy heat demand simulations carried out with SimStadt simulation software, described in Section 3.2.

3.1 Input data

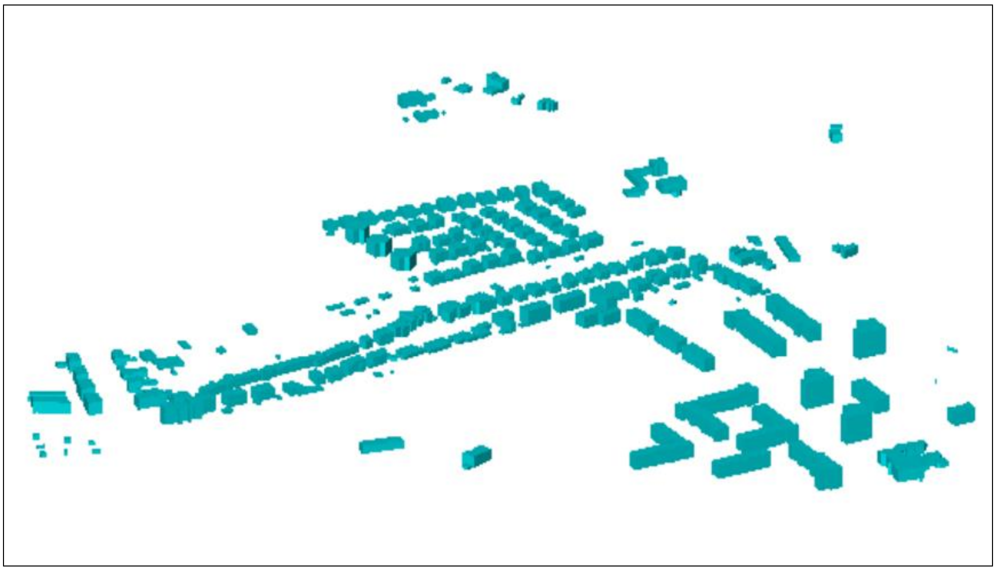
The dataset describing the city of Essen was available as a CityGML file in both LOD1 and LOD2. These two files have already been used in other research approaches like in the WeBest project, where the Essen dataset has been used in order to display in a WebGIS the SimStadt results, consisting of buildings styled in different colours according to their energy demand for heating. Thanks to the work done in the WeBest project, all the preliminary activities needed to check the CityGML compliance of the input data for SimStadt simulation were not needed for this use case. The two input datasets are shown in Figure 4 and Figure 5.

Figure 4. Essen LOD2



Source: Own elaboration, HFT, 2019

Figure 5. Essen LOD1



Source: Own elaboration, HFT, 2019

Each file contains for each building the attributes listed below:

— BezugspunktDach (ReferencePointRoof)
— DatenquelleBodenhoehe (DataSourceGroundLevel)
— Gemeindeschluessel (MunicipalKey)
— DatenquelleLage (DataSourcePosition)
— DatenquelleDachhoehe (DataSourceRoofHeight)
— **Usage (function of the buildings, as ALKIS Codes)**
— **YearOfConstruction**

Written in bold letters are the attributes which are most important for the workflow applied here. The other attributes are not relevant nor used for the processing of thermal behaviours of buildings using SimStadt.

Using a dataset with higher accuracy, it can be assumed that more realistic conditions are considered in the simulation. This increase in the LOD level leading to an approximation of real values has been investigated in this analysis. The purpose of using the dataset described above is to showcase the state of the art of simulating or estimating energy labels and consumption predictions for buildings.

In order to demonstrate how a higher accuracy of the geometries in the Essen dataset is reflected in the results of a heating energy demand simulation, a comparative analysis is carried out. Since two levels of detail are available describing the same buildings, the method incorporated in SimStadt can be evaluated looking into the potential increase of accuracy of the results when using more detailed input data. The two output datasets (estimating the energy heat demand) have been put into comparison with real consumption data per building.

The information of the age per building is not part of the open data model provided by the city of Essen and it has been provided separately for the purpose of this analysis.

The real energy consumption values used to compare the simulation results represent sensitive data if considered at the single building level. However, they can be aggregated for the purpose of this analysis and visualised as far as the values of single buildings can’t be traced back.

3.2 Simulation environment

SimStadt forms the foundation of the approach of the HFT Stuttgart to generate energy labels.

The software has been created with the goal to process data of the actual urban situation and future planning scenarios. Such scenarios include energy demand analysis of single buildings, city quarters, entire cities and regions. Further applications span from simulations of heating demand and photovoltaic potential analysis up to simulations for building refurbishment and renewable energy strategies.

The heating energy demand estimation is structured into 8 steps, described below.

1. **Import CityGML**

This step imports the CityGML file into the workflow and optionally checks if the file is valid against its declared schema definitions. Also, the number of buildings for each level of detail is counted.

2. **Create SimStadt Model**

The city model is analysed in terms of available information per building. This includes a check of already existing EnergyADE (Application Domain Extension) information in the input dataset. After that analysis, each attribute is extracted and stored in a so-called SimStadt-model, which simply enables the software to use all attributes in the following processing steps.

3. **Geometry Pre-processor**

Important geometrical attributes for the calculation of the final monthly energy balance are calculated. This includes the building’s volume, the amount of area covered by neighbouring buildings, the height and other attributes.

4. **Physics Pre-processor**

In this step, the connection to the Building-physics-library is done. Depending on the geometric attributes, a classification into a building type is made, and the predefined parameter set per type is added to each building according to the mandatory YOC attribute (YearOfConstruction) in the input dataset.

5. Usage Pre-processor

Here the assignment of usage related parameters to the building is done based on the usage attribute of each building. It is possible to assume residential usages if the function attribute is left empty in the input dataset. These parameters enrich each building energy model to enable the calculation of the monthly energy balance.

6. Weather Processor

The database of the external software "INSEL" [9] is accessed to be able to include the local outside temperature into the calculation. More specifically the temperature in the sky, on the ground and the irradiance levels (direct, global, and diffuse) in W/m^2 are specified by the information stored inside the database of INSEL.

7. Radiation Processor

For this step, different radiation models are considered to make assumptions on the sun radiation on each surface of the building. Depending on the chosen radiation model, the software is capable to include shadows and reflections coming from the different surfaces in the 3D city model into the final calculation of the energy demand. For each surface, the area, the tilt and the azimuth are evaluated and an irradiance value allocated.

8. Monthly Energy Balance

Here the information of the previous steps is gathered for the final output of an energy demand value at building scale. Depending on the settings chosen before, the simulation process is started, the output includes heating or cooling energy demand or a combined output file, where both calculation results are listed per building.

A repository is created inside the software SimStadt, where the CityGML dataset to be simulated is stored. A new project is also created and the heating energy demand workflow selected. Then clicking on the button "Run" the simulation process starts. Since SimStadt is a modular software where each workflow can be put together using pre-made (or self-made) workflow steps, the results for each step can be shown. These include the distribution of the function attributes in the input dataset, the visualization of the buildings according to their heat transfer coefficient, the visualization according to the year of construction, etc.

3.3 Methodology

As already explained in Section 2, because the introduction of a uniform energy labelling approach for the European building stock brings a series of issues and limitations, the simulation results are also expressed in absolute values. For the generation of the labels, a labelling method based on a classification of energy values also known as the German "Energieausweis" (Energy Certificate) has been applied. The output of an energy simulation obtained with SimStadt is mapped to labels using a simple program written in Java. The column for the specific heating demand, which describes the total yearly heating demand per m^2 of the buildings heated area is considered. This value is given in kWh/m^2.

A geometrical and semantic pre-processing of the CityGML datasets which contain the attributes and geometries of each building is not necessary since the files have been prepared for and used in a previous project (WeBest). Full SimStadt compliancy is therefore already given. The attributes "Usage" (given in the tag "gml:function") and the year of Construction (given as "gml:yearOfConstruction") are declared in a non-generic way according to the CityGML 2.0 schema clarification and are therefore readable by SimStadt. The values of the actual energy demand are given at building scale.

The actual energy consumption data source bears the problem that the heating energy demand hasn't been indicated per m^2. Only the total energy consumption per year per building is provided. Aiming at generating energy labels, which are basically estimated based on a value per m^2 in the unit kWh/m^2/year, the real consumption values are modified by dividing the total energy consumption of a building by the heated area of a building in LOD2 as calculated with SimStadt. Of course, this approach introduces several data uncertainties in the results used for the comparison between LOD1, LOD2 and the actual energy consumption. This is further explained in Section 3.4.3.

3.4 Results

3.4.1 Comparison between LOD1 and LOD2

Figure 6 and Figure 7 have been created by rounding the heating energy demand, which has been calculated per m^2 heated area. The rounded values are then allocated into bins having the size of 10 kW/h. The range from

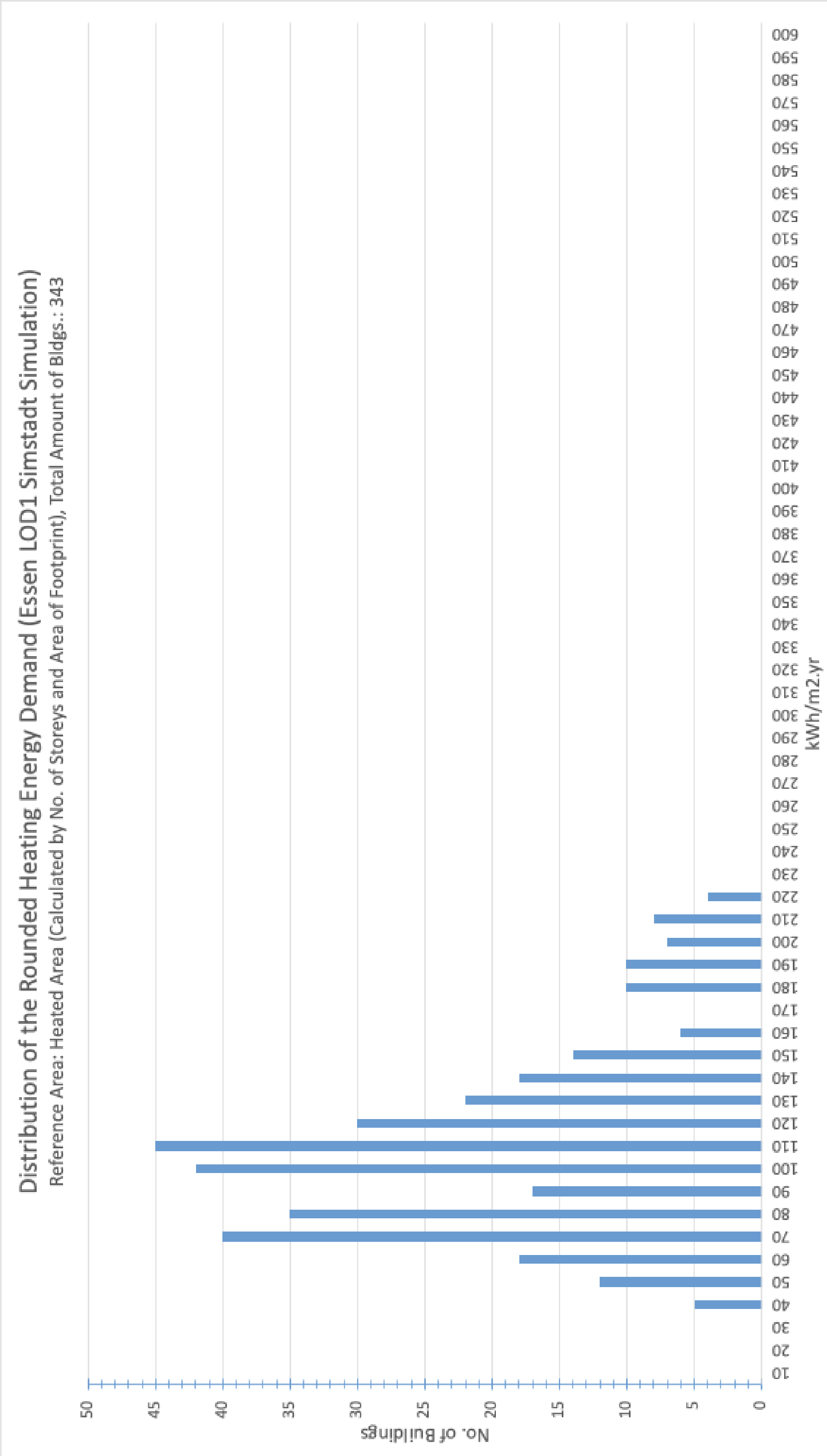
0 to 600 kW/h has been chosen to comply with the range of values in the results using the Zwolle dataset (see Section 4.3).

Whit reference to the Figure 6 and to the Figure 7, it can be observed that the estimated energy demand values in LOD1 are in general lower than those calculated with a LOD2 dataset.

This kind of result is in line with the results of the study “Comparison of building modelling assumptions and methods for urban scale heat demand forecasting, Future Cities and Environment” [10] (see Figure 3 in the paper), where in most cases the LOD2 values are higher than LOD1 values. This circumstance can also be observed in Figure 6 and Figure 7.

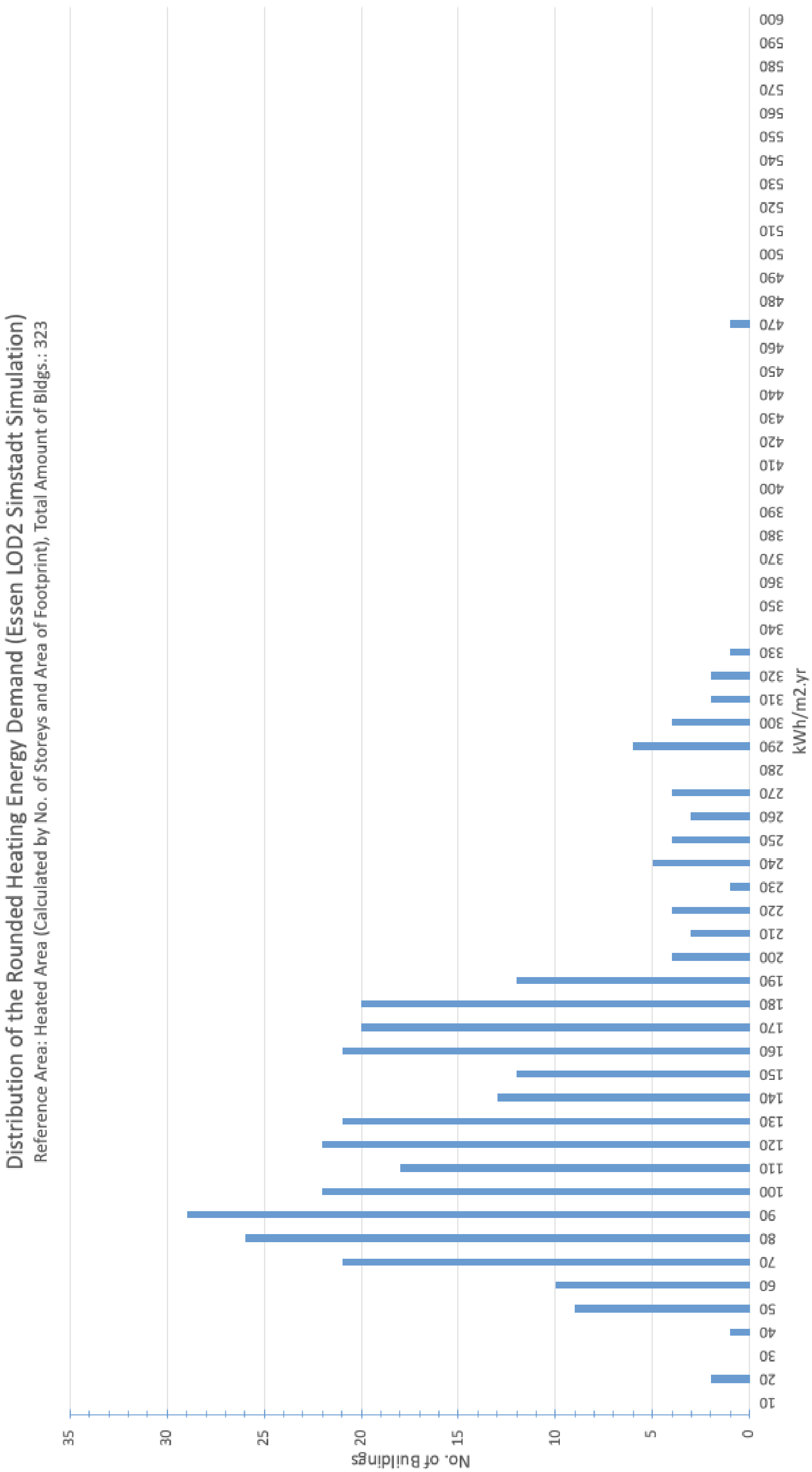
The reason for that is that LOD2 buildings normally have more outer surface areas where heated air can be transmitted to the environment, and therefore need more energy to heat up the interior.

Figure 6. Rounded simulated heating energy demand using the Essen LOD1 Citymodel



Source: Own elaboration, HFT, 2019

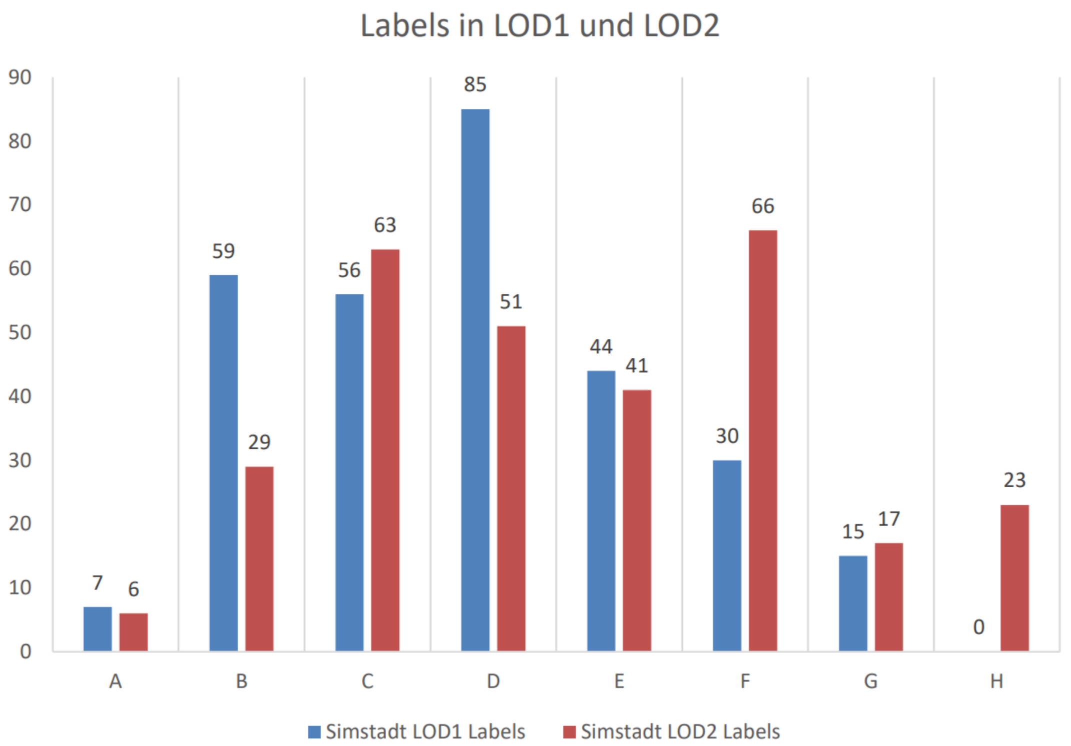
Figure 7. Rounded Heating Energy Demand using Essen LOD2 Citymodel



Source: Own elaboration, HFT, 2019

Figure 8 has been created by generating the labels using the “Energieausweis” method. In total 296 matching buildings occur in both LOD1 and LOD2 datasets. A better matching of the number of labels can be observed for the labels A, E and G.

Figure 8. Comparison of labels estimated when running the simulation with LOD1 and LOD2 datasets



Source: Own elaboration, HFT, 2019

Figure 9 shows absolute values of several example buildings. What can be observed is that the specific heating demand, from which the labels are derived, is always higher in LOD2 than LOD1, whilst, looking at the total heating energy demand per year, the LOD1 values are again higher.

The critical factor here is the heated area, which, in a very simplified form, is derived by the area of the building footprints multiplied by the storey number of each building. Because LOD1 building footprints are less detailed than LOD2 building footprints and have a general tendency to have a larger area than in the LOD2 case, it can be observed in most cases that the heated area for LOD1 buildings is larger than LOD2 buildings. The larger heated area in LOD1 buildings causes, in turn, a higher value of the total yearly heating energy demand when compared to LOD2 buildings.

Looking at the specific heat demand, LOD2 values are in general higher than LOD1 values, because of the larger area of the outward-faced surface areas for LOD2 buildings where air can transmit through, so more energy must be provided to keep the interior at a comfortable temperature.

However, despite the LOD2 higher specific heat demand, the LOD1 higher heated area causes the LOD1 higher total yearly heat demand.

Figure 9. Absolute value comparison for some example Buildings

Gebäude ID	SIMSTADT LOD1	SIMSTADT LOD2	LOD1 Specific Heat Demand [kWh/m².yr]	LOD2 Specific Heat Demand [kWh/m².yr]	Abweichung Specific Heat Demand LOD1 zu LOD2	Heated Area LOD1 [m²]	Heated Area LOD2 [m²]	Totaly Yearly Heat demand LOD1 [kWh/yr]	Total Yearly Heat Demand LOD2 [kWh/yr]
DENW22AL50000rac	E	F	144,8	173,6	-28,8	121,3	86,9	19484	16465
DENW22AL50000qaj	B	B	68,6	72,2	-3,6	700	642,5	59087	56580
DENW22AL50000veK	C	E	98,6	154	-55,4	68	33,8	7787	5733
DENW22AL50000ubO	E	F	141,8	184,6	-42,8	291,8	195,1	46004	39098
DENW22AL50000vSU	C	D	84,7	111,4	-26,7	118,3	75,9	11890	9659
DENW22AL50000sME	E	E	132,3	132,3	0	476	476	70494	70494
DENW22AL50000vt9	B	B	64,1	64,1	0	136,2	136,2	10891	10891
DENW22AL50000v1U	D	D	121,3	112,9	8,4	20,7	19,9	3186	2893
DENW22AL50000oCk	C	C	78,9	88,4	-9,5	42,2	35,4	4711	4293

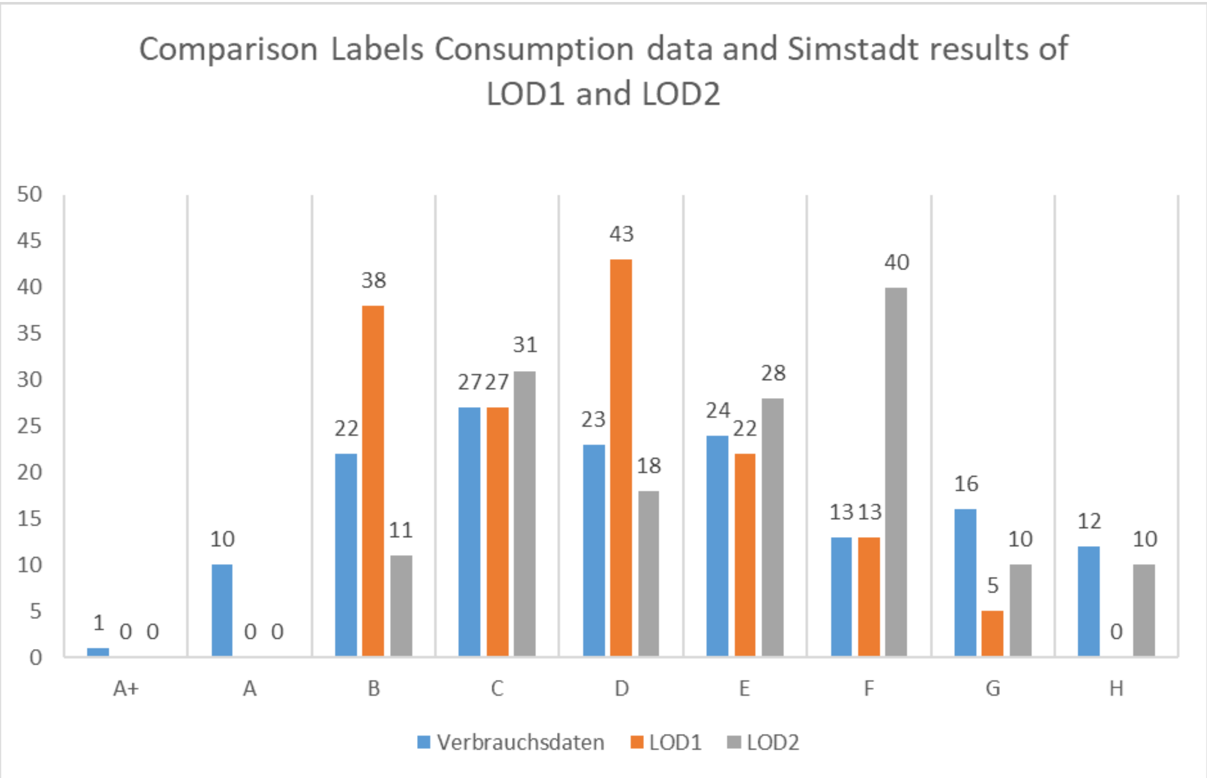
Source: Own elaboration, HFT, 2019

3.4.2 Comparison to real consumption values

As a next step the LOD1 and LOD2 simulation results are compared to the actual available consumption data. The consumption data is given at building level. The energy consumption used for heating by either gas-powered and/or electric heating elements is provided for each building. The sum of both is the total annual energy consumption which is compared to the simulation results.

A comparison between the same set of buildings in the simulation results with LOD1 and LOD2 buildings and the real consumption data is shown in Figure 10 and Figure 11. In order to make this comparison, the three datasets needed to be filtered. The remaining matching pairs are only 148 buildings for which information from all three different sources is available.

Figure 10. Comparison of labels generated from real consumption data and simulated values in LOD1 and LOD2[18]



Source: Own elaboration, HFT, 2019

[18] Verbrauchsdaten corresponds to real consumption values

Figure 11. Direct Comparison between simulated and real consumption labels of some example Buildings

UUID	Real consumption data labels	Simstadt Labels LOD 1	Simstadt Labels LOD2
DENW22AL50000vzW	A	D	E
DENW22AL50000uJi	B	B	B
DENW22AL50000vsX	H	C	D
DENW22AL50000uCv	E	D	F
DENW22AL50000tx8	H	D	F
DENW22AL50000pOJ	B	D	E
DENW22AL50000pRO	H	D	E

Source: Own elaboration, HFT, 2019

3.4.3 Data uncertainties

The comparison between the labels derived from LOD1 and LOD2 simulations and real consumption data, as illustrated in Figure 11, shows relevant differences between the results coming from each data source. Only in the case of labels C and E the number of buildings having such labels is approximately the same.

The reason for the big differences in certain labels can't be identified unequivocally, because of several sources of uncertainties in the data. One reason could be the fact that in the CityGML files there is no information about possible refurbishment scenarios. In this case SimStadt might have calculated a rather poor energy demand value where a refurbishment actually happened, improving the energy efficiency. Another information not present in the CityGML files is whether attics or basements are heated or not, which may lead to wrong simulations.

Moreover, in the results of the simulations obtained with LOD1 and LOD2 data the heated area per building is estimated from the respective 3D model, without considering the influence of users' behaviour in terms of possible special heating habits. Because SimStadt assumes a rather normal heating schedule assigned depending on the function (residential, commercial, etc.) of the buildings, the user behaviour, not modelled inside SimStadt, represents another significant source of uncertainty of the simulations results.

3.4.4 Calculation of labels using a different reference surface

A different way of assuming energy labels per buildings has been also elaborated. The standard way of calculating the total yearly heating energy demand per building is to multiply the specific space heating energy demand (kWh/m^2/yr) by the heated floor area.

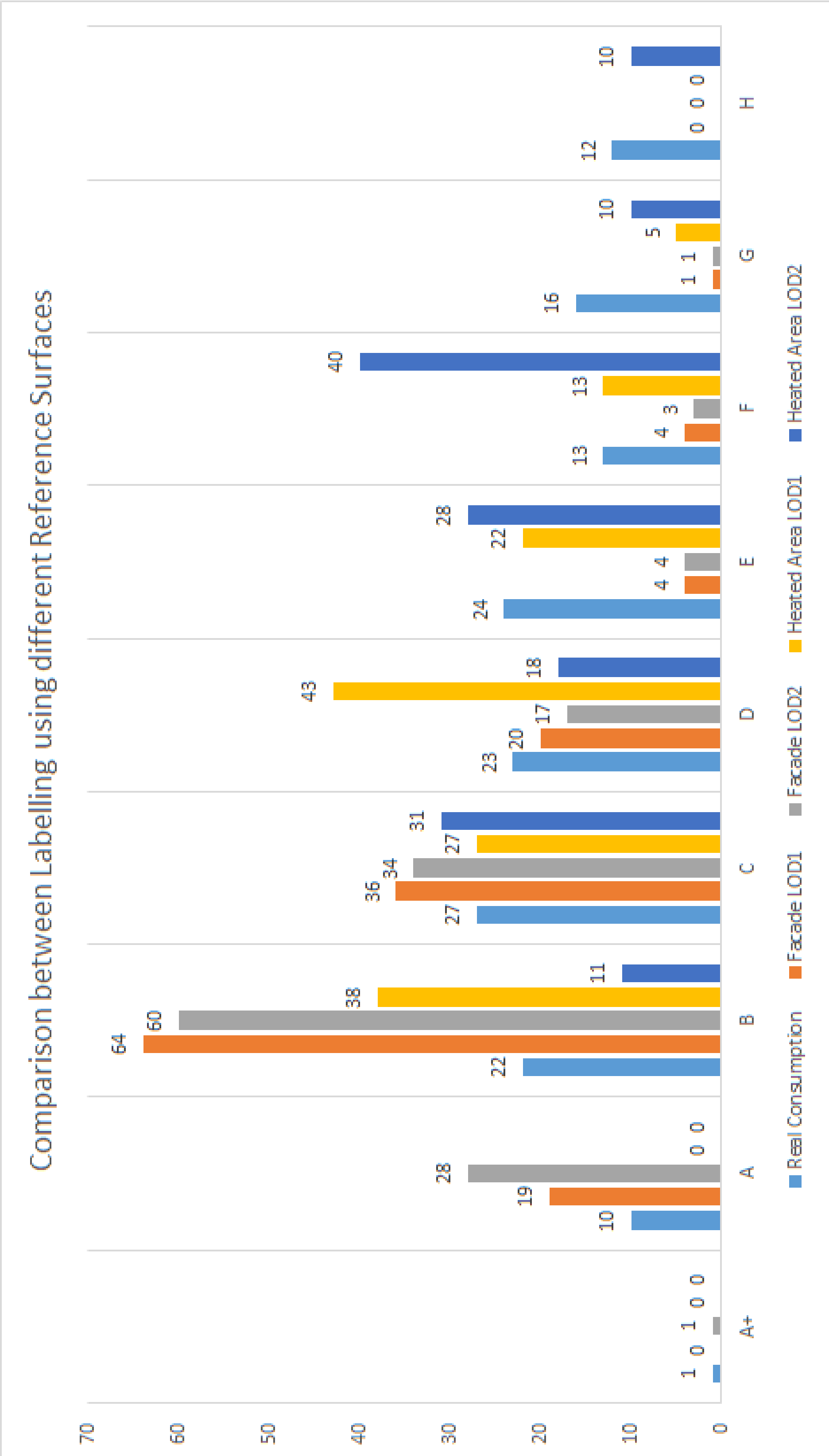
This section examines the impact of the change of the reference area on the label attribution per building.

In order to attribute labels using a new reference area, the specific heating energy demand is multiplied by the facade area, which can be calculated from the 3D building data. The variation of the labels resulting from the two different energy reference surfaces (floor area vs facade area) can be used to evaluate the buildings with regard to the energy demand required to heat the internal floor area on the one hand, and with regard to the demand that would be needed to heat the facade area on the other hand.

Considering the facade area makes sense because a big portion of heat loss is almost always caused by the emission of warm air through the facade of a house. Buildings that have a small facade area therefore perform better in an energy assessment, as less heating energy can be lost. Figure 12 shows, for the same 148 buildings shown in Figure 7, a comparison between the labels as attributed using the real consumption data, the labels calculated using the heated floor area, and the labels as calculated using the facade area of the buildings.

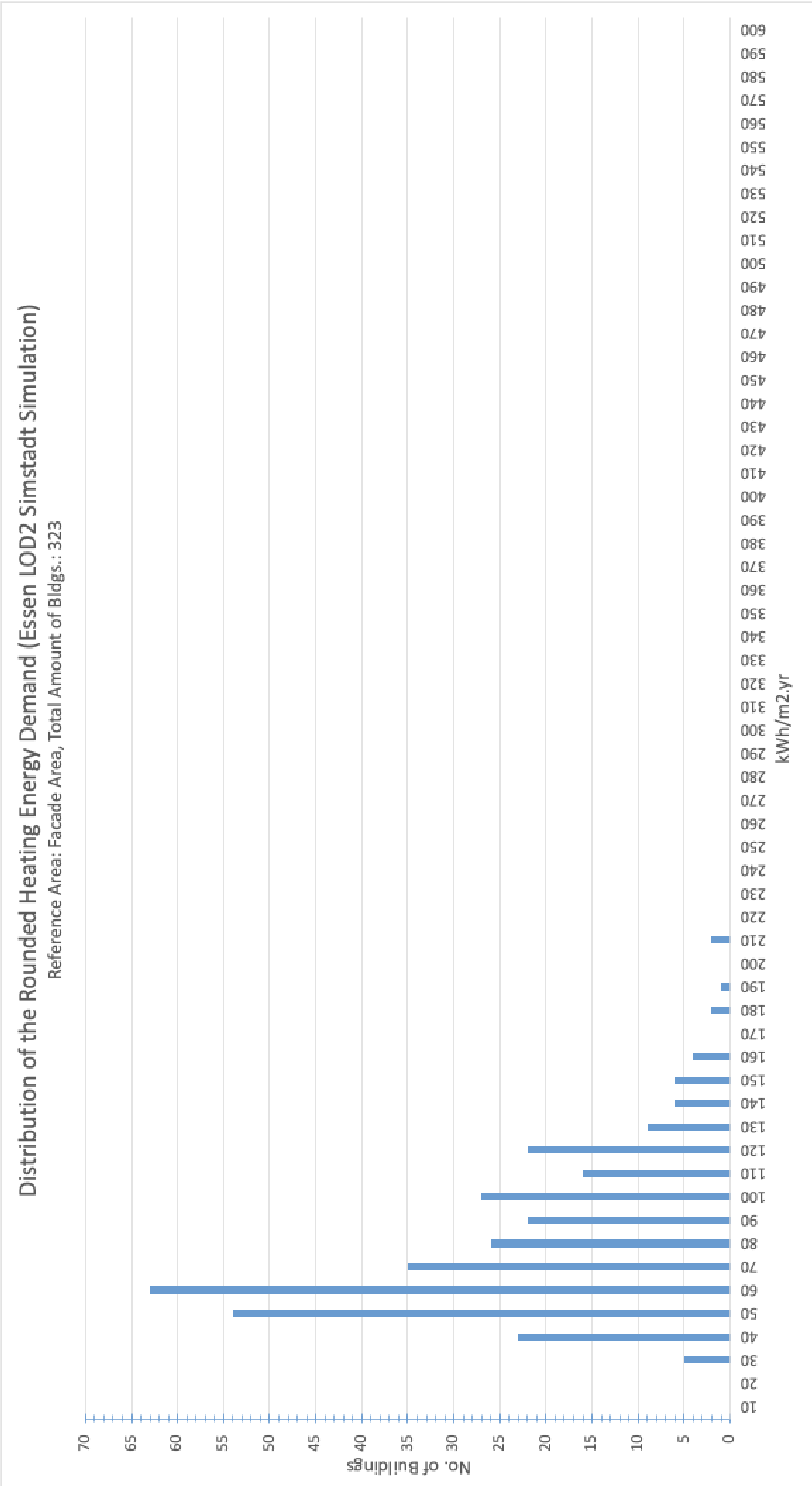
The facade area-based attributions of labels C and D and, to some extent, label A, are surprisingly close to actual consumption data. A general correlation between the labels calculated with the facade surfaces and the labels using the standard energy reference surface cannot be seen except for label C. This circumstance is also confirmed looking at 0, which shows the rounded absolute values calculated per m^2 using the facade area as reference surface.

Figure 12. Comparison of labels using different reference surfaces



Source: Own elaboration, HFT, 2019

Figure 13. Rounded simulated heating energy demand per m^2 of Facade Area using the LOD2 City Model



Source: Own elaboration, HFT, 2019

3.5 Conclusions

Considering the diagrams shown in Section 3.4.1, it can be concluded that using different LODs as input to the calculation in SimStadt introduces changes in the output. The observed trend is that an increase in the LOD causes an increase of the simulated energy values.

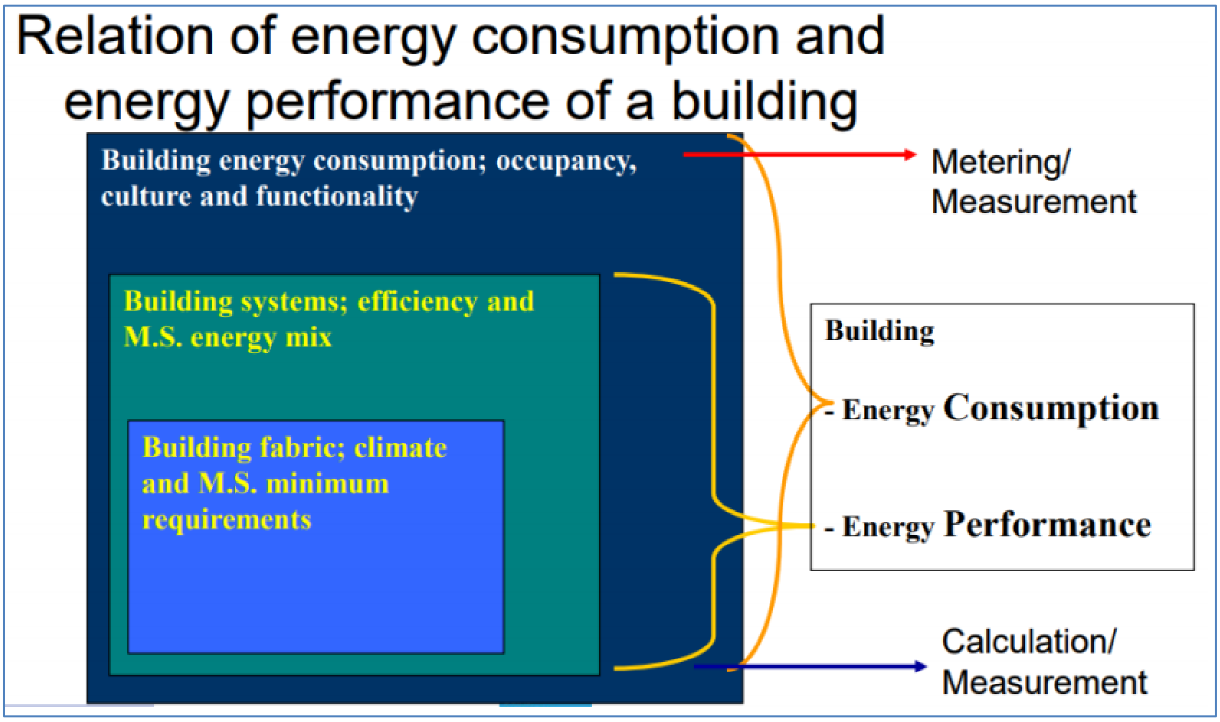
Regarding the comparison with real consumption values, there are several reasons why they do not match. Among them, the most recurrent and significant one is the impossibility to model energy-saving refurbishment measures adopted in the buildings, as well as user behaviour introducing specific heating habits.

In fact, despite a huge part of a building overall energy consumption is made up by the physical parameters (such as the building fabric) which already enable a good estimation of the energy heat demand using state-of-the art assessment methods, the factor which is not predictable is the occupant's behaviour, because only assumptions can be made in this case.

In general, a comparison between simulated energy demand and actual energy consumption is only possible with a limited accuracy.

This circumstance is also shown in Figure 14, where the relationship between energy performance (mostly meaning the energy derived by physical factors such as climate and building fabric) and energy consumption (meaning the consumption of buildings with respect to the occupancy schedules, etc.) is shown.

Figure 14. Relation between energy consumption and energy performance



Source: Presentation "Energy Performance of Buildings - Status and Strategy for using Dynamical Calculation Methods", p. 9, Hans Bloem, 2nd General Consortium Meeting. 26-27.05.2015, JRC, 2015

4 Simulations in NL (test area 1 – Zwolle)

In this section the simulations made in the test area of Zwolle in the Netherlands are described. In this case, an input datasets, corresponding to LOD1 CityGML level of details, has been used as input data for energy heat demand simulations carried out with SimStadt simulation software, described in Section 3.2.

4.1 Input data

The dataset used for the analysis described in this section, represented by 3D buildings in CityGML LOD1, has been provided by the Dutch Cadastre.

The CityGML file is available in the coordinate Reference System EPSG:7415 (Amersfoort / RD New + NAP height).

In Figure 15 the different feature types included in the dataset can be seen.

Figure 15. All of the 3D content in the CityGML dataset

Source: Own elaboration, HFT, 2019

Not only buildings (in turquoise) but also waterbodies (blue), landuse (green, brown), bridges (orange) and roads (grey) are included. In total the stored information has a volume of 2GB and counts 1923 buildings. For each building 23 attributes are present:

- Creation Date
- Min height surface
- shape_area
- shape_length
- gebruiksdo
- **gebruiks_1 (Building Usage/function)**
- aanduidingrecordinactief
- aanduidingrecordcorrectie
- aanduidinginonderzoek
- documentnummer
- einddatum
- officieel

- identificatie (Building ID)
- begindatum
- target_fid
- **bouwjaar (Year Of construction)**
- Status (Status of the building (Is the building currently in use or is it left empty?))
- Objectid (Object ID – For the distinctive differentiation the Object ID is not needed, the Building-ID serves that purpose already)
- documentdatum
- measuredHeight
- lod0FootPrint (List of the vectors describing the ground surface of the building)
- lod0RoofEdge (List of the vectors describing the roof surface)
- lod1Solid (List of the vectors describing the block of the building)

In the list above, the information which is relevant for the energy analysis using the SimStadt software is in bold (the year of construction and the function).

4.1.1 Pre-processing

The input CityGML file needs some pre-processing to be done before being imported into SimStadt.

4.1.1.1 Change of the Coordinate Reference System

SimStadt has been developed in Java and during the creation of the software many libraries have been used to enable the inclusion of prebuilt functionalities into the source code. One of these libraries (PROJ[19]) is responsible for the coordinate transformation of the input CityGML file. The transformation to a global system is done to easily include local environmental data coming from an external database (from the software INSEL in this case) into the workflow. Nevertheless, the transformation of the coordinates is only for internal use and the SimStadt output is still given in the input Coordinate Reference System (CRS). The Java library used for that is called 'proj4j' and is capable to transform coordinates between plenty of systems.

The CRS in the Zwolle CityGML file is given with an EPSG code which is not covered by the used library. Therefore, an alternative code has to be declared. In the CRS declaration part of the CityGML file, the EPSG code is changed from 7415 (Amersfoort / RD New + NAP height) to 28992 (Amersfoort / RD New). This new code is then usable in the library and SimStadt does not throw any errors while importing the dataset. This new EPSG code describes the same CRS as the code 7415, which in turn allows to leave the coordinates in the CityGML-file untouched.

4.1.1.2 Isolation of buildings

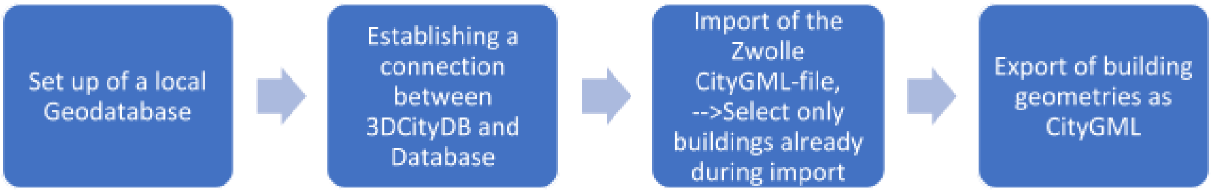
The CityGML file of the Zwolle area has been delivered including geometries for nearly all available modules of CityGML 2.0. Since SimStadt is designed to work only with files containing building geometries, the buildings have to be extracted. For this purpose, the software "3DCityDB"[20] is used, which is basically a PostGIS database which enables to store 3D spatial data. After importing the whole CityGML file, it allows to export only geometries of the type "buildings" in a new CityGML file for the further analysis.

The workflow applied is shown in Figure 16.

Selecting only the buildings from the CityGML file caused a reduction of the file size from 2 GB to 24 MB.

19 https://proj.org/
20 https://www.3dcitydb.org/3dcitydb/

Figure 16. Applied workflow for the extraction of the buildings from the CityGML file

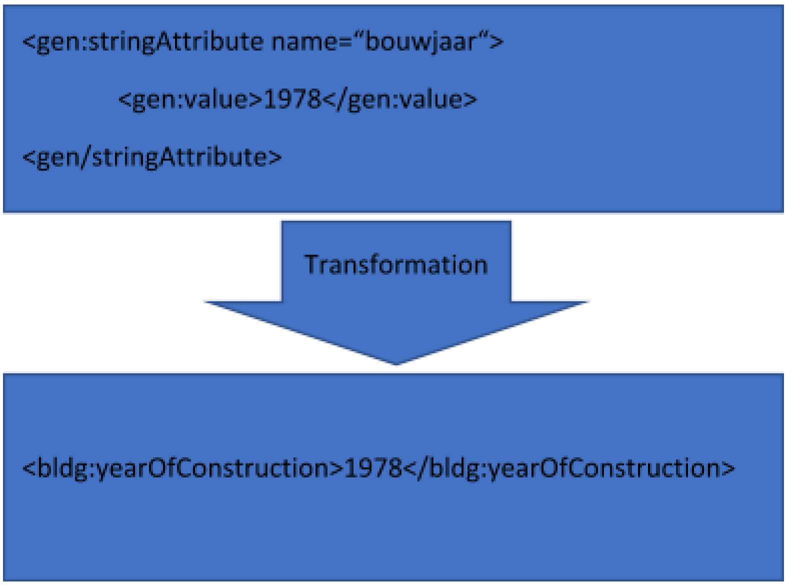


Source: Own elaboration, HFT, 2019

4.1.1.3 Transformation of building attributes

The CityGML file has been created in a way that many attributes have been captured. The Dutch cadastral authority included these attributes in Dutch terms (e.g. "bouwjaar" for year of Construction). These Dutch attributes are included into the CityGML file by using generic attributes. SimStadt needs at least the year of construction and the function attribute. For this purpose, the generic attributes "bouwjaar" and "gebruiks_1" are translated into the official CityGML standard attribute declaration. Some Java code has been written to fulfil this task. Figure 17 shows the functionality of the program.

Figure 17. Schema of the transformation of Dutch building attributes into non-generic CityGML-attributes



Source: Own elaboration, HFT, 2019

The function attribute which is given by the Dutch generic attribute "gebruiks_1" also had to be transformed into an official CityGML-standard-attribute. An obstacle is the fact that the Dutch description of functions is different from the Usage library already implemented in SimStadt.

The library implemented in SimStadt is based on German ALKIS Codes (codes distributed by the German cadastral authority), therefore an allocation from Dutch descriptive names to German ALKIS Codes had to be done, as shown in Table 1.

Table 1. ALKIS Code Allocation

Dutch Function	English translation of the Dutch Function	German ALKIS Code	English translation of the German ALKIS Code
Woonfunctie	Residential buildings	1010	Residential Building
Industriefunctie	Industrial Building	2112	Company Building
Kantoorfunctie	Office	2020	Office Building

Bijeenkomstfunctie	administrative building	3010	Administrative Building
Overige gebruiksfunctie	Other functions	9999	Other functions
Onderwijsfunctie	School	3021	General Educational School
Winkelfunctie	business premises	2050	Commercial Building
Gezondheidszorgfunctie	Hospitals	3051	Hospital
Sportfunctie	Sports facilities	3210	Building for Sports Purposes
Logiesfunctie	Hotel, Motel, Pension	2071	Hotel, Motel, Pension

Source: Own elaboration, HFT, 2019

4.1.1.4 Geometrical Errors in the Dataset

The geometries of the buildings must be as much error-free as possible, in order not to prevent/alter the calculation of several geometrical attributes of each building.

These attributes are:

— footprint area [m^2]

— total wall area above ground [m^2]

— building's volume [m^3]

— area of wall surfaces shared with another building [m^2]

— area of walls facing to the outside [m^2]

— area on the roof [m^2]

— mean height of the building [m]

— heated area derived by the average storey height coming from the usage library and the building height [m^2]

In case there are errors in the buildinggeometry, this could cause the miscalculation of some of the attributes shown above. Those miscalculations have a direct impact on the plausibility of the heating energy demand values. For example, in case that the volume of a building can't be derived because of too many geometrical errors, SimStadt is assuming the bounding box of a building as the new volume.

4.2 Simulation environment

After the successful preparation of the Zwolle CityGML input dataset, it is ready to be imported into the SimStadt. simulation software, already described in Section 3.2.

Out of 1923 buildings, 80 have as function attribute the code 9999 (or "overige gebruiksfunctie", i.e. other function). Using the already implemented German library prevents SimStadt to calculate any heating energy demand values since these are considered not to be heated.

4.3 Results

4.3.1 Label-classification using the German "Energieausweis"-Method

The SimStadt output is a csv file containing columns with the calculated values per building after each simulation step. The most important values generated in the last step of the workflow are the yearly, monthly and specific space heating demand results. The specific space heating demand shows the yearly sum of the heating energy demand per m^2 of the heated area of the building. This column has been used to assign labels according to several labelling methods. One method is to assign labels according to the German "Energieausweis" (in English: "energy certificate"), as shown in Table 2.

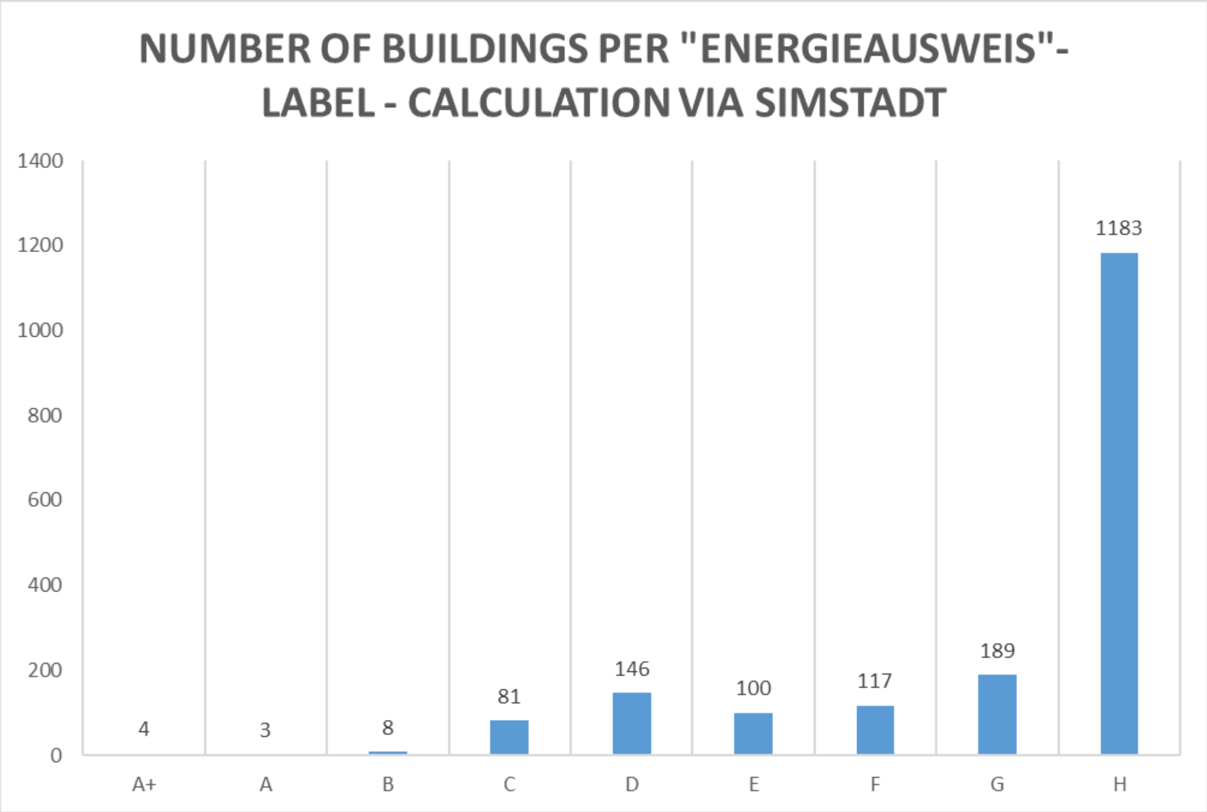
Table 2. German Energieausweis: Labelling of energy values

kWh/m^2/year Interval	Label
0 - <30	A+
30 - <50	A
50 - <75	B
75 - <100	C
100 - <130	D
130 - <160	E
160 - <200	F
200 - <250	G
>=250	H

Source: Own elaboration, HFT, 2019

Figure 18 visualizes the results of “specific heating demand” as calculated by SimStadt and then classified into labels. The values given by SimStadt in kWh/m^2/year are attributed to labels using a custom-made Java program. Most of the buildings are attributed to the label H, consistently with the fact that Zwolle consists of very old buildings in general. In the CityGML dataset describing the city Zwolle, the average year of construction of all buildings is 1925. That means that in the German building physics library the parameters belonging to the oldest available construction year epoch are taken, which are causing very poor heating energy demand values.

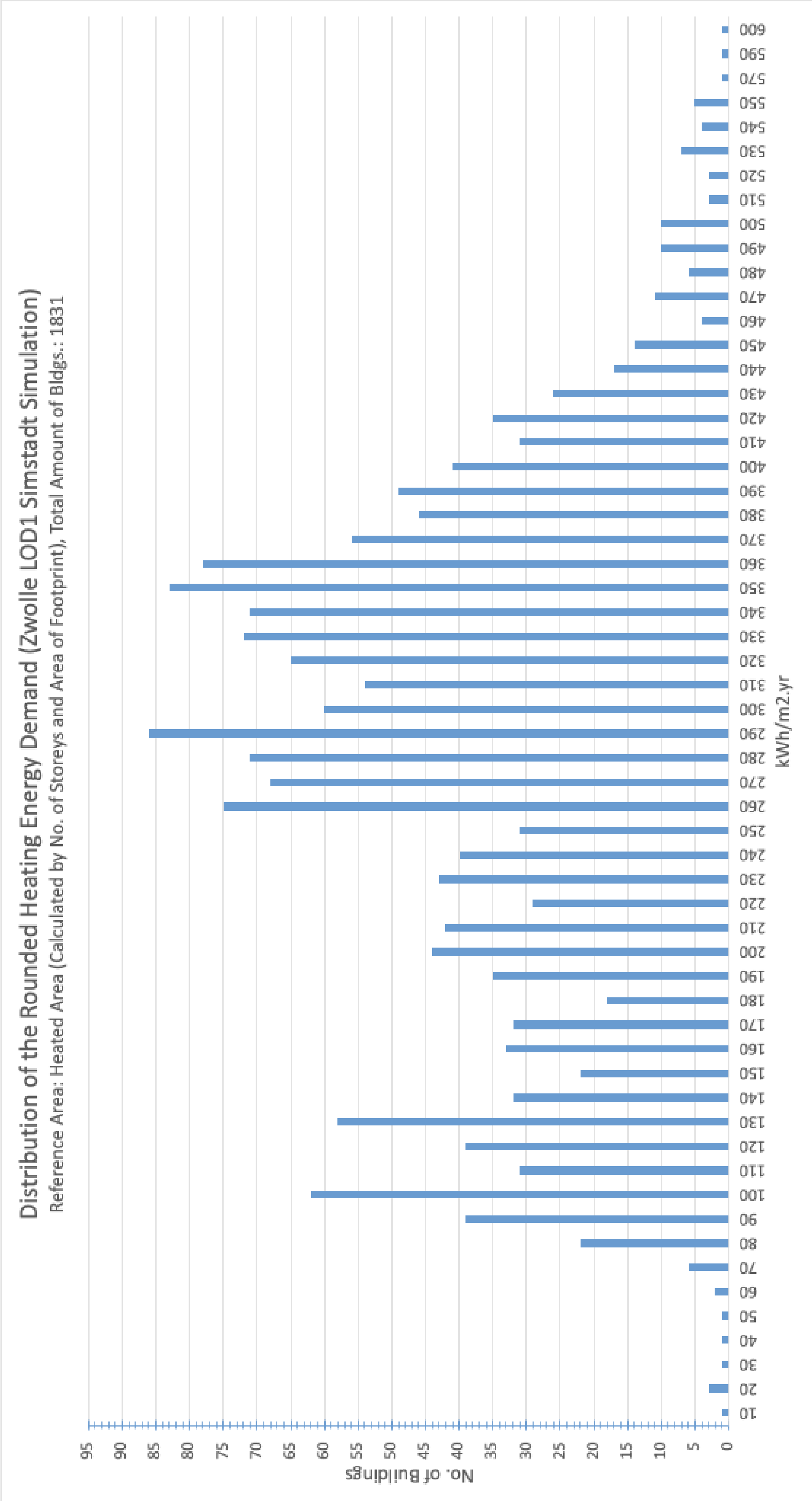
Figure 18. SimStadt results classified using the "Energieausweis"-labels



Source: Own elaboration, HFT, 2019

Since the heat demand of many buildings goes far beyond the limit of 250 kWh/m²/year (Label H in the "Energieausweis" table), the heat demand absolute values shown in Figure 19 provide a better representation than that based on the labels. For the creation of Figure 19 the heating demand of all buildings has been rounded, in order to consider intervals of 10 kWh/m²/year and count the number of buildings within each interval. It can be seen that a consistent number of buildings have a heating energy demand of around 290 kWh/m²/year.

Figure 19. Rounded simulated heating energy demand using the Zwolle LOD1 Citymodel



Source: Own elaboration, HFT, 2019

4.3.2 Results using the Dutch labelling method (RVO Method)

In order to compare the "Energieausweis" labelling method to other approaches, the labels of the buildings are also attributed using the method developed by the Dutch cadastre. This method basically consists of a table (shown in Figure 20) where a label can be looked up depending on the building type and year of construction.

Figure 20. Allocation of the Energy labels - methodology from Dutch cadastre

Dwelling type	Building period									
	up to and including 1945	1946-1964	1965-1974	1975-1982	1983-1987	1988-1991	1992-1999	2000-2005	2006-2013	2014 and further
seperate house	G	F	D	C	C	B	B	B	A	A
Semi detached house	G	F	D	C	C	C	B	B	A	A
Detached house	G	F	D	C	C	C	B	B	A	A
Detached corner house	F	E	C	C	C	C	B	A	A	A
Flat/appartement*	G	E	E	B	C	C	C	B	A	A

Source: Own elaboration, HFT, 2019

Since the SimStadt simulation has been carried out using supporting libraries describing a German building stock, the building types which are automatically assigned in SimStadt have to be mapped to the ones which can be seen in Figure 20.

It has been decided to implement the following mapping of building types:

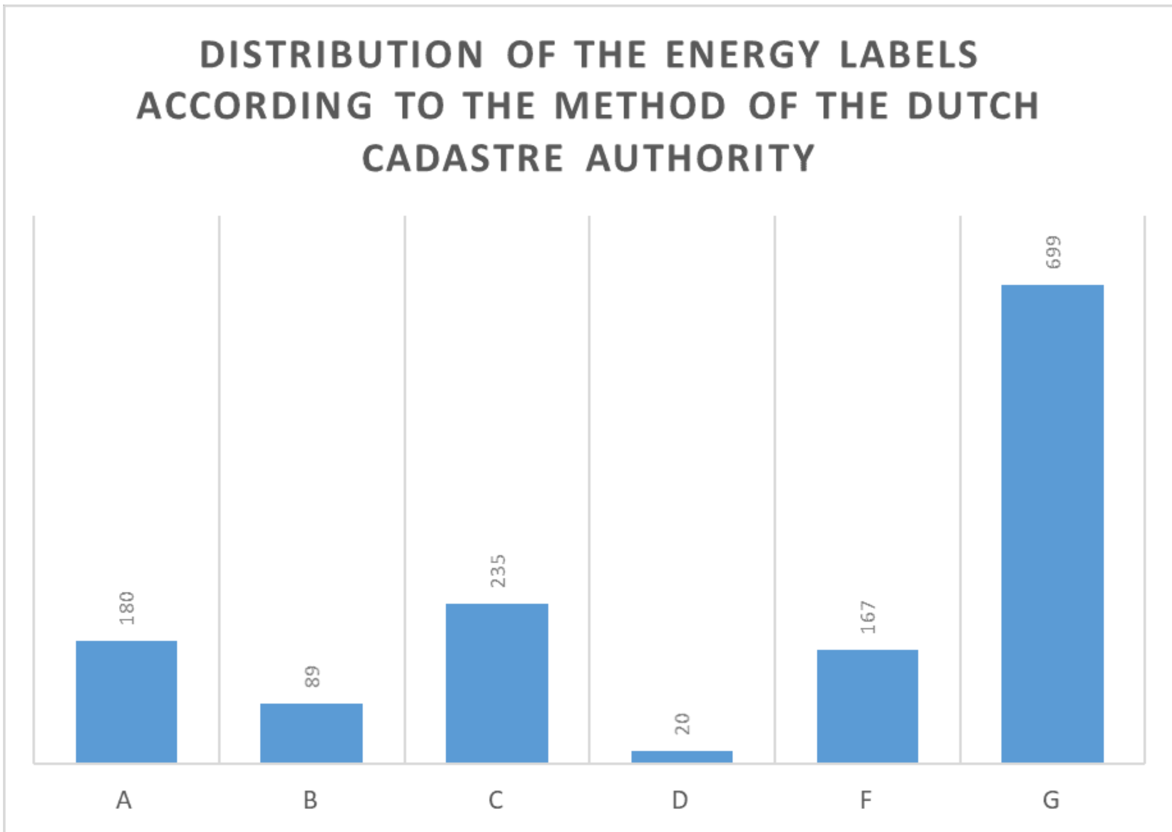
Table 3. Mapping of Dutch building types to German building types

Dutch Building types	German Building types
Separate House	Single Family House
Semidetached House	Multi Family House, Row House, Big Multi Family House
Detached House	Single Family House
Detached Corner House	Single Family House
Flat/Apartment	Is not existing in the Zwolle dataset, but in this context, it would be part of Big Multi Family House

Source: Own elaboration, HFT, 2019

A Java program has been created to attribute the labels according to the Dutch method, checking the assigned building types in SimStadt and the year of construction. The result of this classification is shown in Figure 21.

Figure 21. Distribution of Energy Labels following the method of the Dutch cadastre authority



Source: Own elaboration, HFT, 2019

As can be seen in Figure 21, this method generates a much more distributed attribution of the energy labels per building when compared to the one shown in Figure 18. Nevertheless, most of the buildings are very old (average building year is 1925) and therefore get the worst label G. This result is similar to the result obtained with the German method, which also classifies most of buildings into the label with highest energy demand. Therefore, the majority of the buildings have a low energy efficiency performance in both methods.

4.4 Conclusions

In general, the comparison between the two labelling methods is difficult since a different number of labels are given using the two approaches. One big benefit of the Dutch "RVO"-method is that it is independent of any format of the input information, and energy classifications of buildings can be quickly made. On the other hand, this approach might not be the most accurate, since the actual geometry of a building is disregarded.

But in some way the geometry is represented in the dwelling types listed in Figure 20, although the incorporation of the actual shape of a building is more accurate in the approach using SimStadt.

Moreover, SimStadt uses a huge variety of different input parameters, which can be adjusted in a way to better fit to the actual building stock and the construction patterns a country is following. The fact that more parameters are considered for the estimation of a heating energy demand value, leads to the conclusion that SimStadt method delivers more plausible values as output, with respect to the RVO method.

Nevertheless, the attribution of energy labels following the schema of the "Energieausweis" has some issues, since most of the buildings in the Zwolle Dataset are very old and therefore have a poor heating energy performance. More than 60% of the buildings have been estimated to be classified in the Label H, which means that these buildings consume more than 250 kWh/m^2.yr. For this reason, the absolute values have been included in this report as well, as can be seen in Figure 19. There it can be seen that the actual average of the estimated energy demand is located at around 290-310 kWh/m^2.yr.

5 Simulations in NL (test area 2 – Enschede)

5.1 Introduction

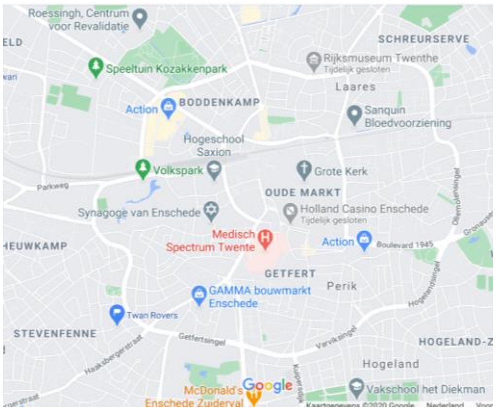
This section describes the results of the development and application of a CityGML model based on open data and the prediction of the energy consumption of a group of buildings in the Dutch city of Enschede using the simulation environment SimStadt.

The chosen case study is a group of residential buildings in the Dutch city of Enschede. The research questions to be answered are:

1. What is a practical workflow to develop a LOD 1 CityGML model from publicly available GIS data?
2. How accurate is a LOD 1 CityGML model for predicting the energy consumption for heating with SimStadt compared with measured energy-use data?

Enschede is a municipality and city in the eastern Netherlands in the province of Overijssel, home of the University of Twente and the Saxion University of Applied Science. The eastern part of the urban area reaches the border with Germany. The municipality of Enschede consisted of the city of Enschede until 1935, when the rural municipality of Lonneker, which surrounded the city, was annexed after the rapid industrial expansion of Enschede which began in the 1860s and involved the building of railways and the digging of the Twentekanaal. The municipality of Enschede counts approx. 160'000 inhabitants. The inner city lies within the ring road called "De Singel" (see Figure 22 and Figure 23). In the late 19th century the city developed into a major Dutch textile manufacturing centre causing the number of inhabitants to quintuple between 1870 and 1900.

Figure 22. Map of Enschede's inner city



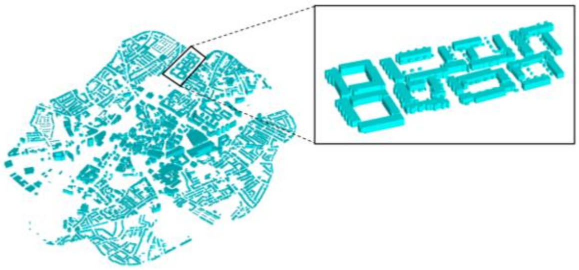
Source: Google Maps, 2020

Figure 23. Inner city of Enschede as 3D model, LOD1

Source: Own elaboration, HFT, 2020

The neighbourhood considered in this study is part of the district "De Bothoven" and lies to the East of the inner-city area. It is situated to the Northeast of the Hoge Bothofstraat (see Figure 24). The neighbourhood is dominated by dwellings, particularly terraced houses and apartments which are to a large extent owned by the local housing association "Domijn".

Figure 24. Case study: neighbourhood near the Hoge Bothofstraat



Source: Own elaboration, HFT, 2020

The considered neighbourhood consists of 113 building blocks and 374 addresses / dwellings. The dwellings were re-built in the early 80'ies in a style corresponding to the historic topology of the early residential areas, housing workers from the nearby textile factories (see Figure 25 and Figure 26).

Figure 25. Benninkburg 1 – 45 Enschede

Source: Own elaboration, HFT, 2020

Figure 26. Brinkhuisburg 1-3 Enschede

Source: Own elaboration, HFT, 2020

5.2 Methodology

To answer the research questions formulated in Section 5.1, a number of aspects have to be addressed such as:

1. Review of available input data for the development of a LOD 1 CityGML model;
2. Definition of a practical workflow to aggregate the data into a model to be simulated with SimStadt;
3. Customization of the SimStadt simulation program for the Dutch context, consisting in the development of a Dutch Building Physics Library for SimStadt and in the collection and analyses of different climate datasets;
4. Identification and pre-processing of measured energy-use data for a comparative analysis with predicted energy-use;
5. Comparative analysis of predicted and measured energy-use.

5.3 Data availability in The Netherlands

The Dutch Cadaster is the main source for data related to the building stock in the Netherlands. Since 2020 the Dutch Cadaster publishes annually three datasets of its country topology: the first being a 3D representation of topographical objects for waters, roads and buildings, the second being a 3D representation of topological objects for buildings only, accounting for differences in height, and thirdly, two dimensional representations of buildings including different statistics for building height[21]. Data with respect to year of construction and dwelling type can be derived from different databases such as BAG, BRK etc. (see Table 4 for more detailed information).

[21] https://www.pdok.nl/3d%20basisvoorziening

Table 4. Overview of data resource for Dutch building stock[22]

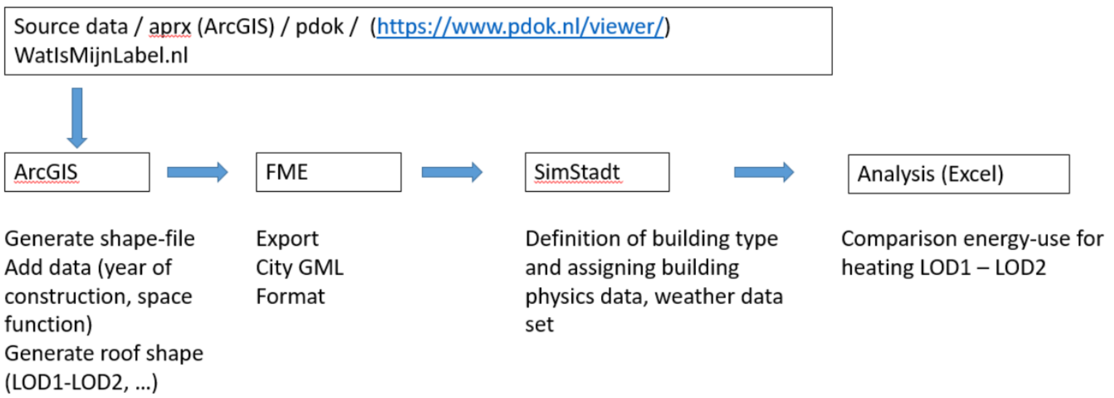
Key registers and other national datasets	Attribute
Key register for addresses & buildings (BAG)	Building use, address, year of construction, dwelling type (derived attribute))
Key register cadastre (BRK)	Transactions, characteristics seller property, characteristic owner property
Key register for large scale topography	Geometry
National energy label database	Energy label (EPC)
Additional dwelling specific information	Various, including implemented energy saving measures
Aerial imagery	Point clouds

Source: Own elaboration, Saxion, 2020

5.4 Workflow

The formulated workflow for the development of a LOD1 CityGML model, shown in Figure 27, is based on the use of ArcGIS and FME.

Figure 27. CityGML model development from GIS data - workflow



Source: Own elaboration, Saxion, 2020

The preparation of a CityGML file from GIS data sources such as the Dutch BAG register is not a straightforward task. The preparation and review process involves several tools and different control loops on aspects such as the assignments of the proper building type, plausibility checks of the building height, and definition of storey heights to estimate the correct number of floors per building.

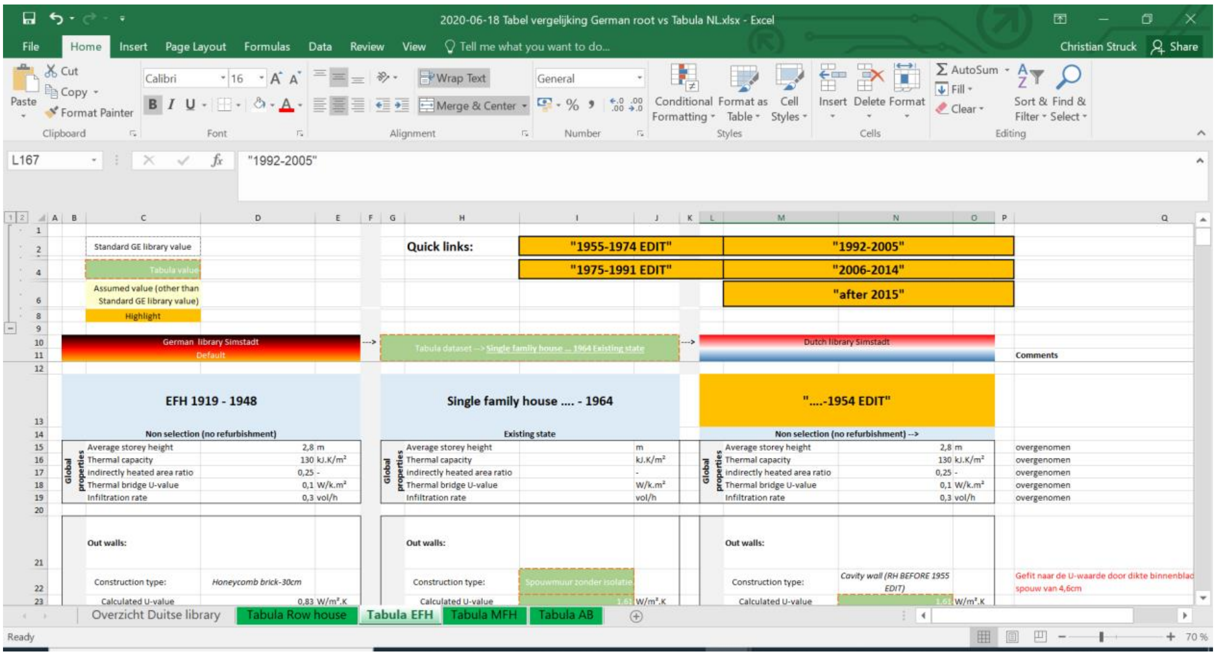
5.5 Developments of Dutch Building Physics Library for SimStadt

The library has been developed for four building types and five construction periods, see Figure 28. The building types are terraced houses, single family houses, detached houses and apartments. The construction periods are before 1955, 1955-1974, 1975-1991, 1992-2005, 2006-2014 and after 2014. The building physical properties originate from the Tabula Webtool database for the Netherlands.[23]

[22] Coors, Vranken, Martirano et al. (2018) *Assessing energy performance of buildings using modeling based on existing administrative and topographical data, Presentation at INSPIRE conference 2018*

[23] https://webtool.building-typology.eu/?c=all#bm

Figure 28. Screenshot of the conversion tool for the preparation of the Dutch Building Physical Library for SimStadt



Source: Own elaboration, Saxion, 2020

5.6 Climate datasets

Four different climate datasets have been considered for the analysis:

- a dataset present in the INSEL database of SimStadt nearest to the City of Enschede, Münster;
- a dataset containing average monthly climate parameters for the year 2019;
- a typical metrological year for the region of Twente originating from ASHRAE[24];
- a standardized climate dataset for energy calculation for the Netherlands published by the Royal Netherlands Standardization Institute (KNMI).

The characteristics of the four datasets are described in the following sub-sections and summarised in Table 5.

5.6.1 INSEL Dataset for Münster (DE)

INSEL v.8 is an integrated simulation environment language based on a block diagram interface for programming applications in the renewable energy sector. INSEL is developed to support the design, analysis and education on concepts for complex energy projects. It allows to synthesize meteorological time series data, model creation and simulation of photovoltaic solar thermal energy systems as well as the computational simulation of buildings integrated with energy systems. SimStadt enables access to different meteorological databases in INSEL such as TMY3, DWD, and others. The dataset for Münster, Germany has been chosen for the comparative analysis as it represents the nearest available location to the selected case study.

5.6.2 NEN 5060 – 2018 / Hygrothermal performance of buildings - Climatic reference data

The Dutch standard NEN 5060:2018 provides climate reference datasets for the determination of comfort and energy performance of buildings as well as specification of heating and air conditioning systems. The standard presents three datasets, one for energy demand prediction and two for the assessment of the overheating risk and comfort in buildings. The datasets consist of annual data for the period 1986 until 2005, in which measured data from representative months for that period are statistically selected to form an average annual dataset consisting of 8784 data points per parameter for performance predictions for the considered period. The dataset is the Dutch implementation of the NEN-EN-ISO 15927 standard. The dataset is supposed to be used for design

[24] American Society of Heating, Refrigerating and Air-Conditioning Engineers

calculations, such as heating load, cooling load as well as annual energy performance. Additionally, degree days are provided.

5.6.3 ASHRAE IWEC2 Weather File for TWENTE

The ASHRAE IWEC2 database contains "typical" weather files for 3012 locations outside the United States and Canada. The files are derived from Integrated Surface Hourly (ISH) weather data originally archived at the National Climatic Data Center. IWEC2 weather files were developed through ASHRAE Research Project RP-1477, "Development of 3012 Typical Year Weather Files for International Locations," by White Box Technologies, Moraga, California, Y. Joe Huang, Principal Investigator [11]. These files are derived from meteorological reports of weather stations around the world that are archived in the Integrated Surface Hourly (ISH) data base maintained by the National Climatic Data Center (NCDC). For these selected locations, the ISH database includes weather observations for, on average, at least four times per day of wind speed and direction, sky cover, visibility, ceiling height, dry-bulb temperature, dew-point temperature, atmospheric pressure, liquid precipitation, and present weather for at least 12 years of record up to 25 years. They are intended to be used for computational performance comparisons of solar energy conversion systems and building systems to alternative system types, configurations and locations in the United States and its territories. They represent typical rather than extreme conditions and are not suited for designing systems to meet the worst-case conditions occurring at a given location.

5.6.4 KNMI Dataset / averaged monthly data based on measurements for 2019

The Royal Dutch Meteorological Institute (KNMI) publishes recorded time-series data for the most important climate parameters, starting from 1951. The datasets are updated daily for 50 locations in The Netherlands. The data used for the analysis originates from Twente and consists of hourly averages data points for the year 2019, which have been processed to be used with SimStadt.

Table 5. Characteristics of weather datasets

Pos.	Item	Purpose	Location	Origin of presented data
1	NEN 5060-2018	Energy performance predictions and relative performance comparison	All of The Netherlands	Statistically composed from measured data 1986 - 2005
2	ASHRAE IWEC2 Twenthe	Energy performance predictions and relative performance comparison	Dutch Region of Twente	Statistically composed from measured data from recordings of 15 – 20 years
3	INSEL Münster	Recorded weather data from DWD	City of Münster, probably Münster Airport	Absolut data, hourly averaged
4	KNMI 2019	Recorded weather data from KNMI	Dutch Region of Twente	Absolut data, hourly averaged

Source: Own elaboration, Saxion, 2020

5.7 Identification and pre-processing of measured energy-use data

Local energy distribution in The Netherlands is in the hands of seven publicly owned companies. The majority of the companies provide standardized open data related to use of electricity and gas. The energy distributor in the North-East of the Netherlands is ENEXIS[25]. ENEXIS provides open data from its supply region to stimulate innovation. The data provided is organised around eight “topics” including energy generation and consumption data of small consumers. For privacy reasons energy use data of at least 10 households are aggregated into one figure.

[25] https://www.enexis.nl/over-ons/wat-bieden-we/andere-diensten/open-data

For simplification the “Postcode 6” areas are used for data aggregation, where possible. Postcode 6 refers to the definition of Dutch postal codes containing four digits and two letters. The first two digits refer to the region, and the latter two to the village or neighbourhood. The letters indicate the street or a section of a street.

A snapshot of the detailed and the resulting aggregated data are shown in Figure 29 and Table 6, respectively.

Figure 29. Format Postcode 6 – energy consumption data for 2019

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	NETBEHEERDER	NETGEBIED	STRAATNAAM	POSTCODE_VAN	POSTCODE_TOT	WOONPLAATS	LANDCODE	PRODUCTSOORT	VERBRUIKSSEGMENT	AANSLUITINGEN_AANTAL	LEVERINGSRICHTING_PERC	FYSIEKE_STATUS_PERC	SOORT_AANSLUITING_PERC	SOORT_AANSLUITING	SJV_GEMIDDELD	SJV_LAAG_TARIEF_PERC	SLIMME_METER_PERC
56899	Enexis B.\	ENEXIS	Bolster	7881 XN	7884 NA	EMMER-COMPAS	NL	GAS	KVB	17	100	100	100	G4	1607,94	0	
56900	Enexis B.\	ENEXIS	Boulevard	7511 AA	7511 AB	ENSCHEDE	NL	ELK	KVB	26	100	100	69,23	1x35	2452,35	59,5	7,69
56901	Enexis B.\	ENEXIS	Boulevard	7511 AC	7511 AC	ENSCHEDE	NL	ELK	KVB	20	100	100	65	1x35	2244,9	49,96	5
56902	Enexis B.\	ENEXIS	Boulevard	7511 AC	7511 AC	ENSCHEDE	NL	GAS	KVB	18	100	100	100	G4	80,83	0	
56903	Enexis B.\	ENEXIS	Boulevard	7511 AD	7511 AD	ENSCHEDE	NL	ELK	KVB	103	100	100	86,41	1x35	3871,85	42,79	4,85
56904	Enexis B.\	ENEXIS	Boulevard	7511 AD	7511 AD	ENSCHEDE	NL	GAS	KVB	101	100	100	99,01	G4	1261,45	0	
56905	Enexis B.\	ENEXIS	Boulevard	7511 AE	7511 AG	ENSCHEDE	NL	ELK	KVB	28	100	100	57,14	1x35	7252,96	46,02	32,14
56906	Enexis B.\	ENEXIS	Boulevard	7511 AG	7511 AG	ENSCHEDE	NL	GAS	KVB	19	100	100	100	G4	271,53	0	
56907	Enexis B.\	ENEXIS	Boulevard	7511 AH	7511 AH	ENSCHEDE	NL	ELK	KVB	25	100	100	72	1x35	1738,64	36,74	12
56908	Enexis B.\	ENEXIS	Boulevard	7511 AH	7511 AH	ENSCHEDE	NL	GAS	KVB	24	100	100	100	G4	157,92	0	
56909	Enexis B.\	ENEXIS	Boulevard	7511 AJ	7511 AJ	ENSCHEDE	NL	ELK	KVB	15	100	100	66,67	1x35	9108,2	26,26	13,33
56910	Enexis B.\	ENEXIS	Boulevard	7511 AJ	7511 AJ	ENSCHEDE	NL	GAS	KVB	14	100	100	85,71	G4	1724,86	0	
56911	Enexis B.\	ENEXIS	Boulevard	7511 AK	7511 AL	ENSCHEDE	NL	ELK	KVB	93	100	100	76,34	1x35	2693,17	49,43	27,96
56912	Enexis B.\	ENEXIS	Boulevard	7511 AL	7511 AL	ENSCHEDE	NL	GAS	KVB	73	100	100	97,26	G4	1418,23	0	
56913	Enexis B.\	ENEXIS	Boulevard	7511 AM	7511 AM	ENSCHEDE	NL	ELK	KVB	31	100	100	96,77	1x40	2090	49,62	6,45
56914	Enexis B.\	ENEXIS	Boulevard	7511 AM	7511 AM	ENSCHEDE	NL	GAS	KVB	31	100	100	100	G4	851,55	0	
56915	Enexis B.\	ENEXIS	Veenstraa	7511 AP	7511 AP	ENSCHEDE	NL	ELK	KVB	40	100	100	100	1x35	2492,23	50,23	2,5
56916	Enexis B.\	ENEXIS	Veenstraa	7511 AP	7511 AP	ENSCHEDE	NL	GAS	KVB	31	100	100	100	G4	28,77	0	
56917	Enexis B.\	ENEXIS	Marthalaa	7511 AR	7511 AR	ENSCHEDE	NL	ELK	KVB	37	100	100	97,3	3x25	2296,73	49,09	8,11
56918	Enexis B.\	ENEXIS	Veenstraa	7511 AS	7511 AS	ENSCHEDE	NL	ELK	KVB	44	100	100	97,73	1x35	1804,32	49,44	70,45
56919	Enexis B.\	ENEXIS	Veenstraa	7511 AT	7511 AT	ENSCHEDE	NL	ELK	KVB	24	91,67	100	70,83	1x35	2365,46	49,51	54,17

Enexis-kleinverbruiksgegevens_0

Ready

Source: ENEXIS, 2020

Table 6. Normalized annual energy use, electricity use and gas consumption (2019)

Enexis (1-1-2019)					Gas		Electricity	
Pos.	Postal code 6 level	Street	House number	N° of adresses	N° of connections	average annual gas use per postal code, normalized [m³] gas	N° of connections	average annual gas use per postal code, normalized [kWh] electricity
1	7511 LB	Kremersmaten	40 - 82	20	20	1306,20	20	2304,30
2	7511 LC	Kremersmaten	84 - 132	23	23	1253,65	23	2722
3	7511 LD	Kremersmaten	134 - 176	21	21	1315,57	21	2362,29
4	7511 LJ	Kremersmaten	143 - 171	15	15	1284,00	15	2115,87
5	7511 LK	Stinsburg	1 - 18	18	18	933,06	18	1841,72
6	7511 LL	Stinsburg	19 - 37	19	18	1219,22	19	2556,42
7	7511 LM	Tijhofburg	1 - 26	13	13	1680,38	13	3039,31
8	7511 LN	Tusveldburg	1 - 24	12	12	1201,08	12	2010,17
9	7511 LP	Tusveldburg	26 - 54	15	15	1100,07	15	1924,4
10	7511 LR	Tusveldburg	1 - 37	19	19	1023,37	19	2097,11
11	7511 LS + 7511 LT	Tusveldburg + Engelsburg	39 - 55, 1 -28	27	27	1278,15	27	2133,48

12	7511 MA	Voogsgerdburg	1 - 35	18	18	973,28	18	2185,56
13	7511 MB	Benninkburg	1 - 45	23	23	922,04	23	2003,61
14	7511 MC	Benninkburg	47 - 107	28	24	965,46	29	1578,9
15	7511 MD	Benninkburg	111 - 159	25	25	901,80	25	1901,88
16	7511 ME	Benninkburg	2 - 26	14	13	1000,08	13	1679,23
17	7511 MG	Brinkhuisburg	2 - 34	17	17	1317,65	17	2065,53
18	7511 MH	Brinkhuisburg	42 - 70	15	15	1359,27	15	2226,33
19	7511 MJ	Brinkhuisburg	1 - 47	24	24	1020,71	24	1815,54
20	7511 MK	Brinkhuisburg	49 - 75	14	14	1479,86	14	2734,71

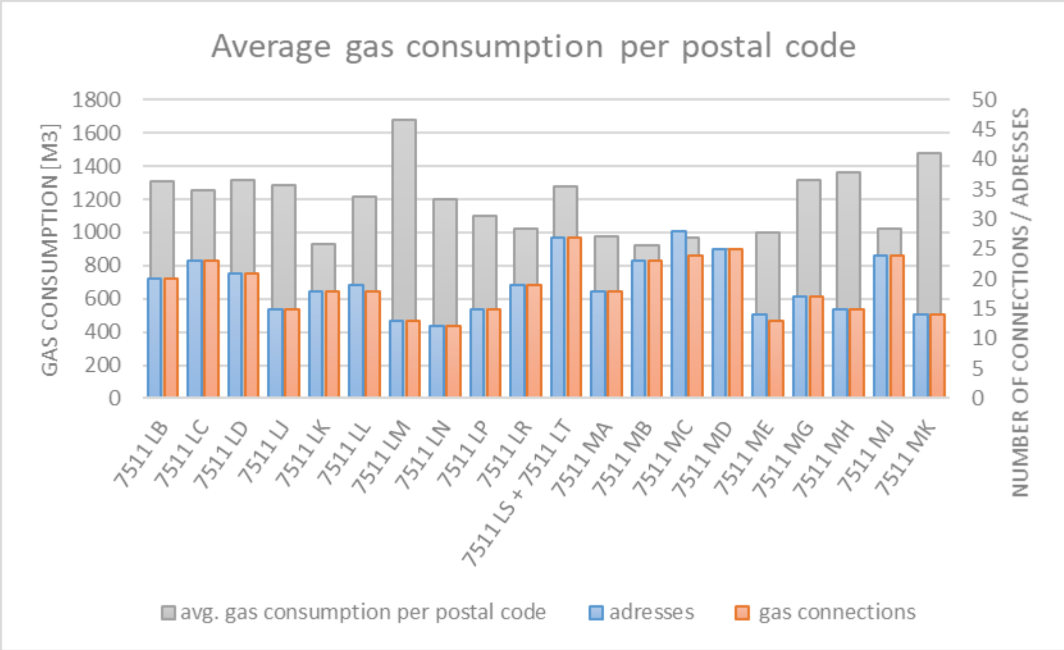
The published standardized annual energy consumption data for electricity and gas is related to a standard year corrected for climate deviations from normal, caloric value of gas and variation g pressure in gas supply [26].

The energy use data for gas and electricity are visualized in Figure 30 and Figure 31, respectively.

It can be noticed that the number of utility connections, gas & electricity, and addresses are largely identical. However, variations are noticeable, as for example for Postcode 7511 ME. For gas connections, the difference amounts to six connections less than addresses. This can be explained by addresses which have been disconnected from the gas network due to all-electric renovation measures.

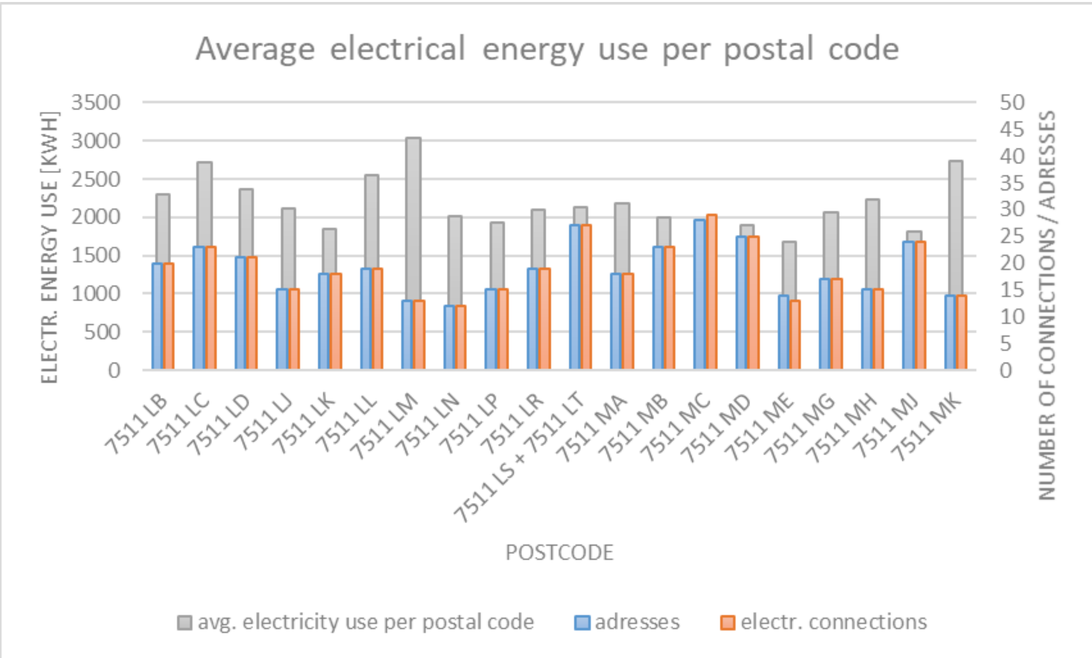
[26] https://www.acm.nl/sites/default/files/old_download/documenten/acm-energie/informatiecode-2015-01-01.pdf

Figure 30. Average gas consumption per postal code for considered neighbourhood (Hoge Bothofstraat)



Source: ENEXIS, 2020

Figure 31. Average electricity use per postal code for considered neighbourhood (Hoge Bothofstraat)



Source: ENEXIS, 2020

To extract the amount of gas used for space heating, the gas consumption data has to be disaggregated into gas consumption heating, cooking and domestic hot water. For doing so, Dutch standard figures have been used according to ECN (Menkveld, 2009). The results are shown in Table 7.

Table 7. Average residential gas consumption per function in NL (2006)

Function	Gas consumption [m^3]	Gas consumption [%]
Total	1652	100
Space heating	1212	73
Domestic hot water	375	23
Cooking	65	4

Source: Own elaboration, Saxion, 2020

From the Postcode 6 dataset, an overall energy use for space heating of 2.9 GWh can be derived for the neighbourhood. An averaged energy use for space heating per gas connection of 7800 kWh has been calculated, as shown in Table 8, whilst the total annual energy use for cooking, hot water and heating is shown in Table 9.

Table 8. Energy use for case study, total & per connection

Pos.	Electricity use [kWh]	Energy use for heating [kWh]
Total	812'665	2'950'683
Per connection (average)	2'139	7'765
Note: The Postcode 6 data shows 6 less gas connections than addresses for the neighbourhood. The analysis makes use of the number of connections, 374.		

Source: Own elaboration, Saxion, 2020

Table 9. Total annual energy use for cooking, hot water and heating based on Postcode 6 data

		Energy use	
Pos.	Function	[m^3 gas]	[kWh]
1	Cooking	13'838	135'189
2	Domestic hot water (DHW)	100'980	937'193
3	Space heating	317'928	2'950'683
Note: Conversion of m^3 to kWh is based on an upper caloric value of 35.17 MJ/m^3 for gas and an average boiler efficiency of 0.95.			

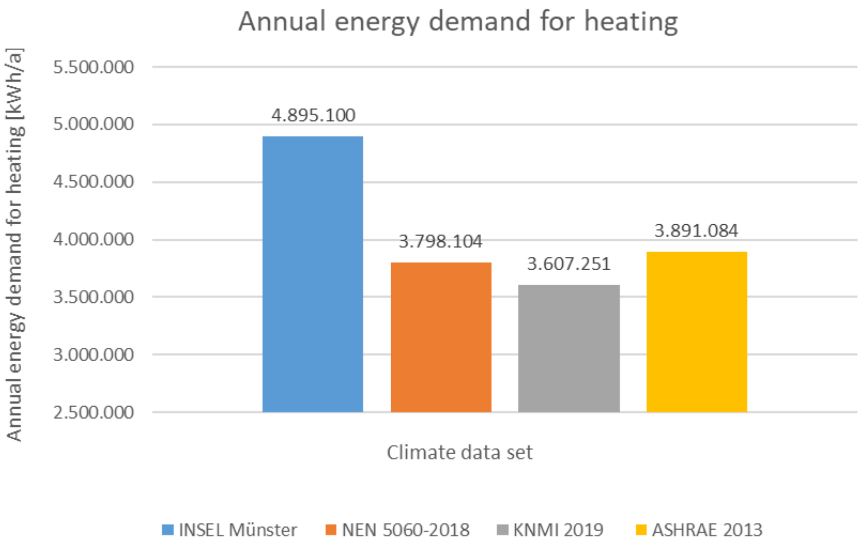
Source: Own elaboration, Saxion, 2020

5.8 Results

5.8.1 Simulation study, sensitivity of predicted energy use for heating

A simulation study has been conducted to determine the resulting deviation of the predicted annual energy demand for heating, as shown in Figure 32. Four climate datasets have been used for the analysis: INSEL Münster, NEN 5060:2018, KNMI 2019 and ASHARE 2013 for Twente.

Figure 32. Predicted annual energy demand for heating, sensitivity to climate datasets



Source: Own elaboration, Saxion, 2020

The predicted heating demand varies between min. 3.6 GWh to max. 4.9 GWh, representing a deviation of +17% and -12% around a mean value of 4.04 GWh, as shown in Table 10, in which the numbers are rounded off after conversion to GWh.

Table 10. Predicted annual energy demand for heating for 2019 using four different climate datasets

Pos.	Climate dataset	Total annual energy demand for heating [GWh]	Deviation from average (%)
1	INSEL Münster	4.9	+ 21
2	NEN 5060 - 2018	3.8	- 7
3	KNMI 2019	3.6	- 11
4	ASHRAE 2013	3.9	- 4
5	Average	4.05	n/a

Source: Own elaboration, Saxion, 2020

To determine the prediction accuracy of CityGML models in LOD 1 format two comparisons have been carried out. First a comparison with the national average energy use for heating and second with consumption figures derived from Postcode 6 data.

5.8.2 Plausibility check: Comparison with national average energy use data

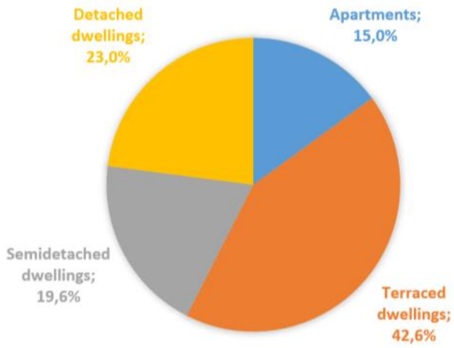
The annual energy use for heating based on Postcode 6 data has been determined to be 2.95 GWh for the neighbourhood, which translates to 7'800 kWh per dwelling per year. That is approximately 69% of the national average energy use for heating from 2017 (11'300 kWh / 2019)[27]. The difference can be quantified to be 31%.

Potential reasons for the difference can be explained by the homogeneity of dwelling types, occupancy use patterns and applied renovation measures. As there is no publicly available data available with respect to occupancy patterns of dwellings, and there is no data available which track the applied renovation measures per dwelling on a national scale, the comparison focuses on the homogeneity of the dwelling types in the considered neighbourhood.

[27] https://www.milieucentraal.nl/energie-besparen/inzicht-in-je-energierekening/gemiddeld-energieverbruik/

The Dutch residential building stock consists of 15% apartments, 42.6% terraced dwellings, 19.6% semi-detached and 23% detached dwellings[28], as shown in Figure 33.

Figure 33. Distribution of dwelling types in NL



Source: Own elaboration, Saxion, 2020

However, the considered neighbourhood does not contain detached dwellings and only a limited number of semidetached dwellings (end-of-terrace), which are typically associated with a higher energy consumption for heating due to their increased facade area. Data obtained for 2014 indicate that terraced dwellings constructed between 1975 and 1991 show a 36% reduced energy use for heating when compared to detached dwellings. Multifamily residences, apartments, show a 64% reduction when compared with detached dwellings[29]. These considerations lead to the conclusion that the national average energy use is only limitedly applicable for a comparison with the Postcode 6 data.

5.8.3 Comparison of predicted energy use for heating and Postcode 6 data

After having established the reliability of the normalized Postcode 6 data, this has been used for comparison with the predicted energy use for the neighbourhood. As the prediction using the Münster dataset did show a deviation from the mean of + 21 %, it has been disregarded for further analysis.

The results presented in Table 11 show a difference between Postcode 6 data and predictions ranging from 20% to 30%. The smallest difference of 20% is observed using a weather dataset (KNMI 2019) based on recorded weather data from the considered period.

Table 11. Comparison of predicted energy use for heating with Postcode 6 data

Pos.	Dataset	Total annual energy demand for space heating [GWh]	Deviation from reference (%)
1	Postcode 6	3.0	reference
2	NEN 5060 - 2018	3.8	+ 27
3	KNMI 2019	3.6	+ 20
4	ASHRAE 2013	3.9	+ 30

Source: Own elaboration, Saxion, 2020

Considering the uncertainty inherent to the analysis from sources such as state of renovation, heating system (natural gas boilers or electrical heating systems), occupancy, effect of normalization of Postcode 6 data, the 16% difference can be considered a good fit. However, it should be considered that the neighbourhood shows an exceptional homogeneity of dwelling types and a rather recent year of construction (1980). When modelling larger neighbourhoods with more different years of construction and dwelling types, increased uncertainty is expected, leading to an increased difference between measured and predicted energy use for space heating.

28 https://www.cbs.nl/nl-nl/nieuws/2016/14/vier-op-de-tien-huishoudens-wonen-in-een-rijtjeshuis
29 http://dspace.library.uu.nl/handle/1874/330473

The predicted deviations show a slight improvement of the prediction accuracy when using locally monitored data (KNMI 2019) from the considered period instead of using statistically composed climate datasets representative for a wider region such as NEN 5060-2018 and ASHRAE 2013. By using locally monitored data differences due e.g. to urban heat island effects can be excluded.

5.9 Conclusions

The main aim of the study was to identify a practical workflow for developing LOD 1 CityGML datasets and to determine the accuracy of SimStadt to predict the annual heating demand for a neighbourhood when compared with measured energy use data.

It was found that the workflow to develop LOD 1 CityGML datasets from publicly available data sources requires advanced skills and knowledge of at least three major software tools such as ArcGIS, FME workbench and SimStadt. These tools need to be used in sequence to develop and simulate LOD 1 CityGML datasets. To be able to use SimStadt with a Dutch case study, a local building physics library was developed, specific for The Netherlands.

The best prediction accuracy for the space heating energy demand was a +20% difference between measurements and predictions. It has to be noted that the measured energy use data originates from a network operator and contains standard normalized data to account for variations in climate conditions and natural gas composition.

It was found that nationally averaged energy use data is less suitable for comparison, as these do not account for the (in)homogeneity of dwelling types in the targeted neighbourhoods. Furthermore, it was found that although simulating on large scale 100+ building blocks, local climate data is best to be used for the analysis. In this study, locally measured climate data over the considered period outperformed a statistically-derived climate dataset representative for a period of 20 years (1986-2005).

The study shows that moving towards more detailed CityGML models such as LOD 2 does not necessarily contribute to an increased accuracy of the predictions, as long as uncertainties with respect to: (1) the availability of absolute and non-normalized gas consumption data for neighbourhoods; (2) the detailed differentiation of gas use for heating, domestic hot water and cooking; and (3) the current information on the actual thermal properties of the buildings (due to applied renovation measures) are taken into account.

If additional performance indicators are necessary to be reported such as kWh/m^2 floor area, more effort is needed to correct the CityGML data storey heights to better estimate floor area data.

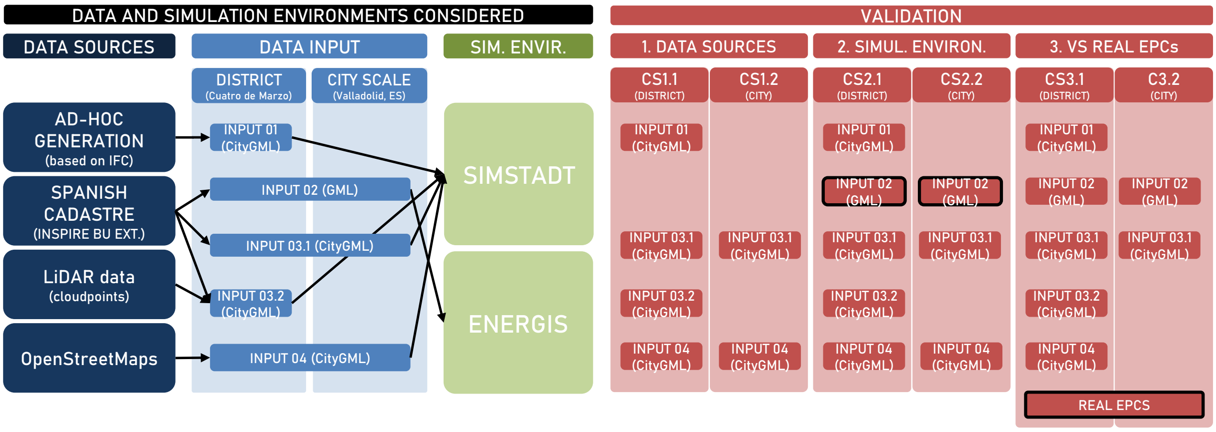
6 Simulations in ES (test area of Valladolid)

6.1 Introduction

In this section, the results of the simulations performed in Spain are presented. In particular, the objectives of these simulations are various, because of different comparison methods based on the available calculation methods and source data. In this line, the following sub-sections provide an overview of the simulation environments (sub-section 6.1.1), used data inputs (sub-section 6.1.2) and, finally, the proposed case studies (section 6.1.3).

Figure 34 illustrates the approach followed in this report and offers an overview of the different case studies in Spain. The left part of the figure provides details of the data and of the simulation environments used, whereas the right part is related to the validation processes carried out.

Figure 34. Overview of data, simulation environments and validation process proposed



Source: own elaboration, CARTIF, 2020

1. Data and simulation environment considered: four different data sources have been identified to generate the necessary data input to perform the calculations described in this section. In particular, (1) ad-hoc generated CityGML models based on IFC, (2) Spanish Cadastre data based on INSPIRE BU extended data model, (3) LiDAR data, as complementary source to detect building heights, and (4) OpenStreetMap[30], as collaborative database with potential to achieve consistent worldwide coverage.

From the processing of these data sources, different inputs in CityGML format (Inputs 01, 03.1, 03.2 and 04), and GML format (Input 02) have been generated, which have been simulated in the two simulation environments considered in this report (namely, SimStadt and ENERGIS [8]). A s it can be observed, all CityGML inputs (in 3D) are calculated with SimStadt, whereas GML input (in 2D) are calculated with ENERGIS (see Figure 34).

It is also worth to note the different scale of each of the inputs, which are at district scale (covering the Cuatro de Marzo district in Valladolid, Spain), or at city scale (covering the whole city of Valladolid).

2. Validation. Based on the input processed at different scales and the two simulation environments considered, a validation in three steps has been done. The first step is related to the impact of the generation of data input from different **data sources**. To compare the results of these case studies (CS1.1 at district scale), only results obtained with SimStadt are considered.

The second step is related to the differences encountered in the results when simulating with different **simulation environments** (i.e. SimStadt vs ENERGIS) in CS2.1 and CS2.2. As it can be seen, to the results already obtained in CS1.1 and CS1.2, two additional results marked in black can be observed: those derived from Input 02 at district and at city level.

In the third step, the results obtained with the simulations are compared with **real Energy Performance Certificates** (EPC), which share similarities with the simulation environments considered. In this case, real EPCs from the Castilla y León region have been considered.

[30] https://wiki.openstreetmap.org/wiki/Main_Page

6.1.1 Overview of simulation and validation environments

An overview of simulation and validation environments used for the simulations in the Spanish test area is provided in Table 12.

Table 12. Overview of simulation environments

	Name	Calculation method description
SIMULATION ENVIRONMENT	1. SimStadt	Tool developed by HFT Stuttgart to simulate energy urban models, where a CityGML model is required as an input (LOD 1 or LOD 2), as well as physical characteristics of the buildings considered in the simulations. To cover the second, a Buildings Physics Library of Germany is provided as a default library. However, in order to more accurately perform the simulations, it is possible to set up additional Building Physics Libraries or adapt the library to the characteristics of the building stock at hand. A Building Physics Library for the Netherlands is available at https://transfer.hft-stuttgart.de/gitlab/SimStadt/building-physics-library-nl.
	2. ENERGIS	Tool developed by CARTIF to simulate in a bottom up approach the energy demand (heating and cooling) of urban settings, by aggregating the results obtained at building level. The main aim of this tool is to deploy Energy Performance Certification tools validated in Spain (in particular CE3X[31]) and automate the process by automatically processing cadastral input (which follows INSPIRE BU extended data model), as well as a building physics library which is based on reference data based on the Building Code, which varies according to the building construction period as well as by climate zone. It is worth mentioning that the tool's scope only covered the residential sector, at the moment of the report submission.
VALIDATION ENVIRONMENT	1. Real EPCs	Real Energy Performance Certificates from the Junta de Castilla y León, even when not being a calculation method per se, are data used to compare the results obtained both from SimStadt and also from ENERGIS. The real EPCs are obtained from a public database offered by the Junta de Castilla y León, in particular by the EREN (Ente Regional de la Energía de Castilla y León). Updated every day, it is possible to consult four main values (CO_2 emissions, primary energy consumption, heating energy demand and cooling energy demand) of the registered EPCs in the Castilla y León region (Spain). These datasets are considered highly useful. However, not all the values of the Energy Performance Certificate are available to the public, which would enable to reproduce the calculations and check the accuracy of the calculations. Nevertheless, it is worth to mention that these datasets are provided at the same level of granularity as the submitted EPCs, that is, they can refer to an individual dwelling, or the whole building block, as well as to different uses (residential, retail, education, etc). These aspects should be considered when establishing comparisons.

Source: own elaboration

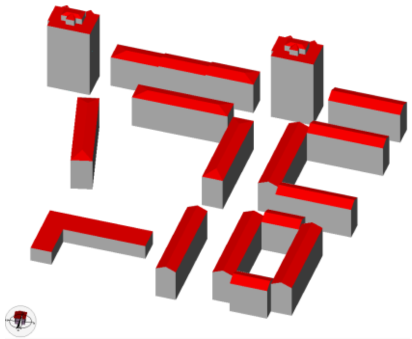
6.1.2 Overview of data inputs

As for the data input used, several data sources and building models have been deployed, which are briefly described below. For an in-depth description, please refer to section 6.2 and its corresponding sub-sections. The objective of selecting different data sources to generate building models is to compare their efficiency in the simulation and also to derive conclusions on how they differ from INSPIRE data. In particular, ad hoc generation of models (Input-01) is compared to INSPIRE-compliant data (Input-02), as well as to data generated based on the processing of INSPIRE-compliant data (Input-03.1 and Input-03.2) and other public sources such as OpenStreetMap (Input-04).

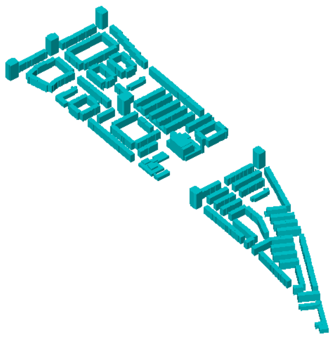
[31] http://www.efinova.es/CE3X

In each of the cases, it is indicated whether the data applies to district level (District marked in green and City marked in red), or whether it also covers the city level (District and City marked in green).

Table 13. Overview of data inputs

<table>
<tr><th colspan="3">Data input and brief description</th></tr>
<tr><td>Input-01: Cuatro de Marzo CityGML model</td><td>District</td><td>City</td></tr>
<tr><td>Figure 35. CityGML LOD2 of Cuatro de Marzo

Source: own elaboration, CARTIF, 2020</td><td colspan="2">Cuatro de Marzo is a district in the city of Valladolid, Spain, located in the Castilla y León region. This CityGML model, in LOD2, represents some of the buildings contained in this district, in particular 27 building blocks and 2 multi-family towers. The model was generated by CARTIF ad hoc for another H2020 project (in particular OptEEmAL [12]), based on the extraction of information from BIM standard data (in IFC format [13]). A more in-depth description on its generation can be found in section 6.2.1.</td></tr>
<tr><td>Input-02: Spanish cadastral input [INSPIRE BU extended]</td><td>District</td><td>City</td></tr>
<tr><td>Figure 36. Valladolid cadastral data

Source: Spanish Cadastre GML, 2020</td><td colspan="2">The Spanish cadastre offers information on the INSPIRE data themes of Buildings, Cadastral Parcels as well as Addresses. These are useful sources to analyse the built environment. In this case, the most useful data theme to be deployed is BU, which is based on the extended version of this data theme, where some modifications have been performed. The data used is directly downloaded from the ATOM services offered by the Spanish Cadastre [14]. However, it is after this download when the data is enriched and ready to be processed with a specific simulation environment. A more in-depth description of how these data is pre-processed can be found in section 6.2.2.</td></tr>
</table>

Input-03.1: Valladolid CityGML model [based on INSPIRE BU extended]	District	City
Figure 37. Valladolid CityGML model based on cadastral data  *Source*: own elaboration, CARTIF, 2020	Using Input-02 data, a CityGML model has been generated using FME software. As a result, a CityGML model with Level of Detail 1 (LOD1) has been generated, considering an average floor height of 2.7m. More information on the generation of this model can be found in section 6.2.3.	
Input-03.2: CityGML model [based on INSPIRE BU extended + LiDAR data]	District	City
Figure 38. Cuatro de Marzo CityGML model from cadastre + LiDAR data  *Source*: own elaboration, CARTIF, 2020	In a similar manner to the process followed in Input-03.1, Input-03.2 has been generated. The basis of this model continues to be the Spanish Cadastral data. However, the main difference to the abovementioned Input-03.1 is the application of mean real heights of the building extracted from the analysis of LiDAR data. LiDAR cloud points were pre-processed to improve the model results. As a result, a CityGML model with Level of Detail 1 (LOD 1) was generated.	
Input-04: Valladolid CityGML model [based on OSM data]	District	City
Figure 39. Valladolid CityGML model from OSM  *Source*: own elaboration, CARTIF, 2020	Using OpenStreetMap data, a CityGML model has been generated using FME software [15]. Because OpenStreetMap is a collaborative approach and not every area is described with the same amount of detail, specific assumptions were made to generate this model. More information on the generation of this model can be found in section 6.2.5.	

Source: own elaboration, CARTIF, 2020

Another highly relevant issue to consider when performing calculations or simulations of the building stock is the enrichment of the model with information of the building use of each building, as well as their thermal characteristics. In this case, since the study is restricted to analysing energy demand and not energy consumption, there is no need to characterise the energy systems contained in the buildings and calculate the fuel consumption for heating or cooling.

If an energy audit of a building was to be performed, specific characteristics of the building would need to be determined (by onsite observations performed by experts or even measurements), in order to obtain relevant parameters such as the thermal transmittance of the building's envelope. If defined well, this is a tedious process which involves knowing for instance, the number of layers a façade has, as well as the building materials and their characteristics (conductivity and density among other).

In contrast to these procedures, the definition of these thermal characteristics in urban energy simulation tools is usually performed through the generation of building typologies and the corresponding allocation of these building typologies to the set of buildings being analysed, according to variables such as year of construction, climate zone, use of the building, etc.

In the simulations reported in this section, two sets of building typology libraries are used: the one used by default in SimStadt (German Building Physics Library), and the one used by the ENERGIS tool. Even when these libraries can be modified and adapted in both tools, the simulations shown in this report use the original ones from each of the tools. This is an important aspect to be highlighted, which affects especially the results obtained in case studies 2 and 3, since in the first case study only models coming from SimStadt are compared among each other.

6.1.3 Overview of case studies proposed

By combining the available data sources, models, calculation methodologies and building physics libraries, the following case studies are proposed. In Table 14 each case study is defined in terms of its (i) objective, (ii) scale, (iii) simulation environment used, and (iv) data input deployed. In section 6.4, the results of these case studies can be observed.

Table 14. Overview of case studies proposed

<table>
<tr><th colspan="10">Case Study 1 (CS1). Different dataset generation</th></tr>
<tr><td colspan="10">Objective Using the same simulation environment, the objective is to compare the results obtained from by using input datasets for the same area, which have been generated following different methods. In particular, the generation of ad-hoc models (Input-01), models based on cadastral data (Input-03.1 and Input 03.2) or models based on OpenStreetMap data (Input-04) have been explored. This has been explored both at the district scale (CS1.1) as well as at the city scale (CS1.2).</td></tr>
<tr><td>Case study code</td><td>Scale</td><td>SimStadt</td><td>ENERGIS</td><td>Input 01</td><td>Input 02</td><td>Input 03.1</td><td>Input 03.2</td><td>Input 04</td><td>Comparison</td></tr>
<tr><td>Case Study 1.1</td><td>District</td><td>X</td><td>-</td><td>X</td><td>-</td><td>X</td><td>X</td><td>(X)</td><td>Section 6.4.1.1</td></tr>
<tr><td>Case Study 1.2</td><td>City</td><td>X</td><td>-</td><td>-</td><td>-</td><td>X</td><td>-</td><td>(X)</td><td>Section 6.4.1.2</td></tr>
<tr><th colspan="10">Case Study 2. Different simulation environments</th></tr>
<tr><td colspan="10">Objective: Once the different dataset generation has been explored, the objective is to compare the results with another simulation environment, in this case with ENERGIS, which only allows as input data the one obtained directly from the cadastre. In this line, the results obtained in CS1 with the SimStadt simulations have been compared to those obtained with ENERGIS, both at the district scale (CS2.1), and at the city scale (CS2.2).</td></tr>
</table>

Case study code	Scale	SimStadt	ENERGIS	Input 01	Input 02	Input 03.1	Input 03.2	Input 04	Comparison
Case Study 2.1	District	X	X	X	X	X	X	(X)	Section 6.4.2.1
Case Study 2.2	City	X	X	-	X	X	-	(X)	Section 6.4.2.2
Case Study 3. Comparison of results with real Energy Performance Certificates									
Objective: Finally, the objective of this case study is to compare the results of CS1 and CS2 with real Energy Performance Certificates (Validation environment 1. Real EPCs). Similarly, this has been tested at the district scale (CS3.1) as well as at the city scale (CS3.2). As it can be seen from the table below, the same inputs and simulation environments are considered.									
Case study code	Scale	SimStadt	ENERGIS	Input 01	Input 02	Input 03.1	Input 03.2	Input 04	Comparison
Case Study 3.1	District	X	X	X	X	X	X	(X)	Section 6.4.3.1
Case Study 3.2	City	X	X	-	X	X	-	(X)	Section 6.4.3.2

Source: own elaboration, CARTIF, 2020

6.2 Data input

In this section, the different data inputs as reflected in Table 13 are explained in more depth, as well as a description of the pre-processing required. This depends on the simulation environment where these datasets have been deployed, as presented in Table 15 and widely explained in section 6.3.

However, the main areas tackled in the data input cover two main scales (district scale and city scale) within the same city in Spain, in particular, Valladolid. In order to provide some context before entering into the explanation of the models used, the following tables are provided.

Table 15. Contextual data of the city of Valladolid, Spain

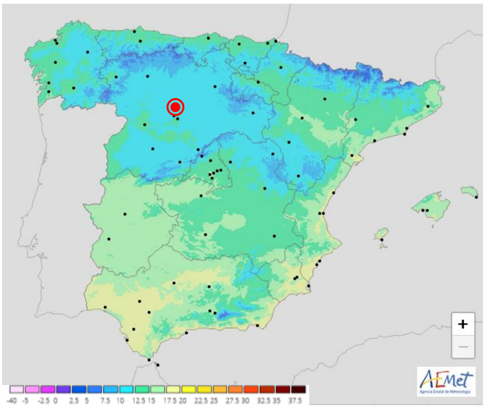
City scale: Valladolid, Spain			
Altitude	690 m	Buildings / population	17.046 / 298.412 inhabitants
Heating deg. day (HDD)	3121	Average winter temp.	5ºC
Cooling deg. day (CDD)	394	Average summer temp.	20,5ºC
The city of Valladolid (41°39′07″N 4°43′43″O) is the capital of the Castilla y León region, located to the north-west of Spain, counts on 298.412 inhabitants and has a total surface of 197,91 km^2 (population density of 1514,4 inhabitants / km^2).			

Figure 40. Location of Valladolid within Spain, its province and with respect to its climate (Köppen)

Source: Wikipedia and AEMET

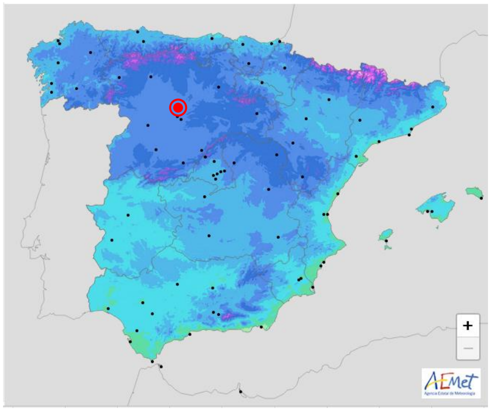
With respect to its climate, according to the Köppen classification[32] it corresponds to Csa and Csb climates. This is due to the fact that the city is located in the valley of the Duero river and it is surrounded by mountain ranges that isolate the area where the city is located from the sea. The climate is extreme and dry. The following figures from AEMET shows the characterisation of the city in terms of its weather parameters can be seen.

Figure 41. Yearly average temperature



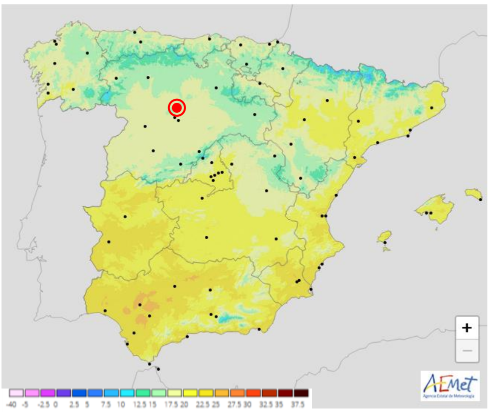
Source: AEMET

Figure 42. Yearly min. average temperature



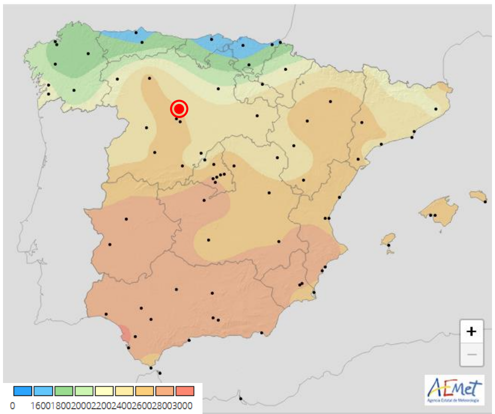
Source: AEMET

Figure 43. Yearly max. average temperature



Source: AEMET

Figure 44. Isolation map



Source: AEMET

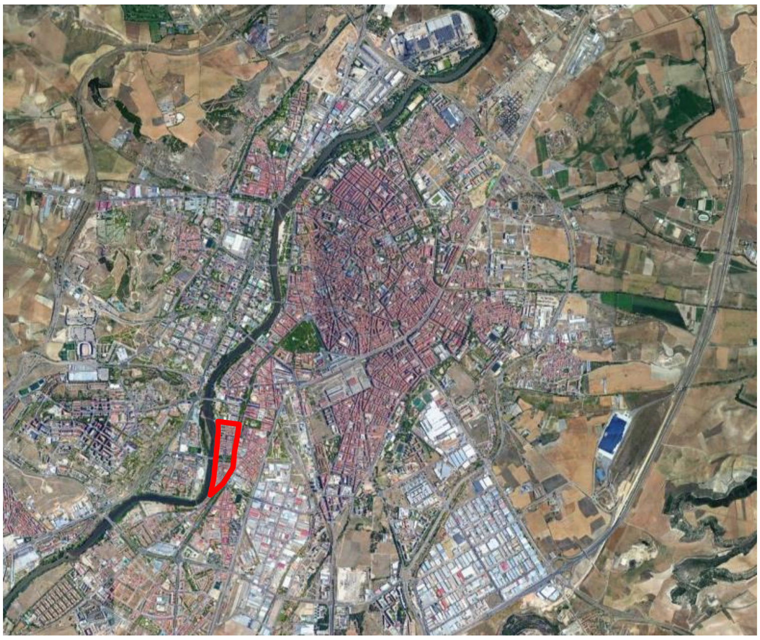
District scale: Cuatro de Marzo District, Valladolid, Spain			
Altitude	690 m	Number of buildings	183 blocks 6 towers

[32] https://www.nationalgeographic.org/encyclopedia/koppen-climate-classification-system/

Population	3750 inh.	Num. of dwellings	1941 dwellings
Construction date	1960-1970	Conditioned area	166.000 m^2

Within the city of Valladolid, Cuatro de Marzo district covers a broad area of the city close to the river Pisuerga (see Figure 44). It corresponds to one of the several districts where Valladolid grew when there was an increased demand of dwellings in the 60's.

Figure 45. Cuatro de Marzo district within Valladolid, Spain



Source: own elaboration based on Google Maps data, CARTIF, 2020

Due to this fact, the buildings contained in the district were built in the same period and following the same building standards. Also, in terms of residential buildings, there are only two main categories of buildings: those which are building blocks, which are placed following different orientations and form lines or clusters of buildings with an internal space; or multi-family towers, which are also located following different orientations. Moreover, some of the buildings of this district have undergone some retrofitting actions within the R2Cities project ("Residential Renovation towards nearly zero energy CITIES") [16], since the district was a demo site of this FP7 project. As a consequence, the study of this area is especially interesting. However, in the input data and calculations provided, only a selection of buildings in the north of the district was considered.

6.2.1 Input-01: Cuatro de Marzo CityGML model

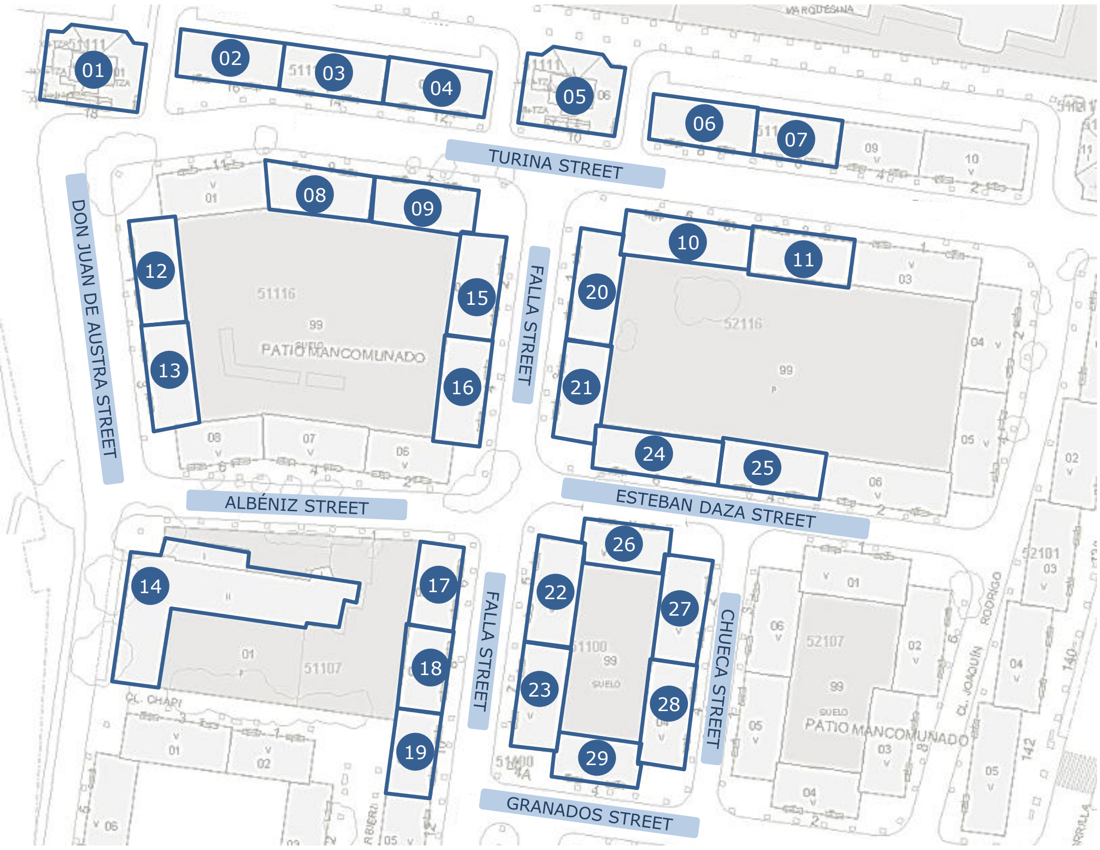
The Cuatro de Marzo CityGML model used in the calculations was created for another project, in particular for the OptEEmAL H2020 project ("Optimised energy efficiency design platform for refurbishment at district level") [12], since Cuatro de Marzo was one of the case studies and the model was to be inserted within the platform to complement the energy calculations performed with it. In particular, to work with the platform and apart from contextual data and objective data inserted by the user, it was necessary to introduce one IFC model per building which was subject of retrofitting, as well as a CityGML model describing all of the buildings subject to retrofitting as well as neighbouring buildings which could cast a shadow over the selected buildings. In this context, the models holding most of the information were the IFC models (BIM standard) [13], whereas the CityGML model was used as a complement to describe the district situation in terms of location, and shadows.

For this process, firstly the IFC models were generated by ACCIONA (partner in the OptEEmAL project), whereas the CityGML model was created by CARTIF based on the inputs from the IFC. The main objective of this transformation was to decrease the level of detail from the IFC (which would correspond more or less to LOD 4 in CityGML terminology) to LOD 2 in CityGML, which was the level required for the purposes within the OptEEmAL platform. This ad hoc process was performed using the software Sketch-Up together with a plug in called CityEditor [17].

6.2.1.1 *Specific description of the dataset*

As mentioned before, the model of the Cuatro de Marzo district does not cover the whole extension of the district, and a few buildings to the north are selected which represent the main typologies found in the district: linear blocks and towers. Figure 45 shows the selection of buildings, marked over the Spanish Cadastre cartography.

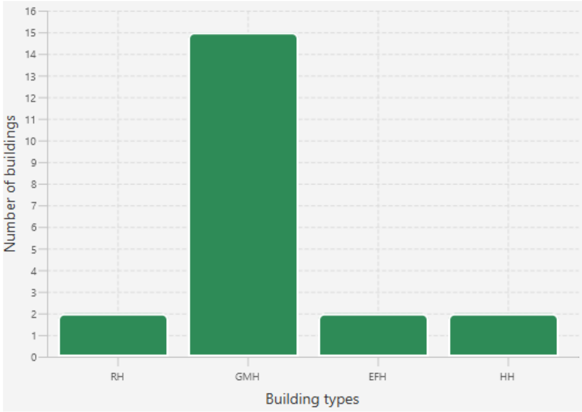
Figure 46. Selected buildings of Cuatro de Marzo district in Input-01



Source: own elaboration, CARTIF, 2020

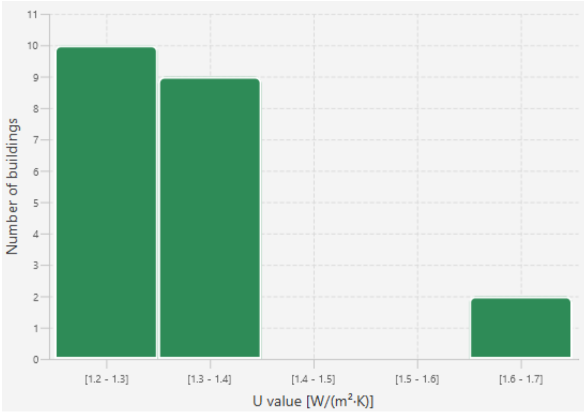
In the model, some of the buildings have been joined in order to form a single building, whereas in three cases (buildings 1, 5 and 14), the building has been subdivided into two parts. The figures below (Figure 47, Figure 48, Figure 49 and Figure 50) provide more information with respect to the main characteristics of the buildings. It must be highlighted that all of the buildings with a unique GMLid are devoted to only one function (ALKIS code = 1000 is residential and 2100 is office and administration), and that all are considered to not have undergone any refurbishment.

Figure 47. Buildings per type - Input01 District



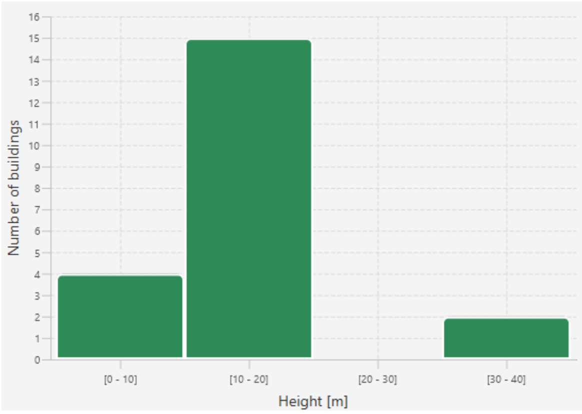
Source: own elaboration, based on SimStadt processing

Figure 48. Buildings per U-value range - Input01 District



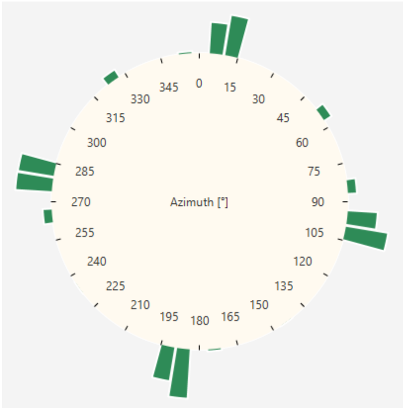
Source: own elaboration, based on SimStadt processing

Figure 49. Buildings per height range - Input01 District



Source: own elaboration, based on SimStadt processing

Figure 50. Wall orientation - Input01 District



Source: own elaboration, based on SimStadt processing

6.2.1.2 Pre-processing required

The required pre-processing of the data was guided by the needs of the SimStadt simulation environment, since it is only in this context that the model has been deployed in the case studies proposed. In order to enter the SimStadt tool, a direct test with the model was performed, which was unsuccessful. To detect the potential mistakes, a checking was performed using the CityDoctor software[33], also developed by HFT Stuttgart. This software implements methods and metrics for analysis, testing and correction of syntax, geometry and semantics of 3D city models. Once these aspects were corrected, the model was simulated appropriately in SimStadt.

Moreover, and even when relating to the results obtained from simulating with SimStadt and not to the inputs per se, it is relevant to comment on the results obtained due to the post-processing that was needed because of how the model was generated. The results obtained from SimStadt are a .csv file with:

— **Buildings identification**: GML id, ParentGMLId, Latitude, Longitude, LOD, year of construction, refurbishment status (original or refurbished).

— **Buildings use**: ALKIS code and primary and secondary usages (as well as their area), building type

— **Building geometry**: total wall area, footprint, shared walls, volume, heated volume, etc.

— Building's thermal characteristics: mean U-value

[33] https://www.citydoctor.eu/index.php/citydoctor_main.html?language=en [last access October, 2020]

— **Energy demand**: specific heating and cooling demand, total heating and cooling demand (yearly and disaggregated per month).

— Domestic hot water demand

These results allow to analyse in detail a selected building area. However, problems were encountered when processing the results, since the buildings in the model were originally grouped or divided into building parts. As a consequence, no individual results per cadastral reference (at building level) were available and this fact impeded their direct comparison. Thus, in order to be able to compare the results derived from this Input 01, it was necessary to fine-tune them by allocating to each real building its corresponding heating and cooling demand. This way, a result was obtained per cadastral reference (at building level).

It should be highlighted that this post-processing of the results obtained was required in Input-01 because the model had been developed for another purpose (OptEEmAL), where it did not matter if individual buildings were grouped with their neighbours.

6.2.2 Input-02: Spanish cadastral input [INSPIRE BU extended]

The Spanish Cadastre offers official and highly relevant information with respect to buildings, cadastral parcels and addresses, following the INSPIRE guidelines of the themes: Buildings (BU), Cadastral Parcels (CP) and Addresses (AD). The easiest way to access the data, apart from the web viewer, is through the ATOM services, which are available through the following link[34]. Additionally, WFS and WMS services are available as well.

In order to understand the content of the potential Spanish Cadastral input, the following tables (Table 16 and Table 17) describing the attributes "Building" and "BuildingPart" are presented. Marked in grey are those attributes that exist in the INSPIRE definition but are not deployed in the Spanish Cadastre.

Table 16. Attributes and contents within Building in the Spanish Cadastre

Building	
Attribute Name	Description of contents in Spanish Cadastre
gml:FeatureCollection	Heading GML object where the extended 2D Building scheme is defined. It has the *gml:id* "2ES.SDGC.BU".
gml:featureMember	Structure containing every building object.
bu-ext2d:Building	Main structure with a *gml:id* composed of values defined in *inspireID* and it is a unique identifier for all the data group.
gml:boundedBy	Structure that defines the rectangle covering the geometry of the object defined by its low-left and above-right coordinates. The coordinates have been defined in the reference system indicated in srsName.
bu-core2d:beginLifespanVersion	Date when the data has been submitted to the cadastral data base.
bu-core2d:conditionofConstruction	Values that expresses the condition of the construction. It can be: "ruin", "declined" or "functional". In case that in the same parcel there are different units, this value has been the best among them.

[34] http://www.catastro.minhap.es/webinspire/index.html [last access October, 2020]

bu-core2d: dateOfConstruction	Structure that defines the date of construction. It is composed by two attributes: *bu-core2d:beginning* and *bucore2d:end*. If there is more than one building unit, in the *beginning* field the oldest date is adopted and in the *end* field the newest. They are always referenced to the 1st of January.
bucore2d: endLifespanVersion	Date when the data has been deprecated. This value is not defined since it does not provide historical information.
bucore2d: externalReference	Structure where the URL to the direct access to the cadastral information in the Sede Electrónica del Catastro is added (field *building2d:informationSystem*. The field *bu-core2d:referenc*" contains the reference to the cadastral parcel.
bu-core2d:inspireId	Unique identifier for all the groups of data in INSPIRE. It is composed by a base:Identifier structure with the two following values.
base:localId	First 14 characters of the cadastral reference.
base:namespace	For buildings it is "ES.SDGC.BU", which corresponds to the acronym of the country, producer entity and the group of data.
bu-core2d:addresses	Address object, through a *xlink:href* the WFS service of the address(es) associated to the building can be accessed.
bu-core2d:cadastralParcels	Cadastral parcel object, through a *xlink:href* the WFS service of the cadastral parcel associated to the building can be accessed.
bu-ext2d:geometry	Geometry of the building in GML. It is a *gml:MultiSurface* structure that can hold several *gml:Surface*. These objects have to have a unique *gml:id* composed by the *gml:id* of the cadastral zoning and a prefix and a suffix. The geometry is defined by exterior ring vertices and holes can exist which are defined in an interior ring structure. The coordinate list of the rings (*gml:postList*) duplicates the first and last vertex. The exterior one is defined clockwise and the interior one counter-clockwise. The reference system is the one defined in *srsName*. It holds the other two attributes defined below.
bu-core2d: horizontalGeometryEstimatedAccuracy	Accuracy in meters, which adopts the value of0.1.
bu- core2d:horizontalGeometryReference	Indicates that the geometry of the building is the footprint of what is built above ground. It has the value: "footprint".
bu-ext2d:currentUse	Predominant use of the building. The value is obtained calculating the use that covers more surface in the cadastral parcel where the building is located. The following values are admitted: “1_residential”, “2_agriculture”, “3_industrial”, “4_1_office”, “4_2_retail”, “4_3_publicServices”.
bu-ext2d:numberOfBuildingUnits	Number of properties of the cadastral parcel that are contained in the building.

bu-ext2d:numberOfDwellings	Number of properties of the cadastral parcel that are contained in the building with a residential use.
buext2d: numberOfFloorsAb oveGround	Number of floors of the building. This data cannot be provided at building level, since in the Spanish cadastral data model the volume cannot be delimited for the complete building, it is a value which is reflected in *BuildingPart*.
bu-ext2d:document	Structure where in the field *bu-ext2d: documentLink* an URL is provided with access to an image of the façade. It is possible that the query will not provide an image if this is not contained in the data base. The structure includes the field *bu-ext2d: format* with the value "jpeg" and the field *bu-ext2d:sourceStatus* with the value "NotOfficial".
bu-ext2d:officialArea	Structure that represents the surface of the building in square meters in the field *buext2d:value* and the type of Surface measured, which will always be "grossFloorArea" in the field *bu-ext2d:officialAreaReference*.

Source: Spanish Cadastre

Table 17. Attributes and contents within Building Part in the Spanish Cadastre

BuildingPart	
Attribute Name	Description of contents in Spanish Cadastre
gml:FeatureCollection	Heading GML object where the extended 2D Building scheme is defined. It has the *gml:id* "ES.SDGC.BU".
gml:featureMember	Structure containing every building part.
Bu-ext2d:BuildingPart	Structure of each part of a building has a *gml:id* composed by the values defined in *inspireID* and it is a unique identifier for all the group of data. Its value is the identifier of the building with the suffix "partX", being X a sequential number.
bucore2d: beginLifespanVersion	Date when the data has been submitted to the cadastral data base.
bucore2d: conditionofConstruction	It has no value for building parts.
bu-core2d:inspireId	Unique identifier for all the groups of data in INSPIRE. It is composed by a *base:Identifier* structure with the two following values.
base:localId	First 14 characters of the cadastral reference and a sequential suffix "partX".
base:namespace	For buildings it is "ES.SDGC.BU", which corresponds to the acronym of the country, producer entity and the group of data.
bu-core2d:addresses	Address object, through a *xlink:href* the WFS service of the address(es) associated to the building can be accessed.

bu-core2d: cadastralParcels	Cadastral parcel object, through a *xlink:href* the WFS service of the cadastral parcel associated to the building can be accessed.
bu-ext2d:geometry	Geometry of the building part in GML. It is a *gml:MultiSurface* structure that can hold several *gml:Surface*. These objects have to have a unique *gml:id* composed by the *gml:id* of the cadastral zoning and a prefix and a suffix. The geometry is defined by exterior ring vertices and holes can exist which are defined in an interior ring structure. The coordinate list of the rings (*gml:postList*) duplicates the first and last vertex. The exterior one is defined clockwise and the interior one counter-clockwise. The reference system is the one defined in *srsName*. It holds the other two attributes defined below.
bucore2d: horizontalGeometryEstimatedAccuracy	Accuracy in meters, which adopts the value of 0.1.
bucore2d: horizontalGeometryReference	Indicates that the geometry of the building is the footprint of what is built above ground. It has the value: "footprint".
buext2d: numberOfFloorsAboveGround	Number of floors above ground.
buext2d: heightBelowGround	Height of the floors below ground in meters. It is an estimated height of 3m/floor.
buext2d: numberOfFloorsBelowGround	Number of floors below ground.

Source: Spanish Cadastre

6.2.2.1 Specific description of the dataset

The dataset to be considered is the one corresponding to the municipality of Valladolid. In particular this dataset contains 17.046 elements in the BU GML and 106.219 elements in the BU PART GML. It is a 2D representation in GML format of the city following the abovementioned specifications, which has been downloaded from the cadastre from their ATOM services.

Figure 51. Input-02: City scale - Valladolid GML model

Source: Spanish Cadastre

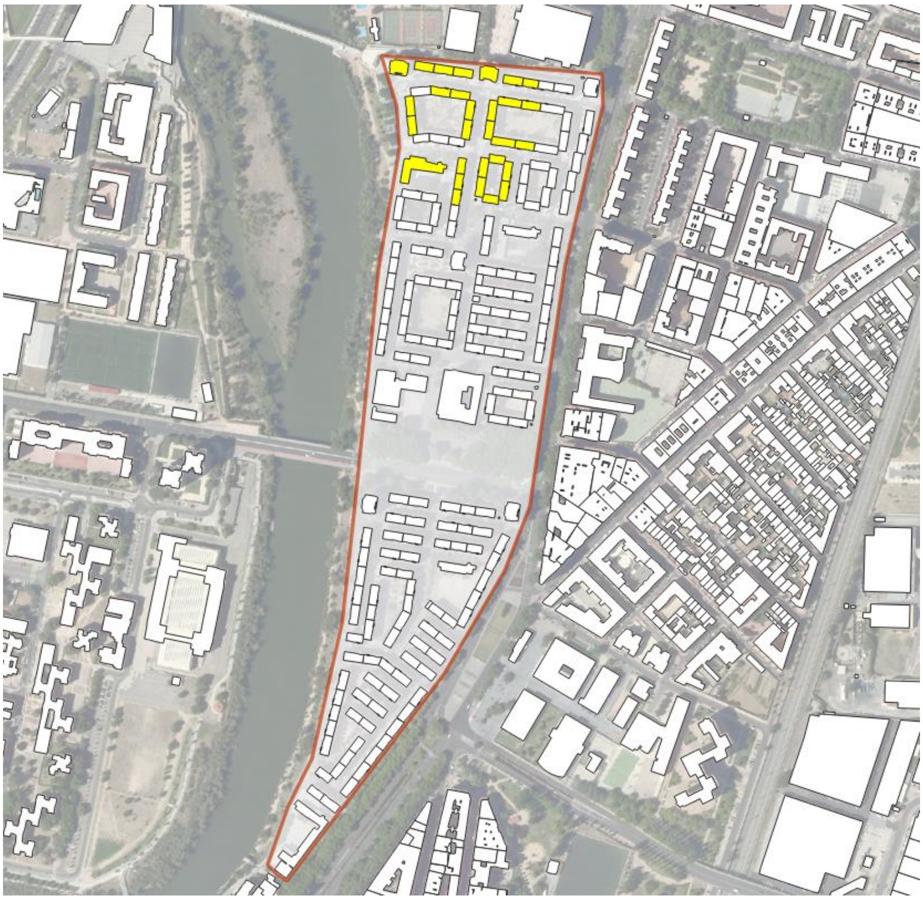
In the cases where the district scale has been analysed, an extraction of this data has been performed, selecting only those datasets relevant to the area, but with no different treatment than that of the city scale (marked in the above Figure 51). In order to exemplify the selected datasets, a figure is shown in the Figure 52 below. This model contains 207 buildings.

As it can be observed, the selected district of Cuatro de Marzo is broader in scope than the selection of buildings shown in Input-01. This is the reason why the selected buildings of Input-01 have been extracted as well, in order to be able to compare these results in the proposed case studies. This is shown in the following Figure 52, containing 29 buildings.

Figure 52. Input-02: District scale – Cuatro de Marzo GML model

Source: own elaboration, CARTIF, 2020

Figure 53. Input-02: District scale (smaller version) - Valladolid GML model



Source: own elaboration, CARTIF, 2020

6.2.2.2 Pre-processing required

These datasets are used in the simulation environment of ENERGIS. Within it, apart from the geometrical characterization of the buildings, the building physical characteristics should be applied. This has been explained in section 6.3.1.2; however, the hypothesis applied to the geometrical parameters are explained in Table 18.

Table 18. Spanish cadastral data – Pre-processing and hypothesis towards it use within ENERGIS

Spanish Cadastral data – Pre-processing and hypothesis towards its use within ENERGIS	
Item	Pre-processing action or hypothesis applied
Generic information about the building	No hypothesis is applied; however, it is worth to mention that buildings are characterised by their address, municipality, region, postal code, etc; and the main identifier for each building is the cadastral reference. Nevertheless, in terms of calculations, building parts are considered, in order to be able to distinguish among different volumes present in each building and their respective heights. This is performed this way due to the fact that the number of floors in the Spanish Cadastre are provided in each building part and not at building level.
Building use	Even when the cadastre provides 7 different types of building uses (1_residential, 2_agriculture, 3_industrial, 4_1_office, 4_2_retail, and 4_3_publicServices), in ENERGIS only residential buildings are considered, that is, those with a building use of 1_residential.
Building type	Even when differentiations among buildings according to their typologies could be performed based on additional geometric processing (to determine for instance if there are individual houses, multi-family building blocks, etc), the differentiation among buildings is only performed based on its use.
Climate zone	Based on the location of the building, its winter and summer climate zones are assigned. This has been consulted in ENERGIS in order to assign the appropriate reference climate data.
Conditioned surface	Obtained from cadastre, assumed to be gross surface.
Number of floors above ground and below ground	Number of floors above ground and below ground are obtained directly from the cadastre, per building part. This data is the basis to calculate the height of the buildings.
Floor height	Assumed to be 2,7m / floor. This is an assumption, as this data is not made explicit in the cadastre.
Roof	All of them are considered flat, since no information of the tilting of the roofs is calculated in this methodology.
External walls and shared walls	The total area is be calculated by multiplying the perimeter of the footprints with the building heights (calculated as explained above). It is worth to mention that the orientation of each wall is extracted, as well as its classification as external or neighbouring wall, due to the huge impact this aspect has on the energy performance of buildings.
Openings	Information about opening is not provided in the cadastre; however, estimated window wall ratios can be established. However, this has not been performed by directly using a percentage depending on the building type, but by analysing each external wall, and estimating the number of pillars that could fit in that specific wall. Then a percentage of windows was calculated by

	considering that windows were 1 meter high and were placed at 1 meter from the floor, with a longitude that varied.
Thermal bridges	Similarly, thermal bridges are calculated, since they are important heat sinks for buildings. Following a similar approach to that of the openings, thermal bridges are calculated for pillars in edges, within the wall, opening contour, etc.

Source: own elaboration, CARTIF, 2020

6.2.3 Input-03.1: Valladolid CityGML model [based on INSPIRE BU extended]

Input-03.1 is based on the same data source used in the previous case, the Spanish cadastre.The main difference is the enrichment process performed on this dataset. In this case, FME has been used in order to:

- **Generate a CityGML model LOD 1** (based on the number of floors defined in the cadastre, an average floor height has been assigned)
- **Enrich the model with building's use and year of construction** (since these are the main parameters for energy demand calculation)

Then, based on this characterisation, it was possible to assign to the model its corresponding typology from the German Buildings Physics library, to be able to simulate it with SimStadt. In particular, following the correspondence shown in the next table:

Table 19. Correspondence of ALKIS codes and Spanish Cadastre uses

Use category	ALKIS Code	Spanish Cadastre uses
residential	1010	1_residential
company building	2112	4_2_retail
office building	2020	4_1_office
administration building	3010	4_3_publicServices
business building	2050	4_1_office
industrial building	2110	3_industrial
agriculture building	1220	2_agriculture
other use	9999	

Source: own elaboration, CARTIF, 2020

6.2.3.1 *Specific description*

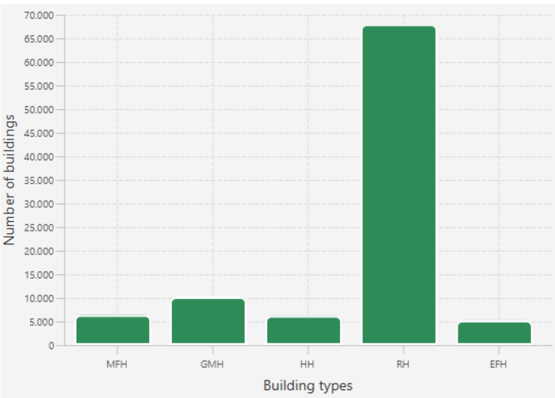
This model covers the whole city of Valladolid (see Figure 54).

Figure 54. Input-03.1: City Scale – Valladolid model based on INSPIRE BU extended (cadastral data)

Source: own elaboration, CARTIF, 2020

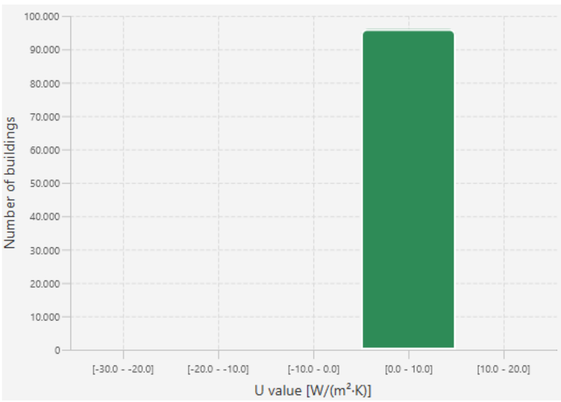
The whole model is formed by **97.670 buildings** with the building typologies observed in Figure 55, and the number of buildings per U-value range (Figure 56). Figure 57 shows that the majority of buildings are below 10 meters high. However, it should be highlighted that this model is generated based on the building parts, that is, the individual volumes that build up each building. This is the reason why this value is so high.

Figure 55. Buildings per type - Input03.1 City



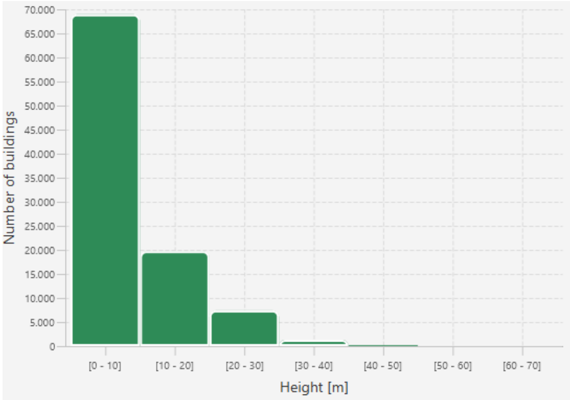
Source: own elaboration, CARTIF, 2020

Figure 56. Buildings per U-value range– Input03.1 City



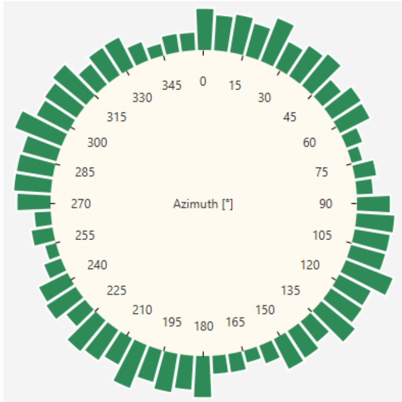
Source: own elaboration, CARTIF, 2020

Figure 57. Buildings per height range- Input03.1 City



Source: own elaboration, CARTIF, 2020

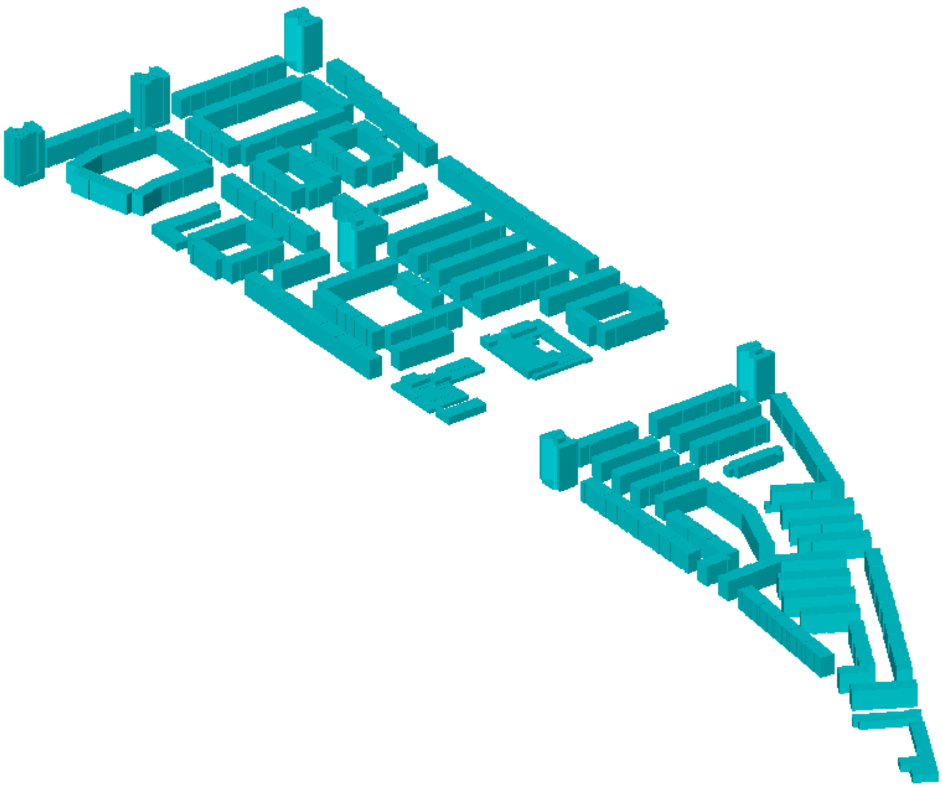
Figure 58. Wall orientation- Input03.1 City



Source: own elaboration, CARTIF, 2020

In order to analyse the district scale, an extraction of this model has been carried out with the Regions Processor tool contained within SimStadt. As a result, the following model has been extracted and used for the district scale comparisons:

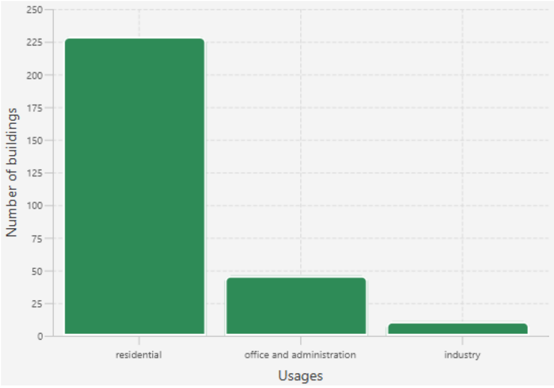
Figure 59. Input-03.1: District Scale – Cuatro de Marzo model based on INSPIRE BU extended (cadastral data)



Source: own elaboration, CARTIF, 2020

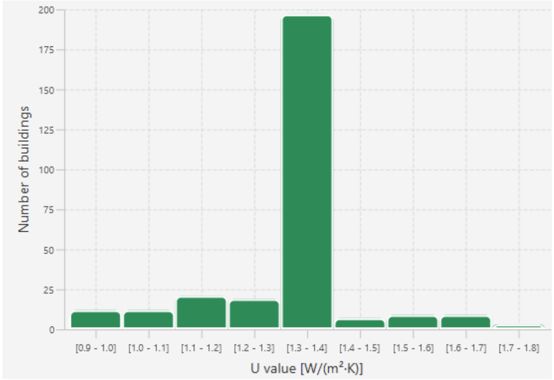
The district model corresponding to Cuatro de Marzo is formed by **289** buildings with the typologies observed in Figure 60, and the number of buildings per U-value range (Figure 61).

Figure 60. Buildings per type - Input03.1 District



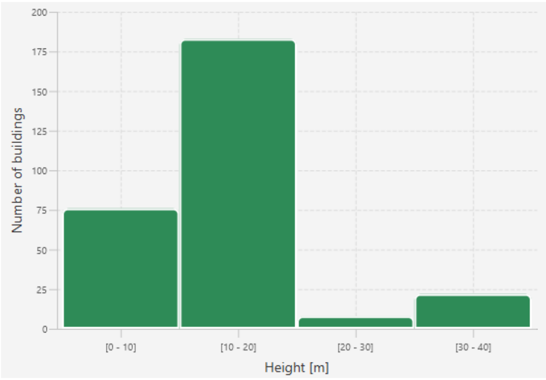
Source: own elaboration, based on SimStadt processing

Figure 61. Buildings per U-value range– Input03.1 District



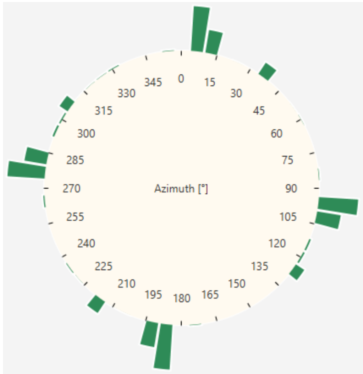
Source: own elaboration, based on SimStadt processing

Figure 62. Buildings per height range- Input03.1 District



Source: own elaboration, based on SimStadt processing

Figure 63. Wall orientation- Input03.1 District



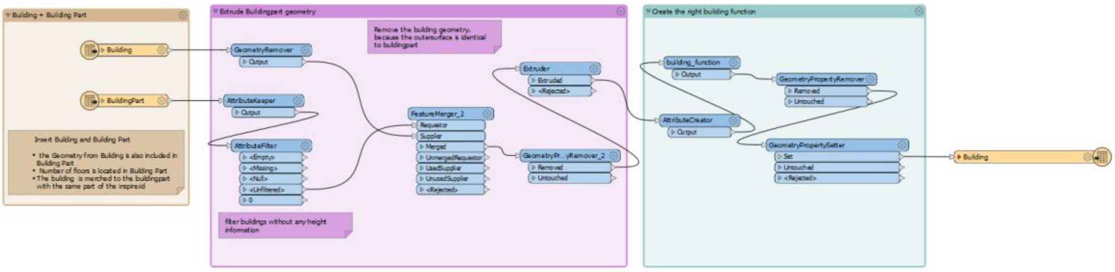
Source: own elaboration, based on SimStadt processing

6.2.3.2 Pre-processing required

This CityGML model in LOD 1 is to be used within SimStadt. In this line, it is necessary to obtain geometric data, as well as information of the building use and its year of construction. Only with these values it is possible to assign a building typology and simulate it with this tool. With this objective in mind, and considering the available inputs in the Spanish cadastre, the FME model shown in Figure 63 is proposed to generate the CityGML model.

As it can be seen, inputs are required from both Building and Building Part. The model merges both inputs using the identifier provided by cadastre once the geometries of the Buildings are removed and the Building Parts are filtered. After that, Building Parts are extruded using the number of floors and are enriched by the building use and its year of construction. Finally, the Level of Detail (LOD 1) is created and the CityGML file is saved.

Figure 64. FME workbench to generate CityGML models based on Spanish Cadastral input



Source: own elaboration using FME

6.2.4 Input-03.2: Cuatro de Marzo CityGML model [based on INSPIRE BU extended + LiDAR data]

Input-03.2 shares the same data source as Input-03.1: the Spanish Cadastre. The transformations proposed are the same as above; however, this model can be considered closer to reality since real heights are applied, instead of an average height based on the number of floors present in the cadastre.

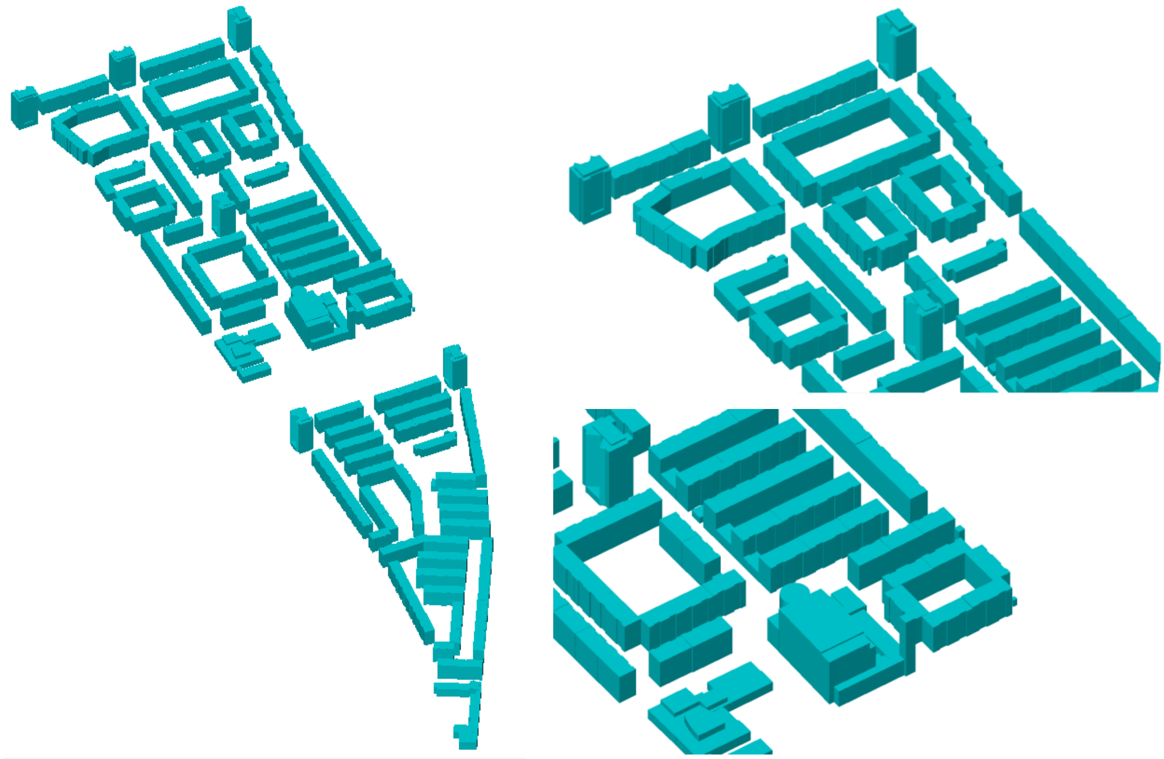
In order to apply real heights to the model, LiDAR (Light Detection and Ranging) data has been processed. These data consist of cloud points that are generated in specific flights and are classified into 20 categories or classification values (e.g. ground, low vegetation, medium vegetation, high vegetation, building or water) according to the ASPRS Standard LiDAR Point Classes[35]. Depending on the density of the cloud points, the accuracy of the model can be higher. More information on how this data has been applied in the specific case of Cuatro de Marzo can be found in section 6.2.4.2.

6.2.4.1 Specific description

This model covers only the Cuatro de Marzo district, due to the amount of pre-processing required to re-classify the cloud points to extract the building heights from LiDAR data. The following Figure 65 provides an overview of the model:

[35] http://www.asprs.org/a/society/committees/lidar/LAS_1-4_R6.pdf

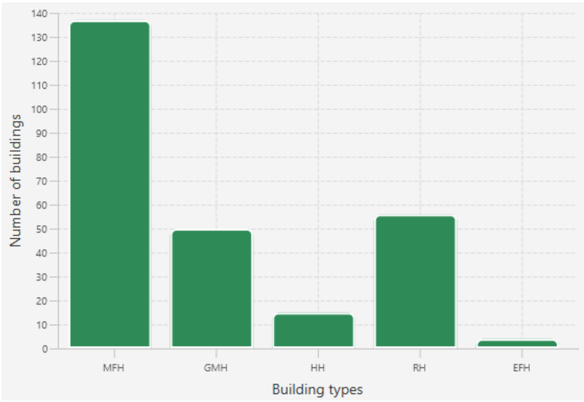
Figure 65. Input-03.2: District Scale – Cuatro de Marzo model based on INSPIRE BU extended (cadastral data) + real building heights based on LiDAR data



Source: own elaboration using FME and ArcGIS

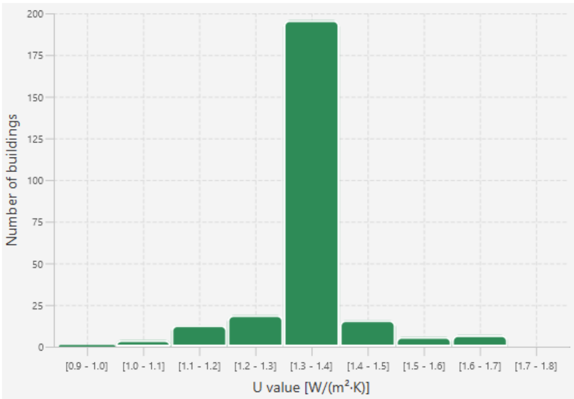
This district model corresponding to Cuatro de Marzo is formed by 263 buildings with the typologies observed in Figure 66, and the number of buildings per U-value range (Figure 67).

Figure 66. Buildings per type - Input03.2 District



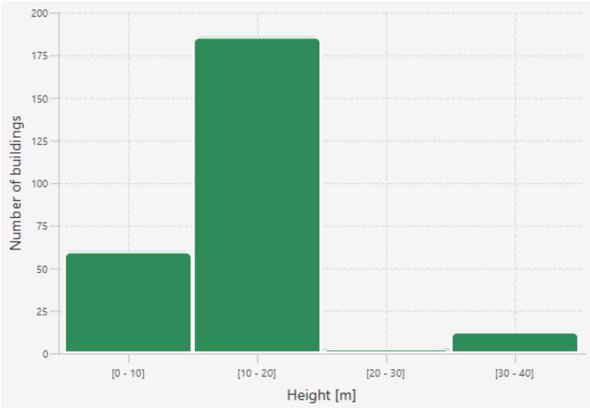
Source: own elaboration, based on SimStadt processing

Figure 67. Buildings per U-value range- Input03.2 District



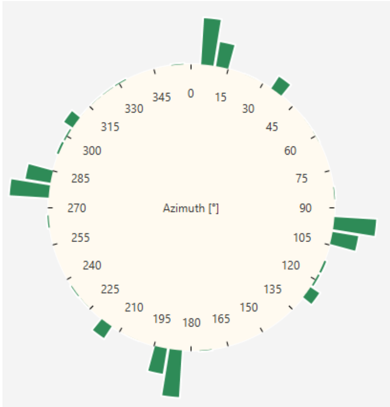
Source: own elaboration, based on SimStadt processing

Figure 68. Buildings per height range - Input03.2 District



Source: own elaboration, based on SimStadt processing

Figure 69. Wall orientation - Input03.2 District



Source: own elaboration, based on SimStadt processing

6.2.4.2 Pre-processing required

The main difference between this model and the previous one is the application of LiDAR data to specify real heights of buildings. This value is an important variable to define urban environments in 3D or calculate energy demand related to the residential sector. However, as mentioned previously, some public datasets such as those coming from the Spanish Cadastre (Building Parts) contain the number of floors of each building, and enable to apply a standard height / floor. However, the deviations of this approach from reality makes the assessment of LiDAR cloud points essential to calculate the height in the most accurate way.

LiDAR data are classified following the classification codes provided by the American Society for Photogrammetry and Remote Sensing (ASPRS). In particular, code 2 (ground) and code 6 (buildings) are the most relevant to calculate building heights. These codes are assigned to elements that can be found on the Earth surface, and enable their later analysis to define building boundaries. In this case LiDAR data are obtained from the National Geographical Institute of Spain and have a resolution of 0.5 to 1 point/m^2.

Figure 70. LiDAR cloudpoints with ASPRS code classification

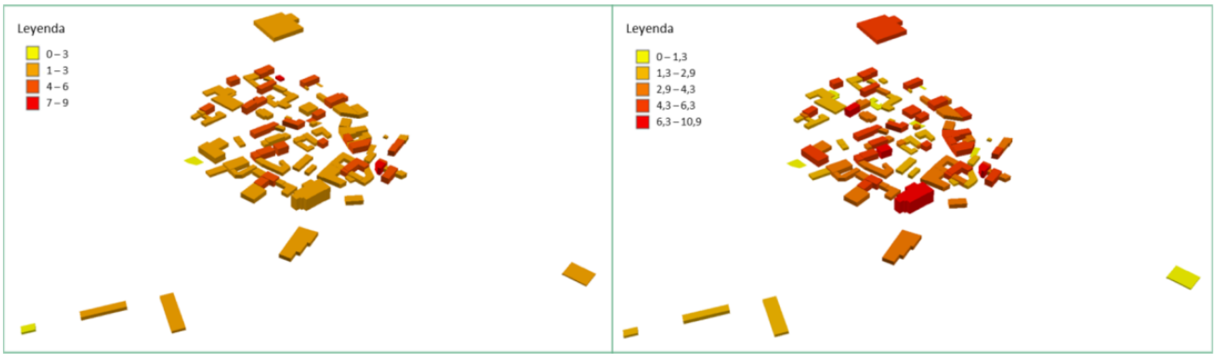
Source: own elaboration

However, this code assignation is performed in an automatic manner and it is prone to errors. As a result, some of these points are incorrectly classified. Their correction and re-classification needs to be performed manually by using specific software such as ArcGIS or similar, to reduce the classification error and avoid uncertainty in the obtained results.

Once data are correctly classified, all of the information needed can be extracted. In this case, a model was developed in ArcGIS to perform this processing, which enabled to obtain the building height to then integrate it with the 2D information from the buildings.

According to the tests performed in other models, the improvement achieved with this approach is considerable, since in many cases (50% on average), the height of the buildings is higher than in reality. In Figure 71 these differences can be appreciated in a test performed in Saldaña (Palencia, Spain). The left figure shows the heights estimated based on the number of floors in the cadastre and an average height / floor of 3 m, whereas the right figure provides the heights according to LiDAR data.

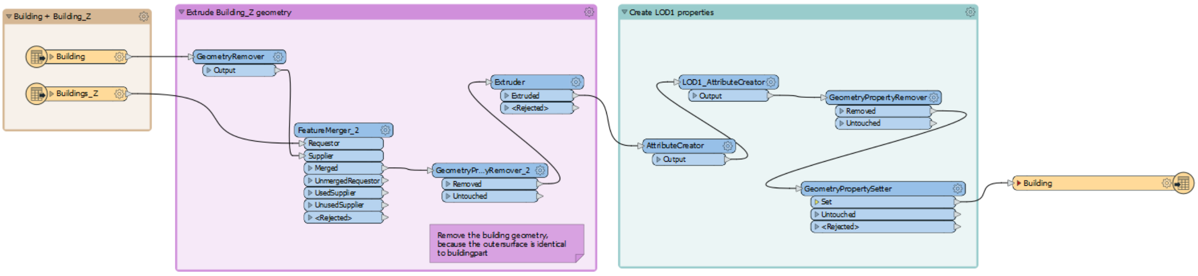
Figure 71. Building height comparison: left – Cadastre average, right – LiDAR data



Source: own elaboration

The same approach has been integrated in the FME workbench, as it can be seen below. Real mean heights have been introduced in an input shapefile (Building_Z) developed by the ArcGIS previously explained model that contains Building Part boundaries. This shapefile is merged with Building attributes to be extruded using the mean height value for each building. After that, Building_Z dataset is enriched by building use and year of construction. Finally, the Level of Detail (LOD 1) is created and the CityGML file is saved.

Figure 72. FME workbench to generate CityGML models based on Spanish Cadastral input and LiDAR data



Source: own elaboration using FME

6.2.5 Input-04: Cuatro de Marzo CityGML model [based on OSM data]

Input-04 is based on a crowdsourced dataset: OpenStreetMap (OSM). OSM is a collaborative project with the aim of developing a geospatial database of vector features for the whole world. This database can be used to develop urban energy models requiring georeferenced data of the cities infrastructures that are mainly represented by the location, function and occupancy of the different buildings. Everyone can contribute to OSM, i.e. add or edit any object available in the database. Consistency and accuracy of the data can vary from region to region. Data can be accessed, among others, through the following services:

- **Overpass API**[36]: This API serves up custom selected parts of the OSM map data. It acts as a database over the web: the client sends a query to the API and gets back the data set that corresponds to the query.
- **OSM planet**[37]: It includes complete copies of the full OpenStreetMap database which are regularly updated.
- **Geofabrik's free download server**[38]: This server has data extracts from the OpenStreetMap project which are normally updated every day. You can select your continent and then your country of interest to download your data.
- **Other sources**: Includes other additional sources that are included in the OpenStreetMap wiki.

In the OSM database, different types of georeferenced objects can be mapped and stored. Such objects are for example streets, buildings, land use and transportation networks (e.g. roads and railways). Data are available by three different data types (elements [39]) representing the most common objects. These types are nodes, ways and relations. Nodes are georeferenced points in space, which are defined by their geographical coordinates. Ways are an ordered collection of connected nodes, which either define a non-closed object such as path or closed objects (e.g. the footprint area of a building). Ways can represent either an empty polygon or an area (a filled polygon). Relations are the most complex data type in OSM and are used to represent objects in relations to each other, such as a bus route including all stops and road sections. In addition to geometry, OSM objects have one or more specific tags, i.e. semantic attributes. These define the meaning of e.g. elements in a street (buildings, constructive elements, urban elements, roads, etc.) and most of their characteristics (especially on buildings, covering the type, usage, height, etc.). There are groups of tag categories and some of them include also sub-categories with pre-set values.

The most common way of mapping objects in OSM is by means of GPS devices or by mapping from a satellite image or a combination of both methods. Due to the ease of use of the online application to insert data in OSM, certain tags are more utilized than others, and that would be reflected into the quantity of data available for certain building characteristics.

However, it should be highlighted that problems arise from both quantity and quality of the data. For the first case, there are still large sections of the municipalities with few or almost no information inserted. For the second case, some valuable characteristics can be missing. In fact, it is not uncommon to find out that there are lots of building references, but these references are limited to only one single tag (e.g., building=yes, as this is the most common one).

36 https://wiki.openstreetmap.org/wiki/Overpass_API
37 https://wiki.openstreetmap.org/wiki/Planet.osm
38 https://download.geofabrik.de/
39 https://wiki.openstreetmap.org/wiki/Elements

Related to the availability of data is the availability of layers. The OSM tools (online and offline) have some layer maps available, depending on the location, at country level. This means that for certain countries there would be good and updated layers, consisting of recent aerial photos with good quality, or a national cadastre layer (invaluable), or other options that could be better or worse depending on what is available in each country. For example, in Spain the cadastre layer enables users define the surface of a building with high precision; meanwhile a user had to use a blurry aerial photo in a country with few layers available.

With these constraints in mind, Input-04 has been created. A specific description can be found below (section 6.2.5.1), as well as an explanation of the pre-processing that was required to obtain this model (section 6.2.5.2).

6.2.5.1 Specific description

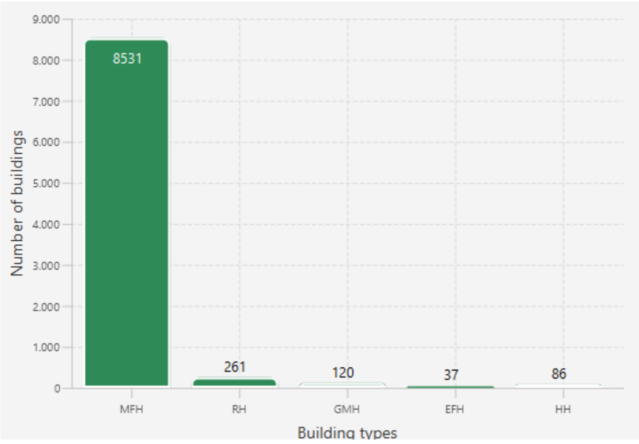
The model generated can be seen in Figure 73. From a first view, it can be observed that the coverage of buildings in the city is vast. However, there is a homogeneity in the building heights that is a consequence of the information found in the data source and the hypothesis applied, as it is shown in section 6.2.5.2.

Figure 73. Input-04: City Scale – Valladolid model based on OSM data

Source: own elaboration using FME

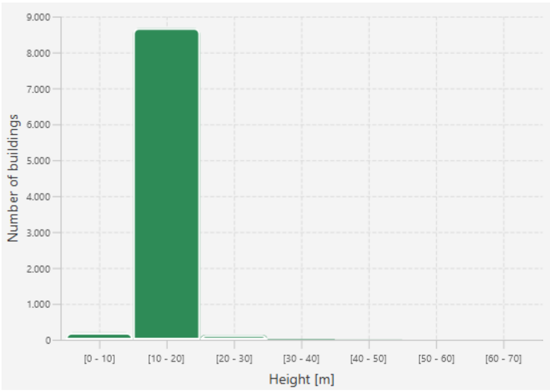
In order to understand the model better, the following figures are provided. As it can be observed, it contains **9.035 buildings**. Most of these buildings (8531) have been classified as multi-family homes, whereas the rest (261 +120 + 37 + 86) correspond to the RH, GMH, EFH, and HH typologies. Also worth to highlight is the homogeneous height present in the whole model, which was identified at a first glance.

Figure 74. Buildings per type - Input04 City



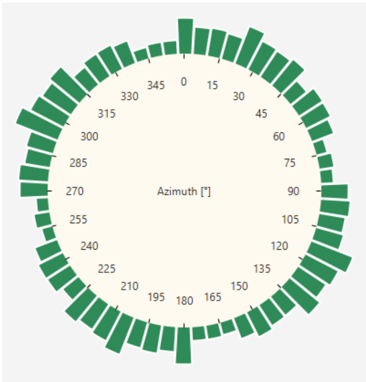
Source: own elaboration, based on SimStadt processing

Figure 75. Buildings per height range- Input04 City



Source: own elaboration, based on SimStadt processing

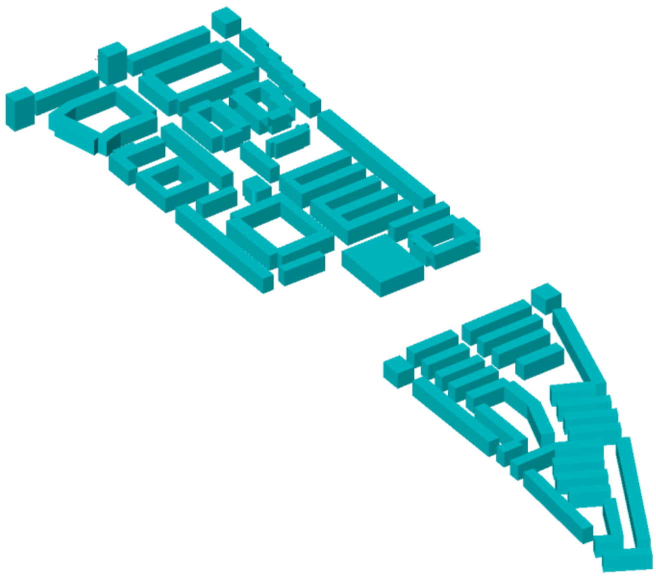
Figure 76. Wall orientation-Input04 City



Source: own elab.based on SimStadt processing

As performed in the case of Input-03.1, an extraction of the Cuatro de Marzo District is provided in Figure 76.

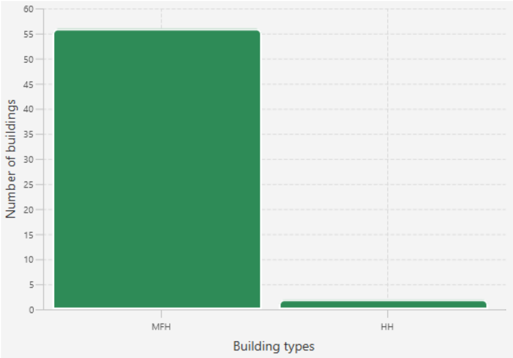
Figure 77. Input-04: District Scale – Cuatro de Marzo model based on OSM data



Source: own elaboration using FME

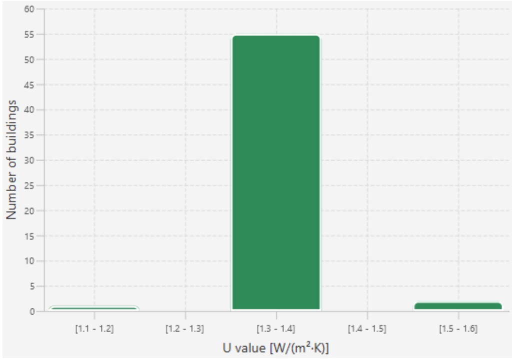
This district model corresponding to Cuatro de Marzo is formed by 58 buildings with the building types observed in Figure 78, and the number of buildings per U-value range (Figure 79).

Figure 78. Buildings per type -Input04 District



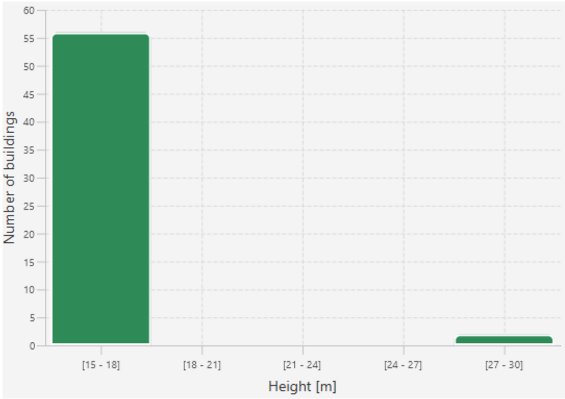
Source: own elaboration, based on SimStadt processing

Figure 79. Bdgs per U-value range-Inp.04 Dist



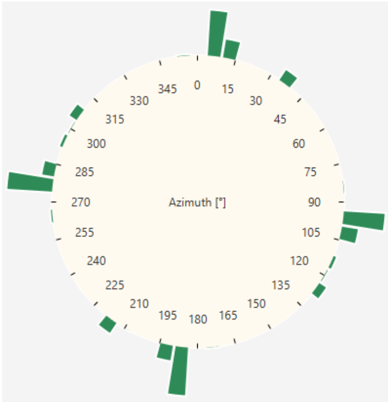
Source: own elaboration, based on SimStadt processing

Figure 80. Buildings per height range- Input04 District



Source: own elaboration, based on SimStadt processing

Figure 81. Wall orientation- Input04 District

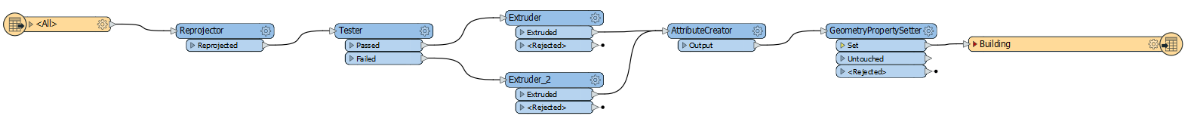


Source: own elaboration, based on SimStadt processing

6.2.5.2 *Pre-processing required*

Following the same approach as in the previous cases, OSM data were integrated in the FME workbench (see Figure 82). In order to determine the height of the building, the attributes present in OSM were first consulted, if there was no available height and only number of floors above ground, the floor height was assumed to be 2,7 meters. If no information was available for this respect, a height of 15 meters was assumed. This assumption was achieved by introducing two extruders in the model. In a first step, OSM data were reprojected and tested to define the two ways of height assumptions. After that an attribute is created to define the building use (the model assumes that all buildings are residential). Finally, the Level of Detail (LOD1) is created and the CityGML file is saved.

Figure 82. FME workbench to generate CityGML models based on OSM input



Source: own elaboration using FME

The process to generate models based on OSM is a promising way to automate the model generation based on a data source that is used worldwide. However, due to its collaborative nature, the data available is very heterogeneous in terms of completeness, and in the majority of occasions the building height or use was not available. This is the reason why the abovementioned hypotheses needed to be applied. In the case of Cuatro de Marzo, these hypothesis are very appropriate, since the district is very homogeneous and matches the characteristics of the buildings, except from the towers which are higher than the building blocks. As shown in

Figure 77, this difference has been captured in two of the towers, where number of floors were available, but not in the rest, where the common hypotheses were applied.

This inaccuracies based on the data generation due to lack of information, together with the fact that buildings are referenced using "ways" (according to OSM), and not to cadastral references, made the direct comparison more difficult. This is the reason why only a preliminary analysis of the results is presented in section 6.4.1.2, but no further comparisons with this dataset are presented in the following case studies.

6.3 Simulation and validation environments

In this section of the report two simulation environments and one validation environment are provided, which are described in the following sections.

6.3.1 Simulation environments

The two simulation environments share a common aim, which is tackling the urban scale by individually analysing each building's heating and cooling demand throughstandardised methods. However, some differences exist among them, which is reported in section 6.3.1.3 after having described both tools.

6.3.1.1 SimStadt

SimStadt [6] is the name of an urban simulation environment developed at HFT Stuttgart[40] and of a project of the same name.

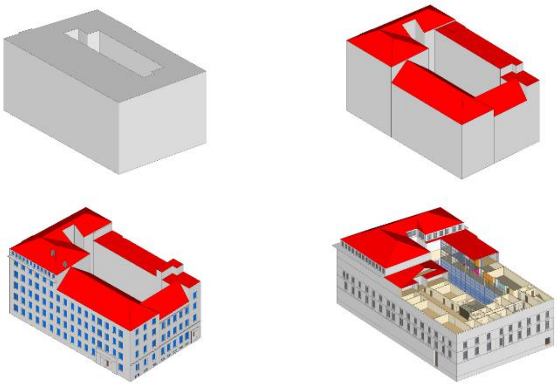
SimStadt in its current stage of development is able to use data of a real urban planning situation or planning state for energy analyses of buildings, city quarters, whole cities and even regions. The application scenarios range from high-resolution simulations of building heating requirements and potential studies for photovoltaics to the simulation of building refurbishment and renewable energy supply scenarios. Thus SimStadt is able to accompany e.g. architects, engineering offices, urban planners and municipalities substantially in integrated planning processes and for the definition of measures towards a sustainable (re)design of buildings and quarters.

This energy analysis method addresses any building (residential, mixed and non-residential) of any urban areas in the world, insofar a **virtual 3D city model** and minimum building parameter inputs are available. In this line, five key pillars should be highlighted [6]:

1. Virtual 3D City Model

The start of the workflow is the virtual 3D city model, modelled in the open standard format CityGML.

Figure 83. Levels of Detail (LOD) in CityGML



Source: HFT Stuttgart

One main advantage of the 3D city model format CityGML is its object modelling specification in different Levels of Details (LoD). The simplest geometric representation of a building for a heating demand evaluation consists of a simple rectangular block. This block model is equivalent to the Level of Detail 1 (LoD1) of CityGML. The Level of Detail 2 (LoD2) adds the roof form to the building level, Level of Detail 3 (LoD3) adds in the

[40] Hochschule für Technik Stuttgart, https://www.hft-stuttgart.de/

positioning of the façade windows, and Level of Detail 4 (LoD4) incorporates the modelling of the indoor space. The SimStadt Simulation Environment handles CityGML LoD 1 and LoD 2 models.

Such 3D City Models are generally created using LiDAR or stereo air photo, and enhanced with available semantic datasets such as building year, building usage, number of storeys, etc.

2. Quality Management

The requirements of SimStadt to a 3D Building Model are specified in a validation plan [18]. In short, the minimal requirements are a solid geometry of every Building and Building Part, and at least the year of construction and the function as mandatory attributes per building. Given the diverse quality levels of incoming virtual 3D city models, the healing module [CityDoctor] [19] offers a method of controlling and repairing the geometrical quality of the 3D City Model, for example, by closing polygons and volumes or separating buildings with common adjacent walls.

3. Energy Simulations

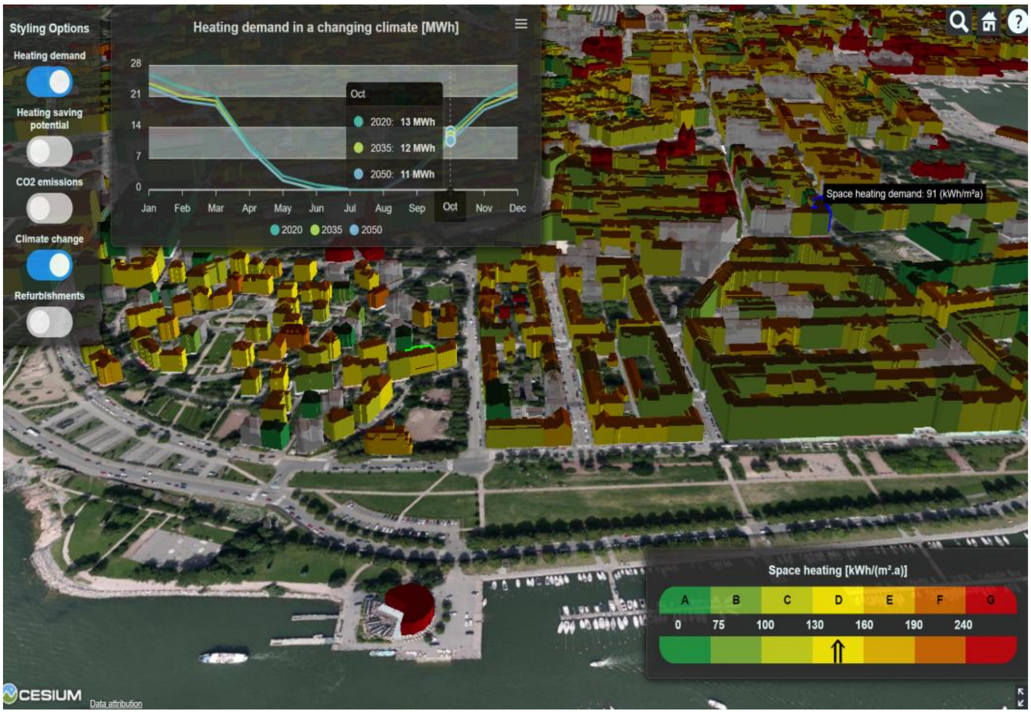
Based on this enhanced virtual 3D city model, the simulation tools INSEL [9], CitySim [20] and Stanet [21], coupled with the SimStadt Simulation Environments, allow for a variety of energy simulations:

- heating/cooling demand calculation (monthly energy balance or hourly dynamical simulation)
- photovoltaic potential calculation
- simulation of renewable energy systems
- simulation of heating/cooling networks

4. Visualisations

Simulation results and performance indices such as **heating demand, CO_2 emissions, primary energy and energy saving potentials** may be visualized in the virtual 3D city model and analysed in a decision-making module (see Figure 83).

Figure 84. Visualization possibilities in SimStadt



Source: [22]

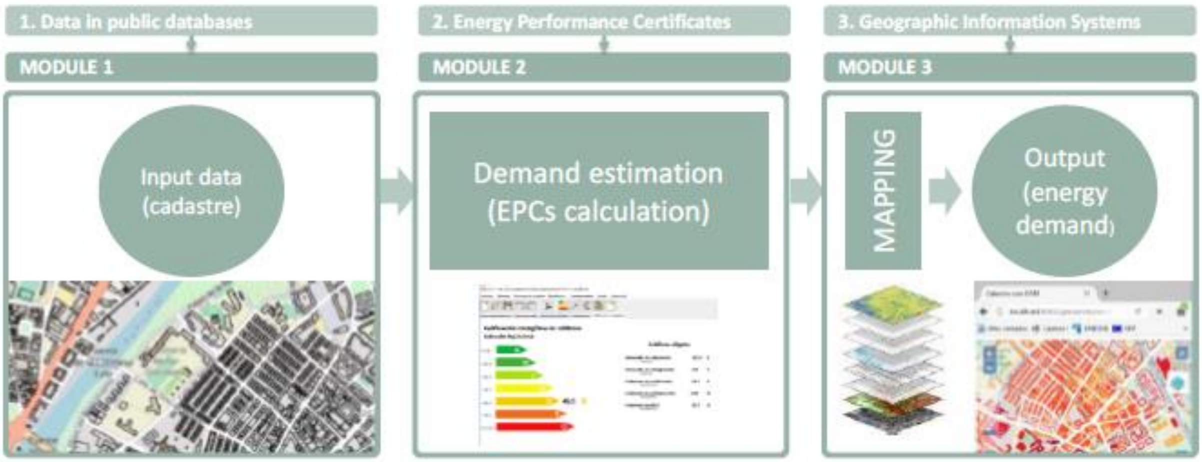
6.3.1.2 ENERGIS

The ENERGIS tool [8] has been developed at CARTIF within a collaboration project with the same name. Its main objective is to provide an easy to use energy decision support tool to map energy demand at urban and regional level calculated through validated methods. To this end, public data is collected, analysed and processed; the energy demand building by building is automatically calculated using a validated Energy Performance Certificate tool; and all of the information is mapped in friendly web maps, making use of the functionalities provided by Geographic Information Systems (GIS). Therefore, the three main pillars of the platform, which are closely related to the modules into which the platform is divided, are:

1. To exploit publicly available repositories
2. To implement a demand calculation method based on validated methodologies
3. To exploit mapping and visualisation capabilities

These main pillars are translated into the three main modules the tool entails, shown in Figure 85 and explained below.

Figure 85. ENERGIS tool main pillars and modules



Source: CARTIF, own elaboration

Module 1: Information processing and treatment

The pillar of the platform is the use of open public data from official sources to be deployed in the estimation of the energy demand. Besides, these public data must be retrieved from different sources automatically. Therefore, after being gathered, these data should be processed and transformed.

There are three main types of data required by the platform: (1) geometry data on buildings, (2) climate zones and (3) building thermal properties, explained below.

For the geometry data on buildings the key data source is the Spanish cadastre [23]. The cadastre provides for each building geometrical information and general semantic information that is used in order to identify and to characterize the building. This information is mainly the geo-located footprint of the building, the number of building floors above ground, and below ground, the year of construction, the current use of the building and the address of the building. The geometry information automatically collected is processed on the one hand to generate the information for the different envelope elements of the building, with their dimensions and the orientation, and on the other hand to produce shadow patterns with the information of the façades of neighbouring buildings.

For climate-related data, the National Code for Building Construction [24] in Spain was queried, since it establishes reference climate zones.

In the case of the building thermal properties, the National Building Code was consulted to identify the characteristic to be used. Based on several studies a catalogue of building elements and materials was generated. The ENERGIS platform is able to use this catalogue in order to consider different building characteristics, where according to the type of element, the year of construction and the climatic zone, some thermal characteristics and other parameters are assigned in the same way that the EPC tools use this catalogue of building elements.

Module 2: Estimation of the energy demand

The estimation of the energy demand is the core of the platform. This estimation is based on the automatic operation in one of the EPC calculation tools recognized by the Spanish government for the energy certification of existing buildings, in particular the CE3X tool [25].

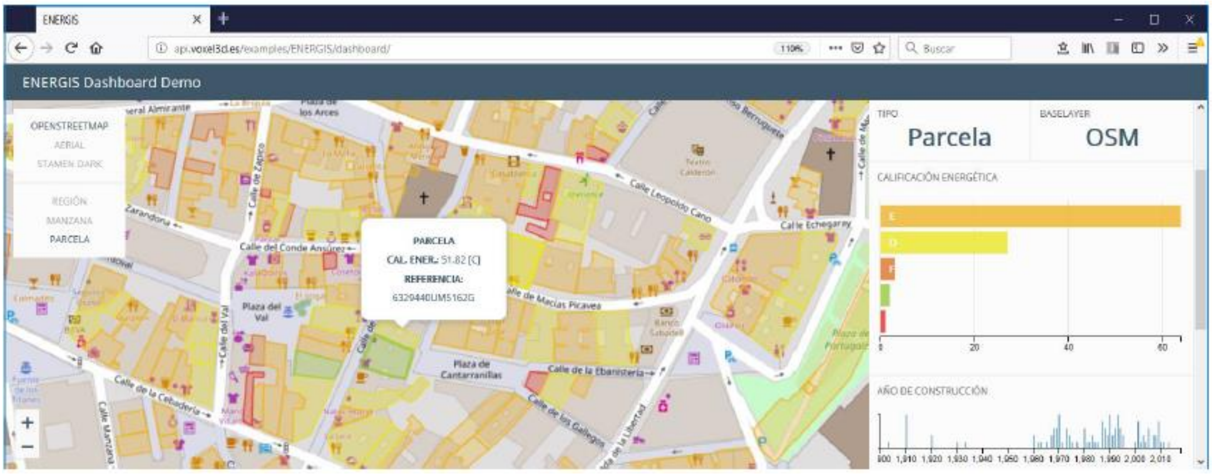
The operation of the estimation engine consists in two steps: (1) the creation of files that are used in the CE3X tool, collecting the information retrieved in the previous stage and parsing them in the right format and (2) the automatic execution of the CE3X tool using the file created in the immediately previous step. The results obtained are the energy demands of the building: cooling, heating and global energy demand.

Module 3: Mapping and visualisation

The output information of the ENERGIS tool is stored in a geodatabase, which is structured in three tables: one table for buildings, one for blocks (groups of buildings) and one for cadastral zones (neighbourhoods), which are the three scales provided by the Spanish Cadastre's online services. These three scales are common definitions set in the INSPIRE Directive [26], which is followed and implemented through the data offered and the services provided by the Spanish Cadastre. For the building demand the information from the previous Module 2 was used, while for the demand in the blocks and the cadastral zones is the result of aggregation operations over the values for the buildings.

The results are shown to the public through the ENERGIS online platform. The data is presented with the geo-referenced values and with a recognisable colour code that corresponds to the Energy Label scale used in Energy Performance Certificates, as shown in Figure 86. The additional information and filtering capabilities available in the online web platform are expected to help the planner in identifying districts or zones with high energy demand.

Figure 86. ENERGIS platform screenshot



Source. ENERGIS platform

In addition, an online web portal with further information of the platform (information about EPCs in the ENERGIS platform, instructions on how to use the tool, etc.) has been implemented in the following link: http://api.voxel3d.es/examples/ENERGIS/portal/.

6.3.1.3 Similarities and differences between SimStadt and ENERGIS

In order to understand the results shown by both tools, it is necessary to observe the differences and similarities between both tools, which are summarised in a brief way in the following table:

Table 20. Similarities and differences between SimStadt and ENERGIS

Topic	SimStadt	ENERGIS
Scope	Simulation for urban environments with the building as minimum unit (which can be analysed also in terms of its parts).	
Approach	To be used by end-users whocan introduce their specific models in CityGML.	To be visualized by the user, the input data is extracted automatically from public sources and pre-calculated before being shown within the platform.
Input geometry	3D: In CityGML format, allowing for different levels of detail.	2D: In GML format, based on INSPIRE BU extended. Automatically obtained from the Spanish Cadastre
Calculation engine	INSEL	CE3X
Buildings Physics Library	Existing default physics building library (German Building Library); however, users can generate their own library.	Existing default building library for Spain. New libraries should, for the moment, be configured by the administrator to cover more countries.
Buildings Usage Library	Existing default characteristics in the tool, but the user can generate their own library.	Existing default building usage for Spain. New libraries should, for the moment, be configured by the administrator to cover other usages.

Source: own elaboration

6.3.2 Validation environment: Real Energy Performance Certificates

Energy Performance Certificates (EPCs) have been in place for several years. Based on the objectives of putting energy efficiency first, achieving global leadership in renewable energies and providing a fair deal for consumers, the European Commission proposed a package of Energy Directives "Clean Energy for All Europeans" [27]. It includes as well eight different legislative proposals that tackle, among others, Energy Efficiency, Energy Performance in Buildings, Renewable Energy and Governance.

In particular, when considering EPCs, and based on the requirements imposed by the recast of 2018 [28] of the Energy Performance of Buildings Directive (EPBD, 2010/31/EU) [29], EU Member States are required submit an EPC for every dwelling, building block, or commercial premise to be leased or sold, as well as for every new construction and public buildings.

These directives set certain objectives to Member States and should be transposed by each country in order to comply with them by establishing plans and strategies. Depending on the administrative structure in each country, the plans can either be established at the national level, or some high-level guidelines at national level can be set and then specific objectives at regional level implemented. After each Member State has carried out his strategies, the results are to be reported at EU level.

In order to assure coherence among the results obtained in each Member State, a methodological framework is described in the annex of the aforementioned Directive. This annex does not exactly set the formulas to be deployed, but instead presents the type of calculations to perform or which aspects to consider (for instance, thermal bridges). Therefore, each Member State has the obligation to transpose this framework in their country and develop either a concrete methodology or develop specific tools to serve this purpose, leading sometimes to inhomogeneous approaches within the EU.

However, grading systems are required in Energy Performance Certificates, which allow for comparisons among countries, as well as contribute to the understanding of the general public. These labels normally use a colour code (from green – most efficient, to red – least efficient) and are usually accompanied by a letter (normally from A to G). The scales that allow to translate a specific energy performance to an energy label are established based on the status of the building stock, the climate zone, the building type, etc. Thus, even though the approaches are varied, **a common understanding of what a building would require to be more energy efficient is provided by means of these energy labels**.

As it can be observed, EPCs are an official document to display energy performance of buildings across the EU and are mandatory in certain cases. As a result, an increasing amount of valuable data is generated to this respect, thus making these results an ideal validation environment for the purposes of this report. Nevertheless, two other key aspects should be mentioned:

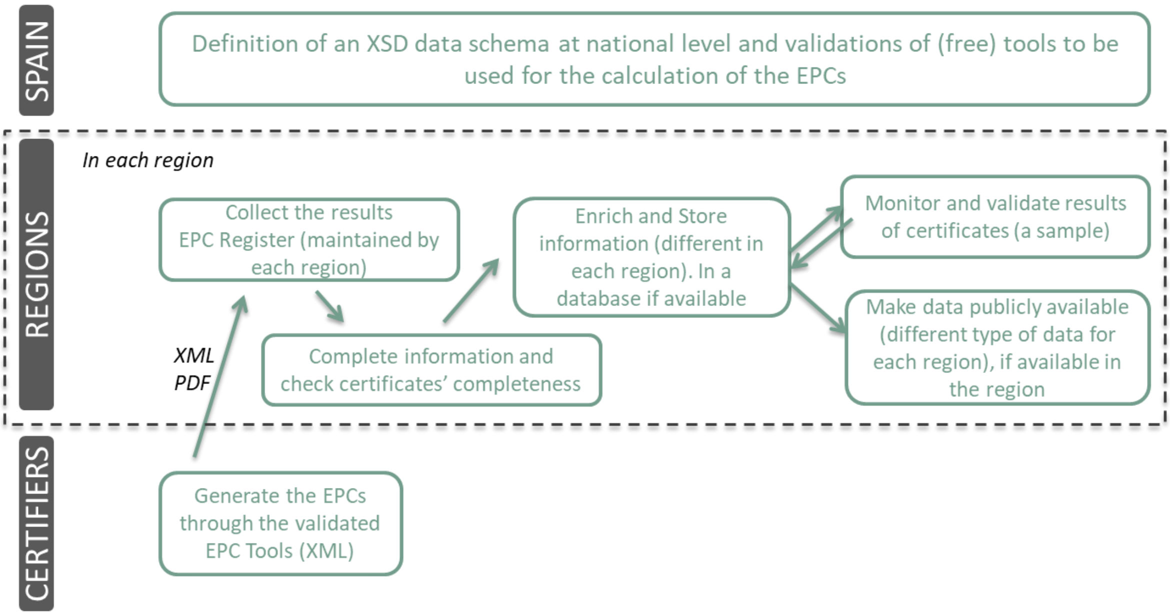
- **Veracity of results and checking mechanisms**: even when being an official document, the veracity of the results should be handled with care. In particular, some of the issues that could influence EPC results are the following:
    - Energy Performance Certificates are submitted by experts. In some countries these experts are certified through a specific exam, whereas in others certain degrees are enough to certify expertise in these matters.
    - Energy Performance Certification tools sometimes allow for the use of default values whenever a specific parameter is not known or has not been specifically measured.
    - Compliance and checking mechanisms are implemented in a varied manner across the EU and not all EPCs are consistently checked.
- **Granularity**: Energy Performance Certificates can be issued for individual elements (commerces, dwellings, offices, etc), or for entire buildings. The level of granularity poses a specific complexity when trying to analyse vast amounts of Energy Performance Certificates.

6.3.2.1 Real Energy Performance Certificates in Spain

In Spain, Energy Performance Certificates are issued by experts such as architects or engineers, which have to calculate the Energy Performance Certificate through a nationally validated tool, or through an equally valid method (which should be appropriately justified). However, normally experts use one of the validated tools at national level, since, in the case of Spain, they are free of charge and are offered by the *Ministerio para la Transición Ecológica y el Reto Demográfico*. In particular, the following tools can be used: *Herramienta unificada LIDER-CALENER (HULC)*, *CE3, CE3X* or *CERMA*. Additionally, from 5th of July of 2018 onwards, *also CYPETHERM HE Plus, SG SAVE* and the *CE3X* complement for new buildings can be used [30].

In terms of competences, regional authorities are in charge of the management of Energy Performance Certificates and should implement appropriate methods for experts to submit their EPCs, and a database to store them. Additionally, they should apply the appropriate validation and checking mechanisms. This can be observed in Figure 87.

Figure 87. Management process for the EPCs followed in Spain.



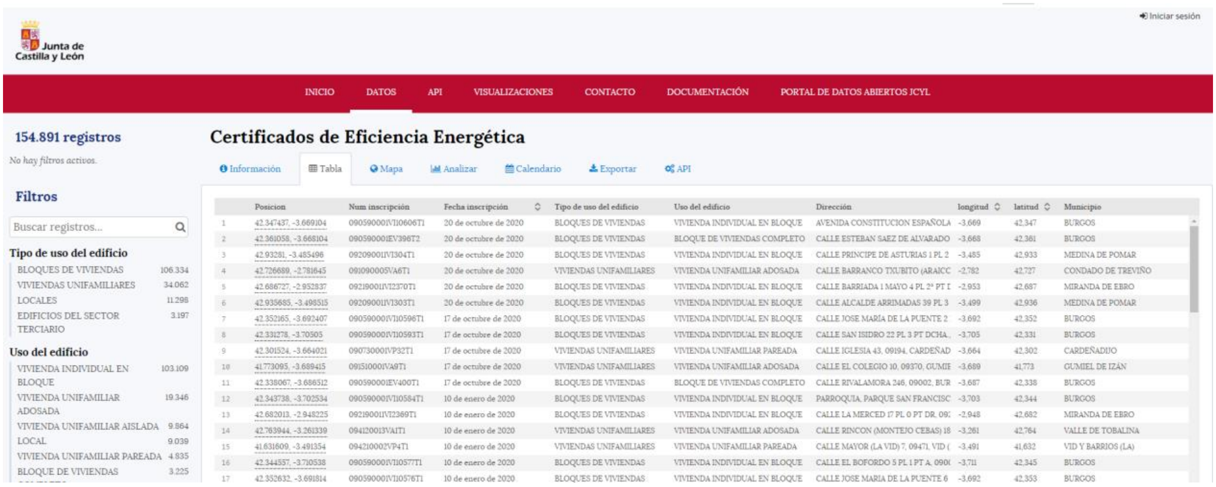
Source: own elaboration, CARTIF, 2019.

The data stored by the regions can be publicly available or not. In the case of the region of Castilla y León, some of the datasets contained in the Energy Performance Certificates are provided through their open data platform. Nevertheless, not all the content of Energy Performance Certificate is offered.

6.3.2.2 Real Energy Performance Certificates in Castilla y León

The regional energy authority in charge of the management of Energy Performance Certificates in Castilla y León region is the Ente Regional de la Energía de Castilla y León [31]. The EPC data provided to the public can be accessed through their open data portal, shown in Figure 88:

Figure 88. Energy Performance Certificates' open data portal in Castilla y León



Source: Castilla y León open data portal

Data can be accessed online, through an API or exported to a file. The data publicly available for each Energy Performance Certificate is the following. However, it should be highlighted that all the data contained in the EPCs is contained in the EREN database, even when not available to the public:

- Inscription number
- Date of submission

- Type of use of building (broad category)
- Building use (specific category)
- Address
- Longitude
- Latitude
- Municipality
- Province
- Specific primary consumption
- Primary energy consumption label
- CO_2 emissions ratio
- CO_2 emissions label
- Specific heating energy demand
- Heating energy demand label
- Specific cooling energy demand
- Cooling energy demand label

6.4 Results

In this section, the results obtained in each use case are presented. As explained in Figure 34 and in Table 14, the Case Studies are divided into three groups, which tackle three different aspects, in particular:

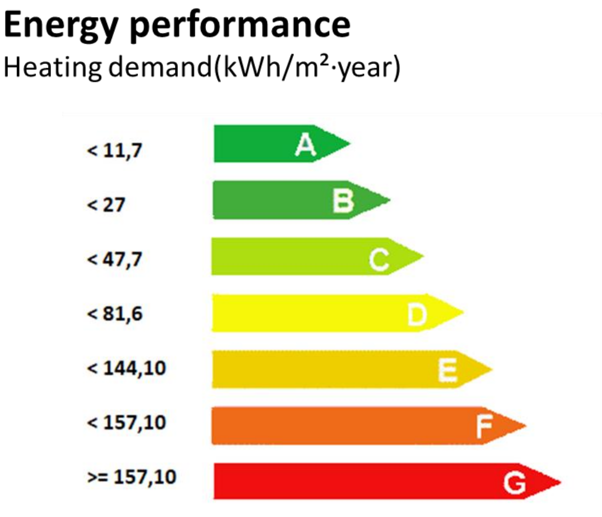
- **How does the generation of datasets affect the final results?** This is observed in Case Study 01, at district scale (CS1.1), where results obtained in the same simulation environment (SimStadt developed by HFT) when using different data inputs are compared with each other.
- **How do the results vary in two different simulation environments that share the same objective?** In this case, a new simulation environment is introduced, in particular the ENERGIS tool developed by CARTIF. One more time, at district (CS2.1) and city level (CS2.2) the results derived have been compared.
- **Are the results comparable to real EPCs?** In order to validate the obtained results with comparable real data available to the public and generated following similar principles than those followed in the previous simulation environments, the results are compared to real Energy Performance Certificates in the Castilla y Léon region. This is performed both at district (CS3.1) and city level (CS3.2).

It is worth to mention, that for all the comparisons presented in the following subsections, the results obtained from the simulations (either coming from SimStadt or from ENERGIS) needed to be post-processed. Post-processing was performed in order to find a common comparable element that was identifiable in an easy manner: the building identified with its corresponding cadastral reference.

In each of the case studies, the same structure is followed. First the data preparation is presented, then data is analysed and the results of each simulation is presented. Finally, comparisons are made between the datasets and conclusions are derived from these comparisons.

To calculate the labels in all the datasets the limits for the values of the heating demand label in Valladolid for residential buildings have been used. This limit is shown in Figure 88.

Figure 89. Limits of heating demand label values



Source: own elaboration based on demand limits from Valladolid according to IDAE calculations

It is important to know these limits in order to interpret the differences that exist between the different datasets. The limits for all the values included within the energy performance certificate, which end up translated into a label, aim to convey the difficulties a building will have towards improving their energy performance. In this line, these limits are established depending on whether the building is residential or not, or whether an individual element of a building (e.g. dwelling) is being certified or if the whole building is. Thus, these values vary depending on the building typology, climate zone and if the building is existing, or if it is new. In this case, as mentioned before, only one set of limits has been used in the comparisons, the one corresponding to existing residential buildings, since it is the most common typology to be compared.

6.4.1 Case Study 1: Different dataset generation

Table 21 defines the main parameters of Case Study 1. It also provides the references to the sections of the document where the individual results of each of the simulations can be seen and the sections where the comparisons among results are presented.

Table 21. Case study 1: main parameters

Case Study 1. Different dataset generation									
Case study code	Scale	SimStadt	ENERGIS	Input 01	Input 02	Input 03.1	Input 03.2	Input 04	Comparison
Case Study 1.1	District	X	-	X	-	X	X	(X)	Sec. 6.4.1.1
Case Study 1.2	City	X	-	-	-	X	-	(X)	Sec. 6.4.1.2

Source: own elaboration

6.4.1.1 CS1.1: Different dataset generation at district scale

6.4.1.1.1 CS1.1 Data preparation

In this case study, three datasets are considered, which have been transformed from BU or BU Parts to BU as follows:

Table 22. Case Study 1.1: data preparation required

Name	Initial	Initial elements	Target	Final elements
01D_small_4marzo.csv	BU	28 buildings	BU	28 buildings
031D_small_4marzo.csv	BUParts	48 BU Parts	BU	30 buildings
032D_small_4marzo.csv	BUParts	37 BU Parts	BU	30 buildings

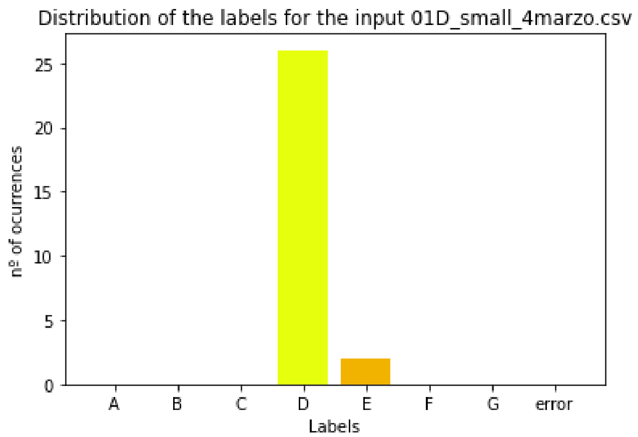
Source: own elaboration

This initial step was necessary in order to count with the same number of elements in the comparison and have a homogeneous reference to compare to.

6.4.1.1.2 CS1.1 Data analysis

This section presents the results from the simulations performed in SimStadt of the three abovementioned inputs, in Table 23, Table 24 and Table 25.

Table 23. Results for Input 01 – Cuatro de Marzo district (CS1.1)

<table>
<tr><th colspan="4">Case Study 1.1: Results for 01D_small_4marzo (28 buildings)</th></tr>
<tr><td rowspan="9">Figure 90. Distribution of the labels for Input 01 – District level


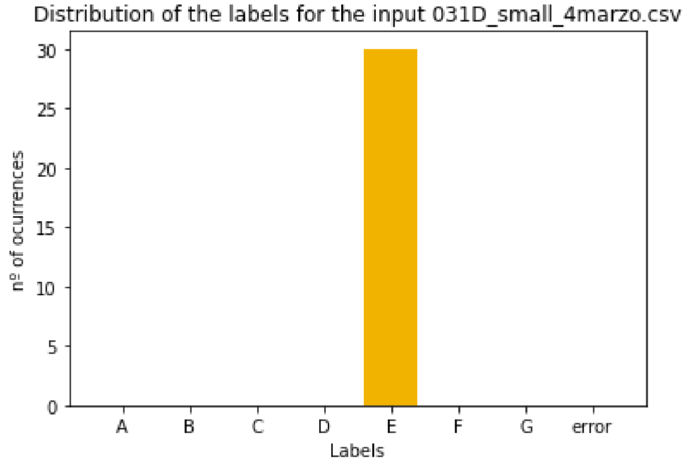
Source: own elaboration based on SimStadt results</td><th>Label</th><th>Buildings</th><th>Percentage</th></tr>
<tr><td>Label A:</td><td>0</td><td>0.00%</td></tr>
<tr><td>Label B:</td><td>0</td><td>0.00%</td></tr>
<tr><td>Label C:</td><td>0</td><td>0.00%</td></tr>
<tr><td>Label D:</td><td>26</td><td>92.86%</td></tr>
<tr><td>Label E:</td><td>2</td><td>7.14%</td></tr>
<tr><td>Label F:</td><td>0</td><td>0.00%</td></tr>
<tr><td>Label G:</td><td>0</td><td>0.00%</td></tr>
<tr><td>Label error:</td><td>0</td><td>0.00%</td></tr>
</table>

Source: own elaboration based on SimStadt results

Table 24. Results for Input 03.1 – Cuatro de Marzo district (CS1.1)

Case Study 1.1: Results for 031D_small_4marzo (30 buildings)

Figure 91. Distribution of the labels for Input 03.1 – District level



Source: own elaboration based on SimStadt results

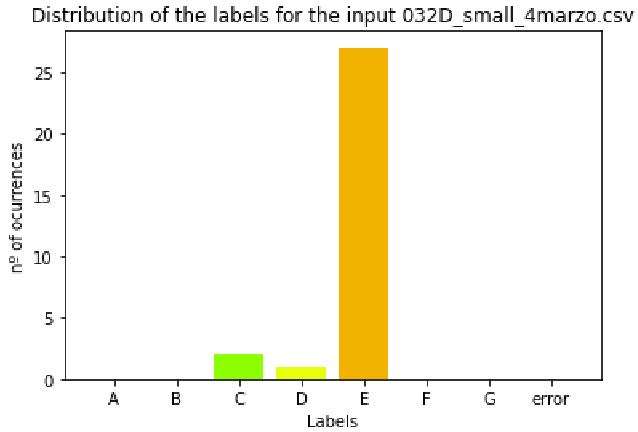
Label	Buildings	Percentage
Label A:	0	0.00%
Label B:	0	0.00%
Label C:	0	0.00%
Label D:	0	0.00%
Label E:	30	100.00%
Label F:	0	0.00%
Label G:	0	0.00%
Label A:	0	0.00%

Source: own elaboration based on SimStadt results

Table 25. Results for Input 03.2 – Cuatro de Marzo district (CS1.1)

Case Study 1.1: Results for 032D_small_4marzo (30 buildings)

Figure 92. Distribution of the labels for Input 03.2 – District level



Source: own elaboration based on SimStadt results

Label	Buildings	Percentage
Label A:	0	0.00%
Label B:	0	0.00%
Label C:	2	6.67%
Label D:	1	3.33%
Label E:	27	90.00%
Label F:	0	0.00%
Label G:	0	0.00%
Label error:	0	0.00%

Source: own elaboration based on SimStadt results

6.4.1.1.3 CS1.1 Data comparison

Data comparison is performed per set of two datasets, considering in every case the buildings they have in common (see Table 27, Table 28, Table 29). In order to provide an overview of the comparisons presented in this section, Table 26 is presented. Then, in section 6.4.1.1.4 conclusions have been extracted based on these comparisons.

Table 26. Label comparison (CS1.1)

Case Study 1.1: Label comparison			
Label	Input 01D_small	Input 031D	Input 032D
Label A:	0.00%	0.00%	0.00%
Label B:	0.00%	0.00%	0.00%
Label C:	0.00%	0.00%	6.67%
Label D:	92.86%	0.00%	3.33%
Label E:	7.14%	100.00%	90.00%
Label F:	0.00%	0.00%	0.00%
Label G:	0.00%	0.00%	0.00%
Label error:	0.00%	0.00%	0.00%

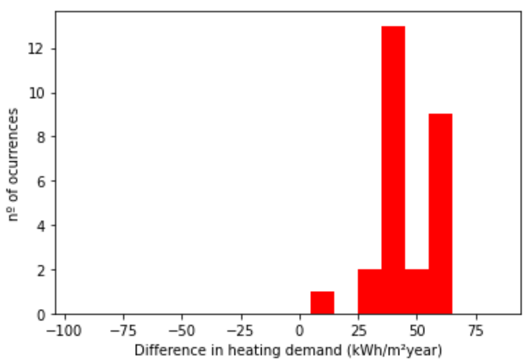
Source: own elaboration based on SimStadt results

The results presented in the above table vary with inputs coming from more detailed modelling: ad-hoc modelling of Input 01-D, and 03.2-D considering real building heights coming from LiDAR.

Table 27. CS1.1: comparison of Input 03.1D (small) and Input 01-D (small)

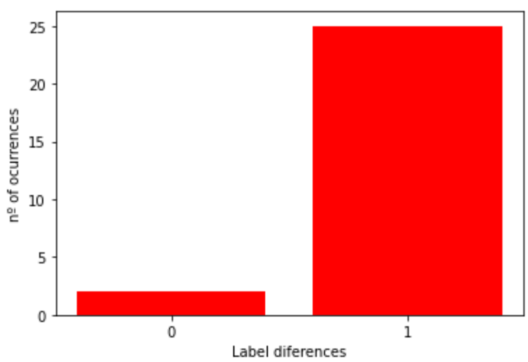
Case Study 1.1: Comparison of Input-031-D (small) and Input 01-D (small)						
Buildings in common	27	Label differences (in % and # buildings)				
Building with reference 5110704UM5151A, 5110007UM5151A, and 5110004UM5151A are not present in in dataset 01D_small_4marzo.csv		-2	-1	0	1	2
		0.00%	0.00%	7.41%	92.59%	0.00%
		0	0	2	25	0

Figure 93. Heating demand differences for datasets: 031-D (small) and 01-D (small)



Source: own elaboration

Figure 94. Label differences for datasets: 031-D (small) and 01-D (small)



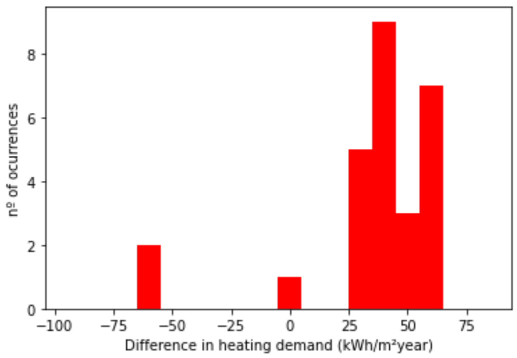
Source: own elaboration

Source: own elaboration

Table 28. CS1.1: comparison of Input 03.2D (small) and Input 01-D (small)

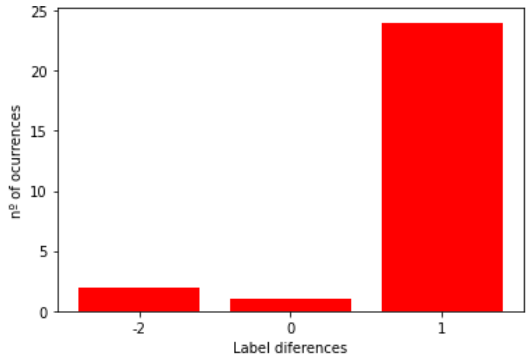
Case Study 1.1: Comparison of Input 03.2D (small) and Input 01-D (small)					
Buildings in common	27	Label differences (in % and # buildings)			
Building with reference 5110704UM5151A ,5110007UM5151A, Building with reference 5110004UM5151A not present in dataset 01D_small_4marzo.csv	-2	-1	0	1	2
	7.41%	0.00%	3.70%	0.00%	88.89%
	2	0	1	0	24

Figure 95. Heating demand differences for datasets: 032-D (small) and 01-D (small)



Source: own elaboration

Figure 96. Label differences for datasets: datasets: 032-D (small) and 01-D (small)



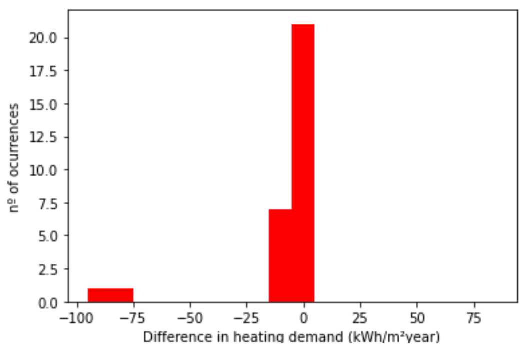
Source: own elaboration

Source: own elaboration

Table 29. CS1.1: comparison of Input 03.2D (small) and Input 03.1D (small)

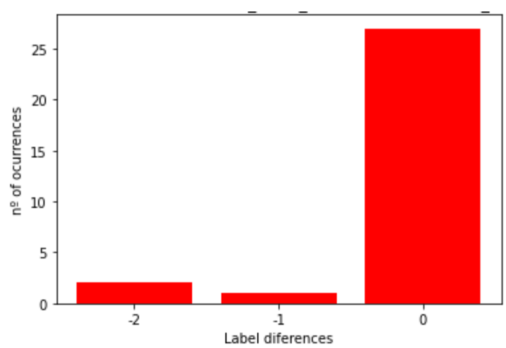
Case Study 1.1: Comparison of Input 03.2D (small) and Input 03.1D (small)					
Buildings in common	30	Label differences (in % and # buildings)			
All buildings are compared	-2	-1	0	1	2
	6.67%	3.33%	90.00%	0.00%	0.00%
	2	1	27	0	0

Figure 97. Heating demand differences for datasets: Input 03.2D (sm) & 03.1D (sm)



Source: own elaboration

Figure 98. Label differences for datasets: Input 03.2D (small) and Input 03.1D (small)



Source: own elaboration

Source: own elaboration

6.4.1.1.4 CS1.1 Conclusions

Three different inputs have been compared in this Case Study 1.1, corresponding to ad-hoc generation of models (Input 01-D), generation of models based on cadastral data (Input 03.1-D), and based on cadastral data with real heights extracted from LiDAR analysis (Input 03.2-D). All of the models correspond to the District level of Cuatro de Marzo "smaller" version, which shares the same scope as Input 01 (28 buildings). Several conclusions can be extracted:

— **Homogeneity of results**: in the case of Inputs 01 and 03.2 the label results were not homogeneous, and different labels were obtained; whereas in the case of Input 03.1-D all of the heating energy labels were the same.

— **Label similarities in Inputs from cadastre**: In the cases of Input 03.1-D and Input 03.2-D, the majority of buildings obtained a Label E (100% and 90%, respectively), whereas Input 01 obtained more efficient values, with 92.86% of the buildings achieving a Label E.

— **Extreme differences in values can be due to outliers**: when observing the results of the pairwise comparisons, the normal situation is that results are shifted as a block either to the left or to the right, especially when analysing the energy demand. However, there are some cases where outliers appear (see e.g. Figure 95 or Figure 97). As a result, major differences appear in the labels, reaching even a two label difference. It would be advisable to detect to which buildings these results correspond, in order to reach a robust conclusion.

— **Higher heating energy demand**: is obtained with inputs from cadastre (Input 03.1-D), then higher energy performance is obtained as a result when applying real heights obtained with LiDAR (Input 03.2-D). Finally input 01-D presents the lowest values for the heating energy demand. This can be due to the clean definition of volumes of this Input, versus the complexity found in Inputs 03.1-D and Input 03.2-D, which leads to increasing the external wall surface and, consequently, the energy demand.

In this case study CS1.1, it would be advisable to:

— Detect buildings offering extreme values. This would allow to determine the reason of why certain results act as outliers in a district where buildings are very homogeneous.

— Calculate external building wall surface, Heated area and/or Volume in order to find whether these parameters are linked to the results obtained.

6.4.1.2 CS1.2: Different dataset generation at city scale

Case study 1.2 is not performed since the generation of Input 04 entailed difficulties for its comparison. However, the results obtained from simulating Input 04 with SimStadt are presented in section 6.4.2.1 (at district scale, see Table 35) and section 6.4.2.2 (at city scale, see Table 44).

6.4.2 Case study 2: Different simulation environments

The following Table 30 defines the main parameters of Case Study 2. It also provides the references to the sections of the document where the individual results of each of the simulations can be seen and the sections where the comparisons among results are presented.

Table 30. Case study 2: main parameters

Case Study 2. Different simulation environments									
Case study code	Scale	SimStadt	ENERGIS	Input 01	Input 02	Input 03.1	Input 03.2	Input 04	Comparisons
Case Study 2.1	District	X	X	X	X	X	X	(X)	Sec. 6.4.2.1
Case Study 2.2	City	X	X	-	X	X	-	(X)	Sec. 6.4.2.2

Source: own elaboration

6.4.2.1 CS2.1: Different simulation environments at district scale

6.4.2.1.1 CS2.1 Data preparation

In this case study, four datasets are considered, which have been transformed from BU or BU Parts to BU as follows:

Table 31. Case Study 2.1: data preparation required

Name	Initial	Initial elements	Target	Final elements
02D_4marzo.geojson	BU	189 buildings	BU	189 buildings
031D_4marzo.csv	BU Parts	286 BU parts	BU	206 buildings
032D_4marzo.csv	BU	205 buildings	BU	205 buildings
04D_4marzo.csv	BU	25 buildings	BU	25 buildings

Source: own elaboration

This initial step was necessary in order to count with the same number of elements in the comparison and have an homogeneous reference to compare to.

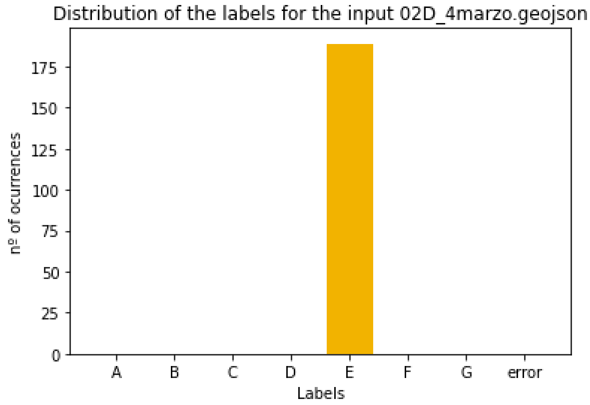
6.4.2.1.2 CS2.1 Data analysis

This section presents the results from the simulations performed in SimStadt of the four abovementioned inputs, in Table 32, Table 33, Table 34 and Table 35.

Table 32. Results for Input 02 – Cuatro de Marzo district (CS2.1)

Case Study 2.1: Results for 02D_4marzo (189 buildings)

Figure 99. Distribution of the labels for Input 02 – District level



Source: own elaboration based on ENERGIS results

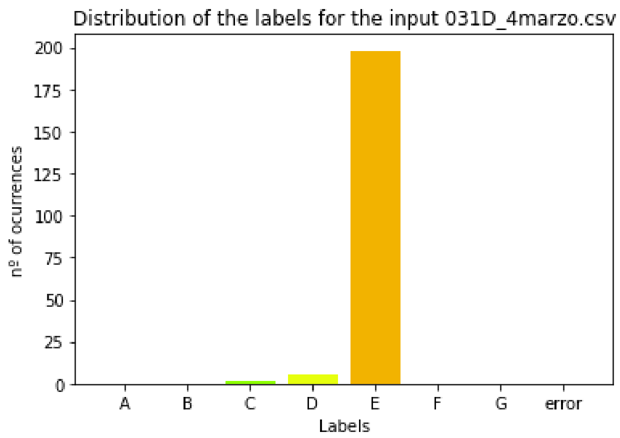
Label	Buildings	Percentage
Label A:	0	0.00%
Label B:	0	0.00%
Label C:	0	0.00%
Label D:	0	0.00%
Label E:	189	100.00%
Label F:	0	0.00%
Label G:	0	0.00%
Label error:	0	0.00%

Source: own elaboration based on ENERGIS results

Table 33. Results for Input 03.1 – Cuatro de Marzo district (CS2.1)

Case Study 2.1: Results for 31D _4marzo (206 buildings)

Figure 100. Distribution of the labels for Input 03.1 – District level



Source: own elaboration based on SimStadt results

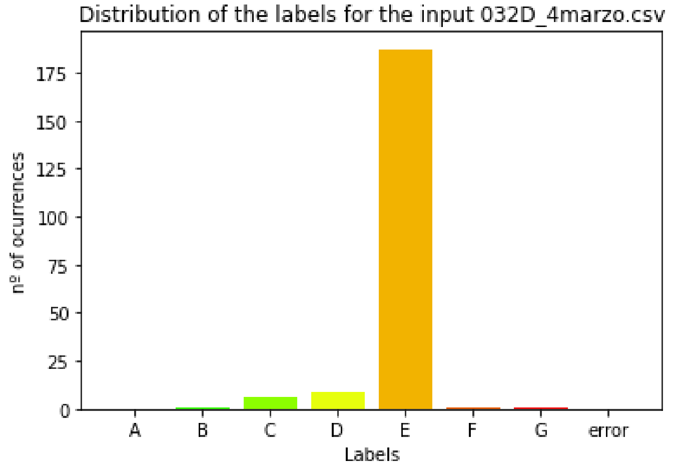
Label	Buildings	Percentage
Label A:	0	0.00%
Label B:	0	0.00%
Label C:	2	0.97%
Label D:	6	2.91%
Label E:	198	96.12%
Label F:	0	0.00%
Label G:	0	0.00%
Label error:	0	0.00%

Source: own elaboration based on SimStadt results

Table 34. Results for Input 03.2 – Cuatro de Marzo district (CS2.1)

Case Study 2.1: Results for 32D _4marzo (205 buildings)

Figure 101. Distribution of the labels for Input 03.2 – District level



Source: own elaboration based on SimStadt results

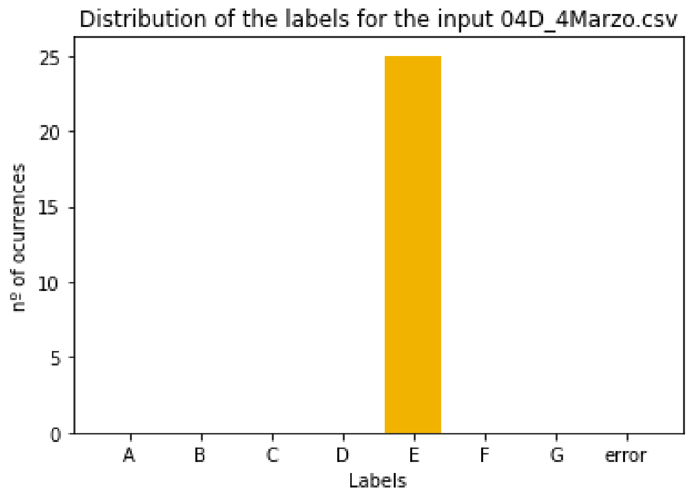
Label	Buildings	Percentage
Label A:	0	0.00%
Label B:	1	0.49%
Label C:	6	2.93%
Label D:	9	4.39%
Label E:	187	91.22%
Label F:	1	0.49%
Label G:	1	0.49%
Label error:	0	0.00%

Source: own elaboration based on SimStadt results

Table 35. Results for Input 04– Cuatro de Marzo district (CS2.1)

Case Study 2.1: Results for 04D_4marzo (25 buildings)

Figure 102. Distribution of the labels for Input 04 – District level



Source: own elaboration based on SimStadt results

Label	Buildings	Percentage
Label A:	0	0.00%
Label B:	0	0.00%
Label C:	0	0.00%
Label D:	0	0.00%
Label E:	25	100.00%
Label F:	0	0.00%
Label G:	0	0.00%
Label error:	0	0.00%

Source: own elaboration based on SimStadt results

6.4.2.1.3 CS2.1 Data comparison

Data comparison is performed per set of two datasets, considering in every case the buildings they have in common (see Table 37, Table 38, and Table 39). It must be highlighted, that comparisons are not performed with Input 04-D, as explained in section 6.2.5. In order to provide an overview of the comparisons presented in this section, the following Table 36 is presented.

Table 36. Label comparison (CS2.1)

Case Study 2.1: Label comparison				
Label	Input 02D	Input 031D	Input 032D	Input 04D
Label A:	0.00%	0.00%	0.00%	0.00%
Label B:	0.00%	0.00%	0.49%	0.00%
Label C:	0.00%	0.97%	2.93%	0.00%
Label D:	0.00%	2.91%	4.39%	0.00%
Label E:	100.00%	96.12%	91.22%	100.00%
Label F:	0.00%	0.00%	0.49%	0.00%
Label G:	0.00%	0.00%	0.49%	0.00%
Label error:	0.00%	0.00%	0.00%	0.00%

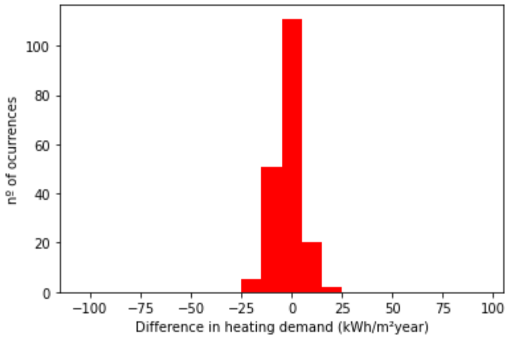
Source: own elaboration based on SimStadt and ENERGIS results

Table 37. CS2.1: comparison of Input 03.1D and Input 02D

Case Study 2.1: Comparison of Input 03.1D and Input 02D						
Buildings in common	189	Label differences (in % and # buildings)				
Buildings with reference 5108601UM5150G, 4903303UM5140D, 5110701UM5151A, 5009404UM5150G, 5109416UM5150G, 5108304UM5150G, 5108407UM5150G 5106609UM5150E, 5111112UM5151A, 5109414UM5150G, 5109202UM5150G, 5008501UM5150G, 5106608UM5150E, 5009407UM5150G, 5110007UM5151A, 5109412UM5150G, 5107608UM5150E are not present in dataset 02D_4marzo.geojson		-2	-1	0	1	2
		0.00%	0.00%	100%	0.00%	0.00%
		0	0	189	0	0

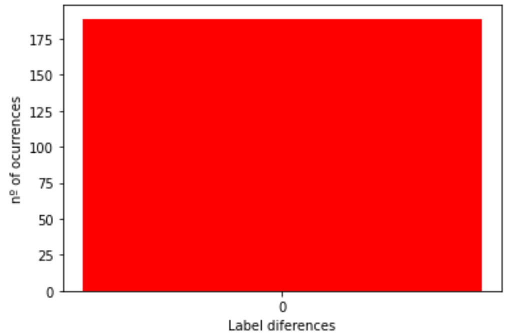
Source: own elaboration

Figure 103. Heating demand differences for datasets: Input 03.1D and Input 02D



Source: own elaboration

Figure 104. Label differences for datasets: Input 03.1D and Input 02D

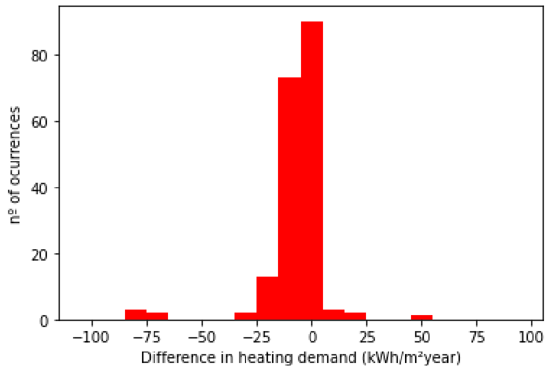


Source: own elaboration

Table 38. CS2.1: comparison of Input 03.2D and Input 02D

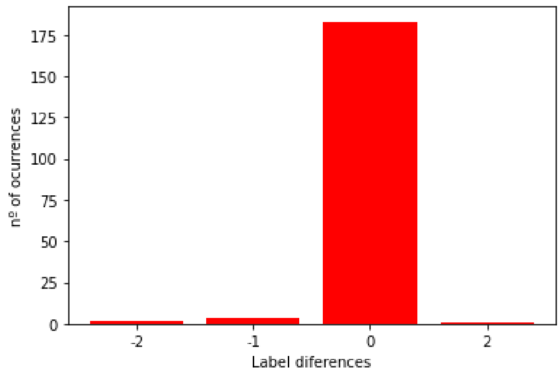
Case Study 2.1: Comparison of Input 03.2D and Input 02D						
Buildings in common	189	Label differences (in % and # buildings)				
Buildings with reference 5108601UM5150G, 4903303UM5140D, 5110701UM5151A, 5009404UM5150G, 5109416UM5150G, 5108304UM5150G, 5108407UM5150G 5106609UM5150E, 5111112UM5151A, 5109414UM5150G, 5109202UM5150G, 5008501UM5150G, 5106608UM5150E, 5009407UM5150G, 5110007UM5151A, 5109412UM5150G, 5107608UM5150E not present in dataset 02D_4marzo		-2	-1	0	1	2
		1.06%	1.59%	96.83%	0.00%	0.53%
		2	3	183	0	1

Figure 105. Heating demand differences for datasets: Input 03.2D and Input 02D



Source: own elaboration

Figure 106. Label differences for datasets: Input 03.2D and Input 02D



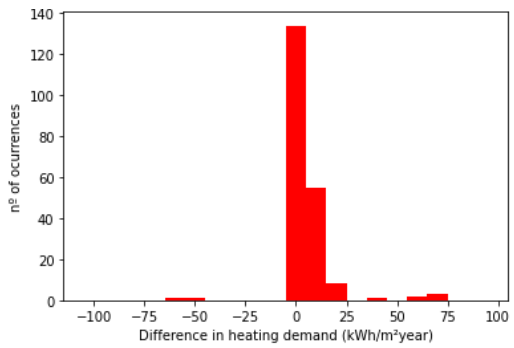
Source: own elaboration

Source: own elaboration

Table 39. CS2.1: comparison of Input 03.1D and Input 03.2D

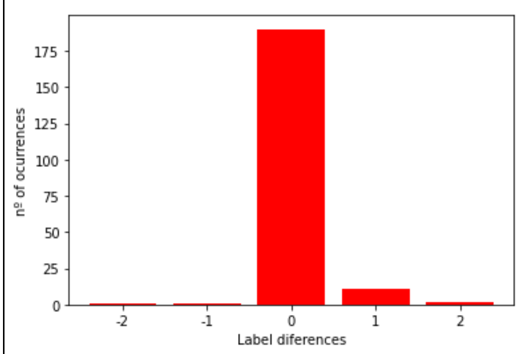
Case Study 2.1: Comparison of Input 03.1D and Input 03.2D						
Buildings in common	205	Label differences (in % and # buildings)				
Building with reference 5111112UM5151A not present in dataset 032D_4marzo.csv		-2	-1	0	1	2
		0.49%	0.49%	92.68%	5.37%	0.98%
		1	1	190	11	2

Figure 107. Heating demand differences for datasets: Input 03.1D and Input 03.2D



Source: own elaboration

Figure 108. Label differences for datasets: Input 03.1D and Input 03.2D



Source: own elaboration

Source: own elaboration

6.4.2.1.4 CS2.1 Conclusions

Three different inputs have been compared in this Case Study 2.1, namely Input-02D (GML coming directly from the cadastre), Input-03.1D (CityGML LOD 1 based on cadastral input), and Input-03.2D (CityGML LOD 1 based on cadastral input and real heights from LiDAR data). Additionally, and even when not compared to the abovementioned inputs, Input-04D (based on OSM data) has been calculated in SimStadt, and general conclusions can be extracted based on results shown in Table 36.

As it has been previously discussed, the objective of this case study is to compare the results obtained with two different simulation environments: SimStadt (results from Input-03.1D and Input-03.2D) and ENERGIS (results from Input-02D). In this case, the comparisons have been made at district level, covering all the buildings in Cuatro de Marzo district in Valladolid, Spain. From these comparisons, several conclusions can be extracted:

— **Homogeneity of results:** it can be observed that results from Input-02 (ENERGIS) and Input-04 (OSM) coincide and are homogeneous (all buildings have been assigned an E Label). This can be due to the hypotheses applied when generating the Input-04 model, which perfectly match the characteristics of the district (5 floors high, same building typology) and which are very similar to the hypotheses applied when generating Input-02D results (extracting number of floors from cadastral data, 5, and multiplying by average floor height, 2.7). In order to check the adequacy of the hypothesis applied in Input 04, the performance of the model should be observed at city scale, where different building typologies and building heights exist.

In contrast, Inputs-03.1D and Input-03.2D show different results, with most of the buildings having an E label (96.12% and 91.22% respectively).

— **Label similarities in Inputs**: when analysing the label comparison of the Input-03.1D (SimStadt) with the Input-02 (ENERGIS), it can be seen that **for the 189 buildings in common** they have the same label for all these buildings. Regarding the label comparison of the Input-03.2D (SimStadt –LiDAR analysis) with the Input-02 (ENERGIS), the main labels are coincident (96.83%), but for some buildings differences in the label appear and in some cases with two steps of difference. In the case of the comparison of the Input-03.1D (SimStadt) with Input-03.2D (SimStadt–LiDAR analysis) the coincident labels are the 92.68%, and it is important to highlight that 5.35% of buildings present a less efficient label for the Input-03.1D (SimStadt) with respect to Input-03.2D (SimStadt –LiDAR analysis).

— **Heating energy demand:** the heating energy demand for the Input-03.1D (SimStadt) and the Input-02 (ENERGIS) are very similar (the differences between them are limited, from -25 kWh/year*m^2 to 25 kWh/year*m^2). Regarding the comparison of Input-03.1D (SimStadt) with respect to Input-03.2D (SimStadt–LiDAR analysis), Figure 107 shows that the heating energy demand is slightly higher in the Input-03.1D for some buildings.

— **Extreme differences in values can be due to outliers**: the outliers that cause the differences in the labels can be detected in Figure 107, showing some occurrences for those values lower than -50 kWh/year*m^2 and those higher than 50 kWh/year*m^2. So, these outliers are from Input-03.2D (SimStadt–LiDAR analysis). More information extracted for these buildings is shown in the Table 40. This table shows that the heated area and heated volume of these building is lower for the model created with the LiDAR analysis. It should be analysed if this is due to a problem of the model or it is instead an improvement given by it.

Table 40. CS2.1: buildings that present divergent values

CS2.1: buildings that present divergent values			
Building ref	Input	Area	Volume
5108601UM5150G	031D_4marzo.csv	2201.00	6877.70
	032D_4marzo.csv	428.20	1338.20
5107601UM5150E	031D_4marzo.csv	2458.10	7681.20
	032D_4marzo.csv	1388.20	4338.10
5111111UM5151A	031D_4marzo.csv	2730.70	8533.20
	032D_4marzo.csv	1578.70	4933.30
5111106UM5151A	031D_4marzo.csv	2685.40	8391.60
	032D_4marzo.csv	1715.40	5360.80
5111106UM5151A	031D_4marzo.csv	2676.70	8364.50
	032D_4marzo.csv	1651.60	5161.30
5107607UM5150E	031D_4marzo.csv	2446.40	7645.00
	032D_4marzo.csv	1373.60	4292.50

In this case study CS2.1, it would be advisable to:

— Analyse what model corresponds better to the reality and check if, as it was expected, the model created with the LiDAR analysis improves the results.

— Evaluate if the improvement obtained using a more complex model is worth the higher additional computational cost related.

6.4.2.2 *CS2.2: Different simulation environments at city scale*

6.4.2.2.1 CS2.2 Data preparation

In this case study, three datasets are considered, which have been transformed from BU or BU Parts to BU as follows:

Table 41. Case study 2.2: data preparation required

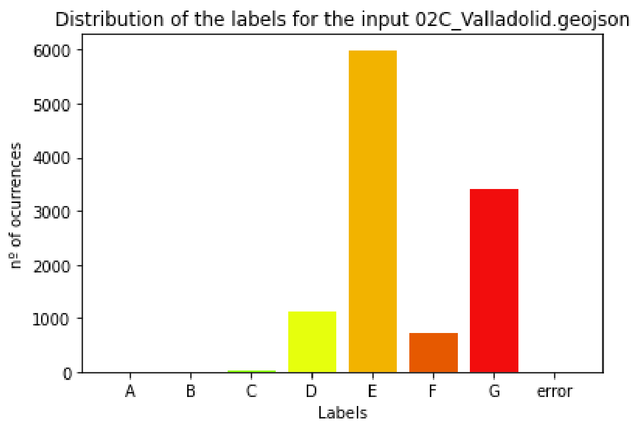
Name	Initial	Initial elements	Target	Final elements
02C_Valladolid.geojson	BU	11,262 buildings	BU	11,262 buildings
031C_Valladolid.csv	BUParts	88,894 BU parts	BU	14,738 buildings
04D_4marzo.csv	BU	8,999 buildings	BU	8,999 buildings

Source: own elaboration

This initial step was necessary in order to count with the same number of elements in the comparison and have a homogeneous reference to compare to.

6.4.2.2.2 CS2.2 Data analysis

Table 42. Results for Input 02 – Valladolid City (CS2.2)

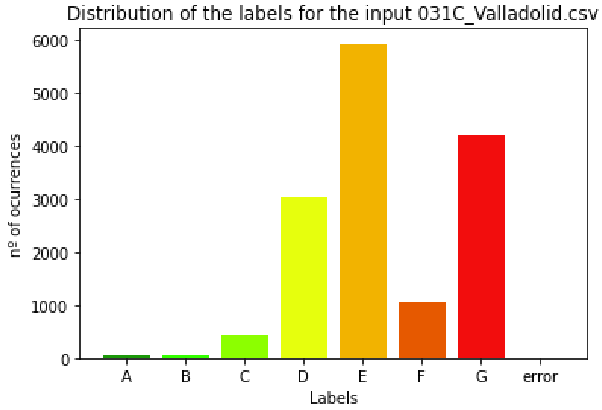
<table>
<tr><th colspan="4">Case Study 2.2: Results for 02C_Valladolid (11.262 buildings)</th></tr>
<tr><td rowspan="9">Figure 109. Distribution of the labels for Input 02 – City level

Source: own elaboration based on ENERGIS results</td><td>Label</td><td>Buildings</td><td>Percentage</td></tr>
<tr><td>Label A:</td><td>0</td><td>0.00%</td></tr>
<tr><td>Label B:</td><td>0</td><td>0.00%</td></tr>
<tr><td>Label C:</td><td>24</td><td>0.21%</td></tr>
<tr><td>Label D:</td><td>1111</td><td>9.87%</td></tr>
<tr><td>Label E:</td><td>5993</td><td>53.21%</td></tr>
<tr><td>Label F:</td><td>726</td><td>6.45%</td></tr>
<tr><td>Label G:</td><td>3408</td><td>30.26%</td></tr>
<tr><td>Label error:</td><td>0</td><td>0.00%</td></tr>
</table>

Source: own elaboration based on ENERGIS results

Table 43. Results for Input 03.1 Valladolid City (CS2.2)

Case Study 2.2: Results for 031C_Valladolid (14738 buildings)

Figure 110. Distribution of the labels for Input 03.1 – City level



Source: own elaboration based on SimStadt results

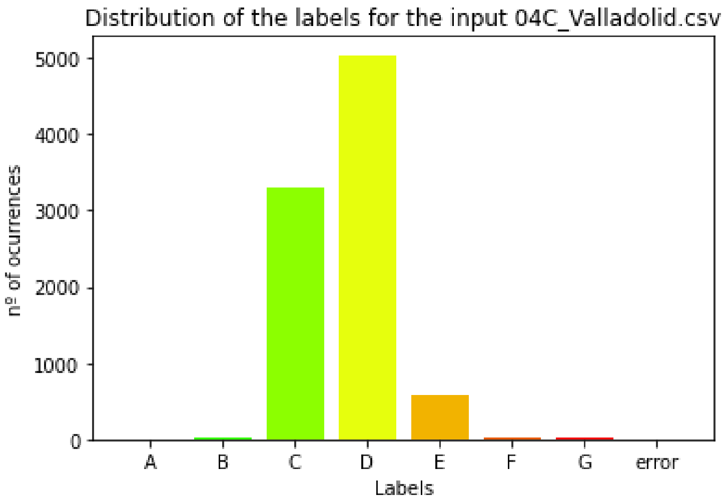
Label	Buildings	Percentage
Label A:	51	0.35%
Label B:	44	0.30%
Label C:	424	2.88%
Label D:	3031	20.57%
Label E:	5928	40.22%
Label F:	1065	7.23%
Label G:	4192	28.44%
Label error:	3	0.02%

Source: own elaboration based on SimStadt results

Table 44. Results for Input 04 Valladolid City (CS2.2)

Case Study 2.2: Results 04C_Valladolid (8999 buildings)

Figure 111. Distribution of the labels for Input 04 – City level



Source: own elaboration based on SimStadt results

Label	Buildings	Percentage
Label A:	0	0.00%
Label B:	28	0.31%
Label C:	3295	36.62%
Label D:	5038	55.98%
Label E:	586	6.51%
Label F:	21	0.23%
Label G:	31	0.34%
Label error:	0	0.00%

Source: own elaboration based on SimStadt results

6.4.2.2.3 CS2.2 Data comparison

Data comparison is performed per set of two datasets, considering in every case the buildings they have in common (see Table 46 and Table 47). As before, Input-04-C is not compared to the other results, as per the explanations provided in section 6.2.5. In order to provide an overview of the comparisons presented in this section, Table 45 is presented.

Table 45. Label comparison (CS2.2)

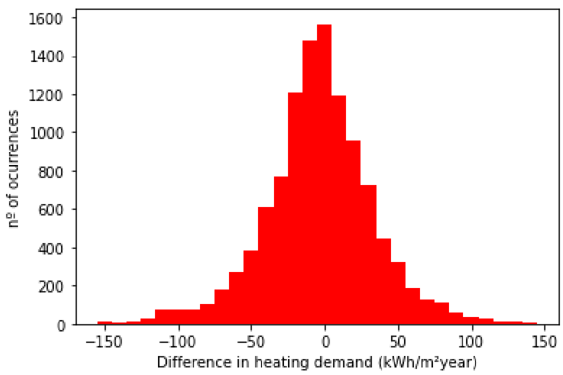
Case Study 2.2: Label comparison			
Label	Input 02C	Input 031C	Input 04C
Label A:	0.00%	0.35%	0.00%
Label B:	0.00%	0.30%	0.31%
Label C:	0.21%	2.88%	36.62%
Label D:	9.87%	20.57%	55.98%
Label E:	53.21%	40.22%	6.51%
Label F:	6.45%	7.23%	0.23%
Label G:	30.26%	28.44%	0.34%
Label error:	0.00%	0.02%	0.00%

Source: own elaboration based on SimStadt and ENERGIS results

Table 46. CS2.2: comparison of 03.1C and Input 02C

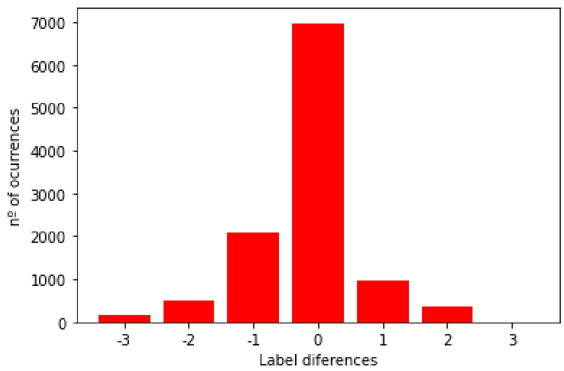
Case Study 2.2: Comparison of 03.1C and Input 02C								
Buildings in common	11,100	Label differences (in % and # buildings)						
All buildings were present in both datasets		-3	-2	-1	0	1	2	3
		1.57%	4.46%	18.84%	62.9%	8.79%	3.38%	0.06%
		174	495	2091	6981	976	375	7

Figure 112. Heating demand differences for datasets: 03.1C and Input 02C



Source: own elaboration

Figure 113. Label differences for datasets: 03.1C and Input 02C



Source: own elaboration

Source: own elaboration

6.4.2.2.4 CS2.2 Conclusions

Two different inputs have been compared in this Case Study 2.2, namely Input-02C (GML coming directly from the cadastre) and Input-03.1C (CityGML LOD 1 based on cadastral input). Additionally, and even when not compared to the abovementioned inputs, Input-04C (based on OSM data) has been calculated in SimStadt, and general conclusions can be extracted based on results shown in 0.

As it has been previously discussed, the objective of this case study is to compare the results obtained with two different simulation environments: SimStadt (results from Input-03.1C) and ENERGIS (results from Input 02C). In this case, the comparisons have been made at city level, covering all the buildings presented in the datasets. It is important to highlight that for Input_02C only results for residential building are present. From these comparisons, several conclusions can be extracted:

— **Homogeneity of results for cadastre datasets:** before commenting the results, one issue to take into account is the number of buildings for each approach, that is very different for each input, so the results have to be analysed carefully. It can be observed that results from Input-02 (ENERGIS) and Input-03.1C (SimStadt) present similar distributions per labels with the main occurrences for the E label for both datasets. For the Input 04C, D label is the label that appears in most buildings.

— **Label similarities in Inputs for an elevate percentage**: when analysing the label comparison of the Input-03.1C (SimStadt) with the Input-02C (ENERGIS), it can be seen that **for the 11,100 buildings in common** they have the same label for 62.9% of the buildings. The percentage of buildings with only one difference of label or without differences is 90.53%. So the buildings with more than one step of difference between the datasets are lower than 10%. It is also important to note that there are more labels identifying higher demands for the Input-02C (ENERGIS) than for Input-03.1C (SimStadt).

— **Similar heating energy demand:** the differences for the heating energy demand for the Input-03.1C (SimStadt) and the Input-02 (ENERGIS) seems a Gaussian function centred near 0 value (but in the negative axis x), confirming the results in the label similarity: the heating energy demand for Input-02C is slightly higher than the one for Input-03.1C.

— **Extreme differences occur for a significant number of buildings**: the buildings that present a difference between datasets lower than -50 kWh/year*m² and those higher than 50 kWh/year*m² are many. More information should be extracted for these buildings in order to identify the origin of these differences.

In this case study CS2.2, it would be advisable to:

— Analyse the buildings that present differences in the labels of more than one step

— Analyse the buildings that present differences in the heating energy demand in absolute value higher than 50 kWh/year*m². It is important to analyse that considering the ranges of each label (the same difference in kWh/year*m² can affect differently depending on the label analysed).

6.4.3 Case study 03: Comparison of results with real EPCs

Table 47 defines the main parameters of Case Study 3. It also provides the references to the sections of the document where the individual results of each of the simulations can be seen and the sections where the comparisons among results are presented.

Table 47. Case study 3: main parameters

Case Study 3. Comparison of results with real Energy Performance Certificates									
Case study code	Scale	SimStadt	ENERGIS	Input 01	Input 02	Input 03.1	Input 03.2	Input 04	Comparison
Case Study 3.1	District	X	X	X	X	X	X	(X)	Sec. 6.4.3.1
Case Study 3.2	City	X	X	-	X	X	-	(X)	Sec. 6.4.3.2

Source: own elaboration

This initial step was necessary in order to count with the same number of elements in the comparison and have a homogeneous reference to compare to. In this line, it is worth to highlight the approach performed with real Energy Performance Certificates, since they can be provided either at the Building Unit level (e.g. dwelling level), or the building level. In order to cope with these differences and achieve a common comparison unit, a simplified approach was implemented to translate real EPCs at building unit level to building level. In particular, since the Junta de Castilla y León does not provide the corresponding cadastral references of the buildings of the energy performance certificates, nor the surface of the buildings they refer to, a simple average of the elements contained in a building was performed. In other words, if there are 5 certificates in the same building, the specific heating energy demand (expressed in kWh/m2*year) was added and divided, in this case, by 5.

In any case, the result obtained for each building should be considered as an estimation of the real EPC, since the behaviour of the whole building is represented considering only a part of it (only the EPCs that are available). In our case study in Valladolid, in many of the buildings the dwellings with available EPCs underrepresent all the dwellings in the building. Therefore, even if the Real EPCs are taken as a dataset for validation, we must consider that they would not always represent the reality completely.

Besides, an additional problem should be highlighted from this approach, consisting in the unknown proportion in which each specific heating demand value contributes to the whole specific heating demand value of the building. This fact is also combined with the lack of information on where the dwelling is located in the building; since the Energy Performance Certificate of a dwelling located at the top of a building block has more contact to the exterior through the roof, this dwelling does not perform in the same way as a dwelling located in the middle of a multi-family building.

Evidently, this is a simplified approach that could lead to inaccuracies, which need to be highlighted at the beginning of this comparison. In any case, only partial solutions could be provided to solve this issue by using cadastral data. Namely, implementing a more complex query to the cadastre to extract the surface corresponding to each dwelling. Nevertheless, the exact location of the dwelling within the building would never be available according to today's available data from the cadastre.

6.4.3.1 CS3.1: Comparison of results with real EPCs at district scale

6.4.3.1.1 CS3.1 Data preparation

In this case study, three datasets are considered, which have been transformed from BU, BU Parts or Building Units to BU as follows. This initial step was necessary in order to count with the same number of elements in the comparison and have a homogeneous reference to compare to.

Table 48. Case study 3.1: data preparation required

Name	Initial	Initial elements	Target	Final elements
01D_small_4marzo.csv	BU	28 buildings	BU	28 buildings
02D_4marzo.geojson	BU	189 buildings	BU	189 buildings
031D_4marzo.csv	BUParts	286 BU Parts	BU	206 buildings
032D_4marzo.csv	BU	205 buildings	BU	205 buildings
RealEPC_D_4marzo.geojson	BU	156 EPCs at BU_Unit	BU	156 EPCs at Building level

Source: own elaboration

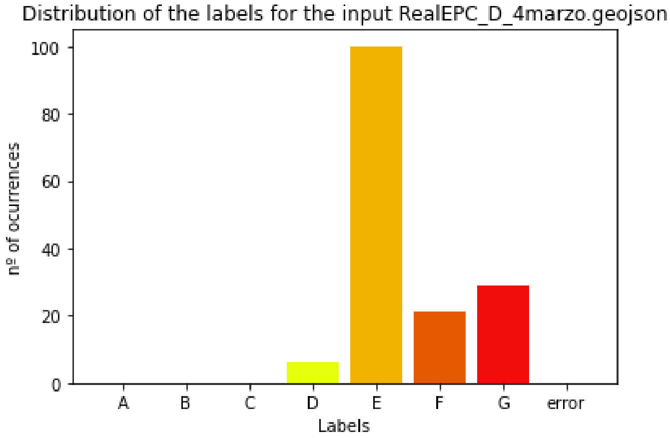
6.4.3.1.2 CS3.1 Data analysis

The analysis of the data used in this case study can be seen in the following tables included in this sub-section. The data have been compared with the following data:

Table 49. Real EPC for Cuatro de Marzo district (CS3.1)

Case Study 3.1: Results for RealEPC_D_4marzo (156 buildings)

Figure 114. Distribution of the labels for Real EPCs – District level



Source: own elaboration based on real EPCs from EREN database

Label	Buildings	Percentage
Label A:	0	0.00%
Label B:	0	0.00%
Label C:	0	0.00%
Label D:	6	3.85%
Label E:	100	64.10%
Label F:	21	13.46%
Label G:	29	18.59%
Label error:	0	0.00%

Source: own elaboration based on real EPCs from EREN database

6.4.3.1.3 CS3.1 Data comparison

Data comparison is performed per set of two datasets, considering in every case the buildings they have in common. In this case, each of the Inputs is compared to the validation environment at hand: Real EPCs. This is shown in 0, Table 52, Table 53 and Table 54. In order to provide an overview of the comparisons presented in this section, the following Table 50 is presented.

Table 50. Label comparison (CS3.1)

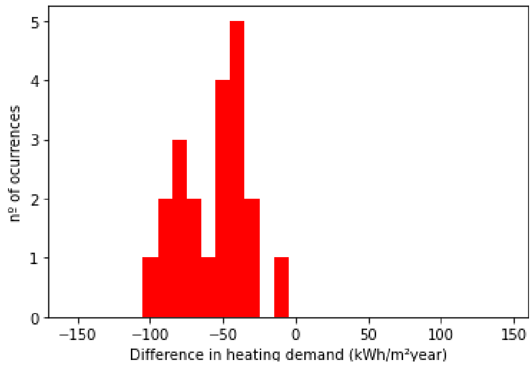
Case Study 3.1: Label comparison					
Label	01D_small	02D	031D	032D	Real EPCs
Label A:	0.00%	0.00%	0.00%	0.00%	0.00%
Label B:	0.00%	0.00%	0.00%	0.49%	0.00%
Label C:	0.00%	0.00%	0.97%	2.93%	0.00%
Label D:	92.86%	0.00%	2.91%	4.39%	3.85%
Label E:	7.14%	100.00%	96.12%	91.22%	64.10%
Label F:	0.00%	0.00%	0.00%	0.49%	13.46%
Label G:	0.00%	0.00%	0.00%	0.49%	18.59%
Label error:	0.00%	0.00%	0.00%	0.00%	0.00%

Source: own elaboration based on SimStadt results

Table 51. CS3.1: comparison of Input 01D (small) and RealEPC_D

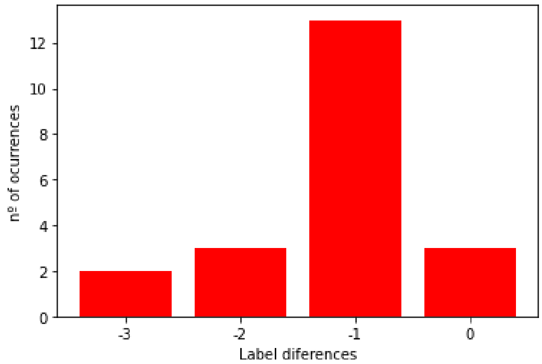
Case Study 3.1: Comparison of Input 01D (small) and RealEPC_D								
Buildings in common	21	Label differences (in % and # buildings)						
All buildings are present in both datasets		-3	-2	-1	0	1	2	3
		9.52%	14.29%	61.9%	14.29%	0%	0%	0%
		2	3	13	3	0	0	0

Figure 115. Heating demand differences for datasets: Input 01D (small) and RealEPC_D



Source: own elaboration

Figure 116. Label differences for datasets: Input 01D (small) and RealEPC_D

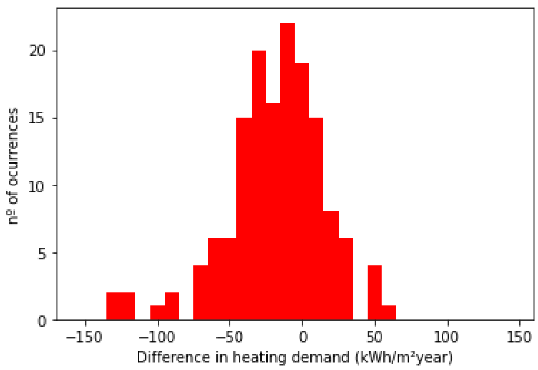


Source: own elaboration

Table 52. CS3.1: comparison of Input 02D and Real EPC_D

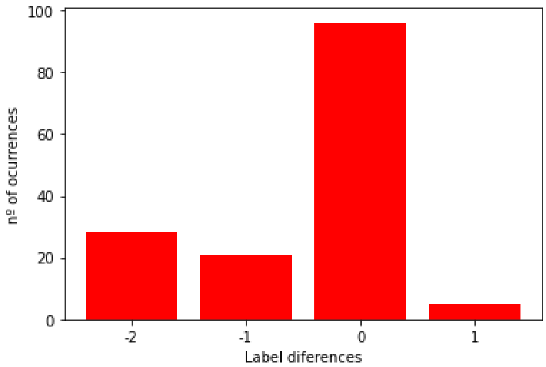
Case Study 3.1: Comparison of Input 02D and RealEPC_D						
Buildings in common	150	Label differences (in % and # buildings)				
All buildings are present in both datasets		-2	-1	0	1	2
		18.67%	14%	64%	3.33%	0.00%
		28	21	96	5	0

Figure 117. Heating demand differences for datasets: Input 02D and RealEPC_D



Source: own elaboration

Figure 118. Label differences for datasets: Input 02D and RealEPC_D



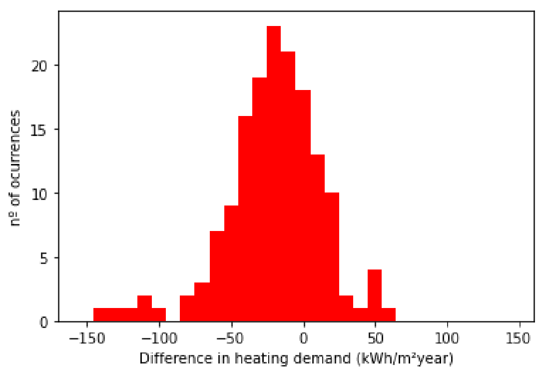
Source: own elaboration

Source: own elaboration

Table 53. CS3.1: comparison of 03.1D and RealEPC_D

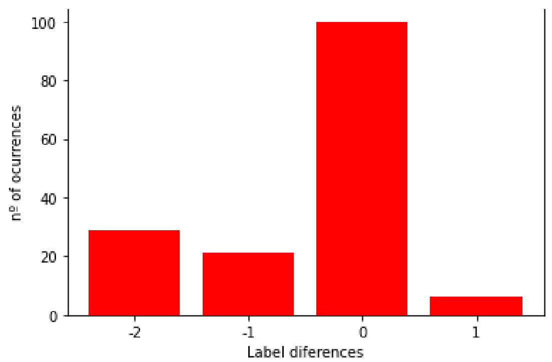
Case Study 3.1: Comparison of Input 03.1D and RealEPC_D						
Buildings in common	150	Label differences (in % and # buildings)				
All buildings are present in both datasets		-2	-1	0	1	2
		18.59%	13.46%	64.10%	3.85%	0.00%
		29	21	100	6	0

Figure 119. Heating demand differences for datasets: Input 03.1D and RealEPC_D



Source: own elaboration

Figure 120. Label differences for datasets: Input 03.1D and RealEPC_D



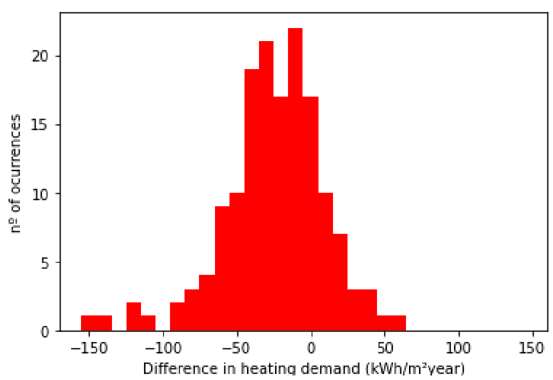
Source: own elaboration

Source: own elaboration

Table 54. CS3.1: comparison of 03.2D and RealEPC_D

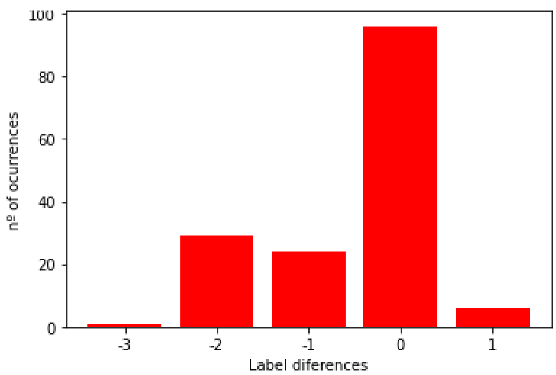
Case Study 3.1: Comparison of Table 1. of Input 03.2D and RealEPC_D						
Buildings in common	156	Label differences (in % and # buildings)				
All buildings are present in both datasets		-3	-2	1	0	1
		0.64%	18.59%	15.38%	61.54%	0.00%
		1	29	24	96	6

Figure 121. Heating demand differences for datasets: Input 03.2D and RealEPC_D



Source: own elaboration

Figure 122. Label differences for datasets: Input 03.2D and RealEPC_D



Source: own elaboration

Source: own elaboration

6.4.3.1.4 CS3.1 Conclusions

Four different inputs have been compared in Case Study 3.1, namely Input 01-D (CityGML generated ad-hoc), Input-02D (GML coming directly from the cadastre), Input-03.1D (CityGML LOD 1 based on cadastral input), and Input-03.2D (CityGML LOD 1 based on cadastral input and real heights from LiDAR data). All of them are compared to the validation environment 1: results from Real EPCs available in the Junta de Castilla y León website, processed as explained in section 6.4.3.1.1.

The number of buildings covered in each Input is the same:; however, in the case of Input 01-D, the district is smaller and a lower number of buildings is considered. This is to be considered when viewing the results shown in Table 50.

From these comparisons, several conclusions can be extracted:

- **Homogeneity of results**: when comparing the results in general (Table 50), the majority of the buildings in Inputs 02-D, 03.1-D and 03.2-D achieved a label E (100%, 96.12% and 91.22%, respectively). In the case of the Real EPCs, this is also the most common label, but with a lower share as the other inputs (64.10%). In general, it is observed that Real EPCs depict a less efficient situation (label E – 64.10%, label F – 13.46%, G – 18,59%) than the results obtained from the simulations, which are normally better (from label E to more efficient).
- **Overall shift of values in one direction**: due to the abovementioned observation that real EPCs show less efficient values than the simulations, when comparing the differences in labels or the differences in energy demand, an overall shift of the values is observed. As a consequence: in Input 01-D (0) label differences of one up to three steps can be observed; in Input 02-D (Table 52) the values are more centred, but again label differences of up to 2 labels are observed; and in Input 03.1-D (Table 53) the results are quite similar in terms of distribution to those of Input 02-D. However, Input 03.2 (Table 54) presents higher differences with up to two and three label steps. These differences correspond to 19,23% of the buildings.
- **Extreme differences in values can be due to outliers**: in addition to the abovementioned general shift, in all of the graphs where the differences in energy demand are compared (Figure 115, Figure 117, Figure 119 and Figure 121), discontinuities in the graphs are observed. In particular, Figure 115 shows two different parts, and an extra set of heating demand difference which is closer to zero. If compared to Figure 116 these differences are subtler. This case is more evident in Figure 117 (Input 02), where the majority of the results are in the heating demand difference range of -75 to +40 kWh/m2*year, and extreme values correspond to differences of around -100 kWh/m2*year and + 50kWh/m2*year. Similarly, Table 53 (Input 03.1-D) and Table 54 (Input 03.2-D) present the same results with extreme values being lower as the previous.
- **Lack of correspondence of the shape of heating demand graphs and label graphs**: these variations highlight the fact that the differences appreciated in labels are not directly comparable to the differences encountered in heating energy demand. This is due to the fact that the ranges established to define the labels are not equivalent among each other and some labels cover a broader range of heating energy demand values than others. For more information to this respect, please refer to the introduction of section 6.4.

The most important conclusion for this specific case study is that the two tools in general obtain results with higher efficiency compared to the results offered by the real EPCs.

6.4.3.2 CS3.2: Comparison of results with real EPCs at city scale

6.4.3.2.1 CS3.2 Data preparation

In this case study, three datasets are considered, which have been transformed from BU, BU Parts or Building Units to BU as follows:

Table 55. Case study 3.2: data preparation required

Name	Initial	Initial elements	Target	Final elements
02C_Valladolid.geojson	BU	11.262 buildings	BU	11.262 buildings
031C_Valladolid.csv	BUParts	88.894 BU parts	BU	14.738 buildings
RealEPC_D_4marzo.geojson	BU_Unit	20.148 EPCs at BU_Unit	BU	5.081 EPCSs at building level

Source: own elaboration

This initial step was necessary in order to count with the same number of elements in the comparison and have a homogeneous reference to compare to.

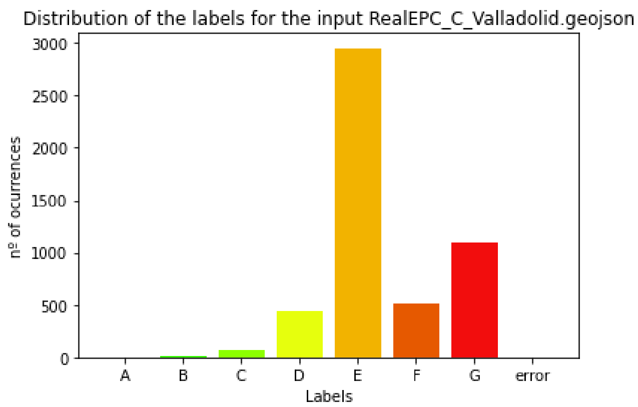
6.4.3.2.2 CS3.2 Data analysis

The analysis of the data used in this case study can be seen in the following Table 56. The data have been compared with the following data:

Table 56. Real EPC for Valladolid (CS3.1)

Case Study 3.2: Results for RealEPC_C_Valladolid (5081 buildings)

Figure 123. Distribution of the labels for Real EPCs – City level



Source: own elaboration based on real EPCs from EREN database

Label	Buildings	Percentage
Label A:	4	0.08%
Label B:	13	0.26%
Label C:	67	1.32%
Label D:	444	8.74%
Label E:	2952	58.10%
Label F:	510	10.04%
Label G:	1091	21.47%
Label error:	0	0.00%

Source: own elaboration based on real EPCs from EREN database

6.4.3.2.3 CS3.2 Data comparison

Data comparison is performed per set of two datasets, considering in every case the buildings they have in common. In this case, each of the Inputs is compared to the validation environment at hand: Real EPCs. This is shown in Table 58 and 0. An overview of the comparisons is presented in Table 57.

Table 57. Label comparison (CS3.2)

Case Study 3.2: Label comparison			
Label	Input 02C	Input 031C	Real EPCs
Label A:	0.00%	0.35%	0.08%
Label B:	0.00%	0.30%	0.26%

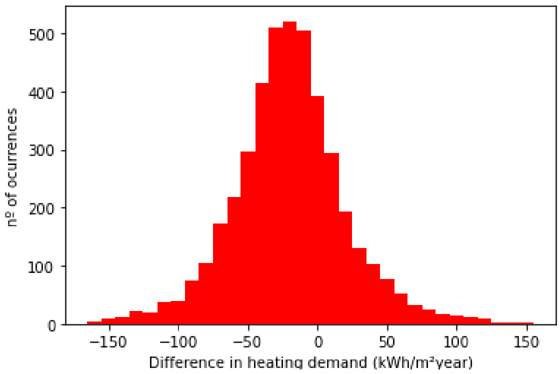
Label C:	0.21%	2.88%	1.32%
Label D:	9.87%	20.57%	8.74%
Label E:	53.21%	40.22%	58.10%
Label F:	6.45%	7.23%	10.04%
Label G:	30.26%	28.44%	21.47%
Label error:	0.00%	0.02%	0.00%

Source: own elaboration based on SimStadt and ENERGIS results

Table 58. CS3.2: comparison of 02C with RealEPC_C

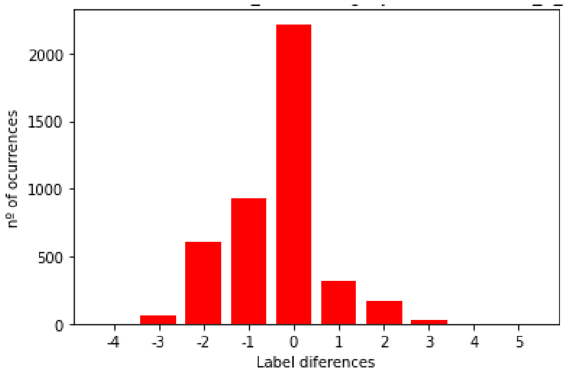
Case Study 3.2: Comparison of 02C with RealEPC_C												
Buildings in common												4.345
Label differences (in % and # buildings)												
-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
0%	0%	0.02%	1.47%	14.06%	21.50%	51.00%	7.25%	3.87%	0.74%	0.07%	0.02%	0%
0	0	1	64	611	934	2.216	315	168	32	3	1	0

Figure 124. Heating demand differences for datasets: Input02-C and RealEPC-C



Source: own elaboration

Figure 125. Label differences for datasets: Input02-C and RealEPC-C



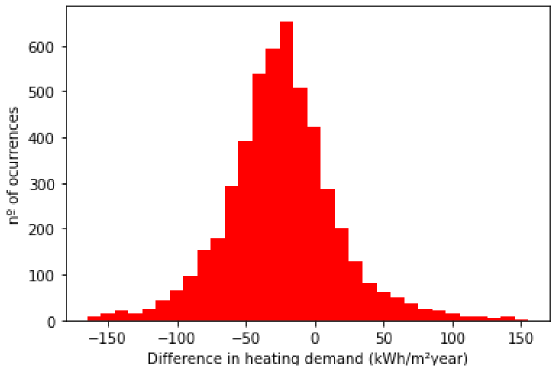
Source: own elaboration

Source: own elaboration

Table 59. CS3.2: comparison of 03.1-C with RealEPC_C

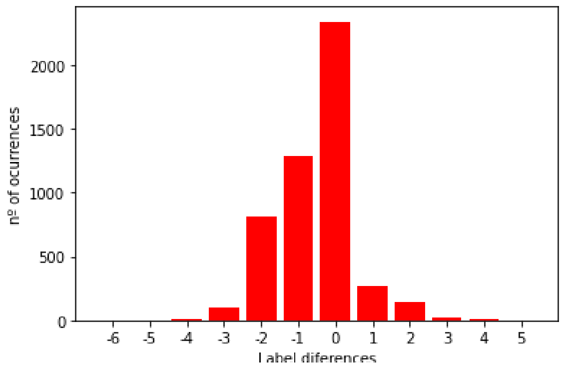
Case Study 3.2: Comparison of 03.1-C with RealEPC_C												
Buildings in common												5.003
Label differences (in % and # buildings)												
-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
0.02%	0.04%	0.12%	2.04%	16.31%	25.82%	46.87%	5.36%	2.80%	0.52%	0.08%	0.02%	0%
1	2	6	102	816	1.292	2.345	268	140	26	4	1	0

Figure 126. Heating demand differences for datasets: Input03.1-C and RealEPC-C



Source: own elaboration

Figure 127. Label differences for datasets: Input03.1-C and RealEPC-C



Source: own elaboration

Source: own elaboration

6.4.3.2.4 CS3.2 Conclusions

Two different inputs have been compared to Real EPCs (refer to section 6.4.3.1.1 for details on their processing) in Case Study 3.2, namely Input 02-D (GML coming directly from the cadastre) and Input-03.1D (CityGML LOD 1 based on cadastral input).

The scale that is compared is the city scale, and the buildings compared in each case have to do with the available Energy Performance Certificates in the city of Valladolid. This is to be considered when viewing the results shown in Table 57. From these comparisons, several conclusions can be extracted:

— **Homogeneity of results**: in general, when observing Table 57 and bearing in mind the abovementioned considerations, it can be seen that the majority of the buildings in Inputs 02-C and 03.1-C achieved an E label (53.21% and 40.22% respectively). This is also true for the real EPCs datasets, with a 58.1% of E label. In the case of the second label with more occurrences also is the same for the three datasets, the G label in this case.

— **Label similarities for an considerable percentage respect to the real EPCs**: when analysing the label comparison of the Input-02C (ENERGIS) with the real EPC **for the 4,345 buildings in common**, it can be seen that they have the same label for 51% of the buildings. Besides, the percentage of buildings with only one step of label or without differences is 79.75%. This is a high percentage but is lower than when comparing the results of the different tools between them. So the buildings with more than one step of difference between the datasets are around 20%, a significant value. In the case of the 5,003 building in common of the Input-03.1D and the real EPCs there is a coincidence in the label for the 46.87% and a percentage of building with less than two steps in the label difference of 78.05% (slightly lower than for the Input-02C). In both cases, for the buildings that do not have the same label than in the Real EPCs dataset, there are more labels reflecting more energy efficiency with respect to the real ECPs dataset

— **Heating energy demand:** the differences for the heating energy demand for the Input-03.1C (SimStadt) and the Input-02 (ENERGIS) respect to real EPCs reflect a Gaussian function not centred in the 0 value, but approximately in -25 kWh/year*m², confirming the results in the label similarity: the heating energy demand for Input-02C and Input-03.1C are lower than those for the real EPCs.

— **Extreme differences occur for a significant number of buildings**: the buildings that present a difference between datasets lower than -75kWh/year*m² and those higher than 25kWh/year*m² are many. More information should be extracted for these building in order to identify the origin of these differences.

6.4.4 Conclusions

This section 6 of the report has aimed to present different methodologies to calculate energy performance labels and compare the results among them in the city of Valladolid (Spain) by both considering a specific district with more or less homogeneous buildings, and the city as a whole. To this end, model by model comparisons have been performed. These comparisons focused on three main axes:

— **Generation of input data**: to understand how the generation of datasets affects the final results. To achieve this objective, several methods to generate models have been explored: generating ad-hoc CityGML models (Input-01), extracting data from public data sources like the Spanish cadastre (Input-02), and generating CityGML models based on the mentioned data (Input-03-1). Also, the latter model was combined with LiDAR data in order to obtain more precise heights (Input 03.2). Finally, a crowdsourced dataset such as OpenStreetMap was also used as a base to generate models of the city (Input-04).

— **Different simulation environments**: to understand what impact they have on results. In this sense, the results from two tools sharing a similar objective were compared with each other: SimStadt and ENERGIS. The differences among them were also explored.

— **Comparison to real Energy Performance Certificates**: to compare the simulations performed to the closest real data that is available at this level. In this line, the publicly available data on Energy Performance Certificates presented by the Junta de Castilla y León in their open data portal was used as a validation environment.

To explore these different axes, first the addressed challenge was presented (section 1). Then the data input used was explained, as well as the method used towards its generation (section 6.2). Special attention was placed to the data processing required and its enrichment, since it can have an impact on the results obtained. Five different Inputs are considered in total, all of them at district scale (Cuatro de Marzo district) and some also at city scale (Valladolid, Spain) (Table 13), on the methodology used to generate the data. The following section 6.3 showed the two simulation environments used in the comparisons (SimStadt and ENERGIS) and explained the validation environment (Energy Performance Certificates of Castilla y León). Finally, in section 6.4 results of the comparisons were presented. Each of the three main case studies was linked to one of the abovementioned axes and was associated to the two scales considered. Case Study 1.1 and Case Study 1.2 tackled the first axis, at district and city scale, respectively. Case 2.1 and 2.2 focused on the axis related to different simulation environments, at district and city scale, respectively. The comparison to real EPCs was performed in Case Studies 3.1 and 3.2, at district and city scale, respectively. In each case study the data preparation, data analysis, data comparison and conclusions were presented. Based on this process, the following conclusions can be extracted:

General considerations based on the scale tackled: the conclusions that could be extracted depending on the scale tackled varied, since it was easier to derive conclusions when a lower number of buildings were considered. In this sense, the selection of a district like Cuatro de Marzo was appropriate, since it counts on a relatively large number of buildings, but the residential buildings only have two typologies: multi-family block and multi-family tower, with the same building heights, number of dwellings etc. This allowed to understand the results more easily.

Conclusions based on axis 1: Generation of different input data: to interpret appropriately the results presented in section 6.4.1, the generation of the city models. (6.2) should be also taken into account. In particular, specific assumptions in the city model generation (such as the consideration of a specific average height, or the assumption of the window-wall-ratio) can lead to an increase in heating demand values. Additionally, an aspect that should be carefully handled is the consideration of the level of granularity of the building data. Indeed, considering a high level of granularity and analysing building parts in contrast to considering “cleaner” models, can result in an increased heating demand, due, predictably, to the increased external wall surface through which heating energy can be lost (refer to 6.4.1.1.4). Certainly, the careful analysis of external wall surfaces, heated volume, window-wall-ratios applied, as well as the analysis of shared walls should be considered as necessary next steps in these analyses.

The attention to granularity in the tests presented in the report was also linked to the possibility to identify the buildings or the building elements. In this line, the identification code used in the cadastre (cadastral reference) was highly beneficial in order to compare results. This fact was not only linked to having a common ID, but also to the fact that they referred to the same geometrical element. This was not the case of OpenStreetMap, since there was no cadastral reference to link data to, but one primitive named "ways". In this case, this could have been solved by previously processing this data and assigning the corresponding cadastral reference to each "way". However, a test should be performed in order to know if discrepancies exist with respect to the geometric definition of the footprints from both data sources (OSM and cadastre).

Another aspect that should be explored is the need or not to increase the accuracy of the generated city models and assess from a cost-benefit perspective whether it is worth to spend resources for improving the accuracy of these models or whether the estimation of certain parameters (such as building height, use, etc.) is good enough. This applies especially to the Inputs generated with OpenStreetMap (Input-04) and those complemented with LiDAR data (Input 03.2). In the first case, a quick approach is presented to deal with the lack of information present in OpenStreetMap in Valladolid. As a result, most of the buildings in the city have the same year of construction (1960), height (15 meters) and use (residential). Potentially, this approach could be easily reused in any city around the world. In contrast, the approach presented in Input-03.2, where LiDAR data was analysed to extract the real height of the buildings entailed a time-consuming process, which was partially performed manually in order to be able to re-classify the point clouds not correctly classified. Thus, it involved a lot of resources. A cost-optimal analysis of both approaches with more cases would be required to be able to determine the appropriateness of using one method or the other, which highly depends on the use of results.

Conclusions based on axis 2: Different simulation environments. . to explore the impact of using different simulation environments, the results presented in section 6.4.2 should be analysed together with the description of the validation environment presented in section 6.3. In this regard, it should be noted that even though a brief comparison of both approaches has been presented in Table 20, a deeper understanding of both tools would be necessary in order to better interpret the results. This could be achieved by simulating smaller city models to understand the different results offered by both tools. In addition, a special focus should be placed on understanding the role of the building physics libraries used by both tools, as well as the building usage libraries. In the comparisons performed, default values have been used for both tools. This implies using the German Building Physics Library for models simulated with SimStadt, and a Spanish Building Physics library for models simulated with ENERGIS. A first next step would be to generate a Spanish Building Physics library to be used in SimStadt to check if there are significant differences. It is important to note that the comparison between the two tools is done only for residential buildings, because ENERGIS simulations are limited only to residential buildings.

In any case, in the comparisons performed between both tools it is worth to highlight that there were not significant differences in the results obtained, in particular when comparing Input-02 (GML from cadastre) and Input-03.1 (CityGML model from cadastre), i.e. inputs originating from the same source and based on similar hypotheses. This can potentially indicate that the tools simulate in a very similar way. However, further tests should be performed with other building physics libraries.

Conclusions based on axis 3: Comparison of results with real EPCs. The third axis focused on the data closest to real data available at this scale, consisting in Energy Performance Certificates, assuming the validity of their content. These comparisons entailed three main challenges: (1) not all buildings have an EPC, (2) the level of granularity of EPCs varied, with some of them related to building elements (dwellings) and others related to buildings and (3) the comparison value used was the energy performance label for heating energy demand. These challenges were tackled by implementing a simple approach based on averages to have an estimated value of the energy performance of buildings where EPCs existed. In order to improve this approach, it would be necessary to link the results to the heated surface they correspond to. However, this would not solve all the problems encountered, as discussed in section 6.4.3.

In any case, the analyses performed offered the same observation both at district and city level: Energy Performance Certificates usually depicted higher heating demand values than those calculated with simulation tools (SimStadt and ENERGIS), with heating energy demand differences in a range from -160 to 160 kWh/year*m² but centred in 25 kWh/year*m². The origin of this difference should be explored in more detail, since the tools used by experts to generate the EPCs are similar to SimStadt and ENERGIS. Therefore, no significant differences should exist in the parameters/variables used by the tools, apart the way how the information is modelled by experts in the tools: both geometric, as well as that related to the building thermal envelope. To analyse this in more depth, it would be advisable to compare how the results have been modelled in EPCs and compare them to the approaches implemented by the SimStadt or ENERGIS tools. However, even though this information is contained within Spanish EPCs, it is not publicly available in Castilla y León. An intermediate solution would be to calculate (by an expert) EPCs of a series of buildings were all the real information is known and compare the results with the approaches proposed in SimStadt and ENERGIS.

As a concluding remark, it should be highlighted that the main aim of this report was to perform model by model comparisons to explore their differences and detect potential improvements. However, when aiming at using the results from these simulations in real life, it is important to calibrate the model with real data. This has been sought the best possible way by resorting to the data source that was closest to reality in this context: real EPCs. In this line, it was assumed that these real EPCs were correctly calculated, even when systematic checks are not performed on all of them[41]. A more in-depth analysis of real energy consumption by analysing monitoring data would be required. However, it should be handled with care as well, since energy consumption goes one step further to energy demand calculation, since it considers not only the efficiency of the HVAC system implemented, but also the behaviour of the user. A separation between the energy consumption derived from (1) climatic conditions and building envelope, (2) functioning of HVAC system and (3) user behaviour and control systems would be required.

In any case, as discussed before, it would be necessary first of all to determine the purpose the results of the models are used for, in order to know the accuracy required for each specific decision to be taken. In this era where humanity is overflown with data is more necessary than ever to know the precise objectives that are sought with each decision and what kind of variables would affect the achievement of those objectives.

[41] Energy Performance Certificates management authorities are required to implement compliance and checking mechanisms to the EPCs. However, this does not imply that all of them are checked, and usually random checks are performed on submitted EPCs.

7 INSPIRE harmonization of input/output data for energy simulations

7.1 Introduction

Several approaches can be applied to assess the energy performance of a building, each of them having different requirements in terms of input data and different methodological complexity, determining different levels of accuracy of the results obtained. Among several energy simulation tools available, the SimStadt software [6], developed by the University of Applied Science Stuttgart, allows to perform energy simulation at city district level and beyond, providing an assessment of energy heat demand at building level. This is expressed in KWh/m^2 per month or year, using as input data CityGML LoD1 or LoD2 3D data of the simulated buildings, together with period of construction, typology and usage of the buildings.

Despite CityGML is an encoding standard widely used for 3D city models, it is worth to explore the possibility to use also 3D building data encoded in a different format, such as INSPIRE 3D building data, as input data for SimStadt and for all the energy simulation tools.

Moreover, it is also worth to explore the possibility to improve the interoperability of SimStadt output data, as well as of all the energy simulation tools providing an assessment of energy demand at building level, with data containing energy related information at building level, such as EPC datasets.

In the frame of this use case, two different scenarios have been considered and described in section 2:

- a mapping exercise between CityGML datasets and INSPIRE Building 3D data models has been performed to enable the use of SimStadt software to assess the energy heat demand of the building stock in all the cities, regions and countries for which building datasets in conformity to INSPIRE Building 3D application schema are available.
- A second scenario has been considered, in which 3D buildings data compliant to INSPIRE Building 3D data model are used as input data for the energy simulation tool.

For both the scenarios described above, a possible extension of the INSPIRE Building 3D data model, based on the EPC4EU data model (representing an extension of the INSPIRE Building 2D data model developed in the frame of other use cases of the Energy & Location Applications to harmonise EPC datasets[42]) has been drafted. This new extension enables the inclusion in the same dataset (INSPIRE compliant) of the energy simulation results generated by SimStadt and the information contained in the EPCs. This part is described in section 3.

A series of resources developed during the execution of the activities reported in this section (e.g. mapping tables, transformation projects, examples of harmonised data) are available for download in the dedicated JoinUp page[43].

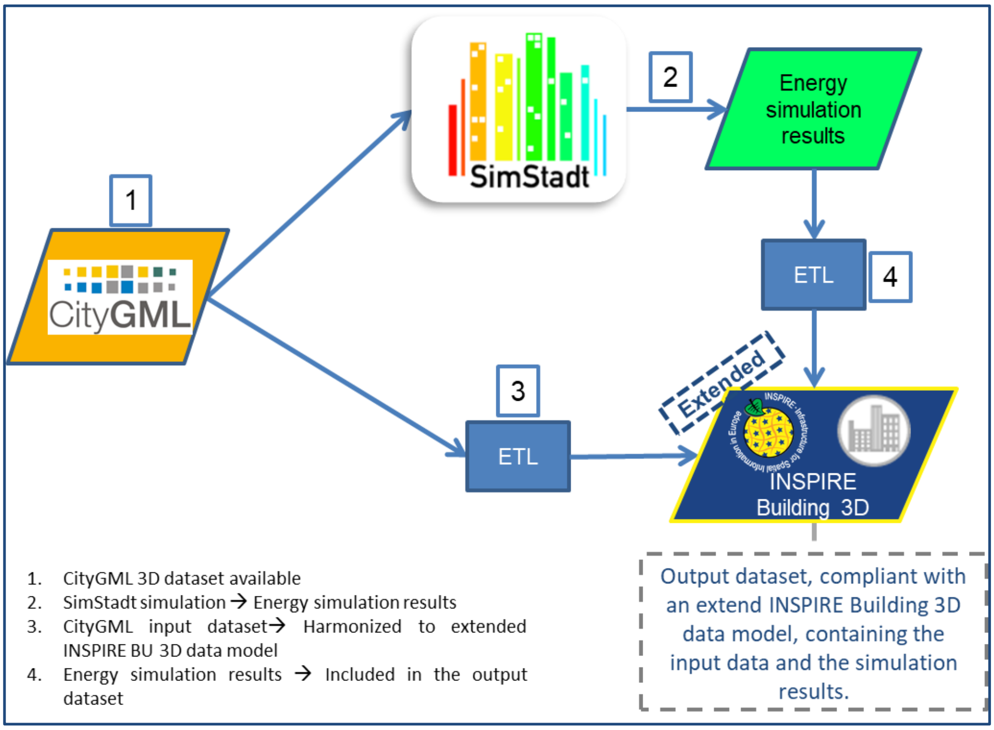
7.2 Data transformation

In the first scenario the availability of a CityGML dataset has been considered as starting point of the process. This type of dataset can be used directly as input data for the energy simulation tool assessing the energy heat demand of the buildings. The CityGML input dataset can be transformed to an extended INSPIRE BU 3D data model, in order to integrate in a unique dataset the original building information and the results of the energy simulation. This scenario is illustrated in Figure 128 below.

[42] https://joinup.ec.europa.eu/node/704567

[43] https://joinup.ec.europa.eu/node/704529

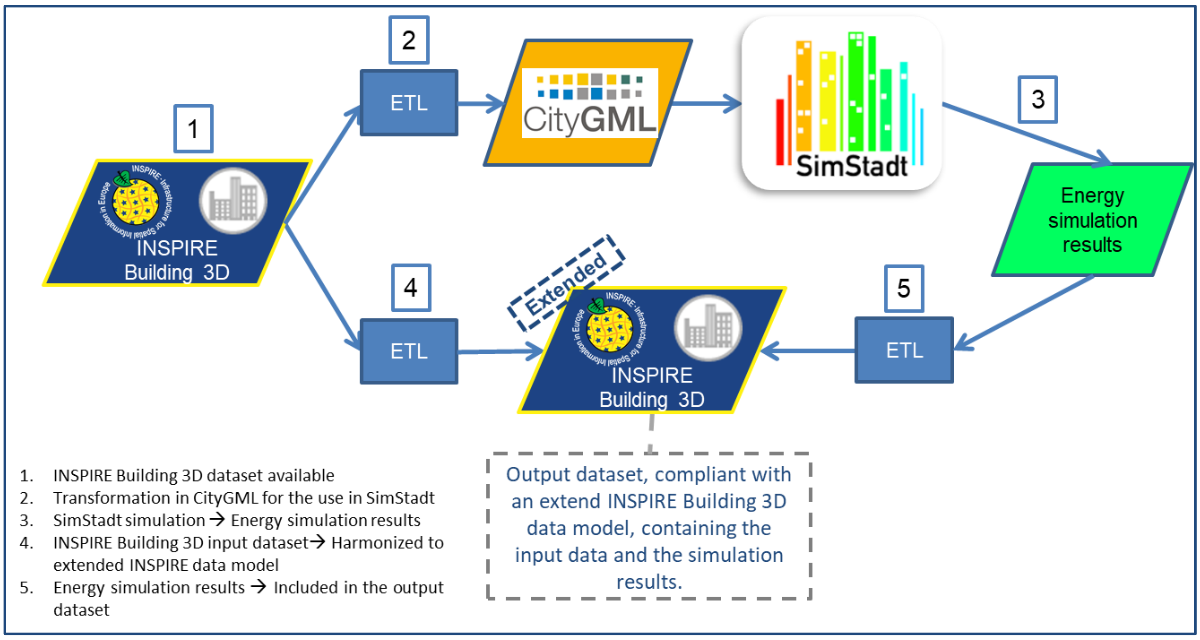
Figure 128. Data transformation - Scenario 1



Source: own creation, JRC, 2020.

In the second scenario the availability of an INSPIRE BU 3D dataset has been considered as starting point of the process. This type of dataset cannot be used directly as input data for the energy simulation tool, at least for SimStadt. In this case a new transformation step is needed to make available the source dataset compliant to the CityGML data model and thus suitable as input data for the energy simulation tool. The INSPIRE BU 3D initial dataset can be transformed to an extended INSPIRE BU 3D data model to integrate in a unique dataset the original building information and the results of the energy simulation. This scenario is illustrated in Figure 129 below.

Figure 129. Data transformation - Scenario 2



Source: own creation, JRC, 2020.

This section explains how the mapping exercise between CityGML and INSPIRE Building 3D data has been performed. After a general description of the methodology of data transformation, the description of how it has been applied to this specific case is provided. Several transformation exercises have been performed between the different versions of a CityGML dataset and the different versions of the INSPIRE Building 3D data model. These transformations are related to the scenario 1 described above.

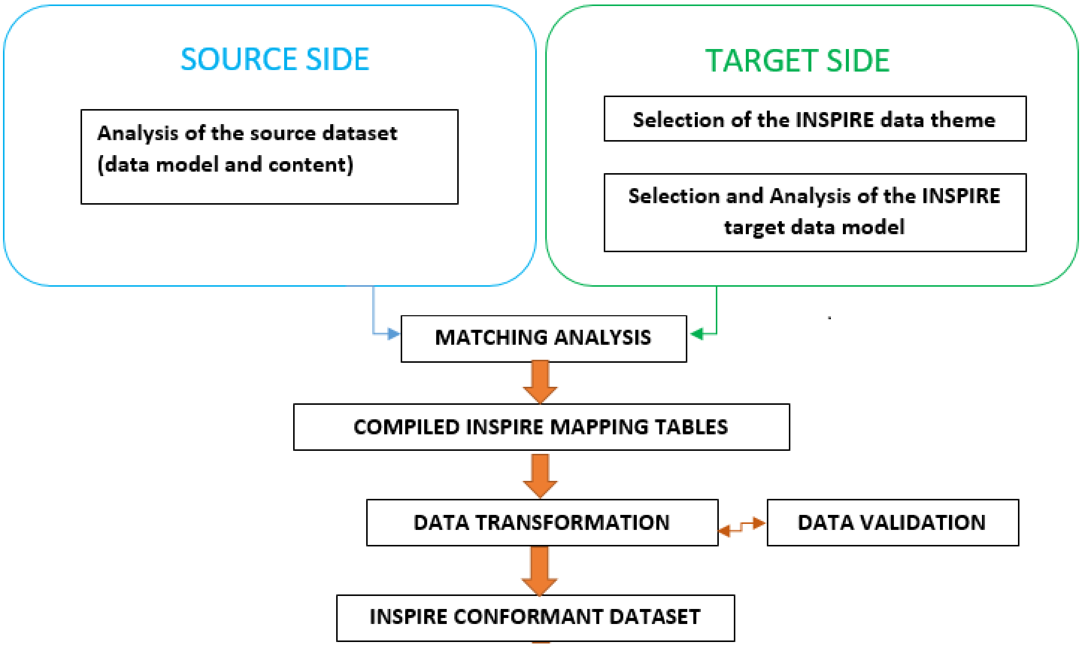
The reverse transformation from INSPIRE Building 3D dataset to CityGML, considered in the scenario 2, has been only investigated from a conceptual point of view. The transformation can be easily executed applying the methodology described in the following paragraphs.

7.2.1 General methodology of data transformation

THE STEPS OF THE INSPIRE HARMONISATION PROCESS

The workflow in Figure 130 illustrates the process through which heterogeneous sources of spatial data can be transformed and validated according to requirements of INSPIRE ISDSS Regulation[44] (and subsequent amendments) and INSPIRE Data Specifications and Technical Guidelines[45] (including latest amendments and corrigenda).

Figure 130. INSPIRE data harmonisation process



Source: own creation, JRC, 2020.

The following subsections describe in detail the three main steps of the INSPIRE data harmonisation process and include references to the tools that can be used to perform the tasks described therein.

Step 1: analysis of source and target data models

The starting point of the INSPIRE data harmonisation process consists in performing a deep analysis of:

— the content and the structure of the source dataset to be harmonised ('source' data model)

— the INSPIRE Implementing Rules as regards interoperability of spatial datasets and services as well as the INSPIRE Data Specification Technical Guidelines (including latest amendments and corrigenda).

The purpose is to identify:

— the INSPIRE spatial data theme under which the dataset falls (this choice is not always straightforward, since a single dataset structure and content could be related to more than one INSPIRE theme, so the analysis is aimed to select the best fitting one),

[44] https://inspire.ec.europa.eu/documents/commission-regulation-eu-no-10892010-23-november-2010-implementing-directive-20072ec-0

[45] https://inspire.ec.europa.eu/data-specifications/2892

— the 'target' data model against which the source data has to be harmonised (usually more than one data model / application schema is available for the same data theme, therefore the analysis is aimed to select the data model /application schema that best suits the source dataset content).

Step 1 is schematised in Figure 131.

Figure 131. Data Harmonisation Step 1- Inputs and Outputs

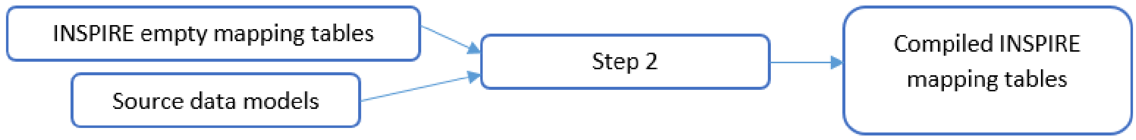
Source: own creation, JRC, 2020.

Step 2: mapping of source data model into target INSPIRE data model

A crucial step of the whole harmonisation process is the identification of the correspondences between the elements belonging to the 'source' data model and the INSPIRE 'target' data model, as illustrated in Figure 132.

This step includes both the '*schema matching*', aimed to identify semantically related elements between the source and target data models, and the '*schema mapping*' targeted to define the relevant transformation rules between matching elements.

Figure 132. Data Harmonisation Step 2 - Inputs and Outputs



Source: own creation, JRC, 2020.

A concrete '*gap analysis*' has to be performed, allowing for the identification of missing or incomplete attributes in the source dataset. Therefore, depending on the results of the *matching* and *mapping* operations, there could be the need for '*pre-processing*' activities on the source dataset, to prepare it for the transformation according to the INSPIRE data model e.g., change the Coordinate Reference System or the attributes data type/data type format, etc. In some cases, there could also be a need for an integration / modification of the source dataset content e.g., if some of the attributes required by INSPIRE are missing in the source data, or whether topological validation issues arise due to the possible geometry inconsistences in the source data. In such a case, the data provider should be contacted and full support and explanations should be given to overcome eventual issues.

To perform the 'schema matching'/'schema mapping' tasks, the ***INSPIRE mapping tables*** available on the INSPIRE website[46] can be used. Figure 133 shows the INSPIRE mapping table for the BU 3D Core data theme. A 'mapping table' for an INSPIRE data theme is an xml table describing the relevant INSPIRE application schema (listing feature types, attributes, data types, associations between feature types, constraints, etc.) and allowing for description of the source schema. With reference to Figure 133, the INSPIRE Application Schema section (on the left) is pre-filled, while fields of the source schema can be mapped using the columns on the left.

The *INSPIRE mapping tables* could be partially customised to better fit the mapping process e.g., adding rows for inline description of complex types structure.

[46] http://inspire.ec.europa.eu/Data-Models/Data-Specifications/2892

Figure 133. INSPIRE mapping table for BU 3D Core

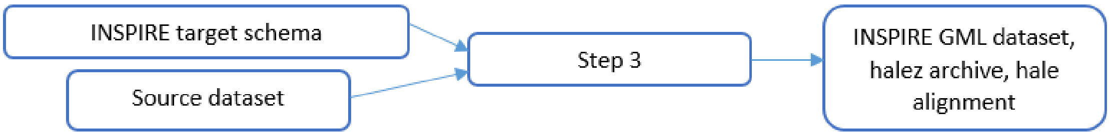
Application Schema 'Buildings3D' (version 3.0)							Application Schema <provide name of source schema>								
Type	Documentation	Attribute Association role Constraint	Attribute / Association role / Constraint	Values / Enumerations	Multiplicity	Voidable / Non-Voidable	Type	Documentation	Attribute / Association role	Attribute / Association	Values / Enumerations	Multiplicity	Voidable / Non-Voidable	Status	Remarks
Building Supertypes:*BuildingAbstractBuildingAbstractConstruction*	-- Name -- Building A Building is an enclosed construction above and/or underground, used or intended for the shelter of humans, animals or things or for the production of economic goods. A building refers to any structure permanently constructed or erected on its site.														
		beginLifespanVersion	-- Name -- Begin lifespan version	DateTime	1	voidable									
		conditionOfConstruction	-- Name -- Condition of	ConditionOfConstructionValue	1	voidable									
		dateOfConstruction	-- Name -- Date of construction	DateOfEvent	0..1	voidable									
		dateOfDemolition	-- Name -- Date of demolition Date	DateOfEvent	0..1	voidable									
		dateOfRenovation	-- Name -- Date of last major	DateOfEvent	0..1	voidable									
		elevation	-- Name -- Elevation	Elevation	0..*	voidable									
		endLifespanVersion	-- Name -- End lifespan version	DateTime	0..1	voidable									
		externalReference	-- Name -- External reference	ExternalReference	0..*	voidable									
		heightAboveGround	-- Name -- Height above ground	HeightAboveGround	0..*	voidable									
		inspireId	-- Name -- inspire id External	Identifier	1										
		name	-- Name -- Name Name of the	GeographicalName	0..*	voidable									
		buildingNature	-- Name -- Building nature	BuildingNatureValue	0..*	voidable									
		currentUse	-- Name -- Current use Activity	CurrentUse	0..*	voidable									
		numberOfDwellings	-- Name -- Number of dwellings	Integer	0..1	voidable									
		numberOfBuildingUnits	-- Name -- Number of building units	Integer	0..1	voidable									
		numberOfFloorsAboveGround	-- Name -- Number of floors above	Integer	0..1	voidable									
		parts	The building parts composing the	BuildingPart	0..*	voidable									
		geometry2D	-- Name -- geometry 2D	BuildingGeometry2D	0..*	voidable									
		geometry3DLoD1	-- Name -- geometry 3D LoD 1	BuildingGeometry3DLoD1	0..1										
		geometry3DLoD2	-- Name -- geometry 3D LoD 2	BuildingGeometry3DLoD2	0..1										
		geometry3DLoD3	-- Name -- geometry 3D LoD 3	BuildingGeometry3DLoD	0..1										
		geometry3DLoD4	-- Name -- geometry 3D LoD 4	BuildingGeometry3DLoD	0..1										
BuildingPart Supertypes:*BuildingPartAbstractBuildingAbstractConstruction*	-- Name -- Building part A BuildingPart is a sub-division of a Building that might be considered itself as a building.NOTE 1: A building part is homogeneous related to its physical, functional and temporal aspects. EXAMPLE: A building may be composed of two building parts having different heights above ground.														
		beginLifespanV	-- Name -- Begin	DateTime	1	voidable									
		conditionOfCon	-- Name --	ConditionOfConstruc	1	voidable									
		dateOfConstruct	-- Name -- Date of	DateOfEvent	0..1	voidable									
		dateOfDemoliti	-- Name -- Date of	DateOfEvent	0..1	voidable									
		dateOfRenovati	-- Name -- Date of	DateOfEvent	0..1	voidable									
		elevation	-- Name --	Elevation	0..*	voidable									
		endLifespanVer	-- Name -- End	DateTime	0..1	voidable									
		externalRefere	-- Name -- External	ExternalReference	0..*	voidable									
		heightAboveGr	-- Name -- Height	HeightAboveGround	0..*	voidable									
		inspireId	-- Name -- inspire	Identifier	1										
		name	-- Name -- Name	GeographicalName	0..*	voidable									
		buildingNature	-- Name -- Building	BuildingNatureValue	0..*	voidable									
		currentUse	-- Name -- Current	CurrentUse	0..*	voidable									
		numberOfDwell	-- Name -- Number	Integer	0..1	voidable									
		numberOfBuildi	-- Name -- Number	Integer	0..1	voidable									
		numberOfFloors	-- Name -- Number	Integer	0..1	voidable									
		geometry2D	-- Name --	BuildingGeometry2D	0..*	voidable									
		geometry3DLoD	-- Name --	BuildingGeometry3D	0..1										
		geometry3DLoD	-- Name --	BuildingGeometry3D	0..1										
		geometry3DLoD	-- Name --	BuildingGeometry3D	0..1										
		geometry3DLoD	-- Name --	BuildingGeometry3D	0..1										
BuildingGeometry3DLoD	-- Name -- Building geometry 3D LoD Data type grouping the 3D geometry of a building or building part and the metadata information attached to this geometry.														
		geometryMultiS	-- Name --	GM_MultiSurface	0..1										
		geometrySolid	-- Name --	GM_Solid	0..1										
		terrainIntersecti	-- Name -- Terrain	GM_MultiCurve	0..1	voidable									
		verticalGeometr	-- Name -- Vertical	ElevationReferenceV	0..1	voidable									
		horizontalGeom	-- Name --	Length	0..1	voidable									
		verticalGeometr	-- Name -- Vertical	Length	0..1	voidable									
BuildingGeometry3DLoD1 Supertypes:BuildingGeometry3DLoD	-- Name -- Building geometry 3D LoD 1 Data type grouping the specific metadata attached to the 3D geometry, when provided by a LoD 1 representation.														
		geometryMultiS	-- Name --	GM_MultiSurface	0..1										
		geometrySolid	-- Name --	GM_Solid	0..1										
		terrainIntersecti	-- Name -- Terrain	GM_MultiCurve	0..1	voidable									
		verticalGeometr	-- Name -- Vertical	ElevationReferenceV	0..1	voidable									
		horizontalGeom	-- Name --	Length	0..1	voidable									
		verticalGeometr	-- Name -- Vertical	Length	0..1	voidable									
		horizontalGeom	-- Name --	HorizontalGeometryR	0..1	voidable									
		verticalGeometr	-- Name -- Vertical	ElevationReferenceV	0..1	voidable									
BuildingGeometry3DLoD2 Supertypes:BuildingGeometry3DLoD	-- Name -- Building geometry 3D LoD 2 Data type grouping the specific metadata attached to the 3D geometry, when provided by a LoD2 representation.														
		geometryMultiS	-- Name --	GM_MultiSurface	0..1										
		geometrySolid	-- Name --	GM_Solid	0..1										
		terrainIntersecti	-- Name -- Terrain	GM_MultiCurve	0..1	voidable									
		verticalGeometr	-- Name -- Vertical	ElevationReferenceV	0..1	voidable									
		horizontalGeom	-- Name --	Length	0..1	voidable									
		verticalGeometr	-- Name -- Vertical	Length	0..1	voidable									
		horizontalGeom	-- Name --	HorizontalGeometryR	0..1	voidable									

Source: own creation, JRC, 2020.

Step 3: transformation of source data according to INSPIRE target data model

Once the conceptual mapping between the source and target schema has been defined (and the possible inconsistencies have been solved during the pre-processing phase), source data can be transformed according to INSPIRE making use of a transformation tool. The step is illustrated in Figure 134. For this mapping exercise the latest version of hale studio[47] (also known as hale studio) has been used, an open source Data Transformation Software and one of the most performant and extensively used. Figure 135 illustrates how hale studio works: the user defines a collection of mapping rules – also referred to as 'alignment' - between the elements of a 'source' and a 'target' schema, then hale studio transforms input data according to the defined alignment, and finally exports transformed data using the specified format (e.g. GML).

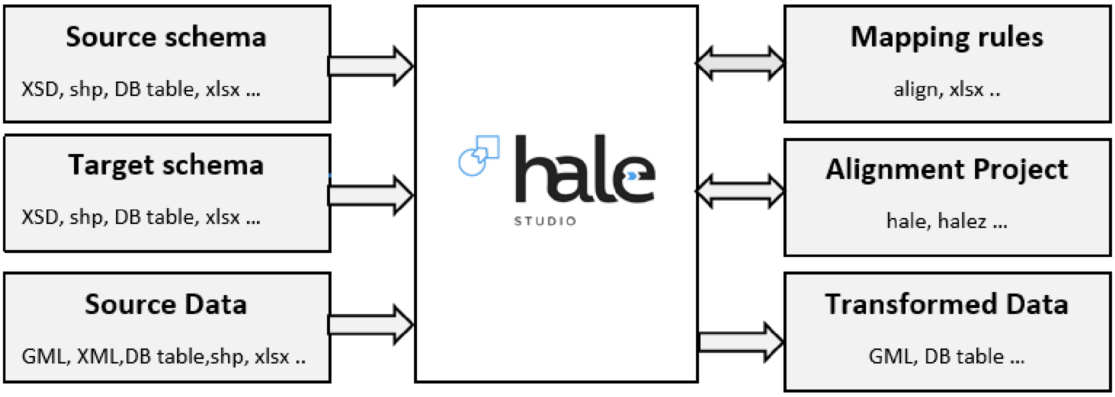
Figure 134. Data Harmonisation Step 3 - Inputs and Outputs



Source: own creation, JRC, 2020.

The compiled mapping table (output of Step 2) can be very useful to set up the hale studio alignment

Figure 135. hale studio



Source: own creation, JRC, 2020.

It is worth highlighting that *hale studio* also enables the user to perform a schema-based validation on the transformed data against the selected target schema, in particular the compliance to the schema structure (mandatory properties, restrictions on property values, etc.)

The output of Step 3 is a GML dataset file that:

- is conformant to the INSPIRE IR requirements and theme-specific requirements contained in the relevant INSPIRE Data Specification;
- includes all information present in the source data, which could be mapped onto the target data model.

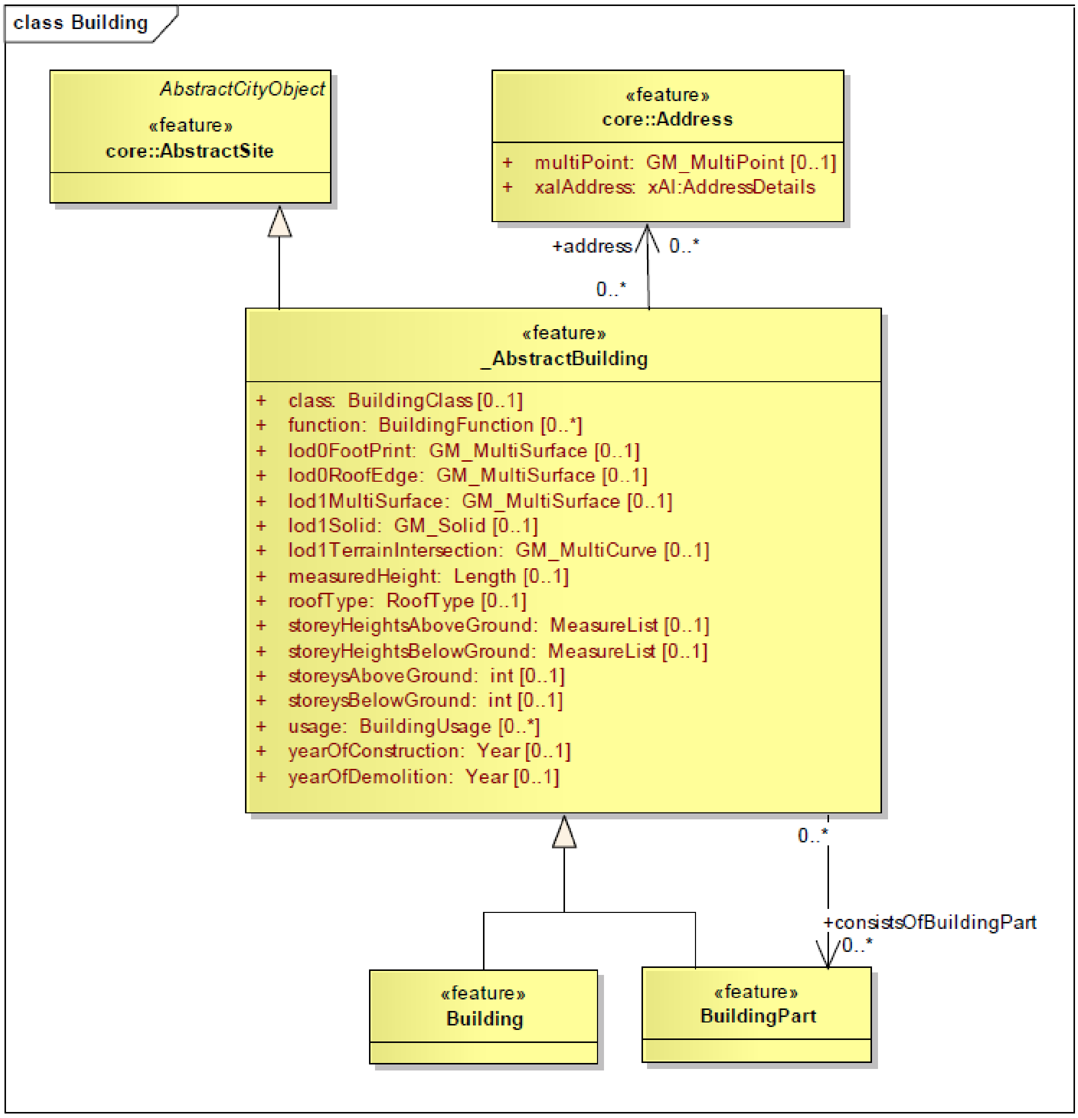
7.2.2 Source data model CityGML

CityGML standard and its level of details have been described in sub-section 2.1

Figure 136 shows how the LOD1 data model is schematized in UML. It contains two main feature types, Building and BuildingPart, which inherit an abstract feature type containing common attributes.

[47] https://www.wetransform.to/products/halestudio/

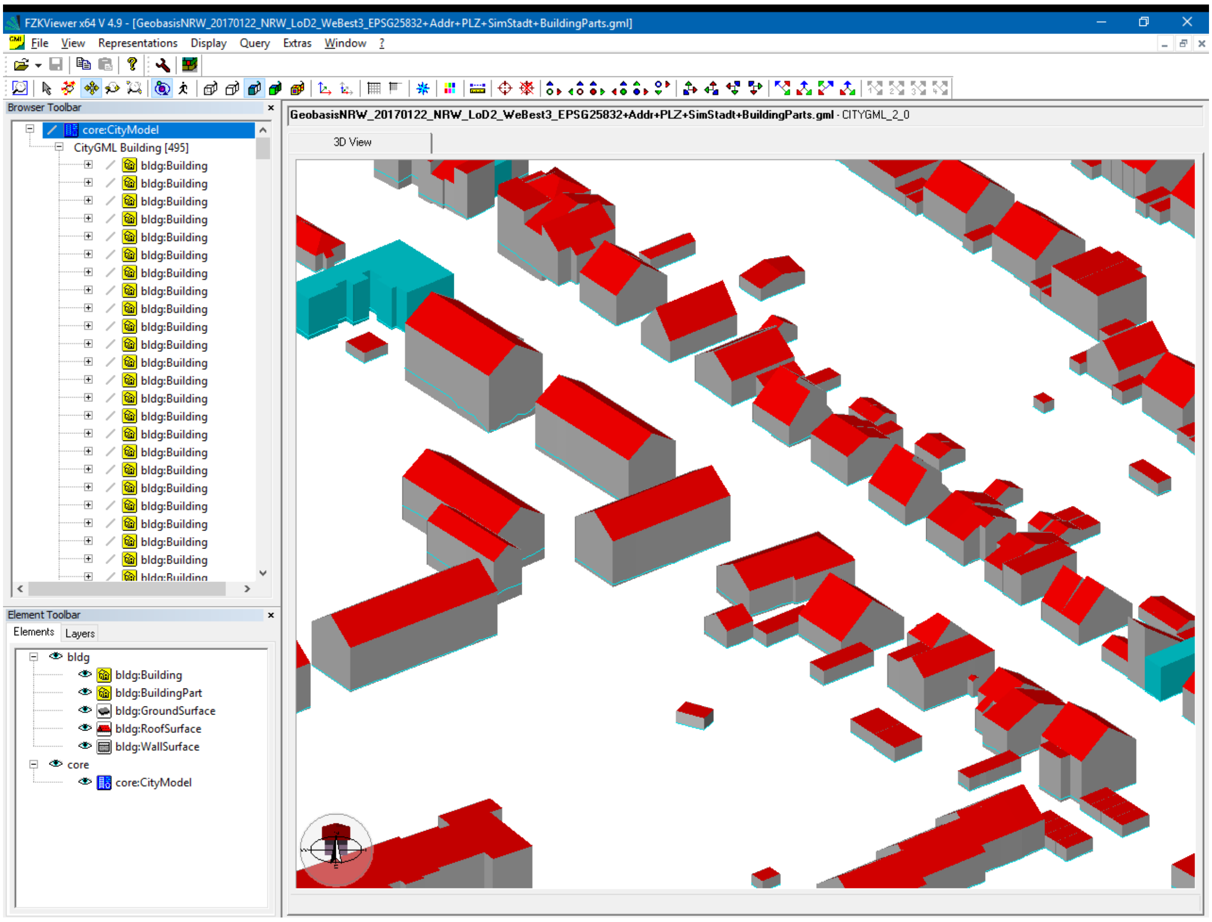
Figure 136. LOD1 UML data model



Source: own creation, JRC, 2020.

The LoD2 data model adds information about the boundary surfaces (roof, wall, ground, etc.). In the Figure 137 there is an example of how a LoD2 dataset appears in a CItyGML viewer.

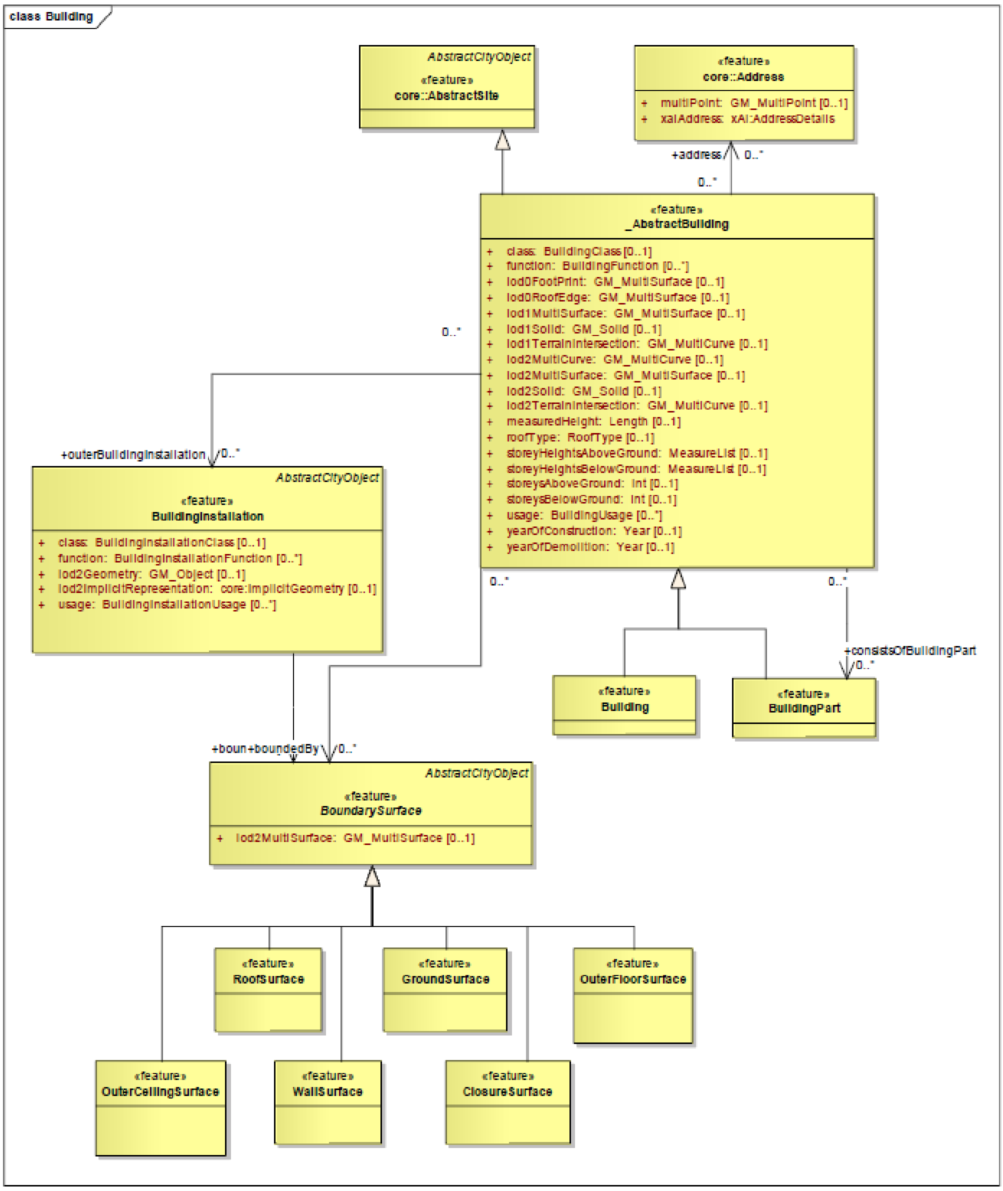
Figure 137. Example of an LoD2 dataset



Source: own creation, JRC, 2020.

Figure 138 shows how the LoD2 data model is schematized in UML. It adds to the feature types defined by LoD1 the associations with two additional feature types, defining the boundary surfaces and the installations of the building.

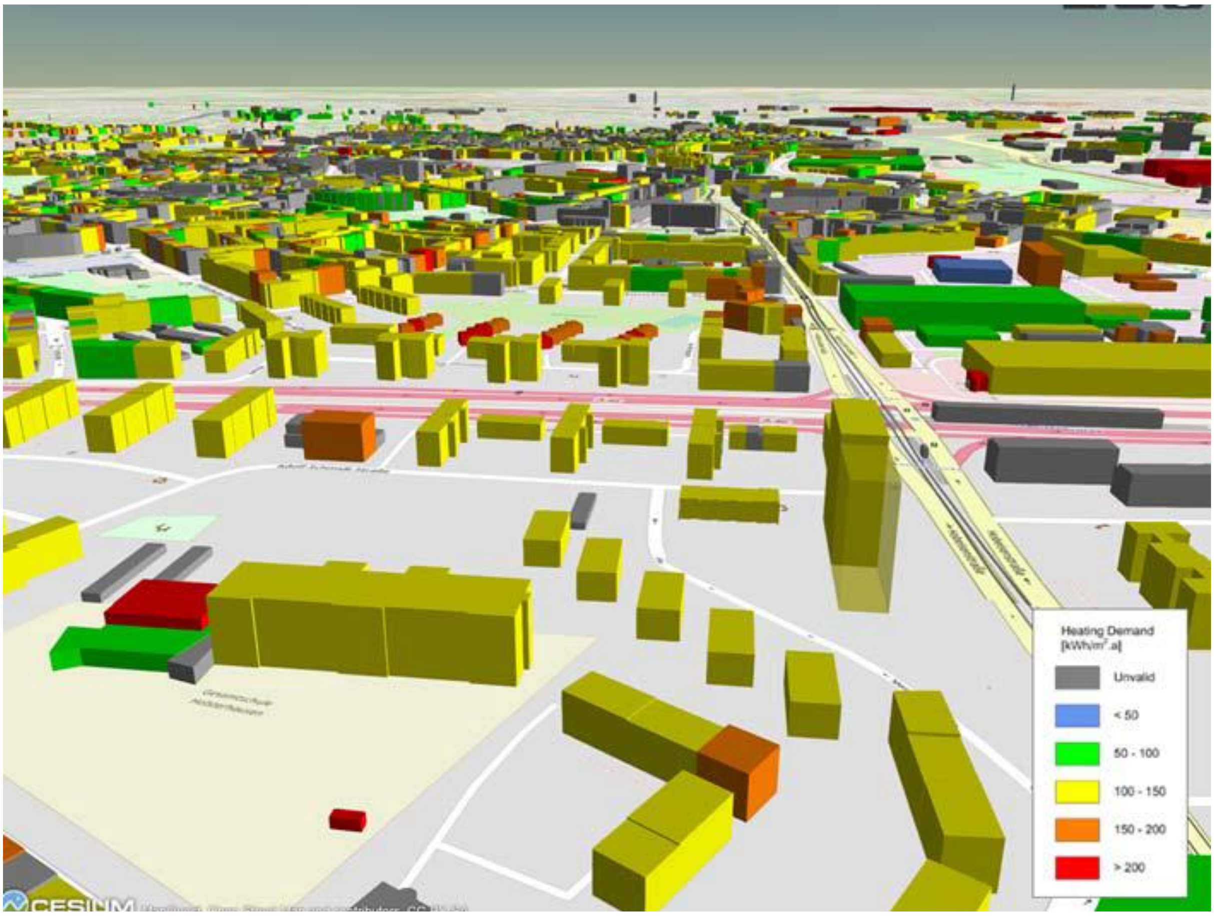
Figure 138. LoD2 UML data model



Source: own creation, JRC, 2020.

Two different 3D buildings datasets in two different test areas have been considered in this mapping exercise. Figure 139 shows the dataset related to the test area in Essen (DE) and Figure 140 the one related to test area of Zwolle (NL).

Figure 139. ESSEN test area



Source: own creation, JRC, 2020.

Figure 140. ZWOLLE test area

Source: own creation, JRC, 2020.

7.2.3 Target data model INSPIRE BU 3D

The natural candidate target data model for CityGML datasets is the INSPIRE Buildings data model, which was developed taking into consideration the CityGML standard.

The INSPIRE Data Specification on Buildings[48] provides six different profiles, covering different levels of detail from the semantic and geometric points of view (core vs. extended and 2D vs. 3D).

Two kinds of semantic profiles are present in this data specification:

— *normative core profile*, which includes both basic topographic data (such as height, number of floors, nature of buildings, date of construction, etc.) and coarse official data (such as current use, number of dwellings or of building units); it aims to fulfil most user requirements.

— *informative extended profile*, which includes more detailed information about buildings and building related objects.

The common semantics used by all profiles has been described in a *base application schema*.

Building data may be available and required either as 2D (or 2,5D) data or as 3D data. The INSPIRE data specification proposes two kinds of geometric profiles:

— *2D profile* (with 2D or 2,5D geometry)

— *3D profile* (with 3D geometry)

It's worth highlighting that two out of the six application schemas are just abstract schemas gathering the concepts that are used in common by the other instantiable schemas.

The delivery of data may be done using four options (profiles) that consist of a combination of application schemas, as explained in Figure 141 and in Figure 142.

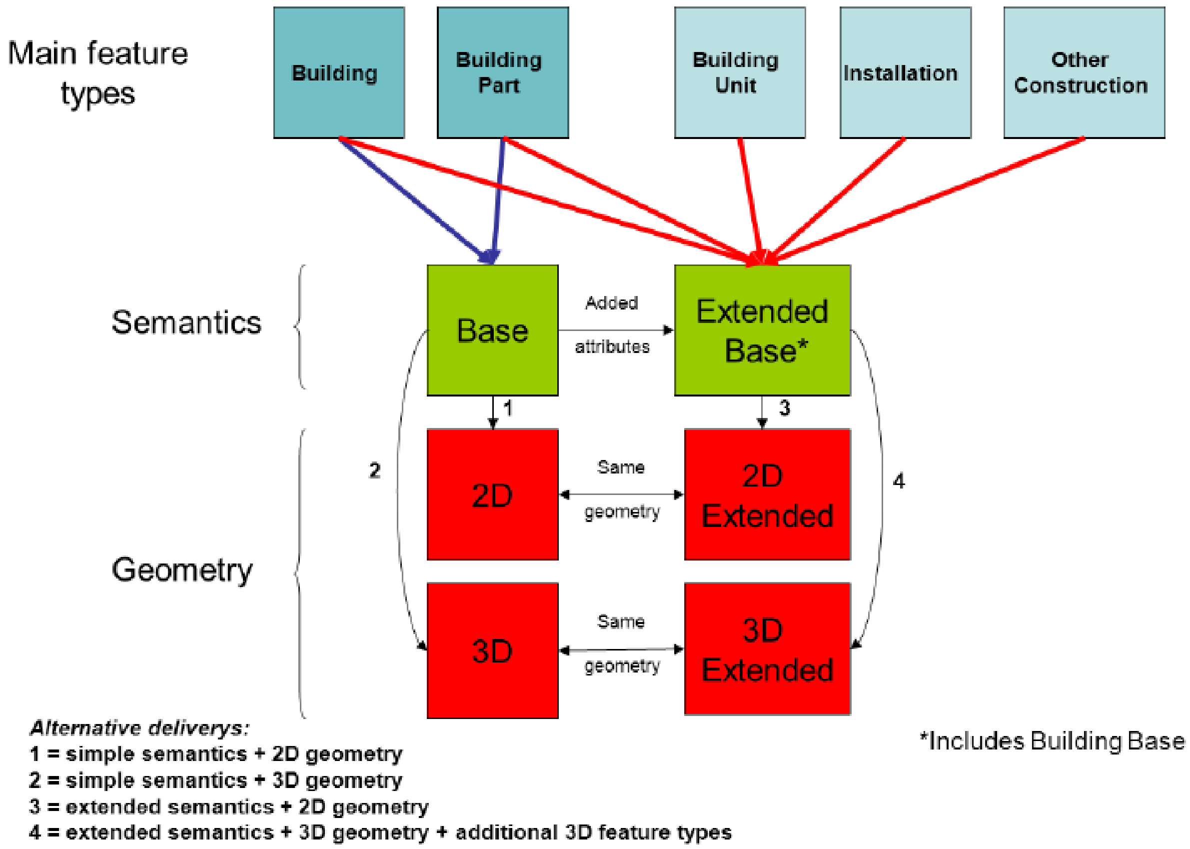
Figure 141. The profile approach for theme Buildings

	Basic semantic	Rich semantic
2D geometry	**Core 2D profile** *uses application schemas:* - *base* - *Buildings2D*	**Extended 2D profile** *uses application schemas:* *- base* *- Buildings2D* *- base extended* *- extended 2D*
3D geometry	**Core 3D profile** *uses application schemas:* - *base* - *Buildings3D*	**Extended 3D profile** *uses application schemas:* *- base* *- Buildings3D* *- base extended* *- extended 3D*

Source: D2.8.III.2 INSPIRE Data Specification on Buildings – Technical Guidelines.

[48] https://inspire.ec.europa.eu/id/document/tg/bu

Figure 142. Content and structure of application schemas for theme Buildings



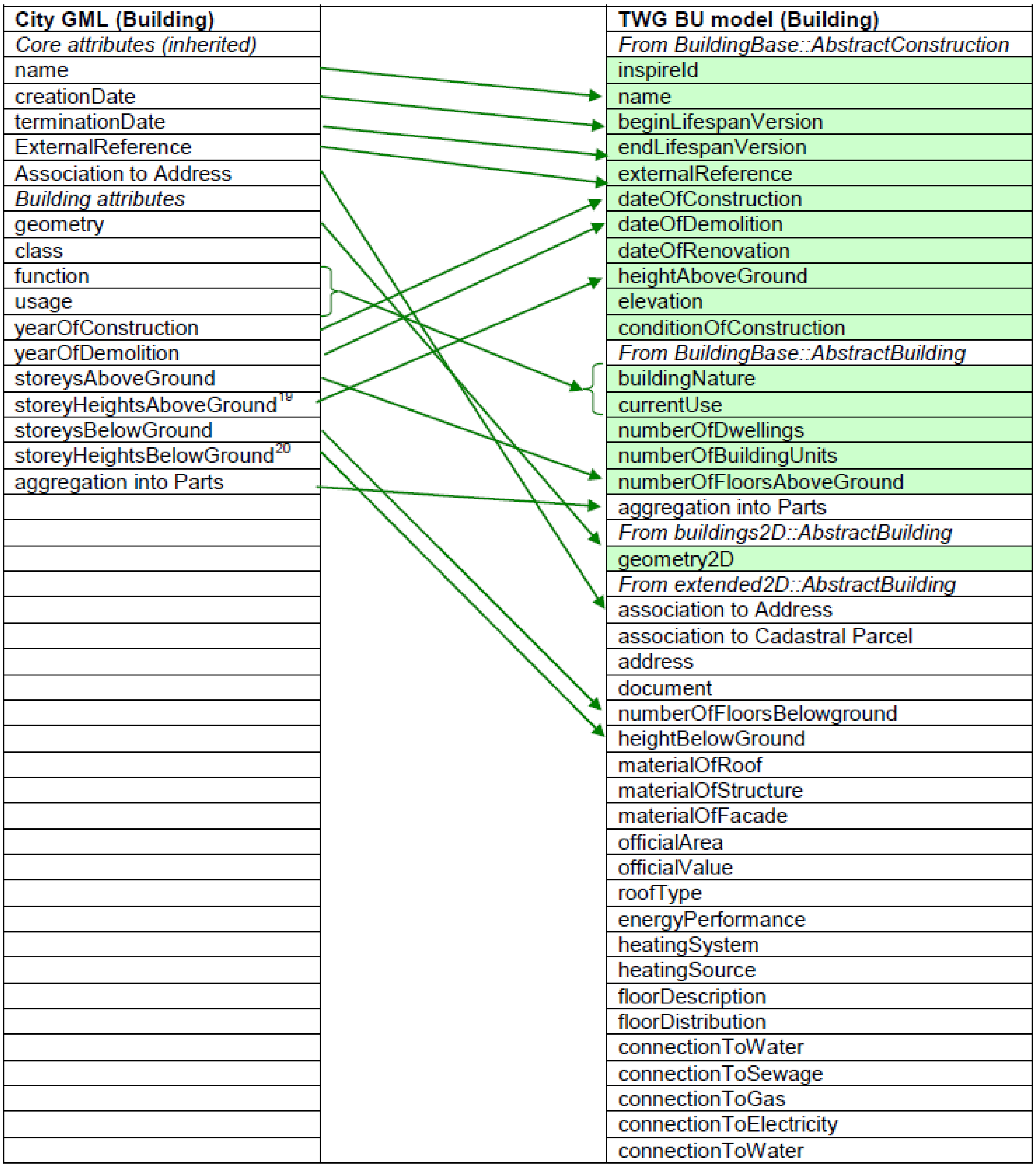
Source: D2.8.III.2 INSPIRE Data Specification on Buildings – Technical Guidelines.

In Figure 142 feature types are represented in blue, abstract application schemas in green and instantiable application schemas in red.

CityGML has strongly influenced the development of the INSPIRE BU model, both for the 2D and 3D profiles. Indeed, many use cases that were considered in the development of the INSPIRE BU data models required a three-dimensional representation of buildings and therefore the building representation in CityGML LoD1 - LoD4 has been added to the INSPIRE BU model as core 3D profile, whereas the whole content of LoD1 - LoD4 (including the features attached to buildings, such as boundaries, openings, rooms, etc.) are the basis of the INSPIRE extended 3D profile.

Figure 143 shows the mapping from CityGML to INSPIRE for the Building feature type.

Figure 143. Correspondence between the two data models



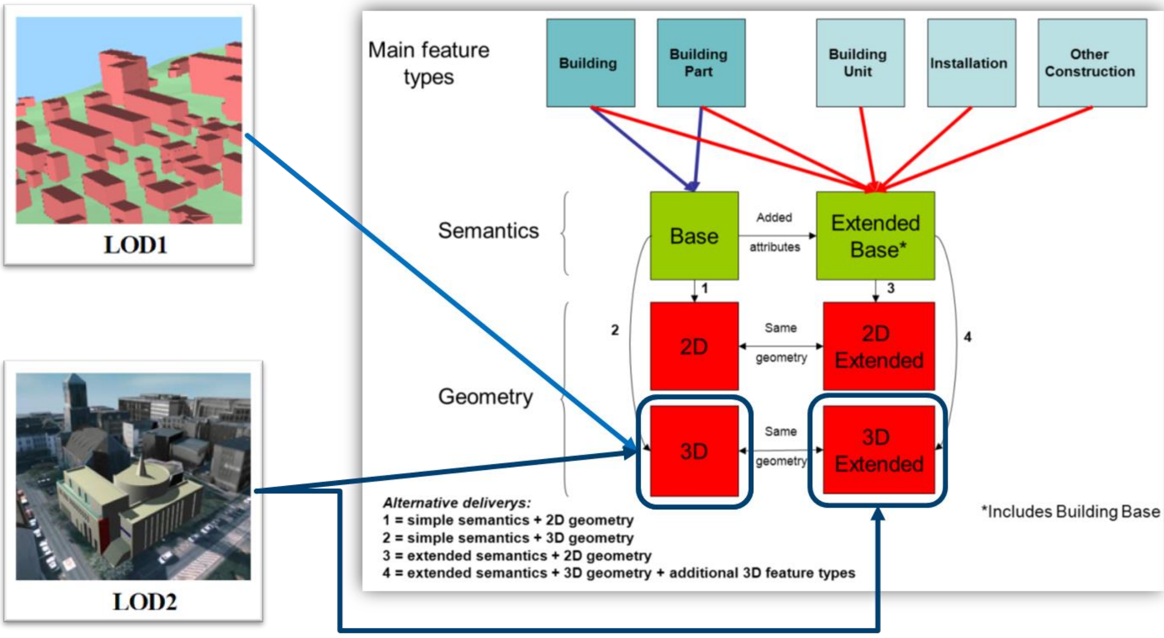
Source: D2.8.III.2 INSPIRE Data Specification on Buildings – Technical Guidelines.

7.2.4 Mapping CityGML to INSPIRE BU 3D

With reference to the scenario 1 described in section 7.2, three different data transformation exercises, schematically shown in Figure 144, have been performed:

- CityGML LoD1 dataset transformed to INSPIRE BU 3D CORE data model
- CityGML LoD2 dataset transformed to INSPIRE BU 3D CORE data model
- CityGML LoD2 dataset transformed to INSPIRE BU 3D EXTENDED data model.

Figure 144. Relation between CityGML and INSPIRE BU 3D data models

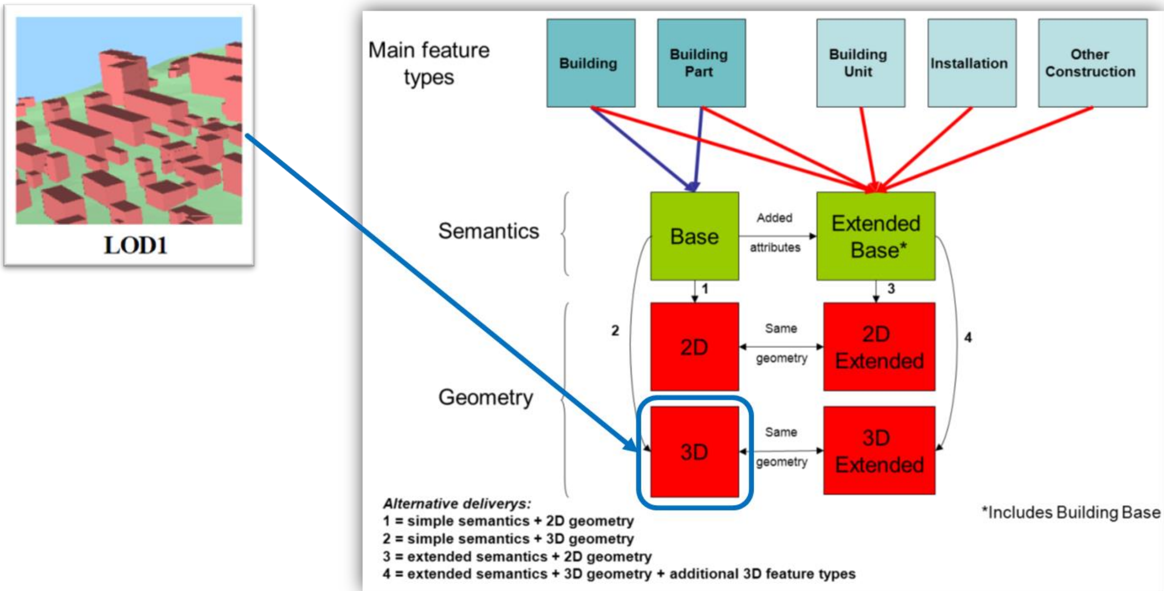


Source: own creation, JRC, 2020.

7.2.4.1 CityGML LoD1 to Buildings - Core 3D

In this section the first data transformation exercise, related to the CityGML LoD1 dataset harmonised according to the INSPIRE Buildings - Core 3D data model, schematically shown in Figure 145, is described.

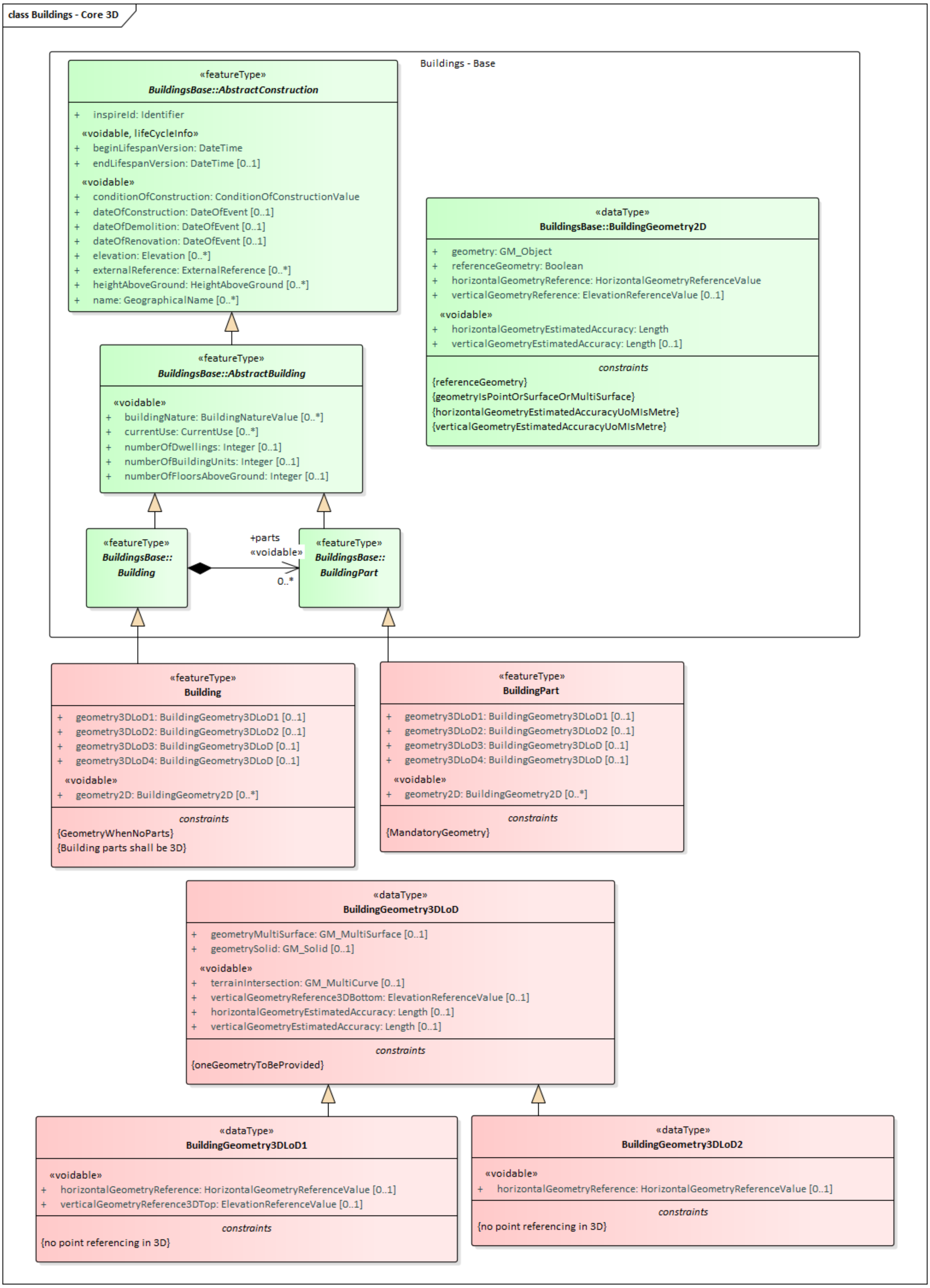
Figure 145. Relation between data models for the first mapping exercise



Source: own creation, JRC, 2020.

As shown in the UML diagram in Figure 146, the Buildings3D application schema describes the 3D geometric representation of the spatial object types defined in the Buildings Base application schema, namely buildings and building parts, inheriting the common semantics of Buildings base.

Figure 146. Buildings - Core 3D UML diagram



Source: own creation, JRC, 2020.

After the analysis of the source and target data model, a matching exercise has been performed in order to derive the correspondences between elements of the source and target schemas. The transformation rules identified have been applied using hale studio as transformation tool to physically transform the dataset and they have been documented using a matching table.

Due to the similarities between the two data models, a simplified version of the matching table, shown in Figure 147 has been used for the mapping.

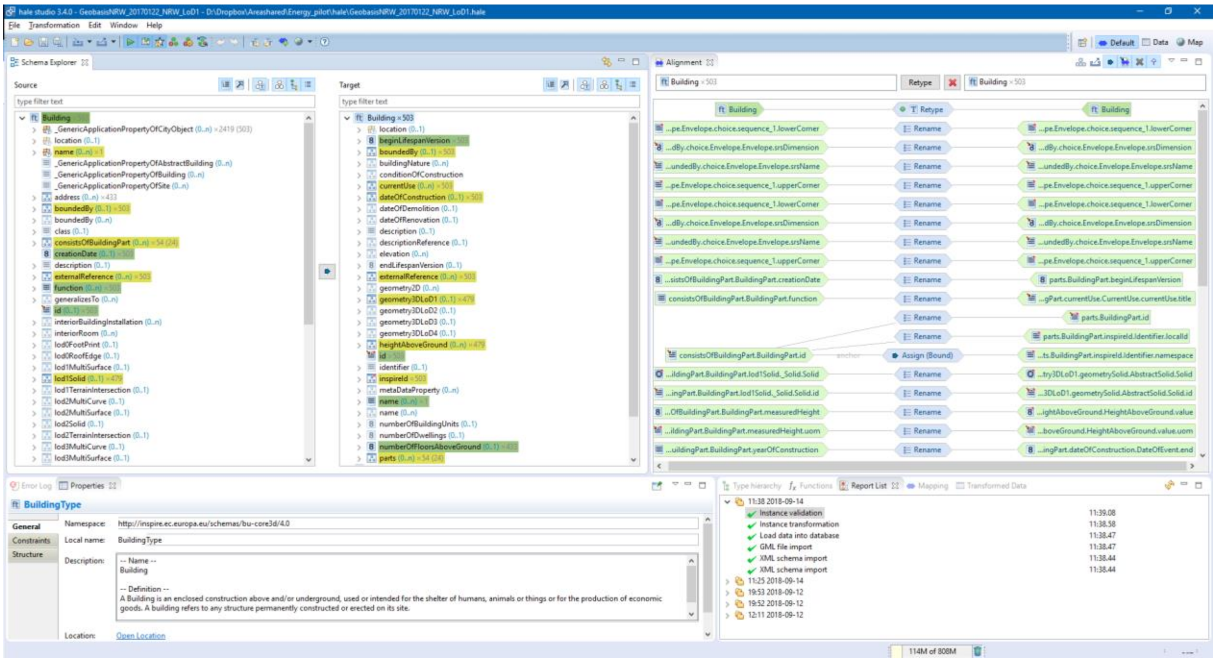
Figure 147. Matching table between LoD1 and INSPIRE BU3d Core

Attribute of Source Data	Attribute of Target Data	Function
Building	Building	Retype
Address		
~~address...Country...Locality.LocalityName~~	~~location.location.choice.LocationString~~	~~Rename~~
~~address...Country...Locality...PostalCodeNumber~~	~~location.location.choice.LocationKeyWord~~	~~Rename~~
~~address...Country...Locality...ThoroughfareName~~	~~location.priorityLocation.choice.LocationString~~	~~Rename~~
~~address...Country...Locality...ThoroughfareNumber~~	~~location.priorityLocation.choice.LocationKeyWord~~	~~Rename~~
~~address...Country...Locality.Thoroughfare.Type~~	~~location.priorityLocation.type~~	~~Rename~~
~~address...Country...Locality.Type~~	~~location.location.type~~	~~Rename~~
boundedBy(0...1)		
boundedBy...Envelope...lowerCorner	boundedBy...Envelope...lowerCorner	Rename
boundedBy...Envelope.srsDimension	boundedBy...Envelope.srsDimension	Rename
boundedBy...Envelope.srsName	boundedBy...Envelope.srsName	Rename
boundedBy...Envelope...upperCorner	boundedBy...Envelope...upperCorner	Rename
consistsOfBuildingPart		
~~consistsOfBuildingPart.BuildingPart~~	~~parts.BuildingPart~~	~~Reproject Geometry~~
other attributes are mapped same as attributes under Building		
creationDate		
creationDate	beginLifespanVersion~~.nilReason~~	Rename
externalReference		
externalReference.externalObject.choice.name	externalReference.ExternalReference.reference	Rename
externalReference.informationSystem	externalReference.ExternalReference.informationSystem	Rename
function		
function	currentUse.CurrentUse.currentUse.title	Rename
id		
id	id	Rename
id	inspireId.identifier.localId	Rename
	inspireId.identifier.namespace	Assign
lod1Solid		
~~lod1Solid._Solid.Solid~~	~~geometry3DLoD1.BuildingGeometry3DLoD1.geometrySolid.AbstractSolid.Solid~~	~~Reproject Geometry~~
~~lod1Solid._Solid.Solid.exterior._Surface.CompositeSurface~~	~~geometry3DLoD1...Solid...surfaceMember...CompositeSurface~~	~~Reproject Geometry~~
lod1Solid._Solid.Solid.exterior._Surface.CompositeSurface.id	geometry3DLoD1...Solid...surfaceMember...CompositeSurface.id	Rename
lod1Solid._Solid.Solid.exterior._Surface.CompositeSurface.surfaceMember	geometry3DLoD1...Solid...surfaceMember...CompositeSurface.surfaceMember	Reproject Geometry
lod1Solid...CompositeSurface.surfaceMember._Surface.Polygon	geometry3DLoD1...Solid...surfaceMember...CompositeSurface.surfaceMember.AbstractSurface.Polygon	Reproject Geometry
~~lod1Solid...CompositeSurface...Polygon...LinearRing~~	~~geometry3DLoD1...Solid...surfaceMember...CompositeSurface...Polygon...LinearRing~~	~~Reproject Geometry~~
lod1Solid...CompopsiteSurface...Polygon...LinearRing...posList	lod1Geometry...Solid...CompositeSurface...Polygon...LinearRing...posList	Rename
lod1Solid...CompopsiteSurface...Polygon...LinearRing...posList.srsDimension	lod1Geometry...Solid...CompositeSurface...Polygon...LinearRing...posList.srsDimension	Rename
lod1Solid...CompopsiteSurface...Polygon.id	lod1Geometry...Solid...CompositeSurface...Polygon.id	Rename
lod1Solid._Solid.Solid.id	geometry3DLoD1.BuildingGeometry3DLoD1.geometrySolid.AbstractSolid.Solid.id	Rename
measuredHeight		
measuredHeight	heightAboveGround.HeightAboveGround.value	Rename
measuredHeight.uom	heightAboveGround.HeightAboveGround.value.uom	Rename
name		
name.name	name	Rename
storeysAboveGround		
storeysAboveGround	numberOfFloorsAboveGround	Rename
yearOfConstruction		
yearOfConstruction	dateOfConstruction.DateOfEvent.~~end.nilReason~~	Rename

Source: own creation, JRC, 2020.

After having set the transformation rules between source and target data model, as shown in the screenshot of the hale studio project in Figure 148, the dataset harmonised according to the target data model has been exported.

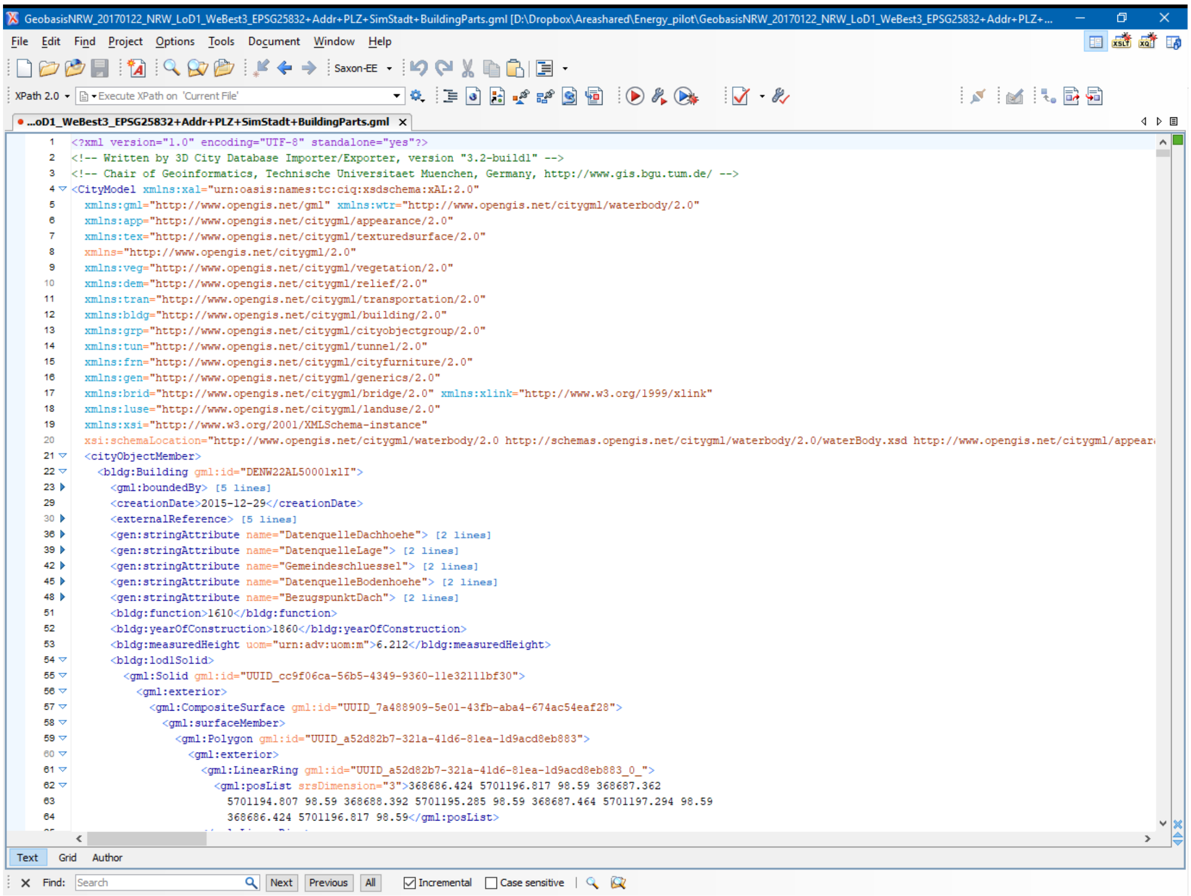
Figure 148. Hale studio project



Source: own creation, JRC, 2020.

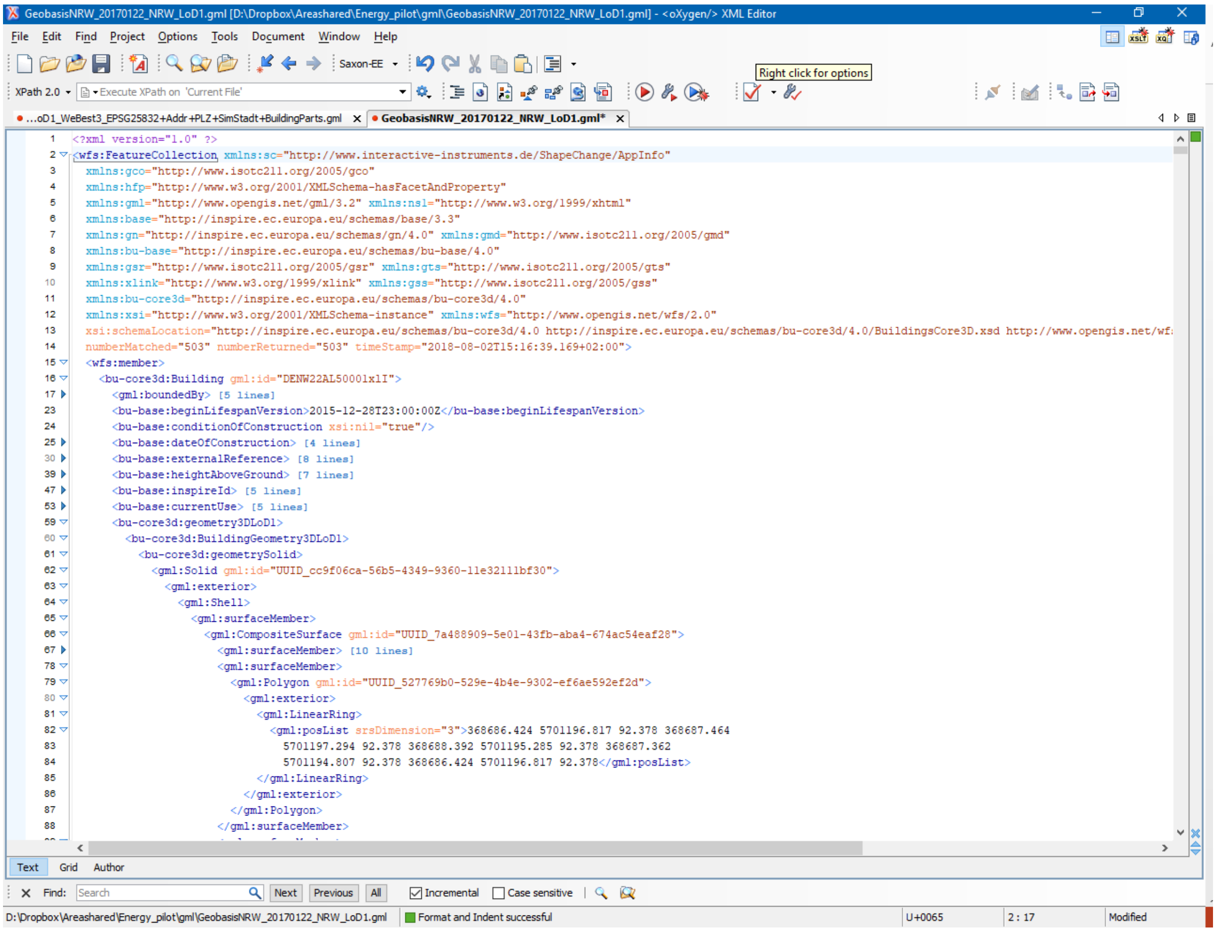
The source CityGML dataset and the harmonized GML datasets in an XML viewer are shown in Figure 149 and Figure 150, respectively.

Figure 149. Source dataset – CityGML LoD1



Source: own creation, JRC, 2020.

Figure 150. Harmonised dataset - INSPIRE BU Core-3D

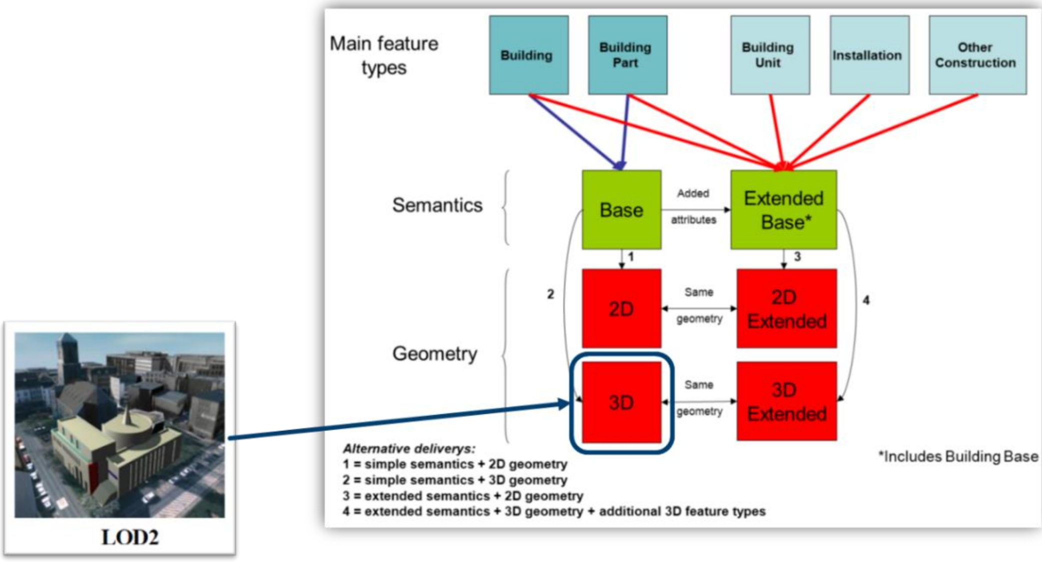


Source: own creation, JRC, 2020.

7.2.4.2 CityGML LoD2 to Buildings - Core 3D

In this section the second data transformation exercise, related to the CityGML LoD2 dataset harmonised according to the Buildings - Core 3D data model, schematically shown in the Figure 151, is described.

Figure 151. Relation between data models for the second mapping exercise



Source: own creation, JRC, 2020.

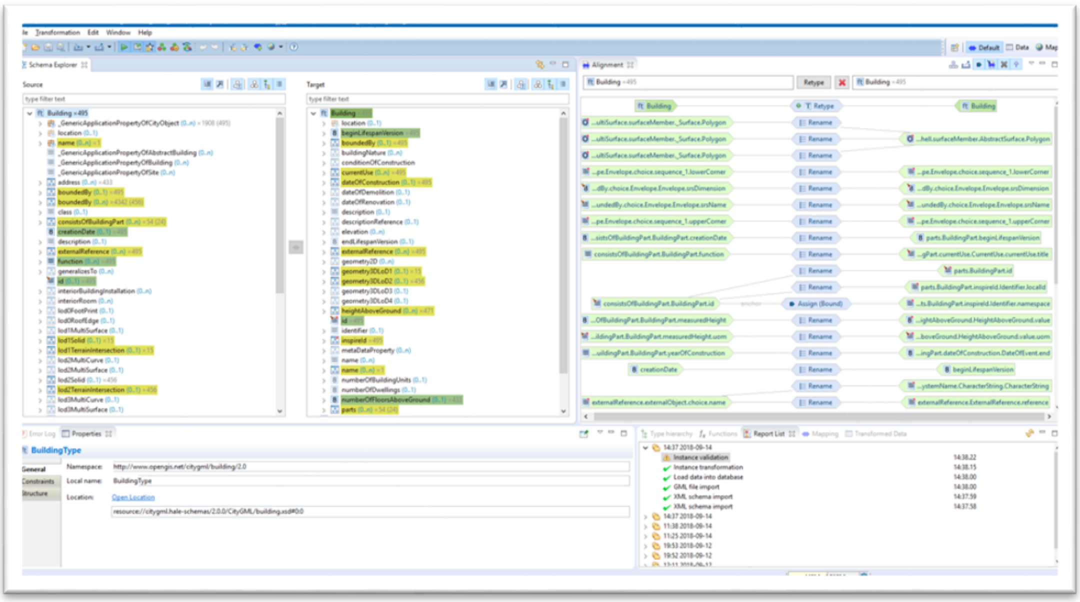
The matching table and the hale project are shown in Figure 152 and Figure 153, respectively.

Figure 152. Matching table between LoD2 and INSPIRFigure 153E BU3D Core

Attribute of Source Data	Attribute of Target Data	Function
Building	Building	Retype
boundedBy(0…1)		
boundedBy…Envelope…lowerCorner	boundedBy…Envelope…lowerCorner	Rename
boundedBy…Envelope.srsDimension	boundedBy…Envelope.srsDimension	Rename
boundedBy…Envelope.srsName	boundedBy…Envelope.srsName	Rename
boundedBy…Envelope…upperCorner	boundedBy…Envelope…upperCorner	Rename
consistsOfBuildingPart		
~~consistsOfBuildingPart.BuildingPart~~	~~parts.BuildingPart~~	~~Reproject Geometry~~
other attributes are mapped same as attributes under Building		
creationDate		
creationDate	beginLifespanVersion.~~nilReason~~	Rename
externalReference		
externalReference.externalObject.choice.name	externalReference.ExternalReference.reference	Rename
externalReference.informationSystem	externalReference.ExternalReference.informationSystem	Rename
function		
function	currentUse.CurrentUse.currentUse.title	Rename
id		
id	id	Rename
id	inspireId.identifier.localId	Rename
	inspireId.identifier.namespace	Assign
lod1Solid		
~~lod1Solid._Solid.Solid~~	~~geometry3DLoD1.BuildingGeometry3DLoD1.geometrySolid.AbstractSolid.Solid~~	~~Reproject Geometry~~
~~lod1Solid._Solid.Solid.exterior._Surface.CompositeSurface~~	~~geometry3DLoD1…Solid…surfaceMember…CompositeSurface~~	~~Reproject Geometry~~
lod1Solid._Solid.Solid.exterior._Surface.CompositeSurface.id	geometry3DLoD1…Solid…surfaceMember…CompositeSurface.id	Rename
~~lod1Solid._Solid.Solid.exterior._Surface.CompositeSurface.surfaceMember~~	~~geometry3DLoD1…Solid…surfaceMember…CompositeSurface.surfaceMember~~	~~Reproject Geometry~~
~~lod1Solid…CompositeSurface.surfaceMember._Surface.Polygon~~	~~geometry3DLoD1…Solid…surfaceMember…CompositeSurface.surfaceMember.AbstractSurface.Polygon~~	~~Reproject Geometry~~
~~lod1Solid…CompositeSurface…Polygon…LinearRing~~	~~geometry3DLoD1…Solid…surfaceMember…CompositeSurface…Polygon…LinearRing~~	~~Reproject Geometry~~
lod1Solid…CompopsiteSurface…Polygon…LinearRing…posList	lod1Geometry…Solid...CompositeSurface…Polygon…LinearRing…posList	Rename
lod1Solid…CompopsiteSurface…Polygon…LinearRing…posList.srsDimension	lod1Geometry…Solid...CompositeSurface…Polygon…LinearRing…posList.srsDimension	Rename
lod1Solid…CompopsiteSurface…Polygon.id	lod1Geometry…Solid...CompositeSurface…Polygon.id	Rename
lod1Solid._Solid.Solid.id	geometry3DLoD1.BuildingGeometry3DLoD1.geometrySolid.AbstractSolid.Solid.id	Rename
lod1TerrainIntersection		
lod1TerrainIntersection.MultiCurve	geometry3DLoD1.BuildingGeometry3DLoD1.terrainIntersection.MultiCurve	Rename
boundedBy(0…n)		
Ground Surface		
boundedBy…GroundSurface.lod2MultiSurface.MultiSurface…Polygon	geometry3DLoD2.BuildingGeometry3DLoD2.geometrySolid...Solid.exterior.Shell.surfaceMember…Polygon	Rename
Roof Surface		
boundedBy…RoofSurface.lod2MultiSurface.MultiSurface…Polygon	geometry3DLoD2.BuildingGeometry3DLoD2.geometrySolid...Solid.exterior.Shell.surfaceMember…Polygon	Rename
Wall Surface		
boundedBy…WallSurface.lod2MultiSurface.MultiSurface…Polygon	geometry3DLoD2.BuildingGeometry3DLoD2.geometrySolid...Solid.exterior.Shell.surfaceMember…Polygon	Rename
lod2TerrainIntersection		
lod2TerrainIntersection.MultiCurve	geometry3DLoD2.BuildingGeometry3DLoD2.terrainIntersection.MultiCurve	Rename
measuredHeight		
measuredHeight	heightAboveGround.HeightAboveGround.value	Rename
measuredHeight.uom	heightAboveGround.HeightAboveGround.value.uom	Rename
name		
name.name	name	Rename
storeysAboveGround		
storeysAboveGround	numberOfFloorsAboveGround	Rename
yearOfConstruction		
yearOfConstruction	dateOfConstruction.DateOfEvent.~~end.nilReason~~	Rename

Source: own creation, JRC, 2020.

Figure 153. Hale studio project

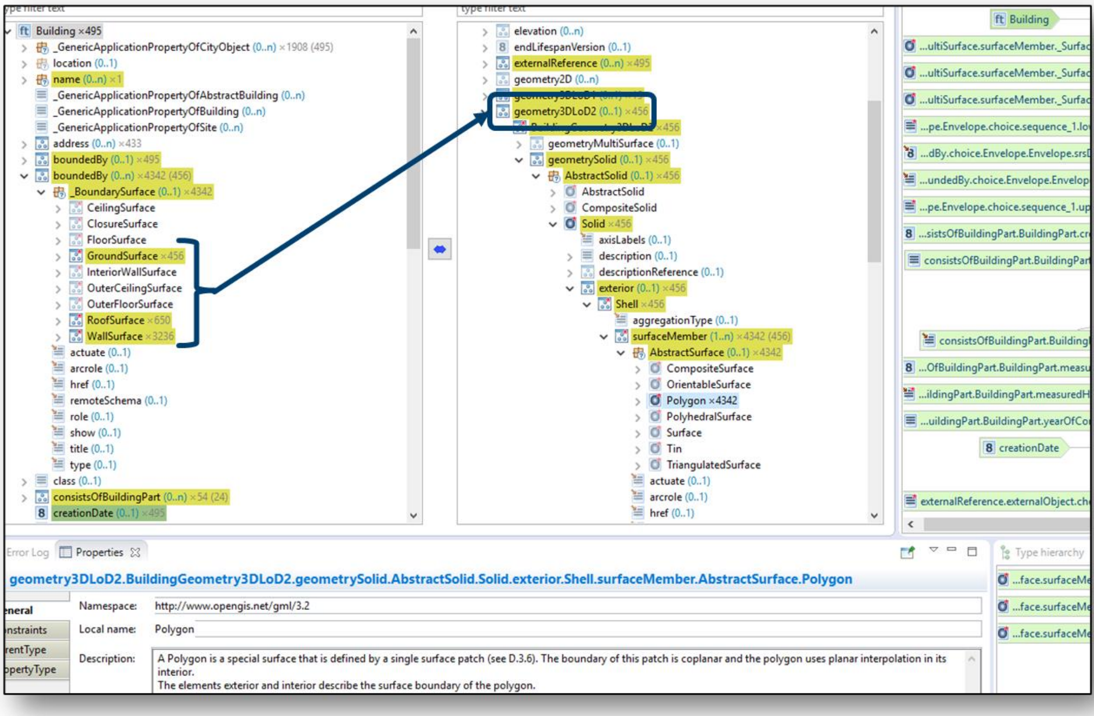


Source: own creation, JRC, 2020.

To implement topology, CityGML uses the XML concept of XLinks provided by the GML. Each geometry object to be shared by different geometric aggregates or different thematic features is assigned a unique identifier, which may be referenced by a GML geometry property using an *href* attribute. CityGML does not deploy the built-in topology package of GML3, which provides separate topology objects accompanying the geometry. This kind of topology is very complex. Nevertheless, it lacks flexibility when datasets, which might include or neglect topology, should be covered by the same data model. Conversely, the *XLink* topology is simple and flexible and nearly as powerful as the explicit GML3 topology model.

Figure 154 shows in detail how the LoD2 geometries have been mapped to the related INSPIRE geometry attribute.

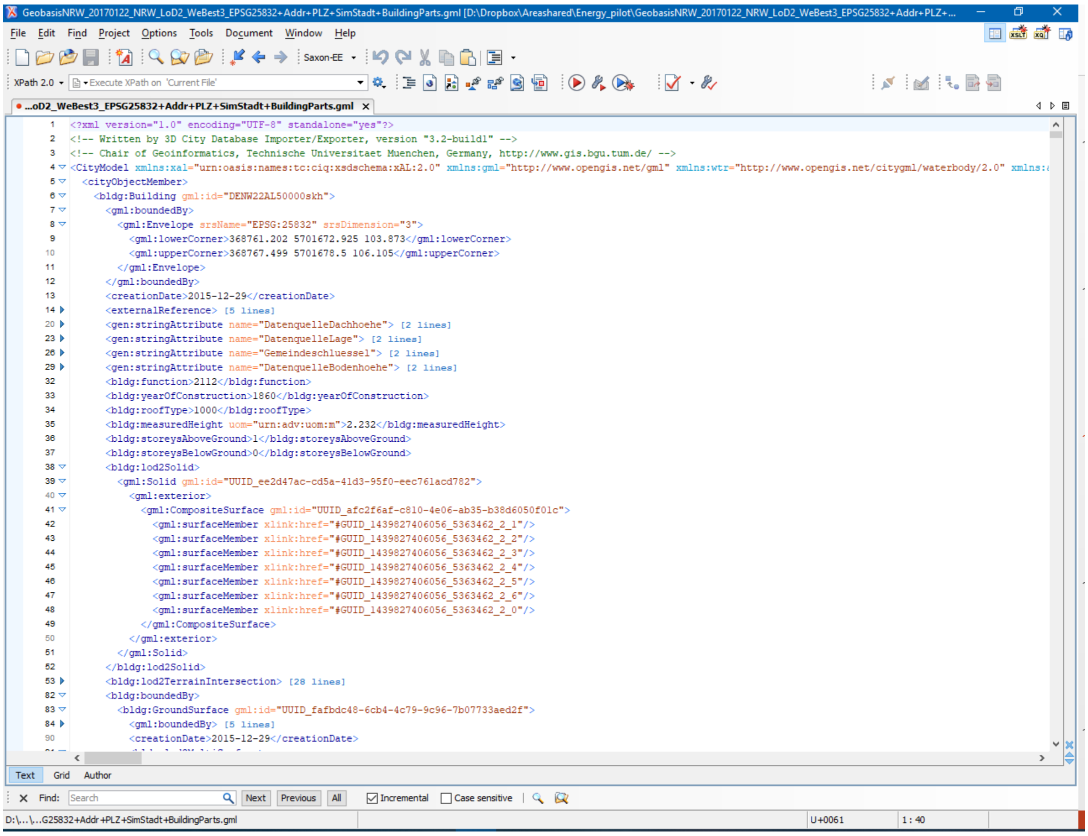
Figure 154. Hale studio project: details about the geometry mapping



Source: own creation, JRC, 2020.

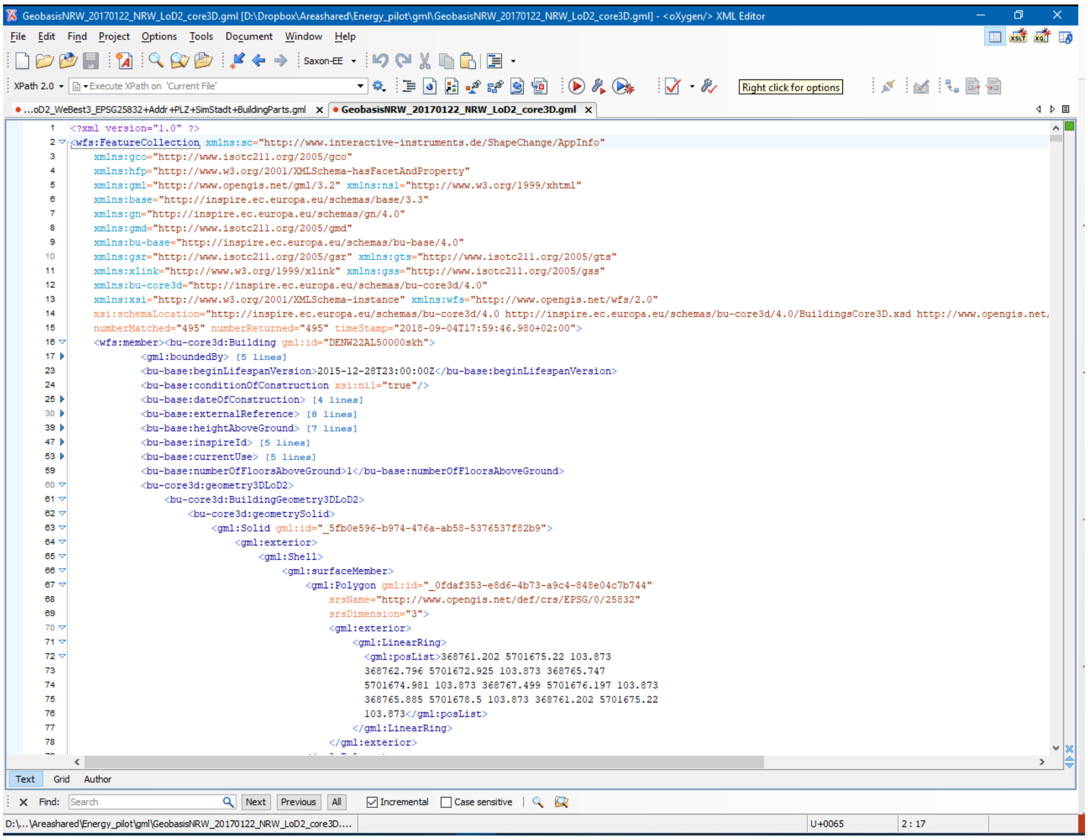
The source CityGML dataset and the harmonized GML dataset in an XML viewer are shown in Figure 155 and in the Figure 156, respectively.

Figure 155. Source dataset – CityGML LoD2



Source: own creation, JRC, 2020.

Figure 156. Target dataset - INSPIRE BU Core-3D

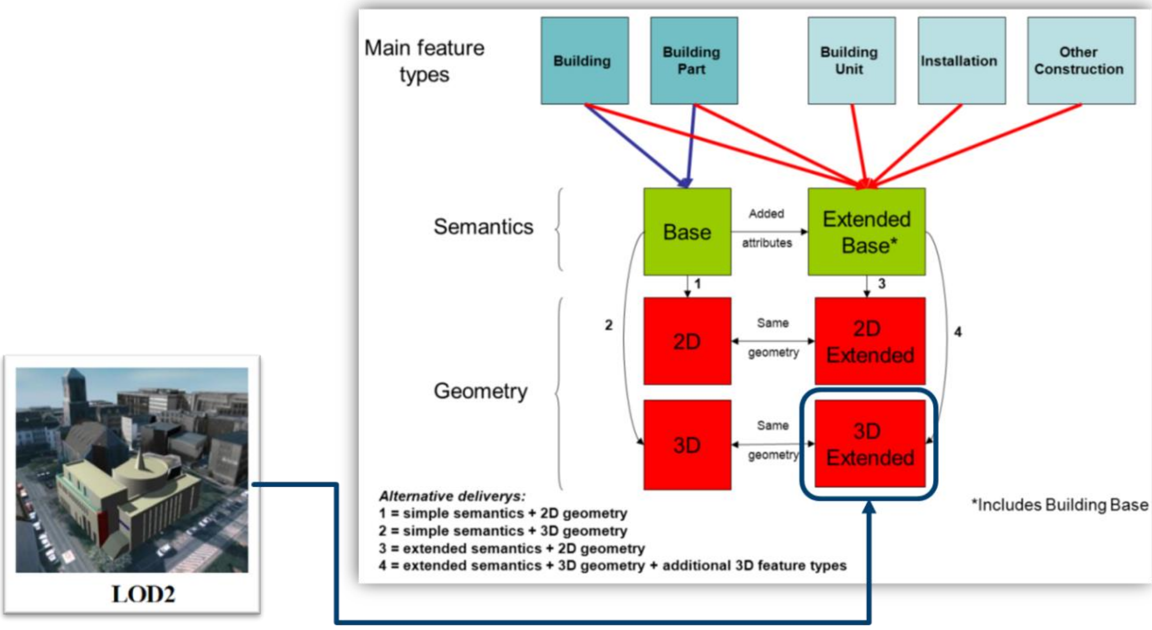


Source: own creation, JRC, 2020.

7.2.4.3 CityGML LoD2 to Buildings - Extended 3D

In this section the third data transformation exercise, related to the CityGML LoD2 dataset harmonised according to the Buildings - 3D Extended data model, schematically shown in Figure 157, is described.

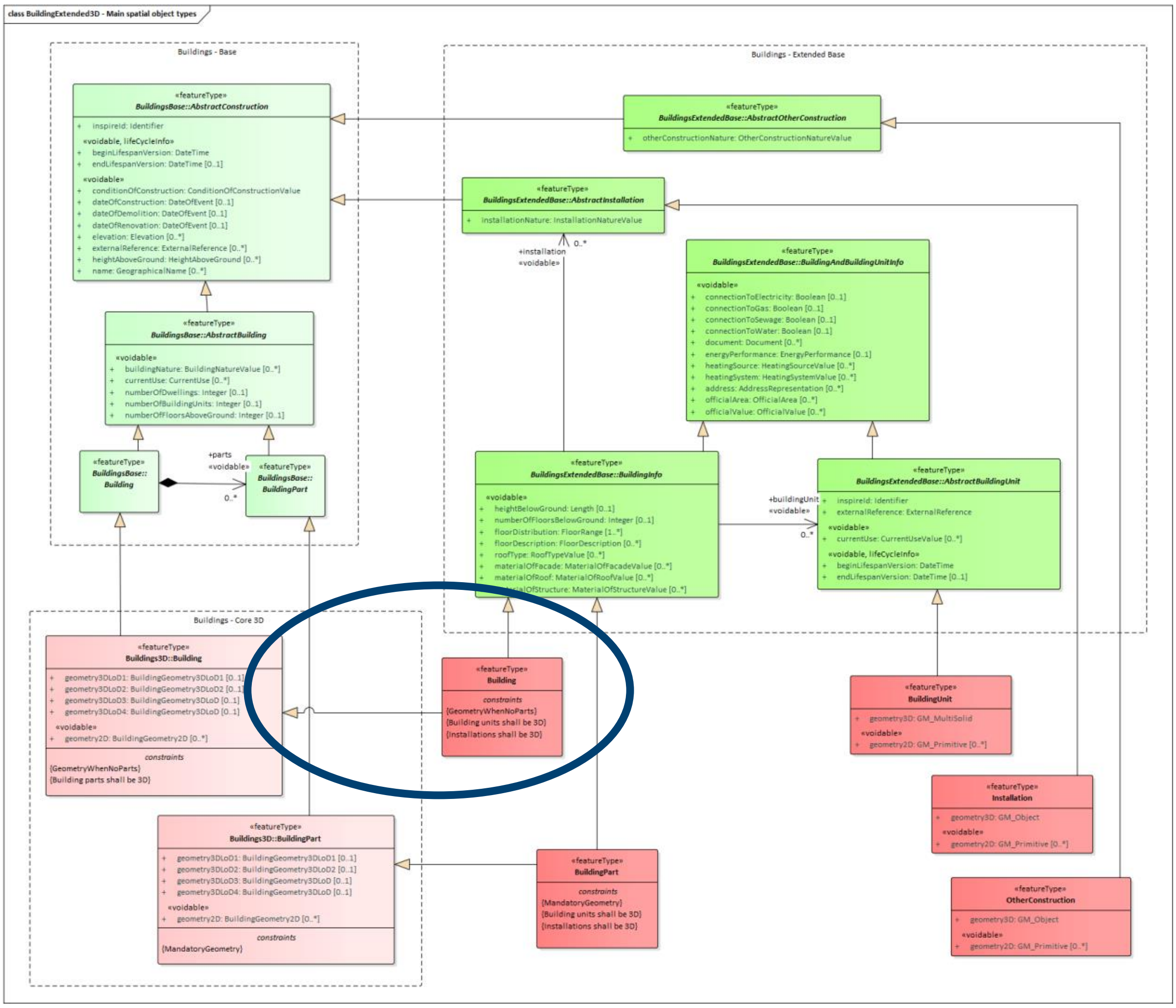
Figure 157. Relation between data models for the third mapping exercise



Source: own creation, JRC, 2020.

It is important to note that all the INSPIRE extended schemas, which are not legally binding, are still in a draft form and, in addition, not always maintained, e.g. in terms of encoding issues. The Buildings Extended 3D application schema contains a double inheritance for the Building and BuildingPart feature types, blue-circled in Figure 158, which creates problems when the physical application schema has to be generated from the logical UML data model, because these two double generalizations are not properly encoded in the relevant application schema.

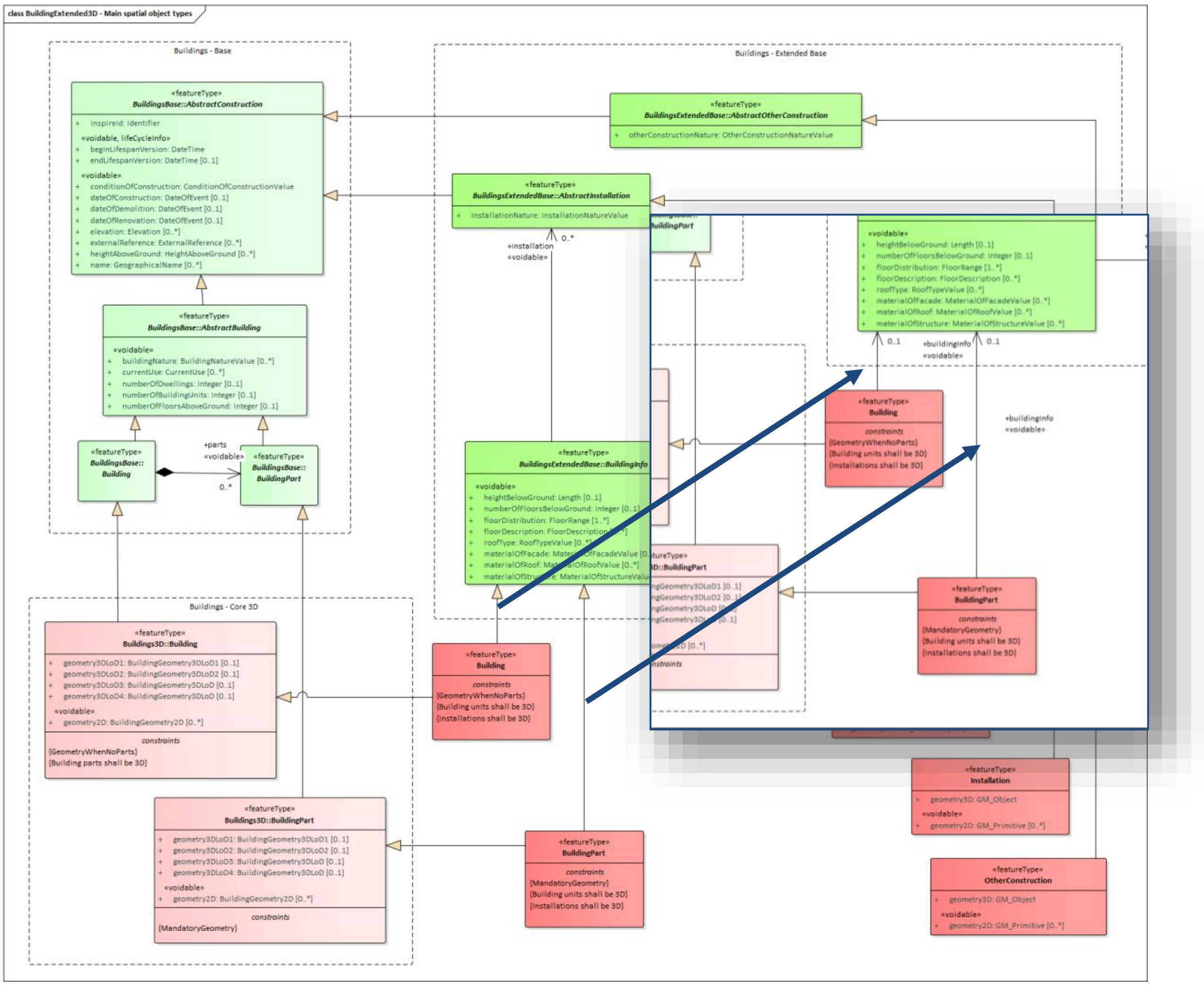
Figure 158. Buildings - Extended 3D UML diagram



Source: own creation, JRC, 2020.

The double inheritance of the *Building* and *BuildingPart* feature types of the INSPIRE Extended3D schema has been therefore modified, maintaining the inheritance of the *Building* and *BuildingPart* feature types of the INSPIRE 3D Core schema and creating an association (with multiplicity 0..1) with the feature type *BuildingInfo*, as shown in Figure 159. This solution, which avoids the double inheritance and substitutes one inheritance with one association, was adopted in order to overcome encoding problems in the generation of the related XSD (GML application schema).

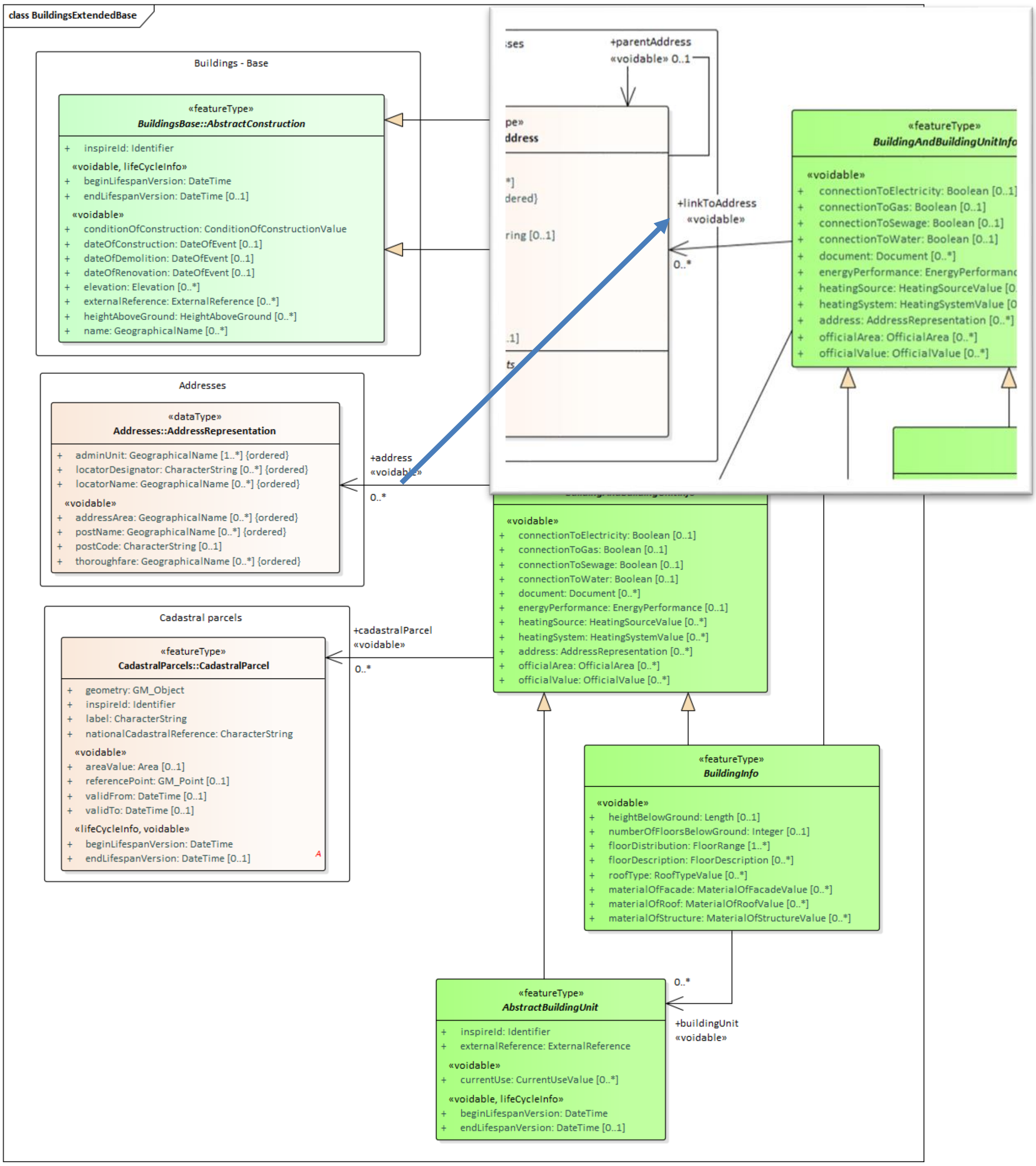
Figure 159. Buildings - Extended 3D modified UML diagram



Source: own creation, JRC, 2020.

Another necessary modification has been introduced in the Buildings - Extended Base schema: the name of the association "address" has been changed to "linkToAddress", because the value “address” was already used for another attribute, and this generated some issues during the creation of the related xsd.

Figure 160. Buildings - Extended base modified UML diagram



Source: own creation, JRC, 2020.

After having introduced the abovementioned modifications, new versions of the Buildings Extended Base and Extended 3D application schemas, shown in Figure 161 and Figure 162, respectively, have been generated and the application schemas have been published in an online repository in order to be easily used by the transformation tool.

Figure 161. Buildings - Extended base modified application schema


```
https://www.epsilon-italia.it/public/EnergyPilot/schemas/uc2/2.0/BuildingsExtendedBase.xsd

<?xml version="1.0" encoding="UTF-8"?>
<!--Generated by Enterprise Architect 12.0.1210 ( Build: 1210 )-->
- <schema xmlns="http://www.w3.org/2001/XMLSchema" xmlns:base2="http://inspire.ec.europa.eu/schemas/base2/2.0"
  xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:cp="http://inspire.ec.europa.eu/schemas/cp/4.0"
  xmlns:ad="http://inspire.ec.europa.eu/schemas/ad/4.0" xmlns:gmlexr="http://www.opengis.net/gml/3.3/exr" xmlns:bu-
  base="http://inspire.ec.europa.eu/schemas/bu-base/4.0" xmlns:base="http://inspire.ec.europa.eu/schemas/base/3.3"
  xmlns:gml="http://www.opengis.net/gml/3.2" version="2.0" xmlns:bu-ext="http://inspire.ec.europa.eu/draft-schemas/bu-ext/2.0"
  targetNamespace="http://inspire.ec.europa.eu/draft-schemas/bu-ext/2.0" elementFormDefault="qualified">
    <import schemaLocation="https://inspire.ec.europa.eu/schemas/base2/2.0/BaseTypes2.xsd"
      namespace="http://inspire.ec.europa.eu/schemas/base2/2.0"/>
    <import schemaLocation="http://schemas.opengis.net/iso/19139/20070417/gmd/gmd.xsd" namespace="http://www.isotc211.org/2005/gmd"/>
    <import schemaLocation="https://inspire.ec.europa.eu/schemas/cp/4.0/CadastralParcels.xsd" namespace="http://inspire.ec.europa.eu/schemas/cp/4.0"/>
    <import schemaLocation="https://inspire.ec.europa.eu/schemas/ad/4.0/Addresses.xsd" namespace="http://inspire.ec.europa.eu/schemas/ad/4.0"/>
    <import schemaLocation="http://schemas.opengis.net/gml/3.3/extdEncRule.xsd" namespace="http://www.opengis.net/gml/3.3/exr"/>
    <import schemaLocation="https://inspire.ec.europa.eu/schemas/bu-base/4.0/BuildingsBase.xsd" namespace="http://inspire.ec.europa.eu/schemas/bu-
      base/4.0"/>
    <import schemaLocation="https://inspire.ec.europa.eu/schemas/base/3.3/BaseTypes.xsd" namespace="http://inspire.ec.europa.eu/schemas/base/3.3"/>
    <import schemaLocation="http://schemas.opengis.net/gml/3.2.1/gml.xsd" namespace="http://www.opengis.net/gml/3.2"/>
  - <annotation>
      <documentation>-- Name -- Building extended base -- Definition -- The extended base application schema for INSPIRE theme buildings.</documentation>
    </annotation>
  - <element type="bu-ext:AbstractBuildingUnitType" abstract="true" substitutionGroup="bu-ext:BuildingAndBuildingUnitInfo" name="AbstractBuildingUnit">
    - <annotation>
        <documentation>-- Name -- Abstract building unit -- Definition -- Abstract spatial object type grouping the semantic properties of building units. A
          BuildingUnit is a subdivision of Building with its own lockable access from the outside or from a common area (i.e. not from another BuildingUnit),
          which is atomic, functionally independent, and may be separately sold, rented out, inherited, etc. -- Description -- Building units are spatial
          objects aimed at subdividing buildings and/or building parts into smaller parts that are treated as seperate entities in daily life. A building unit is
          homogeneous, regarding management aspects. EXAMPLES: It may be e.g. an apartment in a condominium, a terraced house, or a shop inside a
          shopping arcade. NOTE 1: According to national regulations, a building unit may be a flat, a cellar, a garage or set of a flat, a cellar and a garage.
          NOTE 2: According to national regulation, a building that is one entity for daily life (typically, a single family house) may be considered as a
          Building composed of one BuildingUnit or as a Building composed of zero BuildingUnit.</documentation>
      </annotation>
    </element>
  - <complexType abstract="true" name="AbstractBuildingUnitType">
    - <complexContent>
      - <extension base="bu-ext:BuildingAndBuildingUnitInfoType">
        - <sequence>
          - <element type="dateTime" name="endLifespanVersion" maxOccurs="1" minOccurs="0">
            - <annotation>
                <documentation>-- Name -- End lifespan version -- Definition -- Date and time at which this version of the spatial object was superseded
                  or retired in the spatial data set.</documentation>
              </annotation>
            </element>
          - <element type="dateTime" name="beginLifespanVersion">
            - <annotation>
                <documentation>-- Name -- Begin lifespan version -- Definition -- Date and time at which this version of the spatial object was inserted
                  or changed in the spatial data set.</documentation>
              </annotation>
            </element>
          - <element name="externalReference">
            - <annotation>
                <documentation>-- Name -- External reference -- Definition -- Reference to an external information system containing any piece of
                  information related to the spatial object. -- Description -- Typically, the external reference will be established to the information
```


Source: own creation, JRC, 2020.

Figure 162. Buildings - Extended 3D modified application schema


```
https://www.epsilon-italia.it/public/EnergyPilot/schemas/uc2/2.0/BuildingsExtended3D.xsd

<?xml version="1.0" encoding="UTF-8"?>
<!--Generated by Enterprise Architect 12.0.1210 ( Build: 1210 )-->
- <schema xmlns="http://www.w3.org/2001/XMLSchema" xmlns:gmlexr="http://www.opengis.net/gml/3.3/exr" xmlns:bu-ext="http://inspire.ec.europa.eu/draft-
  schemas/bu-ext/2.0" xmlns:bu-core3d="http://inspire.ec.europa.eu/schemas/bu-core3d/4.0" xmlns:base="http://inspire.ec.europa.eu/schemas/base/3.3"
  xmlns:gml="http://www.opengis.net/gml/3.2" version="2.0" xmlns:bu-ext3d="http://inspire.ec.europa.eu/draft-schemas/bu-ext3d/2.0"
  targetNamespace="http://inspire.ec.europa.eu/draft-schemas/bu-ext3d/2.0" elementFormDefault="qualified">
    <import schemaLocation="http://schemas.opengis.net/gml/3.3/extdEncRule.xsd" namespace="http://www.opengis.net/gml/3.3/exr"/>
    <import schemaLocation="http://www.epsilon-italia.it/public/EnergyPilot/schemas/uc2/2.0/BuildingsExtendedBase.xsd"
      namespace="http://inspire.ec.europa.eu/draft-schemas/bu-ext/2.0"/>
    <import schemaLocation="https://inspire.ec.europa.eu/schemas/bu-core3d/4.0/BuildingsCore3D.xsd" namespace="http://inspire.ec.europa.eu/schemas/bu-
      core3d/4.0"/>
    <import schemaLocation="https://inspire.ec.europa.eu/schemas/base/3.3/BaseTypes.xsd" namespace="http://inspire.ec.europa.eu/schemas/base/3.3"/>
    <import schemaLocation="http://schemas.opengis.net/gml/3.2.1/gml.xsd" namespace="http://www.opengis.net/gml/3.2"/>
  - <annotation>
      <documentation>-- Name -- Building extended 3D -- Definition -- The extended 3D application schema for INSPIRE theme buildings.</documentation>
    </annotation>
  - <element type="bu-ext3d:BoundaryGeometry3DType" substitutionGroup="gml:AbstractObject" name="BoundaryGeometry3D">
    - <annotation>
        <documentation>-- Name -- BoundaryGeometry3D -- Definition -- The information related to the boundary geometry as 3D data.</documentation>
      </annotation>
    </element>
  - <complexType name="BoundaryGeometry3DType">
    - <sequence>
      - <element type="gml:LengthType" name="verticalGeometryEstimatedAccuracyLoD4" maxOccurs="1" minOccurs="0">
        - <annotation>
            <documentation>-- Name -- verticalGeometryEstimatedAccuracyLoD4 -- Definition -- The estimated absolute positional accuracy of the Z
              coordinate of the LoD4 boundary representation, in the INSPIRE official Coordinate Reference System. Absolute positional accuracy is
              defined as the mean value of the positional uncertainties for a set of positions where the positional uncertainties are defined as the
              distance between a measured position and what is considered as the corresponding true position. -- Description -- NOTE:This mean value
              may come from quality measures on a homogeneous population of buildings or from an estimation based on the knowledge of the
              production processes and of their accuracy.</documentation>
          </annotation>
        </element>
      - <element type="gml:LengthType" name="verticalGeometryEstimatedAccuracyLoD3" maxOccurs="1" minOccurs="0">
        - <annotation>
            <documentation>-- Name -- verticalGeometryEstimatedAccuracyLoD3 -- Definition -- The estimated absolute positional accuracy of the Z
              coordinate of the LoD3 boundary representation, in the INSPIRE official Coordinate Reference System. Absolute positional accuracy is
              defined as the mean value of the positional uncertainties for a set of positions where the positional uncertainties are defined as the
              distance between a measured position and what is considered as the corresponding true position. -- Description -- NOTE:This mean value
              may come from quality measures on a homogeneous population of buildings or from an estimation based on the knowledge of the
              production processes and of their accuracy.</documentation>
          </annotation>
        </element>
      - <element type="gml:LengthType" name="verticalGeometryEstimatedAccuracyLoD2" maxOccurs="1" minOccurs="0">
        - <annotation>
            <documentation>-- Name -- verticalGeometryEstimatedAccuracyLoD2 -- Definition -- The estimated absolute positional accuracy of the Z
              coordinate of the LoD2 boundary representation, in the INSPIRE official Coordinate Reference System. Absolute positional accuracy is
              defined as the mean value of the positional uncertainties for a set of positions where the positional uncertainties are defined as the
              distance between a measured position and what is considered as the corresponding true position. -- Definition -- NOTE:This mean value
              may come from quality measures on a homogeneous population of buildings or from an estimation based on the knowledge of the
              production processes and of their accuracy.</documentation>
          </annotation>
        </element>
      - <element type="gml:MultiSurfacePropertyType" name="LoD4MultiSurface" maxOccurs="1" minOccurs="0">
```


Source: own creation, JRC, 2020.

After the generation of the modified application schemas, the mapping step has been performed and the related matching table has been compiled, as shown in Figure 163.

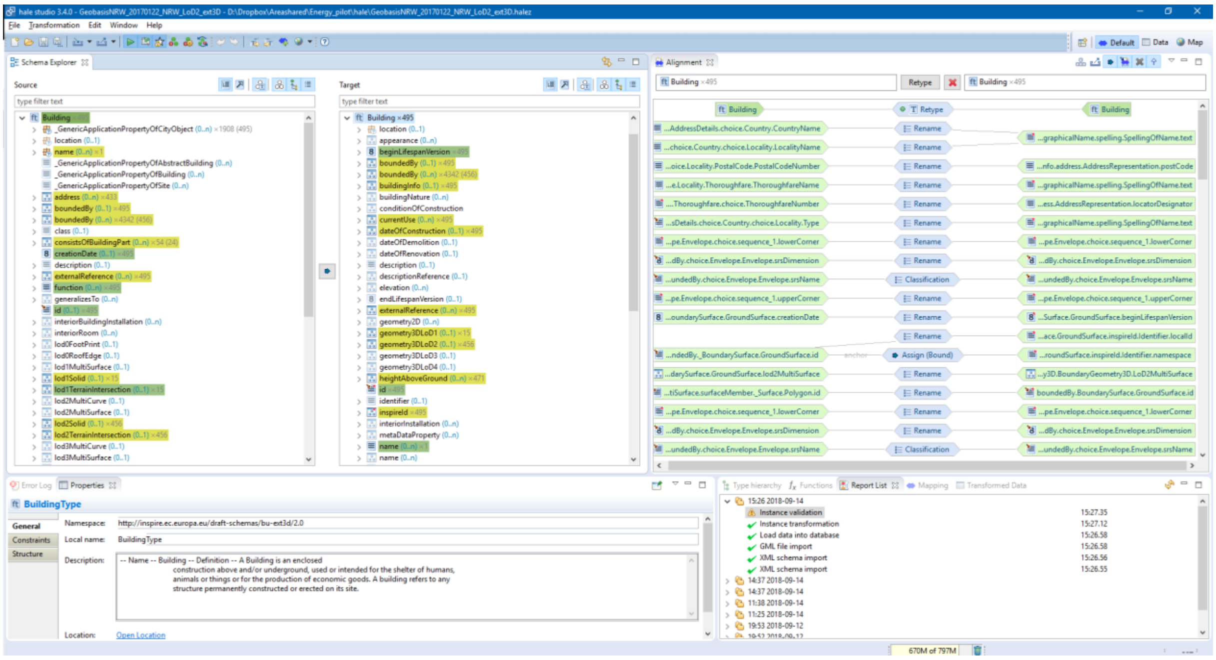
Figure 163. Matching table between LoD2 and INSPIRE BU3D Extended

Attribute of Source Data	Attribute of Target Data	Function
Building	Building	Retype
Address		
address...Country.CountryName	buildingInfo...address...adminUnit...name	Rename
address...Country...Locality.LocalityName	buildingInfo...address...adminUnit...name	Rename
address...Country...Locality...PostalCodeNumber	buildingInfo...address...postCode	Rename
address...Country...Locality...ThoroughfareName	buildingInfo...address...locatorName	Rename
address...Country...Locality...ThoroughfareNumber	buildingInfo...address...locatorDesignator	Rename
address...Country...Locality.Type	buildingInfo...address...addressArea...name	Rename
boundedBy(0...n)		
Ground Surface		
~~boundedBy._BoundarySurface.GroundSurface~~	~~boundedBy.BoundarySurface.GroundSurface~~	~~Reproject Geometry~~
boundedBy...GroundSurface...Envelope...lowerCorner	boundedBy...GroundSurface...Envelope...lowerCorner	Rename
boundedBy...GroundSurface...Envelope.srsDimension	boundedBy...GroundSurface...Envelope.srsDimension	Rename
boundedBy...GroundSurface...Envelope.srsName	boundedBy...GroundSurface...Envelope.srsName	Rename
boundedBy...GroundSurface...Envelope...upperCorner	boundedBy...GroundSurface...Envelope...upperCorner	Rename
boundedBy._BoundarySurface.GroundSurface.id	boundedBy.BoundarySurface.GroundSurface.id	Rename
~~boundedBy._BoundarySurface.GroundSurface.lod2MultiSurface.MultiSurface~~	~~boundedBy._BoundarySurface.GroundSurface.multiSurfaceLod2.MultiSurface~~	~~Reproject Geometry~~
boundedBy...GroundSurface.lod2MultiSurface.MultiSurface...Polygon...LinearRing...posList	boundedBy...GroundSurface.multiSurfaceLod2.MultiSurface...Polygon...LinearRing...posList	Rename
boundedBy...GroundSurface.lod2MultiSurface.MultiSurface...Polygon...LinearRing...posList.srsDimension	boundedBy...GroundSurface.multiSurfaceLod2.MultiSurface...Polygon...LinearRing...posList.srsDimansion	Rename
Roof Surface		
~~boundedBy._BoundarySurface.RoofSurface~~	~~boundedBy.BoundarySurface.RoofSurface~~	~~Reproject Geometry~~
boundedBy...RoofSurface...Envelope...lowerCorner	boundedBy...RoofSurface...Envelope...lowerCorner	Rename
boundedBy...RoofSurface...Envelope.srsDimension	boundedBy...RoofSurface...Envelope.srsDimension	Rename
boundedBy...RoofSurface...Envelope.srsName	boundedBy...RoofSurface...Envelope.srsName	Rename
boundedBy...RoofSurface...Envelope...upperCorner	boundedBy...RoofSurface...Envelope...upperCorner	Rename
boundedBy._BoundarySurface.RoofSurface.id	boundedBy.BoundarySurface.RoofSurface.id	Rename
~~boundedBy._BoundarySurface.RoofSurface.lod2MultiSurface.MultiSurface~~	~~boundedBy._BoundarySurface.RoofSurface.multiSurfaceLod2.MultiSurface~~	~~Reproject Geometry~~
boundedBy...RoofSurface.lod2MultiSurface.MultiSurface...Polygon...LinearRing...posList	boundedBy...RoofSurface.multiSurfaceLod2.MultiSurface...Polygon...LinearRing...posList	Rename
boundedBy...RoofSurface.lod2MultiSurface.MultiSurface...Polygon...LinearRing...posList.srsDimension	boundedBy...RoofSurface.multiSurfaceLod2.MultiSurface...Polygon...LinearRing...posList.srsDimansion	Rename
Wall Surface		
~~boundedBy._BoundarySurface.WallSurface~~	~~boundedBy._BoundarySurface.WallSurface~~	~~Reproject Geometry~~
boundedBy...WallSurface...Envelope...lowerCorner	boundedBy...WallSurface...Envelope...lowerCorner	Rename
boundedBy...WallSurface...Envelope.srsDimension	boundedBy...WallSurface...Envelope.srsDimension	Rename
boundedBy...WallSurface...Envelope.srsName	boundedBy...WallSurface...Envelope.srsName	Rename
boundedBy...WallSurface...Envelope...upperCorner	boundedBy...WallSurface...Envelope...upperCorner	Rename
boundedBy._BoundarySurface.WallSurface.id	boundedBy.BoundarySurface.WallSurface.id	Rename
~~boundedBy._BoundarySurface.WallSurface.lod2MultiSurface.MultiSurface~~	~~boundedBy._BoundarySurface.WallSurface.multiSurfaceLod2.MultiSurface~~	~~Reproject Geometry~~
boundedBy...WallSurface.lod2MultiSurface.MultiSurface...Polygon...LinearRing...posList	boundedBy...WallSurface.multiSurfaceLod2.MultiSurface...Polygon...LinearRing...posList	Rename
boundedBy...WallSurface.lod2MultiSurface.MultiSurface...Polygon...LinearRing...posList.srsDimension	boundedBy...WallSurface.multiSurfaceLod2.MultiSurface...Polygon...LinearRing...posList.srsDimansion	Rename
boundedBy(0...1)		
boundedBy...Envelope...lowerCorner	boundedBy...Envelope...lowerCorner	Rename
boundedBy...Envelope.srsDimension	boundedBy...Envelope.srsDimension	Rename
boundedBy...Envelope.srsName	boundedBy...Envelope.srsName	Rename
boundedBy...Envelope...upperCorner	boundedBy...Envelope...upperCorner	Rename
consistsOfBuildingPart		
~~consistsOfBuildingPart.BuildingPart~~	~~parts.BuildingPart~~	~~Reproject Geometry~~
other attributes are mapped same as attributes under Building		
creationDate		
creationDate	beginLifespanVersion~~.nilReason~~	Rename
externalReference		
externalReference.externalObject.choice.name	externalReference.ExternalReference.reference	Rename
externalReference.informationSystem	externalReference.ExternalReference.informationSystemName...CharacterString	Rename
externalReference.informationSystem	externalReference.ExternalReference.informationSystem	Rename
function		
function	currentUse	Rename
id		
id	id	Rename
id	inspireId.identifier.localId	Rename
lod1Solid		
~~lod1Solid._Solid.Solid.exterior._Surface.CompositeSurface~~	~~lod1Geometry.AbstractSolid.Solid.exterior.Shell.surfaceMember.AbstractSurface.CompositeSurface~~	~~Reproject Geometry~~
lod1Solid...CompopsiteSurface...Polygon...LinearRing...posList	lod1Geometry...Solid...CompositeSurface...Polygon...LinearRing...posList	Rename
lod1Solid...CompopsiteSurface...Polygon...LinearRing...posList.srsDimension	lod1Geometry...Solid...CompositeSurface...Polygon...LinearRing...posList.srsDimension	Rename
lod2Solid		
lod2Solid._Solid.Solid.exterior._Surface.CompositeSurface.surfaceMember.href	lod2Geometry.AbstractSolid.Solid.exterior.Shell.surfaceMember.AbstractSurface.CompositeSurface.surfaceMember.href	Rename
measuredHeight		
measuredHeight	heightAboveGround.HeightAboveGround.value	Rename
measuredHeight.uom	heightAboveGround.HeightAboveGround.value.uom	Rename
name		
name.name	name	Rename
storeysAboveGround		
storeysAboveGround	numberOfFloorsAboveGround	Rename
storeysAboveGround	buildingInfo...floorDistribution...highestFloor	Rename
storeysBelowGround		
storeysBelowGround	buildingInfo...floorDistribution...lowestFloor	Rename if value = 0
storeysBelowGround	buildingInfo...floorDistribution...lowestFloor	FormattedString if value <> 0 ['-' + value]
yearOfConstruction		
yearOfConstruction	dateOfConstruction.DateOfEvent.end.~~nilReason~~	Rename

Source: own creation, JRC, 2020.

The transformation rules identified and documented in the matching table have been applied to the source dataset, using the hale studio transformation tool, as shown in Figure 164.

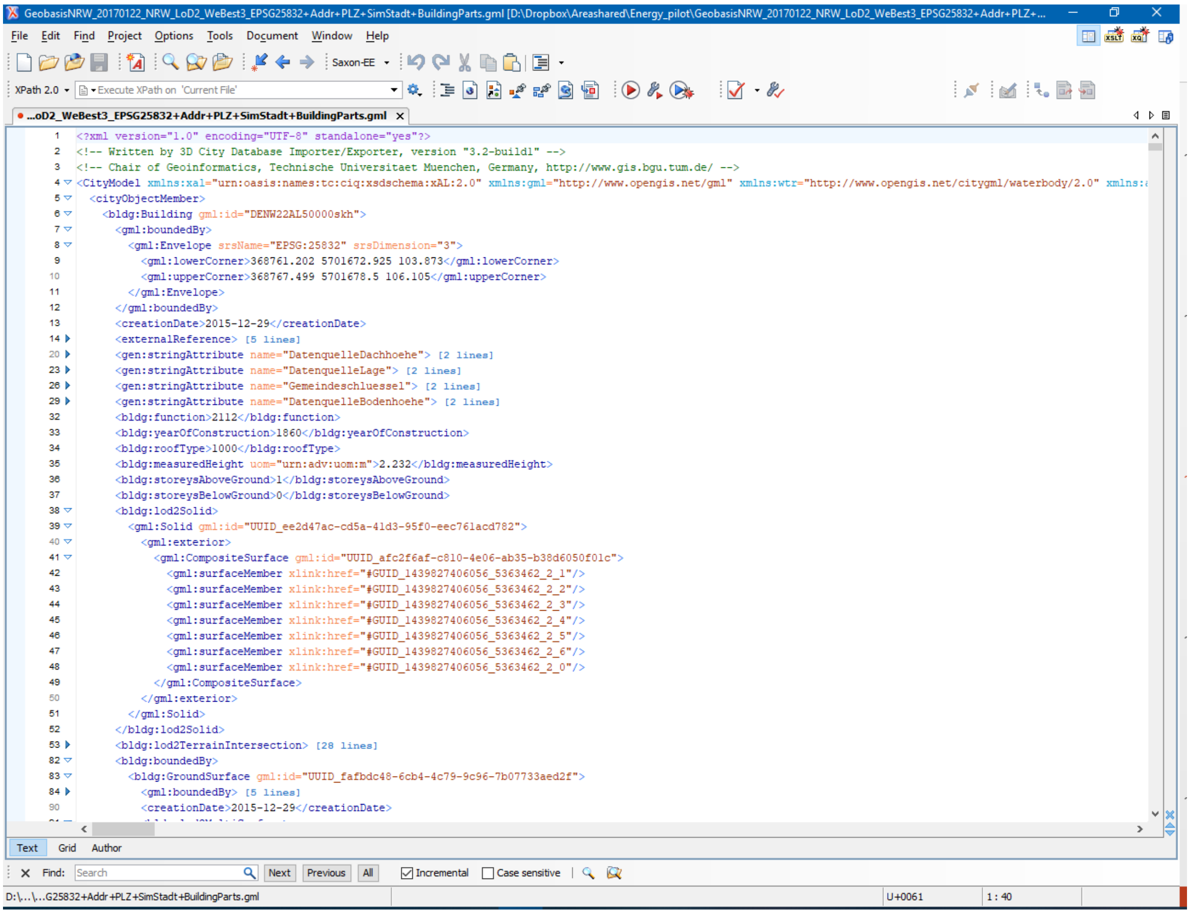
Figure 164. Hale studio project



Source: own creation, JRC, 2020.

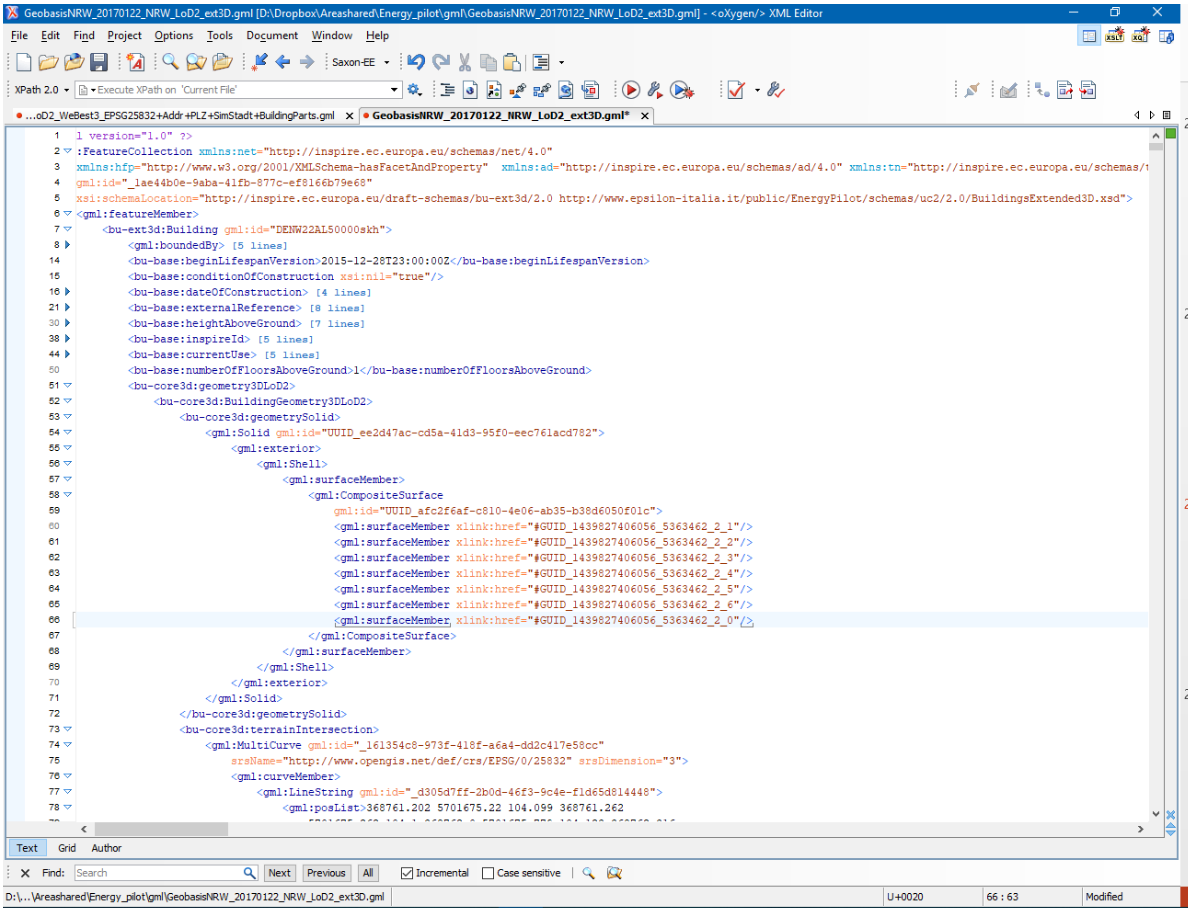
The source CityGML Lod2 dataset and the related harmonized GML dataset in an XML viewer are shown in the Figure 165 and Figure 166, respectively.

Figure 165. Source dataset – CityGML LoD2



Source: own creation, JRC, 2020.

Figure 166. Target dataset - INSPIRE BU Extended-3D



Source: own creation, JRC, 2020.

7.3 Extension of the INSPIRE Building 3D data model

The output of the SimStadt simulations is the "average monthly heating energy demand" calculated for the building or building unit, measured in KWh/m^2/month. In order to provide this information through an output dataset compliant to INSPIRE, an extension is required because there are no attributes suitable for this scope in any of the INSPIRE Buildings data models.

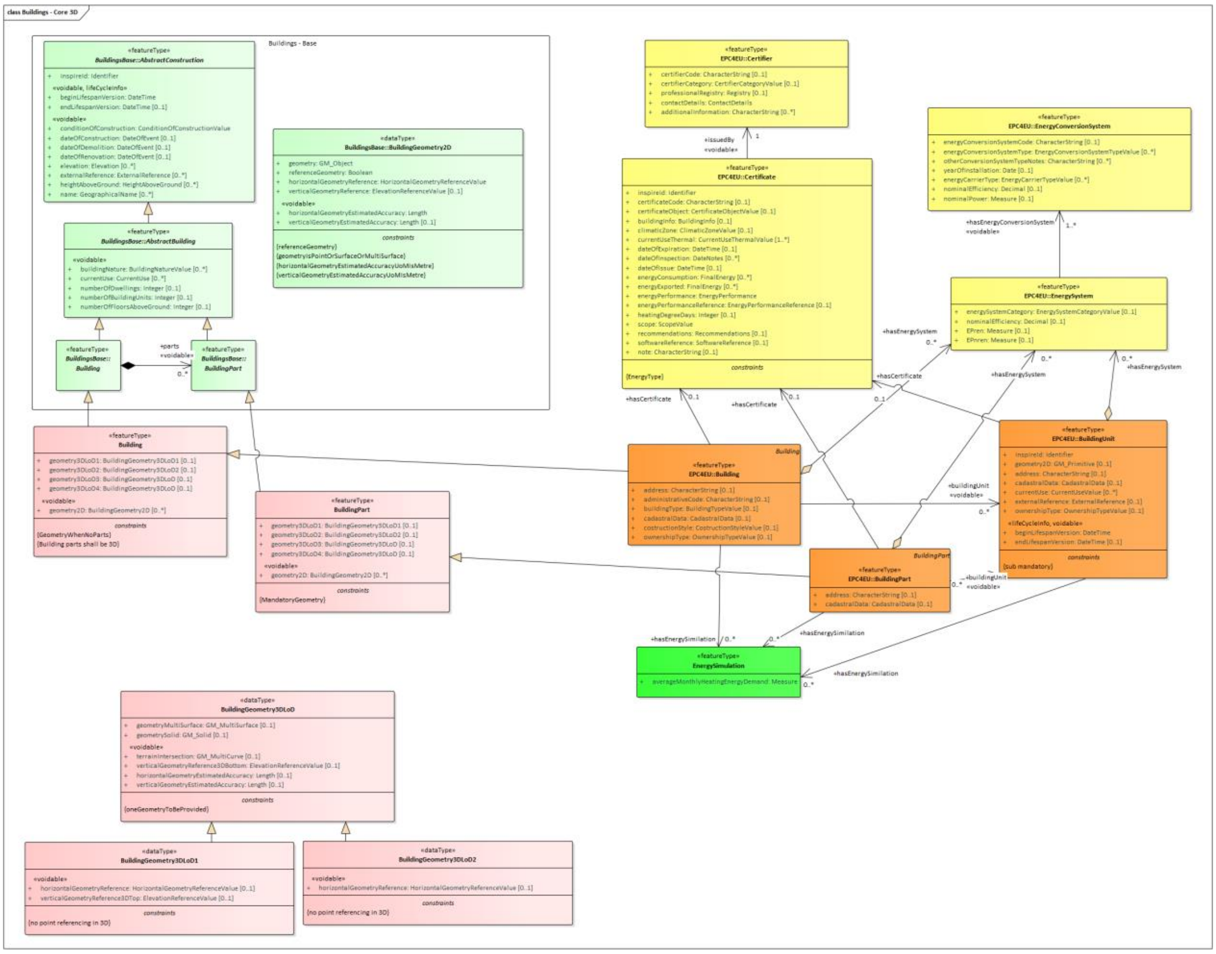
In the frame of the use cases "EPC-IT - INSPIRE Harmonisation of Energy Performance Certificates of buildings datasets in Italy" and "EPC-ES - INSPIRE Harmonisation of Energy Performance Certificates of buildings datasets in Spain", an extension of the INSPIRE BU Core2D data model has been developed in order to map the information contained in the Energy Performance Certificates (EPC) of buildings. The UML diagram of this extended data model, named EPC4EU, is available online[49].

Because it could be useful to have in the same dataset the information contained in the Energy Performance Certificates and in the energy simulation results, a new data model has been drafted for this purpose.

It extends the INSPIRE Building 3D Core data model, adding the same feature types defined in the EPC4EU data model (in orange and yellow in Figure 167) and a new feature type "EnergySimulation" (in green) that can be used to map the results of an energy performance simulation.

[49] https://inspire-sandbox.jrc.ec.europa.eu/energy-pilot/epc4eu/data-model/3.0/html/

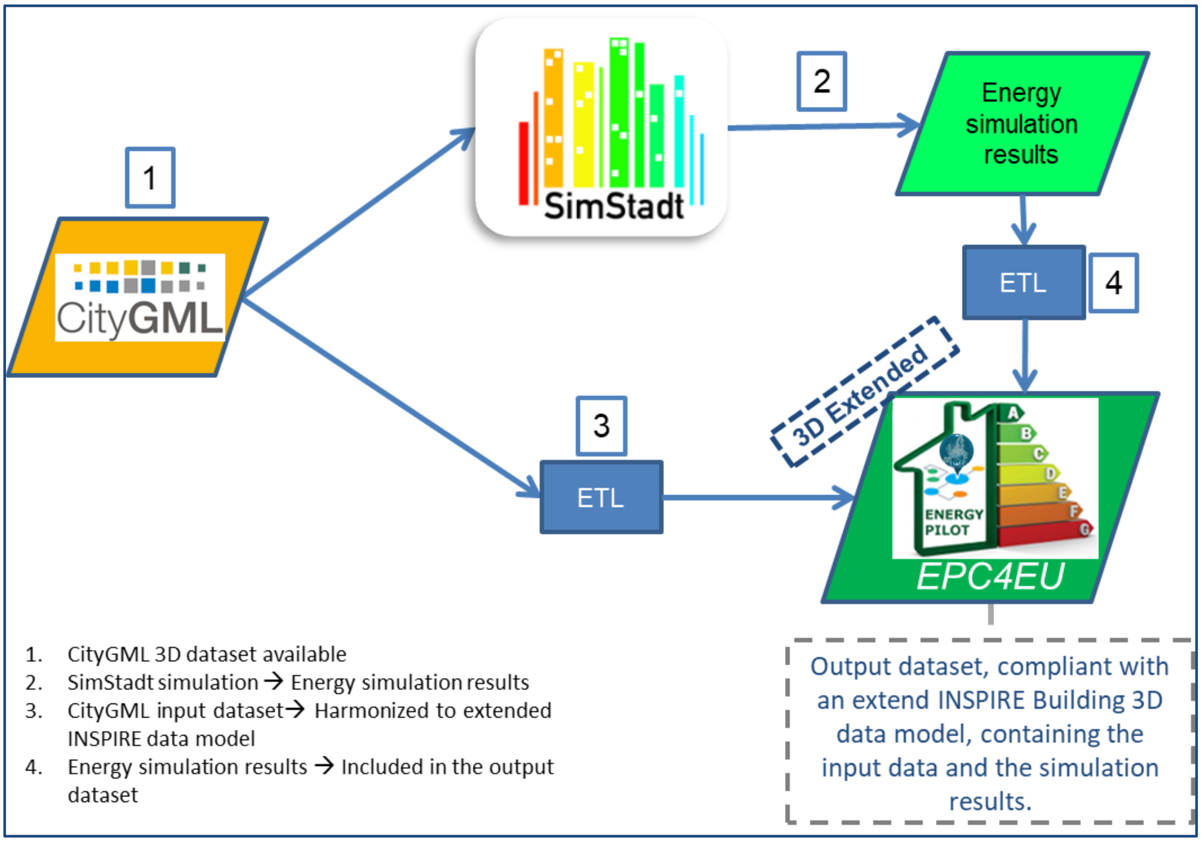
Figure 167. New extension of INSPIRE Building 3D Core data model



Source: own creation, JRC, 2020.

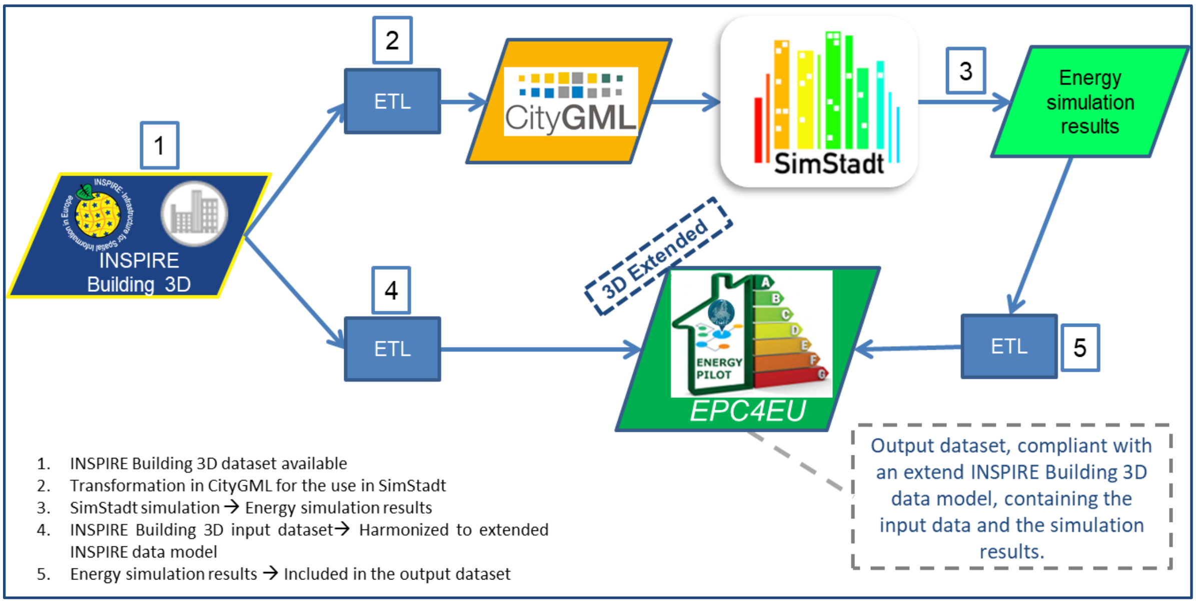
Taking into consideration the proposed extension, the two scenarios described in section 7.2 can be schematized as in the Figure 168 and in the Figure 169.

Figure 168. Final data transformation - Scenario 1



Source: own creation, JRC, 2020.

Figure 169. Final data transformation - Scenario 2



Source: own creation, JRC, 2020.

7.4 Conclusions

The possibility to improve the interoperability of input/output data to/from energy simulation tools, like SimStadt, has been explored.

Regarding the input datasets, it has been illustrated how the energy simulation tools can benefit from the availability of INSPIRE 3D building datasets, because they can be easily transformed according to the CityGML data model, required by SimStadt, following the methodology described in this report.

Regarding the output datasets, the possibility to make the results of the energy simulations available in a unique dataset together with the original building information and the additional information contained in the EPCs has been illustrated. The interoperability of the output dataset can be guaranteed by using an extension of the INSPIRE Building 3D data model, based on the EPC4EU data model (representing an extension of the INSPIRE Building 2D data model developed in the frame of other use cases of the Energy & Location Applications to harmonise EPC datasets).

In terms of practical mapping exercises, easily re-usable in similar data transformation contexts, 3 different mappings between CityGML LoD1/LoD2 and INSPIRE BU 3D have been implemented:

- CityGML LoD1 vs INSPIRE BU 3D CORE
- CityGML LoD2 vs INSPIRE BU 3D CORE
- CityGML LoD2 vs INSPIRE BU 3D EXTENDED

Moreover, during the extension of the INSPIRE BU 3D data model, errors in the INSPIRE draft extended schemas (BuildingExtendedBase and BuildingExtended3D) have been found and fixed.

Finally, it has been highlighted that the reverse mapping from INSPIRE BU 3D to CityGML is easily doable and it can improve the interoperability of energy simulation tools using CityGML as input data.

8 Conclusions

A methodology to perform energy simulations predicting energy heat demand at building level, based on the use of SimStadt software and of input data consisting of CityGML 3D building data and weather data, has been applied in several case studies in 4 test city areas in 3 different Member States (NL, DE and ES) and thoroughly documented in the sections from 3 to 6.

The Table 60 below summarises the main results obtained for each case study (CS), focusing on key elements, such as the simulation scale (district or city), the simulation engine (SimStadt). the input data used for the simulations and the main conclusions.

Table 60. Comparison among all the case studies

DE-Essen	**Scale:** City	**Simulation engine:** SimStadt
INPUTS	CityGML based on German cadastre in different level of detail (LoD 1, LoD 2);	
MAIN CONCLUSIONS:	**Impact of model geometry:** the improved accuracy of the simulation results depending on the better accuracy of the 3D building input data has not been demonstrated. Most important is that the volume and the number of storeys of the building or building part is represented correctly by the model geometry. This can be approximated in CityGML LoD 1 by using the average building height and an additional attribute "eaves height". In CityGML LoD 2 the roof structure is part of the building geometry. However, several comparisons between results obtained with LoD 1 and LoD2 CityGML datasets indicate that the floor area is over-estimated in LoD 2 data sets if it is derived from the 3D building geometry without any further information such as number of floors. However, the heating demand depend on the heating volume of the building, the floor area is only used to calculate the indicator KWh/m^2y. If 3D building geometry is available, a better indicator is KWh/m^3 per year. **Verification of the results:** when comparing energy simulations with real energy consumption data, it is important to highlight that energy simulations do not consider user behaviours, as well as possible energy efficiency interventions made on (parts of) the simulated buildings, which instead have a strong impact on the energy consumption. **Transfer methodology to other regions / EU member states:** the datasets used are available for the entire building stock in Germany, provided by the state survey. The year of construction is missing in this national dataset, but can be integrated from (commercial) data sources. Information about refurbishment of buildings is missing. It has been shown that the data can be converted to the 3D building data of INSPIRE without loss of information.	
NL-Zwolle	**Scale:** City	**Simulation engine:** SimStadt
INPUTS	CityGML from Dutch cadastre (LoD 1)	
MAIN CONCLUSIONS:	**Transfer methodology to other regions:** the datasets used are available for the entire building stock in the Netherlands. Comparison with energy consumption data (time resolution: year) available as open data in the Netherlands is not trivial as the data is aggregated, but not at building level. A mapping of simulated energy demand per building to the available consumption data is not always possible. **Verification of the results:** In any kind of comparison of energy performance of buildings in different Member States, it is much better to compare absolute values expressed in KWh/m^2y rather than comparing the labels, because the interval values the latter refer to are fixed by country-dependant national laws.	
NL-Enschede	**Scale:** District	**Simulation engine:** SimStadt

INPUTS	CityGML from Dutch cadastre (LoD 1)	
MAIN CONCLUSIONS:	**Transfer methodology to other regions:** a practical workflow to develop a LOD 1 CityGML model from publicly available GIS data related to buildings and addresses has been defined and tested. To be able to use SimStadt with a Dutch case study, a local building physics library was developed, specific for the Netherlands. Different climate datasets, needed for the simulations, have been collected and analysed. **Verification of the results:** in this case study a verification approach has been used to compare the simulation results with real energy consumption data. The best prediction accuracy for the space heating energy demand was a +20% difference between measurements and predictions.	
ES - CS 1.1	**Scale:** District	**Simulation engine:** SimStadt
INPUTS	**01** (Ad hoc CityGML), **03.1** (CityGML from cadastre), **03.2** (CityGML from LiDAR).	
MAIN CONCLUSIONS:	**Impact of model geometry:** Inputs 01 (ad-hoc generation) and 03.2 (LiDAR data) are able to capture differences in the building geometries and obtain different labels even in the very homogeneous Cuatro de Marzo district (in particular more accurate heights and accurate wall surfaces). **Energy performance is higher** when performing ad hoc modelling (Input 01). Label D was obtained in comparison to other inputs leading to Label E. **Homogeneity of results:** Input 03.1 (Spanish Cadastre), generated by applying 3m height / floor, offers homogeneous results, since there are only two different building typologies in Cuatro de Marzo District.	
ES - CS 2.1	**Scale:** District	**Simulation engine:** SimStadt + ENERGIS
INPUTS	**02** (Cadastre 2D), **03.1** (CityGML from cadastre), **03.2** (CityGML from LiDAR), **04** (CityGML from OSM).	
MAIN CONCLUSIONS:	**Homogeneity of results:** found using Inputs 02 (cadastre) and 04 (OSM), since the hypothesis applied for OSM (15 m total height / building) is very accurate for the Cuatro de Marzo district, where most of the buildings have 5 floors, corresponding to the hypothesis applied in the case of the cadastre (3m/floor). **Same labels obtained with SimStadt and ENERGIS:** the differences in energy performance were around 25kWh/m2. **Slightly higher heating energy demand obtained with input generated with LiDAR:** it might be due to the more complex geometry and the higher external wall surface. **Some outliers detected in LiDAR generation:** city model needs to be checked.	
ES - CS 2.2	**Scale:** City	**Simulation engine:** SimStadt + ENERGIS
INPUTS	**02** (Cadastre 2D), **03.1** (CityGML from cadastre), **04** (CityGML from OSM).	

MAIN CONCLUSIONS:	**More difficulty to extract conclusions** due to the variety of buildings. **Label similarities for inputs 02 and 03.1:** where the majority of buildings were rated as E or G. **OSM input resulted in higher energy efficiency.** However, analysis of the overall building stock in Valladolid would be necessary to confirm if the hypothesis applied is reasonable (15 m / building). **Extreme differences occur for a significant number of buildings**, which reached more than 50kWh/m2.	
ES - CS 3.1	**Scale:** District	**Comparison with Energy Performance Certificates**
INPUTS	**01** (Ad hoc CityGML), **02** (Cadastre 2D), **03.1** (CityGML from cadastre), **03.2** (CityGML from LiDAR).	
MAIN CONCLUSIONS:	**Homogeneity of results**: most of the labels obtained with the different approaches coincided with the ones obtained in the EPCs, where "label E" was the predominant one. However, in the results derived from inputs 02, 03.1 and 03.2 almost 90-100% of the buildings were rated as E, whereas only 64,10% were rated as E in the real EPCs. **Differences in Input 01**: in the ad hoc model higher efficiencies were obtained as a result and most of the buildings (92.86%) were rated with a label D (instead of E as in the other cases). **Results to be handled with care**: despite these discrepancies, it is worth to mention that the considered EPCs labelled not only buildings as a whole, but also individual dwellings, (where higher discrepancy in results can be found). These data are compared to results obtained with the simulations at building level (inputs **01, 02, 03.1, 03.2**). Thus, some deviations are to be expected. EPCs at dwelling level were also considered in the comparison, in order to have a higher number of "reference" EPCs.	
ES - CS 3.2	**Scale:** City	**Comparison with Energy Performance Certificates**
INPUTS	**02** (Cadastre 2D), **03.1** (CityGML from cadastre)	
MAIN CONCLUSIONS:	**Homogeneity of results**: at city scale the homogeneity in terms of buildings achieving the same label can be observed, since the majority of the buildings obtained an E label in all cases (02 – 53.21%, 03.1 – 40.22%, Real EPCs – 58,10%). **OSM input could not be used** in this comparison due to the difficulty to identify the buildings and overlap with the cadastral geometry.	

A comparative analysis of the simulation results has been done, aiming at providing insight into the following aspects:

- identify the main obstacles to find and pre-process the input data required by the simulations, including the need to adapt the building physical library used by the simulation software to local contexts,
- identify the main factors influencing the accuracy of the simulation results,
- estimate the influence of the accuracy of the CityGML LoD of the input data on the accuracy of the simulations results,
- identify the main sources of mismatch to be considered when comparing the simulation results with real energy consumption data.

For each of the above listed aspects, the following main conclusions can be drawn:

- Despite the availability of 3D city models as open data is increasing, information required by the energy simulations such as building age is often available only under restricted conditions.
- In the case of the simulations for the test area of Enschede (NL) the building physic library natively present in SimStadt and related to Germany has been successfully adapted to the Dutch building typologies, proving the viability of the adaptation.
- The preparation of the 3D building data as input data for the energy simulations requires the use of software tools which in turn require specific skills.
- A verification methodology to guide the interpretation of the results and of their differences has been introduced.
- The improved accuracy of the simulation results depending on the better accuracy of the 3D building input data has not been demonstrated. Several comparisons between results obtained with LOD1 and LOD2 CityGML datasets have shown that there are some aspects of the building fabric which are better considered using LOD1 datasets, e.g. the reduced over-estimation of the floor area.
- When comparing energy simulations with real energy consumption data, it is important to highlight that energy simulations do not consider user behaviours, as well as possible energy efficiency interventions made on (parts of) the simulated buildings, which instead have a strong impact on the energy consumption.
- In any kind of comparison of energy performance of buildings in different Member States, it is much better to compare absolute values expressed in KWh/m^2/y rather than comparing the labels, because the interval values the latter refers to are fixed by country-dependant national laws.
- Despite all the simulations documented in this report have been made with the SimStadt software, in the case of Spain the simulations have been done using also another software (ENERGIS). However, assessing the dependency of the simulation results on the simulation software would require additional investigations which are out of scope of the work undertaken.

Finally, several mapping exercises between CityGML and INSPIRE data models available for 3D Buildings have been executed and documented, improving the interoperability of input and/or output datasets of the simulations.

In conclusion, notwithstanding the above listed issues, the methodology can be re-used in other geographical areas (Member States) by parties aiming to assess the energy performance of their building stock and interested to preliminary assess costs and benefits of applying the same (or similar) methodologies based on the availability of similar datasets, with respect to those used in the comparative analysis presented in this report.

The conclusions above may have potential implications on several policy-related discussions regarding the improvement of the energy efficiency of the building stock. To this end, the following recommendations can be formulated.

Recommendation 1: 3D city models at different levels of detail, including information required by the energy simulations such as building age, should be made available as High Value Datasets[50] and shared according to FAIR[51] principles, possibly within Energy Data Spaces[52].

Recommendation 2: An EU common methodology to assess and document the quality, expressed in terms of different quality components (e.g. accuracy, completeness, up-to-date), of the input/output data used for the simulations of energy heat demand for building, should be developed.

Recommendation 3: Building physic libraries modelling the different building typologies in the different Member States should be developed adopting common semantics and shared under FAIR conditions.

Recommendation 4: An EU common methodology to validate the results of the simulations of energy heat demand for buildings, obtained with different simulation software, should be developed.

[50] High Value Datasets defined in the DIRECTIVE (EU) 2019/1024 on open data and the re-use of public sector information (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024).

[51] Findability, Accessibility, Interoperability, Reusability. The FAIR Guiding Principles for scientific data management and stewardship (https://doi.org/10.1038/sdata.2016.18)

[52] Common European Data Spaces, as defined in the European Strategy for data (https://digital-strategy.ec.europa.eu/en/policies/strategy-data)

Recommendation 5: Adequate digital skills needed for an accurate assessment of the energy performance of the building stock should be formalised at EU level and the set-up of adequate education and training initiatives should be encouraged to fill-in the related skill gaps..

References

[1] C.F. Reinhart, and C.C. Davila: Urban building energy modeling – a review of a nascent field, Build. Envirn. 97 (2016) 196 – 202

[2] F. Xue, L. Wu, W. Lu: Semantic enrichment of building and cityinformation models: a ten-years review, Adv. Eng. Inform. 47 (2021) 101245

[3] R. Nouvel, K.H. Brassel, M. Bruse, E. Duminil, V. Coors, U. Eicker, D. Robinson: SimStadt, a new workflow-driven urban energy simulation platform for CityGML City models, Proc. CISBAT 2015, 889-894

[4] R. Nouvel, M. Zirak, V. Coors, U. Eicker: The influence of data quality on Urban Heating Demand Modeling using 3D city models, Comput. Environ. Urban Syst. (2017) 68-80

[5] Dochev, P. Gorzalka, V. Weiler, J. Estevam Schmiedt, M. Linkiewicz, U. Eicker, B. Hoffschmidt, I. Peters, B. Schröter, Calculating urban heat demands: An analysis of two modelling approaches and remote sensing for input data and validation, Energy Buildings 226 (2020) 110378

[6] SimStadt tool, https://SimStadt.hft-stuttgart.de/en/index.jsp [last viewed, November 2020]

[7] https://www.osti.gov/biblio/90674-international-energy-agency-building-energy-simulation-test-bestest-diagnostic-method

[8] ENERGIS website: http://api.voxel3d.es/examples/ENERGIS/portal/ [last viewed, November 2020]

[9] INSEL: Software for simulation, monitoring, and visualization of energy systems https://www.insel.eu/en/ [last viewed, November 2020]

[10] Monien, D., Strzalka, A., Koukofikis, A., Coors, V., & Eicker, U. (2017). Comparison of building modelling assumptions and methods for urban scale heat demand forecasting. Future Cities and Environment, 3(1), 2.

[11] Huang YJ. Development of 3012 IWEC2 weather files for international locations (RP-1477). ASHRAE Trans 2014; 120: 340–355.

[12] OptEEmAL project, Optimised energy efficient design platform for refurbishment at district level, https://www.opteemal-project.eu/ [last viewed, November 2020]

[13] BuildingSMART, Industry Foundation Classes standard, https://technical.buildingsmart.org/standards/ifc/ [last viewed, November 2020]

[14] Dirección General del Catastro, Servicios INSPIRE de Cartografía Catastral, http://www.catastro.minhap.es/webinspire/index.html [last viewed, November 2020]

[15] FME software, https://www.safe.com/ [last viewed, November 2020]

[16] R2Cities project, Residential Renovation towards nearly zero energy CITIES, http://es.r2cities.eu/ [last viewed, November 2020]

[17] 3DIS, City Editor Sketch-up Plugin: https://www.3dis.de/cityeditor/ [last viewed, November 2020]

[18] V. Coors, M. Betz, E. Duminil: A concept of Quality Management of 3D City Models Supporting Application Specific Requirements, Journal of Photogrammetry, Remote Sensing and Geoinformation Science, Vol 88, pp 3–14 (2020) Springer, 2020 ISSN: 2512-2789, https://doi.org/10.1007/s41064-020-00094-0 , Open access

[19] City Doctor tool, https://www.citydoctor.eu/index.php/citydoctor_main.html?language=en [last viewed, November 2020]

[20] CitySim http://infoscience.epfl.ch/record/148717/files/dr.bsCitySim%20BS09_1083_1090.pdf [last viewed, November 2020]

[21] Stanet network analysis http://stafu.de/en/ [last viewed, November 2020]

[22] http://193.196.52.104:8080/Apps/Helsinki/view.html Rossknecht, M.; Airaksinen, E. Concept and Evaluation of Heating Demand Prediction Based on 3D City Models and the CityGML Energy ADE—Case Study Helsinki. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 602.

[23] Spanish Cadastre: https://www.sedecatastro.gob.es/ [last viewed, November 2020]

[24] Spanish Building Code, Código Técnico de la Edificación https://www.codigotecnico.org/ [last viewed, November 2020]

[25] Efinovatic, CENER, Documento Reconocido para la Certificación Energética de Edificios Existentes, CE3X tool, http://www.efinova.es/CE3X [last viewed, November 2020]

[26] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32007L0002 [last viewed, November 2020]

[27] European Commission, Clean Energy for all Europeans package, https://ec.europa.eu/energy/topics/energy-strategy/clean-energy-all-europeans_en [last viewed, November 2020]

[28] Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency (Text with EEA relevance) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018L0844 [last viewed, November 2020]

[29] Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:32010L0031 [last viewed, November 2020]

[30] Validated Energy Performance Certification tools in Spain https://energia.gob.es/desarrollo/EficienciaEnergetica/CertificacionEnergetica/DocumentosReconocidos/Paginas/procedimientos-certificacion-proyecto-terminados.aspx [last viewed, November 2020]

[31] Energy Performance Certificates open data portal in Castilla y León https://analisis.datosabiertos.jcyl.es/explore/dataset/certificados-de-eficiencia-energetica/table/?flg=es&location=7,40.40761,-1.9281&basemap=jawg.streets [last viewed, November 2020]

List of abbreviations and definitions

ADE	Application Domain Extension
ALKIS	Amtliches Liegenschaftskatasterinformationssystem
API	Application Programming Interface
ASPRS	American Society for Photogrammetry & Remote Sensing
CDD	Cooling Degree Days
CE3X	Tool for energy certification in Spain
CityGML	City Geography Markup Language
CS	Case Study
CSV	Comma Separated Values
DE	Germany
EC	European Commission
EPC	Energy Performance Certificate
ES	Spain
ESCO	Energy Service Company
ETL	Extract, Transform, Load
EU	European Union
FP7	European Commission Seventh Framework Programme
GIS	Geography Information System
GML	Geography Markup Language
GPS	Global Positioning System
HDD	Heating Degree Days
IFC	Industry Foundation Classes
INSPIRE ISDSS	COMMISSION REGULATION (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial datasets and services
ISA2	Interoperability solutions for public administrations, businesses and citizens Programme
ISO	International Organization for Standardization
JRC	Joint Research Centre
LiDAR	Light Detection and Ranging

LOD	Level of Detail
NL	The Netherlands
OGC	Open Geospatial Consortium
OSM	OpenStreetMap
PNOA	Plan Nacional de Ortofotografía Aérea
UML	Unified Modeling Language
WFS	Web Feature Service
WMS	Web Map Service
XML	eXtensible Markup Language

List of figures

List of tables

KJ-NA-30963-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/746342

ISBN 978-92-76-46608-6