Lessons learnt from the integration of open data and semantic 3D city models for urban building energy modelling in the Netherlands

Camilo León-Sánchez[1], Giorgio Agugiaro[1], Jantien Stoter[1]

[1] 3D geoinformation group, Delft University of Technology, Faculty of Architecture and the Built Environment,
Department of Urbanism, Julianalaan 134, 2628BL Delft, The Netherlands,
c.a.leonsanchez,g.agugiaro,j.e.stoter@tudelft.nl

Keywords: open data,CityGML, Energy ADE, UBEM

Abstract

This paper presents the lessons learnt from the integration of open datasets in the Netherlands for the creation of a country-wide enriched semantic 3D city model for urban building energy modelling. Although the Netherlands provides open access to building data up to the dwelling level, several challenges still remain related to data fragmentation, inconsistency, and incompleteness. The resulting dataset uses the CityGML with the Energy ADE data model since they offer a robust framework for integrating geospatial and non-geospatial data for energy applications. Our research highlights the need for significant preprocessing, harmonisation pipelines, and enrichment strategies to address gaps in data completeness and reliability. Finally, we identify critical missing data (e.g., renovation history, thermal zoning, and detailed HVAC specifications) and propose directions for improvement.

1. Introduction

The current demographic trend of the planet's population of moving from rural to urban areas highlights the necessity to accurately quantify both current and future trends in the energy performance of buildings to support adequate energy supply planning, and evidence-based policymaking. Urban Building Energy Modelling (UBEM) provides a computational framework to address these needs by enabling large-scale simulation and analysis of building energy performance (Davila et al., 2016).

UBEM refers to the techniques, methods, and software tools used to simulate the energy performance of buildings at several scales (from local to regional levels). UBEM follows two approaches: top-down and bottom-up. The top-down approach treats the building sector as a single unit, offering aggregate estimates without capturing differences among individual buildings. In contrast, the bottom-up approach models the energy performance of buildings based on their physical characteristics. This results in more detailed energy modelling but also requires a lot of data (on the buildings, but also on the energy supply infrastructure, weather, etc.). As a result, bottom-up UBEM often relies on building stock models and detailed representations of individual structures, usually based on building archetypes.

The practical application of bottom-up UBEM is heavily dependent on data. Detailed information on construction methods, materials, and occupancy patterns is essential, but such data are often fragmented, incomplete, or unavailable. When accessible, these data typically originate from multiple sources, necessitating harmonisation before they can be used in simulation workflows. Additionally, researchers have developed machine learning techniques to infer missing attributes and improve the completeness of building datasets (Seyedzadeh et al., 2019, Wang et al., 2020).

Over the past 15 years, semantic 3D city models (s3DCMs) have emerged as a key asset in supporting bottom-up UBEM. These models represent both the geometry and semantic properties of city objects in a structured three-dimensional environment. By incorporating building 2D (or, better: 3D) geometries, volumes, usage types, and renewable energy availability, s3DCMs can enhance the precision and scalability of energy simulations.

This paper presents the lessons learnt from the integration of national open datasets in the Netherlands and evaluates the potential of the s3DCM data model as a unifying platform for UBEM. Despite the availability of extensive open data, significant preprocessing is required to align these resources with the requirements of simulation workflows. Therefore, the Dutch context can serve as a useful case study to reflect on opportunities and limitations in open data-driven energy modelling.

2. Open data and semantic 3D city models

UBEM requires a lot of coherent spatial and non-spatial data. However, the characteristics of the data may vary depending on the location, the scope, and the method used for their collection. Ideally, UBEM applications would have access to complete information on building characteristics (e.g., cadastral, geometric, and physical properties), local climate data, energy consumption records, and occupant behaviour. However, in practice, such ideal conditions are rarely met. As shown in figure 1, data can be classified into 2 main groups: geospatial (building geometries and climate) and non-geospatial (construction materials, energy-related data – systems, consumption, labels). Figure 1 relies on a similar reasoning as in (Wang et al., 2022) when it comes to the classification of the data acquisition methods for UBEM, although their review article does not look at the implementation of UBEM at the national level.

2.1 Geospatial data

2.1.1 Building data Semantic 3D city models (s3DCMs) have become one of the most prominent and convenient sources of geospatial data of buildings for UBEM. These models, coming from the "GIS world", are typically based on the boundary

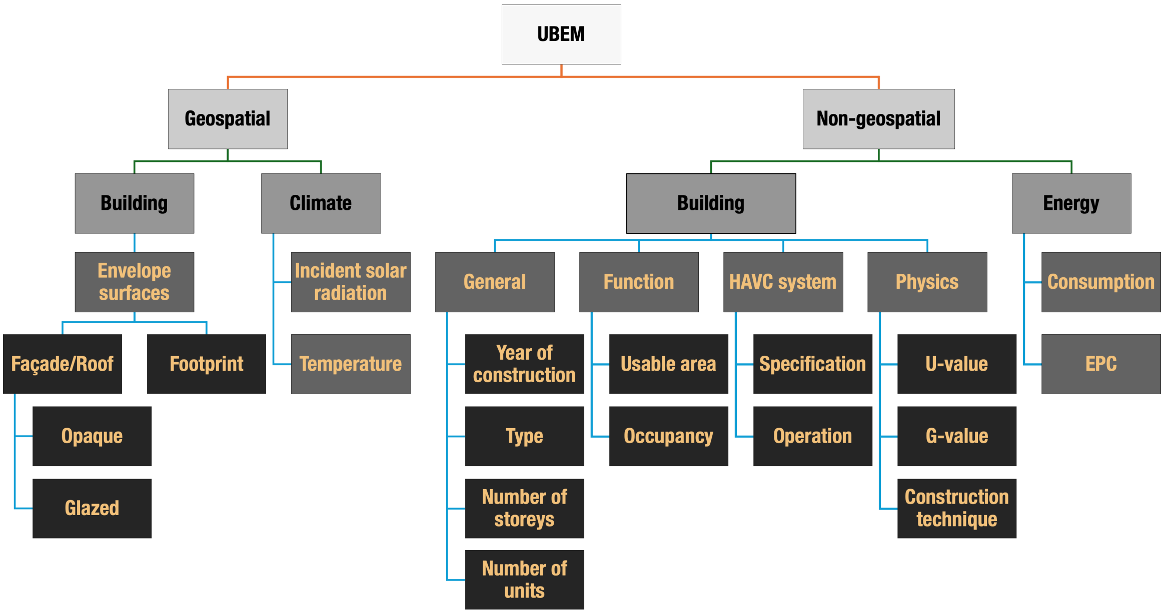


Figure 1. Hierarchical overview of UBEM data requirements. The darker the box, the more detailed and fine-grained the corresponding data.

representation (BRep) for the geometries of all relevant urban objects, therefore making buildings particularly well-suited for applications such as solar potential studies or physics-based energy performance simulations of buildings. Additionally, s3DCM may include additional relevant data for UBEM. One example is represented by the explicit geometrical modelling of party walls (i.e. shared walls between adjacent buildings). However, not all s3DCMs are a perfect fit for UBEM since the detail of the geometrical modelling of the building (Level of Detail, or, in short, LoD) might not represent relevant data. That is the case for opaque and glazed areas of the thermal envelope of a building, the latter consisting of openings such as windows and doors. In such a case, a window-to-wall surface ratio is required instead to approximate an exact distinction.

2.2 Climate data

When it comes to climate data, two main climate parameters are crucial for UBEM: the incident solar irradiation on the thermal envelope surfaces and the outside temperature, which influence the heating and cooling loads. These data can be obtained from weather stations and local sensors. Usually, national meteorological agencies offer these data based on observations collected through ground-based weather stations. When performing simulations, long-term averaged datasets such as typical meteorological year (TMY) files can be used. These datasets provide standardised climate values based on multi-year observations and are suitable for building energy simulation (BES) tools. An example of this type of dataset is the climate.onebuilding.org repository (Lawrie and Crawley, 2023).

2.3 Non-geospatial data

2.3.1 Building data Non-geometric building data include characteristics that are independent or do not vary based on the geometric representation of buildings. These include:

- Building function (e.g., residential, office) influences internal heat gains and operational schedules of a building.
- Occupancy profiles reflect human activity patterns, such as working or opening hours in offices or schools.
- Heating, ventilation, and air conditioning (HVAC) systems, including types of heating and cooling technologies.
- Number of storeys, number of living units, building type (e.g., single-family home, apartment block), and year of construction, which strongly influence energy performance due to evolving construction practices and regulatory standards (Neufert et al., 2021, Rijksoverheid, 2024).

2.3.2 Physics-related data Understanding the thermal behaviour of buildings requires detailed knowledge of construction techniques and material properties. Relevant parameters include:

- Thermal transmittance (U-value) of walls, roofs, and ground floors.
- Solar energy transmittance (g-value) of windows.
- Air infiltration rates, surface reflectance, and material colour.

These attributes directly affect heating and cooling demands and are fundamental to accurate energy modelling.

2.3.3 Energy data Energy-related data serve two prominent roles: as input for data-driven UBEM approaches and as validation for simulation results. These data may be theoretical (e.g., energy performance certificate EPC values) or measured (e.g., electricity and gas consumption). The main difference is that the former tries to exclude the influence of human behaviour by using standard values per m^2, while the latter is directly dependent on it. Due to privacy regulations, detailed consumption records are, however, rarely available at the building level (or at finer resolution, e.g. at dwelling level), prompting the use of aggregated data at the postcode or municipal level.

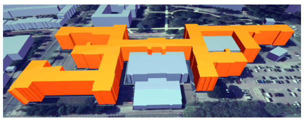
2.4 Data availability

In the case of the Netherlands, key open datasets include:

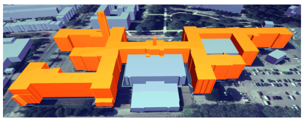
2.4.1 Building data

- BAG (Basisregistratie Adressen en Gebouwen): the Dutch building and address registry. It provides the year of construction, number of living units, and building identifiers. Available as open data from Kadaster (Kadaster, 2024).
- BAG-plus is an enhanced version that is currently available as open data only in Amsterdam. It adds attributes such as building type (single vs. multi-family) and number of storeys (Gemeente Amsterdam, 2022) to the existing data model of the BAG. As of April 2025, it includes data on 197,808 buildings and 568,843 dwellings.
- 3DBAG, a country-wide LoD2 semantic 3D model of all buildings in the Netherlands (Peters et al., 2022), is generated periodically by combining BAG with the national elevation model (AHN) (Rijkswaterstaat, 2024). An example of all geometric representations available in the 3DBAG for the Faculty of Architecture and the Built Environment of the Delft University of Technology is shown in figure 2.

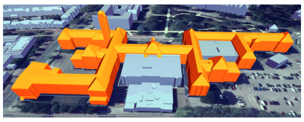
(a) 3D mesh ©2024 Apple Inc.



(b) 3DBAG LoD1.2



(c) 3DBAG LoD1.3



(d) 3DBAG LoD2.2

Figure 2. Several 3D representations of the TU Delft Architecture Faculty building (BK)

2.4.2 Physics-related data No open dataset currently provides building-level physical properties. Instead, archetype-based datasets are used. That is the case of TABULA (EPISCOPE Project, 2017) and the ”Report on example buildings” (in Dutch, “Voorbeeldwoningen”) (RVO, 2023). One of the products of the TABULA project was the definition of building typologies across Europe. The four main building types defined for the participant countries are Single-Family House (SFH), Terraced House (TH), Multi-Family House (MFH), and Apartment Block (in Dutch ”Flatwoning”) (FW).

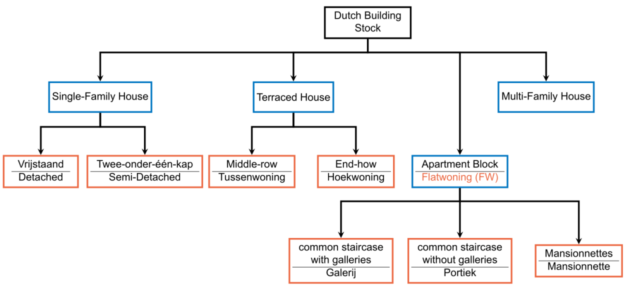


Figure 3. Dutch building types according to TABULA. Blue boxes indicate generic types; Orange boxes indicate the Dutch specific types

This classification was further extended in the case of the Netherlands with six additional types: detached (in Dutch ”Vrijstaand”, VW), semi-detached house (in Dutch ”Twee-onder-één-kap”, TOEK), middle-row (in Dutch ”Tussenwoning”, TW), end-row (in Dutch ”Hoekwoning”, HW), common staircase without galleries (in Dutch ”Portiek”, PW), and common staircase with galleries (in Dutch ”Galerij”, GW), and mansionnette (MW). The hierarchy of the building types in the Netherlands is shown in figure 3. Further details of each of the building types are available in the above-mentioned references.

Both TABULA and ”Report on example buildings” provide standardised heat transmittance coefficient values (u-values) of the different elements that compose the building's thermal envelope, such as walls, roofs and ground surfaces, as well as the solar energy transmittance of windows (g-value).

2.4.3 Energy data EP-online is the official EPC database in the Netherlands (Rijksoverheid, 2025). It provides energy performance indicators of building units through APIs and downloadable files (XML, CSV, XLSX). As of March 2025, it contains 5.5 million records representing 1,200,000 buildings. The dataset includes 42 attributes, but the available data vary based on the computation method used. Therefore, some attributes may be empty when not required by the computation method. For example, the EnergieIndex is not used in the Dutch Technical Agreement (in Dutch: ”Nederlands Technische Afspraak” or, in short, NTA8800), which is the current official method to compute the energy performance in the Netherlands. Although the main purpose of this dataset is to store EPC-related data, it offers an additional second purpose as it contains the building type and the building subtype in the case of apartment blocks based on the living unit.

2.4.4 Weather data The KNMI (Royal Netherlands Meteorological Institute) operates 48 automatic weather stations distributed over the whole Netherlands (KNMI, 2024a). These stations record hourly observations for temperature, wind, humidity, and global solar radiation (KNMI, 2024b). For more detailed radiation modelling (diffuse, direct, solar angles), it offers typical meteorological year (TMY) datasets for each of the weather stations, which are based on at least 30 years of observed data (Klement, 2024). The two datasets (hourly records and typical values) are offered as open datasets.

A summary of the open datasets available in the Netherlands and used in this research is shown in figure 4, the figure follows the same colour scheme as in figure 1. The lowest level (the darkest colour) indicates the attribute used from the corresponding dataset (mentioned at the fourth level).

2.5 Data model

We use the CityGML standard when referring to a s3DCM from the geospatial perspective. In this standard, all concepts are based on the ontological definition of a city. Therefore, there is spatio-semantic consistency across all elements defined by the data model. In our research, we use CityGML 2.0 (Gröger et al., 2012) due to the software support and the feasibility of extending the data model of the standard by means of Application Domain Extensions (ADE). In our case, we use the Energy ADE to support energy-related data. The implemented data pipeline is shown in figure 5.

https://doi.org/10.5194/isprs-annals-X-4-W7-2025-89-2025 | © Author(s) 2025. CC BY 4.0 License.

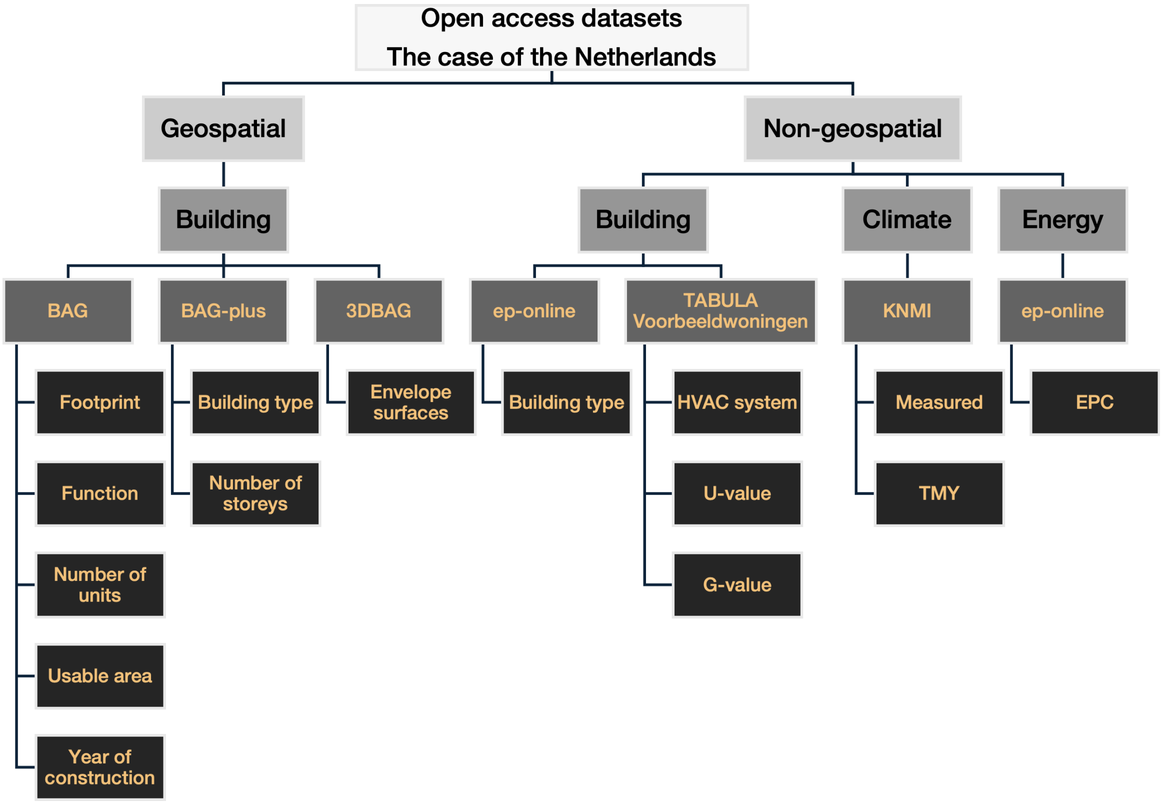


Figure 4. Graphical summary of the open datasets available in the Netherlands.

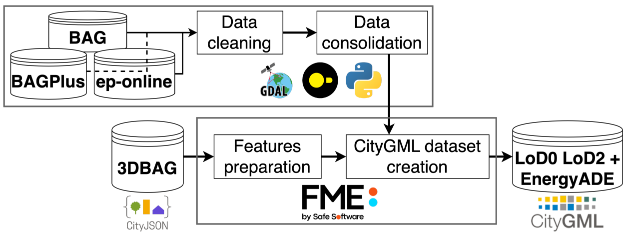


Figure 5. Data pipeline for the creation of the output dataset

The Energy ADE provides a standardised data model to allow single-building energy simulations and country-wide energy assessments focused on the building sector (Agugiaro et al., 2018). However, we are currently testing the updated data model of the Energy ADE 2.0 (Agugiaro and Padsala, 2025). The Energy ADE 2.0 builds upon the Energy ADE, it incorporates the know-how gathered from experiences and feedback collected since 2018 when the Energy ADE was first released.

3. Challenges in using open data for energy modelling

Despite the availability of extensive open data in the Netherlands, several challenges hinder their direct application in UBEM. As discussed in Section 2, relevant datasets are distributed across multiple sources, each one using distinct data schemas. For instance, the 3DBAG dataset uses a variety of data formats (CityJSON, OBJ, GPKG), and it does not have a fixed schema, which has evolved since its first release in 2021 with the addition or modification of attributes. For example, the first version of the 3DBAG has 30 attributes while version 2024.04 has 45 attributes. Similarly, ep-online for energy labels adopts its data formats and model, resulting in a fragmented landscape that requires harmonisation prior to use.

The Netherlands has done remarkable work in the context of open access data and the standardisation of unique identifiers for buildings, living units, addresses, and the links between them. However, issues and ambiguities persist. For example, several living units are registered in multiple buildings—46,257 out of 10,928,296 units in the BAG dataset are registered in multiple buildings, with cases of living units registered in up to eleven buildings. A related issue arises from ep-online, where 3,743,762 of the 5,605,189 EPC records lack a corresponding building ID. These problematic cases require therefore additional spatial processing steps, such as spatial joins, to establish accurate relationships between datasets.

3.1 Missing or incomplete data

Currently, ep-online covers approximately 11% of the national building stock (1,250,078 out of 10,928,296), and many entries are already outdated. In the Netherlands, the energy performance certification (EPC) of buildings is not automatically generated, but manually computed during specific cases, such as property sales or rentals. Several records were generated using former official methods that were replaced by the NTA8800 standard in 2020 (NEN, 2020). Since all records are stored within a unified data model, newer attributes have been simply appended as additional columns rather than creating separate datasets. For example, older records include the ”energy index” attribute. In contrast, more recent records include calculated energy demand values and heating and cooling demand values in detail. Regardless of the computation method, all entries include an energy label and the dwelling id. These examples limit the consistency and usability of the data, affecting its reliability. Additionally, due to privacy regulations in the Netherlands, actual energy consumption data are only available in aggregated form at the postcode level (CBS, 2023), which restricts access to detailed building-level information.

In the case of geometrical representation of buildings in the 3DBAG, this dataset does not provide detailed façade-level information of buildings such as window and door size and location (essentially, a LoD3 according to CityGML), leading to assumptions such as window-to-wall ratios to overcome this data gap. Another example of incomplete data in the BAGis the livings unit information, where 41% of the buildings do not contain records about it and, therefore, no data are available about their function. This issue limits the applicability of our dataset on UBEM workflows which depend on building usage, such as the computation of the energy demand of buildings based on the energy balance method according to the NTA8800. The spatial distribution of this case is shown in figure 6.

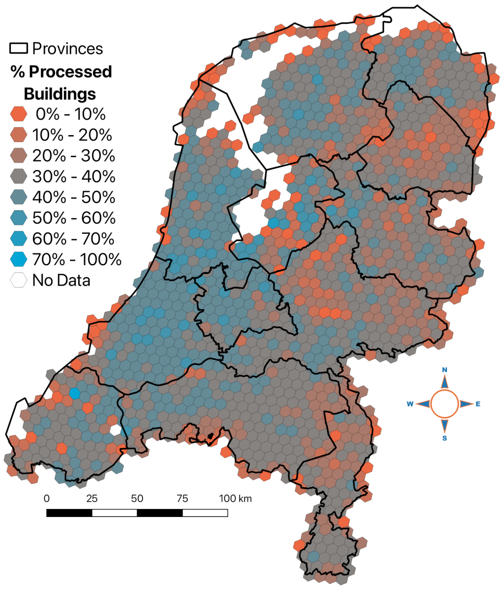


Figure 6. Spatial distribution of buildings without living units information

3.2 Data quality issues

Another challenge corresponds to data quality issues. Each dataset requires independent data cleaning before it can be used as input for UBEM, followed by consistency checks after integration. Since some attributes are present in multiple datasets—sometimes with differing values—prioritisation rules must be defined. For example, both 3DBAG and BAG-Plus (in Amsterdam) provide the number of storeys. However, 3DBAG's values are generated through a machine learning model (Roy et al., 2023), whereas BAG-Plus is based on municipality records. For this case, during the data fusion process, we assign BAG-Plus higher precedence due to its likely higher reliability, however, this applies only to the buildings of Amsterdam, as the BAG-Plus is not available elsewhere.

Furthermore, the machine learning method used in the 3DBAG to infer the number of storeys per building achieves an accuracy of 94.5% for buildings up to five storeys but drops to 52.3% for taller buildings (Roy et al., 2023). Therefore, the 3DBAG does not include this attribute for buildings with more than five storeys. However, in our data analyses we have found high discrepancy errors regarding this attribute, as shown in figure 7, complemented by table 1:the building is assigned 1 storey, but, in reality, it has 5. Additional checks should be implemented to highlight possible errors in the input datasets. Another limitation of the implemented model to infer the number of storeys is the partial ones, such as mezzanines or "half-floors" which are common in the Netherlands.

Figure 7. Google street view for building with Pand ID 0599100000672213

Table 1. Summary data for Pand ID 0599100000756485

Footprint area	Number storeys	3DBAG height	3DBAG volume
$58.8m^2$	1	15.56m	$977.3m^{32}$

In the Netherlands, during the computation of the EPC of buildings, specialists assign the energy labels along with the corresponding building type class. However, for certain classes, the classification rules are vague and lead to misclassification. That is the case of multi-family buildings, which have multiple possibilities based on factors such as the number of storeys of the living unit, the type of entrance to the building or the access corridors to the living units. Therefore, in the ep-online dataset, a single building may contain multiple living units with different categories assigned to each unit. That is the case shown in figure 8, where units within the same building are labelled with different building types.

Figure 8. Examples of buildings in Rotterdam being assigned multiple, different classes of building type

To overcome the issue with the building type data from ep-online, (Poon, 2024) investigated the use of 3DBAG as an input dataset to classify the Dutch building stock into the building types according to TABULA. Poon's method involves attribute extraction from the BAG and 3DBAG while using ep-online as the ground truth. The classification algorithms selected are support vector machine (SVM) (Cristianini and Shawe-Taylor, 2000) and random forest (RF) (Breiman, 2001). The results are not accurate enough to be used to classify apartment blocks into the subcategories: common staircase galleries (galerij), common staircase without galleries (portiek) and Mansionnettes. For the modelling and prediction part, 80% of the data are used for training the classifiers, and the remaining 20% serves for model evaluation. The method was applied to eight case studies; the first two correspond to the municipalities of Rijssen-Holten and Delft. However, there is a severe class imbalance with four classes representing each of them less than 1% of the dataset: flatwoning 0.38%, maisonnette 0.02%, portiek 0.05% and galerij 0.03%. The additional six case studies focus on specific building types. In most of the cases, RF performs better than SVM; however, the accuracy of the classifiers varies from 51.6% to 92% based on the class and the case study. Therefore, we did not replicate these models to classify whole Dutch building stock available in the 3DBAG.

Usable area data also shows significant inconsistencies. In the BAG, 3,578 buildings have a reported usable area of just 1 m^2, while their footprint is at least 48% larger. To address this, we implemented a checker that estimates a theoretical usable area based on 80% of the footprint multiplied by the number of storeys. The 20% reduction of the footprint area accounts for non-heated spaces like stairwells or storage areas, in line with (Dochev et al., 2020, Johari et al., 2023). Buildings where the reported usable area deviates by more than 30% from the estimated value were tagged for further inspection. This tag resulted in 1,726,288 buildings (16.6%) identified as potentially inconsistent. The spatial distribution of the flagged buildings in the Netherlands is shown in figure 9.

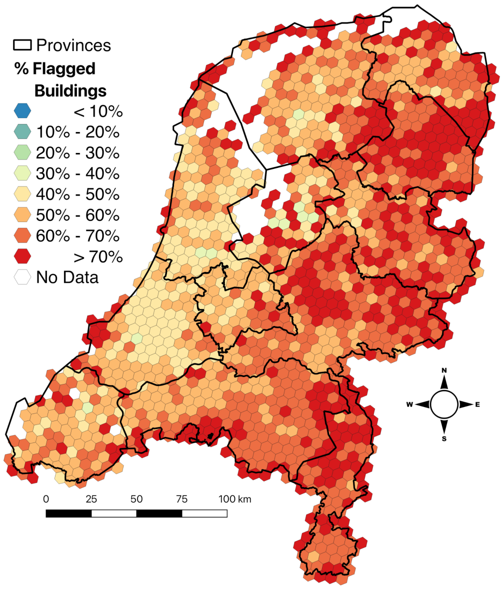


Figure 9. Spatial distribution of flagged buildings in the Netherlands

Additional inconsistencies arise when comparing values across datasets. One example is the building with ID 1742100000006419, for which the BAG lists a usable area of 5,835 m^2—27 times greater than its footprint extracted from its geometric representation in the 3DBAG. Table 2 presents the conflicting values. Such discrepancy indicates potential data entry errors across registries. Another common issue relates to function ambiguity. In BAG, 2.6% of living units (273,118 out of 10,234,487) have multiple registered functions. Without information on how usable area is distributed among functions, assumptions are required during analysis—introducing further uncertainty.

Finally, discrepancies in temporal resolution might contribute to data misalignments. For example, the 3DBAG is generated from AHN (aerial LiDAR) data collected over multi-year periods—AHN5 will span 2023–2025. In contrast, the BAG is updated on a daily basis. This temporal mismatch can lead to inconsistencies when comparing building geometry and attributes derived from these sources.

3.3 Testing the new dataset

The resulting enriched s3DCM has been used to compute the net heat demand of buildings based on the NTA8800 (NEN, 2024). Due to the data requirements of the method, among them the building's function, only 40% (4,164,949 buildings) of the Dutch building stock could be processed. The computation uses the physics-related data from the corresponding archetype categories. However, we can only assume an as-built simulation scenario since we lack information on the history of the buildings or the renovation processes that might have taken place already. This decision leads to the overestimation of the energy demand of older buildings. This problem is indeed well-known in literature (Buckley et al., 2021, Lin et al., 2023). Basic statistical metrics of the computed net heat demand are presented in table 2. Additional details are provided in table 3, which shows the mean annual net heat demand by building type. These data are complemented by figure 10, which displays a letter-value plot of the computed net heat demand.

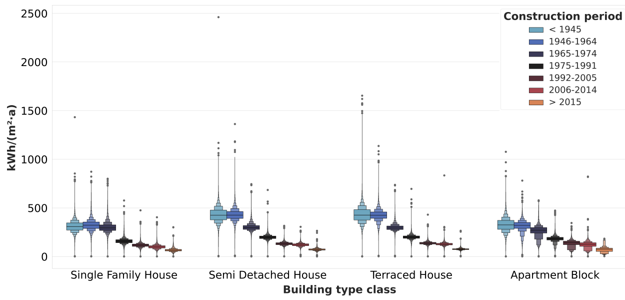


Figure 10. Letter-value plot of the computed net heat demand by building type and construction period in the Netherlands

Table 2. Computed annual net heat demand basic statistics in the Netherlands, all values are expressed in kWh/(m^2·a)

Min.	Mean	Median	Std Dev.	Max.	Mode
0.7	260.9	238.6	125.5	2,461.1	194.9

Table 3. Mean annual net heat demand per building type class in the Netherlands

Building type class*	No. Buildings	Building Stock %	kWh (m^2·a)
SFH	773,048	18.16%	222.4
TOEK	670,009	15.74%	289.2
TH	2,696,494	63.34%	265.8
FW	117,658	2.76%	241.5

*Following the classes defined in section 2.4.2

The results indicate that Terraced House category shows the highest variability of the computation results in all classes based on the period of construction of the building. This behaviour is higher for buildings built before 1945. The spatial distribution in the Netherlands of the average net heat demand is shown in figure 11.

https://doi.org/10.5194/isprs-annals-X-4-W7-2025-89-2025 | © Author(s) 2025. CC BY 4.0 License.

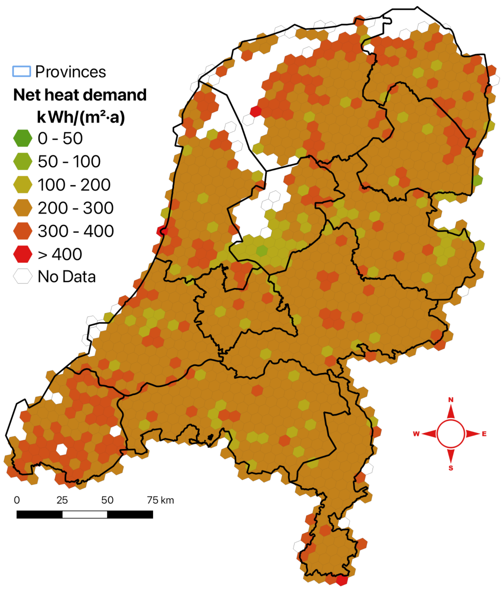


Figure 11. Computed mean net heat demand

4. Conclusion

This paper has presented the lessons learned from the integration of multiple open spatial and non-spatial datasets in the Netherlands, with the purpose of creating a national database to be used for Urban Building Energy Modelling (UBEM). Although the availability of open datasets in the country is relatively high, their direct application in UBEM is still limited by challenges related to fragmentation, inconsistent standards, and data incompleteness. Addressing some of these challenges still requires significant preprocessing pipelines and data enrichment strategies to ensure completeness and consistency, while some issues still remain unresolved. For example 40.2% of buildings in the BAG lack information on registered living units. As a result, their function cannot be determined, which prevents their inclusion in applications such as net heat demand calculations based on the NTA8800 method.

UBEM depends on a clear and coherent spatial and semantic representation of the built environment. A critical component of this is the hierarchical decomposition of space, where elements such as doors belong to walls, walls to buildings, and buildings to broader urban contexts. This hierarchy enables structured relationships between geometry and semantics, which are essential to link spatial entities with their physical and operational attributes. Without a formal data model, such as that offered by CityGML, it would neither feasible to connect geometric components to energy-related properties—nor to interpret elements unambiguously, such as recognising a wall as part of the thermal envelope.

In this regard, s3DCMs provide a strong foundation for energy modelling. When enhanced with additional data, they offer the structure and flexibility required to support bottom-up approaches to UBEM. The CityGML data model, and particularly the Energy ADE, provides a semantically rich and geometrically explicit framework tailored to UBEM requirements. Its use of unique identifiers and object hierarchies ensures consistency, for example, by indicating that a BuildingPart cannot exist without a corresponding Building and by linking geometry to attributes such as usage type or energy demand.

However, the creation of s3DCM suitable for UBEM requires rigorous data quality management throughout the data pipeline. This includes input control, intermediate consistency checks during the integration and, when possible, verification against other sources to detect conflicts. Data heterogeneity across formats and scales further influences the importance of harmonisation mechanisms and error detection protocols.

The data requirement analysis conducted in this research also highlighted areas where currently available datasets remain insufficient. For example, energy simulations involving solar irradiance require an additional 3D context, including vegetation, terrain morphology, and surface reflectivity (albedo). Heating demand simulations depend on detailed properties of the building envelope (construction materials, thicknesses, U-values, g-values), specifications of HVAC systems, occupancy patterns, and information on energy sources. Crucially, the absence of open data on building renovation history constrains the ability to model both the current performance of the building stock and future refurbishment scenarios at a high level of accuracy. Future efforts should focus on enhancing the semantic and temporal granularity of open data and improving standardisation across datasets. This research could be further extended in several ways. One possibility is the development of robust methods for identifying of data inconsistencies. Other possibility refers to the use of other datasets; for instance, Building Information Modelling (BIM) which might provide indoor information, although, only available for a limited number of buildings. Finally, the use of remote sensing methods to extract information of elements of the building such as materials or shape and size of doors and windows.

Acknowledgements

Part of the work presented in this paper has been developed within the European project DigiTwins4PEDs (Utilisation of urban digital twins to co-create flexible positive energy systems for districts). The project is funded by the European Commission under the Horizon Europe Partnership scheme (Funding Indicator: 03EN3081A). The DUT Call 2022 also contributes to the Urban Transition Mission of Mission Innovation as part of the MICall 2022 initiative.

References

Agugiaro, G., Benner, J., Cipriano, P., Nouvel, R., 2018. The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations. *Open Geospatial Data, Software and Standards*, 3(1).

Agugiaro, G., Padsala, R., 2025. A proposal to update and enhance the CityGML Energy Application Domain Extension. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), 5–32.

Buckley, N., Mills, G., Reinhart, C., Berzolla, Z. M., 2021. Using urban building energy modelling (UBEM) to support the new European Union's Green Deal: Case study of Dublin Ireland. *Energy and Buildings*, 247, 111115.

CBS, 2023. Energielevering aan woningen en bedrijven naar postcode. Last Modified: 2023-11-15T09:00:00+01:00.

Cristianini, N., Shawe-Taylor, J., 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.

Davila, C. C., Reinhart, C. F., Bemis, J. L., 2016. Modelling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. 117, 237–250.

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

EPISCOPE Project, 2017. IEE project TABULA webtool. https://webtool.building-typology.eu.

Gemeente Amsterdam, 2022. BAG-plus. https://api.data.amsterdam.nl/v1/wfs/bag/.

Gröger, G., Kolbe, T., Nagel, C., Häfele, K.-H., 2012. OGC city geography markup language (CityGML) encoding standard. Open Geospatial Consortium. https://www.opengeospatial.org/standards/citygml. ISSN:18632246.

Johari, F., Shadram, F., Widén, J., 2023. Urban building energy modeling from geo-referenced energy performance certificate data: Development, calibration, and validation. *Sustainable Cities and Society*, 96, 104664.

Kadaster, 2024. Over Basisregistratie Adressen en Gebouwen - BAG. https://www.kadaster.nl/zakelijk/registraties/basisregistraties/bag/over-bag.

Klement, M., 2024. Meteorological data. The Dutch PV portal. https://www.tudelft.nl/en/ewi/over-de-faculteit/afdelingen/electrical-sustainable-energy/photovoltaic-materials-and-devices/dutch-pv-portal/meteorological-data.

KNMI, 2024a. Automatic weather stations. https://www.knmi.nl/kennis-en-datacentrum/uitleg/automatische-weerstations.

KNMI, 2024b. Uurgegevens van het weer in nederland. klimatologie. https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens.

Lawrie, L. K., Crawley, D. B., 2023. Development of global typical meteorological years (TMYx). Publication Title: http://climate.onebuilding.org.

Lin, Z., Hong, T., Xu, X., Chen, J., Wang, W., 2023. Evaluating energy retrofits of historic buildings in a university campus using an urban building energy model that considers uncertainties. *Sustainable Cities and Society*, 95, 104602.

NEN, 2020. NTA 8800:2020 nl. Royal Netherlands Standardization Institute. https://www.nen.nl/nta-8800-2020-nl-272896.

NEN, 2024. NTA 8800:2024 nl. Royal Netherlands Standardization Institute. https://www.nen.nl/nta-8800-2024-nl-320123.

Neufert, E., Kister, J., Lohmann, M., Merkel, P., Brockhaus, M., 2021. *Bauentwurfslehre*. 43 edn, Im Bauwelt Verlag.

Peters, R., Dukai, B., Vitalis, S., van Liempt, J., Stoter, J., 2022. Automated 3D Reconstruction of LoD2 and LoD1 Models for All 10 Million Buildings of the Netherlands. *Photogrammetric Engineering and Remote Sensing*, 88(3), 165–170. eprint: 2201.01191.

Poon, H.-K. C., 2024. Inferring the residential building type from 3dbag.

Rijksoverheid, 2024. Bouwbesluit. Besluit bouwwerken leefomgeving. https://wetten.overheid.nl/BWBR0041297/2024-01-01.

Rijksoverheid, 2025. EP-online. Rijksdienst voor Ondernemend. https://www.ep-online.nl/. Publication Title: Resources, tools and inspiration for buildings.

Rijkswaterstaat, 2024. Dataset: Actueel hoogtebestand nederland (AHN). https://www.pdok.nl/introductie/-/article/actueel-hoogtebestand-nederland-ahn.

Roy, E., Pronk, M., Agugiaro, G., Ledoux, H., 2023. Inferring the number of floors for residential buildings. *International Journal of Geographical Information Science*, 37(4), 938–962.

RVO, 2023. Voorbeeldwoningen 2022 bestaande bouw. Rijksdienst voor Ondernemend. https://www.rvo.nl/onderwerpen/wetten-en-regels-gebouwen/voorbeeldwoningen-bestaande-bouw. Rijksdienst voor Ondernemend Nederland.

Seyedzadeh, S., Pour Rahimian, F., Rastogi, P., Glesk, I., 2019. Tuning machine learning models for prediction of building energy loads. 47, 101484.

Wang, C., Ferrando, M., Causone, F., Jin, X., Zhou, X., Shi, X., 2022. Data acquisition for urban building energy modeling: A review. 217, 109056.

Wang, Z., Hong, T., Piette, M. A., 2020. Building thermal load prediction through shallow machine learning and deep learning. 263, 114683. Publisher: Elsevier.