

Enhancing urban energy applications through semantic 3D city models and open data

The case of the Netherlands

Camilo Alexander León Sánchez

Enhancing urban energy applications through semantic 3D city models and open data

The case of the Netherlands

Enhancing urban energy applications through semantic 3D city models and open data

The case of the Netherlands

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,

chair of the Board for Doctorates

to be defended publicly on

Monday 06, October 2025 at 12:30 o'clock

by

Camilo Alexander LEÓN SÁNCHEZ

Master of Science in Information and Communication Sciences,
Universidad Distrital Francisco José de Caldas, Colombia

Master of Science in Geodesy and Geoinformation Science,
Technische Universität Berlin, Germany

born in Bogotá, Colombia

This dissertation has been approved by the promotor and copromotor.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. J.E. Stoter,	Delft University of Technology, <i>promotor</i>
Dr. G. Agugiaro,	Delft University of Technology, <i>copromotor</i>

Independent members:

Prof. dr. V. Coors,	Hochschule Für Technik Stuttgart
Dr. J.H. Kämpf,	Idiap Research Institute
Prof. dr. ir. H.J. Visscher,	Delft University of Technology
Prof. dr. L.C.M. Itard,	Delft University of Technology

Other members:

Prof. dr. A. van Timmeren, Delft University of Technology



Keywords: UBEM, NTA 8800, open data, s3DCM, CityGML, Energy ADE

Printed by: Ridderprint | <https://www.ridderprint.nl>

Cover by: Luz Adriana Sarmiento Oliveros, Ink on paper & María José León Garnica, digital design.

Copyright © 2025 by C.A. León Sánchez (ORCID [0000-0002-9696-7229](https://orcid.org/0000-0002-9696-7229))

ISBN 978-94-6518-115-8

An electronic copy of this dissertation is available at
<https://repository.tudelft.nl/>.

to Adriana

CONTENTS

Summary	xi
Samenvatting	xiii
Resumen	xv
Acknowledgements	xvii
1 Introduction	1
1.1 Motivation	1
1.2 Problem statement	3
1.3 Research objective and scope	5
1.4 Outline of the thesis	6
1.4.1 Structure of the thesis	6
1.4.2 Personal pronouns	8
2 Theoretical Framework–Literature Review	9
2.1 Generalities of Urban Building Energy Modelling	9
2.1.1 Physics-based models	12
2.1.2 Data-driven models	13
2.1.3 Reduced-order models	15
2.1.4 Current challenges of UBEM	17
2.2 Data Sources for UBEM	19
2.2.1 Geospatial data: Semantic 3D city models	19
2.2.2 Non-geospatial data	29
2.3 Data Quality	33
2.4 Energy Balance Method	34
2.5 NTA 8800	37
2.5.1 Introduction	37
2.5.2 Net heat demand calculation	38
2.5.3 Total heat transfer for heating	40
2.5.4 Total heat gains for heating	47
2.5.5 Heat gains through incident solar radiation	51
2.5.6 Effective internal heat capacity of a zone	57
2.5.7 Utilisation factors	59
2.5.8 Hot tap water	61
2.6 Existing Dutch models	62
2.7 Conclusion remarks	64

3	Aligning UBE data requirements for national s3DCM	67
3.1	General data considerations	67
3.1.1	Geospatial Data	68
3.1.2	Non-Geospatial Data	70
3.2	Data availability	71
3.2.1	Geospatial Data	71
3.2.2	Non-geospatial Data	78
3.3	Conclusion remarks	84
4	Data pipeline design	85
4.1	Testbed creation	86
4.2	Databases creation	90
4.2.1	Building Physics data	90
4.2.2	EnergyBAG DB: the creation of a dataset for the whole Netherlands	93
4.2.3	Data indexing	98
4.3	Solutions to enhance data	102
4.3.1	Solar analysis	103
4.3.2	Building type inferring using machine learning algorithms	114
4.3.3	Calculation of the number of storeys per building	119
4.4	Conclusion remarks	120
5	Implementation	123
5.1	General considerations	123
5.2	Conceptual decisions	125
5.2.1	C4-model system context	125
5.2.2	C4-model system container	126
5.3	Implementation of the energy simulation system	129
5.4	Challenges in the implementation phase	132
5.5	Implementation decisions	133
5.5.1	Modelling	133
5.5.2	Implementation workflow	136
5.6	Focused review: specific building cases	138
5.6.1	Pand ID 1742100000100986	139
5.6.2	Pand ID 1742100000006419	140
5.6.3	Pand ID 1742100000096311	141
5.6.4	Pand ID 0150100000010860	142
5.6.5	Additional inconsistencies	143
5.6.6	Lessons learnt	144
5.7	Conclusion remarks	148
6	Results and Discussion	151
6.1	Case study: Rijssen-Holten	152
6.1.1	Processed data	152
6.1.2	Computed net heat demand	154
6.1.3	Analysis of specific cases	160

6.1.4	Results comparison	165
6.1.5	Thermal hull refurbishment scenario	168
6.2	The Netherlands	171
6.2.1	Computed net heat demand	173
6.2.2	Analysis of specific cases	178
6.2.3	Results comparison	181
6.2.4	Thermal hull refurbishment scenario	183
6.3	Conclusion remarks	185
7	Conclusions, reflections and future research	187
7.1	Conclusions	188
7.2	Reflections, limitations and contributions	192
7.2.1	Limitations	193
7.2.2	Contributions	194
7.3	Future work	195
	Bibliography	197
	List of Publications	211
	Curriculum Vitæ	213

SUMMARY

The development of society has led to a dramatic change in urban areas. As of 2025, at least 56% of the world's population lives in cities, and it is projected to reach 70% by 2050. Despite occupying only 3% of the Earth's surface, urban areas account for 60% to 80% of global "energy consumption". This intensifies the need for accurate and reliable energy demand models to support carbon reduction and energy transition goals.

Urban Building Energy Modelling (UBEM) provides a structured framework for simulating building energy performance at multiple spatial scales. However, UBEM depends heavily on detailed and high-quality data, which is often fragmented or unavailable as open data. Semantic 3D city models (s3DCMs) are one promising data source. These models offer standardised geometric and semantic representations of urban elements in a three-dimensional environment. This thesis investigates the use of s3DCMs and open data to enhance urban energy applications, focusing on the Netherlands as a case study.

The first part of the thesis addresses the models and data requirements of UBEM, with an emphasis on the Dutch official method for calculating energy performance. It evaluates CityGML as a data model for energy-related applications and analyses the availability and suitability of open datasets in the Netherlands for UBEM use.

The second part focuses on the implementation of the corresponding datasets and simulation solutions to compute the energy performance of buildings. It describes the input data sources, their entities, and the relevant attributes, as well as the enrichment of the s3DCM by linking multiple datasets. A CityGML-based testbed for energy-related applications was published as part of this work, representing the municipality of Rijssen-Holten with Buildings, trees and a digital terrain model (DTM). The enriched s3DCM has been used to perform solar analysis.

Subsequently, the thesis outlines the design and implementation of a building energy simulation (BES) solution for computing the net heat demand of buildings. Due to data limitations at country level, the focus remains on net heat demand rather than full primary energy demand. Required parameters for primary energy calculation were unavailable without introducing additional assumptions.

The simulation results cover two case studies: the municipality of Rijssen-Holten and the national building stock of the Netherlands. Outputs are classified by building type and construction period and compared against available Energy Performance Certificate (EPC) data. Although the comparison must be interpreted with caution,

it offers a contextual benchmark for the results.

This thesis highlights the value of open data and the procedures required to enhance existing s3DCM for energy-related use. It also proposes additional research directions, including the integration of solar energy simulation results with the created BES to implement hybrid approaches that better reflect the current characteristics of the Dutch building stock.

SAMENVATTING

De ontwikkeling van de samenleving heeft geleid tot dramatische veranderingen in stedelijke gebieden. In 2025 wonen minstens 56% van de wereldbevolking in steden, en naar verwachting zal dit aantal stijgen tot 70% in 2050. Deze stedelijke gebieden beslaan slechts 3% van het aardoppervlak, maar zijn verantwoordelijk voor 60-80% van het energieverbruik. Dit verbruik onderstreept de kritieke behoefte aan nauwkeurige modellering van de energievraag op basis van betrouwbare modellen, zodat aan toekomstige energiebehoeften en CO₂-reductiedoelstellingen voldaan kan worden.

Urban Building Energy Modelling (UBEM) biedt een omvattend kader om te voldoen aan de behoefte aan betrouwbare modellen door de energieprestaties van gebouwen op verschillende analyseschalen te simuleren. Toch is UBEM sterk afhankelijk van gedetailleerde, hoogwaardige gegevens—een uitdaging die wordt versterkt door de beperkte beschikbaarheid en fragmentatie van open databronnen. Een van deze databronnen zijn semantische 3D-stadsmodellen (s3DCMs), die gebaseerd zijn op gestandaardiseerde gegevensmodellen en een geometrisch en semantisch rijke representatie bieden van de elementen die een stad vormen binnen een driedimensionale omgeving. Het proefschrift bespreekt de verbetering van stedelijke energietoepassingen door s3DCM en open data, met een focus op Nederland als case study.

De eerste focus van het proefschrift ligt op UBEM, inclusief de modellen en methoden ervan, en op basis daarvan op de gegevensbehoeften. Aangezien de casestudy in Nederland plaatsvindt, biedt het een overzicht van de officiële methode om de energieprestatie van gebouwen te berekenen vanuit een gegevensperspectief. Het onderzoek omvat tevens een analyse van CityGML als gegevensmodel voor het beheer van energiegerelateerde stadsgegevens. Daarnaast presenteert het onderzoek een analyse van de in Nederland beschikbare open data die geschikt is voor UBEM.

De tweede focus van het proefschrift ligt op de implementatie van de benodigde datasets en simulatieoplossingen voor het berekenen van de energievraag van gebouwen. Het onderzoek beschrijft het proces van het creëren van de datasets die later in het proefschrift worden ingezet voor diverse energietoepassingen. De beschrijving omvat zowel de specificatie van de gegevensbronnen, hun entiteiten en relevante attributen als de methoden voor het verbinden van de verschillende datasets en het vaststellen van de kenmerken van het verrijkte s3DCM die uit dit proces voortvloeien. Als onderdeel van dit proces is een op CityGML gebaseerd

3D-stadsmodel, dat als testbed voor energiegerelateerde toepassingen dient, gepubliceerd. Dit testbed is een handmatig verrijkt CityGML-model van de gemeente Rijssen-Holten, met gegevens over gebouwen, bomen en een digitaal terreinmodel (DTM). Daarnaast presenteert het onderzoek het gebruik van het verrijkte s3DCM als invoergegevensset voor het uitvoeren van zonne-analyses.

Vervolgens presenteert het proefschrift de overwegingen en beslissingen die zijn gemaakt voor het creëren van een energie gebouwsimulatie (BES)-softwareoplossing die de energievraag van gebouwen berekent. Het proefschrift richt zich op de nettowarmtebehoefte in plaats van berekeningen van de primaire energievraag vanwege gegevensbeperkingen op nationaal niveau. Verschillende parameters die vereist zijn voor de berekening van de primaire energie zijn niet openbaar beschikbaar, en het overbruggen van deze gegevenskloof zou verdere aannames vereisen.

Het onderzoek presenteert de resultaten van de berekening van de nettowarmtebehoefte voor twee casestudies. De eerste casestudy betreft de gemeente Rijssen-Holten, het studiegebied van het gepubliceerde testbed, en de tweede betreft heel Nederland. De resultaten zijn geclassificeerd per gebouwtype en bouwperiode en worden vergeleken met de beschikbare energieprestatiecertificaat (EPC)-gegevens in het land. Hoewel de uitkomsten van deze vergelijking met de nodige voorzichtigheid moeten worden geïnterpreteerd, is dit gedaan om context te bieden ten opzichte van de officiële energiegegevens per gebouw in Nederland.

Het proefschrift benadrukt het belang van open data en de stappen die nodig zijn om bestaande s3DCM's geschikt te maken voor energietoepassingen. Daarnaast wordt aanvullend onderzoek en toekomstig werk voorgesteld, zoals de integratie van zonne-energiesimulatiere resultaten met de ontwikkelde BES, om hybride methoden te implementeren die de huidige kenmerken van het Nederlandse gebouwenbestand beter weerspiegelen.

RESUMEN

El desarrollo de la sociedad ha transformado profundamente las zonas urbanas. En el año 2025, al menos el 56% de la población mundial vive en ciudades, porcentaje que podría alcanzar el 70% para el 2050. Aunque ocupan sólo el 3% de la superficie del planeta, concentran entre el 60% y 80% del consumo de energético. Este contexto exige modelos fiables que estimen con precisión la demanda de energía primaria y que respalden las metas de descarbonización actuales.

El modelado energético de edificios a escala urbana (UBEM, por sus siglas en inglés) ofrece un marco para esa necesidad al simular al desempeño energético de edificios en diversas escalas. No obstante, depende de datos detallados y de alta calidad, cuya obtención se ve limitada por la escasez y fragmentación de las fuentes abiertas.

Entre dichas fuentes destacan los modelos semánticos de ciudades en 3D (s3DCM). Son modelos estandarizados que representan, con riqueza geométrica y semántica, los elementos urbanos en tres dimensiones. Esta disertación explora cómo los s3DCM junto con datos abiertos mejoran las aplicaciones energéticas urbanas, con los Países Bajos como caso de estudio.

En primer lugar, se examina UBEM, sus modelos y métodos, y los consiguientes requerimientos de datos. Se presenta una visión general del método oficial neerlandés para calcular el desempeño energético de edificios, desde la perspectiva de los datos. Por lo cual, se revisan los datos abiertos disponibles en los Países Bajos pertinentes para UBEM. Esto concluye con el análisis a CityGML como modelo para gestionar información energética urbana.

En segundo lugar, se detalla la implementación de conjuntos de datos y herramientas de simulación para calcular el desempeño energético de edificios. Se describe el proceso de creación de dichos conjuntos de datos, sus fuentes, entidades y atributos relevantes. También se detalla la integración a partir de múltiples fuentes de datos y las características del s3DCM enriquecido resultante. En este marco se publicó un banco de pruebas basado en CityGML para aplicaciones energéticas, el cual consiste en un modelo enriquecido y depurado del municipio de Rijssen-Holten con edificios, árboles y un modelo digital del terreno (MDT). El banco de pruebas resultante ha sido empleado como insumo para análisis solares.

Posteriormente, se exponen las decisiones efectuadas para desarrollar una solución de software de simulación energética de edificios (BES, siglas en inglés). La tesis se centra en la demanda neta de calor y no en la energía primaria, debido

a limitaciones de datos a escala nacional. Varios parámetros necesarios para la energía primaria no están disponibles abiertamente y suponen asumir hipótesis adicionales.

La investigación presenta demanda neta de energía para dos casos: Rijssen-Holten y los Países Bajos. Los resultados se clasifican por tipología edificatoria y periodo de construcción. Se comparan con los datos del certificado de eficiencia energética (EPC) neerlandés. Esta comparación debe ser interpretada con cautela y sólo sirve para contextualizar los resultados frente a los datos oficiales disponibles por edificio.

La tesis subraya la relevancia de los datos abiertos y los pasos para enriquecer s3DCM aptos para aplicaciones energéticas. Además, propone integrar resultados de simulación solar con la BES desarrollada. El objetivo es implementar métodos híbridos que representen mejor el parque edificatorio neerlandés.

ACKNOWLEDGEMENTS

*Gracias a la vida que me ha dado tanto.
Me dio el corazón que agita su marco
cuando miro el fruto del cerebro humano;
cuando miro el bueno tan lejos del malo
cuando miro el fondo de tus ojos claros.*

Violeta Parra

I would like to express my gratitude to everyone who supported, guided, and challenged me throughout my PhD research at the 3D geoinformation group of the Technische Universiteit Delft (TU Delft). I am grateful for every conversation, piece of advice, and word of encouragement, all of which were essential to this journey.

My deepest thanks go to my promoter, **Prof. dr. J.E. (Jantien) Stoter**, for her scientific advice, commitment, and dedication to this research. Also, I am very thankful for her concerns about my wife's and my well-being during our times of sorrow. I am profoundly thankful to my supervisor, **Dr. Giorgio Agugiaro**, for his constant support and encouragement, his constructive criticism, and his valuable remarks. He has taught me through his actions that the meaning of the German word "Doktorvater" runs deeper than "doctoral supervisor." He is a true inspiration for my future actions.

I wish to thank the selection committee for the PhD position in *3D city models for energy applications*. I will forever be grateful for the chance to pursue my dream, and I hope I have honoured your trust.

My thanks also go to Hugo Ledoux for his invaluable support with my early scholarship applications before I joined the group. I remain grateful for his help at a time when it was most needed.

To my (former) colleagues and friends at the 3D geoinformation group, it has been a privilege and an honour to meet you and to work by your side. Thank you for sharing the most valuable item that everyone has: your time. To my *partners in crime*, thank you for the wonderful non-academic activities that help me distract my mind and connect with each of you. I hope we will meet again soon for a bubble tea and, of course, a hot pot.

I extend my sincere gratitude to the municipality of Rijssen-Holten for providing access to their dataset, which was fundamental to my research. I am particularly thankful to Joop Voortman for his availability and patience in answering my numerous questions about the data.

My gratitude extends to the MSc students in Geomatics whom I supervised or co-supervised during my PhD. Their dedication and hard work played a critical role in advancing my research. In the digital version of this dissertation, each student's name is linked to their publicly accessible thesis, celebrating their contributions concretely. [Dennis Giannelli](#), [Yuzhen Jin](#), [Özge Tufan](#), [Konstantinos Pantelios](#), [Carolín Bachert](#), [Tendai Mbwanda](#), [Chris Poon](#), [Gabriela Koster](#), [Bing-Shiuan Tsai](#), [Longxiang Xu](#).

Finally, I thank the doctoral committee members for accepting our invitation to be part of it, for reading my dissertation, for the feedback provided, and for the academic discussion during my public defence.

* * *

Gracias a mis abuelos, víctimas, resilientes y luchadores perseverantes que lo dieron todo por sus familias. A Gladys Sánchez, mi madre, y Pedro León, mi padre, les agradezco por todo lo que hicieron por y para mis hermanos y por mí. Por llevar en silencio el dolor de tener a un hijo lejos, mientras este buscaba cumplir sus sueños.

A mis hermanos, les agradezco por compartir la vida conmigo cada día a través de una pantalla, manteniéndonos unidos a pesar de la distancia. A mis sobrinos por sus mensajes de apoyo y los abrazos en los pocos días juntos durante estos años. A mis tíos y primos, por sus mensajes y apoyo durante estos años.

A Rosa Oliveros y a Luis Sarmiento, mis suegros, gracias por el apoyo incondicional que han brindado a nuestro hogar, por visitarnos y acompañarnos en todo momento.

A mi amada esposa, Luz Adriana: tú le das un brillo especial a mi vida que espero que alumbre eternamente. A Ónix y Haku, quienes con una simple mirada son capaces de arrancarme de la oscuridad de mis pensamientos para llevarme a la alegría de caminar juntos.

A mis amigos en Colombia, gracias por estar allá y aquí, por pensar en mí y estar pendientes de mi porvenir. Escribir sus nombres significaría darles un orden que no hace ninguna justicia y no tiene ningún sentido; gracias por reír conmigo y acompañarme en mis tristezas. Pero sobre todo, gracias por estar ahí cuando los he necesitado. Y a todos aquellos, en diferentes lugares, que me han demostrado su interés por mí y mi camino con una llamada, un correo o un mensaje, mi más sincero agradecimiento.

* * *

To Gustavo Arciniegas and Robin Berghuijs, my paranympths: you and your families opened your homes to mine, making us feel happy and welcome in the Netherlands. I am so grateful to you for that. We truly wish we can stay here longer. I selected you for this role in tribute to everything you have done for us.

1

INTRODUCTION

1.1. MOTIVATION

Since the industrial revolution, the place where people live has changed dramatically from rural to urban areas (Lampard, 1955). Although this trend started centuries ago, it still occurs; as of today, 56% of the global population lives in urban areas, and it is expected to increase to 70% by 2050 (The World Bank Group, 2024). Since cities are responsible for 60-80% of all "energy consumption"¹ (United Nations, 2015), the current global political agenda emphasises enhancing living standards in urban areas, including optimisation of energy usage.

Accurately quantifying current and expected building energy demand is the basis to ensure adequate energy supply planning and informed decision-making. Therefore, the development of scalable and realistic models that consolidate building energy demand is essential for long-term energy planning. These models are essential for addressing the challenges of climate change and population growth. Due to these challenges, governments have implemented new legislation and incentives to promote building renovations in accordance with the climate and economic goals.

The variety of buildings, techniques, and methods involved in modelling the energy demand of the building stock at the urban scale creates a challenging research context. The limited availability of open measured data to validate these models further complicates efforts in this area. Thus, a robust building stock model plays an important role for the correct modelling of the energy performance of buildings at urban scale.

¹The term *energy consumption* is widely used, even by institutions such as the United Nations and the International Energy Agency. However, since energy can only be transformed (e.g., between potential and kinetic forms) and is always conserved –neither created nor destroyed– the term is scientifically incorrect. I use it in this document between quotation marks due to its widespread adoption, and when cited, because it appears in the terminology of the corresponding authors.

Urban Building Energy Modelling (UBEM) provides a suitable framework for addressing societal challenges of climate change and growing population needs by enabling computational analysis and simulation of building energy performance at the city scale (Davila *et al.*, 2016). UBEM refers to the techniques, methods, and software tools used to simulate the energy performance of buildings at several scales.

UBEM utilises physical models of heat and mass flows in and around buildings. These models can quantify operational energy use and the environmental conditions (indoor, outdoor) for groups of buildings (Reinhart and Cerezo Davila, 2016). To enhance the calculation of building energy performance, the development of UBEM techniques has increasingly centred on simulation and multi-scale analysis using flexible approaches (Abbasabadi and Mehdi Ashayeri, 2019). These approaches must be adaptable to various building characteristics, including size, geometry, economic function, construction materials, and thermophysical properties.

UBEM follows two main approaches: top-down and bottom-up. The former uses the whole building sector as one variable in the analysis and, for that reason, it does not account for differences among buildings when calculating "energy consumption". The latter focuses on individual buildings and their physical characteristics. Achieving accurate UBEM outcomes through bottom-up methods relies on comprehensive building stock modelling and detailed representations of individual structures, often based on building archetypes.

Data dependency in UBEM represents a complex challenge due to the need for detailed information about buildings, such as construction techniques and materials. Furthermore, if such data exist, they are often not publicly available. In addition, the limited open data available come from several sources, requiring harmonisation processes, so the data are suitable for UBEM. This harmonisation step adds another layer of complexity for modelling the energy performance of buildings. In cases where no data are available or accessible, machine learning techniques have been developed and implemented to address the lack of these data (Seyedzadeh *et al.*, 2019; R. Wang *et al.*, 2020).

Semantic 3D city models (s3DCMs) have become increasingly relevant over the past 15 years in bottom-up Urban Building Energy Modelling (UBEM) approaches. Although the acronym 's3DCM' does not commonly appear in the literature –which typically only uses '3DCM' instead– this PhD thesis adopts s3DCM to highlight the significance of semantics. Specifically, s3DCM refers to standardised data models (e.g., CityGML), whereas 3D city models denote other data formats (e.g., OBJ, BLEND) without semantic information. S3DCMs represent the geometry and semantic information of city objects in a three-dimensional environment, providing essential input for physics-based energy simulations. By incorporating details such as building geometry, occupancy, and the availability of local renewables, these models facilitate more accurate analyses of building energy performance.

1.2. PROBLEM STATEMENT

The International Energy Agency (IEA) states that in 2023, the total final "energy consumption" (TFC) was 442 EJ², which is split between industry (167 EJ), buildings (133 EJ), transport (116 EJ), and other end-uses (27 EJ) (IEA, 2023). These values indicate that buildings account for 30% of the total "energy consumption". Furthermore, the energy demand is expected to increase, primarily driven by the rise in households from 2.2 billion in 2023 to 3 billion in 2050. Figure 1.1 shows the energy demand of the building sector in 2022, classified by source and by end-use. It also contains several projections of the energy demand of the building sector by 2030 and 2050, named STEPS, APS, and NZE. These acronyms correspond to the different transition scenarios, which are:

- STEPS: Stated Policies Scenario. It is based on the latest policy settings, including energy, climate, and industry
- APS: Announced Pledges Scenario. It assumes that national governments meet all national and energy targets in full and on time
- NZE: Net Zero Emissions by 2050, which limits global warming to 1.5°C, requiring additional progress by society and industry

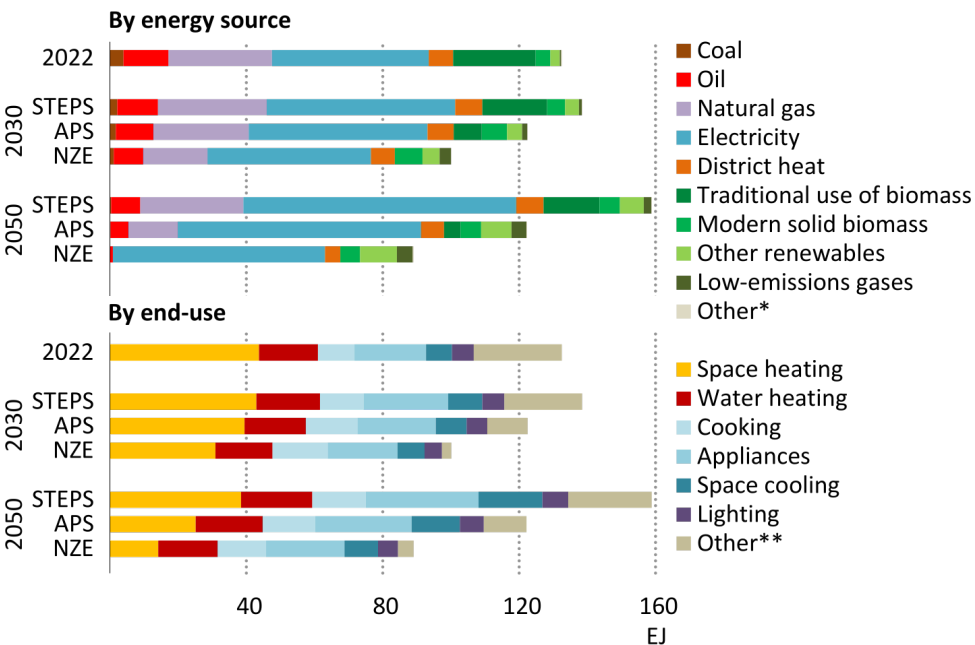


Figure 1.1: Buildings sector energy demand by source and end-use, 2022-2050. Source (IEA, 2023)

²One exajoule (EJ) is equal to 277.8 terawatt-hour (TWh)

Based on [figure 1.1](#) data, it can be concluded that more effort is needed to follow the stated policies to decrease the energy demand of buildings. To achieve meaningful reductions, renovating building envelopes and adopting new operational technologies are essential. Under the Announced Pledges Scenario (APS), the energy demand for buildings is expected to decrease by 12% by 2030 compared to the Stated Policies Scenario (STEPS).

Most of the energy consumed in buildings is used for space heating. However, the increasing demand for water heating is driven by population growth (IEA, [2023](#)). These trends in building energy demand add pressure on the energy system and have negative environmental impacts.

In response to these challenges, there is a trend towards a sustainable building sector, employing local renewable energy sources to meet existing energy demands. One example is photovoltaic (PV) technology in buildings, which among the rest, is more and more adopting 3D city models as one of the input data sources (Bensehla *et al.*, [2021](#); Mohammed *et al.*, [2024](#); Santhanavanich *et al.*, [2023](#)).

Other challenges affecting the accuracy of energy predictions include changes in weather patterns and rising temperatures as a result of climate change phenomena; the difference in national or local construction codes that introduce heterogeneity in the thermal characteristics of buildings; occupancy schedules, and the thermal comfort preferences, as well as the limited access to building data such as the insulation levels or renovation history. In particular, it is a significant challenge for governments to collect detailed data on the renovation of buildings. However, the availability of such data greatly improves the accuracy of building energy analyses.

To address the challenge of data availability for accurate building energy analysis, the European Union (EU) in 2024, released the fourth revision of the energy performance of buildings directive –EPBD IV– (European Parliament, [2024](#)), aiming to increase the rate of renovation of buildings in the EU. The directive also supports the digitalisation of energy systems for buildings and allows national governments to decide on the renovation measurements that best suit their contexts. At the same time, in the Netherlands, there is a preliminary work by the Ministry of Internal Affairs (in Dutch "*Ministerie van Binnenlandse Zaken*"), which delegated the task to Kadaster Netherlands to perform an exploration of a National Facility for Building Data (in Dutch "*Landelijke Voorziening Gebouwgegevens*") as a support for decision-making in the building sector (Noordegraaf and van Haaften, [2024](#)).

Building Energy Simulation (BES) tools are essential in assessing the building energy performance. These tools play a significant role in engineering and architectural design by providing usage patterns, identifying cost-effective renovations or calculating payback periods for energy-saving measures, such as photovoltaic panel installation (van den Brom, [2020](#)). However, these tools face significant challenges at the city level, pointing to the need for strategies to

improve the accuracy of energy performance calculations at different scales.

Moreover, previous studies have highlighted that existing BES tools often produce theoretical consumption values that can diverge significantly from actual measured values (Glasgo *et al.*, 2017; van den Brom, 2020). This discrepancy suggests inherent limitations in the approaches followed by the BES tools. Additionally, these limitations show the need for more detailed information on building physics, as no national database records city-wide construction techniques, materials, or changes over their lifespan as Building Information Modelling (BIM) for individual buildings or projects, except for a few new buildings.

Adopting building archetypes overcomes this data uncertainty partially. For example, the Netherlands Enterprise Agency (in Dutch *"Rijksdienst voor Ondernemend Nederland"*–RVO–) publishes a brochure of stereotype buildings for energy analyses (RVO, 2023). These archetypes provide generalised models that represent common building types, aiding in estimating energy performance where specific data is lacking. However, RVO states that this classification should be considered only for characterisation purposes, not individual building energy analyses. Currently, no research confirms or further elaborates on this statement from RVO.

The use of a BES tool tailored to the whole Netherlands in tandem with s3DCM as a data source for the energy performance of buildings has yet to be done. Moreover, both the method and the reliable open data have limitations that must be identified and specified in detail. Therefore, there needs to be an understanding of the data requirements of the tools that can be supplied or derived from a s3DCM in relation to current open datasets.

1.3. RESEARCH OBJECTIVE AND SCOPE

The goal of this PhD thesis is to contribute to enhancing urban energy applications through semantic 3D city models (s3DCM) and open data. The research focuses on the Netherlands as the primary use case. Accordingly, in chapter 2, I will review the Dutch Technical Agreement (in Dutch *"Nederlands Technische Afspraak"*–NTA–) 8800 –hereafter referred to as NTA 8800– as the framework for calculating the energy performance of buildings using an energy balance method. The scope of analysing this norm is the identification of a standardised calculation workflow of the energy demand of buildings and their data requirements.

A s3DCM partially meets the data requirements, as it represents just *one* of the available information sources. Therefore, my investigation focuses on leveraging country-wide *open* data and evaluating how the s3DCM data model can serve as a data integration platform. The goal is to simplify and streamline the selection and integration of multiple datasets for energy applications.

Before using an s3DCM for building energy performance, it is necessary to

explore the underlying data model and the semantic data.

As a data source for building energy performance, s3DCM involves exploring the underlying data model and the semantic data according to the level of detail (LoD). My analysis consequently includes identifying additional data that need to be collected from other sources. One example is the heat transfer coefficient (U-value) of the surfaces (roofs, walls) that define the building's thermal hull. However, not all required data are available in the desired detail for all buildings. For example, the size of openings (windows or doors) in the façade of a building may not always be available. To address these data gaps, I make design decisions, such as using a ratio (percentage of the surface area) for the size of openings, which is discussed in detail in [chapter 3](#).

The software, data, and design decisions are tailored to the Netherlands. [Chapter 4](#) presents the design decisions regarding the data preparation of the open data available to comply with the data requirements of the NTA 8800; [chapter 5](#) provides detailed information on the decisions taken when creating a BES tool with the scope to calculate the net heat demand of buildings in the Netherlands based on the NTA 8800. The goal is, while giving a detailed description of the decisions taken during my implementation, is to provide insights to the opportunities of s3DCM for BES to support the transition to sustainable urban environments.

This PhD thesis will address the following research questions:

- What are the requirements for UBEM that s3DCM should meet?
- What are the challenges in data preparation for UBEM?
- How could existing s3DCM based on CityGML be improved when to serve as input data for UBEM?
- How can artificial intelligence methods improve the quality and content of s3DCM for UBEM?

Finally, I will reflect on the role of s3DCM when it comes to policies for the building sector.

1.4. OUTLINE OF THE THESIS

1.4.1. STRUCTURE OF THE THESIS

The thesis is organised into seven chapters and two main parts, as shown in [Figure 1.2](#). Besides this introduction ([chapter 1](#)) and the conclusions ([chapter 7](#)), there are two main sections. The first one refers to the framework ([chapters 2 and 3](#)) and the second one to the implementation part ([chapters 4 to 6](#)).

Framework

[Chapter 2](#) provides the main concepts of UBEM, including data-driven, physics-

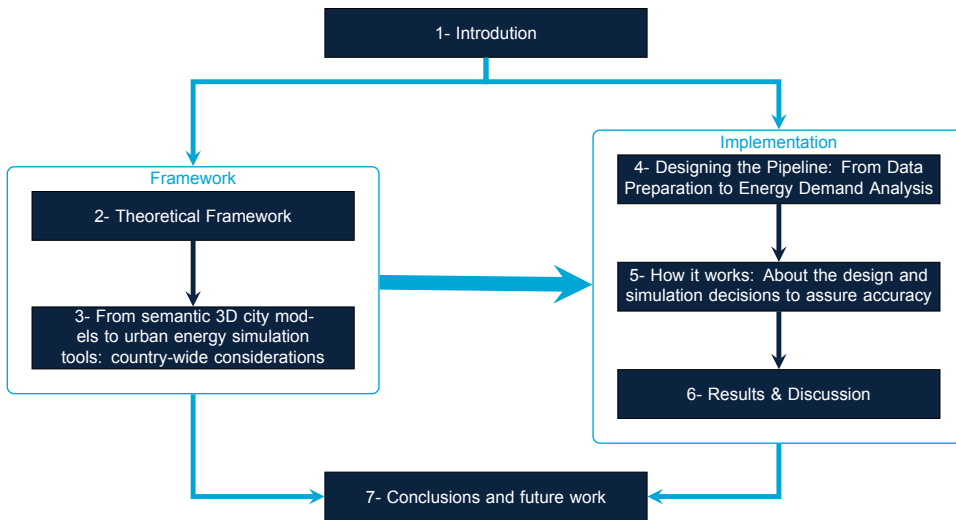


Figure 1.2: Organisation of the thesis

based, and reduced-order modelling approaches. These concepts form the theoretical framework for my research. It also contains a literature review pertinent to my study.

Chapter 3 builds upon the framework defined in **chapter 2** to discuss the requirements of the to-be-developed BES. This chapter contains the general considerations and data requirements according to the Dutch norm NTA 8800, which is the method followed in this thesis for computing the energy performance of buildings.

Chapter 4 details the design of the pipeline for the development of a BES tool for the Netherlands. This chapter consolidates the framework and concepts from the previous two chapters, highlighting the integration of multiple public open datasets and s3DCM. It describes each stage, from data ingestion and transformation to the computation of the heat energy demand of buildings in the Netherlands.

Implementation

Chapter 5 describes the challenges and decisions taken, as well as the software diagrams that explain the implemented BES. The chapter includes several examples of data inconsistencies found during this stage of my PhD research and the decisions taken to overcome some of such inconsistencies.

Chapter 6 presents the output of the simulations performed using the BES tool described in **chapter 5** at two scales: first, a case study in the municipality of Rijssen-Holten, then at the country level for the whole Netherlands. It includes a data comparison against the data available from the energy label dataset and an example of calculation of a refurbishment scenario for old buildings.

Chapter 7 discusses the challenges, limitations and contributions of the thesis on the use of open data and semantic 3D city models to enhance urban energy applications.

1.4.2. PERSONAL PRONOUNS

In this thesis, the personal pronoun 'I' is used to refer to the work done by myself under the supervision of my promotor and copromotor. I use the pronoun 'we' when referring to work carried out in collaborations in which others were involved. The respective publications are mentioned and cited in this document.

2

THEORETICAL FRAMEWORK—LITERATURE REVIEW

This chapter provides the conceptual framework, context, and definitions of Urban Building Energy Modelling—UBEM, which will be fundamental to the upcoming chapters. Additionally, the chapter includes a literature review of UBEM approaches. Then, it describes the Dutch norm for the calculation of the energy performance of buildings, which is the main method implemented in this research.

This chapter is structured into five sections. First, I present an overview of urban building energy modelling (UBEM), including a literature review and description of the current challenges ([section 2.1](#)). Second, I discuss the data modelling for UBEM with specifying between geospatial and non-geospatial data ([section 2.2](#)) with a focus on semantic 3D city models (s3DCM) ([section 2.2.1](#)). Third, I present the energy balance method ([section 2.4](#)). Fourth, I describe the NTA 8800, which is the Dutch norm for the calculation of the energy performance of buildings ([section 2.5](#)). The chapter closes with the key findings of the literature review and their importance in the development of the following chapters ([section 2.7](#)).

2.1. GENERALITIES OF URBAN BUILDING ENERGY MODELLING

Urban Building Energy Modelling (UBEM) involves the development of techniques, methods, and software tools aimed at simulating the energy performance of buildings in urban areas across multiple scales of analysis (Hong, Chen, X. Luo *et al.*, 2020; Kamel, 2022; Malhotra, Bischof *et al.*, 2022;

Oraopoulos and Howard, 2022; Reinhart and Cerezo Davila, 2016). UBEM evaluates individual building characteristics and dynamics within the broader context of urban microclimate interactions. The goal of UBEM is to provide detailed and accurate information on energy demand and usage across various time scales, which can inform sectors such as urban design, building operations, and energy policy development and implementation.

The performance indicators or scale of analysis depends on the specific use case, i.e., the questions to be answered by the analysis. For instance, the timeframe and time resolution (hourly, daily, yearly) may be used to evaluate the responsiveness to energy demand fluctuations in a few buildings or an entire district network within a city. Additionally, some applications may need to fully account for urban microclimate variations, which often rely on weather data obtained through historical aggregation models. Consequently, the choice of scale, model, and data quality depends on the questions being addressed and the available resources. Figure 2.1 shows a general overview of the critical components of a UBEM system, which can be split into three main components: datasets, simulation workflow and the analysis of the results from multiple perspectives such as stakeholder analysis, support for policymakers, metrics evaluation.

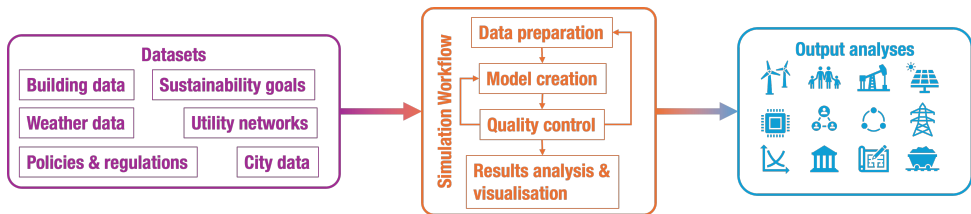


Figure 2.1: Overview of UBEM components

Residential "energy consumption" modelling techniques can be classified into two main approaches: top-down and bottom-up. Swan and Ugursal (2009) propose a widely adopted hierarchical framework that organises these techniques according to the level and type of input data used, as shown in Figure 2.2. Building upon this work, Langevin *et al.* (2020) extended the framework by introducing a flexible hierarchy that links different approaches through hybrid models, adds sub-layers for key energy use determinants (e.g., people, building stock, environment). Furthermore, it incorporates additional descriptive dimensions (e.g., system boundary, spatio-temporal resolution, dynamics, uncertainty) that should be specified alongside the four high-level quadrants (Q1-Q4). The proposed scheme is shown in figure 2.3.

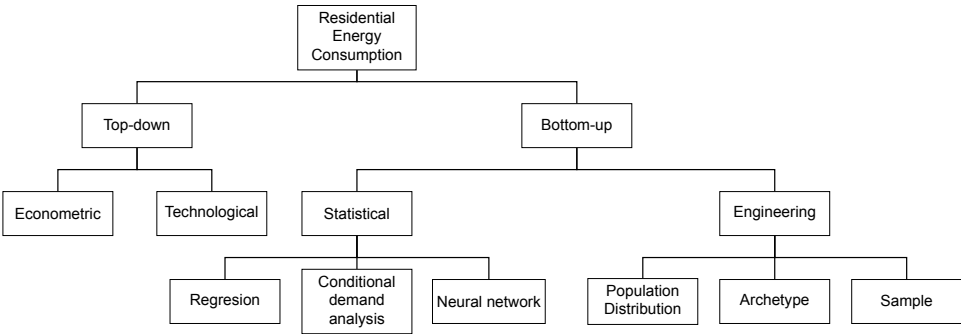


Figure 2.2: Hierarchy of modelling techniques for UBE, from (Swan and Ugursal, 2009)

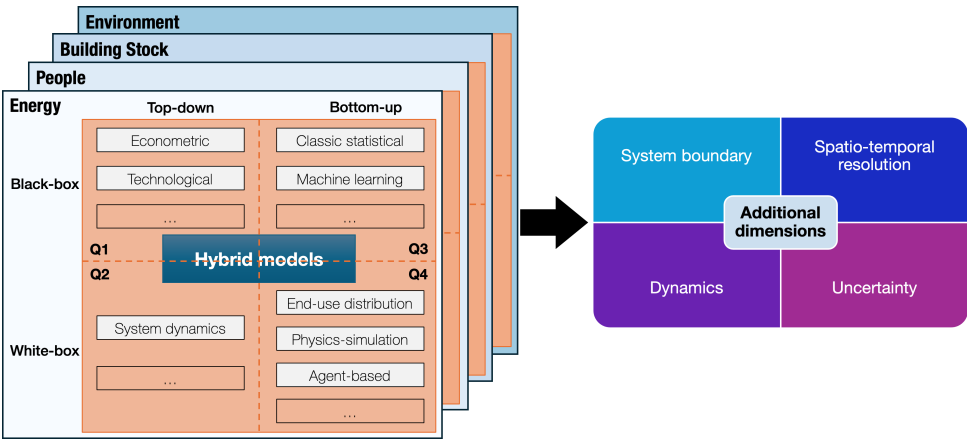


Figure 2.3: Updated classification of UBE models proposed by Langevin *et al.* (2020)

Top-down models collectively consider the entire building sector to calculate total "energy consumption". While effective for macroscale analysis, these models offer limited insight into the energy performance of specific buildings or groups within an urban context (Ali, Shamsi, Hoare *et al.*, 2021). These models can be classified into econometric and technological. Econometric top-down models focus on macroeconomic trends and past relationships rather than the physical factors in buildings that influence energy demand. These models can also not account for sudden technological changes, leading to a lack of adaptability when addressing climate change, where future conditions may differ significantly from historical patterns. Meanwhile, technological top-down models consider factors such as saturation effects, technological progress, and structural changes influencing energy use. However, these factors are only sometimes explicitly

detailed within the models.

In contrast, *bottom-up* models provide a detailed assessment of energy use in a building or cluster of buildings, using physics and engineering principles to address the interactions between buildings and their environment. These models then extrapolate their findings to larger geographic areas, offering high spatial and temporal resolution in their analysis. This detailed approach supports a deeper understanding of energy dynamics at the building and community levels.

Bottom-up UBEM transcends the method of aggregating single building values to urban scale levels by incorporating the dynamics and influences of buildings and their environment. This symbiotic relationship between buildings and their environment is evident in the urban microclimate, encompassing factors such as shading, radiant heat exchange, and solar reflection, all of which contribute, for example, to urban heat island effects (Johari, Peronato *et al.*, 2020).

The outcomes of UBEM provide valuable insights into energy demand, greenhouse gas (GHG) emissions, and the financial implications of building retrofitting under various scenarios. Given the significance of these results in policy-making, ensuring their accuracy is paramount. However, this complex and multidimensional process involves several considerations, including spatial and temporal resolution, error metrics, and the scale of analysis. For example, aggregated annual data at the district level may achieve accuracies within 1% (D. Wang *et al.*, 2018), while monthly data after Bayesian calibration can show variability of up to 15% (Sokol *et al.*, 2017). Furthermore, the accuracy of hourly values for individual buildings can reach errors of up to 120% (Kristensen *et al.*, 2018).

2.1.1. PHYSICS-BASED MODELS

The physics-based methods, also called engineering or simulation methods, use simulation techniques, building characteristics, construction, climate, and system data to calculate end-use "energy consumption". Swan and Ugursal (2009) state that physics-based models may use distribution, sample, or archetype-based approaches.

The *distribution-based* approach analyses the regional distribution of building energy use to calculate end-use energy consumption (Swan and Ugursal, 2009). Howard *et al.* (2012) developed a model to estimate building energy end-use intensity (in kWh/m²) in New York City for space heating, hot water, cooling, and other electricity uses. Their model focuses on the building function and is calibrated using ZIP code-level energy data and end-use ratios from U.S. surveys.

The *sample-based* approach starts with input data from individual sample buildings. This method requires an extensive database to represent buildings accurately (Swan and Ugursal, 2009). Once the sample is established, it is

the foundation for generating the building stock. The spatial distribution of this stock is then determined using aggregate datasets and techniques like iterative proportional updating. Nägeli *et al.* (2022) use this approach to create synthetic spatial building stock for the spatial analysis of the energy demand in Ireland and Austria. Their outcomes indicate that the resulting building stock energy demand can be analysed for energy planning purposes when no building-level data are available.

The *archetype-based* approach categorises buildings by type, size, climate, and construction year (Lavagna *et al.*, 2018). These categories are used to model the "energy consumption" in simulation engines to assess the impact of energy strategies on a regional or national scale (Ali, Shamsi, Hoare *et al.*, 2021). This approach, dependent on quantitative building physics data, aggregates these models to represent the broader building stock. Inputs such as thermal properties, heating patterns, and occupancy schedules are essential. However, the approach makes several assumptions about occupant behaviour and requires significant technical data to produce accurate "energy consumption" estimates (Swan and Ugursal, 2009). Energy Plus (Crawley *et al.*, 2001) is a well-known *physics-based* BES that several researchers have used due to the quality of its models and results (Ali, Bano *et al.*, 2024a; Faure *et al.*, 2022; Seyedzadeh *et al.*, 2019; Souza *et al.*, 2024).

2.1.2. DATA-DRIVEN MODELS

Data-driven modelling uses statistical and AI methods to analyse building "energy consumption" based on available data, such as building stock, billing records, and socio-economic variables (Ali, Shamsi, Hoare *et al.*, 2021). There are two approaches under this kind of models: statistical and artificial intelligence (AI). Statistical approaches apply regression to correlate energy use with design and operational parameters. On the other hand, AI approaches train on ground-truth data, usually historical datasets, to identify relationships between "energy consumption" and influential factors like building and urban characteristics and occupancy patterns. The general method for these models starts with data collection and preparation. Then, data are used for feature engineering, followed by data splitting between training and testing purposes. The next step is the model development and, finally, its performance evaluation.

Ali, Bano *et al.* (2024a) implement a method that performs the energy performance of urban residential buildings using ensemble-based machine learning and end-use demand segregation models that include heating, cooling, lighting, energy appliances and domestic hot water consumption of buildings. Their method is evaluated using a synthetic building dataset of one million buildings through the parametric modelling of 19 key variables for residential building archetypes (Detached, Terraced, Semi-detached, Bungalow) in Ireland. The study implemented an end-use demand segregation method, including heating, lighting, equipment, photovoltaics, and hot water, to predict the energy

performance of buildings on an urban scale. The results achieve an accuracy of 76% without segregation, which increases to 91% when using end-use demand segregation.

Mosteiro-Romero, Quintana *et al.* (2024) propose an agent-based model of building occupants' activities and thermal comfort in an urban district in Singapore. This research explores the influence of district operation strategies on building energy performance in the context of increased workspace flexibility. Results show a decrease between 6%-15% of the energy demand for cooling by implementing occupant-driven ventilation and set points.

Li *et al.* (2023) propose an energy-flow-based ensemble calibration model to reduce the discrepancies in building performance simulations to within 6%. Their model features three layers: input neurons, complex heat and electrical transfer models, and output neurons. By replacing hidden neurons with these transfer models, the approach accounts for deep energy retrofit interactions while minimising computational demands.

Ali, Shamsi, Bohacek *et al.* (2020) propose an approach that combines bottom-up, data-driven, and spatial modelling techniques for mapping building energy performance across multiple scales using a Geographic Information System (GIS). They use data from approximately 650,000 Irish energy performance certificates. Their approach predicts the energy performance of over 2 million buildings. Their results indicate that the model performance obtains 76% for the energy rating classification in Ireland (A1, A2, ..., F, G). In the case of performing an aggregated classification of labels (A, B, CD, EFG), the model performance increased to 88% accuracy. In both cases, they use deep-learning algorithms. Their results were used for spatial modelling with open data for energy planning and to support decision-making. Additionally, the proposed model identifies clusters of buildings that could be used for potential energy saving by retrofitting processes.

Van den Brom 2020 investigates the "energy consumption" in dwellings, its characteristics and the importance of the occupancy behaviour. In this research, van den Brom uses data from the social rental sector audit and evaluation of energy saving results (in Dutch "Sociale Huursector Audit en Evaluatie van Resultaten Energiebesparing" –SHAERE"), Dutch statistics data and registry data from the Netherlands and Denmark. The findings of this research state that energy efficiency and household behaviour interact in such a complex way that they cannot be fully explained by occupant characteristics alone.

R. Wang *et al.* (2020) propose a stacking model to forecast building "energy consumption" by combining various base prediction algorithms into meta-features. This combination allows the final model to analyse datasets from different spatial and structural perspectives. The model performance is evaluated by comparing its results with existing prediction models, including K-Nearest Neighbour, Support Vector Machine, Gradient Boosted Decision Tree, Random Forest, and Extreme Gradient Boosting. The comparison indicates that the proposed stacking method outperforms the other models.

Yoon (2020) proposes a virtual in-situ calibration method that combines Bayesian inference with an auto-encoder to address sensor errors affecting air handling units' operation and energy use (AHUs). The method employs a three-step process for constructing input variables and introduces a novel distance function in the auto-encoder to ensure effective calibration under various malfunction conditions. A case study showed that this approach eliminates a cooling coil sensor error, reducing energy waste by 38% and demonstrating its potential for self-repair, diagnostics, and automation in building energy management.

Vázquez-Canteli *et al.* (2019) present an integrated simulation environment combining CitySim (Mutani *et al.*, 2018) and TensorFlow (Martín Abadi *et al.*, 2015). They implement a deep reinforcement learning controller for the HVAC systems to analyse its response when applying changes in the building energy model. One of the case studies is about energy savings of a single building, proposing three scenarios: base case and sensitivity analysis, installation of photovoltaic PV panels and building retrofitting.

One common characteristic of these models is the requirement of a lot of reliable data for training of the models including the consumed energy values.

2.1.3. REDUCED-ORDER MODELS

Reduced-order models, including resistance-capacitance (RC) methods, are commonly used for faster assessment of building energy performance and require fewer inputs compared to fully physics-based or data-driven methods. These approaches rely on standards set by organisations such as the International Organization for Standardisation (ISO), which set parameters and simplify complex physical processes. For instance, ISO 52016-1:2017 (ISO, 2017) provides normative guidelines for calculating the building energy performance by considering essential physical parameters and system types. In the European Union, reduced-order methods are widely applied to determine energy performance ratings, as seen in the Netherlands with the NTA 8800 norm (NEN, 2024). In the case of the NTA 8800, this is considered as reduced-order model since it uses monthly average values for most of its calculations (in the case of hot tap water, it uses yearly values). The norm also applies specific correction factors to account for dynamic effects, and therefore it cannot perform hourly or sub-hourly calculations

Examples of reduced-order models Urban Building Energy Simulation tool –BES– include CitySim (Emmanuel and Jérôme, 2015), SimStadt (Nouvel, Brassel *et al.*, 2015), City Energy Analyst (CEA) (Fonseca *et al.*, 2016), and TEASER (Remmen *et al.*, 2018). While these tools provide valuable simulation capabilities, each presents limitations that affects their direct applicability in the Dutch context and for large-scale urban energy modelling. For instance, CitySim is recognised for its accurate physical models, however, its computational cost remains high—computing even small urban areas (e.g., a 1.2-km-side square)

can take several days, as experienced in León-Sánchez, Giannelli *et al.* (2025). SimStadt incorporates specific archetypes and decisions calibrated for German municipalities and specific cities like New York or Rotterdam. Therefore, limiting its generalisability to countries with different building stock compositions and regulations. CEA adopts a block-based geometry and uses ISO 52016-based equations, yet its default parameterisation follows Swiss standards, making it less suitable without substantial reconfiguration.

Guen *et al.* (2018) use CitySim to evaluate the improvement of the energy sustainability of the village of Hemberg in Switzerland. In their method, CitySim is the BES that computes the hourly energy demand of buildings based on retrofitting according to the Swiss energy labels and the solar potential of the boundary surfaces of the building envelope. They use three scenarios that consider the integration of different energy sources, such as PV panels, wind energy and efficient energy systems.

CitySim has also been used to compute the cooling demand of buildings in hot regions. Mohammed *et al.* (2024) use CitySim as the urban energy engine to assess cooling demand. Their study evaluates the impact of adopting cool roofs and modifying urban reflectivity in Dubai. The results indicate that increasing the albedo has a positive correlation with the reduction of the building cooling demand, demonstrating the effectiveness of these strategies in subtropical desert areas.

Rossknecht and Airaksinen (2020) evaluate heating demand predictions in Helsinki using semantic s3DCM based on CityGML and its Energy ADE (see section 2.2.1.4), with the energy simulation tool SimStadt (Coors *et al.*, 2021). Their simulation results for heating demand were validated against measured consumption data from 1915 city-owned buildings in Helsinki. The analysis revealed a mean absolute percentage error (MAPE) of 25.6% when using default parameters. However, substituting the heated area (computed by the simulation tool) with the net floor area data from the Helsinki Energy and Climate Atlas reduced the error to 19.6%, highlighting the impact of accurate spatial input data on model performance.

Using a case study approach, Vargova *et al.* (2023) examines the benefits of green roofs as part of green transformation processes. The research involved comparing pre- and post-establishment (roof modification) data with simulations. Building energy demand simulation models were assessed using SimStadt, where the software library was adapted, with the roof's U-value being the most significant parameter change. The results indicate a 15%-40% reduction in the roof's U-value, depending on the type of greening. Consequently, the computed annual building heat demand of buildings shows an average reduction of 4%.

Oraiopoulos, Hsieh *et al.* (2023) use CEA as the computational framework to assess the energy demand of Swiss communities based on their spatial archetypes (urban, suburban, and rural). They evaluate three urban development scenarios for these archetypes. Their findings indicate that by 2060, the

urban archetype will see an increase in space cooling demand, reaching levels comparable to space heating. The suburban archetype shows an increase in overall energy demand due to urban development, while the rural archetype consistently presents high space heating demand across all scenarios.

2.1.4. CURRENT CHALLENGES OF UBEM

The UBEM uncertainties are based on factors related to inaccurate input data such as the construction materials of buildings, occupancy behaviours, equipment and their schedules, energy systems, weather conditions and refurbishment history of buildings. Some authors have researched and categorised these sources of uncertainty from different perspectives. In the case of physics-based models, Ding *et al.* (2015) identifies two sources, one based on the researchers' biases and a second one based on the simulation tools. Yu *et al.* (2022) classify building uncertainties into four: specifications, models, materials and scenarios. In this section, I classify and discuss the uncertainties from the data perspective. The main factors are humans, buildings and weather. The hierarchy diagram is shown in figure 2.4 followed by the explanation of each of the factors ordered from left to right. Other sources of uncertainties such as the model are shown in gray in figure 2.4. However, they are not discussed here since this PhD thesis is focused on the use of open data for UBEM.

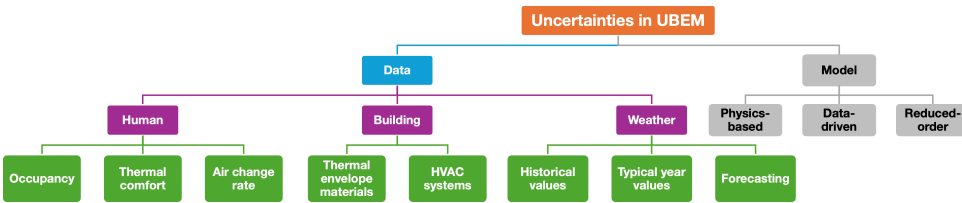


Figure 2.4: Sources of uncertainties in UBEM

Human behaviour strongly influences uncertainties in UBEM (Faure *et al.*, 2022). This influence can be classified into occupancy, thermal comfort and the air change rate. Any variance in those factors leads to variations in energy consumption (Fu and Miller, 2022). One way to reduce human influence is at the base of the Energy Performance Certificate (EPC) method (Heidenthaler *et al.*, 2023), which is based on a “standard behaviour”.

Several building factors influence the UBEM uncertainties, including U-values, g-value, emissivity, absorption, infiltration rate, material thickness and construction techniques. Although the U-value of materials has already been defined, they can vary according to weather conditions (O’Hegarty *et al.*, 2021; Ohlsson *et al.*, 2022; Y. Wang *et al.*, 2022). In the case of building materials, emissivity and absorption are physical properties that determine how constructions reflect or absorb solar energy. Silva and Ghisi (2014) state that these parameters

significantly influence the energy performance simulations but, conversely, are very seldom known and available at the city scale; hence, they are available only for some single buildings.

When it comes to weather data, UBEM methods, specifically physics-based models, use weather station datasets as input for their models (Ascione *et al.*, 2022; Eames *et al.*, 2016). For example, Mahdy *et al.* (2022) simulate the “energy consumption” of residential buildings using typical meteorological year (TMY) data in Egypt. They predict an annual “energy consumption” of 1508 kWh per flat by 2080. Nevertheless, relying on a single weather dataset leads to uncertainty since that dataset might not correctly represent the current and future weather conditions, including conditions such as global warming (Hong, Chang *et al.*, 2013).

Another challenge that authors face in UBEM projects is the need for more data standardisation (Kong *et al.*, 2023); multiple data sources lead to using several data formats for data sharing. Regarding occupant behaviour, sharing between building performance simulation programs is supported by XML, JSON or YAML format (Hong, Chen, Belafi *et al.*, 2018). However, these formats still need to overcome the lack of a standard data model to standardise the representation of occupant behaviour objects and increase the interoperability between simulation platforms.

Several governments already have energy-related data available through web portals, for example, Berlin (Berlin Partner / Senatsverwaltung für Wirtschaft, 2024), Bogotá (IDECA, 2024), the city of New York 2024, the Netherlands (Kadaster, 2024d). However, there are still discrepancies between government agencies within the same country in adopting consistent data formats for those data portals while implementing taxonomies that facilitate access to UBEM-oriented dataset types.

The scale of analysis raises multiple challenges. The time resolution, spatial size (of the study area), and the level of detail are only three elements for consideration when discussing the scale of analysis. Del Ama Gonzalo *et al.* (2023) evaluate several BES tools for heating and cooling “energy consumption” based on the early stages of the design of a new building. Among their findings, authors indicate that only aggregated annual values are comparable with the building codes and energy efficiency standards. Regarding monthly values, their results surpass discrepancies of 30% compared to the ASHRAE standard.

In large study areas, simulating every building is often impractical. To address this, researchers aggregate the building stock into a few representative classes by developing building archetypes that capture the key characteristics of each group. They then simulate these representative models and extrapolate the results to calculate the energy performance of the entire building stock. Examples of these cases include Carnieletto, Bella *et al.* (2024) for several European cities; Sood *et al.* (2023) for Dublin, Ireland; Deng, Chen *et al.* (2022) for Changsha, China; Roth *et al.* (2020) for New York City, USA; Carnieletto, Ferrando *et al.*

(2021) for Italy.

2.2. DATA SOURCES FOR UBEM

The performance of UBEM is dependent on how data from multiple sources is selected and prepared. Each source plays a critical role in the accuracy of the model outcome. The following sections discuss key data sources that could serve as input data for UBEM.

2.2.1. GEOSPATIAL DATA: SEMANTIC 3D CITY MODELS

Creating, manipulating, and using 3D digital representations of real-world objects, including a mix of geometric, topological, and semantic elements, are essential aspects of 3D modelling (Arroyo Ohori *et al.*, 2024). Ideally, the modelling of real-world objects should follow a general approach that satisfies the information needs of various application fields, even though the modelled features can be combined arbitrarily. S3DCM facilitate many applications, such as urban wind and dispersion simulations, energy studies, noise studies, and various analyses that require a planned architectural design to be placed in its context (Stoter *et al.*, 2020). All definitions and contextualisation in this section focus on using s3DCM as standardised source of integrated information to manage and share data between several stakeholders for energy applications.

In my PhD research, I use the CityGML standard when referring to s3DCM from the geospatial perspective. In this standard, all concepts are based on the ontological definition of a city and the elements that compose it. Therefore, there is semantic consistency across all elements defined by the data model of CityGML. Additionally, the standard has been widely used for the representation of cities. Several examples of cities adopting the CityGML data model can be found in "awesome CityGML" (Wysocki, Schwab *et al.*, 2024) or "open cities" (3D geoinformation group, 2024) for examples around the world about its adoption. Furthermore, the Infrastructure for Spatial Information in Europe (INSPIRE) directive bases several elements of data specification of buildings on the concepts defined by CityGML (Laurent *et al.*, 2024).

2.2.1.1. CITYGML

CityGML is an OGC standard designed as an open data model for the storage and exchange of virtual 3D city models, implemented as an application schema of the Geography Markup Language 3 (GML3) (Gröger, T. Kolbe *et al.*, 2012). Initially published as an encoding standard in 2008 (CityGML 1.0), version 2.0 was released 2012

Besides the conceptual data model, several encodings exist *CityJSON* (Ledoux, Arroyo Ohori *et al.*, 2019), a JSON-based format for the CityGML data model

(current version 2.0.1), and *cjdb* (Powalka *et al.*, 2024) a database solution for storing CityJSON files into a PostGIS database; *3DCityDB* (Yao, Nagel *et al.*, 2018), a SQL-based 3D geo-database for CityGML-based 3D city models (supports XML and JSON encodings). In my research, I use CityGML 2.0 since most of the software tools currently available support this data model.

2.2.1.1.1 Modularisation

In CityGML 2.0, the core module comprises the main components and classes of the data model. It includes the *CityObject* class, which is the basic for all thematic modules. The twelve thematic extension modules that expand the core module are Appearance, Bridge, Building, CityFurniture, CityObjectGroup, Generics, LandUse, Relief, Transportation, Tunnel, Vegetation, WaterBody (Gröger, T. Kolbe *et al.*, 2012).

In particular, the *Building* module is fundamental to my research because it represents my research's main object of study: buildings. The UML class diagram of this thematic module is shown in figure 2.5. Classes *Building* and *BuildingPart* are specialisations of abstract class *AbstractBuilding*. Furthermore, the distinction between classes indicates that a *Building* can be composed of multiple *BuildingParts*.

This modelling decision brings flexibility to scenarios where two structures, such as a house and a shed on the same cadastre plot, can be classified as *BuildingParts* of a *Building* feature, which is a common case in the Netherlands. Hence, each building part is characterised autonomously, which is relevant, especially in case of the major differences in geometry, use or year of construction, to mention some examples.

Regarding the geometric representation of buildings, CityGML utilises the boundary representation method, which considers only surfaces. Buildings can have either a *Solid* or *MultiSurface* geometric representation, which is not mutually exclusive. The *MultiSurface* representation allows additionally the classification of boundary surfaces into categories such as *RoofSurface*, *OuterCeilingSurface*, *WallSurface*, *OuterFloorSurface*, *GroundSurface*, *ClosureSurface*, *CeilingSurface*, *InteriorWallSurface*, and *FloorSurface*.

This detailed semantical classification is unavailable across all Levels of Detail (LoD); for example, a LoD1 model does not include any thematic surface; a LoD3 model allows thematic surfaces such as *RoofSurface* or *WallSurface* to contain finer elements such as openings, including *Doors* and *Windows*; only LoD4 supports interior representations.

2.2.1.1.2 Multi-LoD Modelling

CityGML supports the concurrent representation of the same object (*CityObject*) using multiple LoDs. This capability allows simultaneous analysis and visualise

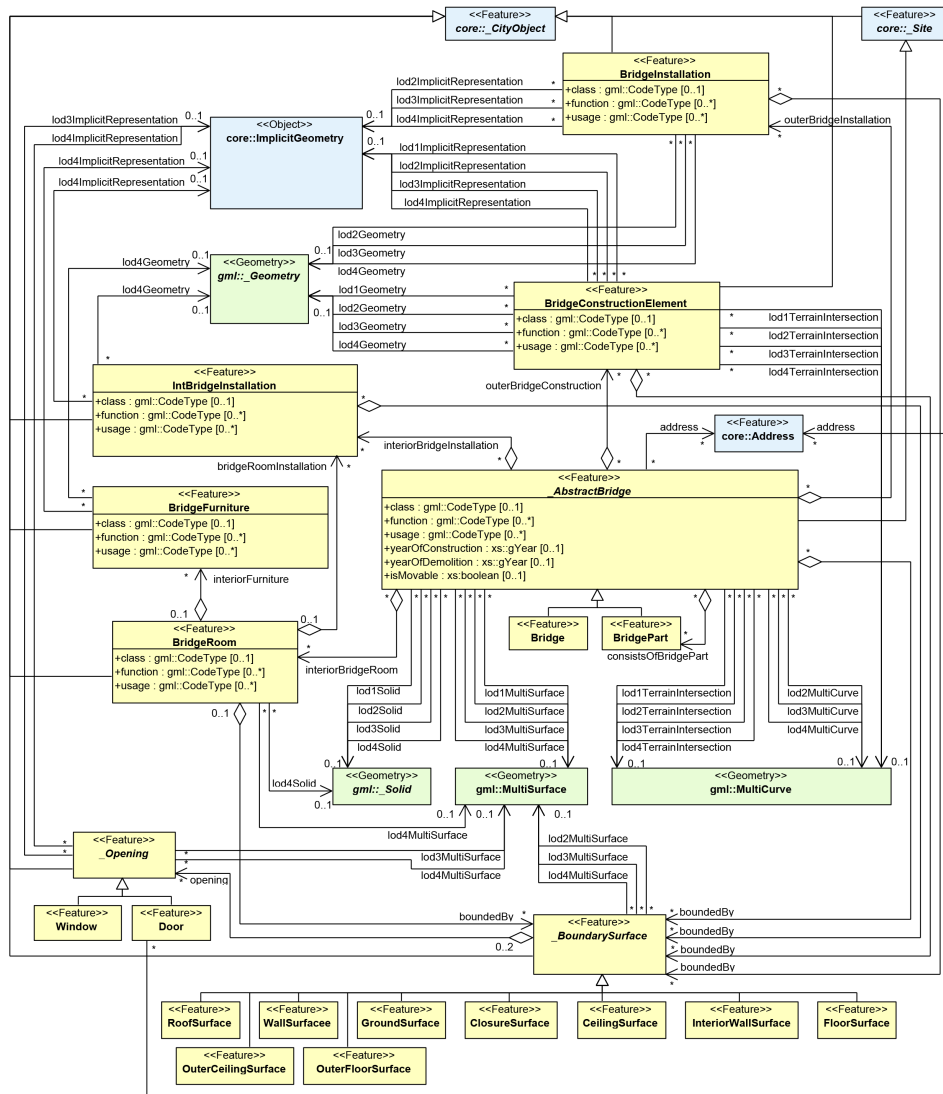


Figure 2.5: UML model representing the CityGML 2.0 *Building*, from Gröger, T. Kolbe *et al.* (2012)

the object at various degrees of resolution. The LoD classification indicates that the higher the number, the more detailed the geometric and semantic representation of the feature will be.



Figure 2.6: Example of LoDs specified in CityGML v2.0, from Biljecki, Ledoux *et al.* (2016)

The representation of *Buildings* in CityGML happens across five LoDs (see figure 2.6). *LoD0* represents the most basic form, showing only the footprint or projected rooftop of the building. *LoD1* displays buildings as simple block models. *LoD2* allows *buildings* to be represented as solids or multisurfaces. *LoD3* further refines the representation by including openings such as doors and windows. Finally, *LoD4* contains all the features of *LoD3* while adding internal details, such as rooms, stairs, and furniture.

This representation model has been extended by Biljecki, Ledoux *et al.* (2016). Although it is not officially part of the standard, it is fully supported by CityJSON (see section 2.2.1.2), and this is the geometric representation implemented by the 3DBAG (see section 3.2.1.2). However, I stick in my research to the official LoD concept of the standard since the data model to manage energy-related is based on the CityGML encodings in XML and SQL that do not support the extended LoDs (see in section 2.2.1.4).

2.2.1.1.3 Extensibility

CityGML provides two different approaches to support the exchange of data that is not defined by the existing thematic classes (Gröger, T. Kolbe *et al.*, 2012):

1. Generic objects and attributes

Generic Attributes and *Generic CityObjects* offer a flexible extension mechanism to the CityGML data model without altering its schema. *Generic CityObjects* allow the representation of urban features not specifically included in the CityGML data model, like for example a windmill. Similarly, a *generic attribute* is helpful to add additional data to any feature, such as the number of occupants of a *building* or the owner of the windmill, without modifying the CityGML structure. However, only six types are available: string, integer, double, date, uri, and measure.

2. Application Domain Extensions (ADE)

The ADE mechanism provides a robust framework for extending the standard model with explicitly modelled feature types, leading to well-defined object semantics, attributes, and relationships. This approach enhances interoperability among various stakeholders by allowing extensions to be

formally specified in a schema definition file (XSD) created by specific domain communities that refer to the CityGML XSD.

The ADE mechanism includes additional spatial and non-spatial attributes, as well as their relations and associations. When defining new feature types, it is recommended to base them on the CityGML abstract base class *CityObject*. To extend existing CityGML classes, the best practice is to create subclasses using the "hook" mechanism in its XML schema definition, ensuring seamless integration and consistency within the model. Biljecki, Kumar *et al.* (2018) list in their paper review, 44 published ADEs.

2.2.1.2. CITYJSON

CityJSON is a JSON-based encoding of a subset of the CityGML data model. It is an OGC community standard from 2023 (Ledoux and Dukai, 2023). The CityGML data model has been flattened to reduce its hierarchy, complexity and storage size. Therefore, there are two kinds of City Objects, *1st-level* City Objects that cannot have a parent, and *2nd-level* City Objects that require a parent to exist figure 2.7.

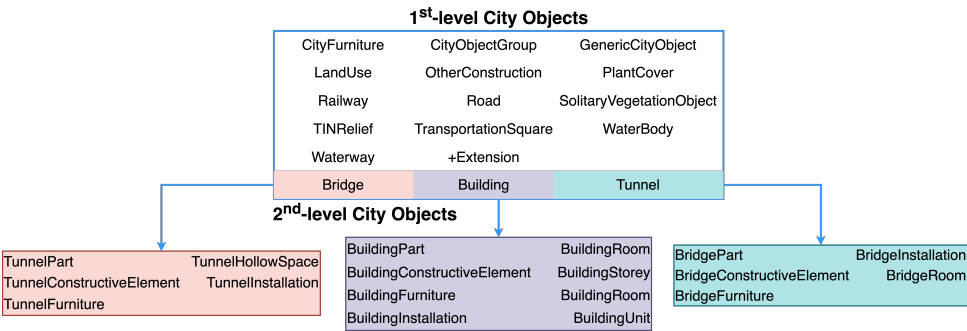


Figure 2.7: Implemented CityJSON classes

CityJSON 1.x supports a subset of CityGML v2.0, and CityJSON 2.0.0 partially supports a subset of CityGML v3.0, which is presented as a graphical overview in figure 2.8. Although this encoding does not modify the definition of LoDs, CityJSON editors encourage the use of the refined LoDs by Biljecki, Ledoux *et al.* (2016) (Figure 2.9).

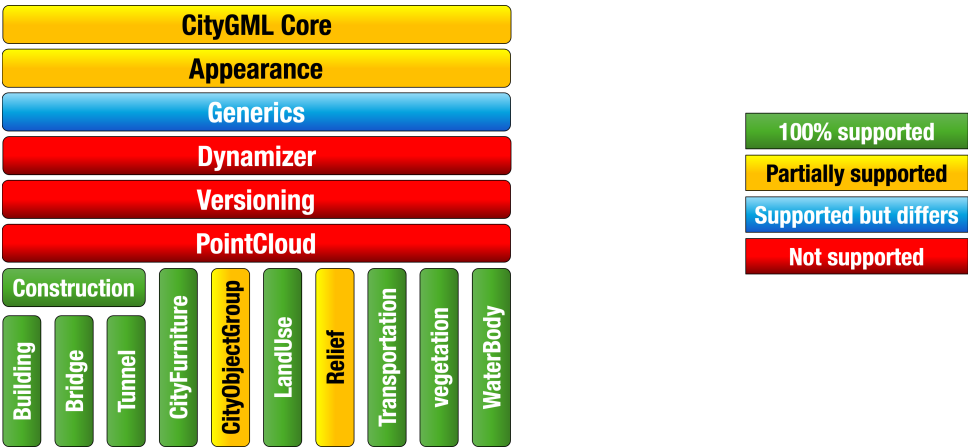


Figure 2.8: Graphical summary of the CityGML modules supported by CityJSON 2.0

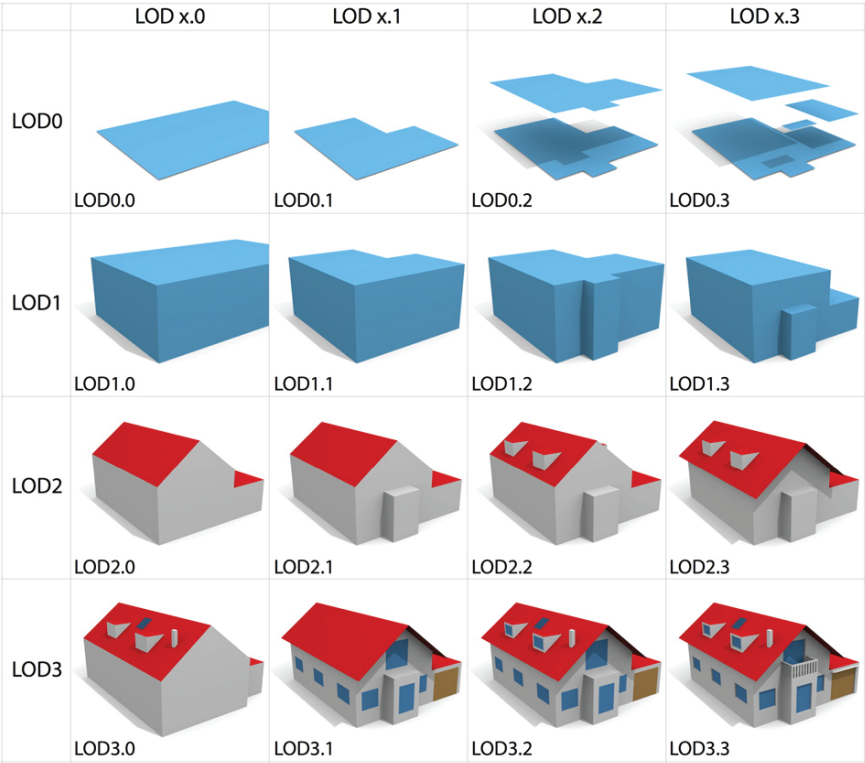


Figure 2.9: LoDs supported by CityJSON, from Biljecki, Ledoux *et al.* (2016).

2.2.1.2.1 Extensions

One of the design philosophies of JSON is being "schema-less," meaning that new properties can be defined for JSON objects without requiring documentation (Ledoux and Dukai, 2024). However, CityJSON supports a conceptual extension mechanism that is similar to the one in CityGML. A CityJSON Extension results in a JSON file that details how the core data model of CityJSON is extended and is used to validate CityJSON files. Unlike the ADE mechanism, CityJSON restricts schema extensions to the following cases:

1. Defining new properties at the root of a document
2. Adding attributes to existing City Objects
3. Defining a new Semantic Object
4. Defining a new City Object or extending an existing City Object

According to (Ledoux and Dukai, 2024), an *Extension* is used to enforce specific properties, attributes, or *CityObjects* in CityJSON objects. Just like the case of ADEs, the authors suggest that Extensions are beneficial in cases of data exchange between multiple data producers and consumers.

2.2.1.3. SQL ENCODING

The 3DCityDB consists of a database schema designed for spatially enhanced relational database management systems (ORACLE Spatial/Locator, PostgreSQL/PostGIS, PolarDB/Ganos). It includes a set of database procedures and software tools that allow users to import, manage, analyse, visualise, and export virtual 3D city models according to the CityGML standard (Yao, Nagel *et al.*, 2018).

The implemented database schema is the result of the manual mapping of CityGML classes and data types onto tables. This approach reduces database complexity, improves operating performance, and enhances semantic interoperability. By following this method, the authors mapped the CityGML class hierarchy onto tables.

Additionally, the "3DCityDB Tools"¹ is a QGIS plugin designed for connecting to a 3DCityDB (v. 4.x) instance utilising PostgreSQL/PostGIS. This plugin enables the importation of all CityGML 2.0 LoDs from multiple 3DCityDB schemas into a QGIS project. Additionally, it supports editing of feature attributes and offers functionalities for deleting features and cleaning up database schemas.

2.2.1.4. ENERGY APPLICATION DOMAIN EXTENSION – ENERGY ADE

The CityGML Energy Application Domain Extension –hereafter referred to as Energy ADE– extends the CityGML 2.0 data model to support energy-related data. It provides a standardised data model to allow single-building energy

¹It has been published as an MSc thesis in Geomatics at TU Delft titled "Development of a QGIS plugin for the CityGML 3D City Database" by Pantelios (2022)

simulations and country-wide energy assessments focused on the building sector (Casper *et al.*, 2018). The data model of the Energy ADE is split into seven functional modules, which are shown (along with their dependencies) in figure 2.10. The colour code per package of the diagram is followed as well in following the figures:

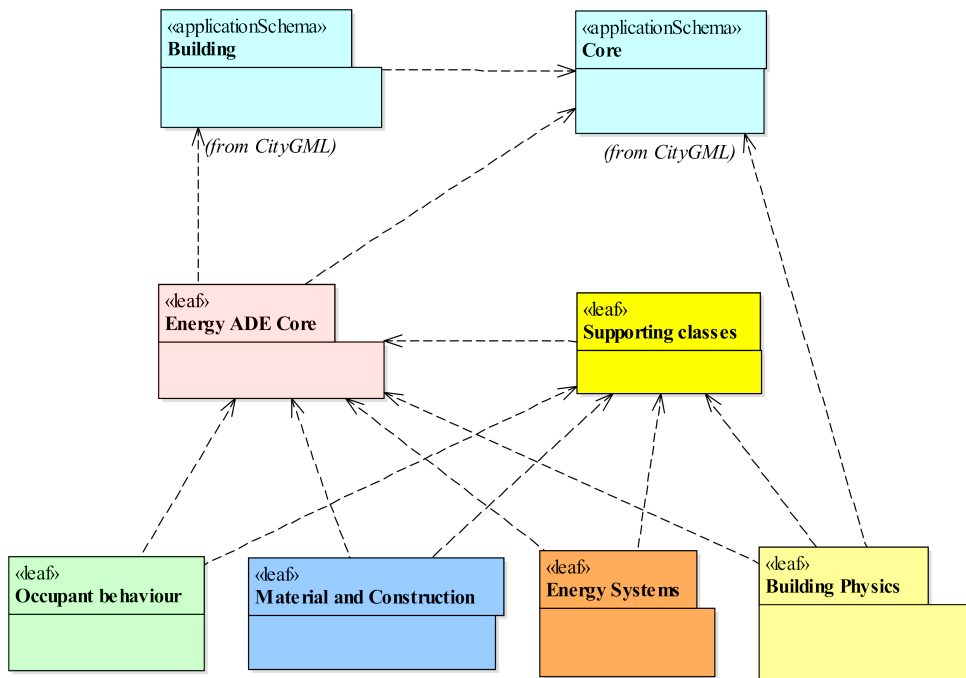


Figure 2.10: Colour-coded modular structure of the Energy ADE UML diagram.
From (Bachert *et al.*, 2024)

- **Core Module:**

It is the foundation module of the entire Energy ADE. It extends CityGML classes with energy-relevant properties and provides abstract base classes for the other functional modules. The Core module defines data types, enumerations, and codelists used across multiple modules. It extends the CityGML classes *_AbstractBuilding* and *_CityObject* with new properties. Furthermore, it facilitates the connection of city objects with energy demand parameters (*EnergyDemand*) and weather data (*WeatherData*). The UML diagram is shown in figure 2.11.

- **Building Physics module:**

This module focuses on providing essential input data for detailed simulations of a building's thermal behaviour. It introduces the classes *ThermalZone*, *ThermalBoundary*, and *ThermalOpening*, which can be related to CityGML classes like *Room*, *_BoundarySurface*, and *_Opening*. A *ThermalZone* represents a reference volume for heating and cooling

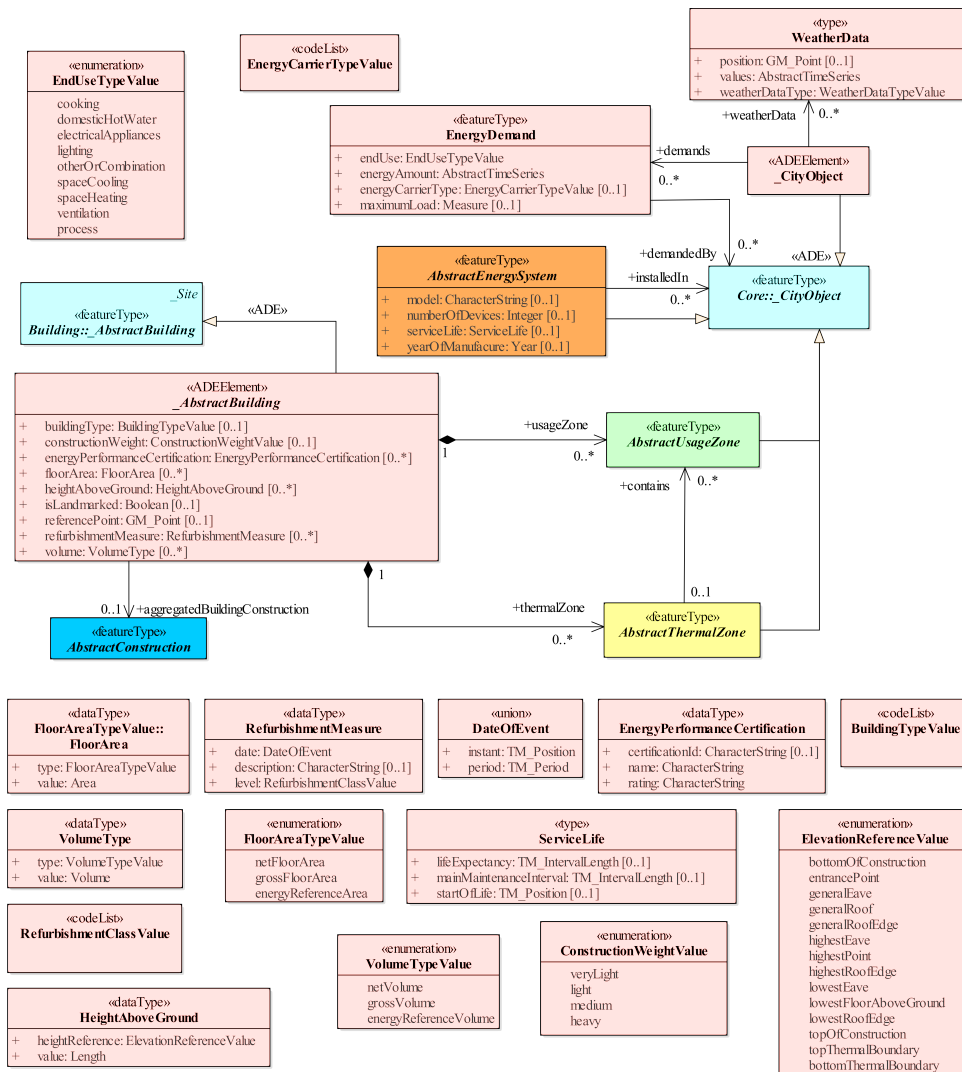


Figure 2.11: Energy ADE core UML diagram. From (Bachert *et al.*, 2024)

demand calculations. *ThermalBoundary* objects separate thermal zones from each other and the environment, while *ThermalOpening* objects model parts of a thermal boundary that allow radiation energy transfer.

- **Occupant behaviour module:**

This module enables the modelling of building occupants and their behaviour relevant to energy simulation. The main class, *UsageZone*, represents regions of homogeneous usage and is referenced by a *ThermalZone*. A *UsageZone* can be subdivided into *BuildingUnit* objects holding ownership

information.

- **Material and Construction module:**

This module characterises the physical properties of building construction parts, detailing their structure and thermal/optical properties. The central class is *construction*, which can model a base construction or a *ReverseConstruction* involves an inverted order of layers. Each layer can be further categorised by the physical properties of the materials it is made of.

- **Energy Systems module:**

This module represents energy forms and systems for energy demand and supply analyses. It relates to the Energy ADE and CityGML through the *EnergyDemand* class and the *AbstractEnergySystem* class, which can be connected to any *_CityObject*. *AbstractEnergySystem* serves as the base for energy conversion systems, distribution systems, storage systems, and emitters. The module includes special classes for specific conversion systems like boilers, heat pumps, photovoltaic systems, and solar thermal systems.

- **Supporting Classes module:**

In addition to these modules, the Energy ADE includes supporting classes for representing time series of physical data (*AbstractTimeSeries*), schedules (*AbstractSchedule*), and weather data. These supporting classes enhance the ability to model dynamic aspects of urban energy systems.

A deeper description of the Energy ADE is beyond the scope of this thesis. Further details are available in (Agugiaro, Benner *et al.*, 2018). The Energy ADE has already been used in several national and international projects (Geiger *et al.*, 2022; Malhotra, Shamovich, Frisch *et al.*, 2019; Rossknecht and Airaksinen, 2020). Due to the completeness of the Energy ADE data model and the several experiences gathered since its release in 2018, I chose it to manage the data generated during my PhD research.

2.2.1.4.1 3DCityDB Extension

The 3DCityDB supports CityGML ADE; among the ADEs that are fully supported by this encoding as well from the Importer/Exporter tool are:

- Energy ADE (Yao and Nagel, 2022). The support is tailored to the Karlsruhe Institute of Technology (KIT) profile (Benner, 2018, pages 19-29).
- Quality ADE (Nagel, 2024b). It adds support for managing the Quality ADE version 0.1.4.
- i-UR ADE (Nagel, 2024a). It adds support for managing the i-Urban Revitalization version 1.4

2.2.1.4.2 CityJSON Energy Extension²

Tufan et al. 2022 propose a conversion mapping of the Energy ADE KIT profile (section 2.2.1.4). However, it is not a direct translation due to the structural differences between the XML and JSON data formats. There are several reasons behind this difference: first, a simplification of hierarchical structure where feasible; this means the exclusion of abstract classes since most of them do not include any attributes to pass on to their subclasses. However, that is not the case for the *AbstractTimeSeries* object, which has a *variableProperties* attribute to pass on to *RegularTimeSeries* and *RegularTimeSeriesFile* objects. Only in this case do both objects have the same attribute in the Extension schema.

Second, an alignment with data requirements for the energy heat demand calculation (test case) has influenced the model's design. The data types were simplified, including the removal of the units of measurement due to the GML measure type not being supported, and new attributes were added to comply with the use test case.

Third, CityJSON extension mechanism is limited to the definition of new CityObjects, attributes, and root properties. Consequently, non-CityObjects had to be incorporated under the *extraCityObjects* property. However, this is a mechanism intended solely for CityObject-type elements. Our decisions do not cause validity errors, but the design decision results in less clarity in the Extension schema since the types of stored objects may not always match the name of the container property in the Extension.

2.2.2. NON-GEOSPATIAL DATA

2.2.2.1. BUILDING PHYSICS DATA

Building physics data are vital for the characterisation of energy-related aspects of buildings. These data include construction assemblies and details about Heating, Ventilation, and Air Conditioning (HVAC) systems (Ali, Bano et al., 2024b). Building physics data provide insights into the thermal properties of buildings and the characteristics of their systems, all of which are necessary for accurate "energy consumption" predictions. For example, information on the heat transfer coefficient (U-value) of roofs, walls, and windows and the solar energy transmittance of windows (g-values) directly influence the building's energy performance (Malhotra, Shamovich, Raming et al., 2022).

One of the significant challenges in UBEM is data collection and representation of building physics, particularly given the diverse building compositions within an urban area (C. Wang, Ferrando et al., 2022). Each building type has varying construction assemblies and HVAC systems; their variation comes

²This section has been published in the MSc Thesis in Geomatics at the TU Delft titled "Development and Testing of the CityJSON Energy Extension for Space Heating Demand Calculation" by Tufan (2022)

from multiple reasons, such as the construction/renovation year, building size, economic destination, and regulations. This diversity makes data collection and harmonisation complex and time-consuming since UBEM requires detailed building information.

Therefore, archetypes are a common "plan B", i.e. a fallback when data are not available for UBEM. Archetypes group buildings with similar properties, simplifying the data collection by allowing researchers to define the physical characteristics of buildings for entire classes rather than each building (Ali, Bano *et al.*, 2024a). The creation of archetypes involves segmentation, where buildings are divided into classes based on characteristics like age, type, and climate zone, followed by characterisation, where detailed data such as U-values and HVAC characteristics are assigned to each group.

For example, the Intelligent Energy Europe (IEE) TABULA project, hereafter referred to as TABULA, was established to improve the refurbishment of the European housing sector. The project began by defining uniform European-wide building archetypes. This classification was extended by elaborating building stock models to assess the refurbishment process and forecast future "energy consumption". The project performed case studies in 16 countries to track the implementation of energy-saving measures and their influence on consumption in practice. One of the project's outputs is the TABULA WebTool, a web platform that contains the physical characteristics of the residential building stock (EPISCOPE Project, 2017).

In the Netherlands, the characterisation of the housing stock for the TABULA project was in charge of the Ministry of the Interior and Kingdom Affairs (in Dutch "*Ministerie van Binnenlandse Zaken en Koninkrijksrelaties*") (Agentschap NL, 2011). In 2022, the Netherlands Enterprise Agency (in Dutch "*Rijksdienst voor Ondernemend Nederland*"–RVO–) published an updated version of this characterisation to support the improvement of the energy performance of the Dutch housing stock (RVO, 2023). Both characterisations follow the same approach and define classes based on the building type and period of construction. The resulting classes are considered in the example homes for existing construction (in Dutch "*Voorbeeldwoningen bestaande bouw*") report and provide an insight into the building stock and the performance of buildings based on archetypes. Data have been published for each of the building classes, including U-values, g-values, airflow, ventilation type, heating, and hot water systems. However, for the latter three ones, these are generic for all building archetypes and only vary when based on the type of renovation. Furthermore, neither the TABULA project nor the example homes for existing construction have proposed or published a standardised data model for managing these types of data.

In 2013, the SHAERE report was published (Delft University of Technology, 2016). This dataset has been used to evaluate the possible paths to reach the Dutch targets for TABULA in improving the energy performance of buildings.

For the national target, the Net Zero Energy Building (NZEB) level has been considered as the scope for the construction and renovation of buildings.

2.2.2.2. OCCUPANCY DATA

Occupancy highly influences the energy demand of a building. People's behaviour plays a significant role in indoor conditions based on their comfort preferences (Page, 2007). For example, in an office, people use lightning and electrical devices for their activities. Furthermore, they add internal heat to the thermal zone and might consume hot water (Ren *et al.*, 2024). Driven by indoor preferences, occupants may adjust the ventilation for thermal comfort and health reasons, and these actions influence HVAC energy use. Additionally, occupants adjust shadings or lighting appliances for visual comfort.

The most common method to include the occupants' influence in BES is the definition of occupancy schedules and patterns that are represented by profiles. These profiles are usually represented as a time-dependent series of zero to one (0 to 1) values, representing the variation of a total value of the variation over the period based on the presence of people. For example, the indication of internal heat gains from people and the use of appliances and lighting. Schedules refer to diversity in the usage hours based on the day of the week (working days, weekends) or the season (winter, summer) (Doma and Ouf, 2023). However, these schedules and patterns need more detail and accuracy since they represent, in a general manner, the characteristics of human behaviour.

Dabirian *et al.* (2023) propose a stochastic archetype modelling of the building stock on a city level based on occupant-related schedules extracted from time-series of electricity consumption data of buildings. They used the data collected from 3053 energy meters from 19 locations in Europe and North America between 2016 and 2017 with hourly time resolution. The data contains information such as gross floor area, year of construction, and the building use. They then clean and standardise the data. Their method derived a plug load profile for each building type in specific climate zones. The BES uses these values for the building archetype and district scale.

Samareh *et al.* 2024 implement a method to create an occupancy model based on signal data derived from Wi-Fi sensing technology by deriving the number of people in a building. Their results are used as input parameters in their BES tool. Their method uses various machine learning techniques and achieves a test accuracy of 77%. However, they use data augmentation techniques due to the limited availability and diversity of data in their original dataset. After the augmentation, their ML models achieved an accuracy of 91%.

Mosteiro-Romero, Hischer *et al.* (2020) propose a method to evaluate the impact of building occupants on district-scale energy models. Their population-based approach (PopAp), inspired by agent-based transportation modelling, draws on class and employee registers to simulate occupant presence and its effect on

district-level energy demand. When compared to conventional deterministic and stochastic approaches, PopAp yielded a 33% higher maximum occupant density in the studied area. The difference in the annual energy demand values between models remained within 10% for all demand types, while hourly values increased up to 15%. This hourly-level discrepancy is particularly significant, given the critical role of temporal granularity in predicting peak energy demand.

2.2.2.3. WEATHER DATA AND CLIMATE DATA

In a given location, short-term (minutes, hours, days, seasons) meteorological state of the atmosphere refers to weather data. Long-term (~ 30 years)) average weather conditions refer to climate data (Herman and Golberg, 1978). Weather data are a crucial input for UBEM, and their accuracy influences the reliability of building-energy simulations. Factors such as outdoor temperature, relative humidity, wind, and solar radiation directly affect the energy demand of buildings. Nevertheless, the kind of data used varies according to the scope of the research. For monitoring, managing, or validating the actual energy performance of a specific building, researchers employ observed meteorological data, often called actual meteorological year (AMY files) (Bianchi and Smith, 2019; Hong, Chang *et al.*, 2013). By contrast, baseline studies and typical UBEM performance assessments generally rely on *typical meteorological year* (TMY) datasets (Mohajeri *et al.*, 2016; Romero Rodríguez *et al.*, 2017). The TMY files are derived from long-term climate records; in the case of the Netherlands, these files are generated from the observations of the KNMI automatic weather stations from an average 30 years observations (KNMI, 2024a). Therefore, each simulation adopts the same “typical” weather data associated with its nearest station.

Yang *et al.* (2023) research the impact of using multiple weather datasets and the urban morphology in urban building energy simulation for a study area of 41 urban blocks. The input weather data was collected from one weather station and 16 climate sensors, with fewer sensors installed across the study area, which were later compared to a multi-source TMY dataset (EPW file). The difference between the measured weather station data and the microclimate in the study area was 4.9°C. The weather data performed better than EPW data in reflecting the urban microclimate and predicting the building “energy consumption”.

N. Luo *et al.* (2022) propose a framework to integrate CityBES, a UBEM platform, and CityFFD and urban microclimate model (UCM) tool to facilitate data exchange between the two simulation tools. By doing so, they integrate microclimate parameters into UBEM simulations. Their case study, conducted in a district of 97 buildings in San Francisco, USA, demonstrates the influence of using urban microclimate data compared to TMY data. The results show that using UCM data yields an average building facade temperature difference of 5.3°C, 8.9°C higher air temperature, and a 19.5% increase in cooling energy demand.

UBEM is a valuable tool to evaluate the impact of climate change in urban areas; one example is the research of Deng, Javanroodi *et al.* (2023), which investigates how climate change will affect the energy performance of buildings in Geneva, Switzerland. For this purpose, they used a BES tool named Automated Building Performance Simulation (AutoBPS) to simulate the thermal "energy consumption" for space heating and cooling. The models were calibrated using annual heating consumption data available per building and evaluated four different climate change scenarios for the year 2050. Their results indicate a space heating consumption decrease of 21%-31%, while the space cooling consumption increases by 95%-144%.

2.3. DATA QUALITY

Previous sections have discussed the different data requirements in UBEM. However, there is no mention of data quality and its relevance. Among the challenges to obtaining accurate and complete datasets at the urban scale are:

- **Data acquisition**
While geometric information can be captured for individual buildings through design documents or field measurements (C. Wang, Ferrando *et al.*, 2022), these techniques are impractical city-wide. A small city in the Netherlands such as Delft ($\approx 54,594$ buildings, of which $\approx 37,371$ are residential, for a population of $\approx 110,173$ (van Bijsterveld, 2025)) shows the logistical barrier to manual data collection.
- **Impact on model reliability**
The usefulness of UBEM for design and policy is based on reliable simulation results. Inaccuracies in input data propagate through the model: at hourly resolution, prediction of errors for single buildings can exceed 40%. However, when annual resolution, city-wide errors are acceptable (Nouvel, Zirak *et al.*, 2017).
- **Data completeness**
Open databases are available for relatively few countries, leading to a limited broader applicability. Furthermore, these datasets are not always complete some reasons include legal restrictions (limited licensing) or data sensibility.
- **FAIR principles**
Despite data availability, another element to consider is the following of the FAIR (Findable, Accessible, Interoperable, Re-usable) practices (Utrilla Guerrero and Barrera-Coronel, 2025). Not including these practices adds another layer of complexity on data quality.
- **Generalisation of dynamic elements**
Simplified occupant schedules and assumed HVAC systems introduce systematic bias; general-purpose assumptions often diverge from real

behaviour and installed equipment (C. Wang, Ferrando *et al.*, 2022). Missing refurbishment status can lead to heating-demand overestimation of 70-100% (Nouvel, Zirak *et al.*, 2017). Similarly, incorrectly managed attics or basements may cause $\approx 20\%$ error.

- **Automated validation**
Both geometries and attributes require validation to ensure their usability. For 3D geometries, high geometrical accuracy leads to good solids, which lead to geometrical validity. Tools such as CityDoctor (Pries *et al.*, 2021), FME (Safe Software, 2024), and val3dity (Ledoux, 2018) detect and, in some cases, repair geometric errors in 3D city models. Their diagnostic reports aid in prioritising data-cleaning efforts before energy simulation. For attributes,
- **Geometric detail and solar gains**
LoD1 or LoD2 geometric representation and uncertain window-to-wall ratios affect solar-gains computations and transmission losses, directly influencing heating-demand calculations.
- **Scarcity of high-resolution validation data**
Hourly measured energy data are rarely available (C. Wang, Duplessis *et al.*, 2024), hindering rigorous validation. Techniques such as multivariate imputation or weighting schemes can improve robustness when survey data are partial.
- **Sensitivity, error propagation, and continuous improvement**
Sensitivity analyses identify which input parameters most affect simulation accuracy and therefore merit targeted data collection (Nouvel, Zirak *et al.*, 2017). Ongoing assessments of error propagation support systematic correction of outdated or inconsistent datasets (Geske *et al.*, 2023).

2.4. ENERGY BALANCE METHOD

The energy balance method for the computation of the energy demand of buildings is based on the *law of conservation of energy*, which states that, for a closed system, the total energy inflow equals the total energy outflow (Britannica, 2024). In UBEM, a building is treated as a closed system: energy entering must match energy leaving. If net energy outflow is negative, the building requires additional heat to maintain the set-point temperature; if it is positive, excess heat must be removed by cooling (Van Bueren *et al.*, 2012, p. 123).

The system boundary is defined by the building's external walls, roof, and ground-floor surfaces. Where detailed zoning data are unavailable, the entire building is assumed to behave as a single thermal zone with uniform temperature and humidity. All conduction, convection, and radiation processes are consolidated into four principal heat flows (presented in figure 2.12): solar gains, infiltration

and ventilation, internal gains and transmission through the envelope. For analyses using coarse time steps (monthly or longer), short-term heat storage and release within building elements can be treated in a more simplified way, as they are "high-frequency", circadian phenomena that are hard to model with the low-frequency monthly computations.

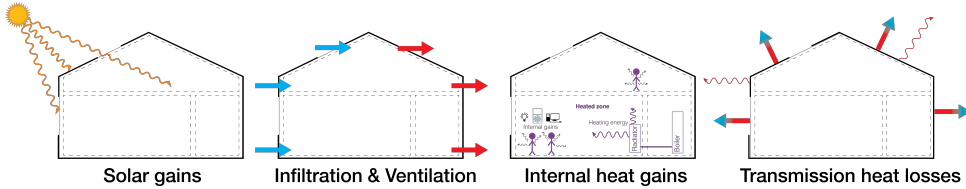


Figure 2.12: Energy balance of a building related to the main energy transfer processes

Energy performance calculations for buildings using this method are conducted over discrete time intervals, such as an hour, month, heating/cooling season, and year. For example, processes like domestic hot water are computed for an entire year, using fundamental constants defined for a typical year (ISO, 2017).

The total energy exchange for space heating $Q_{H,nd}$ of a building is the difference between heat losses (transfers) and heat gains (Figure 2.13) and is computed using equation (2.1).

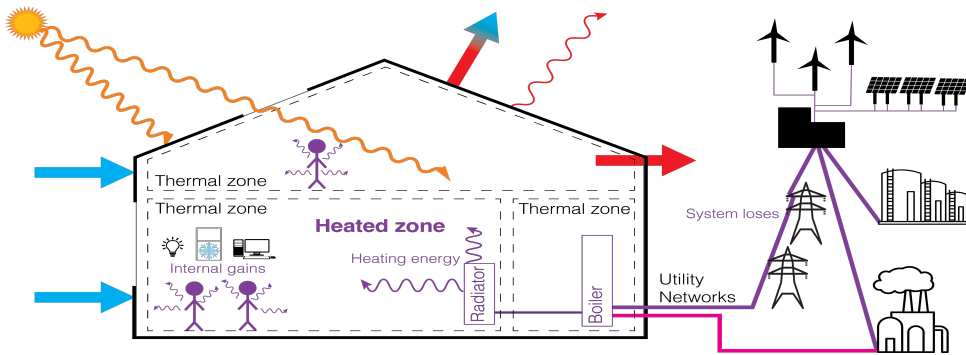


Figure 2.13: Graphical overview of the energy balance method

$$Q_{H,nd} = (Q_{H,tr} + Q_{H,ve}) - \eta_{H,g} \cdot (Q_{int} + Q_{sol}) \quad (2.1)$$

Where:

- Calculation of heat losses
 - Through the boundary surfaces ($Q_{H,tr}$)
 - Heat flows from areas of higher temperature to areas of lower

temperature until they reach equilibrium. This movement occurs through three methods: convection, radiation and conduction. In the case of convection, heat transfers via liquids or gases. For radiation, any object with a temperature above 0 K emits heat through electromagnetic waves. When these waves reach another object, they are converted into heat.

The amount of heat radiated depends on the temperature of the emanating object. Finally, heat conduction occurs as molecules vibrate in solids. As the temperature rises, molecular vibrations become faster, transferring heat to the neighbouring colder molecules. Due to the poor conductivity of gases, solids and liquids are the major means of transportation.

– Periodical airing ($Q_{H,ve}$)

Van der Linden *et al.* (2018) highlight that ventilation is essential for maintaining good indoor air quality, which directly impacts occupant health. Therefore, building regulations indicate minimum ventilation requirements. Ventilation can be provided either naturally or mechanically, and it is designed to ensure an adequate exchange of indoor and outdoor air.

In contrast, infiltration refers to unintentional air exchange through leaks in the building envelope. Furthermore, it can result in irregular air distribution, causing draughts and energy waste. The extent of infiltration depends on the air permeability of the building's outer walls, ground and roof surfaces. While ventilation systems are controlled, infiltration is a consequence of construction quality and materials.

- The calculation of heat gains is based on the solar irradiation of the building, its internal heat sources and the utilisation factor for the heat sources. These elements are explained as follows:

– Solar irradiation (Q_{sol})

Several factors influence the amount of solar gains in a building, including the orientation of the outer wall and the shading resulting from the building itself and the surrounding objects. Other factors that influence the potential for solar gains are the physical characteristics of the walls and the use of protection elements such as sunscreens (Van der Linden *et al.*, 2018).

– Internal heat sources (Q_{int})

Heat gains are internal loads of sensible heat. Among the internal heat sources are lighting, occupants, machines, equipment, and all kinds of processes in the room (Etheridge, 2010; Moser *et al.*, 2001). For simulation purposes, the time-dependent energy released from these sources is defined by schedules.

– Utilisation factor for heat gains ($\eta_{H,g}$)

It is a reducing factor of the total heat gains in the monthly calculation method to obtain the resulting reduction of the building energy needed for heating. The inclusion of this reducing factor considers the dynamic effects of the gain utilisation factor for heating and the heat transfer utilisation factor for cooling. In the case of the cooling mode, if the losses are negative, there is no utilisation. Norm-based computations such as the NTA 8800 define this factor as a function of the heat-balance ratio of the heat gains H_{gn} (NEN, 2024). However, the effect of inertia in the case of intermittent heating or cooling or the case of switch-off is taken into account separately.

More details about each one of the above-mentioned items will be provided in the following section, with particular attention to how the NTA 8800 prescribes the computation of each item.

2.5. NTA 8800

2.5.1. INTRODUCTION

As of 1st January 2021, all new buildings in the Netherlands are required to comply with the Almost Energy Neutral Buildings directive (in Dutch *"Bijna Energie Neutrale Gebouwen"* – BENG–) (RVO, 2024). BENG provides a foundation for the principles that define the Dutch Technical Agreement (in Dutch *"Nederlands Technische Afspraak"* –NTA–) 8800. The NTA 8800 norm embodies the official method for determining the energy performance of dwellings in the Netherlands. Since its first release in July 2020 (NEN, 2020), the norm has been updated annually, with the current version, at the time of writing, being the NTA 8800:2024, published on 1st January 2024 (NEN, 2024).

For the calculation of the energy performance of a building, the norm first defines how to compute the net energy demand and then how to obtain the primary energy demand, which is eventually used to assess the energy performance of the building.

According to the NTA 8800, the net energy demand is heat that must be supplied to a calculation zone to maintain the desired temperature during a specific period and depends on the physical characteristics of the building. The primary energy demand is defined as the energy used to produce the power supplied to a building that has not been subjected to any conversion or transformation process. Also, It depends on the quantities of energy purchased and exported by the energy carriers, using conversion factors. The following requirements have been stated by the norm for the calculation of the primary energy demand of buildings:

- "Energy consumption" per building function
- Energy delivered by each energy carrier, e.g. gas, oil, electricity
- Energy collected [solar, wind] on-site (location of the building)

- Renewable energy used for heating, cooling and hot tap water
- Primary energy factors (fixed values by the norm) per energy carrier, including the produced on-site

Although the NTA 8800 presents the method for the calculation of the primary energy demand of buildings, in my PhD research, I focus on the net heat demand mainly due to the lack of data at the country level, which affects the calculation of the net energy demand and, even more, of the primary energy demand. In the following sections of this thesis, I discuss the reasons for my decision. However, no modifications of the method presented by the norm have been done during my PhD research. I have taken the simplifications presented by the norm for those cases in which there is no access to detailed data of the building, i.e., access to the clean energy sources such as solar panels, the number of calculation zones inside a building or the number of tabs (for water) per calculation zone.

As a remark, several of the parameters mentioned above are not open and available at the country level. I am using [figure 2.14](#) as a visual guidance to facilitate the navigation through the workflow and formulas in [section 2.5.2](#).

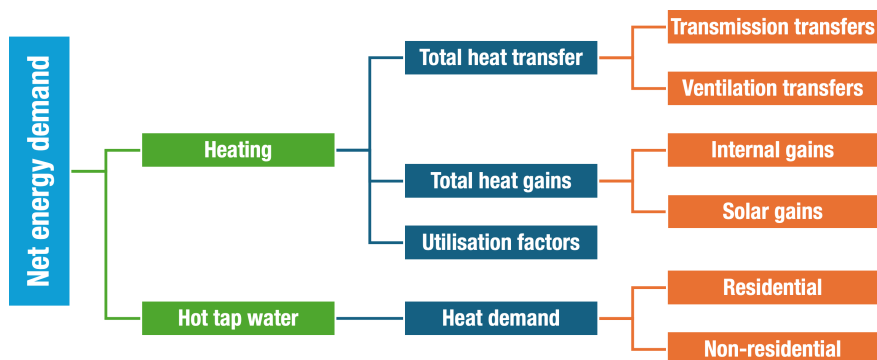


Figure 2.14: Simplified hierarchical diagram of the NTA8800

2.5.2. NET HEAT DEMAND CALCULATION

According to the NTA 8800, a calculation zone z_i is a building or part of a building that is considered as one entity for the calculation of "energy consumption" for heating, cooling, humidification, dehumidification and ventilation. [Figure 2.15](#) shows some examples of calculation zones for the same building. Each section begins with a colour-adjusted version of [figure 2.14](#) highlighting its correspondence with the NTA 8800 net energy demand overview to provide better context of the complete workflow.

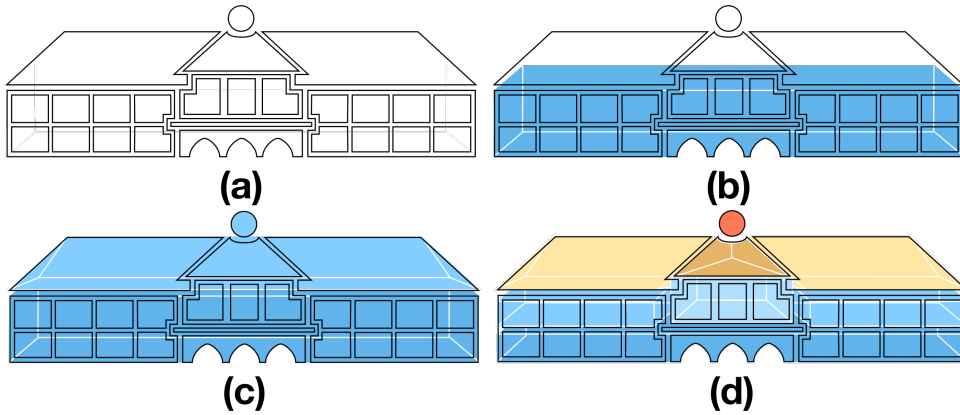


Figure 2.15: Examples of calculation zones in a building. (a) No calculation zones; (b) one calculation zone for the first two storeys; (c) one calculation zone for the whole building; (d) multiple calculation zones, each differentiated with a different colour

For each calculation zone z_i and month m_i , the monthly energy requirement for heating, $Q_{H;nd;z_i;m_i}$, in kWh, is calculated using [equation \(2.2\)](#):
If:

$$Q_{H;nd;z_i;m_i} = \begin{cases} 0, & \text{if } \gamma_{H;z_i;m_i} \leq 0 \text{ and } Q_{H;gn;z_i;m_i} > 0 \\ 0, & \text{if } \gamma_{H;z_i;m_i} \leq 2 \end{cases} \quad (2.2)$$

$Q_{H;nd;z_i;m_i}$ Monthly energy required for heating for the calculation zone z_i and month m_i as defined below, in kWh

$\gamma_{H;z_i;m_i}$ Dimensionless heat balance ratio for heating, see [equation \(2.37\)](#)

In other cases:

$$Q_{H;nd;z_i;m_i} = Q_{H;ht;z_i;m_i} - \eta_{H;gn;z_i;m_i} \cdot Q_{H;gn;z_i;m_i} - \Delta\eta_{H;gn;z_i;m_i} \cdot Q_{H;gn;z_i;m_i} - \eta_{H;gn;z_i;m_i} \cdot (Q_{H;ls;rbl;z_i;m_i} - Q_{C;ls;rbl;z_i;m_i}) + Q_{W;nd;z_i;m_i} \quad (2.3)$$

With the lower limit as: $Q_{H;nd;z_i;m_i} \geq 0$

$Q_{H;ht;z_i;m_i}$ Total heat transfer for heating in kWh, see [section 2.5.3](#)

$\eta_{H;gn;z_i;m_i}$ Dimensionless utilisation factor for the heat gain, see [section 2.5.7](#)

$Q_{H;gn;z_i;m_i}$ Total heat gain for heating in kWh, see [section 2.5.3](#)

$\Delta\eta_{H;gn;z_i;m_i}$	Difference in utilisation factor for the heat gain when calculated with and without taking into account the internal heat gain due to recoverable losses from or to the space heating system. Simplified as 0
$Q_{H;ls;rbt;z_i;m_i}$	Sum of all recoverable losses from or to the space heating system in calculation zone z_i , and month m_i . Simplified as 0
$Q_{C;ls;rbt;z_i;m_i}$	Sum of all recoverable losses from or to the space cooling system in calculation zone z_i , and month m_i . Simplified as 0
$Q_{W;nd;z_i,m_i}$	Sum of the net heat demand for hot tap water, see section 2.5.8

European norms use the term *heat transfer* instead of *heat loss* when heat loss is negative due to a higher average outdoor temperature than the set point temperature for heating for the calculation zone. The following sections will provide the details on each of the terms of [equations \(2.2\)](#) and [\(2.3\)](#).

2.5.3. TOTAL HEAT TRANSFER FOR HEATING

The main elements of the total heat transfer for heating are shown in [figure 2.16](#). Additionally, the heat transfer through heating is defined in [equation \(2.4\)](#).

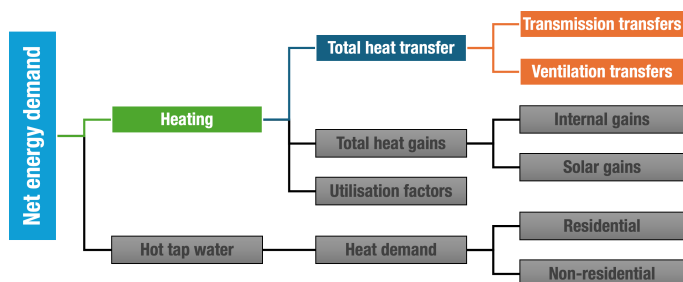


Figure 2.16: Heat transfer through transmission

$$Q_{H;ht;z_i;m_i} = Q_{H;tr;z_i;m_i} + Q_{H;ve;z_i;m_i} \quad (2.4)$$

Where:

$Q_{H;tr;z_i;m_i}$	Total heat transfer by transmission for heating in kWh, determined using equation (2.5)
$Q_{H;ve;z_i;m_i}$	Total heat transfer through ventilation for heating in kWh, determined using equation (2.15)

2.5.3.1. HEAT TRANSFER THROUGH TRANSMISSION

The total heat transfer by transmission for heating $Q_{H;tr;z_i;m_i}$ in kWh is determined by [equation \(2.5\)](#).

$$Q_{H;tr;z_i;m_i} = \left[(H_{H;tr(excl.gf);z_i;m_i} + H_{g;an;z_i,m_i}) (\theta_{int;calc;H;z_i;m_i} - \theta_{e;avg;m_i}) \right] \cdot 0.001 \cdot t_{m_i} \quad (2.5)$$

In which:

$$H_{H;tr(excl.gf;m_i);z_i;m_i} = H_{H;D;z_i;m_i} + H_{H;U;z_i;m_i} + H_{H;A;z_i;m_i} + H_{H;p;z_i} \quad (2.6)$$

Where for each calculation zone z_i and month m_i :

$H_{H;tr(excl.gf;m_i);z_i;m_i}$	Total heat transfer coefficient by transmission for heating except for the ground floor, in W/K
$H_{g;an;z_i,m_i}$	heat transfer coefficient for building elements in thermal contact with the ground in W/K, determined using equation (2.7)
$H_{H;D;z_i;m_i}$	Direct heat transfer coefficient between the heated space and the outside air in W/K according to equation (2.10)
$H_{H;U;z_i;m_i}$	Heat transfer coefficient via adjacent unheated spaces in W/K
$H_{H;A;z_i;m_i}$	Heat transfer coefficient via adjacent heated spaces in W/K
$H_{H;p;z_i}$	Heat transfer coefficient of zone z_i through vertical pipes passing through the thermal envelope and in direct communication with outside air in W/K, determined using equation (2.13)
$\theta_{int;calc;H;z_i;m_i}$	Calculation temperature of the calculation zone for heating in °C. Available at table 2.1
$\theta_{e;avg;m_i}$	Average outdoor temperature in month m_i in °C. Available at table 2.1
t_{m_i}	Calculation value for the length of the month under consideration in h. Available at table 2.1

The climate data for computations is presented in [Table 2.1](#). Additionally, The norm states the following simplifications: $H_{H;U;z_i;m_i} = 0$, $H_{H;A;z_i;m_i} = 0$.

Table 2.1: NTA 8800 climate data

Month	t_{m_i} h	$\theta_{e;avg;m_i}$ °C	$\theta_{e;argII;m_i}$ °C	$u_{site;m_i}$ m/s	$\theta_{ODA;preh;WTWC;m_i}$ °C
January	744	2.61	-	3.04	-

Continued on next page

Table 2.1 – continued from previous page

Month	t_{m_i} h	$\theta_{e;avg;m_i}$ °C	$\theta_{e;argII;m_i}$ °C	$u_{site;m_i}$ m/s	$\theta_{ODA;preh;WTWC;m_i}$ °C
February	672	4.82	13.97	4.15	-
March	744	5.91	13.00	2.99	-
April	720	9.32	13.70	3.06	-
May	744	14.73	14.56	2.97	25.63
June	720	16.12	15.62	2.78	27.49
July	744	18.05	16.17	2.63	26.34
August	744	18.48	16.90	2.51	27.29
September	720	15.63	15.11	2.71	25.30
October	744	10.40	15.04	2.78	-
November	720	7.99	13.43	2.83	-
December	744	4.00	-	2.83	-

Where:

t_{m_i}	Length of the month
$\theta_{e;avg;m_i}$	Monthly average outdoor air temperature
$\theta_{e;argII;m_i}$	Monthly average outdoor air temperature for ventilated cooling
$u_{site;m_i}$	Monthly average wind speed
$\theta_{ODA;preh;WTWC;z_i;m_i}$	Monthly average temperature of the supply air before the heat recovery system during the period that there is cold recovery via the heat recovery system

2.5.3.1.1 Heat transfer coefficient for building elements in thermal contact with the ground

H_g is computed when the ground surface of the building lies directly to the terrain or when the ground surface is above crawl spaces or unheated basements. It is computed using [equation \(2.7\)](#).

$$H_g = A_{fl} \cdot U_{fl} + \sum_j (\ell_j \cdot \psi_{gr;j}) \quad (2.7)$$

Where:

A_{fl}	Surface area of the floor directly on the ground in m ²
U_{fl}	Heat transfer coefficient of the floor surface, in W/(m ² K)
ℓ_j	Length of the linear thermal bridge of the floor perimeter, j , in m
$\psi_{gr;j}$	Linear heat transfer coefficient of part j of the floor perimeter to the ground, in W/(m · K)

A simplified method, *flat-rate allowance*, is possible for the linear thermal bridges of the building, making the output stationary heat transfer coefficient through the

ground surfaces, $H_{g;for}$, in W/K:

$$H_{g;for} = A_{fl} \cdot U_{fl} + 0.5 \cdot P \quad (2.8)$$

Where: P is the length of the perimeter in m. In the case of buildings with a basement, $H_{g;for}$ shall be computed using:

$$H_{g;for} = A_{fl} \cdot U_{fl} + 0.5 \cdot P + \sum A_{T;bW} \cdot (U_{bW;j} + \Delta U_{for}) \quad (2.9)$$

Where:

$A_{T;bW}$ Total area of the basement walls in m^2

$U_{bW;j}$ Stationary heat transfer coefficient of the area below ground level wall part j in $W/(m^2 \cdot K)$, e.g. thermal transmittance

2.5.3.1.2 Direct heat transfer coefficient between the heated space and the outside air

The direct heat transfer coefficient between the heated space and the outside air H_D is computed as in [equation \(2.10\)](#).

$$H_D = \sum_i (A_{T,i} \cdot U_{G,i}) + \sum_k (\ell_k \cdot \psi_k) + \sum_j \chi_j \quad (2.10)$$

Where:

$A_{T,i}$ Surface area of the opaque element i of the external separation construction in m^2

$U_{C,i}$ Heat transfer coefficient of the flat element i of the external separation construction in $W/(m^2K)$

ℓ_k Length of the linear thermal bridge, k , in m , e.g. the length size of the surface sides

ψ_k Linear heat loss coefficient of the thermal bridge, k , in $W/(mK)$

χ_j Heat loss coefficient of the point thermal bridge, j , in W/K

However, the norm states a simplified *flat-rate allowance* for linear thermal bridges can be computed using [equation \(2.11\)](#).

$$H_{D,for} = \sum_i (A_{T,i} \cdot (U_{C,i} + \Delta U_{for})) \quad (2.11)$$

In which:

$$\Delta U_{for} = \max \left\{ 0; 0.1 - 0.025 \cdot \left(\frac{\sum_i (A_{T,ntr;i} \cdot U_{C,ntr;i})}{\sum_i A_{T,ntr;i}} - 0.4 \right) \right\} \quad (2.12)$$

Where:

$A_{T,ntr;i}$	Surface area of the non-transparent element i , not being a floor above a crawl space or directly on the ground or a roof in m^2
$U_{C,ntr;i}$	Heat transfer coefficient of the non-transparent element i , not being a floor above a crawl space or directly on the ground or a roof in $W/(m^2 \cdot K)$

ΔU_{for} may only be applied when done for the whole building. Furthermore, mixing methods between *flat-rate* and *non-flat-rate* is not allowed. Except for the building ground surface, all surfaces are included. Table 2.2 shows an example of values of ΔU_{for} .

Average U-value of building envelope not touching the ground $[W/(m^2K)]$	ΔU_{for} $[W/(m^2K)]$
0.8	0.00
0.6	0.05
0.4	0.10

Table 2.2: Example values for standardised addition for the calculation of linear thermal bridges (NEN, 2024)

2.5.3.1.3 Heat transfer coefficient through vertical pipes

The heat transfer coefficient through vertical pipes $H_{H;p;z_i}$ is computed as equation (2.13).

$$H_{H;p;z_i} = \sum_j N_{storeys;j} \cdot H_{H;p;spec;j} \tag{2.13}$$

Where:

$H_{H;p;z_i}$	Heat transfer coefficient of zone z_i through vertical pipes passing through the thermal envelope and in direct communication with outside air, in W/K
j	Number of vertical pipes in the calculation area that pass through the thermal envelope and are in direct communication with outside air
$N_{storeys;j}$	Number of storeys of the calculation zone where vertical pipe j is located.
$H_{H;p;spec;j}$	Heat transfer coefficient per building layer for vertical pipe j , according to table 2.3 in W/K

Table 2.3: Fixed values for the heat transfer coefficient via vertical pipes per building layer

Type of pipe	$H_{H;p;spec;j}$
Uninsulated vertical pipe through thermal shell	1.8
Insulated vertical pipe through thermal shell ^a	0.5
No penetrations through thermal shell	0

^aA pipe may only be considered insulated if it is over its entire length isolated.

In the case of residential buildings, It is assumed that there is at least one uninsulated fictitious vertical pipe per residential unit.

$$N_{storeys;j} = \frac{H}{3} \quad (2.14)$$

Where:

$N_{storeys;j}$ number of storeys assumed for the fictitious vertical pipe j
 H External building height measured in m

The number of storeys is rounded down to an integer with a minimum of 1.

2.5.3.2. HEAT TRANSFER THROUGH VENTILATION

The heat transfer through ventilation is defined by [equation \(2.15\)](#).

$$Q_{H;ve;z_i;m_i} = H_{H;ve;z_i;m_i} \cdot (\theta_{int;calc;H;z_i} - \theta_{e;avg;m_i}) \cdot 0.001 \cdot t_{m_i} \quad (2.15)$$

Where:

$H_{H;ve;z_i;m_i}$ Total heat transfer coefficient through ventilation for heating in W/K, determined using [equation \(2.16\)](#)
 $\theta_{int;calc;H;z_i}$ Calculation temperature of the calculation zone for heating in °C, according to [table 2.1](#)
 $\theta_{e;avg;m_i}$ Average outdoor temperature in month m_i in °C, according to [table 2.1](#)
 t_{m_i} Calculation value for the length of the month under consideration in h, determined using [table 2.1](#)

2.5.3.2.1 Heat transfer coefficient for ventilation

The heat transfer coefficient for ventilation $H_{H;ve;z_i;m_i}$ is defined by equation (2.16).

$$H_{H;ve;z_i;m_i} = \rho_A \cdot C_A \cdot \frac{\sum_k (q_{v;k;H;z_i;m_i} \cdot b_{v;k;H;z_i;m_i} \cdot f_{v;dyn;k;z_i;m_i})}{3600} \quad (2.16)$$

Where for each calculation zone z_i and month m_i :

$\rho_A \cdot C_A$	Heat capacity of air per volume, in $J/(m^3 \cdot K)$
ρ_A	Density of air = $1.205 kg/m^3$
C_A	Heat capacity of air = $1005 J/kgK$
$b_{v;k;H}$	Supply temperature correction factor for air volume flow k , as determined by equation (2.17)
$f_{v;dyn;k}$	Dynamic correction factor for air volume flow k , = 1
$q_{v;k;H}$	Air volume flow k , in m^3/h , as determined by table 2.5

2.5.3.2.1.1 Supply temperature correction factor

The correction factor for supply temperature for air flow k is defined by equation (2.17).

$$b_{v;k;H;m_i} = \frac{(\theta_{int;set;H;stc;z_i;m_i} - \theta_{sup;k;H;m_i})}{(\theta_{int;set;H;stc;m_i} - \theta_{e;avg;m_i})} \quad (2.17)$$

Where for each month m_i :

$\theta_{int;set;H;z_i;m_i}$	Set-point temperature for the thermally conditioned zones of the adjacent calculation zone z_i for heating, in $^{\circ}C$, according to table 2.4
$\theta_{sup;k;H;m_i}$	Supply temperature of airflow k for heating, in $^{\circ}C$
$\theta_{e;avg;m_i}$	Average outside temperature per month in $^{\circ}C$, according to table 2.1

The supply temperature correction factor has a value other than one (1) if the temperature of the air supplied to the calculation zone $\theta_{sup;k;H;m_i}$ is not equal to the outdoor temperature.

Table 2.4: Set-point temperature for thermally conditioned zones

Calculation zone function	$\theta_{int;set;H;stc;z_i}$ $^{\circ}C$	$\theta_{int;set;C;stc;z_i}$ $^{\circ}C$
Childcare facility	21	

Continued on next page

Table 2.4 – continued from previous page

Calculation zone function	$\theta_{\text{int;set;H;stc;z}_i}$ °C	$\theta_{\text{int;set;C;stc;z}_i}$ °C
Other meeting facility	21	
Prison	21	
Healthcare function with beds	22	
Other healthcare facility	21	
Offices	21	
Lodging	21	
Educational	21	
Sports	16	
Retail	21	
Residential	20	

2.5.3.2.1.2 Determining Air Volume Flow

The calculation values for the specific air permeability $q_{v10;spec;cal}$ and the correction factor f_{type} are defined per building type in [table 2.5](#). The Dutch building types are described in detail in [section 3.2.2.1](#) with graphical examples in [figures 3.8](#) and [3.12](#).

Table 2.5: Calculation value for the specific air permeability per building type and their corresponding correction factor

Building Type	q_v	Version variant	f_{type}
Ground-based buildings			
Single family house with a pitched roof and one-storey non-residential buildings with a pitched roof	1.0	Intermediate location	1.0
		End, edge, or corner location	1.2
		Detached building, sloped roof	1.4
		Detached building, partially flat roof	1.2
		Detached building, partially flat roof	1.2
Single family house with a flat roof and other one-storey non-residential buildings	0.7	Intermediate location	1.0
		End, edge, or corner location	1.2
		Detached building, partially flat roof	1.4
Multi-storey buildings			
Living unit in multi-storey non-residential and apartment buildings	0.5	Intermediate location on ground or intermediate floor	1.0
		End, edge, or corner location on ground or intermediate floor	1.3
		Intermediate location on the top floor	1.2
		End, edge, or corner location on the top floor	1.4

2.5.4. TOTAL HEAT GAINS FOR HEATING

The hierarchical workflow for the calculation of the heat gains for heating are presented in [figure 2.17](#). This parameter is based on the internal gains and solar

gains as expressed in [equation \(2.18\)](#).

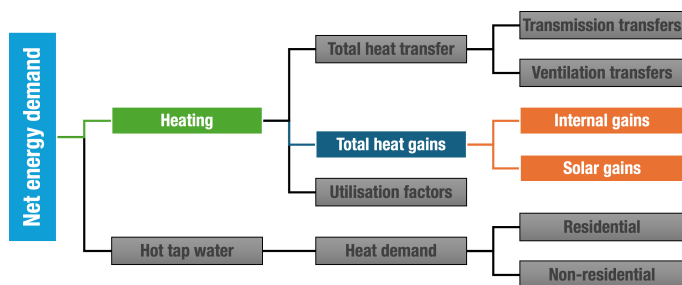


Figure 2.17: Total heat gains

$$Q_{H;gn;z_i;m_i} = Q_{H;int;z_i;m_i} + Q_{H;sol;z_i;m_i} \quad (2.18)$$

Where:

$Q_{H;int;z_i;m_i}$ Total internal heat gain for heating in kWh, determined either by [equation \(2.19\)](#) for residential buildings or [equation \(2.23\)](#) for non-residential buildings

$Q_{H;sol;z_i;m_i}$ Total solar gain for heating in kWh, determined using [equation \(2.29\)](#)

2.5.4.1. INTERNAL HEAT GAINS

The Internal heat gain concerns the contribution to the heat management by internal sources other than the deliberate supply of heat for space heating, space cooling or hot tap water (HTW) preparation. Only the internal heat gains in the calculation zone itself are included in the calculation. The norm separates the calculation for residential ([equation \(2.19\)](#)) and non-residential ([equation \(2.23\)](#)) buildings.

2.5.4.1.1 Residential Building

$$Q_{H;int;dir;z_i;m_i} = 180 \cdot N_{res;z_i} \cdot N_{P;res;z_i} \cdot 0.001 \cdot t_{m_i} \quad (2.19)$$

Determine the number of residents per calculation zone per residential building $N_{P;res;z_i}$ based on the average usable area per dwelling as follows:

If:

$$\frac{A_{g;z_i}}{N_{res;z_i}} \leq 30m^2 : N_{P;res;z_i} = 1 \quad (2.20)$$

$$30m^2 < \frac{A_{g;z_i}}{N_{res;z_i}} \leq 100m^2 : N_{P;res;z_i} = 2.28 - \frac{1.28}{70} \times \left(100 - \frac{A_{g;z_i}}{N_{res;z_i}} \right) \quad (2.21)$$

$$\frac{A_{g;z_i}}{N_{res;z_i}} \geq 100m^2 : N_{P;res;z_i} = 1.28 + 0.01 \times \frac{A_{g;z_i}}{N_{res;z_i}} \quad (2.22)$$

2

Where:

$N_{res;z_i}$	Number of residential units in calculation zone z_i
$N_{P;res;z_i}$	Average number of residents per calculation zone per residential building
$A_{g;z_i}$	Usable area of the calculation zone in m^2
t_{m_i}	Calculation value for the length of the month under consideration in h , according to table 2.1

2.5.4.1.2 Non-Residential Building

$$Q_{H;int;dir;z_i} = \left(\Phi_{H;int;oc;z_i;m_i} + \Phi_{H;int;A;z_i;m_i} + \Phi_{H;int;L;z_i;m_i} + \Phi_{H;int;W;z_i;m_i} + \Phi_{H;int;V;z_i;m_i} + \Phi_{H;int;proc;z_i;m_i} \right) \cdot 0.001 \cdot t_{m_i} \quad (2.23)$$

Where:

$\Phi_{H;int;oc;z_i;m_i}$	Heat flow as a result of heat production by people for heating in W, determined using equation (2.24)
$\Phi_{H;int;A;z_i;m_i}$	Heat flow due to heat production by equipment for heating in W, determined using equation (2.25)
$\Phi_{H;int;L;z_i;m_i}$	Heat flux through recoverable losses from lighting in W determined using equation (2.26)
$\Phi_{H;int;W;z_i;m_i}$	Heat flow through recoverable losses of the HTW system for heating in W, determined using equation (2.28)
$\Phi_{H;int;V;z_i;m_i}$	Heat flow through recoverable losses of the ventilation system in W. Set to 0
$\Phi_{H;int;proc;z_i;m_i}$	Heat flow through recoverable losses from or to processes and goods for heating in W. Set to 0
t_{m_i}	Calculation value for the length of the month under consideration in h , using table 2.1

2.5.4.1.2.1 Heat flow through people

The heat flow through people $\Phi_{int;Oc;z_i}$ is defined by [equation \(2.24\)](#).

$$\Phi_{int;Oc;z_i} = q_{Oc;usi} \times f_{\tau;usi} \times A_{g;z_i} \quad (2.24)$$

Using $q_{OC;usi}$, $f_{\tau;usi}$ fixed values from [table 2.6](#).

$q_{OC;usi}$ Specific internal heat production by people in W/m^2 , weighted according to the usable area. From [table 2.6](#)

$f_{\tau;usi}$ Correction factor for the occupancy time. From [table 2.6](#)

$A_{g;z_i}$ Usable area of the calculation zone in m^2

2.5.4.1.2.2 Heat flow through equipment

The heat flow through equipment $\Phi_{int;A;z_i}$ is defined by [equation \(2.25\)](#).

$$\Phi_{int;A;z_i} = q_{A;usi} \times A_{g;z_i} \quad (2.25)$$

Where:

$q_{A;usi}$ Specific internal heat production due to the average power of equipment. From [table 2.6](#)

$A_{g;z_i}$ Usable area of the calculation zone in m^2

2.5.4.1.2.3 Heat flux through lighting

The heat flow through lighting $\Phi_{H;int;L;z_i;m_i}$ is defined by [equation \(2.26\)](#).

$$\Phi_{H;int;L;z_i;m_i} = q_{L;usi} \times A_{g;z_i} \quad (2.26)$$

Where:

$q_{L;usi}$ Specific internal heat average production due to lighting. From [table 2.6](#)

$A_{g;z_i}$ Usable area of the calculation zone in m^2

Table 2.6: Specific internal heat production value per function

Calculation zone function	q_{OC} W/m^2	$f_{\tau;usi}$	q_A W/m^2	$q_{L;usi}$ W/m^2
Childcare facility	10	0.30	1	2.5
Other meeting facility	10	0.15	1	2.5
Prison	3	0.80	2	2.25
Healthcare function with beds	5	0.80	4	2.5
Other healthcare facility	5	0.30	3	1.25
Offices	5	0.30	2	1.25
Lodging	3	0.40	4	1.75
Educational	10	0.30	2	1
Sports	3	0.30	1	3
Retail	3	0.40	3	3

2.5.4.1.2.4 Heat flow through recoverable losses from lighting

The heat flow through recoverable losses from lighting $\Phi_{int;L;z_i}$ is defined by equation (2.27).

$$\Phi_{int;L;z_i} = \frac{f_L}{x} W_t \times 1000 \cdot t_{an} \quad (2.27)$$

Where:

Dimensionless reduction factor whose value is:

- 0.3 if the total installed power (P_n) is determined for lighting (flat-rate)
 - 0.5 if at least 70% of the luminaires, weighted by the total installed power (P_n), is extracted
 - 1.0 in other cases
- f_L
- W_t "energy consumption" for lighting to provide the necessary lighting levels per year in kWh;
- t_{an} Calculation value for the total length of the year in h. Determined using table 2.1

2.5.4.1.2.5 Heat flow through recoverable losses of the hot tap water system

The heat flow through recoverable losses of the hot tap water system $\Phi_{int;W;z_i;m_i}$ is defined by equation (2.28).

$$\Phi_{int;W;z_i;m_i} = \frac{\sum_{si} Q_{W;ls;rb;si;z_i;m_i} \cdot 1000}{t_{m_i}} \quad (2.28)$$

Where:

- $\Phi_{int;WA;z_i;m_i}$ Heat flow through recoverable losses from or to the HTW system in calculation zone z_i and month m_i in W;
- $Q_{W;ls;rb;si;z_i;m_i}$ Recoverable loss of HTW system si in calculation zone z_i , in month m_i in kWh
- t_{m_i} Calculation value for the length of the month under consideration in h, determined using table 2.1

2.5.5. HEAT GAINS THROUGH INCIDENT SOLAR RADIATION

The method is derived from NEN-EN-ISO 52016-1:2017 (ISO, 2017). Heat gains due to solar radiation result from the solar radiation received by the surfaces that define the thermal envelope of the building. The radiation is influenced by the orientation, inclination, sun shading and solar transmission, solar absorption, and the heat transfer properties of the surfaces. The solar heat gains are defined by equation (2.29).

$$Q_{H;sol;dir;z_i;m_i} = \sum_k Q_{H;sol;w_i,k;m_i} + \sum_k Q_{H;sol;op,k;m_i} \quad (2.29)$$

Where, for each element k and month m_i :

- $Q_{H;sol;dir;z_i;m_i}$ Monthly solar gain of the calculation zone in kWh for heating
- $Q_{H;sol;w_i,k;m_i}$ Monthly solar heat gain through transparent element w_i , k in kWh for heating. Determined using [equation \(2.30\)](#)
- $Q_{H;sol;op,k;m_i}$ Monthly solar heat gain by non-transparent element op , k in kWh for heating. Determined using [equation \(2.34\)](#)

2.5.5.1. HEAT FLUX DUE TO INCIDENT SOLAR RADIATION THROUGH TRANSPARENT PARTS

The heat flux due to incident solar radiation through transparent parts of the building envelope (from now on called windows) w_i for heating, $Q_{H;sol;w_i,k;m_i}$ in kWh, is calculated for each element k using [equation \(2.30\)](#):

$$Q_{H;sol;w_i,k;m_i} = g_{gl;w_i,k;m_i} \cdot A_{w_i,k} \cdot (1 - F_{fr;w_i,k}) \cdot F_{sh;obst;w_i,k;m_i} \cdot I_{sol;w_i,k;m_i} \cdot 0.001 \cdot t_{m_i} - Q_{sky;w_i,k;m_i} \quad (2.30)$$

Where, for each window w_i and month m_i :

- $Q_{H;sol;w_i,k;m_i}$ Solar heat gain through transparent element w_i , k in kWh, for heating
- $g_{gl;sol;op,k;m_i}$ dimensionless average effective total solar factor of window w_i , k , per month m_i , for heating, determined using [2.5.5.1.2](#), where all glazing is calculated using the method for non-diffusing glazing

The transparent element can consist of clear glazing but also (permanent) diffusing glazing, glass bricks or (permanent or movable) sun protection. However, the dimensionless mean effective total solar gain factor, $g_{gl;w_i,k;m_i}$, is also determined for diffusing glazing or glass bricks according to the method for non-diffusing glazing.

- $A_{w_i,k}$ Area of window w_i , k , in m^2
- $F_{fr;w_i,k}$ Frame fraction of window w_i , k , the ratio of the frame area to the total area of the glazed portion of window w_i , k , determined using [2.5.5.1.3](#)
- $F_{sh;obst;w_i,k;m_i}$ Dimensionless shading reduction factor for external impediments of window w_i , k , determined using [table 2.9](#)
- $I_{sol;w_i,k;m_i}$ Monthly average total incident solar radiation per m^2 area of window w_i , k , at a given angle of inclination β_{w_i} and orientation γ_{w_i} in W/m^2 determined using [table 2.8](#)

t_{m_i}	Calculation value for the length of the month under consideration in h, determined using table 2.1
$Q_{sky;w_i,k;m_i}$	Monthly extra heat flow due to heat radiation to the sky from window w_i, k in kWh, determined using 2.5.5.1.1

External obstructions are nearby obstructions of building parts on the own lot, such as grooves, side ribs, overhangs or adjacent building parts.

2.5.5.1.1 Heat radiation to the sky

The heat radiation to the sky $Q_{sky;k;m_i}$ is defined by [equation \(2.31\)](#).

$$Q_{sky;k;m_i} = 0.001 \cdot F_{sky;k} \cdot R_{se;k} \cdot U_{c;k} \cdot A_{c;k} \cdot h_{lr;e;k} \cdot \Delta\theta_{sky;m_i} \cdot t_{m_i} \quad (2.31)$$

Where, for each element k and month m_i :

$Q_{sky;k;m_i}$	Extra heat flow due to heat radiation from building envelope element k to the sky, in kWh
$F_{sky;k}$	Visibility factor between building envelope element k and the sky, determined using table 2.10
$R_{se;k}$	Heat transfer resistance on the outside of element k in m^2K/W , as determined using table 2.11
$U_{c;k}$	Heat transfer coefficient of element k in $W/(m^2 \cdot K)$
$A_{c;k}$	Area of element k
$h_{lr;e;k}$	Heat transfer coefficient for long-wave radiation on the outside of the structure, for which the following numerical value applies: $h_{lr;e} = 4.14 W/(m^2 \cdot K)$
$\Delta\theta_{sky;m_i}$	Average difference between the apparent sky temperature and the outside temperature, for which the following numerical value applies: $\Delta\theta_{sky;m_i} = 11K$
t_{m_i}	Length of the month under consideration in h. Determined using table 2.1

2.5.5.1.2 Sun access factor: Windows with non-diffusing glazing

The total solar access factor depends on the angle of incidence (height and azimuth) of the incident solar radiation. The (time-weighted average) value required for the calculations is lower than the solar accession factor for radiation perpendicular to the glazing, $g_{gl;n}$. The total solar gain factor (corrected for the angle of incidence) is calculated according to [equation \(2.32\)](#):

$$g_{gl;w_i} = F_w \cdot g_{gl;n;w_i} \quad (2.32)$$

Where:

F_w	Correction factor for non-scattering glazing: 0.90
$g_{gl;n;w_i}$	Sun factor with a perpendicular incidence of solar radiation, determined using NEN-EN 410 or according to table 2.7

2

The numerical value for $g_{gl;n;w_i}$ must be rounded down to a multiple of 0.05.

Table 2.7: Standard values for the total solar factor at perpendicular incidence, $g_{gl;n}$, for common types of glazing

Type	$g_{gl;n}$ ^a
Only glass	0.85
Double glass	0.75
Double glass with spectral (low) selective and low emissive coating (HR++)	0.60
Triple glazing without or with one spectrally (low) selective and low emissivity coating	0.50
Triple glass with two spectrally (low) selective and low emissivity coatings	0.40
Single glass with single glass front window or rear window without coating	0.75

^aAssuming a clean surface and normal, clear and non-scattering glass.

2.5.5.1.3 Frame fraction

The area of the glazing can be determined with the geometric data or window dimensions (Method A) or derived from a fixed frame fraction (Method B). The same choice must be made for all windows in a building.

- **Method A:** The frame fraction of window w_i , $F_{fr;w_i}$, should be calculated according to the following formula:

$$F_{fr;w_i} = 1 - \frac{A_{gl;w_i}}{A_{w_i}} \quad (2.33)$$

Symbol	Description
$F_{fr;w_i}$	Frame fraction
$A_{gl;w_i}$	Area of the glazed part of window w_i in m^2
A_{w_i}	Area of window w_i in m^2

- **Method B:** If the frame fraction is unknown when determining the transmission losses, e.g., because fixed values for the heat transfer

coefficient are used for windows, the following numerical value must be used for the frame fraction:

$$F_{fr;w_i} = 0.25$$

2

2.5.5.2. HEAT FLUX DUE TO INCIDENT SOLAR RADIATION THROUGH NON-TRANSPARENT PARTS

For each non-transparent surface of the thermal envelope of the building ($op; k$), the heat flux for heating $Q_{H/C;sol;op,k;m_i}$ in kWh, in month m_i , is calculated using equation (2.34):

$$Q_{H;sol;op,k;m_i} = \alpha_{sol} \cdot R_{se} \cdot U_{C;op,k} \cdot A_{C;op,k} \cdot F_{sh;obst;op,k;m_i} \cdot I_{sol;op,k;m_i} \cdot 0.001 \cdot t_{m_i} - Q_{sky;op,k;m_i} \quad (2.34)$$

Where, for every non-transparent element k and month m_i :

$Q_{H;sol;op,k;m_i}$	Solar heat gain by non-transparent element op, k for heating, in kWh
α_{sol}	Dimensionless absorption coefficient for solar radiation. Fixed value = 0.6
R_{se}	Heat transfer resistance on the outside in m^2K/W . Defined by table 2.11
$U_{C;op,k}$	Heat transfer coefficient of non-transparent element op, k in $W/(m^2 \cdot K)$
$A_{C;op,k}$	Area of non-transparent element op, k in m^2 .

The remaining variables are described for equation (2.30) (replacing index w_i by index op). Additionally, the dimensionless shading reduction factor for external impediments of a non-transparent element op, k is $F_{sh;obst;w_i,k;m_i} = 1$.

Table 2.8: Excerpt of the monthly average total incident solar radiation, $I_{sol;m_i}$, averaged over all hours; ground reflection coefficient $\rho = 0.2$. Taken from (NEN, 2024)

β	0°		30°						
	γ	180° (Z)	225° (SW)	270° (W)	315° (NW)	360° (N)	45° (NE)	90° (E)	135° (SE)
Month		$I_{sol;m_i}$							
January		28.0	50.5	44.4	29.0	16.2	14.9	15.8	26.9
February		49.3	69.1	61.2	46.2	32.9	27.2	34.5	49.4
March		96.6	122.5	109.3	87.7	66.7	56.4	72.8	97.6

Continued on next page

Table 2.8 – continued from previous page

β	0°	30°								
γ	-	180° (Z)	225° (SW)	270° (W)	315° (NW)	360° (N)	45° (NE)	90° (E)	135° (SE)	
Month		$I_{Sol;m_i}$								
	April	160.5	189.5	174.5	146.5	115.6	104.6	125.1	158.9	184.1
	May	197.0	211.1	201.5	179.9	155.8	148.5	160.6	186.3	206.3
	June	209.3	211.2	210.7	199.4	180.6	171.0	173.0	189.7	204.4
	July	191.0	196.1	193.2	180.2	162.1	153.0	156.9	175.0	190.0
	August	177.2	197.9	198.3	178.4	147.6	125.8	127.5	152.8	179.3
September	123.9	154.0	146.2	121.1	91.6	73.7	86.5	113.7	140.1	
	October	73.2	102.4	91.5	68.8	47.3	36.3	48.9	71.6	93.6
	November	34.3	54.8	47.7	32.9	20.5	18.6	20.9	33.8	48.6
	December	21.0	38.3	32.6	20.6	12.5	12.2	12.5	21.2	33.1

Table 2.9: Excerpt of shading reduction factor ($F_{sh;obst;m_i}$) for heat demand calculation with complete obstruction. Taken from (NEN, 2024)

South													
Orientation	Vert.	Diagonally turned up						Diagonally turned down					
Tilt compared to horizontal													
Month	90°	75°	60°	45°	30°	15°	0°	105°	120°	135°	150°	165°	180°
Month	I _{sol;m_i}												
January	0.19	0.20	0.22	0.25	0.30	0.39	0.55	0.18	0.19	0.23	0.33	0.82	1.00
February	0.30	0.31	0.32	0.36	0.40	0.47	0.58	0.30	0.33	0.42	0.69	0.99	1.00
March	0.35	0.33	0.34	0.35	0.38	0.42	0.49	0.38	0.46	0.67	0.98	1.00	1.00
April	0.36	0.32	0.30	0.30	0.31	0.33	0.36	0.44	0.62	0.98	1.00	1.00	1.00
May	0.46	0.39	0.35	0.33	0.32	0.33	0.34	0.61	0.92	1.00	1.00	1.00	1.00
June	0.56	0.47	0.41	0.38	0.37	0.37	0.37	0.75	1.00	1.00	1.00	1.00	1.00
July	0.56	0.47	0.43	0.40	0.39	0.39	0.41	0.72	0.98	1.00	1.00	1.00	1.00
August	0.42	0.37	0.34	0.33	0.33	0.35	0.37	0.53	0.77	1.00	1.00	1.00	1.00
September	0.34	0.32	0.32	0.32	0.34	0.38	0.43	0.39	0.50	0.77	1.00	1.00	1.00
October	0.28	0.27	0.29	0.31	0.34	0.39	0.48	0.29	0.34	0.45	0.82	0.99	1.00
November	0.24	0.25	0.27	0.30	0.35	0.43	0.56	0.24	0.25	0.30	0.46	0.95	1.00
December	0.19	0.19	0.19	0.25	0.30	0.39	0.55	0.18	0.19	0.23	0.33	0.82	1.00

Table 2.10: Form factor values between the structure k and the sky $F_{sky;k}$

Value	Condition
1	Horizontal structure, the angle of inclination from the horizontal of which is less than or equal to 5°
0.75	Inclined structures with an angle of inclination from the horizontal less than or equal to 75° , but greater than 5°
0.5	Vertical structure with an angle of inclination from the horizontal greater than 75°
0	External partition structures adjacent to the outside that lean over (facing the ground)
0	Partition constructions between a calculation zone and an adjacent unheated greenhouse

Table 2.11: Heat transfer resistances at different heat flow directions

Heat transfer resistance	Direction of the heat flow		
$m^2 \cdot K/W$	Up	Horizontal	Down
R_{si}	0.10	0.13	0.17
R_{se}	0.04	0.04	0.04

2.5.6. EFFECTIVE INTERNAL HEAT CAPACITY OF A ZONE

The effective internal heat capacity of the calculation area (air, furniture and building elements) represents the total heat capacity seen from the inside. The method is derived from NEN-EN-ISO 52016-1:2017 (ISO, 2017) and calculated using equation (2.35).

$$C_{m;int;eff;z_i} = D_{m;int;eff;z_i} \cdot 1000 \cdot a_{g;z_i} \quad (2.35)$$

$C_{m;int;eff;z_i}$ Effective internal heat capacity of the calculation zone, in J/K ;

$D_{m;int;eff;z_i}$ Specific internal heat capacity of the calculation zone, determined using tables 2.12 to 2.14, in kJ/m^2K ;

$a_{g;z_i}$ Usable area of calculation zone z_i in m^2 .

Table 2.12: Fixed values for the specific internal heat capacity

Type of construction Floors	Type of construction Walls	$D_{m,int;eff;z_i}$ [kJ/(m ² K)]	
		Close ceiling ^a	None or open ceiling ^b
Light	Light	55	80
Light	Heavy	110	180
Heavy	Light		
Very heavy	Light		
Heavy	Heavy	180	360
Light	Very heavy		
Heavy	Very heavy	250	450
Very heavy	Heavy		
Very heavy	Very heavy		

^aIn non-residential buildings, the column "closed or lowered ceiling" should be used unless at least 15% net of the ceiling surface of a free-hanging ceiling in the occupied area, evenly distributed over the ceiling, is open.

^bFor residential buildings, the column "no or open ceiling" should be used as a starting point.

Table 2.13: Specification of the type of floor construction for the determination of the specific internal heat capacity

Type of construction	Flooring
Light	Wooden floors
	Timber frame construction (hsb) floors
	Steel frame construction (sfb) floors
	Floors of any type that are insulated on the inside
Heavy	Steel-concrete floors
	Non-solid concrete floors, such as hollow core slab and cassette floors
Very heavy	Massive concrete floors

Table 2.14: Specification of the type of wall construction for the determination of the specific internal heat capacity

Type of construction	Walls
Light	Timber frame construction (hsb)
	Steel frame construction (sfb)
	Steel skeleton construction
	Walls of any type that are insulated on the inside
Heavy	Load-bearing masonry
Very heavy	Concrete column-girder skeleton construction
	Concrete wall-floor skeleton construction

2.5.7. UTILISATION FACTORS

Factors reducing the total monthly heat gains in the monthly calculation method are used to obtain the resulting reduction of the building energy needed for heating. The method is derived from ISO 52016-1:2017 (ISO, 2017). The dimensionless utilisation factor for heat gain, $\eta_{H,gn}$, is a function of the heat balance ratio for heating, $\gamma_{H;z_i;m_i}$, and a numerical parameter, $\alpha_{H;z_i;m_i}$, which depends on the thermal inertia of the building. Figure 2.18 highlights this section's focus areas.

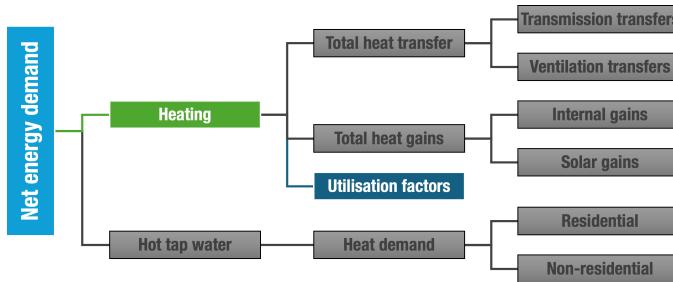


Figure 2.18: Utilisation factors

The utilisation factor is calculated for each zone and month using equation (2.36):

$$\eta_{H;gn;z_i;m_i} = \begin{cases} \frac{1 - (\gamma_{H;z_i;m_i})^{\alpha_{H;z_i;m_i}}}{1 - (\gamma_{H;z_i;m_i})^{\alpha_{H;z_i;m_i} + 1}}, & \gamma_{H;z_i;m_i} > 0 \text{ and } \gamma_{H;z_i;m_i} \neq 1 \\ \frac{\alpha_{H;z_i;m_i}}{\alpha_{H;z_i;m_i} + 1}, & \gamma_{H;z_i;m_i} = 1 \\ \frac{1}{\gamma_{H;z_i;m_i}}, & \gamma_{H;z_i;m_i} \leq 0 \text{ and } Q_{H;gn;z_i;m_i} > 0 \\ 1, & \gamma_{H;z_i;m_i} \leq 0 \text{ and } Q_{H;gn;z_i;m_i} \leq 0 \end{cases} \quad (2.36)$$

Whereby:

$$\gamma_{H;z_i;m_i} = \frac{Q_{H;gn;z_i;m_i}}{Q_{H;ht;z_i;m_i}} \quad (2.37)$$

Where:

$\eta_{H;gn;z_i;m_i}$	Dimensionless utilisation factor for heat gain
$\gamma_{H;z_i;m_i}$	Dimensionless heat balance ratio for heating
$\alpha_{H;z_i;m_i}$	Dimensionless numerical parameter. Determined using equation (2.38)
$Q_{H;gn;z_i;m_i}$	Total heat gains for heating, in kWh. Determined using 2.5.4
$Q_{H;ht;z_i;m_i}$	Total heat transfer for heating, in kWh. Determined using section 2.5.3

The dimensionless numerical parameter $\alpha_{H;z_i;m_i}$ is calculated by the following formula:

$$\alpha_{H;z_i;m_i} = \alpha_{H;0} + \frac{\tau_{H;z_i;m_i}}{\tau_{H;0}} \quad (2.38)$$

$\alpha_{H;0}$	Dimensionless numerical reference parameter, which has the following number value: $\alpha_{H;0} = 1.0$
$\tau_{H;z_i;m_i}$	Time constant of the heat requirement in h. Determined using equation (2.39)
$\tau_{H;0}$	Reference time constant, for which the following number value holds: $\tau_{H;0} = 15$, in h

2.5.7.1. CALCULATION ZONE TIME CONSTANT

The calculation zone time constant, τ , characterises the internal thermal inertia of the calculation zone. The time constant can be different for heating and cooling calculations. It can vary from month to month depending on the variation of the two variables on which the time constant depends, namely H_{tr} and H_{ve} . The time constant, distinguished for heat demand (H) and cooling demand (C) in hours, is calculated as follows:

$$\tau_{H;z_i;m_i} = \frac{C_{m;int;eff;z_i}}{3600 \cdot H_{H;tr(excl.grf);z_i;m_i} + H_{H;g;adj;z_i} + H_{H;ve;z_i;m_i}} \quad (2.39)$$

Where:

$\tau_{H;z_i;m_i}$	Time constant of calculation zone z_i for the heat demand and the cooling demand calculation respectively in h
$C_{m;int;eff;z_i}$	Effective internal heat capacity of the calculation zone in J/K. Determined using equation (2.35)
$H_{H;tr(excl.grfl);z_i;m_i}$	Total heat transfer coefficient by transmission for heating excluding the ground floor in W/K. Determined using equation (2.6)
$H_{H;g;adj;z_i}$	Seasonal average of the total heat transfer coefficient by transmission through the ground floor in W/K
$H_{H;ve;z_i}$	Total heat transfer coefficient through ventilation, for month m_i in W/K. Determined using equation (2.16)

2.5.8. HOT TAP WATER

The net heat demand for hot tap water (HTW) for the residential buildings is determined based on a prescribed demand per resident in kWh per year, where the number of residents is dependent on the usable floor area. For the utility buildings category, the net heat demand for hot tap water is determined by a prescribed demand per calculation zone in kWh per m^2 of usable floor area per year. [Figure 2.19](#) highlights this section's focus areas.

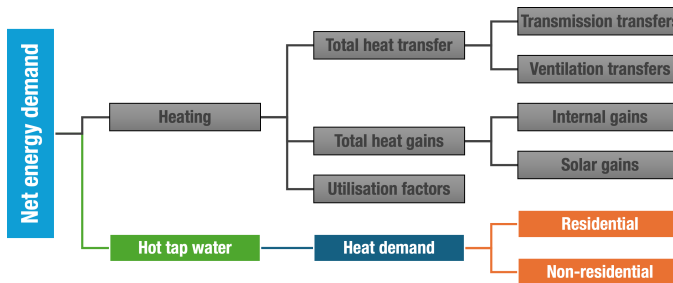


Figure 2.19: Hot tap water

2.5.8.1. RESIDENTIAL

$$Q_{W;nd;z_i,m_i} = N_{woon;z_i} \cdot N_{P;woon;z_i} \cdot Q_{W;nd;spec;p} \cdot \frac{t_{m_i}}{t_{an}} \quad (2.40)$$

The number of occupants per calculation zone $N_{P;woon;z_i}$ based on the average usable floor area per residential unit per calculation zone as determined using [equations \(2.20\) to \(2.22\)](#). Heat demand for HTW for residential buildings is 856 kWh/person, with a specific usage of HTW of 60°C water of 40.29 l/d/person. The distribution of HTW between the kitchen and bathroom are fixed values as 0.2 for kitchen ($C_{W;nd;k}$) and 0.8 for bathroom ($C_{W;nd;b}$).

2.5.8.1.1 Utility Buildings

$$Q_{W;nd;z_i,m_i} = Q_{W;nd;spec;us_i} \times A_{g;z_i} \times \frac{t_{m_i}}{t_{an}} \quad (2.41)$$

Where:

- $Q_{W;nd;spec;us_i}$ Monthly heat demand for HTW weighted according to the function in kWh/m^2 per year. Determined using [table 2.15](#)
- $A_{g;z_i}$ Usable floor area of calculation zone z_i in m^2
- t_{m_i} Length of month m_i in h. Determined using [table 2.1](#)
- t_{an} Length of the year in h

The calculation value for $Q_{W;nd;spec}$ per unit of usable floor area for utility buildings is given in [table 2.15](#).

Table 2.15: Annual specific $Q_{W;nd;spec}$ fixed values per usage function

Calculation zone function	$Q_{W;nd;spec}$ $kWh/m^2/year$	$V_{W;nd;spec;day}$ $l/m^2/day$
Childcare facility	2.8	0.131
Other meeting facility	2.8	0.131
Prison	4.2	0.196
Healthcare function with beds	15.3	0.719
Other healthcare facility	2.8	0.131
Offices	1.4	0.065
Lodging	12.5	0.588
Educational	1.4	0.065
Sports	12.5	0.588
Retail	1.4	0.065

2.6. EXISTING DUTCH MODELS

There are several datasets and data models currently available in the Netherlands, in this section I will discuss the most relevant ones. **SAWEC** (Vethman *et al.*, 2019), the Simulation and Analysis model for the Explanation and Prediction of Housing-Related Energy Use and CO₂ Emissions "Dutch Simulatie en Analyse model voor verklaring en voorspelling van het Woninggebonden Energieverbruik en CO₂-emissie". Its scope is to simulate the building-energy-related energy use of living units for future projections and monitoring of energy systems. It is used to produce the annual estimates within the building stock of the Climate and

Energy Outlook, in Dutch *Klimaat-en Energieverkenning –KEV–* Hammingh *et al.* (2024).

Vesta MAIS (PBL, 2025), this spatial energy model has been used since 2011 for the calculation of the energy use and CO2 emissions of the built environment for the period up to 2050. It evaluates the potential and cost of building-level measures, analysing their impact on emissions reduction, energy use, investment costs and financial returns. It supports national and regional decision-making. Since 2017, it has been open source, allowing it to be used by municipalities and private companies.

Currently, **Hestia 1.0** (van der Molen, Poorthuis *et al.*, 2023) is the open-source, spatially-explicit energy model for every dwelling in the Netherlands, which has been developed to replace SAWEC and Vesta MAIS. Its results have been used already for the annual Climate and Energy Exploration, in Dutch *Klimaat- en Energieverkenning –KEV–* (Hammingh *et al.*, 2024) as well as the municipal heat transition vision, in Dutch *Transitievisie Warmte –TVV–* (van der Molen, Langeveld *et al.*, 2023). Hestia is a multi-source platform since it combines data modelling, simulations engines and forecasting of the energy demand of the Dutch building stock. Figure 2.20 shows the schematic representation of the different components of Hestia.

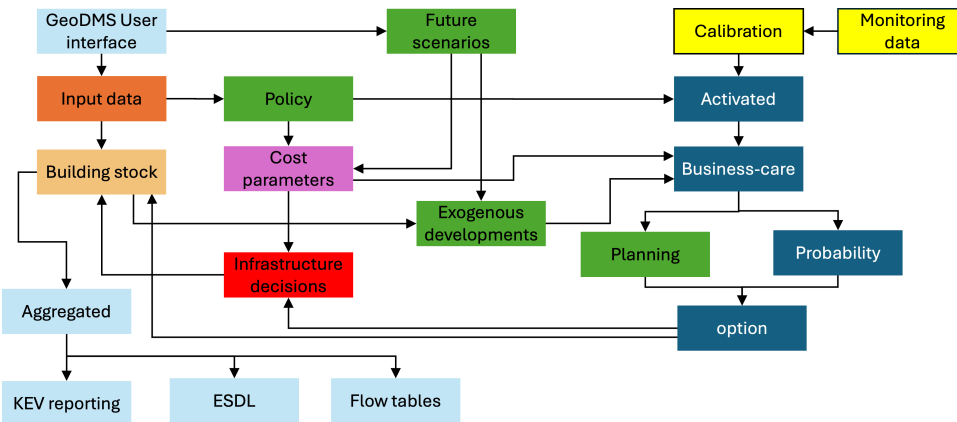


Figure 2.20: Schematic representation of the different components of Hestia. Adapted and translated from van der Molen, Poorthuis *et al.* (2023), colours follow the same pattern as in the mentioned report

The building stock is defined by the cadastre Netherlands (Kadaster) INSPIRE database Kadaster (2025). This dataset corresponds to the Basic Registration Addresses and Buildings –BAG– (see section 3.2.1.1). However, in this case, the dataset is structured according to the European INSPIRE guidelines. The data from new buildings are defined by using predefined building archetypes (see section 3.2.2.1), including average floor area and HVAC systems. According

to the authors, building geometric data are extracted from the 3DBAG (see [section 3.2.1.2](#)). However, no clear details are available in the documentation about the use of this dataset. The physics data of the surfaces that define the thermal envelope of buildings are provided by the WoON 2018 dataset Stuart-Fox *et al.* (2022) (see [section 3.2.2.1](#)). The data collected from the WoON dataset have been extrapolated to find the distribution of building areas for each building type-year-of-construction combination (see [table 3.5](#)). The energy performance certificate data is taken from ep-online (see [section 3.2.2.2](#)); authors indicate that for The building stock is defined by the cadastre Netherlands (Kadaster) INSPIRE database Kadaster (2025). This dataset corresponds to the Basic Registration Addresses and Buildings –BAG– (see [section 3.2.1.1](#)). However, in this case, the dataset is structured according to the European INSPIRE guidelines. The data from new buildings are defined by using predefined building archetypes (see [section 3.2.2.1](#)), including average floor area and HVAC systems. According to the authors, building geometric data are extracted from the 3DBAG (see [section 3.2.1.2](#)). However, no clear details are available in the documentation about the use of this dataset. The physics data of the surfaces that define the thermal envelope of buildings are provided by the WoON 2018 dataset Stuart-Fox *et al.* (2022) (see [section 3.2.2.1](#)). The data collected from the WoON dataset have been extrapolated to find the distribution of building areas for each building type-year-of-construction combination (see [table 3.5](#)). The energy performance certificate data is taken from ep-online (see [section 3.2.2.2](#)); authors indicate the development of a simplified NTA 8800 method for the computation of the energy performance of buildings. However, no documented details are available.

2.7. CONCLUSION REMARKS

The literature review in this chapter leads to the following conclusions. UBEM approaches require a lot of data from multiple sources, that could be feasible to obtain for specific buildings but impractical when thinking of city- or country-wide analyses. Therefore, archetypes are being used for those scales of analyses.

Machine learning in energy simulations is used primarily to identify the relationship between building parameters and actual "energy consumption" data. Also, it is used to predict future consumption patterns based on statistical and historical energy data. Data-driven approaches for energy simulation typically classify buildings into well-defined archetypes based on their function and year of construction. This approach allows for the classification of most buildings based on these representative archetypes.

In several researches, when authors discuss multi-scale or city-scale analyses, building-scale simulations are performed on selected buildings that best represent defined archetypes. The results of these representative simulations are then applied to all other buildings in the corresponding categories. This strategy

significantly reduces the computational time compared to the simulation of every building in the study area.

The literature also indicates that most UBEM approaches are file-based, which can limit data accessibility, scalability, and performance. Incorporating database management systems (DBMS) enhances data storage, retrieval, and processing, especially when dealing with large datasets, which is typical of UBEM. Furthermore, integrating DBMS improves the interaction between the data and BES.

Finally, The NTA 8800 specifies the energy balance method tailored to the Netherlands. This method depends on the geometrical and semantical representation of buildings, including their physical characteristics.

The analysis of the NTA 8800 carried out in this chapter lays the foundation for the analysis on data requirements characteristics (availability, quality) carried out in [chapter 3](#), and that will also influence the second part of this PhD Thesis regarding the implementation ([chapters 4 to 6](#)).

Other Dutch initiatives have been implemented to compute the energy demand of the building stock. However, their work relies on official statistical data that is not open access and therefore has not been considered in this PhD research.

3

ALIGNING DUTCH OPEN SPATIAL AND NON-SPATIAL DATA WITH UBEM REQUIREMENTS

The introduction and theoretical framework chapters have provided the knowledge to discuss the data requirements for Urban Building Energy Modelling. They provided the conceptual reasoning for the data collection based on the scope of using open spatial and non-spatial data at the country level using the Netherlands as a concrete case study.

The theoretical framework presented in [chapter 2](#) indicates that Urban Building Energy Modelling (UBEM) requires extensive data from multiple heterogeneous sources. However, these requirements are not standardised and can vary depending on factors such as the scale of analysis. The chapter is structured around two primary data types: geospatial and non-geospatial sources. First, I discuss general data considerations for UBEM and demonstrate how semantic 3D city models (s3DCM) and open data in the Netherlands can address these requirements ([section 3.1](#)). Then, I describe the country-wide open datasets available in the Netherlands that will be used in the following chapters ([section 3.2](#)). The chapter closes with the key findings of the data requirements of UBEM ([section 3.3](#)).

3.1. GENERAL DATA CONSIDERATIONS

UBEM requires lots 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 in the case study. In an ideal scenario, all necessary data would be available, including all building characteristics (cadastral, geometric, and physics), localised weather data, "energy consumption" records, and information on occupants' behaviour. In practice, however, data collection limitations, access restrictions, or incomplete datasets often prevent such an ideal scenario. Figure 3.1 outlines the essential data required for UBE, which fall into three major categories: building, climate, and energy. Data types are classified as discussed in section 2.2. Therefore, *Buildings* are the essential element of UBE, and their data requirements vary depending on the type of data.

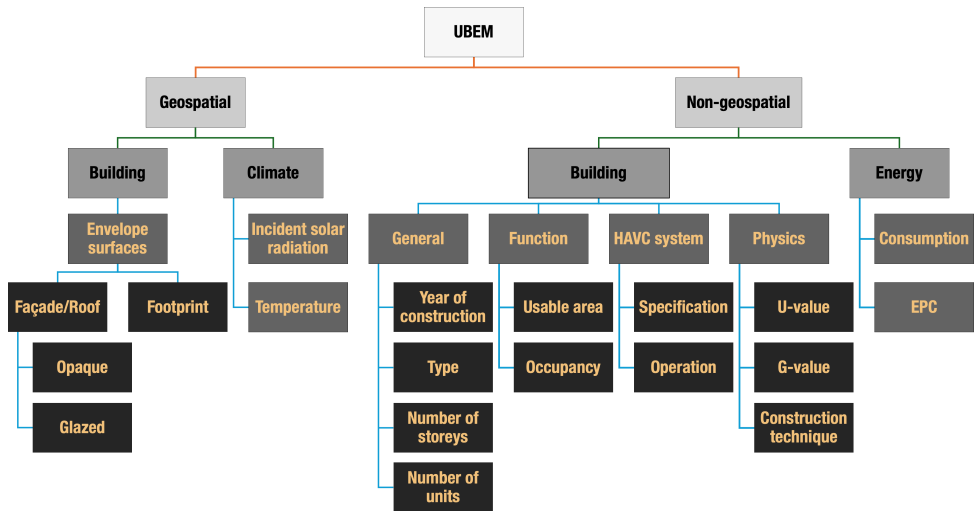


Figure 3.1: General overview of UBE data requirements, from (León-Sánchez, Augiario *et al.*, 2025)

3.1.1. GEOSPATIAL DATA

3.1.1.1. BUILDING DATA

Geometric data encompass all data about the building that has a spatial relationship with the earth; these data define the shape of the building while having a spatial connotation, including an associated coordinate reference system. The scope of this PhD thesis is to enhance energy applications using semantic 3D city models (s3DCM). Therefore, s3DCMs are the most prominent and convenient geometric data source. As explained in section 2.2.1, buildings are modelled using the boundary representation method. This approach is appropriate for energy applications such as quantifying the solar potential or simulating the energy performance of buildings. When following the CityGML standard, information such as surface area, inclination, and orientation can be computed from the *ThematicSurfaces* that define the building's envelope. Furthermore, inclination and orientation are required for processes such as solar

irradiance simulation of the surfaces of the building.

Information that is relevant for UBEM may be included in the s3DCM. For example, neighbouring buildings that share boundary surfaces (*PartyWalls*) with other buildings have shared surface areas that are essential when performing heat transfer calculations. That is also the case when there are multiple thermal zones inside the building. However, this information is not present in all 3DCMs. Furthermore, not all LoDs are a perfect fit for energy applications since the thermal envelope of a building shall be classified between opaque and glazed due to their corresponding physical properties. However, LoD1 and LoD2 building representations in CityGML do not support this data. Therefore, ratios like window-to-wall or window-to-roof (or, more generally, opening-to-surface) are required instead to account for this distinction, which is also the case for window frames. The *GroundSurface* of a building or its footprint (in cases such as the 3DBAG the roof's print [figure 3.6](#)) details the area touching the ground and the linear thermal boundary of the building on the ground.

3.1.1.2. CLIMATE AND WEATHER 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. The most accurate method to measure solar irradiance over a surface is through pyranometers; however, this is impractical at urban scale as the sensor only measures the irradiance at specific locations. Other options are satellite remote sensing observations, simulations-based approaches using geospatial and weather data, and statistical data. According to NTA 8800, a shading reduction factor must be included when using statistical climate data.

The Royal Netherlands Meteorological Institute (KNMI) serves as the Dutch national weather service. It operates a comprehensive nationwide observation network, a key component of which includes 48 automatic weather stations—34 on land and 14 at sea (KNMI, [2024a](#)). These stations are equipped to measure various meteorological variables such as air temperature, humidity, air pressure, wind speed and direction, precipitation amount, type of precipitation (rain or snow), cloud height and quantity, solar radiation, and horizontal visibility. Hourly observations are available to download as open data in the KNMI website (KNMI, [2024b](#)).

However, these stations record global solar radiation. It is necessary to consume the typical year data sets when requiring diffuse and direct solar radiation measurements, as well as data on solar elevation and azimuth, which are derived from scientific models and consolidate the average of observed data collected over a minimum period of 30 years (Klement, [2024](#)).

A source of open dataset is the *climate.OneBuilding.Org* project (Lawrie and Crawley, [2023](#)), a repository of building simulation climate data. This platform offers climate data specifically designed to support building simulations, including

EnergyPlus files (EPW) from various organisations and countries. In the Netherlands, this dataset includes data from the same weather stations as the KNMI.

3.1.2. NON-GEOSPATIAL DATA

3.1.2.1. BUILDING DATA

Non-geometric data encompass all data that characterise the building, and these are independent of or do not vary based on the geometric representation of the buildings. In UBE, the building's *function* is essentially non-geometric information, and several parameters depend on this attribute. In the case of NTA 8800, these parameters have standard values per square meter, as shown in [table 2.6](#). *Occupancy* data correspond to the human activities that influence energy consumption, such as opening hours for offices or educational buildings. Information about heating, ventilation, and air conditioning (HVAC) systems, which include water pumps, heaters and boilers, is fundamental due to the "energy consumption" of these systems, as well as their operational capacity, which is crucial for maintaining occupant health and comfort inside buildings.

The *number of storeys* affects the total floor area, influencing the internal loads due to occupancy, lighting, and equipment. Knowing the *number of living units* within a building is equally important; for example, a residential building with multiple living units will have specific characteristics, such as the *number of kitchens and bathrooms*, which vary depending on the building *usage*. The building type and construction characteristics significantly affect its energy performance. For example, the physical properties of a two-storey residential house differ from those of a twelve-storey apartment block, as the construction materials and techniques vary. Similarly, the year of construction influences the construction methods used, as these are characterised by regulations and state-of-the-art of construction techniques that were in place at the time of the construction (Neufert *et al.*, 2021; Rijksoverheid, 2024a).

3.1.2.2. PHYSICS DATA

The performance of a building strongly affects its *use* in terms of indoor climate, thermal comfort, air quality, and lighting, which directly influence its "energy consumption". Therefore, knowing the construction materials and techniques used for each part of the building is essential. This information, for example, includes critical data on thermal transmittance (*U-value*), the solar energy transmittance (*g-value*) for windows of the surfaces that form the thermal envelope, and the building's *infiltration rate*. Additionally, factors like the building's infiltration rate, colour, and surface reflectance play a significant role when determining heat absorption, heat loss, and overall energy demand.

3.1.2.3. ENERGY DATA

Energy data on theoretical energy demand or actual data on "energy consumption" serve two primary purposes: input for data-driven UBEM models and as ground truth for accuracy assessment of the UBEM results. These data can be obtained from energy performance certificates and actual "energy consumption" values, typically disaggregated into electricity, gas, and heating. Due to privacy restrictions, direct access to detailed "energy consumption" data is often limited. As a result, aggregated data at the postal code, district, or city level is commonly used as an alternative.

In the case of the Netherlands, these data are published by the Dutch Central Bureau of Statistics (CBS). However, these are aggregated at the postcode-6 level (the highest spatial resolution by postcode in the Netherlands). This dataset contains average supplies of natural gas and electricity to homes and businesses (CBS, 2023).

3.2. DATA AVAILABILITY

The analysis of the data requirements identified the data that are needed for UBEM for my thesis. However, some datasets have a multipurpose usage (i.e., the s3DCM) and, therefore, may contain different information than is required for UBEM. Hence, I discuss only those entities and attributes that directly relate to the research questions from section 1.3, with the scope of using open data.

3.2.1. GEOSPATIAL DATA

3.2.1.1. BAG

The basic registration addresses and buildings (in Dutch "*Basisregistratie Adressen en Gebouwen*"—BAG), contains information on all addresses and buildings in the Netherlands, such as year of construction, footprint gross floor area, purpose and their location. Municipalities are source owners of the BAG and are responsible for data quality and registration. Kadaster publishes the National Address Data System (in Dutch "*Landelijke Voorziening BAG*"—LV BAG), as open data available to the public (Kadaster, 2024c).

This dataset is updated monthly, among the available data formats, including OGC services, APIs, and ATOM services published by Kadaster (2024b). I use the XML format ATOM service since it provides the complete dataset. The BAG is a concurrent dataset, meaning that its tables contain historical records. For the dataset updated to 8th August 2024, table *verblijfsobject* contains 22'956.576 records corresponding to 10'234.487 unique living units (in Dutch "*verblijfsobject*")—hereafter referred to as vbo—. A graphical summary of the data available in the BAG is shown in figure 3.4. Table 3.1 presents the attributes available in all entities of the BAG. A package diagram of the entities available in the BAG is shown in figure 3.2. A simplified Entity Relation diagram of the BAG is shown in

figure 3.5; the attributes highlighted in colour indicate those that relate entities to each other.

Table 3.1: Description of common BAG attributes among its entities

Attribute	Description	Data type	Unit
ogc_fid	Unique database identifier	serial	-
identificatie	Cadastre identifier	text	-
status	This attribute indicates the situation of the element for the corresponding mutation. Examples of this attribute include: permit granted; in use; renovation; demolished; out of use	text	-
voorkomen	Sequential number that increases for any mutation that happens to a cadastre element. This value is unique per identificatie	serial	-
wkb_geometry	It contains the geometry representation of the element	geometry	-

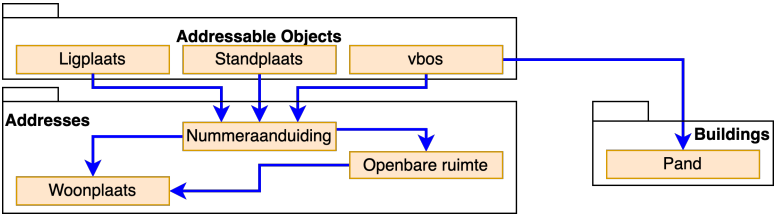


Figure 3.2: Overview of the BAG entities

Where:

Ligplaats	Places in the water. Floating objects suitable for residential, commercial, or recreational purposes
Standplaats	Designated piece of land intended to place an element not directly connected to the ground permanently, e.g., a camp. It is suitable for residential, commercial, or recreational purposes
Verblijfsobject (vbo)	The smallest unit of use located within one or more buildings and suitable for residential, commercial, or recreational purposes, accessed via its lockable entrance from the public road, a yard, or a shared traffic area, can be the subject of property rights transactions and is functionally independent
Nummeraanduiding	Address indicator assigned by the municipality to Verblijfsobject, Standplaats or Ligplaats
Openbare ruimte	Outdoor space located within a single residential area
Pand	Building, it is an independent unit connected to the ground, accessible and lockable
Woonplaats	Part of the territory of a municipality designated as a residential area

The distribution of buildings based on their construction period is presented in figure 3.3, 73.78% of the building stock was constructed before the 21st century, indicating a high potential for renovation. Furthermore, only 12.54% of buildings have been built since 2012, meaning they comply with the latest construction regulations in the Netherlands (Rijksoverheid, 2024a). However, the building

dataset is incomplete; for instance, function data is missing for 41% of buildings. Among the 59% with available function data, [figure 3.4b](#) shows the distribution across vbos. The detail attributes per entity are defined in [table 3.2](#).

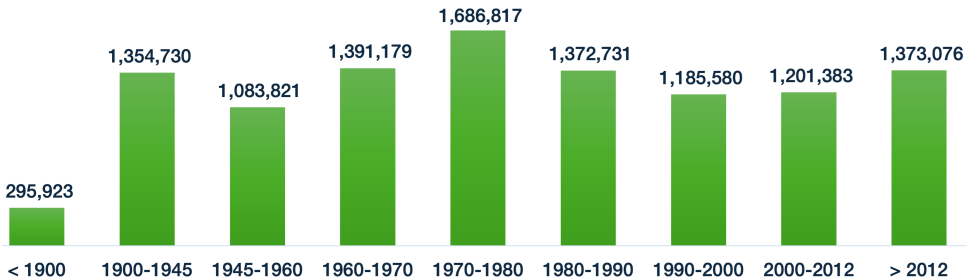


Figure 3.3: Properties by the year of construction

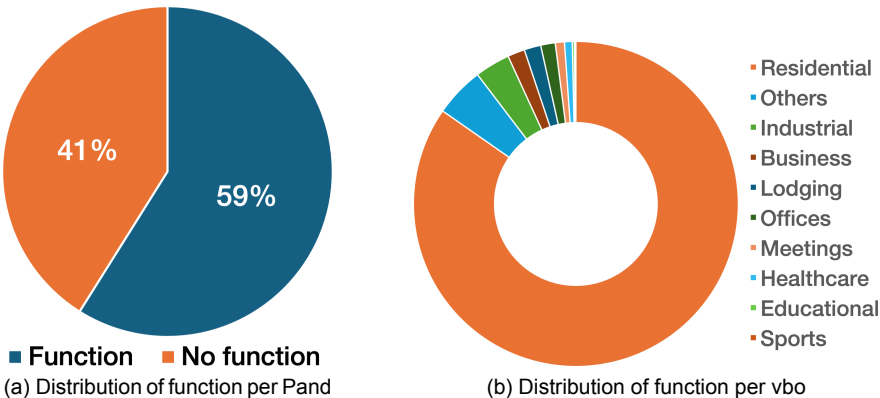


Figure 3.4: Basic characteristics of the BAG data

Table 3.2: Description of the attributes kept by BAG entity

Attribute	Description	Data type	Unit
Pand			
oorspronkelijk bouwjaar	year of construction	Date	-
vbo			
gebruiksdoel	Function. It is a categorization of the purpose of use of the vbo as included in the permit or determined by survey (Kadaster, 2024e)	text	-
oppervlakte	Usable area. In terms of the BAG, a vbo is first delineated, then its usable area is determined (Kadaster, 2024e). It uses the NEN 2580:2007 standard to measure it. This method is also used for the "Valuation of immovable property" (WOZ in Dutch). Among the indications, it consists of the sum of the usable living area and the usable area of other indoor spaces. It applies if the other indoor space is an internally accessible indoor area of a vbo; Indoor spaces that are not internally accessible must be reachable from outdoors; it is considered a usable area if the indoor area has a ceiling higher than 1,5m (NEN, 2007)	Real	m ²
hoofd adres nummer aanduiding ref	main address number indication reference. This attribute is the foreing key of the nummeraanduiding entity	text	-
pandref	This attribute is the foreing key of pand entity. However, in this case it is an array of values since one vbo can be located in multiple buildings, that is the case of 224.521 vbos	array	-
Nummeraanduiding			
huisnummer	building number	integer	-
huisletter	building letter	text	-
huisnummer toevoeging	building number addition, a further addition to a building number or a combination of building number and building letter	text	-
postcode	post code	text	-
openbareruimte			
naam	Street name	text	-
type	Street tyupe	text	-
woonplaats			
naam	Name of the residential area	text	-

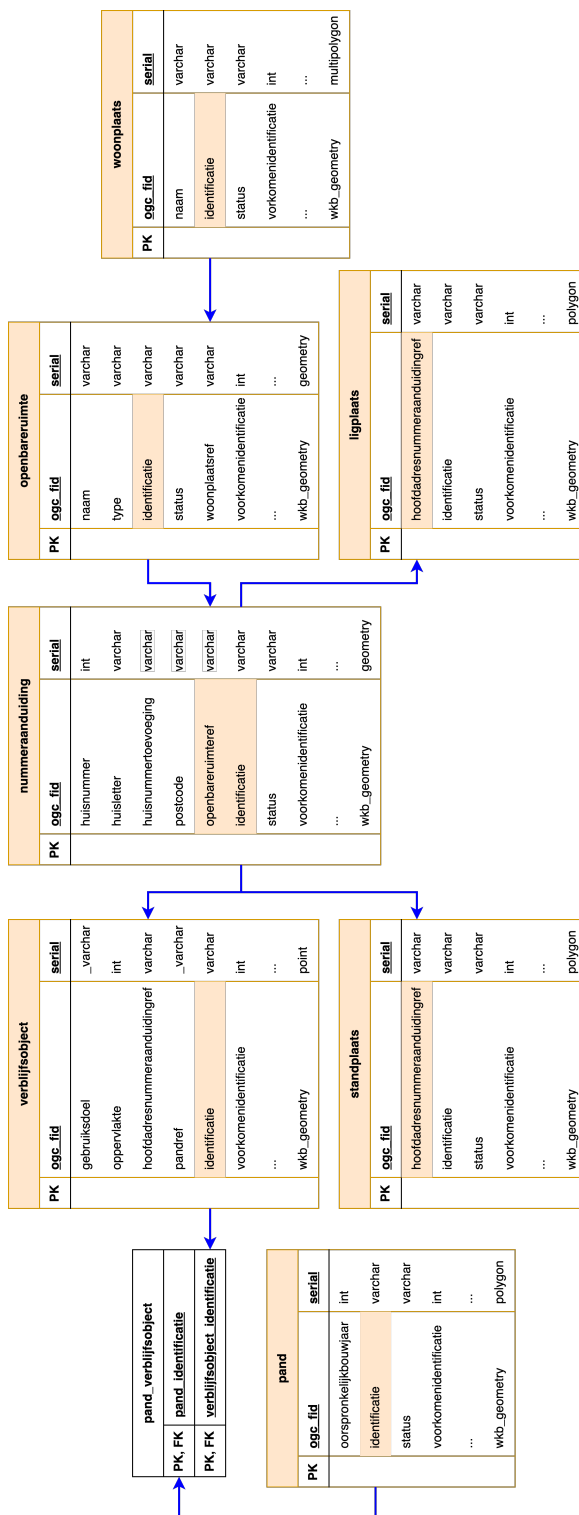


Figure 3.5: BAG Entity Relation diagram

3.2.1.2. 3DBAG

In 2021, the 3D Geoinformation Group at TU Delft published a country-wide s3DCM dataset, called 3DBAG, containing all buildings in the Netherlands in three different levels of detail (LoD1.2, LoD1.3 and LoD2.2) (Peters *et al.*, 2022). It combines two main datasets, the BAG and the current elevation dataset of the Netherlands (in Dutch "*Actueel Hoogtebestand Nederland*"), hereafter referred to as AHN, which is the digital height map of the Netherlands that contains detailed and precise height data (point clouds lidar) with an average of at least 8 points per square meter (Rijkswaterstaat, 2024). A sketch of the representations available in the 3DBAG is shown in figure 3.6. It is important to mention that the building's footprint in the 3DBAG represents the area covered by the projected roof on the ground since the 2D BAG building polygons, on which the 3DBAG is based, represent the outer lines of building roofs as seen from above and not the footprint. The attributes available from the 3DBAG, as detailed in table 3.3, are kept for the processes discussed in the following chapters of this PhD thesis. Some of these attributes come from the BAG, while others are added during the 3D reconstruction process.

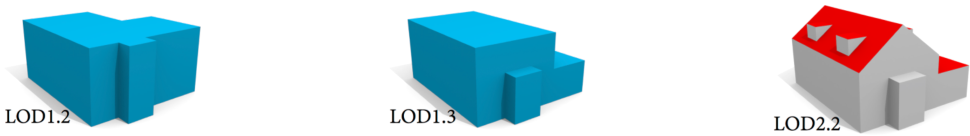


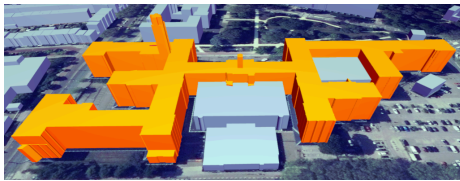
Figure 3.6: LoDs available in the 3DBAG. Sample images from (Biljecki, 2017)



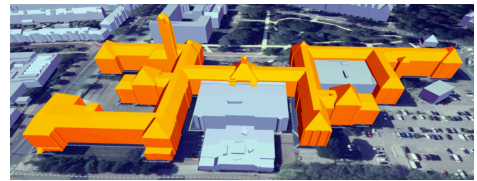
(a) 3D mesh ©2024 Apple Inc.



(b) 3DBAG LoD1.2



(c) 3DBAG LoD1.3



(d) 3DBAG LoD2.2

Figure 3.7: Several 3D representations of the TU Delft Architecture Faculty building (BK)

Table 3.3: Description of the 3DBAG attributes keep for the energy applications

Attribute	Description	Data type	Unit
identificatie*	ID of a building	text	-
oorspronkelijk bouwjaar*	Year of construction of the building	date	YYYY
b3 bouwlagen	An estimation of the number of floors in the building. Estimates are only available for buildings with up to 5 estimated floors	Integer	-
b3 dak type	Roof type of the building. Values: <ul style="list-style-type: none"> • slanted: Roof with at least one slanted surface • multiple horizontal: Roof with multiple, only horizontal surfaces • single horizontal: Roof with a single horizontal surface • no points: No point was found for the building • could not detect: Could not detect a roof surface, even though points were found 	categorical	-
b3 h dak 70p	The 70th percentile elevation on roof surface based on reconstructed 3D model in LoD2.2. Given as elevation above sea level (Amsterdam Ordnance Datum)	Real	m
b3 h maaiveld	Elevation above sea level at the ground level of the building. Calculated as the 5th percentile of the ground points found within a 4-meter radius of the building	Real	m
b3 kas warenhuis	The building is greenhouse or warehouse (according to TOP10NL) or has an area of over 70000m ²	categorical	true/false
b3 opp buitenmuur	Total area of the exterior walls	Real	m ²
b3 opp dak plat	Total area of the flat parts of the roof	Real	m ²
b3 opp dak schuin	Total area of the sloped parts of the roof	Real	m ²
b3 opp grond	Total ground floor area	Real	m ²
b3 opp scheidings- muur	Total area of the party walls	Real	m ²
b3 val3dity lod22	Val3dity error codes for the LoD2.2 3D model. Empty list means valid geometry	list	-
b3 volume lod22	Volume of the LoD2.2 model	Real	m ³

3.2.1.3. BAG-PLUS

This is a suggested complementary dataset to the existing BAG data for the whole Netherlands. However, only the municipality of Amsterdam publishes the BAG-plus as open data to support its processes (Gemeente Amsterdam, 2022). The additional attributes include information on the building type and the number of storeys of the building. The published dataset contains the data from the municipalities of Amsterdam and Weesp; however, data from Weesp are limited, as it was incorporated into Amsterdam on 24th March 2022. Relevant attributes are listed in [table 3.4](#).

Table 3.4: BAG-plus Additional Attributes

Attribute	Data Type	Description
ligging omschrijving	string	Property type
type_woonobject	string	Type of residential unit: One, multiple
aantal_bouwlagen	Integer	Number of floors of the property
hoogste_bouwlaag	Integer	The Highest storey of the property
laagste_bouwlaag	Integer	The Lowest storey of the property

3

As of September 2024, it contains records of 1.212.214 vbos, corresponding to 555.079 buildings.

3.2.2. NON-GEOSPATIAL DATA

3.2.2.1. PHYSICS-RELATED DATA

Currently, no open dataset contains detailed physics-related data of buildings in the Netherlands. The only dataset that I am aware that keeps a record of the renovation of buildings is the Nederland Housing under research (in Dutch "*WoonOnderOnderzoek*"–WoON), which is a large-scale survey performed by CBS. This survey provides data on several components of housing conditions, including energy-related building characteristics such as HVAC systems, solar panels, and building materials (e.g., insulation and windows). Conducted between August 2020 and September 2021, the survey collected responses from 46,700 participants (Stuart-Fox *et al.*, 2022). However, due to privacy regulations, these data are only accessible via CBS servers, and only aggregated values can be exported to protect respondents' data. Therefore, these data will not be used in this research due to access restrictions, which is out of the scope of this thesis.

Therefore, the alternative is to use physics-related data based on building archetypes, that is the case of the Episcopo TABULA, and the example buildings report (in Dutch "*Voorbeeldwoningen*"). One of the products of the TABULA project, as discussed in section 2.2.2.1, is the definition of building typologies across Europe. The project defines four main building types for the participating countries: *Single-Family House* (SFH), *Terraced House* (TH), *Multi-Family House* (MFH), and *Apartment Block* (in Dutch "Flatwoning") (FW).

In the Netherlands, this classification is extended 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).

Building types can be described as follows:

- **Vrijstaand (single family house):** Detached buildings with 2 to 4 storeys that do not share walls with neighbouring buildings

- **Twee-Onder-Een-Kap (semi detached house):** Buildings that share a common wall with one neighbour; the literal translation from Dutch is "two under one roof." These buildings typically have 2 to 4 storeys
- **Tussenwoning (middle-row house):** Buildings constructed in a row of joined houses, sharing at least two common side walls with neighbouring units. These buildings typically have 3 to 5 storeys
- **Hoekwoning (end-row house):** Buildings located at the end or corner of a row of joined houses. These buildings typically have 3 to 5 storeys
- **Maisonnette:** living units spread across several storeys within multi-storey residential buildings
- **Portiek (common staircase without galleries):** One-storey living units within multi-storey residential buildings, accessible through a closed porch
- **Galerij (common staircase with galleries):** One-storey living units in multi-storey residential buildings with elevator access, where units are accessible via an open gallery
- **Flatwoning (other apartment building):** A broad category for apartment blocks that do not fit into the other described categories

The definitions and specifications per building type are based on the example buildings report (in Dutch "*Voorbeeldwoningen*") 2022 (RVO, 2023). Oddly, it does not specify the cases for 1-storey buildings. A hierarchical diagram of these building types classes is shown in [figure 3.9](#).

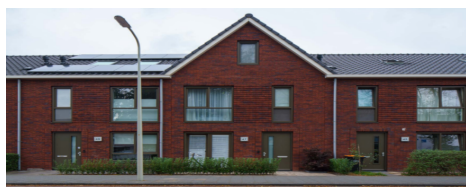
TABULA contains building physics data for three refurbishment scenarios: existing state, typical refurbishment, and Net Zero Energy Building (NZEB). These data correspond to walls, roofs, windows, and doors and have an associated U-value per each refurbishment scenario. Additionally, TABULA includes data on HVAC, ventilation, and DHW systems for each scenario. These data are used in the implementation described in [chapter 5](#). The classification of building types is further refined by construction year as building regulations — particularly those related to energy performance — have evolved. This distinction is shown in [table 3.5](#).



(a) Vrijstaand



(b) Twee-Onder-Een-Kap



(c) Tussenwoning



(d) Hoekwoning



(e) Flatwoning



(f) Maisonnette



(g) Portiek



(h) Galerij

Figure 3.8: Examples of the building types extracted from (RVO, 2023)

The TABULA classification of buildings in the Netherlands is based on the “Example buildings” report–Voorbeeldwoningen 2011– (Agentschap NL, 2011) by the Ministry of Royal and Internal Affairs (in Dutch “*Ministerie van Binnenlandse Zaken en Koninkrijksrelaties*”) RVO (2023) has updated this report, incorporating significant changes. The updated version includes the use of the Energy section of the WoON database, the adoption of the NTA 8800 as a method for computation of the Energy performance of buildings and the addition of two new construction year categories: 2006-2014 and 2015-2018 as shown in table 3.5.

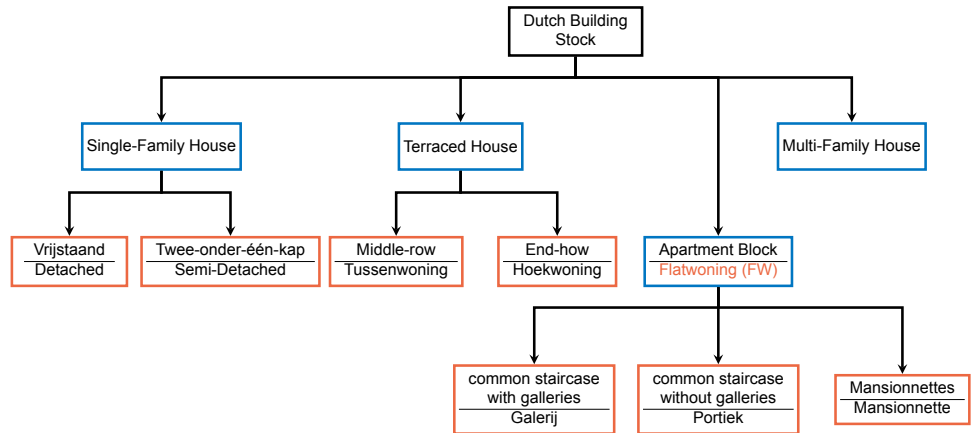


Figure 3.9: Dutch building types according to TABULA. Blue boxes indicate generic types; Orange boxes indicate the Dutch specific types

Table 3.5: Classification of the building types by year of construction

Building Type	up to 1945	1946-1964	1965-1974	1975-1991	1992-2005	2006-2014	after 2015
Vrijstaand		X	X	X	X	X	X
Twee-Onder-Een-Kap		X	X	X	X	X	X
Tussenwoning	X	X	X	X	X	X	X
Hoekwoning	X	X	X	X	X	X	X
Maisonnette		X	X	X	X	X	X
Galerij		X	X	X	X	X	X
Portiek	X	X	X	X	X	X	X
Flatwoning		X	X	X	X	X	X

3.2.2.2. ENERGY DATA

Since 1st January 2008, it is mandatory for sale/renting or to hand over any vbo in the Netherlands (not only for residential units) to have a registered energy performance certificate (EPC) (Rijksoverheid, 2024b). These data are collected by ep-online, which is the official Dutch database that contains the EPC and energy performance indicators of buildings. Furthermore, ep-online offers public access. This access is offered by Rest API and file formats: XML, CSV and XLSX. Due to the nature of the database, it is constantly updated, so file-based updates are available daily, and every first day of the month, RVO publishes a total consolidation of the database to download.

The open access dataset follows a simplified data model, one registry per row and when available or required, the corresponding attribute is filled. Currently, it contains 42 attributes (February 2025), and it will grow if future versions

require it. Furthermore, it contains information on several energy performance computation methods. Therefore, multiple attributes from the ep-online dataset are empty since the computation method does not require them; one example is the attribute *EnergieIndex* (performance index), which is not required for registration using the NTA 8800 method. As of September 2024, the ep-online dataset contains 5,587,693 records, corresponding to 981.940 buildings. The distribution of the EPC by vbo is shown in figure 3.10 and the distribution of the EPC by building type is shown in figure 3.11.

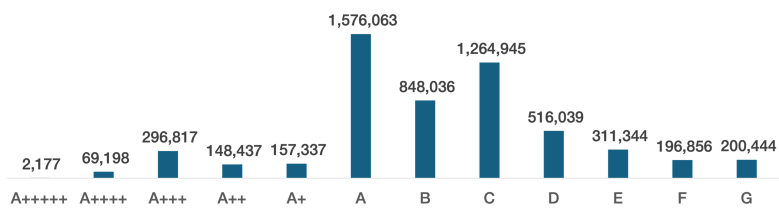


Figure 3.10: Distribution of the energy performance certificates by the vbo

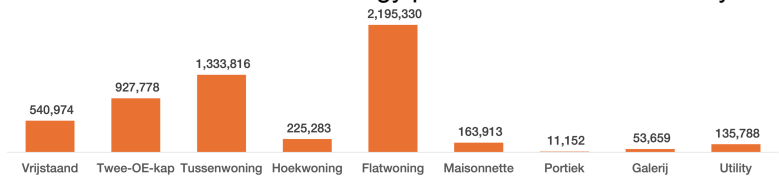


Figure 3.11: Distribution of the energy performance certificates by building type

Although the main purpose of this dataset is the EPC-related data, it offers a second purpose of my research as it contains the building type and subtype of the vbos; this only applies to flats shown in figure 3.8. Furthermore, in the 2022 report, the building type categories for Apartment Blocks (flatwoningen) were simplified by stopping classifying living units into Portiek, Mansionnette or Galerij. For the subtype of vbo, the classification is based on its location inside the apartment block as shown in figure 3.12.

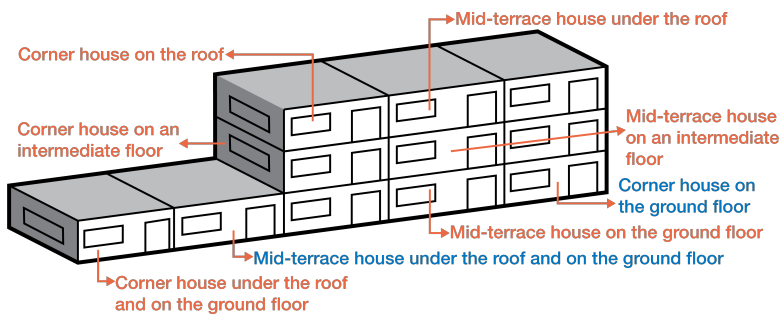


Figure 3.12: Building subtypes. Text colour is used for visual distinction and has no interpretative significance

Table 3.6: Excerpt of the attributes used in the ep-online dataset

Attribute	Description
<i>Gebouwklasse</i>	The type of building: residential or utility building.
<i>Gebouwtype</i>	Type of dwelling.
<i>Gebouwsubtype</i>	Dwelling subtype: the location of the apartment in the residential building.
<i>Gebruiksoppervlakte_thermische_zone</i>	Usable area of the thermal zone in m ² .
<i>Energieklasse</i>	The letter of the energy label (label class).
<i>Energiebehoefte</i>	The energy requirement in kWh/(m ² · y)
<i>PrimaireFossieleEnergie</i>	The primary fossil energy use in kWh/(m ² · y)
<i>Aandeel_hernieuwbare_energie</i>	The share of renewable energy in %.
<i>Warmtebehoefte</i>	Net heat demand for the energy performance compensation (EPV) in kWh/m ² · y
<i>BerekendeCO2Emissie</i>	The calculated CO2 emission in kg/(m ² · y)
<i>BerekendeEnergieverbruik</i>	The calculated total "energy consumption" in kWh/(m ² · y)

A summary of the open datasets available in the Netherlands and used in this research is shown in [figure 3.13](#), the figure follows the same colour scheme as in [figure 3.1](#). The lowest level (the darkest colour) indicates the attribute used from the corresponding dataset (mentioned at the fourth level).

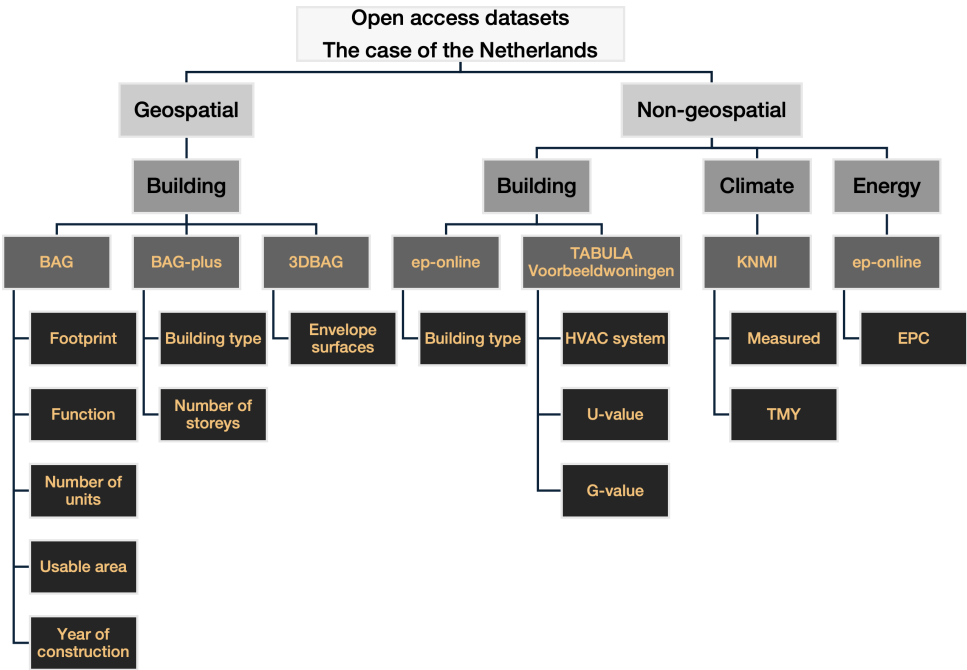


Figure 3.13: Graphical summary of the open datasets available in the Netherlands, from (León-Sánchez, Agugiaro *et al.*, [2025](#))

3.3. CONCLUSION REMARKS

The extensive data requirements of Urban Building Energy Modelling (UBEM), as outlined in [chapter 2](#), served as the foundation for identifying and presenting relevant open datasets available in the Netherlands. While specific requirements—such as geometric representations of buildings—are partially met by existing datasets, critical gaps remain. For instance, 3D building models lack classifications of boundary surfaces (e.g., glazed vs. opaque), limiting their utility for detailed energy simulations.

Regarding physics-related data, no nationally comprehensive datasets exist for building characteristics or renovation histories. Therefore, I will use standardised archetype-based data (e.g., construction periods, typologies), as discussed in [chapter 4](#). This chapter details the method and workflow to enrich semantic 3D city models (s3DCM) with the purpose of enhancing energy applications by integrating geospatial and non-geospatial data sources.

4

DESIGNING THE PIPELINE: DATA PREPARATION FOR ENERGY DEMAND ANALYSIS

This chapter describes the processes to prepare the data required for urban building energy modelling for my case study. First, the data preparation is detailed, focusing on compliance with the minimum UBEM requirements while dealing with the issues faced during data curation (e.g., inconsistencies, missing attributes, and scalability constraints). Second, some examples of computational solutions to resolve identified data gaps are introduced; these examples show some of the steps towards how several datasets can be integrated into a cohesive workflow.

This chapter presents the data pipeline of my PhD research. First, I describe the creation of the testbed for energy applications, which serves as the city-level case study ([section 4.1](#)). Next, I detail the development of country-level databases, including the definition of entities and their attributes ([section 4.2](#)). Third, I highlight key findings from various academic projects and studies conducted during my research ([section 4.3](#)). These projects focus on evaluating solar energy potential using enriched s3DCMs and implementing solutions to address data gaps identified in [section 4.2](#). Finally, I close the chapter with the key findings this chapter, including reflections on the testbed, the EnergyBAG DB and the implemented energy-related solutions ([section 3.3](#)).

4.1. TESTBED CREATION¹

The geographical scope of this PhD thesis is the Netherlands. It could be considered a big dataset since the Dutch building stock has more than 10.000.000 buildings. Therefore, we created a small dataset in the municipality of Rijssen-Holten, which is located in the eastern part of the Netherlands (figure 4.1); it has approximately 38,000 inhabitants and around 23,000 buildings. The creation of this dataset required the consolidation of multiple data sources with manual work to guarantee the quality of the dataset, enabling its use in various follow-up applications. The resulting dataset is the case study of several sections of this PhD thesis since I have stricter control of the data available than the country-wide dataset.

4



Figure 4.1: Rijssen-Holten, thick red line, location in the Netherlands

The sole source of geospatial information on buildings is the 3DBAG using its LoD2.2 geometric representation. I processed the data from the 3DBAG, including the following considerations: Buildings with more than one part are hierarchically organised using CityGML classes, such as Building and BuildingPart. The unique building ID is assigned only to the root Building object containing the parts. The former contains the general semantic information for the whole Building, while the latter contains the detailed data as well as the geometries. Buildings(Parts) contain the address objects according to the BAG. The coplanar surfaces from the same object (Building or Building Part) and thematic definitions

¹This section has been published in "Testing the New 3D BAG Dataset for Energy Demand Estimation of Residential Buildings". In ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-4/W1-2021:69-76. 2021; by (León-Sánchez, Giannelli *et al.*, 2021) and "Creation of a CityGML-Based 3D City Model Testbed for Energy-Related Applications". In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48:97-103. 2022; by (León-Sánchez, Aguiaro *et al.*, 2022)

are merged to reduce the number of surfaces. Further tests were carried out regarding the thematic surfaces composing the LoD2 model (i.e. GroundSurfaces, WallSurfaces, RoofSurfaces). Some errors were found (and corrected), such as:

- Missing GroundSurfaces in a (very) limited number of Building(Parts). The missing geometries were generated and added by projecting and dissolving the sloped roof surfaces of the affected Building
- Thematic surfaces missing a classification or being classified incorrectly (e.g. as InteriorWallSurfaces instead of WallSurfaces). (Re)classification rules were defined based on the normal vector of each surface.

After the geometrical computations and adjustments, we joined the s3DCM with a survey done by the municipality of Rijssen-Holten that included the number of storeys per building (parts). From the dataset joins, only 60% of buildings obtained function information; in the case of the number of storeys, only around 50% of buildings were surveyed. Therefore, the remaining data required manual collection using Google Street View (Google, 2022) and Mapillary (Mapillary, 2022). Those platforms offer street-level-based imagery. Figures 4.2a and 4.2b show a sample of the exact location at the municipality of Rijssen around the EPSG:4326 coordinates $\varphi = 52^{\circ}18'40.1''$ $\lambda = 6^{\circ}31'42''$ for both datasets.



(a) Google Street View sample



(b) Rijssen Mapillary sample

Figure 4.2: Excerpt of the street view data

The building reconstruction method of the 3DBAG does not split or subdivide each *Building* into smaller *BuildingParts* since there is no open data available for all buildings that support such process. Therefore, we collect the highest value of the number of storeys of a building, both above and below the ground.

Figure 4.3 shows an example of a residential building in Rijssen with two methods for determining its number of storeys. Figure 4.3b presents Rijssen-Holten's survey approach: the building is split into parts according to each part's storey count, with the part outlined in magenta representing one storey and the section outlined in green indicating two. Figure 4.3a shows our approach. Because we rely on the 3DBAG as our source for building geometry, we do not subdivide the model; instead, we assign a single storey count equal to the maximum value across all parts, which is in this case 2. Although this is a simplification for the

workflow, it may lead to an overestimation of any energy-related calculations, which is the case of the building's energy demand when using reduced-order models because the attribute is affectively over-assigned and a close to reality value would be a decimal between 1 and 2.

4

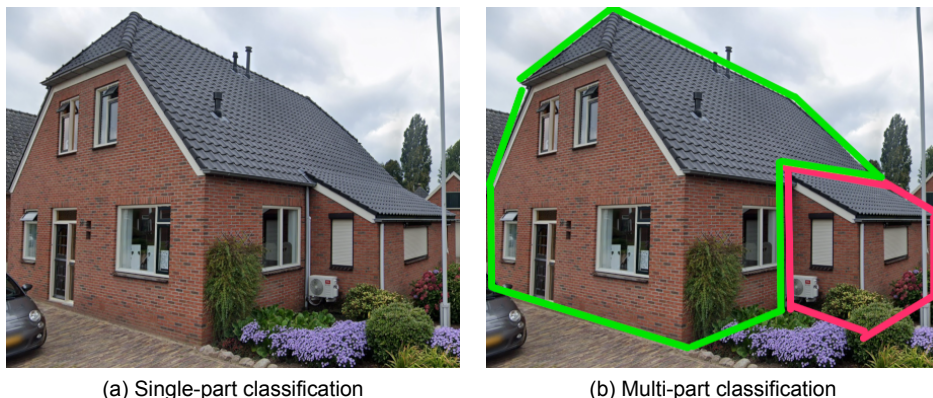


Figure 4.3: Example of a 2-storey building

For the building type classification, we used simplified rules to speed up the classification process, e.g. a *Building(Part)* without adjacent ones and one vbo can be classified as Single Family House. However, when trying to use the attribute available in the ep-online dataset, we found classification inconsistencies as the example in presented in [figure 4.3](#) in which an apartment block has associated three different building type categories: galerij, portiek, flatwoning (apartment block).



(a) Google Street View



(b) Energy label dataset

4

Figure 4.3: Example of a multi familiar residential building with several classification types

Therefore, only in the case of buildings assigned with one building type class were assumed correct, and for the remaining ones, this attribute was collected manually. These time-consuming and manual processes allowed me to get a deeper insight and understanding of the issues that characterise the to-be-integrated datasets.

We created this dataset to test heterogeneous Building Energy Simulation –BES– tools with a reference dataset that is sufficiently rich and reliable in terms of attributes, semantics, and geometries. In this process, we also applied manual edits. However, for the EnergyBAG DB, manual data collection is not feasible, given the size of the Dutch building stock (circa 10 million buildings). Consequently, data cleaning processes have been automated based on the lessons learnt during the testbed creation in Rijssen-Holten. For example, we do not perform validations on the input geometries of the 3DBAG since they are already validated using *val3dity* (Ledoux, 2013, 2018) and its result is already available as a generic attribute per *BuildingPart*. Nevertheless, the same pre-processing steps are applied to the whole Netherlands.

The software tools used in the design of the pipeline include:

- **DuckDB²**: The current version does not support 3D geospatial data. I

²This is a fast open-source and multi-platform database system. I used version 1.0; further details at

used it for alphanumeric and 2D geospatial data operations, such as reading and aggregating the data from BAG, BAG-Plus, and ep-online and assigning the H3 index to each building³

- **PostgreSQL and PostGIS**⁴: It is used for data management of all databases
- **FME Form**⁵: It is used to read and post-process the 3DBAG CityJSON-based files, including adding *Building(Part)* and *thematicSurfaces* feature extraction.
- **Python**⁶ It is used for data cleaning and preparation from the resulting process from DuckDB and FME and for updating the data stored in the 3DCityDB.

4

4.2. DATABASES CREATION

4.2.1. BUILDING PHYSICS DATA

As specified in [chapters 2 and 3](#), the physical properties of the building thermal envelope are essential for conducting energy performance analyses. While such data may be accessible for individual buildings, compiling detailed information for all buildings at a national scale remains unfeasible. To address this challenge, I have created a Building Physics Library Database (LibraryDB) containing standardised building physics parameters for predefined archetypes. These archetype-specific values can then be applied in workflows such as energy performance certification, retrofit planning, or large-scale energy demand simulations.

The TABULA project ([section 3.2.2.1](#)) serves as the foundational source of data for this work, from which the initial data model is derived. In developing the library data model, I emphasise ensuring compatibility with heterogeneous data sources and multiple geographical contexts. However, this approach means the generalisation of buildings according to the archetypes defined by the TABULA project ([figure 3.9](#)). The three main schemas of the LibraryDB are presented in [figure 4.4](#) followed by a description of the tables per schema in [table 4.1](#).

The entity relationship diagram resulting from the creation of the LibraryDB is presented in [figure 4.5](#). I designed the data model to support the storage of data from multiple sources. For instance, data available from voorbeeldwoningen (RVO, [2023](#)) has been incorporated into the database I use for my implementation.

(Raasveldt and Muehleisen, [2024](#))

³This is done using the duckdb h3 extension (Brodsky and Piovesan, [2022](#))

⁴I used versions 16.4 and 3.5, respectively. Further details at (PostGIS Project Steering Committee, [2023](#); PostgreSQL Global Development Group, [2024](#))

⁵I used version 2024.1.0. Further details at (Safe Software, [2024](#))

⁶I used version v3.11.5. However, I tested it using versions from v3.10.X to v3.12.X. Further details at (Python Software Foundation, [2024](#))

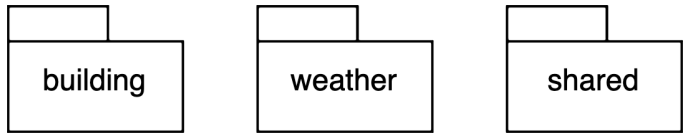


Figure 4.4: Schemas of the LibraryDB

Table 4.1: Description of the schemas and tables of the LibraryDB

Name	Type	Description
shared	schema	Contains the root tables that define the basic parameters for the database
libraries	table	Contains the definition parameters of the data sources (libraries) stored in the database, for example, EPISCOPE TABULA (EPISCOPE Project, 2017), Voorbeeldwoningen (RVO, 2023)
orientation_class	table	Contains the orientation classes for surfaces based on the azimuth
parameter_info	table	Defines the type of elements available in the database
building	schema	This schema deals with the building-related data
age_class	table	Contains the classification parameters of buildings based on the year of construction
building_class	table	Contains the classification parameters of buildings based on the building type
building_function	table	Contains the classification parameters of buildings based on the function
element	table	Contains the elements to which the parameter (i.e., U-value) is assigned, for example, a Whole building, Ground shell, Outer wall, Inner Wall, and inner wall
energy_class	table	Contains the definition parameters of the energy label classes. The table includes the range of values of the class and an RGB colour definition for visualisation purposes. It is based on the Dutch regulation (Rijksoverheid, 2024c), with classes specified according to the building's function
layer_material	table	Describes construction layers composed of multiple materials
layer	table	Contains the different materials that define the building envelope (Wall, Roof, Ground)
material	table	Contains the data on construction materials
opening	table	Contains the data regarding windows and doors
parameter	table	A consolidation table that relates the following: age_class, building_class, building_function, refurbishment, element, parameter, layered_material, opening
profile	table	Contains the occupancy profiles of buildings
refurbishment	table	Stores parameters describing refurbishments
type_class	table	Contains the building type classes
weather	schema	It deals with the weather-related data
climate_parameter	table	Contains typical statistical climate parameters
weather_station	table	Stores descriptive attributes of weather stations
weather_parameter	table	Contains the time series data for a given weather station

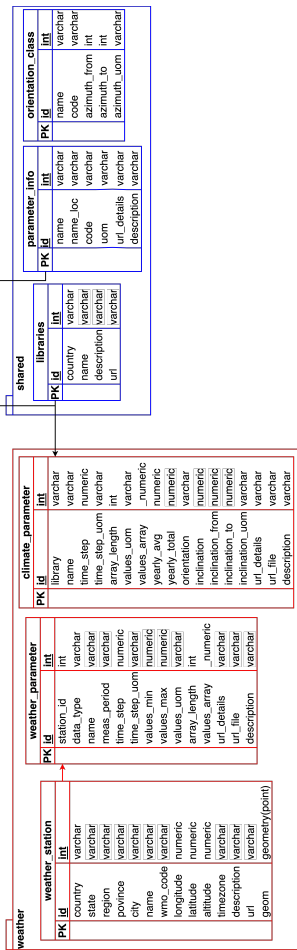


Figure 4.5: Entity relationship of the LibraryDB

4.2.2. ENERGYBAG DB: THE CREATION OF A DATASET FOR THE WHOLE NETHERLANDS

The pipeline shown in figure 4.6, is a further development of the work described in section 4.1. It includes the software tools used to enhance the existing semantic 3D city model with the open datasets to create the database used in this research: the CityGML+Energy ADE dataset (hereafter referred to as the EnergyBAG DB), which is managed by a 3DCityDB instance. Figure 4.7 shows a simplified entity relationship diagram linking all datasets and having the buildings in the 3DBAG as the reference object of integration of data.

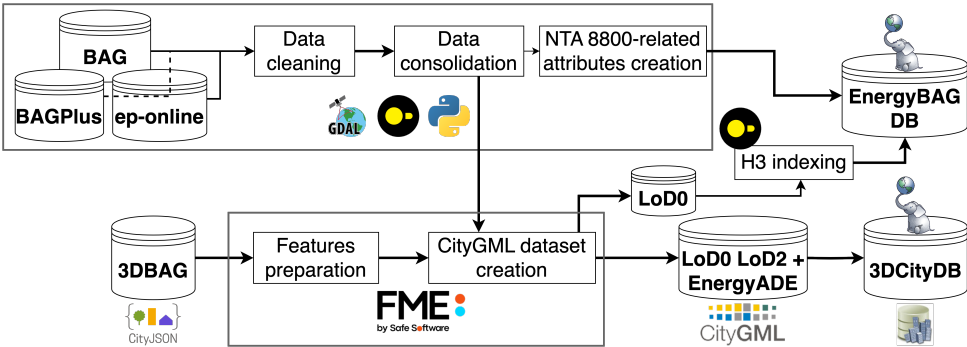


Figure 4.6: Data pipeline for the EnergyBAG DB creation

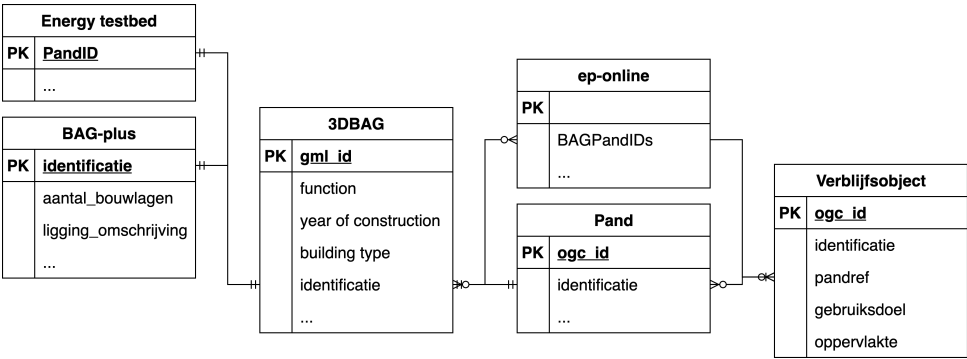


Figure 4.7: Entity relation diagram of the multiple datasets used in my research

The various input datasets may contain overlapping data for the same building, most notably in the building type and the number of storeys. Both attributes are available in the energy-tested dataset and in BAG-plus, while ep-online do not provide storey-related data. To resolve data conflicts, I applied the following hierarchy criteria when assigning these attributes:

1. Testbed dataset, because it was manually collected and verified.

- 2. BAG-plus, which is continually updated according to a consistent set of criteria
- 3. ep-online, in which new records are appended, but existing entries are not modified

The hierarchy indicates that in those cases when more than one dataset is available, it follows the strict order. Additionally, since the primary source of this PhD research is the BAG, I only fall on the other datasets in cases where an attribute is not available from the BAG.

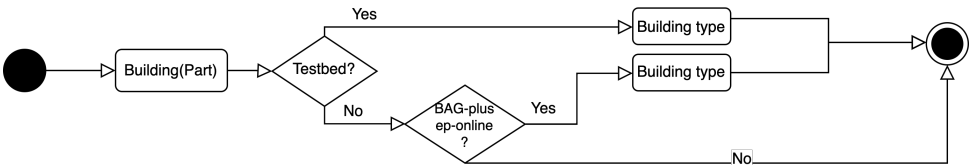


Figure 4.8: Workflow for assigning the building type attribute

A summary of the attributes available for the *Building CityObject* is shown in [table 4.2](#). This table describes the additional *generic attributes*. The table also includes extra attributes extracted from the geometries of the s3DCM (shown in [figure 4.10](#)), which are required for my research. In the case of *BuildingParts*, this *CityObject* is present if the *Building* contains more than one construction. For my PhD thesis, I use the term *Building(Part)s* when referring to either both *Building(s)* and *BuildingPart(s)*. The attributes available for the *BoundarySurface* of the *Boundary(Part)s* are shown in [table 4.3](#).

Table 4.2: *Building(Parts)* attributes available in the resulting CityGML dataset

Attribute	Description	Unit	Data Type
Building			
pand_id	Pand identifier		String
3dbag_tile	Name of the 3DBAG tile		String
lod_max	Maximum level of detail of the building		Integer
is_single_part	Indicate if the building is single- or multi-part		String
list_adjacent_buildings	List of the gml_id of the adjacent buildings		String
n_adjacent_buildings	Number of adjacent buildings		Integer
property_area	This attribute corresponds to the Pand m^2 area. Extracted from BAG		Measure
usable_area	Usable area of the building	m^2	Measure
vbos	number of vbos registered in the building		Integer

Continued on next page

Table 4.2 – continued from previous page

Attribute	Description	Unit	Data Type
(function)_vbos	Number of vbos per function. The text (function) shall be replaced by the corresponding function defined in the BAG. This attribute only appears when the building has a vbo with the corresponding function		String
(function)_area	Total usable area per function. It is defined as the (function)_vbos attribute	m^2	Measure
is_greenhouse	Indicates if the building is a greenhouse		String
status	Property status. Extracted from the BAG		String
roofArea	Area of the roof surfaces of the building	m^2	Measure
compactness	Relation between the volume and the enveloped area: volume/envelope		Real
roof_envelope_ratio	Ratio between the roof and envelope areas		Real
usable_area	Total usable area of the building	m^2	Measure
building_type	Building type according to the TABULA specification		String
dutch_building_type	Building type according to the Voorbeeld-woningen specification		String
energy_class	Energy labels of the building according to ep-online		String
envelope_area	Enveloped area of the building	m^2	Measure
ExteriorWalls_area	Exterior walls area	m^2	Measure
footprint_area	Area of the footprint of the building	m^2	Measure
footprint_height	Height of the footprint of the building	m	Measure
footprint_perimeter	Perimeter of the footprint of the building	m	Measure
PartyWalls_area	Area of the party-walls (shared surface) between adjacent buildings	m^2	Measure
roof_footprint_ratio	Ratio between the roof and footprint areas		Real
RoofFlatPart_area	Area of the flat parts of the roof	m^2	Measure
RoofSloppedPart_area	Area of the slopped parts of the roof	m^2	Measure
val3dity	Code according to val3dity of the reconstructed geometry of the building		String
3dbag_storeys	number of storeys of the building according to the 3DBAG		Integer
volume	Gross value of the building	m^3	Measure

Table 4.3: *BoundarySurface* attributes available in the resulting CityGML dataset

Attribute	Description	Unit	Data Type
inclination	Inclination angle of the surface measured upwards from the horizontal plane	degrees	Measure
azimuth	Azimuth angle of the surface measured clockwise from North; -1 is NULL	degrees	Measure
surf_normal	Normal surface value, as a comma separated value of n_x , n_y , n_z		String
direction	Direction of the surface. Possible values are: N, NE, E, SE, S, SW, W, NW		String
lod2_area	3D area of the surface	m^2	Measure

4

The presence of *ThermalZone*, an Energy ADE *CityObject*, indicates that the corresponding *Building(Part)* has registered at least one vbo. The number of *ThermalZones* corresponds to the distinct functions present in the respective *Building(Part)*. However, these *CityObjects* lack geometrical representation, as there is no open dataset in the Netherlands providing the interior representation of buildings. Therefore, a *ThermalZone* contain only alphanumeric data, which are shown in [table 4.4](#). In the case of the *ThermalZone UsageZone* Energy ADE *CityObjects*, I only store data if they are coming from actual data sources. Other values are not used even if they may be set at runtime. When it comes to the thermal hull of the *ThermalZone*, I assume only one thermal zone per *Building(Part)*. The *infiltrationRate* attribute is not set, as it will be assigned at a running time from the libraryDB.

Table 4.4: *ThermalZone* attributes available in the resulting CityGML dataset

Attribute	Description	Unit	Data Type
floorArea	Total usable area for that function in the <i>Building(Part)</i>	m^2	Measure
isHeated	This attribute characterises the thermal zone condition, indicating if it should be processed (when isTrue)		Boolean

As with the *ThermalZone*, the presence of *UsageZone* Energy ADE *CityObject* exists in the resulting dataset when the *Building(Part)* contains at least one vbo. Its importance lies in its relationship to an occupied *ThermalZone*, which must connect to a *UsageZone* that defines the boundary conditions for the calculation of the heating and cooling demands (Agugiaro, Benner *et al.*, 2018). The attributes available are shown in [table 4.5](#). The attributes available for the *BoundarySurfaces* of the *Building(Part)*s are shown in [table 4.3](#). The consolidated values of the produced dataset are shown in [table 4.6](#).

Table 4.5: *UsageZone* attributes available in the resulting CityGML dataset

Attribute	Description	Unit	Data Type
floorArea	Total usable area for that function in the <i>Building(Part)</i>	m^2	Measure
usageZoneType	Indicates the usage of the corresponding part of the <i>Building(Part)</i>		String

Despite the use of multiple data sources, ~90% of the resulting dataset does not contain information regarding the building type, as shown in [figure 4.9](#). These values correspond to buildings; thus, the values per class remain low. Later in this chapter I will show, how this issue can be addressed.

However, Object Vision ([2023a](#)), a company based in the Netherlands, published its GeoDMS framework for modelling geographic datasets (Object Vision, [2023b](#)), which includes the estimation of the residential type of buildings (*woonpand_type*) based on neighbouring buildings (*Pand*) and the number of registered residential living units (*vbos*). The classification categories in the method include: *Vrijstaand*, *Twee-Onder-Een-Kap*, *Tussenwoning*, *Hoekwoning*, *multi-family building*. Additionally, Esri Netherlands offers an online feature layer with the classification of the residential buildings based on the BAG (esri, [2025](#)). This feature service is offered as subscriber content. The classes available in this dataset include: *Vrijstaande*, *Twee-Onder-Een-Kap*, *Tussenwoning*, *Hoekwoning*, *Appartement*.

The reasons for not using the aforementioned datasets twofold. First, the PBL/TNO Hestia dataset was unknown to us at the time we began our experiments to calculate this attribute in 2022-2023, which led to the development of the MSc thesis by Chris Poon (see [section 4.3.2](#)). Second, the data published by Esri is subject to licensing restrictions, which is against of the scope of this PhD thesis about the use open access data.

Table 4.6: Summary of the s3DCM created in my research

CityObject	Count
Building	10,413,738
BuildingPart	28,099
GroundSurface	10,422,993
RoofSurface	29,641,949
WallSurface	104,895,065
Energy ADE ThermalZone	6,454,421
Energy ADE UsageZone	6,454,421
Surface geometries	299,139,932

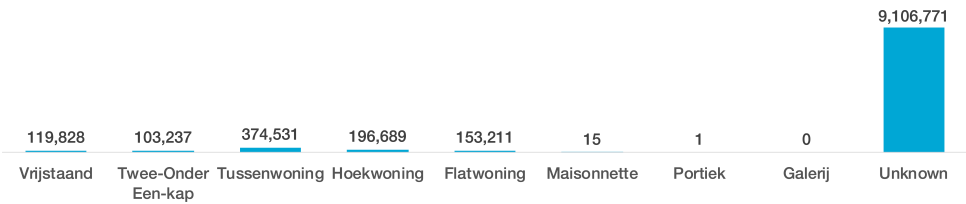


Figure 4.9: Buildings from 3DBAG classified by type

4

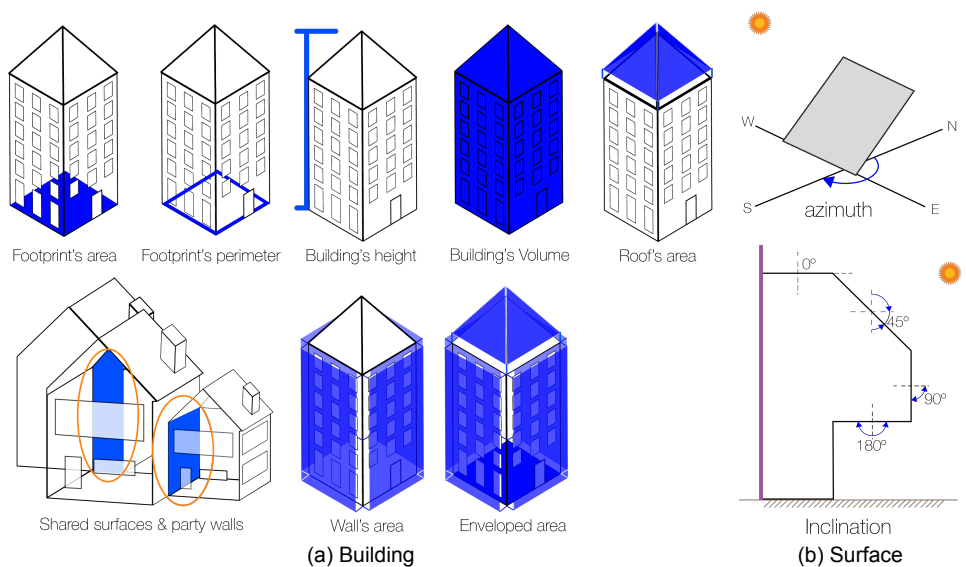


Figure 4.10: Graphical examples of the features extracted from the *CityObjects* geometries

4.2.3. DATA INDEXING

The EnergyBAG DB cannot be used as a single piece in UBEM due to its large size, as summarised in [table 4.6](#). Managing such a vast amount of geospatial data is a common challenge in the geospatial community, and it is geometrically taken care of by splitting the data into smaller regions, which makes handling the data more feasible. That is also the case of the 3DBAG, which is shared using a quadtree data structure to index its dataset (Samet, 1984), facilitating data sharing and processing. [Figure 4.11](#) shows the Netherlands covered by 3DBAG tiles.

Unlike the 3DBAG, I decided to use an indexing method that is not affected by the number of neighbours, and global indexes meet this criterion. These types of indexes have been studied since the 1960s.

The proposal of Sadourny *et al.* (1968) divide the globe into 20 spherical triangles that form a spherical icosahedron⁷ figure 4.12(A). Their goal is to represent the globe in nearly equal areas and shapes. Each of these faces is further subdivided into smaller triangles, as shown in figure 4.12(B).

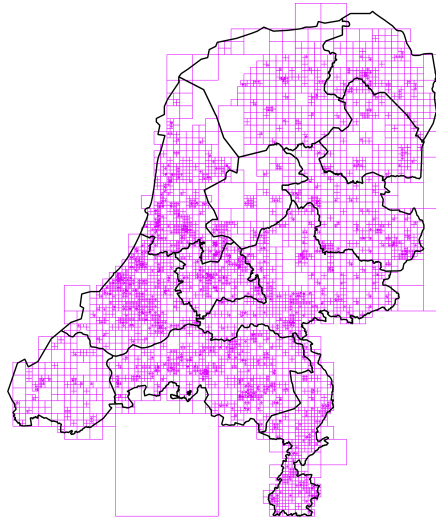
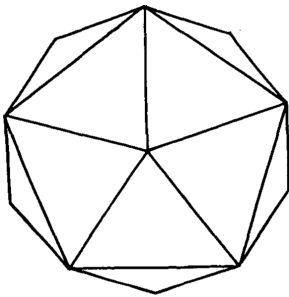
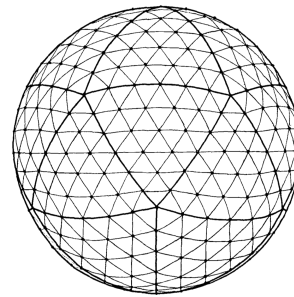


Figure 4.11: Provinces of the Netherlands overlaid with the boundary of the 3DBAG tiles in purple



(a) Icosahedron



(b) Representation of the icosahedron-hexagonal grid split into 6 equal arcs

Figure 4.12: Examples of an icosahedron (a) and an icosahedral-hexagonal grid (b) covering the sphere. From (Sadourny *et al.*, 1968).

⁷A polyhedron with 20 faces

A current method that follows a similar approach is the H3 index, which is a 2D hierarchical global index developed by Uber to index the planet into a hexagonal grid (Uber Technologies, 2018). This index supports sixteen resolutions, from level 0 to 15. In this method, the higher the level number, the finer the hexagon cell size. Furthermore, each finer resolution corresponds to one-seventh the area of its coarser resolution. However, a significant disadvantage of this algorithm is that finer cells are only partially contained within a parent cell. Nevertheless, additional reasons for the selection of this tiling method include:

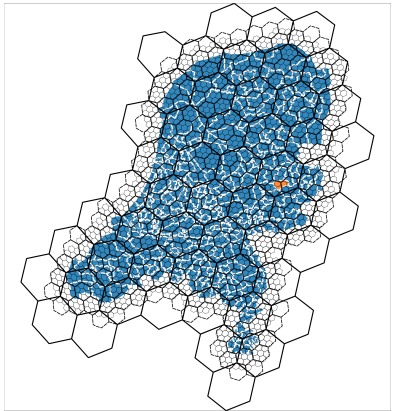
- Projection independence: It does not require cartographic projections to process the data, allowing its application in several locations.
- A uniform border-to-centre distance: The consistent distance from tile borders to cell centre points ensures reliability in horizon analyses, a critical factor for solar potential calculations (section 4.3.1).
- Consistent neighbouring distances: The distance between the centre point of one tile and its neighbours remains uniform, which is a feature not available in other tiling geometries such as square or triangular.

Therefore, the use of this hierarchical index allows a more balanced representation of the study area. Table 4.7 shows a summary of the indexes for the EnergyBAG data.

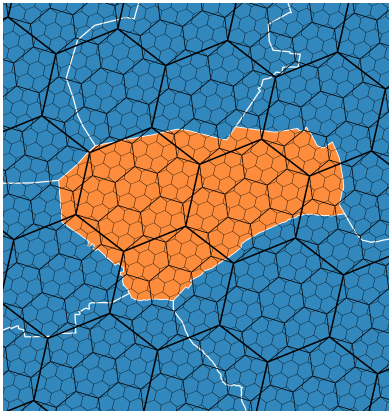
Table 4.7: Statistical summary of Building(Part)s per H3 tiling level

Level	Average edge Length (km)	Number of Hexagons	Number of Building(Part)s		
			Avg.	Min.	Max.
1	483.057	3	1.477.938	3.474.331	7.299.004
2	182.513	5	249	2.084.599	6.952.998
3	68.979	12	314	868.583	3.846.589
4	26.072	43	3	242.395	902.822
5	9.854	235	1	44.353	239.095
6	3.725	1.353	1	7.704	67.289
7	1.406	8.382	1	1.243	19.086
8	0.531	48.812	1	214	4.745
9	0.201	204.943	1	51	1.024
10	0.076	711.024	1	15	215

Figure 4.13 shows two examples of the Netherlands covered by three hierarchical levels of the H3 index represented by black lines in which the coarser the level, the darker the line. The whole Netherlands is shown in figure 4.13a, and the study area of the testbed is shown in figure 4.13b. To ensure the efficiency of the execution of the EnergyBAG BES, I computed the H3 tile per index level for all *Building(Part)s*. Finally, I stored this relation in the EnergyBAG DB.



(a) The Netherlands (in blue) covered by H3 indexes in levels 4,5,6



(b) Rijssen-Holten (in orange) covered by H3 indexes in levels 6,7,8

Figure 4.13: Example of geographical areas covered by three levels of H3 tiles from bigger to smaller

Figure 4.14 shows an excerpt of Rijssen 3DCM filtered by five (5) H3 tiles at level 7; the resulting dataset contains 15.777 *Building(Part)s* (red) corresponding to 240.007 *ThematicSurfaces*, 398 tiles of *Relief* (light green), and 25.358 *SolitaryVegetationObject* (dark green).

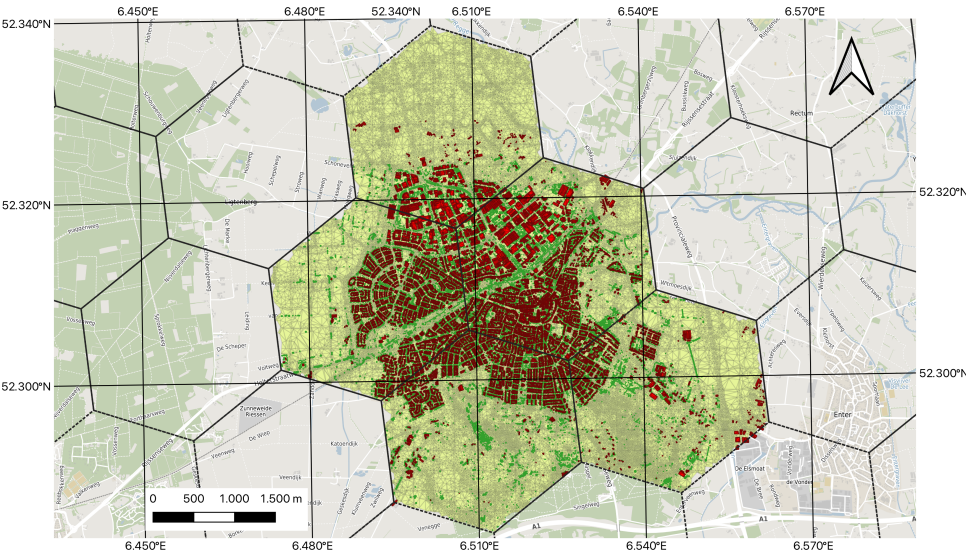


Figure 4.14: Excerpt of Rijssen 3DCM using H3 index tiles level 7, base map BGT OpenBasisKaart

To define the location of vegetation CityObjects, we followed the approach of Voortman (2021), which is based on the use of the AHN3 point cloud dataset to identify vegetated areas in Rijssen-Holten. The method filters vegetation points higher than 2m above ground level to exclude small shrubs and hedges. The remaining points are buffered by 50cm and dissolved to aggregate the green areas. From these regions, we extract a location every 5m and use the average height at each location to assign this parameter to the FME transformer. This is used to model the implicit geometry of trees, without incorporating semantic information such as species or age. As a result, we do not distinguish trees by leaf density or shading capacity across different weather seasons.

Voortman (2021) published the methodology implemented by the municipality of Rijssen-Holten for the identification of the vegetation areas in their jurisdiction using the AHN3 point cloud dataset Rijkswaterstaat (2024). We use Voortman work as the starting point for our trees modelling approach. Voortman's calculation is done using FME and removes all points from the ANH3 point cloud that were not both vegetation and their height difference compared to the ground level is higher than 2m so small shrubs and hedges are not included; The remaining points are buffered 50cm and dissolved. The result is a projection of the green volumes of the municipality in 2D. A final step in their process correspond to the clean up by cheking it against the high resolution aerial photo available by the Dutch Kadaster services Kadaster (2024d). We decided to use a simplified geometrical representation of trees based by using the *ImplicitGeometry*⁸ defined by the CityGML standard Gröger, T. Kolbe *et al.* (2012).

4.3. SOLUTIONS TO ENHANCE DATA

The first section focuses on solar analysis in urban areas with a comparison of solar simulation tools section 4.3.1.1 and the presentation of a solar simulation tool based on s3DCM section 4.3.1.2. These projects are followed by an exploration of the use of machine learning to classify buildings into predefined archetypes, which could address the lack of openly available building physics data for the Dutch building stock section 4.3.2. Sections 4.3.1.2 and 4.3.2 correspond to MSc theses in Geomatics at TU Delft. Finally, section 4.3.3 summarises the findings from an MSc Thesis in Geomatics to infer the number of storeys in the Dutch building stock.

⁸"Type for the implicit representation of a geometry. An implicit geometry is a geometric object, where the shape is stored only once as a prototypical geometry, e.g. a tree or other vegetation object, a traffic light or a traffic sign. This prototypic geometry object is re-used or referenced many times, wherever the corresponding feature occurs in the 3D city model." Gröger, T. Kolbe *et al.* (2012, p. 159)

4.3.1. SOLAR ANALYSIS

The scope is to report on research work that has developed with topics that are relevant for my thesis: How to enhance existing data, e.g., solar irradiance values and how to scope of this section is the development of a computational tool that performs accurate simulations of solar radiation on building surfaces. From the methodological point of view, the work is divided into two phases:

1. Evaluation of existing tools: An assessment of current solar irradiation tools, focusing on data requirements, usability, and output accuracy.
2. Tool development and implementation: The design and implementation of a simulation tool that calculates solar irradiation using semantic 3D city models (s3DCM) in CityGML format.

4.3.1.1. ANALYSIS OF SOLAR SIMULATION TOOLS⁹

Obtaining solar irradiance measurements on a building surface presents several challenges due to its complexity. This task is possible at an urban scale using remote sensing. However, satellites capture data mostly from roofs. Hence, solar irradiation simulation tools offer a possible alternative to overcome the lack of solar irradiance data over the building's facade at the city scale. As a first step, we have performed a qualitative and quantitative analysis of GIS-based software simulation tools for solar potential via solar irradiation on roofs to evaluate the quality of these simulation tools to overcome the lack of solar irradiation on the boundary surface of buildings.

A total of seven simulation tools have been evaluated. We have compared them in terms of input and output data, as well as their accuracy, based on local weather data as ground truth. The tools, listed in alphabetical order, are ArcGIS Pro v3.1 (Esri, 2024), CitySim v2023.06 (Mutani *et al.*, 2018), GRASS GIS v7.8.7 (Hofierka *et al.*, 2007), Ladybug v1.6.33 (Sadeghipour Roudsari and Pak, 2013), SAGA GIS v8.5.1 (Conrad, 2010), SimStadt v0.10.0 snapshot 20230307 (Duminil *et al.*, 2022), and Urban Multi-scale Environmental Predictor v4.0.2 (UMEP) (Lindberg *et al.*, 2023). Due to the variety of the simulation tools used, the input datasets for this research include raster and vector data. Our method, shown in figure 4.15, is designed to analyse and compare the GIS software tools in a way as standardised as possible despite their different characteristics or data requirements.

Our case study is located at the weather station of Heino in the municipality of Raalte in the Netherlands. Table 4.8 shows a summary of the input datasets used

⁹This section presents a shorter version of publications "Comparative analysis of geospatial tools for solar simulation" (León-Sánchez, Giannelli *et al.*, 2025) in Transactions in GIS vol. 29, no. 1 (2025). Data, results and analyses are shown only the study area in the Netherlands. "Comparison and Evaluation of Different GIS Software Tools to Estimate Solar Irradiation". ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 5, no. 4 (2022): 275–82. (Giannelli *et al.*, 2022)

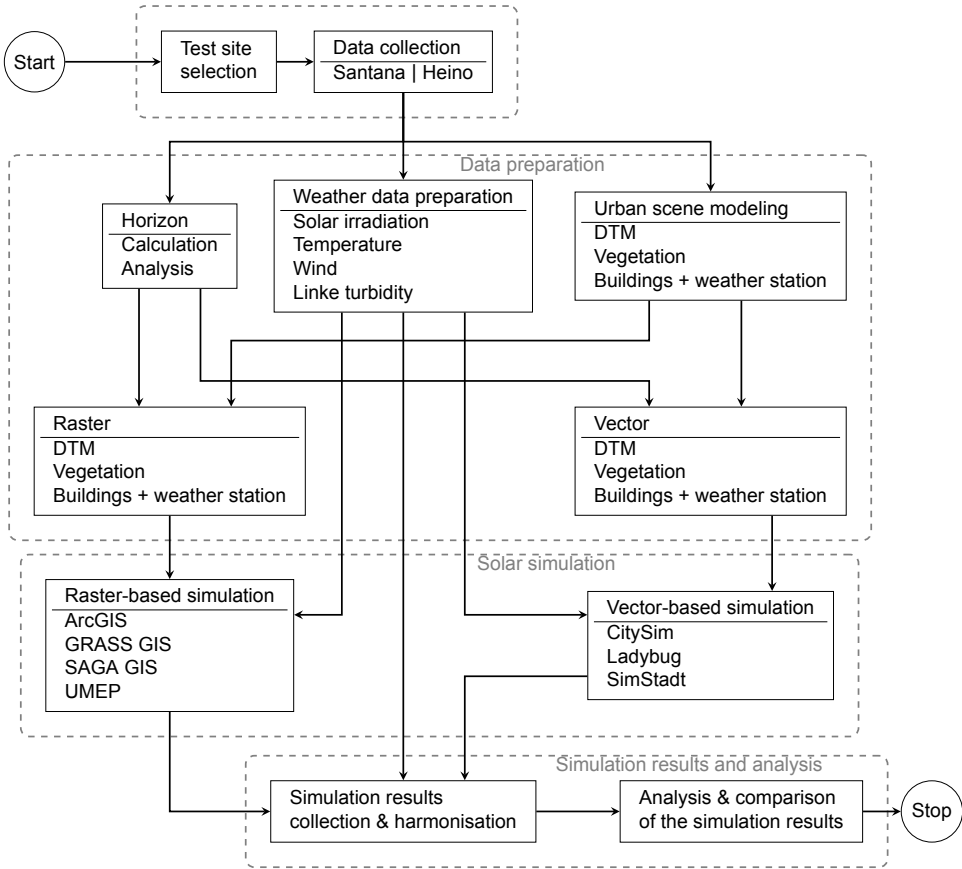


Figure 4.15: General workflow of the comparative method described in this article

in our research. Incorporating Linke Turbidity data influences the computation of the solar potential. Both GRASS GIS and SAGA GIS employ clear-sky models for their computations. Buildings in the study area are extracted from the 3DBAG. To model the vegetation, we represent trees as CityGML *implicitGeometry*. Vegetation is used as shadowing objects.

Table 4.8: Summary of the input data sets in our research

Raalte, Netherlands	
Geospatial	3DBAG: 3D Buildings Dutch elevation database (AHN3): Aerial point cloud Raster-based DSM: <ul style="list-style-type: none">• AHN3: 0.5m, 5m
Weather	Royal Netherlands Meteorological Institute (KNMI) Heino weather station: Typical yearly values of global, direct, diffuse solar irradiation, ambient and ground temperature, wind speed, cloud coverage, pressure, and rainfall.

In [figure 4.16](#), we present the results of all simulation tools applied to the case study. From a user's perspective, superimposing the results from all simulation tools in a single graph represents the magnitude of differences that can be expected when using one software tool over another. This visual comparison illustrates the variations in simulation outputs for the same dataset.

Although all simulation tools tend to follow the same pattern as the ground data—which is expected since the ground data serves as the source for the computations—it is meaningful to mention that raster-based simulation tools implement a clear-sky model, which means that the simulated results are expected to differ from the weather station data used as ground truth, generally producing higher values. Nevertheless, this was only the case for some simulation tools; only SAGA GIS and GRASS GIS produced higher solar irradiance values than the ground truth data.

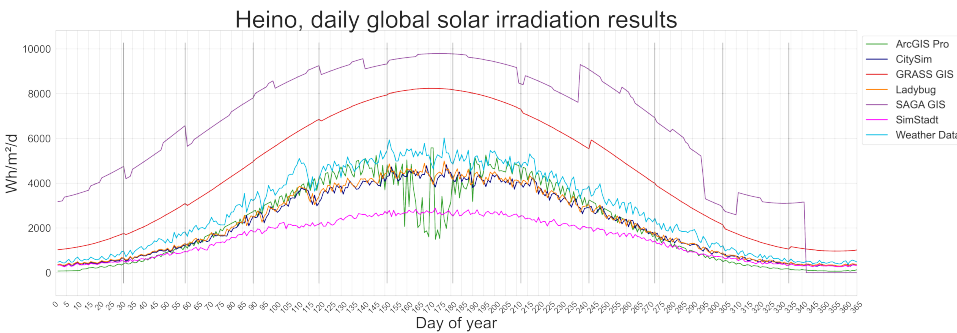


Figure 4.16: Time series of the simulation results and the weather data for Heino

The differences between the various software tools are likely due to the specific solar radiation models they implement. Some tools, particularly ArcGIS Pro, act as a sort of black box because the details of their models are not publicly disclosed due to commercial restrictions. This absence of transparency limits the ability to tune further or adjust the models to match ground truth data better, making it challenging to understand the reasons for discrepancies in the simulation results.

It is relevant to remember that direct comparison of results coming from between raster-based and vector-based models can be misleading since the former adopts a clear-sky model. In contrast, the latter does not, although, in both cases, the results are expressed in $\text{Wh/m}^2/\text{d}$.

[Table 4.9](#) and [figure 4.17](#) contain the yearly aggregated values of the global irradiation. Ladybug produces the closest value, with a difference of -19.75% , and the lowest RMSE of 0.63. The former is followed by CitySim, with -21.88% and 0.726, respectively. On the other hand, SAGA GIS produces the furthest results, with a difference of 131.58% and a RMSE of 3.972, followed

by SimStadt, with a -48.82% and 1.614 , respectively.

Although clear-sky models should consistently deliver higher values compared to ground truth, it is interesting to note that this is not the case for the results presented in our research. The output values of ArcGIS Pro output are consistently lower, while the ones from GRASS GIS have the opposite behaviour. Vector-based simulation results offer better results than raster-based ones. Ladybug and CitySim can produce accurate results, although they require more extended data preparation and computational time. However, the results indicate the importance of adding ground-truth weather data to calibrate the simulation results to achieve greater accuracy.

4

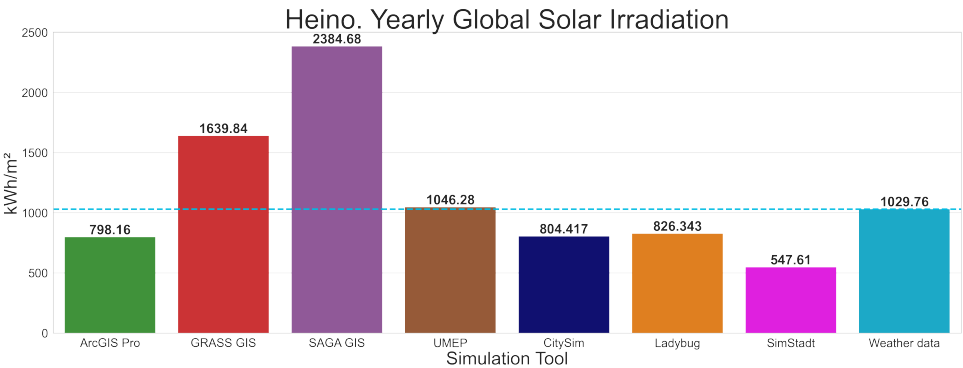


Figure 4.17: Yearly global irradiation values obtained from the simulation results and the ground truth at Heino. The cyan dotted line indicates the reference yearly value

Table 4.9: Heino, yearly global irradiation values from the simulation results compared to the ground truth [$kWh/m^2/a$]

	Raster-based				Vector-based		
	ArcGIS Pro	GRASS GIS	SAGA GIS	UMEP	CitySim	Ladybug	SimStadt
Total	798.16	1,639.84	2,384.68	1,046.28	804.42	826.34	547.61
Diff.	-22.49%	59.25%	131.58%	37.30%	-21.88%	-19.75%	-46.82%
RMSE	0.904	1.875	3.972		0.726	0.63	1.614

A comparison overview of all parameters we considered for our qualitative analyses is shown in [table 4.10](#). We evaluate not only the accuracy of the simulation's output. Except for SimStadt, a terrain model is fundamental input data for all simulation tools.

Table 4.10: Comparison of the tested simulation tools for solar irradiation. Running times were calculated using a desktop computer with the following hardware specifications: Windows 11, Intel i7-9700K CPU, 64 GB RAM, AMD Radeon VII GPU and NVMe PCIe M.2 2280 SSD

	Raster-based				Vector-based		
	ArcGIS Pro	GRASS GIS	SAGA GIS	UMEP	CitySim	Ladybug	SimStadt
	Raster-based				Vector-based		
	ArcGIS Pro	GRASS GIS	SAGA GIS	UMEP	CitySim	Ladybug	SimStadt
License	Commercial	FOSS	FOSS	FOSS	FOSS	FOSS but Rhinoceros 3D/Grasshopper	FOSS
Minimum input data requirements	Raster-based DSM	Raster-based DSM	Raster-based DSM	Raster-based DSM, Weather data	Vector-based 3D scene (Buildings, relief, vegetation), weather data	Vector geometries as Brep / Mesh, weather	Vector-based 3D scene (Buildings in CityGML LoD1/2), weather
Optional input files	N/A	Slope, aspect, Linke turbidity, albedo maps	Sky View Factor, Water Vapour Pressure, Linke turbidity	Vegetation Canopy, Vegetation Trunk zone albedo value	Horizon file	N/A	N/A
Interaction	GUI / Python	GUI / Python / shell	GUI/shell	GUI / Python	GUI/shell	GUI / Python	GUI/shell
Urban features	All features represented in the DSM	All features represented in the DSM	All features represented in the DSM	Buildings, trees and ground represented in DSM	Buildings, Relief, Vegetation	Buildings, Relief, Vegetation	Buildings
Santana. Timing (HH:MM)	03:18	08:40	00:12	00:11	16:44	00:26	00:06
Heino. Timing (HH:MM)	31:34	17:19	11:51	30:24	268:45	00:18	00:11
Results: Type	2.5D surfaces	2.5D surfaces	2.5D surfaces	2.5D surfaces	3D surfaces	3D surfaces	3D surfaces
Output	Raster file	Raster file	Raster file	Raster file	TSV file	Data tree	out file

Continued on next page

Table 4.10 – continued from previous page

	Raster-based				Vector-based		
	ArcGIS Pro	GRASS GIS	SAGA GIS	UMEP	CitySim	Ladybug	SimStadt
Minimum temporal resolution	Hourly	Second / daily	Hourly	Yearly	Hourly	Hourly	Hourly

4.3.1.1.1 Lessons learnt Here are the lessons learnt from this research:

- Raster-based simulation tools are preferred when there are data access restrictions since they only require a Digital Surface Model.
- Vector-based simulation tools offer better results than raster-based ones but have more complex data requirements. Furthermore, this type of data is less commonly available, which increases data preparation time.
- There is no single tool that consistently produces the most accurate results across all locations. Nevertheless, all simulation tools follow the same pattern as the ground truth.
- In our study areas, Ladybug, CitySim, UMEP, and ArcGIS Pro provide good accuracy against the ground truth. However, simulation results should be calibrated to achieve better accuracy.

4.3.1.2. SOLAR IRRADIANCE COMPUTATION¹⁰

Section 4.3.1.1 explains the data requirements to perform a solar simulation. This knowledge is the starting point for the design of a workflow to compute solar irradiance at the city level at hourly temporal resolution using s3DCM. We use the 3DBAG LoD2.2 as the input dataset, although the workflow supports any CityJSON-encoded dataset. Therefore, we avoid assumptions about urban morphology.

4.3.1.2.1 Method

Our proposed method for the computation of the solar irradiance at a given location is shown in figure 4.18. The workflow starts by extracting the centre coordinates of the input s3DCM. These coordinates are required to calculate the position of the sun over a typical year. To reduce the computational load, we keep only the records when the position of the sun is above the horizon since

¹⁰This section consolidates the MSc thesis in Geomatics at TU Delft titled “High-resolution, large-scale, and fast calculation of solar irradiance with 3D City Models” by Xu (2024), partial results were published in Recent Advances in 3D Geoinformation Science, 31-47 (2024) (Xu, León-Sánchez *et al.*, 2024b) and in International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 167-174 (2024) (Xu, León-Sánchez *et al.*, 2024a)

they will be used in the following steps. The four key components of our method are:

1. **Point grid generation:** Generate sampling points on *CityObject* surfaces via recursive triangle splitting. The split level depends on target density (e.g., $16\text{m}^2 = \text{one split}$, $18\text{m}^2 = \text{two splits}$). Non-uniform sampling occurs due to area-based split logic.
2. **Viewshed calculation:** For each grid point, trace a hemisphere of rays to compute the Sky View Factor (SVF). The results are stored as (N, M) vectors for reuse in reflective calculations. Where N represents the azimuth steps and M represent the elevation steps.
3. **Shadow calculation:** Compute binary shadow masks using sun position rays. Optimized with Bounding Volume Hierarchy (BVH) to reduce complexity from $\mathcal{O}(NM)$ to $\mathcal{O}(N \log M)$.
4. **Solar irradiance calculation:** Solar irradiance is split into several elements:

- **Direct beam:** $I_{S,dir,\beta} = M_{shadow} \cdot DNI \cdot \cos \delta$
Here, M_{shadow} represents the binary shadow mask, and δ represents the angle between the surface normal and the solar vector.
- **Sky diffuse:** It uses an isotropic model using $I_{S,diff,\beta} = DHI \cdot SVF$ with Ground View Factor–GVF as $GVF = 1 - SVF$
- **Reflective irradiance:** Our method uses a cached semantic viewsheds and a voxelized irradiance field to reduce the computational complexity from exponential to linear
 - **Semantic scene voxelization:** We aggregate the previously calculated direct beam and sky diffuse solar irradiance through voxelization to create a light field within the scene. We treat the reflective irradiance from urban objects as irradiance emitted from the voxels.
 - **Iterative irradiance propagation:** Using the voxelized irradiance field and the constructed viewshed map for each grid point, we begin simulating reflective solar irradiance. For each grid point, we reference all pixels in its stored viewshed map, where the voxel IDs and incident angles are recorded, allowing us to identify the three visible faces of each voxel.

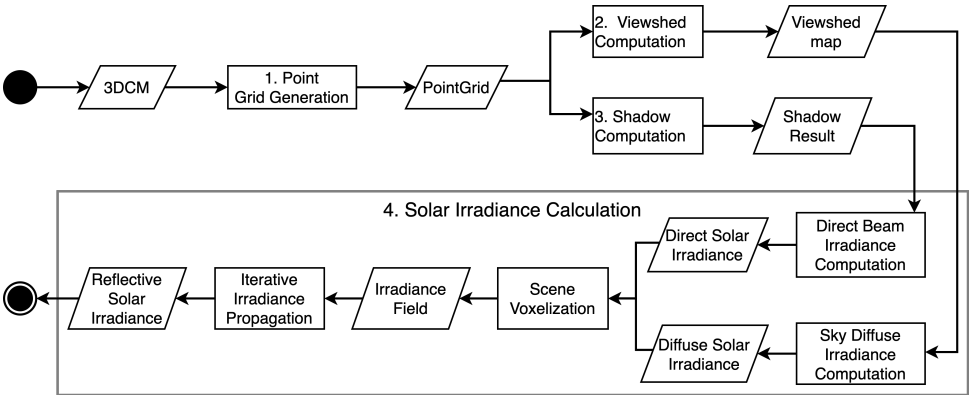


Figure 4.18: Implemented Solar irradiance workflow. Based on (Xu, 2024)

4

4.3.1.2.2 Results

This study uses two ground truth datasets. The first one corresponds to Heino, while the second dataset corresponds to the data of two monitoring stations (S1 and S2) with solar irradiance measurements collected by TU Delft’s Photovoltaic Materials and Devices (PVMD) group on a campus building Andres *et al.* (2023) and Calcabrini (2023). Simulations are conducted for four dates—21 March, 21 June, 22 September, and 22 December (equinoxes and solstices)—using hourly KNMI weather data. Scenarios with 0, 1, and 2 light bounces are evaluated alongside two simplified baseline cases.

A comparison of results from CitySim, Ladybug, and SimStadt versus our tool for 2 February is shown in figure 4.19. The line plot colours for the simulation tools match those used in section 4.3.1.1. Our results indicate a general underestimation of our solution compared to the weather data and the other simulation tools. Furthermore, the difference is higher from 10:30 am to 2:30 pm.

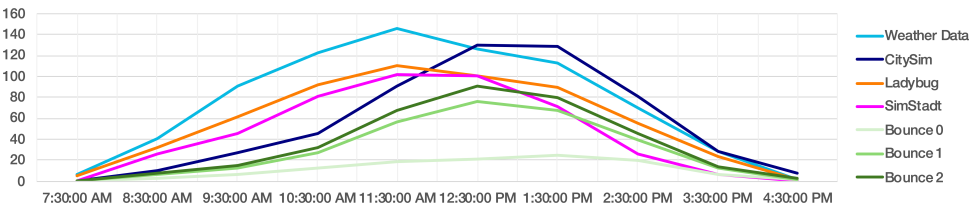


Figure 4.19: Heino weather station at 2 February

The results of the PVMD stations on 2 February 2020 are shown in figures 4.20 and 4.21. For sensor S1 (figure 4.20), the computed values closely follow the

observed data pattern with minimal differences. However, this consistency is not visible in sensor S2 (figure 4.21), which shows significant discrepancies that grow as the number of bounces increases. This is a sign that while the direct beam follows the rules, the bouncing part needs further investigation due to some possible errors.

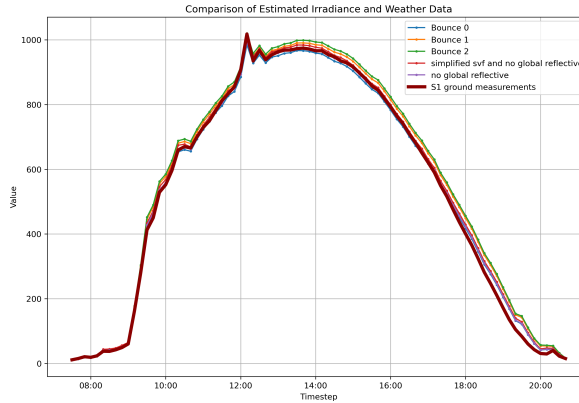


Figure 4.20: Line plot of the results obtained in PVMD monitoring station **S1** on 21 August 2020

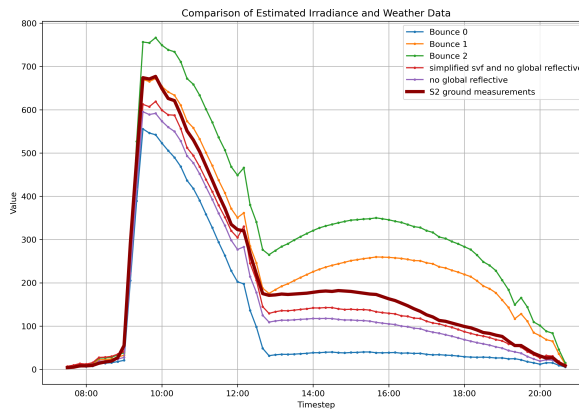


Figure 4.21: Line plot of the results obtained in PVMD monitoring station **S2** on 21 August 2020

The computation times for solar irradiance simulation in the study areas are presented in table 4.11. The total computational time varies based on the number of the simulation time steps. These values correspond to a hemisphere sampling resolution of 5 degrees and around a million of sample points. Due to the implemented method, the computation time increases linearly according to the time steps and the number of sample points. Also, the shadowing and direct

solar irradiance calculation requires less than 1% of the computational time while the remaining is spent on the reflective solar irradiance calculation.

Date	Epochs	Shadow computation	SVF computation	Direct + Diffuse	Reflective	Total	Reflective ratio
2024/08/21	80	0.62	51.03	1.47	12319.02	12372.14	99.57
2024/02/01	45	0.22	26.76	0.68	8017.02	8044.68	99.66

Table 4.11: Computation time for PVMD monitoring station with the proposed method, all values are in seconds

4

Additionally, we conducted further testing of our model on a larger area in Rijssen-Holten to evaluate its scalability and robustness [figure 4.22](#). We generated a CityJSON file containing buildings, terrain, and vegetation, covering an area of a $\sim 35km^2$. We use the H3 tiling index discussed in [section 4.2.3](#) to split the study area. Details of the study area are provided in [table 4.12](#).

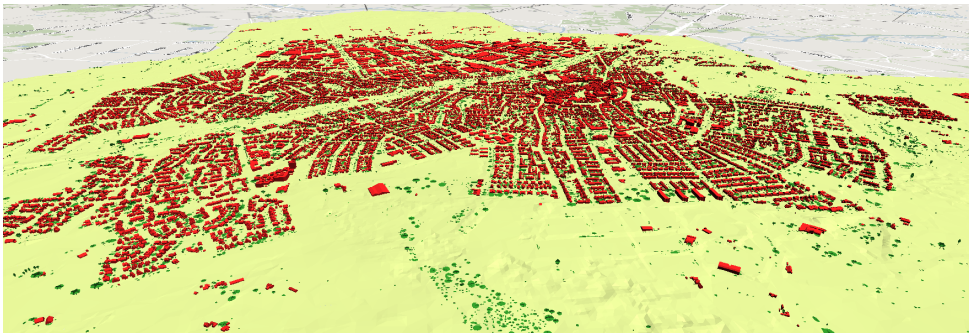


Figure 4.22: Rijssen input dataset Sample

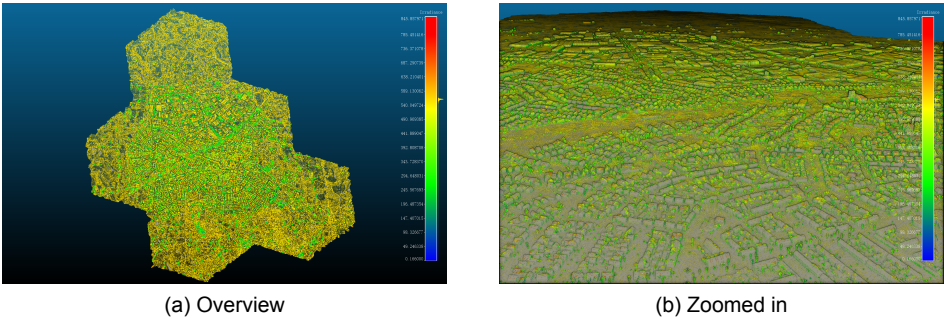


Figure 4.23: General visualization solar irradiance results in Rijssen

Attribute	Value
Geographic center	6.52E, 52.31N
H3 indexes	871f16cf1ffffff, 871f16cf5ffffff, 871f16cc4ffffff, 871f16ce6ffffff, 871f16ce2ffffff
Spatial extent	6100 * 6200 * 50
<i>Buildings</i>	15,744
<i>ThematicSurfaces</i>	717,747
<i>RoofSurfaces</i>	49827
<i>WallSurfaces</i>	174436
Trees objects and Surfaces	25,358/608,592
Terrain objects and Surfaces	398/100,091
Total Surfaces	1426430
Number of sample points	6457,806

Table 4.12: Dataset details for four different cities

Without simulating reflective solar irradiance, the entire simulation would take less than 7 minutes, a summary of the computational time is shown in [table 4.13](#).

Epochs	Shadow computation	SVF computation	Direct + Diffuse	Reflective	Total	Reflective ratio
54	1.02	151.44	5.97	42941.12	43093.58	99.65

Table 4.13: Computation time for Rijssen with the proposed method, all values are in seconds

4.3.1.2.3 Lessons learnt

Here are the lessons learnt from this research:

- The incorporation of BVH in our method removes the need for geometrical assumptions. BVH handles complex urban landscapes by integrating digital terrain models. Moreover, it allows precise shadow calculations on minor features, such as corners
- Our method requires further refinement. The calculation results for the direct element of the solar irradiance are sufficiently in line with the ground truth. However, the results obtained from the first and second light bounces are worst since they increase at each bounce which should not be the case.
- A more detailed representation of the building stock in terms of physical properties of the *CityObjects*, directly influences the calculation of the solar irradiance in buildings, as these are affected by the albedo of the surrounding environment.

- The current implementation requires refactoring in the bouncing section to support large study areas. We planned to compute the solar irradiation of buildings in Rijssen-Holten, which covers an area of approximately 20km by 20km. However, this was not feasible because the model would require approximately 1.5TB of data storage and about 400 hours of computation time on personal computers.

4.3.2. BUILDING TYPE INFERRING USING MACHINE LEARNING ALGORITHMS¹¹

This section summarises the MSc thesis in Geomatics, "Inferring the residential building type from 3DBAG". This research was carried out to evaluate the use of 3DBAG as an input dataset for the classification of the Dutch building stock into the existing building archetypes defined by RVO (2023). The classification of the Dutch building stock is still an issue because this information is not always available. Poon (2024) combines feature engineering (attributes extraction and selection) and several machine learning algorithms to infer the building type classes for the Dutch building stock, according to TABULA. The proposed 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). SVM is effective in high-dimensional spaces and is valuable when dealing with many attributes¹². Furthermore, these algorithms do not require heavy hyperparameter tuning. Figure 4.24 shows in a general way the implemented method for inferring the building types from the 3DBAG using machine learning.

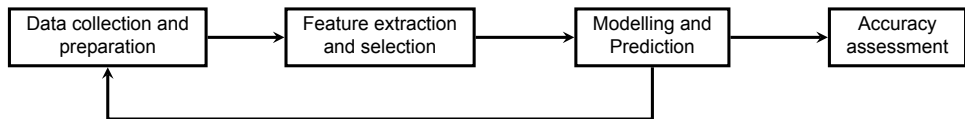


Figure 4.24: Flowchart of the implemented method

The first case study is based on the testbed section 4.1. We evaluate the extracted attributes of buildings from the 3DBAG to eliminate redundancy while keeping an accurate prediction. 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 in several case studies. The attribute of the building type available in the testbed is used as ground truth since the building type data was manually collected, requiring no further data preparation. However, there is a severe class imbalance, with four classes representing each

¹¹This section has been published in the MSc Thesis in Geomatics at TU Delft titled "Inferring the residential building type from 3DBAG" by Poon (2024).

¹²In data science and artificial intelligence, the correct term is "feature". However, I use in this PhD the term attribute to avoid confusion with the term "feature" from the geospatial point of view

of them less than 1% of the dataset: Flatwoning 0.38%, Maisonnette 0.02%, Portiek 0.05% and Galerij 0.03%.

Table 4.14: Rijssen-Holten dataset classification by building type

Building type	count	percentage
Vrijstaand (VW)	12,746	48.99%
Twee-Onder-Een-Kap (TOEK)	6,414	24.65%
Tussenwoning (TW)	3,738	14.37%
Hoekwoning (HW)	2,997	11.52%
Flatwoning (FW)	99	0.38%
Maisonnette (MW)	5	0.02%
Portiek (PW)	12	0.05%
Galerij (GW)	7	0.03%

Due to this imbalance, Poon chose seven additional case studies. All of these cases consider ep-online data as ground truth. However, the dataset requires further preprocessing (data cleaning). The second case study corresponds to Delft. Delft's housing characteristics differ from with Rijssen-Holten, with 33% of buildings containing one dwelling and 67% containing more (AlleCijfers.nl, 2024). Although Delft's class distribution is also imbalanced, minority classes are better represented (table 4.15).

Rijssen-Holten's and Delft's case studies correspond to urban areas delimited by the municipalities. However, both datasets present a high imbalance with Maisonnette, Galerij, and Portiek. Therefore, additional six case studies focus on specific building types were chosen. Case Study 3 at Duivendrecht/Venserpolder focuses on Flatwoning; Case Study 4 at Bijlmer-Oost focuses on Galerij; Case Study 5 at Borneo-Sporenburg focuses on Maisonnette; Case Study 6 at Laakkwartier focuses on Portiek; Case Study 7 at Oud-Diemen/Steigereiland focuses on Twee-Onder-Een-Kap and Terrace houses (Tussenwoning, Hoekwoning); and Case Study 8 at Laren focuses on Vrijstaand. Table 4.15 summarises the case studies; it is fundamental to point out the discrepancy in the classification of buildings.

Table 4.15: Building type classification datasets

Case Study	VW	TOEK	TW	HW	FW	GW	PW	MW
Delft	253	552	5,999	1,437	1,476	63	118	681
3	9	5	621	196	420	17	3	24
4	8	1	331	64	38	12	-	9
5	4	19	376	28	236	-	4	69
6	4	-	86	5	649	5	6	96
7	88	69	287	103	49	-	-	6
8	163	25	2	2	7	-	-	-

Two methods are used for attributes selection for each of the ML algorithms: ANOVA-F and Mutual Information (MI) for SVM and impurity-based and permutation-based methods for Random Forest (RF). Using these approaches, he applied four models to each municipality-based case study based on the ML algorithm and the attribute selection method. We consider Rijssen-Holten and Delft as the primary case studies because their sampling is not biased by balancing the input dataset.

4.3.2.1. RESULTS

A summary of the performance of the training models of Rijssen-Holten and Delft respectively are presented in [tables 4.16](#) and [4.17](#). In the case of Rijssen-Holten, the accuracy of the models is higher than 91%. However, that is not the case with balanced accuracies. The accuracy of Delft models is considerably lower, although the RF models have higher balanced accuracies.

Table 4.16: Summary of Rijssen-Holten case study

Model	Tuning time (s)	Training time (s)	Accuracy	Balanced accuracy
SVM ANOVA-F	25.2	1.42	0.918	0.583
SVM MI	20.43	0.35	0.917	0.571
RF impurity	426.31	2.87	0.911	0.741
RF permutation	344.27	2.48	0.918	0.652

Table 4.17: Summary of Delft case study

Model	Tuning time (s)	Training time (s)	Accuracy	Balanced accuracy
SVM ANOVA-F	112.9	7.04	0.672	0.395
SVM MI	240.39	8.75	0.389	0.330
RF impurity	996.91	14.16	0.711	0.737
RF permutation	1,105.94	17.31	0.733	0.737

[Table 4.18](#) shows the accuracy of the models. The first row shows the values obtained using 20% of the input datasets (80% used for training). The second row corresponds to the results when using the entire input dataset in the other case study. The class distribution imbalance influences the models' low accuracy.

Table 4.18: Overview of Model Performances on Case Studies 1 and 2 Applied to Various Datasets

Case Study	Building Type	Case Study 1 Models				Case Study 2 Models			
		SVM ANOVA-F	SVM MI	RF Impurity	RF Perm.	SVM ANOVA-F	SVM MI	RF Impurity	RF Perm.
Test dataset 20%	Accuracy	91.8%	91.7%	91.1%	91.8%	67.2%	38.9%	71.1%	73.3%
	Balanced accuracy	58.3%	57.1%	74.1%	65.2%	39.5%	33.0%	73.7%	73.7%
Other dataset	Accuracy	55.3%	55.6%	56.6%	56.1%	60.3%	38.3%	81.2%	84.0%
	Balanced accuracy	42.0%	40.4%	40.4%	44.7%	47.6%	23.8%	58.0%	62.5%
3	Accuracy	75.8%	79.5%	87.9%	82.1%	85.4%	43.7%	80.2%	90.4%
	Balanced accuracy	49.2%	50.6%	53.9%	51.6%	36.7%	24.9%	74.9%	66.7%
4	Accuracy	85.5%	86.0%	87.7%	87.3%	80.3%	30.0%	90.3%	89.4%
	Balanced accuracy	59.6%	60.4%	64.5%	60.5%	32.3%	11.3%	78.2%	81.5%
5	Accuracy	60.1%	59.9%	63.5%	61.1%	75.0%	53.4%	67.5%	69.8%
	Balanced accuracy	54.7%	49.6%	62.3%	65.5%	39.6%	26.0%	60.5%	60.6%
6	Accuracy	17.0%	18.1%	26.7%	63.1%	82.1%	34.0%	68.5%	71.2%
	Balanced accuracy	42.8%	43.0%	44.8%	51.6%	40.2%	33.0%	68.6%	62.6%
7	Accuracy	89.4%	83.2%	89.2%	89.0%	60.8%	27.9%	80.2%	86.0%
	Balanced accuracy	71.8%	62.8%	67.7%	67.4%	35.4%	23.9%	74.9%	78.9%
8	Accuracy	98.0%	98.5%	81.9%	98.5%	86.9%	83.4%	93.0%	93.0%
	Balanced accuracy	90.6%	94.2%	66.6%	91.4%	43.4%	28.6%	72.0%	64.2%

RF performs better than SVM in classifying building types despite case study 8, which is flatwoning-focused. However, the lower accuracy values in case study 6 (Portiek-focused) emphasise the importance of reliable ground truth. This case dataset has no manual correction from the data extracted from ep-online, leading to mislabelling issues. Energy labelling of dwellings in the Netherlands is not an automatic or systematic process but a manual task initiated by specific situations, such as when properties are rented or sold. The specialist who computes the energy labelling must assign a building type classification to the construction of the dwelling. However, the criteria for classifying buildings vary depending on the agent in charge. These multiple criteria characterise the challenge of the classification of a building type. Therefore also having an impact on the reliability of the training data. From [figure 4.25](#) it can be seen that, in general, buildings can have multiple classes assigned .

[Figure 4.25a](#) shows two Flatwoning buildings with the same situation; both of them can be classified as Galerij due to the galleries in the facade or apartment. However, the former class is a specification of the latter two. [Figure 4.25b](#) shows several examples of Flatwoning located over Schieweg in Rotterdam. Although no differences can be observed from the facade, dwellings are labelled

Flatwoning, Portiek, and Maisonnette.

Figure 4.25c shows two examples in Amerongenstraat in The Hague. The building in the corner with Tenswoudelaan (upper right part of the 2D map) is labelled Twee-Onder-Een-Kap and Hoekwoning. Both labels are logical since the building shares one thermal envelope with another; however, both buildings are in a row, so they should actually be considered Terrace Houses.

These multiple labelling raise challenges when performing accuracy assessments, pointing out the importance of having a precise and clean dataset as ground truth, which we do not have. Based on the results obtained and the hit-and-miss analysis, collecting additional attributes per building will increase the correct classification of the input data: number of storeys per vbo for Maisonnette, open porch presence for Portiek, or presence of galleries in the facade for Galerij. Only the number of storeys per vbo is available for these three attributes in the BAG-plus (Gemeente Amsterdam, 2022).

The results indicate a high potential in using ML methods to overcome this data gap with an accuracy of $> 51.6\%$. However, the results are not accurate enough to be used to classify complex buildings as Galerij, Maisonnette, or PK. This misclassification happens due to the unclear classification rules of the buildings like the building in The Hague figure 4.25c.



Figure 4.25: Examples of buildings assigned with multiple building type classes

4.3.2.2. LESSONS LEARNT

The lessons learnt from this project are:

- The topic is relevant since there is no open data available at the Dutch national level about the building types.
- Attributes such as adjacency to other buildings, building width, and LoD2.2 volume have a strong correlation with Dutch residential building types. These attributes can be extracted from s3DCM. However, other critical attributes—such as the number of storeys and the presence of open porches or galleries, which directly relate to the definition of specific building types—are available as open data, nor can they be extracted from s3DCM.
- There is significant variation in the accuracy of the implemented models. These also exhibit different performances in the selected case studies when distinguishing between single-vbo and multi-vbos buildings. Furthermore, in most cases, Random Forest models outperform Support Vector Machine models in both suitability and accuracy.
- The classification rules of buildings are ambiguous, and this leads to misclassification by both humans and AI models. For example, in the results shown in [figure 3.11](#), none of the buildings that have a vbo classified as Galerij are uniquely classified as such ([figure 4.9](#)). This ambiguity indicates a significant challenge and adds complexity to the process.
- There are limitations on the ground truth. All AI methods rely heavily on labelled data for training, so discrepancies in classification rules across several datasets impact negatively the model performance.
- The uncertainty found in the results does not allow us to choose a single model that is accurate enough to be applied to the whole building stock of the Netherlands. Therefore, this is an unsolved data gap for the use of CityGML-based s3DCM for UBEM.

Nevertheless, in [section 5.5.1.3](#) present a workaround to reduce the impact of this lack of data.

4.3.3. CALCULATION OF THE NUMBER OF STOREYS PER BUILDING

I briefly summarise the research done by Roy *et al.* (2023), which investigates inferring the number of storeys in buildings in the Netherlands using CityJSON-based s3DCM data due to the importance of the results for my PhD thesis. Their research uses machine learning models trained on features derived from the BAG, 3DBAG, census data from the Dutch Central Bureau of Statistics (CBS), and training datasets from municipalities, including Amsterdam (BAG-Plus), Rotterdam, and Rijssen-Holten. The prediction models used were Random Forest (RF), Gradient Boosting (GB), and Support Vector Regression (SVR).

The study categorised 25 attributes into four groups:

- Cadastral: Construction year, building function, net internal area, number of units, and building type
- 2D geometric: Attributes derived from building footprints, such as area, perimeter, number of vertices, and neighbouring buildings
- 3D geometric: Attributes derived from 3D models, including building height (at various percentiles), roof shape, ridge-eave height difference, roof and wall areas, and building volume
- Census-based: Population density, percentage of multi-household buildings, and the average number of cafés within a specific radius

The Gradient Boosting (GB) model outperformed the common geometric approach of dividing building height by a predefined storey height. Building height was identified as the most influential predictor, while the study found no significant improvement in using LoD2 over LoD1 building geometries, suggesting that complex 3D models are unnecessary for storey predictions.

However, the model's accuracy decreases significantly for buildings with more than five storeys, from 94.5% for buildings up to five storeys to 52.3%. Another limitation of this model is the “half-floor,” which is commonly found for example in old buildings in historical city centre Amsterdam. Following its publication, the model has been refined. Since February 2024, its output has been incorporated as an additional attribute in the 3DBAG, keeping the limitation to reliable data outcomes for buildings with up to five storeys. Therefore, I will introduce in [section 5.5.1.3](#) a workaround to overcome this data gap.

4.4. CONCLUSION REMARKS

The pipeline implemented in this chapter prepares the data for energy analyses. It involves multiple steps and several decision-making processes using various software tools. In the case of the testbed, I performed lots of manual data collection that is not feasible to replicate when creating the country-wide dataset. Additionally, the data indexing method used to classify the Dutch building stock proved its functionality when used to filter the data in the computation of solar irradiance (see [section 4.3.1.2](#)).

The results of the machine learning projects to classify the Dutch building stock (see [section 4.3.2](#)) and to infer the number of storeys per building (see [section 4.3.3](#)) underscore the importance of high-quality input data for AI approaches. For the former, due to variability in the accuracy of the results, these values could not be reliably used as input data for future simulation processes [chapter 5](#). For the latter, the use of the calculation results is restricted to buildings up to five storeys.

As a final remark, despite basic data cleaning processes (e.g., removing negative years of construction of the buildings), no additional comparisons or validations

were performed so that I remove as little as possible the data provided by the input datasets. Further analyses into the input datasets will take place in [chapter 5](#).

5

IMPLEMENTATION

In this chapter, I provide more details on the software-side workflow of the urban Building Energy Simulation I have developed; here, I focus on the system design, the computations and tests to meet the objectives of my research. The chapter describes the underlying software processes that support my modelling and simulation decisions. It contains the details of the implemented software components, some of the challenges encountered, and some solutions to overcome them.

This chapter outlines the context and key decisions made for the implementation of the EnergyBAG Building Energy Simulation (BES). First, I introduce the general considerations for developing this BES ([section 5.1](#)). Next, I describe the software-based conceptual decisions, including C4 diagrams and corresponding tables that define the system components ([section 5.2](#)). Then, I present the software-side UML diagrams that illustrate the required processes for the EnergyBAG BES ([section 5.3](#)). Following this, I discuss the challenges related to data availability and the NTA 8800 requirements ([section 5.4](#)), along with the implementation decisions taken to address dataset limitations and regulatory constraints ([section 5.5](#)). Additionally, I review specific building cases with illogical results in early stages of the implementation; these cases result in adjustments in the implementation ([section 5.6](#)). This section also details a data validation check introduced as a result of these findings. Finally, the chapter concludes with reflections on the implementation process and key takeaways ([section 5.7](#)).

5.1. GENERAL CONSIDERATIONS

The main goal of this implementation is to develop and validate a methodological framework rather than to produce a polished software product. All decisions concerning data handling, model structure, and computational routines focus on Urban Building Energy Modelling (UBEM) in conjunction with semantic 3D city models (s3DCM). The code thus serves as a means to test the

viability of open datasets for computing building heat demand at various scales, from city to country level.

The scope of this chapter is to investigate the feasibility of implementing the design decisions based on the data requirements. Consequently, the code serves as a proof of concept rather than a user-focused tool for end users. My priority is to confirm that the logical workflow can produce consistent outcomes, which are then evaluated in [chapter 6](#).

As a proof of concept, the software will demonstrate the applicability of the proposed method at different scales of analysis. The software is written to handle input data from multiple data sources from individual buildings to the entire Dutch building stock, as discussed in [chapter 3](#). Although the following sections in this chapter do not explicitly track the iterative cycle of coding, testing, validation, and refinement, they detail the rationale behind the choices made during this phase of the research.

5

Since the goal is to validate design decisions and compute heat demand for the Dutch building stock, neither code optimisation nor robust software architecture became a priority for the implementation. Instead, the scope is to ensure that real-world datasets yield meaningful results under operational constraints.

I selected Python as the programming language for my implementation, although it has known performance limitations (Pereira *et al.*, 2021). Execution speed or resource usage inefficiencies were considered acceptable trade-offs, allowing more time to concentrate on the theoretical issues of my method and unexpected data gaps that require further work. Thus, the user interface remains minimal.

I calculate several parameters in advance, such as the building footprint areas and gross volume, surface areas, and number of storeys. The reason for these pre-calculations is to distribute the computational load and to minimise repetitive operations each time a simulation is run. It thereby increases the consistency across multiple executions and reduces the likelihood of discrepancies based on code updates.

For flexibility, intermediate results are stored so that only specific sections of the code can be executed to perform particular analyses (e.g., testing the influence of different air exchange rates on the net heat demand). This approach originates from my experience with SimStadt and CitySim, where any data change requires repeating the entire workflow, which is a computationally expensive and time-consuming process independent of the size of the study area.

Finally, the definitions given in [section 2.1](#) and [section 2.5](#) form the conceptual theoretical framework for my implementation. The data requirements explained in [chapter 3](#) guide each element of the pipeline design, indicating the need to integrate s3DCM with building and usage archetypes, weather data, and occupancy characteristics. Since the goal extends to country-level analyses, assumptions from the NTA 8800 (NEN, 2024) are fully adopted to ensure consistency and completeness of my method according to Dutch regulations.

5.2. CONCEPTUAL DECISIONS

The theoretical framework presented in [chapter 2](#), the reflections based on the data requirements from [chapter 3](#), and the design decisions made in [chapter 4](#) are consolidated here through the implementation of a building energy software tool based on the BAG, which is therefore named *EnergyBAG BES*.

I use the C4 model to represent the architecture and functionalities of the EnergyBAG BES. C4 stands for (System) Context, Containers, Components, and Code (Brown, 2023). This model provides a hierarchical set of diagrams that represent the software system at multiple levels of abstraction. Specifically, I use the system context and system container diagrams to introduce my implementation.

5.2.1. C4-MODEL SYSTEM CONTEXT

The *system context* diagram serves as a starting point by showing the software system within its environment. It helps to understand the scope of the system, who uses it, and what the critical system dependencies are. [Figure 5.1](#) shows the EnergyBAG BES context diagram; the description of its components is available in [table 5.1](#).

Table 5.1: Description of EnergyBAG BES system context diagram

Element	Definition
Data Preparation	This system is in charge of the data collection for the energy simulation. It could involve the calculation of missing parameters from the corresponding source
Data Manipulation	This system takes care of data necessary for the energy simulation. It deals with several data sources and interacts with the other systems in EnergyBAG BES
Energy Simulation	This system takes care of the energy simulation

The design decision of the EnergyBAG BES allows users to compute the net heat demand of buildings based on three main processes: data preparation, energy simulation and data management. A user communicates with the data preparation and energy simulation systems when they want to perform a calculation. The data preparation system checks whether all the data required for a specific energy simulation are available.

If any data are missing, it performs the necessary calculations to generate the missing data. Then, the energy simulation system uses these data to perform the requested calculations. Both the data preparation and energy simulation systems read from and write to the data manipulation system, which manages data across multiple data sources.

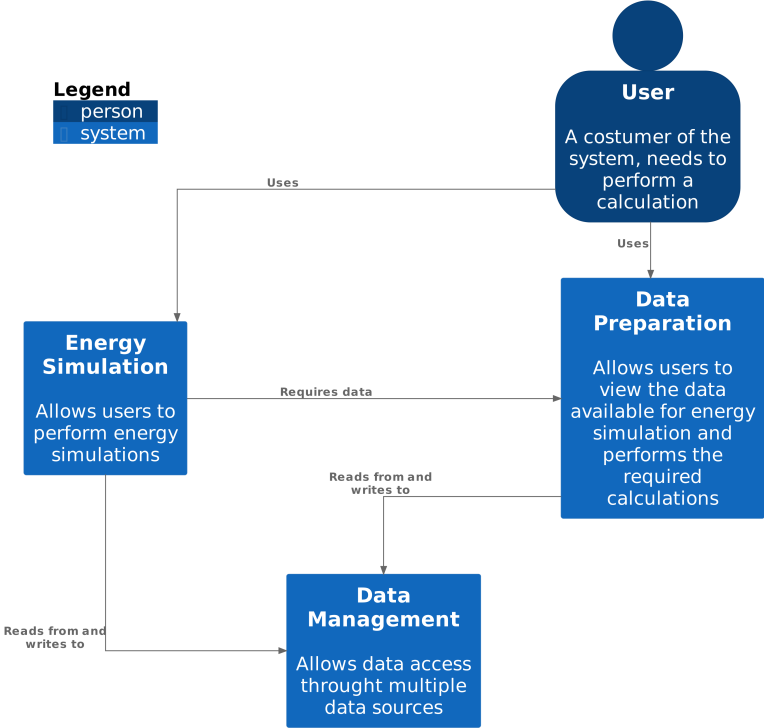


Figure 5.1: C4-model system context diagram for EnergyBAG BES

5.2.2. C4-MODEL SYSTEM CONTAINER

A simplified definition of the *system containers* indicates that it is a zoom to the software system in scope. This diagram shows the high-level technical components (containers) and their interactions. Therefore, it serves as a detailed specification of the systems defined in the system context and provides a more granular view of the system’s architecture. Table 5.2 describes the elements of the system container diagram for the EnergyBAG BES that are shown in Figure 5.2.

Table 5.2: Description of EnergyBAG BES system container diagram

Element	Definition
Business Rule	It defines the business rules to perform the energy simulations, for example, solar irradiance model, shadowing conditions, refurbishment scenario
Study Area	It deals with the definition of the area of interest for the calculations. In the case of big areas, it split them into tiles

Continued on next page

Table 5.2 – continued from previous page

Element	Definition
Transmission	Deals with the computation of the heat transfers by transmission
Ventilation	Deals with the computation of the heat transfers through air flows from outside
Internal gains	It deals with the computation of the energy gains values by people and appliances
Solar gains	It provides the solar gains of the building
Solar irradiance	Calculates the solar irradiance for a given study area
Parameters creation	Calculates the necessary parameters for the available energy simulations, i.e., the number of storeys of a building part according to the NTA 8800 (equation (2.14))
Parameters collection	Allows the users to query the data available for the energy simulation
Data Manipulation	This base system provides the data manipulation methods and deals with the interaction with several databases
3DCityDB	Database management system (DBMS) for the s3DCM. Described in section 2.2.1.3
LibraryDB	DBMS for the building physics data. Described in section 4.2.1
EnergyBAG DB	It contains the business rules, user parameters and results

The user interacts with two containers: *business rule* and *study area*. Business rule sets the rules and configuration parameters for the energy simulation. Examples of business rules include the selection of refurbishment conditions, the solar irradiance method, or how to proceed when a building(part) has no or insufficient vbos data. The study area handles the definition of the area of interest for the calculation and splits it into tiles for better efficiency of the simulation tool.

The core of the energy simulation system is composed of several containers that perform specific calculations according to the energy balance method defined by NTA 8800 [section 2.5](#). However, the design decisions of the energy simulation system do not follow a hierarchical approach as shown in [figure 2.14](#) since the user can directly execute the computation of the solar irradiance at the study area. For example, the solar irradiance values might come from an external data source, as the computation performed in [section 4.3.1.2](#) or CitySim, Ladybug –to mention some– instead of the NTA 8800 standard values.

As already mentioned, the energy simulation system interacts with the Data Preparation system. However, the energy simulation system is split into Parameters Creation and Parameters Collection, which, as their names indicate, are in charge of either creating the corresponding parameters required for the computation of the energy simulation or collecting them from the corresponding data sources by means of the Data Manipulation system. All these containers communicate through API(s).

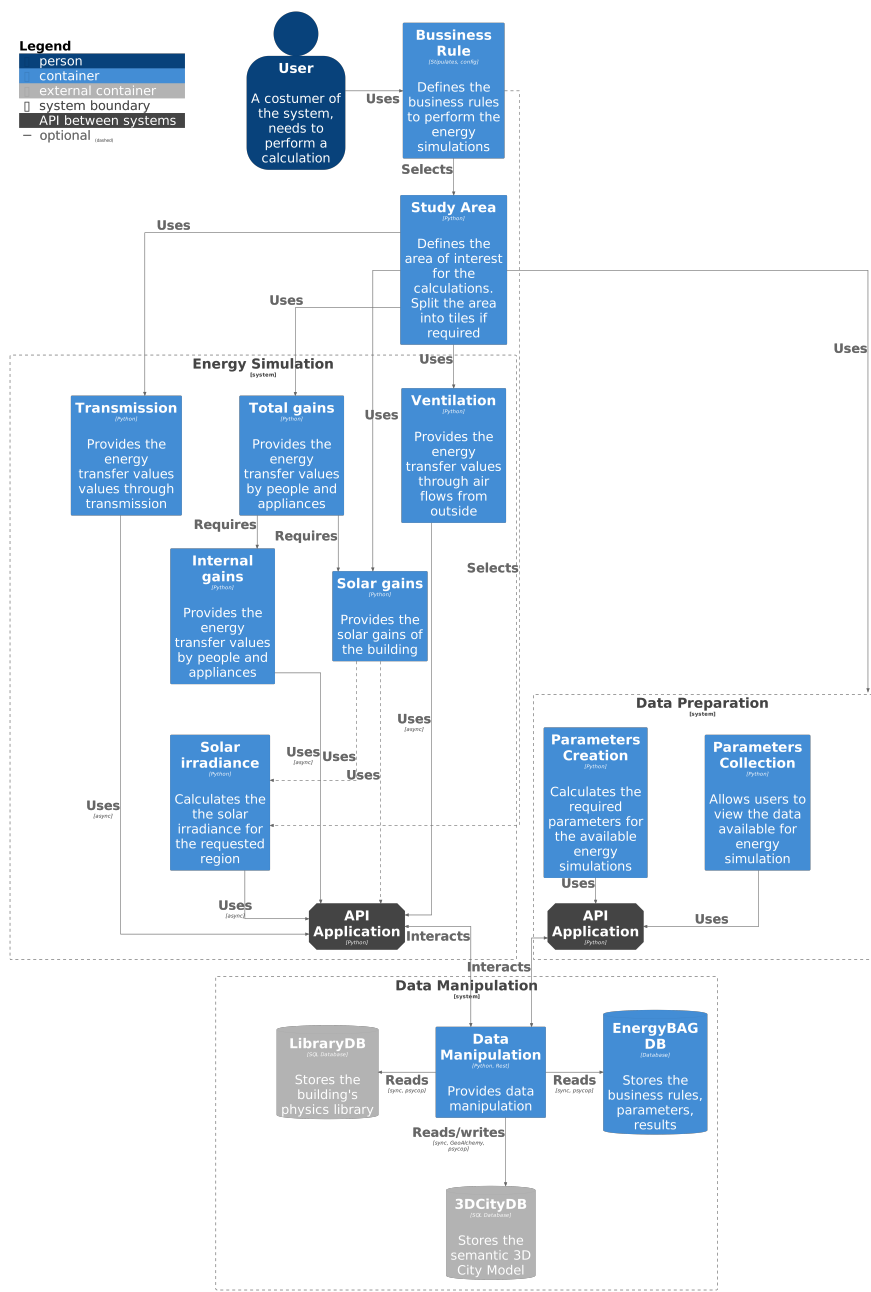


Figure 5.2: C4-model system container diagram for the EnergyBAG BES

The remaining system corresponds to Data Manipulation. It is responsible for managing the interaction with the several databases used in the process. The

communication between the systems and the data is done using the Python Database API specification (Lemburg, 1999). Data are stored in multiple databases as already introduced in chapter 4.

The data manipulation system, shown in figure 5.2 is split into two main components: data manipulation and the databases. For the data manipulation, I use Python v3 and the library *SQLAlchemy*, which is a Python toolkit and Object Relational Mapper (ORM) that provides a suite of well-known persistence patterns (Bayer, 2012); and *psycopg* which is the PostgreSQL adapter for Python (Varrazzo, 2023).

The data manipulation system is crucial for the energy simulation since it retrieves the necessary data from the databases. Therefore, it is in charge of storing the computing energy simulation results in both the 3DCityDB and EnergyBAG DB. Grouping all the application data manipulation processes into this system enables efficient data management and retrieval from the other systems.

5.3. IMPLEMENTATION OF THE ENERGY SIMULATION SYSTEM

The Unified Modelling Language (UML) sequence diagram of the Energy Simulation System is shown in figure 5.3. The sequence flows from left to right. The main components are Python and the databases LibraryDB, EnergyBAG DB and 3DCityDB. The application is initiated by the Python script, which is the central orchestrator of the system. It starts by validating the connection to all databases. In case of a non-working connection, the application will not continue. After the connection to the databases is validated, the Python script queries the physical data, e.g., U-values and g-values, based on the building type. The following step corresponds to the query of the NTA 8800 parameters. Among these parameters are set point temperature per function, length of the month, standard climate data, and internal heat gains values per building function. These data are queried at the beginning since they are used in the following processes and values and never vary. The EnergyBAG DB also contains business rules that the user can customise. Some examples of them are the refurbishment scenario of the computation and the management of data inconsistencies between the input datasets. The latter is explained in detail in section 5.6.6.

After receiving the response of the EnergyBAG DB for the NTA 8800 parameters and business rules, the implementation loop starts. The following parts are repetitive and correspond to:

- **Spatial Filter:** As mentioned in section 4.2.3, I chose the H3 Indexing to split the study area into equal size smaller chunks, which are more manageable in terms of data flow and data processing between the DB servers and the computer that performs the calculations. This filter obtains

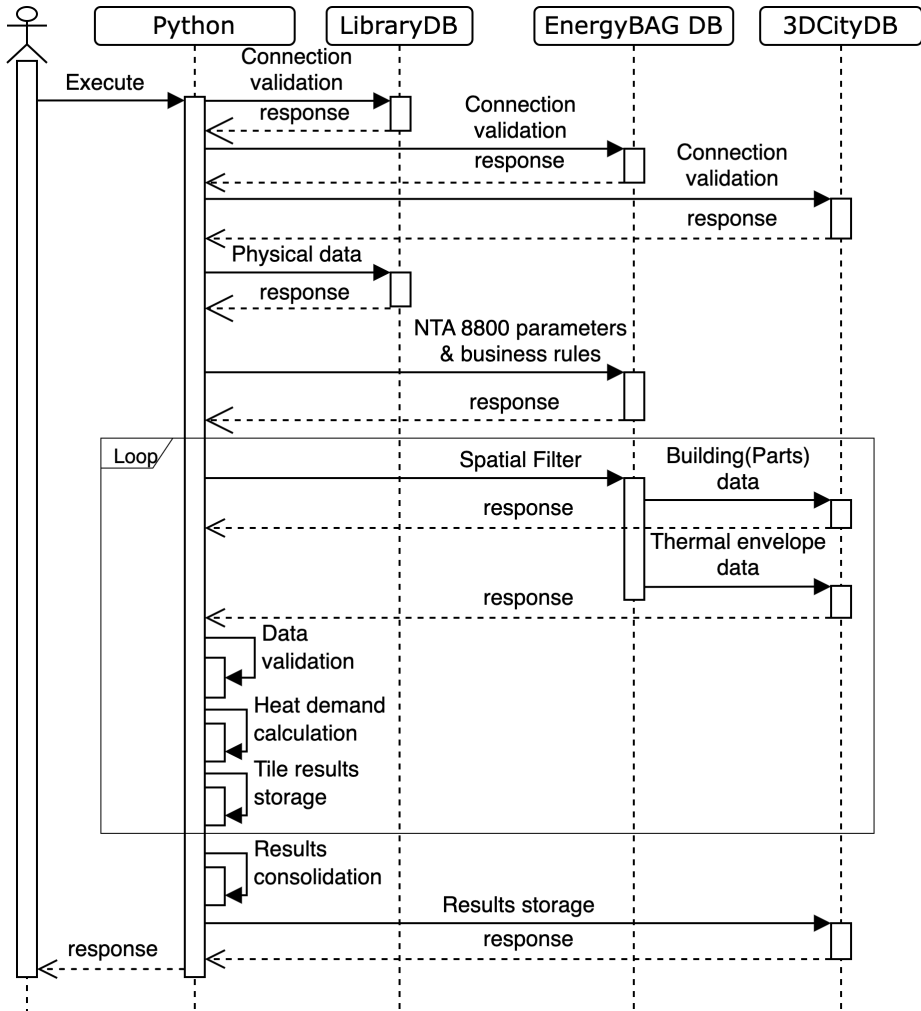


Figure 5.3: Sequence diagram of EnergyBAG BES

the *Building(Part)s* inside the corresponding tile with their corresponding attributes and thermal envelope data.

- **Data Validation:** The scope is to guarantee that the data comply with the minimal requirements to perform the corresponding calculations based on the established business rules.
- **Heat demand calculation:** The resulting data are used for the computation of the net heat demand of buildings according to the NTA 8800 and the selected business rules.
- **Tile results storage:** The final step corresponds to the storage of the

calculation results, which is managed per H3 tile. Results are first stored in external files to avoid the saturation of the DB server when executing multiple loops simultaneously. Then, the data from the output files are aggregated into a single file. Finally, computation results are stored in the input 3DCityDB instance.

The remaining processes correspond to the consolidation of the results per tile into one, which is stored in the 3DCityDB instance using the Energy ADE (KIT profile) data model.

Figure 5.4 shows the activity diagram of the EnergyBAG BES. It starts by establishing the connection to the databases mentioned in the Data Manipulation System (figure 5.1). After validating the database connections, data are filtered by the corresponding H3 index that covers the study area. The filtered data from the Building(Part)s, and their corresponding ThematicSurfaces are handled by Pandas DataFrames (McKinney, 2010), which are two-dimensional tabular data with mutable size.

The following process consists of the Basic Calculations, which compute the parameters necessary for the net heat demand calculation. The resulting data are merged to create a single DataFrame that contains all the data required for the final simulation steps. Based on the data requirements, I first calculate the number of occupants per residential zone using equations (2.20) to (2.22). This value is later used to calculate the internal gains and the heat demand for hot tap water, implementing the respective formulas given in section 2.5.1.

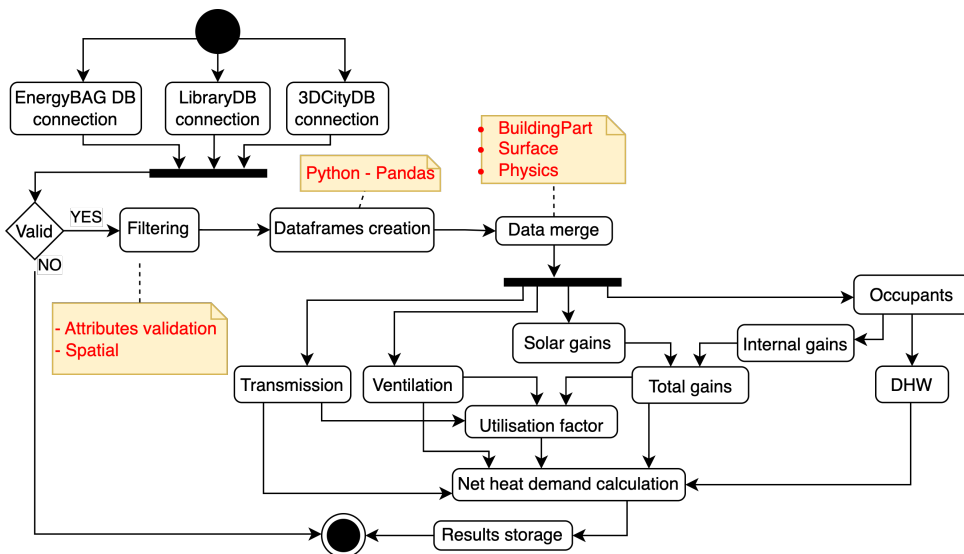


Figure 5.4: Activity diagram of EnergyBAG BES

The remaining decision is on the solar irradiance model to use for the calculation of solar gains. The default parameter is the statistical data available in the NTA 8800; the other possibility is the use of values (if available or already precomputed) computed by external simulation tools as discussed in [section 4.3.1.2](#), which are stored in a 3DCityDB+Energy ADE (KIT profile) instance. The remaining processes follow the logic introduced in [section 2.5](#). The final process is to store the results in the databases using the Data Manipulation system.

5.4. CHALLENGES IN THE IMPLEMENTATION PHASE

The most persistent problem encountered during the research is the lack of data. In [chapter 4](#), I discuss this issue and propose some strategies to overcome it based on the experience gathered on the preliminary tests. However, during the implementation of EnergyBAG BES, those data gaps became more significant. One example is the absence of usable area data. As already shown in [figure 3.4](#), 59% of buildings lack vbo information, meaning that they do not have information on function and usable area. As a consequence, it is not possible to compute the net heat demand since the NTA 8800 specifies several of its parameters based on this attribute. Currently, the only way to solve this challenge is to make assumptions that introduce additional uncertainties to the calculations.

Additional challenges come from the specifications of the NTA 8800. The norm has no indications for processing industrial buildings. Therefore, buildings with this function cannot be processed at the moment in my implementation, or it is required to include additional data sources or methods. In the case of healthcare buildings, for example, the norm distinguishes between healthcare facilities with beds and healthcare facilities without beds. These classes have different values for specific parameters, that is, the case of the set-point temperature with values of 22°C and 21°C, respectively. However, such granularity is not present in the BAG and can not be identified geometrically from a LoD2 model as the 3DBAG. Hence, these mismatches in categorising buildings lead to an approximation in which all healthcare buildings will be associated with the same parameter values.

Further challenges and the steps taken to resolve them or minimise their impact are outlined in [section 5.6](#). They mainly involve data issues, such as mismatched data epochs ([section 5.6.1](#)) or inconsistencies within the input datasets ([sections 5.6.1](#) and [5.6.6](#)). Moreover, specific NTA 8800 parameters can introduce errors if not handled correctly. One example is the set-point temperature per function compared to the monthly average outdoor air temperature values. Sports buildings, for instance, have a heating set-point of 16°C ([table 2.4](#)), which is below the monthly average temperature from June (16.12°C) through August (18.48°C). Overlooking these data will lead to inconsistencies in the calculation results.

5.5. IMPLEMENTATION DECISIONS

The following implementation decisions were made based on the empirical knowledge acquired in the course of this research. They are based on the input s3DCM and the calculation zone approach on which the NTA 8800 is based. As mentioned in [section 5.1](#), the scope of this research is to perform the calculations at the country level. Therefore, data handling should be sufficiently simplified and generalised to enable the management of large datasets.

5.5.1. MODELLING

5.5.1.1. THERMALZONE(S)

The 3DBAG, like many other 3D city models, provides the geometrical representation of buildings. There is no information about their interior. Consequently, any internal division introduces additional assumptions. [Figure 5.5](#) shows the decision taken for my implementation. The input building ([figure 5.5a](#)) is represented as a LoD2 MultiSurface geometry ([figure 5.5b](#)). I decided to use, for the moment, a single *ThermalZone* per building as it is common also in other similar cases found in literature (Johari, Shadram *et al.*, 2023; Nutkiewicz *et al.*, 2021). Therefore, the *ThematicSurfaces* of the LoD2 geometrical representation of Buildings are the same as the *ThermalBoundaries* of the *ThermalZone* [figure 5.5b](#). However, my implementation supports multiple *ThermalZones* per building (as [figure 5.5c](#)) by querying from the input s3DCM all *ThermalZones* per building including their *ThermalBoundaries* (when present). All computations are already done separately up to the *ThermalZones* level and later aggregated to the *Building(Part)*.

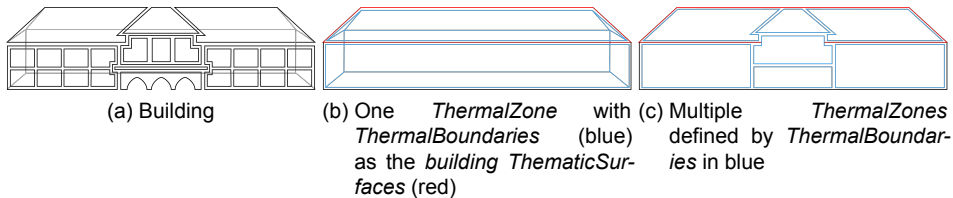


Figure 5.5: Energy ADE *ThermalZone* approach of my implementation

5.5.1.2. OPENINGS

The input geometries derived from the 3DBAG LoD2 representation do not have openings (e.g., needed for windows). To address this gap, I adopted a window-to-wall ratio of 0.3 and a window-to-roof ratio of 0.1 based on data exploration conducted during my research and are in line with similar values found in literature. These parameters can be adjusted and further refined as necessary for subsequent simulations, for example, by using typical values per building type when available.

5.5.1.3. NUMBER OF STOREYS

Another decision in modelling relates to the number of storeys, a required parameter for several processes in the NTA 8800. However, as mentioned in [section 4.3.3](#), this attribute is not available for the entire country. Consequently, as a fallback, I use the NTA 8800 specification ([equation \(2.14\)](#)). The workflow for creating this parameter is shown in [figure 5.6](#).

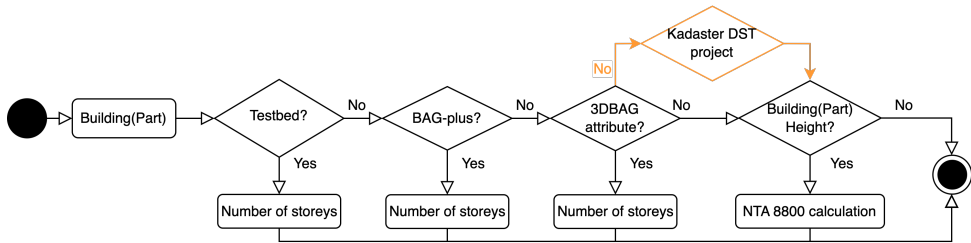


Figure 5.6: Logical workflow for defining the number of storeys in the EnergyBAG

The workflow includes an orange step representing a dataset created by the Data Science Team (DST) of Kadaster Netherlands. This team processed the entire 3DBAG to compute the number of storeys per building (Kadaster, 2022). The used method developed by (Roy *et al.*, 2023), briefly presented in [section 4.3.3](#) predictors classify results into two categories: 1 to 5 storeys (accuracy: 91.94%) and 6 to 47 storeys (accuracy: 42.99%), highlighting the lower reliability of results for taller buildings.

Additionally, the online dataset was actually inaccessible for an extended period during my research (2023–2024). Given the indicated data quality and accessibility challenges, I decided not to include this dataset. Instead, I used the NTA 8800 specification as a fallback to assign the number of storeys to buildings lacking this data. From December 2024 onwards, the number of storeys is available as a *GenericAttribute* in the 3DBAG, and therefore, this can be resolved in future developments of my workflow.

5.5.1.4. BUILDING TYPE

As already written in [section 4.3.2](#) a major issue due to lacking data is the building type since ~ 90% of buildings do not have this information. Due to the low accuracy of the classifiers in the method developed by Poon (2024), I could not use the results as input data for the EnergyBAG BES. Therefore, I defined simplified classification rules for building types using the current data available to increase the number of buildings that can be processed. This method is also used by other software solutions, such as SimStadt, to fill similar data gaps.

The rules that classify the building types consider factors such as the number

of vbos registered to the *Building(Part)*, the number of storeys, and the number of adjacent *Building(Part)*s (see figure 5.7 for the adjusted workflow). The "function" attribute is used to categorise *Building(Part)*s into non-residential (when none of the "function" values in the *Building(Part)* is residential); in the case of Mixed-used buildings, these are categorised into the remaining classes. However, the resulting classes do not include the Dutch-specific types shown in figure 3.9.

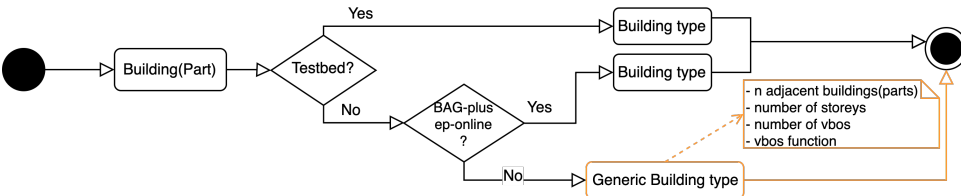


Figure 5.7: Adjusted Workflow for assigning the building type attribute

Figure 5.8 presents the aggregated results of the EnergyBAG dataset by building type. For comparison, it also includes the previous classification of the Dutch building stock based on the input datasets as shown in figure 4.9. The decrease in the number of semi-detached houses can be attributed to the computed number of storeys. According to RVO (2023), these buildings should have up to four storeys, leading to fewer *Building(Part)*s being classified in this category.

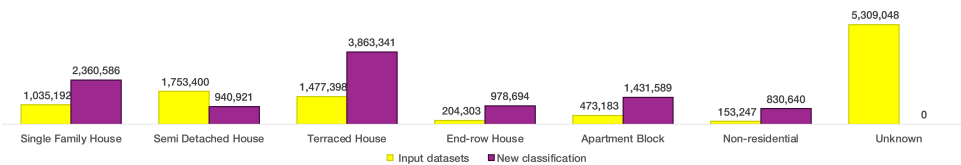


Figure 5.8: Distribution of classes before (yellow) and after (purple) my classification method

5.5.1.5. HEAT TRANSFERS THROUGH VENTILATION

Calculating the air volume flow presents another challenge, as it is fundamental to determine heat transfers by ventilation. Due to the input data available, I decided to use the method based on the building type, this method bases the values for the Air volume flow (q_v) and Dynamic correction factor (f_v) on the building type categories, the number of storeys of the building and the roof type of the building. However, this approach still has uncertainties since the values are specified for one-storey buildings and the location of the vbo (dwelling) inside the apartment block as presented in figure 3.12. Based on this, I consider all buildings with more than one storey as corner locations on the top floor table 2.5 since this category assigns the highest dynamic correction factor ($f_v = 1.4$).

Consequently, it leads to higher values of the net heat demand through ventilation losses.

5.5.1.6. HEATING PERIOD

The NTA 8800 specifies that the winter period, or heating season, spans 212 days (NEN, 2024, p. 1084), corresponding to the interval from 1st October to 30th April, inclusive. While my scripts can perform computations for all twelve months of the year, the yearly aggregated values used in my research and presented in [chapter 6](#) focus specifically on the winter period, as NTA 8800 specifies.

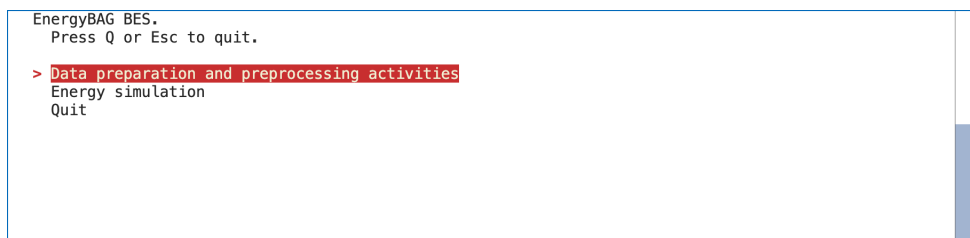
5.5.1.7. WEATHER DATA

I use the climate parameters defined by the NTA 8800 for my computations. Therefore, solar irradiance and shading conditions are assigned based on the inclination and orientation of the thermal boundary surfaces. For the shading conditions, I use the values named as "full obstruction conditions" (NEN, 2024, table 17.13).

5.5.2. IMPLEMENTATION WORKFLOW

The implemented workflow is divided into several functionalities, each with a specific purpose. The software-based processes required for the pipeline design ([chapter 4](#)) are consolidated into the first main section of my solution. The second section correspond to the Energy Simulation. To facilitate the processing and deal with large datasets, I use Pandas Dataframes for data manipulation. All custom functions are vectorised, following Python data management principles, to enhance performance.

I implemented an interactive menu using the Python library Simple Terminal Menu¹. The hierarchical menu of EnergyBAG BES is shown in [figure 5.9](#). The workflow is split into two main elements: *Database set-up* and *Energy simulation*.



```

EnergyBAG BES.
Press Q or Esc to quit.
> Data preparation and preprocessing activities
  Energy simulation
  Quit
  
```

Figure 5.9: EnergyBAG BES main menu

¹Version 1.6.4. Further details in (Meyer, 2023)

Data preparation and preprocessing activities focuses on data preparation and preprocessing tasks that are not required for every simulation run. By precomputing static variables (e.g., number of storeys, heated area per building, building volume) and reusing them across simulations, computational efficiency is significantly enhanced. These variables remain unchanged between simulations and are configured once prior to execution. The available preprocessing options, shown in [figure 5.10](#), include:

```
Data preparation and preprocessing activities.
Press Q or Esc to back to main menu.

> Data load
Data cleaning
Data preprocessing
Database enrichment
Data quality checks
Back to Main Menu
```

Figure 5.10: EnergyBAG BES options

- *Data load*: Loads the multiple input datasets to the centralised EnergyBAG DB
- *Data cleaning*: It applies basic data cleaning methods to remove logical inconsistencies in the input datasets.
- *Data preprocessing*: It corresponds to the data preparation to perform the net heat demand calculation based on the NTA 8800. Its scope is to guarantee that the minimum data requirements are met. Therefore, it includes the number of storeys and building type classification as described in [section 5.5.1.3](#), usable area validator [section 5.6.6](#) and the assignment of the corresponding H3 index tiles to each of the *Building(Part)s* in the input dataset.
- *Database enrichment*: It stores the computed parameters into the corresponding 3DCityDB instance.
- *Data quality checks*: It performs validations to the input datasets among the several sources. One example is the validation of the usable area as described in [section 5.6.6](#)

The options available for the *Energy simulation* are presented in [Figure 5.11](#). These include *Net heat demand calculation*, *Results consolidation* and *Results analysis*. Each option in the menu performs the following tasks respectively:

```

Energy simulation Menu.
Press Q or Esc to back to main menu.

> Heat demand calculation
Results consolidation
Results post-processing
Back to Main Menu

```

Figure 5.11: EnergyBAG BES options

- *Heat demand calculation*: It contains all scripts to compute the net heat demand per *Building(Part)*. It follows the sequence diagram presented in [figure 5.3](#). The script is structured for multi-threading execution; the number of threads that perform the computation is based on the number of physical cores of the hosting computer. Each thread processes the *Building(Part)s* within its assigned H3 tile.

The output per thread is a CSV file with all calculations aggregated by *Building(Part)*. This approach ensures interoperability with database users who have read-only access permissions. Furthermore, CSV files are a handy way to analyse data locally, including for people with less database experience.

- *Results consolidation*: It consolidates the results of all tiles of the input dataset into a single file that is used in the following options of the menu. It contains the scripts to store the aggregated results into a 3DCityDB+Energy ADE (KIT profile) instance.
- *Results post-processing*: It contains all the scripts to generate the figures and tables that are presented in [chapter 6](#).

5.6. FOCUSED REVIEW: SPECIFIC BUILDING CASES

Here, I present a small selection of problematic cases found when preprocessing the data prior the calculation of the net heat demand of buildings. These examples highlight common patterns that lead to noticeable incorrect calculations without requiring direct comparison to ground truth data. That is the reason for having additional *data quality checks* in the implementation workflow prior to processing a building to deal with such cases. All samples included in this section are related to the usable area (in Dutch "*gebruiksdoel*") attribute

5.6.1. PAND ID 1742100000100986

According to the BAG (Kadaster, 2024a), this single family house residential building was constructed in 2023 and has a usable area of **595m²** (figure 5.12). The computed yearly heat demand for this building is 5.3 kWh/(m²·a), which is a very low value even for a new building. The features extracted from the EnergyBAG DB provide the data available in table 5.3:

Table 5.3: Summary data for Pand ID 1742100000100986

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
65.4m ²	1	14.5m	152.8m ³	1	1924m ²

5

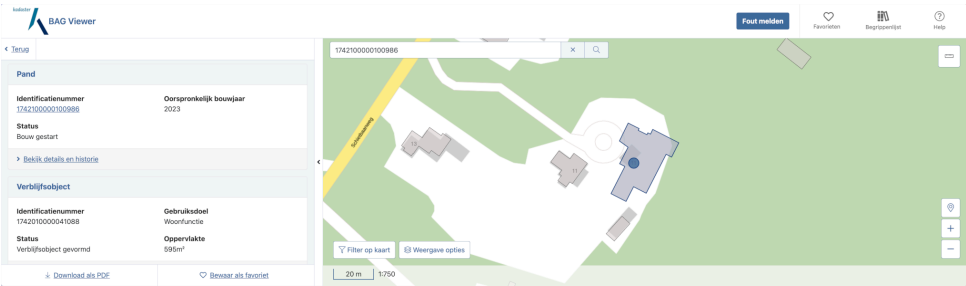


Figure 5.12: BAG viewer data for building with Pand ID 1742100000100986

Although no inconsistencies are evident when each dataset is evaluated independently, they may emerge when the datasets are combined. For example, it is unrealistic for a building with a footprint area of **65.4m²** and **1** storey to have a usable area of **595m²**. Furthermore, figure 5.13 highlights the discrepancy between the geometric representation from the BAG dataset (black outline) and the footprint from the 3DBAG dataset (red polygon). Both figures 5.12 and 5.13 use the official map from the Basic Registry of Large-Scale Topography (BGT) as their base, which closely aligns with the footprint from the 3DBAG. These observations highlight the need for additional validation steps before conducting any computation if such a case happens.

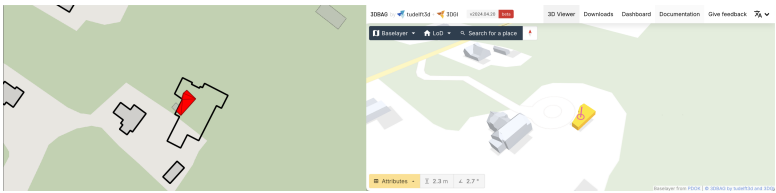


Figure 5.13: Geometrical representation of Pand ID 1742100000100986 at December 2024

5.6.2. PAND ID 1742100000006419

This industrial building, constructed in 1989 (Kadaster, 2024a), is two storeys above ground and contains one vbo, as shown in figure 5.14b. The properties extracted from the EnergyBAG database are summarised in table 5.4. The computed yearly heat demand is 9.5 kWh/(m²·a), which is remarkably low for a building from this construction period, even if it was purely residential.

Table 5.4: Summary data for Pand ID 1742100000006419

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
215.5m ²	2	20.72m	1950.7m ³	1	5835m ²

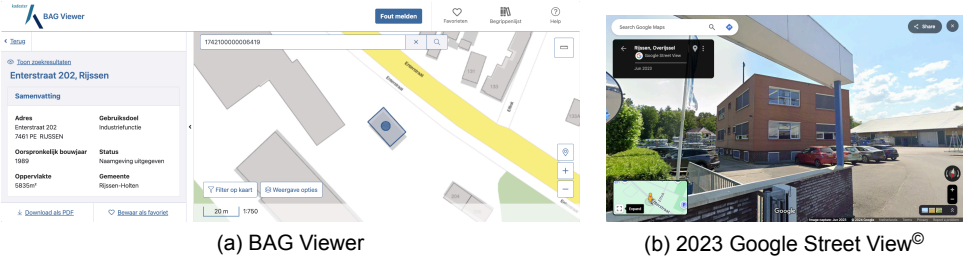


Figure 5.14: Basic visualisation of building 1742100000006419

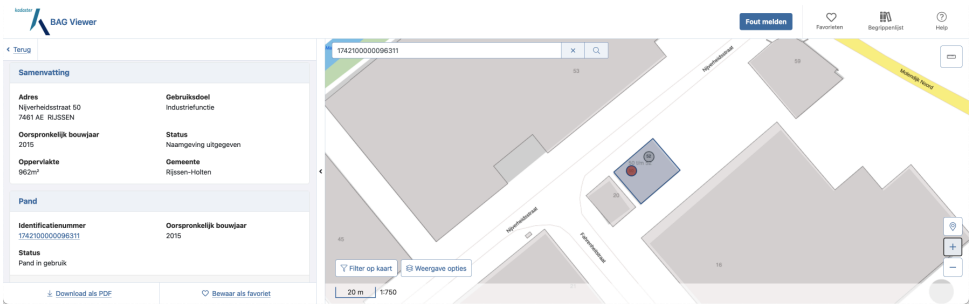
However, the values in table 5.4 show evident inconsistencies for the usable area (5835m²) since there is no physical space for such a value: the usable area, according to the BAG, is 27 times the size of the footprint area of the building, which is not in line with the two storeys that can be identified in figure 5.14b.

5.6.3. **PAND ID 1742100000096311**

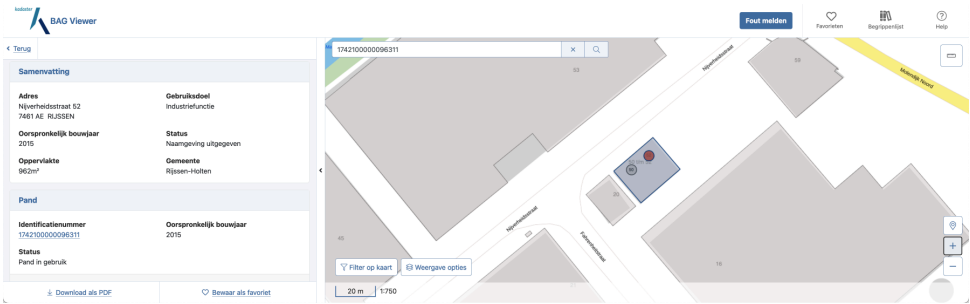
The BAG dataset (Kadaster, 2024a) indicates that the building was constructed in 2015 with an industrial function; therefore, its building type is non-residential. The basic description of the BAG is shown in figure 5.15. The features extracted from the EnergyBAG DB provide the data available in table 5.5:

Table 5.5: Summary data for Pand ID 1742100000096311

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
526.7m ²	2	16.59m	3591.8m ³	2	1924m ²



(a) Nijverheidsstraat 50



(b) Nijverheidsstraat 52

Figure 5.15: BAG viewer data for building with Pand ID 1742100000096311

Although the building height may suggest more than two storeys, Google Street View (figure 5.16) confirms that the building has only two storeys. Furthermore, based on the collected data, this building has a usable area of more than 3,5 times the area of the building footprint, which seems unlikely.



Figure 5.16: 2023 Google Street View[©] Pand ID 1742100000096311

5

The computed yearly net heat demand for this building is $8.9 \text{ kWh}/(\text{m}^2 \cdot \text{a})$, which would be a very low value if it was a residential building. The year of construction is almost a decade old compared to the current legislation for the construction of new buildings, which means that it does not follow the latest construction specifications in the Netherlands for this type of building.

5.6.4. PAND ID 0150100000010860

The computed yearly heat demand for this residential building is **11 MWh** $/(\text{m}^2 \cdot \text{a})$, which is a very high value for a building with one registered vbo and a usable area of **72m²** [figure 5.17](#). Furthermore, its 3DBAG footprint area is significantly larger, i.e. **1,994m²**. The features extracted from the EnergyBAG DB provide the data available in [table 5.6](#):

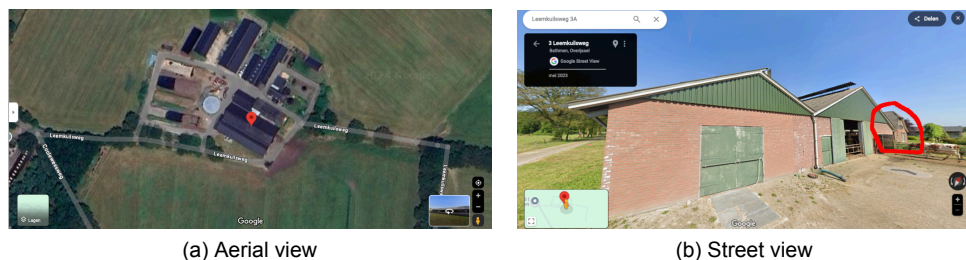
Table 5.6: Summary data for Pand ID 0150100000010860

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
1993.9m ²	2	16.64m	10373.3m ³	1	72m ²

[Figure 5.18](#) shows the Google imagery of Pand ID 0150100000010860. It is visible that the data from the BAG do not correspond to the ground. Furthermore, based on [figure 5.18b](#), it can be concluded that the BAG data corresponds to the building highlighted by the red circle.



Figure 5.17: BAG viewer data for building with Pand ID 0150100000010860



(a) Aerial view

(b) Street view

Figure 5.18: Google imagery for building with Pand ID 0150100000010860. Taken from <https://www.google.com/maps>

The previous examples serve to illustrate the widely erroneous results after data integration processes. These errors not only reveal the impact of data inconsistencies but also serve as indicators of the underlying building typology.

5.6.5. ADDITIONAL INCONSISTENCIES

Out of 10,553,631 buildings, 3,578 have a registered usable area of **1m²**, while their 3DBAG footprint is at least **48% larger**. This discrepancy highlights possible issues with the data. One example is the building with ID **0160100001372603**. As detailed in [section 3.2.1.1](#), I use the BAG data from epoch 8th August 2024; it records this building with one vbo and a usable area of **1m²** but the 3DBAG has a footprint area of **12,890.66m²**. However, historical records (now expired) show different usable area values, as shown in [figure 5.19](#). Furthermore, as of December 2024, the BAG viewer [figure 5.20](#) lists this building with a usable area of **13,370m²** (Kadaster, [2024a](#)).

A2 pand_identificatie	A2 vbo_identificatie	123 usable_area	documentdatum	A2 documentnummer	123 voorkomenidentificatie
0160100001372603	0160010000041907	1	2023-02-08	469950	6
0160100001372603	0160010000041907	8,176	2019-05-28	2456196	5
0160100001372603	0160010000041907	8,176	2018-06-14	V2018-0203-01	4
0160100001372603	0160010000041907	6,536	2015-05-21	1468028	3
0160100001372603	0160010000041907	6,536	2014-12-08	1408143	2
0160100001372603	0160010000041907	6,536	2003-01-10	BV_HDB2001:2352	1

Figure 5.19: BAG historical data for building with Pand ID 0160100001372603



Figure 5.20: BAG viewer data for building with Pand ID 0160100001372603 at December 2024

5.6.6. LESSONS LEARNT

Based on the typologies of errors for which examples have been shown before, I implemented a data quality checker focused on the usable area per building. This validator calculates a theoretical usable area by considering 80% of the building's footprint area and the number of storeys. The remaining 20% factor accounts for non-heated spaces (e.g., stairwells, attics) to avoid overestimation (Dochev *et al.*, 2020; Johari, Shadram *et al.*, 2023). The computed value is then compared to the BAG usable area, and buildings with discrepancies exceeding 30% are flagged as shown in equation (5.1). These flagged buildings are subsequently excluded from the analyses²

$$usable_area_{calc} = footprint_area \cdot 0.8 \cdot num_storeys \tag{5.1}$$

If:

$$0.7 \cdot usable_area_{calc} \leq usable_area_{BAG} \leq 1.3 \cdot usable_area_{calc}$$

Then:

$$is\ computable? = False$$

²This is a heuristic decision based on the data exploration I have done during my PhD research. Additionally, the rationale is that the usable area means those areas of a building that are heated. Using a percentage of the footprint provides a logical and systematic approach at the city level. Moreover, a deviation exceeding 30% between the official usable area and the calculated value suggests inconsistencies in the data sources for the corresponding building or indicates the need for an alternative approach.

The resulting EnergyBAG DB indicates that 4,184,177 *Building(Part)s* (40.2%) have no vbos, therefore cannot be considered as residential, and 1,726,288 buildings (16.6%) should be excluded from the analysis due to inconsistencies in the usable area data. One of the reasons for the discrepancy of values between datasets, could be related to the epoch of the input datasets; for example, 3DBAG is derived from AHN lidar data (section 3.2.1.2), which is collected over multiple years using aerial surveys. The upcoming AHN5 is scheduled for data collection between 2023 and 2025, with Rijssen-Holten expected to be scanned in 2025 (Rijkswaterstaat, 2023). In contrast, the registry data, from which the BAG is updated and released monthly, is updated on a daily basis. Figure 5.21 shows the distribution of the EnergyBAG DB, the processed 43.2% of the BAG corresponds to 4,502,969 *Building(Part)s*.

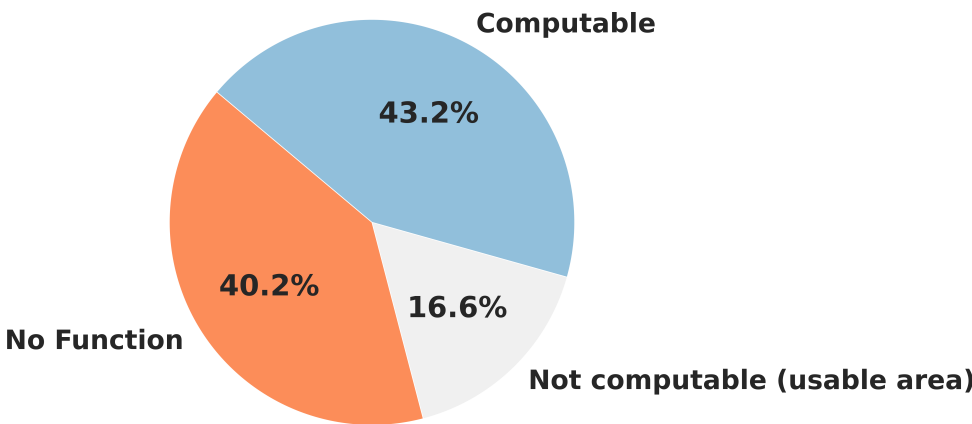


Figure 5.21: Suitability of Dutch building stock for EnergyBAG BES processing

Additionally, the processed 4,502,969 *Building(Part)s* can be further categorised based on their class (an aggregation derived from their function). Among these, 40% are residential buildings, while 1.2% (128,354 buildings) are classified as mixed-use. The distribution is presented in figure 5.22.

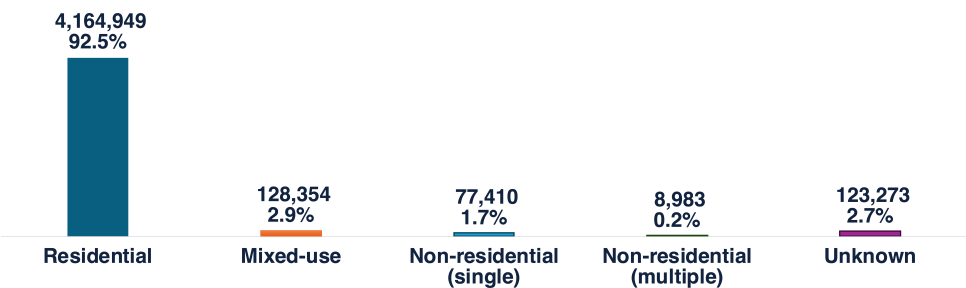


Figure 5.22: Processed building stock classified by class in the Netherlands

The spatial distribution of the computable Building(Part)s in the Netherlands is shown in [figure 5.23](#), where darker shades of blue indicate a lower number of processed elements. The spatial distribution of the Building(Part)s with no vbos data in the Netherlands is shown in [figure 5.24](#). Also, the spatial distribution in the Netherlands of the flagged *Building(Part)s* from [section 5.6.6](#) are shown in [figure 5.25](#).

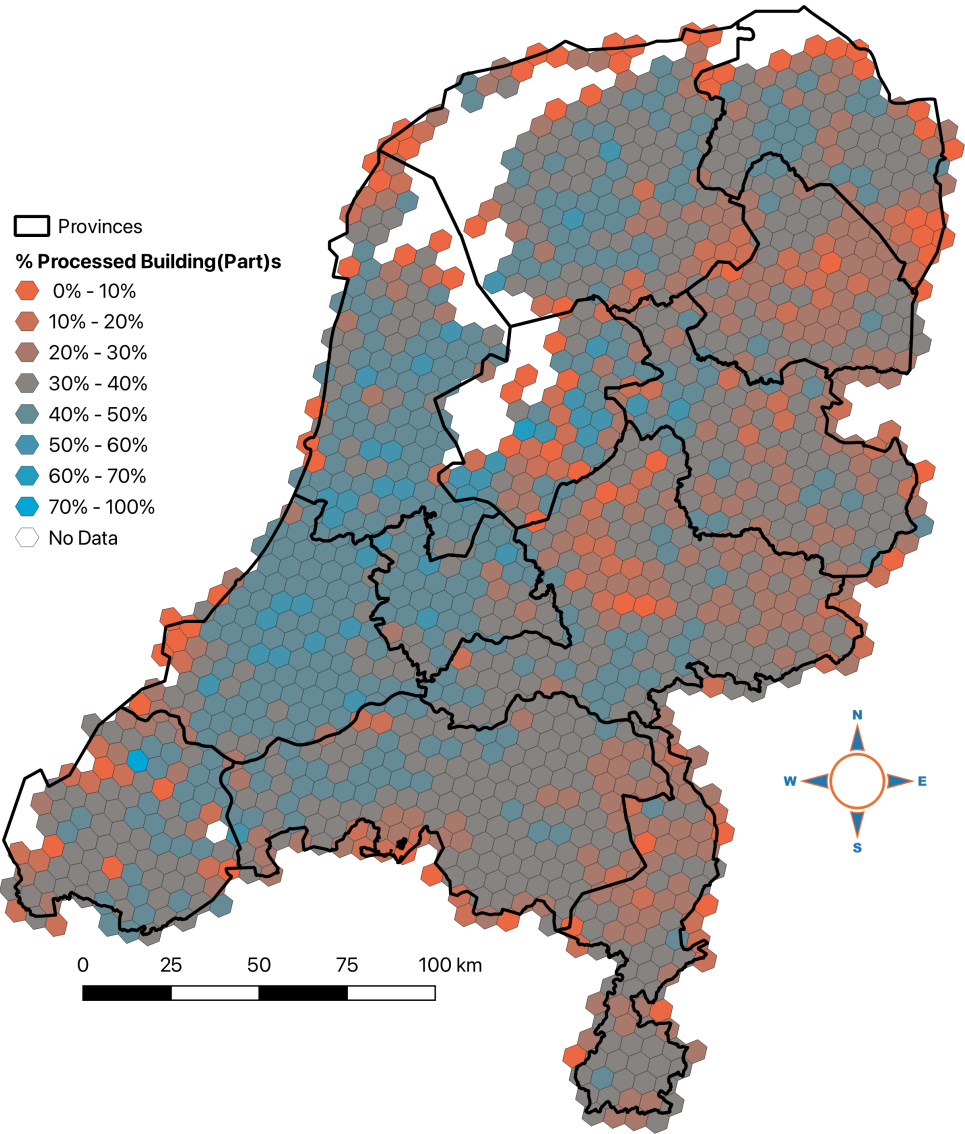


Figure 5.23: Spatial distribution of unprocessed Building(Part)s in the Netherlands

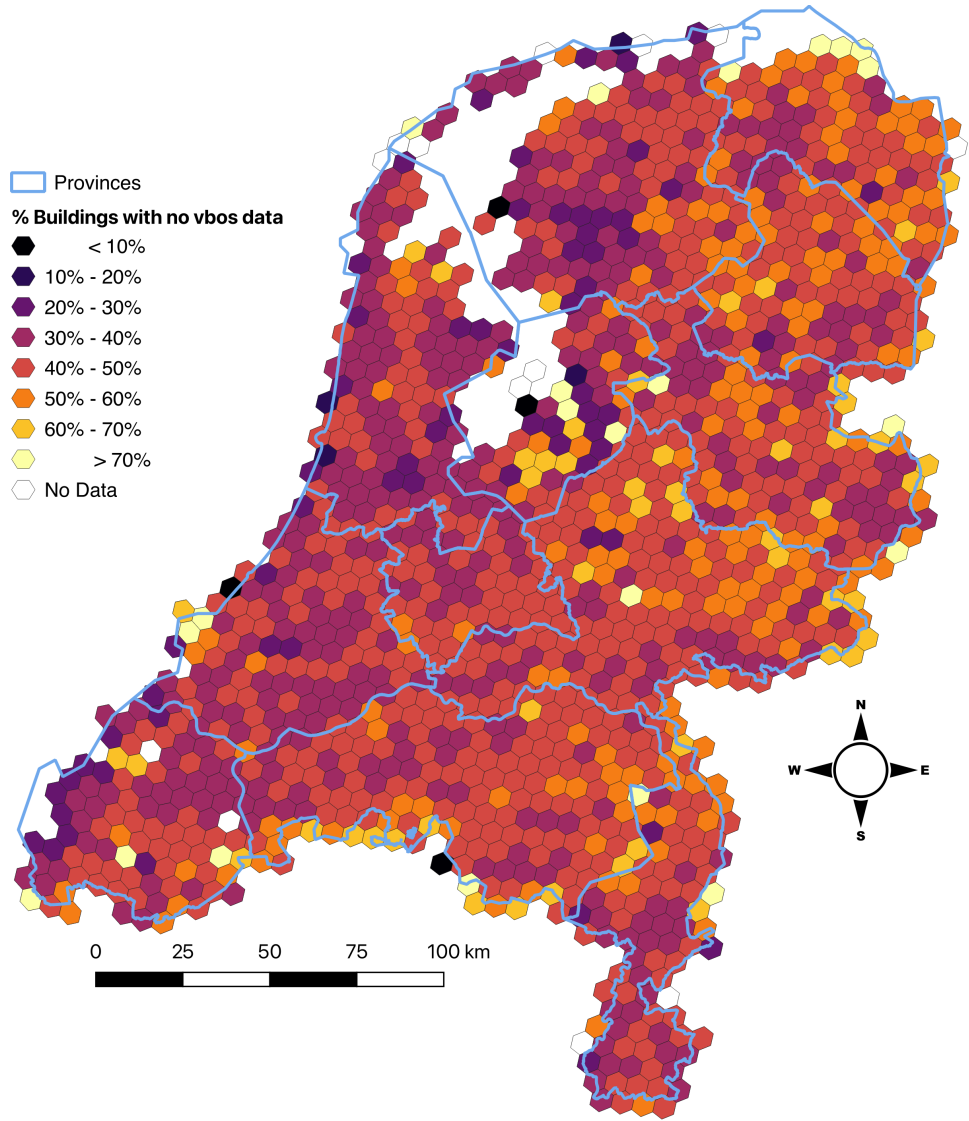


Figure 5.24: Spatial distribution of Building(Part)s with no vbos data in the Netherlands

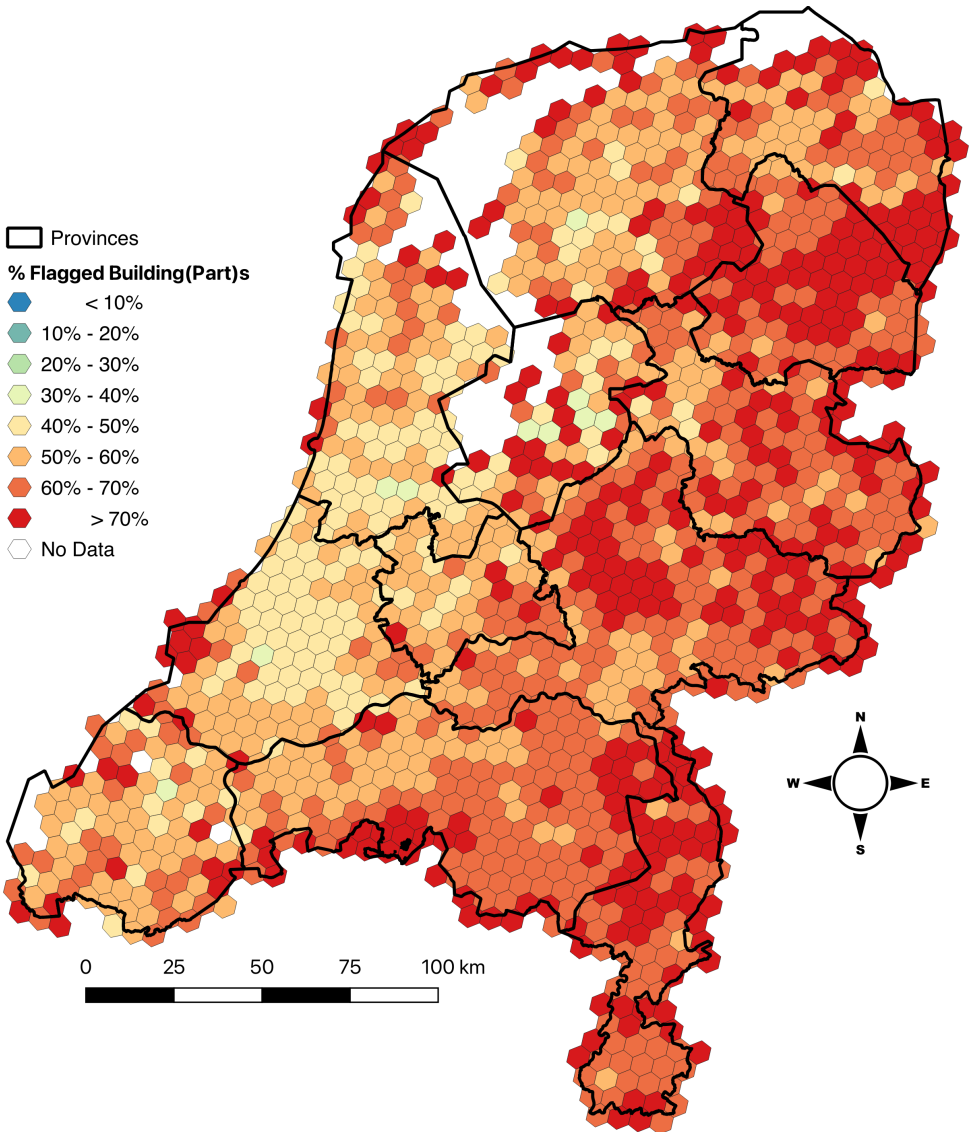


Figure 5.25: Spatial distribution of flagged Building(Part)s in the Netherlands

5.7. CONCLUSION REMARKS

This chapter documents the decisions made during the implementation of the EnergyBAG BES in alignment with the NTA 8800. As outlined in [chapters 3](#) and [4](#), UBEM requires extensive input data. However, not all data requirements could be met using available open datasets, necessitating strategic decisions to

address these gaps.

During the implementation, the need for robust data validation became clear to ensure consistency across datasets. While the Netherlands exemplifies strong open data practices—such as the use of unique identifiers to facilitate cross-dataset integration—opportunities for improvement remain. These include resolving inconsistencies in semantic classifications and enhancing metadata completeness, as shown in [figures 5.23 to 5.25](#).

The simulation results are presented in [chapter 6](#), which are separated for the testbed and the Netherlands. The analyses of the results are based on light of what have been discussed in [chapter 4](#) and this chapter.

6

RESULTS AND DISCUSSION

In this chapter, I present and discuss the results of the implemented EnergyBAG BES. The following sections provide a description of the results. The results are compared against ep-online values since this dataset offers country-wide building level energy open data.

This chapter is divided into two main sections: Rijssen-Holten ([section 6.1](#)) and the Netherlands ([section 6.2](#)). Both case studies follow the same structure. First, I describe the input dataset, including basic statistical analyses of the results. Then, each case study is analysed in the following subsections:

- Computed net heat demand: This section presents and discusses the computed heat demand results for Rijssen-Holten ([section 6.1.1](#)) and the Netherlands ([section 6.2.1](#)).
- Analysis of specific cases: Here, I examine buildings with extreme values, providing a description and discussing possible reasons for the calculated results. The corresponding sections are Rijssen-Holten ([section 6.1.3](#)) and the Netherlands ([section 6.2.2](#)).
- Results comparison: This section compares the computed results with the energy data available from the energy performance certificate database. The corresponding sections are Rijssen-Holten ([section 6.1.4](#)) and the Netherlands ([section 6.2.3](#)).
- Thermal hull refurbishment scenario: This section evaluates a refurbishment scenario and its impact on heat demand. The corresponding sections are Rijssen-Holten ([section 6.1.5](#)) and the Netherlands ([section 6.2.4](#)).

The chapter concludes with final remarks in [section 6.3](#). The computer used for all calculations has the following specifications:

- CPU: AMD Ryzen 9 7950X (16 cores @ 4,5GHz)
- GPU: Inno3D GeForce RTX 4070 Twin X2 (12GB GDDR6X)

- RAM: 128GB DDR5 @ 5.600MT/s
- Storage: 1TB M.2 @ 3.500MB/s

Although this is a high-end personal desktop computer, it is not a mainframe or supercomputer. Despite not having cutting-edge hardware (by 2025 standards), the results demonstrate the practicality of the designed pipeline and implementation.

6.1. CASE STUDY: RIJSSSEN-HOLTEN

This section presents the computed values for Rijssen-Holten extracting the data from the EnergyBAG DB chapter 4 and the computed model based on the NTA 8800 detailed in section 2.5, and implemented in chapter 5.

6.1.1. PROCESSED DATA

For Rijssen-Holten, the results of the usable area check indicate that 5,053 Building(Part)s (20.94% of the building stock) are not computed due to discrepancies higher than the 30% rule as explained in the previous chapter (i.e. when the official usable area and the calculated usable area differ more than 30%). Therefore, only 9,430 Building(Part)s (39.09%) are processed. Figure 6.1 shows the distribution of the building stock in Rijssen-Holten for computing the net heat demand based on the NTA 8800 and it is detailed in table 6.1.

6

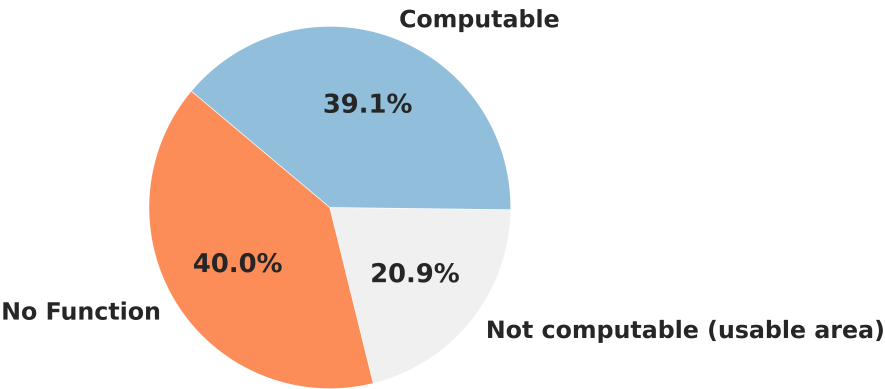


Figure 6.1: Processing distribution of the building stock in Rijssen-Holten

The basic statistics for the computed annual net heat demand per m² are shown in table 6.2. The values in this section, tables, and figures correspond to the refurbishment scenario “as-built”, which corresponds to the physical conditions

of the archetype building during the construction period and as if it was never refurbished.

Table 6.1: Rijssen-Holten EnergyBAG dataset

Building(Part)s		
Full dataset	24,170	
Usable area discrepancy ¹	5,138	21.26%
Non-processed (no vbos) ²	9,544	39.49%
Processed ³	9,430	39.26%
Residential ⁴	8,054	33.32%

¹ Usable area discrepancy indicates differences beyond tolerance thresholds.

² A *Building(Part)* without vbos lacks function data, requiring assumptions for computations.

³ Includes both residential and non-residential functions.

⁴ Refers to *Building(Part)s* containing at least one residential vbo.

From [table 6.1](#), the processed *Building(Part)s* that are further classified as residential, indicate that the *Building(Part)* has at least one vbo as residential. I evaluate the influence of the *PartyWalls* in the calculation of the net heat demand because, for Rijssen-Holten, I have this information (Agugiaro, Zwamborn *et al.*, 2022). This is useful to get a feeling of what it will be for the Netherlands, since we do not have detailed *PartyWalls* information for all buildings in the Netherlands. The basic statistics of the results for Rijssen-Holten are shown in [table 6.2](#); this table is complemented by [figure 6.2](#). These figures present the heatmaps of the mean values of the results classified by building type and construction period. Except for the single family house class (as expected), all other classes are influenced by the inclusion of the *PartyWalls* in the calculation of the net heat demand of the building. This influence is higher for older construction periods, with the highest difference between classes for apartment blocks constructed in the period 1965-1974 with an average of 99.9 kWh/(m²·a), almost 50% lower when including this parameter into the calculations.

Table 6.2: Rijssen-Holten computed annual net heat demand basic statistics. All values are expressed in kWh/(m²·a).

	Minimum	Mean	Median	Std Deviation	Maximum	Mode
Processed	28.8	304.2	303.2	158.3	1835.5	144.2
Residential						
<i>PartyWalls</i> as exterior walls	21.2	252.8	251.0	126.7	1478.5	284.1
<i>PartyWalls</i> as abadiatic walls	1.9	196.0	208.8	100.3	734.0	172.4

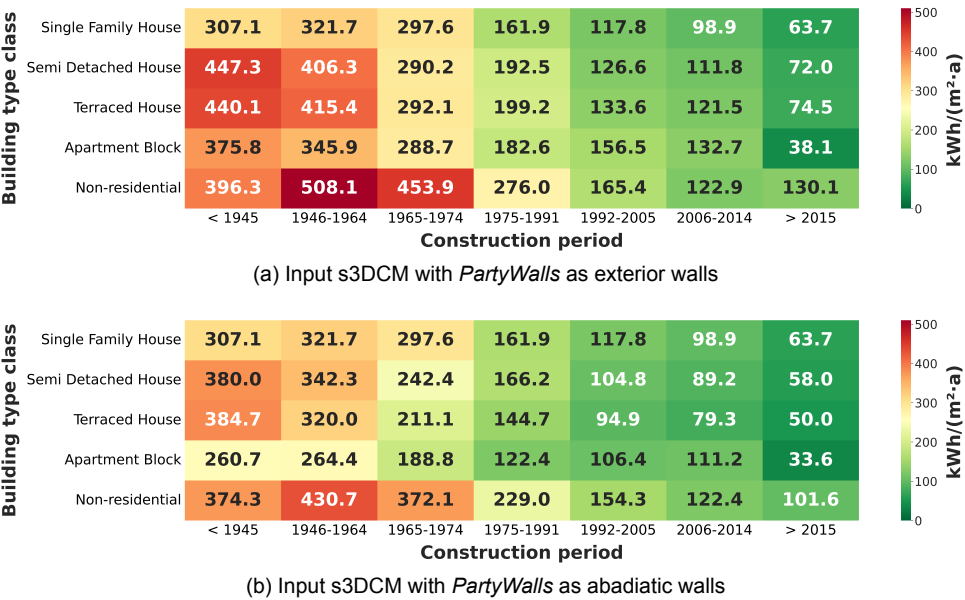


Figure 6.2: Annual mean net heat demand per building type and construction period classes in Rijssen-Holten

6

6.1.2. COMPUTED NET HEAT DEMAND

The distribution of the computed net heat demand by building type (refer to figure 5.8) and construction period (refer to table 3.5) is shown in figure 6.3 using a letter-value plot. The results indicate high variability, particularly for non-residential building(Part)s, with outliers exceeding 1,400 kWh/(m²·a) in the 1946–1964 construction period. Non-residential buildings encompass a wide range of functions, each requiring different assumptions about heating, ventilation, and operational hours. Therefore, these results must be taken with particular care.

Although I compute also non-residential buildings, I focus on *Building(Part)s* classified as residential in table 6.1 (residential and mix-use) Restricting the dataset of analysis to the residential buildings reduces the uncertainties by the additional assumptions mentioned above. The computed net heat demand of Rijssen-Holten is shown in figure 6.4, values are aggregated by building type and construction period classes. The figure is further complemented by tables 6.3 and 6.4, which provide the mean net heat demand by building type and construction period, respectively.

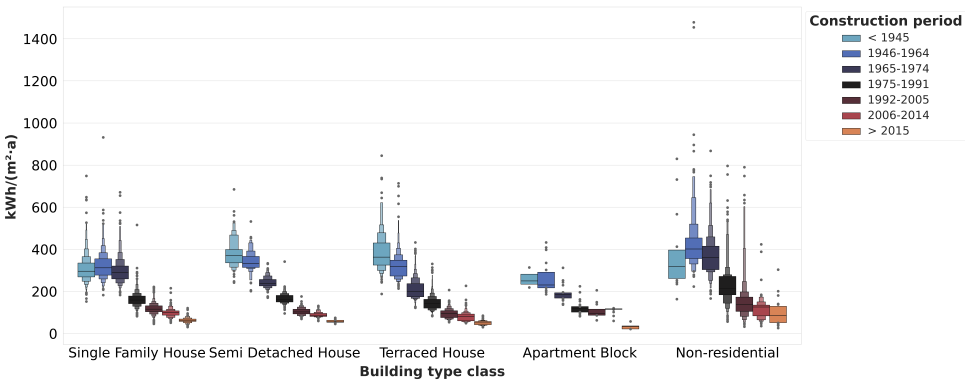


Figure 6.3: Letter-value plot of the computed net heat demand by building type and construction period in Rijssen-Holten

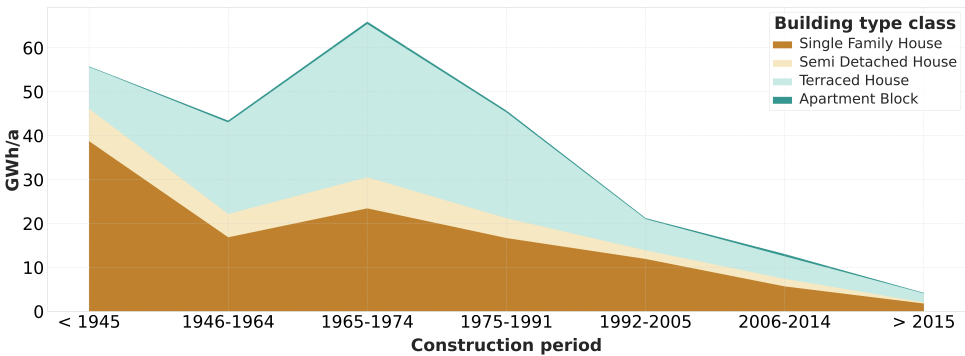


Figure 6.4: Computed total net heat demand in Rijssen-Holten aggregated by building type classes and construction periods

Table 6.3: Mean annual net heat demand per building type class

Building type class	No. Building(Part)s	Building stock %	kWh/(m ² ·a)
Single Family House	2476	30.74%	266.2
Semi Detached House	946	11.75%	266.3
Terraced House	4523	56.16%	215.2
Apartment Block	109	1.35%	179

Table 6.4: Mean annual net heat demand per construction period

Construction period	No. Building(Part)s	Building stock %	kWh/(m ² ·a)
< 1945	1,004	12.5%	389.4
1946–1964	1,063	13.2%	389.4
1965–1974	1,959	24.3%	276.9
1975–1991	1,935	24.0%	176.3
1992–2005	988	12.3%	120.7
2006–2014	688	8.5%	100.5
> 2015	417	5.2%	64.6

Among all construction periods in Rijssen-Holten, buildings constructed before 1964 show the highest annual net heat demand of 389.4 kWh/(m²·a). This category aggregates two classes that account for 25.7% of the local building stock, thereby indicating a pronounced effect on the aggregated results. As expected, newer construction periods show lower net heat demand values, which could indicate the use of modern building materials with enhanced thermal properties.

Using figures 6.5 to 6.9, I conducted an initial evaluation of the results. Figure 6.5 shows the distribution of net heat demand across all residential building type and construction period classes using a letter-value (boxen plot) graph. Buildings constructed before 1945 have the highest mean net heat demand in most building type classes.

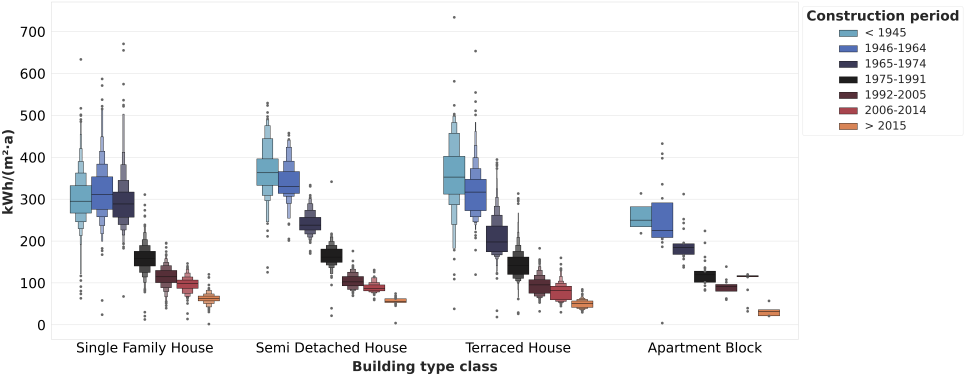


Figure 6.5: Distribution of the computed mean net heat demand in Rijssen-Holten by building type and construction period

Semi Detached Houses and *Single Family Houses* show the highest mean values, on the opposite *Terraced Houses* and *Apartment Blocks* show lower mean values

across all construction periods. For these two building type classes, buildings constructed between 1946 and 1974 show reduced mean heat demand compared to pre-1945 buildings. Post-1975 construction periods reveal a notable decline in annual heat demand variability across all building types, with further reductions in variability observed in newer periods. Buildings constructed after 2015 display minimal value ranges and few outliers.

Across all periods, *Semi Detached Houses* consistently record the highest median heat demand, while, *Apartment Blocks* present the lowest means and reduced variability compared to other building types. Outliers are more prevalent in older periods (< 1945 and 1946–1964 classes), particularly among terraced houses, which frequently show anomalously high values.

The mean net heat demand by building type and construction period classes is shown in [figures 6.6](#) and [6.7](#). In these figures, *Single Family Houses* and *Semi Detached Houses* indicate the higher net heat demand across all construction periods, with the < 1945 period having the peak value (368.2 kWh/(m²·a)). As expected, the computed net heat demand decreases significantly for more recent construction periods. Furthermore, in [figure 6.7](#), the error bar of apartment blocks in the period 1945–1964 indicates a very high variability of the computed net heat demand.

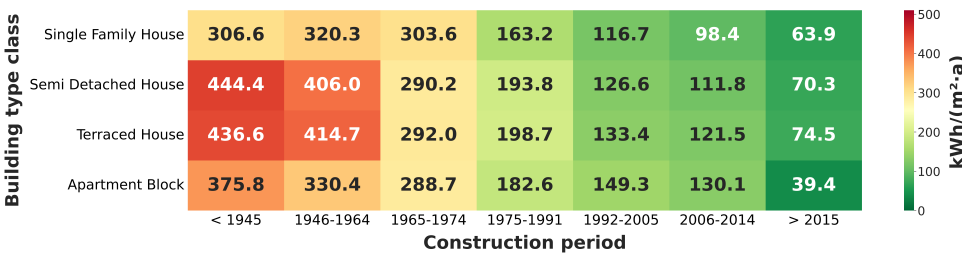


Figure 6.6: Heatmap of the computed mean net heat demand by building type class and construction period in Rijssen-Holten

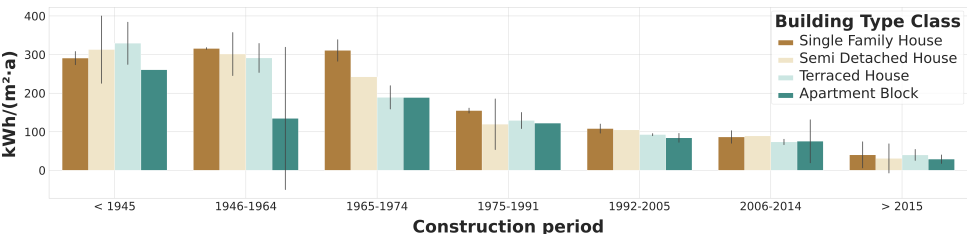


Figure 6.7: Distribution of computed annual net heat demand across building types and construction periods in Rijssen-Holten, including their associated standard deviation (error bars) for variability

Additional insights into the computed net heat demand across building type classes are provided by [Figures 6.8 and 6.9](#). The scatter plot in [Figure 6.8](#) relates usable area (m^2) to net heat demand ($\text{kWh}/(\text{m}^2 \cdot \text{a})$), indicating a high concentration of smaller buildings with relatively low heat demand. In addition, several *Semi Detached* and *Terraced Houses* show unusually high demand values given their size, some examples of high values are presented in [section 6.1.3](#).

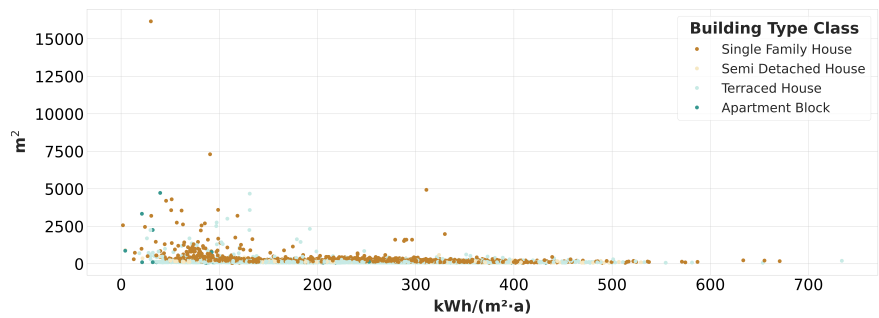


Figure 6.8: Distribution of the mean computed annual net heat demand and the building usable area per building type in Rijssen-Holten

6

Most Building(Part)s have computed net heat demand values between 100 and 500 $\text{kWh}/(\text{m}^2 \cdot \text{a})$ as shown in [Figure 6.9](#). *Terraced house* buildings(Part)s have the highest frequency, with a sharp density peak around 180 $\text{kWh}/(\text{m}^2 \cdot \text{a})$. *Single Family House* class shows two main peaks around 120 and 300 $\text{kWh}/(\text{m}^2 \cdot \text{a})$.

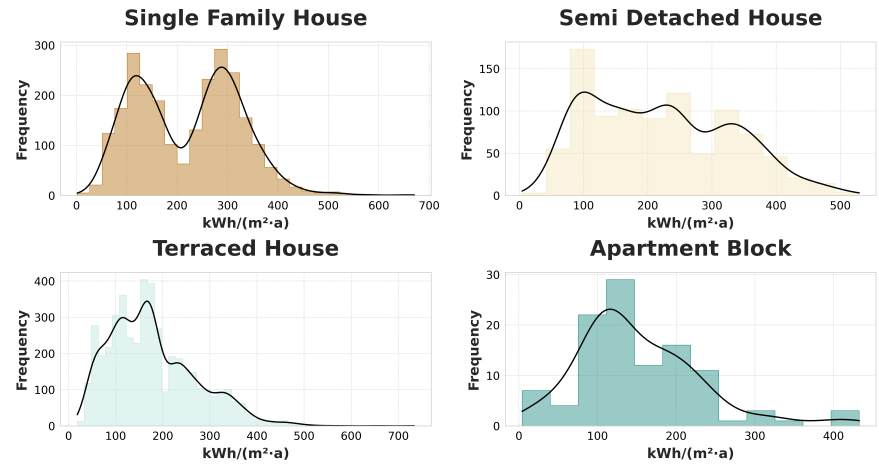


Figure 6.9: Histogram of the computed annual net heat demand by building type class in Rijssen-Holten

The spatial distribution of the computed net heat demand for Rijssen and Holten

is shown in figures 6.10 and 6.11 respectively. Building(Part)s in gray indicate they were not processed, either due to missing function data or being flagged for discrepancies in the usable area data. In Rijssen-Holten, several non-processed buildings are located within urban blocks without direct street access with small footprint size. These locations suggest they are likely barns or storage rooms, which are typically unheated and used for storage purposes.



Figure 6.10: Spatial distribution of the computed net heat demand in Rijssen



Figure 6.11: Spatial distribution of the computed net heat demand in Holten

6.1.3. ANALYSIS OF SPECIFIC CASES

This section discusses cases with extreme high values to explore the underlying causes of these outcomes. I analyse the buildings with the extreme values of the computed annual net heat demand per square meter. Finally, I assess potential outliers to determine whether they represent genuine peculiarities or methodological artefacts.

6.1.3.1. PAND ID 1742100000008646

This mix-used building, has a computed annual net heat demand of 734kWh/(m²·a). The basic data of the building is shown in table 6.5. BAG data indicates that it has one vbo with two functions, residential and industrial. However, the data available are not explicit about the corresponding area per function (as shown in figure 6.12); this situation leads to errors since the only possible standardised solution is to divide the total usable area of the vbo by the number of functions registered to it (as I did in my PhD research). This building is located in the rural area of Rijssen-Holten, where several agricultural businesses are located, as can be seen from the 3DBAG viewer in figure 6.13.

Table 6.5: Summary data for Pand ID 1742100000008646

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
312.4m ²	1	18.96m	1850.1m ³	1	201m ²

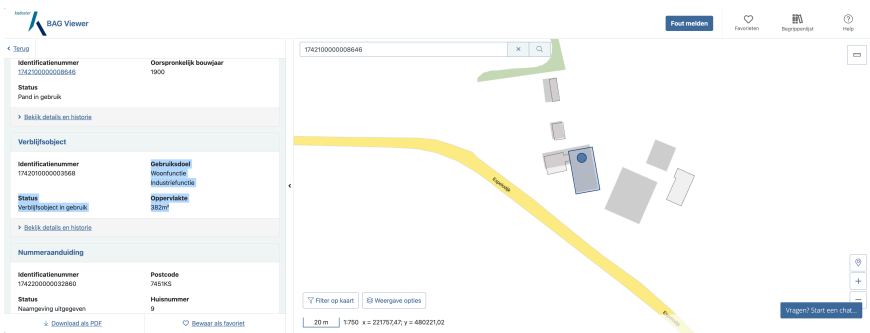


Figure 6.12: BAG viewer data for building with Pand ID 0150100000010860

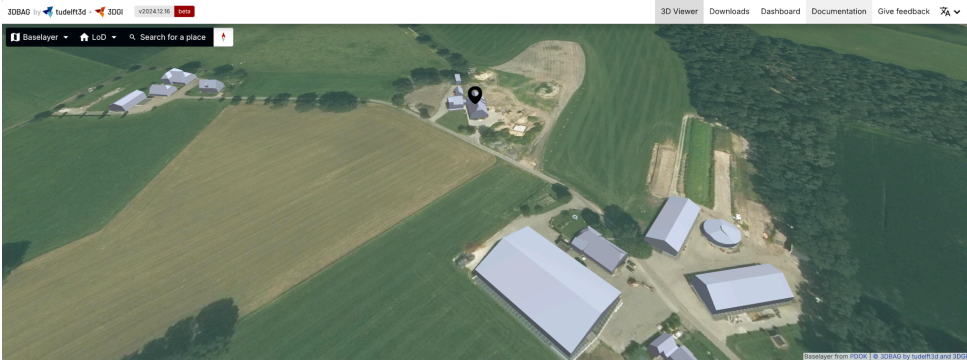


Figure 6.13: 3DBAG viewer for building with Pand ID 0150100000010860

6.1.3.2. PAND ID 1742100000014961

This mix-used building, has a computed annual net heat demand of $670.7\text{kWh}/(\text{m}^2 \cdot \text{a})$. The basic data of the building is shown in [table 6.6](#). This building has two vbos, one is residential, the other vbo has two function registered: residential and industrial (as shown in [figure 6.14](#)). As previously mentioned, it is not explicit the area of the multi-function vbo. Also, it is located in the rural area of Rijssen-Holten as shown in [figure 6.15](#).

6

Table 6.6: Summary data for Pand ID 1742100000014961

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
304.1m ²	1	18.61m	1613.3m ³	2	177m ²

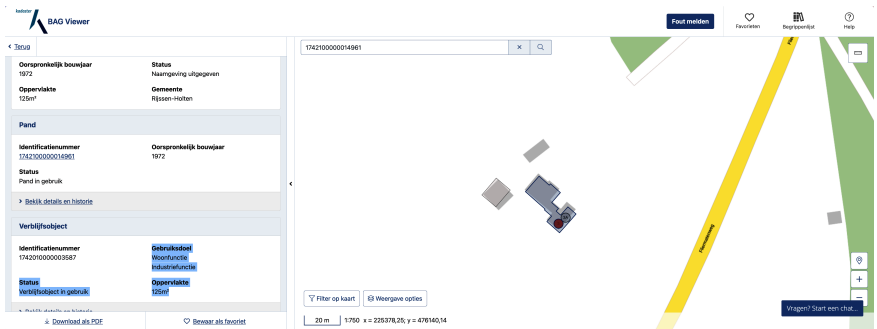


Figure 6.14: BAG viewer data for building with Pand ID 1742100000014961



Figure 6.15: 3DBAG viewer for building with Pand ID 1742100000014961

6.1.3.3. **PAND ID 1742100000010671**

This mix-used building, has a computed annual net heat demand of 655.2kWh/(m²·a). The basic data of the building is shown in table 6.7. This building has the same characteristics as the building with Pand ID 1742100000014961: two vbos, one residential and the other vbo has two funcions registered, residential and industrial (as shown in figure 6.16). This building is also located in the rural area of Rijssen-Holten as shown in figure 6.17.

Table 6.7: Summary data for Pand ID 1742100000010671

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
386.1m ²	1	24.57m	1909m ³	2	217m ²

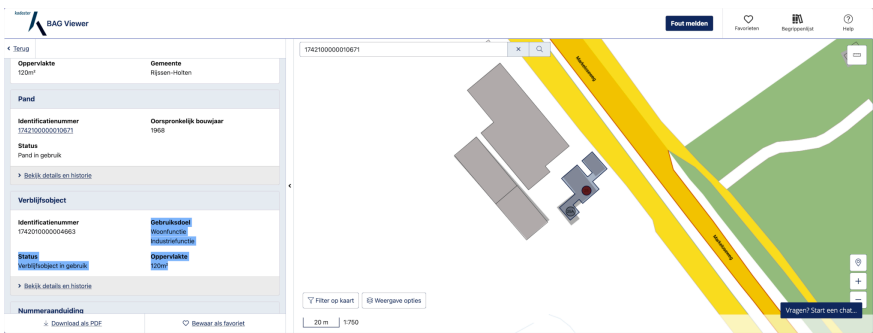


Figure 6.16: BAG viewer data for building with Pand ID 1742100000010671

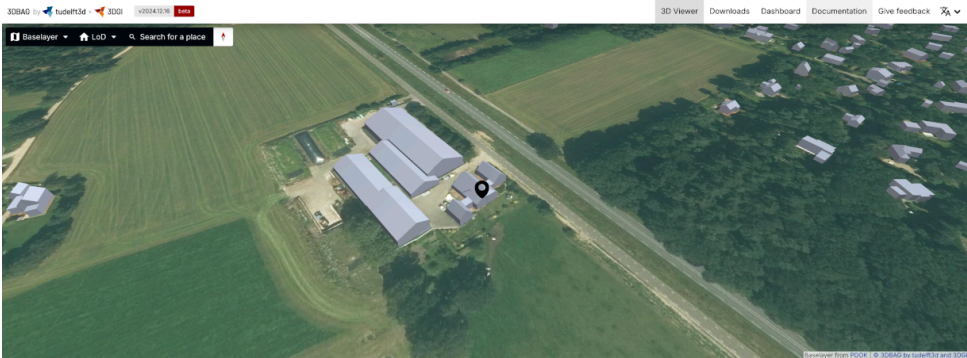


Figure 6.17: 3DBAG viewer for building with Pand ID 1742100000010671

6.1.3.4. PAND ID 1742100000096511

This mix-used building, has a computed annual net heat demand of $1.9\text{kWh}/(\text{m}^2 \cdot \text{a})$. The basic data of the building is shown in [table 6.7](#). This building has the same characteristics as the building with Pand ID 1742100000010671: two vbos, one residential and the other industrial (as shown in [figure 6.18](#)). This building is also located in the rural area of Rijssen-Holten as shown in [figure 6.19](#).

6

Table 6.8: Summary data for Pand ID 1742100000096511

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
1185.5m^2	3	21.13m	12165.4m^3	2	2571m^2

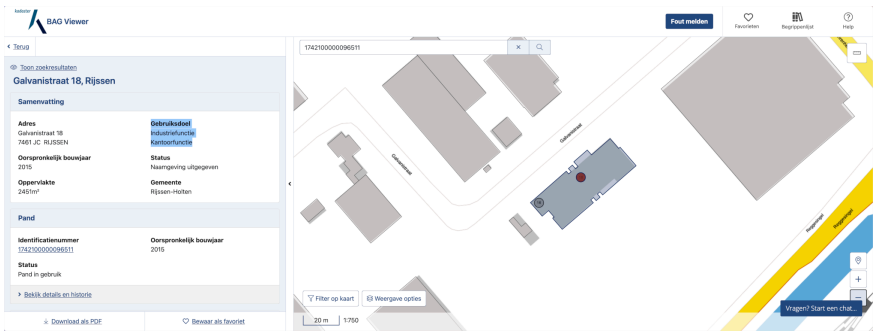


Figure 6.18: BAG viewer data for building with Pand ID 1742100000096511



Figure 6.19: 3DBAG viewer for building with Pand ID 1742100000096511

6

The examples presented in this section prompted a review of the input BAG dataset. The analysis revealed that 273,118 out of 10,234,487 vbos (2.6%) have multiple registered functions. The inability to determine the usable area distribution across these functions introduces uncertainty, as the allocation of this attribute could be inaccurate. To mitigate this data gap, the total usable area was divided equally among the registered functions; for example, a vbo with two functions assigned 50% of the total area to each. While pragmatic, this approach risks errors in cases of imbalanced distributions (e.g., 30%–70% splits or greater).

Additionally, buildings classified as semi-detached or terraced houses may be misrepresented. Although these structures are adjacent to neighbouring buildings, the physical connections are often minimal (e.g., shared walls at limited points). Current classification rules account for adjacency (binary presence/absence) and not the extent of shared surfaces.

6.1.3.5. PAND ID 1742100000000155

This mix-used building has a calculated annual net head demand of **30.3kWh/(m²·a)** for its residential zone. The basic data of the building is shown in [table 6.7](#). I checked this building due to its big usable area since it looks like an outlier in [figure 6.8](#). However, this value seems to be correct based on the information available from the BAG viewer [figure 6.20](#).

This building, constructed in 1982, has 75 vbos registered with several functions, including residential and business. Additionally, none of the vbos has multiple functions registered to them. Although, most the facts seems correct, the results are very low compared to the mean values presented in [figures 6.6](#) and [6.7](#).

Table 6.9: Summary data for Pand ID 1742100000000155

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
6714.7m ²	3	23.16m	57804.4m ³	75	16168m ²

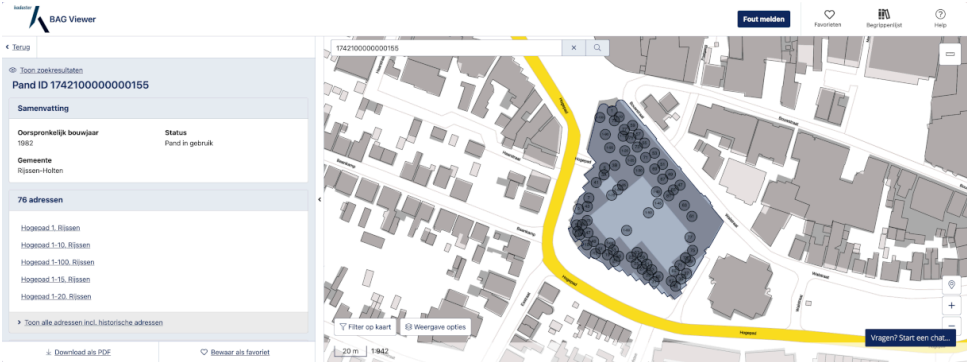


Figure 6.20: BAG viewer data for building with Pand ID 1742100000000155

6.1.4. RESULTS COMPARISON

Ep-online is the only open energy-related data source at the building level in the Netherlands, making validation of the computed net heat demand challenging, because this dataset contains the *primary* energy demand per vbo and not the *net* heat demand, which is the value I calculated. Therefore, a direct comparison between the *primary* energy demand and the *net* heat demand is not correct. Although it can be expected that "*ceteris paribus*" a vbo will have a value with *primary* energy demand bigger than the *net* demand. I perform this comparison to assess my results using the official open data values in the Netherlands.

Since ep-online stores information per vbo, the comparison can include only those buildings where the number of vbos matches the ones in BAG. Additionally, I keep for the analysis only those buildings in which the usable area value from the two datasets is within 30%, to align with [section 5.6.6](#). From the processed 9,430 *Building(Part)s*, 1,661 (17.6%) match both criteria. I compare the percentage difference between the computed net heat demand per building and the values available in ep-online, using the ep-online data as the ground truth following [equation \(6.1\)](#).

$$Difference = \left(\frac{Computed_{nHD} - ep-online_{pED}}{ep-online_{pED}} \right) \cdot 100 \quad (6.1)$$

where:

- nHD: *net* heat demand
- pED: *primary* energy demand

In theory, the difference should be negative, as it results from subtracting the *primary* energy demand from the *net* heat demand. The distribution of these differences is shown in figure 6.21 and detailed in table 6.10. The pie chart categorises the differences by percentage, with each slice representing the proportion of buildings within a given class.

As can be seen from figure 6.21 and table 6.10, the results are rather different from what one would expect. The percentage difference classes in figure 6.21 indicate that my implementation generally overestimates the net heat demand compared to ep-online, with 77.28% of the buildings showing higher values. A more detailed breakdown is presented in figure 6.22, which shows the mean percentage difference by building type and construction period. Except for apartment blocks, the net heat demand for buildings constructed before 1991 is higher. One of the reasons is that my implementation does not account for renovations or refurbishment of the buildings, which typically reduces heat demand. Instead, it is based on "as-built" conditions. Consequently, the computed values tend to be higher than those recorded in ep-online, which uses the conditions of the vbo at the time of performing the analysis. For this reason the results of the comparison must be interpreted with caution.

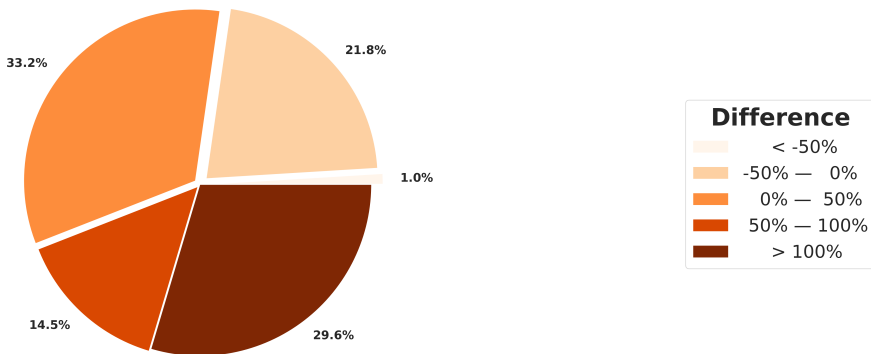


Figure 6.21: Distribution of *Building(Part)s* based on the percentage difference between ep-online values and computed net heat demand

Table 6.10: Distribution of *Building(Part)s* based on the percentage difference between ep-online values and computed net heat demand

Difference	No. Buildings
< - 50%	14
-50% – 0%	313
0% – 50%	478
50% – 100%	208
>100%	426

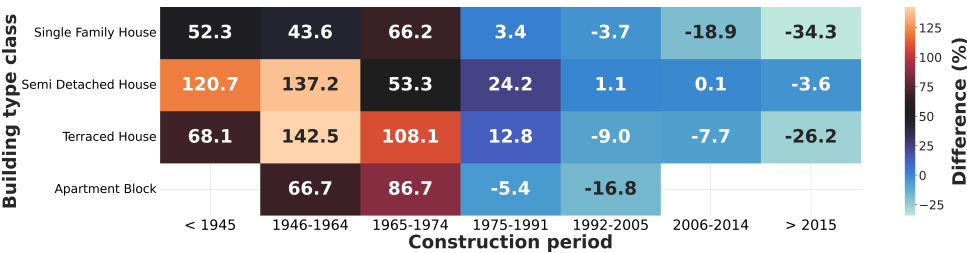


Figure 6.22: Variation in computed net heat demand to ep-online reference values by building class and construction period

Nevertheless, as shown in [figure 6.23](#), newer buildings have a lower computed net heat demand compared to the ep-online data. This behaviour which is more in line with the expected values, also indicates that, as expected, the building archetypes of TABULA better represent newer buildings than old buildings if no refurbishment information is included in the analysis.

[Figure 6.24](#) shows the distribution of percentage differences between the computed net heat demand and ep-online data, grouped by building type. As previously shown, *Single Family Houses* have the highest number of buildings with lower computed net heat demand values compared to ep-online. This class shows as well a high variability with percentage difference values from -100% to 250% values. *Semi Detached Houses* and *Terraced Houses* show a large range, indicating a higher variability in the physical characteristics or renovation status for these building type classes. In conclusion, these histograms reinforce the need for refined specification of buildings characteristics, for example, their refurbishment status.

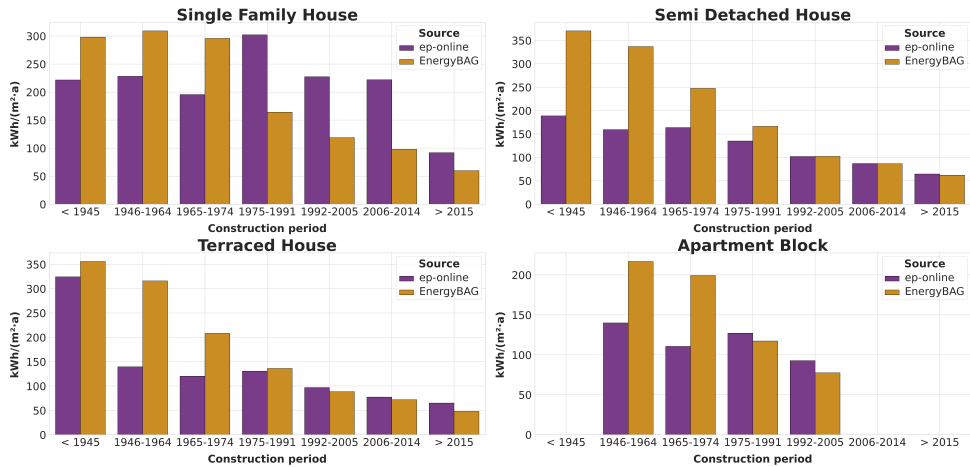


Figure 6.23: Comparison of the computed net heat demand to ep-online reference values by building type and construction period

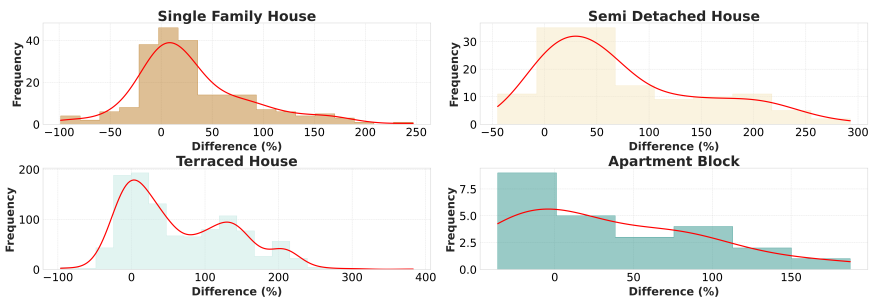


Figure 6.24: Distribution of the variation of the computed net heat demand to ep-online reference values by building type

6.1.5. THERMAL HULL REFURBISHMENT SCENARIO

Another test to the EnergyBAG BES is the execution of a fictitious renovation scenario. For this case, all buildings are treated as renovated to the physical characteristics of building archetypes built after 2015 following a net zero energy building (NZEB) approach. Note that this, it is just a hypothetical case that we use to evaluate the approach. The basic statistical values of the results of the refurbishment scenario are presented in [table 6.11](#) and further classified by building type ([table 6.12](#)) and construction period ([table 6.13](#)). The classified computation results according to the building type and construction period classes are shown in [figures 6.25](#) and [6.26](#) using a heatmap and a letter-value plot graph respectively.

Table 6.11: Refurbishment scenario computed annual net heat demand basic statistics in Rijssen-Holten, all values are expressed in kWh/(m²·a).

Minimum	Mean	Median	Std Deviation	Maximum	Mode
1.7	46.2	45.75	11.4	129.7	35.7

Table 6.12: Refurbishment scenario computed mean annual net heat demand per building type class in Rijssen-Holten

Building type class	No. Building(Part)s	Building stock %	kWh/(m ² ·a)
Single Family House	2476	30.74%	52.9
Semi Detached House	946	11.75%	47.8
Terraced House	4523	56.16%	42.3
Apartment Block	109	1.35%	45.8

Table 6.13: Refurbishment scenario computed mean annual net heat demand per construction period in Rijssen-Holten

Construction period	No. Building(Part)s	Building stock %	kWh/(m ² ·a)
< 1945	1,004	12.5%	51.5
1946–1964	1,063	13.2%	48.4
1965–1974	1,959	24.3%	43.1
1975–1991	1,935	24.0%	45.3
1992–2005	988	12.3%	46.7
2006–2014	688	8.5%	45.0
> 2015	417	5.2%	48.1

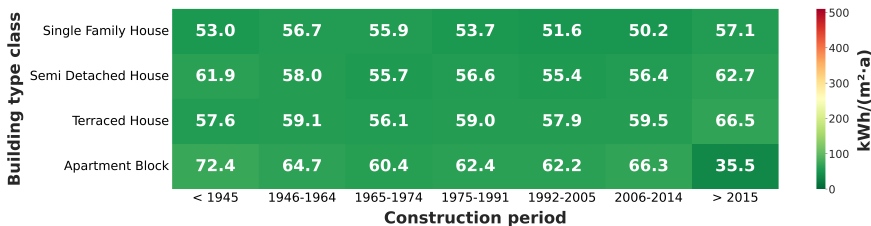


Figure 6.25: Heatmap of the computed mean net heat demand by building type class and construction period in Rijssen-Holten for the refurbishment scenario

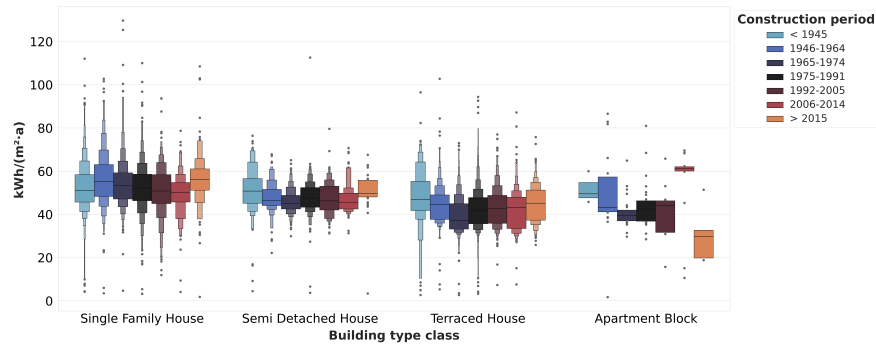


Figure 6.26: Letter-value plot of the computed net heat demand by building type and construction period in the Netherlands for the refurbishment scenario

The comparison of the computed net head demand values under the refurbishment scenario to ep-online values are presented in [figure 6.27](#). This comparison is classified according to the building type and construction period categories. For all building types, the computed net heat demand values are lower than ep-online values. However, the difference reduces significantly as the construction period is more recent which is reasonable. This is an expected behaviour since the current renovation status of each building in the ep-online database is better represented for the recently built buildings compared to older construction periods.

6

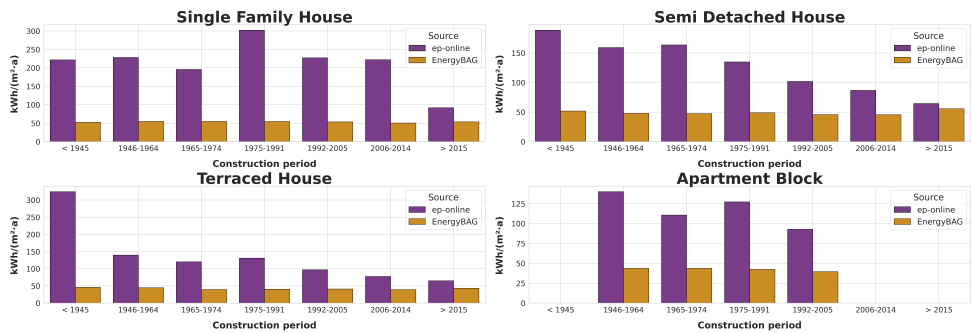


Figure 6.27: Comparison of computed net heat demand under the refurbishment scenario to ep-online reference values

The spatial distribution of the calculated net heat demand for the mentioned refurbishment scenario is shown in [figure 6.28](#).



Figure 6.28: Spatial distribution of the computed net heat demand in Rijssen for the refurbishment scenario

6.2. THE NETHERLANDS

As a final test of the implementation, in order to test the scalability and overall simulation time, I compute the net heat demand for the entire EnergyBAG database, covering the whole of the Netherlands. Therefore, I split the Netherlands into smaller, manageable areas using the H3 index at level 6. This subdivision results in 1,353 tiles ([table 4.7](#)). However, not all tiles are processable due to the exclusion of buildings for the circumstances mentioned previously: the lack of vbo data ([section 5.4](#)) or the discrepancies in the usable area between datasets ([section 5.6.6](#)).

The computation time for the 1353 tiles of the EnergyBAG DB is **05h 03**. However, processing time varies at each execution as stated in [section 5.1](#). The implementation was done in Python and followed a multithreaded approach in which a different thread treats each H3 index tile. Furthermore, Python does not optimize host resources and is limited in multiprocessing, resulting in threads executed in parallel in the same CPU's core. Nevertheless, when summing the execution time per H3 tile, the resulting time is **157,5h**. Despite these limitations, the results demonstrate the feasibility of performing country-wide analyses for the Netherlands in terms of needed computation time and required hardware specifications. As an indicator, [table 6.14](#) shows an excerpt of the execution time per H3 tile, showing the 20 slowest and 20 fastest tiles.

Table 6.14: Excerpt of the H3 tiles processed with their number of processed *Building(Part)s* and execution time

Tile Index	Number of Building(Part)s	Execution time
861969537fffff	29,777	1:30:26
86196b85fffff	35,127	1:28:53
86196b84fffff	28,843	1:16:52
861969527fffff	25,729	1:08:54
861968257fffff	29,866	1:08:12
861969497fffff	27,072	1:05:26
861969c9fffff	20,215	1:00:53
861969087fffff	23,605	0:56:05
86196b847fffff	14,455	0:55:08
86196bb17fffff	25,969	0:53:10
86196bb57fffff	22,229	0:52:44
86196dccfffff	23,025	0:50:39
861969697fffff	29,801	0:50:30
86196d85fffff	22,961	0:50:22
861969197fffff	19,290	0:48:48
86196ba27fffff	22,559	0:48:13
86196832fffff	24,690	0:47:55
86196822fffff	24,100	0:47:40
861969507fffff	19,975	0:46:48
861fa4847fffff	21,058	0:45:48
⋮	⋮	⋮
861fa5467fffff	6	0:00:05
861f164cfffff	13	0:00:05
86196b367fffff	25	0:00:05
861968b67fffff	48	0:00:05
861969a0fffff	5	0:00:04
861fa4a87fffff	2	0:00:04
861969137fffff	49	0:00:04
86196d78fffff	7	0:00:04
861968ba7fffff	6	0:00:04
86196c22fffff	71	0:00:04
86196d61fffff	12	0:00:03
861fa554fffff	6	0:00:03
86196dc9fffff	1	0:00:03
861f16457fffff	18	0:00:03
86196b147fffff	3	0:00:02
86196892fffff	4	0:00:02
861f14c07fffff	4	0:00:02
86196c35fffff	5	0:00:02
86196b167fffff	1	0:00:02
86196b2b7fffff	1	0:00:02

Basic statistics of the normalised timing per-tile are shown in [table 6.15](#)

and [figure 6.29](#). Additionally, the normalised timing per-building are shown in [table 6.15](#) and [figure 6.29](#)

Table 6.15: Basic statistics of the execution time per tile in seconds

Minimum	Mean	Median	Std Deviation	Maximum	Mode
1.94	426.1	226.2	600.5	5425.6	1.94

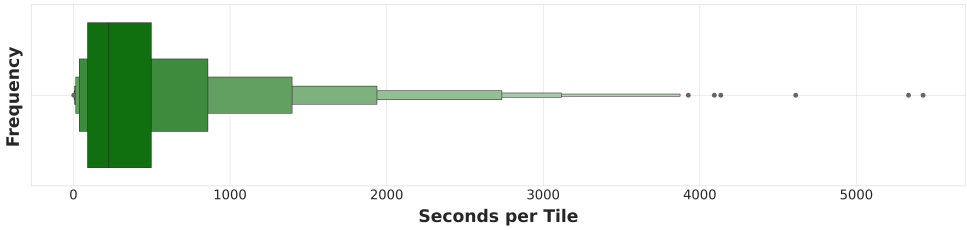


Figure 6.29: Distribution of the execution time per tile in seconds

Table 6.16: Basic statistics of the execution time per *Building(Part)s* in seconds

Minimum	Mean	Median	Std Deviation	Maximum	Mode
0.005	7.13	7.01	4.7	104.8	0.005

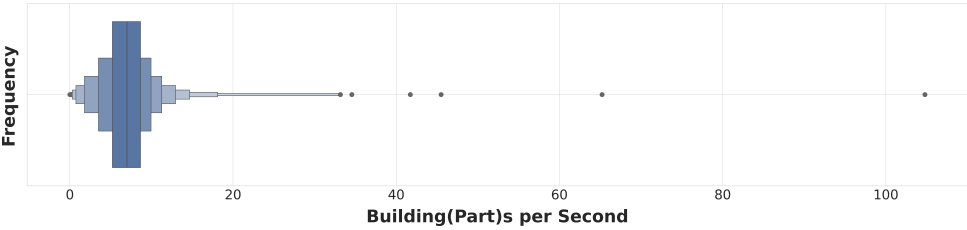


Figure 6.30: Distribution of the execution time per building in seconds

6.2.1. COMPUTED NET HEAT DEMAND

The computed net heat demand values for buildings in the Netherlands are summarised in [table 6.17](#), which provides basic statistical metrics. In addition, the computed annual net heat demand values for processed buildings in the Netherlands are shown in [figure 6.31](#).

For granular analysis, mean values classified by building type and construction period are available in [table 6.18](#) and [table 6.19](#), respectively, with additional insights provided by the bar plot in [figure 6.32](#). Finally, it is important to note that these values are higher because there are no data regarding shared walls in the EnergyBAG DB for the whole the Netherlands—a limitation discussed earlier—which also impacts the thermal performance of the *Building(Part)*.

Table 6.17: Computed annual net heat demand basic statistics in the Netherlands, all values are expressed in kWh/(m²·a)

Minimum	Mean	Median	Std Deviation	Maximum	Mode
0.7	260.9	238.6	125.5	2,461.1	194.9

Table 6.18: Mean annual net heat demand per building type class in the Netherlands

Building type class	No. Building(Part)s	Building stock %	kWh/(m ² ·a)
Single Family House	773,048	18.16%	222.4
Semi Detached House	670,009	15.74%	289.2
Terraced House	2,696,494	63.34%	265.8
Apartment Block	117,658	2.76%	241.5

Table 6.19: Mean annual net heat demand per construction period in the Netherlands

Construction period	No. Building(Part)s	Building stock %	kWh/(m ² ·a)
< 1945	797,752	19%	393.5
1946–1964	609,059	14%	405
1965–1974	736,005	17%	298.3
1975–1991	1,074,842	25%	193.5
1992–2005	594,259	14%	131.6
2006–2014	242,886	6%	121.3
> 2015	202,406	5%	73.8

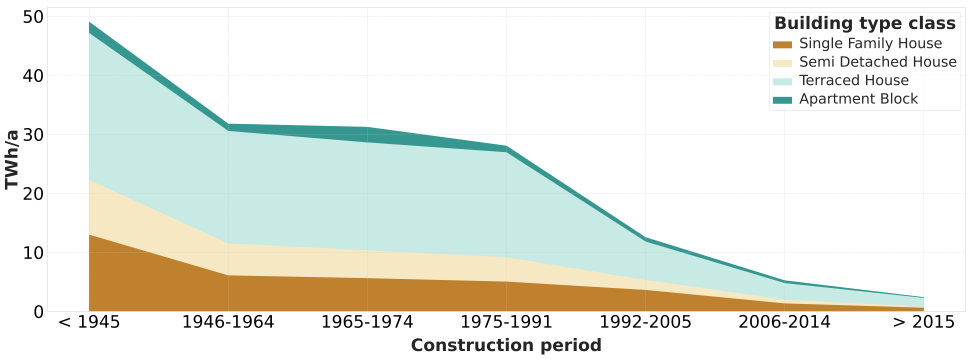


Figure 6.31: Computed annual net heat demand in the Netherlands aggregated by building type and construction periods

Based on figures 6.7 and 6.32, the Dutch building stock behaves differently compared to Rijssen-Holten. In the Netherlands, *Apartment Block* class shows the lowest net heat demand values across most periods (with the exception of pre-1945 constructions, where *Single Family House* class is lower), with mean values below $300 \text{ kWh}/(\text{m}^2 \cdot \text{a})$ for older constructions. In contrast, the *Terraced House* class consistently has the highest net heat demand values, with buildings constructed pre-1964 peaking at around $400 \text{ kWh}/(\text{m}^2 \cdot \text{a})$. Additionally, this class exhibits the highest variability on the computation results as shown in figure 6.33 particularly in buildings constructed before 1945.

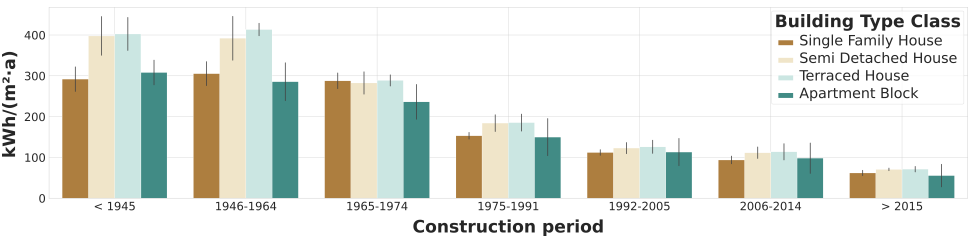


Figure 6.32: Distribution of computed annual net heat demand across building types and construction periods in the Netherlands, including their associated standard deviation (error bars) for variability

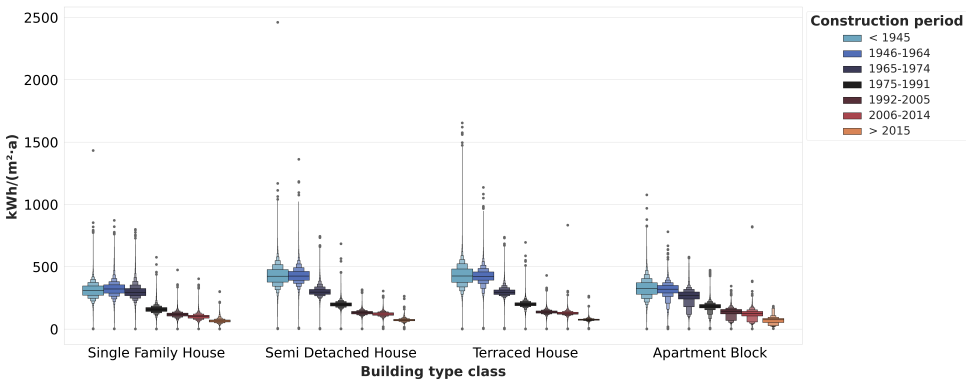


Figure 6.33: Letter-value plot of the computed net heat demand by building type and construction period in the Netherlands

The frequency distribution of the computed annual net heat demand grouped by building type in the Netherlands is shown in [figure 6.34](#). *Single family houses* show two distinct peaks, one centred around 150 kWh/(m²·a) and another at approximately 300 kWh/(m²·a), which is in line with the results of Rijssen-Holten ([Figure 6.9](#)). *Semi detached houses* indicate a broader range of computed values, with the two highest peaks near 170 and above 200 kWh/(m²·a). *Terraced houses* present multiple peaks, most notably around 200 kWh/(m²·a). Apartment blocks have the lowest computed net heat demand, with a central peak near 1900 kWh/(m²·a).

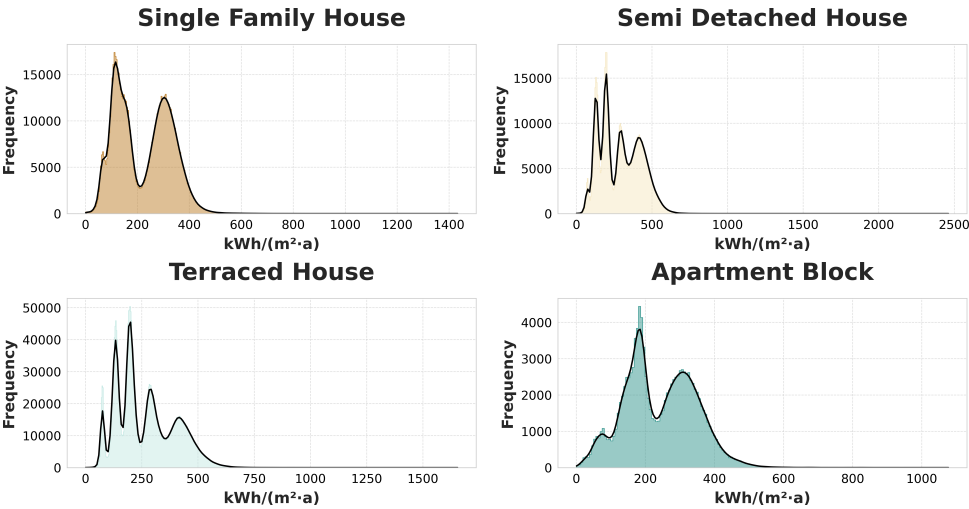


Figure 6.34: Distribution of computed annual net heat demand across building types and construction periods in the Netherlands, including their associated standard deviation (error bars) for variability

Finally, [figure 6.35](#) shows a heatmap of the computed mean net heat demand by building type and construction period in the Netherlands. Compared to Rijssen-Holten, *Semi Detached Houses* and *Terraced Houses* exhibit the highest discrepancy with higher values for the Dutch building stock. Additionally, the newest construction period exhibits higher values in the Netherlands and in the case of *Apartment Blocks*, the values of the Netherlands are double the mean value of Rijssen-Holten. However, this building type hardly occurs in Rijssen-Holten which indicates that it is not representative in this municipality.

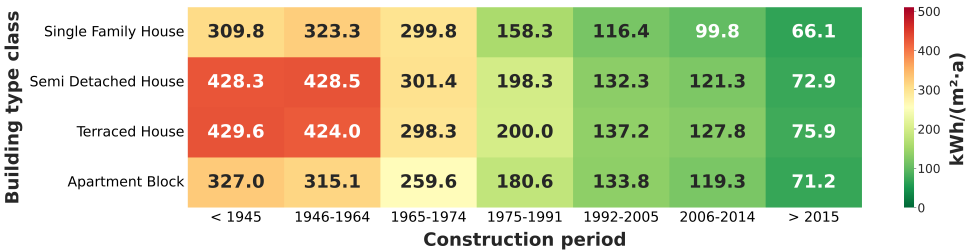


Figure 6.35: Heatmap of the computed mean net heat demand by building type class and construction period in the Netherlands

[Figure 6.36](#) shows the computed mean net heat demand per m² in the Netherlands aggregated at the H3 index level 6.

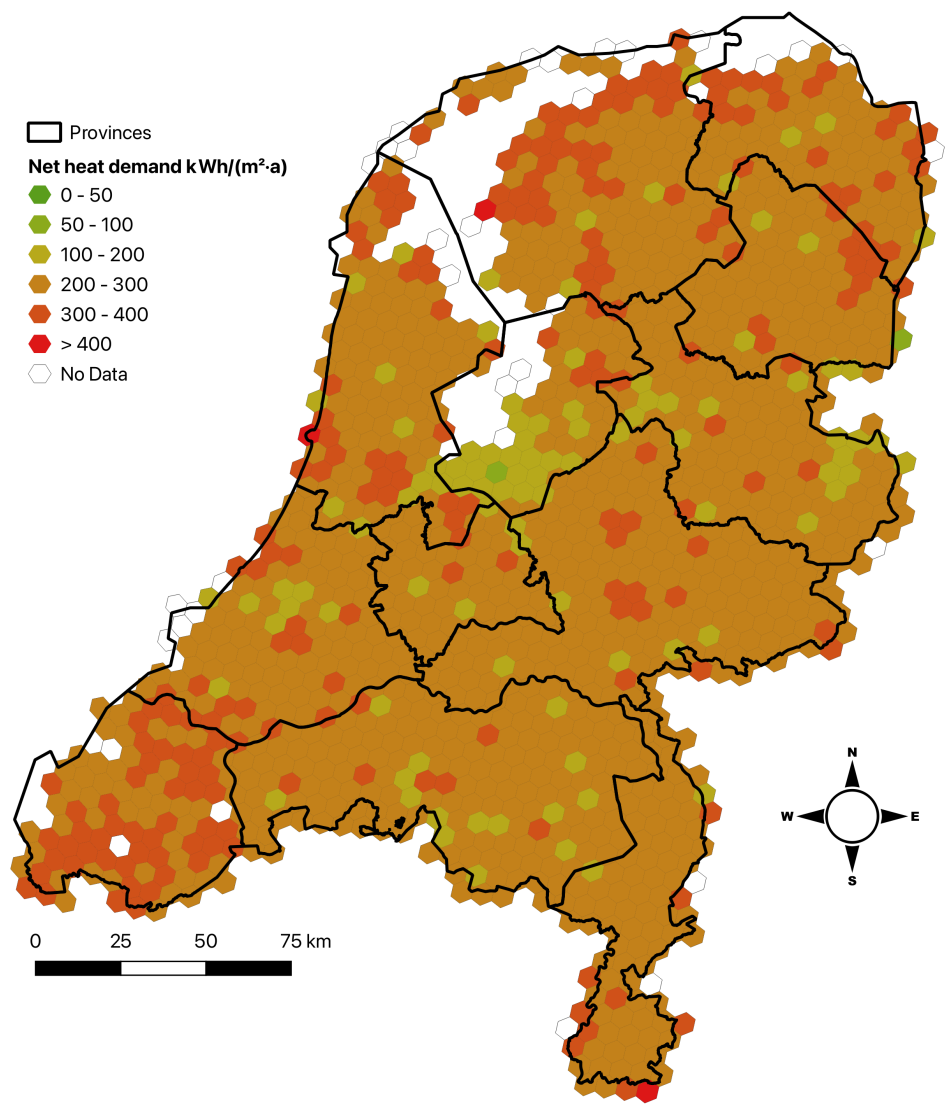


Figure 6.36: Mean net heat demand by H3 Tile

6.2.2. ANALYSIS OF SPECIFIC CASES

6.2.2.1. PAND ID 0599100000672213

This residential building has a computed annual net heat demand of 2.4 MWh/(m²·a). The basic data for the building is provided in [table 6.7](#). Inconsistencies become apparent when comparing the building data from the EnergyBAG DB with that from the BAG (see [figure 6.37](#)), the 3DBAG (see

figure 6.38), and Google Street View imagery (see figure 6.39). Based on the Google Street View image, this is a five-storey building.

Table 6.20: Summary data for Pand ID 0599100000672213

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
35.8m ²	1	15.48m	552.6m ³	1	33m ²

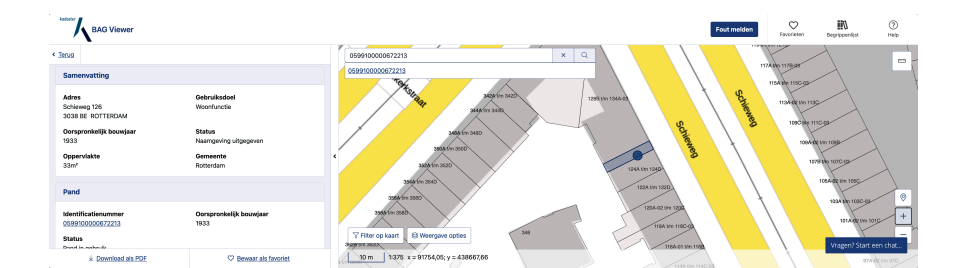


Figure 6.37: BAG viewer data for building with Pand ID 0599100000672213

6

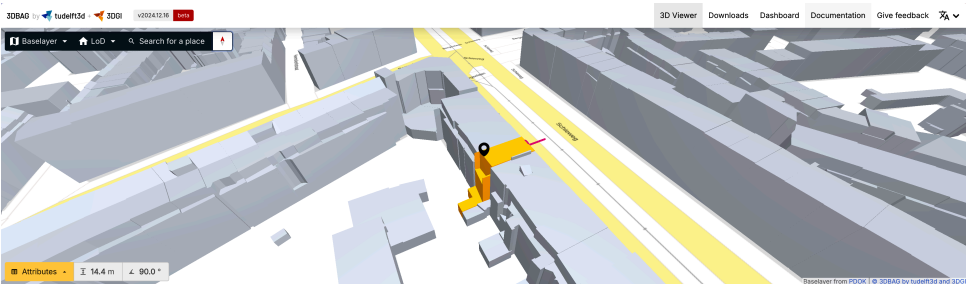


Figure 6.38: 3DBAG viewer for building with Pand ID 0599100000672213



Figure 6.39: Google street view for building with Pand ID 0599100000672213

Most likely, I assumed that it is a one-storey vbo within a *Terraced House*; however, based on the information available from the BAG clearly indicates that this hypothesis is incorrect, since there is no overlap in the building footprints with the adjacent buildings.

6.2.2.2. PAND ID 0599100000756485

This residential building has a computed annual net heat demand of 1,6 MWh/(m²·a). The basic data for the building is provided in [table 6.7](#). This building has the same pattern as the case of building with Pand ID 0599100000672213 in [section 6.2.2.1](#). Inconsistencies become apparent when comparing the building data from the EnergyBAG DB with that from the BAG (see [figure 6.40](#)), the 3DBAG (see [figure 6.41](#)), and Google Street View imagery (see [figure 6.42](#)). Based on the Google Street View image, this is a five-storey *Terraced House*.

As in the case of [section 6.2.2.1](#) there is a data inconsistency between in the input data of the EnergyBAG, the number of storeys attribute is provided by the 3DBAG. If I had followed the NTA 8800 simplification, this would be a five-storey since the norm indication for buildings without the number of storeys is to divide the height of the building by 3 (see [equation \(2.14\)](#)).

Table 6.21: Summary data for Pand ID 0599100000756485

Footprint area	Number storeys	3DBAG height	3DBAG volume	Number vbos	Total usable area
58.8m ²	1	15.56m	977.3m ³	1	58m ²

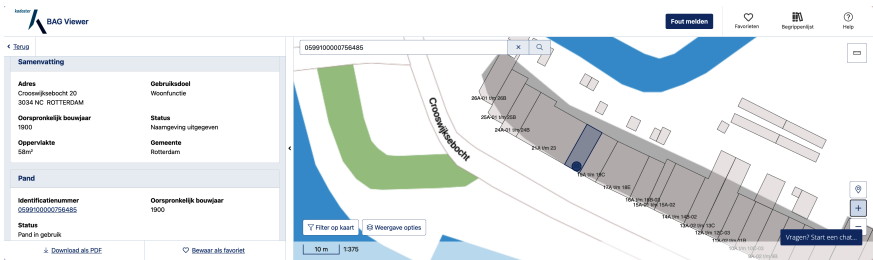


Figure 6.40: BAG viewer data for building with Pand ID 0599100000756485

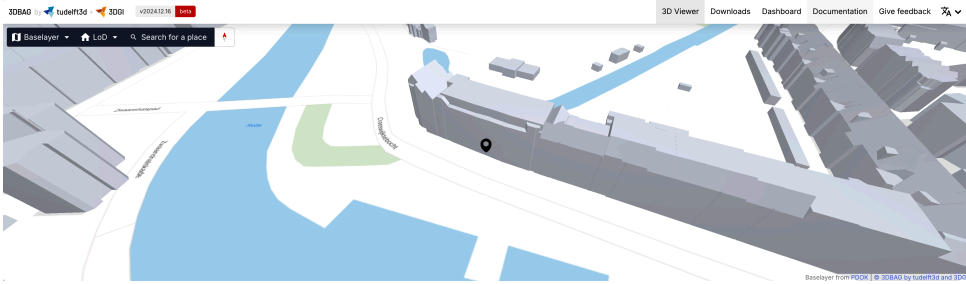


Figure 6.41: 3DBAG viewer for building with Pand ID 0599100000756485



Figure 6.42: Google street view for building with Pand ID 0599100000756485

Although the examples show the same pattern, this was not intentional but rather the result of exploring the computation outcomes based on their magnitude. However, these cases highlight the need to define additional validators to minimise the occurrence of such errors. A simple validator could involve comparing the number of storeys from the 3DBAG and the computed value according to the NTA 8800 ($\frac{\text{Building height}}{3}$). Building(Part)s with a difference greater than 2 storeys should adopt the value derived from the NTA 8800. The selection is based on the official regulations in the Netherlands for calculating the energy performance of buildings.

6.2.3. RESULTS COMPARISON

To compare the implementation results for the Netherlands, I follow the same approach used for Rijssen-Holten. However, it should be noted that, even more than in Rijssen-Holten, the results must be interpreted with caution because the data have not been polished or manually verified as in the case study, which would have been impossible given the high number of buildings. From the processed Dutch building stock dataset, only 601,626 buildings (14.13%) are used for this comparison against the ep-online dataset. Figure 6.43 shows the

distribution of computed energy demand relative to ep-online values, categorised by percentage difference, and this is further detailed in [table 6.22](#).

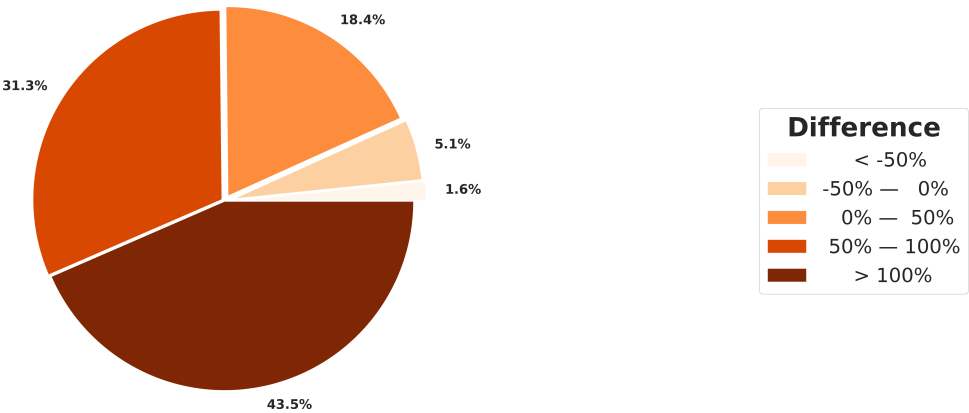


Figure 6.43: Distribution of *Building(Part)s* based on the percentage difference between ep-online values and computed net heat demand in the Netherlands

6

Table 6.22: Distribution of *Building(Part)s* based on the percentage difference between ep-online values and computed net heat demand

Difference	No. Buildings
< - 50%	9,793
-50% - 0%	30,794
0% - 50%	110,857
50% – 100%	188,560
100%	261,622

The implementation results are poorer at the country level compared to Rijssen-Holten. This discrepancy is not surprising, as I had greater control over the testbed set during its preparation. Attributes such as the number of storeys and building type are two examples of possible inconsistencies. Additionally, I computed the *PartyWalls* per surface for adjacent buildings in Rijssen-Holten but not for the whole Netherlands, directly influencing the net heat demand calculations. [Figure 6.44](#) shows a bar plot with the comparison of the computed net heat demand classified by the building type.

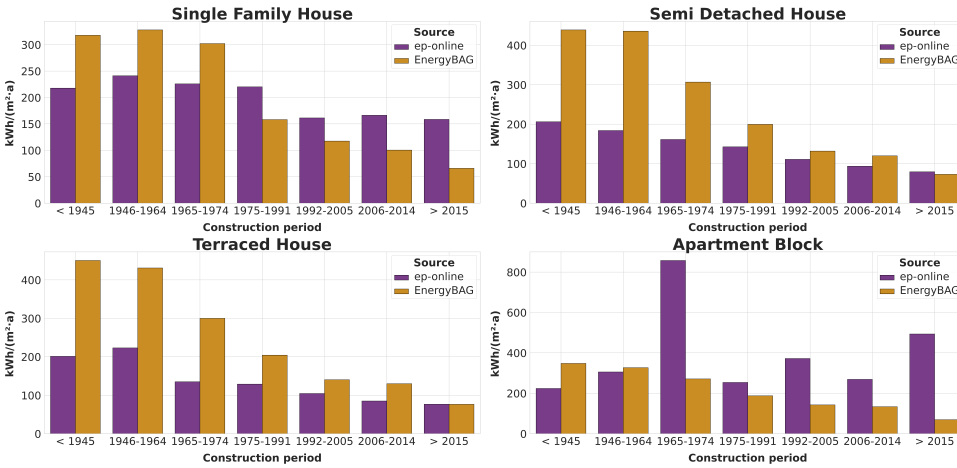


Figure 6.44: Comparison of the computed mean net heat demand against ep-online at country level

The variation in computed net heat demand against ep-online by building type and construction period is shown in figure 6.45. At the country level, *Single Family Houses* have lower overestimation compared to the other building types, particularly for pre-1991 buildings. However, the differences between ep-online values are slightly lower than those of Rijssen-Holten. *Semi Detached Houses* show higher differences at the country level for all construction periods compared to the values of Rijssen-Holten; in the case of buildings pre-1945, the difference is more than double that of the case study.

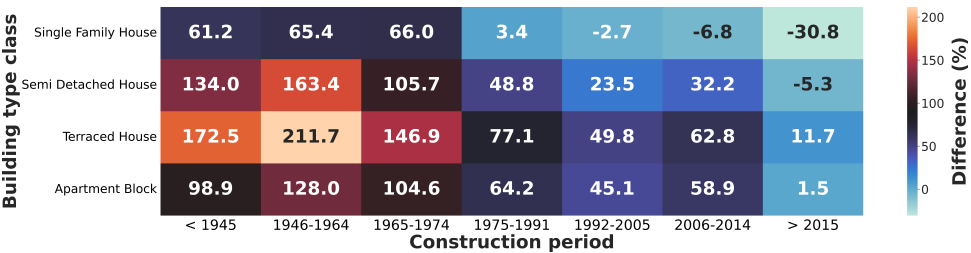


Figure 6.45: Variation in computed net heat demand with the reference values(ep-online) by building class and construction period

6.2.4. THERMAL HULL REFURBISHMENT SCENARIO

Similar to Rijssen-Holten refurbishment scenario, I perform the calculation of the net heat demand for the Netherlands following the same criteria as of Rijssen-Holten (section 6.1.5). The basic statistics of the results are presented in table 6.23.

Table 6.23: Computed annual net heat demand basic statistics of the Netherlands for the refurbishment scenario, all values are expressed in kWh/(m²·a).

Minimum	Mean	Median	Std Deviation	Maximum	Mode
0.3	59.1	70.4	10.8	420	57.2

The comparison of the computed net head demand values under the renovation scenario to ep-online values are presented in [figure 6.46](#). This comparison is classified according to the building type and construction period categories. Like in Rijssen-Holten, the computed net heat demand values are lower than ep-online values and is according to what one would expect.

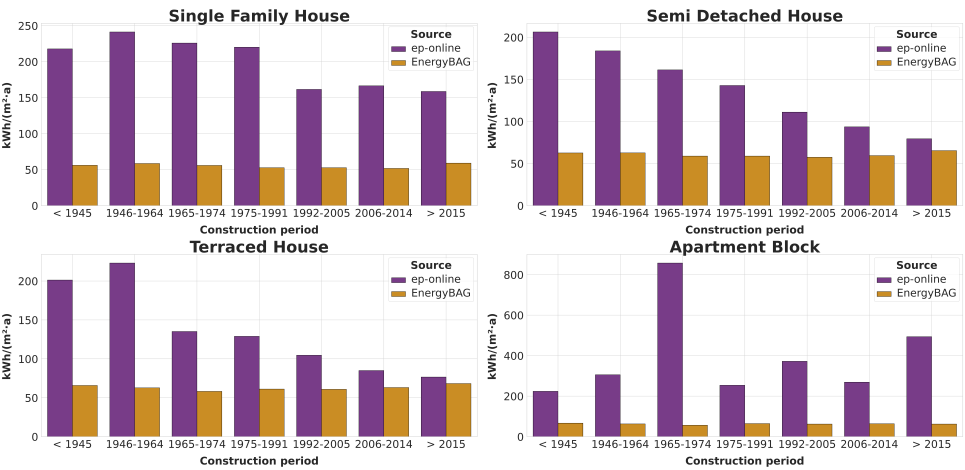


Figure 6.46: Comparison of computed net heat demand under the refurbishment scenario to ep-online reference values

The computed mean net heat demand per m² for the refurbishment scenario in the Netherlands aggregated at the H3 index level 6 is shown in [figure 6.36](#).

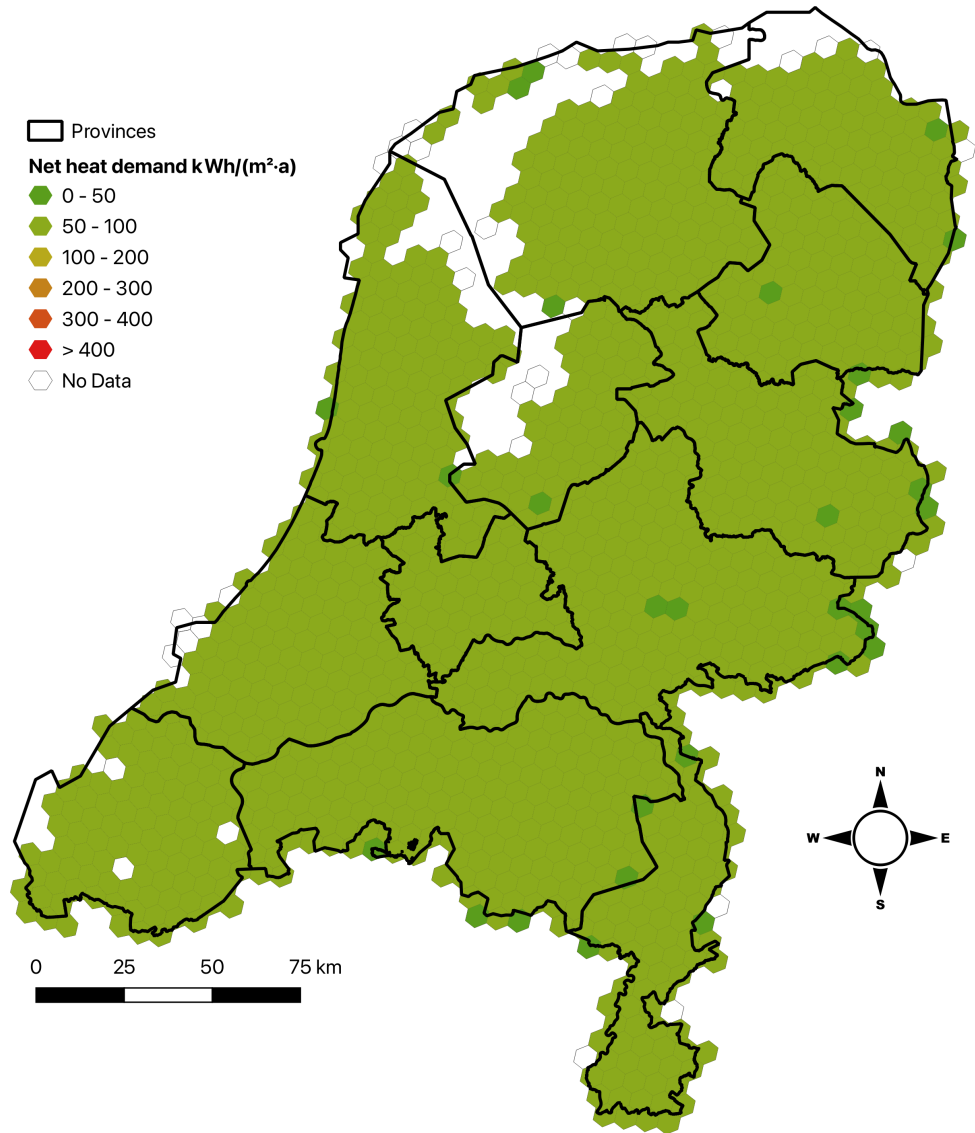


Figure 6.47: Mean net heat demand by H3 Tile

6.3. CONCLUSION REMARKS

This chapter presented the adaptability of the EnergyBAG BES for several use cases. The first case study corresponds to the use of the testbed for energy applications as the input dataset for the calculation of the net neat demand. After presenting the results, I discussed some cases with extremely high values of the

computed net heat demand, trying to identify the reasons for such cases. The pattern for these examples is vbos with multiple functions associated without information about the distribution of the usable area per "function".

I continue the analysis by comparing the results with the energy data available from ep-online. As already mentioned, this comparison should be taken with caution since I am comparing net heat demand and primary energy demand values. The EnergyBAG BES showed a lousy performance for old buildings, which is logical since the input physics-related data is based on archetype "as built" conditions. That is not the case with the ep-online dataset since the energy performance certificate is calculated based on the current conditions of the building at the time of processing.

I use the BES to simulate a thermal hull refurbishment scenario in which the physical conditions of the buildings were updated following the current specifications for new constructions. The results, as expected, show a dramatic reduction of the computed net heat demand, indicating the usability of BES to apply customised renovation scenarios, e.g. renovating all the windows to triple glazing.

The second case study corresponds to the whole of the Netherlands. I followed the same approach as in Rijssen-Holten. However, the input datasets differ since the Netherlands does not have manually cured data. Also, the *PartyWalls* data available for the Netherlands is a value per building. Consequently, this attribute cannot be used in the calculation without making additional assumptions. As expected, the country-wide results are worse than in Rijssen-Holten. However, it is also an indicator of the potentiality of the EnergyBAG BES to support different input datasets without hardcoding parameters or tailored to specific data.

Finally, the comparison of the simulation results against the energy data available country-wide (ep-online) shows research paths to follow to improve the quality of the implementation, on the one hand, but also provide insight on some issues on the input data and then the effect on the overall pipeline results.

The final remarks and reflections about my PhD research are discussed in [chapter 7](#).

7

CONCLUSIONS, REFLECTIONS AND FUTURE RESEARCH

This chapter presents the conclusions of the thesis on the use of semantic 3D city models to enhance urban energy applications based on open data. It contains my reflections to answer the research questions presented in the chapter of this document. Then I discuss the challenges, limitations and contributions of my work, followed by reflections on future perspectives on the use of semantic 3D city models in urban building energy modelling.

This chapter reflects on the key findings of my PhD research. Its content is split into the conclusions ([section 7.1](#)), which is organised to answer the research questions presented in [section 1.3](#). First, I reflect about urban building energy modelling (UBEM), its data requirements and challenges that open data through semantic 3D city models (s3DCM) can fulfil. This is followed by the reflections on the improvement of s3DCM to support energy applications. Then, I discuss on the influence of artificial intelligence methods on the quality s3DCM for energy applications. This section closes with my reflections on the role of s3DCM data models in facilitating European and Dutch directives that are related to the energy efficiency of the building stock.

The second part chapter ([section 7.2](#)) focuses with the description of the limitations, my contributions and future work of my implementation. In the final section ([section 7.3](#)), I discuss future perspectives and how my research work align with to them.

7.1. CONCLUSIONS

The principal objective of this thesis was to enhance urban energy applications through s3DCM, using the Netherlands as the case study. To achieve this, I enriched these s3DCM with open geospatial and energy-related open data to support the characterisation of the building stock. Therefore, I analysed UBEM, its scope, and its challenges, leading to the identification of critical data requirements for using and integrating multiple country-wide open datasets.

WHAT ARE THE REQUIREMENTS FOR UBEM THAT S3DCM SHOULD MEET?

From my research, it can be concluded that UBEM requires a clear and coherent spatial and semantic representation of urban features. A fundamental aspect of this is the hierarchical decomposition of space. For example, a door is part of a wall, a wall is part of a building, and a building is located in a city. This hierarchy not only defines the relationships among individual components but also ensures that each element is integrated into a more extensive system. For example, in the EnergyBAG Building Energy Simulation (BES), knowing the surfaces that define the thermal envelope of a thermal zone (or computation zone in terms of the NTA 8800) is essential to computing its net heat demand and, subsequently, the corresponding building. Without a formal data model, it would not be feasible to link well-defined geometric elements with their physical properties, and consistency in data interpretation—such as unambiguously recognizing a wall as a thermal boundary—would be compromised.

Based on the work carried out during my PhD research, CityGML data-model has been proven to be a reliable framework to support the data requirements of UBEM. It offers a well-defined semantic and geometric representation of urban entities, with structured relationships that ensure, for instance, that a *BuildingPart* cannot exist independently of a *Building* and all of them are linked unequivocally by a set of unique IDs. Furthermore, CityGML does not only model buildings but allows the representation of other entities (spatial and non-spatial) with the required properties for UBEM, e.g., a *WallSurface* can be associated with *WeatherData* such as solar irradiance values or a *Building(Part)* can be further decomposed into several *ThermalZone(s)* with specific parameters such as infiltration rate.

WHAT ARE THE CHALLENGES IN DATA PREPARATION FOR UBEM?

For this research question, several conclusions can be drawn. Firstly, in the case of the Netherlands, the open Basic Registration Addresses and Buildings (BAG) dataset provides the basis for data collection and centralisation. This dataset contains essential characteristics—such as building footprints, usage functions, registered living units (vbos) with unique IDs, and private areas—and has been instrumental in the semantic 3D reconstruction of buildings (the 3DBAG).

The differences in the scope, update frequency, and quality of each dataset require the creation of basic validators to maintain consistency. For example, the BAG is updated daily and released monthly, whereas the 3DBAG is updated approximately twice a year. These differences might lead to discrepancies and errors, such as changes in the usable area of a vbo, as discussed in [chapter 5](#).

Additional challenges were identified within the BAG dataset. Approximately 40.2% of Dutch buildings lack information on registered living units, and 16.6% show possible errors in their private area attribute (called "usable area" in this PhD thesis). These issues emphasise the requirement of rigorous data quality checks across the data pipeline process (upon data input, during the data preparation alone and when verifying against other available sources) and, ideally, to reduce redundancy.

From the data requirement analysis, the basic building geometry and registry attributes, as well as energy applications such as solar irradiance and heating demand simulations, require further data. For solar analyses, additional information about the 3D urban scene on trees, terrain, and the reflectivity (albedo) of surfaces is essential. Heating demand calculations further depend on detailed data regarding thermal zones, physical characteristics of building envelopes (including construction materials, thickness, U-values, and g-values), heating, ventilation, and air conditioning HVAC systems, occupancy, and energy sources. The lack of open data on building renovation history hinders the computation of the status quo and advanced refurbishment scenarios.

In light of the building renovation data gaps, I adopted the building archetypes defined by the Dutch government via the TABULA project. Although pragmatic, this approach leads to misclassifications due to the unclear classification rules regarding the generalisation of building characteristics. These archetypes categorise buildings based on type, construction period, and—for apartment blocks—the location of the vbo within the building. However, this limited classification oversimplifies building characteristics and still is not free from errors. For example, the *maisonnette* category is intended for vbos with more than one storey. In practice, several *Terrace Houses* are multifamily buildings with one vbo on the ground floor and another spanning the first and second floors, suggesting the vbo on the first and second floor should be classified as *maisonnette*, which is a challenge since from the vbo perspective, the classification is correct. However, from the building perspective, it should be aggregated into a more general category. Therefore, the misclassification of buildings emphasises the need for a classification framework underpinned by clear classification rules to ensure (cross-dataset) consistency, particularly for parameters like renovation histories, which are critical for energy simulations but remain largely inaccessible or, even worse, undocumented at all.

Standardised values about the renovations in buildings play a significant role in energy performance calculation. As presented in [chapter 6](#), the computation results for old buildings have less accurate values since the "as built" building

physics properties available from the building archetype have a higher discrepancy with the current status of the buildings. Therefore, detailed records of building renovations are critical to ensure the accuracy of the simulation results.

The NTA 8800 defines standardised usage profiles for buildings, assigning predefined values to occupancy, operating hours, and hot water consumption to perform asset rating energy analyses. Consequently, the norm does not use actual occupant behaviour and the specific characteristics of the energy appliances; for example, HR boilers are categorised into fixed classes to limit variation. This simplification facilitates the computation of the energy performance of the building, but it might not represent the actual conditions, which is pertinent when performing building-scale analyses.

HOW COULD EXISTING S3DCM BASED ON CITYGML BE IMPROVED TO SERVE AS INPUT DATA FOR UBEM?

The Energy Application Domain Extension (Energy ADE) already provides a comprehensive schema for managing and sharing energy-related data within CityGML-based 3DCMs, which is the reason why I chose this approach instead of “pure” CityGML and for the reasons that were given to answer the first question in [section 7.1](#). However, the current version (EnergyADE v1.0) dates back to 2018 and could be improved. Based on my PhD research, I propose the following improvements to the current version.

7

First, the model does not natively support urban areas based on functionality. Although my work primarily groups buildings using the H3 spatial indexing tile system for computational efficiency, the addition of standardised classes for energy districts, e.g., geographic clusters of buildings with shared energy systems or policies—could benefit interoperability for collaborative planning and district-scale simulations.

Second, the Energy ADE omits the opening-to-surface ratio attribute, a proxy metric essential for approximating façade areas (e.g., windows, doors) when detailed geometric representations are unavailable. This value has a direct influence on multiple parameters for the energy demand calculation, such as solar gains, heat transfer through ventilation and heat loss. Therefore, since it is indispensable for urban-scale energy simulations based on LoD1 and LoD2 building models, I would propose adding it to the classes that represent the thermal hull.

Third, Energy ADE does not provide the number of living units attribute, which serves as a pragmatic alternative when detailed interior subdivision information (e.g., individual apartments) is absent. This parameter is critical for applying standardised energy balance methods such as the Dutch NTA 8800, where predefined values for occupancy and heat demand rely on residential unit counts. By adding this attribute, the calculation of the energy performance of residential buildings could be more accurate, especially in national contexts like

the Netherlands, where multi-dwelling buildings such as Terrace Houses and Apartment Blocks dominate the building stock.

Finally, CityGML 3.0 introduced fundamental changes to the data model—including modularisation and enhanced semantics—which require a comprehensive revision of the EnergyADE data model to adjust it to the new data model. In 2024, we published the journal article “Mapping the CityGML Energy ADE to CityGML 3.0 Using a Model-Driven Approach” (Bachert *et al.*, 2024). In this publication, we tested a method to map the existing data model of the Energy ADE for CityGML 2.0 to CityGML 3.0 by following the “integrate as much as possible” approach with guidelines for uniform logical and conceptual alignment.

By leveraging CityGML v3.0’s refined space and geometry concepts, our approach reuses geometries directly from other CityObjects rather than explicitly defining them in the ADE classes. Thus, it enables a multi-geometry representation with different LoDs for the Energy ADE classes derived from AbstractCityObject or its subclasses while expanding geometric applicability to other Energy ADE classes.

HOW CAN ARTIFICIAL INTELLIGENCE METHODS IMPROVE THE QUALITY AND CONTENT OF S3DCM FOR UDEM?

Artificial intelligence (AI) is driving a technological revolution across virtually every discipline, including urban building energy modelling (UBEM). For my PhD research, AI serves as a data provider, as discussed in [section 4.3](#). However, no matter how sophisticated AI becomes, its effectiveness is limited by the quality of the input data, which is needed for training the models. High-quality, well-labelled data are essential for reliable outcomes.

In my research, we used machine learning to infer residential building types from 3DBAG data, mapping them to building archetypes defined in the Example building (in Dutch “*voorbeeldwoningen*”) report (RVO, 2023). During this process, inconsistencies in the ground truth (ep-online) and the fuzzy definitions of the building type classes adversely affected model performance and complicated result analysis. The observed variation in model accuracy, from 61.1% to 98.5% according to the building type class and case study, affects both the potential of AI in UDEM and its strong dependency on reliable input data. Therefore, It was not a reliable classifier to replicate country-wide and use its results for other calculations.

Another example from my research is the cases presented in [section 6.2.2](#); both of them followed the same pattern with unrealistic values for the number of storeys of the building, which is an attribute generated from a machine learning project (Roy *et al.*, 2023). Therefore, It is necessary to adopt validation frameworks to ensure the quality of the input datasets.

7.2. REFLECTIONS, LIMITATIONS AND CONTRIBUTIONS

The implementation of the EnergyBAG, both its database (DB) and the building energy simulation (BES), demonstrates the feasibility of leveraging open datasets and s3DCM to compute building net heat demand at a national scale based on the NTA 8800. In this section, I discuss the relevance of my work in the context of the governmental initiatives in Europe and the Netherlands. Followed by the limitations and contributions of my work.

THE ROLE OF S3DCM WHEN IT COMES TO POLICIES FOR THE BUILDING SECTOR

The latest revision of the Energy Performance of Buildings Directive –EPBD IV–(EU/2024/1275) mandates that EU member states achieve a fully decarbonised building stock by 2050 (European Parliament, 2024). The directive defines several aspects as crucial: standardised data collection, the creation of logbooks to keep a digital record of the history of the building, and building renovation passports.

The EPBD IV further aims to accelerate energy efficiency and sustainable building practices by promoting faster renovations of the existing building stock and updating requirements for new constructions. Although the directive calls for digital building logbooks and renovation passports, it provides little guidance on the management methods, data models and formats to use. In the Netherlands, the Ministry of Housing and Urban Planning advocates a consolidated national building data framework that serves as an effective data hub for building data.

For this reason, Kadaster Netherlands carried out some preliminary work on the National Facility for Building Data (in Dutch "*Landelijke Voorziening Gebouwgegevens*" (Noordegraaf and van Haaften, 2024)), which in its goals has plans to operationalise the EPBD IV through a national digital building logbook. This logbook should act as a centralised repository for building-related data, including energy performance certificates, renovation records, smart meter data, and energy ratings, similar to what the *woningpas* in Belgium (Vlaamse overheid, 2025).

However, implementing both the European and the national initiatives raise critical concerns regarding data security and governance. Key open questions include: Where will private-sensitive data be stored? Who will manage access and be responsible, reliable for? How will citizen anonymity be preserved? These are questions that should be answered in a joint effort between the government, the private sector and academia. For example, the *WoonOnderOnderzoek Nederland* (WoON) dataset, managed by the Dutch Central Bureau of Statistics (CBS), is not accessible at the building level to protect the residents' privacy, which in turn limits the availability of valuable granular data for UBE.

One potential approach is to ensure that sensitive data remains under state control rather than being managed by private entities. This concept is based on

the belief that privacy and fair data usage are public goods best safeguarded by regulations by the legislative power rather than by voluntary guidelines. My research does not provide specific answers, but the geospatial community is already involved in providing standards and techniques to deal with some of the challenges in terms of data management, including integration and storage, access management, and preserving citizens' anonymity. Also, geospatial data enables the spatial contextualisation of buildings, their energy performance metrics, and lifecycle information. As shown in this PhD thesis, s3DCM shows a valid starting point of a reliable and mature data model that supports complex geometric and semantic data and serves as an effective interchange format among stakeholders.

RENEWABLE ENERGY SOURCES CADASTRE

Two of the significant challenges that our society is facing (energy transition and climate change) have driven the adoption of renewable energy sources, for example, solar photovoltaic panels or geothermal heat pumps. However, there are only local initiatives to consolidate records of renewables information, that is the case of The Hague or Rotterdam, which have municipality records of green roofs and installed solar panels. Extending these records into a single national cadastre repository, ideally with real-time monitoring of system status, provides a better understanding of the territory and therefore to support more accurate modelling.

7.2.1. LIMITATIONS

Although the results show expected patterns regarding net heat demand, the results are still questionable. However, these are limited to the input data. One example is the *PartyWalls* data, which was available for the Rijssen-Holten dataset; some building classes showed a reduction of up to 35% of the mean net heat demand. However, the EnergyBAG BES is suffering from a lack of data, e.g., no renovation record of buildings or no data about openings. The BES is not restricted to a specific dataset. Hence, it will work based on the provided data, which is demonstrated in the two test cases documented in this PhD thesis: Rijssen-Holten and the Netherlands.

The second limitation is that I only compute the net heat demand instead of the primary energy demand due to a lack of data. As I have already pointed out in [section 2.5](#), the lack of refurbishment information directly influences the calculation of the net heat demand and, consequently, further computations based on this value. Other physics-related data that is missed is the energy systems available per building; the operation of such systems has a direct impact on the energy performance of a building. Additionally, data about the connectivity to supply networks allows the use of the correct conversion factors according to the energy supplied by each provider and reflects the environmental impact of

the "energy consumption" of the building. Closely related is the applicability of energy renewables such as solar panels, thermal collectors, wind turbines, and geothermal. When available, they offset the demand for conventional fossil-based energy and, consequently, reduce the primary energy demand due to the lower conversion factor of renewables compared to fossil fuels. Therefore, it will support an accurate calculation of the potential for decarbonisation in the built environment.

Another limitation of my implementation is the lack of validation possibilities for the output of my calculations. Using the ep-online values to perform a data comparison can only provide a simplified or qualitative indication of computed values. However, I decided to use this dataset to compare the output of the EnergyBAG BES because it follows the same calculation method in a controlled scenario that collects the input data for the calculation directly from the vbo.

Despite the mentioned limitations of the implementation and the datasets, my research has shown the feasibility, opportunities and challenges of open data for energy calculations, testing their completeness and accuracy. Additionally, to avoid additional assumptions, I did not use aggregated statistical data like energy supply per postcode (CBS, 2023). This decision has a direct influence on the final results. Since I could only compute the net heat demand instead of the primary energy demand of buildings.

7

7.2.2. CONTRIBUTIONS

There are several contributions from my research. First, the EnergyBAG DB consolidates multiple open datasets into a reliable s3DCM with energy-related data, which support energy applications such as the solar irradiance simulation (see [section 4.3.1.2](#)). This process includes the definition of data checks to ensure data consistency across multiple datasets (see [sections 5.6.5](#) and [6.1.4](#)).

Second, I created a data model for the storage of the physics-related data of buildings when there is no detailed information (the LibraryDB, see [section 4.2.1](#)). Although my case study is the Netherlands, this database support multiple input datasets and I was tested already with the data from several data sources like the TABULA project and the "example homes for existing construction" (in Dutch "*Voorbeeldwoningen bestaande bouw*").

Finally, although my proof-of-concept approach prioritised a functional workflow over polished software, the EnergyBAG BES was successfully implemented. A key principle of my PhD research was to adopt a simplified approach that can effectively cope with common data limitations. The design decisions include the definition of single thermal zones per building or the use of ratios to overcome the absence of detailed data about openings. However, these decisions do not indicate that my implementation is tailored to a specific dataset, but they provide a scalable framework for large-scale energy simulations.

7.3. FUTURE WORK

My research could be extended and improved in several ways. First, the implementation of a more elaborated system for the identification of data inconsistencies, as I have developed for checking the usable area between datasets, could automate error detection and suggest corrections based on multi-source cross-referencing.

I base my analyses on the case study in Rijssen-Holten. It exemplifies the designed and implemented method of my PhD research. However, the EnergyBAG BES should be tested with different cities and ideally using other datasets, even if the 3D representation of buildings is not watertight.

The scope of the EnergyBAG BES was a proof of concept based on Python's programming flexibility. Future versions should adopt compiled languages such as C++ or Rust and parallel multi-core computing frameworks to enhance performance for country-scale analyses. Additionally, the modularisation of the codebase and the implementation of rigorous unit tests would improve the maintainability and reproducibility of the BES. This modularisation could be implemented to support a finer and more advanced scenario mechanism, such as decarbonisation pathways and retrofit prioritisation policies.

From the geomatics point of view, geospatial data gaps require further work and integration with other datasets that provide additional information. Geomatics can provide the methods and techniques required to perform such integration. For example, remote sensing can be applied to extract elements of the facade such as windows or doors (Wysocki, Xia *et al.*, 2023; Xia, 2023). S3DCMs will move towards automatic LoD3 representation of buildings in the future. Hence, one of the major limitations and simplifications will be reduced, and the energyBAG BES tool is already ready for that.

Although it will be rare, indoor information from BIM data is expected to be integrated with city models (El Yamani *et al.*, 2023; van der Vaart *et al.*, 2024). Therefore, there will be an improvement of the models of the building stock, and based on this, a multi-thermal zone approach could be followed, for which the EnergyBAG BES already brings support.

Furthermore, it is feasible to go beyond the scope of the NTA 8800 by using solar simulation tools, as we have developed, to compute the solar irradiance values of the boundary surfaces of the building. Therefore, the solar gains will represent the situation in the building more accurately.

My work has shown that UBEM can enhance urban energy applications through semantics 3D city models and open data that can be extended to a country level. However, there are still several challenges that need to be resolved. It is reasonable to assume that the current trend and developments that we are observing in the digitalisation of the built environment - at all scales - will further develop solutions to address the opportunities and limitations I have shown in this thesis.

BIBLIOGRAPHY

- 3D geoinformation group (2024). *Cities/regions around the world with open datasets*. Open Data / Open Cities. url: <https://3d.bk.tudelft.nl/opendata/opencities/> (visited on 13/02/2025).
- Abbasabadi, N. and J. K. Mehdi Ashayeri (2019). 'Urban energy use modeling methods and tools: A review and an outlook'. In: *Building and Environment* 161 (July). Publisher: Elsevier, pp. 106–270. issn: 03601323. doi: [10.1016/j.buildenv.2019.106270](https://doi.org/10.1016/j.buildenv.2019.106270).
- Agentschap NL (2011). *Voorbeeldwoningen 2011 bestaande bouw*. Verantwoordingsrapportage nr.2KPWB1034. Rijksdienst voor Ondernemend Nederland. Sittard: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (Visited on 16/07/2024).
- Agugiaro, G., A. Zwamborn, C. Tigchelaar, E. Matthijssen, C. León-Sánchez, F. Van Der Molen and J. Stoter (2022). 'on the Influence of Party Walls for Urban Energy Modelling'. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 48.4, pp. 9–16. issn: 16821750. doi: [10.5194/isprs-archives-XLVIII-4-W5-2022-9-2022](https://doi.org/10.5194/isprs-archives-XLVIII-4-W5-2022-9-2022).
- Agugiaro, G., J. Benner, P. Cipriano and R. Nouvel (2018). 'The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations'. In: *Open Geospatial Data, Software and Standards* 3.1. issn: 2363-7501. doi: [10.1186/s40965-018-0042-y](https://doi.org/10.1186/s40965-018-0042-y).
- Agugiaro, G., K. Pantelios, C. León-Sánchez, Z. Yao and C. Nagel (2024). 'Introducing the 3DCityDB-Tools Plug-In for QGIS'. In: *Recent Advances in 3D Geoinformation Science*. 3DGeoInfo 2023. Ed. by T. H. Kolbe, A. Donaubaauer and C. Beil. Munich: Springer Nature Switzerland, pp. 797–821. isbn: 978-3-031-43699-4.
- Ali, U., S. Bano, M. H. Shamsi, D. Sood, C. Hoare, W. Zuo, N. Hewitt and J. O'Donnell (Jan. 2024a). 'Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach'. In: *Energy and Buildings* 303, p. 113768. issn: 03787788. doi: [10.1016/j.enbuild.2023.113768](https://doi.org/10.1016/j.enbuild.2023.113768). (Visited on 23/01/2024).
- (Apr. 2024b). 'Urban residential building stock synthetic datasets for building energy performance analysis'. In: *Data in Brief* 53, p. 110241. issn: 23523409. doi: [10.1016/j.dib.2024.110241](https://doi.org/10.1016/j.dib.2024.110241). (Visited on 01/09/2024).
- Ali, U., M. H. Shamsi, M. Bohacek, K. Purcell, C. Hoare, E. Mangina and J. O'Donnell (2020). 'A data-driven approach for multi-scale GIS-based building energy modeling for analysis, planning and support decision making'. In: *Applied Energy* 279 (May). Publisher: Elsevier Ltd, p. 115834. issn: 03062619. doi: [10.1016/j.apenergy.2020.115834](https://doi.org/10.1016/j.apenergy.2020.115834).
- Ali, U., M. H. Shamsi, C. Hoare, E. Mangina and J. O'Donnell (Sept. 2021). 'Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis'. In: *Energy and Buildings* 246, p. 111073. issn: 03787788. doi: [10.1016/j.enbuild.2021.111073](https://doi.org/10.1016/j.enbuild.2021.111073). (Visited on 24/08/2024).
- AlleCijfers.nl (19th Sept. 2024). *Gemeente Delft in cijfers en grafieken (bijgewerkt 2024!)* AlleCijfers.nl. url: <https://allecijfers.nl/gemeente/delft/> (visited on 23/09/2024).
- Andres, C., C. Ruben, G. David, B. Pavel, M. Patrizio, Z. Miro and I. Olindo (Feb. 2023). 'Time-varying, ray tracing irradiance simulation approach for photovoltaic systems in complex scenarios with decoupled geometry, optical properties and illumination conditions'. In: *Progress in Photovoltaics: Research and Applications* 31.2, pp. 134–148. issn: 1062-7995, 1099-159X. doi: [10.1002/pip.3614](https://doi.org/10.1002/pip.3614). (Visited on 22/11/2024).
- Arroyo Otori, K., H. Ledoux and R. Peters (2024). 'Introduction to 3D modelling of the built environment: reality, data models and data structures'. In: *3D modelling of the built environment. Lecture notes*. 0.9. Section: 1. Delft, NL. isbn: 978-1-4051-4601-2. doi: [10.1002/9781444310795.ch49](https://doi.org/10.1002/9781444310795.ch49). url: <https://github.com/tudelft3d/3dbook>.
- Ascone, F., R. F. D. Masi, A. Gigante and G. P. Vanoli (2022). 'Resilience to the climate change of nearly zero energy-building designed according to the EPBD recast: Monitoring, calibrated energy

- models and perspective simulations of a Mediterranean NZEB living lab'. In: *Energy and Buildings* 262, p. 112004. issn: 0378-7788. doi: [10.1016/j.enbuild.2022.112004](https://doi.org/10.1016/j.enbuild.2022.112004).
- Bachert, C., C. León-Sánchez, T. Kutzner and G. Agugiaro (2024). 'Mapping the CityGML Energy ADE to CityGML 3.0 Using a Model-Driven Approach'. In: *ISPRS International Journal of Geo-Information* 13.4. issn: 2220-9964. doi: [10.3390/ijgi13040121](https://doi.org/10.3390/ijgi13040121). url: <https://www.mdpi.com/2220-9964/13/4/121>.
- Bayer, M. (2012). 'SQLAlchemy'. In: *The Architecture of Open Source Applications Volume II: Structure, Scale, and a Few More Fearless Hacks*. Ed. by A. Brown and G. Wilson. aosabook.org. url: <http://aosabook.org/en/sqlalchemy.html>.
- Benner, J. (27th June 2018). *From CityGML to Energy ADE - Workflow*. Karlsruhe, DE: Institut für Automation und Angewandte Informatik. url: <https://en.wiki.energy.sig3d.org/images/upload/EnergyADE-Generation-Workflow.pdf>.
- Bensehla, S., Y. Lazri and M. C. Brito (June 2021). 'Solar potential of urban forms of a cold semi-arid city in Algeria in the present and future climate'. In: *Energy for Sustainable Development* 62, pp. 151–162. issn: 09730826. doi: [10.1016/j.esd.2021.04.004](https://doi.org/10.1016/j.esd.2021.04.004). (Visited on 13/02/2024).
- Berlin Partner / Senatsverwaltung für Wirtschaft (2024). *Berlin 3D Download portal*. url: <https://www.businesslocationcenter.de/en/economic-atlas/download-portal> (visited on 02/09/2024).
- Bianchi, C. and A. D. Smith (July 2019). 'Localized Actual Meteorological Year File Creator (LAF): A tool for using locally observed weather data in building energy simulations'. In: *SoftwareX* 10, p. 100299. issn: 23527110. doi: [10.1016/j.softx.2019.100299](https://doi.org/10.1016/j.softx.2019.100299). (Visited on 20/06/2025).
- Biljecki, F. (2017). 'Level of detail in 3D city models'. PhD thesis. Delft, NL: Delft University of Technology. 279 pp. url: <https://resolver.tudelft.nl/uuid:6fe1dea8-53b3-4734-9e0c-ff01ed393d79>.
- Biljecki, F., K. Kumar and C. Nagel (2018). 'CityGML Application Domain Extension (ADE): overview of developments'. In: *Open Geospatial Data, Software and Standards* 3.1. Publisher: Open Geospatial Data, Software and Standards, pp. 1–17. issn: 2363-7501. doi: [10.1186/s40965-018-0055-6](https://doi.org/10.1186/s40965-018-0055-6).
- Biljecki, F., H. Ledoux and J. Stoter (2016). 'An improved LOD specification for 3D building models'. In: *Computers, Environment and Urban Systems* 59, pp. 25–37. issn: 01989715. doi: [10.1016/j.compenvurbsys.2016.04.005](https://doi.org/10.1016/j.compenvurbsys.2016.04.005).
- Breiman, L. (1st Oct. 2001). 'Random Forests'. In: *Machine Learning* 45.1, pp. 5–32. issn: 1573-0565. doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- Britannica (5th July 2024). *conservation of energy*. In: *Encyclopedia Britannica*. url: <https://www.britannica.com/science/conservation-of-energy> (visited on 04/09/2024).
- Brodsky, I. and C. Piovesan (2022). *h3-duckdb*. Version 1.1.1. Publication Title: Github. url: <https://github.com/isaacbrodsky/h3-duckdb> (visited on 30/08/2024).
- Brown, S. (2023). *The C4 model for visualising software architecture*. Leanpub book. 110 pp. url: <https://c4model.com/>.
- Calcabrini, A. (26th Oct. 2023). 'Solar resource modelling and shading tolerant modules for the urban environment'. PhD thesis. Delft, NL: Delft University of Technology. 113 pp. url: <https://doi.org/10.4233/uuid:02485fcb-1445-4479-9308-518b579e3a6d>.
- Carnieletto, L., A. D. Bella, D. Quaggiotto, G. Emmi, A. Bernardi and M. De Carli (June 2024). 'Potential of GSHP coupled with PV systems for retrofitting urban areas in different European climates based on archetypes definition'. In: *Energy and Built Environment* 5.3, pp. 374–392. issn: 26661233. doi: [10.1016/j.enbenv.2022.11.005](https://doi.org/10.1016/j.enbenv.2022.11.005). (Visited on 22/01/2024).
- Carnieletto, L., M. Ferrando, L. Teso, K. Sun, W. Zhang, F. Causone, P. Romagnoni, A. Zarrella and T. Hong (2021). 'Italian prototype building models for urban scale building performance simulation'. In: *Building and Environment* 192 (December 2020). Publisher: Elsevier Ltd, p. 107590. issn: 03601323. doi: [10.1016/j.buildenv.2021.107590](https://doi.org/10.1016/j.buildenv.2021.107590).
- Casper, E., R. Nouvel and J. Benner (2018). *CityGML Energy ADE*. The CityGML Wiki. url: https://www.citygmlwiki.org/index.php/CityGML_Energy_ADE (visited on 01/09/2024).
- CBS (15th Nov. 2023). *Energielevering aan woningen en bedrijven naar postcode*. Centraal Bureau voor de Statistiek. Last Modified: 2023-11-15T09:00:00+01:00. url: <https://www.cbs.nl/nl-nl/maatwerk/2023/46/energielevering-aan-woningen-en-bedrijven-naar-postcode> (visited on 14/02/2025).
- City of New York (2024). *NYC Open Data*. url: <http://nycod-wpengine.com/> (visited on 03/09/2024).

- Conrad, C. (2010). *Tool Potential Incoming Solar Radiation*. Publication Title: SAGA-GIS Tool Library Documentation (v8.4.2). url: https://saga-gis.sourceforge.io/saga_tool_doc/8.4.2/ta_lighting_2.html.
- Coors, V., P. Rodrigues, M. Betz, S. Schneider, V. Weiler, E. Duminil, B. Schröter, A. Klöber, D. Holweg, T. Brüggemann, K. Bohn, M. Rein, L. Goll, B. Balbach, F. Spath and B. Gärtner (2021). *En-Eff:Stadt – SimStadt 2.0*. 03ET1459. Hochschule für Technik Stuttgart (HFT), p. 155. url: https://simstadt.hft-stuttgart.de/attachments/SimStadt2_Schlussbericht.pdf (visited on 25/08/2024).
- Crawley, D. B., L. K. Lawrie, F. C. Winkelmann, W. Buhl, Y. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte and J. Glazer (Apr. 2001). 'EnergyPlus: creating a new-generation building energy simulation program'. In: *Energy and Buildings* 33.4, pp. 319–331. issn: 03787788. doi: [10.1016/S0378-7788\(00\)00114-6](https://doi.org/10.1016/S0378-7788(00)00114-6). (Visited on 11/05/2024).
- Cristianini, N. and J. Shawe-Taylor (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press. 189 pp. isbn: 978-0-511-80138-9. url: <https://doi.org/10.1017/CBO9780511801389>.
- Dabirian, S., N. Rahimi and U. Eicker (4th Sept. 2023). 'Stochastic-based occupant-centric building archetype modelling using plug loads'. In: 2023 Building Simulation Conference. doi: [10.26868/25222708.2023.1381](https://doi.org/10.26868/25222708.2023.1381). (Visited on 07/11/2024).
- Davila, C. C., C. F. Reinhart and J. L. Bemis (2016). 'Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets'. In: *Energy* 117, pp. 237–250. issn: 0360-5442. doi: [10.1016/j.energy.2016.10.057](https://doi.org/10.1016/j.energy.2016.10.057).
- Del Ama Gonzalo, F., B. Moreno Santamaría and M. J. Montero Burgos (19th Jan. 2023). 'Assessment of Building Energy Simulation Tools to Predict Heating and Cooling Energy Consumption at Early Design Stages'. In: *Sustainability* 15.3, p. 1920. issn: 2071-1050. doi: [10.3390/su15031920](https://doi.org/10.3390/su15031920). (Visited on 22/01/2024).
- Delft University of Technology (23rd Feb. 2016). *Energetische verbeteringsmaatregelen in de sociale-huursector – enkele uitkomsten van de SHAERE-monitor 2010-2013*. Delft: OTB - Research for the built environment. url: <https://episcopo.eu/monitoring/case-studies/nl-the-netherlands/>.
- Deng, Z., Y. Chen, J. Yang and Z. Chen (2022). 'Archetype identification and urban building energy modeling for city-scale buildings based on GIS datasets'. In: *Building Simulation*. Vol. 15. Springer, pp. 1547–1559. doi: [10.1007/s12273-021-0878-4](https://doi.org/10.1007/s12273-021-0878-4).
- Deng, Z., K. Javanroodi, V. M. Nik and Y. Chen (Sept. 2023). 'Using urban building energy modeling to quantify the energy performance of residential buildings under climate change'. In: *Building Simulation* 16.9, pp. 1629–1643. issn: 1996-3599, 1996-8744. doi: [10.1007/s12273-023-1032-2](https://doi.org/10.1007/s12273-023-1032-2). (Visited on 07/11/2024).
- Ding, Y., Y. Shen, J. Wang and X. Shi (2015). 'Uncertainty Sources and Calculation Approaches for Building Energy Simulation Models'. In: *Energy Procedia* 78, pp. 2566–2571. issn: 1876-6102. doi: [10.1016/j.egypro.2015.11.283](https://doi.org/10.1016/j.egypro.2015.11.283).
- Doan, T. Q., C. León-Sánchez, R. Peters, G. Agugiaro and J. Stoter (2021). 'Volume Comparison of Automatically Reconstructed Multi-Lod Building Models for Urban Planning Applications'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXIV ISPRS Congress. Vol. V-4-2021. Copernicus Publications, pp. 169–176. doi: [10.5194/isprs-annals-v-4-2021-169-2021](https://doi.org/10.5194/isprs-annals-v-4-2021-169-2021). url: <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/V-4-2021/169/2021/>.
- Dochev, I., P. Gorzalka, V. Weiler, J. Estevam Schmiedt, M. Linkiewicz, U. Eicker, B. Hoffschmidt, I. Peters and B. Schröter (Nov. 2020). 'Calculating urban heat demands: An analysis of two modelling approaches and remote sensing for input data and validation'. In: *Energy and Buildings* 226, p. 110378. issn: 03787788. doi: [10.1016/j.enbuild.2020.110378](https://doi.org/10.1016/j.enbuild.2020.110378). (Visited on 15/02/2025).
- Doma, A. and M. Ouf (Feb. 2023). 'Modelling occupant behaviour for urban scale simulation: Review of available approaches and tools'. In: *Building Simulation* 16.2, pp. 169–184. issn: 1996-3599, 1996-8744. doi: [10.1007/s12273-022-0939-3](https://doi.org/10.1007/s12273-022-0939-3). (Visited on 07/11/2024).
- Duminil, E., K. Brassel, R. Nouvel, A. Benoit, M. Bruse, M. Betz, N. Alam, H. Dastageeri, P. Wate, P. Debue, S. Köhler, V. Weiler, P. Monsalvete, M. Zirak, K. Bao, S. Schneider and V. Coors (2022). *SimStadt*. url: <https://simstadt.hft-stuttgart.de/> (visited on 25/09/2024).

- Eames, M. E., A. P. Ramallo-Gonzalez and M. J. Wood (2016). 'An update of the UK's test reference year: The implications of a revised climate on building design'. In: *Building Services Engineering Research and Technology* 37.3, pp. 316–333. doi: [10.1177/0143624415605626](https://doi.org/10.1177/0143624415605626).
- El Yamani, S., H. Rafika and R. Billen (25th Aug. 2023). 'IFC-CityGML Data Integration for 3D Property Valuation'. In: *ISPRS International Journal of Geo-Information* 12. Publisher: MDPI AG. doi: [10.3390/ijgi12090351](https://doi.org/10.3390/ijgi12090351).
- Emmanuel, W. and K. Jérôme (2015). 'A verification of CitySim results using the BESTEST and monitored consumption values'. In: *Building Simulation Applications*. ISBN: 9788860460745, pp. 215–222. issn: 25316702.
- EPISCOPE Project (2017). *TABULA Webtool*. IEE Project TABULA (2009 - 2012). url: <https://webtool.building-typology.eu> (visited on 30/08/2024).
- Esri (2024). *Area Solar Radiation (Spatial Analyst)*. url: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/area-solar-radiation.htm> (visited on 01/03/2024).
- esri (2025). *Woningtypering - Overview*. url: <https://services.arcgis.com/nSZVuSjZjHpEZZbRo/arcgis/rest/services/Woningtypering/FeatureServer> (visited on 05/07/2025).
- Etheridge, D. W. (2010). '3 - Ventilation, air quality and airtightness in buildings'. In: *Materials for Energy Efficiency and Thermal Comfort in Buildings*. Ed. by M. R. Hall. Woodhead Publishing Series in Energy. Woodhead Publishing, pp. 77–100. isbn: 978-1-84569-526-2. doi: [10.1533/9781845699277.1.77](https://doi.org/10.1533/9781845699277.1.77).
- European Parliament (24th Apr. 2024). *Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the energy performance of buildings*. Doc ID: 32024L1275, Doc Sector: 3, Doc Type: L. url: <https://eur-lex.europa.eu/eli/dir/2024/1275/oj>.
- Faure, X., T. Johansson and O. Pasichnyi (2022). 'The Impact of Detail, Shadowing and Thermal Zoning Levels on Urban Building Energy Modelling (UBEM) on a District Scale'. In: *Energies* 15.4. issn: 1996-1073. doi: [10.3390/en15041525](https://doi.org/10.3390/en15041525).
- Fonseca, J. A., T.-A. Nguyen, A. Schlueter and F. Marechal (2016). 'City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts'. In: *Energy and Buildings* 113, pp. 202–226. issn: 0378-7788. doi: [10.1016/j.enbuild.2015.11.055](https://doi.org/10.1016/j.enbuild.2015.11.055).
- Fu, C. and C. Miller (2022). 'Using Google Trends as a proxy for occupant behavior to predict building energy consumption'. In: *Applied Energy* 310, p. 118343. issn: 0306-2619. doi: [10.1016/j.apenergy.2021.118343](https://doi.org/10.1016/j.apenergy.2021.118343).
- Geiger, A., A. Nichersu, K.-H. Häfele and V. Hagenmeyer (20th Sept. 2022). 'Usage profile enrichment of CityGML models for urban building energy modeling'. In: *BauSim Conference 2022*. doi: [10.26868/29761662.2022.28](https://doi.org/10.26868/29761662.2022.28). (Visited on 19/02/2025).
- Gemeente Amsterdam (4th Nov. 2022). *BAG-plus*. url: <https://api.data.amsterdam.nl/v1/wfs/bag/> (visited on 18/09/2024).
- Geske, M., M. Engels, A. Benz and C. Voelker (4th Sept. 2023). 'Impact of different Input Data on Urban Building Energy Modeling for the German Building Stock'. In: *Building Simulation Conference Proceedings*. 2023 Building Simulation Conference. Vol. 18. ISSN: 2522-2708. IBPSA. doi: [10.26868/25222708.2023.1253](https://doi.org/10.26868/25222708.2023.1253). url: https://publications.ibpsa.org/conference/paper/?id=bs2023_1253 (visited on 23/07/2025).
- Giannelli, D., C. León-Sánchez and G. Agugiaro (2022). 'Comparison and evaluation of different GIS software tools to estimate solar irradiation'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXIV ISPRS Congress. Vol. 5. Nice, France: Copernicus Publications, pp. 275–282. doi: [10.5194/isprs-Annals-V-4-2022-275-2022](https://doi.org/10.5194/isprs-Annals-V-4-2022-275-2022).
- Glasgo, B., C. Hendrickson and I. L. Azevedo (Oct. 2017). 'Assessing the value of information in residential building simulation: Comparing simulated and actual building loads at the circuit level'. In: *Applied Energy* 203, pp. 348–363. issn: 03062619. doi: [10.1016/j.apenergy.2017.05.164](https://doi.org/10.1016/j.apenergy.2017.05.164). (Visited on 12/05/2024).
- Google (2022). *Google Street View*. Accessed: 2025-02-01. <https://www.google.com/streetview/>.
- Gröger, G., T. Kolbe, C. Nagel and K.-H. Häfele (2012). *OGC City Geography Markup Language (CityGML) Encoding Standard*. In collab. with G. Gröger, T. Kolbe, C. Nagel and K.-H. Häfele. Version 2.0. ISSN: 18632246. url: <https://www.opengeospatial.org/standards/citygml>.

- Gröger, G., T. H. Kolbe, A. Czerwinski and C. Nagel (2008). *OpenGIS City Geography Markup Language (CityGML) Encoding Standard*. 08-007r1. 234 pp. url: <http://www.opengis.net/spec/citygml/2.0>.
- Guen, M. L., L. Mosca, A. Perera, S. Coccolo, N. Mohajeri and J.-L. Scartezzini (Jan. 2018). 'Improving the energy sustainability of a Swiss village through building renovation and renewable energy integration'. In: *Energy and Buildings* 158, pp. 906–923. issn: 03787788. doi: [10.1016/j.enbuild.2017.10.057](https://doi.org/10.1016/j.enbuild.2017.10.057). (Visited on 09/11/2024).
- Hammingh, P., M. van Seville, M. Hoff, C. Volkers, P. Koutstaal and M. Menkveld (24th Oct. 2024). *Klimaat- en Energieverkenning 2024*. 5490. PBL Planbureau voor de Leefomgeving, p. 183.
- Heidenthaler, D., Y. Deng, M. Leeb, M. Grobbauer, L. Kranzl, L. Seiwald, P. Mascherbauer, P. Reindl and T. Bednar (Sept. 2023). 'Automated energy performance certificate based urban building energy modelling approach for predicting heat load profiles of districts'. In: *Energy* 278, p. 128024. issn: 03605442. doi: [10.1016/j.energy.2023.128024](https://doi.org/10.1016/j.energy.2023.128024). url: <https://linkinghub.elsevier.com/retrieve/pii/S0360544223014184> (visited on 23/01/2024).
- Herman, J. R. and R. A. Golberg (1978). *Sun, Weather, and Climate*. Vol. 426. Scientific, Technical Information Office, National Aeronautics and Space Agency.
- Hofierka, J., M. Suri and T. Huld (2007). *r.sun*. Publication Title: GRASS GIS 8.2.2dev Reference Manual. url: <https://grass.osgeo.org/grass82/manuals/r.sun.html>.
- Hong, T., W.-K. Chang and H.-W. Lin (2013). 'A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data'. In: *Applied Energy* 111, pp. 333–350. issn: 0306-2619. doi: [10.1016/j.apenergy.2013.05.019](https://doi.org/10.1016/j.apenergy.2013.05.019).
- Hong, T., Y. Chen, Z. Belafi and S. D'Oca (Feb. 2018). 'Occupant behavior models: A critical review of implementation and representation approaches in building performance simulation programs'. In: *Building Simulation* 11.1, pp. 1–14. issn: 1996-3599, 1996-8744. doi: [10.1007/s12273-017-0396-6](https://doi.org/10.1007/s12273-017-0396-6). (Visited on 03/09/2024).
- Hong, T., Y. Chen, X. Luo, N. Luo and S. H. Lee (2020). 'Ten questions on urban building energy modeling'. In: *Building and Environment* 168, p. 106508. issn: 0360-1323. doi: [10.1016/j.buildenv.2019.106508](https://doi.org/10.1016/j.buildenv.2019.106508). (Visited on 22/01/2024).
- Howard, B., L. Parshall, J. Thompson, S. Hammer, J. Dickinson and V. Modi (Feb. 2012). 'Spatial distribution of urban building energy consumption by end use'. In: *Energy and Buildings* 45, pp. 141–151. issn: 03787788. doi: [10.1016/j.enbuild.2011.10.061](https://doi.org/10.1016/j.enbuild.2011.10.061). (Visited on 25/08/2024).
- IDECA (2024). *Datos Abiertos Bogotá*. url: <https://datosabiertos.bogota.gov.co/> (visited on 02/09/2024).
- IEA (2023). *World Energy Outlook 2023*. International Energy Agency, p. 355. url: <https://www.iea.org/reports/world-energy-outlook-2023>.
- ISO (June 2017). *Energy performance of buildings — Energy needs for heating and cooling, internal temperatures and sensible and latent heat loads — Part 1: Calculation procedures (ISO Standard No. 52016-1:2017)*. Publication Title: International Organization for Standardization. Geneva, Switzerland. url: <https://www.iso.org/standard/65696.html>.
- Johari, F., G. Peronato, P. Sadeghian, X. Zhao and J. Widén (Aug. 2020). 'Urban building energy modeling: State of the art and future prospects'. In: *Renewable and Sustainable Energy Reviews* 128, p. 109902. issn: 1364-0321. doi: [10.1016/j.rser.2020.109902](https://doi.org/10.1016/j.rser.2020.109902). (Visited on 23/01/2024).
- Johari, F., F. Shadram and J. Widén (Sept. 2023). 'Urban building energy modeling from geo-referenced energy performance certificate data: Development, calibration, and validation'. In: *Sustainable Cities and Society* 96, p. 104664. issn: 22106707. doi: [10.1016/j.scs.2023.104664](https://doi.org/10.1016/j.scs.2023.104664). (Visited on 15/02/2025).
- Kadaster (2022). *bouwlagen-hoogste-gebouw*. Inzichten Bouwlagen By Data Science Team (DST). url: <https://data.labs.kadaster.nl/dst/-/stories/inzichten-bouwlagen> (visited on 12/01/2025).
- (2024a). *BAG Viewer*. Zoeken in Basisregistratie Adressen en Gebouwen (BAG). url: <https://bagviewer.kadaster.nl/lvbag/bag-viewer/?zoomlevel=1> (visited on 12/12/2024).
- (2024b). *Dataset: Basisregistratie Adressen en Gebouwen (BAG)*. pdok. url: <https://www.pdok.nl/introductie/-/article/basisregistratie-adressen-en-gebouwen-ba-1> (visited on 12/09/2024).
- (2024c). *Over BAG*. Basisregistratie Adressen en Gebouwen - BAG. url: <https://www.kadaster.nl/zakelijk/registraties/basisregistraties/bag/over-bag> (visited on 12/09/2024).

- Kadaster (2024d). *PDOK*. Publieke Dienstverlening Op de Kaart (PDOK). url: <https://www.pdok.nl/> (visited on 03/09/2024).
- (2024e). *Praktijkhandleiding BAG*. url: <https://imbag.github.io/praktijkhandleiding/attributen> (visited on 12/09/2024).
- (2025). *INSPIRE datasets*. url: <https://www.kadaster.nl/zakelijk/producten/adressen-en-gebouwen/inspire-datasets> (visited on 05/08/2025).
- Kamel, E. (18th Nov. 2022). 'A Systematic Literature Review of Physics-Based Urban Building Energy Modeling (UBEM) Tools, Data Sources, and Challenges for Energy Conservation'. In: *Energies* 15.22, p. 8649. issn: 1996-1073. doi: [10.3390/en15228649](https://doi.org/10.3390/en15228649). (Visited on 22/01/2024).
- Klement, M. (2024). *Meteorological data*. The Dutch PV Portal. url: <https://www.tudelft.nl/en/ewi/over-de-faculteit/afdelingen/electrical-sustainable-energy/photovoltaic-materials-and-devices/dutch-pv-portal/meteorological-data> (visited on 11/10/2024).
- KNMI (2024a). *Automatic Weather Stations*. url: <https://www.knmi.nl/kennis-en-datacentrum/uitleg/automatische-weerstations> (visited on 01/03/2024).
- (2024b). *Uurgegevens van het weer in Nederland*. Klimatologie. url: <https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens> (visited on 11/11/2024).
- Kong, D., A. Cheshmehzangi, Z. Zhang, S. P. Ardakani and T. Gu (Aug. 2023). 'Urban building energy modeling (UBEM): a systematic review of challenges and opportunities'. In: *Energy Efficiency* 16.6, p. 69. issn: 1570-646X, 1570-6478. doi: [10.1007/s12053-023-10147-z](https://doi.org/10.1007/s12053-023-10147-z). (Visited on 03/09/2024).
- Kristensen, M. H., R. E. Hedegaard and S. Petersen (Sept. 2018). 'Hierarchical calibration of archetypes for urban building energy modeling'. In: *Energy and Buildings* 175, pp. 219–234. issn: 03787788. doi: [10.1016/j.enbuild.2018.07.030](https://doi.org/10.1016/j.enbuild.2018.07.030). (Visited on 23/01/2024).
- Lampard, E. E. (1955). 'The history of cities in the economically advanced areas'. In: *Economic development and cultural change* 3.2. Publisher: University of Chicago Press, pp. 81–136. issn: 00130079, 15392988. url: <https://www.jstor.org/stable/1151815> (visited on 18/07/2024).
- Langevin, J., J. Reyna, S. Ebrahimigharehbaghi, N. Sandberg, P. Fennell, C. Nägeli, J. Laverge, M. Delghust, É. Mata, M. Van Hove, J. Webster, F. Federico, M. Jakob and C. Camarasa (Nov. 2020). 'Developing a common approach for classifying building stock energy models'. In: *Renewable and Sustainable Energy Reviews* 133, p. 110276. issn: 13640321. doi: [10.1016/j.rser.2020.110276](https://doi.org/10.1016/j.rser.2020.110276). (Visited on 03/06/2025).
- Laurent, D., K.-G. Johansson, S. Barlow, E. Bergström, Z. Ferencz, G. Gröger, F. Kooij, F. Mortier, K. Skjelbo, F. Taucer, A. Velasco, E. Wysocka, J. Gaffuri and M. Lutz (31st Jan. 2024). *D2.8.III.2 Data Specification on Buildings – Technical Guidelines*. Technical Guidelines D2.8.III.2_v3.1.0. INSPIRE Infrastructure for Spatial Information in Europe. url: https://knowledge-base.inspire.ec.europa.eu/publications/inspire-data-specification-buildings-technical-guidelines_en.
- Lavagna, M., C. Baldassarri, A. Campioli, S. Giorgi, A. D. Valle, V. Castellani and S. Sala (2018). 'Benchmarks for environmental impact of housing in Europe: Definition of archetypes and LCA of the residential building stock'. In: *Building and Environment* 145, pp. 260–275. issn: 0360-1323. doi: [10.1016/j.buildenv.2018.09.008](https://doi.org/10.1016/j.buildenv.2018.09.008).
- Lawrie, L. K. and D. B. Crawley (27th Mar. 2023). *Development of Global Typical Meteorological Years (TMYx)*. Publication Title: <http://climate.onebuilding.org>. url: <http://climate.onebuilding.org>.
- Ledoux, H. (2013). 'On the validation of solids represented with the international standards for geographic information'. In: *Computer-Aided Civil and Infrastructure Engineering* 28.9, pp. 693–706. doi: <https://dx.doi.org/10.1111/mice.12043>.
- (2018). 'val3dity: validation of 3D GIS primitives according to the international standards'. In: *Open Geospatial Data, Software and Standards* 3.1, p. 1. doi: <https://dx.doi.org/10.1186/s40965-018-0043-x>.
- Ledoux, H., K. Arroyo Ohori, K. Kumar, B. Dukai, A. Labetski and S. Vitalis (2019). 'CityJSON: a compact and easy-to-use encoding of the CityGML data model'. In: *Open Geospatial Data, Software and Standards* 4.1. issn: 2363-7501. doi: [10.1186/s40965-019-0064-0](https://doi.org/10.1186/s40965-019-0064-0).
- Ledoux, H. and B. Dukai (20th Oct. 2023). *CityJSON Community Standard*. Version 2.0.0. url: <https://docs.ogc.org/cs/20-072r5/20-072r5.html>.
- (11th Apr. 2024). *CityJSON Specifications 2.0.1*. Version 2.0.0. Publication Title: CityJSON. A JSON-based encoding for 3D city models. url: <https://www.cityjson.org/specs/2.0.1/> (visited on 30/08/2024).

- Lemburg, M.-A. (12th Apr. 1999). *PEP 249 – Python Database API Specification v2.0*. Python Enhancement Proposals 249. url: <https://peps.python.org/pep-0249/> (visited on 23/10/2024).
- León-Sánchez, C., G. Agugiaro and J. Stoter (Oct. 2022). 'Creation of a CityGML-based 3D city model testbed for energy-related applications'. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. 7th International Conference on Smart Data and Smart Cities (SDSC). Vol. 48. Issue: 4/W5-2022. Sydney, Australia: Copernicus Publications, pp. 97–103. isbn: 16821750. doi: [10.5194/isprs-archives-XLVIII-4-W5-2022-97-2022](https://doi.org/10.5194/isprs-archives-XLVIII-4-W5-2022-97-2022).
- (Sept. 2025). 'Lessons learnt from the integration of open data and semantic 3D city models for urban building energy modelling in the Netherlands'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 20th 3D GeoInfo & 9th Smart Data Smart Cities Conference. Tokyo, Japan: Copernicus Publications. Best Young Researcher Award at the 9th International Smart Data and Smart Cities conference 2025.
- León-Sánchez, C., D. Giannelli, G. Agugiaro and J. Stoter (2021). 'Testing the new 3D BAG dataset for energy demand estimation of residential buildings'. In: *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. 6th International Conference on Smart Data and Smart Cities. Vol. XLVI-4/W1-2021. Stuttgart: Copernicus Publications, pp. 69–76. doi: [10.5194/isprs-archives-XLVI-4-W1-2021-69-2021](https://doi.org/10.5194/isprs-archives-XLVI-4-W1-2021-69-2021).
- (2025). 'Comparative analysis of geospatial tools for solar simulation'. In: *Transactions in GIS* 29.1, e13296. doi: [10.1111/tgis.13296](https://doi.org/10.1111/tgis.13296).
- Li, Z., J. Ma, Y. Tan, C. Guo and X. Li (2023). 'Combining physical approaches with deep learning techniques for urban building energy modeling: A comprehensive review and future research prospects'. In: *Building and Environment* 246, p. 110960. issn: 0360-1323. doi: [10.1016/j.buildenv.2023.110960](https://doi.org/10.1016/j.buildenv.2023.110960).
- Lindberg, F., T. Sun, S. Grimmond, Y. Tang and N. Wallenberg (2023). *UMEP Manual — UMEP Manual documentation*. url: <https://umep-docs.readthedocs.io/en/latest/index.html> (visited on 01/03/2024).
- Luo, N., X. Luo, M. Mortezaazadeh, M. Albetar, W. Zhang, D. Zhan, L. (Wang and T. Hong (9th Nov. 2022). 'A data schema for exchanging information between urban building energy models and urban microclimate models in coupled simulations'. In: *Journal of Building Performance Simulation*, pp. 1–18. issn: 1940-1493, 1940-1507. doi: [10.1080/19401493.2022.2142295](https://doi.org/10.1080/19401493.2022.2142295). (Visited on 07/11/2024).
- Mahdy, M., I. Elwy, S. Mahmoud, M. Abdelalim and M. Fahmy (2022). 'The impact of using different weather datasets for predicting current and future energy performance of residential buildings in Egypt'. In: *Energy Reports* 8, pp. 372–378. issn: 2352-4847. doi: [10.1016/j.egy.2022.01.052](https://doi.org/10.1016/j.egy.2022.01.052).
- Malhotra, A., J. Bischof, A. Nichersu, K.-H. Häfele, J. Exenberger, D. Sood, J. Allan, J. Frisch, C. Van Treeck, J. O'Donnell and G. Schweiger (Jan. 2022). 'Information modelling for urban building energy simulation—A taxonomic review'. In: *Building and Environment* 208, p. 108552. issn: 0360-1323. doi: [10.1016/j.buildenv.2021.108552](https://doi.org/10.1016/j.buildenv.2021.108552). (Visited on 22/01/2024).
- Malhotra, A., M. Shamovich, J. Frisch and C. Van Treeck (2019). 'Parametric Study of the Different Level of Detail of CityGML and Energy-ADE Information for Energy Performance Simulations'. In: *Building Simulation 2019*. Rome, Italy, pp. 3429–3436. doi: [10.26868/25222708.2019.210607](https://doi.org/10.26868/25222708.2019.210607). (Visited on 19/02/2025).
- Malhotra, A., M. Shamovich, S. Raming, J. Frisch and C. Van Treeck (20th Sept. 2022). 'AN OPEN-SOURCE CITYGML ENRICHMENT TOOL (CITYENRICH)'. In: *BauSim Conference 2022*. doi: [10.26868/29761662.2022.56](https://doi.org/10.26868/29761662.2022.56). (Visited on 01/09/2024).
- Mapillary (2022). *Mapillary*. Accessed: 2025-02-01. <https://www.mapillary.com/>.
- Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu and Xiaoqiang Zheng (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. url: <https://www.tensorflow.org/>.
- McKinney, W. (2010). 'Data Structures for Statistical Computing in Python'. In: *Proceedings of the 9th Python in Science Conference*. Ed. by S. van der Walt and J. Millman, pp. 56–61. doi: [10.25080/Majorsa-92bf1922-00a](https://doi.org/10.25080/Majorsa-92bf1922-00a).

- Meyer, I. (15th Dec. 2023). *simple-term-menu: A Python package which creates simple interactive menus on the command line*. Version 1.6.4. url: <https://github.com/IngoMeyer441/simple-term-menu> (visited on 25/11/2024).
- Mohajeri, N., G. Upadhyay, A. Gudmundsson, D. Assouline, J. Kämpf and J.-L. Scartezzini (Aug. 2016). 'Effects of urban compactness on solar energy potential'. In: *Renewable Energy* 93, pp. 469–482. issn: 09601481. doi: [10.1016/j.renene.2016.02.053](https://doi.org/10.1016/j.renene.2016.02.053). (Visited on 13/02/2024).
- Mohammed, A., A. Khan, H. S. Khan and M. Santamouris (Apr. 2024). 'On the energy impact of cool roofs in Dubai'. In: *Solar Energy* 272, p. 112447. issn: 0038092X. doi: [10.1016/j.solener.2024.112447](https://doi.org/10.1016/j.solener.2024.112447). (Visited on 12/05/2024).
- Moser, A., A. Schöulin, L. Davison, V. Corrado, V. Dorer and M. Koschenz (2001). 'DESIGN WITH MODELING TECHNIQUES'. In: *Industrial Ventilation Design Guidebook*. Ed. by H. Goodfellow and E. Tähti. Vol. 11. San Diego: Academic Press, 1025–p3. isbn: 978-0-12-289676-7. doi: [10.1016/B978-012289676-7/50014-X](https://doi.org/10.1016/B978-012289676-7/50014-X).
- Mosteiro-Romero, M., I. Hischer, J. A. Fonseca and A. Schlueter (Aug. 2020). 'A novel population-based occupancy modeling approach for district-scale simulations compared to standard-based methods'. In: *Building and Environment* 181, p. 107084. issn: 03601323. doi: [10.1016/j.buildenv.2020.107084](https://doi.org/10.1016/j.buildenv.2020.107084). (Visited on 07/11/2024).
- Mosteiro-Romero, M., M. Quintana, R. Stouffs and C. Miller (June 2024). 'A data-driven agent-based model of occupants' thermal comfort behaviors for the planning of district-scale flexible work arrangements'. In: *Building and Environment* 257, p. 111479. issn: 03601323. doi: [10.1016/j.buildenv.2024.111479](https://doi.org/10.1016/j.buildenv.2024.111479). (Visited on 25/08/2024).
- Mutani, G., S. Cocco, J. Kämpf and M. Bilardo (2018). *CitySim Guide : Urban Energy Modelling*. Torino, Italia: CreateSpace Independent Publishing Platform. 114 pp. isbn: 978-1-987609-73-8. url: <https://www.amazon.com/CitySim-Guide-Urban-Energy-Modelling/dp/1987609735>.
- Nagel, C. (19th Sept. 2024a). *i-UR ADE extension for the 3D City Database*. Version 2.1.0. Publication Title: Github. Berlin. url: <https://github.com/3dcitydb/plugin-ade-manager> (visited on 30/08/2024).
- (19th Sept. 2024b). *Quality ADE extension for the 3D City Database*. Version 0.1.4. Publication Title: Github. Berlin. url: <https://github.com/3dcitydb/plugin-ade-manager> (visited on 30/08/2024).
- Nägeli, C., L. Thuvander, H. Wallbaum, R. Cachia, S. Stortecky and A. Hainoun (2022). 'Methodologies for Synthetic Spatial Building Stock Modelling: Data-Availability-Adapted Approaches for the Spatial Analysis of Building Stock Energy Demand'. In: *Energies* 15.18. issn: 1996-1073. doi: [10.3390/en15186738](https://doi.org/10.3390/en15186738).
- NEN (2007). *NEN 2580:2007*. Version 2007. url: <https://www.nen.nl/nen-2580-2007-nl-113982> (visited on 03/04/2024).
- (1st July 2020). *NTA 8800:2020 nl*. Version 2020. url: <https://www.nen.nl/nta-8800-2024-nl-320123> (visited on 03/04/2024).
- (1st Jan. 2024). *NTA 8800:2024 nl*. Version 2024. url: <https://www.nen.nl/nta-8800-2024-nl-320123> (visited on 03/04/2024).
- Neufert, E., J. Kister, M. Lohmann, P. Merkel and M. Brockhaus (2021). *Bauentwurfslehre*. 43rd ed. Im Bauwelt Verlag. 615 pp. isbn: 978-3-658-34236-4.
- Noordegraaf, L. and R. van Haaften (26th July 2024). *Landelijke Voorziening Gebouwgegevens*. Kadaster NL, p. 76. doi: [10.1007/978-94-015-3600-4_8](https://doi.org/10.1007/978-94-015-3600-4_8).
- Nouvel, R., K.-H. Brassel, M. Bruse, E. Duminil, V. Coors, U. Eicker and D. Robinson (2015). 'SimStadt, a new workflow-driven urban energy simulation platform for CityGML city models'. In: *Proceedings of International Conference CIBAT 2015 Future Buildings and Districts Sustainability from Nano to Urban Scale*. LESO-PB, EPFL, pp. 889–894. url: <https://infoscience.epfl.ch/server/api/core/bitstreams/e06c9c38-35f3-4bcc-9a3c-154d1b430776/content>.
- Nouvel, R., M. Zirak, V. Coors and U. Eicker (July 2017). 'The influence of data quality on urban heating demand modeling using 3D city models'. In: *Computers, Environment and Urban Systems* 64. Publisher: Elsevier Ltd, pp. 68–80. issn: 01989715. doi: [10.1016/j.compenvurbsys.2016.12.005](https://doi.org/10.1016/j.compenvurbsys.2016.12.005).
- Nutkiewicz, A., B. Choi and R. K. Jain (2021). 'Exploring the influence of urban context on building energy retrofit performance: A hybrid simulation and data-driven approach'. In: *Advances in Applied Energy* 3 (May). Publisher: Elsevier Ltd, p. 100038. issn: 26667924. doi: [10.1016/j.adapen.2021.100038](https://doi.org/10.1016/j.adapen.2021.100038). url: <https://doi.org/10.1016/j.adapen.2021.100038>.

- O'Hegarty, R., O. Kinnane, D. Lennon and S. Colclough (2021). 'In-situ U-value monitoring of highly insulated building envelopes: Review and experimental investigation'. In: *Energy and Buildings* 252, p. 111447. issn: 0378-7788. doi: [10.1016/j.enbuild.2021.111447](https://doi.org/10.1016/j.enbuild.2021.111447).
- Object Vision (2023a). *Contact*. url: <https://www.objectvision.nl/about/> (visited on 05/07/2025).
- (2023b). *Home · ObjectVision/BAG-Tools Wiki*. BAG-Tools. url: <https://github.com/ObjectVision/BAG-Tools/wiki> (visited on 05/07/2025).
- Ohlsson, K. E. A., G. Nair and T. Olofsson (2022). 'Uncertainty in model prediction of energy savings in building retrofits: Case of thermal transmittance of windows'. In: *Renewable and Sustainable Energy Reviews* 168, p. 112748. issn: 1364-0321. doi: [10.1016/j.rser.2022.112748](https://doi.org/10.1016/j.rser.2022.112748).
- Oraopoulos, A. and B. Howard (Apr. 2022). 'On the accuracy of Urban Building Energy Modelling'. In: *Renewable and Sustainable Energy Reviews* 158, p. 111976. issn: 1364-0321. doi: [10.1016/j.rser.2021.111976](https://doi.org/10.1016/j.rser.2021.111976). (Visited on 22/01/2024).
- Oraopoulos, A., S. Hsieh and A. Schlueter (2023). 'Energy futures of representative Swiss communities under the influence of urban development, building retrofit, and climate change'. In: *Sustainable Cities and Society* 91, p. 104437. issn: 2210-6707. doi: [10.1016/j.scs.2023.104437](https://doi.org/10.1016/j.scs.2023.104437).
- Page, J. (19th Oct. 2007). 'Simulating occupant presence and Behaviour in Buildings'. PhD thesis. Lausanne: EPFL. 134 pp. url: <https://infoscience.epfl.ch/handle/20.500.14299/10000> (visited on 01/10/2024).
- Pantelios, K. (Jan. 2022). 'Development of a QGIS plugin for the CityGML 3D City Database'. Master. Delft: Technische Universiteit Delft. 91 pp. url: <https://repository.tudelft.nl/islandora/object/uuid%3Afb532bef-81b9-482b-921a-e7ce907cb544>.
- PBL (2025). *Vesta*. 4583. Den Haag: Planbureau voor de Leefomgeving. url: <https://www.pbl.nl/modellen/vesta> (visited on 04/08/2025).
- Pereira, R., M. Couto, F. Ribeiro, R. Rua, J. Cunha, J. P. Fernandes and J. Saraiva (May 2021). 'Ranking programming languages by energy efficiency'. In: *Science of Computer Programming* 205, p. 102609. issn: 0167-6423. doi: [10.1016/j.scico.2021.102609](https://doi.org/10.1016/j.scico.2021.102609). (Visited on 29/01/2025).
- Peters, R., B. Dukai, S. Vitalis, J. van Liempt and J. Stoter (2022). 'Automated 3D Reconstruction of LoD2 and LoD1 Models for All 10 Million Buildings of the Netherlands'. In: *Photogrammetric Engineering and Remote Sensing* 88.3. _eprint: 2201.01191, pp. 165–170. issn: 00991112. doi: [10.14358/PERS.21-00032R2](https://doi.org/10.14358/PERS.21-00032R2).
- Poon, H.-K. (18th Jan. 2024). 'Inferring the residential building type from 3DBAG'. MSc Thesis. Delft University of Technology. 70 pp. url: <https://resolver.tudelft.nl/uuid:3ef77acb-b38b-4fa5-9b1e-31813b00b739> (visited on 22/09/2024).
- PostGIS Project Steering Committee (2023). *PostGIS*. Version 3.5. url: <https://postgis.net/>.
- PostgreSQL Global Development Group (23rd Nov. 2024). *PostgreSQL*. Version 16.4. url: <https://www.postgresql.org/> (visited on 23/11/2024).
- Powalka, L., C. Poon, Y. Xia, S. Meines, L. Yan, Y. Cai, G. Stavropoulou, B. Dukai and H. Ledoux (2024). 'cjdb: A Simple, Fast, and Lean Database Solution for the CityGML Data Model'. In: *Recent Advances in 3D Geoinformation Science*. 3DGeoInfo. Ed. by T. H. Kolbe, A. Donaubauer and C. Beil. Springer Nature Switzerland, pp. 781–796. isbn: 978-3-031-43699-4. doi: [10.1007/978-3-031-43699-4_47](https://doi.org/10.1007/978-3-031-43699-4_47).
- Pries, M., V. Coors and M. Betz (2021). *Projekt: CityDoctor2*. Version 3.4.0. url: <https://projekt.bht-berlin.de/citydoctor2/> (visited on 02/08/2025).
- Python Software Foundation (2024). *The Python Language Reference*. Python documentation. url: <https://docs.python.org/3/reference/index.html> (visited on 23/11/2024).
- Raasveldt, M. and H. Muehleisen (2024). *duckdb*. Version 1.1.1. url: <https://github.com/duckdb/duckdb>.
- Reinhart, C. F. and C. Cerezo Davila (2016). 'Urban building energy modeling - A review of a nascent field'. In: *Building and Environment* 97. Publisher: The Authors, pp. 196–202. issn: 03601323. doi: [10.1016/j.buildenv.2015.12.001](https://doi.org/10.1016/j.buildenv.2015.12.001).
- Remmen, P., M. Lauster, M. Mans, M. Fuchs, T. Osterhage and D. Müller (2018). 'TEASER: an open tool for urban energy modelling of building stocks'. In: *Journal of Building Performance Simulation* 11.1, pp. 84–98. doi: [10.1080/19401493.2017.1283539](https://doi.org/10.1080/19401493.2017.1283539).
- Ren, J., X. Zhou, X. Jin, Y. Ye, F. Causone, M. Ferrando, P. Li and X. Shi (Oct. 2024). 'A systematic review of occupancy pattern in urban building energy modeling: From urban to building-scale'. In: *Journal of Building Engineering* 95, p. 110307. issn: 23527102. doi: [10.1016/j.jobe.2024.110307](https://doi.org/10.1016/j.jobe.2024.110307). (Visited on 07/11/2024).

- Rijksoverheid (2024a). *Bouwbesluit*. Besluit bouwwerken leefomgeving. url: <https://wetten.overheid.nl/BWBR0041297/2024-01-01> (visited on 11/09/2024).
- (2024b). *EP-Online*. Publication Title: Resources, tools and inspiration for buildings. url: <https://www.ep-online.nl/> (visited on 15/09/2024).
- (12th Nov. 2024c). *Regeling energieprestatie gebouwen*. Overheid.nl. Last Modified: 2024-02-28. url: <https://wetten.overheid.nl/BWBR0020921/2022-06-01> (visited on 12/11/2024).
- Rijkswaterstaat (22nd Dec. 2023). *Eerste deel van AHN 5 is beschikbaar!* AHN. Publisher: AHN. url: <https://www.ahn.nl/eerste-deel-van-ahn-5-is-beschikbaar> (visited on 17/12/2024).
- (2024). *Dataset: Actueel Hoogtebestand Nederland (AHN)*. Introductie - pdok. url: <https://www.pdok.nl/introductie/-/article/actueel-hoogtebestand-nederland-ahn> (visited on 01/03/2024).
- Romero Rodríguez, L., E. Duminil, J. Sánchez Ramos and U. Eicker (Apr. 2017). 'Assessment of the photovoltaic potential at urban level based on 3D city models: A case study and new methodological approach'. In: *Solar Energy* 146, pp. 264–275. issn: 0038092X. doi: [10.1016/j.solener.2017.02.043](https://doi.org/10.1016/j.solener.2017.02.043). url: <https://linkinghub.elsevier.com/retrieve/pii/S0038092X17301445> (visited on 13/02/2024).
- Rosknecht, M. and E. Airaksinen (2020). 'Concept and evaluation of heating demand prediction based on 3D city models and the CityGML energy ADE-case study Helsinki'. In: *ISPRS International Journal of Geo-Information* 9.10, pp. 1–19. issn: 22209964. doi: [10.3390/ijgi9100602](https://doi.org/10.3390/ijgi9100602).
- Roth, J., A. Martin, C. Miller and R. K. Jain (2020). 'SynCity: Using open data to create a synthetic city of hourly building energy estimates by integrating data-driven and physics-based methods'. In: *Applied Energy* 280 (October). Publisher: Elsevier Ltd, p. 115981. issn: 03062619. doi: [10.1016/j.apenergy.2020.115981](https://doi.org/10.1016/j.apenergy.2020.115981). url: <https://doi.org/10.1016/j.apenergy.2020.115981>.
- Roy, E., M. Pronk, G. Agugiaro and H. Ledoux (2023). 'Inferring the number of floors for residential buildings'. In: *International Journal of Geographical Information Science* 37.4, pp. 938–962. doi: [10.1080/13658816.2022.2160454](https://doi.org/10.1080/13658816.2022.2160454).
- RVO (2023). *Voorbeeldwoningen 2022 bestaande bouw*. Verantwoordingsrapportage W/E 30751 | RVO EPBD 190035. Rijksdienst voor Ondernemend Nederland. Eindhoven: Rijksdienst voor Ondernemend Nederland. url: <https://www.rvo.nl/onderwerpen/wetten-en-regels-gebouwen/voorbeeldwoningen-bestaande-bouw> (visited on 16/07/2024).
- (2024). *Energieprestatie - BENG*. url: <https://www.rvo.nl/onderwerpen/wetten-en-regels-gebouwen/beng> (visited on 12/05/2024).
- Sadeghipour Roudsari, M. and M. Pak (2013). 'Ladybug: A Parametric Environmental Plugin For Grasshopper To Help Designers Create An Environmentally-conscious Design'. In: *Building Simulation Conference Proceedings*. Chambéry France, pp. 3128–3135. url: https://www.ibpsa.org/proceedings/bs2013/p_2499.pdf.
- Sadourny, R., A. Arakawa and Y. Mintz (June 1968). 'Integration of the nondivergent barotropic vorticity equation with an icosahedral-hexagonal grid for the sphere'. In: *Monthly Weather Review* 96.6, pp. 351–356. issn: 0027-0644, 1520-0493. doi: [10.1175/1520-0493\(1968\)096<0351:IOTNBV>2.0.CO;2](https://doi.org/10.1175/1520-0493(1968)096<0351:IOTNBV>2.0.CO;2). (Visited on 24/10/2024).
- Safe Software (2024). *About Workspaces*. FME Workbench. url: https://docs.safe.com/fme/html/FME_Desktop_Documentation/FME_Desktop/Workbench/workspaces_about.htm (visited on 20/11/2024).
- Samareh Abolhassani, S., A. Zandifar, N. Ghourchian, M. Amayri, N. Bouguila and U. Eicker (June 2024). 'Occupant counting model development for urban building energy modeling using commercial off-the-shelf Wi-Fi sensing technology'. In: *Building and Environment* 258, p. 111548. issn: 03601323. doi: [10.1016/j.buildenv.2024.111548](https://doi.org/10.1016/j.buildenv.2024.111548). (Visited on 07/11/2024).
- Samet, H. (June 1984). 'The Quadtree and Related Hierarchical Data Structures'. In: *ACM Computing Surveys* 16.2, pp. 187–260. issn: 0360-0300, 1557-7341. doi: [10.1145/356924.356930](https://doi.org/10.1145/356924.356930). (Visited on 13/11/2024).
- Santhanavanich, T., R. Padsala, M. Rosknecht, S. Dabirian, M. M. Saad, U. Eicker and V. Coors (5th Dec. 2023). 'ENABLING INTEROPERABILITY OF URBAN BUILDING ENERGY DATA BASED ON OGC API STANDARDS AND CITYGML 3D CITY MODELS'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* X-1/W1-2023, pp. 97–105. issn: 2194-9050. doi: [10.5194/isprs-annals-X-1-W1-2023-97-2023](https://doi.org/10.5194/isprs-annals-X-1-W1-2023-97-2023). (Visited on 12/05/2024).

- Seyedzadeh, S., F. Pour Rahimian, P. Rastogi and I. Glesk (2019). 'Tuning machine learning models for prediction of building energy loads'. In: *Sustainable Cities and Society* 47, p. 101484. issn: 22106707. doi: [10.1016/j.scs.2019.101484](https://doi.org/10.1016/j.scs.2019.101484).
- Silva, A. S. and E. Ghisi (2014). 'Uncertainty analysis of user behaviour and physical parameters in residential building performance simulation'. In: *Energy and Buildings* 76, pp. 381–391. issn: 0378-7788. doi: [10.1016/j.enbuild.2014.03.001](https://doi.org/10.1016/j.enbuild.2014.03.001).
- Sokol, J., C. Cerezo Davila and C. F. Reinhart (Jan. 2017). 'Validation of a Bayesian-based method for defining residential archetypes in urban building energy models'. In: *Energy and Buildings* 134, pp. 11–24. issn: 0378-7788. doi: [10.1016/j.enbuild.2016.10.050](https://doi.org/10.1016/j.enbuild.2016.10.050). (Visited on 23/01/2024).
- Sood, D., I. Alhindawi, U. Ali, J. A. McGrath, M. A. Byrne, D. Finn and J. O'Donnell (Sept. 2023). 'Simulation-based evaluation of occupancy on energy consumption of multi-scale residential building archetypes'. In: *Journal of Building Engineering* 75, p. 106872. issn: 23527102. doi: [10.1016/j.jobe.2023.106872](https://doi.org/10.1016/j.jobe.2023.106872). (Visited on 22/01/2024).
- Souza, A. C. D. B. d., F. M. d. Vasconcelos, J. C. d. N. Azevedo, L. P. Muse, G. A. M. Moreira, J. V. d. R. Alves, M. E. d. L. Tostes, C. C. M. d. M. Carvalho and A. A. d. Nascimento (2024). 'Analysis of Energy Efficiency Measures in Integrating Light-Duty Electric Vehicles in NZEB Buildings: A Case Study in an Educational Facility in the Brazilian Amazon'. In: *Energies* 17.17. issn: 1996-1073. doi: [10.3390/en17174343](https://doi.org/10.3390/en17174343).
- Stoter, J., K. Arroyo Otori, B. Dukai, A. Labetski, K. Kumar, S. Vitalis and H. Ledoux (2020). 'State of the Art in 3D City Modelling - Six Challenges Facing 3D Data as a Platform'. In: *GIM International*, pp. 4–7. url: <https://www.gim-international.com/content/article/state-of-the-art-in-3d-city-modelling-2>.
- Stuart-Fox, M., T. Kleinepijper, D. Ligthart and B. Blijie (Oct. 2022). *Wonen langs de meetla: Resultaten van het WoonOnderzoek Nederland 2021*. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, p. 125. url: <https://www.woononderzoek.nl/> (visited on 13/09/2024).
- Swan, L. G. and V. I. Ugursal (Oct. 2009). 'Modeling of end-use energy consumption in the residential sector: A review of modeling techniques'. In: *Renewable and Sustainable Energy Reviews* 13.8, pp. 1819–1835. issn: 1364-0321. doi: [10.1016/j.rser.2008.09.033](https://doi.org/10.1016/j.rser.2008.09.033).
- The World Bank Group (2024). *World Development Indicators | DataBank*. url: <https://databank.worldbank.org/Urban-2023/id/1623846f> (visited on 10/05/2024).
- Tufan, Ö., K. Arroyo Otori, C. León-Sánchez, G. Agugiaro and J. Stoter (2022). 'Development and testing of the CityJSON energy extension for space heating demand calculation'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 17th 3D GeoInfo Conference. Vol. XLVIII-4/W4-2022. Place: Sydney, Australia. Sydney, Australia: Copernicus Publications, pp. 169–176. doi: [10.5194/isprs-archives-XLVIII-4-W4-2022-169-2022](https://doi.org/10.5194/isprs-archives-XLVIII-4-W4-2022-169-2022).
- Tufan, Ö. (2022). 'Development and Testing of the CityJSON Energy Extension for Space Heating Demand Calculation'. Master. Delft: Technische Universiteit Delft. url: <https://repository.tudelft.nl/islandora/object/uuid:0061d94c-46b6-4910-bb73-b6cdc9b85f84>.
- Uber Technologies (23rd June 2018). *H3: A Hexagonal Hierarchical Geospatial Indexing System*. url: <https://github.com/uber/h3> (visited on 20/10/2024).
- United Nations (2015). *Goal 11: Make cities inclusive, safe, resilient and sustainable*. Cities - United Nations Sustainable Development Action. url: <https://www.un.org/sustainabledevelopment/cities/> (visited on 12/11/2023).
- Utrilla Guerrero, C. and M. Barrera-Coronel (2025). *RDM-Module 3: Documentation and Metadata*. The Research Data Management. url: <https://collegerama.tudelft.nl/Mediasite/Play/d7bd16131769465d8b1307ec27358b3e1d> (visited on 02/08/2025).
- van Bijsterveld, W. (1st Aug. 2025). *Gemeente Delft in cijfers en grafieken | AlleCijfers.nl*. AllCharts.info. url: <https://allecijfers.nl/gemeente/delft/> (visited on 02/08/2025).
- Van Bueren, E., H. Van Bohemen, L. Itard and H. Visscher, eds. (2012). *Sustainable Urban Environments: An Ecosystem Approach*. Dordrecht: Springer Netherlands. isbn: 978-94-007-1293-5 978-94-007-1294-2. doi: [10.1007/978-94-007-1294-2](https://doi.org/10.1007/978-94-007-1294-2). url: <https://link.springer.com/10.1007/978-94-007-1294-2> (visited on 24/06/2025).
- van den Brom, P. (2020). 'Energy in Dwellings A comparison between Theory and Practice'. ISBN: 9789463662536, ISSN: 22123202. PhD thesis. Delft University of Technology. doi: [10.7480/abe.2020.3](https://doi.org/10.7480/abe.2020.3).
- Van der Linden, A. C., I. M. Kuijpers-Van Gaalen and A. Zeegers (2018). *Building Physics*. 2nd ed. isbn: 978-90-06-10612-1.

- Van der Molen, F., J. Langeveld, C. Herbstritt, W. Poorthuis and N. Hoogervorst (2023). *Overzicht Transitievisies Warmte: Signalen, Obstakels & Potentieel*. 5051. Den Haag: PBL Planbureau voor de Leefomgeving.
- Van der Molen, F., W. Poorthuis, A. Zwamborn, C. Tigchelaar, R. Niessink and V. Rovers (2023). *Functioneel ontwerp HESTIA 1.0*. 5196. PBL Planbureau voor de Leefomgeving.
- Van der Vaart, J., J. Stoter, A. Diakité, F. Biljecki, K. A. Ohori and A. Hakim (2024). 'Assessment of the LoD Specification for the Integration of BIM-Derived Building Models in 3D City Models'. In: *Recent Advances in 3D Geoinformation Science*. Ed. by T. H. Kolbe, A. Donaubauer and C. Beil. Cham: Springer Nature Switzerland, pp. 171–191. isbn: 978-3-031-43699-4.
- Vargova, A., S. Köhler, S. Hötzel, B. Schröter, Z. Vranayova and D. Kaposztasova (19th Dec. 2023). 'Transformation of Urban Spaces: The Impact of Green Roofs in Košice, Slovakia'. In: *Sustainability* 16.1, p. 22. issn: 2210-1050. doi: [10.3390/su16010022](https://doi.org/10.3390/su16010022). (Visited on 12/05/2024).
- Varrazzo, D. (2023). *psycpg*. Postgresql driver for Python - Psycpg. url: <https://www.psycpg.org> (visited on 21/10/2024).
- Vázquez-Canteli, J. R., S. Ulyanin, J. Kämpf and Z. Nagy (2019). 'Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities'. In: *Sustainable Cities and Society* 45, pp. 243–257. issn: 2210-6707. doi: [10.1016/j.scs.2018.11.021](https://doi.org/10.1016/j.scs.2018.11.021).
- Vethman, P., C. Volkers, C. Tigchelaar and M. Menkveld (2019). *Korte modelbeschrijving SAWEC*. 3839. Den Haag: Planbureau voor de Leefomgeving.
- Vlaamse overheid (2025). *Woningpas*. Vlaanderen.be. url: <https://www.vlaanderen.be/bouwen-wonen-en-energie/bouwen-en-verbouwen/woningpas> (visited on 17/02/2025).
- Voortman, J. (2021). *GIS laag groenvolume (2D) uit AHN3 lidar (3D)*. Accessed: 2022-03-18. url: <https://www.linkedin.com/pulse/gis-laag-groenvolume-2d-uit-ahn3-lidar-3d-joop-voortman>.
- Wang, C., M. Ferrando, F. Causone, X. Jin, X. Zhou and X. Shi (June 2022). 'Data acquisition for urban building energy modeling: A review'. In: *Building and Environment* 217, p. 109056. issn: 03601323. doi: [10.1016/j.buildenv.2022.109056](https://doi.org/10.1016/j.buildenv.2022.109056). (Visited on 01/09/2024).
- Wang, C., B. Duplessis, E. Peirano, P. Schetelat and P. Riederer (2024). 'Addressing Data Scarcity in UDEM Validation: Application of Survey Sampling Techniques'. In: .
- Wang, D., J. Landolt, G. Mavromatidis, K. Orehounig and J. Carmeliet (June 2018). 'CESAR: A bottom-up building stock modelling tool for Switzerland to address sustainable energy transformation strategies'. In: *Energy and Buildings* 169, pp. 9–26. issn: 0378-7788. doi: [10.1016/j.enbuild.2018.03.020](https://doi.org/10.1016/j.enbuild.2018.03.020). (Visited on 23/01/2024).
- Wang, R., S. Lu and W. Feng (2020). 'A novel improved model for building energy consumption prediction based on model integration'. In: *Applied Energy* 262, p. 114561. issn: 03062619. doi: [10.1016/j.apenergy.2020.114561](https://doi.org/10.1016/j.apenergy.2020.114561).
- Wang, Y., K. Liu, Y. Liu, D. Wang and J. Liu (2022). 'The impact of temperature and relative humidity dependent thermal conductivity of insulation materials on heat transfer through the building envelope'. In: *Journal of Building Engineering* 46, p. 103700. issn: 2352-7102. doi: [10.1016/j.jobbe.2021.103700](https://doi.org/10.1016/j.jobbe.2021.103700).
- Wysocki, O., B. Schwab, C. Beil, C. Holst and T. H. Kolbe (2024). 'Reviewing Open Data Semantic 3D City Models to Develop Novel 3D Reconstruction Methods'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 48.4, pp. 493–500. doi: [10.5194/isprs-archives-XLVIII-4-2024-493-2024](https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-493-2024).
- Wysocki, O., Y. Xia, M. Wysocki, E. Grilli, L. Hoegner, D. Cremers and U. Stilla (June 2023). 'Scan2LoD3: Reconstructing semantic 3D building models at LoD3 using ray casting and Bayesian networks'. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver, BC, Canada: IEEE, pp. 6548–6558. isbn: 9798350302493. doi: [10.1109/CVPRW59228.2023.00696](https://doi.org/10.1109/CVPRW59228.2023.00696). (Visited on 24/02/2025).
- Xia, Y. (2023). 'A data-driven approach to add openings to 3D BAG building models'. Master. Delft University of Technology. 64 pp. url: <https://resolver.tudelft.nl/uuid:eaff38b9-65f4-4c1c-8cec-aebd8fc4b18f> (visited on 05/02/2025).
- Xu, L. (29th Oct. 2024). 'High-resolution, large-scale, and fast calculation of solar irradiance with 3D City Models'. MSc Thesis. Delft University of Technology. 88 pp. url: <https://resolver.tudelft.nl/uuid:d97174a3-3e1b-4e64-8de0-33a6ec2c48b2> (visited on 13/11/2024).

- Xu, L., C. León-Sánchez, G. Agugiaro and J. Stoter (2024a). 'High resolution solar potential computation in large scale urban areas by means of semantic 3D city models'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-4/W11-2024, pp. 167–174. doi: [10.5194/isprs-archives-XLVIII-4-W11-2024-167-2024](https://doi.org/10.5194/isprs-archives-XLVIII-4-W11-2024-167-2024).
- Xu, L., C. León-Sánchez, G. Agugiaro and J. Stoter (2024b). 'Shadowing Calculation on Urban Areas from Semantic 3D City Models'. In: 18th 3D GeoInfo Conference. Ed. by T. H. Kolbe, A. Donaubauer and C. Beil. Recent Advances in 3D Geoinformation Science. Cham: Springer Nature Switzerland, pp. 31–47. isbn: 978-3-031-43699-4. doi: [10.1007/978-3-031-43699-4_2](https://doi.org/10.1007/978-3-031-43699-4_2).
- Yang, Y., Q. Gu, H. Wei, H. Liu, W. Wang and S. Wei (Sept. 2023). 'Transforming and validating urban microclimate data with multi-sourced microclimate datasets for building energy modelling at urban scale'. In: *Energy and Buildings* 295, p. 113318. issn: 03787788. doi: [10.1016/j.enbuild.2023.113318](https://doi.org/10.1016/j.enbuild.2023.113318). (Visited on 07/11/2024).
- Yao, Z. and C. Nagel (15th Dec. 2022). *ADE Manager Plugin*. Version 2.2.0. Publication Title: Github. Munich. url: <https://github.com/3dcitydb/plugin-ade-manager> (visited on 30/08/2024).
- Yao, Z., C. Nagel, F. Kunde, G. Hudra, P. Willkomm, A. Donaubauer, T. Adolphi and T. H. Kolbe (2018). '3DCityDB - a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML'. In: *Open Geospatial Data, Software and Standards* 3.1. ISBN: 4096501800467. issn: 2363-7501. doi: [10.1186/s40965-018-0046-7](https://doi.org/10.1186/s40965-018-0046-7).
- Yoon, S. (2020). 'In-situ sensor calibration in an operational air-handling unit coupling autoencoder and Bayesian inference'. In: *Energy and Buildings* 221, p. 110026. issn: 0378-7788. doi: [10.1016/j.enbuild.2020.110026](https://doi.org/10.1016/j.enbuild.2020.110026).
- Yu, J., W.-S. Chang and Y. Dong (2022). 'Building Energy Prediction Models and Related Uncertainties: A Review'. In: *Buildings* 12.8. issn: 2075-5309. doi: [10.3390/buildings12081284](https://doi.org/10.3390/buildings12081284).

LIST OF PUBLICATIONS

12. C. León-Sánchez, G. Agugiaro and J. Stoter (Sept. 2025). 'Lessons learnt from the integration of open data and semantic 3D city models for urban building energy modelling in the Netherlands'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 20th 3D GeoInfo & 9th Smart Data Smart Cities Conference. Tokyo, Japan: Copernicus Publications. Best Young Researcher Award at the 9th International Smart Data and Smart Cities conference 2025
11. C. León-Sánchez, D. Giannelli, G. Agugiaro and J. Stoter (2025). 'Comparative analysis of geospatial tools for solar simulation'. In: *Transactions in GIS* 29.1, e13296. doi: [10.1111/tgis.13296](https://doi.org/10.1111/tgis.13296)
10. C. Bachert, C. León-Sánchez, T. Kutzner and G. Agugiaro (2024). 'Mapping the CityGML Energy ADE to CityGML 3.0 Using a Model-Driven Approach'. In: *ISPRS International Journal of Geo-Information* 13.4. issn: 2220-9964. doi: [10.3390/ijgi13040121](https://doi.org/10.3390/ijgi13040121). url: <https://www.mdpi.com/2220-9964/13/4/121>
9. L. Xu, C. León-Sánchez, G. Agugiaro and J. Stoter (2024a). 'High resolution solar potential computation in large scale urban areas by means of semantic 3D city models'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-4/W11-2024*, pp. 167–174. doi: [10.5194/isprs-archives-XLVIII-4-W11-2024-167-2024](https://doi.org/10.5194/isprs-archives-XLVIII-4-W11-2024-167-2024)
8. G. Agugiaro, K. Pantelios, C. León-Sánchez, Z. Yao and C. Nagel (2024). 'Introducing the 3DCityDB-Tools Plug-In for QGIS'. in: *Recent Advances in 3D Geoinformation Science*. 3DGeoInfo 2023. Ed. by T. H. Kolbe, A. Donaubaue and C. Beil. Munich: Springer Nature Switzerland, pp. 797–821. isbn: 978-3-031-43699-4
7. L. Xu, C. León-Sánchez, G. Agugiaro and J. Stoter (2024b). 'Shadowing Calculation on Urban Areas from Semantic 3D City Models'. In: 18th 3D GeoInfo Conference. Ed. by T. H. Kolbe, A. Donaubaue and C. Beil. Recent Advances in 3D Geoinformation Science. Cham: Springer Nature Switzerland, pp. 31–47. isbn: 978-3-031-43699-4. doi: [10.1007/978-3-031-43699-4_2](https://doi.org/10.1007/978-3-031-43699-4_2)
6. G. Agugiaro, A. Zwamborn, C. Tigchelaar, E. Matthijssen, C. León-Sánchez, F. Van Der Molen and J. Stoter (2022). 'on the Influence of Party Walls for Urban Energy Modelling'. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 48.4, pp. 9–16. issn: 16821750. doi: [10.5194/isprs-archives-XLVIII-4-W5-2022-9-2022](https://doi.org/10.5194/isprs-archives-XLVIII-4-W5-2022-9-2022)
5. C. León-Sánchez, G. Agugiaro and J. Stoter (Oct. 2022). 'Creation of a CityGML-based 3D city model testbed for energy-related applications'. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. 7th International Conference on Smart Data and Smart Cities (SDSC). vol. 48.

Issue: 4/W5-2022. Sydney, Australia: Copernicus Publications, pp. 97–103. isbn: 16821750. doi: [10.5194/isprs-archives-XLVIII-4-W5-2022-97-2022](https://doi.org/10.5194/isprs-archives-XLVIII-4-W5-2022-97-2022)

4. Ö. Tufan, K. Arroyo Otori, C. León-Sánchez, G. Agugiaro and J. Stoter (2022). 'Development and testing of the CityJSON energy extension for space heating demand calculation'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 17th 3D GeoInfo Conference. Vol. XLVIII-4/W4-2022. Place: Sydney, Australia. Sydney, Australia: Copernicus Publications, pp. 169–176. doi: [10.5194/isprs-archives-XLVIII-4-W4-2022-169-2022](https://doi.org/10.5194/isprs-archives-XLVIII-4-W4-2022-169-2022)
3. D. Giannelli, C. León-Sánchez and G. Agugiaro (2022). 'Comparison and evaluation of different GIS software tools to estimate solar irradiation'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXIV ISPRS Congress. Vol. 5. Nice, France: Copernicus Publications, pp. 275–282. doi: [10.5194/isprs-Annals-V-4-2022-275-2022](https://doi.org/10.5194/isprs-Annals-V-4-2022-275-2022)
2. C. León-Sánchez, D. Giannelli, G. Agugiaro and J. Stoter (2021). 'Testing the new 3D BAG dataset for energy demand estimation of residential buildings'. In: *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. 6th International Conference on Smart Data and Smart Cities. Vol. XLVI-4/W1-2021. Stuttgart: Copernicus Publications, pp. 69–76. doi: [10.5194/isprs-archives-XLVI-4-W1-2021-69-2021](https://doi.org/10.5194/isprs-archives-XLVI-4-W1-2021-69-2021)
1. T. Q. Doan, C. León-Sánchez, R. Peters, G. Agugiaro and J. Stoter (2021). 'Volume Comparison of Automatically Reconstructed Multi-Lod Building Models for Urban Planning Applications'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. XXIV ISPRS Congress. Vol. V-4-2021. Copernicus Publications, pp. 169–176. doi: [10.5194/isprs-annals-v-4-2021-169-2021](https://doi.org/10.5194/isprs-annals-v-4-2021-169-2021). url: <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/V-4-2021/169/2021/>

CURRICULUM VITÆ

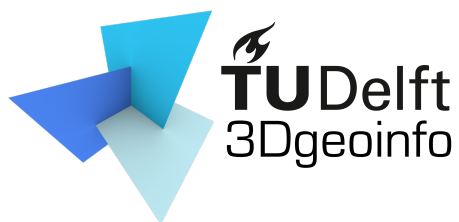
Camilo Alexander León Sánchez was born on 6 March 1983 in Bogotá, Colombia. He obtained his high school diploma in 1999 at Externado Nacional Camilo Torres and earned a BSc in Cadastre Engineering and Geodesy from the Francisco José de Caldas District University in 2006.

In 2013, Camilo completed an MSc in Geodesy and Geoinformation Sciences at the Berlin University of Technology, with a thesis titled *Estimation of Electric Energy Demand using 3D City Models*. He also holds an MSc in Information Sciences and Communications from the Francisco José de Caldas District University, where his thesis was titled *Evaluation of digital techniques for coffee crops identification from multispectral imagery*.

He has worked as a GIS specialist for the Colombian Coffee Growers Federation within the Coffee Information System (SICA) and for the Spatial Data Infrastructure of Bogotá (IDECA). From 2015 to 2020, he served as a university lecturer in Geomatics at several institutions, including the University of Applied and Environmental Sciences (Universidad de Ciencias Aplicadas y Ambientales—UDCA—), the Military University (Universidad Militar Nueva Granada), Sergio Arboleda University, and the National University of Colombia in Bogotá.

In October 2020, Camilo began his PhD research on semantic 3D city models for energy applications, supervised by Prof. Jantien Stoter and Dr. Giorgio Agugiaro at Delft University of Technology.

In 2025, Camilo started working as a postdoctoral researcher within the European project DigiTwins4PEDs (Utilisation of urban digital twins to co-create flexible positive energy systems for districts) at Delft University of Technology. 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.



ISBN 978-94-6518-115-8



