Sustainable Cities and Society 101 (2024) 105075

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Data-driven classification of Urban Energy Units for district-level heating and electricity demand analysis

Luis Blanco [a,b,*], Alaa Alhamwi [c], Björn Schiricke [a], Bernhard Hoffschmidt [a,b]

[a] German Aerospace Center (DLR), Institute of Solar Research, Cologne, Germany
[b] RWTH Aachen University, Chair of Solar Components, Aachen, Germany
[c] German Aerospace Center (DLR), Institute of Networked Energy Systems, Oldenburg, Germany

ARTICLE INFO

Keywords:
Urban Energy Units
Energy district
Urban planning
Machine learning
Open-source
GIS

ABSTRACT

The building sector is a significant contributor to global energy consumption and accounts for approximately one-third of total greenhouse gas emissions. While building energy analysis has traditionally focused on individual buildings, analyzing larger settlements, such as districts or neighbors, offers additional opportunities. The objective of this study is to define and classify typical urban areas for energy analysis, referred to in this paper as Urban Energy Units (UEUs), which represent geographical regions within a city with specific building's characteristics, settlement patterns and energy demand. Sixteen different UEUs were classified using literature and open data. The proposed methodology leverages open-source data and uses a random forest model to enhance missing building properties of the building stock such as building age and construction type. It further subdivides the study area into geographically defined sections, and deploys a decision tree model to classify these sections into the sixteen different UEUs. These UEUs enable the creation of energy districts in a modular manner and flexible for its use in any given area. This study demonstrates the practical implications related to the 2023 german municipality heating plan. The methodology was applied in Oldenburg, a mid-sized German city. The city was subdivided into a total of 8249 UEUs, with the detailed results for energy demand presented in this report.

1. Introduction

Cities are responsible for up to 80% of the global energy consumption (DESA, 2019). A third of the total greenhouse gas (GHG) emissions are related to the building sector, making it an important target for reducing urban energy consumption. In Germany, heating and hot water systems account for 84% of the final energy consumption of the residential sector and contribute to almost a third of the country's GHG emissions (IEA, 2020; BPIE, 2015; Statista Search Department, 2022a, 2022b). Germany recently presented a new building energy law for the municipality heating plan (BMWK, 2020; BMWSB, 2023), where municipalities are to submit plans for the conversion of the heating infrastructure to achieve neutrality by 2045. But drawing up a heating plan takes time, requires a large amount of highly qualified staff and incurs high costs. Many cities are still in the early stages of municipal heating planning and are analyzing how high the heat demand is in the neighborhoods, where a strategic expansion of district heating and where a decentralized supply, for example via heat pumps, could be applied (Städtetag, 2023). However, many small communities lack the resources mentioned above and are trying to find more effective solutions.

Deploying energy efficiency strategies can help reduce energy-related emissions from the building sector and accelerate the community heating plan in Germany. One of these strategies is to move the energy performance targets away from individual buildings towards a district level. This is because energy analysis of larger settlement units or districts offers more opportunities than the isolated energetic refurbishment of individual buildings (Ahlers, Driscoll, Wibe, & Wyckmans, 2019; Dettmar, Drebes, & Sieber, 2020; Konstantinou & Knaack, 2011; Shnapp, Paci, & Bertoldi, 2020). Moving from individual buildings towards a district level approach could bring more opportunities in the areas of: energy consumption analysis, urban planning, cost-effective methods for high energy efficiency, coupling of the energy grid with renewable energy systems and consumption behavior analysis. Further, energy efficiency measures on an individual building will reduce the overall demand of the district, coupling these options with technologies, such as renewable energy systems, local energy networks and

* Corresponding author at: German Aerospace Center (DLR), Institute of Solar Research, Cologne, Germany.
E-mail addresses: luis.blancobohorquez@dlr.de (L. Blanco), Alaa.Alhamwi@dlr.de (A. Alhamwi), Bjoern.Schiricke@dlr.de (B. Schiricke), Bernhard.Hoffschmidt@dlr.de (B. Hoffschmidt).

https://doi.org/10.1016/j.scs.2023.105075
Received 17 August 2023; Received in revised form 13 October 2023; Accepted 16 November 2023
Available online 24 November 2023

energy storage can offer overall cost-effective solutions (Shnapp et al., 2020).

This set-up requires holistic and smart planning. The problem arises when trying to define the concept of a district or neighborhood within a city. There is no general definition of a district as the concept may vary according to quantitative and/or qualitative properties (such as administrative areas, construction building types, and social or cultural characteristics). The conventional thinking of a district – as an administrative region within a city – is not enough when trying to integrate sustainable energy strategies and energy management approaches in urban planning in order to minimize overall energy consumption and CO_2 emissions.

The following study focuses on geographically defining existing urban areas based on shared building and architectural traits. Instead of defining future energy districts, we will focus on subdividing the existing urban space into geographical units that contain typical building characteristics, settlement patterns, and energy demands. These geographical units will be referred to in this paper as **Urban Energy Units (UEUs)**. Each UEU characterizes a specific geographical area with distinct building features, settlement patterns, and energy demands. The UEUs will serve as the building blocks for creating energy districts as they can be flexibly combined to define larger geographic areas, or districts, each possessing its own geographic boundaries, building features, settlement patterns, and energy requirements.

The main objective of this study is to develop an automated GIS-based methodology for subdividing any study area into UEUs based on large-scale open-source data. Statistical and machine learning (ML) tools will be used to classify urban areas into 16 different UEUs, each with its own typical building characteristics and settlement patterns. The main advantage of such methodology is, that it provides a standardized framework for future energy district analysis by establishing standardized areas (UEUs) that can be combined in a modular way to create various district types, adaptable to different geographic contexts and operational needs. The study's objectives can be summarized in the following tasks:

- Generate a building stock database for a study area using available open-source data.
- Employ various ML and statistical methods to divide the study area into 16 different UEUs using the generated building stock and assign corresponding energy demand values.
- Combine UEUs to construct energy districts within the study area.
- Understand the relevance of the UEUs in real-world scenarios, particularly concerning the municipality heating plan of Germany.

2. Literature review

There is no universally agreed-upon definition for the term "district"[1] in the urban context. Accurately defining districts based on energy consumption in a standardized manner poses challenges, yet it is a critical task for modern urban planning. It provides a structured framework for efficient resource allocation, infrastructure development, and policy-making (Shnapp et al., 2020). The ultimate goal is to reduce CO_2 emissions, enhance urban livability, and promote sustainability. In this section, we offer a concise overview of the existing literature on the concept of districts in urban planning. We investigate GIS-based and data-driven approaches proposed by other authors. Furthermore, we underscore the importance of districts in the context of recent legal changes in Germany, highlighting their central role in contemporary urban planning.

[1] In the urban context the term "district" is often interchanged with the term "neighborhood", in this paper we will only use the first.

2.1. Urban context

The urban environment is the physical and social habitat for almost half of mankind in the present day world. In many regions of the world almost three quarters of the population live in towns and cities. The urban environment is highly diverse and dynamic and has strongly contributed to many aspects of our lives and culture such as social units, market trade, democracy and in today's perspective our energy consumption as well (Boerefijn et al., 2010; Bruegmann, 2019; Couch, Petschel-Held, & Leontidou, 2008; Hipp, Faris, & Boessen, 2012).

Many cities around the world struggle nowadays to transform their urban spaces in order to achieve their sustainability goals. Achieving sustainable urban development is a hugely difficult task to accomplish due to its complex and continuously-evolving nature. Initially, the focus of sustainability assessments within the built environment was individual buildings, the smallest urban units of a human settlement (Lützkendorf & Balouktsi, 2017). However, the sustainability of the urban environment poses much more complex issues than solely the performance of single buildings. Thus, the need to expand the sustainability scope and metrics to larger scales of the built environment leading to a more targeted focus on neighborhoods or districts as the appropriate geographic scale for intervention. Authors like Adams (1994), Berardi (2013), Elci, Delgado, Henning, Henze, and Herkel (2018), Lützkendorf and Balouktsi (2017) have highlighted the advantages of planning and implementing sustainability principles at the district scale. They argue that planning at this scale allows for the formulation of efficient strategies that integrate various factors, including, designing of public and private spaces, optimal transportation and energy supply, and the dynamics of community interaction. Balouktsi, Lützkendorf, Kopfmüller, and Steltzer (2017) explain how the subdivision of the urban system into smaller units, such as districts, can facilitate more targeted sustainability transformations. Nevertheless, they also argue that district-scale sustainable development is an ongoing process that demands continuous engagement, monitoring, assessment, and adaptation.

In terms of energy analysis, a district turns out to be of great interest. Districts align well with the scale of small or medium-sized energy supply and distribution systems, making them a logical unit for urban development plans in the context of energy planning (Elci et al., 2018; Koch, 2010). A comprehensive review of the challenges and obstacles in energy planning at the urban scale has been presented by Cajot et al. (2017). Understanding energy demand at the district scale enables the design of efficient energy management systems, which have gained relevance with the increasing of decentralized, fluctuating renewable energy systems, such as photovoltaic modules, solar–thermal systems and heat pumps.

The concept of districts in urban planning represents a pivotal shift in addressing the complexities of sustainability within the urban environment. It acknowledges the multifaceted nature of urban development and provides a structured framework for sustainable transformation, from efficient resource allocation to energy-conscious planning.

2.2. GIS-based and data driven approaches

Urban planning is one of the main applications of Geographical Information Systems (GIS). Urban planners use GIS both as a spatial database and as an analysis and modeling tool (Yeh, 1999). With the increase in user-friendliness and functions of GIS software and the marked decrease in the prices of GIS hardware, GIS is an operational and affordable information system for planning. The main constraints in the use of GIS in urban planning today are not technical issues, but the availability of data, organizational change, and staffing.

In terms of Urban Energy Modeling, GIS-based and data-driven models can be effectively used to assess and optimize energy-related processes within urban areas (Manfren, Nastasi, Groppi, & Garcia, 2020). These tools enable the integration of geospatial data such as

Table 1
GIS-based and data-driven building energy modeling studies compared in this paper in terms of urban scale, approaches and application.

Authors	Urban scale	Approaches	Application
Koch (2010)	Districts	Engineering-based, Statistical	Energy efficient planning
Hipp et al. (2012)	District	Machine learning	Social district boundaries
Nouvel et al. (2015)	Building, District, City	Engineering-based, Statistical	Heat consumption models
Quan, Li, Augenbroe, Brown, and Yang (2015)	Building, City	Engineering-based	Energy modeling
Li, Quan, and Yang (2016)	District, City	Engineering-based	Energy performance simulation
Ma and Cheng (2016)	City	Machine learning	Energy use intensity
Yamamura, Fan, and Suzuki (2017)	City	Engineering-based	Energy performance
Chen, Hong, and Piette (2017)	Building, City	Engineering-based	Retrofit analysis
Moghadam, Toniolo, Mutani, and Lombardi (2018)	Building, City	Statistical	Built environment energy use
Groppi, de Santoli, Cumo, and Garcia (2018)	Building, City	Statistical	Energy consumption and solar potential
Nutkiewicz, Yang, and Jain (2018)	Building, District	Machine learning	Urban energy simulation
Zheng and Weng (2019)	Counties	Engineering-based	Energy demand
Ahn and Sohn (2019)	District, City	Statistical	Energy use
Ali et al. (2020)	Building, City	Machine learning, Statistical	Planning and decision making
El Kontar, Polly, Charan, Fleming, Moore, Long, and Goldwasser (2020)	Building, District	Statistical	Software development
Wurm et al. (2021)	Building, City	Machine learning	Heat demand modeling
Garbasevschi et al. (2021)	Building, City	Machine learning	Building stock generation

building locations, land use, topography, and climate information, which are crucial for accurately modeling energy consumption and distribution patterns. Table 1 shows a comparison of some GIS-based and data-driven studies reviewed for this paper. Engineering-based methods use synthetic experimental data, implement a limited number of typologies, and there are numerous assumptions embedded in energy simulations which directly affect the accuracy of results. Data-driven approaches, on the other hand, do not require detailed knowledge about the building as these approaches estimate building energy performance based on historical data either using statistical or machine learning models (Abbasabadi & Ashayeri, 2019), nevertheless there is no standardization of such methods, and hence limit their use in policy-making. While statistical models use sample data about buildings to build a mathematical relationship between the building's energy consumption and characteristics, machine learning models implement algorithms that learn from data to predict building energy performance with minimal assumptions.

This study implements both data-driven (ML) and a GIS-based approach at a large scale, few studies do that as seen in Table 1. The methodology is explained in detail in the following sections. However, three important sources of the literature were key for this research. The first study is that from Wurm et al. (2021), in this study the authors developed a workflow for deep learning-based building stock modeling using aerial images at city scale for heat-demand modeling, this study showed the advantages of using data-driven approach, specifically the use of Convolutional Neural Network (CNN), in large urban scale in Germany, however the authors encounter one key problem with the particular data needed, the information about the German building stock used was found in a grid format, which meant that the outcome of the heat models was also in grid format. The second study was that from Garbasevschi et al. (2021), here, the authors developed a ML model, specifically a Random Forest (RF) model, to take the same data found in grid format a disaggregate the information for the individual buildings depending on the geometrical features of the buildings, they focused on the parameter of the building age or construction period and showed how the prediction of building individual age influenced the heat demand, for the heat demand they used a simple calculation due to the fact that the focus was on the ML disaggregate model.

The third study, out of which this paper takes great inspiration, is that from Dettmar et al. (2020). In their work, the authors define three big groups for types of urban units: Settlement areas, open spaces, and single elements. Due to a historically similar urban development, settlement areas in Germany show homogeneous and recurring structural and technical characteristics. The authors classify these areas into 14 definable building structures from 1 to 10, with subdivisions such as 1.a, 1.b etc. Open spaces are the second main group, a total of 15 different open spaces identified by the authors, and these are undeveloped areas in the urban space such as: public green spaces, traffic areas, water bodies and rural areas. Single elements such as special buildings in the urban context can also play a significant role from an energy perspective, the authors identified 23 different single elements such as educational buildings, hospitals, cinemas, sport fields etc. These are, mostly, non-residential buildings used for mono-functional purposes. The authors define morphological characteristics for all of the urban units of type settlement areas and this characteristics present a great value for the present study. The settlement areas and open spaces were denominated by Dettmar et al. (2020) as ESTs and the individual elements as EE. More about these urban units and their characteristics is explained in Section 3.

2.3. Significance in today's German context

The German government has ambitious targets for reducing emissions related to final energy consumption in the building sector. They aim to cut emissions by 50% by 2030 and make the building stock nearly climate-neutral by 2045 (BMWK, 2019b). This involves the goal of feeding 30% of heat networks from renewables or waste heat by 2030 and increasing it to 80% by 2040. To achieve this, the government employs two main instruments: the first instrument is the Building Energy Act (known as Gebäudeenergiegesetz in German), which includes the guidelines for energy retrofitting and energy performance standards for the building stock (BMWK, 2019a, 2020, 2023). The second instrument is the municipal heat planning introduced in August 2023, requiring municipalities to outline their heating infrastructure conversion plans for neutrality by 2045 (BMWSB, 2023).

Municipal heat plans are to be developed on a mandatory basis and comprise four phases. The inventory analysis phase assesses the current state of buildings, energy infrastructure, and consumption. The potential analysis phase explores technical-economic climate protection possibilities like district heating and heat pumps. The 2040 target scenario phase establishes conditions for a climate-neutral building stock by 2040, and the heat transition strategy phase focuses on concrete measures. Developing these plans is time-consuming, demanding qualified personnel and incurring higher costs. Many cities and municipalities are still in the early stages and overwhelmed by the task (Städtetag, 2023).

To effectively address energy consumption in the German building sector and expedite community heating plans, realistic models that consider urban areas and energy consumption are crucial. This paper introduces a GIS and data-driven methodology that speeds up the initial phase of German municipal heat planning by using available data to divide the urban space into smaller urban units.

3. Urban energy units

As stated in the previous sections, although there is no universally accepted definition of the term “district”, smart urban planning needs to integrate new ideas and concepts for districts analysis in order to develop sustainable energy strategies and minimize CO_2 emissions in urban areas.

This study takes a step back from the conventional approach of energetically defining the future state of a district. Instead, our focus shifts to the geographical delineation of pre-existing regions within a city. These regions share common building and architectural parameters, enabling the assignment of typical energy consumption values. Our approach transitions from an individual building-oriented viewpoint to an encompassing district perspective. These geographically defined regions, characterized by their shared parameters, serve as standardized units. They can be flexibly assembled to create energy districts, forming what we call Urban Energy Units (UEUs).

As stated in the literature review, the definition of urban units presented by Dettmar et al. (2020) serve as the base for our data-driven classification model and definition of the UEUs. Dettmar et al. (2020) identified 14 different settlement areas, 15 open spaces and 23 single elements. The authors focused on understanding the structural and morphological characteristics of each one of the urban units, by calculating the statistical parameters of the building stock. The values of these parameters were the result of on site studies and statistical analysis of the german building stock. However, the study stops there and in order to classify a new region, the urban-planer needs to manually draw the boundaries and look at aerial imagery to decide the urban unit class.

This paper, on the other hand, focuses on the identification of such urban units by using data-driven approaches and establishing a clear methodology that can be applied to any given region in an automatized manner. However, because of the complexity of the subject, the problem of data availability and the purpose of applying this methodology in different German cities, the 25 different types of urban units identified by Dettmar et al. (2020) needed to be simplified, and the corresponding parameters and values that characterize them, needed to be re-adjusted. This is due to the fact that for a first approach it was better to have a simpler version of the classes (for example avoiding the subdivisions like 1.a and 1.b) and/or avoiding classification of urban units that with information from other sources such as land use or official databases could be better classified, for example open spaces are easy to identify as the main information is usually found in OSM land use database and for single elements the building function of the 3D GML building models.

Table 2 provides a description of each UEU along with the simplifications made in comparison to Dettmar et al. (2020). In cases where more than one EST was presented, we calculated the statistical mean of the different parameters. This table also presents the typical structural concept of these urban spaces as blueprints for each UEU. Crucially, it includes values for the total heat and electricity demand. These values are categorized into four scenarios: No refurbishment, total refurbishment, partial refurbishment, and passive housing for the entire building stock presented in the corresponding UEU. Table 2 specifically displays the energy demand values for the worst-case scenario, which assumes no refurbishment of the building stock for the German census year class 1949–1978. It is important to acknowledge that the energy values presented in Table 2 and Appendix B Table B.2 have been statistically adjusted to match the building age classes found in the German census. However, the validation of these values goes beyond the scope of this paper. Future research conducted at our institute will concentrate on refining the accuracy of UEU energy demand calculations, focusing on higher temporal and spatial resolutions via the development of load profiles.

Table 3 shows the parameter values for the classification of the UEUs based on the values calculated by Dettmar et al. (2020). In their work, they define 47 parameters, from which in this work just 23 were selected (use distribution and construction type have different classes), the decision was based in data availability, parameters such as number of people living in the building, proportion of windows and type of materials, are difficult to obtained, and that is why the number of parameters was reduced. More detailed information about the UEU-types, parameters and energy demand values is found in Tables A.1, A.2 and A.3 of Appendix A, and Tables B.1 and B.2 of Appendix B.

In urban planning, the study area’s boundaries can change due to property rights or development plans, potentially excluding important energy elements or urban spaces. To account for this, flexible study area boundaries are essential for incorporating existing energy options and enhancing the overall energy balance. The method shown here is based on the premise that the urban space can be subdivided into morphological units called UEUs. By identifying types and abstracting the urban structure, the urban fabric can be reduced to the parameters that are essential for energy balance. Based on UEUs, the energy related properties of an area are calculated in a simplified form and because of the clear demarcation among each other, the UEUs can be combined in a modular way. This approach makes an appropriate approximation of the building stock, making the urban planning adaptable to the changes of the energy demand and supply. Fig. 1 illustrates how any city area can be divided into urban units and then be classified into the 16 UEUs.

4. Methodology

In this study, we applied the general methodology outlined in Fig. 2. This involved selecting a study area, gathering pertinent data, creating databases for the building stock, defining urban units spatially, categorizing them into UEUs, and assembling these UEUs to form energy districts. Subsequent subsections provide a comprehensive breakdown of each step in the methodology and detail the processing, integration, and analysis of various datasets.

4.1. Study area

The first step of the general methodology is the selection of the study area. This can be any given area within Germany. The following study is focused on the city of Oldenburg in the state of Lower Saxony, Germany (see Fig. 3). This location was selected because required datasets were already available for this city and others were provided by the project partner in Oldenburg. Next subsections introduce the collected datasets.

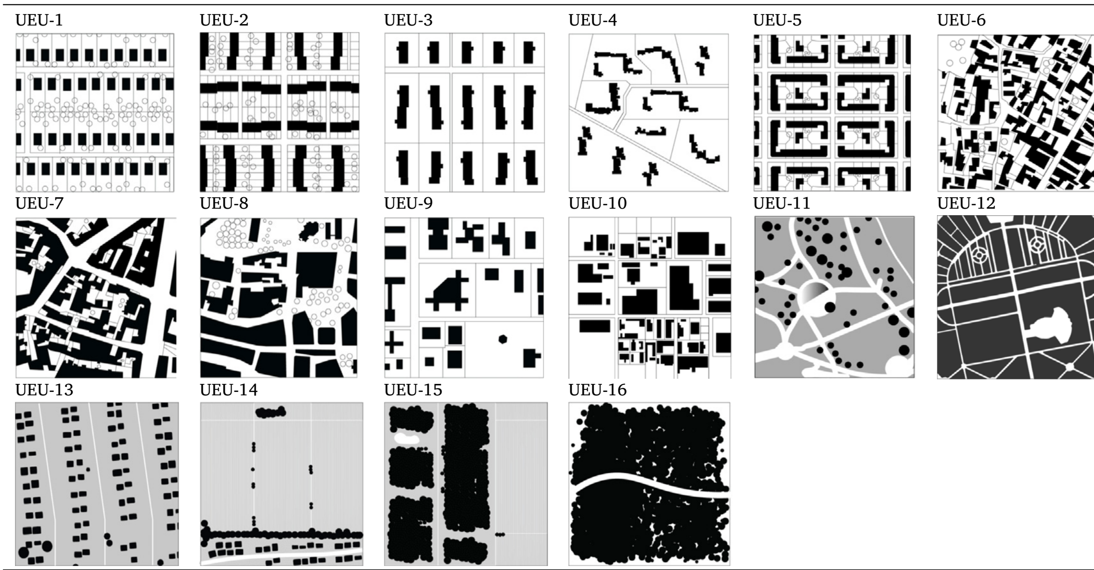
4.2. Data collection

The second step of the general methodology is the respective data collection of the study area. Three main data sources are needed: Building 3D Models, Census data, and GIS data.

Table 2
Short description of UEU-types, subdivided into settlement areas with different predominant use and the main differences with the urban units (ESTs) defined by Dettmar et al. (2020). Heat and electricity demand values under the scenario of **no refurbishment** for the building stock and for the census year class **1949–1978.** Values given in MWh/ha × a.

UEU	Description	Difference from Dettmar et al. (2020)	Heat demand	Electricity demand
	Settlement areas with predominant residential use			
UEU-1	Small-scale detached housing	EST-1, 1.a and 1.b	433	78
UEU-2	Terraced house-like development	EST-2	1040	116
UEU-3	Low to medium-height row development	EST-3	1420	268
UEU-4	Large-scale residential development	EST-4, 4.a and 4.b	1281	395
UEU-5	Perimeter block development	EST-5	3243	694
	Settlement areas with predominant mix use			
UEU-6	Village development	EST-6	1738	312
UEU-7	Historic old town	EST-7	3126	751
UEU-8	Inner city	EST-8, 8.a and 8.b	4924	693
	Settlement areas with administrative and commercial use			
UEU-9	Business, office and administrative area	EST-9	–	–
UEU-10	Industrial area	EST-10, 10.a and 10.b	–	–
	Open space areas			
UEU-11	Public parks	EST-11	–	–
UEU-12	Cemeteries	EST-12	–	–
UEU-13	Allotment gardens	EST-13	–	–
UEU-14	Arable land	EST-23	–	–
UEU-15	Permanent grassland	EST-24	–	–
UEU-16	Forest	EST-25	–	–

Typical concept of the structure of each urban space



[1]EST 14–22 were not included in the analysis as they represented water bodies and street infrastructure. Single elements (EE) were also not included as the individual classification of single buildings is done with other data sources.
[2]The images of the typical concept of the structure of each urban space are an adaptation from the images found in Dettmar et al. (2020)
[3]The UEU-types 9 and 10 variate a lot due to the fact that they are the administrative, commercial and industrial areas and the heat demand is very heterogeneous. UEU-types 11–16 are open spaces and the heating demand is effectively almost zero in relationship with the settlement areas.

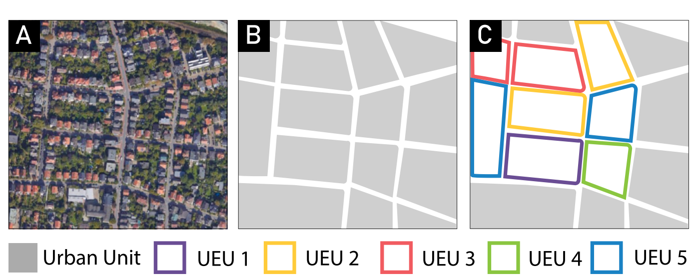


Fig. 1. Graphical representation of urban units. A: Satellite image of a part of the city of Oldenburg. B: Basic representation of urban units in a GIS-space. C: Classification of the urban units into the different UEU-types.

Table 3
Parameter values of the corresponding building stock found within each UEU-type and input for the UEU classification model.[a,b]

Parameter	Description	Unit
Floor area ratio	Ratio between the total footprint area of the buildings within the section and the total area of the section	–
Storeys area ratio	Ratio between the total storey area of the buildings within the section and the total area of the section	–
Num. buildings ratio	Number of buildings within the section divided by the area of the section	1/ha
Avg. num. of storeys	Average number of storeys of all the buildings within the section	–
Avg. footprint area	Average buildings footprint area within the section	m^2
Residential area	Residential area divided by the total area of the section in ha	m^2/ha
Industrial area	Industrial area divided by the total area of the section in ha	m^2/ha
Brutto living space	Brutto living space divided by the total area of the section in ha	m^2/ha
Brutto footprint area	Brutto footprint area divided by the total area of the section in ha	m^2/ha
A/V-ratio	The average quotient of the exterior surface areas (A) and the heated volume of the buildings (V) within the section	m^2/m^3
Avg. envelope surface	Average enveloping surface of the buildings within the section divided by the area of the section in ha	m^2/ha
Avg. façade	Average of the proportion of total façade of the buildings within the section divided by the are of the section in ha	m^2/ha
Use distribution	Ratio of the following usages of the buildings within the section: Industrial, Trade, Services, Residential, Other	–
Construction type	Ratio of the following construction types of the buildings within the section: SFH, MFH, Office, Special, Hall, Other	–
Free space	Free space of the section, it can be calculated as 1− total building footprint area	–

[a] This is a scaled down version of the data input as for each parameter the information of minimum, maximum and quartiles is available.
[b] In the case of open spaces labeled UEU 11–16, we determined their classification based on official data and OSM. Since these areas have minimal or no human settlement infrastructure, all the parameters are effectively zero.

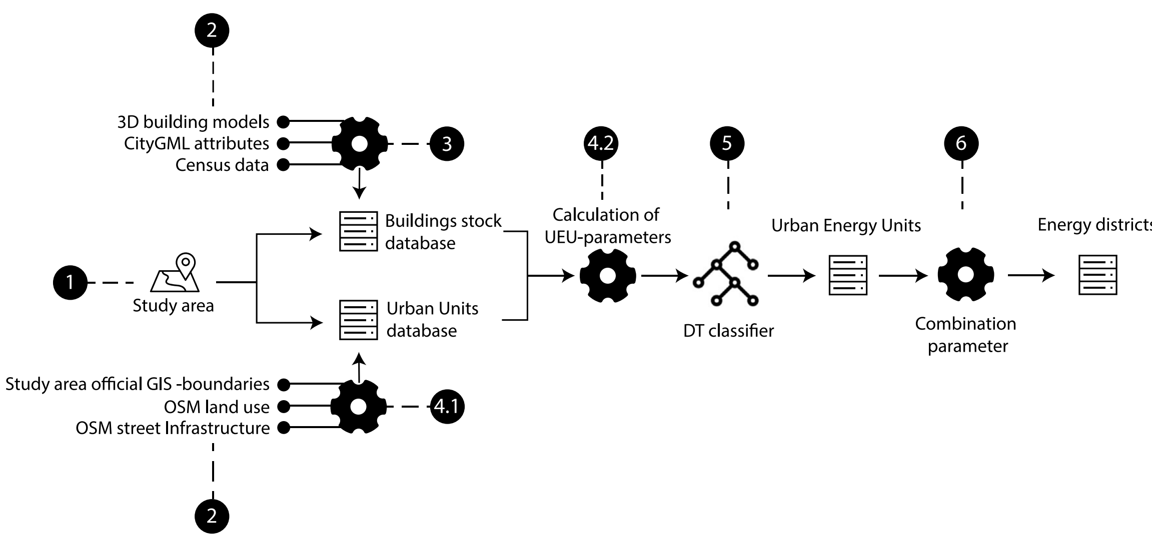


Fig. 2. General methodology for the classification of Urban Energy Units and construction of energy districts. **1:** Selection of study area. **2:** Data collection. **3:** Generation of a building stock database. **4:** Generation of an Urban Units database and calculation of UEU-parameters. **5:** DT model for UEU classification. **6:** Construction of energy districts through a combination parameter. The respective numbers represent each one of the following subsections of the methodology.

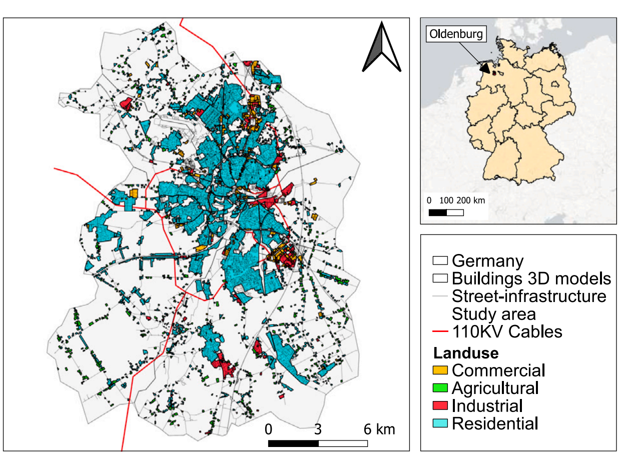


Fig. 3. Study area: Graphical representation of the GIS data extracted by FlexiGIS and showcased for the city of Oldenburg, Germany.

4.2.1. Building 3D models

A building information model is a digital representation of a built facility. It includes the geometry of building components at various Levels of Detail (LoD) (Borrmann, König, Koch, & Beetz, 2015). LoD defines different object complexities (see Fig. 4). LoD0 indicates footprints, LoD1 involves extruded footprints (blocks), LoD2 encompasses volumes with simplified roof shapes, LoD3 provides volumetric models with more architectural details (e.g., windows, roof overhangs, façade details), and LoD4 extends LoD3 by adding indoor features like rooms and furniture (Biljecki, Ledoux, & Stoter, 2016).

The source data for Oldenburg's building models is accessible on Lower Saxony's data portal (LGLN, 2021). These models are in CityGML-LoD2 format, containing information about building geometry, including geographical coordinates, footprint, perimeter, area, ground surfaces, height, walls, roof height, and standardized roof shapes. Oldenburg has a total of 56,749 buildings, with approximately 80% (42,875) for residential use. The remaining 20% consists of industrial, commercial, agricultural, and educational buildings. However, please note that the database may have incomplete entries and classification errors.

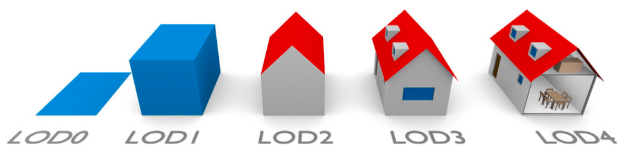


Fig. 4. The five different LoDs for building models in CityGML 2.0 (Biljecki et al., 2016). Licensed under CC BY-NC-ND 4.0.

4.2.2. Census data

Apart from 3D building models, another data source for Germany's building stock is the 2011 census, a national population and housing statistical report (Census, 2011). Due to data protection regulations, individual data points are not publicly available. Instead, the published data is aggregated in the INSPIRE-compliant 100 m grid format INSPIRE (2017), with the highest spatial resolution being 100 m (BKG, 2019). The aggregated data includes various building characteristics, with the ones used in this study being: **Building Age**: It is categorized into ten age classes, representing unequal intervals of building construction years. **Building Form**: It is classified into four categories, distinguishing between free-standing, semi-detached, row houses, or other structures. **Building Ownership**: The information is provided in eight different classes. **Building Use**: It distinguishes between buildings exclusively for residential purposes and those used for both residential and non-residential purposes. **Building Size**: The census categorizes it into ten classes, ranging from single-family houses to structures with over 13 apartment units. **Heating System**: It classifies buildings based on their heating methods, including district heating, floor heating, block heating, central heating, stoves, or no heating system. For more detailed information, you can refer to the Census (2011) database. Fig. 5 depicts a graphical representation of the 100 m × 100 m INSPIRE grid format and the building footprints. The figure shows the difference of data structure between the census and the building models.

4.2.3. GIS data

Most of the GIS data used in this study was obtained through the HCMGIS-plugin in QGIS, which allows direct data retrieval from OpenStreetMap (OSM) (Geofabrik GmbH Karlsruhe, 2020; HCMGIS, 2018). Another method for accessing urban infrastructure datasets is by using the FlexiGIS plugin for QGIS, developed by co-author Alhamwi (Alhamwi, Medjroubi, Vogt, & Agert, 2017). This plugin filters, clusters, and characterizes data on buildings, land use, and highways from OSM. GIS data was specifically collected for the city of Oldenburg, as indicated in Fig. 3. The primary GIS datasets utilized in this paper encompass street infrastructure, land use, and administrative boundaries. The street infrastructure data adopts a vector-multiline format, with streets categorized based on their significance within the road network. Land use data is presented in polygon-format, with each polygon tagged according to the land's designated use. An important dataset of this study is the location and capacities of distribution transformers. For this dataset, we relied on data provided by the local network operator, EWE NETZ, in Oldenburg, this is the only data-set which is not of open-access but as later explained in the methodology section, this is a dataset which can vary depending on the project. Due to copyright and data privacy regulations, this specific data cannot be shared or published.

4.3. Buildings stock database through ML integration models

The third step of the general methodology is generating a complete database of the building information within the study area by integrating the main datasets. First, the Building 3D GML models with the census data. As previously discussed, the 3D GML models and the census database have inherently distinct data structures; therefore, a critical data cleaning and processing step to create a unified database is needed. The 3D GML models convey the spatial characteristics of each building using a specific coordinate system, and other building attributes, like function and roof shape. In contrast, the census data is organized in a multi-tabular structure. Each row corresponds to a specific 100 m × 100 m grid cell in the INSPIRE-compliant grid format, and each column represents an attribute for that cell.

The challenge is to merge these distinct structures into a single, coherent, and comprehensible format. This involves effectively associating each census attribute with every individual 3D building model. In the census dataset, each parameter is presented as an aggregated mean value for a 100 m × 100 m grid cell, as shown in Fig. 5. Many buildings can naturally fall within the same grid cell or overlap multiple grid cells simultaneously. This situation presents a classic data disaggregation problem, where statistical data is deconstructed into various individual variables based on specific characteristics. In the census data each 100 m × 100 m grid cell is assigned a specific label, along with a measure of accuracy. The accuracy is defined as the relationship between the total number of buildings in a cell and the number of buildings that possess a particular attribute within that cell. Leveraging these two data characteristics (label and accuracy), we developed a Random Forest (RF) model to allocate the attributes of building age and construction type to each of the 3D building models. Notably, the attributes "building use" and "building form" are already present in the CityGML files. However, certain attributes, such as "total number of buildings", "number of apartments", and "property", were not taken into account because they are not relevant to the investigation.

RF is a supervised learning algorithm known for its robustness to noise, computational efficiency, built-in importance estimation, and its capability to handle both categorical and continuous data. It is particularly effective for tasks involving high data dimensionality and multicollinearity of features. Moreover, RF is commonly used in classifying remote sensing and geographic data. To construct the RF models, we selected classification features based on previous research, particularly focusing on building age and construction type prediction, and incorporated the census accuracy for each grid cell. These features include geometric attributes such as building height, roof angle, number of storeys, volume, building footprint characteristics, and corresponding grid cell accuracy (ranging from 0% for no buildings to 100% for all buildings with the same attribute). Qualitative attributes like building use, roof type, and building function were integrated into the model during the learning process.

The learning process of the RF models consisted of two main components. First, city-specific learning features were used. Second, a separate dataset from another city was employed as an independent sample where the classification features of building age and construction type were already known. We used the city of Wuppertal, one of the few German cities with available open-source data on building ages.

In this model, classification attributes encompass both the building's geometric and descriptive attributes, as well as information about the census grid cell from the INSPIRE-compliant grid format where each building is situated. The learning dataset was trained using buildings located in census grid cells with only one attribute value. A dataset structure similar to the one used by Garbasevschi et al. (2021) in Wuppertal was employed. To address potential issues like over-classification, we used a sub-set of 70% of the Oldenburg building stock database for training and tested it on the remaining 30% (test dataset). We also performed cross-validation with data from Wuppertal. Over-classification occurs when the training data is dominated by a specific class (for building age, the class 1949–1978 is prevalent due to post-war construction, and for construction type is the single family housing). To counteract this bias, we employed random oversampling, ensuring an equal representation of all classes. For hyper-parameter optimization, we utilized Grid Search Cross-Validation with the 'GridSearchCV' function in Python. This technique is particularly valuable

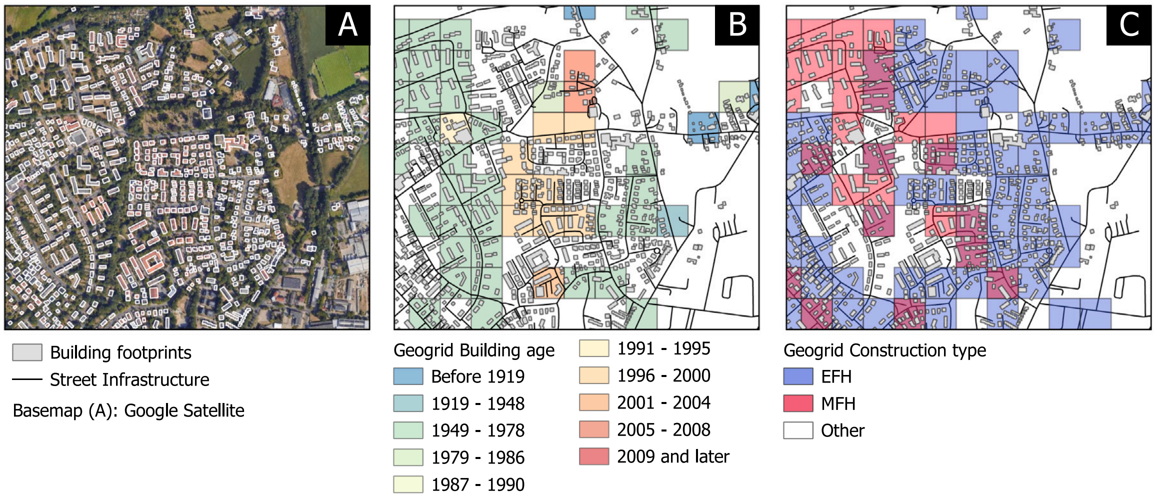


Fig. 5. GIS and census data visual representation. A: Part of the city of Oldenburg, showing the building footprints from the 3D GML models with Google Satellite Basemap. B: Census data aggregated in INSPIRE 100 m grid for the parameter of building age and subdivided into the ten different classes. C: Census data aggregated in INSPIRE 100 m grid for the parameter of construction type. Colored grid cells represent a space of 100 m × 100 m in the same area of the city of Oldenburg. Building footprints are also shown.

in machine learning, especially for tuning hyper-parameters in Random Forest (RF) models. RF hyper-parameters, such as the number of trees or the maximum tree depth, can significantly impact model performance. The 'GridSearchCV' function automates this process by systematically exploring numerous RF configurations, allowing us to select the best-performing model. Last but not least, a Mean Decrease in Impurity (MDI) information gain was employed in order to quantify the importance of each feature in the decision-making of the model.

The performance of a classification algorithm is generally evaluated by its accuracy or success rate, defined as the ratio of correctly labeled observations to the size of input data. For problems of multi-class classification, the overall accuracy can be misleading when the representation of classes in the sample is unequal. This is the case with most of the building datasets under analysis. After generating a prediction for the building age and building construction type for all the buildings within the study area, we were able to construct a general database, with 39 different attributes as a result of the integration of the GIS, census and 3D CityGML datasets.

4.4. Urban units database

The fourth step in the general methodology is the generation of a database with the information of the spatially defined sections within the study area and the relevant attributes for the classification of UEU found in Table 3. This database is generated by the pre-processing and integration of the OSM land use and street infrastructure datasets, creating the GIS-space of the spatially defined urban units, and for each urban unit the calculation of the UEU-parameters is done by analyzing the statistical distribution of the individual buildings found within each urban unit.

4.4.1. Spatial boundaries through interpolation of OSM data

The street-infrastructure dataset, includes streets categorized by their importance within the road network. Streets labeled as "service" and "tertiary" were removed, leaving only the more significant ones. To enhance data processing, the street data was transformed into a multi-polygon format by applying a 1.5 m buffer and then dissolving the geometries. This process, shown in Fig. 1B, aided in sectioning the study area. After dividing the study area into these sections, the land use attributes from the OSM database were assigned to each section. The land use data is organized as vector multipolygons, with each polygon indicating land usage. These polygons were cropped to fit the sections mentioned earlier, and their associated attributes were transferred accordingly.

4.4.2. Calculation of UEU parameters

The calculation of the 23 UEU-parameters are needed for the future classification of the UEUs (see Tables 2 and 3). These parameters (detailed in explained in Appendix B) are calculated by processing of the buildings database into the spatially defined sections or urban units. The integration was done through statistical methods for each one of the parameters, either by averaging, adding or listing the respective values of all buildings within the urban units. Due to the fact that the building database refers to the individual buildings within the study area, and the urban units database refers to the subdivision of the study area into different urban units, the integration of these two databases is done through: first, aggregation of the parameters of the individual houses through statistical methods and second, allocation of such parameters into each one of the urban units of the study area. In order to have a common unit of reference, all parameters are given per hectare of urban unit.

The parameters refer to geometrical and structural characteristics of the buildings within each urban unit. Firstly, the geometrical characteristics such as mean height, total ground area, average number of storeys etc, are calculated through analysis of the 3D GML models of the buildings, and secondly, the structural characteristics such as form of the building, age of the building, type of land use, etc., are the result of ML models that classify the corresponding individual buildings and geospatial analysis of the different data sources. At the end of this process, we obtain a complete database of the study area subdivided into urban units and attributes explaining the statistical information of all individual buildings within each one of these urban units.

4.5. Decision tree model for UEU classification

At this moment we have calculated the 23 UEU-parameters for every urban unit of the city of Oldenburg. However, everyone of these urban units needs to be classified into one of the corresponding UEU-types (see Table 2). Here comes the fifth step of the general methodology, namely, the UEU classification.

We employed a Decision Tree (DT) model to classify the urban units into UEU types. A DT serves as a decision-support tool, representing decisions and their potential outcomes in a tree-like structure. It is essentially a collection of conditional statements. Each case, characterized by specific attributes, falls into one of several classes, with attributes that can be either continuous or discrete (Quinlan, 1987). The main advantages of choosing a DT for the classification model include its transparency. Decision trees are easy to understand as

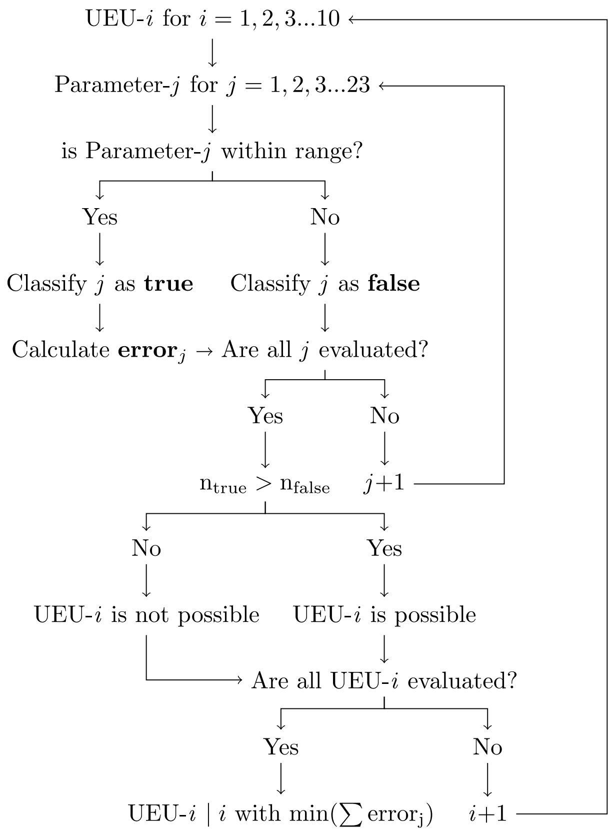


Fig. 6. Graphical representation of the Decision Tree (DT) model.

their structure resembles a flowchart with a series of if–else conditions, making them intuitive for interpretation. Furthermore, they can effectively capture nonlinear relationships between features and the target variable, making them suitable for modeling complex decision boundaries. Lastly, decision trees can rank the importance of features by assessing their contribution to the model's decision-making process, helping identify the most relevant features for predictions. However, it is essential to note that DTs can be sensitive to data variations, leading to variations in the optimal decision tree structure and outputs with even minor changes in the data.

The DT classification process relied on a set of if–else conditions that compared specific parameters of each urban unit with established thresholds for each UEU type. These thresholds, encompassing minimum, average, and maximum values, were derived from literature (see values in Appendix B Table B.1). The DT's operation unfolded as follows: The algorithm began by selecting an urban unit from the urban unit database created in prior steps. For each parameter, it checked if the value fell within the specified range (defined by the minimum and maximum literature values). If the value was within the range, the parameter was labeled as 'true', and the algorithm calculated the deviation from the mean literature value. If the value was outside the range, the parameter was labeled as 'false.' This process was repeated for all parameters, with a count of 'true' and 'false' for each UEU class. If the majority of parameters were labeled 'true' for a particular class, it became a possible classification for the selected UEU. The algorithm iterated through all potential classes and selected the one with the minimum classification error. You can see a visual representation of this DT model in Fig. 6. The algorithm was implemented in Python.

4.6. Construction of energy districts

The last step in our general methodology involves the creation of energy districts through the aggregation of various UEUs within a city. Let us break down this process. Initially, UEUs are defined based on literature and open-source data, then classified according to key parameters describing settlement patterns and building characteristics. These UEUs are subsequently stored in a database and serve as the fundamental analysis units for future urban energy planning. They not only provide a standard way to delineate energy areas in a city but also possess the unique feature of being combined in a modular manner to assess larger settlements. In essence, they act as building blocks for energy districts.

However, a critical question arises: How do we aggregate UEUs to create these energy districts? The answer is a flexible compromise between standardization and adaptability to special conditions. Each UEU serves as a standard building block for energy analysis, and urban planners or energy analysts are given the freedom to select parameters for combining UEUs into energy districts. Constructing energy districts relies on the specific requirements of energy planners, projects, administrative bodies, architects, or any stakeholders focused on energetically assessing large areas with a standardized approach. An energy district can be formed by summing the UEUs. This is achieved by selecting a UEU combination parameter, which can be customized based on the project or study area's specific needs. If no combination parameter is chosen, each UEU functions as an independent energy district. This level of flexibility demonstrates the adaptability of our methodology while still maintaining a standardized core.

In our study, we exemplify how to create energy districts in Oldenburg. We have selected a parameter that we believe effectively bridges the current energy system with future energy planning: the 'network transformers'. This choice is driven by two primary reasons. Firstly, network transformers serve as a crucial link between electricity and heating demand. In future scenarios, renewable sources will generate electricity for the heating demand, and these transformers will play a key role in distributing this energy. Secondly, network transformers are spatial points, allowing for geographically referenced point-vector data, which aligns with our methodology's adaptability. By utilizing this parameter, we illustrate how to construct energy districts in the context of our study in Oldenburg, highlighting the valuable interconnection between the existing system and future energy planning. It is important to note that these network transformers' data, provided by the local network operator in Oldenburg, is restricted due to copyright and data privacy concerns and cannot be publicly disclosed. To combine UEUs using network transformers and create geographically defined energy districts, we follow these steps:

1. Voronoi Diagram Generation: We start by generating a Voronoi diagram, a geometric division of the plane based on the network transformers. Each transformer becomes the center of a Voronoi polygon.
2. Overlap with Classified UEUs: Next, we overlap these Voronoi grids with the pre-classified UEUs. Both the Voronoi polygons and UEUs are in a multi-polygon vector format, which allows for this overlap.
3. UEU Allocation: We allocate each UEU to the Voronoi polygon that covers more than 50% of its area. Meaning that each Voronoi cell has a set of corresponding UEUs. Each UEU belongs to one Voronoi grid, even though geographically it may intersect with multiple Voronoi grids.
4. Energy District Aggregation: We aggregate all UEUs that belong to each network transformer, creating a single energy district for each transformer.

Combining UEUs based on their proximity to a network transform is just one of many ways to aggregate them. Urban planners have the flexibility to choose from various methods. The visual explanation can be found in Fig. 7.

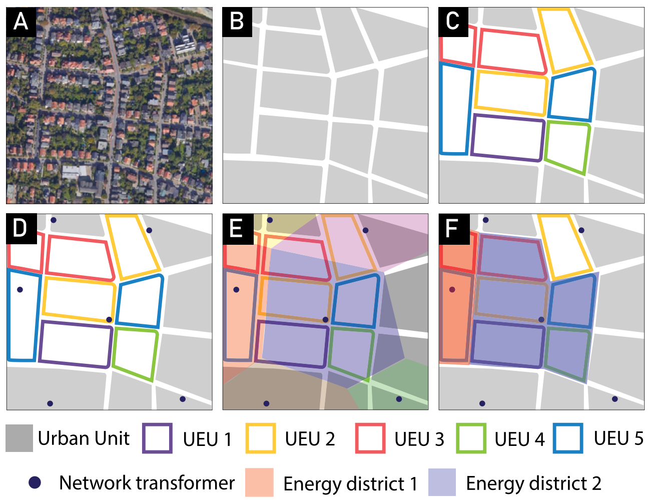


Fig. 7. Expanded methodology. A: Satellite image of a part of the city of Oldenburg. B: Basic representation of urban units in a GIS-space. C: Classification of the urban units into the different UEU-types. D: Geographical location of network transformers in the GIS-space. E: Illustration of the Voronoi polygons according to the location of the network transformers. F: Construction of two energy districts by adding different UEUs together that fall into the same Voronoi polygon.

5. Results

The following paragraphs show the results of this study, this section follows the general methodology shown in Fig. 2 plus the construction of energy districts in Oldenburg. It shows first the results for the construction of the building stock database, the construction of the Urban Units database, how the two databases are combined with an integration model and the calculation of the UEU attributes, in order to use the decision tree classifier and obtain the final UEUs. Last, the combination of the UEUs to construct the energy districts are shown.

5.1. Building stock database through ML integration models

The Building stock database constructed in this study was the result of the integration of the building 3D models, CityGML attributes, the geolocation and the census data for the study area. This database has a tabular structure with a total of 56,749 records (individual buildings) and 37 attributes describing the geometry, building type and census data of each building. According to the Administration of Oldenburg, there are 45,438 residential buildings and about 2956 non-residential buildings, the difference with the number of records is allocated to the individual small constructions like garden houses or garages.

Attributes such as building ID, number of neighbors, grid ID allocation, among others, were allocated due to geolocalization process. This geolocalization process made possible the integration of the 3D building models with the CityGML attributes and GIS data by using the same Coordinate Reference System *ETRS89 Lambert Azimuthal Equal Area*. Other attributes such as centroids, areas, perimeters, height-area ratio etc., were calculated with mathematical processes using python and the GeoPandas library. Last but not least, the allocation of census data for each one of the buildings was the result of a combination of a geolocalization process with a ML data-disaggregation model.

The learning phase had an accuracy of 91% for the classification of the building age just in the city of Wuppertal after optimization of the model with the 'GridSearchCV' function of scikit-learn. When we added the building information of Oldenburg and expanded the learning dataset including now the buildings in Oldenburg that fell in those grid cells with just one classification of the building age, the total accuracy with optimization dropped to 84%. This values is still higher than the accuracy of other models, and includes more information and learning parameters from the census data, making it a high-accuracy classification of the building age of buildings when no more information is available. When comparing the general aggregated classification of all the respective classes we see a 84% accuracy in the classification of the buildings in Oldenburg, meaning that overall in the city 84% of the buildings were classified correctly and each individual building has a probability of being classified between 84%–100% depending whether or not it falls into one of those unique grid cells. Fig. 8 shows the main results of the classification model. In the first place, it shows the importance of the first 13 features ranked from top to bottom, this ranking is derived from the Mean Decrease in Impurity (MDI) information gain. Here we see that the most important feature for classification of building age are the centroid or location of the corresponding building, height, and how many buildings are in the same grid cell. Fig. 8 also shows the confusion matrix for the final classification of the building age for the buildings in Oldenburg, showing higher number of predicted values on the matrix's diagonal and an overall accuracy of 84%. And finally we can show an overall aggregated histogram per building age class in Oldenburg, comparing the predicted values with those of the general census data. Fig. 9 shows an illustration on a GIS-space of how the buildings look before and after their classification with the RF model.

For the construction type the model classified all buildings in Oldenburg into one of three major classes (SFH, MFH and others). According to aggregated 2011 census data, 81% of residential buildings in Oldenburg are SFH, approximately 16% are MFH, and the remaining 2% are other residential buildings such as garages and small gardens. The presented model classifies with 82% accuracy the buildings between SFH and MFH. However, it misclassifies 8 times more buildings into the class of 'others'. This is because the 3D CityGML models

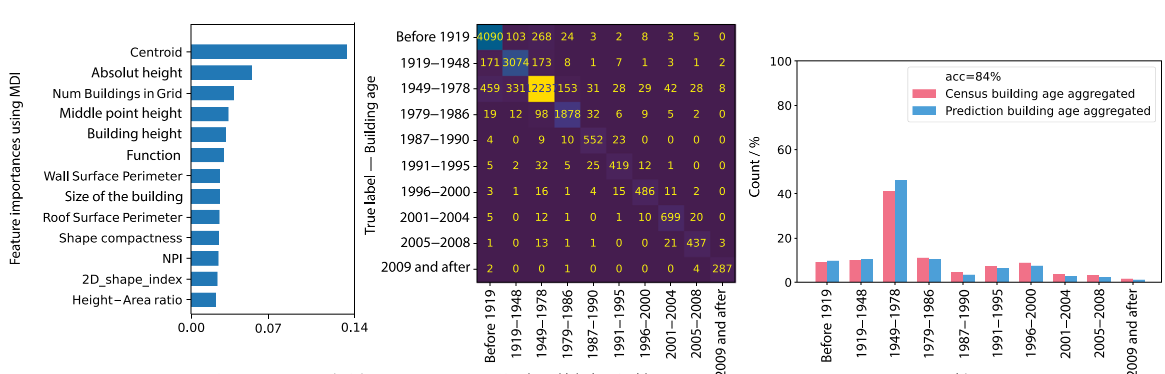


Fig. 8. Results of the RF classification model for Oldenburg. Left: Feature importance of the model. Middle: Confusion matrix of each possible class for the building age showing true and predicted labels. Right: Histograms of the aggregated census data compared to the building age prediction showing a 84% accuracy.

includes all of the buildings within the study area (including the non-residential like administrative buildings and others) and the census database focuses only in residential buildings. Even after filtering the buildings, there is still significant misclassification, likely because the 3D building models only contain geometry information and may lack other relevant parameters beyond just the building's geometry.

Fig. 9. Graphical representation of the *building age* and *construction type* before and after the classification of the buildings in Oldenburg on a GIS-space with the INSPIRE-compliant 100 m grid format.

5.2. Urban units database

The first result was the construction of the urban units database. This database contains the information of the spatially defined sections within the study area and the relevant attributes for the future classification into specific UEUs found in Table 3. This database was generated by the pre-processing and integration of the OSM land use and street infrastructure data sets described in Section 4.4.1 and the calculation of the UEU attributes from Dettmar et al. (2020) described in Section 4.4.2. A total of 8249 urban units were spatially defined, these urban units build a database with all 23 parameters needed for the UEU classification process. The database has a size of 8249 × 23 and for this reason is not presented here. However the results can be appreciated in Fig. 10.

5.3. Decision tree model for UEU classification

Each of the previously spatially defined urban units was classified into one of the 16 UEUs mentioned in Tables 2, A.1, A.2 and A.3 by applying the decision tree classifier described in Section 4.5. Fig. 10 shows the classification of the spatially defined UEUs for the city of Oldenburg. Every single UEU contains all of the parameters of Table 3 and is also linked to the building stock database, accessing to the corresponding information of the 3D building models located within that UEU. The classification shows clearly how the old town and industrial areas around it are well classified, also the residential buildings mostly corresponding to UEU-types 1 and 2. On the upper right corner we can see the whole study area of Oldenburg and surroundings exported by FlexiGIS and classified by this model, the residential parts of the city are clearly distinguishable as well as the arable lands and forest surrounding the city. This shows that classification model works on large scales.

As explained before, the classification was done via a DT classifier. The DT classifier is based on predetermined reference values taken out of the literature (Dettmar et al., 2020); therefore, there is no training process and the validation of the is done by comparing the classified UEU with the corresponding satellite image, and seeing whether or not the building's infrastructure corresponds to the information of the UEU. This process was done with a random number of UEUs and the model classified all of them well. As an example of this process, in Fig. 10 we can see how the city center of Oldenburg is clearly delimited by the UEUs 7–8, which are the historic old town and inner city respectively, and the UEUs 1–5 are predominantly regions of residential use, which is also seen in the corresponding satellite image.

It is important to emphasize that if this methodology is to be applied in other projects, we will be able to construct a larger database with more information and therefore; the model could be adapted to increase the classification accuracy. As of now, this is the first ever classification of UEU done with computational methods and therefore the inconsistencies with reality and classification errors are to be expected. These inconsistencies are seen in the *unclassified* UEUs of the model, and are mainly cause because of: Low quality information about the building stock of that region and/or high mixture of commercial, residential and industrial use.

The classification was done for the city of Oldenburg and the administrative areas shown in Fig. 3. Table 4 shows the percentages of classification for each UEU-type for the 8249 urban units, here we can see that the largest amount of area for the study area is of class UEU-14 (Arable land) and UEU-16 (Forest) with a total of 55%. It makes sense due to the fact that the study area incorporates a large amount of rural area around the main urban city. The urban city of Oldenburg is characterized mainly by UEU-1 (small-scale, detached residential development of low to medium storey height) and UEU-10 (Industrial area) which accurately corresponds to the data provided by the city administration and the census data with approximately 65% of the buildings being single family houses. A total of 11% of the area was not classified by the model, this could be because of the following reasons. First, the corresponding spatially defined urban spaces either had a high mix of industrial and residential buildings, making it difficult for the model to differentiate between residential and non-residential. Second, the information of the 3D building models was not complete or non existent, meaning that although there are existing buildings within that area, the model was not able to allocate the building properties given by the 3D model (such as height, area, etc.) and therefore could not classified the corresponding urban unit. The last reason for the unclassified area, is because part of the open urban space does not fall into one of the established categories, examples are open parking lots or construction areas, future versions of the model could implement such open spaces, but for simplification of this study just the open spaces mentioned in Table 4 were classified.

5.4. Construction of energy districts

Construction of energy districts is made through the combination of different UEUs within a city. The UEU combination is made with help of the selection of the UEU combination parameter as shown in the main methodology. The UEU combination parameter chosen for this study was the geolocation of network distribution transformers of the city of Oldenburg, because they act as a link between the current electricity capacity and the heating demand in urban energy planing.

In order to combine UEUs through network transformers and geographically defined energy districts, we follow the procedure explained in Section 4.6 through the calculation of Voronoi polygons. A total of 853 network transformers are located in the study area, and their location was provided by the local network operator EWE NETZ in Oldenburg. These 853 network transformers have different capacities of 200, 250, 400, 630 and 800 kVA, which in first instance and for the purpose of generating the Voronoi polygons is irrelevant; however, in the future, such information can be integrated in the decision making process in order to better construct energy districts according to the specific needs.

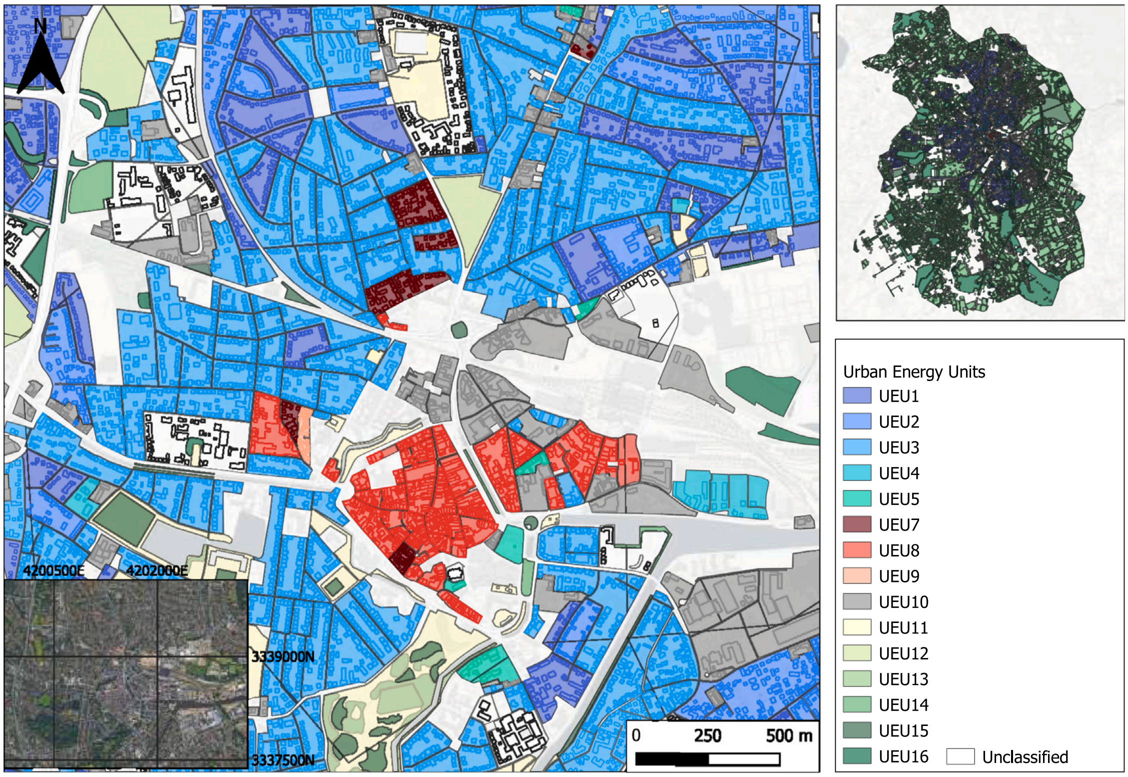


Fig. 10. Part of the city of Oldenburg divided into the specific UEU-types 1–16. Coordinates of the specific area are given in the satellite image. City center, surroundings of residential and commercial areas, as well as unclassified UEUs are defined by the UEU-types.

	UEU	Classification acc (%)	Mean Building age class	Area UEU (Ha)	Number Buildings	SFH/MFH/Other (%)	Heat demand (MWh/(ha*a))	Electricity demand (MWh/(ha*a))
1	7	76.2	Before 1919	0.56	39	51/10/39	1057 - 4675	751
2	8	74.5	Before 1919	0.42	31	42/29/29	1951 - 6520	693
3	8	76.3	Before 1919	0.38	10	20/50/30	1951 - 6520	693
4	8	71.1	Before 1919	0.38	15	7/26/67	1951 - 6520	693

	Heat demand (MWh/a)	Electricity demand (MWh/a)
Energy District	2894 - 10311	1238

Fig. 11. Energy districts of Oldenburg. A: General view of Oldenburg divided into energy districts. B: Part of the city of Oldenburg divided into energy districts, the energy districts were constructed around the network transformers shown as red points. (C, D): Energy analysis of a single energy district constructed with four UEUs is shown alongside the satellite image and the summary of the total energy demand in the table.

Table 4
Percentage distribution for each UEU-type. A total of 8249 UEUs were classified in an area of 258.10 km^2.

Urban energy unit	Description	Percentage %
	Settlement areas with predominant residential use	
UEU-1	Small-scale, detached residential development	14.7
UEU-2	Terraced house-like development	2.2
UEU-3	Row development of low to medium storey height	3.9
UEU-4	Large-scale residential development	0.2
UEU-5	Perimeter block development	0.1
	Settlement areas with predominant mix use	
UEU-6	Village development	0.0
UEU-7	Historic old town	0.1
UEU-8	Inner city	0.5
	Settlement areas with administrative and commercial use	
UEU-9	Business, office and administrative area	0.1
UEU-10	Industrial area	7.4
	Open space areas	
UEU-11	Public parks	2.6
UEU-12	Cemeteries	0.3
UEU-13	Allotment gardens	0.4
UEU-14	Arable land	36.9
UEU-15	Permanent grassland	1.5
UEU-16	Forest	18.0
Unclassified	–	11.1

A total of 798 different energy districts where constructed in this study, and these energy districts, process the information of: the building stock, the respective UEU, energy demand, and network transformer capacity. Fig. 11 shows the constructed energy districts for the same area of Fig. 10. It also shows the analysis of one specific energy district chosen at random and a summary of the characteristics of each UEU of this district. More about the energy demand shown in this table is explained in the following paragraphs.

As previously mentioned, the configuration of energy districts relies on a combination parameter. For this study, the chosen parameter is the geolocation of network transformers. However, it is important to note that this parameter can be adjusted to align with the specific requirements of energy planners. While this adaptability is advantageous, it also comes with limitations. On the positive side, it provides flexibility in parameter selection. Nevertheless, it is essential to acknowledge that data related to the location of network transformers may not always be readily available, presenting a potential limitation.

5.5. Heat demand at district level

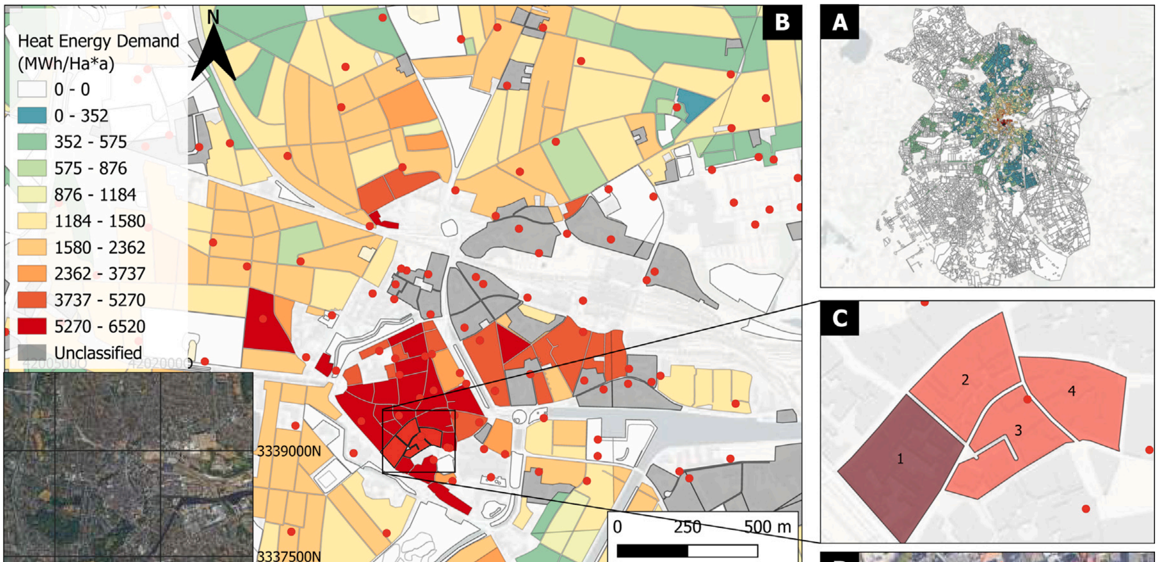
As we can see, the selected energy district is built around a single network transformer, a total of four units were grouped in order to built this energy district. One unit was classified as UEU-7 and the remaining three of type 8. Information about the mean building age, distribution of building type (SFH, MFH or other), number of buildings in the unit, total area of the unit, as well as the classification accuracy are found on the corresponding table. The most important information is the prediction of the total heat and electricity demand for the four UEU and the overall energy district, the values for the energy demand are taken from Dettmar et al. (2020), adapted according to the german year construction classes of the census and the information is found in Appendix B Table B.2. There are four possible cases, according to the level of refurbishment. The worst case is the assumption that the entire building stock is not refurbished and the best case scenario is the assumption that the entire building stock is entirely refurbished. The information can also be presented as a range for the heating demand, due to the fact that the information about the level of refurbishment of the individual buildings is not known, so by assuming a total refurbishment and no refurbishment of the buildings, we obtain the minimum and maximum heat demand values per hectare per year. This is one of the main limitations of the UEUs, because there is no information about the state of refurbishment of the buildings, we have to assume for conservative reasons, the worst case scenario. However, there are some studies that statistically classify the heat demand according to the refurbishment status of different European countries (Loga, Diefenbach, Stein, & Born, 2012). The selected energy district has a total final predicted heating demand between 2800–10,300 MWh/a (depending on the best or worst case refurbishment of the buildings) and electricity demand of 1238 MWh/a.

The outcome of this methodology is the mapping of UEUs with their respective heat demands. In Fig. 12, the energy demands of Oldenburg's UEUs are showcased. The total energy demand, denoted in MWh/(ha $\times$ a), is specified for each UEU. It is important to note that these energy values are based on the best-worst-case scenario (with or without refurbishment of the building stock) as explained in Section 3. Furthermore, in Fig. 11, you can see the energy analysis of the same district. For a broader perspective, Fig. C.1 in Appendix C illustrates the energy mapping for the entire city of Oldenburg, demonstrating the rapid spatial analysis of energy demand across large areas and emphasizing the advantages of employing UEUs as the primary unit of spatial measurement.

6. Practical application of the UEUs

The practical application of the UEUs can be summarized by the fact that because many municipalities in Germany are still in the early stages of the municipality heating plan (as stated in the introduction and literature review) and because there is lack of personal and standardized processes, this methodology and its results can accelerate the first phase of the heating plan which is to understand the building stock and energy demand in neighborhoods.

Small municipalities, because of their lack of qualified personnel to generate such energy analysis, subcontract specialized engineering offices, from which there are not so many in the country in order to meet the demand, and are therefore currently under high pressure, a classical high demand/low offer scenario. It is clear that this methodology presents a complex mixture of ML models, GIS-based approaches and new definitions that may not be used directly by engineering offices and energy planners. However; the concept is already defined and the process is established, further development of a user friendly interface or of a general UEU-based energy map of the country could accelerate the acceptance and use of the UEUs, such as the energy map of Oldenburg presented in Appendix C. The main result being, to locate the areas with high and low heat demand to see where new



	UEU	Classification acc (%)	Mean Building age class	Area UEU (Ha)	Number Buildings	SFH/MFH/Other (%)	Heat demand (MWh/(ha*a))	Electricity demand (MWh/(ha*a))
1	7	76.2	Before 1919	0.56	39	51/10/39	1057 - 4675	751
2	8	74.5	Before 1919	0.42	31	42/29/29	1951 - 6520	693
3	8	76.3	Before 1919	0.38	10	20/50/30	1951 - 6520	693
4	8	71.1	Before 1919	0.38	15	7/26/67	1951 - 6520	693

	Heat demand (MWh/a)	Electricity demand (MWh/a)
Energy District	2894 - 10311	1238

Fig. 12. Showcase of Energy District Analysis of Oldenburg. The respective UEUs are displayed according to their heat energy demand in MWh pro constructed area in hectares (ha) pro year given a worst case of refurbishment state of the building stock. A: General heat map view of Oldenburg. B: Part of the city of Oldenburg divided into UEUs, with the location of the network transformers shown as red points. (C, D): Energy analysis of a single energy district constructed with four UEUs. Alongside the satellite image is a table with a summary of the UEU characteristics of the district with the total energy demand, here the values for heat demand of best and worst case are displayed.

technologies should be built (for example heat pumps, short or long distance heat networks) and integrated into the mix. The concept has been applied in two pilot-projects with two different engineering offices in Germany. The results have been of interest to the engineering offices as they are able to present a documented version of the heating plan in form of maps to the municipalities using, among other information, the UEUs; however, further development of is still needed

7. Conclusion and outlook

In this paper, we introduced an innovative methodology for subdividing urban areas into sixteen different morphological units known as Urban Energy Units (UEUs). These UEUs are characterized by specific building attributes, settlement patterns, and energy demands, which are detailed in various tables throughout this study. Our primary data sources included 3D-CityGML building models, census data, OpenStreetMap, and official administrative GIS data. We applied this methodology in Oldenburg, a mid-sized German city. Our analysis revealed that crucial information for effective energy analysis in Germany is often missing, specifically parameters related to building age, construction type, and refurbishment status. To address this gap, we developed a Random Forest model to estimate the first two parameters with an accuracy of 84%. For the third parameter, we established a worst-best-case scenario. Oldenburg was divided into 8249 distinct UEUs, each classified into one of sixteen different classes using a Decision Tree model. This covered a total area of 258.10 km^2, from which 89% was successfully classified.

The UEUs were subsequently combined to form 798 distinct energy districts centered around network transformers. This was due to the fact that the energy transformers act as a link between the current electricity capacity and the future heating demand in urban energy planing. The energy analysis of an individual district was showcased with results presented using maps that highlight the effectiveness of GIS in energy planning. The respective values for heat and electricity demand were adapted statistically to align with the building age classes of the German census, drawing from literature sources. However, it is important to note that the validation of such values is out of the scope of this paper

Future research will have a dual focus. Firstly, we aim to typify the heat and electricity demand profiles of UEUs by validating them against aggregated building energy demand models at higher spatial and temporal resolutions. We also plan to expand the number of UEU classes by applying our methodology to other urban areas like municipalities and smaller cities. On the practical side, we intend to ground the UEUs' application in forthcoming pilot projects in collaboration with engineering firms responsible for creating municipal heating plans. Additionally, our focus will be on developing user-friendly software that facilitates the seamless integration of our findings. This approach acknowledges that city planners typically do not employ complex machine learning models in their decision-making processes.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Luis Armando Blanco Bohorquez reports a relationship with German Aerospace Center DLR that includes: employment. Luis Armando Blanco Bohorquez has patent KLASSIFIZIERUNG VON URBANEN ENERGIEEINHEITEN UND ENERGETISCHEN QUARTIEREN MIT TECHNIKEN DER KÜNSTLICHEN INTELLIGENZ pending to German Aerospace Center DLR. This article is part of the doctoral studies of the main author. The doctoral studies are conducted in the RWTH Aachen University in Germany.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to acknowledge the local grid operator EWE NETZ for giving the opportunity to test the developed model by using a few datasets of the locations of distribution transformers in the study area.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.scs.2023.105075.

References

Abbasabadi, N., & Ashayeri, M. (2019). Urban energy use modeling methods and tools: A review and an outlook. *Building and Environment*, *161*, Article 106270.

Adams, D. (1994). *Urban planning and the development process*. Psychology Press.

Ahlers, D., Driscoll, P., Wibe, H., & Wyckmans, A. (2019). Co-creation of positive energy blocks. *IOP Conference Series: Earth and Environmental Science*, *352*(1), Article 012060.

Ahn, Y., & Sohn, D.-W. (2019). The effect of neighbourhood-level urban form on residential building energy use: A GIS-based model using building energy benchmarking data in seattle. *Energy and Buildings*, *196*, 124–133.

Alhamwi, A., Medjroubi, W., Vogt, T., & Agert, C. (2017). GIS-based urban energy systems models and tools: Introducing a model for the optimisation of flexibilisation technologies in urban areas. *Applied Energy*, *191*, 1–9.

Ali, U., Shamsi, M. H., Bohacek, M., Purcell, K., Hoare, C., Mangina, E., et al. (2020). A data-driven approach for multi-scale GIS-based building energy modeling for analysis, planning and support decision making. *Applied Energy*, *279*, Article 115834.

Balouktsi, M., Lützkendorf, T., Kopfmüller, J., & Steltzer, V. (2017). Creating methods, procedures and tools for a more sustainable neighbourhood development–experiences from Germany. In *Proceedings of the world sustainable built environment conference* (pp. 2272–2278).

Berardi, U. (2013). Sustainability assessment of urban communities through rating systems. *Environment, Development and Sustainability*, *15*, 1573–1591.

Biljecki, F., Ledoux, H., & Stoter, J. (2016). An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, *59*, 25–37.

Boerefijn, W. N. A., et al. (2010). *The foundation, planning and building of new towns in the 13th and 14th centuries in Europe: An architectural-historical research into urban form and its creation*. Universiteit van Amsterdam [Host].

Borrmann, A., König, M., Koch, C., & Beetz, J. (2015). *Building information modeling: Technologische Grundlagen und industrielle Praxis*. Springer-Verlag.

Bruegmann, R. (2019). *Sprawl: A compact history*. University of Chicago Press.

Bundesamt für Kartographie und Geodäsie (2019). *Geographische Gitter für Deutschland in Lambert-Projektion (GeoGitter Inspire)*. *©GeoBasis-DE / BKG*. Bundesamt für Kartographie und Geod¨asie, Retrieved from https://gdz.bkg.bund.de/index.php/default/open-data/geographische-gitter-fur-deutschland-in-lambert-projektion-geogitter-inspire.html, BKG.

Bundesministerium für Wirtschaft und Klimaschutz (BMWK) (2019a). Das neue Gebäudeenergiegesetz – kurz zusammen gefasst. Retrieved from https://www.bmwi.de/Redaktion/DE/Downloads/G/gebaeudeenergiegesetz-zusammen-gefasst.pdf?_blob=publicationFile&v=6, BMWK.

Bundesministerium für Wirtschaft und Klimaschutz (BMWK) (2019b). Klimaschutzprogramm 2030 der Bundesregierung zur Umsetzung des Klimaschutzplans 2050. *Hg. v. Die Bundesregierung. Berlin.*, *21*, Retrieved from https://www.bundesregierung.de/breg-de/schwerpunkte/klimaschutz/massnahmenprogramm-klima-1679498, BMWK.

Bundesministerium für Wirtschaft und Klimaschutz (BMWK) (2020). Gebäudeenergiegesetz (GEG) - Energy Saving Ordinance for Buildings. Retrieved from https://www.gesetze-im-internet.de/gebaeudeeeg_2020/.

Bundesministerium für Wohnen, Stadtentwicklung und Bauwesen (BMWSB) (2023). Entwurf eines gesetzes für die wärmeplanung und zur dekarbonisierung der wärmenetze (1. Referentenentwurf). Retrieved from https://www.bmwsb.bund.de/SharedDocs/gesetzgebungsverfahren/Webs/BMWSB/DE/kommunale-waermeplanung.html, BMWSB.

Bundestag, D. (2023). Anschlussdrucksache 20-25-426 zum Entwurf eines Gesetzes zur Änderung des Gebäudeenergiegesetzes. Retrieved from https://www.bundestag.de/resource/blob/956254/3320714bc5fad9e22d6e5de31f28e9d7/A-Drs-20-25-426-FH-des-BMWK-fuer-einen-AeA-der-Koafraktionen-data.pdf, BMWK.

Cajot, S., Peter, M., Bahu, J.-M., Guignet, F., Koch, A., & Maréchal, F. (2017). Obstacles in energy planning at the urban scale. *Sustainable Cities and Society*, *30*, 223–236.

Chen, Y., Hong, T., & Piette, M. A. (2017). Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. *Applied Energy*, *205*, 323–335.

Couch, C., Petschel-Held, G., & Leontidou, L. (2008). *Urban sprawl in europe: landscape, land-use change and policy*. John Wiley & Sons.

DESA, U. (2019). United nations, department of economic and social affairs, population division. World urbanization prospects: The 2018 revision (ST/ESA/SER. a/420). New york: United nations. Retrieved from https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

Dettmar, J., Drebes, C., & Sieber, S. (2020). *Energetische stadtraumtypen. Strukturelle und energetische kennwerte von stadträumen*. Stuttgart: Fraunhofer IRB Verlag, ISBN: 978-3-7388-0342-6, S 2., überarbeitete und erweiterte Auflage.

El Kontar, R., Polly, B., Charan, T., Fleming, K., Moore, N., Long, N., et al. (2020). *URBANopt: An open-source software development kit for community and urban district energy modeling*: *Tech. rep.*, National Renewable Energy Lab.(NREL), Golden, CO (United States), Retrieved from https://www.osti.gov/biblio/1677416.

Elci, M., Delgado, B. M., Henning, H.-M., Henze, G. P., & Herkel, S. (2018). Aggregation of residential buildings for thermal building simulations on an urban district scale. *Sustainable Cities and Society*, *39*, 537–547.

Garbasevschi, O. M., Schmiedt, J. E., Verma, T., Lefter, I., Altes, W. K. K., Droin, A., et al. (2021). Spatial factors influencing building age prediction and implications for urban residential energy modelling. *Computers, Environment and Urban Systems*, *88*, Article 101637.

Geofabrik GmbH Karlsruhe (2020). *Geofabrik downloads*. Geofabrik GmbH, Retrieved from https://download.geofabrik.de/.

Groppi, D., de Santoli, L., Cumo, F., & Garcia, D. A. (2018). A GIS-based model to assess buildings energy consumption and usable solar energy potential in urban areas. *Sustainable Cities and Society*, *40*, 546–558.

HCMGIS - Center for Applied GIS of HoChiMinh City (2018). HCMGIS - geometry processing and field calculation plugin for QGIS. GitHub repository, GitHub, HCMGIS, Retrieved from https://github.com/hcmgis/HCMGIS.git.

Hipp, J. R., Faris, R. W., & Boessen, A. (2012). Measuring 'neighborhood': Constructing network neighborhoods. *Social Networks*, *34*(1), 128–140.

IEA (2020). *Germany 2020. energy policy review*. Paris.: International Energy Agency, Retrieved from https://www.iea.org/reports/germany-2020.

INSPIRE Thematic Working Group Coordinate Reference Systems & Geographical Grid Systems (2017). *D2.8.I.2 data specification on geographical grid systems – technical guidelines*: *INSPIRE Thematic Working Group Coordinate Reference Systems & Geographical Grid Systems*, Elsevier, Retrieved from https://inspire.ec.europa.eu/id/document/tg/gg, INSPIRE.

Koch, A. (2010). Heat demand modeling at the scale of urban neighborhoods. In *Next generation infrastructure systems for eco-cities* (pp. 1–6). IEEE.

Konstantinou, T., & Knaack, U. (2011). Refurbishment of residential buildings: A design approach to energy-efficiency upgrades. *Procedia Engineering*, *21*, 666–675.

Landesamtes für Geoinformation und Landesvermessung Niedersachsen (2021). OpenGeoData Niedersachsen. Retrieved from https://opengeodata.lgln.niedersachsen.de/, LGLN, Data licence Germany – attribution – Version 2.0 / dl-de/by-2-0.

Li, Z., Quan, S. J., & Yang, P. P.-J. (2016). Energy performance simulation for planning a low carbon neighborhood urban district: A case study in the city of Macau. *Habitat International*, *53*, 206–214.

Loga, T., Diefenbach, N., Stein, B., & Born, R. (2012). Tabula: Further development of the German residential building typology. Darmstadt: Institut Wohnen und Umwelt GmbH, Retrieved from.

Lützkendorf, T., & Balouktsi, M. (2017). Assessing a sustainable urban development: Typology of indicators and sources of information. *Procedia Environmental Sciences*, *38*, 546–553.

Ma, J., & Cheng, J. C. (2016). Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology. *Applied Energy*, *183*, 182–192.

Manfren, M., Nastasi, B., Groppi, D., & Garcia, D. A. (2020). Open data and energy analytics-an analysis of essential information for energy system planning, design and operation. *Energy*, *213*, Article 118803.

Moghadam, S. T., Toniolo, J., Mutani, G., & Lombardi, P. (2018). A GIS-statistical approach for assessing built environment energy use at urban scale. *Sustainable Cities and Society*, *37*, 70–84.

Nouvel, R., Mastrucci, A., Leopold, U., Baume, O., Coors, V., & Eicker, U. (2015). Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support. *Energy and Buildings*, *107*, 204–212.

Nutkiewicz, A., Yang, Z., & Jain, R. K. (2018). Data-driven urban energy simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Applied Energy*, *225*, 1176–1189.

Quan, S. J., Li, Q., Augenbroe, G., Brown, J., & Yang, P. P.-J. (2015). Urban data and building energy modeling: A GIS-based urban building energy modeling system using the urban-EPC engine. In *Planning support systems and smart cities* (pp. 447–469). Springer.

Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, *27*(3), 221–234.

Rapf, O., Faber, M., Marian, C., & Fata, F. (2015). *Renovating Germany's building stock*. Brussels: Buildings Performance Institute Europe (BPIE), BPIE.

Shnapp, S., Paci, D., & Bertoldi, P. (2020). *Enabling positive energy districts across Europe: energy efficiency couples renewable energy*. Luxembourg: Publication Office of the European Union.

Städtetag, D. (2023). *So läuft die wärmeplanung in den städten – umfrage des deutschen städtetages*. Retrieved from https://www.staedtetag.de/files/dst/docs/Presse/2023/Umfrage-des-Deutschen-Staedtetages-zur-Waermeplanung-in-den-Mitgliedsstaedten.pdf.

Statista Search Department (2022a). Greenhouse gas (GHG) emissions of the building sector in Germany from 1990 to 2020, by building type. Retrieved from https://www.statista.com/statistics/1297313/ghg-emissions-by-buildings-in-germany-by-type/.

Statista Search Department (2022b). Share of final energy consumption in residential buildings in Germany in 2020, by end use. Retrieved from https://www.statista.com/statistics/1297381/final-energy-consumption-in-residential-buildings-in-germany-by-end-use/.

Statistische Ämter des Bundes und der Länder (2011). Zensus 2011 – Gebäude und Wohnungen. Retrieved from https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html?nn=3806618#Gitter, Census.

Wurm, M., Droin, A., Stark, T., Geiß, C., Sulzer, W., & Taubenböck, H. (2021). Deep learning-based generation of building stock data from remote sensing for urban heat demand modeling. *ISPRS International Journal of Geo-Information*, *10*(1), 23.

Yamamura, S., Fan, L., & Suzuki, Y. (2017). Assessment of urban energy performance through integration of BIM and GIS for smart city planning. *Procedia Engineering*, *180*, 1462–1472.

Yeh, A. G. (1999). Urban planning and GIS. *Geographical Information Systems*, *2*(877–888), 1.

Zheng, Y., & Weng, Q. (2019). Modeling the effect of climate change on building energy demand in Los Angeles county by using a GIS-based high spatial-and temporal-resolution approach. *Energy*, *176*, 641–655.