Classification of Building Properties from the German Census Data for Energy Analysis Purposes

Luis Blanco[1,2], Megha Aditya[1,3], Björn Schiricke [1], Bernhard Hoffschmidt[1,2]
[1]German Aerospace Center (DLR), Institute of Solar Research, Cologne/Jülich, Germany
[2]RWTH Aachen University, Chair of Solar Components, Aachen, Germany
[3]Paderborn University, Paderborn, Germany

Abstract

The building sector is an important target for reducing urban energy consumption. Detailed data on the building stock is needed for modelling urban building energy demands but its availability is often insufficient. In Germany, the largest available public database about the building stock is the census national database, containing critical attributes for building characterization on a national scale, such as building age, construction type, and number of residents. However, the detailed information about the individual buildings is restricted by national data privacy laws and the information is found in an aggregated format.

This study shows statistical and machine learning approaches to take the census data and disaggregate its information to each individual building in order to use that information for urban building energy modelling. This study presents a classification model for the following parameters of the 2011 census: building age, building form and heating type, and Number of residents. A study case was conducted in Oldenburg, Germany.

Highlights

- Generation of high-quality detailed building stock data while keeping the data privacy policies intact.
- Classification models for different relevant parameters for modelling urban building energy demands.
- Integration of different tools such as machine learning and geoanalysis to obtained building stock data.
- Combination of 3D building models and census data into a single database.

Introduction

The building sector uses over 40% of energy in developed countries, with heating and hot water systems in German residential buildings contributing to 84% of final energy use and almost one-third of greenhouse gas emissions. Detailed building parameters are necessary to construct energy models, as factors like environmental conditions, occupant behavior, and building regulations, impact the energy demand. Residential buildings are often the focus of energy modeling studies due to the fact that non-residential buildings are more complex and often lack of statistical information (Rapf et al. (2015); Economidou et al. (2020); IEA (2020); Statista Search Department (2022a,b); Loga et al. (2012)).

Residential building data is acquired by various methods such as census data collection, formal building and dwelling registers, surveys, or remote sensing (Mata et al. (2014); van den Brom et al. (2019)). In Germany, the largest available public database about the building stock is the national census database (Statistische Ämter des Bundes und der Länder (2011)). However, the detailed information about each individual building is restricted by the national data privacy laws. The census database contains critical attributes for building characterization on a national scale, such as building age, construction type, and number of people per building. Still, the information is found in an aggregated format to comply with the data security laws. Other spatial data sources like OpenStreetMaps or CityGML have increased in popularity due to the growing number of open city data initiatives. Simultaneously, it is of interest to national and local administrations to have an updated state of their building stock, which provides enhanced opportunities for energy building modeling by integrating different open-source databases. In Germany, local administrations provide open data of the 3D models of their building stock.

The main objective of this study is to show statistical, GIS and machine learning based models that can take the census data of Germany and allocate its aggregated information to the individual 3D building CityGML models. By doing so, we are able to complement the building stock by not having only the geometrical properties of the buildings but also relevant characteristics like building age, typology, morphology and number of residents. These parameters are relevant for urban energy modelling and would otherwise be difficult to obtain due to data privacy laws.

In previous works, methods for estimating the energy

characteristics of individual buildings were developed with the aid of machine learning and with the use of GIS, census, and statistical data. This is our basis for the development of more advanced approaches for the energy analysis of buildings. Authors such as Garbasevschi et al. (2021) and Wurm et al. (2021) have already encountered the problem of data aggregation in the census database. Wurm et al. (2021) used a convolutional neural network to build a model which uses aerial images and integrates the census information of construction type and building age. Garbasevschi et al. (2021), focused on developing a Random Forest (RF) model to predict the building age of the census data for all individual buildings depending on their geometrical properties. The newest developments that we present in this publication have focused on refined RF models trying to integrate more characteristics of the census data beyond the building age and construction type attributes such as the number of people per building and the building's heating systems.

Study Area and Data

This study focuses on the city of Oldenburg, Germany, as illustrated in Figure 1. The city was chosen due to the availability of necessary datasets at the outset of the project, some of which were provided by local project partners. The following subsections provide an overview of the collected data.

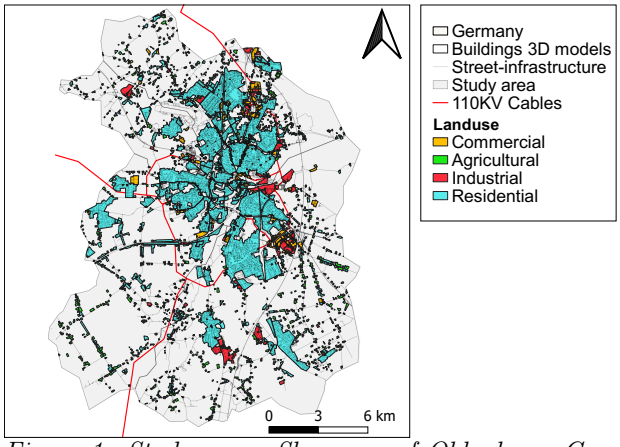

Figure 1: Study area. Showcase of Oldenburg, Germany.

Census Database

A census is the procedure of systematically acquiring, recording and calculating population information about the members of a given population. This term is used mostly in connection with national population and housing censuses (United Nations (2008)). In Germany, the entity in charge of the national statistical census is the Federal and State Statistical Office (Statistische Ämter des Bundes und der Länder). This office conducted a population, building and housing census in 2011. The 2011 census is an important cornerstone for the overall system of population and building statistics, on which other parts of the system are based. Due to data privacy concerns, the 2011 census needed to ensure data protection and information security of individuals and their properties. Because of this reason, the 2011 census public data is published in an aggregated grid format ensuring confidentiality, integrity and authenticity of the data.

The 2011 census public data is presented in the INSPIRE (2014) compliant 100 m grid format of the German Federal Agency of Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie. (2019)). The highest level of spatial resolution is a grid of 100 m × 100 m and every single grid cell contains the information of the 2011 census data respective to that specific area.

The 2011 census database consists of three main datasets: *Population*, *families and households*, and *buildings and apartments*. The population census dataset contains information about the amount of people living in the corresponding 100 m × 100 m grid cell, the amount is given as a integer number strictly greater than 0, grid cells with value −1 mean uninhabited or to be kept secret.

The families and households dataset contains information about the family structure and the living situation of the households. The values are aggregated to each corresponding 100 m × 100 m grid cell. The values are divided into three main parameters: type of family household, type of living arrangement and size of the household. Each of this parameters contains different classes upon which the corresponding grid cell is given a value. Lastly, the buildings and apartments dataset contains information about different parameters of the buildings and apartments in Germany. The values are aggregated to each corresponding 100 m × 100 m grid cell. The values are divided into seven different parameters: building age, building form, building's ownership, building use, building size, heating system and number of apartments in the building.

Figure 2 shows a graphical representation of the INSPIRE-grid format and how the 2011 census data is aggregated in each grid cell. It shows the 10 different classes of the census building age parameter. Each grid cell contains one value for each one of the building age classes. Figure 2 also shows the footprints of some buildings allocated in the shown area, this means that the census values apply for all of the buildings within a each specific grid cell. The same happens for all of the census parameters of the three main datasets. All of this translates into a multidimensional problem in order to disaggregate and allocate the census values to each specific building.

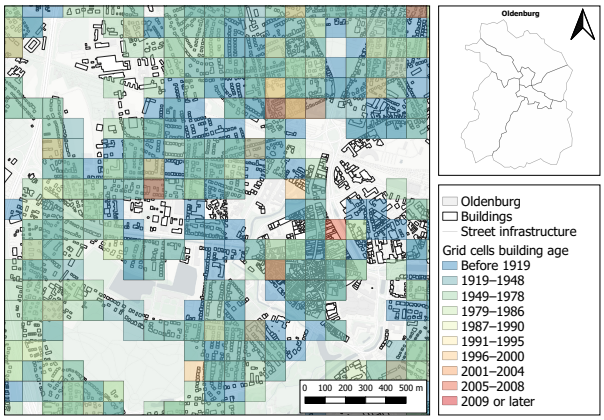

Figure 2: Study area. Showcase of the INSPIRE-compliant 100 m grid format for a 2.5 km^2 area in Oldenburg, Germany. Shown is the building age parameter from census dataset. Footprints of the buildings within that area and the street infrastructure are also shown.

Building 3D Models

A building information model is a comprehensive digital representation of a built facility. It typically includes the geometry of the building components at a defined Level of Detail (LoD) (Borrmann et al. (2015)). LoD is a concept that allows the representation of an object in different complexities (see Figure 3). LoD0 covers footprints. LoD1 contains blocks (i.e. extruded footprints). LoD2 describes volumes with generalized roof shapes. LoD3 specifies volumetric models with greater architectural details including windows (as well as other openings), roof overhangs, and more façade details. Finally, LoD4 extends LoD3 with additional indoor features like rooms or furniture (Biljecki et al. (2016)).

Figure 3: The five different LoDs for building models in CityGML (Biljecki et al, 2016). Licensed under CC BY-NC-ND 4.0.

The source information about the building models of the city of Oldenburg is available on the data portal of Lower Saxony (Landesamt für Geoinformation und Landesvermessung Niedersachsen (2021)). The building models are given in format CityGML-LoD2, this includes detailed information about the building and its geometry such as: geographical coordinates, footprint, perimeter, area, ground surfaces, height, walls, roof height and roof shapes. A total of 56 749 building models were exported for the city of Oldenburg. About 75% (42 875) of them are residential buildings. The other 25% are distributed among industrial, commercial, agricultural and educational buildings; Nevertheless, the database is incomplete and errors in such classifications are expected.

Methodology

This study aims to disaggregate the information of the 2011 German census database and allocate its parameters to each specific 3D building model. Not all of the parameters given in the census database are essential for energy calculations, while others, such as building age and building form, are crucial but only available in the census database, making it essential to find new methodologies to disaggregate the information and give the individual buildings their respective characteristics. Another parameter which this study takes great focus on, is the number of residents in the building which can help understand energy consumption behavior. Table 1 lists the parameters considered in this study and their respective classes. A disaggregation model for each one of these parameters is described in the following sections, based on statistical, GIS and machine learning models.

Table 1: Parameters of the 2011 census database that are investigated in this study, showing the respective descriptions and labels that will be used for the classification models.

Parameter	Description	Pred. Label
Building age	Before 1919	1
	1919–1948	2
	1949–1978	3
	1979–1986	4
	1987–1990	5
	1991–1995	6
	1996–2000	7
	2001–2004	8
	2005–2008	9
	2009 and later	10
Building form	Detached Semi-detached Terraced house	SFH
	Detached Semi-detached Terraced house 3–6 apt. 7–12 apt. $>$ 13 apt. Other building types	MFH
Heating type	District heating	1
	Single-storey heating	2
	Block heating	3
	Central heating	4
	Furnaces	5
	No heating	6
Number of residents	Number of people	>0
	Uninhabited or secret	-1

Building Age

The year of construction influences the energy performance of the building. With the introduction of thermal regulations over the past decades, newly constructed buildings are more energy-efficient than old

ones (Aksoezen et al. (2015)). Building age is divided into ten classes in Table 1, representing different time periods of construction. A building cannot belong to two different classes, resulting in a multi-classification problem. Although in reality the determination of building ages is more complex as the buildings can be refurbished partially or completely; however, this is not part of the scope and the buildings will be classified into a single census class.

To classify buildings into their respective age class, we adopt Garbasevschi et al. (2021)'s approach, which utilized a RF classification model. The model was trained using building geometric features (such as height volume, perimeter, etc.) and from street and block metric features (such as centrality, distance to road, intersections, etc.), for the training data, available building age data from the city of Wuppertal was used. Our approach is similar but differs in some steps and includes the following.

First, we import all of the building 3D LoD2 models into a single database, the information included in those models are geometric features (consisting of all geolocalized points of the building), function of the building (residential, administrative, industrial, etc.) and roof type (14 possible types). Second, we isolate the building age parameter from the census database. Third, we geolocalized all buildings within each $100\,\mathrm{m} \times 100\,\mathrm{m}$ grid cell, by calculating the building's centroid thereby, avoiding that one building be located into more than one grid cell. Fourth, the census data contains a parameter called q which tells how many buildings were surveyed in that grid cell, and how many fall into each class. With this parameter q, we can calculate a relative *probability of being*, and because we already know how many buildings are in each grid cell, we can make sure that the relative class proportion be maintained after the classification. Fifth, like Garbasevschi et al. (2021), we make use of a RF model and building age training data of Wuppertal and also those buildings which fall into a grid cell with only one class. Lastly, the RF was built with the python package *scikit-learn*, the model was trained, corrected for oversampling (most of the buildings in Germany were built after the war, making the class 1949–1978 the majority class and resulting in class imbalance) and applied to the remaining buildings of the city of Oldenburg. The RF model learned from geometric features of the buildings, function of the building, roof type, and probability of being according to the census dataset. A detailed list is found in the results.

Building Form

The building form or construction type impacts the thermal behavior of buildings, e.g., a freestanding (semi-)detached house is more exposed to energy loss due to the higher portion of exterior walls in relation to building volume than terraced houses or multi-family houses (Kaden and Kolbe (2013); Ma and Cheng (2016); Wurm et al. (2021)). The 2011 census database provides information about the building form and distinguishes between 10 different classes as seen in Table 1.

To simplify the classification process, building form classes are grouped into three major categories: single-family houses (SFH), multi-family houses (MFH), and others. The proposed methodology for classifying Oldenburg's buildings into their respective forms includes: First, importing 3D LoD2 models into a database, including information regarding geometric features, building's function and roof type. Second, we isolate the building form parameter from the census database. Third, we geolocalized all buildings within each grid cell. Fourth, grouping building forms into the three major classes; the original census classes 1–3 are SFH, the original census classes 4–9 are MFH and the original class 10 is kept as others. Fifth, the census parameter q or *probability of being* was calculated. Sixth, because no training data for Oldenburg or any nearby city was found, implementing a random distribution classification based on probability and condition statements for geometry and building types.

- If the area of the building is $\geq 250\,\mathrm{m}^2$ then Building is classified as MFH.
- If the number of floors ≥ 3 then the building classified as MFH.
- If the function of the building is not residential then building is classified as other.
- If the area of the building is $< 250\,\mathrm{m}^2$ and number of storeys ≥ 4 then the building is classified as MFH. ($250\,\mathrm{m}^2$ is a reference area that distinguishes between SFH and MFH, taken from Loga et al. (2012).)

By randomly assigning building forms while considering the probability distribution of each grid cell, the model can classify all Oldenburg buildings into the three main classes using the four condition statements.

Heating Type

The 2011 German census database includes information on the predominant form of heating used in the building, which presents an opportunity to understand energy supply, demand, and consumption in a given area. However, the heating type parameter in the census does not translate into a classification problem because it only indicates whether a building is connected to a central or district heating system, contains Single-storey heating technology, or has at least one apartment without a heating system. Central heating typically refers to a heating system that is located within a building and serves that building alone, while district heating is a system that serves multiple buildings in a local area. It is possible for a

building to be connected to both a central and district heating system or to have Single-storey heating independent of its connection to a central or district heating system.

To disaggregate the census information for Oldenburg's buildings, a GIS-based model is implemented to visualize the information and correlate which grids have one or more heating types. The model considers the census *probability of being* for each heating type and decision rules, such as: First, no building can fall into the no heating class and another class. Second, all buildings with Single-storey heating or single/multi-room furnaces will be codependent from either central or district heating and finally, buildings with block heating can also be connected to a central and/or district heating system.

Number of Residents

The quantity of people living in a building significantly influences the overall energy demand for both electricity and heating. For heating, factors such as building size, insulation, climate, and heating system type affect the energy demand. For electricity, the number and types of appliances used, lighting, and heating and cooling systems affect energy demand. Therefore, predicting the number of residents in a building is crucial for designing and managing energy-efficient buildings. This study uses the census housing and population datasets. Table 1 shows the classes for the number of residents per grid cell, indicating the number of people residing per grid cell.

Training datasets were generated for each building type class, specifically Single-Family House (SFH), Multi-Family House (MFH), and Other. Homogeneous grid cells of the census data with the same building type were selected in order to equally distribute the amount of residents to each one of the buildings within that grid cell according to the relative volume proportion of the building (footprint area and number of storeys). Non-linear relationships were observed among the dataset attributes, prompting the selection of an appropriate machine learning model. XGBoost, which is gradient-boosted decision tree machine learning algorithm known for its effectiveness with tabular data, was chosen as the final model.

The model's objective was to allocate residents per grid cell to the buildings within heterogeneous grid cells, considering the proportional building area and number of storeys. The model takes takes into consideration the amount of buildings within each grid cell and their geometric characteristics (3D CityGML LoD2 models), the number of stories of each building within each grid cell (assuming that each floor is approximately 3.0 m and the amount of residents at the grid cell level.

Results

Building age

The building age classification model was trained on a 70% subset of Oldenburg's building stock database, tested on the remaining 30% and cross-validated with data from Wuppertal and grid cells with a single class value. The random oversampling method was used to address the over-classification problem during the learning phase (Shelke et al. (2017)). With this approach we achieve an accuracy of 91% for Wuppertal buildings alone. However, after adding Oldenburg data and expanding the learning dataset, the accuracy dropped to 84% even after optimization with the *GridSearchCV* function. This value is still higher than the accuracy of other models, because it includes and learns from the census data, making it a high-accuracy classification of the building age of buildings when no more information is available. Figure 4 shows the main results of the classification model. In the first place, it shows the importance of the first 12 features using the Mean Decrease in Impurity (MDI) information gain. The main features for classification are building location, height, and the number of buildings in the same grid cell. The confusion matrix for the final classification of the building age for the buildings in Oldenburg shows a higher number of predicted values on the matrix's diagonal and an overall accuracy of 84%. An aggregated histogram per building age class in Oldenburg comparing predicted values with the aggregated count of the census data is also shown.

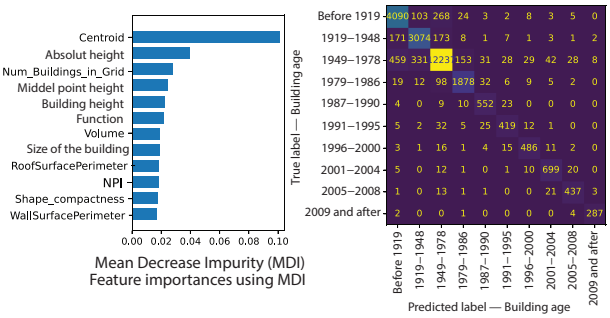

Figure 4: Results of the RF classification model for Oldenburg. Left: Feature importance of the model. Right: Confusion matrix of each possible class for the building age showing true and predicted labels.

The final results provide a comparison of the aggregated total values. Based on the 2011 census buildings and apartments dataset, the municipality of Oldenburg has 42 875 residential buildings, distributed according to their year of construction as shown in the green histogram in Figure 5. The red histogram displays the total sum for each building age class for the same 42 875 demonstrating a high correlation in the total distribution.

Building form

By implementing four condition statements and randomly assigning building forms while also taking into

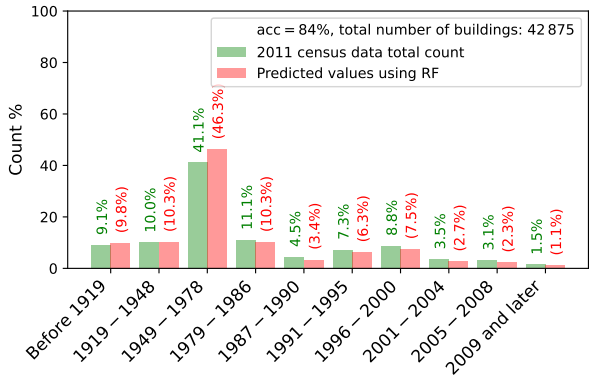

Figure 5: Histograms of the aggregated 2011 building age census data compared to the building age prediction of the RF model showing a 84% accuracy.

account the probability distribution of each grid cell, the model can classify all buildings in Oldenburg into one of three major classes. The histograms in Figure 6 summarize the main results. 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 relatively good buildings between SFH and MFH. However, it misclassifies 8 times more buildings into the class of 'other'. This is because the 3D CityGML models 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.

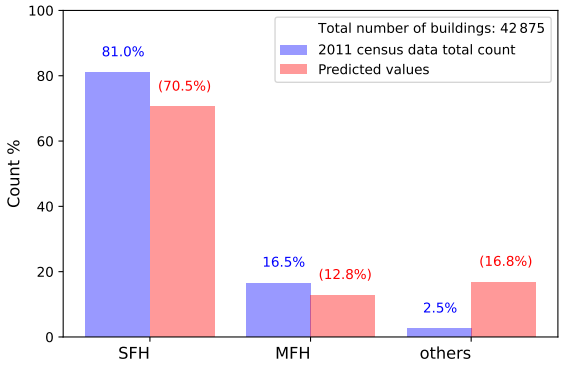

Figure 6: Histograms of the aggregated 2011 building form census data compared to the building form prediction.

Heating type

By applying a GIS-based model, we have correlated 3D building models in Oldenburg with one or more heating system types. The majority of buildings in Oldenburg are connected to a central heating type, meaning that all the residential units in a building are heated by a central heating point located inside the building (usually in the basement). This is evidenced by almost all of the geographical grid cells (86.7%) in Oldenburg belonging to this class as shown in Figure 7. The remaining 13.3% are distributed among other heating classes. It should be noted that this represents the total number of grid cells with this class and not the grid cells with only this class. While census data trends these classes from each other (because of data structure similarity with the other census parameters), our GIS-based model allocates buildings knowing that this overlapping of parameters is physically possible. Our model predicts that a total of 92% of buildings are connected to a central heating system (see Table 2).

Table 2: Aggregated percentages per class of the 2011 building form census data compared to the heating type allocation prediction.

Heating Type	Census %	Predicted %
1	1.7	0.8
2	9.6	6.6
3	0.6	0.2
4	86.7	92.0
5	1.0	0.2
6	0.3	0.1

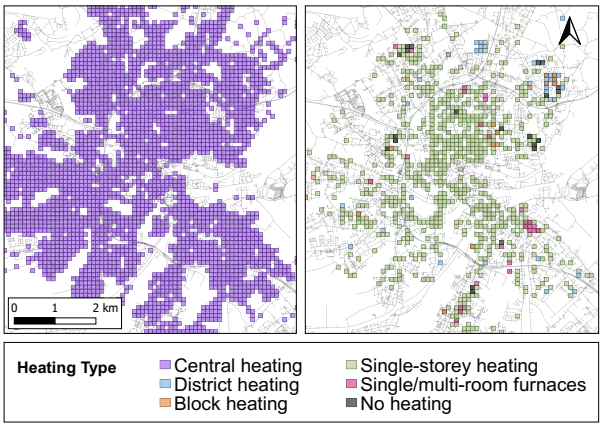

Figure 7: GIS visualization of the 100 m×100 m grid cells of the 2011 census heating type parameter for Oldenburg. Left map shows the area of Oldenburg where central heating is allocated. Right map shows where the other classes are allocated.

Number of residents

A comprehensive model was constructed by integrating three separate models, each dedicated to predicting the number of residents for a specific building form class (refer to Table 1). To mitigate overfitting, a hyper-parameter optimization process was executed. The model's accuracy can be evaluated at the grid cell level, providing a measure of the precision in predicting the number of residents for each building within the respective grid cell. The accuracy of grid cells spans from 10% to 98%, with an average of 39.8%. Approximately 25% of all buildings were classified with an accuracy of 80% or higher. The

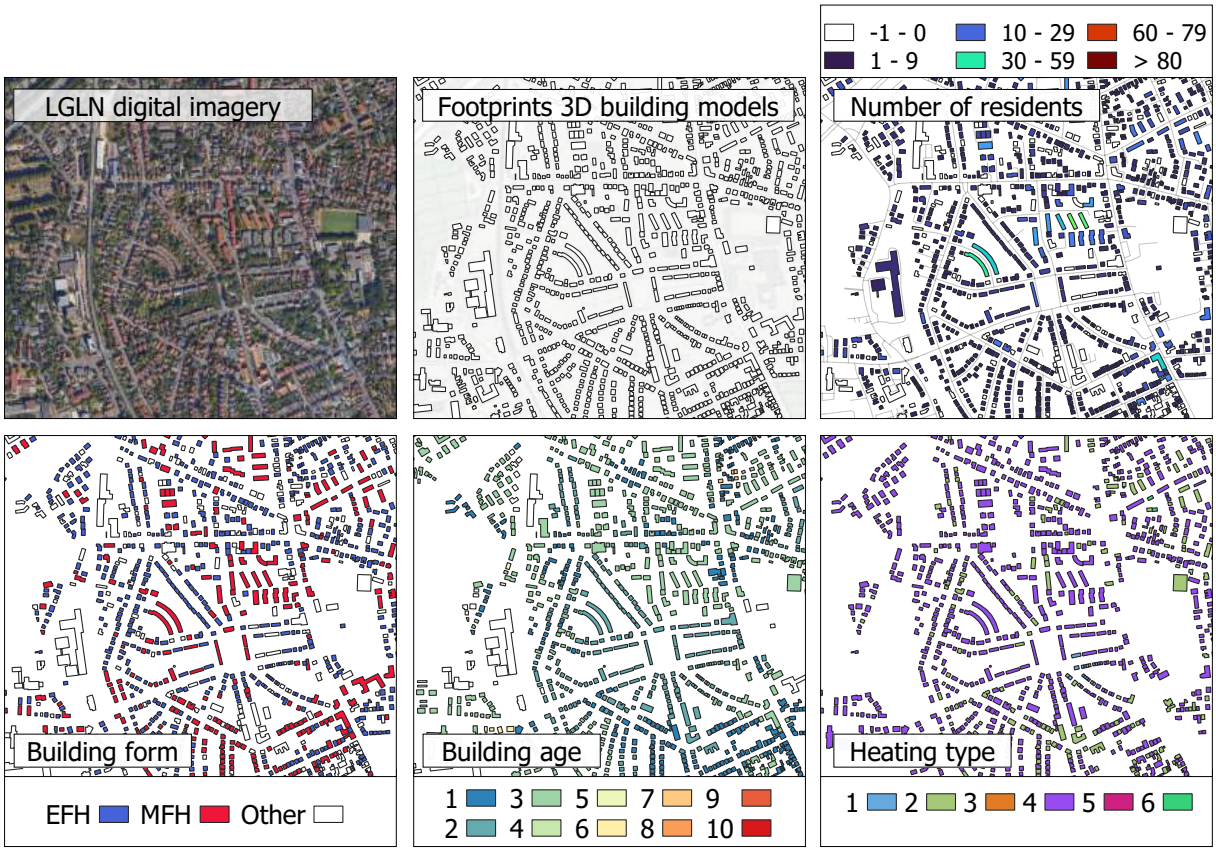

Figure 8: Representation of the results from the different classification models and their visualization into the respective classes. Top left: General view of the selected area for Oldenburg using LGLN digital imagery from the OpenGeoData Niedersachsen (dl-de/by-2-0). Top center: Footprints of the 3D building models. Top right: preliminary classification of the number of residents per building (model under development). Bottom left: Classification of the buildings into the major classes of the building form. Bottom center: Classification of the buildings into the ten building age classes. Bottom right: Classification of the buildings into the heating type classes taking into account that a single building can have multiple heating systems.

performance of the model is closely tied to the accuracy of the building type classification. A better building type classification will lead to a better prediction of the number of residents. It is important to note that census data contains a lot of private or secured data labeled as -1, these were not include in the models. These grid cells are considered sensitive due to potential discrimination and confidentiality concerns. Examples of sensitive areas include prisons and rehabilitation centers. The exclusion of these grid cells ensures the protection of individuals' privacy and prevents any discriminatory impact.

Table 3: Statistical distribution of the residents of Oldenburg for building form.

Building Form	Share of Residents %
SFH	55.4
MFH	40.8
Other	3.8

Conclusion

The 2011 census database consists of three main datasets: *Population, families and households*, and *buildings and apartments*, and the data is presented in the INSPIRE-compliant 100 m grid format with the highest resolution of 100 m × 100 m grid cells. This study investigated 5 different parameters out of these datasets and created a statistical, machine learning and GIS-based model in order to disaggregate the information and classify all buildings within the city of Oldenburg into the respective classes.

This study showed a RF model for the building age reaching 84% accuracy. To classify the building form into the three major classes we developed a statistical approach that classifies with 81% overall accuracy between SFH and MFH but when taking into account the other building types, the model misclassifies 8 times more than expected because of missing information in the building 3D models. For the heating type we used a GIS-based approach and classified all buildings with the possible combinations of heating systems, a weighted accuracy of 89% between the real and predicted aggregated values. For the number of residents a gradient-boosted decision tree model was developed showing an overall accuracy of 39%, where the models are closely related to the building form classification. A general visualization of the classifica-

tion results for all of the models is found in Figure 8, where for a specific area of Oldenburg all buildings are classified into the respective parameters.

In conclusion, census data provides useful information that can be used to parameterize building energy models. This study shows the possibilities of using statistical, machine learning and GIS-based models in order to classify buildings and generate a detailed national building stock while still complying with the data privacy laws. The results obtained can be enhanced by developing a larger machine learning model that incorporates and learns from more parameters.

References

Aksoezen, M., M. Daniel, U. Hassler, and N. Kohler (2015). Building age as an indicator for energy consumption. *Energy and Buildings 87*, 74–86.

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

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

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

Economidou, M., V. Todeschi, P. Bertoldi, D. D'Agostino, P. Zangheri, and L. Castellazzi (2020). Review of 50 years of eu energy efficiency policies for buildings. *Energy and Buildings 225*, 110322.

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

IEA (2020). *Germany 2020. Energy Policy Review*. International Energy Agency, Paris.

INSPIRE (2014). *INSPIRE Data Specification on Geographical Grid Systems – Technical Guidelines*. European Comission: Thematic Working Group Coordinate Reference Systems & Geographical Grid Systems. https://inspire.ec.europa.eu/id/document/tg/gg.

Kaden, R. and T. H. Kolbe (2013). City-wide total energy demand estimation of buildings using semantic 3d city models and statistical data. In *Proc. of the 8th International 3D GeoInfo Conference*.

Landesamt für Geoinformation und Landesvermessung Niedersachsen (2021). OpenGeoData Niedersachsen. Data licence Germany – attribution – Version 2.0 / dl-de/by-2-0.

Loga, T., N. Diefenbach, B. Stein, and R. Born (2012). Tabula: Further development of the german residential building typology.

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

Mata, É., A. S. Kalagasidis, and F. Johnsson (2014). Building-stock aggregation through archetype buildings: France, germany, spain and the uk. *Building and Environment 81*, 270–282.

Rapf, O., M. Faber, C. Marian, and F. Fata (2015). Renovating germany's building stock.

Shelke, M. S., P. R. Deshmukh, and V. K. Shandilya (2017). A review on imbalanced data handling using undersampling and oversampling technique. *Int. J. Recent Trends Eng. Res 3*(4), 444–449.

Statista Search Department (2022a). Greenhouse gas (ghg) emissions of the building sector in germany from 1990 to 2020, by building type.

Statista Search Department (2022b). Share of final energy consumption in residential buildings in germany in 2020, by end use.

Statistische Ämter des Bundes und der Länder (2011). Zensus 2011 – Gebäude und Wohnungen.

United Nations (2008). *Principles and Recommendations for Population and Housing Censuses*, Volume Series M No. 67/Rev. 2. p. 8. 2011-05-14 at the Wayback Machine. ISBN 978-92-1-161505-0.

van den Brom, P., A. R. Hansen, K. Gram-Hanssen, A. Meijer, and H. Visscher (2019). Variances in residential heating consumption–importance of building characteristics and occupants analysed by movers and stayers. *Applied Energy 250*, 713–728.

Wurm, M., A. Droin, T. Stark, C. Geiß, W. Sulzer, and H. Taubenböck (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.