Decarbonisation potential of multi-family Swedish Homes

Hanxiao Qi

Master thesis in Energy-efficient and Environmental Buildings

Faculty of Engineering | Lund University

Lund University

Lund University, with eight faculties and a number of research centres and specialized institutes, is the largest establishment for research and higher education in Scandinavia. The main part of the University is situated in the small city of Lund which has about 112 000 inhabitants. A number of departments for research and education are, however, located in Malmö. Lund University was founded in 1666 and has today a total staff of 6 000 employees and 47 000 students attending 280 degree programmes and 2 300 subject courses offered by 63 departments.

Master Programme in Energy-efficient and Environmental Building Design

This international programme provides knowledge, skills and competencies within the area of energy-efficient and environmental building design in cold climates. The goal is to train highly skilled professionals, who will significantly contribute to and influence the design, building or renovation of energy-efficient buildings, taking into consideration the architecture and environment, the inhabitants' behaviour and needs, their health and comfort as well as the overall economy.

The degree project is the final part of the master programme leading to a Master of Science (120 credits) in Energy-efficient and Environmental Buildings.

Examiner: Niko Gentile (Division of Energy and Building Design)

Supervisor: Ricardo Bernardo (Division of Energy and Building Design), Henrik Davidsson (Division of Energy and Building Design)

Keywords: Building Energy, OpenStreetMap, EPCs, BETSI, Machine Learning, Automated process, carbon neutral buildings, net zero energy, decarbonisation.

Publication year: 2022

Abstract

Global warming is mainly caused by greenhouse gas emissions. The building sector is responsible for a significant share of energy use and greenhouse gas emissions. Most of the existing buildings in Europe were built before 2001 and the vast majority will remain in place after 2050 when EU aims to achieve climate neutrality. Thus, renovation of the European building stock is needed to reduce operational energy and, therefore, emissions. In order to decide upon an energy renovation, one needs first to model the existing building, typically using energy simulations. Urban energy modeling for large building stocks is necessary in order to increase the rate of renovation; this requires large sets of data. However, personal inspection of the whole building stock is not realistic; on the other hand, large-scale open-source databases lack the required level of detail.

This thesis project investigates a methodology for semi-automatically generating building energy models at urban scale. The energy models can be created based on open-access databases: OpenStreetMap, BETSI database, and Energy Performance Certifications database. OpenStreetMap is an open-access map database, and it contains building footprints. BETSI database is based on building statistics for Sweden, and it includes detailed construction information and thermal properties. Energy Performance Certifications database contains general construction information for specific buildings such as heated floor area, floor numbers, and energy performance of buildings.

The developed methodology can derive the building footprint data from the OpenStreetMap database. 3D building models are then created with geometry data from Energy Performance Certificates and BETSI databases. Thermal properties can be determined from the BETSI database to create building energy models. The global warming potential of building operational energy is later calculated by the climate impacts of heating and electricity use. This methodology was illustrated in several case-studies building blocks from different geographical locations in Sweden and construction periods.

Results on the case studies show that it is possible to semi-automatically generate building energy models that predict the energy performance without any input data from a user. The accuracy compared to measurements of space heating from the energy performance certificates was between 3% and 21%. However, more data is required to calibrate the building energy model for higher accuracy. This can be done by, for example, adjusting the simulation input data to fit the actual monthly energy use or by other input data from the user, such as the window-to-wall ratio.

Acknowledgement

First, I would like to express my sincere appreciation to my supervisor Ricardo Bernardo. Without his patient guidance and profound knowledge. Then I would like to thank my co-supervisor, Henrik Davidsson. His solid knowledge helped me a lot in the design of methodology. I would also like to thank Rafael Campamà Pizarro. We had many discussions and he inspired me through the whole process of the thesis project.

No beautiful words can express my gratitude, but I would like to dedicate this ancient Chinese verse to everyone who has helped me：

桃李不言, 下自成蹊.

Peach/plum blooms need not blow their own horns. Spontaneously sightseers come to them in droves.

Last but not least, I would like to thank my parents. They have given me the best support since I was born. I love you!

Thank you all!

Table of content

1 Introduction

1.1 Background

Global warming is a growing problem for human beings. It is mainly caused by greenhouse gas emissions due mainly to consumed energy produced from burning fossil fuels (United States Environmental Protection Agency, 2020). The Paris Agreement sets a goal of limiting temperature rise by two degrees Celsius (United Nations, 2015). As one of the first signatories, Sweden aims to achieve climate neutrality by 2045 (Bonde et al., 2020), while twenty-five municipalities anticipated the date to 2030.

The building sector is responsible for 40% of the energy consumption in the EU, and this leads to 30% of carbon emissions (European Commission, 2005). Residential buildings account for 66% of energy use and 64% of greenhouse gas emissions in the building sector (Lopez-Gonzalez et al., 2016).

In Sweden, almost half of the multi-family buildings were fabricated before 1970, and these buildings may perform poorly in the energy perspective (Swedish National Board of Housing Building and Planning, 2016). Thus, for multi-family building blocks, it is critical to reduce operational energy use by renovation.

1.2 Literature Review

In order to decide which renovations are viable, one needs first to understand what the current performance of the building is. This is normally done by energy simulations. However, collecting detailed building information is time-consuming and difficult to carry out in large scale. As a simplification, sample buildings or archetypes can be used to represent building stocks (Swan and Ugursal, 2009). In a previous study, Érika Mata and colleagues used building regulations and codes from France, Spain, and other countries to create archetypes (Mata et al., 2014). Still, the geometries of buildings were not considered when using archetypes to the estimate the performance of the national building stock.

The Swedish building code *Boverkets byggregler* (BBR) does not contain the detailed thermal requirement for building components, it sets an overall target instead (Boverket, 2013). Thus, thermal properties for different archetypes in Sweden must be defined from other databases.

There are two approaches used by building databases to describe buildings: *Top-down* and *Bottom-up*.

A *top-down* database describes buildings at a general level. It commonly includes addresses, aggregated energy use, economic information, and other information. Database using this approach lacks details of buildings (Johnston, 2011).

The Energy Performance certifications (EPCs) database uses this approach. The energy performance certificate is a mandatory requirement for all member states in the EU when constructing, selling, or renting buildings (European Parliament, 2002). EPCs should contain general construction information such as heated floor area and floor numbers, energy performance of buildings, location of buildings, energy improvement suggestions, and current standards and benchmarks. The EPCs are valid for ten years if no significant renovations or other changes are made to the home. In Sweden, selected EPC data is provided to the public by searching for location of buildings, such as addresses or postcodes (Li et al., 2019). Information for around 1.5 million buildings is included in Swedish EPCs. Information contained in EPCs is shown in Figure 1 below.

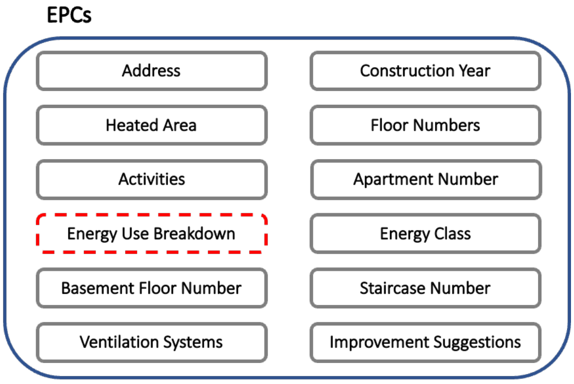

Figure 1 Content of Energy Performance Certifications

A *bottom-up* database contains detailed information on building constructions such as materials used in walls and roofs, window structure, basement construction, etc. This database is suitable for generating building energy models. However, expertise in energy modeling and extensive empirical data are needed. Besides, creating an energy model can be difficult and time-consuming even for an expert.

BETSI (*Bebyggelsens Energianvändning, Tekniska Status och Innemiljö*) is a bottom-up database. To gain a more comprehensive understanding of the technical characteristics of the national building stock, the Swedish government started the BETSI study in 2006. The study was commissioned by the National Board of Swedish Housing, Building and Planning with assistance from Statistics Sweden (Boverket, 2022). Around 1400 residential buildings selected from 30 municipalities in different climate zones (Boverket, 2010b) by Statistics Sweden (SCB) were investigated, including multi-family houses. BETSI study is anonymous, so it is impossible to identify the building data with locations. Information contained in BETSI is shown in Figure 2 below.

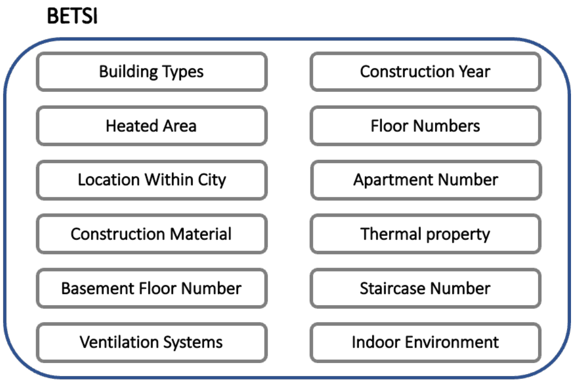

Figure 2 Content of BETSI

1.3 Purpose

As illustrated in Figure 3, this project exploits building energy modelling between *Top-down* and *Bottom-up* methods using open-access databases. Instead of presenting several archetypes to represent large segments of building stocks, this project developed a method to semi-automatically build energy simulation models of multi-family residential buildings without the need of any input data from the user. The simulation results can be used for the Global Warming Potential calculation.

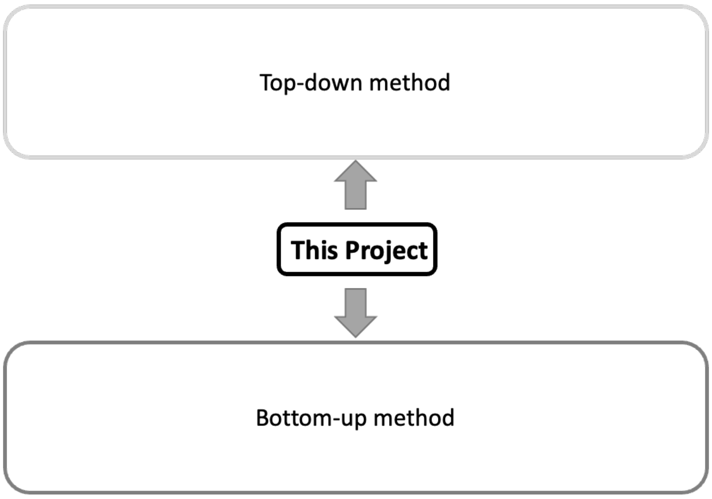

Figure 3 Position of the Master Thesis Project

1.4 Methodology development

An overall view of the semi-automated generation of building energy models is presented in Figure 4 below. Firstly, the building footprint can be derived from OpenStreetMap database (OSM). The floor number of the building can be determined using the EPCs database that contains addresses of buildings. A method was developed to get building construction details from BETSI using information from EPCs that is specific to the investigated building. Thus, floor height can be derived, and a 3D building model can be generated. Thermal properties can also be determined by the same method; therefore, the building energy model can be built. Figure 4 below represents the single building energy model generation, while it is also suitable for building blocks or entire neighbourhoods.

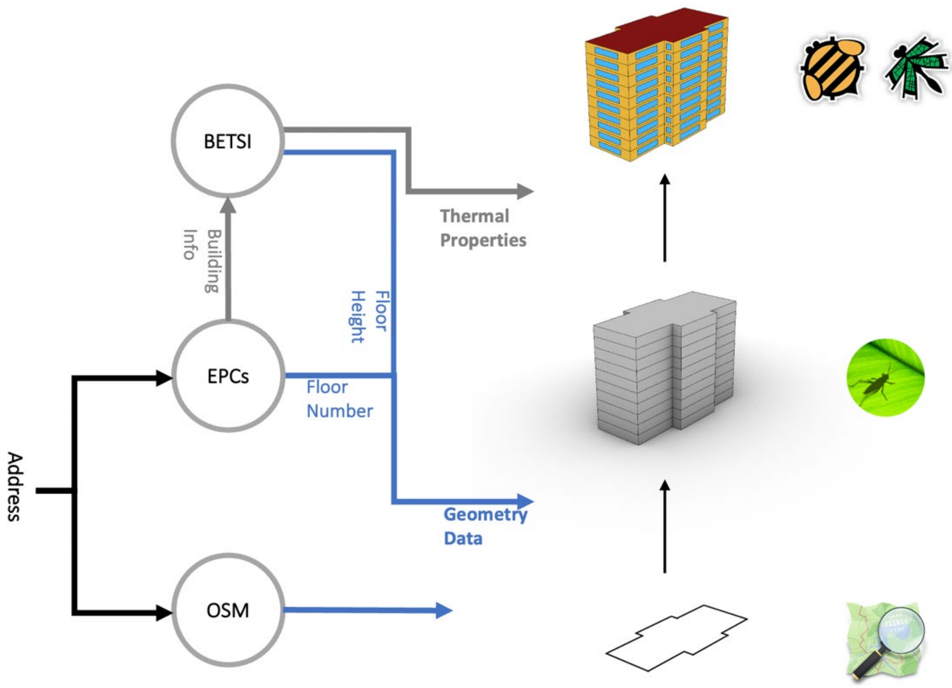

Figure 4 Flow Chart of the Thesis Project

Three key issues need to be addressed:

- How to process and derive thermal property data from BETSI?
- How to get building footprints from OpenStreetMap by addresses?
- And how to conduct the entire semi-automated building energy model generation process through databases, OpenStreetMap, and Grasshopper simulation?

These issues will be described in the next sections.

1.5 Software, website, and licensing

The open-source programming language Python is used to process data and implement automated functions. Python is a high-level combination of interpreted, compiled, interactive, and object-oriented scripting language, and it is designed to be highly readable. PyCharm was selected as the Integrated Development Environment for programming. Individual licenses for students and teachers are used to obtain the Professional Edition.

This project also used Rhino3D and Grasshopper. Grasshopper enables diverse types of calculations focused on geometry. The license was obtained through the Department of Architecture and Built Environment at Lund University. The mainly used plugin in Grasshopper is Ladybug Tools, including Honeybee and Dragonfly, to perform energy simulations for single buildings and building blocks. Ladybug Tools is a set of free computer applications. It integrates various simulation tools like Radiance, EnergyPlus, OpenStudio.

1.6 Limitations

This project relies heavily on the government databases BETSI and EPCs, as well as OpenStreetMap, so the completeness and accuracy of the databases and maps directly affects the accuracy and process usability of the default energy model input data. Besides, the derived data from BETSI can only perform a ‘qualified guess’ for building stocks if no technical data was updated for specific buildings.

2 Data processing in BETSI

The first attempt was to use machine learning to determine specific thermal properties for any building. By using several general information like construction years and building locations, the aim is to predict some energy simulation input data such as U-values of building envelope. This approach was later proven to be unfeasible. Machine learning generated models fluctuated too much in predicting the values of thermal properties. However, this first approach showed that two parameters correlate best with the U-value of the building envelope: the building construction year and building type. Since U-value is one of the most important parameters for building energy simulation, such an outcome was used to inform the second approach.

The second approach consisted in taking the energy simulation input based on groups of buildings divided by construction period and building type. Since building typology is not part of any of the databases, a machine learning model was trained to recognize building types based on general building information from EPCs.

2.1 Machine Learning

Machine learning is one type of artificial intelligence that encompasses an extensive range of concepts and methods (Joshi, 2020). This thesis project refers specifically to algorithms that can automatically build computational models of these complex relationships by processing the available data and maximizing the performance of the model (Baştanlar and Özuysal, 2014). Python plugins Scikit-learn (Pedregosa et al., 2011) and XGBoost (Chen and Guestrin, 2016) were used. Scikit-learn is a powerful machine learning library provided by a third party in Python, covering all aspects from data pre-processing to training models. Using Scikit-learn, the amount of code can be reduced so that the user can focus on analysing the data distribution, tuning the model, and modifying the hyperparameters. XGBoost is an optimized distributed gradient enhancement library designed for efficiency, flexibility, and portability.

The data contains 'feature' and 'label'. Features are normally considered as the input for the model, and labels are the output. The following descriptive sentence shows the difference between 'features' and 'labels':

*An animal with four legs and a tail that makes a "meow" sound is a **cat***

Where 'animal', 'four legs', 'a tail', and 'meow sound' are features, and 'cat' is label.

As for a building, 'features' and 'labels' can be demonstrated as:

*For a building built before 1960, with a heated floor area of 1000 m^2, located in the center of the city, the U-value of the exterior wall is **0.5 W/m^2K***

Where 'before 1960', '1000 m^2', 'center of the city' are features, '0.5 W/m^2K' is label.

As shown in Figure 5, the data used for building the model is called 'training data', and data for validating the model is called 'testing data'. It should be noted that, in this section, training data counts for 75% of the entire selected data to train the model, while test data is 25% for testing the model. After the model was trained, it will predict the values by inputting features from test data, then compare the predicted values (predicted labels) with the actual value (actual labels from the database).

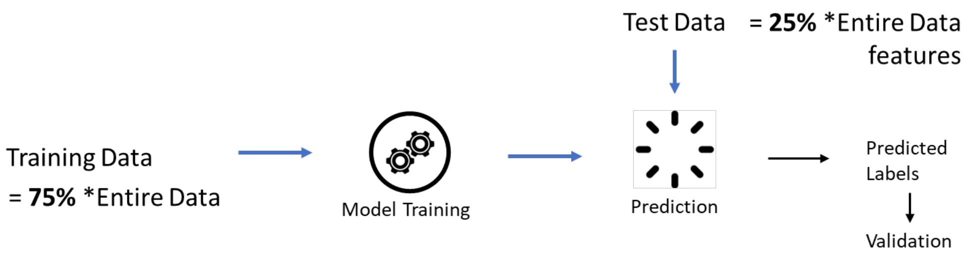

Figure 5 Division of training and testing data

Different machine learning models can deal with different types of problems. Most common three of which are regression, classification, and clustering. Regression, as explained in two-dimension in Figure 6, is the creation of a curve or line to fit a given data point and make a prediction, for instance, U-values of envelope components. The coefficient of determination ($0 < R^2 < 1$) is used to estimate the regression model. It describes how much the real values can be explained by the model. 0 means no explanation, while 1 means 100% explained.

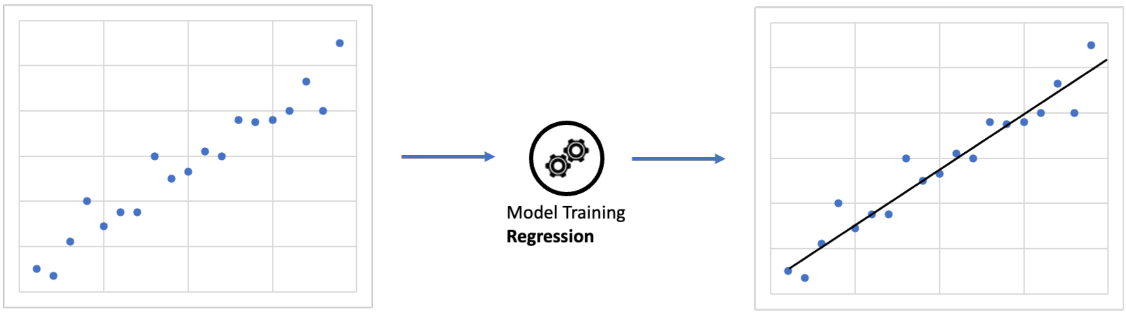

Figure 6 Diagram of Regression Method

As an example presented in Figure 7, classification also creates one or more curves. But the curves are used to split two or more data types. This type of model can be used for determining building types for instance.

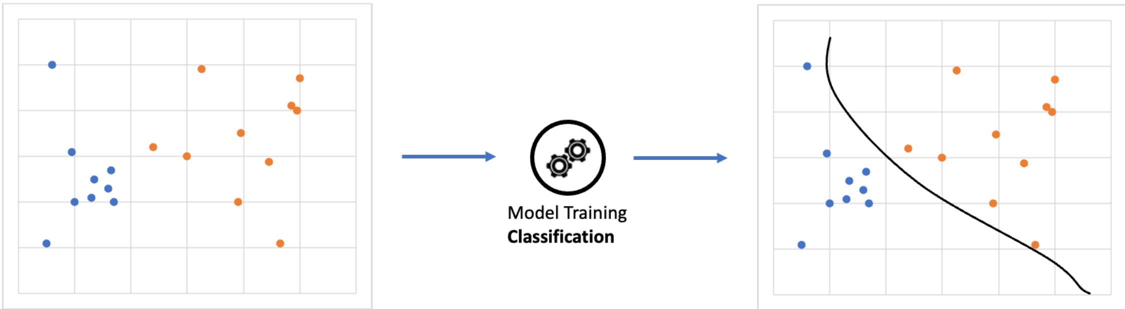

Figure 7 Diagram of Classification Method

The regression method was applied in the first attempt to determine specific thermal properties for any building while it was proven to be unfeasible. Then, the classification method was used in the second approach to determine building types that are not included in any open-access building database in Sweden. The two approaches will be demonstrated below.

2.2 Approach 1 - Regression for Thermal Properties

The first attempt used the regression method to predict thermal properties for each specific building. The target was to predict energy simulation input by simply entering general building information such as construction year, the number of floors, etc. This is illustrated in Figure 8. The labels of machine learning was thermal properties like U-values.

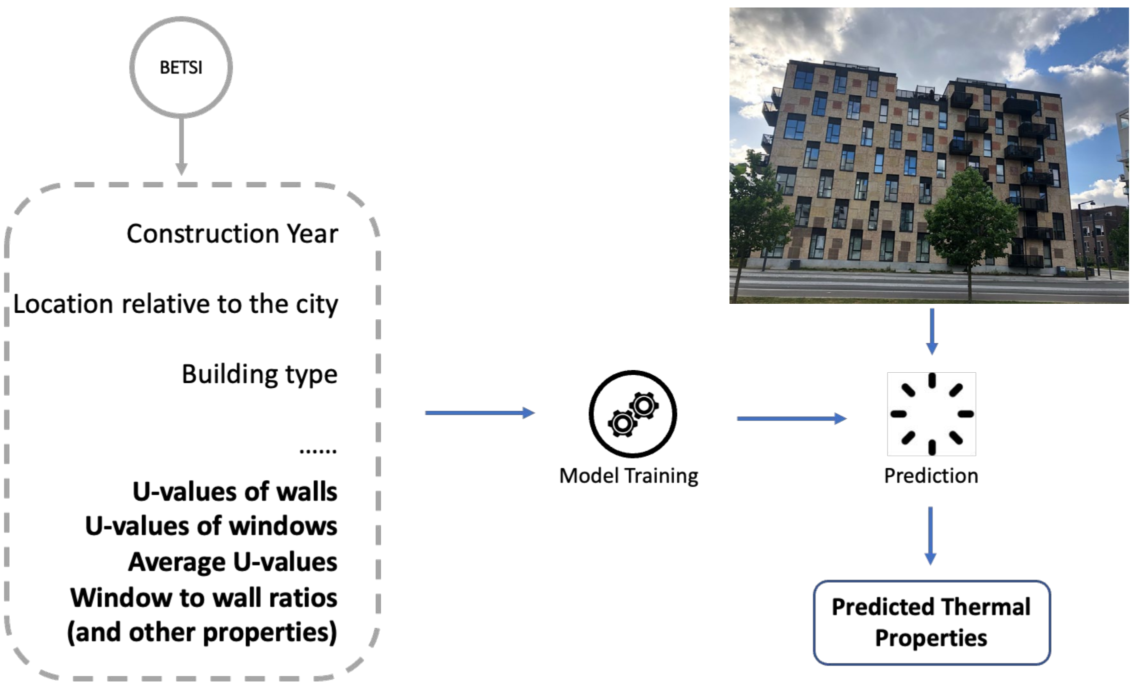

Figure 8 Regression for Thermal Properties

BETSI data was firstly pre-processed. All unrealistic values, i.e., U-values higher than 2 W/m^2K or lower than 0.1 W/m^2K, were discarded. The following is an example of the prediction of the U-value, which is one of the most critical parameters in building energy simulation.

Several features were chosen to predict the U-value of walls and windows, and the average U-value of the envelope. Features are building years, locations in cities, and building types (shown in Table 1).

Table 1 Features used in regression and their values/Classes

Features	**Content of features**
Building years	Pre – 2005
Location in cities	Förort, Innerstad, Villakvarter, Glesbygd
Building types	Lamellhus, Skivhus, Punkthus, Loftgångshus, Flerbostadsvilla, AnnanVilken, Radhus

(Only one Sutteränghus in multi-family houses, thus excluded)

Since the regression method cannot handle strings as input features, encoding, which means transferring strings to numbers, was applied to feature 'Location in cities' and 'building type'. These features are not ordinal, so the one-hot encoding rather than label encoding was used to avoid logistic errors. These are presented in Figure 9 below.

Lamellhus	0	1000000
Skivhus	1	0100000
Punkthus	2	0010000
Loftgångshus	3	0001000
Flerbostadsvilla	4	0000100
AnnanVilken	5	0000010
Radhus	6	0000001
String	**Label Encoding**	**One-hot Encode**

Figure 9 Ordinary strings and two different encoding methods.

There are several methods for regression: Linear Regressor, SGD Regressor, Ridge, and Random Forest Regressor, etc. These methods were compared to perform the best fitting of actual values. The most efficient regressor was Random Forest Regressor, and the best-fitting results are U-values of exterior walls, which are shown in Figure 10 and Figure 11 below. The fitting performs reasonably in the training set with an R^2 value of 0.74, while significant fluctuations can be observed. This directly leads to the poor prediction performance on the test set, whose R^2 value is only 0.13, which means the model can only explain 13% of the variation within the data. This indicates that the model is not able to predict specific energy simulation input data for any building.

It was also found that the construction year and the building type were parameters with the highest correlation with U-value which is the most important input for building energy simulation. These parameters were used in next section.

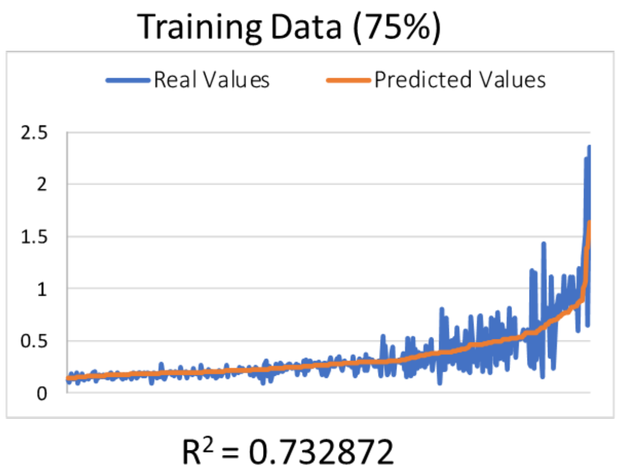

Figure 10 Charts of real values and prediction values for training set.

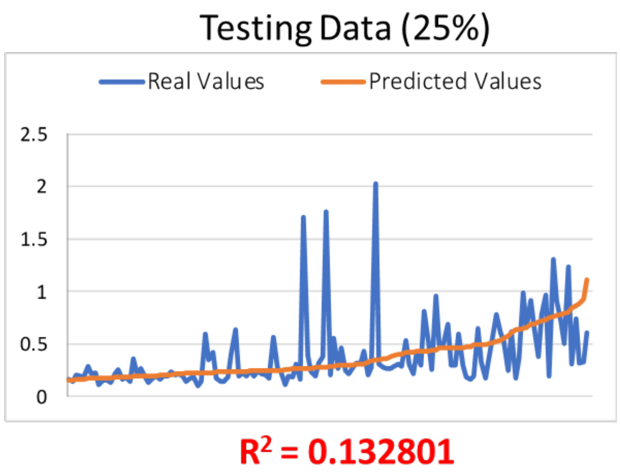

Figure 11 Charts of real values and prediction values for testing set.

2.3 Approach 2 – Grouping for Thermal Properties

The second method is to take the average of the eligible data within all categories. For each type, parameters will be average values to present all building stocks within that group.

All buildings were divided into five construction periods based on the BETSI report (Boverket, 2010a), which are pre-1960, 1961-1975, 1976-1985, 1986-1995, and 1996-2005. For each period, there are seven different building types: Lamellhus, Skivhus, Punkthus, Loftgångshus, Flerbostadsvilla, AnnanVilken, and Radhus. Thus, thirty-five groups of input datasets were determined, as shown in Figure 12. These two parameters were found most correlated with the U-value.

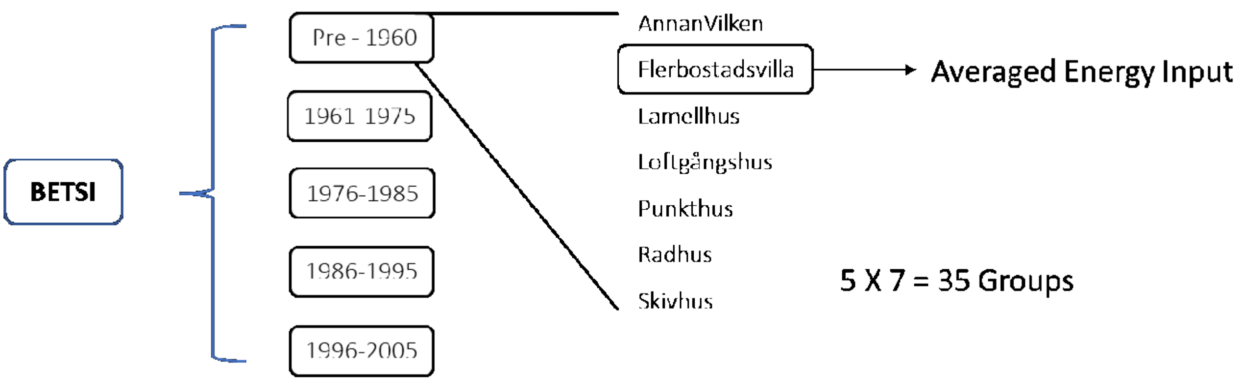

Figure 12 Groups of BETSI dataset for energy model input

Eighteen parameters were selected from the BETSI database, presented in Table 2 below. The values used in the simulation of each group are shown in Appendix.

Table 2 Parameters derived from BETSI database.

Parameters	
U-wall	Floor height Basement
U-window	Floor height Ground
U-ground	Floor height Typical
U-roof	Floor height Attic
A-roof	Average floor height
WWR_N	Ratio of Crawl Space
WWR_S	Ratio of Basement
WWR_W	U-wall base under ground
WWR_E	U-wall base above ground

2.4 Building Type Recognition

The EPCs includes construction years while it does not contain building types. Therefore, the classification method in machine learning was used to recognize building types by some general building information.

Several classification methods were applied to achieve better accuracy. DecisionTree and RandomForest methods were chosen from Scikit-learn to conduct the building type recognition. However, methods from Scikit-learn could not handle the 'None' value. If the 'None' value is simply replaced by the number 0, this could result in logical uncertainties. For example, some buildings may not have staircases or basements, and the numbers should be 'none', while in Scikit-learn, they must be transferred to '0'. Thus, XGBoost, which could calculate the 'none' value, was compared with the aforementioned methods.

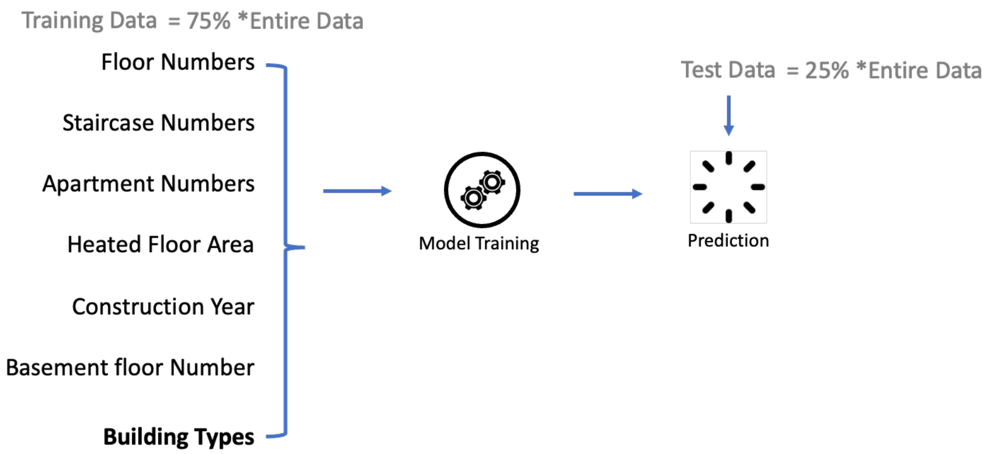

Figure 13 Flow chart of building type recognition method

Based on the definition and description of the building types (Vidén, 1985), six features were put into the model for training: construction year, apartment number, staircase number, floor number, heated basement floor number, and heated floor area (Figure 13). Testing data was the count for 25% of the entire BETSI database, as mentioned before. The highest accuracy rate was achieved by the XGBoost classification method, which is 84%. The classification tree can be found in Appendix.

Therefore, for any building existing in EPCs, it can be categorized as one group, and then thermal properties can be applied to the correspondent archetype (Figure 14).

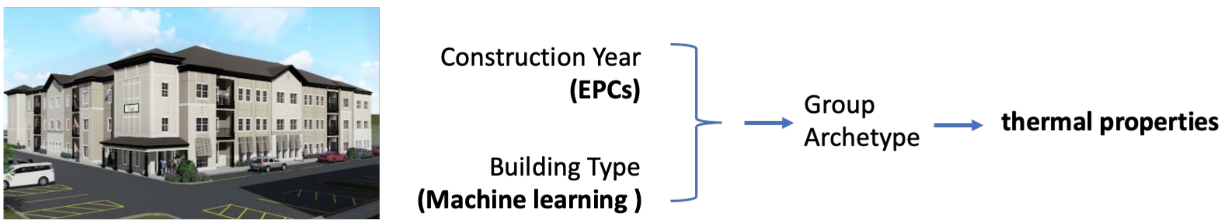

Figure 14 Getting building thermal properties

3 Other Input and Assumptions

3.1 Thermal Bridges and Air leakage

There is a significant thermal loss through thermal bridges and air leakage of the building envelope. However, determining the positions and values of these two types of thermal loss is too complex and time-consuming. To estimate these values, an assumption was made that an additional loss counts for 15% of the thermal loss (excluding thermal bridges and air leakage) should be added. Molander and Olofsson (2012) found that thermal bridges are responsible for one-third (5% from 15%) of the additional losse. The simulated energy use was raised by 5% to present thermal bridge losses in the later section.

Zou (Zou, 2010) provides a method to predict the infiltration rate at pressure difference at 50 Pa by building information such as year of construction and ventilation type, while the value used in EnergyPlus should be at the natural condition, approximately 4 Pa pressure difference (Sherman, 1980). Therefore, Zou's model of infiltration rate should be lowered by a factor determined by the Swedish standard SS 24300 (Swedish Institute for Standards, 2020). However, the model was proven not to perform ideally due to significant deviation and uncertainties (Berge, 2011). To avoid overestimating the airtightness of Swedish buildings, the air leakage value of 0.9 l/sm^2 was assumed for all the models in the simulation later.

3.2 SHGC

The solar heat gain coefficient (SHGC) represents how much solar energy can penetrate the window. Determining the SHGC can be complex. There are several methods to calculate the value (McCluney, 2002). In this thesis project, the SHGC value was assumed as 0.5, and it can be updated by field inspection of buildings of interest.

3.3 Ventilation

Ventilation from BETSI contains F-system, FVP-system, FT-system and FTX system. F-system means that the ventilation flow is driven by exhaust fans. FVP-system is a type of exhaust system that uses a heat pump to lower the temperature and thereby extract heat from the exhaust air before it leaves the building. FT-system involves both supply and exhaust fans with associated ducting. FTX-system is a type of FT system that uses a heat exchanger to transfer a large part of the exhaust air heat to the supply air (Boverket, 2007).

However, ventilation data of multi-family houses was not included in the BETSI public available database. Thus, the Ideal Air system was used in honeybee and dragonfly simulations. Ventilation rates of the multi-family houses are between 0.28 and 0.35 $l/(s \cdot m^2)$ in BETSI and 0.35 $l/(s \cdot m^2)$ was used as default in the process.

3.4 Internal load

Internal loads consist of heat gains from people, equipment and lighting. The equipment and lighting schedule was derived from the project report held by SBUF (Westin, 2019) and are shown in Figure 15. The occupancy schedule was adjusted by Swedish values (Deru et al., 2011) and is shown in Figure 16.

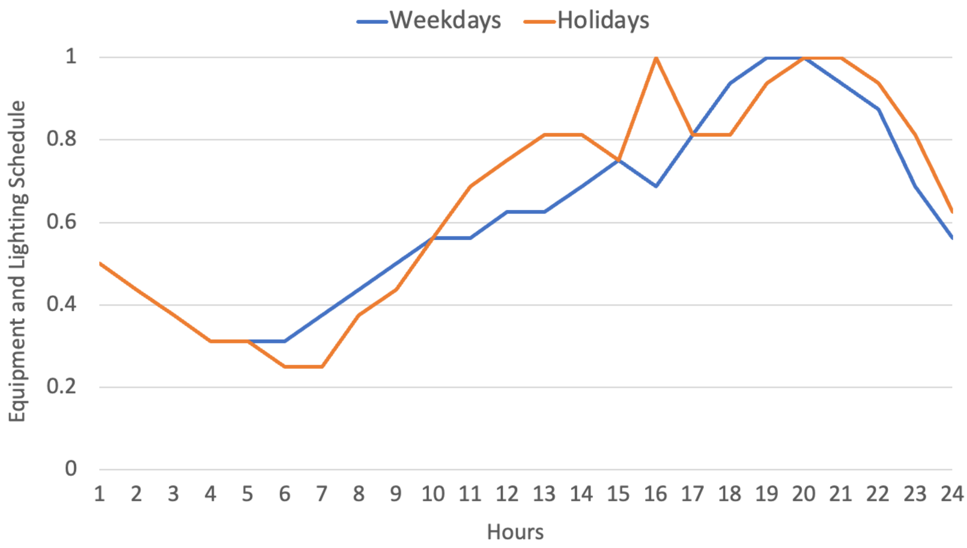

Figure 15 Equipment and lighting schedule

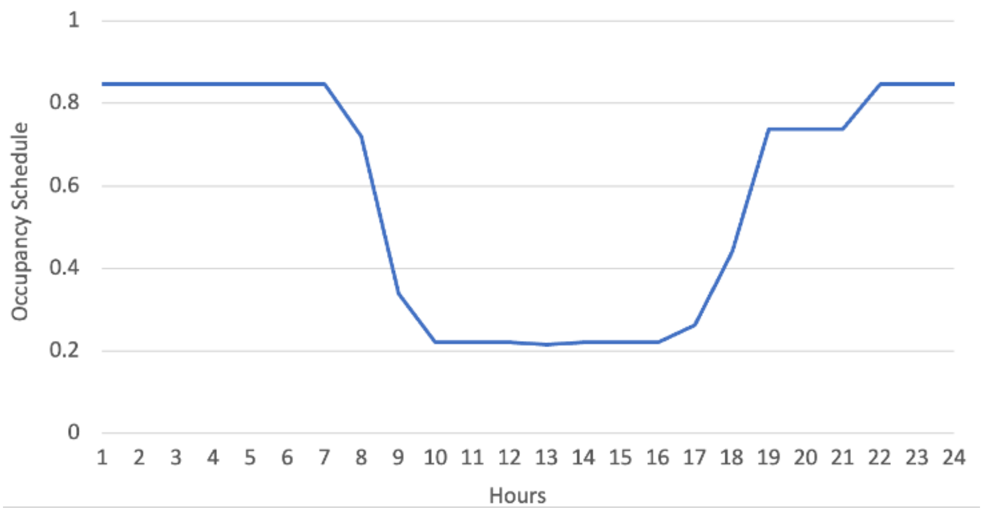

Figure 16 Occupancy schedule

Most internal load values were derived from SVEBY (Svebyprogrammet, 2012). People power was set as 80 W/m^2 as suggested. The total household electricity should be 30 kWh/m^2, and the proportion of lighting was 21% while 79% for equipment (Svebyprogrammet, 2012).

3.5 Global Warming Potential

Greenhouse gases are the components of gases in the atmosphere that contribute to the greenhouse effect. The Global Warming Potential (GWP) measures the impact of greenhouse gases on global warming. It is the relative ability to cause global warming by comparing a specific gas with the same mass of carbon dioxide (CO_2). The higher the GWP value, the greater the ability of the gas to raise the atmospheric temperature. This ability is measured in terms of equivalent CO_2 emissions (United States Environmental Protection Agency). According to NollCO_2, 22 kgCO_2 /MWh was considered for electricity and 60 kgCO_2 /MWh for district heating (Sweden Green Building Council, 2020).

4 Footprint extraction from OSM

OpenStreetMap is a free, editable map of the whole world created and maintained by nearly 5 million registered users and more than 1 million map contributors in every country in the world. It contains the most extensive building data (OpenStreetMap contributors, 2017).

As presented in Figure 17, OpenStreetMap consists of three levels: nodes, ways, and areas. Nodes are specific points defined by their latitudes and longitudes. Nodes also contain some information such as addresses and names. Ways are made of nodes as polygons representing linear elements like roads and rivers. Closed ways are boundaries of areas, for example, buildings or forests. The relation among these three elements is also presented in Figure 17.

This project collected closed ways building footprints from OpenStreetMap. Based on the address of the building, the building node was firstly found. Then the nearest building polygon was searched which is the building footprint. Building context can also be found by increasing the searching distance.

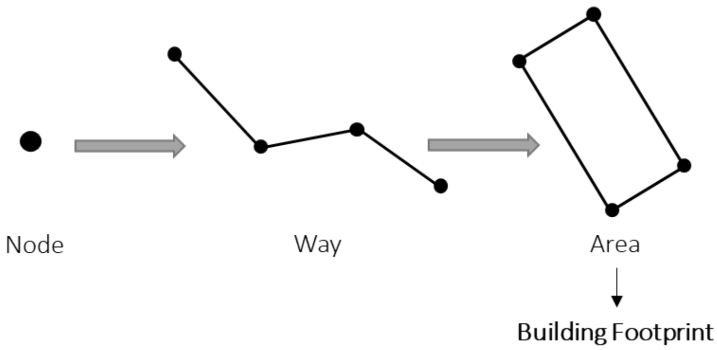

Figure 17 Components in OpenStreetMap

Data from OpenStreetMap was derived by a Python plugin 'OSMnx' (Boeing, 2017). OSMnx is a Python package that allows download of geospatial data from OpenStreetMap and models, projects, visualizes, and analyses real-world street networks and any other geospatial geometries. Information can be easily downloaded and analysed via the tool.

Firstly, the address of the chosen building was converted to the node's position with latitude and longitude by Nominatim (Nominatim Contributors). Nominatim is a tool to search OSM data by name and address (geocoding) and generate synthetic addresses of OSM points (reverse geocoding). The command in Python is 'nominatim.geocode()'. The process is shown in Figure 19.

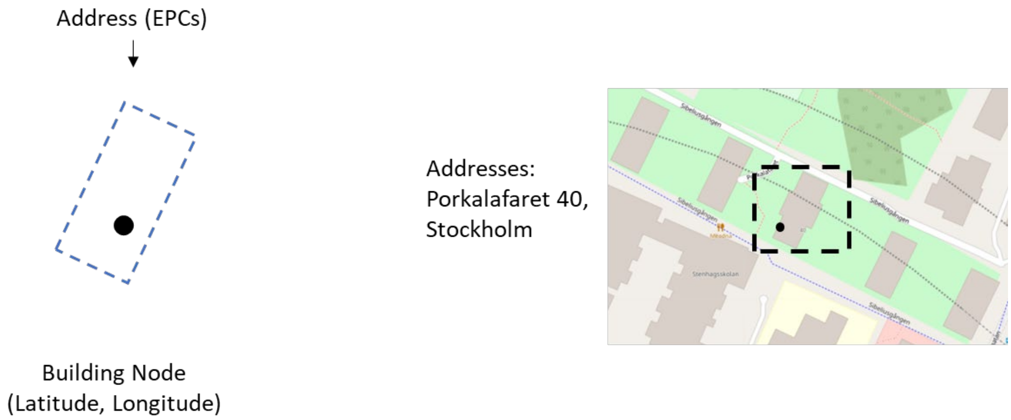

Figure 18 Get building node from address.

Based on the latitude and longitude of the building node, building footprint can be conveniently derived from OpenStreetMap. This was achieved by searching for the nearest polygon with 'building' tag, as shown in Figure 19 below. The command of OSMnx in Python is *'OSMnx.geometries_from_point()'*. It should be noted that the searching distance is defined as a constant in the command. Thus, a 'while' statement was used to define the best search distance. The searching distance started from 1 meter and would be increased by 0.5 meter if no polygons were found.

Scale factors were determined to adjust the footprint area to fit the ground floor heated area which is measured data found in EPC.

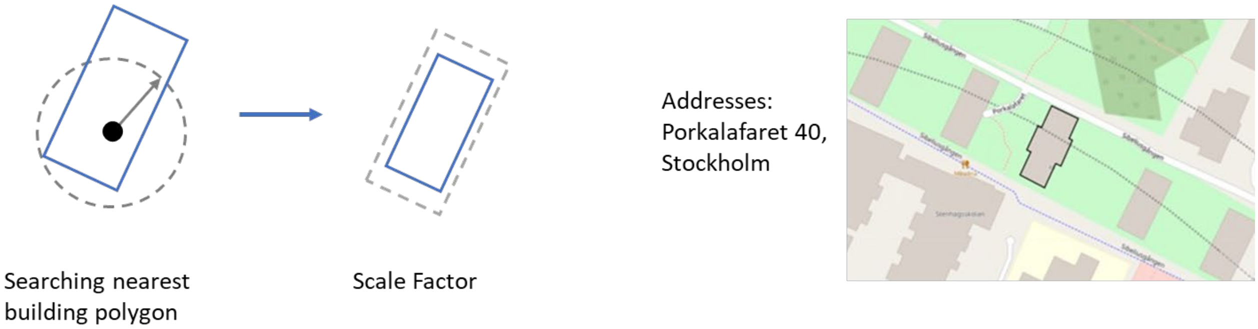

Figure 19 Searching nearest building polygon from the building node.

As illustrated in Figure 20, footprints for building context can be added by increasing the searching distance. The distance was set as 100 meters.

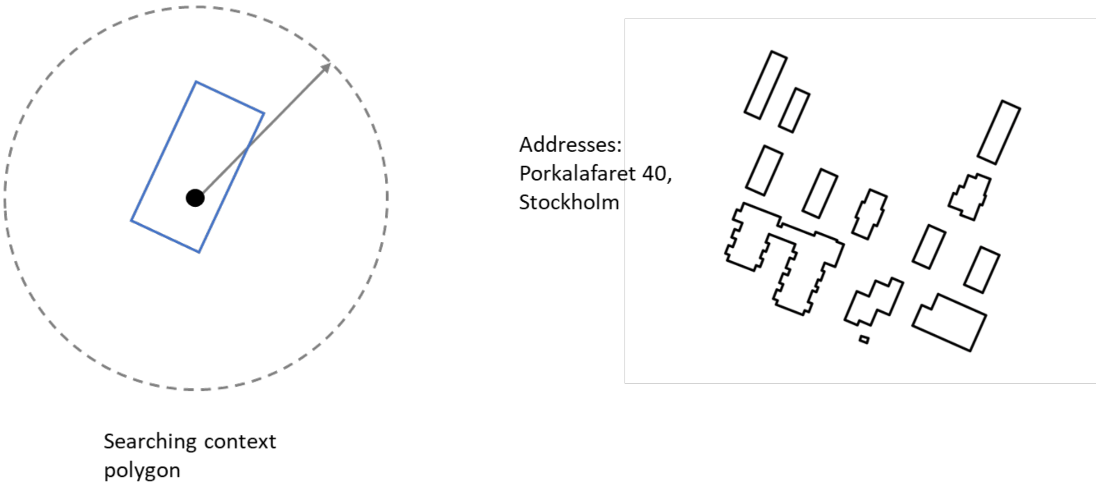

Figure 20 Searching polygon data of building context

This method can also be used to extract building blocks. Based on several addresses for the building block, several building nodes can be searched, and therefore the building polygons for each building within the neighbourhood can be derived. Using this method, all buildings illustrated in Figure 21 were selected while only four building nodes were shown.

Building context was not included when estimating building blocks. This is because the buildings were the building context for each other in the neighbourhood.

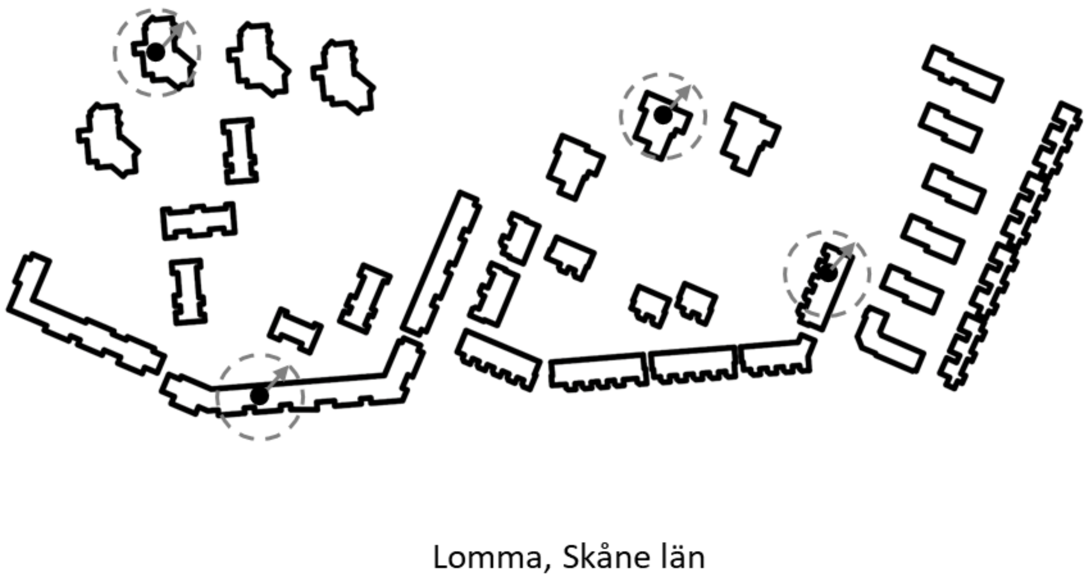

Figure 21 Getting building footprint for building blocks

Building footprint data (polygon data) was converted from latitude and longitude coordinates to metric coordinates, then was normalized and inserted into Grasshopper. This will be presented in the next section.

5 Implementation of the Semi-automated Process

The semi-automated methodology was successfully carried out by the XLS files and custom Grasshopper components. XLS files can store information about buildings, including geometry and energy simulation data. The custom Grasshopper component can process and extract data to Grasshopper for building 3D models and energy models (Figure 22 and Figure 23). The code for the custom grasshopper component for single buildings can be found in Appendix.

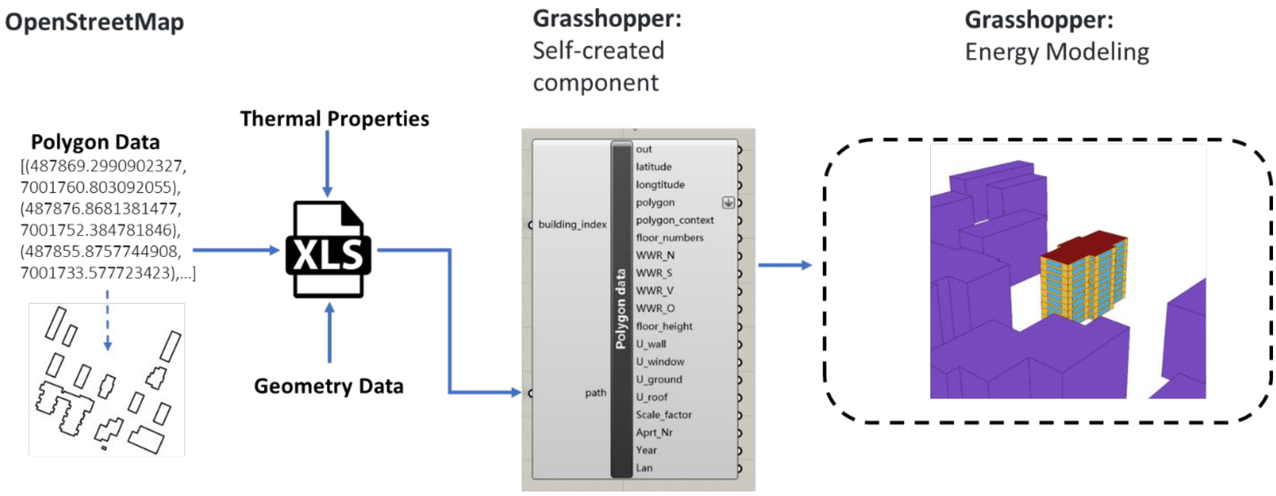

Figure 22 Automated modelling process for single buildings

The process is also the same for building block simulations (Figure 23). The code of the component for building blocks is presented in Appendix.

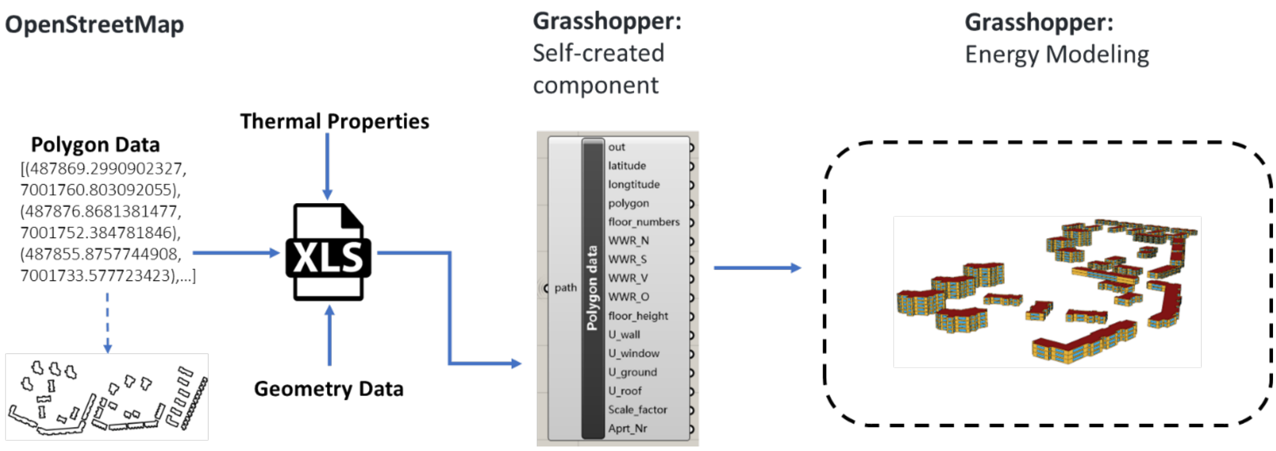

Figure 23 Automated modelling process for building blocks

The polygon data from OSMnx consists of multiple point coordinates, which are directly converted from latitude and longitude to metric data, so the data values are large and need to be normalized. With polygon data, floor numbers, floor height, and scale factor, shoe boxes can be generated as floors to represent the building as described in Figure 24. With polygon data as the bottom surface and the floor height as the height of the shoebox, a floor-by-floor 3D building model can be created.

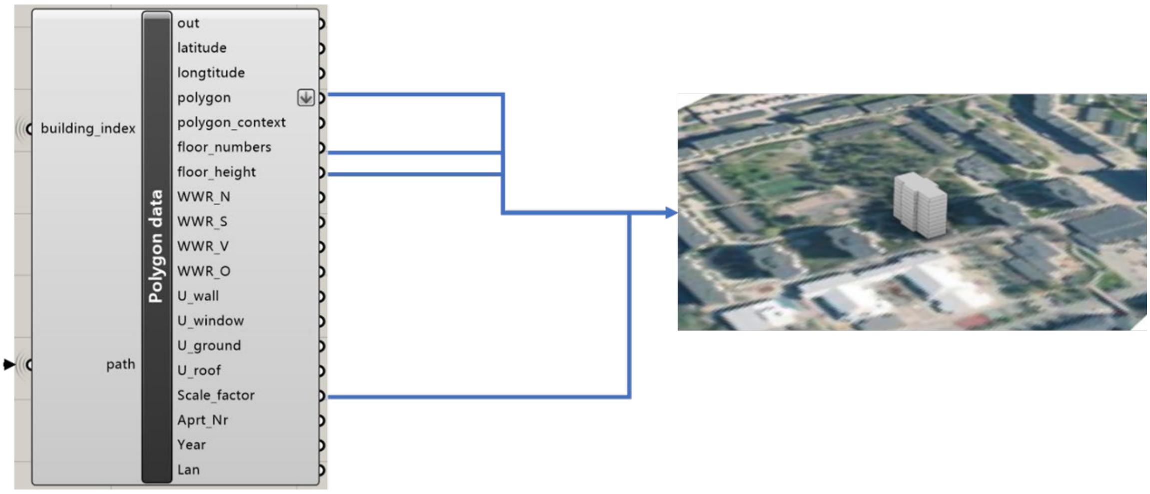

Figure 24 Generating building geometry

Window to wall ratio of each façade and U-values of envelope components can be extracted by the custom Grasshopper component to build energy models after creating the 3D model of the building. Grasshopper plugin 'Honeybee' was used for single building energy models.

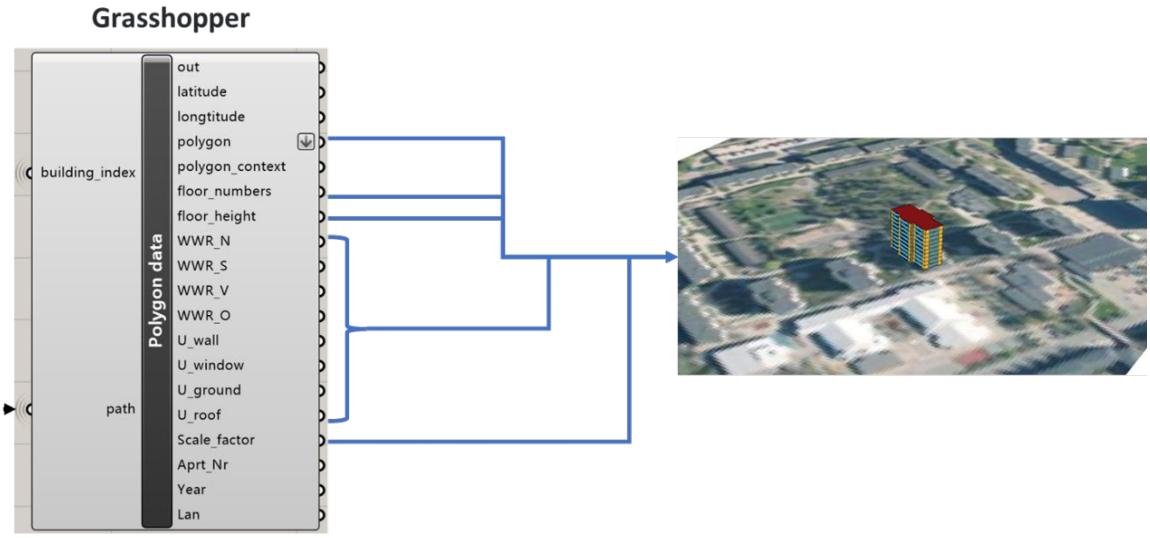

Figure 25 Generating energy model

Building context can also be modelled by the polygon of context as Figure 26 shown. Heights of building context were set as the same as building as default.

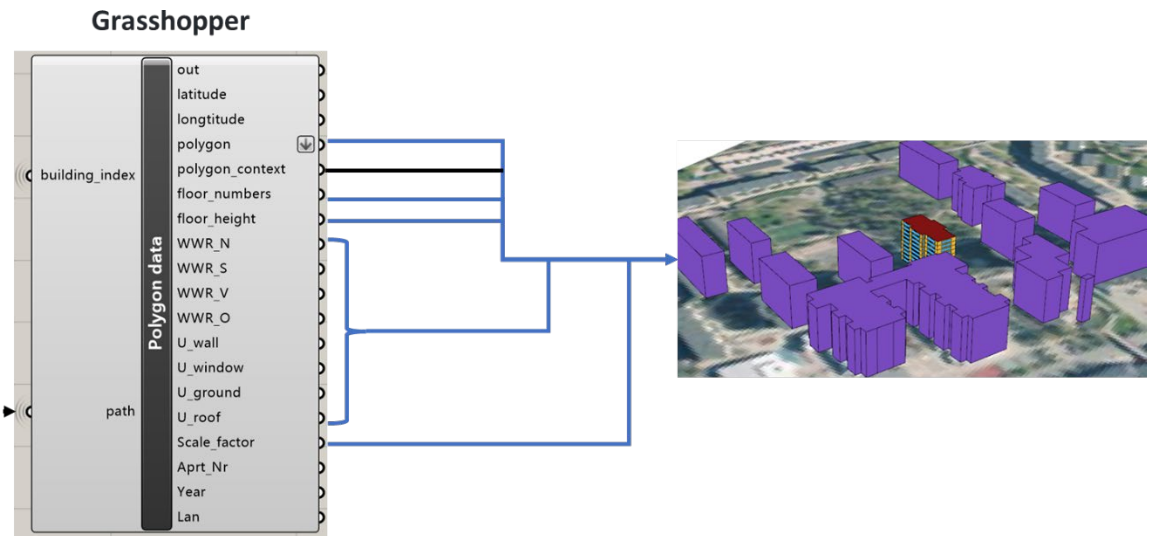

Figure 26 Generating building context

The building blocks model can be created in the same manner (Figure 27). Dragonfly was used to perform the energy simulation.

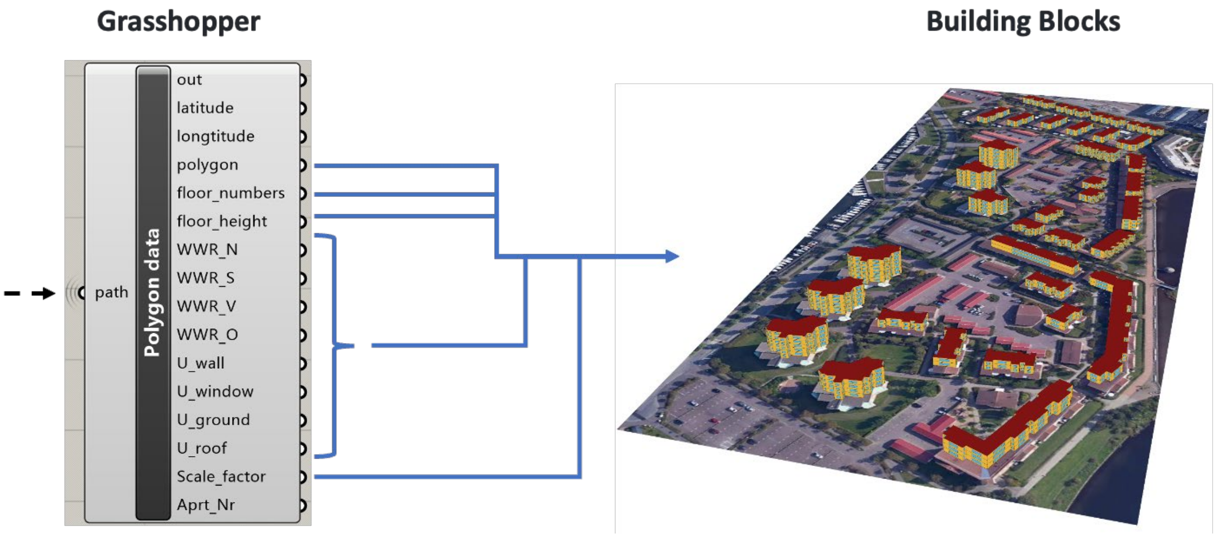

Figure 27 Generating building block models.

6 Cases study and Comparison

Five buildings from different construction periods and climate zones were selected to illustrate the process and results.

Figure 28 Positions of selected buildings

As shown in Figure 28, the selected buildings are from three cities. One building was from Östersund, Jämtlands county, two buildings were from Stockholm, Stockholm county, and two buildings were from Malmö, Skåne county. These three cities represent three different climate zones. All five buildings are heated by district heating. Since parts of the content of EPCs are confidential, the exact location of the building is not specified.

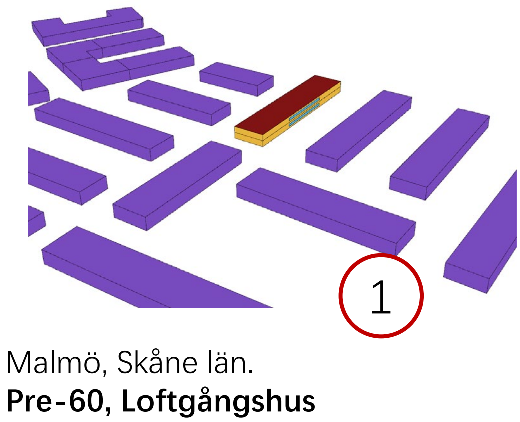

Figure 29 Building one

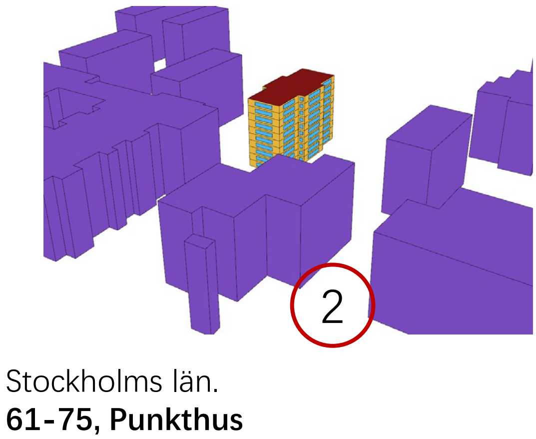

Figure 30 Building two

Building 1 located in Malmö, Skåne county and was constructed in 1935. The building type is Loftgångshus (Figure 29). Building 2 is located in Stockholm and was constructed in 1975. The building type Punkthus (Figure 30).

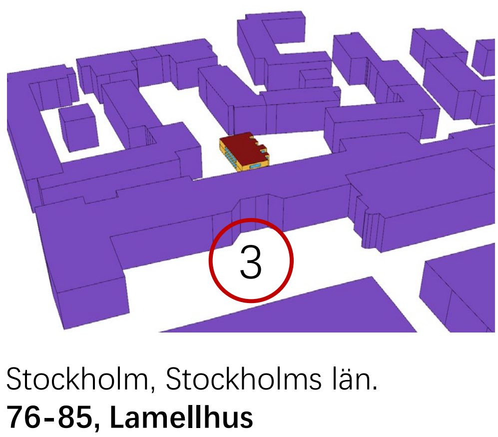

Figure 31 Building three

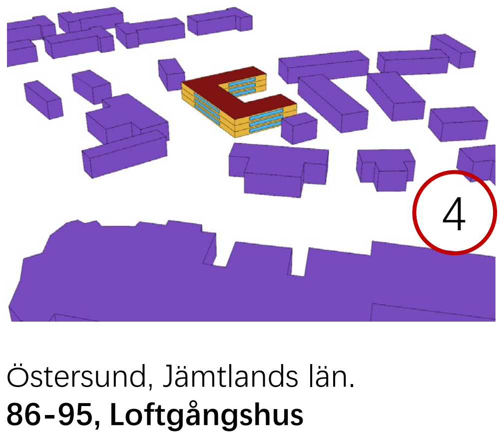

Figure 32 Building four

Building 3 is located in Stockholm and was constructed in 1985. The building type is Lamellhus (Figure 31). Building 4 is located in Östersund, Jämtlands county and was constructed in 1992.The building type is Loftgångshus (Figure 32). Building 5 is located in Malmö, Skåne county and was constructed in 2003. The building type is Punkthus (Figure 33).

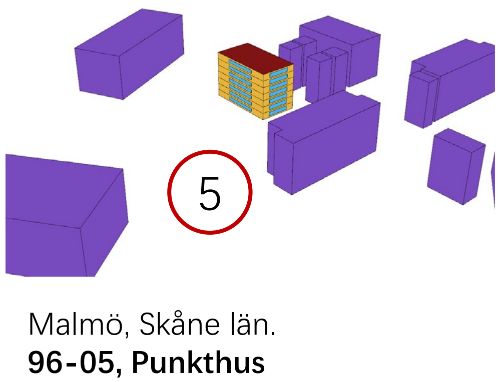

Figure 33 Building five

Five neighborhoods were also selected to test the simulation procedure at small-scale urban level. These blocks were selected from Skåne county and Värmlands country (Figure 34).

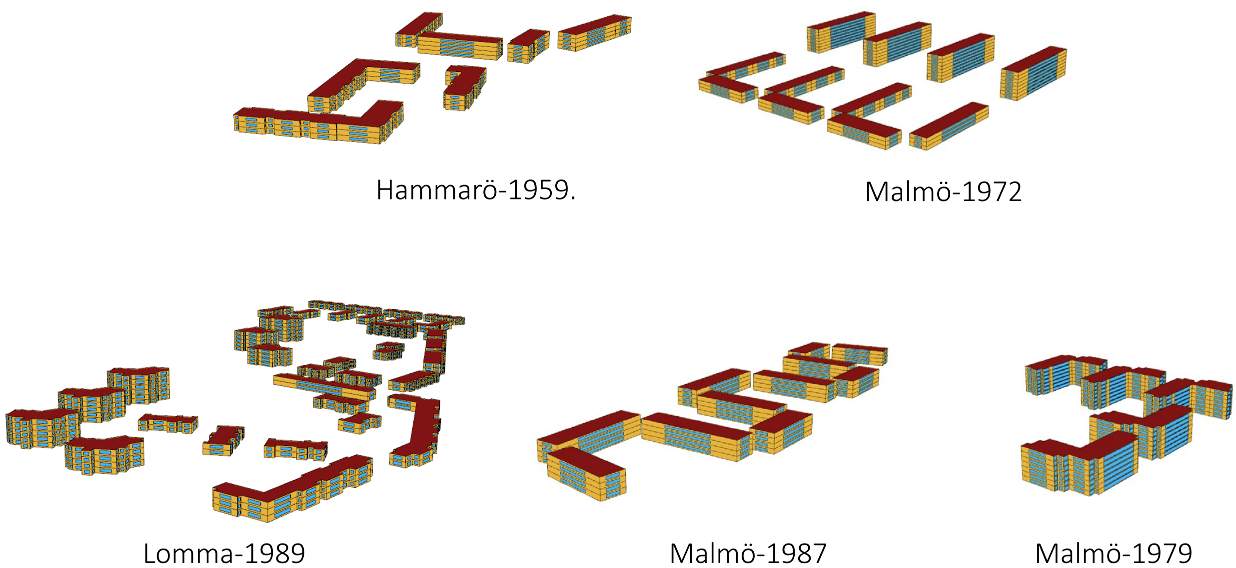

Figure 34 Five building blocks generated by the process

Comparing the heating energy in the EPCs with the results obtained from the automatic model, the maximum error was less than 21%, and the minimum error was 3%. The average error was 11.62%. This indicates that, only with the existing database, a reasonable building energy model can be automatically generated by the process developed in this thesis, without any specific input data from a user. Electricity was not compared because the electricity use depends on activities, equipment types, PV systems, and so on. Global Warming Potential is also shown in Table 3 (in CO_2 equivalent).

Table 3 Building information from EPCs and simulation results

Building	Construction Year	EPCs Heating Energy / (kWh/ m^2/y)	Simulation Heating Energy / (kWh/m^2/y)	Errors / %	GWP/ (kgCO_2eq/m^2/y)
Building 1	1935	91	110	21	7.23
Building 2	1975	121	107	12	7.06
Building 3	1985	108	128	19	8.33
Building 4	1992	116	111	4	7.35
Building 5	2003	79	76	3	5.24

Simulation results for five building blocks is shown in Table 4 below. It is impossible to compare the simulation data with EPCs energy data because the data from EPCs is aggregated. This will be discussed in the next section.

Table 4 Building block simulation results

Building Blocks	Construction Year	Simulation Heating Energy / (kWh/m^2/y)	GWP/ (kgCO_2eq/m^2/y)
Hammaro_1959	1959	99	6.62
Malmo_1972	1972	88	5.91
Lomma_1989	1989	90	6.06
Malmo_1987	1987	78	5.36
Malmo_1979	1979	86	6.10

7 Discussion and Conclusions

A methodology to generate building energy models in a semi-automated manner without any input data was developed in the project. Such methodology has potential but needs further development to be more accurate and useful in practice. The methodology was achieved by using open-access databases that contain building information. Several case studies building blocks were used to compare the simulated heating energy against measurements included in the Energy Performance Certifications (EPCs). Results of the case studies showed an average error of 11.6% as far as the heating energy use for the building is concerned. However, even if the energy performance of the generated building energy model matches exactly the measured heating energy from the EPCs, it is not possible to conclude that the building is properly modelled. Most probably, that is just a result of chance. More input data is required to calibrate the energy models and increase the accuracy of the results.

OpenStreetMap, BETSI, and Energy Performance Certifications were the databases that were used. BETSI database was used to derive 35 sets of building energy simulation input data categorized by the construction period and building types. A machine learning model to recognize the building type by general building information was developed, and the accuracy of which is 80%.

The most critical constraints with the model that leads to large inaccuracies are the data quality in the databases. The Energy Performance Certifications database is not perfect, and it leads to difficulties in comparing data with simulation results. For example, some buildings only contain summarized data for heating and electricity, so it is impossible to define which heat source the building uses. Heat source, which makes a difference in heat efficiency and equivalent carbon dioxide emission, contains district heating, biofuel, and several types of heat pumps. Furthermore, some buildings share the same energy value (in kWh/m^2) within the entire neighbourhood made of different buildings. This suggests that heating measured data is probably averaged for some buildings which makes the comparison between measurements and simulations of little use. Besides, some building information is also incorrect. For instance, different building areas are built with varying floors, while there is only one floor number in EPCs. Although OpenStreetMap is the most prominent building database globally, there are still some mistakes and a lack of information. Some buildings lack addresses, so addresses cannot access the building polygons. But because OpenStreetMap is an open map that the users can edit, it is simple to add addresses and other information to the buildings. The vast majority of building footprints are generated directly from satellite images by image recognition technology (Figure 35). Thus, in some cases, the building footprints are the shape of roofs. Figure 35 illustrates some building footprint shapes and related buildings' satellite pictures. This may lead to the inaccuracy of the building geometry. As mentioned before, users can also easily modify the shape of the building footprints.

Figure 35 Building polygons and satellite picture.

Besides, the other energy simulation input data, such as ventilation type and air leakage rate, should also be updated. The used values were based on literature review and assumptions, and this could not be accurate for all buildings.

To increase “matching” between the energy model and real building, the energy model can be adjusted and calibrated by two approaches (Figure 36). The first is to incorporate input data from the user. Users can determine thermal properties or geometry information by inspecting the building of interest, such as the window area. The more input data, the better the model. The second is to calibrate the model with monthly energy uses. An algorithm could be developed to automatically adjust the inputs to the energy simulation based on the BETSI database. The most likely combination of thermal properties can be obtained by matching the simulated monthly heating energy use with the actual use.

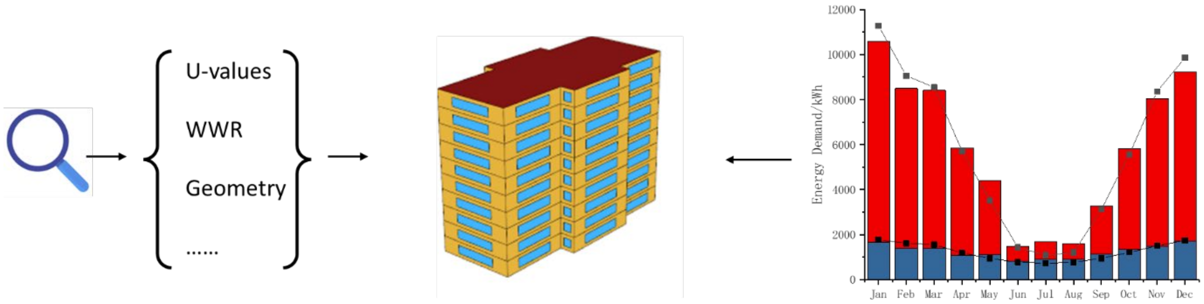

Figure 36 Approaches to improve the model

By using the developed methodology, building energy models can be simply created only by entering the addresses of buildings. The energy models can be used to support decision making regarding renovation alternatives and, therefore, to help reduce the building's operational energy use and greenhouse gas emissions. In other words, this project provided a convenient way to determine the potential of decreasing carbon emissions due to energy consumption.

8 References

BAŞTANLAR, Y. & ÖZUYSAL, M. 2014. Introduction to Machine Learning. *In:* YOUSEF, M. & ALLMER, J. (eds.) *miRNomics: MicroRNA Biology and Computational Analysis.* Totowa, NJ: Humana Press.

BERGE, A. 2011. Analysis of methods to calculate air infiltration for use in energy calculations.

BOEING, G. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems,* 65**,** 126-139.

BONDE, I., KUYLENSTIERNA, J., BÄCKSTRAND, K., ECKERBERG, K., KÅBERGER, T., LÖFGREN, Å., RUMMUKAINEN, M. & SÖRLIN, S. 2020. 2020: Report of the Swedish Climate Policy Council.

BOVERKET 2007. Bilaga 15 Instruktioner till Flerbostadshusprotokoll ver 2.

BOVERKET 2010a. Energi i Bebyggelsen—Tekniska Egenskaper Och Beräkningar—Resultat Från Projektet BETSI. Boverket Karlskrona.

BOVERKET 2010b. Teknisk status i den svenska bebyggelsen: resultat från projektet BETSI. Boverket Karlskrona.

BOVERKET 2013. *Optimala kostnader för energieffektivisering*.

BOVERKET. 2022. *Öppna data - Betsi är en undersökning om byggnaders tekniska status* [Online]. Available: https://www.boverket.se/sv/om-boverket/publicerat-av-boverket/oppna-data/betsi-oppna-data/ [Accessed 2022-05-08].

CHEN, T. & GUESTRIN, C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016. 785-794.

DERU, M., FIELD, K., STUDER, D., BENNE, K., GRIFFITH, B., TORCELLINI, P., LIU, B., HALVERSON, M., WINIARSKI, D. & ROSENBERG, M. 2011. US Department of Energy commercial reference building models of the national building stock.

EUROPEAN COMMISSION 2005. *Doing More with Less: Green Paper on Energy Efficiency*, Luxembourg: Office for Official Publications of the European Communities.

EUROPEAN PARLIAMENT 2002. Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings. *Official Journal of the European Communities,* 46**,** 65-71.

JOHNSTON, D. 2011. A physically-based energy and carbon dioxide emission model of the UK housing stock.

JOSHI, A. V. 2020. Machine learning and artificial intelligence. Springer Nature Switzerland AG.

LI, Y., KUBICKI, S., GUERRIERO, A. & REZGUI, Y. 2019. Review of building energy performance certification schemes towards future improvement. *Renewable and Sustainable Energy Reviews,* 113**,** 109244.

LOPEZ-GONZALEZ, L. M., LOPEZ-OCHOA, L. M., LAS-HERAS-CASAS, J. & GARCÍA-LOZANO, C. 2016. Energy performance certificates as tools for energy planning in the residential sector. The case of La Rioja (Spain). *Journal of Cleaner Production,* 137**,** 1280-1292.

MATA, É., SASIC KALAGASIDIS, A. & JOHNSSON, F. 2014. Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. *Building and Environment,* 81**,** 270-282.

MCCLUNEY, R. 2002. Suggested methodologies for determining the SHGC of complex fenestration systems for NFRC Ratings. *Cocoa Beach, FL: Florida Center for Solar Research*.

NOMINATIM CONTRIBUTORS Nominatim.

OPENSTREETMAP CONTRIBUTORS. 2017. *Planet dump retrieved from https://planet.osm.org* [Online]. Available: https://www.openstreetmap.org [Accessed].

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R. & DUBOURG, V. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research,* 12**,** 2825-2830.
SHERMAN, M. H. 1980. Infiltration-pressurization correlation: Simplified physical modeling.
SVEBYPROGRAMMET 2012. Brukarindata Bostäder Version 1.0. *Sveby: Stockholm, Sweden*.
SWAN, L. G. & UGURSAL, V. I. 2009. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and sustainable energy reviews,* 13**,** 1819-1835.
SWEDEN GREEN BUILDING COUNCIL 2020. NollCO2, Nettonoll klimatpåverkan, Manual 1.0. Sweden Green Building Council.
SWEDISH INSTITUTE FOR STANDARDS 2020. Energy performance of buildings — Part 1: Classification of power need for space heating.
SWEDISH NATIONAL BOARD OF HOUSING BUILDING AND PLANNING 2016. Underlag till den andra nationella strategin för energieffektiviserande renovering. *In:* BOVERKET (ed.). Karlskrona.
UNITED NATIONS. Paris agreement. Report of the Conference of the Parties to the United Nations Framework Convention on Climate Change (21st Session, 2015: Paris). Retrived December, 2015. HeinOnline, 2017.
UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. *Understanding Global Warming Potentials* [Online]. Available: https://www.epa.gov/ghgemissions/understanding-global-warming-potentials [Accessed 2022].
UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. 2020. *Sources of Greenhouse Gas Emissions* [Online]. Available: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions [Accessed 2022].
VIDÉN, S. 1985. Flerbostadshusen i Sverige: kvaliteter och brister, ombyggnadsbehov och möjligheter.
WESTIN, R. 2019. HUSHÅLLSEL I NYBYGGDA FLERBOSTADSHUS.
ZOU, Y. 2010. Classification of buildings with regard to airtightness.

LUND UNIVERSITY

Divisions of Energy and Building Design, Building Physics and Building Services

Department of Building and Environmental Technology