PREFACE

The basis for this research is my personal goal for working in something that impacts positively in societies and the environment. It is my passion to increase my understanding of energy systems and how new technologies and data science can be used to increase energy efficiencies and become more environmentally friendly. The thesis has supposed a challenge as I started it without any knowledge about data analysis, Python or Geographical Information Systems. I faced critical moments due to data manipulation and spent most of the time in data collection and preprocessing of the databases. However, I can say I am proud to have learned and applied these tools and techniques, and I am sure they will bring me new opportunities in years to come.

I could not have achieved these results without strong support from my supervisor Helge Brattebrø and co-supervisor Nina Holck Sandberg. Thank you for all the technical but also the motivation you provided me during the last year.

I want to express my gratitude to Simon James Loveland, from Trondheim Municipality to Åmund Utne, from Statkraft Varme and Johnny Tangen, Roar Skauge and Rolf Hilstad, from TronderEnergi, for meeting me and give me access to the necessary data as well as showing me the business and political perspective for my thesis.

Thanks to Ruslan Zhuravchak, who inspired me with his incredible work and passion for data and to Carine Lausselet who made me believe I was doing something useful when I was through some moments of doubt. Also, thank you to Karl Henning Omre from the Department of Mathematical Sciences for his time expended in helping me to develop further ideas for the study.

I would also thank all my friends; it has been incredible to share this experience with you. Also, thanks to my friends in Spain who have supported me and have shown me that the distance is not a barrier in friendship.

Especial thanks to my aunt, my grandmother, my cousin and my father who have always believed in me and have taught me the power of patience to achieve my personal goals. Also, thank you for always be there when I need you. Thank you Ester, Vicente and Jesus for all your help to make this came true and for all the incredible trips we made together in Norway.

But… I am especially grateful to my boyfriend Rodrigo for being my major personal support and for these two last years enjoying incredible moments together.

ABSTRACT

Reducing emissions and increasing the efficiency of human activities are at the top of research and political agendas. Changes in the energy system are required if we want to reach the goals and targets set to reduce the effects of climate change. Norway has an environmentally friendly energy mix, with hydropower accounting with more than 95% of power production. However, it is also a country with one of the highest energy demand per person, and it appears to continue increasing in the years to come with the rise of electric vehicles (EV). In this context, capacity problems might arise if we are not able to balance demand and supply and increase efficiencies in all sectors.

Norwegian building stocks consumes significantly more electricity than other Nordic countries, especially for heating purposes. Developing, and put into effect, measures to further reduce the power demand from buildings will come together with more electricity capacity for other purposes. The deployment of Variable Renewable Energy Sources (VRES), such as wind power or photovoltaics, the extension or implementation of district heating networks in urban areas, building refurbishments and Near Zero Energy Buildings (NZEB) might, in combination, lead to this target.

The study and understanding of building stocks and their energy consumption are needed in order to implement the abovementioned solutions efficiently and smartly. Recently, technical models have been developed to understand the actual performance of residential and non-residential buildings. Also, dynamic models have presented future forecasting of how different scenarios might change the energy consumption of this sector. In parallel, Geographical Information Systems (GIS) have emerged as a tool that allows spatial visualisation and the creation of energy maps. Moreover, data analysis allows for obtaining relevant information about consumption patterns.

The present study aims to provide a methodology and a tool that combines GIS and data analysis for the study of the spatial dimension of the energy use of building stocks. We apply it to the case of non-residential buildings in Trondheim, Norway.

A georeferenced dataset for the building stock and two non-georeferenced datasets of electricity and heat consumptions were collected from different sources to be the inputs of the model. After cleaning and processing the databases, we calculate the energy intensities by type-cohort archetypes and assign them to all non-residential buildings of Trondheim municipality. With this information, we develop a visualisation tool that allows the creation of energy maps in different resolutions and the analysis of the energy consumption distribution. Finally, we perform a comparative analysis to examine the difference between real and calculated energy intensities from engineering models.

The data analyses and energy intensity calculations were carried out by developing several scripts and functions in Python 3.0. ArcGIS Pro is the GIS employed for the manipulation of the georeferenced dataset and the visualisation of energy maps.

Results from the case study show that heat and electricity consumptions in non-residential buildings are not linearly correlated with neither the year of construction nor the technical requirements of buildings. However, prebound and rebound effects are identified after a comparative analysis of real and calculated energy intensities, especially in old and new buildings. Therefore, user behaviour has important impacts on the energy consumption of non-residential buildings. Finally, we see an unequal energy distribution in Trondheim. Areas with high aggregated energy consumptions are within the district concession area, thus its expansion may help to decrease the electricity demand of buildings located in these areas. On the contrary, more remote regions tend to have higher mean energy intensities and lower aggregated energy consumptions, thus implementing refurbishment plans appear as the best solution to increase the efficiency of buildings and reduce electricity demand.

Even if the data collection and processing is time-consuming and there is uncertainty after data cleaning and merging, the model allows for the systematic combination of energy suppliers databases and georeferenced building stock data to proceed with spatial energy analyses. Finally, the thesis explains how the model may provide a fundamental tool for decision making and more effective solutions.

Data and spatial analysis of the energy use of building stocks. Case study of non-residential buildings in Trondheim

Raquel Alonso Pedrero[a*], Nina Holck Sandberg[a], Helge Brattebø[a]

[a] *Industrial Ecology Program and Department of Energy and Process Engineering. Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway*
**Corresponding author*

Keywords: Building stocks, Energy analysis, Spatial analysis, GIS, Data analysis

ABSTRACT

A methodology is developed for data and spatial analyses of energy consumption in building stocks and to examine their importance in decision making. The work draws upon georeferenced datasets of building stocks and measured energy datasets reported by energy suppliers. It is applied to Trondheim to analyse non-residential buildings and evaluate its energy distribution. In Norway, electricity availability has become an issue due to capacity problems, therefore we study on possible power efficiencies and the extension of district heating networks. Results from the case study show that heat and electricity consumptions in non-residential buildings are not linearly correlated with neither the year of construction nor the technical requirements of buildings. However, prebound and rebound effects are identified after a comparative analysis of real and calculated energy intensities. Therefore, user behaviour is an important factor for real energy use in non-residential buildings. Finally, we see unequal energy distributions within Trondheim municipality. Areas with highly aggregated energy consumptions lie within the district heating concession area, thus its expansion may help to decrease their electricity demand. Remote regions tend to have higher mean energy intensities and lower aggregated energy consumptions, so energy-upgrading is the best solution to reduce their electricity demand.

1. Introduction

1.1. Background and context

There is a widespread agreement on human activities being the main reason for the increasing emissions of greenhouse gases (GHG) (IPCC, 2014), especially, those in which combustion of fossil fuels is required. In Europe, fossil fuels continue dominating as the primary energy source, with 72% of share in 2016 (European Environment Agency, 2019). Hence, improvements in the energy system are needed to reduce environmental impacts. In this context, and with increased awareness in society, European policymakers agreed upon the establishment of the 2020 climate and energy packages and the 2030 climate and energy framework. One of the key sectors in Europe, contributing to the release of 36% of CO2 emissions is the building sector, which consumes close to 40% of the final energy (European Commission, 2018). Increasing the energy efficiency of buildings and the share of renewables in the power supply system are some of the measures to reduce the environmental impact of building stocks dramatically.

Norway, as part of the European Economic Area (EEA), is anchored to the Climate Act and committed to becoming carbon neutral by 2030. The country has a significant low-cost, highly flexible and zero-carbon generation of power due to its high share of hydropower (40% of total primary energy supply) (IEA, 2017). Clean electricity has given access to a higher level of electricity use for space and water heating compared with other countries (IEA, 2017). Several measures are taking place in the building sector in order to increase energy security and have a more-efficient and climate-friendly energy use. Some of these measures are the development and research on Near Zero Energy Buildings (NZEB), the promotion of refurbishments on actual building stocks, and the approval of stricter technical codes. Their combination might have the potential to lower the need for electricity in buildings and decrease capacity issues so that there is more place for the deployment of electricity for other sectors. Also, they may reduce the appearance of peak loads. By this way, Norway may be less dependent on imports or the use of fuels in the future while leading to the de-carbonisation of other high-pollutant sectors. Along with demand-side projects, the supply-side is also facing substantial improvements, especially by the expansion of district heating and the adoption of renewable energy sources (RES).

Therefore, the country is rapidly facing changes in its energy system and its building structure that requires tools and further research for a better understanding of their synergies. Sartori et al. (2009) developed an archetype model to study the effectivity of thermal carriers, heat pumps and conservation measures in reducing the Norwegian building stock energy demand. Results show that conservation measures on a large scale might drive to a reduction of energy consumption. Dynamic dwelling stock models have also been used to explain historical energy use and forecast future scenarios for the Norwegian building stock. Sandberg et al. (2016) use a segmented dynamic dwelling stock model to explain its historical changes. The model is combined with archetype-specific energy intensities to estimate total energy demand. Its application to the Norwegian dwelling stock has shown the importance of shifting to more efficient energy carriers and heating systems for energy savings. The authors also conclude that user behaviours might offset improvements. Sandberg et al. (2017) use the same dynamic stock model to quantify future energy savings. Renovation rates are used to explain the development of the building stock. The research shows a potential decrease in total delivered energy of 52% for the most optimistic scenario, a reduction achieved mainly through advanced renovations and extensive deployment of heat pumps or photovoltaics.

Although these energy performance analyses provide useful insights at an aggregated level, comprehensive energy performance analyses at urban scales may provide more detail comprehension on building energy demand. That is the reason why the spatial dimension has recently appeared as an essential building stock attribute for energy analysis and planning at regional levels. Geographical information systems (GIS) are flexible tools that allow implementing spatial attributes by managing geodata, boosting new modelling techniques and promoting new ways of visualising results and analysing a wide variety of urban features (Li, 2017). Furthermore, GIS has shown to be a key tool in the assessment of RES generation potential, the construction and maintenance of distribution systems as well as in the understanding of energy consumption and heat transfer (Resch et al., 2014).

In recent years, energy and heat maps of different building stocks have been developed using GIS. Möller (2008) explains the necessity to quantify and localise end-use energy consumption and develops a heat map for the Danish building stock. The author combines the national dwelling register with a spreadsheet model for physical heat loss of buildings for the design of the heat atlas. It establishes a general method for future maps and improvements and proposes possible future applications. Other studies rely on statistical methodologies. For instance, a tool for policymakers and sustainable urban planning was developed using regression analyses for the city of Rotterdam (Mastrucci et al., 2014). In general, the methodologies applied in order to develop a heat map rely on a large extent on the data and resource availability, and may differ from each other in area extension and the resolution provided (Möller et al., 2018; Skujevska et al., 2016; Wyrwa et al., 2017).

We can identify some characteristics from buildings stocks that should be covered if we aim to understand their energy performance. Almost no research has been done on the energy use of non-residential buildings or their spatial distributions in urban areas. Moreover, most energy maps are based on calculated energy intensities from engineering-based models, instead of using real measured data. Although, these models can give technical estimations, the effect from human behaviours in energy consumption cannot be represented. Finally, the methodologies proposed in the literature focus on the creation of either energy maps or detailed stock and energy studies, but they do not combine both procedures. In this article, we explain the generic methodology created having in mind all the gaps presented above. Also, a case study of the municipality of Trondheim, in Central Norway, is presented. The main research questions are:

i) What characterises the energy intensities of different types of non-residential buildings in Trondheim? To what extent do they vary across age cohorts, and what factors explain these variations?
ii) Is there a significant difference between calculated and measured energy intensities, and if so, why?
iii) How is the current energy use of Trondheim’s building stock spatially distributed, regarding electricity and heat for different types and age cohorts of non-residential buildings?

iv) What is the appropriate methodology in order to calculate and visualise the above energy use aspects for urban non-residential building stock?

v) Why should energy maps become valuable tools when analysing energy systems and in decision making?

2. Materials and methods

2.1. Methodology framework

A conceptual outline of the GIS model and its application for spatial energy analyses is presented in Figure 1. The main idea behind the model is to investigate the current energy consumption and distribution of building stocks. For this purpose, the proposed model combines a GIS model, to include the spatial dimension, with data analysis procedures that provide insights into the consumption patterns. The main model inputs are energy databases and a georeferenced building stock database. The last deals with the identification of buildings, their location, typology and age as well as technical properties. The rest provide the measured energy consumption of individual buildings or units for different energy carriers. Hereinafter, we will refer to units as the equivalent of dwellings for non-residential buildings. All databases are processed, cleaned and merged to analyse the individual energy consumption from a certain sample from building stock. These individual values are merged in a building stock-energy database so that we can proceed to analyse the data and characterise the energy intensities. Finally, once these processes are completed, we can develop comparative analyses with external engineering-based databases and a visualisation tool to provide the desired energy maps.

2.1.1. Tools

The tools used for the study are ArcGIS Pro and Python 3.0. The former is a professional desktop GIS software that allows the creation of projects and the generation, exploration, edition and sharing of maps and georeferenced data. Python 3.0 is a programming language that contains packages for data processing and analysis. ArcGIS Pro has an application programming interface (API) which allows using Python libraries to perform spatial visualisations, analyses, data management and GIS system administration tasks just by the use scripts. Therefore, the models developed in Python can be directly connected with ArcGIS Pro interface. In this way, results obtained from the models written in Python can be directly exported to a georeferenced dataset and manipulated for the creation of maps and visualisations.

2.1.2. Data collection, preparation and merging

Data collection and preparation are critical steps as they set the limitations and the scope of the analysis. Depending on the case study, different databases can be retrieved from energy suppliers, government organisms or can be found open-sourced. The selected energy carriers would also depend on the case study as they differ from place to place.

When preparing the databases, our final goal is to find a feature that may allow accomplishing a correct merge between them. The first step is to format the georeferenced and non-georeferenced databases in a way we can efficiently work with them and extract the results. The databases may be provided in different file formats and with information that is not relevant for the scope of the study and with some data errors. Consequently, it is necessary to clean and standardise them before performing any analysis. Due to the size of the databases and to be able to have an automatized system, we might be interested in customised data cleaning models for each dataset. These models may be able to connect the energy databases with the georeferenced building stock dataset through different key features, such as building address, building ID or other types of individual building identification codes. As a result, we link the energy records, from different energy carriers, with their respective buildings and allow a way to proceed with the data analysis and electricity characterisation. The analysis can be performed by individual buildings as well as by smaller components such as dwellings and units. It will depend on the data availability and structure and the goal of the study.

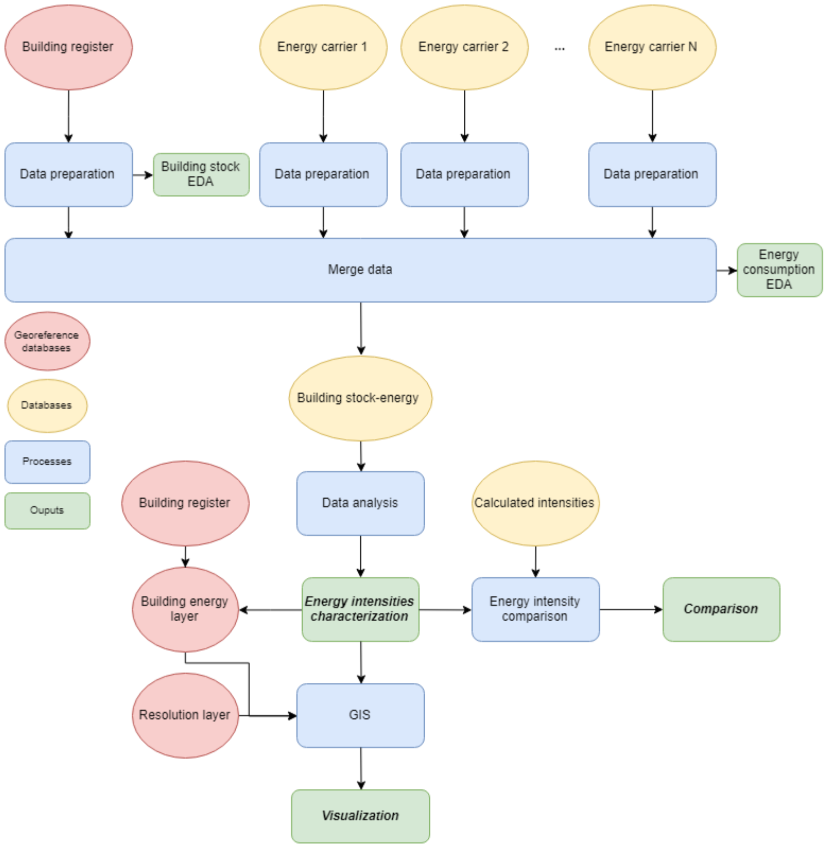


Figure 1. Logical structure of the methodology followed with its four main components: georeferenced databases (in red), non-georeferenced databases (in yellow), processes (in blue) and outputs (in green). The final outputs that provide answers to the research questions are in bold and italic. N refers to the number of energy carriers selected for the analysis.

2.1.3. Assumptions for the building stock

The building stock georeferenced database is an essential source of information in our methodology as it provides the identification and technical characteristics as well as the spatial attributes of each building of the studied area. The main features to be used for the energy analysis are the building typology, the construction year and the total useful floor area. The building register databases usually provide a large number of building categories that may hinder the analysis and comparisons. A solution is to form bigger groups based on general purposes or physical properties. Furthermore, we can assume that some building subcategories are not noteworthy for the energy analysis for not having regular consumption patterns; thus, we propose to exclude them. This can be the case of categories such as garages, animal housing or warehouses, among others.

Furthermore, the year of construction helps to categorise each building according to the year the building was built or the technical code in force at the moment of construction. The building stock is, in this way, divided in age cohorts. In combination with building typologies, we can calculate the energy intensities for each building typology and cohort. Depending on the databases, we might take other specific and customised assumptions to categorise and divide the building stock.

2.1.4. Energy intensities calculation and data analysis

The merged database, henceforth called building stock-energy database, includes the energy intensities for each building record that contains yearly energy consumption. The number of building records with this information varies from study to study as would depend on the quality of the previous steps. We calculate the individual energy intensities of buildings, *b,* using Equation 1 where E is the energy consumption for a particular energy carrier, *ec*, in kWh/m^2, and A is the total useful floor area of the building for the building category purpose in m^2.

$$EI_{b,ec} = \frac{E_{ec,b}}{A_b} \tag{1}$$

Afterwards, we group the resulted energy intensities for individual buildings by building typologies (*t*) and cohorts (*c*) in order to characterise them. We calculate the energy intensity of types and cohorts with an average and dispersion values. When choosing these averages, we should first look at the energy intensities distributions. These can be skewed or non-skewed. In the case of non-skewed distributions, the energy intensity for each group would be the mean energy intensity contained in that group (see Equation 2).

$$EI_{c,t,ec} = \frac{\sum_{b=1}^{n} EI_{b,c,t,ec}}{n} \pm \sigma \tag{2}$$

For skewed distributions, the best approximation for the central tendency is the median or second quartile (Q_2). Using the median will avoid the influence of outliers and skewed data. The range can be defined using the first quartile (Q_1) and the third quartile (Q_3) as in Equation 3.

$$EI_{c,t,ec} = Q_2\,(Q_1, Q_3)$$

where $Q_1 = X_{n+1/4}$, $Q_2 = X_{n+1/2}$ and $Q_1 = X_{3(n+1)/4}$ (3)

being X $=\{EI_{1,c,t,ec}, EI_{2,c,t,ec}, \dots, EI_{b,c,t,ec}\}$ an ordered series of data

Finally, we can proceed to the calculation of energy densities after connecting every building with their corresponding energy intensity given their type and cohort. Equation 4 describes the energy density formula used for comparing regions. As regions do not have the same size, the energy densities (kWh/m^2) are normalized by the aggregated total useful floor area (*A*). Therefore, the energy density reflects the mean energy intensity of a region. This assumption is taken because, with some resolutions, such as districts or political distributions, their land area can differ considerably between them, so comparisons may lead to misunderstandings.

$$ED_{r,ec} = \frac{\sum_{b=1}^{n} EI_{b,r,ec}\, A_{b,r}}{\sum_{b=1}^{n} A_{b,r}} \quad (4)$$

For the building stock and energy data analysis, we perform exploratory data analyses (EDA). This type of statistical analysis is defined as *a combination of numerical and visualisation techniques that allows us to understand different characteristics of datasets, features and the potential relationships between them* (Fuentes, 2018). We propose to develop at least two EDA, one for a better understanding of the building stock, and another for the building-energy dataset for a better comprehension of the energy system. For more details about EDA theory see Section B5 of the Supplementary Material.

2.1.5. Visualisation of the energy map

An energy map is a tool where energy consumptions are aggregated in different spatial resolutions. In the methodology, we develop a model that enables to automatize the process of aggregating results in the areas of study. One model input is the shapefile that contains the spatial vector data with the areas of interests. The other input layer is the georeferenced building-energy database containing the ranges of energy intensities for each building record. We combined both to calculate the aggregated results, as shown in Figure 2. The aggregation is needed as it is not possible to show individual values due to privacy regulations from the energy suppliers. The possibility of selecting the resolution layer brings flexibility to the model as we can modify the areas of aggregation depending on our interests. The energy map can be then modified, edited and visualised in ArcGIS.

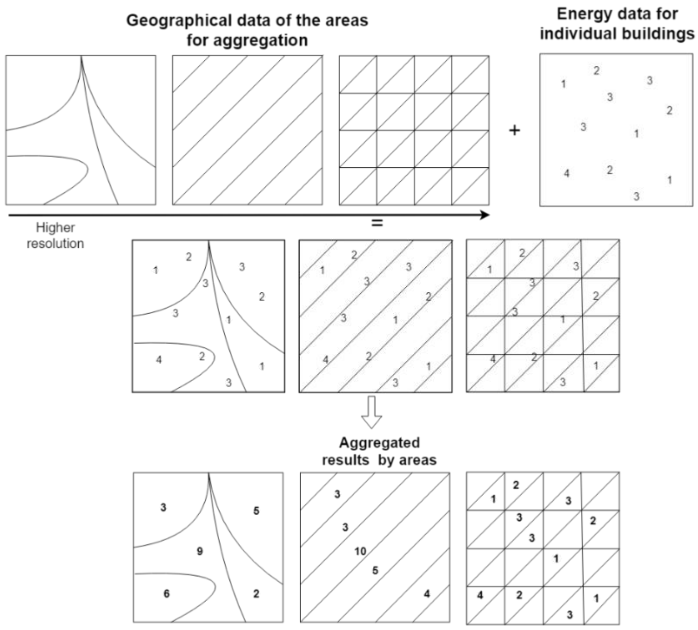


Figure 2. Concept of aggregation and resolutions used for the visualisation of the heat map

2.2. Case study

Trondheim is the third most populated municipality in Norway with around 190 000 inhabitants and an area of around 340 000 km^2. It is located in the region of Central Norway (Midt-Norge) and it is known for being considered as Norway's technological capital. As such, ambitious climate targets have been set to go ahead in the Norwegian's green shift for 2030. Most of the power supply in Trondheim comes from hydropower plants, enabling the municipality to have a de-carbonised building stock. Since 2007, Trondheim is also connected to a district heating plant that covers close to 70% of heating demand (Statkraft Varme, 2018a). Its grid and extension area is presented in Figure 3. Although the power grid capacity is in its majority designed to cover all demand, there are still some areas that may experience some capacity issues with the increase of electric vehicles. As a solution, more distributed energy resources must be locally optimised. Also, various projects have been developed to introduce energy-efficient neighbourhoods to establish zero-emission areas and refurbish the existing building stock (Trondheim Municipality, 2017). The chosen locations for these climate-friendly neighbourhoods are based on energy analyses that consider energy surplus or deficit, bottlenecks and desired RES in each district (Trondheim Municipality, 2017). According to Loveland (2019), spatial and data analyses of the energy consumption of the non-residential buildings of Trondheim would be helpful for the planning of future modifications in the energy system.

When applying the methodology to the case of Trondheim, we analyse two energy carriers: electricity and heat (from district heating). Therefore, three model inputs are used: one georeferenced dataset for the building stock and two energy databases providing measured consumptions. A summary of these databases is given in

Table 1, with data sources, the number of records contained before and after the data preparation process, year and location of the data as well as a brief explanation of the information they contain. A fourth one is used for the comparative analysis. None of these databases is open source. We retrieved the georeferenced dataset from Geodata, a Norwegian firm that provides geographic data for private and public sectors. In the case of the heat and electricity consumption databases, they were provided directly by Statkraft Varme and TronderEnergi under specific confidential contracts.

To allow correct data cleaning and merging, we create two different models for the electricity and the heat consumption databases. Appendix E of the Supplementary Material contains in more detail the two models created in Python. These models are customised for the deployed energy supplier's databases.

The building stock database contains the building typology for each record based on Standards Norway (2013). As mentioned before, the analysis becomes difficult with a larger number of categories. Thus, the categories provided in the dataset are considered sub-typologies, and we group them into building typologies (see Table 3). A more detailed explanation of both excluded and deployed sub-typologies and typologies can be found in Section C1 from the Supplementary Material.

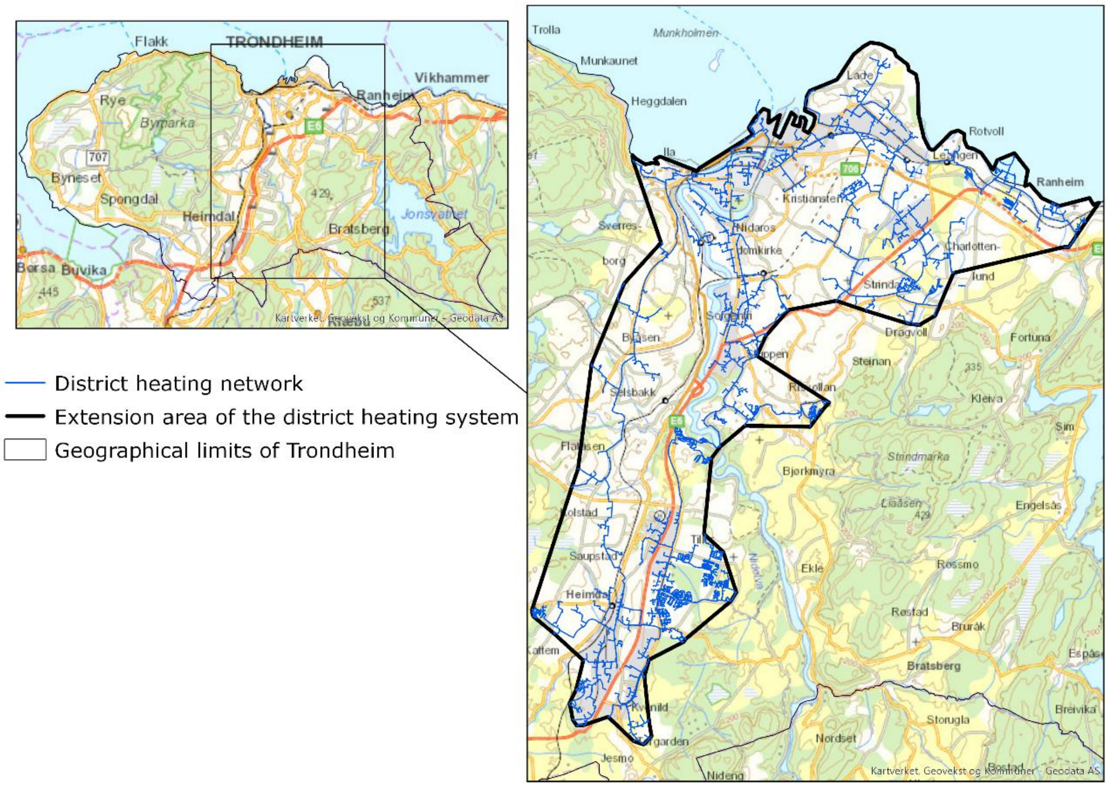


Figure 3. Visualisation of the district heating network and the geographical limits of Trondheim

Table 1. Essential information for the four databases used as inputs for the model.

Database name	Information	Number of records (before and after the cleaning process)	Year(s) / Location	Source
Building stock database	Specific information for each dwelling/unit in Trondheim	Before: 42 099 (for Trondheim) After: 39 224 (for Trondheim)	2018 / Norway	(Geodata, 2018)
Heat consumption database	Measured heat consumption from the building stock covered by the district heating supplier	Before: 2 262 After: 1 090 (628 non-residential)	2018/Norway and Sweden	(Statkraft Varme, 2018b)
Electricity consumption database	Measured electricity consumption	Before: - After: 29 076 (1 370 non-residential)	2013-2018 / Trondheim	(TronderEnergi, 2018)
Calculated energy intensities database	Calculated energy intensities for office, businesses and educational buildings	-	2018 / Oslo	(Sandberg, 2019)

Moreover, the construction year is a missing value for 10% of the records in the georeferenced building stock database. In order to fill these missing values, we assume that the missing construction years are the same as the year from the closest building belonging to the same typology. We use the year of construction to connect each building to their corresponding age cohort. In Norway, the construction codes (TEK) are published by the Building Quality Directorate, and they establish the minimum requirements for a building to be legally built in Norway. As such, we create age cohorts based on TEKs as they will represent buildings with similar technical requirements. Hence, buildings from the same cohort are expected to have similar energy losses and architectural characteristics. Table 2 shows the eight cohorts we defined for the case study and the years they were in force based on Stavset et al. (2015).

Table 2. Technical requirements (TEK) used to categorise the building stock according to the construction year.

TEK	Years
<1949	-1949
TEK49	1950-1968
TEK69	1969-1986
TEK87	1969-1996
TEK97	1997-2006
TEK07	2007-2009
TEK10	2010-2016
TEK17	2017-2018

Once we successfully finish with the merging process (see Section E3 from Supplementary Material), the resulted building stock energy database contains a sample of 1 370 non-residential buildings with their electricity consumption for the year 2018 and 487 with the heat consumption. Table 3 indicates the number of non-residential records associated with their energy consumption. Some are connected to both electricity and heat values, while others only have either electricity or heat. We can distinguish between buildings that only use electricity as the primary energy source in their heating system (EE) and those who are connected to the district heating network and do not use electricity for water or space heating (EH). We also assume that all those buildings that are not linked with heat information are EE buildings.

Table 3. Total useful floor areas and their percentages over the total analysed in the case study for each typology. On the right, the number of records with energy data for non-residential categories in the building stock energy database.

Building type	Area analysed (m^2)	Electricity only records	Heat only records	Electricity and heat records
Business	1 376 227 (88%)	430	107	83
Cultural/Sport	260 966 (67%)	149	58	39
Education	720 283 (69%)	236	113	69
Health	380 081 (77%)	67	31	22
Industry	499 567 (66%)	155	70	51
Office	887 315 (77%)	236	84	64
Service	26 632 (9%)	97	24	19
Total	4 183 018 (73%)	1370	487	347

When doing the energy intensity characterisation, we assume that buildings out of the range of 40-500 kWh/m^2 of electricity intensities are outliers, so they are excluded from the analysis. We established this range based on the calculated energy intensity database (Sandberg, 2019) and other similar studies (Choudhary, 2012). As they are right-skewed (see Figure 11 and Figure 14 from Section B7 in the Supplementary Material), the best approximation for the central tendency of the electricity and heat intensities. Therefore, we will use Equation 3 to characterise the energy intensities of the typologies and cohorts. These ranges are going to be the input energy intensities of the energy maps. Finally, we perform a comparative analysis of our results with those obtained from the engineering model performed in Sandberg (2019). This last model only contains results for office, business and educational buildings; thus, these are the only three categories to be compared.

3. Results

3.1. Building stock analysis

Dwellings in Trondheim conform 92% of the buildings in Trondheim, so only 8% are for non-residential purposes. However, non-residential building typologies covers 34% of the total useful floor area in the municipality. Regarding the spatial dimension, non-residential buildings tend to be concentrated in specific areas while residential buildings are more spread in the municipality (see Figure 4). We can see a higher density of non-residential buildings in the northern part of the municipality where the city centre is situated and where there is a large concentration of offices and business units. To see more detailed maps of the dispersion of the building stock in Trondheim see Section D1 from the Supplementary Material.

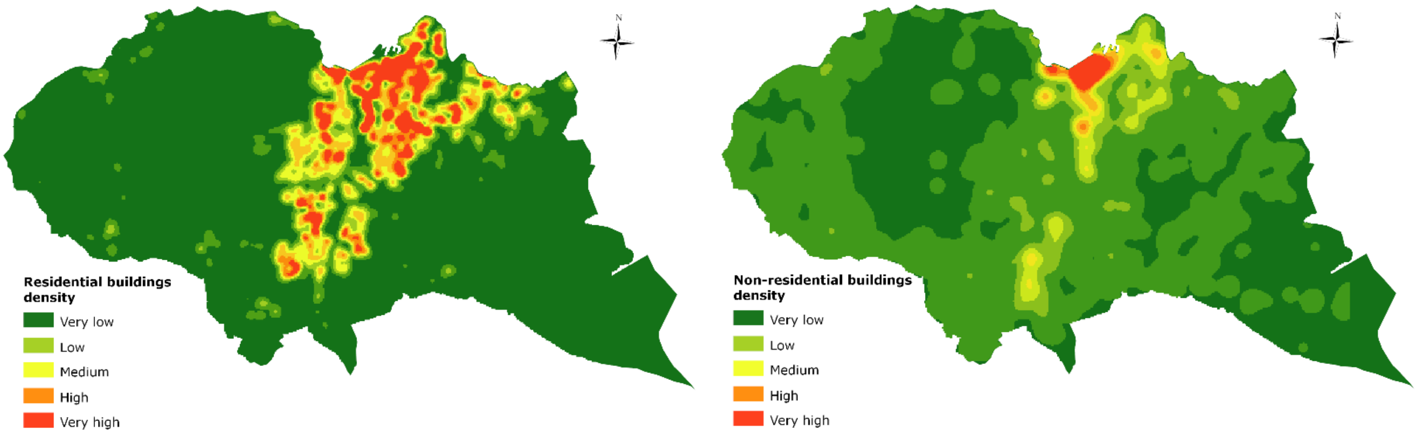


Figure 4. Kernel building density maps for residential (left) and non-residential buildings (right).

Around 20% of the non-residential buildings in Trondheim were constructed before 1949. The building stock in Trondheim is relatively new compared with those in other European cities. Moreover, between 1969 and 2007, almost 50% of the current non-residential buildings were constructed. Figure 5 shows how during the time in which TEK07 was in force, there was a dramatic reduction in the construction activity. Figure 5 also shows the evolution of each building category. Most of the business units are in buildings constructed before 1949. The explanation behind this is that they are located in the city centre where buildings are older than in other areas of the municipality. The same occurs with office and service units. Industries and health buildings were mainly constructed between 1969 and 1997 while education and emergency are the typologies with larger share of buildings constructed with newer technical codes.

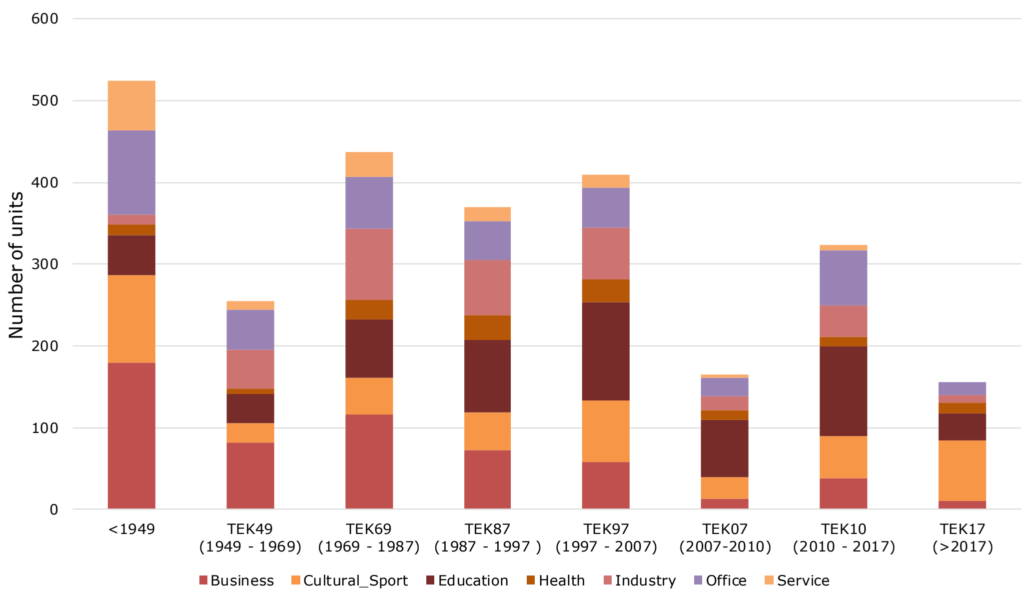
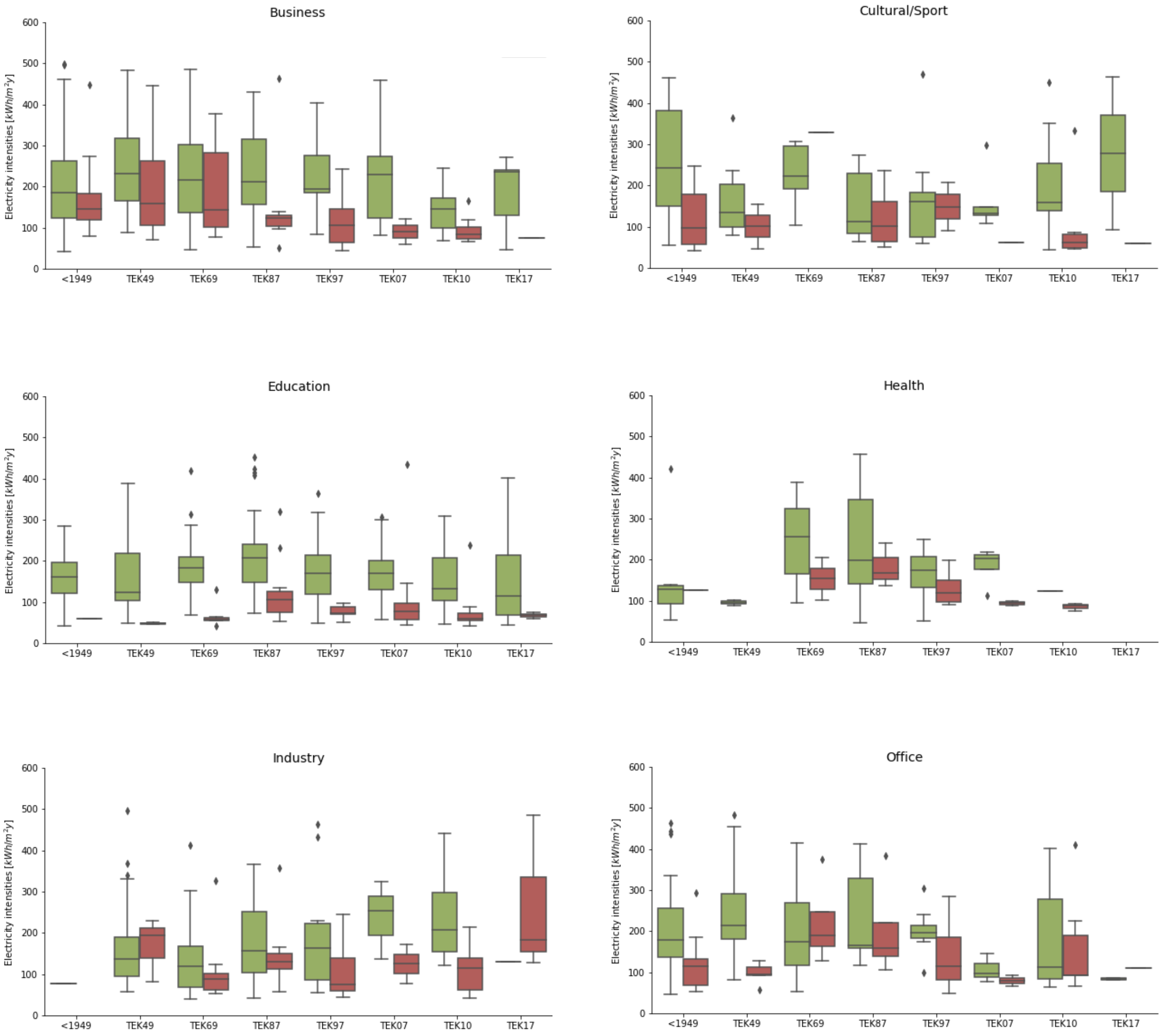


Figure 5. Number of non-residential units in Trondheim by age cohort and typology.

3.2. *Electricity and heat intensities characterisat*ion

The ranges obtained for the electricity intensities for each typology and age cohort are shown in Figure 6. Although we do not find reduction patterns due to stricter building codes, there is a general pattern in which old (*<1949* and *TEK49*), and new buildings (*TEK10* and *TEK17*) have lower electricity intensities than cohorts between them. Comparing the median intensities for each category, we identify that business, offices and service buildings have higher electricity intensities than the rest of the categories. Results from service buildings do not follow the same patterns as the rest of typologies, but this can be influenced by the small number of samples analysed (See Table 3)



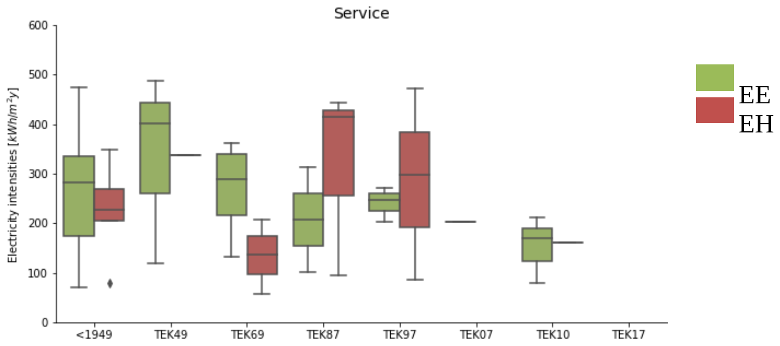


Figure 6. Electricity intensity ranges by typology and age cohort. The EE (green) - buildings with electricity for heat, EH (red) - buildings with district heating for heat.

Furthermore, as expected, buildings connected to the district heating system have lower electricity intensities. To analyse the degree of reduction, we calculated the median electricity intensity for each typology. The absence of a correlation between construction year and electricity intensities makes reliable the use of a unique value of electricity intensity for each building typology, without distinction of cohorts. Looking at Figure 7, we can distinguish the reduction of electricity consumption when connecting the buildings to the district heating network. The degree of decrease varies from typology to typology. Education and cultural/sports buildings are the typologies with a higher percentage of reduction (almost 60%) while for service and industry buildings they barely reach 20% of improvement. We assume that the percentage of reduction between electricity use from EE and EH corresponds to the potential electricity intensity saves we reach when connecting a building to the district heating network.

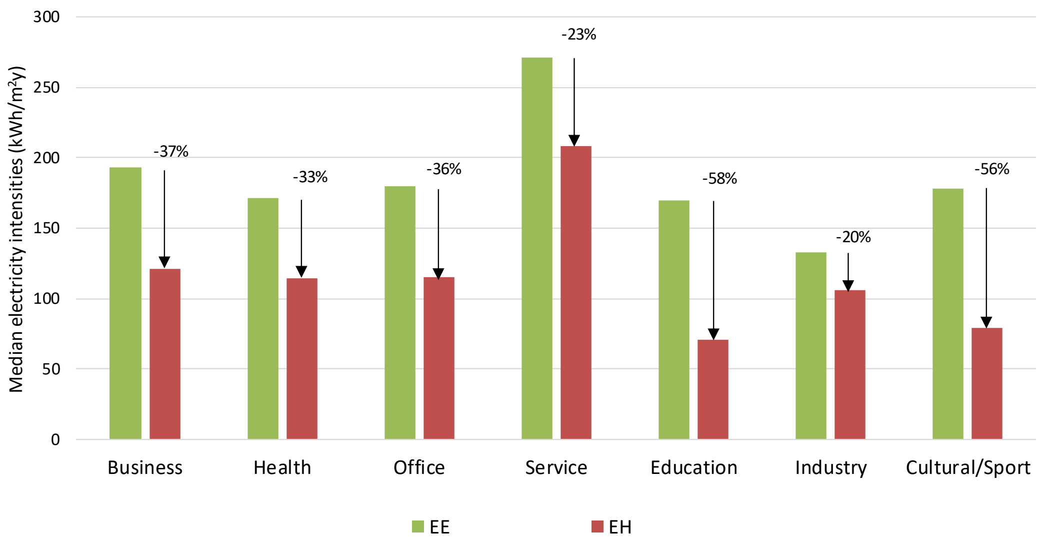
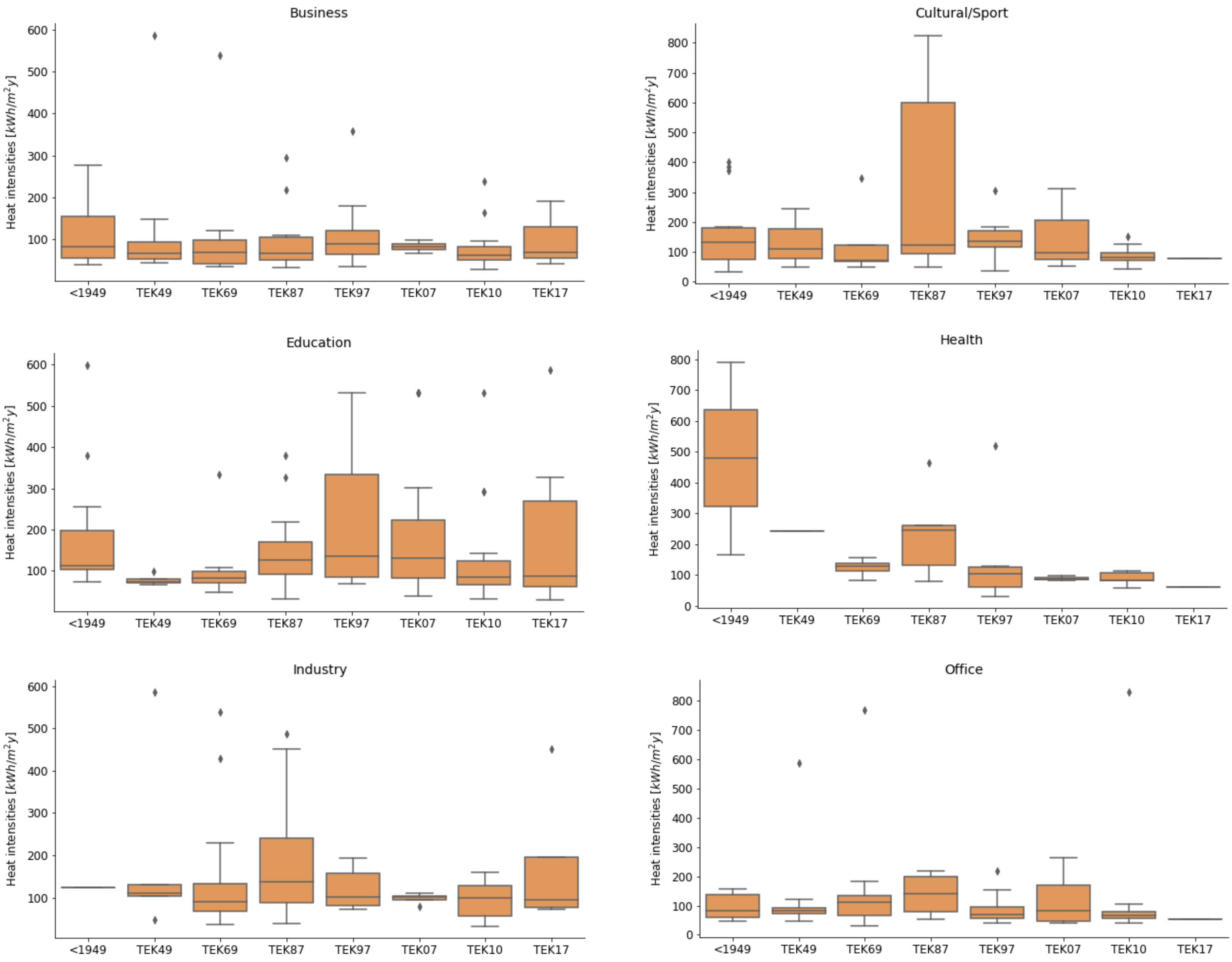


Figure 7. Comparison of electricity intensities between buildings connected to the district heating network (red) and those which are only supplied with electricity (green)

For heat intensities, we obtain similar results. Figure 8 shows the ranges by typology and age cohorts. In general, we can see lower dispersions than for electricity inside each group. However, cultural buildings from cohort TEK89 have the broadest range of heat intensities. Looking into this group in more detail, we find that sports halls are the principal responsible for it as they have high heat demand. The heat intensities for most of the cohorts in every building typology do not tend to surpass the 300 kWh/m^2, excluding education and health units constructed before 1945 and the sports halls mentioned before. Except for some specific cases, new and old buildings are the cohorts with lower heat intensities in every building typology.

The electricity and heat intensities presented in this section are used for the development of the visualisation tool as well as for the comparison with calculated energy intensities. Sections C2 and C3 from the Supplementary Material contain the tables with the exact values used.



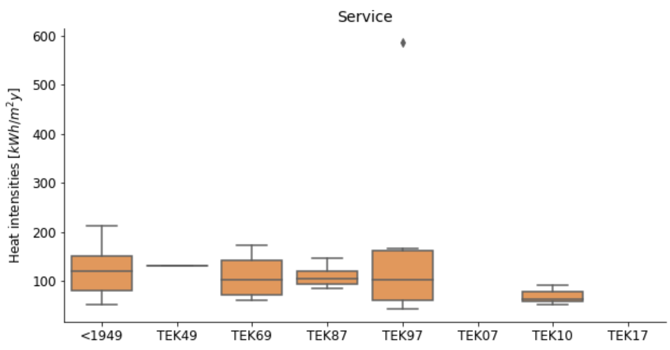


Figure 8. Heat intensity ranges by typology and age cohort

3.3. Measured vs calculated energy intensities

Figure 9 aims to compare the results obtained from measured energy intensities with those calculated from Sandberg (2019). We only make the comparison for educational, office and business buildings. The 45º line in black indicates the situation where calculated and measured intensities are equal. If values fall above this line, theoretical estimations are overestimated compared with measured intensities. On the contrary, values situated bellow indicate overestimation.

When analysing the electricity intensities from the three typologies, we see that in most of the cases theoretical estimations are underestimated. We must highlight the considerable distance of the median values to the 45º line for buildings belonging to older (*<1949, TEK49*) and newer (TEK17) cohorts. Moreover, we can also appreciate that measured intensities and calculated estimations are pretty the same for buildings belonging to the cohorts TEK69, TEK87 and TEK97. Finally, while calculations consider buildings from TEK17 as those with lower average electricity intensities, measured results show that they may be one of the cohorts with higher electricity intensities.

On the other hand, heat intensities have not such explicit patterns. For business buildings, there is a tendency to overestimate the heat intensities. However, for educational and office buildings, calculated intensities are underestimated for most cohorts. As with electricity, heat intensities for TEK69 and TEK97 result in being almost equal in both studies. However, calculated heat intensities are dramatically overestimated for buildings built before 1945. Contrary to the case of electricity, both studies agree upon new buildings being one of the cohorts with lower heat intensities.

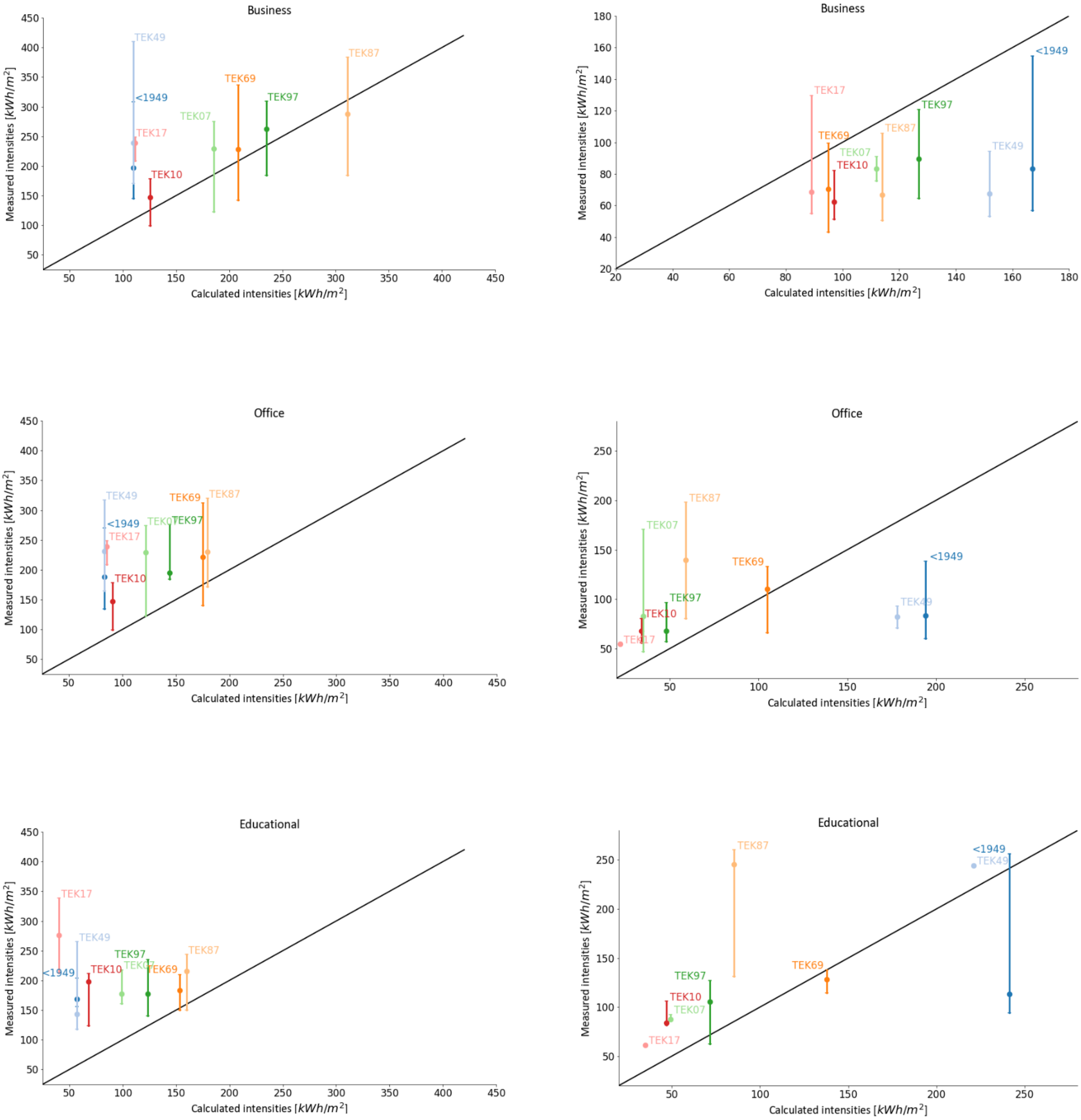


Figure 9. Comparison between measured and calculated electricity (left) and heat (right) intensities for business, office and educational buildings. The calculated measures are extracted from Sandberg (2019). Lower points indicate the 1st quartile, middle dots the 2nd quartile and higher points the 3rd quartile.

3.4. *Energy maps*

The last step of the methodology covers the visualisation of energy maps. On the left side of Figure 10, the distribution of heat and energy demand in the municipality are shown per district zones. It is interesting to see how regions with high and low energy demand are almost the same for both energy carriers. The northern part, corresponding to the city centre of Trondheim, has higher energy demand. Also, the southern part has considerably higher consumptions than the rest of districts which in most cases do not surpass 4.5 GWh for heat and 7 GWh for electricity. The district with the larger heat consumption is located in the area of Midtbyen. The large concentration of office buildings and cultural/sports units constructed during the 90s are the main responsible for this high heat consumption

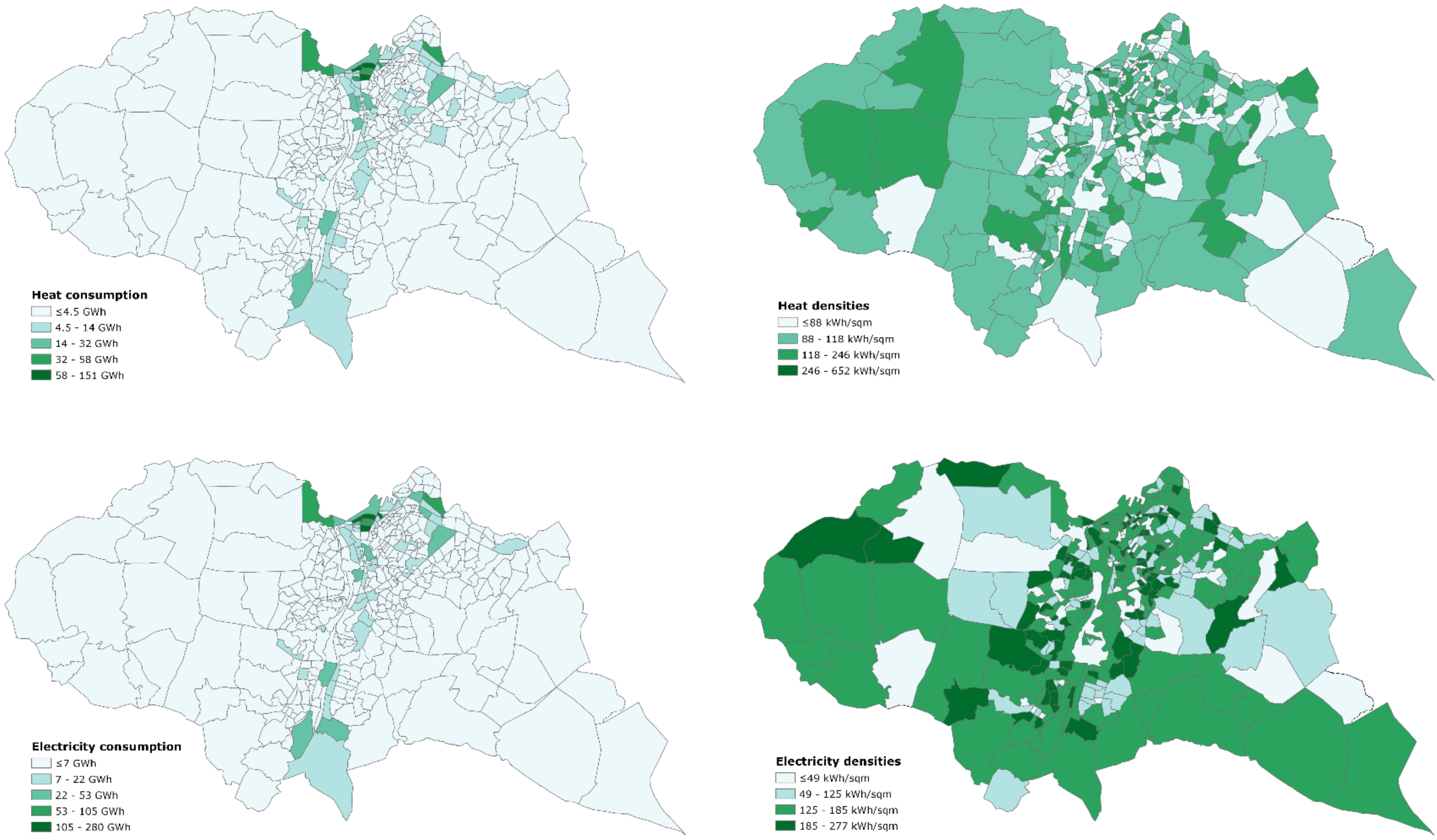


Figure 10. On the left, heat and electricity consumptions per district in GWh. On the right, heat and electricity consumption per total useful floor areas per district.

On the right side of Figure 10, densities for both energy carriers are given. In this case, there is no similarity between energy carriers. We should remind that in this study the energy density indicates the mean energy intensity of the region. We can identify districts far away from the district heating areas with high electricity densities. The building typology mix in these districts is compound by high energy demanding building archetypes. However, as most of them are very low populated, in aggregation (maps on the left) do not reach the levels of consumption of smaller areas that are densely populated.

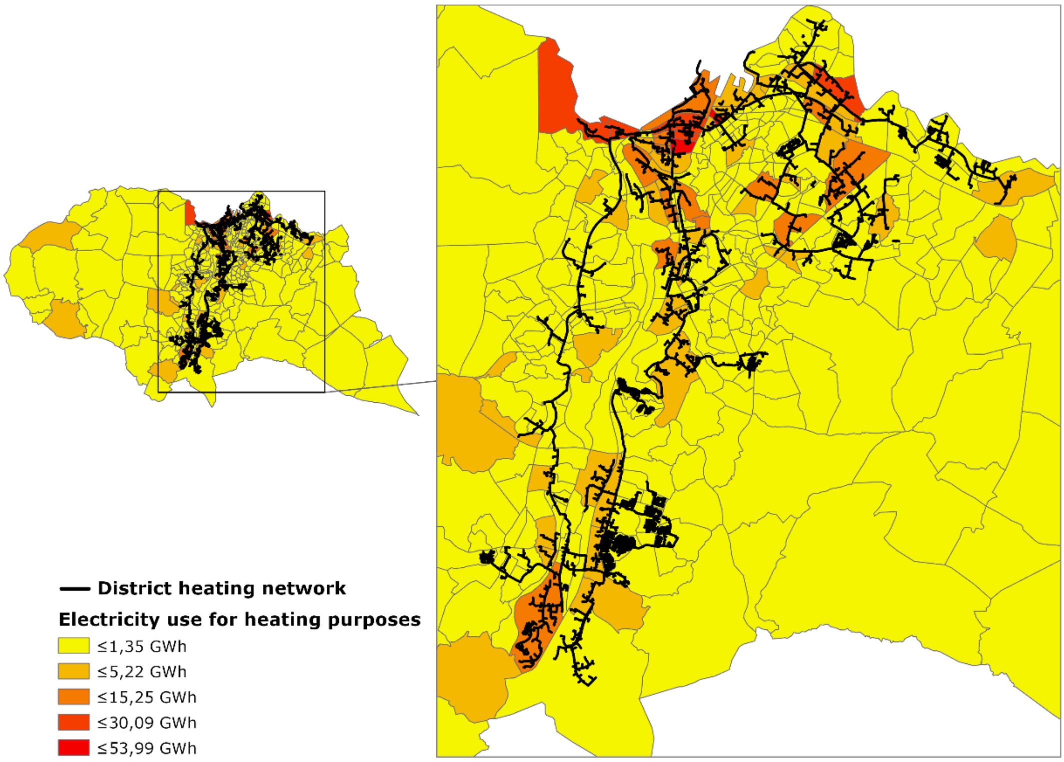


Figure 11. Electricity used for heating purposes in non-residential buildings per square meter of constructed buildings. The resolution is district areas.

Figure 11 shows the the aggregated electricity use for heating purposes. The figure can help to visualise the regions where there is more potential for district heating to substitute the use of electricity. Most of the areas are currently supplied by district heating, but there are still some potential districts without supply.

The visualisation tool allows developing further and more detailed analysis with different resolutions and information. The Supplementary Video allows to contextualize the possible results we can obtain from it.

4. Discussion

The energy intensities of non-residential buildings in Trondheim depends more on their purpose than on their construction year. No linear correlation was found between electricity and heat consumption and construction year. However, significant variations inside the same typology were found from building to building, what is in line to the findings of Søgnen (2002). Besides, we discovered a similar pattern in every building typology and energy carrier: old and new buildings tend to have lower consumptions than buildings constructed from 1969 to 2007.

One explanation for this pattern is the effect on refurbishment of old buildings. Refurbishments of most old buildings have led to a considerable reduction of their energy needs, reaching similar consumption levels as new buildings. In contrast, refurbishments of units belonging to other age cohorts have not taken place yet in large scale. Large deviations in energy consumption in these cohorts are caused by the significant differences between refurbished and non-refurbished buildings.

Furthermore, by performing the comparative analysis we identify important underestimations and overestimations of technical calculations, mainly, in buildings from <1949 and TEK17 cohorts. In electricity, these two age cohorts tend to be underestimated, meaning that real electricity consumption is much higher than expected from the technical requirements. This is especially unusual for new buildings where technical specifications are much stronger. When the real energy use of buildings constructed with strict technical codes are higher than the calculated estimations, it is known as 'rebound effect' (Hertwich, 2005; Sandberg et al., 2017; Sunikka-Blank et al., 2012). In contrast, the resulted real heat intensities of old buildings are lower than technical estimations. This phenomenon is known as "prebound effect", and it can be explained as a result of heating less share of the buildings than what it is assumed in the engineering models (Sandberg et al., 2017). For the rest of the cohorts, technical estimations agree with the results of our model. We highlight how technical models tend to provide energy intensities with significant differences between building cohorts. However, our study shows that the degree of these differences is not very significant. Further research on refurbishments and the prebound-rebound effects might clarify their role-playing in the abovementioned tendencies for old and new non-residential buildings.

Additionally, the methodology promotes the examination of each typology in more detail. Business and office units have the highest median electricity and heat intensities from all the non-residential categories. Their requirement of large numbers of energy demanding appliances and lighting (e.g. computers, freezers, special machinery) concentrated in relatively small areas leads to this high energy intensities. On the contrary, industries are not such energy demanding when normalised by used area due to having large areas employed for warehousing or other non-energy related purposes. In the case of Trondheim, educational buildings have significant energy intensities. The large portion of areas dedicated to higher education institutions in Trondheim explains the energy relevance of educational buildings. This sub-typology holds equipment in laboratories that is as energy demanding as industrial machinery. Therefore, energy intensities of educational buildings might vary considerably in other municipalities with lower or no used area for these purposes. Moreover, cultural/sports and service units seem to have very dispersed electricity and heat intensities. An explanation is the lack of enough samples, so outliers have great influence on calculations.

Analysing the resulting energy maps, the areas with high electricity and heat aggregated consumption are within the district heating extension area. Moreover, more concentration of building units leads to higher aggregated energy consumptions. That explains why large districts may have lower electricity and heat consumption than smaller districts. More expansion of the district heating network within these regions may be a good measure to decrease their electricity consumption. We also identified that regions far away from the district heating concession areas tend to have high energy demand but lower aggregated consumptions. This means not only that district heating extensions are not viable but also it would not be that effective. Alternative measures such as the deployment of VRES or promotion of refurbishment plans might be better solutions in these areas.

Moreover, the building typology mix of an area has considerable influence on the aggregated energy consumption. As mentioned in Section 3.4, those districts situated in the northern part of the municipality are the most energy demanding. This is especially boosted by the large concentrations of business, office and educational buildings, typologies with energy intensities above the rest of the categories. On the contrary, the Southern part is less populated by non-residential units, being most of them used for industrial purposes. However, the large areas that characterise the industry typology conduct to high energy demand.

Moreover, when deploying and creating energy maps, it is of vital importance to understand the purpose of the analysis. For instance, if we are more interested in the extension of district heating, aggregated consumption or energy per square meter of land give better insights of the most interesting areas, as we will like to identify high concentrations of energy in the smaller area. However, if we want to develop refurbishment plans, we might be more interested in the reduction of energy intensities of buildings, therefore the energy density map will help us to identify the areas with the more energy demanding buildings. Also, resolutions are of vital importance. When choosing administrative areas as the regions for aggregations, we have the advantage of better identification and more straightforward analysis of the results. However, they can also bring issues as the land areas are not equal among them. When we are comparing or analysing the results, we have to keep this in mind and consider it so that we do not end in false conclusions. On the contrary, the use of grids might make the visualisation of maps less intuitive and manageable; however, as land areas are equal, we do not have to worry about region land areas.

We have developed a methodology that builds a bridge between GIS and two types of databases to perform more complete energy analyses of the building stock. The methodology presented in this paper is used to analyse the non-residential buildings of Trondheim. The methodology, however, is generic and could easily be applied to other municipalities, regions or countries. Also, it can be extended to residential buildings when energy data is available. Moreover, the bottom-up approach, in which individual buildings are associated with their energy intensities, allows performing the analysis by aggregated consumptions in different resolutions or by individual buildings. This last analysis would be possible to be done if databases are standardised and energy suppliers have more synergies in data collection processes. In this way, data cleaning and merging processes would be facilitated, and individual buildings can easily be connected to their real energy consumptions, without losing too many records as happens in our case study. One drawback of not being able to analyse by individual buildings is the different uncertainty in energy intensities calculations due to differences in the number of records for each typology. Another adjustment to be made in further work is to add weather adjustment procedures that allow the inclusion of databases from different years and that can lead to improved calculations of the energy intensities.

We have demonstrated that energy maps can give very insightful information about the distribution of energy intensities and the actual situation of the energy system. The methodology has also guided us to demonstrate the lack of importance of the construction year when talking about non-residential buildings. A fact that should be considered in engineering models. Moreover, the development of the energy maps set the basis for further studies in which building stock energy consumption is of relevance. For instance, it is possible to perform more technical or social studies by combining energy maps with other layers (i.e. emissions, demographic, solar radiation, transport dynamics).

The development of automating models for the merging processes would allow updating the energy consumptions continuously and gives the possibility of adding new future energy databases. However, the creation of these models can be quite time-consuming and has the disadvantage of being customised for each energy database. Also, it does not remove the need for some manual cleaning and manipulation. As presented above, the standardisation of data collection could improve the cleaning and merging models and could increase the efficiency of their creation.

Finally, the model requires to have an available building stock georeferenced database. This is possible in Norway, a country with a high level of data availability; however, other places might not have enough information recorded. Also, energy databases by individual customers are not publicly available, so data collection takes time as direct contact with energy suppliers is required.

5. Conclusion

Energy consumption in building stocks have been extensively analysed in studies with aggregated and non-spatial approaches. However, GIS has arisen as a new tool which can help to increase the accuracy as well as the resolution of the models. The methodology presented above combines it with data analysis procedures in order to provide further insights that can be done before and after plotting energy maps. In this way, we do not only provide spatial analyses of building stocks but also better explanations of our results. Moreover, it has the advantage of connecting data to individual buildings, so that we are able to get access to more detailed information. The flexibility of GIS in the creation of new tools and its link with programming languages make possible the incorporation of spatial resolution in building stock energy models.

Energy systems are changing due to environmental issues so political measures are emerging and the deployment of other VRES or other energy sources are promoted to reach more environmentally friendly energy mixes. Policymakers and energy suppliers are aware of the importance behind a more complete and holistic understanding of energy consumption in buildings, and new plans are currently under development to increase their efficiency and reach energy and climate goals. Energy maps and their interaction with other geospatial layers will help energy-related actors to develop smart urban plans and localise VRES. Our methodology is developed to promote its adoption by different actors and to provide a tool for energy and urban planning projects. Also, the model has the advantage that can be applied to a large number of regions. It also aims to increase the awareness of the importance of synergies between energy actors, as it is vital for the understanding of energy systems and the development of smart solutions. This is especially needed for data collection, as it can facilitate efficient and alternative use of their data. To conclude, the introduction of their data in similar models can bring them opportunities such as identification of bottlenecks, identification of best locations for RES and a better understanding of energy densities to extend efficiently and profitably their facilities.

6. References

Choudhary, R. (2012). Energy analysis of the non-domestic building stock of Greater London. *Building and Environment, 51*, 243-254. doi:https://doi.org/10.1016/j.buildenv.2011.10.006

European Commission. (2018). Buildings. Retrieved from https://ec.europa.eu/energy/en/topics/energy-efficiency/buildings

European Environment Agency. (2019). Primary energy consumption by fuel. Retrieved from https://www.eea.europa.eu/data-and-maps/indicators/primary-energy-consumption-by-fuel-6/assessment-2

Fuentes, A. (2018). *Hands-On Predicitive Analytics with Python*: Packt Publishing.

Geodata. (2018). *Eiendomsdata (Building stock database).*

Hertwich, E. G. (2005). Consumption and the rebound effect: An industrial ecology perspective. *Journal of Industrial Ecology, 9*(1-2), 85-98. doi:10.1162/1088198054084635

IEA. (2017). *Norway. 2017 Review*. Retrieved from https://www.iea.org/publications/freepublications/publication/EnergyPoliciesofIEACountriesNorway2017.pdf

IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Retrieved from Geneva, Switzerland:

Li, C. (2017). GIS for Urban Energy Analysis. In *Reference Module in Earth Systems and Environmental Sciences*.

Loveland, S.-J. (2019, May 2019) *Interview meeting with Trondheim Kommune/Interviewer: R. Alonso, N. Sandberg, & H. Brattebø.*

Mastrucci, A., Baume, O., Stazi, F., & Leopold, U. (2014). Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam. *Energy & Buildings, 75*(C), 358-367. doi:10.1016/j.enbuild.2014.02.032

Möller, B. (2008). A heat atlas for demand and supply management in Denmark. *Management of Environmental Quality: An International Journal, 19*(4), 467-479. doi:10.1108/14777830810878650

Möller, B., Wiechers, E., Persson, U., Grundahl, L., & Connolly, D. (2018). Heat Roadmap Europe: Identifying local heat demand and supply areas with a European thermal atlas. *Energy, 158*, 281-292. doi:10.1016/j.energy.2018.06.025

Resch, B., Sagl, G., Törnros, T., Bachmaier, A., Eggers, J.-B., Herkel, S., . . . Gündra, H. (2014). GIS-Based Planning and Modeling for Renewable Energy: Challenges and Future Research Avenues. *ISPRS International Journal of Geo-Information, 3*(2), 662-692. doi:10.3390/ijgi3020662

Sandberg. (2019). Paper under preparation.

Sandberg, Sartori, Vestrum, & Brattebø. (2016). Explaining the historical energy use in dwelling stocks with a segmented dynamic model: Case study of Norway 1960–2015. *Energy and Buildings, 132*, 141-153. doi:https://doi.org/10.1016/j.enbuild.2016.05.099

Sandberg, Sartori, I., Vestrum, M. I., & Brattebø, H. (2017). Using a segmented dynamic dwelling stock model for scenario analysis of future energy demand: The dwelling stock of Norway 2016–2050. doi:https://doi.org/10.1016/j.enbuild.2017.04.016

Sartori, I., Wachenfeldt, B. J., & Hestnes, A. G. (2009). Energy demand in the Norwegian building stock: Scenarios on potential reduction. *Energy Policy, 37*(5), 1614-1627. doi:https://doi.org/10.1016/j.enpol.2008.12.031

Skujevska, A., Rosa, M., & Kamenders, A. (2016). Evaluation of Energy Consumption of Municipal Buildings by Heat Energy Demand Mapping. *Energy Procedia, 95*, 444-450. doi:https://doi.org/10.1016/j.egypro.2016.09.061

Søgnen, O.-G. (2002). Modellbyggprosjektet- Måling av formålsdelt energibruk i 26 bygninger. In E. r. n. 1-02 (Ed.): Enova byggoperatør.

Standards Norway. (2013). Klassifikasjon av byggverk. Del 3: Bygningstyper. In (Vol. NS 3457-3:2013). Norway.

Statkraft Varme. (2018a). District heating. Trondheim. Retrieved from https://www.statkraftvarme.no/om-statkraftvarme/vare-anlegg/norge/trondheim/

Statkraft Varme. (2018b). *Statkraft Varme's private server. Heat delivered to users (Heating consumption database).*

Stavset, O., & Kauko, H. (2015). *Energy use in non-residential buildings - possibilities for smart energy solutions*. Retrieved from Norway: https://www.sintef.no/globalassets/sintef-energi/interact/tr-a7510-energy-use-in-non-residential-buildings--possibilities-for-smart-energy-solutions-2015.pdf

Sunikka-Blank, M., & Galvin, R. (2012). Introducing the prebound effect: The gap between performance and actual energy consumption. *Building Research and Information, 40*(3), 260-273. doi:10.1080/09613218.2012.690952

TronderEnergi. (2018). *Electricity consumption of TronderEnergi's users (Electricity database)*.

Trondheim Municipality. (2017). *Municipality plan: Energy and climate 2017–2030* Retrieved from Trondheim: https://www.trondheim.kommune.no/globalassets/10-bilder-og-filer/10-byutvikling/miljoenheten/klima-og-energi/kommunedelplan-energi-og-klima-20172030-trondheim-kommune-060618.pdf

Wyrwa, A., & Chen, Y.-k. (2017). Mapping urban heat demand with the use of GIS-based tools. *Energies, 10*(5), 720.

Supplementary Material

Spatial and data analysis of non-residential building stocks. Case study of Trondheim.

Contents

Appendix A. Supplementary data information

Section A1. Building stock georeferenced database

The building stock database was retrieved from Geodata which is a Norwegian firm that provides geographic data for private and public sectors. Specifically, the following database is information from the land registry, which has been connected to their geographical location. The database contains information of all the buildings registered in Norway. However, we filter it on the buildings registered in the municipality of Trondheim.

The database is the source of the characteristics of the buildings analysed. Moreover, it gives the possibility to localise the buildings and their energy consumption, allowing us to analyse the spatial dimension. Therefore, the building database is a crucial resource to be able to gather information and visualise the results of the analysis.

The database contains information such as identification of buildings, location, typology, age as well as physical and technical characteristics. For the analysis, the physical and technical characteristics are not deployed.

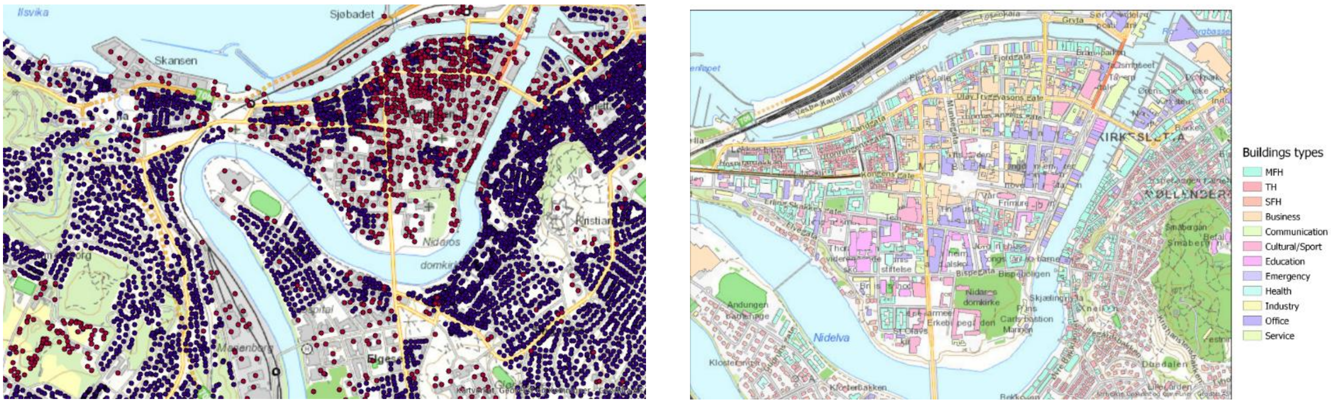


Figure 1. Different visualisations of the building stock georeferenced database in ArcGIS

Metadata

The database contains 54 features and 42 099 records of buildings in Trondheim. The attributes are described as following (Geodata, 2018; Statens Kartverk, 2006, 2014):

- **Byggid and Bygningsnr**: Statistics Norway's numbering of buildings. The number is nationwide, and the municipalities have been allocated to number series. The first 8 digits are serial numbers, and the ninth is a modulus-11 control digit. Its purpose is to be able to uniquely identify the building.
- **Lopenr**: serial number within the floor on which the separately occupied unit lies
- **Bygningstatuskode** and **Bygningstatus:** Information on the building's current status Note: In special instances, a building may get a start-up permit and a certificate of practical completion for a part of the building, and then a start-up permits for another part.
- Naringsgruppekode and Naringsgruppe: Industry groups.
- **Vannforsyningskode** and **Vannforsyning**: Statement of which type of water supply the building has.
- **Avlopskode** and **Avlop**: statement of which type of discharge the building has.

- **Harheis**: statement of whether a building has an elevator.
- **Opprinnelseskode** and **Opprinnelse**: statement of origin and shows how the building has been registered in GAB or the cadastre.
- **Bebygdareal** and **Bebygdarealkilde** and **Fkbareal**: base area of the building, i.e. the area on the property occupied by the building
- **Bruksarealtilbolig**: The floor's total use area for residential purposes Note: In residential buildings, garage is considered a utility area for housing when the garage belongs to a dwelling, also at /on the garage. The area survey follows NS 3940. Exceptions are:
  - The areas shall be stated in square meters.
  - Area for large pipes and channels should not be subtracted.

  At floor level, the utility area comprises all available space. Inspection hatch in an otherwise inaccessible area does not make the area available in this context. Opening for stairs, elevators and the like, and walls between the utility units must be withdrawn. When the plane is a loft, the area is measured out to 0.6m outside the height of 1.9m.
- **Bruksarealtilannet**: Utility area for non-residential use according to NS 3940 with various exceptions Note: Exceptions from NS 3940 are:
  - The areas shall be stated in square meters.
  - Area for large pipes and ducts should not be sustracted.

  At floor level, the area covers all available area. Inspection hatch to an otherwise unavailable area does not make the area available in this context. Opening for stairs, lifts and the like, and walls between utility units must be withdrawn. If the plane is a loft, the area is measured out to 0.6m outside the height of 1.9m.
- **Bruksarealtotalt**: Total utility area is the sum of utility area for residential and utility space for others.
- **Alternativtareal**: Alternative area of the building, stated in whole square meters summed up for the whole building *Note: The municipality is free to enter the area it wants, for example — area for calculating municipal taxes (e.g. rent area).*
- **Alternativtareal2**: The field is at the disposal of the municipalities Note: Coding of the area to be withdrawn or added to the home tax, used with AlternativeAreal.
- **Fylkeid**: County ID
- **Omradeid**: Area over which an object extends.
- **Kommunenr**: Municipality number
- **Verifisert**: Statement of the quality of the coordinates.
- **Ufullstendigareal**: Indicating whether the usable area buksenhetene is fully registered for building (angivelse av om bruksareal på buksenhetene er fullstendig registrert for bygningen) (Incomplete area)
- **Harsefrakminne**: Indicates whether the building is registered in the SEFRAK register. SEFRAK contains buildings built before 1900, possibly in 1945 in parts of Northern Norway.
- **P25statuskode** and **p25statuskode**: Indicates if it was constructed before 1850. (No information)
- **Harkulturminne**: Indicates whether the building has registered cultural monuments from Askeladden or not.
- **Antallboenheter:** Number of residential units on the floor. Calculated automatically by the system for use units of type housing on the floor. A utility unit is included on the floor if it has its entrance in that floor.
- **Bygningstypekode** and **Bygningstype**: Description of what the building is actually used for.

- **Bygningsendringkode** and **Bygningsendring**: Codes for how building change is in the constructional context of the main building.
- **Utenbebygdareal**: Flag indicating that the building does not have the built-up area. When this is set, it is correct that the built-up area is not filled. It is not defective registration.
- **Forstedato** and **Sistedato**: Identical to the construction date. Specifies the date of the first registered building case processing.
- **Antallreg**, **Antallferdig, Antalletasjer, Antallhovedetasjer, Antallkjelleretasjer**, **Antallloft** and **Antallunderetasjer**: number registered, number of floors, number of main floors, number of basement floors, number of lofts, number of lower floors. Detailed properties of the building.
- **Kommunenavn** and **Fylkesnavn**: Municipality and county names.
- **Geometrikilde**: Geometry source.
- **Year**: Year of construction. Taken from *forstedato*. In case it is not available we make the assumption that the building was constructed in the same year than their neighbours' buildings from the same type.

Section A2. Heat consumption database

The database for the district heating consumption was retrieved from the Statkraft Varme's web system. This system contains detailed information on every customer in Norway and Sweden. Information for the year 2018 is downloaded as an Excel file (see **¡Error! No se encuentra el origen de la referencia.**) for all customers. It contains data of the district heating consumption (kWh), the volume, and specific information as addresses or names for each customer. It does not contain information for customers living in single-family houses and terraced houses.

The database is the main source of information for heat consumption in Trondheim, primarily of non-residential buildings.

10 features form the raw database. Five of them identifies the user, one indicates the city and the rest aim to indicate the consumption. The database contains the information in a way in which is not possible to connect it with the Geodata database. Therefore, it needed preparation before the merging process which is described in Section B3. Data cleaning and processing – Heat consumption database.

After the pre-processing step, we end up with a database of 1090 records of users from Statkraft Varme, from which 628 are non-residential buildings. The number of features is 14, but some are added reaching 29 in the final version:

- **MålepunktID**: ID registration in Statkraft system
- **Anleggsnavn**: complete name of the address (adressenavn + nr + postnavn + postnr + bokstav + by)
- **Adresse:** address + nr
- **Adressenavn:** address name
- **Nr:** number of the building
- **Postnanv:** postal name
- **Postnr:** postal code
- **Bokstav:** building's letter
- **By:** city
- **Konsern:** customer information

- **Kundenr:** customer information
- **Kundenavn:** customer information
- **Energi (kWh)**: yearly energy consumption**Makseffekt (kW)**: maximum power**Volum (m3):** volume delivered**Special (y/n):** special record with the information given differently (use for cleaning purposes)**New nr:** new number of building given from the building stock database (use when processing the data)
- **Address text:** feature used when processing the data
- **Address id:** address id from the building stock database
- **Sum bruksareal**: it gives the sum of the areas of the dwellings/units with the same address id. It does not mean that it is the complete building
- **Bruksarealtotalt**: it gives the total area of the building that contains the address id.
- **Bygningstype** and **Bygningstypekode**: building type
- **Year:** year of construction

Section A3. Electricity consumption database

TronderEnergi provided a text file with the electricity consumption of individual customers from 2011 to 2019. The database contains electricity consumption in kWh per year and gives some information from the customers as well as their addresses and coordinates. The database is going to be used as the main source of electricity information. The database was structured in a way that a pre-processing step was needed in order to link it with the building stock database. It contains 29 076 records with 17 different features:

- **Aar**: year of consumption
- **Objectid_1**: identification in the TronderEnergi system
- **Kommunenr**: number of the municipality
- **Beskrivelse:** description. It can contain information of every type, from customers' telephone numbers to building types.
- **Gatenavn:** street name
- **Husnr:** letter of the building
- **Postnr:** postal code number
- **Poststed:** the place of the postal code.
- **Karthenv_x and Karthenv_y**: coordinates of the dwelling/unit
- **Aarsforbruk:** total electricity consumption in kWh.
- **Direccion:** the direction of the building once linked with the building database
- **Bygnignsnr:** building number once linked with the building database
- **Bruksarealtotalt:** total useful floor area of the building
- **Type:** the typology of the building
- **Year:** the construction year.

Section A4. Building stock-energy database

The three databases explained in previous sections are merged into one single database in .csv format from which the energy analysis is done. The script *Merging* was created to allow an automatic creation of the database (see Section E3).

The database contains all the necessary information for building identification and the electricity and heat consumption of a large percentage of non-residential buildings for the year

2018. Information for heat is only available for non-residential building; thus this group of buildings is going to be our focus on the analysis.

Metadata

The database contains 39 319 records that correspond to the number of registered buildings in Trondheim (both residential and non-residential buildings) and 26 features. The information of the first 19 features are explained in the building stock analysis (see **Section C1.** Buildings typologies and sub-typologies.)

- **Type**: type of buildings
- **Year:** year of construction of the buildings
- **Residential:** 1 if is residential, 0 for non-residential buildings
- **TEK:** cohorts use for defining the different buildings technical requirements. These cohorts are chosen as they are used in the database for the calculated energy intensities
- **Electricity:** total electricity consumption in 2018 (kWh/y)
- **Electricity intensity:** electricity intensity (kWh/(m^2 y))
- **Heat intensity:** heat intensity (kWh/(m^2/y)

From the 39 319 records for the merged database, 29 208 contain energy information. From this, 95% are residential buildings while 5% non-residential. Figure 2 indicates the number of buildings by the different typologies as well as their percentages over the total. In total, the dataset contains 29 068 records with their corresponding electricity intensity, 161 records with heat consumption data and 102 buildings with data from both energy carriers.

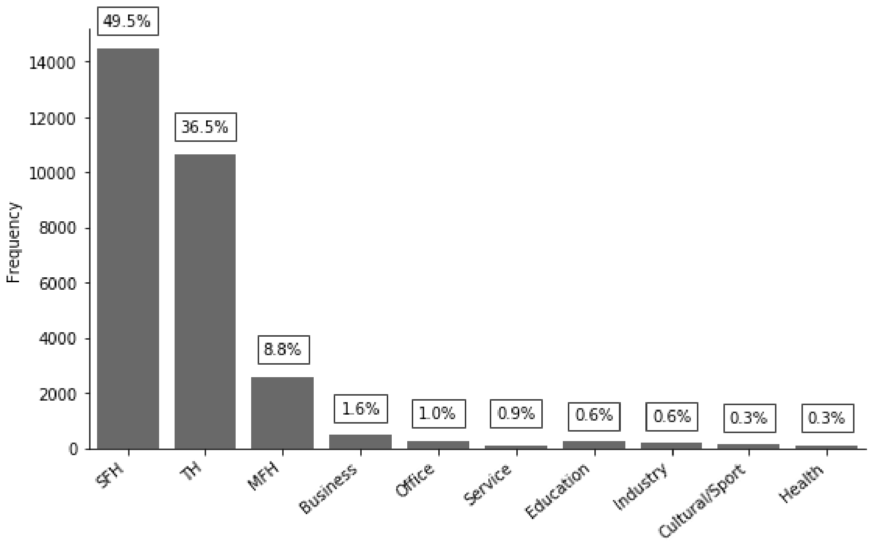


Figure 2. Number of buildings by typologies and their percentages in Trondheim

Table I. Total useful floor areas and their percentages over the total analysed in the case study for each typology. On the right, the number of records with energy data for non-residential categories in the building stock energy database.

Building type	Area analysed (m^2)	Electricity only records	Heat only records	Electricity and heat records
Business	1 376 227 (88%)	430	107	83
Cultural/Sport	260 966 (67%)	149	58	39

Education	720 283 (69%)	236	113	69
Health	380 081 (77%)	67	31	22
Industry	499 567 (66%)	155	70	51
Office	887 315 (77%)	236	84	64
Service	26 632 (9%)	97	24	19
Total	4 183 018 (73%)	1370	487	347

Table I shows the number of records for each building category. Unquestionably, all buildings that use heat from district heating also consume electricity for other purposes. However, due to the complex pre-processing of the data, almost 29% of the electricity values have been lost.

Appendix B. Supplementary methods

Section B1. Data cleaning and processing – Theory

Data collection and preparation are fundamental steps for our methodology. That is why it is of critical importance to understand the type and meaning of the features from our databases. We have two types of features, categorical and numerical.

Categorical variables are those that can take on one of a limited number of possible categories such as gender, name, country, and so forth. There are two sub-types: ordinal variables which have some natural ordering as age groups; and nominal variables that do not have meaningful order.

One of the most typical data cleaning processes of categorical features is what is known as *encoding categorical features*. In our case, we applied **one-hot encoding techniques** to ease data analysis. These types of variables, also known as indicator variables, are mathematically and computationally very convenient for many types of models. They indicate the presence of the attribute (1) or the absence of it (0).

Finally, we apply feature engineering methods that aim to create features from raw data or combining other features in order to use them in the analysis and the predictive model. An example of a feature engineering process used in our model is the addition of the cohorts and building typologies in the databases.

Numerical variables are those whose values can vary in some defined interval and are also divided into two subgroups: *continuous* which can take any value, and an *integer* that can only take integer values.

One of the main purposes when analysing the numerical variables of a dataset is to identify **outliers**. An outlier can be defined as a data point that is way out of keeping with the others or data that does not fit in what it is expected. It is essential to identify them and deal with them. The two options are to delete them or transform. When we have just a few outliers, we may decide to delete them as they might not have big impact on your model. However, in the case of finding several outliers, transforming them is a better option. The general idea is to change the outlier value to the next highest/lowest non-outlier number or change it based upon some other logic.

Section B2. Data cleaning and processing – Building stock georeferenced database

After understanding the building stock georeferenced dataset, we can proceed to exclude the features that are not useful or interesting for the analysis. In this way, we make it easier to manipulate. In result, the database came to have 23 features instead of 53.

In this database, categorical variables have an important repercussion in our analysis as they are the main source of information about the building's characteristics. Several feature engineering processes are needed in order to execute the analysis. A feature named as *typology* is added in order to group the building categories into fewer groups. In other cases, encoding was needed to ease the process of grouping by building types. As an example, an encoded feature named as *residential* indicates if it is a residential or non-residential building. Another important modification in the database is made for building typologies. The feature *bygningstypekode*, as defined before, indicates the type of building according to Standards Norway (2013). For our study 11 categories have been created to facilitate the description of the building stock. From them, 9 describe buildings for non-residential purposes.

Similarly, a revision of numerical feature is performed to identify the outliers that might impact negatively in our results as well as those records which may not be relevant. Specifically, some building subcategories have purposes with different energy behaviour that are out of the scope of the analysis. As an illustration of these subcategories we find garages, animal housing or warehouses, that although they consume electricity, they do not usually consume for heat purposes. For this reason, they are eliminated from the study. For more details on the excluded sub typologies as well as the typologies created, see Table VII of Section C1.

Another categorisation made in the building database is the inclusion of cohorts. We define the cohorts are based on the technical code in force at the time of construction (see Table II). Each year of construction is associated with a specific technical requirement based on Stavset et al. (2015). This categorisation of the buildings based on their technical requirements enables to group buildings with similar energy consumption patterns.

Table II. Age cohort groups based on the year of construction of the building

TEK	Years
Eldre	-1949
TEK49	1950-1968
TEK69	1969-1986
TEK87	1969-1996
TEK97	1997-2006
TEK07	2007-2009
TEK10	2010-2016
TEK17	2017-2018

Section B3. Data cleaning and processing – Heat consumption database

Pre-processing

Due to the complexity of the data, the pre-process was not possible to be directly done in a spreadsheet program. Consequently, several scripts and functions are created in Python and ArcGIS for this purpose. The combination of these scripts allowed us to introduce the Excel file downloaded from the web system and directly modify it. There are two main scripts for the pre-processing of the raw data: *Create files* and *Connect Statkraft with ArcGIS* (see Section E1). These scripts are formed by different functions which are part of Python packages also specially designed for the databases used in the case study.

Table III. Main scripts, packages and functions created for pre-process heat consumption database

Main script	Functions	Package
Create files	Add sheets	Add_Sheets.py
	Modify	modify.py
	Separate grouped values	Separate_grouped.py
	Add lines	add_lines.py

Connect Statkraft with ArcGIS	Connect Statkraft tables	connect_statkraft_tables.py
	Total useful area calculation	bruksareal_calculation.py
	Import to Excel	import_to_excel.py
	Total useful area calculation 2	bruksareal_calculation2.py
	Import to Excel 2	import_to_excel2.py

The tool *Create files* aims to add the necessary sheets to classify and filter the records. First, it filters the records located in Trondheim. As the records belong to customers and not buildings, each record of heat consumption can refer to either individual or groups of buildings. The script is programmed to separate the records of isolated buildings from those which contains information of groups of buildings. The last category contains very different ways of identifying the buildings, thus different classification methods are used depending on the way the information is given. The result is a division between records for isolated buildings, records of groups of buildings with special characters in the data and records of groups of buildings with easy identification.

Lastly, *Connect Statkraft with ArcGIS* links the resulted records with their building identification retrieved from the building stock database. The results are not only imported into the excel file, but they are also exported into ArcGIS Pro, allowing the localisation of the cleaned in ArcGIS for future analysis.

These tools presented are customised for the databases deployed for this study. Their scripts are shown in Section E1 from Appendix D. However, they might be considered as examples of how data can be cleaned and process to be able to proceed with the rest of the methodology

Cleaning and processing

Along with the process followed in the building stock database, some preparation and cleaning of the database were needed.

Some records were seemed not to be correctly linked with their corresponding building identification feature from the building database after the pre-processing. This error is due to the incorrect data values in the localisation features such as addresses or number of building. Without their identification, we do not have enough information to localise them in ArcGIS. Consequently, these records cannot be used to calculate energy intensities.

Moreover, there are just 9 for communication and 2 for emergency typologies in the heat consumption databases. This hinders the calculation of their heat intensities as we have not enough records to perform a proper statistical study. That is the reason why we decided that it is better not to continue analysing these two categories. Their share in the non-residential building stock do not reach 3%. Thus, the impact in the total aggregation is not extremely relevant.

After the cleaning process, the heat database contains 487 records of non-residential buildings, which will give us information about their heat consumption from the district heating system.

Section B4. Data cleaning and processing – Electricity consumption database

The electricity database is provided as a text file in which information is not well structured and some missing values are found. The text file is converted into a .csv to ease the pre-processing and cleaning of the data. Some modifications to the data format are performed to enable the export of the data into ArcGIS. Once the .csv file is exported, it may be linked with the building stock database using the tool *Join with buildings.* Different from the heat database, in this case, the total consumption of a building is the sum of several records, so there is a need for summing up the consumptions by building. To combine these records, we created the tool *Add total consumption* is programmed in Python to automatize the process. To see in more details these two scripts are presented in Section E2 from Appendix E.

The connection with the building stock database is a complicated task as the information given is not organised as well as there are some errors when introducing the identification data. Therefore, after the pre-processing the database used contains 29 076 records linked with their correspondent buildings from the 39 344 we have in Trondheim. This corresponds to approximately 74% of the total building stock.

After the connection with the building stock database, we may know the number of records for each building typology. From the 29 076 records, 1 370 are non-residential buildings. Communication and emergency buildings are also excluded in this case due to the lack of enough data. **¡Error! No se encuentra el origen de la referencia.** shows more in detail the total number of records contained in the database by their building typology. Not further cleaning or data exclusion is needed.

Table IV. Number of records and percentages of non-residential buildings

Building typology	Number of records	Percentage
Business	430	31%
Office	236	17%
Service	97	7%
Industry	155	11%
Cultural/Sport	149	11%
Education	236	17%
Health	67	5%

Section B5. Exploratory Data Analyses - Theory

The exploratory data analysis (EDA) is a combination of numerical and visualization techniques that allow us to understand different characteristics of a dataset, its features, and the potential relationships between them. The process usually is (Fuentes, 2018):

1. Apply standard techniques – understand features.
2. Hypothesis about some aspects of the dataset
3. Apply EDA techniques to begin confirming/rejecting your hypothesis.
4. Stop when you feel comfortable with the understanding you have got.

There are two types of complementary EDA techniques:

- Numerical calculations
- Visualisations

Moreover, there are univariate EDA and bivariate EDA depending on the number of features you are analysing. The first is when the EDA is applied to a single variable, and the aim is to understand each of the features individually. On the other hand, bivariate EDA, by analysing two features, we can gain understanding of the relationships between them.

In the study, we make two different EDA. The first focus on understanding the actual building stock in Trondheim. For it, the building stock database from Geodata will be used. The second EDA is to analyse energy deployment of the building stock, regarding their electricity and heat consumption. The data comes from the building stock-energy dataset.

Section B6. Exploratory Data Analyses – Building stock database

The EDA for this database helps us to understand the complete building stock in Trondheim and gain useful insights into the actual situation.

In Trondheim, 87% of properties, a concept that includes dwellings and units, have residential purposes. Single-family houses (SFH) are the most usual residential building type, followed closed by terraced houses but far away from multi-family houses (MFH), which almost reach 8%. In residential buildings, industrial units are the most dominant group (7%). Business, educational and office business have very similar shares. We must highlight the very few units used for emergency purposes (only 16).

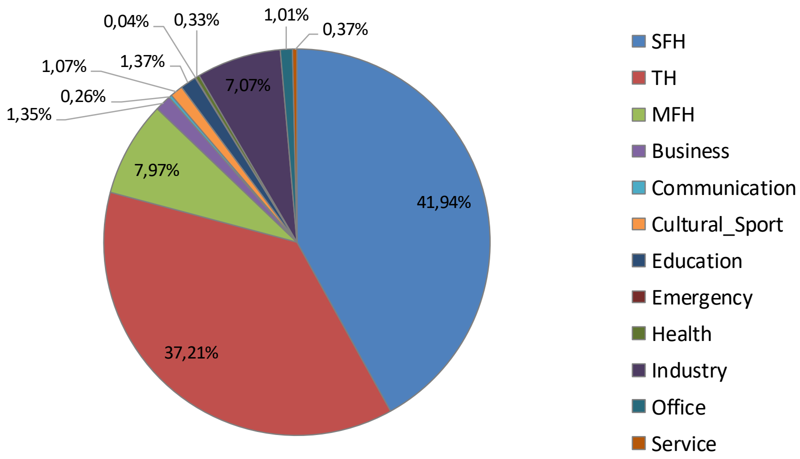


Figure 3. Percentages of units and dwellings in the complete building stock of Trondheim.

Analysing the age of the building stock we find that almost 75% were constructed before 1960 (see Figure 4). This means that we are working with a building stock that is not as old as in most European cities. Moreover, the older building dates of 1 974 and the newest from 2 018 (the study is until this year). The average building in Trondheim was built around the first years of the 70s.

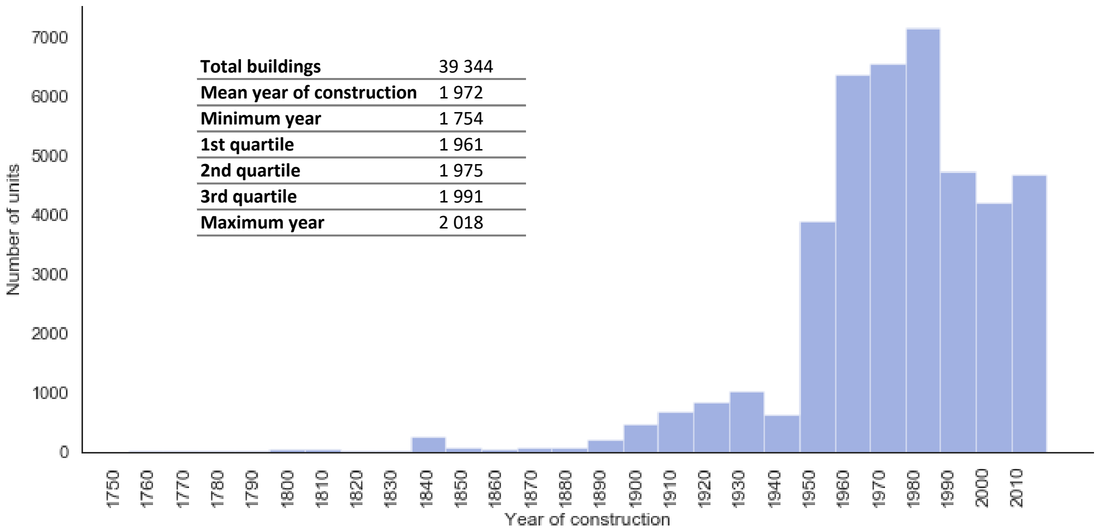


Figure 4. Histogram and analysis of the year of construction for the complete building stock

In Figure 5 and Figure 6 we can see the distribution of the the total useful area of residential and non-residential buildings. SFH and TH have more normalised distributions than TH. The last distribution is right-skewed. On the contrary, the distribution for MFH is uniform. This indicates that the category of MFH includes buildings of very diverse sizes, while the other two residential types are more similar in useful floor area between sub typologies. In general, non-residential types also have uniform distributions, what is expected as inside a typology the buildings may have more diverse purposes. We can see that Industry is the exception. This is because this typology contains a large number of subcategories such as animal housing, warehouses or greenhouses which tend to be quite small.

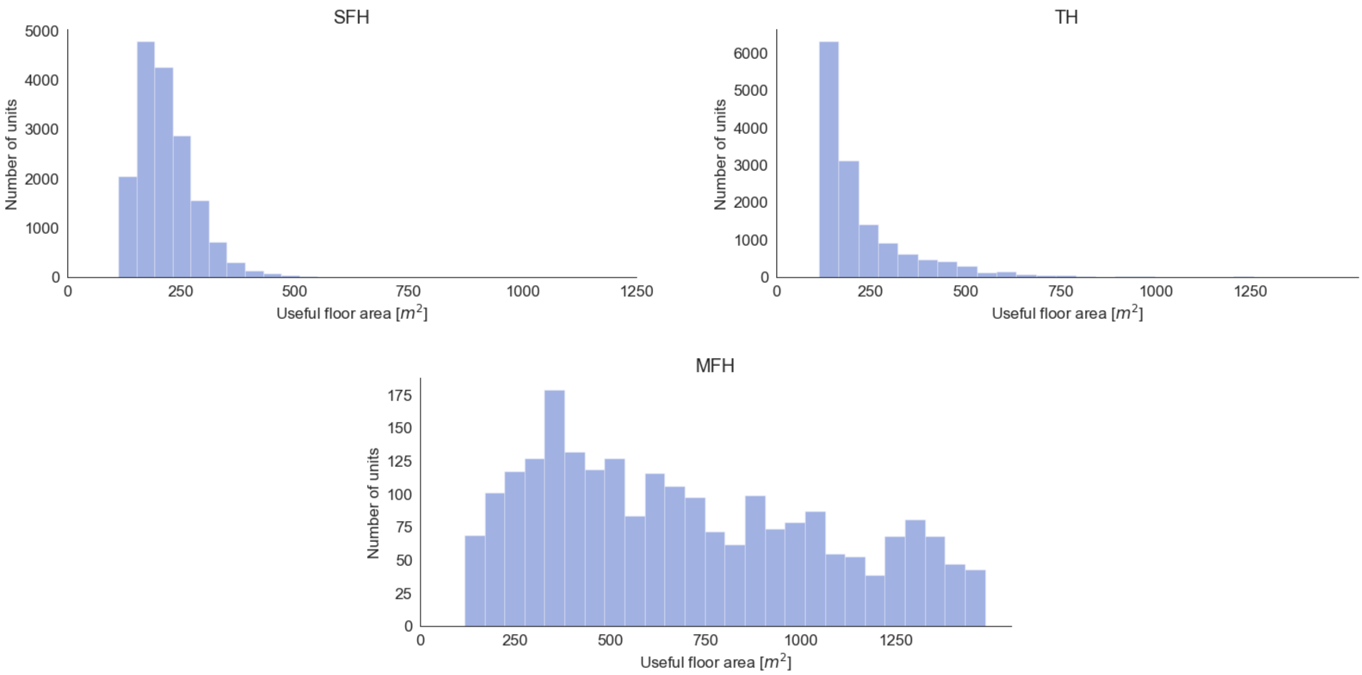


Figure 5. Histograms of the useful floor area of residential buildings

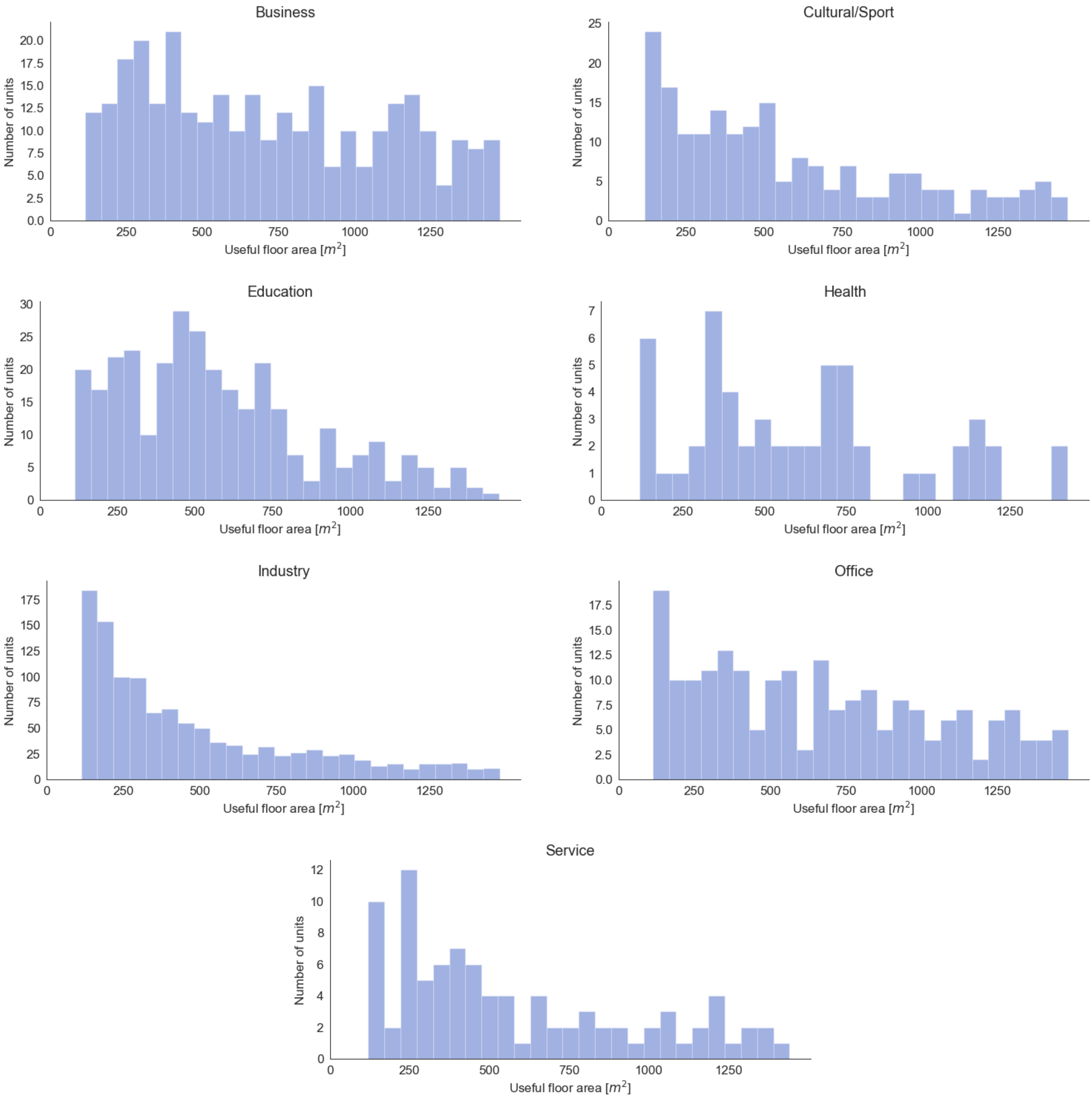


Figure 6. Histograms of the useful floor area of non-residential buildings

These results have to be considered when doing analysis. While for SFH or TH it is correct to introduce average values of their useful floor area in models, for the rest of categories will have considerable uncertainty.

Another interesting characteristic of the building stocks is the land area occupied. In total 9.5 km^2 of land is destined to buildings infrastructure. This corresponds to only almost 3% of the total land in the municipality; thus we can conclude that the building density in Trondheim is very low.

Section B7. Exploratory Data Analyses – Building stock-energy database

Building stock analysed

Although Figure 3 in **Section B6.** Exploratory Data Analyses – Building stock database helped us to understand the actual and general situation of the building stock in Trondheim, we have excluded some buildings due to the reasons explained in **Section B2.** Data cleaning and processing – Building stock georeferenced database. In short, we excluded several sub typologies, especially from industry and communication; and we do not analyse residential buildings in this study. In Figure 7, we indicate the share of the non-residential building types that we are analysing. As we see, industry ceases its predominance and instead we will end up with a large share of business buildings. The main reason behind this decrease of industrial buildings is because several of their sub typologies do not require energy studies (For more details see **Section C1.** Buildings typologies and sub-typologies.). No communication nor emergency buildings are analysed. In this case, we exclude them because of lack of energy data.

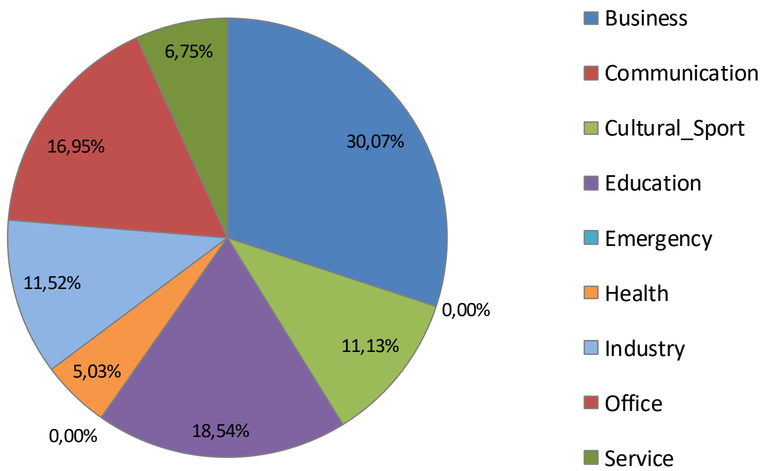


Figure 7. Shares of non-residential buildings analysed in the study. The percentages is for the data after going through the data cleaning process

Electricity intensities

The electricity intensities are analysed for non-residential buildings and a distinction is made between buildings that are connected to the district heating system and those which are not.

We establish that buildings with less than **40 kWh/m2 or higher than 500 kWh/m2** do not represent the building type appropriately, thus are excluded for the analysis. The limits are selected based on the results obtained in the database provided by Sandberg (2019) and Choudhary (2012).

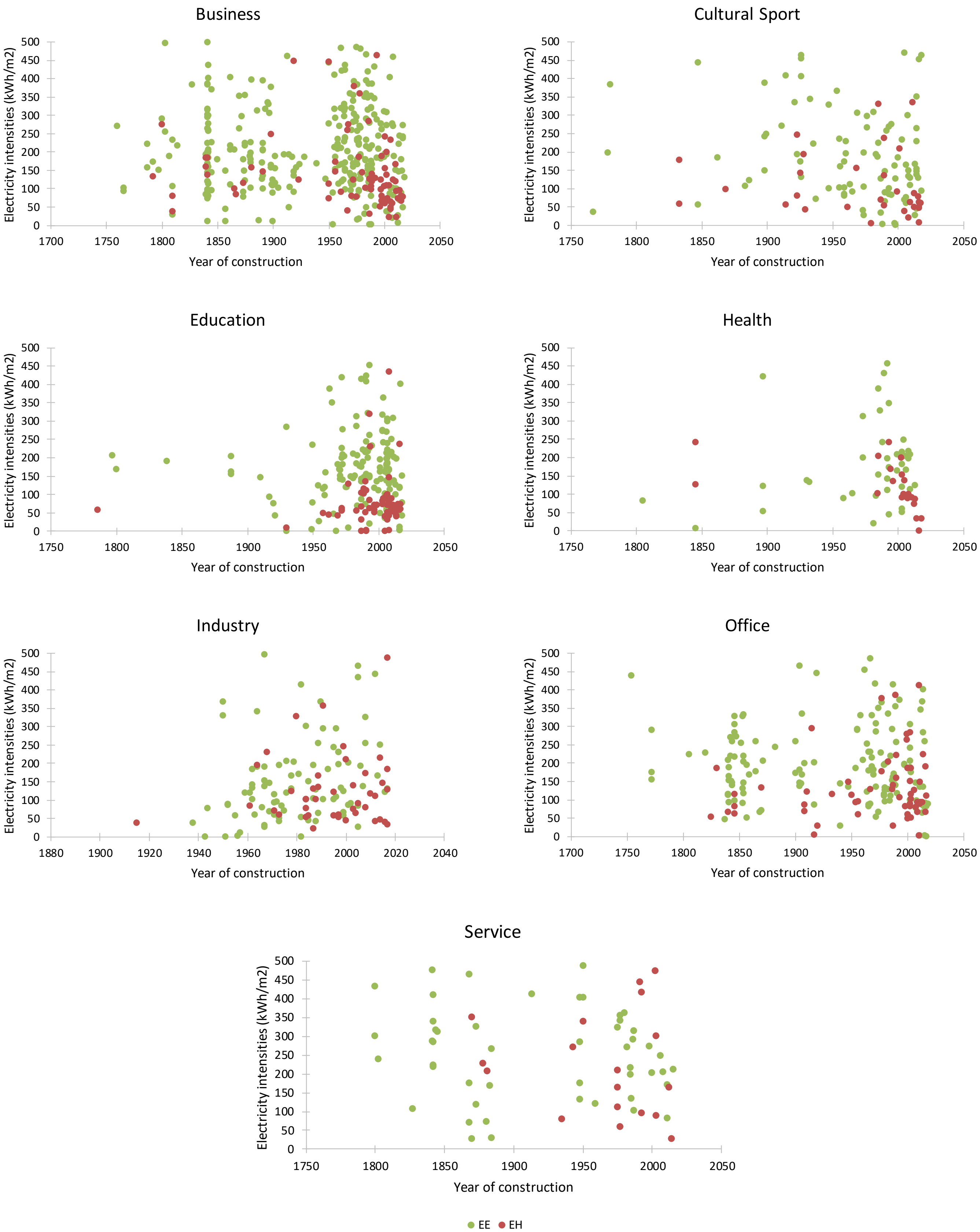


Figure 8. Scatter plot of energy intensity and year of construction for different typologies distinguishing between buildings connected to the DH system and those which are not.

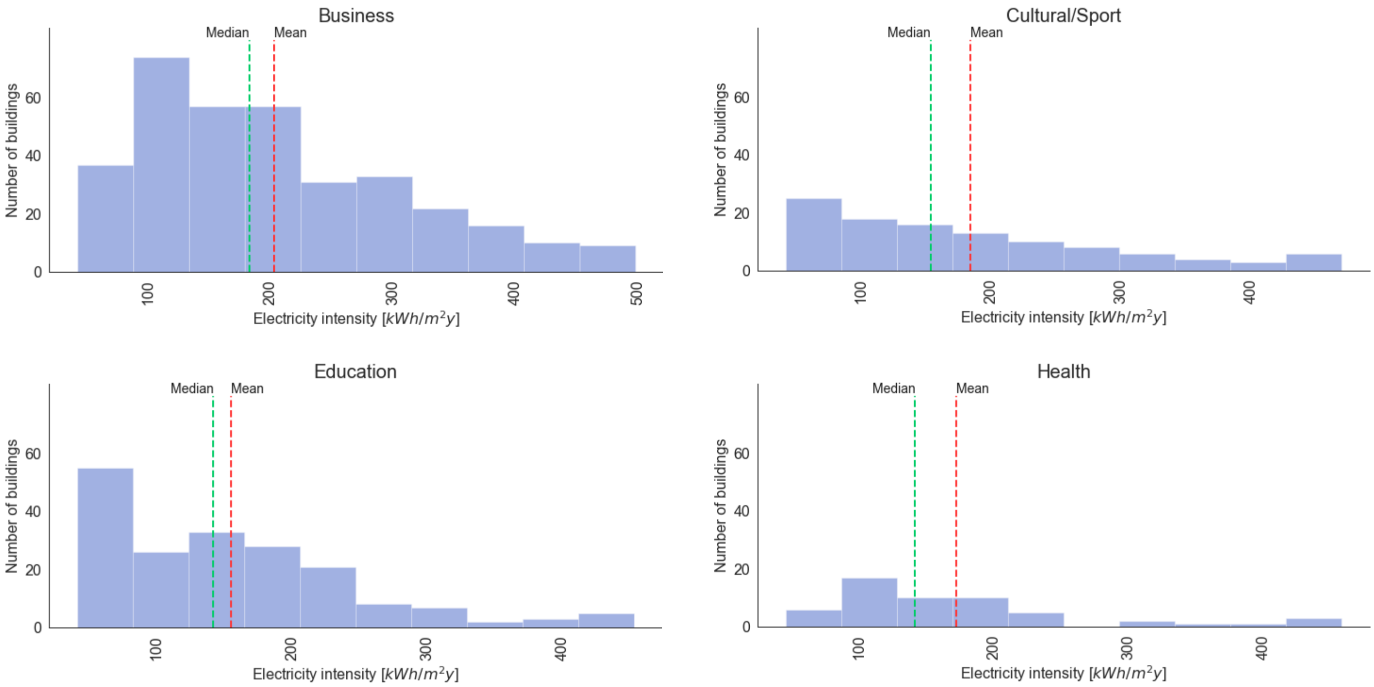
From Figure 8 we can get two different insights. The first is that there is not a clear linear correlation between the year of construction and energy intensities in any type. Moreover, the large amount of noise hinders the visualisation of a pattern. It also means that in case of existing another kind of correlation; it is not strong. However, we can perceive a tendency to have less deviation on the right side of the scatter plots, which may mean that buildings in those buildings may have similar consumptions. We can conclude that a new building does not necessarily come together with a reduction in electricity intensity. Table V shows the Pearson correlation coefficients to indicate the strength of a linear association. They confirm the absence of linear correlation.

Table V. Pearson coefficient to show the relation between construction year and energy intensity

Building type	Pearson coefficient
Business	-0.048
Office	-0.057
Service	-0.044
Education	-0.044
Industry	0.103
Cultural/Sport	-0.173
Health	-0.009

The other conclusion we can gain from the charts is the confirmation that the connection to the district heating system comes together with a reduction in electricity consumption.

Figure 9 shows the distribution of the electricity intensities as well as the mean and median values for each building type. We can see that the histograms are right-skewed in most cases. This means that for our case study, the median value will be more accurate than mean energy intensities. Although for services, the difference between these two ways of calculating the central tendency is not so large, for the rest of categories, the mean will be too overestimated.



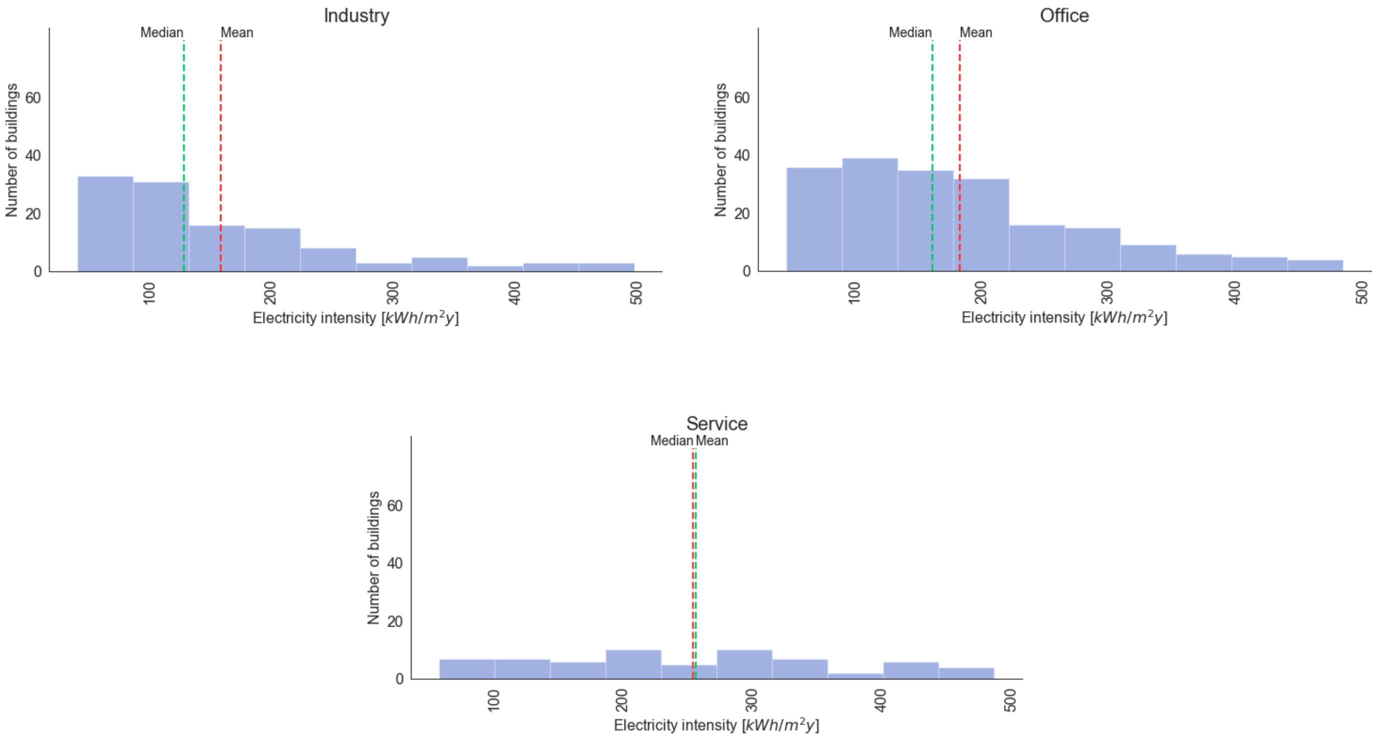


Figure 9. Histograms of electricity intensity by building type

This absence of correlation makes us assume that for non-residential purposes, it is correct to give unique values for electricity intensities and not to group them by technical groups or cohorts. Figure 10 shows the median, 1st quartile and 3rd quartile of electricity intensities without distinguishing between cohorts and differentiating between buildings connected to district heating (EH) and those who are not (EE). The exact values are presented in Table VIII. We can see that Service buildings have the higher electricity intensities but also higher deviations. On the contrary, educational buildings, especially those connected to the district heating system, have the lowest electricity demands. In all categories, EH buildings have lower median intensities than EE buildings.

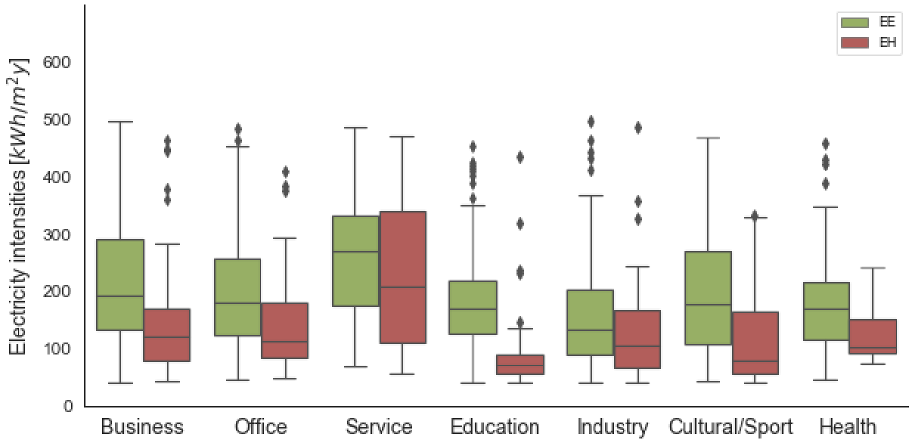


Figure 10. Boxplots of the electricity intensities by each building type

Heat intensities

For the heat intensities, buildings connected to the DH grid are analysed. Figure 11 shows a similar pattern to the electricity consumption: there is no clear correlation between heat intensities of non-residential buildings and their construction year. This is also supported by Table VI. It shows that Pearson's coefficients are close to zero. It is an unexpected result as stricter technical requirements are focused on reducing the heat losses in buildings.

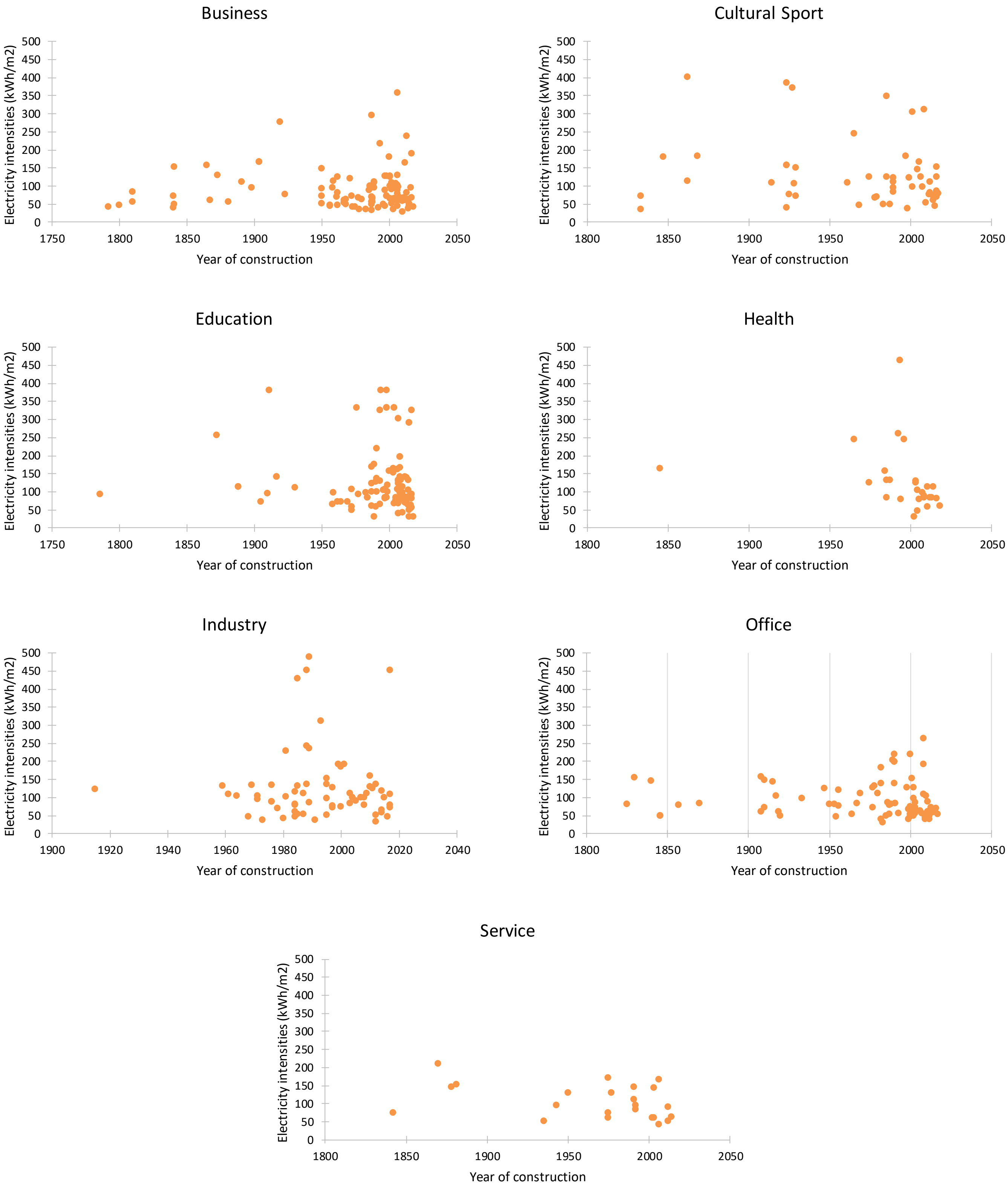


Figure 11. Scatter plots of the heat intensity and year of construction for each non-residential type

Table VI. Pearson coefficient for correlation between year of construction and heat intensities

Building type	Pearson coefficient
Business	0.099
Office	0.068
Service	0.151
Education	0.072
Industry	-0.092
Cultural/Sport	0.050
Health	-0.049

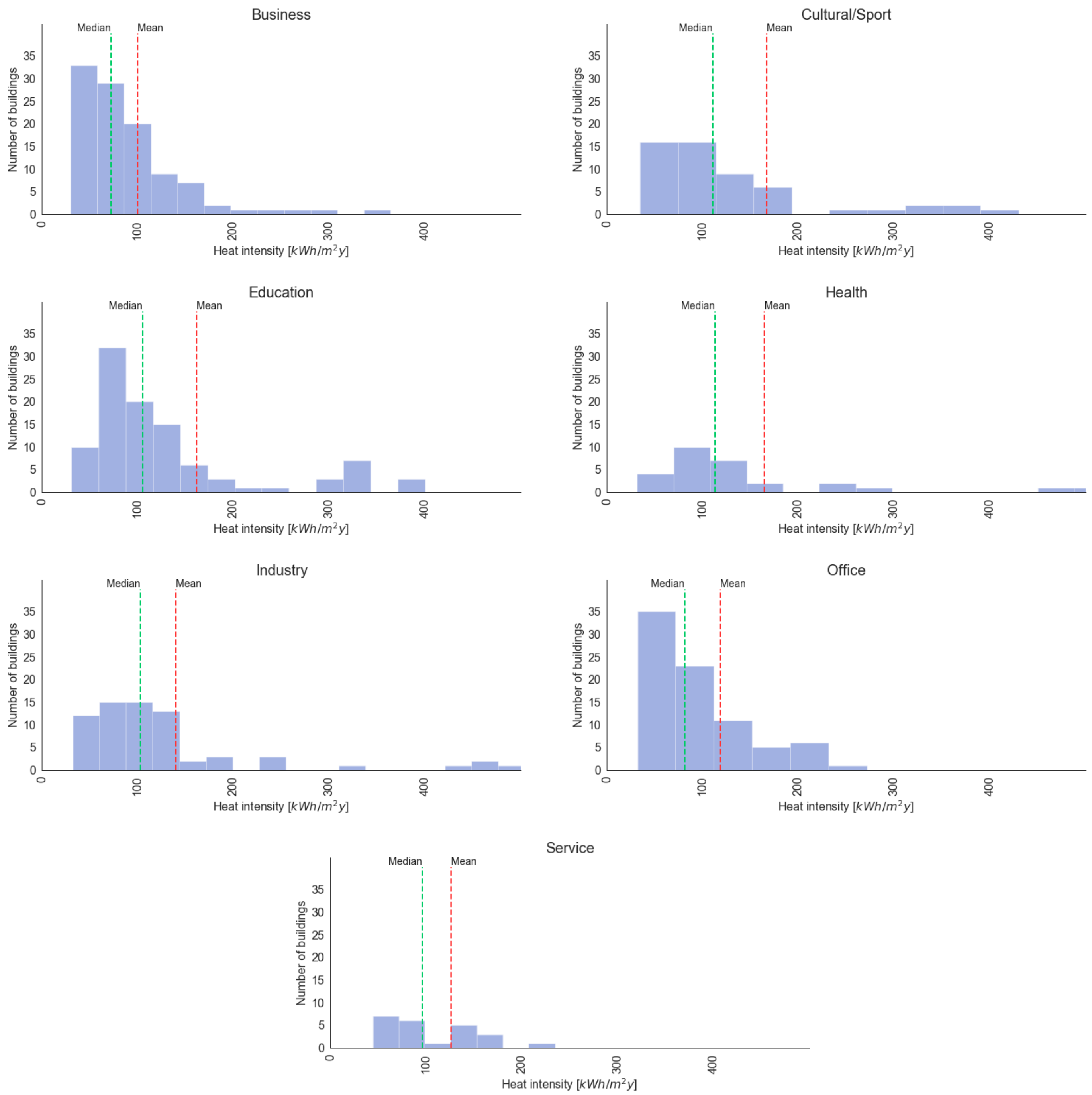


Figure 12. Histograms of heat intensities for each building typology.

Figure 12 presents the histograms of the heat intensities, to decide which type of statistical value should be used. In general, all buildings types have very right-skewed distributions. Thus, the median value is the average to be used for representing heat intensities of non-residential buildings.

The low correlations between heat intensities with age cohorts or year of construction allow us to represent the heat intensities as median values. We indicate them in Figure 13 and Section C2.

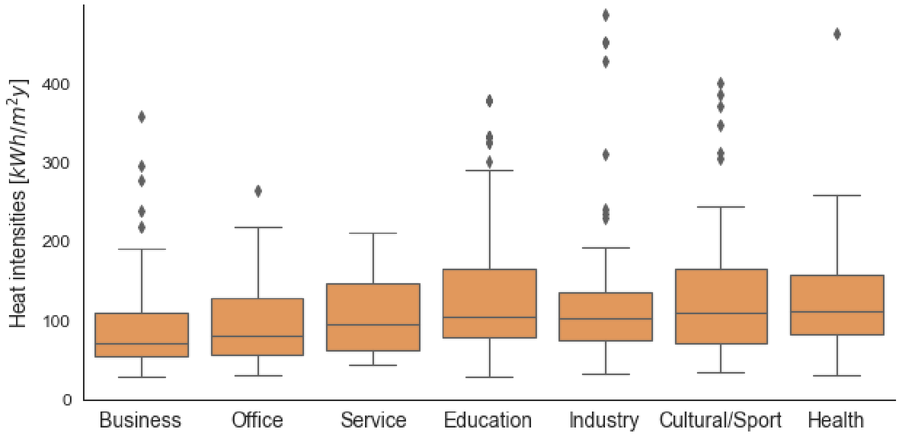


Figure 13. Boxplots of unique heat intensities for non-residential typologies

Section B8. Kernel Distribution Estimation– Theory

Kernel Distribution Estimation (KDE) is one method used for analysing point event distributions, which in our case are building points. It has been used widely for analysis and detection of hot spots events such as traffic accidents or crimes (Levine, 2017; Xie et al., 2008). We selected to use KDE as it is a method of first-order effects which measure the variation in the mean value of the process as well as it easy to understand and implement (Xie & Yan, 2008). KDE aims to *produce a smooth density surface of point events over space by computing event intensity as density estimation* (Xie & Yan, 2008). Its mathematical explanation and algorithm is out of the scope of this study. However, we can perform the estimation by using the KDE tool from ArcGIS Pro.

The KDE allows us to visualise the distribution of building stocks withing the municipality and to see regions more densely occupied. The maps are presented in Section D1.

Section B9. Visualisation methodology

One of the outputs of the methodology described in the study is the possibility of visualize aggregated results of the energy consumption in different spatial resolutions.

The first tool developed to merge and create the building stock-energy database is called *Add energy intensities to buildings.* This model will allow us to add the median, first quartile and third quartile calculated in previous steps to each record. This link is made based on their cohort and building typologies, thus the energy intensities presented in Section C3 are incorporated. The inputs for this scrip are the excel files containing the energy intensities, one per each energy carrier, and two shapefiles for the total building stock. Both shapefiles contain building records, but their difference is that one represents the buildings connected to the district heating system (EH buildings) while the other contains those with electricity for heating purposes (EE buildings).

The tool *Energy aggregation in areas* allows us to perform the necessary aggregations in the desired resolution. The tool has as main inputs the different shapefiles of the desired resolutions and Excel files containing the energy intensities resulted from the previous steps, one excel file for each energy carrier. The model will aggregate by the energy system of the building (EE – buildings with electricity for heating purposes – and EH – buildings connected to the district heating system) and by different typologies. It may be expanded to also aggregate by cohorts. A tool called *Area aggregation in areas* have the same purpose but for area calculations.

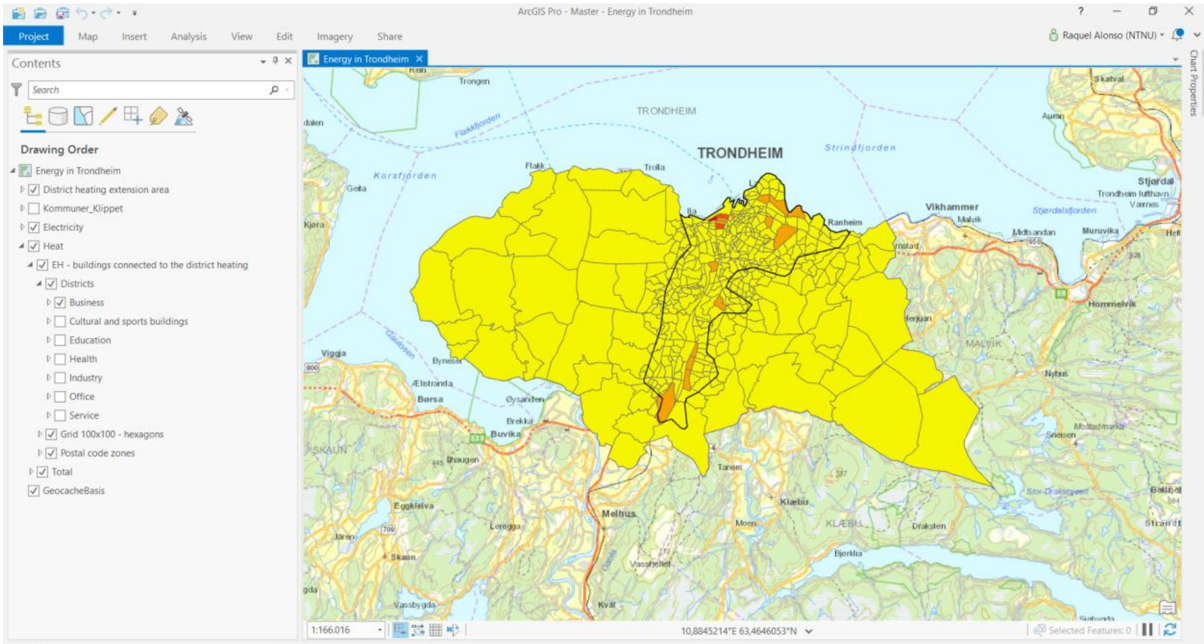


Figure 14. Visualisation tool in ArcGIS Pro

Figure 14 shows how the result looks like in ArcGIS Pro. In the tab on the left, we can select the different layers and the resolutions to be analysed. Afterwards, we can interact with the map. A Supplementary Video is created for a better understanding of the result. Moreover, we can also modify the symbology, see detailed information (e.g. consumption sum, energy densities), create static maps, import extra layers, share our project, add the map to a web page, etc.

The scripts for the visualisation model are shown in Section E4.

Appendix C. Supplementary tables

Section C1. Buildings typologies and sub-typologies.

Table VII. Explanation of sub-typologies contained in each typology, their codes according to XXXX and the number of records included in the building stock database. Sub typologies highlighted in grey are discarded.

SFH

Sub-typologies	Description	Code	Records
Enebolig	Detached building with only one residential unit	111	12758
Enebolig m/hybel/sokkelleilighet	Detached building that additionally contains more apartments	112	4183
Våningshus	Detached house on farm, can also include detached house with apartment apartment, shelf unit etc.	113	666
Våningshus	Detached house on farm, can also include detached house with apartment apartment, shelf unit etc.	113	666

TH

Sub-typologies	Description	Code	Records
Tomannsbolig, vertikaldelt	This applies to residential buildings with two dwellings, including farmhouses on farms.	121	5776
Tomannsbolig, horisontaldelt	This applies to residential buildings with two dwellings, including farmhouses on farms.	122	2133
Våningshus tomannsbolig - vertikaldelt	This applies to residential buildings with two dwellings, including farmhouses on farms.	123	104
Våningshus tomannsbolig - horisontaldelt	This applies to residential buildings with two dwellings, including farmhouses on farms.	124	37
Rekkehus	Row houses have vertical through-going common partitions and each dwelling unit must be located on the ground	131	4893
Kjede-, atriumhus	Chain and atrium houses	133	1011
Terrassehus	Terraced house	135	81
Andre småhus med 3 boliger eller flere	Other small houses with 3 homes or more	136	1549

MFH

Sub-typologies	Description	Code	Records
Stort frittliggende boligbygg på 2 etg	Large detached building on 2 floors	141	276
Stort frittliggende boligbygg på 3 og 4 etg	Large detached residential building of 3 and 4 floors	142	1337
Stort frittliggende boligbygg på 5 etg el mer	Large detached residential building of 5 floors or more	143	228
Store sammenbygde boligbygg på 2 etg	Large built-up residential building of 2 floors	144	141
Store sammenb. boligbygg på 3 og 4 etg.	Large built-up residential buildings of 3 and 4 floors	145	608
Store sammenbygde boligbygg på 5 el mer	Large built-up residential buildings of 5 or more floors	146	139

Bo- og servicesenter	Living and service centre. For the elderly, the disabled, the disabled etc.	151	37
Studenthjem/studentboliger	Dormitory / student housing	152	217
Annen bygning for bofellesskap	Other building for communities or building which is closely related	159	38

Industrial buildings

Sub-typology	Description	Code	Records
Annen industribygning	Other industry buildings	219	106
Annen lagerbygning	Other warehouse	239	660
Annen landbruksbygning	Other agricultural building	249	60
Bygn.for vannfors.bla. pumpest	Building for water supply, pump station, etc.	216	50
Bygning for renseanlegg	Building for treatment plant	214	9
Driftsb. fiske/fangst/oppdr	Commercial building fishing / catching / farming	244	1
Fabrikkbygning	Factory building	211	53
Hus for dyr/landbr.lager/silo	House for animals, feed storage, straw storage, agricultural	241	120
Kjøle- og fryselager	Refrigeration and freezer storage	241	6
Lagerhall	Storehouse	231	196
Naust/redskapshus for fiske	Boathouse / gearbox for fishing	245	14
Veksthus	Greenhouses	243	31
Verkstedbygning	Building for special production or repair	212	136

Office buildings

Sub-typology	Description	Code	Records
Kontor-og administrasjonsbygning , rådhus	Office and administration building is a building for administration, planning and other service work, both public and private.	311	21
Bankbygning, posthus	Bank building, post office	312	5
Mediabygning	House for production and distribution / broadcasting of media products (newspapers, radio and TV)	313	3
Annen kontorbygning	Other office building, or building that is closely linked to office services	319	389

Business buildings

Sub-typology	Description	Code	Records
Kjøpesenter, varehus	A building containing several different shops	321	39
Butikk/forretningsbygning	Shop Building / Business Building	322	265
Bensinstasjon	Fuel stations	323	38
Annen forretningsbygning	Other business buildings	329	224

Communication buildings

Sub-typology	Description	Code	Records
Eksp.bygn. flyterm. kontr.tårn	Expedition building, aircraft terminal, control tower	411	2
Jernbane- og T-banestasjon	Railway and subway station	412	9
Godsterminal	Cargo terminal	415	3
Postterminal	Postterminal	416	3

Annen ekspedisjons- og terminalbygning	Other expedition and terminal building	419	18
Parkeringshus	Parking	431	55
Annen garasje- /hangarbygning	Other garage and hangar building, or building that is closely connected to / serves such building (s)	439	11
Trafikktilsynsbygning	A building where driver card issuance, technical control and registration of motor vehicles take place. The building can also contain an administration part associated with the car inspection	441	1
Annen veg- og trafikktilsynsbygning	-	449	7

Service buildings

Sub-typology	Description	Code	Records
Hotellbygning	Hotel	511	29
Motellbygning	Motel	512	2
Annen hotellbygning	Other building for accommodation (approved by the hotel law or building that is closely linked to / serves such building (s).	519	1
Hospits, pensjonat	Affordable, simply equipped accommodation, usually also served food	521	1
Vandrer-, feriehjem	Original tourist station, mountain lodge, ski station, often combined with other industries such as farms or inns. Also used for cabins connected to the tourist association's trail network, where most are serviced, but also some unattended cabins.	522	1
Camping/utleiehytte	Camping hut: A simpler accommodation hut preferably intended for car tourists. As a rule, they are linked to a camping site. Guests usually keep linens themselves. Rental cabin: A small house with high, medium or low standard, for temporary stay. Mostly larger and better standard than a camping cabin, often not located on a campsite, but more scattered in the terrain. Modern rental cabins are often of a high standard and several are often linked to a centre with guard, kiosk / business and other facilities.	524	10
Annen bygning for overnatting	-	529	7
Restaurantbygning, kafebygning	Restaurant and cafe building	531	76
Sentralkjøkken, kantinebygning	Building for kitchen or canteen attached to larger unit, but where the building is for itself	532	5
Gatekjøkken, kioskbygning	Street Food: Small serving area, with simple dishes - most often semi-finished products - is delivered over the counter. Kiosk: Small building for sale of goods, leaves, sausages and mineral water	533	18
Annen restaurantbygning	-	539	6

Educational buildings

Sub-typology	Description	Code	Records
Lekepark	Place for stay for children 1 - 5 years. May be full day and half day, and there is a certain educational content attached to it	611	13
Barnehage	Place for stay for children 1 - 5 years. May be full day and half day, and there is a certain educational content attached to it	612	193

Barneskole	School building for school classes 1-7, for children aged 6-12 years.	613	88
Ungdomsskole	School building for school classes 8-10, for children aged 13-15 years.	614	20
Kombinert barne- og ungdomsskole	School building for school classes 1-10, for children aged 6-15 years.	615	2
Videregående skole	High school	616	25
Annen skolebygning	Other schools that do not fit in with the above categories, or building that is closely linked to / serves of such buildings	619	143
Universitet, høgskole m/auditorie,lesesal mv.	University and college building with integrated functions, auditorium, reading room, etc.	621	48
Laboratoriebygning	Laboratory buildings (eg in industry, hospitals and universities)	623	17
Annen universitet, høgskole og forskningsbygning	Other university and college buildings that do not fit in with the above categories	629	27

Cultural and sport buildings

Sub-typology	Description	Code	Records
Museum, kunstgalleri	Museum: Building for display of special objects and mention of these. Art gallery: Building for exhibition and sale of art	641	62
Bibliotek, mediatek	Building for lending of books, audio books, films, newspapers and the like. Modern library / media library often has available PC with Internet connection	642	3
Annen museums- og biblioteksbygning	Buildings that do not fit in with the above categories, or building that is closely linked to / serves such building	649	50
Idrettshall	Building primarily for sports purposes, usually also has wardrobe facilities and a kiosk	651	38
Ishall	Building primarily used as a skating rink for long and ice hockey, usually also has wardrobe facilities and a kiosk. The ice rinks are often not more than half the year, and the rest of the time is often used for other sports events, athletics, various tournaments and the like. or it can be used for fairs	652	2
Svømmehall	Indoor swimming pool building used for education, exercise, training and competitions.	653	3
Tribune og idrettsgarderobe	Tribune: Built-in stand for outdoor sports facilities. Sports wardrobe: Building for wardrobe adjacent to sports facilities	654	11
Helsestudio	Compact training facility run on a commercial basis - not public. Modern studios often have options for device training, squash, aerobics, massage and other organized activities	655	6
Annen idrettsbygning	-	659	119
Kinobygning, teaterbygning, opera/konserthus	Building for performing cinema, theatre, opera and concerts	661	8
Samfunnshus , grendehus	Central activity house for the village / hamlet, multifunctional buildings used for everything from sports events to parties and other social gatherings	662	23
Annen kulturhus	-	669	44
Kirke , kapell	Christian place	671	44
Bedehus, menighetshus	Christian assembly house	672	22

Krematorium, gravkapell, bårehus	Crematorium: Building for cremation. Excavation chapel and cabins: Building for storing dead people until burial and cremation	673	3
Synage og moske	Synagogue and Muslim God House	674	1
Kloster	Building where men or women have retreated to realize the religious ideal, in a closed society, and on a particular rule	675	1
Annen bygninger for religiøse aktiviteter	-	679	8

Emergency buildings

Sub-typology	Description	Code	Records
Brannstasjon, ambulansestasjon	Fire station, ambulance station	822	6
Offentlig toalett	This applies to buildings that by definition are classified under main type 8 (prison building, emergency building etc.), but which do not naturally belong to one of the above-mentioned building types.	840	8
Politistasjon	Police station	821	2

Section C2. Energy intensities by building typologies

Table VIII. Electricity intensities by building typologies. We distinguish between buildings connected to the district heating system (EH), those who are not (EE), and provide the energy intensities combining both (General).

EE

Type	Median (*kWh/m²*)	1st quartile (*kWh/m²*)	3rd quartile (*kWh/m²*)
Business	221	155	319
Health	198	138	242
Office	200	148	291
Service	310	209	453
Education	190	147	260
Industry	192	130	331
Cultural/Sport	256	161	396

General

Type	Median (*kWh/m²*)	1st quartile (*kWh/m²*)	3rd quartile (*kWh/m²*)
Business	193	134	290
Health	171	115	217
Office	180	123	258
Service	271	174	332
Education	170	127	219
Industry	133	90	204
Cultural/Sport	178	115	217

EH

Type	Median (*kWh/m²*)	1st quartile (*kWh/m²*)	3rd quartile (*kWh/m²*)
Business	121	79	170

Health	114	93	165
Office	115	84	181
Service	208	111	339
Education	71	57	91
Industry	106	68	168
Cultural/Sport	79	56	166

Table IX. Heat intensities by building typologies.

Type	**Median (*kWh/m²*)**	**1st quartile (*kWh/m²*)**	**3rd quartile (*kWh/m²*)**
Business	123	79	193
Health	101	89	165
Office	115	82	187
Service	208	103	344
Education	68	51	94
Industry	118	62	212
Cultural/Sport	91	53	287

Section C3. Energy intensities by building typologies and age cohorts.

Table X. Electricity intensities by building typologies and cohorts. We distinguish between buildings connected to the district heating system (EH), those who are not (EE), and provide the energy intensities combining both (General).

EE

	Business			Office			Service		
	Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile
Eldre	186	123	263	178	138	256	283	174	335
TEK49	231	165	317	215	180	291	401	260	444
TEK69	215	137	303	174	117	270	289	215	339
TEK87	211	156	315	166	158	330	208	155	261
TEK97	195	184	276	196	184	214	247	225	260
TEK07	229	123	275	97	88	122	203	203	203
TEK10	146	98	171	113	84	278	170	125	190
TEK17	236	129	241	84	82	86	271	174	332

Education			Industry			Cultural/Sport			Health		
Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile
162	121	197	76	76	76	242	149	383	128	92	137
123	103	217	138	95	190	135	98	204	95	92	99
182	149	209	119	69	167	224	192	296	257	166	325
208	148	242	156	103	252	112	85	229	199	142	348
171	119	213	163	87	222	161	75	184	175	132	207
170	130	201	253	195	289	133	129	148	204	177	212
133	104	207	208	154	298	160	139	255	125	125	125
115	69	213	130	130	130	277	185	370	171	115	217

General

	Business			Office			Service		
	Median	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**
Eldre	183	122	258	167	115	229	270	174	332
TEK49	224	152	314	188	108	261	370	284	422
TEK69	190	132	301	178	126	270	215	164	321
TEK87	179	119	297	162	149	306	314	102	415
TEK97	146	84	207	175	93	201	260	214	292
TEK07	155	97	271	86	78	96	203	203	203
TEK10	99	80	150	92	87	241	166	142	180
TEK17	183	89	240	87	84	99	310	209	453

Education			Industry			Cultural/Sport			Health		
Median	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**
159	90	194	76	76	76	192	109	333	130	112	164
118	71	180	140	92	193	132	93	190	95	92	99
166	125	205	102	67	150	245	193	302	203	141	317
156	115	232	130	102	244	112	82	213	186	140	269
135	78	189	123	70	205	161	83	195	158	99	200
139	73	185	172	137	253	131	113	144	155	103	207
73	58	153	133	112	202	136	62	227	90	85	100
76	64	133	156	129	259	92	76	277	198	138	242

EH

	Business			Office			Service		
	Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile
Eldre	145	118	183	115	68	132	228	206	270
TEK49	159	107	262	94	93	112	339	339	339
TEK69	144	101	283	190	164	246	137	98	175
TEK87	123	104	130	159	140	220	415	255	429
TEK97	106	65	146	115	82	186	298	192	385
TEK07	91	75	107	80	73	86	271	174	332
TEK10	84	73	102	92	92	190	162	162	162
TEK17	76	76	76	111	111	111	271	174	332

Education			Industry			Cultural/Sport			Health		
Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile	Median	1st quartile	3rd quartile
159	59	59	59	106	68	168	98	58	178	184	155
118	48	47	49	194	139	212	101	74	127	114	93
166	57	55	62	88	62	102	328	328	328	154	128
156	105	76	125	130	112	150	102	64	160	169	153
135	74	71	87	74	60	138	149	120	178	120	96
139	77	57	98	125	102	149	62	62	62	94	92
73	60	55	72	114	62	138	62	48	81	88	81
76	67	64	71	183	155	335	59	59	59	114	93

Table XI. Heat intensities by building typologies and cohorts.

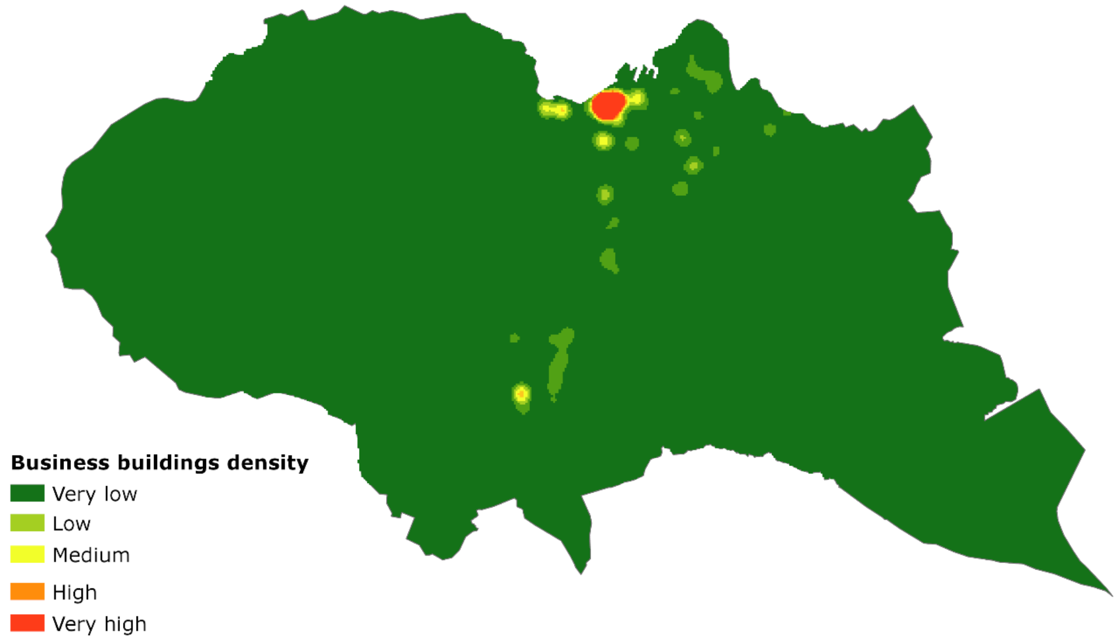
	Business			Office			Service		
	Median	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**
Eldre	83	57	155	83	60	139	121	152	81
TEK49	67	53	94	82	71	93	130	130	130
TEK69	70	43	100	110	66	133	103	141	73
TEK87	67	51	106	139	80	198	104	121	93
TEK97	89	65	121	68	57	97	102	161	61
TEK07	83	76	91	83	47	170	208	103	344
TEK10	62	51	82	68	56	80	63	78	58
TEK17	69	55	130	55	55	55	208	103	344

Education			Industry			Cultural/Sport			Health		
Median	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**	**Median**	**1st quartile**	**3rd quartile**
162	113	94	256	122	122	122	132	76	181	792	635
123	73	71	79	110	104	131	110	79	177	244	244
182	83	72	98	91	67	132	71	68	125	128	114
208	126	92	171	137	88	240	122	94	599	246	131
171	135	84	333	100	80	157	137	117	171	106	63
170	132	84	224	100	95	103	99	76	205	87	86
133	84	66	123	100	55	128	81	72	99	84	81
115	87	62	268	94	77	195	80	80	80	61	61

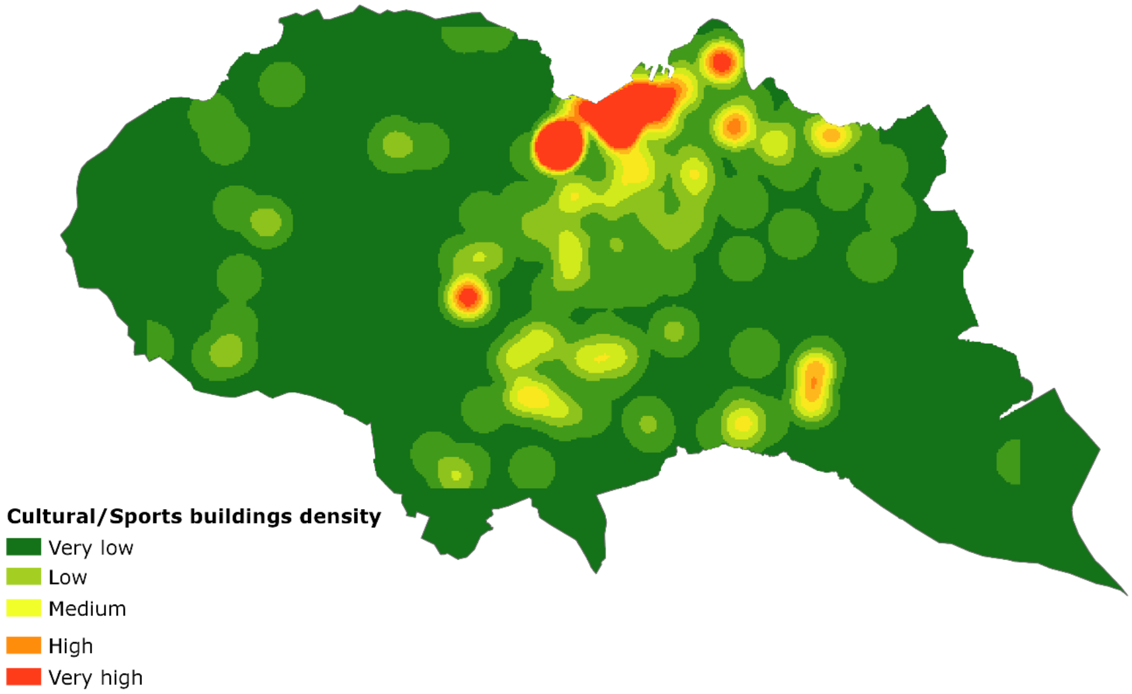
Appendix D. Supplementary maps

Section D1. Building stock dispersion[1].

Business

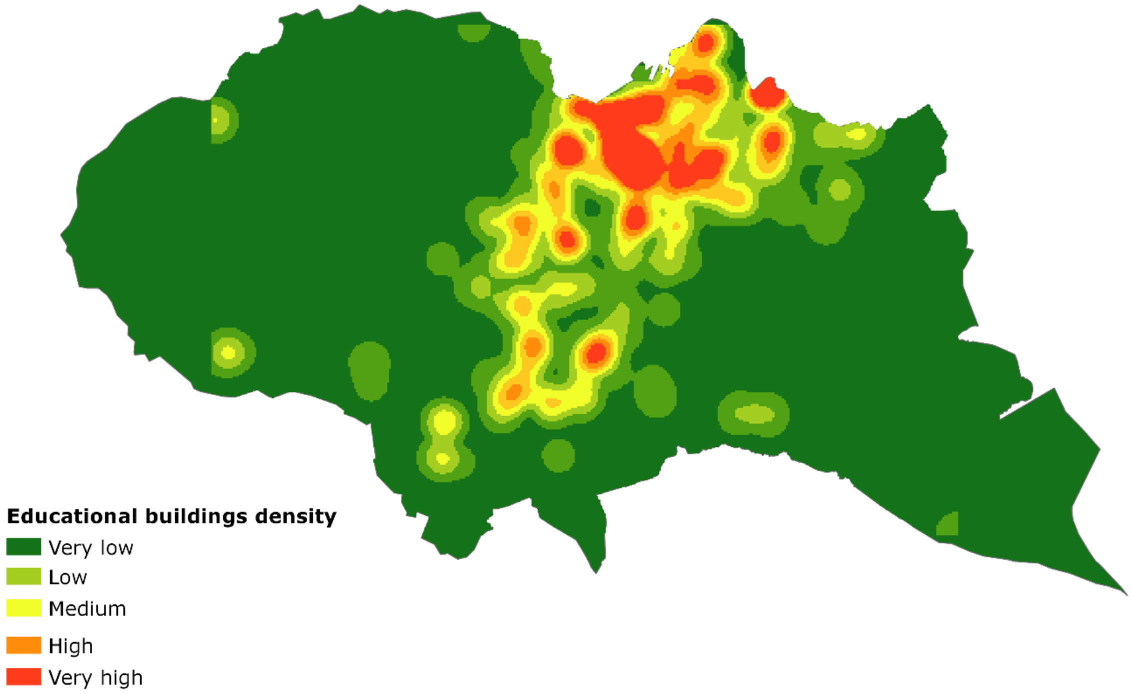


Cultural/Sports

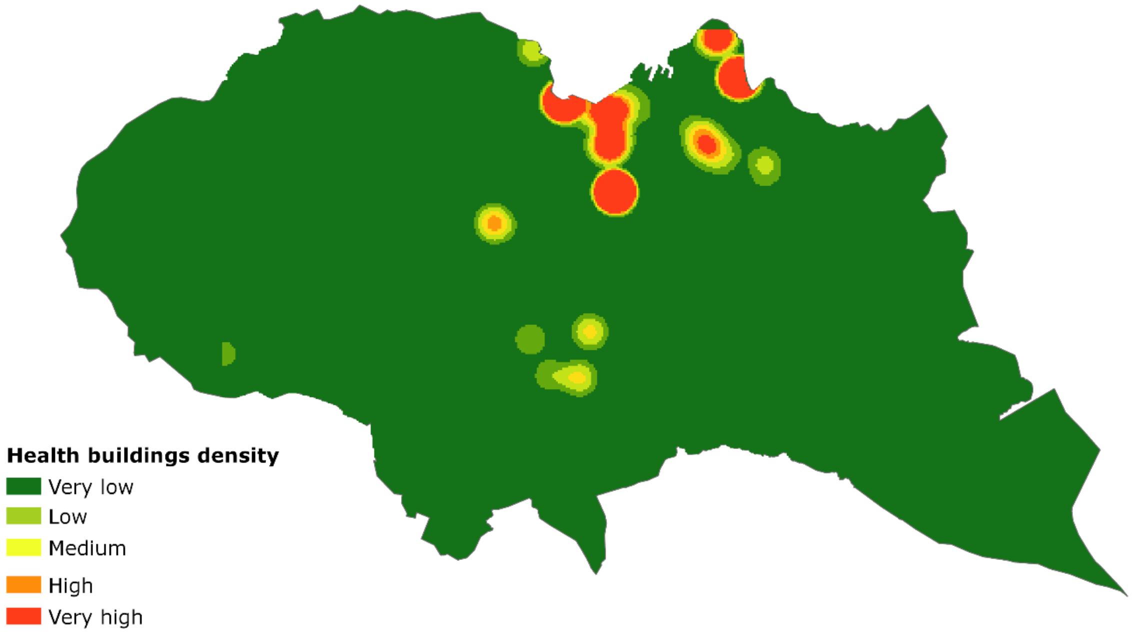


[1] The maps cannot be compared between them. Individual analysis should be made as the calculations are made without taking into consideration with the rest of categories.

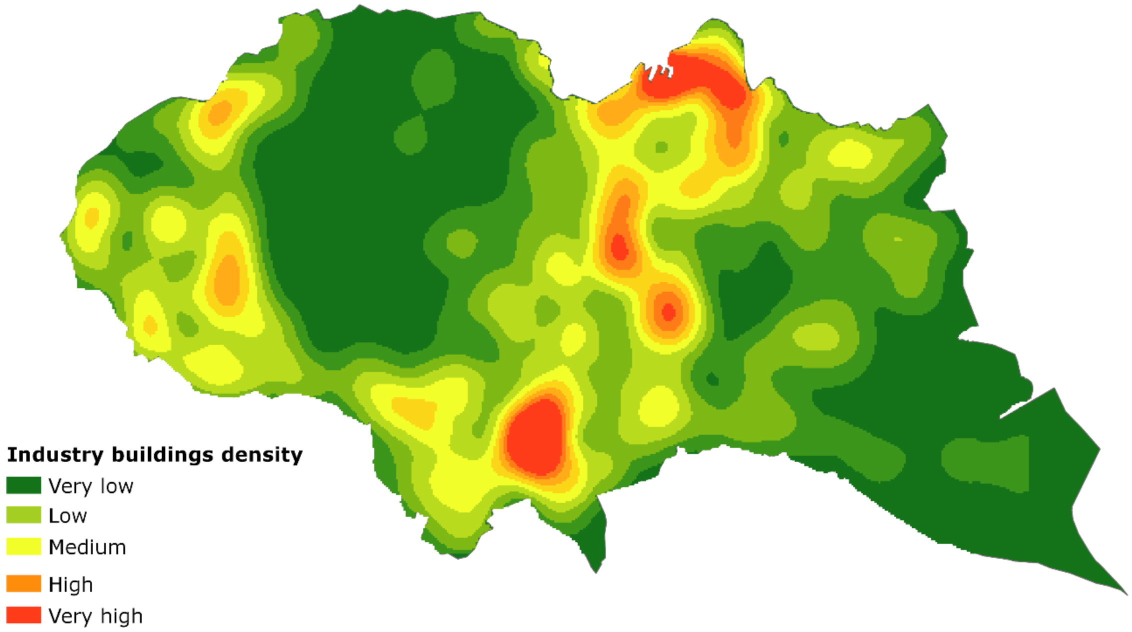
Education



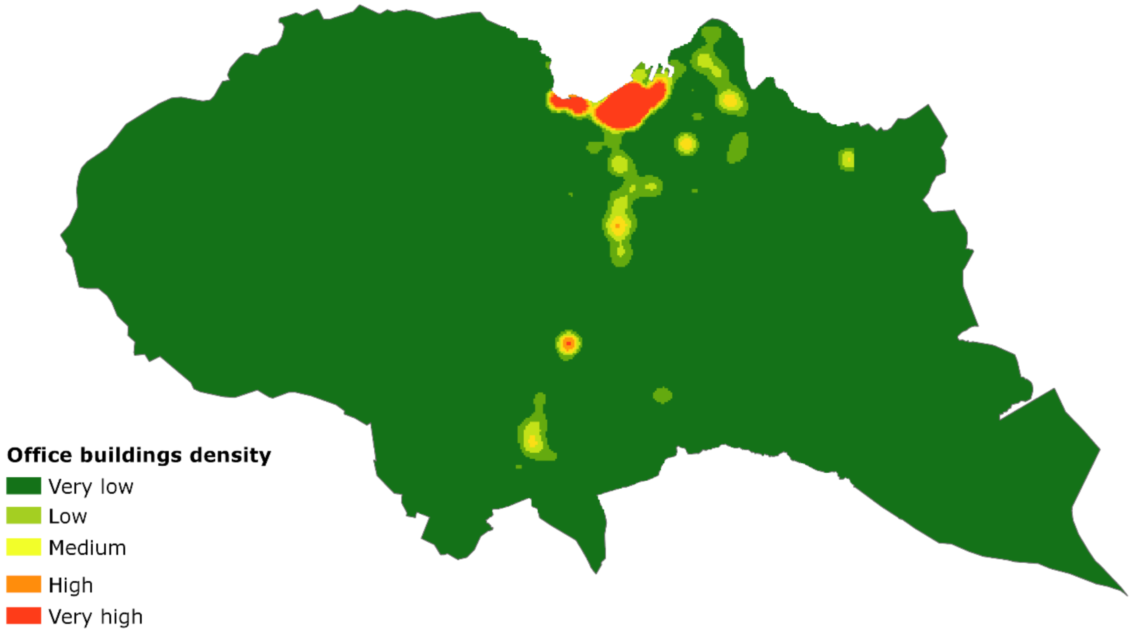
Health



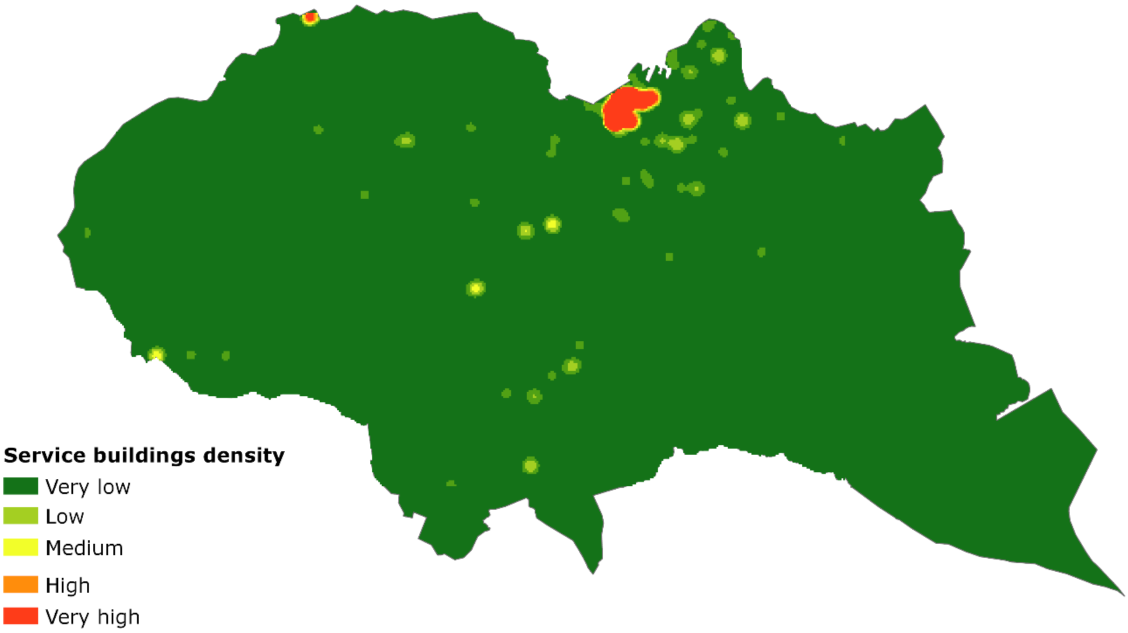
Industry



Office



Service


Service buildings density
Very low
Low
Medium
High
Very high

Appendix E. Supplementary scripts

Section E1. Statkraft Varme pre-processing model

The model for cleaning Statkraft Varme database consists of 2 main scripts, which in turn are formed by different packages that we created.

Table XII. Scripts and functions for the model

Main script	Functions
Create files	Add sheets
	Modify
	Separate grouped values
	Add lines
Connect Statkraft with buildings database	Join Statkraft
	Total useful floor area calculation
	Import to Excel
	Total useful floor area calculation 2
	Import to Excel 2

Code 1. *Create file* script

```
#This module is designed to create a proper excel sheet document from the
data downloaded in Statkraft Varme
from package.Add_Sheets import add_sheets
from package.modify import modify
from package.Separate_grouped import separate_grouped
from package.add_lines import add_lines

#Introduce the name of the file you want to convert. Make sure it is in
the same folder as this module.
#Also add the name of the sheet which contains raw data
filename="excel_files\Fjernvarme_raw_copy.xlsx"
print("Modifying the document: " + filename + "\n")
sheet="Sheet1"
print("Raw data in sheet: " + sheet + "\n")

print("--------Adding proper sheets---------\n")
add_sheets(filename,sheet)

print("--------Obtaining proper information---------\n")
modify(filename)

print("--------Separating grouped records-----------\n")
separate_grouped(filename)

print("--------Adding lines to non special grouped records---------\n")
add_lines(filename)

print("The document has been filled."
```

Code 2. Function *Add sheets*

```
#Module to add the sheets: Trondheim, Nongrouped (add_firstsheets) and
Grouped_nospecial (add_grouped)

def add_sheets(filename, sheet):
        from openpyxl import load_workbook
        doc = load_workbook(filename=str(filename))
        rawdata=doc[str(sheet)]
        #Create sheets
        trondheim=doc.create_sheet("Trondheim")


        #Obtain number of records in Sheet1
        counter = 0
        for row in rawdata:
                counter+=1

        #Add titles to columns
        for c in range(1, 10):
                if c<3:
                        _=trondheim.cell(row=1, column=c,
value='{0}'.format(rawdata.cell(row=1, column=c).value))
                else:
                        _=trondheim.cell(row=1, column=c+6,
value='{0}'.format(rawdata.cell(row=1, column=c).value))

        trondheim.cell(row=1, column=3, value='{0}'.format("adresse"))
        trondheim.cell(row=1, column=4,
value='{0}'.format("adressenavn"))
        trondheim.cell(row=1, column=5, value='{0}'.format("nr"))
        trondheim.cell(row=1, column=6, value='{0}'.format("postnavn"))
        trondheim.cell(row=1, column=7, value='{0}'.format("postnr"))
        trondheim.cell(row=1, column=8, value='{0}'.format("bokstav"))

        #Create copy sheets: Non_grouped, Grouped, Grouped_special,
Grouped_nonspecial
        nongrouped=doc.copy_worksheet(trondheim)
        nongrouped.title="Non_grouped"
        grouped=doc.copy_worksheet(trondheim)
        grouped.title="Grouped"
        grouped.cell(row=1, column=16, value='{0}'.format("special
(y/n)"))
        grouped_special=doc.copy_worksheet(grouped)
        grouped_special.title="Grouped_special"
        grouped_nonspecial=doc.copy_worksheet(grouped)
        grouped_nonspecial.title="Grouped_nonspecial"
        grouped_nonspecial.cell(row=1, column=17,
value='{0}'.format("new_nr"))

        #Go record by record in Sheet 1 to cheek they are in Trondheim
        lines=2 # indicates the line number to add the information in the
sheet
        for row in range(2, counter+1):

                if rawdata.cell(row=row, column=3).value == "Trondheim ":
                        for c in range(1, 10):
                                if c<3:
                                        _=trondheim.cell(row=lines,
column=c, value='{0}'.format(rawdata.cell(row=row, column=c).value))
                                else:
```

```
                                        _=trondheim.cell(row=lines,
column=c+6, value='{0}'.format(rawdata.cell(row=row, column=c).value))

                            lines+=1
                    else:
                            continue

            #Guardar cambios
            doc.save(str(filename)
```

Code 3. Function *Modify*

```
###### Modify excel sheet from Statkraft to obtain address so that it can
be linked with GIS

def modify(filename):
        from openpyxl import load_workbook
        doc = load_workbook(filename=str(filename))
        prueba=doc["Trondheim"]
        grouped=doc["Grouped"]
        nongrouped=doc["Non_grouped"]

        # Calculate maximum row
        counter = 0
        for row in prueba:
                counter+=1
        nr_lineas_counter=0

        #Separate information
        for row in range(2,counter+1):
                #addresse completo
                anleggsnavn=prueba.cell(column=2, row=row).value
                trocitos_anleggsnavn=anleggsnavn.split(",")
                _=prueba.cell(column=3, row=row,
value='{0}'.format(trocitos_anleggsnavn[0].strip()))

                #print(trocitos_anleggsnavn[:]) #linea para ver que
elementos tenemos separados

                #separación calle y nr
                adresse=[]
                nr=[]

                for i in trocitos_anleggsnavn[0]:

                        if i == "0" or i == "1" or i == "2" or i == "3" or
i == "4" or i == "5" or i == "6" or i == "7" or i == "8" or i == "9" or
i=="-":
                                break
                        adresse.append(i)

                adresse="".join(adresse)

                nr=trocitos_anleggsnavn[0].replace(adresse, "")
                nr="".join(nr)

                _=prueba.cell(column=4, row=row,
value='{0}'.format(adresse.strip()))
                _=prueba.cell(column=5, row=row, value='{0}'.format(nr))

                #Post code and postnavn
                post_code=[]
```

```
                postnavn=[]

                for i in trocitos_anleggsnavn[1]:
                        if i == "0" or i == "1" or i == "2" or i == "3" or
i == "4" or i == "5" or i == "6" or i == "7" or i == "8" or i == "9":
                                post_code.append(i)
                        else:
                                postnavn.append(i)

                postnavn="".join(postnavn)
                post_code="".join(post_code)
                postnavn=postnavn.strip()
                postnavn=postnavn.capitalize()

                _=prueba.cell(column=6, row=row,
value='{0}'.format(postnavn))
                _=prueba.cell(column=7, row=row,
value='{0}'.format(post_code))

        #Add info in grouped and non_grouped
        counter = 0
        for row in prueba:
                counter+=1

        lines_g=2 # indicates the line number to add the information in
the sheet
        lines_ng=2
        for row in range(2, counter+1):
                if prueba.cell(row=row, column=5).value.find("-") != -1 or
prueba.cell(row=row, column=5).value.find("+") != -1:
                        for c in range(1, 16):
                                _=grouped.cell(row=lines_g, column=c,
value='{0}'.format(prueba.cell(row=row, column=c).value))
                        lines_g += 1
                else:
                        for c in range(1, 16):
                                _=nongrouped.cell(row=lines_ng, column=c,
value='{0}'.format(prueba.cell(row=row, column=c).value))
                        lines_ng+=1

        #Guardar cambios
        doc.save(str(filename))
```

Code 4. Function *Separate grouped*

```
###### Separate grouped records

def separate_grouped(filename):
        from openpyxl import load_workbook
        # Example: modificar('Trial.xlsx', "Sheet 1", "y"/"n")
        doc = load_workbook(filename=str(filename))
        prueba=doc["Trondheim"]
        grouped=doc["Grouped"]
        grouped_special=doc["Grouped_special"]
        grouped_nonspecial = doc["Grouped_nonspecial"]

        #Add info into grouped_special and grouped_nonspecial
        counter = 0
        for row in grouped:
                counter+=1

        lines_s = 2
```

```
        lines_ns = 2

        for row in range (2, counter+1):
                nr_compro = grouped.cell(row=row, column=5).value
                nr_compro = nr_compro.replace(" -", " ")
                nr_compro = nr_compro.replace("- ", " ")
                nr_compro = nr_compro.replace("+", " ") #ADDED
                nr_compro = nr_compro.replace(" ", "-")
                nr_comprobation = nr_compro.split("-")

                comprobator=[]
                for i in nr_comprobation:
                        comprobator.append(i.isalpha())

                comprobator=str(comprobator)
                if comprobator.find("True") != -1 or
len(nr_comprobation)>2:
                        _=grouped.cell(row=row, column=16,
value='{0}'.format("y"))
                        for c in range(1, 17):
                                _=grouped_special.cell(row=lines_s,
column=c, value='{0}'.format(grouped.cell(row=row, column=c).value))
                        lines_s+=1
                else:
                        _=grouped.cell(row=row, column=16,
value='{0}'.format("n"))
                        for c in range(1, 17):
                                _=grouped_nonspecial.cell(row=lines_ns,
column=c, value='{0}'.format(grouped.cell(row=row, column=c).value))
                        lines_ns+=1

        #Guardar cambios
        doc.save(str(filename)
```

Code 5. Function *Add sheets*

```
#Module to add the sheets: Trondheim, Nongrouped (add_firstsheets) and
Grouped_nospecial (add_grouped)

def add_sheets(filename, sheet):
        from openpyxl import load_workbook
        doc = load_workbook(filename=str(filename))
        rawdata=doc[str(sheet)]
        #Create sheets
        trondheim=doc.create_sheet("Trondheim")


        #Obtain number of records in Sheet1
        counter = 0
        for row in rawdata:
                counter+=1

        #Add titles to columns
        for c in range(1, 10):
                if c<3:
                        _=trondheim.cell(row=1, column=c,
value='{0}'.format(rawdata.cell(row=1, column=c).value))
                else:
                        _=trondheim.cell(row=1, column=c+6,
value='{0}'.format(rawdata.cell(row=1, column=c).value))

        trondheim.cell(row=1, column=3, value='{0}'.format("adresse"))
```

```
        trondheim.cell(row=1, column=4,
value='{0}'.format("adressenavn"))
        trondheim.cell(row=1, column=5, value='{0}'.format("nr"))
        trondheim.cell(row=1, column=6, value='{0}'.format("postnavn"))
        trondheim.cell(row=1, column=7, value='{0}'.format("postnr"))
        trondheim.cell(row=1, column=8, value='{0}'.format("bokstav"))

        #Create copy sheets: Non_grouped, Grouped, Grouped_special,
Grouped_nonspecial
        nongrouped=doc.copy_worksheet(trondheim)
        nongrouped.title="Non_grouped"
        grouped=doc.copy_worksheet(trondheim)
        grouped.title="Grouped"
        grouped.cell(row=1, column=16, value='{0}'.format("special
(y/n)"))
        grouped_special=doc.copy_worksheet(grouped)
        grouped_special.title="Grouped_special"
        grouped_nonspecial=doc.copy_worksheet(grouped)
        grouped_nonspecial.title="Grouped_nonspecial"
        grouped_nonspecial.cell(row=1, column=17,
value='{0}'.format("new_nr"))

        #Go record by record in Sheet 1 to cheek they are in Trondheim
        lines=2 # indicates the line number to add the information in the
sheet
        for row in range(2, counter+1):

                if rawdata.cell(row=row, column=3).value == "Trondheim ":
                        for c in range(1, 10):
                                if c<3:
                                        _=trondheim.cell(row=lines,
column=c, value='{0}'.format(rawdata.cell(row=row, column=c).value))
                                else:
                                        _=trondheim.cell(row=lines,
column=c+6, value='{0}'.format(rawdata.cell(row=row, column=c).value))

                        lines+=1
                else:
                        continue

        #Save changes
        doc.save(str(filename)
```

Code 6. Script Connect Statkraft with building database

```
import arcpy
from package.conectstatkrafttables import conect_statkraft_tables
from package.bruksarealcalculation import bruksareal_calculation
from package.Import_Table_to_Excel import import_to_excel
from arcpy import env

env.workspace = "data\ArcGIS\Geodatabase_MasterThesis.gdb"
env.overwriteOutput = True

# Paths
out_path = "data\ArcGIS\Geodatabase_MasterThesis.gdb"
statkraft_table =
r"data\ArcGIS\Geodatabase_MasterThesis.gdb\statkraft_nongrouped_table"
excel_doc = "data\Statkraft_Varme\excel_files\Fjernvarme_raw_copy.xlsx"

# Layers
unit_points = "Unit_Point"
```

```
buildings_trondheim = "Buildings_Trondheim"
statkraft_nongrouped = "statkraft_nongrouped_table"
statkraft_nongrouped_extra = "statkraft_nongrouped_table_extra"
statkraft_grouped = "statkraft_grouped_table"
statkraft_grouped_extra = "statkraft_grouped_table_extra"

# Fields
join_field = "adresse_text"
join_field_extra= "adresse_text_1"

#---------------------------------------PART 1---------------------------
-----------------------------------------------
# Create table from Excel file
arcpy.ExcelToTable_conversion(excel_doc, statkraft_nongrouped,
"Non_grouped")

# Make in both the table and unit_points the join field
arcpy.AddField_management(statkraft_nongrouped,join_field, "TEXT")
arcpy.CalculateField_management(statkraft_nongrouped, join_field,
"!adresse!.lower()", "PYTHON3")

arcpy.AddField_management(unit_points,join_field, "TEXT")
expression = "GetAddress(!adressenavn!, !nr!)"
codeblock = """
def GetAddress(adresse,nr):
    final_adresse= "{0} {1}".format(adresse.lower(), nr)
    return final_adresse"""
arcpy.CalculateField_management(unit_points, join_field, expression,
"PYTHON3", codeblock)

#Execute Join Field to add information from the building database to the
records
adresseid = 'adresseid'
arcpy.JoinField_management(statkraft_nongrouped, join_field, unit_points,
join_field, [adresseid])

#---------------------------------------PART 2---------------------------
-----------------------------------------------

#Añadir aquellos con bokstav
arcpy.TableToTable_conversion(statkraft_nongrouped, out_path,
statkraft_nongrouped_extra, """ adresseid IS NULL """)

#Eliminate fields to be able to add the new ones
arcpy.DeleteField_management(statkraft_nongrouped_extra, [adresseid])

#Make in both the table and unit_points a field with the adresse all in
minusculas para conectarlo
arcpy.AddField_management(statkraft_nongrouped_extra,join_field_extra,
"TEXT")
arcpy.CalculateField_management(statkraft_nongrouped_extra,
join_field_extra, "!adresse_text!.replace(' ','')", "PYTHON3")

arcpy.AddField_management(unit_points,join_field_extra, "TEXT")
expression = "GetAddress(!adressenavn!, !nr!, !bokstav!)"
codeblock = """
def GetAddress(adresse,nr,letter):
    final_adresse= "{0}{1}{2}".format(adresse.lower(), nr,
letter.lower())
    return final_adresse"""
```

```
arcpy.CalculateField_management(unit_points, join_field_extra,
expression, "PYTHON3", codeblock)

#Execute Join Field to add information from the building database to the
records
arcpy.JoinField_management(statkraft_nongrouped_extra, join_field_extra,
unit_points, join_field_extra, [adresseid])

#---------------------------------------PART 3----------------------------
-----------------------------------------------
conect_statkraft_tables()

#---------------------------------------PART 4----------------------------
-----------------------------------------------
bruksareal_calculation()

#---------------------------------------PART 4----------------------------
-----------------------------------------------
import_to_excel()

## FOR GROUPED BUILDINGS
#---------------------------------------PART 1----------------------------
-----------------------------------------------
# Create table from Excel file
arcpy.ExcelToTable_conversion(excel_doc, statkraft_grouped,
"Grouped_nonspecial")

# Make in both the table and unit_points the join field
arcpy.AddField_management(statkraft_grouped,join_field, "TEXT")
arcpy.CalculateField_management(statkraft_grouped, join_field,
"!adressenavn!.lower()+ ' ' + !new_nr!", "PYTHON3")


#Execute Join Field to add information from the building database to the
records
adresseid = 'adresseid'

arcpy.JoinField_management(statkraft_grouped, join_field, unit_points,
join_field, [adresseid])

#---------------------------------------PART 4----------------------------
-----------------------------------------------
bruksareal_calculation2()

#---------------------------------------PART 5----------------------------
-----------------------------------------------
import_to_excel2()
```

Code 7. Function *Connect Statkraft tables*

```
def conect_statkraft_tables():
    import arcpy
    from arcpy import env

    env.workspace =
"F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb"
    env.overwriteOutput = True

    # Paths
    out_path =
"F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb"
```

```
    statkraft_table =
r"F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb\statkraf
t_nongrouped_table"

    # Layers
    statkraft_nongrouped = "statkraft_nongrouped_table"
    statkraft_nongrouped_extra = "statkraft_nongrouped_table_extra"
    join_field = "MålepunktID"

    # Join both tables
    adresseid = 'adresseid'
    arcpy.JoinField_management(statkraft_nongrouped, join_field,
statkraft_nongrouped_extra, join_field, [adresseid])

    #Calculate Field in statkraft_nongrouped

    with arcpy.da.UpdateCursor(statkraft_nongrouped, [adresseid,
"adresseid_1"]) as cursor:
        for row in cursor:
            if str(row[0]).isalpha():
                row[0]=row[1]
                cursor.updateRow(row)

    # Delete the adresseid field joined
    arcpy.DeleteField_management(statkraft_nongrouped, ["adresseid_1"])
```

Code 8. Function *Total useful floor area calculation*

```
def bruksareal_calculation():
    import arcpy
    from arcpy import env

    env.workspace = r"###.gdb" # Geodatabase containing building
information
    env.overwriteOutput = True

    # Layers
    unit_points = "Unit_Point"
    statkraft_nongrouped = "statkraft_nongrouped_table"

    # Introduce new field in statkraft_nongrouped
    arcpy.AddField_management(statkraft_nongrouped, "sum_bruksareal",
"DOUBLE",field_scale=0)
    arcpy.AddField_management(statkraft_nongrouped, "bruksarealtotalt",
"DOUBLE", field_scale=0)
    arcpy.AddField_management(statkraft_nongrouped, "bygningstypekode",
"FLOAT")
    arcpy.AddField_management(statkraft_nongrouped, "type", "Text")
    arcpy.AddField_management(statkraft_nongrouped, "year", "FLOAT")

    # Import total area and type code from Building_Trondheim
    with arcpy.da.UpdateCursor(statkraft_nongrouped,
["adresseid","sum_bruksareal","bruksarealtotalt","bygningstypekode",
"type", 'year']) as cursor_statkraft:
        for s_x in cursor_statkraft:
            byggid = []
            bruksareal=0
            bruksarealtotalt=0
            type_code=0
            type=""
            year=0
```

```
            with arcpy.da.SearchCursor(unit_points, ["adresseid",
"byggid", "bruksareal" ,"bruksarealtotalt", "bygningstypekode", "TYPE",
'YEAR']) as cursor_units:
                for u_x in cursor_units:
                    if u_x[0]==s_x[0]:
                        byggid.append(u_x[1])
                        bruksareal=bruksareal + u_x[2]
                        bruksarealtotalt = u_x[3]
                        type_code = u_x[4]
                        type=u_x[5]
                        year=u_x[6]
                s_x[1]=bruksareal
                s_x[2]=bruksarealtotalt
                s_x[3]=type_code
                s_x[4]=type
                s_x[5]=year

                print("For addreseid = {0}, we have {1} "
                      " numbers of id and the area can be bruksareal ={2}
or bruksarealtotalt={3} and type {4}".format(s_x[0], byggid,bruksareal,
bruksarealtotalt, type))
                cursor_statkraft.updateRow(s_x)
```

Code 9. Function *Import to Excel*

```
def import_to_excel():
    import arcpy
    from openpyxl import load_workbook

    doc = load_workbook(filename="###.xlsx") # Excel file to create
    hoja = doc["Nongrouped_Linked"]

    arcpy.env.workspace = r"###.gdb"
    arcpy.env.overwriteOutput = True

    #Import tables and select all fields
    table = r"###\statkraft_nongrouped_table"
    field_list = arcpy.ListFields(table)
    list_fields = []
    for x in field_list:
        list_fields.append(x.name)


    #Copy features in ExcelFile
    row=2
    with arcpy.da.SearchCursor(table, list_fields) as cursor:
        for x in cursor:
            for i in range(1,len(list_fields)+1):
                _ = hoja.cell(column=i, row=row, value='{0}'.format(x[i-
1]))
            row+=1

    doc.save("###.xlsx")
```

Functions *Total useful area calculations 2* and *Import to Excel 2* are similar to Code 8 and Code 9. The only difference between them is that the ones shown above are focus on the non-grouped buildings.

Section E2. TronderEnergi pre-processing model

For the TronderEnergi database, only three tools were developed. They are more straightforward to be used as the data provided was easier to link with the building stock database. These tools are *Cleaning*, *Join with buildings* and *Add total consumption*.

First, we will run the tool *Cleaning* to remove the records that are not part of Trondheim Municipality (See Code 10). Then we add the script manually into a feature class (in the codes called TronderEnergi_2018) in ArcGIS to proceed with the rest of the model.

Code 10. Tool *Cleaning*

```
import pandas as pd
import numpy as np

#Import database from Tronder Energy
TronderEnergi =
pd.read_csv('F:\MASTER_THESIS\Data\TronderEnergi\TEnergi.csv', sep=';')
df = TronderEnergi[TronderEnergi['AAR']==2018] #Only year 2018

#Delete unnecesary postnr. We should only have 73 not 76
df = df[(df['POSTNR'] != 7003)]
df = df[(df['POSTNR'] != 7459)]
df = df[(df['POSTNR'] != 7462)]
df = df[(df['POSTNR'] != 7540)]

#Introduce the categories to eliminate
cat = pd.read_csv('F:\MASTER_THESIS\Data\TronderEnergi\Categories.csv',
sep=';')
cat_other = pd.Series(cat['Others'])

#Clean the database
df = df[~df['BESKRIVELSE'].isin(cat_other)]

#Sum the energy intensities for each post zone
df_postnr = df.groupby('POSTNR', as_index=False).sum()
```

Code 11. Tool *Join with buildings*

```
#import packages
import arcpy
from arcpy import env
import numpy as np

env.workspace =
"F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb"
env.overwriteOutput = True

#Layers
unit_point =
r'F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb\Unit_Poi
nt'
tronder_energi =
r'F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb\TronderE
nergi_2018'
buildings =
r'F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb\Building
s_Trondheim'

####Union layer
direccion = 'direccion'
```

```
#Create layer for union
print("Empezando el programa")
print("Añadiendo field: direccion")
#arcpy.AddField_management(tronder_energi, direccion, "TEXT")
#arcpy.AddField_management(unit_point, direccion, "TEXT")

    # tronderEnergi layer
print("Calculating direccion field for tronderEnergi layer")
#arcpy.CalculateField_management(tronder_energi, direccion, "(!GATENAVN!
+ str(!HUSNR!)).lower()")

    # unit_point layer
print("Calculating direccion field for unit_point layer")
#arcpy.CalculateField_management(unit_point, direccion, "(!adressenavn! +
str(!nr!)).lower()")

######Add interesting fields to TronderEnergi
fields = ["bygningsnr","bruksarealtotalt", "TYPE", "YEAR"]
types = ["LONG","LONG", "TEXT", "FLOAT"]
#for i in range(0,len(fields)):
    #arcpy.AddField_management(tronder_energi, fields[i], types[i])


#### Create UpdateCursors for joining Variables

with arcpy.da.UpdateCursor(tronder_energi, [direccion,
"bruksarealtotalt", "TYPE", "YEAR", "OBJECTID", "bygningsnr"]) as
TE_cursor:
    for row_te in TE_cursor:
        print("Modificando record number {}".format(row_te[4]))
        bruk = 0
        Type = ""
        Year = 0
        Byg = 0
        with arcpy.da.SearchCursor(unit_point, ["bygningsnr"],
where_clause=""" direccion = '{}' """.format(row_te[0])) as UP_cursor:
            for row_up in UP_cursor:
                with arcpy.da.SearchCursor(buildings,
["bruksarealtotalt", "TYPE", "YEAR", "bygningsnr"], where_clause = """
bygningsnr = {} """.format(row_up[0])) as b_cursor:
                    for row_b in b_cursor:
                        bruk = row_b[0]
                        Type = row_b[1]
                        Year = row_b[2]
                        Byg = row_b[3]
                        break
                break
        row_te[1] = bruk
        row_te[2] = Type
        row_te[3] = Year
        row_te[5] = Byg
        print("Modificado record number {} \n".format(row_te[4]))
        TE_cursor.updateRow(row_te)
```

Code 12. Tool *Add total consumption*

```
#import packages
import arcpy
from arcpy import env
import numpy as np
```

```
env.workspace =
"F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb"
env.overwriteOutput = True

#Layers
unit_point =
r'F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb\Unit_Poi
nt'
tronder_energi =
r'F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb\TronderE
nergi_2018'
buildings =
r'F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb\Building
s_Trondheim'

#1. Create TE_2018 copying tronderEnergi_2018
#arcpy.Copy_management(tronder_energi, "TE_2018")
te_2018 =
r'F:\MASTER_THESIS\Data\Data_ArcGIS\Geodatabase_MasterThesis.gdb\TE_2018'

#2. Calculate AARSFORBRUK_sum
#arcpy.AddField_management(te_2018, "AARSFORBRUK_sum", "INTEGER")

def unique_values(table , field):
    with arcpy.da.SearchCursor(table, [field]) as cursor:
        return sorted({row[0] for row in cursor)

myValues = unique_values(te_2018 , 'bygningsnr')
print(len(myValues))
print("{}".format(myValues[29265]))

for i in range(myValues.index(182407194), len(myValues)):
    print("Bygnr = {}".format(myValues[i]))
    print("Number ={}".format(i))
    arcpy.MakeTableView_management(te_2018,'prueba_v',where_clause= '
bygningsnr = {} '.format(myValues[i]))
    someValue = sum([r[0] for r in
arcpy.da.SearchCursor("prueba_v",["AARSFORBRUK"])])
    arcpy.CalculateField_management('prueba_v',
"AARSFORBRUK_sum",someValue)
    arcpy.Delete_management('prueba_v')
```

Section E3. Merging model

The building stock-energy database is formed by merging the heat consumption, electricity consumption and the building georeferenced dataset. Code 13 shows the script used for this process.

Code 13. Script *Merging databases*

```
# import packages

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import os
```

```
# ### Functions
# Create cohorts according to TEK
TEK = ['Eldre', 'TEK49', 'TEK69', 'TEK87', 'TEK97', 'TEK07', 'TEK10',
'TEK17']
TEK_years = ['-1949', '1950-1968', '1969-1986', '1987-1996', '1997-2006',
'2007-2009', '2010-2016', '2017-2018']

def clasify_years(df):
    df['TEK'] = pd.Series('Nan', index = df.index)

    df.loc[(df['year'] <= 1949), 'TEK'] = TEK[0]
    df.loc[(df['year'] >= 1950) & (df['year'] <= 1968), 'TEK'] = TEK[1]
    df.loc[(df['year'] >= 1969) & (df['year'] <= 1986), 'TEK'] = TEK[2]
    df.loc[(df['year'] >= 1987) & (df['year'] <= 1996), 'TEK'] = TEK[3]
    df.loc[(df['year'] >= 1997) & (df['year'] <= 2006), 'TEK'] = TEK[4]
    df.loc[(df['year'] >= 2007) & (df['year'] <= 2009), 'TEK'] = TEK[5]
    df.loc[(df['year'] >= 2010) & (df['year'] <= 2016), 'TEK'] = TEK[6]
    df.loc[(df['year'] >= 2017) & (df['year'] <= 2018), 'TEK'] = TEK[7]


def clasify_typologies(df):
    df['type'] = pd.Series('Nan', index = df.index)

    df.loc[(df['bygningstypekode'] <= 113), 'type'] = 'SFH'
    df.loc[(df['bygningstypekode'] <= 136) & (df['bygningstypekode'] >=
121), 'type'] = 'TH'
    df.loc[(df['bygningstypekode'] <= 159) & (df['bygningstypekode'] >=
141), 'type'] = 'MFH'
    df.loc[(df['bygningstypekode'] <= 319) & (df['bygningstypekode'] >=
311), 'type'] = 'Office'
    df.loc[(df['bygningstypekode'] <= 330) & (df['bygningstypekode'] >=
321), 'type'] = 'Business'
    df.loc[(df['bygningstypekode'] <= 249) & (df['bygningstypekode'] >=
211), 'type'] = 'Industry'
    df.loc[(df['bygningstypekode'] <= 449) & (df['bygningstypekode'] >=
411), 'type'] = 'Communication'
    df.loc[(df['bygningstypekode'] <= 539) & (df['bygningstypekode'] >=
511), 'type'] = 'Service'
    df.loc[(df['bygningstypekode'] <= 629) & (df['bygningstypekode'] >=
611), 'type'] = 'Education'
    df.loc[(df['bygningstypekode'] <= 679) & (df['bygningstypekode'] >=
641), 'type'] = 'Cultural_Sport'
    df.loc[(df['bygningstypekode'] <= 739) & (df['bygningstypekode'] >=
719), 'type'] = 'Health'
    df.loc[(df['bygningstypekode'] <= 840) & (df['bygningstypekode'] >=
821), 'type'] = 'Emergency'


# ### Datasets

# #### Statkraft

# Statkraft_varme import
# loading the data

DATA_DIR = '../data'
FILE_NAME = 'Fjernvarme.xlsx'
data_path = os.path.join(DATA_DIR,FILE_NAME)
statkraft = pd.read_excel(data_path, sheet_name='Non_residential',
index_col=0)
```

```
# Statkraft_varme processing

statkraft = statkraft[statkraft['bruksareal'] != 0] #Delete records with
use area = 0
statkraft = statkraft[statkraft['energy_intensity']>20] #Delete records
with energy_intensity lower than 20 as they are supposed to have errors
or not be representative of the building type
statkraft =
statkraft[(statkraft['energy_intensity']>statkraft['energy_intensity'].qu
antile(0.05)) &
(statkraft['energy_intensity']<statkraft['energy_intensity'].quantile(0.9
5))] #Percentiles
statkraft = statkraft[(statkraft['type']!='Communication') &
(statkraft['type']!='Emergency')] #Delete emergency and communication
buildings
clasify_years(statkraft) # Add cohorts by TEK categories


# #### Buildings
# Buildings import
# loading the data

DATA_DIR = '../data'
FILE_NAME = 'Building_info.csv'
data_path = os.path.join(DATA_DIR,FILE_NAME)
buildings_t = pd.read_csv(data_path, sep=';', decimal = ',',
index_col='OBJECTID')

# Buildings processing

buildings_t = buildings_t.drop(['lopenr',
'vannforsyningskode','bebygdarealkilde',
'ufullstendigareal','vannforsyning', 'avlopskode', 'avlop', 'harheis',
'opprinnelseskode', 'opprinnelse', 'fkbareal', 'alternativtareal',
'alternativtareal2', 'fylkeid', 'omradeid', 'kommunenr','harsefrakminne',
'p25statuskode','p25status', 'verifisert',
'harkulturminne','utenbebygdareal','antallreg', 'antallferdig',
'antalletasjer', 'antallhovedetasjer', 'antallkjelleretasjer',
'antallloft', 'antallunderetasjer', 'kommunenavn', 'fylkesnavn',
'geometrikilde'], axis = 1)
buildings_t.rename(columns=lambda x: x.lower(), inplace=True) #Convert
the labels to lower case
buildings_t['residential'] = (buildings_t['naringsgruppekode'] ==
'X').astype('int') # create residential column
buildings = buildings_t[buildings_t['naringsgruppekode'] != 'A']

clasify_typologies(buildings)
clasify_years(buildings)

cat_drop = [239, 249, 216, 241, 231, 245, 411, 415, 416, 419, 431, 439,
441, 449, 524, 522]
for i in cat_drop: # Delete records that belongs to the cat_drop
categories
    buildings = buildings[buildings['bygningstypekode'] != i]

# #### TronderEnergi
# TronderEnergi import
# loading the data

DATA_DIR = '../data'
```

```
FILE_NAME = 'TronderEnergi_2018_bygningsnr.csv'
data_path = os.path.join(DATA_DIR,FILE_NAME)
te = pd.read_csv(data_path, sep=';', decimal = ',', index_col='OBJECTID')

# TronderEnergi processing
te = te[te['YEAR'] != 0]
te.rename(columns=lambda x: x.lower(), inplace=True) #Convert the labels
to lower case


# ### Merging databases

# #### Statkraft+Unit Points

# import unit points database to join the addresseid with bygningsnr
# loading the data

DATA_DIR = '../data'
FILE_NAME = 'unit.csv'
data_path = os.path.join(DATA_DIR,FILE_NAME)
unit_points = pd.read_csv(data_path, sep=';', decimal = ',')

unit_points.rename(columns=lambda x: x.lower(), inplace=True) #Convert
the labels to lower case

# Look for the adresseid from statkraft in the unit_points layer
statkraft = pd.merge(statkraft, unit_points, on = 'adresseid')
statkraft = statkraft.drop(['xcoord', 'ycoord'], axis=1)
#statkraft = statkraft.drop_duplicates('adresseid', keep='first')
statkraft.head()

# Statkraft + Buildingdatabase
# Caution! The Statkraft database has been processed before, thus
although we have 1000 records, it is referred to dwellings and we only
need buidings. The energy intensities are calcu
# calculated for the complete building (various records with the same
energy intensity and same adresseid). We just need to

bygnr_list = statkraft['bygningsnr'].unique().tolist() # We introduce the
bygningsnrs in a list
heat = []
for i in bygnr_list:
    df_2 = statkraft[statkraft['bygningsnr'] == i]
    ht = df_2['energy_intensity'].mean()
    heat.append(ht)

dictionary = {'bygningsnr': bygnr_list, 'heat': heat}
bygnr_heat = pd.DataFrame.from_dict(dictionary)

# TronderEnergi + Buildingdatabase
bygnr_list = te['bygningsnr'].unique().tolist() # We introduce the
bygningsnrs in a list
electr = []
for i in bygnr_list:
    df = te[te['bygningsnr'] == i]
    elec = df['aarsforbruk'].sum()
    electr.append(elec)
dictionary = {'bygningsnr': bygnr_list, 'electr': electr}
bygnr_elect = pd.DataFrame.from_dict(dictionary)

## Merge the three databases
```

```
df = pd.merge(buildings, bygnr_elect, how='left', on='bygningsnr')
df = pd.merge(df, bygnr_heat, how='left', on='bygningsnr')

# Create electricity_intensity column
df['electricity_intensity'] = df['electr']/df['bruksarealtotalt'] # add
electricity_intensity column
df.loc[df['electricity_intensity'] == np.inf, 'electricity_intensity'] =
df['electr']/df['bebygdareal']
#df = df[(df['electricity_intensity'] <
df['electricity_intensity'].quantile(0.95)) &
(df['electricity_intensity'] >
df['electricity_intensity'].quantile(0.05))]

# Final dataFrame
df['byggid']=df.byggid.astype(int)

# ### Export it to csv
# Export to the folder data

DATA_DIR = '../data'
FILE_NAME = 'Energy_Heat_Database.csv'
data_path = os.path.join(DATA_DIR,FILE_NAME)
export_csv = df.to_csv (data_path, sep=';', index = None, header=True,
decimal =',')
```

Section E4. Visualization model

Code 14. Add energy intensities to buildings

```
# -*- coding: utf-8 -*-
"""
Add fields for energy consumption in units_dh and units_nodh. Before running this script is necessary to
add the fields in the feature layers.
"""

import arcpy
from arcpy import env
import numpy as np
import pandas as pd

env.workspace = "...\data\ArcGIS\Geodatabase_MasterThesis.gdb"
env.overwriteOutput = True

# import data
# data for electricity intensities
general = pd.read_excel(r'excel\energy_intensities.xlsx', sheet_name='elect_general')
no_dh = pd.read_excel(r'excel\energy_intensities.xlsx', sheet_name='elect_EE')
dh = pd.read_excel(r'excel\energy_intensities.xlsx', sheet_name='elect_EH')
heat = pd.read_excel(r'excel\energy_intensities.xlsx', sheet_name='heat_general')

# feature layers from ArcGIS
statkraft = 'units_dh'
```

```
    points = 'units_nodh'

    # add values to each row

    with arcpy.da.UpdateCursor(points, ['TYPE', 'COHORT', 'm_heat', 'f_heat', 't_heat']) as cursor:
        for row in cursor:
            m_value, f_value, t_value = 0, 0, 0
            m_value = heat.loc[(heat['Type']== row[0]) & (heat['Cohort']== row[1]), ['Median']].iat[0,0] # get median value
            f_value = heat.loc[(heat['Type']== row[0]) & (heat['Cohort']== row[1]), ['1st interquantile']].iat[0,0] # get first interquartile
            t_value = heat.loc[(heat['Type']== row[0]) & (heat['Cohort']== row[1]), ['3rd interquantile']].iat[0,0] # get third interquartile

            row[2], row[3], row[4] = m_value, f_value, t_value

            # update cursor
            cursor.updateRow(row)
```

Code 15. Energy aggregation in areas

```
# -*- coding: utf-8 -*-
"""
Add fields for energy consumption in units_dh and units_nodh. Before running this script is necessary to
add the fields in the feature layers.
"""

import arcpy
from arcpy import env
import numpy as np
import pandas as pd

env.workspace = "...\data\ArcGIS\Geodatabase_MasterThesis.gdb"
env.overwriteOutput = True

# import data
# data for electricity intensities
general = pd.read_excel(r'excel\energy_intensities.xlsx', sheet_name='elect_general')
no_dh = pd.read_excel(r'excel\energy_intensities.xlsx', sheet_name='elect_EE')
dh = pd.read_excel(r'excel\energy_intensities.xlsx', sheet_name='elect_EH')
heat = pd.read_excel(r'excel\energy_intensities.xlsx', sheet_name='heat_general')

# feature layers from ArcGIS
statkraft = 'units_dh'
points = 'units_nodh'

# add values to each row

with arcpy.da.UpdateCursor(points, ['TYPE', 'COHORT', 'm_heat', 'f_heat', 't_heat']) as cursor:
    for row in cursor:
```

```
        m_value, f_value, t_value = 0, 0, 0
        m_value = heat.loc[(heat['Type']== row[0]) & (heat['Cohort']==
row[1]), ['Median']].iat[0,0] # get median value
        f_value = heat.loc[(heat['Type']== row[0]) & (heat['Cohort']==
row[1]), ['1st interquantile']].iat[0,0] # get first interquartile
        t_value = heat.loc[(heat['Type']== row[0]) & (heat['Cohort']==
row[1]), ['3rd interquantile']].iat[0,0] # get third interquartile

        row[2], row[3], row[4] = m_value, f_value, t_value

        # update cursor
        cursor.updateRow(row)
```

References of the supplementary material

Choudhary, R. (2012). Energy analysis of the non-domestic building stock of Greater London. *Building and Environment, 51*, 243-254. doi:https://doi.org/10.1016/j.buildenv.2011.10.006

Fuentes, A. (2018). Hands-On Predicitve Analytics with Python: Packt Publishing.

Geodata. (2018). *Eiendomsdata*. Retrieved from

Levine, N. (2017). CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents. In S. Shekhar, H. Xiong, & X. Zhou (Eds.), *Encyclopedia of GIS* (pp. 381-388). Cham: Springer International Publishing.

Sandberg. (2019). Paper under preparation.

Standards Norway. (2013). Klassifikasjon av byggverk. Del 3: Bygningstyper. In (Vol. NS 3457-3:2013). Norway.

Statens Kartverk. (2006). *SOSI standard - generell objektkatalog. Fagområde: Bygninger*. Retrieved from https://www.kartverket.no/globalassets/standard/sosi-standarden-del-1-og-2/sosi-standarden/bygg.pdf

Statens Kartverk. (2014). SOSI Generell objecktkatalog Bygning. Versjon 4.5. In.

Stavset, O., & Kauko, H. (2015). *Energy use in non-residential buildings - possibilities for smart energy solutions*. Retrieved from Norway: https://www.sintef.no/globalassets/sintef-energi/interact/tr-a7510-energy-use-in-non-residential-buildings--possibilities-for-smart-energy-solutions-2015.pdf

Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, environment and urban systems, 32*(5), 396-406.