



**Politecnico  
di Torino**

Master's degree program in  
Territorial, Urban, Environmental and Landscape Planning  
Curriculum: Planning for the Global Urban Agenda

Master's Thesis

**Data-driven urban building energy modeling in Satom (CH):  
The energy savings potential and use of available renewable  
energy sources.**

Supervisor

Prof. Guglielmina Mutani

Candidate

Ahad Montazeri

Co-supervisors

Prof. Jérôme H. Kämpf

Academic Year 2022/2023

Thesis Supervisor: Associate Professor Guglielmina Mutani

## Abstract

The thesis delves deeply into innovative methodologies aimed at enriching our comprehension of urban building energy dynamics. By merging the principles of Urban Building Energy Modeling (UBEM) with the strength of Machine Learning (ML) techniques, the study achieves substantial advancements in evaluating potential energy savings and harnessing renewable resources within Satom.

The research journey involves developing a sturdy building model, employing Geographic Information System (GIS) software to enhance modeling precision and data aggregation. Furthermore, by incorporating state-of-the-art ML algorithms like LightGBM and Random Forest through a bottom-up strategy, the study offers accurate forecasts of energy patterns and effective renewable energy utilization.

Additionally, this study pioneers a forward-looking trajectory spanning three decades, meticulously assessing the energy-saving potential of buildings. This initiative intricately weaves together physical attributes, energy efficiency, and socio-economic context. By designing tailored renovation scenarios and implementing meticulous selection processes, the study identifies buildings suitable for integration into the District Heating Network (DHN). This iterative approach systematically optimizes the network's capacity, encapsulating a pioneering strategy that harmonizes innovation, environmental concerns, and infrastructural enhancement.

In sum, the current research underscores the pivotal role of Data-driven techniques in refining Energy Demand (ED) and offers insights to enhance Energy Efficiency (EE) and nurture a greener and more sustainable urban future.

**Keywords:** *Energy efficiency model, Urban building energy modeling, Data-driven models, Machine learning, Place-based approach, Geographic Information System (GIS)*

## Table of contents

|   |      |
|---|------|
| Abstract.....   | ii   |
| Acknowledgment.....                                   | v    |
| List of figures .....                                 | vi   |
| List of tables.....                                   | viii |
| Glossary of terms.....                                | ix   |
| 1. INTRODUCTION .....                                 | 1    |
| 2. LITERATURE REVIEW .....                            | 3    |
| 2.1. Introduction.....                                | 3    |
| 2.2. District heating network's history .....         | 3    |
| 2.3. Technical design of DHN.....                     | 6    |
| 2.4. District heating regulatory policy.....          | 10   |
| 2.4.1. Denmark .....                                  | 12   |
| 2.4.2. Switzerland .....                              | 13   |
| 2.4.3. Italy .....                                    | 13   |
| 2.5. Energy prediction models.....                    | 15   |
| 2.5.1. UBEM, USEM, and USBEM models.....              | 16   |
| 2.5.2. Top-Down and Bottom-Up models.....             | 20   |
| 2.5.3. Black-box, White-box, and Grey-box models..... | 22   |
| 2.6. Researches in the literature.....                | 25   |
| 3. METHODOLOGY.....                                   | 31   |
| 3.1. Introduction.....                                | 31   |
| 3.2. Urban building energy modeling.....              | 31   |
| 3.3. Black-box methodology .....                      | 36   |
| 3.3.1. Random Forest (RF) regression .....            | 36   |
| 3.3.2. Genetic Algorithm (GA).....                    | 38   |

|  |     |
|--|-----|
| 3.3.3. Cross Validation (CV) .....                     | 39  |
| 3.3.4. Light Gradient Boosting Machine (LightGBM)..... | 42  |
| 3.3.5. Optuna .....                                    | 43  |
| 3.4. Energy saving scenario .....                      | 44  |
| 4. DISCUSSION.....                                     | 48  |
| 4.1. Introduction.....                                 | 48  |
| 4.2. Case study .....                                  | 48  |
| 4.3. Buildings' physical model .....                   | 53  |
| 4.4. Black-box method.....                             | 59  |
| 4.4.1. Dataset preparation .....                       | 59  |
| 4.4.2. Tuning hyperparameters.....                     | 62  |
| 4.4.3. Annual heating demand processing.....           | 65  |
| 4.5. DHN's optimal expansion .....                     | 72  |
| 5. CONCLUSION.....                                     | 86  |
| 5.1. Limitations of the Study .....                    | 87  |
| 5.2. Recommendation for future research .....          | 89  |
| 6. REFERENCES .....                                    | 90  |
| APPENDIX A .....                                       | 100 |
| APPENDIX B .....                                       | 102 |
| APPENDIX C .....                                       | 104 |
| APPENDIX D .....                                       | 107 |
| APPENDIX E .....                                       | 109 |

## Acknowledgment

I am profoundly grateful for the completion of this master's thesis, and I wish to express my heartfelt thanks to those who have supported and inspired me throughout this remarkable journey.

First and foremost, I extend my deepest gratitude to my thesis supervisors, Professor Guglielmina Mutani and Professor Jérôme Henri Kämpf, for their unwavering guidance, patience, and insightful feedback. Their mentorship has been invaluable in shaping my research and enhancing my academic growth.

I am grateful to the faculty members of the Inter-university Department of Regional & Urban Studies and Planning (DIST) at Politecnico di Torino, whose expertise and dedication have enriched my understanding of the subject matter and provided me with a solid foundation for my research. I am also grateful to Idiap Research Institute for their support, which enabled me to carry out this research and pursue my academic aspirations.

My sincere appreciation goes to my family and friends for their constant encouragement and belief in my abilities. Your support has been my driving force and a source of strength, especially during the challenging moments.

Last but not least, I extend my appreciation to all those who might not be mentioned here but have played a part in my academic journey. Your influence, no matter how small, has contributed to my growth and accomplishments. Completing this thesis has been a challenging yet rewarding endeavor, and I am humbled by the collective effort that has gone into making it a reality. Thank you, from the bottom of my heart.

I genuinely hope that this endeavor plays a role in fostering a more sustainable future, igniting a constructive transformation within the domain.

Ahad Montazeri

September 2023

## List of figures

|   |    |
|---|----|
| <b>Figure 2.1.</b> Four generations of DH systems.....  | 5  |
| <b>Figure 2.2.</b> main components of a DH network: (A) standard, (B) advanced. (FF: Fossil Fuels, HS: Heat Source, RES: Renewable Energy Sources, TES: Thermal Energy Storage) .....   | 7  |
| <b>Figure 2.3.</b> Example of basic layout for an indirect (left) and a direct (right) space heating substation. (A) Controller (B) Temperature sensors (C) Actuator and control valve (D) Differential pressure controller (E) Non-return valve..... | 8  |
| <b>Figure 2.4.</b> Example of basic layout a DHW exchanger (left) and a DHW storage tank (right). (A) Controller (B) Temperature sensor (C) Actuator and control valve (D) Differential pressure controller (E) Bypass valve                          | 9  |
| <b>Figure 3.1.</b> Flowchart of the applied methodology: Energy Efficiency model utilizing an Urban Building Energy Model as its foundation .....   | 33 |
| <b>Figure 3.2.</b> Schematic illustration of how CV technique is being employed in ML algorithms .....  | 41 |
| <b>Figure 3.3.</b> Minergie modernization system for energy retrofitting of residential buildings .....   | 45 |
| <b>Figure 4.1.</b> Classification of land use for the sampled buildings. ....   | 50 |
| <b>Figure 4.2.</b> Construction period categories for the sampled buildings. ....   | 51 |
| <b>Figure 4.3.</b> Spatial distribution of connected and connectable buildings.....   | 52 |
| <b>Figure 4.4.</b> Representation of buildings' 3D roof model.....  | 55 |
| <b>Figure 4.5.</b> (A) The distribution of "Residential" buildings by volume (B) The distribution of "Non-residential" buildings by volume .....  | 57 |
| <b>Figure 4.6.</b> Spatial distribution of building subsamples in the analysis.....   | 58 |
| <b>Figure 4.7.</b> Calibration of XML file building-wise according to measured data .....   | 60 |
| <b>Figure 4.8.</b> Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Residential Buildings with Occupancy Profile .....   | 65 |
| <b>Figure 4.9.</b> Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Residential Buildings without Occupancy Profile .....  | 66 |
| <b>Figure 4.10.</b> Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Non-residential Buildings with Occupancy Profile .....  | 67 |
| <b>Figure 4.11.</b> Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Non-residential Buildings without Occupancy Profile.....  | 68 |
| <b>Figure 4.12.</b> Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Multifunction Buildings with Occupancy Profile .....  | 69 |
| <b>Figure 4.13.</b> Simplified model of existing District Heating Network (DHN) .....   | 74 |
| <b>Figure 4.14.</b> Building density and energy intensity of buildings within each mesh .....   | 75 |
| <b>Figure 4.15.</b> Mesh prioritization for optimal DHN expansion.....  | 76 |
| <b>Figure 4.16.</b> Distribution of retrofitting technologies for connectable buildings considering socio-economic status of context.....   | 78 |

|  |    |
|--|----|
| <b>Figure 4.17.</b> Real connectable buildings considering different retrofitting scenarios .....              | 80 |
| <b>Figure 4.18.</b> Expansion roadmap of district heating network over next 30 years in 3-year intervals ..... | 81 |
| <b>Figure 4.19.</b> Optimal expansion pathway for District Heating Network to connect feasible buildings.....  | 83 |

## List of tables

|  |    |
|--|----|
| <b>Table 2.1.</b> Characteristics of top-down and bottom-up approaches.....  | 22 |
| <b>Table 2.2.</b> The advantages and disadvantages of the three major building energy prediction approaches  | 25 |
| <b>Table 3.1.</b> Available data sources in UBEM.....  | 35 |
| <b>Table 4.1.</b> Main input parameters in building of energy-use models.....  | 54 |
| <b>Table 4.2.</b> (A) The thermal conductivity of different building elements before retrofitting. (B) The thermal conductivity of different building elements after retrofitting..... | 61 |
| <b>Table 4.3.</b> Hyperparameters of LightGBM model.....   | 63 |
| <b>Table 4.4.</b> Hyperparameters of RF regression model .....   | 64 |
| <b>Table 4.5.</b> Comparing LightGBM and RF regression performance in predicting Urban-Scale heating demand with different renovation technologies .....                               | 70 |
| <b>Table 4.6.</b> Maximum real connectable buildings considering various energy-saving scenarios .....   | 79 |
| <b>Table 4.7.</b> The annual heating demand and peak demand of buildings connected after DHN expansion....   | 84 |

## Glossary of terms

|                       |                                     |
|-----------------------|-------------------------------------|
| <b>AI</b>             | Artificial Intelligence             |
| <b>ANN</b>            | Artificial Neural Network           |
| <b>BEM</b>            | Building Energy Models              |
| <b>C</b>              | Space Cooling                       |
| <b>CHP</b>            | Combined Heat and Power             |
| <b>CityGML</b>        | City Geography Markup Language      |
| <b>CO<sub>2</sub></b> | Carbon Dioxide                      |
| <b>CV</b>             | Cross validation                    |
| <b>DH</b>             | District Heating                    |
| <b>DHC</b>            | District Heating and Cooling        |
| <b>DHN</b>            | District Heating Network            |
| <b>DHW</b>            | Domestic Hot Water                  |
| <b>DUE-S</b>          | Data-driven Urban Energy Simulation |
| <b>EC</b>             | Energy Consumption                  |
| <b>ED</b>             | Energy Demand                       |
| <b>EE</b>             | Energy Efficiency                   |
| <b>EU</b>             | European Union                      |
| <b>GA</b>             | Genetic Algorithm                   |
| <b>GBDT</b>           | Gradient Boosting Decision Tree     |
| <b>GHG</b>            | Greenhouse Gas                      |
| <b>GIS</b>            | Geographic Information Systems      |
| <b>GWh/Y</b>          | Gigawatt hour per Annual            |
| <b>H</b>              | Space Heating                       |
| <b>HD</b>             | Heating Demand                      |
| <b>HVAC</b>           | Heating, Ventilation, and Cooling   |
| <b>IEA</b>            | International Energy Agency         |

|                            |  |
|----------------------------|--|
| <b>kWh/m<sup>3</sup>/Y</b> | Kilowatt hour per Cubic meter per Annual |
| <b>LiDAR</b>               | Light Detection and Ranging              |
| <b>LightGBM</b>            | Light Gradient Boosting Machine          |
| <b>LR</b>                  | Linear Regression                        |
| <b>MILP</b>                | Mixed-Integer Linear Programming         |
| <b>MINLP</b>               | Mixed-integer Nonlinear Programing       |
| <b>ML</b>                  | Machine Learning                         |
| <b>Mm<sup>3</sup></b>      | Mega cubic meter                         |
| <b>MST</b>                 | Minimum Spanning Tree                    |
| <b>PB</b>                  | Physical-based Models                    |
| <b>RES</b>                 | Renewable Energy Sources                 |
| <b>RF</b>                  | Random Forest                            |
| <b>RMSE</b>                | Root Mean Square Error                   |
| <b>RNN</b>                 | Recurrent Neural Network                 |
| <b>SVM</b>                 | Support Vector Machine                   |
| <b>S/V</b>                 | Surface-to-volume ratio                  |
| <b>TES</b>                 | Thermal Energy Storage                   |
| <b>TIN</b>                 | Triangular irregular networks            |
| <b>TMY</b>                 | Typical Meteorological Year              |
| <b>UBEM</b>                | Urban Building Energy Modeling           |
| <b>USBEM</b>               | Urban-Scale Building Energy Modeling     |
| <b>USEM</b>                | Urban-Scale Energy Model                 |
| <b>W/m<sup>2</sup>/k</b>   | watts per square meter per kelvin        |
| <b>XGBoost</b>             | Extreme Gradient Boosting                |
| <b>XML file</b>            | Extensible Markup Language file          |

# 1. INTRODUCTION

The relationship between sustainability and environmental concerns is closely intertwined with energy consumption and carbon dioxide (CO<sub>2</sub>) emissions, factors contributing to both global warming and air pollution (Ang, Choong & Ng, 2015). The impact of built environments on global climate change has a profound influence on the quality of urban living and overall livability. Urban activities serve as the primary sources of greenhouse gas (GHG) emissions. Notably, transportation and buildings emerge as pivotal contributors, collectively accounting for 75% of global CO<sub>2</sub> emissions (United Nations Environment Program, 2016).

An unprecedented drop of 5.4% in global CO<sub>2</sub> emissions in 2020 was experienced due to the COVID-19 pandemic. Though no reliable data was reported for GHG emissions in 2020, the speculations illustrated that reduction in CO<sub>2</sub> was more significant rather than GHG emissions. However, the reoccurrence of an increasing trend in emissions was anticipated for 2021 on (UNDP, 2021).

To be able to tackle energy crisis the basic formula is to enhance energy efficiency (EE) and to use renewable energy sources (RES) (Delmastro, Mutani, & Corgnati, 2016). Enhancing EE is crucial to reducing GHG emissions (UNDP, 2016). An optimal EE policy is built upon two main pillars: the reduction of energy use and the efficient exploitation of the available sources (Bordin, Gordini & Vigo, 2015). Moreover, Resource depletion and climate change caused by raising CO<sub>2</sub> levels have raised interest in RES (van der Zwan & Pothof, 2020).

From the above-mentioned viewpoint, District Heating Network (DHN) gains higher importance in contemporary cities to achieve sustainability and EE (Bordin, Gordini & Vigo, 2015). In many territories with cold and moderate temperatures, the need for space heating (H) and domestic hot water (DHW) in buildings still receives considerable attention as a sizable portion of the overall energy demand (Sperling & Möller, 2011). Around one-third of all main energy resources are consumed by buildings, making them a key area for the implementation of EE measures (Shaikh, Nor, Nallagownden, Elamvazuthi & Ibrahim, 2014).

Returning to the previously mentioned information, the objective of current thesis is to implement a methodology for optimizing the connection of additional buildings to the existing DH network. To achieve it, it was essential to build energy use model to be able to speculate the potential heating demand (HD) of potentially connectable buildings to the existing DH network, and investigate the feasible conditions that allows to supply maximum amount of thermal energy to users. The significant achievement of DH network optimization is being influential in the reduction of the CO<sub>2</sub> emissions, the fact that is also investigated in the final step of the thesis.

What the thesis seeks to answer is that what are the most effective methods and promising optimization algorithms for conducting an analysis of DH network optimization. Additionally, it aims to identify the specific conditions under which each algorithm can exhibit optimal efficiency. There are a number of methodologies which are popular among the experts and in this thesis the attempt is to choose an optimal technique considering the characteristics of the case study and conduct the analysis. In further step, the attempt was to implement the optimization algorithm to analyze the scope and the outcomes of the implementation.

Following the aim of the thesis, research flow carried out by exploring the numerous works published in the literature. Optimization of the DH network has gained a crucial consideration in the literature of RES use in the last few decades and still a vast group of researchers are conducting optimization analysis to develop its accuracy and scope of implementations. The first chapter followed by the penetration on more details about the theories supporting the DH system optimization. The methodologies and the materials used to carry out theories in hand are described in the later chapter. The discussion on the implementation of the methods and findings are reported on the Fourth chapter. At last but not least, the conclusion and the suggestions for the further research in the literature are discussed.

## **2. LITERATURE REVIEW**

### **2.1. Introduction**

Resource depletion and climate change force modern communities to hardly focus on the matter of energy consumption (EC). To tackle those issues a low-carbon sustainable energy system has been suggested which requires reducing unnecessary energy uses, shifting from fossil fuels toward RES, and increasing the EE to be guaranteed (Dou et al., 2021). DH network has gained a critical importance as one of the principal strategies in the transition toward low-carbon sustainable urban development. In this regard, numerous researches have been conducted and still is being conducted in order to study the competencies of the DH network in urban contexts, optimization, and potential development of those systems. Back to this notion, the aim of this chapter is to investigate in depth about the establishment, evolution, operation, and researches in the literature of DH networks.

### **2.2. District heating network's history**

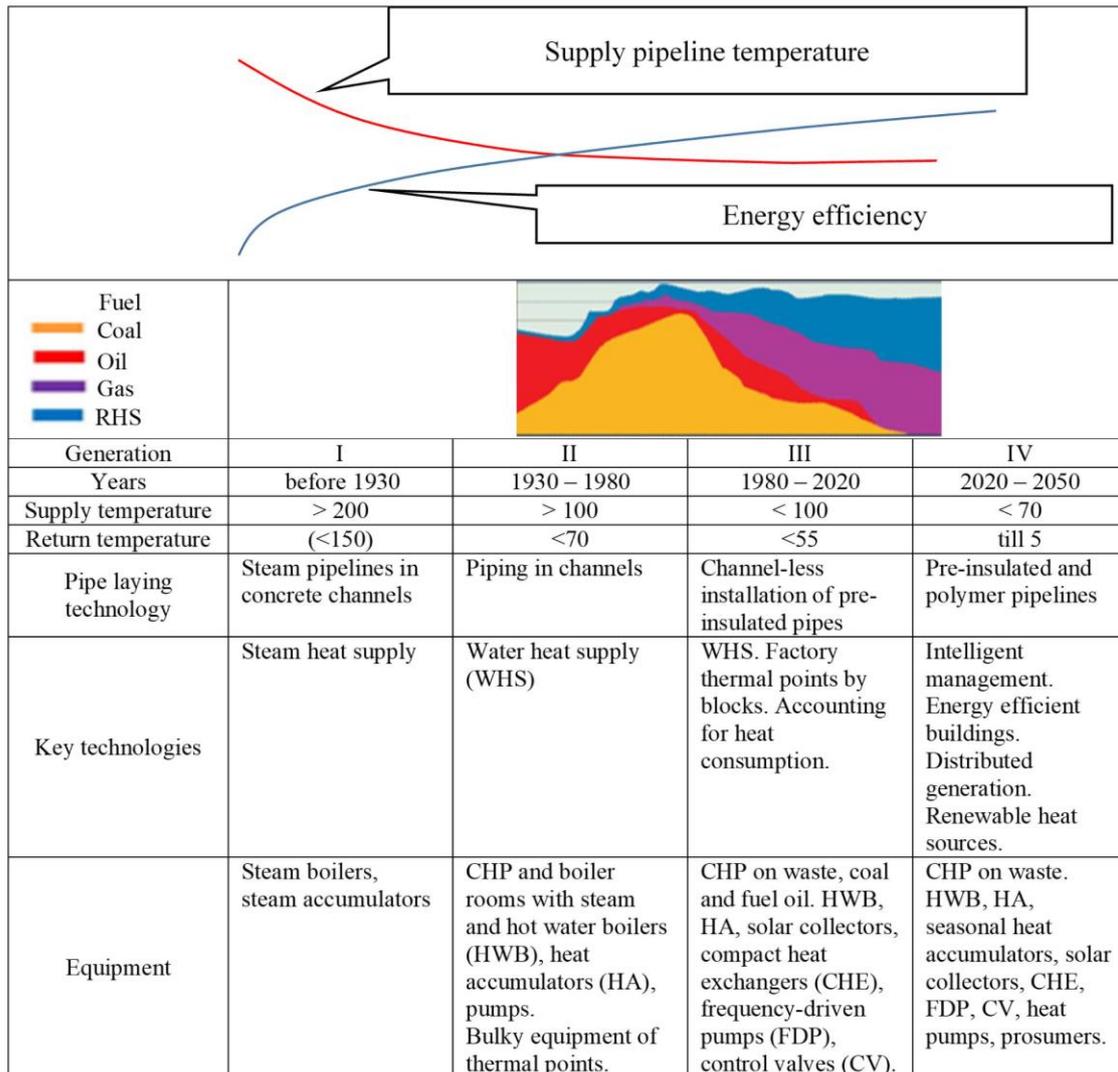
Historically, supplying fuels ranging respectively from wood, charcoal, coal, coke, oil, and natural gas, and burning that fuel was supporting house heating. From 1900 onward, in most of the industrial nations, new technology, namely "Central Heating", has introduced that was burning fuel to heat water and produce steam, pumped into the piping system surrounding buildings, and was supplying H by using radiators. This technology at first implemented in large buildings and continued to be provided for low-density buildings (Wiltshire, 2015).

Although providing central heating networks put forth in the UK at a high speed enjoying the availability of natural gas, in countries such as Denmark, Sweden, Germany, and Many Eastern European countries, H supply turned over from "Central Heating" to "District Heating" (DH). According to historians, the earliest DH network including the hot water distribution construction relying on geothermal energy had been established in Chaudes-Aigues in France. The United States also largely contributed to the evolution of the DH networks in the 1870s with a

breakthrough in the development of the steam distribution system for the first time in the world. The latter technology traveled over the borders and gained a great interest in Europe. Though in Europe later on the modern heat distribution technology is modified based on an alternative to using hot water for supplying H and DHW (Wiltshire, 2015).

Starting from the 19th century on, three generations of the DH networks are possible to recognize. Illustrated in Figure 2.1 the initial DH generation relied upon high-temperature steam (200 °C) for heat distribution. The first generation as mentioned above originated in the US and was the dominant technology in use before 1930 in almost every location where the DH network had developed. The steam-based DH network was designed by using concrete supply and return pipelines in which steam was traveling, steam traps, and compensators. Corrosion as a result of condensation is very common in return pipes that make high losses and reduce EE. As a consequence of significant heat losses by steam along with exposure to the potential risk of steam explosion that may cause severe damage, such systems have now become obsolete (Lund, Werner, et al., 2014; Lund, Østergaard et al., 2018).

High-pressure fluid, namely water, heated over 100 °C and flowing inside the pipelines underpinned the second generation of the DH network. The shift toward the water as a carrier in the heating network was witnessed between the 1930s-1970s and in that specific period of time, numerous cities adopted the new technology in their heating network. Concrete water pipes, bulky tube heat exchangers, and control valves were the main components of the second DH generation. Energy efficiency in the aforementioned heating network is rather poor due to the fact that in such systems it is possible to only have control over temperature adjustment at the supply plant, hence heat consumption and performance of the system are unobservable and uncontrollable (Lund, Werner, et al., 2014; Lund, Østergaard et al., 2018).



**Figure 2.1.** Four generations of DH systems<sup>1</sup>

The third generation is the dominant DH technology all around the world and was introduced in the 1970s. In the latter technology, the water is still the heat carrier inside the supply network, though contrasting with the former generation water temperature was set below 100 °C. Since the suppliers of equipment in the latter generation are from Scandinavian countries, sometimes the third generation is called “The Scandinavian Heating System Technology”. The constituents are pre-fabricated and pre-insulated pipes, trenchless laying technology, compact thermal

<sup>1</sup> (A) Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J. E., Hvelplund, F., & Mathiesen, B. V. (2014). 4th Generation District Heating (4Gdomestic): Integrating smart thermal grids into future sustainable energy systems. *Energy*, 68, 1-11.

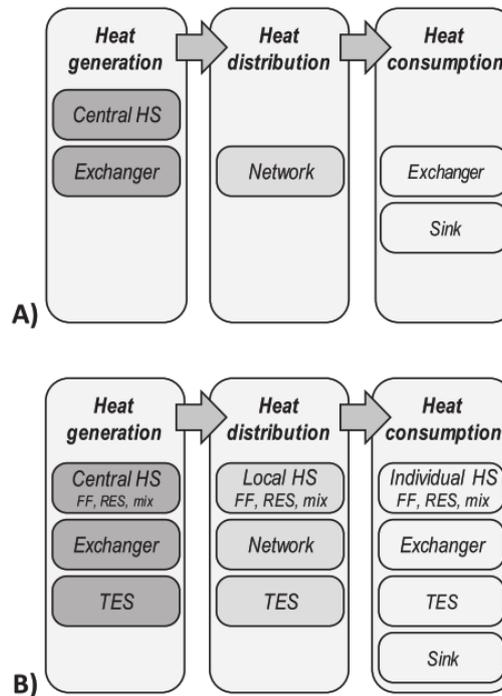
(B) Lund, H., Østergaard, P. A., Chang, M., Werner, S., Svendsen, S., Sorknæs, P., ... & Möller, B. (2018). The status of 4th generation district heating: Research and results. *Energy*, 164, 147-159.

points using plate heat exchangers made of stainless steel, and low volume materials (Lund, Werner, et al., 2014; Lund, Østergaard et al., 2018).

The next generation of the DH networks, will seek to reduce the temperature of the return and supply fluid respectively. The temperature reduction inside the DH network is a key action to lower the cost in the further energy system. In order to actualize this, a variety of interventions at the building level are required. These interventions encompass diverse H systems, suppliers of DHW (Domestic Hot Water), building retrofits, and potential additions like substations or supplementary heat pumps if deemed necessary. In line with aforementioned interventions, power plants and DHN itself should be improved to ensure that the supply at low temperature would still perform efficiently materials (Lund, Werner, et al., 2014; Lund, Østergaard et al., 2018).

### **2.3. Technical design of DHN**

The three primary components of a classic DH network are heat generation, distribution, and consumption. According to the Figure 2.2.A the main components of the DH network consist of the central heat source and heat exchanger for generation, a pipeline network for heat distribution and a heat exchanger as essential part of consumer substation and a heat sink for H and DHW. In the further evolutions of the DH network, in order to fulfilling the energy demand for H and DHW along with handling environmental and climate challenge, advanced components have been added. As shown in Figure 2.2.B, in the energy generation phase, besides adding thermal heat storage, central heat source has been improved to make it possible to use RES including heat pumps. In the distribution phase, thermal heat storage and local heat sources using RES was added, the consumer section enhanced with the installment of individual heat sources besides thermal heat storage (Sayegh et al., 2018).

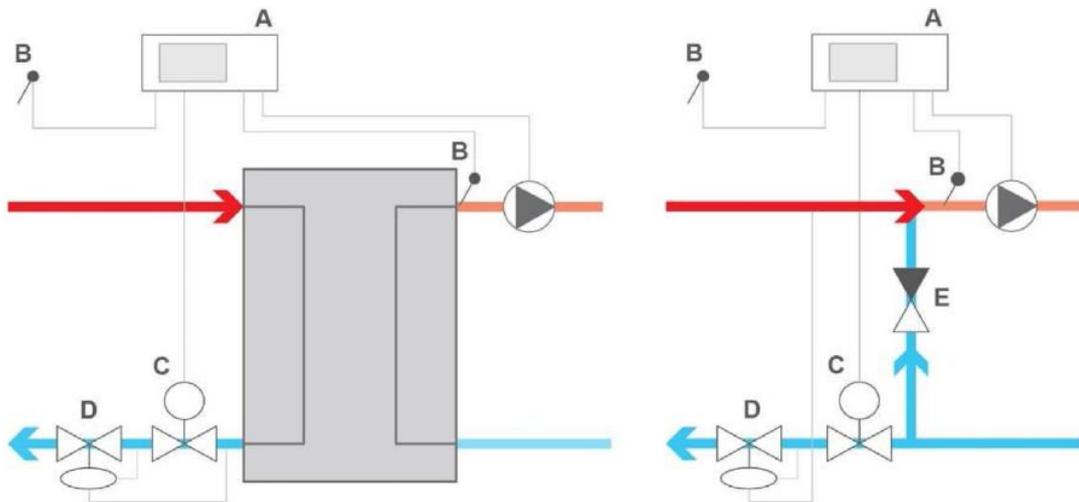


**Figure 2.2.** main components of a DH network: (A) standard, (B) advanced. (FF: Fossil Fuels, HS: Heat Source, RES: Renewable Energy Sources, TES: Thermal Energy Storage)<sup>2</sup>

Customer substations installed in the connected buildings deliver heat from the network to the building heating systems. One system to heat the radiators and one system to distribute DHW are the minimum number of internal distribution systems that need to be heated in a building (Werner, 2013). In order to transport heat from the DH network to the internal space heating system, space heating substations either include a central heat exchanger or a direct connection where the DH network simply provides hot water to the pipes of the internal space heating system. According to Figure 2.3 the substation should include: (A) a controller that enables weather-compensated control to decrease heating network supply temperatures as outdoor temperatures increase and demands for H decrease (Euroheat & Power, 2008); (B) temperature sensors in the outdoor air and on the supply pipe to realize weather-compensated control; (C) a valve that controls the supply of DH network to regulate the heating network supply temperatures according to the weather-compensated set-points; (D) To guarantee suitable operational parameters for the control valve, perhaps the inclusion of a differential

<sup>2</sup> Sayegh, M. A., Jadwiszczak, P., Axcell, B. P., Niemierka, E., Bryś, K., & Jouhara, H. (2018). Heat pump placement, connection and operational modes in European district heating. *Energy and Buildings*, 166, 122-144.

pressure controller could be considered; and (E), if directly connected, a non-return valve to mix the H return water into the supply water to achieve the desired supply temperature set-point. (Østergaard, Smith, Tunzi, & Svendsen, 2022).



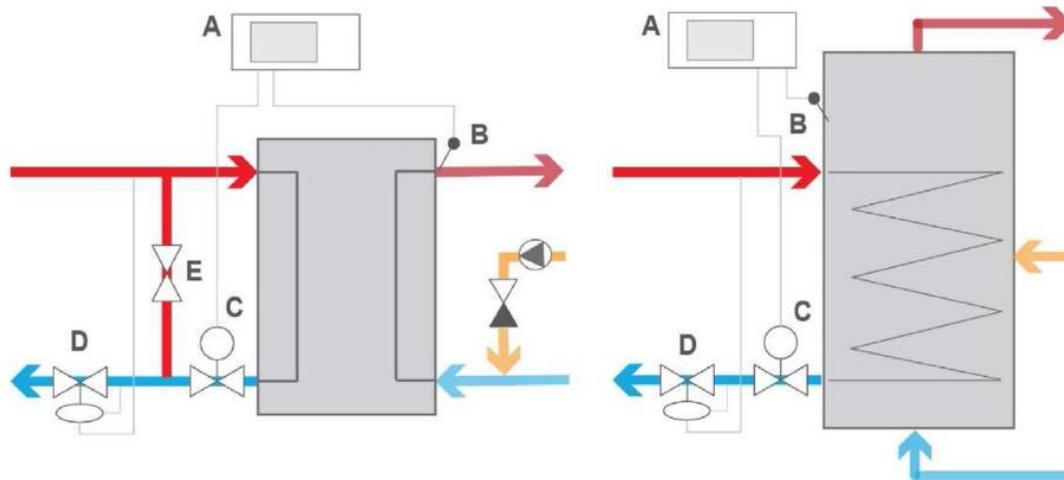
**Figure 2.3.** Example of basic layout for an indirect (left) and a direct (right) space heating substation. (A) Controller (B) Temperature sensors (C) Actuator and control valve (D) Differential pressure controller (E) Non-return valve<sup>3</sup>

Figure 2.4 depicts a sample configuration comprising a DHW exchanger and a DHW storage tank. Instantaneous heat exchangers or buffer tanks are used in the majority of home hot water installations. In each instance, a heat exchanger physically divides the DH water from the fresh drinking water (Østergaard, Smith, Tunzi, & Svendsen, 2022). By instantly heating DHW when tapped, instantaneous heat exchangers lower the amount of DHW that must be kept and the chance of Legionella growth. On the other hand, in single-family homes or terraced homes, they must instantly meet the peak demand, which necessitates larger heat exchangers and service pipes (i.e., the pipes connecting the DH network to the building) (Thorsen, Christiansen, Brand, Olesen, & Larsen, 2011).

Local DHW storage in buffer tanks necessitates careful consideration of Legionella development. Tank systems work best in situations requiring significant bypass flows to maintain high enough temperatures in the pipes (due to long service pipes or low H demands) or in facilities with simultaneous tapings (such as sports

<sup>3</sup> Østergaard, D. S., Smith, K. M., Tunzi, M., & Svendsen, S. (2022). Low-temperature operation of heating systems to enable 4th generation district heating: A review. *Energy*, 123529.

facilities) (Euroheat & Power, 2008). Due to a continual low charging rate, buffer tanks can lower the peak heat demand, but they also cause more heat to escape from DHW networks (Østergaard, Smith, Tunzi, & Svendsen, 2022).



**Figure 2.4.** Example of basic layout a DHW exchanger (left) and a DHW storage tank (right). (A) Controller (B) Temperature sensor (C) Actuator and control valve (D) Differential pressure controller (E) Bypass valve<sup>4</sup>

The citywide distribution network of pipes erected in basements of buildings or buried beneath streets, sidewalks, and park lawns is the most typical component of a DH network. Typically, heat is transferred through water in networks of heat distribution. For high-temperature industrial heat demands and in systems established prior to 1930, steam is employed entirely or in part. The pipes are made to allow for thermal expansion while preventing outside corrosion. The techniques used to build pipes that are buried in the ground have changed over time. Today, prefabricated steel pipes with polyethylene casing and polyurethane foam insulation are the most often used approach. Cheap heat losses, low distribution costs, and high reliability are benefits of this technology (Werner, 2013).

The supply pipelines' water temperature ranges from 70 to 150 C, with an annual average of 80 to 90 C. The lower temperature is utilized in the summer to facilitate the preparation of DHW, while the greater temperature is employed at extremely low outdoor temperatures. The return water's temperature ranges from 35 to 70 C, with an average of 45 to 60 C per year. The linear heat density, pipe size, supply

<sup>4</sup> Østergaard, D. S., Smith, K. M., Tunzi, M., & Svendsen, S. (2022). Low-temperature operation of heating systems to enable 4th generation district heating: A review. *Energy*, 123529.

and return temperatures, and thermal resistance in the pipe insulation all affect how much heat is lost in the distribution system (Werner, 2013).

Rankine cycles with steam superheating are often the foundation of combined heat and power (CHP) plants. In said cycle, a steady pressure and a temperature over the saturation point, the steam from the boiler is superheated. The heat exchanger from the condensing steam turbine uses cooling water to condense the steam into water if the process is solely producing power. The amount of moisture in the steam after the turbine restricts the steam's ability to expand inside the steam turbine. In cases where steam is required for an industrial process, an extraction turbine is employed. Although the CHP's output is reduced by the extraction steam, overall efficiency is still very good (Sipilä, 2016).

Four separate levels of control equipment are used to control the operation of the DH network. The DH operator is in charge of two of these, and the other two are situated in the connected buildings. The first level is the regulation of the heat demand through mixing valves for DHW and thermostatic valves at the radiators. The flow control in the customer substations is the second level. Valve adjustments to the DH water flow regulate the heat transfer. The third stage involves controlling the pressure differential between the network's supply and return pipes. This is done by modifying the distribution pumps' speed. Since the pressure differential is what causes the flow to move through the customer substations, this control level enables all of the customer substations to receive heat. The final stage involves regulating the capacity in the heat generation facilities to manage the supply temperature (Werner, 2013).

## **2.4. District heating regulatory policy**

There are too much to learn from widely various status of DH networks with CHP plants among different countries. From the list of the major institutional factors that differ among the countries are:

- National Energy Policy
- Building regulations
- Price regulation
- Competition
- Feed-in tariffs for CHP and RES
- Emission trading scheme
- Carbon tax
- Investment grants

It can be worthwhile and helpful for a country to learn lessons from the other countries (Nuorkivi, 2016).

The EU is a big proponent of increasing the use of effective DHC and CHP systems to lower emissions and primary energy use as well as to market RES. The rules and measures to achieve the renowned 20-20-20 target aims to reduce primary energy consumption, emissions, and the share of renewable energy sources are introduced in the Directive 2012/27/EU stated on the 25th of October 2012. Under the Directive each Member State must establish an indicative national EE target, which may be based on primary or final energy consumption, primary or final energy savings, or energy intensity (Nuorkivi, 2016).

Previously, with the aim of improving EE in buildings stock across the European Community, a directive namely 2002/91/EC on the energy performance of buildings had been stated. The directive was looking for:

- the broad framework for a method of calculating a building's integrated energy performance;
- the implementation of minimal criteria for new buildings' energy performance;
- the implementation of minimal standards for large existing buildings undergoing significant renovations in terms of their energy performance;
- energy certification of buildings; and
- Buildings should have regular inspections of their heating and cooling systems, as well as assessments of any heating installations with boilers that are more than 15 years old.

The amending Directive on Energy Efficiency (2018/2002) was approved in 2018 as a component of the "Clean energy for all Europeans package" to update the regulatory framework through 2030 and beyond. The energy savings duty in end use, introduced in the 2012 directive, is further extended by the amending directive. According to the amending regulation, all EU nations except for Cyprus and Malta will have to make new energy savings of 0.8% of final energy consumption year for the years 2021–2030. Cyprus and Malta will have to achieve 0.24% each year instead. Except for metering and billing provisions, which have a different deadline (25 October 2020), the directive entered into force in December 2018 and required to be transcribed into national legislation by Member States by 25 June 2020 (European union, 2018).

### **2.4.1. Denmark**

An example of a country with widespread DH supported by strict government regulation is Denmark. The rule gives DH precedence in densely populated building zones. As a result, DH enjoys a sizable market share and is a widely used and favored heating mode across the nation. CHP has been developed to provide high-efficiency service to the DH networks and the national power grid (Nuorkivi, 2016).

Through the establishment of a new public planning process that rationalized heat supply, the first Heat Supply Law of 1979 played a significant role in the expansion of the CHP/DH market. Denmark's reliance on coal and oil as a source of energy decreased with the development of CHP and DH in the 1980s and 1990s. Initially, some oil was gradually replaced by CHP powered by coal and natural gas as well as by a greater use of renewable energy sources; from the mid-1990s, coal has also been phased out (Nuorkivi, 2016).

Danish Energy Agreement for 2012-2020 described the course that Denmark will take until 2050 and created a framework for climate and energy policy up to 2020. According to the Agreement, in 2020, 35% of total energy consumption shall come from renewable sources, with half of that amount coming from wind power. Additionally, compared to 2006, the energy usage should drop by more than 12%

by 2020. By prohibiting the installation of new oil-fired and natural gas boilers in new buildings beginning in 2013, and the installation of new oil-fired boilers in existing buildings in areas with access to DH or natural gas beginning in 2016, the Agreement supports the phase-out of oil-fired boilers in existing buildings (IEA, 2020).

### **2.4.2. Switzerland**

The Federal Assembly of the Swiss Confederation stated Energy Act (EnG) in September 30<sup>th</sup>, 2016 to set requirements for the transition towards RES supply, specifically indigenous renewable energies. According to the directive, at least 50 percent of the global contribution allocated to a canton must be used to promote private measures including connection to existing or a new local and district heating networks. The Energy Act came into effect from 1<sup>st</sup> of the January 2018, with the statement of the Energy Regulation (EnV) by the Swiss Federal Council.

### **2.4.3. Italy**

The Directive 2009/28/EC regulations on promoting the use of energy from renewable sources are transposed into Italian law by Decree No. 28 of March 3, 2011. It starts making modifications to Italy's renewable energy support programs (including some related to RES H&C networks). In order to meet the overall national goal of using 17% renewable energy in gross final energy consumption in 2020, it sets particular sectoral goals. The Decree also established unique administrative procedures for each type of installation in order to govern the development and operation of renewable energy facilities. An expert committee has been assembled ad-hoc to examine the energy potential of biomass and to assist in the creation of pertinent laws (IEA, 2019).

Additionally, specific provisions for authorizations and modifications are included for the development of district heating and cooling as well as to the natural gas and electricity systems. A guarantee fund is funded in part by a tax (0.05€/Sm<sup>3</sup>) levied on natural gas use (paid by the final consumers), which aids in the construction of

DH networks. Additionally, Article 22, Section 3 mandates that all municipalities with more than 50,000 residents develop district heating and cooling networks in coordination with provincial authorities and in accordance with regional energy plans, with the aim of maximizing the use of energy generated from RES (IEA, 2019).

The revised Annexes, which partially replace and supplement the Annexes of the Decree No. 20/2007 that implements Directive 2004/8/EC on the promotion of cogeneration (CHP), are contained in the ministerial decree from August 4, 2011. This metric establishes a new formula for recognizing highly efficient cogeneration. The ministerial order from September 5, 2011, specifies the cogeneration incentive mechanisms. For generation plants and integrated DH plants, it created an incentive system based on the White Certificates Scheme, which is recognized for a term of 10 years and 15 years respectively (IEA, 2013).

The Decree-Law No. 63 of 2013 on "Urgent measures for the transposition of Directive 2010/31/EU of the European Parliament and of the Council on the energy performance of buildings" was amended by Law of 3 August 2013, No. 90. It replaces Legislative Decree 192/2005, which dealt with the implementation of Directive 2002/91/CE on building energy performance.

Objectives of the Decree are as followings:

- encourage improvements in the energy performance of buildings;
- encourage the development, enhancement and integration of renewable energy sources in buildings;
- to support energy diversification;
- promote the competitiveness of the domestic industry through technological development; achieve the national targets for energy and the environment;
- extend and enhance the system of tax deductions from 55 % to 65% (10 equal yearly amounts) for the redevelopment of Buildings (Eco- bonus), building renovations, the installation of RES-H technologies (e.g. 60% for heat pumps, solar thermal collectors, 50% for biomass-fueled heat generators (since 2017(-Art. 3, par. 4 L 205/17)). For energetic requalification works aimed at improving the winter and summer energy

performance of common buildings, the tax deduction will amount to 70% or 75% (IEA, 2019).

## **2.5. Energy prediction models**

The evolution of buildings is closely tied to energy consumption patterns (Sikder, Nagarajan & Koetter, 2018). For instance, a well-designed urban framework can exhibit energy-efficient practices and potentially encourage the adoption of cleaner energy sources as cities expand (Gassar & Cha, 2020). Gaining a deeper comprehension of the spatial and cause-and-effect connections among buildings, infrastructure, energy consumption patterns, and the community's way of life holds significant importance (Resch, Bohne, Kvamsdal, & Lohne, 2016). Hence, to achieve a shift towards clean energy and carbon neutrality by 2050, it is vital to employ comprehensive approaches and techniques that integrate detailed spatial and temporal data from various urban sources (Perera et al., 2023).

Making energy projections, whether for the entire lifespan of a building or for gauging energy consumption performance across various levels of constructions, holds the potential to enhance energy utilization in particular structures by implementing retrofits or integrating advanced renewable energy technologies. This process also presents an opportunity to uncover avenues for reducing energy demand and formulating more effective strategies in urban settings (Ahmad, Mourshed, & Rezgui, 2017; Reinhart, & Davila, 2016). These prognostications aid in appraising diverse design alternatives, devising energy efficiency strategies, and refining supply-demand management, while considering factors like building attributes, weather conditions, socio-economic characteristics, installed equipment, occupancy patterns, and geographical locations, all of which influence energy consumption rates (Kwok, & Lee, 2011; Amasyali, & El-Gohary, 2018). Furthermore, energy projections shed light on the consequences of comprehensive energy retrofit initiatives and changes to energy supply infrastructures. Typically, energy prediction models establish energy requirements as a function of input parameters (such as socio-demographic, economic, climatic, appliance, and building characteristics). As a result, these predictive energy

models emerge as invaluable tools for energy managers, urban designers/architects, and policymakers to assess national/regional energy supply needs and shifts in energy demand for specific structures resulting from upgrades or the integration of new technologies (Gassar & Cha, 2020).

### **2.5.1. UBEM, USEM, and USBEM models**

Urban building energy modeling (UBEM) stands out as an effective approach for simulating and assessing the EE of neighborhood-scale buildings. However, the capability of UBEM to support sustainable development within constructed environments is restricted due to its reliance on traditional physics-based inputs (Heidelberger & Rakha, 2022). This challenge becomes particularly crucial for broader studies encompassing districts or entire cities, where the availability of data also raises concerns (Mutani, Vocale, & Javanroodi, 2023).

UBEM stands as a foundation for the decision-making processes mentioned earlier, offering a bottom-up, physics-rooted method to simulate the thermal behavior of multiple buildings (Reinhart & Davila, 2019). It essentially extends the concept of individual building energy models (BEM) to encompass a considerable number of buildings, ranging from hundreds to even tens of thousands. These BEMs entail intricate calculations involving heat transfer and mass flow, with their core simulation techniques tracing back to the 1970s (Mills, 2004; Swan, & Ugursal, 2009). Over time, BEMs have grown more cost-effective (Roth, 2016), becoming integral tools during the planning and design phases of energy-efficient buildings (Ang, Berzolla, & Reinhart, 2020). To account for the intricate interaction between buildings and their surroundings, the adoption of UBEM and urban-scale energy model (USEM) has gradually increased, enabling extensive static and dynamic simulations of diverse building types and urban configurations (Ang, Berzolla & Reinhart, 2020; Basu, Bale, Wehnert, & Topp, 2019). While the fundamental physics principles remain consistent between BEM and UBEM, the latter necessitates substantial automation procedures and computational capabilities throughout data input, model generation, and analysis (Ang, Berzolla, & Reinhart, 2020).

An UBEM necessitates the fusion of multiple datasets encompassing climate information, building characteristics, construction standards, and usage schedules (Reinhart & Davila, 2019). Climate datasets for simulating building performance have been accessible for some time, following the establishment of a viable data format called the typical meteorological year (TMY) (Hall, Prairie, Anderson, & Boes, 1978; Crawley, Hand, & Lawrie, 1999), coupled with the availability of data in this format for various global regions. The input data related to building geometry in UBEM entails details like building envelope shapes, window opening proportions, and terrain characteristics. Depending on whether the focus is on a new or existing neighborhood, this data can either be sourced from existing datasets or generated anew (Reinhart & Davila, 2019). Moreover, beyond the physical structure, non-geometric aspects such as construction assemblies and HVAC systems must also be defined. At the individual building level, this phase often accounts for roughly a third of the modeling process (Cerezo, Dogan, & Reinhart, 2014) and introduces significant differences between simulated and actual energy consumption due to uncertainties surrounding factors like infiltration rates, equipment loads, and occupant behavior (ASHRAE, 2009). While these parameters can be measured for a small sample of existing buildings, gathering such detailed data becomes infeasible for larger urban areas. Consequently, an UBEM needs to simplify a building stock into "building archetypes," which are essentially building definitions representing groups of structures with similar attributes (Reinhart & Davila, 2019).

UBEM has undergone significant advancement, resulting in a wealth of robust urban data streams originating from diverse sources like geographic information systems (GIS), light detection and ranging (LiDAR), and tax assessor databases. These data streams culminate in synthetic hourly profiles detailing building energy demand under present and potential future circumstances. Various modeling, simulation, and calibration strategies have emerged based on the availability of historical building energy consumption data (Reinhart & Davila, 2019). Alongside this, a range of UBEM applications have surfaced in both academic research and practical implementation (Ang, Berzolla, & Reinhart, 2020).

Urban-Scale Energy Modeling (USEM) serves as a crucial method for simulating energy usage on a city-wide level, encompassing not only individual building attributes but also the larger urban environment (Mutani & Todeschi, 2019). USEMs can be classified into two categories, namely top-down and bottom-up, based on the level of input data (Mutani & Todeschi, 2017). The reliability of an energy model largely depends on the accuracy and completeness of the input dataset, along with the ability to compare results against extensive measured energy consumption data for validation (Reinhart & Davila, 2016).

The primary challenge faced by these models at an urban scale is the management of a vast volume of data, which can vary in accuracy and scope when describing the diverse attributes of buildings and populations across a region (Ryan & Sanquist, 2012). Additionally, due to the intricate nature of the problem, these tools often exhibit prolonged simulation times, which escalate as more elements are introduced into the scenario (Todeschi, Boghetti, Kämpf, & Mutani, 2021). Furthermore, these energy models often overlook several variables that impact consumption, particularly concerning the urban context (Wenjing, Yanuar, & Perry, 2004). Nonetheless, current models and tools encounter difficulties in representing a realistic urban energy distribution capable of evaluating energy performance at the neighborhood scale (Abbasabadi & Ashayeri, 2019).

USEMs make a significant contribution by evaluating the energy efficiency of buildings at the urban scale, encompassing the analysis of energy consumption, renewable energy production, and productivity (Sola, Corchero, Salom, & Sanmarti, 2020). These models find utility in supporting urban planning for both new and existing neighborhoods, conducting retrofit analyses of building stocks, enhancing building energy performance through smart green technologies, and designing and optimizing district energy networks (Johansson, Olofsson, & Mangold, 2017; Ben, & Steemers, 2020). Beyond energy consumption simulation, USEMs also possess the capability to visualize and replicate the influence of the surrounding urban context on buildings. Generally, USEMs intend to cover the urban environment by simulating how a collection of buildings perform in terms of energy usage over different time periods (hourly, daily, monthly, and annually) and spatial scales (individual building, block, neighborhood, and district), thus aiding

energy retrofit strategies by assessing their territorial impact (Nutkiewicz, Yang, & Jain, 2018). Numerous tools and methods for energy simulations (like CitySim, UrbanSim, and Urban Modeling Interface) can be utilized to calculate the energy requirements of a collection of buildings, all while taking into account the urban climate and physical structure (Bruse, Nouvel, Wate, Kraut, & Coors, 2019; Sola, Corchero, Salom, & Sanmarti, 2018, 2020).

By expanding the scope of UBEM and USEM to encompass districts or territories, a novel localized energy model can be introduced, referred to as urban-scale building energy modeling (USBEM) (Abolhassani, Amayri, Bouguila, & Eicker, 2022). In the realm of USBEM, the central approach involves quantifying the energy efficiency of structures within an urban setting across varying spatial and temporal scales. Furthermore, these models have the potential to inform future urban planning and the enhancement of existing as well as planned regions. Grasping the daily and seasonal energy consumption patterns across each city location provides authorities with a deeper understanding of how to harmonize energy supply and demand, averting instabilities and shortages in the energy infrastructure. Such models also facilitate scenario-based strategizing and performance assessment for retrofitting buildings and integrating renewable energy solutions within a city's energy systems. Furthermore, the process of planning and evaluating new city districts becomes less formidable with the utilization of appropriate models. On the whole, USBEM serves as an effective instrument to guide stakeholders, urban planners, and decision-makers in comprehending urban energy systems. This, in turn, empowers them to shape energy strategies, propose sustainable initiatives, and formulate constructive policies (Moghadam, Delmastro, Corgnati, & Lombardi, 2017; Reinhart, & Davila, 2016; Hong et al., 2015). USBEM operates based on two primary methodologies: top-down models, and bottom-up models (Ferrando, Causone, Hong, & Chen, 2020).

### **2.5.2. Top-Down and Bottom-Up models**

Regarding the arrangement of input details and the overall modeling approach, investigations into urban energy dynamics can be broadly classified into two main categories: top-down models and bottom-up models (Swan & Ugursal, 2009; Li et al., 2017; Kavgić et al., 2010). Top-down modeling involves employing aggregated data at the municipal or national scale to depict the relationship between energy consumption and influencing factors like socio-economic variables and climate conditions. These models, known for their simplicity and reliance on historical aggregated data, have found widespread use in urban energy research, as seen in studies (Sailor & Lu, 2004; Tornberg & Thuvander, 2005). However, their reliance on historical macroeconomic energy trends and the absence of intricate technological descriptions limit their suitability for analyzing technological shifts in present and future development studies.

Top-down models can be broadly classified into two categories: econometric models and technological models (Kavgić et al., 2010). Econometric top-down models primarily rely on variables such as income, fuel prices, and gross domestic product to establish the connection between the energy sector and economic output. These models may also incorporate general climatic conditions, such as population-weighted temperature, on a national scale. However, they often lack specific details about current and future technological options, as their focus lies more on macroeconomic trends and historical relationships rather than individual physical factors in buildings that influence energy demand (MIT, 1997). On the other hand, technological top-down models encompass a range of other factors that impact energy usage, such as saturation effects, technological advancements, and structural changes. However, these factors are not explicitly described within the models (Johnston, 2003).

Conversely, bottom-up approaches are constructed using data on various disaggregated components, which are subsequently combined based on estimated individual impacts on energy usage (Rivers & Jaccard, 2005). Bottom-up models operate with a detailed and individualized approach, requiring extensive databases of empirical data to comprehensively describe each component (Shorrock & Dunster, 1997). Based on the level of detail in end-use data and the

applied methodology, bottom-up models can be categorized into three types: statistical, engineering (physical), and hybrid models. Statistical bottom-up models establish connections between individual end-use energy and building characteristics, as well as socio-economic indicators (Kontokosta & Tull, 2017; Moghadam, Toniolo, Mutani, & Lombardi, 2018; Howard et al., 2012). In contrast, engineering models utilize the physical and technological attributes of individual buildings to compute energy requirements, offering a high degree of flexibility for assessing technological advancements and energy efficiency scenarios. However, engineering models demand extensive empirical data and are susceptible to uncertainties in underlying assumptions, especially concerning human behavior and occupancy patterns. This leads to the utilization of hybrid models, which blend characteristics of both statistical and engineering approaches. In these hybrid models, building attributes are defined using physical characteristics (akin to engineering models), while essential data, particularly occupant-related information, is drawn from historical energy use analysis (similar to statistical models). This amalgamation aims to address the limitations of both models and attain a more sophisticated representation (Mutani & Todeschi, 2017; Nouvel et al., 2015; Nutkiewicz, Yang, & Jain, 2017; Déqué, Ollivier, & Poblador, 2000).

In existing literature, the methodology for conducting bottom-up urban-scale energy modeling for buildings, encompassing physical models of heat and mass transfer within and around structures, is commonly referred to as "urban building energy modeling" (Reinhart & Davila, 2016). This approach takes into account the intricate interplay of various factors to create a comprehensive representation of urban energy dynamics, bridging the gap between detailed physical aspects and real-world energy consumption patterns. Table 2.1 comprehensively elaborates on the attributes of top-down and bottom-up methodologies.

**Table 2.1.** Characteristics of top-down and bottom-up approaches<sup>5</sup>

| Approaches                             | Advantages   | Limitations  |
|--|--|--|
| <b>Top-Down Models</b>                 | <ul style="list-style-type: none"> <li>Both long-term socio demographic and market economic effects considered</li> <li>Detailed technology description and actual energy consumption not required</li> <li>Limited input information often with aggregated economic data</li> </ul> | <ul style="list-style-type: none"> <li>Past energy-economy interactions used to predict future energy consumption</li> <li>Long term historical data required</li> <li>Lack in technological details</li> </ul>  |
| <b>Bottom-Up: Statistical Models</b>   | <ul style="list-style-type: none"> <li>Both socio-demographic and marketeconomic effects considered</li> <li>Simulation of energy use at end-use and/or building level</li> <li>Variations in individual end uses considered</li> </ul>  | <ul style="list-style-type: none"> <li>Billing, weather, and/or survey data required</li> <li>A larger number of sampling subjects required</li> <li>Possible multicollinearity to be addressed</li> <li>Simulation results highly dependent on historical consumption trend; prediction well outside of bounds of training data not reliable</li> </ul> |
| <b>Bottom-Up: Physics Based Models</b> | <ul style="list-style-type: none"> <li>Socio-demographic and economic information not required</li> <li>Simulation of energy use at different temporal scales</li> <li>Variations in individual end uses considered</li> </ul>   | <ul style="list-style-type: none"> <li>Detailed physical and technological measures required</li> <li>Socio-demographic and market economic trends not captured</li> <li>Intensive computational effort required</li> </ul>  |

### 2.5.3. Black-box, White-box, and Grey-box models

Various scientific reviews concur on categorizing energy scale modeling into three primary categories (Mutani, Vocale, & Javanroodi, 2023): data-driven (black box), which relies on statistical models (Malhotra et al., 2022) and AI models (e.g., Sun, Haghghat, & Fung, 2020; Liu et al., 2022); process-driven (white-box), which builds upon process-driven models (Gassar & Cha, 2020); and hybrid (grey-box) models, which combine data-driven and process-driven models (Sun et al., 2022; Wang, Lee, & Yuen, 2018) and are primarily employed by environmental and urban planners, governmental bodies, and policymakers (Mutani, Vocale, & Javanroodi, 2023).

The rapid advancement of sensing technologies and the emergence of smart city initiatives have resulted in an abundance of structured and unstructured data streams that describe buildings and their urban surroundings. Concurrently, the field of artificial intelligence is making rapid progress, developing new machine

<sup>5</sup> Li, W., Zhou, Y., Cetin, K., Eom, J., Wang, Y., Chen, G., & Zhang, X. (2017). Modeling urban building energy use: A review of modeling approaches and procedures. *Energy*, 141, 2445-2457.

learning models that utilize these data streams to predict and analyze various physical phenomena within cities, such as air pollution dynamics (Singh, Gupta, & Rai, 2013), traffic flow (Lv, Duan, Kang, Li, & Wang, 2014), and energy consumption (Jain, Smith, Culligan, & Taylor, 2014). A novel framework called Data-driven Urban Energy Simulation (DUE-S) aims to bridge the gap between traditional engineering-based energy simulation models and the emerging data-driven machine learning models (Natkiewicz, Yang, & Jain, 2018).

Data-driven models do not necessitate building thermal balance equations, thereby requiring less or even no detailed physical information about the building. These models rely on historical data to uncover hidden relationships between building energy consumption and input variables (such as weather, building characteristics, occupants' behavior, and equipment schedules) using mathematical techniques. Data-driven methods are particularly well-suited for buildings lacking intricate physical parameters, such as those in the design phase (Chen, Guo, Chen, Chen, & Ji, 2022). The simplicity and flexibility of data-driven models have led to their increasing popularity in building energy prediction (Wang & Srinivasan, 2017).

The promising "black-box" methods employed in data-driven modeling comprise linear regression (LR), support vector machine (SVM), extreme gradient boosting (XGBoost), random forest (RF), recurrent neural network (RNN), and artificial neural network (ANN) (Chen, Guo, Chen, Chen, & Ji, 2022).

White-box-based models, also known as physical-based models (PB) or Process-driven models (Mutani, Vocale, & Javanroodi, 2023), utilize fundamental physical principles to calculate the thermodynamic and energy behavior of entire buildings or specific components within them (Zhao & Magoulès, 2012). PB is the only approach capable of fully estimating the energy usage of the building sector without relying on historical energy consumption data (Gassar & Cha, 2020). Typically, these models rely on energy balance equations that account for the interactions between buildings and the surrounding outdoor environment, taking into consideration building geometry, characteristics, human behavior, urban environment, and climate data. Simulation engines are employed to represent the energy and heat transfer processes within buildings, offering significant versatility across various application fields. Once validated, these energy balance equations

can be customized to fit any context and spatial-temporal scales effectively (Mutani & Todeschi, 2020).

This technique has seen significant development in recent years and is categorized into two main methods: the simplified method, which involves estimating space heating based on climate using heating degree days, and the detailed comprehensive method, which encompasses an extensive thermodynamic and heat transfer analysis for all end-uses within the building. Due to their foundation in end-use physics, physical methods exhibit the highest flexibility and effectiveness in modeling novel technologies that lack historical consumption data (Gassar & Cha, 2020).

The estimation of building energy usage often involves the application of black-box (data-driven) and white-box (physical-based) methods. Each technique has its limitations, especially the white-box approach, which assumes thorough knowledge of all thermal and geometric characteristics of buildings. Gathering such comprehensive information can be challenging, particularly for existing buildings in megacities. Conversely, the black-box approach heavily relies on data, which needs to be abundant and extensive. However, these limitations can be overcome by combining both approaches, leveraging the advantages of one to compensate for the drawbacks of the other (Foucquier, Robert, Suard, Stéphan, & Jay, 2013). This integration of physics and statistics is known as a grey-box-based approach or hybrid approach (Gassar & Cha, 2020).

Hybrid modeling is a powerful approach that amalgamates the advantages of both data-driven and process-driven modeling to achieve optimal results. This synthesis harnesses the computational efficiency of data-driven models and the capability of process-driven models to reveal the physical relationships between variables (Boghetti et al., 2020). Implementing hybrid techniques requires a high level of expertise in formulating appropriate modeling equations and estimating the relevant parameters. In many large-scale applications, building properties are represented as an energy path analogy, where a simplified "resistance-capacitance" circuit describes and characterizes the energy performance of the building sector (Ahmad, Chen, Guo, & Wang, 2018). Recent studies show that hybrid models outperform individual approaches, exhibiting high accuracy in

handling complex situations and unexpected trends (Todeschi et al., 2022). Table 2.2 presents a quick overview of the benefits and drawbacks associated with the three primary methodologies for predicting building energy-use models in.

**Table 2.2.** *The advantages and disadvantages of the three major building energy prediction approaches*<sup>6</sup>

| Item                 | Black-box based approach   | White-box based approach  | Grey-box based approach   |
|----------------------|--|---|---|
| <b>Advantages</b>    | <ul style="list-style-type: none"> <li>Inputs to model are historical data</li> <li>High running speed, except support vector machine</li> <li>Inclusion of macro socio-economic effects</li> <li>High capacity to deal with both linear and nonlinear problems of input variables except for regression</li> </ul>                                      | <ul style="list-style-type: none"> <li>Inputs to model are physical information</li> <li>Results can be interpreted in physical terms</li> <li>No training data are required</li> <li>Explicit representation of end-energy uses</li> <li>High accuracy.</li> </ul>                   | <ul style="list-style-type: none"> <li>Inputs to model are physical information and historical data</li> <li>Results can be interpreted in physical terms</li> <li>Inclusion of macro socio-economic effects</li> <li>Explicit representation of end-energy uses</li> <li>Very high accuracy</li> </ul> |
| <b>Disadvantages</b> | <ul style="list-style-type: none"> <li>Reliance on historical data</li> <li>Several difficulties to interpret results in physical terms</li> <li>Large amount of training data is required</li> <li>Low running speed with support vector machine</li> <li>Low accuracy score with regression</li> <li>No explicit representation of end-uses</li> </ul> | <ul style="list-style-type: none"> <li>Detailed input physical information is required</li> <li>Representative buildings</li> <li>Assumption of occupant behavior</li> <li>No economic factors</li> <li>Not easy to use, needs experience</li> <li>Running speed is medium</li> </ul> | <ul style="list-style-type: none"> <li>An approximate description of the building is required</li> <li>Not easy to use, needs experience</li> <li>Low running speed</li> </ul>  |

## 2.6. Researches in the literature

The supply chain of H and DHW can be energy efficient and play a significant role in energy savings and emission reductions by exploiting the existing DH network (Nagota, Shimoda & Mizuno, 2008; Rezaie & Rosen, 2012). Principally, optimized plants that have a capacity to produce electricity and heat, along with both use of RES (Lindenberger, Bruckner, Groscurth, & Kümmel, 2000; Salomón, Gómez Galindo, & Martin, 2014) and recycling the waste heat through specific industrial processes (Fang, Xia, Zhu, Su, & Jiang, 2013; holmgren, 2006; Sun, F., Fu, Zhang, & Sun, J, 2012) can put energy saving and emission reductions in real. In the case

<sup>6</sup> Gassar, A. A. A., & Cha, S. H. (2020). Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review. *Energy and Buildings*, 224, 110238.

that DH network supplies energy to the high dense urban communities, their efficiency is higher than autonomous heating systems and is the maximum (Dou et al., 2021; Guelpa et al., 2016).

To achieve high efficiency in the DH networks, a number of aspects can be taken into account. Among them on the consumption side evaluating energy-use models as well as building retrofitting policies, coupling with the optimization of the performance of the DH network in the generation and distribution phases are significant.

In (Mutani, Todeschi, Kämpf, Coors, & Fitzky, 2018) three different energy-use models including a GIS-based tool, a 3D energy model with the use of CityGML, and an engineering tool with the use of CitySim is presented to analyze their attributes and determine the best aspects of an optimal model in the assessment process of energy resources, future scenarios, EE solutions and best energy policies. The objective of the work was to find a smarter way to consume energy, compatible with the available and more efficient energy sources.

In a study by Todeschi, Boghetti, Kämpf, and Mutani (2021), they utilize bottom-up urban-energy models to assess the precision and adaptability of energy simulations. These models are based on the hourly heating consumption of residential buildings. The study involves a comparison between two prevalent energy-use models: a machine learning model and a GIS-based engineering model. This comparison is carried out by contrasting the models with anonymous monitoring data. Additionally, a sensitivity analysis using the Morris method is performed on the GIS-based engineering model. This analysis aims to gauge how input variables influence heating consumption and pinpoint potential opportunities for refining the existing model.

In 2019, Lidberg, Olofsson, and Ödlund conducted a study where they evaluated the effects of four distinct energy refurbishment strategies on reducing the heating demand (HD) and domestic hot water (DHW) system requirements in a multi-family building. Additionally, they examined how the DHW circulation system influenced the return temperature. To achieve the study's goals, they employed simulations of a multi-family building within the TRNSYS 17 software environment.

In 2020, Mutani and Todeschi crafted an urban energy balance leveraging EC data from the district heating (DH) network. They not only took into account influential urban contextual features that impact energy performance (EP), but they also adopted a quasi-steady state approach on a monthly timeframe. This method was applied to develop an energy consumption model for residential buildings operating at the neighborhood level.

In a study by Mutani and Todeschi (2021), the energy performance (EP) certification database is harnessed to assess the EP of current residential structures and pinpoint enhanced retrofitting strategies. Achieving this, they implement an urban-level energy model. The primary objective of the research is to establish an updated retrofitting database, enhancing the outcomes of an urban building energy model by incorporating key attributes of the constructed surroundings.

Mutani, Fabiano, Garcia, & Mancini, in 2021, used a bottom-up place-based methodology through the use of statistical analysis to evaluate the spatial distribution of the EC of residential buildings at the urban scale. Also, the EP of buildings for H, space cooling (C), DHW production, and electricity are estimated by using significant linear regressions. The objective of the work was to identify the main variables on which EC depends.

In (Madsen, Sejling, Søgaaard, & Palsson, 1994) stochastic modeling, prediction, and control methods are utilized to optimize the control of the flow and the supply temperature in DH networks following the objective of reducing heat production costs and heat losses in the transmission and distribution net particularly if the heat production takes place at a CHP plants.

In 1995, Benonysson, Bohm, and Ravn used a mathematical model which incorporates the consumers, the DH network, and the production plant to formulate the problem of the selection of supply temperatures in a DH network following the objective of minimizing the operational costs. The so-called node method is described and applied to simulate the flow and temperature development of a given DH network due to the consumers' heat loads and supply temperatures from the plant.

In a study conducted by Guelpa et al. (2016), a geographical information system (GIS) centered model is employed to assess the technical viability of potential extensions to current district heating (DH) networks. The analysis takes into account multiple technical constraints, and a fluid dynamics model is utilized to simulate the distribution mass flow rates for both the present and prospective scenarios.

In (Mutani, Todeschi, Guelpa, & Verda, 2020) enjoying the full benefits of GIS-based models, combined effect of DH expansion with different building retrofitting scenarios is proposed. The aim of the paper is to assess the implementation of the energy policies at urban level to optimize energy demand and supply of buildings and to evaluate effectiveness of future trends of building retrofit measures.

In a 2011 study by Sperling and Möller, an examination was conducted on the current regional energy setup and the immediate integration of renewable energy resources. The research aimed to assess the consequences of district heating (DH) expansion and the implementation of energy-saving measures on end-use. Employing a 'transition perspective,' the study aimed to apply this approach to the growth of local renewable energy systems. To achieve this, the heat atlas methodology was adapted and implemented to analyze the existing building stock, ultimately evaluating the implications for both present and future energy systems.

In (Roland and Schmidt, 2020) a mixed-integer nonlinear programming (MINLP) model is conducted to compute the optimal expansion of an existing tree-shaped DH network given a number of potential new consumers. To this end, the Euler momentum and the thermal energy equation are considered. Also, a novel polynomial approximation is developed for the optimization model. The expansion decisions are modeled by binary variables for which additional valid inequalities are derived.

In 2020, van der Zwan and Pothof introduced what they claim to be a robust and expedient model-predictive control strategy. This approach apparently leverages building thermal mass as a makeshift daily storage solution while attempting to avoid breaching temperature constraints. The paper emphasizes its operational control strategy that purportedly accommodates temperature-restrained renewable

energy sources (RES). They assert that they've contorted the optimization problem into an almost convex form – an apparent prerequisite they emphasize for implementing model-predictive control in practical scenarios.

In 2021, Bjørnskov and colleagues employed the Monte Carlo Tree Search initialization technique in conjunction with a branch and bound solver to tackle the production planning of a combined heat and power (CHP) unit. They cast the optimization problem as a non-convex mixed-integer program within a sliding time window framework. Their objective revolved around discovering improved solutions compared to random initialization, particularly within expansive district heating (DH) networks.

In a study by Résimont, Louveaux, and Dewallef (2021), they introduce a multi-period mixed-integer linear programming (MILP) framework intended for the optimal design and dimensions of a district heating (DH) network. The primary objective was to maximize net cash flow through the utilization of a geographic information system. The focus was on factoring in the year-round temporal patterns of heating demands, which led to the strategic inclusion of thermal storage solutions within the optimized solutions.

A MILP optimization model with the endogenous decision on the potential interconnectors is developed to conduct an economic and environmental assessment of increasing the efficiency of the DH networks by interconnecting adjacent systems in (Dominković, Stunjek, Blanco, Madsen, & Krajačić, 2020). To initialize the optimization a hierarchy of network growth from a tree-like structure to a ring-like and at the end, a meshed structure is developed. A further attempt is applied to investigate the impacts of thermal energy storage (TES) placement with the goal of lowering costs in a form of smaller pipe sizes for interconnecting multiple DH networks.

In (Guelpa, Mutani, Todeschi, & Verda, 2017) a GIS-based model is used to analyze the technical feasibility of possible expansion of existing DH networks. The aim of the paper is to analyze the reduction of fuel demands and CO<sub>2</sub> emissions as well as cost by combining DH expansion with end-use energy savings.

Guelpa, Mutani, Todeschi, and Verda (2018) present an approach for optimizing the incorporation of new buildings into pre-existing extensive district heating (DH) networks. The focal point of the study involves seeking the most efficient expansion strategy with the objective of minimizing pumping expenditures. To accommodate technical constraints effectively, they employ a fluid dynamic network model.

Bordin, Gordini & Vigo, 2016, developed a mathematical model to support DH system planning. The objective of the work was to the selection of an optimal set of new users to be connected to an existing thermal network, maximizing revenues and minimizing infrastructure and operational costs. Steady-state conditions of the hydraulic system and the main technical requirements of the real-world application are taken into account and the optimization problem is modeled through the application of the graph theory and MILP algorithm.

## **3. METHODOLOGY**

### **3.1. Introduction**

To achieve a shift towards clean energy and carbon neutrality by 2050 (European Commission, 2023), it is vital to employ comprehensive approaches and techniques that integrate detailed spatial and temporal data from various urban sources (Perera et al., 2023). In the context of this study, a robust energy efficiency model is carefully devised, encompassing a bottom-up Urban Building Energy Modeling (UBEM) framework alongside an innovative energy-saving scenario analysis.

By integrating these components, the research aimed to construct a highly sophisticated model that goes beyond conventional boundaries. Leveraging a black-box methodology, the study harnessed the potential of advanced Machine Learning algorithms. These algorithms are expertly employed to accurately predict both real-time and future heating demands, consequently facilitating the thoroughly planning of potential expansions within the District Heating Network.

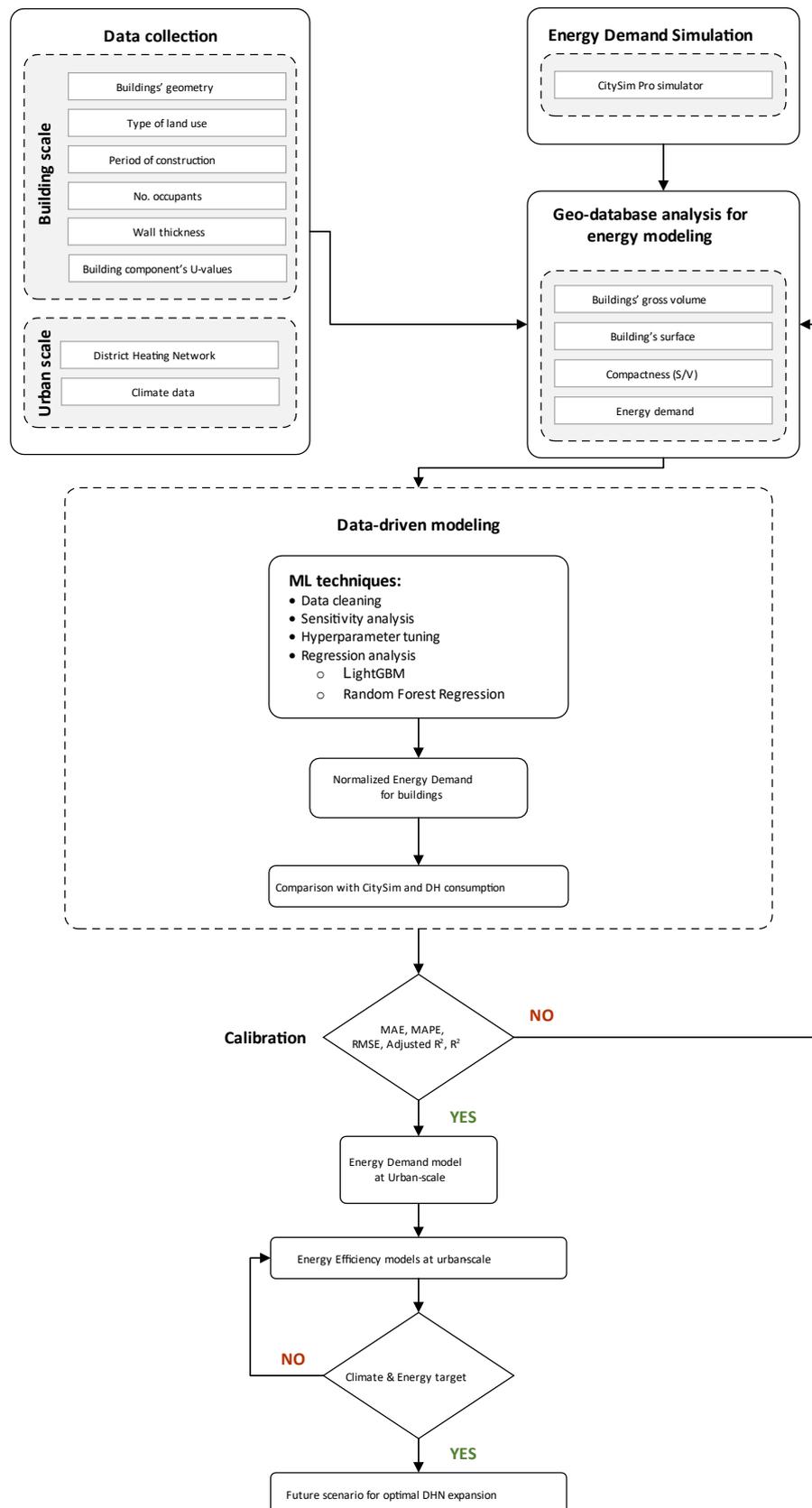
### **3.2. Urban building energy modeling**

In this study, the main method adopted is a bottom-up approach, which entails considering the detailed characteristics, components, and behaviors of individual buildings to estimate aggregated heating demand. To attain this objective, the research design is focused on implementing Urban Building Energy Modeling (UBEM) coupled with a black-box method, a technique that employs place-based strategies enjoying the benefits of Machine Learning (ML) models to replicate the energy efficiency of buildings within a given location. This strategy involved a thorough examination of the HD patterns of the entire building stock supported by Satom SA (An energy company that covers the municipalities of Monthey, Collombey-Muraz), taking into account the intricate attributes of each individual building. This combination of approaches allowed for a detailed analysis of HD and efficiency while considering the nuanced variations among different buildings within the urban context.

In the bottom-up process of UBEM, the following steps are involved:

- **Individual Building Modeling:** Detailed information about each building in the urban area is collected and represented in the model. This includes architectural features, insulation, occupancy patterns, and other factors that influence HD.
- **Incorporation of External Factors:** The model indirectly integrated external factors, such as weather conditions, solar radiation, and urban morphology, which influence HD patterns, by employing the CitySim Pro simulator to simulate HD for a sample population.
- **Energy-demand Simulations:** Heating demand of each building is simulated based on the specific characteristics of the buildings. With this objective in mind, the approach of black-box methodology is adopted, harnessing the advantages offered by machine learning models in its implementation.
- **Aggregation:** The HD data from individual buildings are aggregated to estimate the total HD for a group of buildings within a hypothetical network of meshes, or the entire urban area.
- **Temporal and Spatial Considerations:** The model took into account the variations in energy usage over different time scales (hourly and annual) and spatial scales (individual buildings, neighborhoods, and urban-scale).
- **Validation and Calibration:** The model's accuracy is validated and calibrated using historical HD data. This helped to ensure that the model's predictions align with real-world HD patterns.

Figure 3.1 serves as a detailed and intricate overview of the methodological framework meticulously implemented in the current research endeavor. This graphical representation offers a comprehensive understanding of the step-by-step evolution of the research journey. It encapsulates the complexities involved in data collection, preprocessing, algorithm integration, optimization strategies, and the subsequent evaluation process. This figure assumes the role of a navigational marker, underscoring the interconnectedness of the core elements within the research and providing a visual narrative of the research expedition.



**Figure 3.1.** Flowchart of the applied methodology: Energy Efficiency model utilizing an Urban Building Energy Model as its foundation

Table 3.1 presents the existing input data along with their respective utilization scopes, and it also outlines the tools employed to harness the potential of these input data. A Geographic Information System (GIS) is employed primarily to construct the physical representation of the structures and, secondarily, to pinpoint socio-economic factors. Among the various physical aspects, the compactness of the buildings (expressed as the surface-to-volume ratio, S/V) is computed using a robust GIS method. This involves creating a 3D model of the building roofs using a TIN extension. Through the "Polygon Volume (3D analyst)" tool, the software automatically determined the volume and roof area of each building. By adding the wall and slab surfaces, the overall building surface area is derived. This reliable technique allows for the accurate calculation of a pivotal variable in UBEMs – the S/V ratio – with a high degree of standardization and precision.

Moreover, the utilization of the CitySim Pro platform proved invaluable in the development of the UBEM. In the case study, the CitySim Pro's simulation of HD for connected buildings served as a reference point for enhancing the energy-use model of buildings that have the potential for connection. The overarching objective is to train the model effectively. To accomplish this objective, the cornerstone was "The Calibrated CitySim XML file for Satom". This file played a pivotal role in introducing essential physical and socio-economic parameters of the buildings. These parameters are coupled with the climate data file for the year 2021. The combination of these inputs within the CitySim Pro platform enabled the accurate incorporation of the case study's climatic conditions for the specified year.

In pursuit of the study's goal – the simulation of HD for individual buildings based on their distinct characteristics – the strategy of employing a black-box methodology is embraced. This approach leverages the strengths of machine learning models to achieve accurate predictions. From an array of predictive machine learning algorithms available, the Random Forest (RF) algorithm and Light Gradient Boosting Machine (LightGBM) are judiciously selected as the prime candidates due to their outstanding qualities and proven track record in handling complex, multidimensional datasets that align with the research's requirements.

**Table 3.1.** Available data sources in UBEM

| <b>Data</b>                                      | <b>Data type</b> | <b>Usage</b>   | <b>Source</b>   |
|--|------------------|--|---|
| <b>The surveyor's data (GRB)</b>                 | GeoPackage       | <ul style="list-style-type: none"> <li>Used as building index to link buildings' footprint to the right monitoring data</li> </ul>   | Idiap research institute  |
| <b>swissBUILDINGS3D_2.0</b>                      | MultiPatch       | <ul style="list-style-type: none"> <li>Used to model the buildings' roof 3D model</li> <li>Used to extract minimum z value, maximum z value and buildings' wall surface</li> </ul>   | <a href="https://www.swisstopo.admin.ch/">https://www.swisstopo.admin.ch/</a> |
| <b>swissBUILDINGS3D_3.0 Beta</b>                 | MultiPatch       | <ul style="list-style-type: none"> <li>Used to model the buildings' roof-top-surface and footprint.</li> <li>Used to extract minimum z value, maximum z value and building wall surface</li> </ul>   | <a href="https://www.swisstopo.admin.ch/">https://www.swisstopo.admin.ch/</a> |
| <b>The Calibrated CitySim XML file for SATOM</b> | XML file         | <ul style="list-style-type: none"> <li>Used to simulate HD of building stock</li> <li>Used to extract useful information listed below: <ul style="list-style-type: none"> <li>Buildings' land use</li> <li>Buildings' age</li> <li>Buildings' occupants</li> <li>Wall thickness</li> <li>Building components' U-value</li> </ul> </li> </ul> | Idiap research institute  |
| <b>Climate files for 2021</b>                    | CLI file         | <ul style="list-style-type: none"> <li>Used to simulate heating demand of building stock</li> </ul>  | Idiap research institute  |
| <b>Satom DHN</b>                                 | Raster file      | <ul style="list-style-type: none"> <li>Used to understand the boundaries of DHN service area</li> </ul>  | <a href="https://satomsa.ch/">https://satomsa.ch/</a>                         |
| <b>Satom DHN</b>                                 | CSV file         | <ul style="list-style-type: none"> <li>Used to draw a geo-localized DHN</li> </ul>   | Idiap research institute  |

It is important to highlight that throughout different stages of the analysis; the dataset underwent consistent testing using various methodologies to identify and eliminate outliers. This practice significantly aids in data-driven analysis, ultimately enhancing the accuracy of the results achieved.

Moving forward, an extensive breakdown of the RF and LightGBM algorithms, alongside their specific applications within this study, will be provided.

### **3.3. Black-box methodology**

The choice to employ a black box methodology in this study is influenced by two crucial factors: the availability of historical aggregated data and the remarkable predictive potential of Machine Learning (ML) techniques.

Firstly, the utilization of a black box approach is driven by the wealth of historical aggregated data of case study. This data, encompassing past heating demand patterns and related variables, forms a valuable foundation for modeling future HD. Secondly, the deliberate emphasis on integrating ML techniques stemmed from their unparalleled ability to predict future heating demand. ML algorithms possess an exceptional capacity to uncover intricate patterns, correlations, and non-linear relationships within data that may not be readily apparent through traditional methods.

By embracing these techniques, the study is positioned to capitalize on their predictive power, allowing to forecast HD with a level of accuracy that can greatly inform strategic decision-making and long-term planning, especially critical in the context of DHN expansion. This dual focus on historical insights and future predictability enhanced the robustness of the findings and equipped the proposed UBEM with a potent tool for shaping sustainable energy strategies within dynamic urban environments. Subsequently, a thorough breakdown of the Machine Learning techniques applied is provided.

#### **3.3.1. Random Forest (RF) regression**

Random Forest stands out as one of the most widely utilized and potent machine learning algorithms (Brownlee, 2016). The RF algorithm, introduced by Breiman (Breiman & Cutler, 2001), is a black-box approach founded on the decision tree, a pivotal technique for both classification and regression tasks (Gassar & Cha, 2020).

RF is an ensemble learning technique that comprises three predictors. In this approach, trees are constructed using diverse random features. It generates multiple decision trees by randomly selecting data and variables, ultimately

determining the dependent variable's class based on the consensus of numerous trees (Dalipi, Yildirim Yayilgan, & Gebremedhin, 2016). Averaging is employed to enhance predictive accuracy and manage overfitting. The size of these subsets is regulated by the "max\_samples" hyperparameter when "bootstrap=True" (which is the default setting), or the entire dataset is employed to build each tree if bootstrap is set to False (scikit-learn developers, n.d.).

It belongs to the category of ensemble machine learning techniques referred to as Bootstrap Aggregation or bagging (Brownlee, 2016). Bagging is an ensemble technique that involves training multiple models on distinct subsets of a training dataset and subsequently amalgamating the predictions generated by these individual models (Brownlee, 2020). Beyond that, the RF algorithm adds an additional layer of randomness during the tree growth process. Instead of seeking the optimal feature for node splitting as conventionally done, it looks for the best feature within a randomly chosen subset of features. This approach enhances the diversity of trees, leading to a trade-off between higher bias and reduced variance, ultimately resulting in a more improved model overall (Géron, 2022).

As each tree is constructed through the random selection of both data and variables, they become random trees. The combination of numerous such random trees forms a random forest. The term "forest" implies the utilization of multiple decision trees to enhance the classification of the dependent variable (Veza, Comoglio, Rosso, & Viglione, 2010).

In the scope of this research, the UBE benefited from the predictive capabilities of the RandomForestRegressor class found in the SciKit-learn library. This approach is geared towards conducting accurate predictive analysis within the urban energy context. To optimize the performance of the RF Regression, a Genetic Algorithm (GA) optimizer is applied, which fine-tunes the hyperparameters of the model. By utilizing the GA, the RF Regression is tailored to provide the best possible predictive accuracy while also mitigating the risk of overfitting.

To ensure the robustness and reliability of the model, a Cross-validation (CV) technique is integrated into the genetic optimization process. Cross-validation aided in assessing the generalization performance of the model by partitioning the dataset into training and validation subsets, enhancing the model's ability to

perform well on unseen data. This approach not only enhanced the predictive accuracy of the UBEM but also contributed to its ability to handle a variety of real-world scenarios and unforeseen variations.

During the concluding phase, a comprehensive sensitivity analysis is performed on the dependent variables. This analysis capitalized on the advantages provided by the "feature\_importances\_" attribute of the RandomForestRegressor class. Through this critical step, the influential dependent parameters crucial for constructing accurate energy-use models are successfully discerned. This process involved identifying the most impactful features that contribute to the predictive power of the models, ultimately enhancing the overall reliability and relevance of the outcomes.

### **3.3.2. Genetic Algorithm (GA)**

Searching for optimal hyperparameters is often a challenging endeavor within machine learning endeavors. As the intricacy of deep learning methods expands alongside their popularity, the need for a streamlined and effective automated process for hyperparameter tuning has become more imperative than before (Akiba, Sano, Yanase, Ohta, & Koyama, 2019).

When it comes to hyperparameter tuning, random and grid search strategies are presently the prevailing techniques for refining the hyperparameters of numerous machine learning models (Bergstra & Bengio, 2012; Bergstra, Yamins, & Cox, 2013). Yet, these methods are considered simplistic as they explore all data points within the hyperparameter space without considering the training setup. As a result, these search methodologies are deemed inefficient, resulting in elevated computational complexity and increased costs associated with hyperparameter optimization (Raji et al., 2022).

The genetic algorithm (GA) is an optimization technique inspired by natural selection (Michalewicz, 1992). GA imitates the principle of natural selection known as survival of the fittest. It was introduced by J.H. Holland in the year 1992 (Katoch, Chauhan, & Kumar, 2021). They strive to replicate the evolutionary process of solutions across multiple generations in order to ultimately discover an optimal or

close-to-optimal solution for optimization problems (Di Francescomarino et al., 2018).

This approach involves iteratively applying genetic operators to individuals within the population to generate new populations. Key elements of GA include chromosome representation, selection, crossover, mutation, and fitness function calculation. The selection, crossover, and mutation processes are repeated on the current population until the new population is complete (Katoch, Chauhan, & Kumar, 2021).

In this present study, a Genetic Algorithm (GA) is utilized, with predefined parameters of a population size of 20, 10 generations, and a mutation rate of 0.1. This GA is employed to identify a set of hyperparameters for RF regression that holds potential, encompassing 'n\_estimators', 'max\_depth', 'min\_samples\_split', and 'min\_samples\_leaf'.

### **3.3.3. Cross Validation (CV)**

A core challenge in machine learning revolves around achieving a precise approximation of the generalization error for a model trained on a limited dataset. The accurate assessment of a model's accuracy holds paramount importance not only for evaluating its capacity to generalize but also for selecting the most suitable algorithm from a diverse array of learning algorithms (Rao, Fung, & Rosales, 2008). Cross-validation (CV) serves multiple purposes, functioning as a tool to either approximate the generalization error of a specific model or to aid in model selection by identifying the model among several options with the lowest estimated generalization error (faqs.org, 2014).

According to scikit-learn developers: “learning the parameters of a predictive function and subsequently testing it on the same dataset represents a methodological flaw. Relying on such a model could yield a flawless score due to its ability to merely replicate the provided labels, yet it would lack the capability to make meaningful predictions on new, unseen data. This phenomenon is referred to as overfitting. To counter this, in the context of supervised machine learning

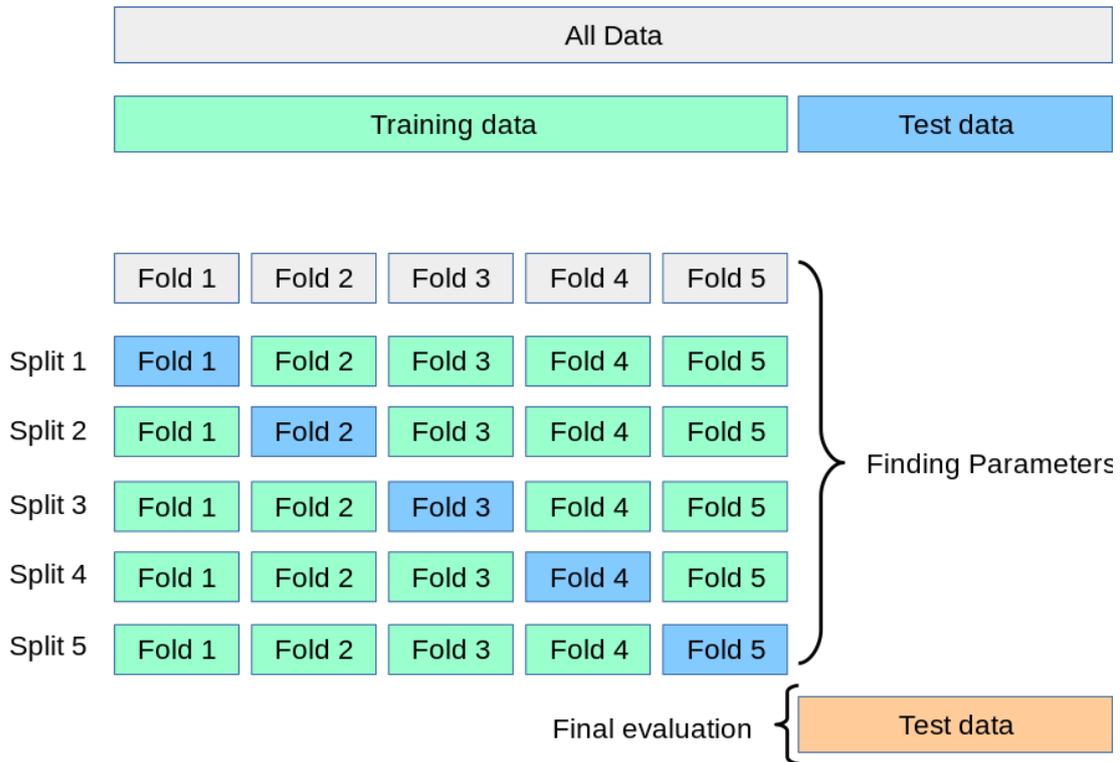
experiments, it is customary to segregate a portion of the available data to form a distinct test set, denoted as  $X_{\text{test}}$  and  $y_{\text{test}}$ .

However, the risk of overfitting persists even when evaluating diverse configurations, known as "hyperparameters," for estimators. The parameters can be adjusted until the estimator exhibits optimal performance, inadvertently leading to the leakage of information about the test set into the model. Consequently, evaluation metrics may no longer accurately reflect generalization performance. To address this, another portion of the dataset can be designated as a "validation set." Here, training occurs on the training set, and evaluation is carried out on the validation set. If the experiment proves successful, a final evaluation is performed on the test set.

Yet, dividing the available data into three sets significantly reduces the number of samples available for training the model, and the outcomes can be influenced by the specific random selection of the (train, validation) sets. To mitigate this issue, a technique called cross-validation is employed. While a test set remains reserved for ultimate evaluation, the validation set becomes unnecessary when conducting CV."

CV is a method for detecting bias and variance in which instead of manipulating hyperparameter values, the focus is on varying the quantity of training data. (Kyriakides & Margaritis, 2019). Within the realm of regression analysis, CV's utility encompasses tasks like determining the optimal count of underlying features and gauging the average prediction error. (Bro, Kjeldahl, Smilde, & Kiers, 2008).

K-fold CV entails partitioning the data into  $k$  segments of roughly equal size. The model is trained  $k$  times, with each iteration excluding one of the segments from training, employing the omitted segment solely for calculating the desired error criterion. When  $k$  is equivalent to the sample size, this practice is referred to as "leave-one-out" CV. "Leave- $v$ -out" represents a more intricate and resource-intensive variant of CV, where all conceivable subsets of  $v$  cases are omitted (faqs.org, 2014). Figure. 3.2 offers a clear depiction of how the Cross-Validation (CV) technique operates on data within the algorithm.



**Figure 3.2.** Schematic illustration of how CV technique is being employed in ML algorithms<sup>7</sup>

The performance metric derived from k-fold CV is the average of the values computed during the iterations. While this approach can be computationally intensive, it minimizes data wastage (unlike when using a fixed validation set), which proves advantageous in scenarios with limited sample sizes, such as inverse inference problems (scikit-learn developers).

Within this present study, the CV technique is explored through the segmentation of the training dataset into 3, 5, and 10 folds, which is determined based on the dataset's frequency. By comparing the outcomes of these different fold sizes, the optimal K value is determined. It's important to highlight that in situations where the training dataset exhibited a scarcity in terms of frequency, introducing energy-saving scenarios to the training dataset could play a pivotal role in augmenting its size. Consequently, this augmentation would potentially lead to heightened accuracy in predicting unseen data points.

<sup>7</sup> scikit-learn developers, n.d., [https://scikit-learn.org/stable/modules/cross\\_validation.html#cross-validation-iterators](https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation-iterators)

### 3.3.4. Light Gradient Boosting Machine (LightGBM)

Gradient Boosting Decision Tree (GBDT) (Friedman, 2001) stands as a prevalent machine learning algorithm, recognized for its efficiency, precision, and comprehensibility. GBDT demonstrates remarkable achievements in diverse machine learning endeavors, encompassing multi-class classification (Li, 2012), click prediction (Richardson, Dominowska, & Ragno, 2007), and learning to rank (Burgess, 2010).

LightGBM (Light Gradient Boosting Machine) is a highly optimized implementation of the gradient boosting algorithm, a machine learning technique that leverages a collection of weak learners, often in the form of decision trees, to tackle regression or classification challenges. Unlike conventional ensemble methods, gradient boosting introduces a sequential addition of these weak learners to the model. This approach ensures that each subsequent learner is tailored to the residuals left by its predecessor, enhancing the model's overall predictive power in a systematic manner (Todeschi, Boghetti, Kämpf, & Mutani, 2021). To enhance implementation efficiency, LightGBM employs a leaf-wise growth strategy for its decision trees, as opposed to examining all preceding leaves for every new leaf. This implementation technique is referred to as the histogram approach (Al Daoud, 2019). According to Microsoft Corporation, LightGBM presents the following merits:

- Accelerated training velocity and heightened efficiency.
- Diminished memory consumption.
- Enhanced precision.
- Compatibility with parallel, distributed, and GPU-based learning.
- Proficient in managing expansive datasets.

In this study, the LightGBM algorithm, developed by the Microsoft team, is utilized for predictive analysis within the UBEM. To achieve this goal, the Optuna optimizer is employed to fine-tune the hyperparameters of the LightGBM model. In order to optimize the performance of LightGBM regression and prevent overfitting, a Cross-validation technique similar to that used in RF regression is incorporated into the Optuna algorithm. This approach aimed to strike a balance between maximizing

accuracy and avoiding overfitting while optimizing the LightGBM model for predicting HD in urban buildings.

Similarly, mirroring the procedure for RF, a sensitivity analysis is also carried out on the dependent variables. This analysis is executed by making use of the advantages offered by the "feature\_importances\_" attribute within the "LightGBM" class. In the process of implementing LightGBM, two distinct training APIs are taken into consideration. For the hyperparameter tuning phase, the "train" method is employed, as it allowed for the integration of a pruning callback. Conversely, when conducting the sensitivity analysis, the "LGBMRegressor" method is utilized, primarily due to its inclusion of the "feature\_importances\_" attribute.

### **3.3.5. Optuna**

OPTUNA, a sophisticated framework for parameter optimization, represents a significant advancement in hyperparameter tuning. Introduced by Akiba, Sano, Yanase, Ohta, and Koyama in 2019, the Optuna framework excels in both the searching and pruning phases, demonstrating remarkable efficiency (Ekundayo, 2020).

Optuna is a dedicated software framework designed for autonomous hyperparameter optimization in machine learning. Its user interface operates on an imperative, define-by-run approach. With the define-by-run API, the code structure developed using Optuna is notably modular, and the construction of search spaces for hyperparameters can adapt dynamically (Akiba, Sano, Yanase, Ohta, & Koyama, 2019). OPTUNA employs its internal memory data structure as the default storage repository unless an alternative choice is indicated. A key advantage of OPTUNA, particularly from a user's perspective, is its seamless applicability for lightweight tasks (Srinivas & Katarya, 2022).

Optuna frames hyperparameter optimization as the process of minimizing or maximizing an objective function that takes a set of hyperparameters as input and yields a validation score. Each optimization procedure in Optuna is referred to as a "study," and each evaluation of the objective function is termed a "trial." Optuna incrementally constructs the objective function through interactions with the trial

object. The trial methods dynamically establish search spaces during the runtime of the objective function. To generate hyperparameters for each trial, users employ the 'suggest API' within the objective function, which statistically samples hyperparameters based on the history of prior trial evaluations. This process enables users to express a broad range of parameter spaces using standard Python syntax, employing loops and conditional statements. This mechanism allows for the representation of even complex parameter spaces in a straightforward and intuitive manner (Akiba, Sano, Yanase, Ohta, & Koyama, 2019).

The benefits provided by Optuna's hyperparameter optimization encompass the straightforward determination of optimization duration and seamless integration of outcomes. It efficiently conserves computational resources by early pruning of trials with below average performance. Additionally, its ease of implementation and compatibility with various machine learning frameworks make it a preferable choice for utilization in this research (Ekundayo, 2020).

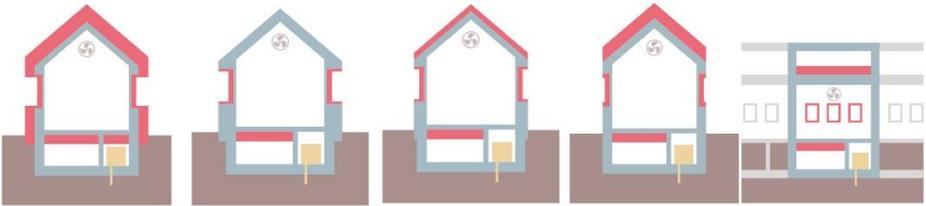
In this ongoing research, the Optuna optimizer is employed to explore a potentially effective collection of hyperparameters for LightGBM regression. These encompassed parameters like 'num\_leaves', 'max\_depth', 'min\_data\_in\_leaf', 'feature\_fraction', 'bagging\_fraction', 'bagging\_freq', 'lambda\_l1', 'lambda\_l2', and 'max\_bin', and the exploration spanned across 500 trials. Furthermore, a pruning callback is strategically implemented within the optimization loop to bolster the precision and efficiency of the tuning process.

### **3.4. Energy saving scenario**

The UBEM in this study is applied across three distinct time-based stages. Initially, the analysis embarked on an examination of the existing status of the DH system, focusing on the buildings already connected, to determine the most efficient UBEM approach. Moving to the second phase, the study delved into evaluating the near-term potential for expanding the district heating system, with a specific emphasis on optimizing the utilization of CHP capacity without necessitating additional enlargement of the DHN. Finally, in the third phase, the investigation shifted

towards assessing the viability of long-term DHN expansion, considering the expansion of the network through the incorporation of new pipes, while ensuring maximum coverage of the CHP capacity.

Returning to the concept of exploring the feasibility of expanding the DHN in the case study, the retrofitting methods adhered closely to Minergie's established guidelines outlined in Figure 3.3 Minergie presents a streamlined and simplified procedure for validating the energy retrofitting of residential structures, utilizing five distinct system approaches that facilitate the attainment of Minergie certification with minimal complexity. All variations offer the additional benefits inherent in Minergie's approach: the elimination of reliance on fossil fuels, enhanced energy efficiency, and elevated comfort standards. Furthermore, most regions provide incentives to encourage the adoption of these practices. These technologies introduced on various typologies of buildings, categorized by their ages, with the primary aim of enhancing the heat transmission of at least one component of the buildings, through the strategic application of technology.



|                                    | System 1  | System 2   | System 3   | System 4   | System 5   |
|------------------------------------|---|--|--|--|--|
| <b>Building envelope CECE</b>      | <b>B</b>  | <b>C</b>   | <b>C</b>   | <b>C</b>   | <b>C</b>   |
| <b>U value (W/m<sup>2</sup>/K)</b> | Roof ≤ 0.17<br>External wall ≤ 0.25<br>Window ≤ 1.0<br>Flooring ≤ 0.25  | Roof ≤ 0.30<br>External wall ≤ 0.40<br>Window ≤ 1.0<br>Flooring ≤ 0.25   | Roof ≤ 0.25<br>External wall ≤ 0.50<br>Window ≤ 1.0<br>Flooring ≤ 0.25 | Roof ≤ 0.17<br>External wall ≤ 0.70<br>Window ≤ 1.0<br>Flooring ≤ 0.25 | Roof ≤ 0.17<br>External wall ≤ 1.10<br>Window ≤ 1.0<br>Flooring ≤ 0.25 |
| <b>Heat production</b>             | Renewable energies (e.g. heat pump, district heating, wood)   |  |  |  |  |
| <b>Air exchange</b>                | Basic ventilation allowed, heat recovery (RC) recommended<br><a href="http://www.minergie.ch/good-quality-ambient-air">www.minergie.ch/good-quality-ambient-air</a> |  |  | Mandatory RC   |  |
| <b>Comfort stowage</b>             | Check summer thermal protection<br><a href="http://www.minergie.ch/summer-thermal-protection">www.minergie.ch/summer-thermal-protection</a>                         |  |  |  |  |
| <b>Electricity</b>                 | Photovoltaic system recommended   | 40% reduction in consumption for household appliances and lighting or photovoltaic system with min. 5Wp per m <sup>2</sup> of energy reference surface |  |  |  |

**Figure 3.3.** Minergie modernization system for energy retrofitting of residential buildings<sup>8</sup>

<sup>8</sup> <https://www.minergie.ch/it/standard/ammodernare/ammodernamento-di-sistema/>

In the pursuit of finding the best incentives for retrofitting that would result in the most efficient use of energy, this study employed a promising approach to create a roadmap for the ideal expansion of DHN by the year 2050.

This approach centered on introducing energy saving scenario for individual buildings taking into account their connection status, together with the socio-economic status of the context. It also factored in a worst-case scenario characterized by the peak demand of the buildings based on their scope of renovation. This demand is then compared against the network's maximum capacity to support HD during that specific hour, with the most favorable scenario being selected based on the premise that the existing CHP infrastructure along with pipes could accommodate the maximum feasible number of connectable buildings.

To achieve abovementioned goal, a grid network with mesh dimensions of 500m by 500m established. Within this framework, the energy intensity of the buildings situated in each mesh is computed on an hourly basis, coupling with the density of the buildings in each mesh. Afterward, meshes were given priority to determine the direction of DHN expansion. The buildings that could be connected within each mesh were also selected randomly, and following this selection process, while considering an annual renovation rate of 1%, the optimal buildings to be connected were identified within three-year intervals.

To find the promising retrofitting scenario, a total of 5 distinct scenarios are introduced, guided by the following principles aimed at determining the peak HD of buildings.

- The initial scenario involved connected buildings without retrofitting alongside connectable buildings that were also not retrofitted.
- The second scenario consisted of connected buildings without retrofitting paired with connectable buildings that underwent retrofitting.
- In the third scenario, connected buildings underwent retrofitting while the connectable buildings remained non-retrofitted.
- The fourth scenario encompassed connected buildings that were retrofitted alongside connectable buildings that also underwent retrofitting.

- Lastly, the fifth scenario investigated an extreme situation involving both connected buildings and those that could be connected, all of which had undergone global retrofitting.

## **4. DISCUSSION**

### **4.1. Introduction**

This research endeavors to leverage the capabilities of Machine Learning (ML) techniques to enhance the predictive accuracy of energy demand models. The focus is on utilizing ML's potential for superior performance in predicting energy usage patterns. However, to extract the maximum benefits from ML techniques, certain preliminary measures are essential to ensure their optimal functioning. These preparatory steps are crucial to attain the highest predictive performance levels.

Moreover, the application of ML techniques extends the horizons of energy use models, enabling the development of comprehensive approaches to devise energy-saving scenarios. By integrating ML, the research aims to create a more encompassing methodology that offers a broader perspective on energy conservation.

The primary objective of this chapter is to introduce a strategic pathway that capitalizes on the comprehensive approach employed in this research. It emphasizes the significance of aligning with the prerequisites of ML techniques to yield more accurate and meaningful results. This strategic alignment ensures that the full potential of these advanced techniques is harnessed, leading to a nuanced understanding of energy demand patterns and the formulation of effective energy-saving strategies.

### **4.2. Case study**

Established in 1972 and operational since 1976, Satom SA is an energy company that stands as a trailblazer in its field. Recently, it has introduced a groundbreaking and eco-conscious initiative at its Monthey and Villeneuve sites. This innovative approach taps into the potential of biomass waste from the Monthey location and plastics from the Villeneuve site. By transforming this substantial waste material into water vapor, the company propels a steam turbine to produce clean, renewable electricity. This not only addresses the negative environmental

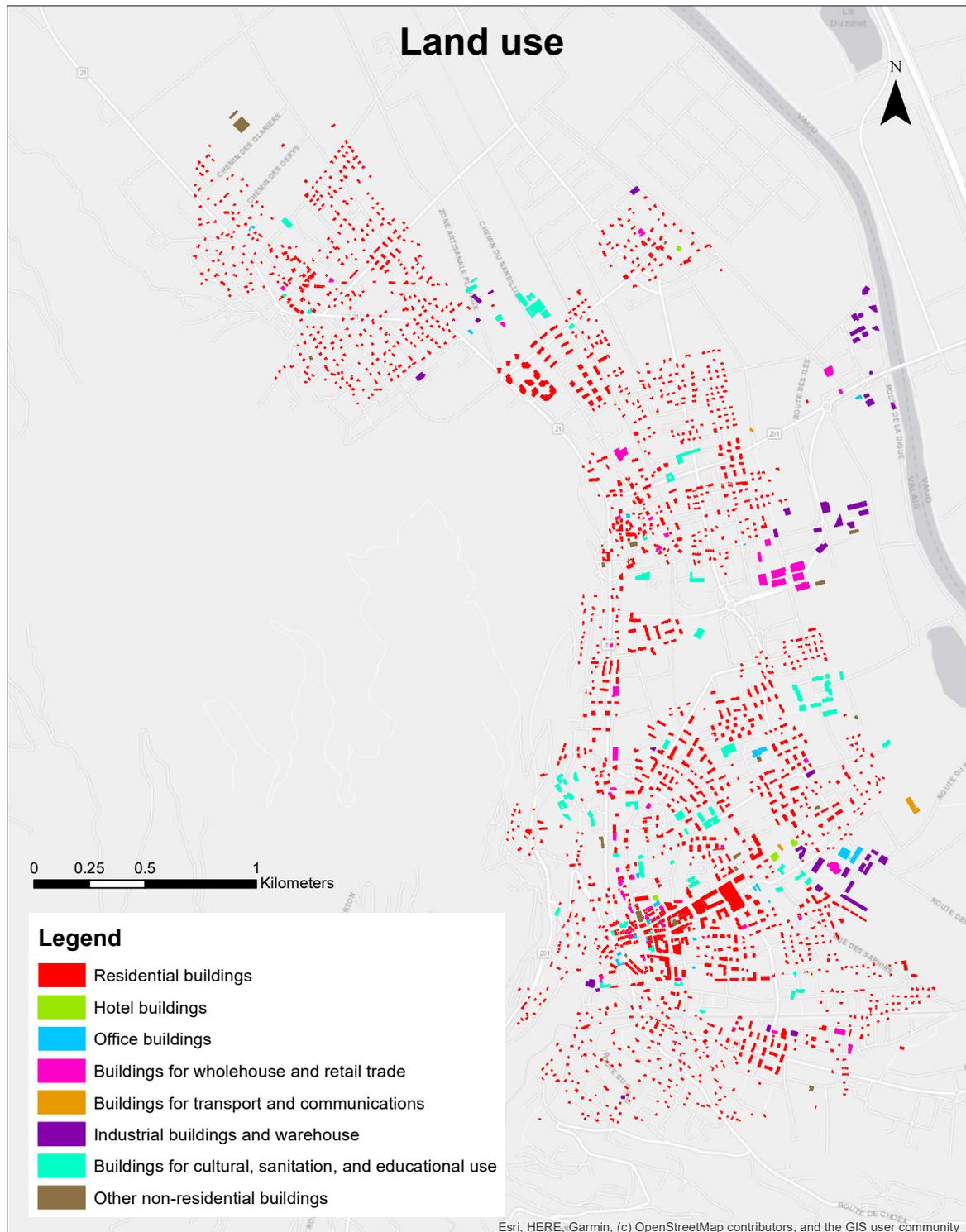
consequences of waste disposal but also plays a significant role in shaping a more sustainable energy trajectory.

A central component of Satom's energy strategy involves their waste heat recovery mechanism. In traditional setups, a significant amount of valuable heat would escape into the environment during electricity generation. However, Satom has innovatively developed a thermal network that captures steam drawn from the turbine and collected from emissions. This network effectively redirects this otherwise lost heat. Through the utilization of this thermal network, the recuperated waste heat is efficiently directed into neighboring buildings, providing them with an eco-conscious heating alternative. This approach not only maximizes energy efficiency but also markedly curbs the collective carbon footprint of the local community.

Satom has established an extensive underground network designed strategically to distribute thermal energy, covering a substantial area that includes the municipalities of Collombey-Muraz and Monthey. This impressive subterranean infrastructure empowers Satom to deliver heating services to numerous residential areas, spanning several kilometers and effectively meeting the heating demands of a significant population.

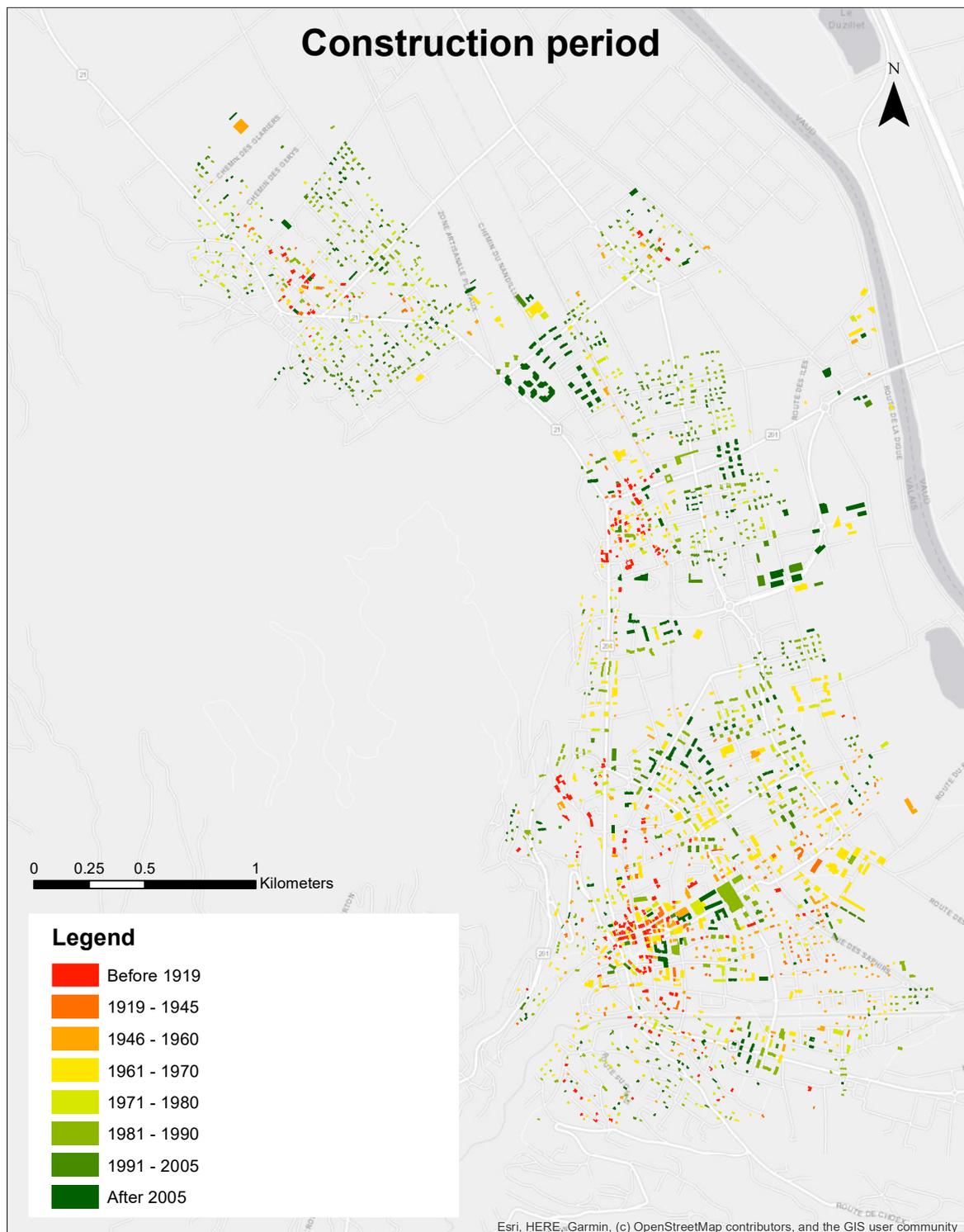
Satom's energy system is technically characterized by injecting 91 GWh of energy into the network annually. This network consists of 79 kilometers of pipes, including 484 operational substations. With this considerable capacity, Satom SA is capable of heating a total living space of 850,000 m<sup>2</sup>. The research involved addressing a primary limitation which is economic factors. To address economic constraints, the study included residential buildings with S/V values under 0.8 m<sup>-1</sup> and non-residential buildings with volume over 600 m<sup>3</sup>. In the targeted municipalities, a total of 2603 buildings were investigated (around 8.74 Mm<sup>3</sup>), out of which 2283 (approximately 7.37 Mm<sup>3</sup>) were economically viable for network connection. The building types based on usage are presented in figure 4.1, and the distribution of buildings' ages is shown in figure 4.2. Within this eligible cluster, there were 2057 residential buildings (about 5.88 Mm<sup>3</sup>) and 226 non-residential buildings (approximately 1.49 Mm<sup>3</sup>), with 370 (around 2.74 Mm<sup>3</sup>) and 61 (approximately 0.54 Mm<sup>3</sup>) successfully connected, respectively (Shown in Figure 4.3). When combined,

the annual heating demand for all connected buildings of both types, totals about 80.57 GWh/Y.



**Figure 4.1.** Classification of land use for the sampled buildings.

Emphasizing the significance of the matter, it should be noted that the volume distribution of residential structures was subjected to statistical analysis, leading to the classification of the buildings into two distinct groups. The initial group comprises 2018 buildings as normal buildings, collectively encompassing a volume



**Figure 4.2.** Construction period categories for the sampled buildings.



Furthermore, considering that diverse information from multiple databases is employed to construct the physical models of buildings, variations arise in terms of accessible dependent variables. The key factor distinguishing these variables is the availability of data regarding building occupants. Within this context, residential and nonresidential buildings themselves are segregated into distinct clusters. Consequently, they are treated independently when forecasting energy demand. There is a disparity in the occurrence of available occupancy profiles for residential and nonresidential buildings. Specifically, there are 1026 residential buildings and 179 nonresidential buildings with these profiles accessible. Correspondingly, there are 1031 residential buildings and 47 nonresidential buildings that lack occupancy profile data.

### **4.3. Buildings' physical model**

A crucial aspect of adopting urban building energy models is the development of a comprehensive physical representation of buildings, encompassing their distinctive attributes. This aspect gains even greater significance when integrating black-box methodologies into these models. This is due to the strong reliance of such techniques on historical data for approximating overall HD. It's evident that, in this scenario, having a wealth of data to inform the analysis would substantially enhance the precision of constructed urban energy models. Consequently, this heightened accuracy would lead to more closely aligned predictions of future heating demands with real-world outcomes.

Keeping this perspective in mind, the aim was to integrate and compile data from various comparable sources to validate the accuracy of the physical model of the buildings. To be more precise, this involved procuring data pertaining to the buildings' geometric attributes, including factors like "Gross heated volume", "Gross heated Surface", and "Compactness Ratio". The main input data contributed in predicting energy use model of buildings in current research is listed in table 4.1.

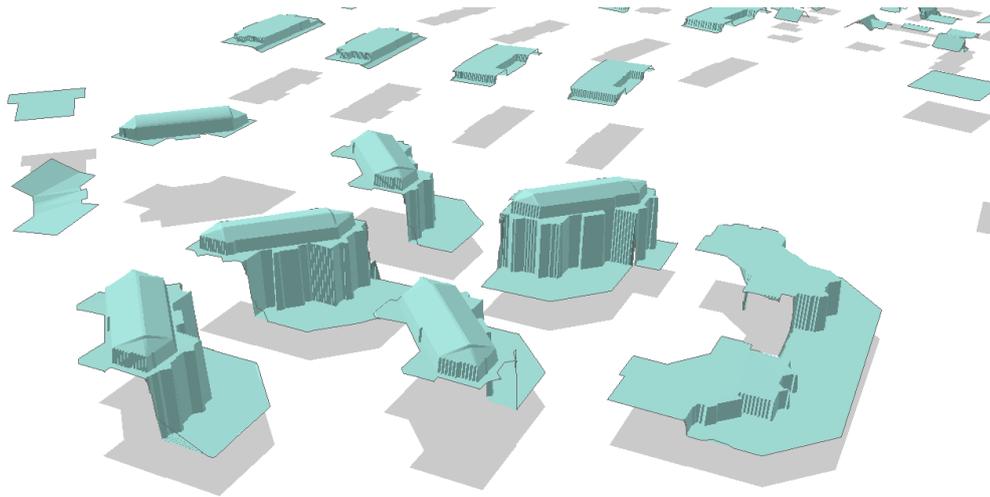
*Table 4.1. Main input parameters in building of energy-use models*

| <b>Input data</b>   | <b>Source</b>       | <b>Tools</b> |
|---|---------------------|--------------|
| <b>Construction period</b>  | The surveyor's data | GIS          |
| <b>Use type</b>   | The surveyor's data | GIS          |
| <b>Building's volume</b>  | Calculated          | GIS          |
| <b>Building's surface</b>   | Calculated          | GIS          |
| <b>S/V ratio</b>  | Calculated          | GIS          |
| <b>Building's occupants</b>   | XML file for SATOM  | None         |
| <b>Wall thickness</b>   | XML file for SATOM  | None         |
| <b>U<sub>roof</sub>, U<sub>wall</sub>, U<sub>slab</sub>, U<sub>Window</sub></b> | XML file for SATOM  | None         |

As indicated within table 4.1, the primary databases employed for extracting the physical dimensions of buildings consist of surveyor's data, swissBUILDINGS3D\_2.0, and swissBUILDINGS3D\_3.0 beta. The surveyor's data was present in the shapefile format and contained details regarding the construction period, land usage of buildings, and most crucially, a unique identifier for each plot of land, which served as the central point for attaching supplementary information to the dataset. On the other hand, swissBUILDINGS3D\_2.0 and swissBUILDINGS3D\_3.0 beta are multi-patch files that store 3D information in a collection of patches, summarizing the boundaries of a 3D object in a single record. The ArcGIS software furnished a specialized toolkit that greatly facilitated the manipulation and extraction of information from such databases.

Since multi-patch files offer better precision and versatility in extracting the physical geometries of buildings, it is valuable to examine the techniques and software employed for calculating the gross heated volume and surface of buildings. In terms of volume computation, the 3D representation of the buildings' roofs shown in figure 4.4 created, which follows a specific sequence ("multi-patch to raster (Conversion)", "raster to point (Conversion)", and "create TIN (3D Analyst)"). Afterward, the "add z information (3D Analyst)" function is applied, assigning the minimum z value of buildings to the layer representing their footprints. In the final stage, a specialized feature within ArcGIS, known as the "polygon volume" tool, is employed to determine the volume of buildings. The computation process of this tool involves initially identifying the shared boundaries between the building's footprint and its 3D roof model, followed by calculating the enclosed volume situated between the footprint and the roof surface. This particular tool proves

highly effective in accurately computing the gross heated volume of all buildings, especially those with multiple levels and varying configurations. Moreover, this tool is intuitively designed to also calculate in this case the roof surface of buildings using the provided TIN data. By incorporating the footprint surface, the total building surface can be computed, contingent upon the prior calculation of wall surfaces. This semi-automated approach to volume and surface calculation results in remarkably precise estimations of these parameters.



*Figure 4.4. Representation of buildings' 3D roof model*

When determining the surface area of building walls, it's essential to consider that for buildings with recesses, as depicted in figure 4.4, a portion of the wall surface can be automatically calculated if it's visible in the 3D representation of the building's roof surface. To prevent redundant computations in such instances, it's sufficient to compute the wall surface only for the sections of walls that are in contact with the outer perimeter of the buildings. This can be achieved by assigning the maximum z value of the buildings to the outer walls, which can be derived from the points obtained by converting the raster model of the roof surface.

The information about building occupants, wall thickness, and U-values for various building components was obtained directly by parsing in "XML file for the buildings in Satom." Subsequently, this data was linked to the building footprint layer through the utilization of the unique key identifier.

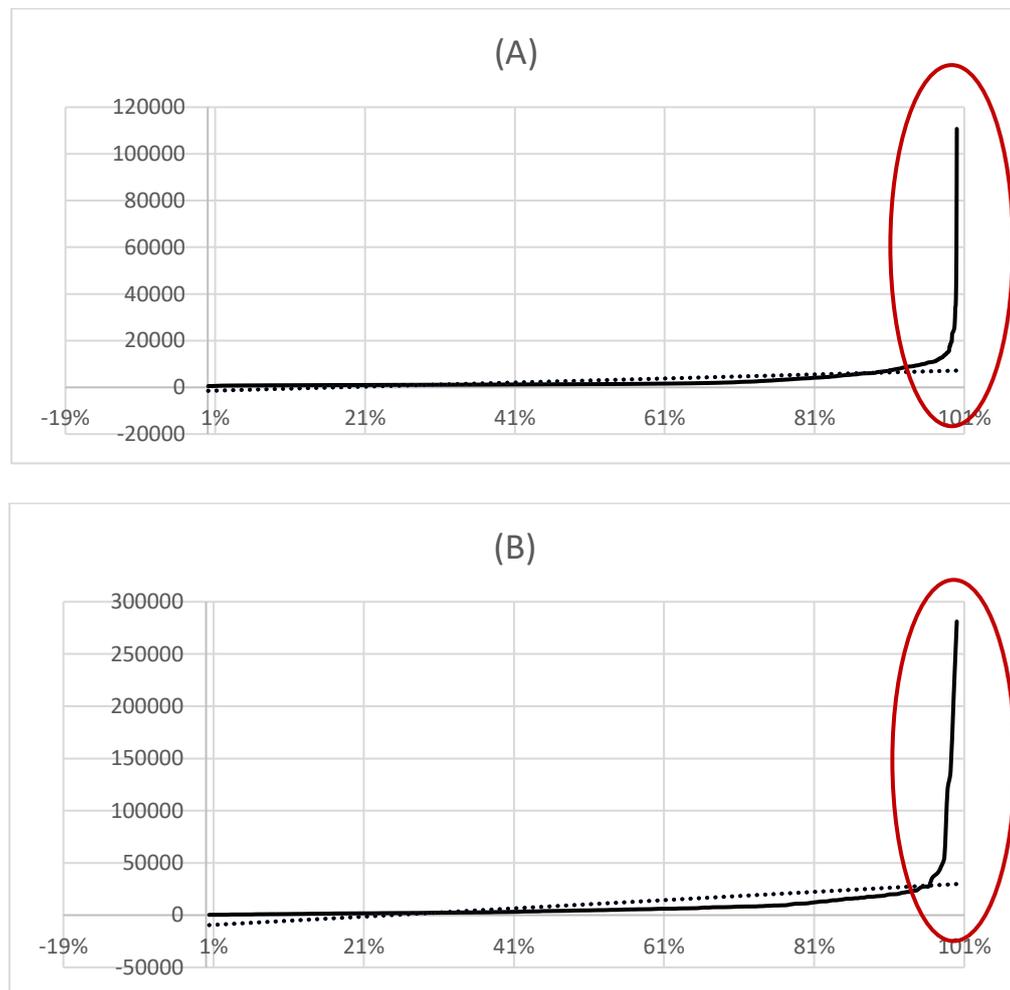
XML (Extensible Markup Language) is a file format that is commonly used for storing and structuring data in a hierarchical format. It employs a set of tags to label and organize pieces of data, making it easily readable for both humans and machines. In this context, the XML file serves as a primary input within CitySim Pro software and contains specific information about the buildings neatly organized within relevant tags, and by parsing it, relevant data can be extracted and utilized for various purposes, such as energy modeling or analysis.

Prior to any analysis being conducted on the sample dataset, the dataset needs to be cleaned of any outliers that could hinder the precision of the analysis and future predictions. Therefore, in this initial phase, buildings were first examined based on their "Compactness Ratio". The relation between the S/V ratio and building typology is such that high-rise buildings are characterized by a lower S/V ratio, while buildings in the form of detached houses for single families have a higher S/V ratio. To ensure that the connection of potentially connectable buildings adds value to the DHN expansion from an economic perspective, constraints were introduced for both residential and nonresidential buildings based on the S/V ratio. In this light, residential buildings with an S/V ratio above 0.8 were deemed unfeasible for connection to the system and were consequently excluded from the analysis. As for nonresidential buildings, this constraint was applied to buildings with a volume of less than 650 m<sup>3</sup>, taking into consideration both the compactness ratio and the volume of connected buildings.

Furthermore, the distribution of buildings based on their volume was analyzed for both residential and nonresidential categories. As shown in Figure 4.5, there are outliers that should be excluded to avoid potentially negative impacts on the final precision of the energy use model.

For non-residential buildings, outliers were excluded since none of the abnormal non-residential buildings were already connected to the DHN. Conversely, concerning abnormal residential buildings, there were a total of 39 buildings, out of which 31 were already connected, and only 8 were not yet connected. These buildings mainly have a dual purpose, serving both residential and commercial activities. Considering the majority were already connected, they were grouped

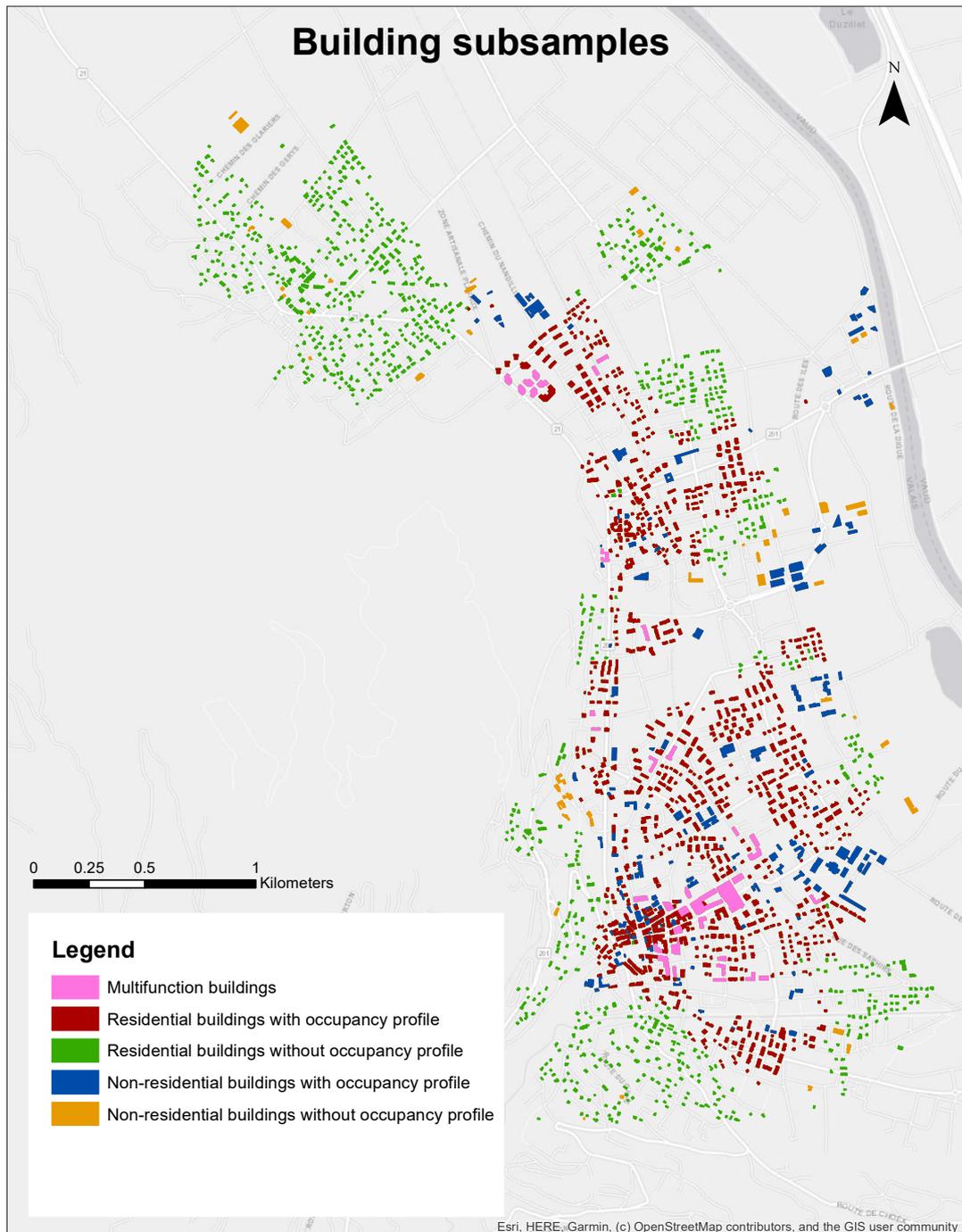
under a specific category labeled as "multifunctional buildings," and their energy use model was also trained using machine learning models.



**Figure 4.5.** (A) The distribution of "Residential" buildings by volume (B) The distribution of "Non-residential" buildings by volume

Ultimately, it's worth noting that the surveyors' database and, in a corresponding manner, the XML file for buildings in Satom primarily encompassed buildings situated in proximity to the existing DHN. To analyze the potential expansion of the DHN in the distant future, buildings were inserted from swissBUILDINGS3D\_3.0 beta. As a result, buildings near the network had occupancy profiles, as the XML file contained information regarding the number of occupants. However, this critical information, which serves as a pivotal dependent variable in building energy use models, was unavailable for buildings sourced from swissBUILDINGS3D\_3.0 beta.

Consequently, since machine learning model architectures depend on both dependent variables and sets of hyperparameters, there were two subsamples for each cluster of residential and nonresidential buildings in energy use modeling: one with information about occupants and one without such information. Consequently, there were a total of five building subsamples in energy use modeling, as illustrated in Figure 4.6.



**Figure 4.6.** Spatial distribution of building subsamples in the analysis

## 4.4. Black-box method

### 4.4.1. Dataset preparation

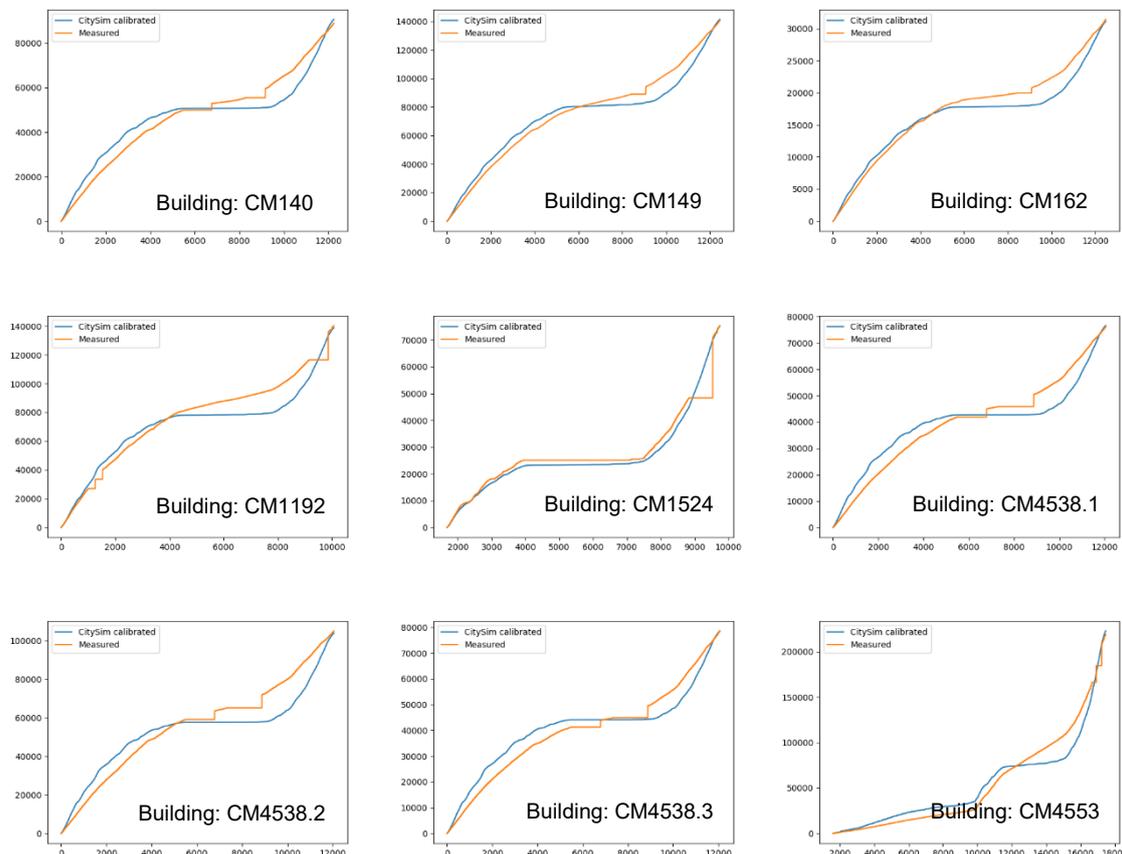
Prior to exploring the black-box methods employed in the present study, the initial step involved simulating the heating demand of buildings using CitySim Pro. To execute this simulation, it was imperative to adjust the XML file to match the measured data provided by Satom SA. The measured data encompassed various substations, each potentially serving one or more buildings, and was collected at different time intervals, typically in 3-hour increments. However, for the purposes of this study, it was crucial to have hourly HD data for buildings since energy-saving scenarios required considering the peak demand within an hour.

Returning to this concept, as illustrated in Figure 4.7, the XML file was calibrated on a building-by-building basis. The calibrated parameters included “Air changes per hour” and the “Minimum outside temperature”, both adjusted within the XML file based on the calibration results. This calibration process enabled the subsequent simulation of buildings in CitySim Pro, allowing for the extraction of their hourly energy requirements for the year 2021, serving as the benchmark for further analysis.

It's important to highlight that, in certain subsets of buildings, particularly multifunctional and nonresidential ones, the number of buildings connected to the network was limited. This limitation had a significant impact on the predictive accuracy of the machine learning models. In the instance of multifunctional buildings, where only 31 were connected, it wasn't even statistically significant to train an energy use model. This situation was exacerbated by the necessity in ML models to split data into training and testing datasets, resulting in an extremely limited number of buildings included in the training set.

To address this limitation and expand the dataset used to train energy demand models, an approach was employed. This approach involved introducing various retrofitting technologies to the buildings and then pairing the connected buildings without renovations to the same buildings that had undergone different retrofitting measures within each subset. The rationale behind this approach stemmed from the fact that, in each retrofitting scenario, there was at least one dependent

variable, namely the U-value determined by the retrofitting technology, together with the independent variable that differed from the other scenarios. Implementing this initiative allowed for a significant increase in the size of the training dataset and effectively enhanced the accuracy of the predictive models.



**Figure 4.7.** Calibration of XML file building-wise according to measured data

The current research incorporates various retrofitting scenarios, including Roof Insulation, Slab Insulation, Wall Insulation, Windows Insulation, and Global Retrofitting. These scenarios are introduced following the guidelines provided by Minergie, as previously depicted in figure 3.30. According to the table, Minergie offers five renovation systems. System 1 is recommended for buildings that have not undergone significant renovation since their initial construction or have only undergone superficial improvements. The primary objective here is to ensure proper thermal insulation. Systems 2, 3, and 4 are suitable for relatively newer or previously renovated buildings. The distinction between these systems lies in the specific combination of thermal insulation values applied to the roof and the exterior

walls. System 5 is intended for buildings that directly about neighboring structures, and where the existing facade should remain unaltered either due to design considerations or practical constraints.

As depicted in table 4.2, considering the construction period of the buildings and the extent of renovation outlined in the XML file for the buildings in Satom, Minergie systems were applied to different clusters of buildings. This was done in a manner ensuring that at least one component of the buildings experience an enhancement in their thermal insulation. Subsequently, based on the U-values presented in table 4.2 (B), adjustments were made to the XML file. For each energy-saving scenario, simulations were conducted using CitySim Pro to estimate the HD of connected buildings that had undergone renovation.

Gathering detailed heating demand information from renovated buildings made the dataset for training the ML model much bigger and more detailed. This dataset grew not just in size but also included many important details. As a result, it gave the ML model the precise information it needed to learn well and accurately predict energy usage.

**Table 4.2.** (A) The thermal conductivity of different building elements before retrofitting. (B) The thermal conductivity of different building elements after retrofitting.

| <b>(A) Before renovation</b>  |  |  |  |  |
|-------------------------------|--|--|--|--|
| <b>Period of construction</b> | Roof (Wm <sup>-2</sup> K <sup>-1</sup> ) | Wall (Wm <sup>-2</sup> K <sup>-1</sup> ) | Slab (Wm <sup>-2</sup> K <sup>-1</sup> ) | Window (Wm <sup>-2</sup> K <sup>-1</sup> ) |
| <b>Before 1919</b>            | 0.7                                      | 1.22                                     | 1.6                                      | 1.7 - 2.3                                  |
| <b>1919-1945</b>              | 0.7                                      | 1.22                                     | 1.6                                      | 1.7 - 2.3                                  |
| <b>1946-1960</b>              | 0.7                                      | 1.53                                     | 1.5                                      | 1.7 - 2.3                                  |
| <b>1961-1970</b>              | 0.65                                     | 1.02                                     | 1.3                                      | 1.7 - 2.3                                  |
| <b>1971-1980</b>              | 0.6                                      | 0.87                                     | 1.1                                      | 1.7 - 2.3                                  |
| <b>1981-1990</b>              | 0.43                                     | 0.89                                     | 0.68                                     | 1.7 - 2.3                                  |
| <b>1991-2000</b>              | 0.31                                     | 0.69                                     | 0.49                                     | 1.7 - 2.3                                  |
| <b>2001-2010</b>              | 0.25                                     | 0.51                                     | 0.35                                     | 1.7 - 2.3                                  |
| <b>After 2010</b>             | 0.22                                     | 0.51                                     | 0.25                                     | 1.7 - 2.3                                  |

| <b>(B) After renovation</b>   |                            |  |  |  |  |
|-------------------------------|----------------------------|--|--|--|--|
| <b>Period of construction</b> | <b>Retrofitting system</b> | Roof (Wm <sup>-2</sup> K <sup>-1</sup> ) | Wall (Wm <sup>-2</sup> K <sup>-1</sup> ) | Slab (Wm <sup>-2</sup> K <sup>-1</sup> ) | Window (Wm <sup>-2</sup> K <sup>-1</sup> ) |
| <b>Before 1919</b>            | System 1                   | 0.17                                     | 0.25                                     | 0.25                                     | 1  |
| <b>1919-1945</b>              | System 1                   | 0.17                                     | 0.25                                     | 0.25                                     | 1  |
| <b>1946-1960</b>              | System 1                   | 0.17                                     | 0.26                                     | 0.25                                     | 1  |
| <b>1961-1970</b>              | System 4                   | 0.17                                     | 0.72                                     | 0.25                                     | 1  |
| <b>1971-1980</b>              | System 4                   | 0.17                                     | 0.72                                     | 0.25                                     | 1  |
| <b>1981-1990</b>              | System 4                   | 0.17                                     | 0.72                                     | 0.25                                     | 1  |
| <b>1991-2000</b>              | System 3                   | 0.25                                     | 0.51                                     | 0.25                                     | 1  |
| <b>2001-2010</b>              | System 2                   | 0.25                                     | 0.4                                      | 0.25                                     | 1  |
| <b>After 2010</b>             | System 2                   | 0.22                                     | 0.4                                      | 0.25                                     | 1  |

#### 4.4.2. Tuning hyperparameters

To determine the annual HD of individual buildings and the combined energy needs at the urban scale, two machine learning algorithms were selected for implementation. The chosen algorithms for this study are LightGBM and RF regression. The reason for employing both of these ML algorithms is to evaluate their performance and select the most effective energy use model for accurate HD predictions. However, as explained in the methodology section, there are critical preliminary steps when using ML algorithms that must be taken into account to ensure optimal performance while preventing overfitting.

The crucial preliminary step in preparing ML algorithms for training is tuning of their hyperparameters. In this study, the Optuna optimizer for tuning the hyperparameters of LightGBM and the GA optimization algorithm for RF regression is employed. Since hyperparameters govern the architecture of ML models and different subsets of buildings exhibit distinct energy behaviors along with specific characteristics introduced through dependent variables, it is essential to fine-tune the hyperparameters of each algorithm separately for each subset of buildings. Tables 4.3 and 4.4 provide detailed information on the tuned hyperparameter sets for LightGBM and RF regression, the range of values explored for each hyperparameter, and the optimized values for each.

It's important to highlight that in the process of tuning the hyperparameters of the ML algorithms, 80% of the datasets were allocated for training purposes, while the remaining was set aside for testing. Additionally, the random state for both training and testing data was specifically set to 42. The choice of random state 42 ensures that the data split into training and testing sets remains consistent each time the process is repeated. This helps maintain the reproducibility of the results and allows for a fair comparison of different hyperparameter configurations.

Additionally, to fine-tune the hyperparameters effectively, a 5-fold cross-validation technique is implemented. This involves partitioning the dataset into separate 'folds,' training the model on various fold combinations, and then evaluating its performance, as explained in the methodology chapter. This method brings several advantages, including the elimination of the requirement for a dedicated validation set, the ability to assess the model's generalization across different data partitions,

a reduction in the risk of overfitting, and the provision of a more dependable estimate of model performance, all of which play a crucial role in hyperparameter tuning. Furthermore, the effectiveness of hyperparameter tuning were assessed by evaluating the Root Mean Square Error (RMSE), as indicated in tables 4.3 and 4.4. This evaluation helped to measure the accuracy of the model's predictions.

*Table 4.3. Hyperparameters of LightGBM model*

|   |                    |              |                      |                |              |                 |
|---|--------------------|--------------|----------------------|----------------|--------------|-----------------|
| Residential buildings<br>with<br>occupancy profile        | Dataset size: 2034 |              | Tuning time: 123.36s |                | RMSE: 1.11   |                 |
|   | Hyperparameter     | Tested range | Optimized Value      | Hyperparameter | Tested range | Optimized value |
|   | num_leaves         | 50 - 100     | <b>66</b>            | bagging_freq   | 1 - 10       | <b>1</b>        |
|   | max_depth          | 1 - 5        | <b>5</b>             | lambda_l1      | 0 - 0.6      | <b>0.1</b>      |
|   | min_data_in_leaf   | 1 - 10       | <b>1</b>             | lambda_l2      | 0 - 0.6      | <b>0.04</b>     |
|   | feature_fraction   | 0.4 - 1      | <b>0.99</b>          | max_bin        | 500 - 1000   | <b>821</b>      |
| Residential buildings<br>without<br>occupancy profile     | Dataset size: 2034 |              | Tuning time: 127.48s |                | RMSE: 1.5    |                 |
|   | Hyperparameter     | Tested range | Optimized Value      | Hyperparameter | Tested range | Optimized value |
|   | num_leaves         | 50 - 100     | <b>87</b>            | bagging_freq   | 1 - 10       | <b>4</b>        |
|   | max_depth          | 1 - 5        | <b>5</b>             | lambda_l1      | 0 - 0.6      | <b>0.21</b>     |
|   | min_data_in_leaf   | 1 - 10       | <b>1</b>             | lambda_l2      | 0 - 0.6      | <b>0.54</b>     |
|   | feature_fraction   | 0.4 - 1      | <b>0.42</b>          | max_bin        | 500 - 1000   | <b>798</b>      |
| Multifunction<br>buildings                                | Dataset size: 186  |              | Tuning time: 118.63s |                | RMSE: 1.58   |                 |
|   | Hyperparameter     | Tested range | Optimized Value      | Hyperparameter | Tested range | Optimized value |
|   | num_leaves         | 50 - 100     | <b>82</b>            | bagging_freq   | 1 - 10       | <b>3</b>        |
|   | max_depth          | 1 - 5        | <b>4</b>             | lambda_l1      | 0 - 0.6      | <b>0.04</b>     |
|   | min_data_in_leaf   | 1 - 10       | <b>2</b>             | lambda_l2      | 0 - 0.6      | <b>0.41</b>     |
|   | feature_fraction   | 0.4 - 1      | <b>0.8</b>           | max_bin        | 500 - 1000   | <b>501</b>      |
| Non-residential<br>buildings with<br>occupancy profile    | Dataset size: 366  |              | Tuning time: 129.48s |                | RMSE: 2.99   |                 |
|   | Hyperparameter     | Tested range | Optimized Value      | Hyperparameter | Tested range | Optimized value |
|   | num_leaves         | 50 - 100     | <b>85</b>            | bagging_freq   | 1 - 10       | <b>4</b>        |
|   | max_depth          | 1 - 5        | <b>5</b>             | lambda_l1      | 0 - 0.6      | <b>0.39</b>     |
|   | min_data_in_leaf   | 1 - 10       | <b>2</b>             | lambda_l2      | 0 - 0.6      | <b>0.26</b>     |
|   | feature_fraction   | 0.4 - 1      | <b>0.57</b>          | max_bin        | 500 - 1000   | <b>631</b>      |
| Non-residential<br>buildings without<br>occupancy profile | Dataset size: 366  |              | Tuning time: 120.17s |                | RMSE: 2.75   |                 |
|   | Hyperparameter     | Tested range | Optimized Value      | Hyperparameter | Tested range | Optimized value |
|   | num_leaves         | 50 - 100     | <b>58</b>            | bagging_freq   | 1 - 10       | <b>8</b>        |
|   | max_depth          | 1 - 5        | <b>5</b>             | lambda_l1      | 0 - 0.6      | <b>0.15</b>     |
|   | min_data_in_leaf   | 1 - 10       | <b>1</b>             | lambda_l2      | 0 - 0.6      | <b>0.51</b>     |
|   | feature_fraction   | 0.4 - 1      | <b>0.49</b>          | max_bin        | 500 - 1000   | <b>767</b>      |
| Non-residential<br>buildings with<br>occupancy profile    | Dataset size: 366  |              | Tuning time: 120.17s |                | RMSE: 2.75   |                 |
|   | Hyperparameter     | Tested range | Optimized Value      | Hyperparameter | Tested range | Optimized value |
|   | num_leaves         | 50 - 100     | <b>58</b>            | bagging_freq   | 1 - 10       | <b>8</b>        |
|   | max_depth          | 1 - 5        | <b>5</b>             | lambda_l1      | 0 - 0.6      | <b>0.15</b>     |
|   | min_data_in_leaf   | 1 - 10       | <b>1</b>             | lambda_l2      | 0 - 0.6      | <b>0.51</b>     |
|   | feature_fraction   | 0.4 - 1      | <b>0.49</b>          | max_bin        | 500 - 1000   | <b>767</b>      |

Based on the findings presented in tables 4.3 and 4.4, it's evident that the Optuna optimizer considerably outperformed the Genetic Algorithm in terms of hyperparameter tuning speed. This outcome is consistent with the expected performance difference, as Genetic Algorithms are typically known for their computational intensity and prolonged execution times. The speed advantage of Optuna in this context is particularly beneficial, as it streamlines the hyperparameter tuning process, reducing computational resources and time requirements.

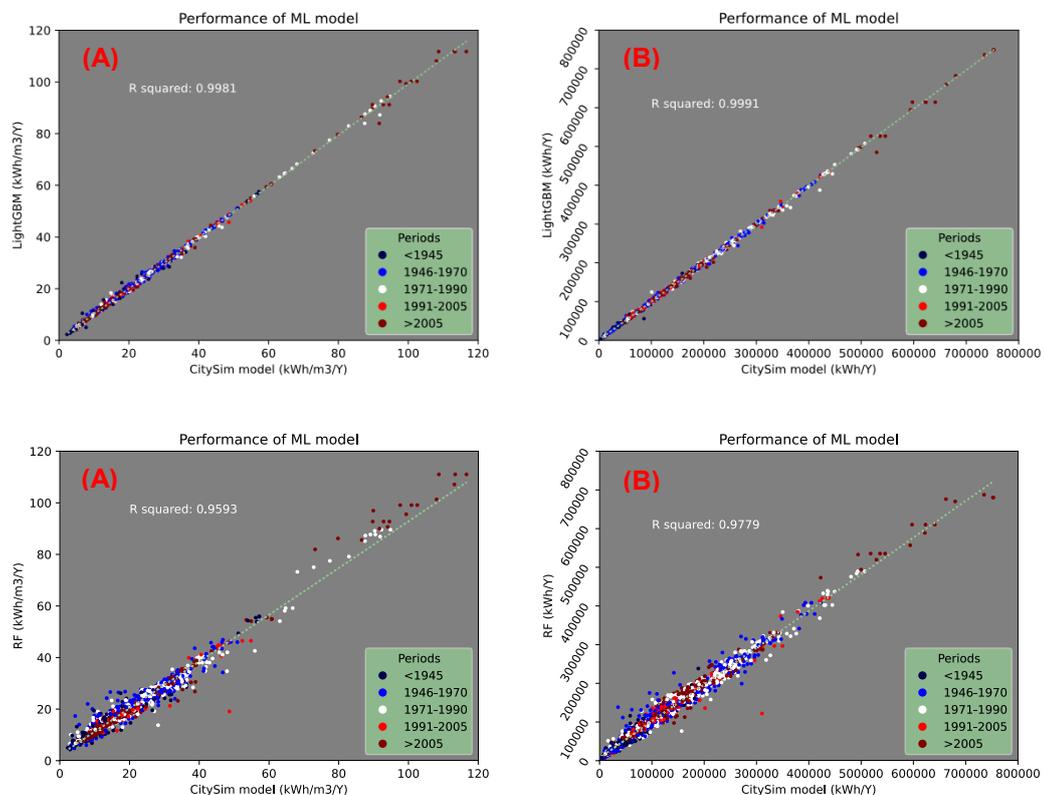
*Table 4.4. Hyperparameters of RF regression model*

|   | Dataset size: 2034   |              | Tuning time: 728.68s |                   | RMSE: 3.55   |                 |
|---|--|--------------|----------------------|-------------------|--------------|-----------------|
|   | Hyperparameter   | Tested range | Optimized Value      | Hyperparameter    | Tested range | Optimized value |
| Residential buildings with occupancy profile        | N_estimators   | 100 - 150    | <b>106</b>           | min_samples_split | 2 - 5        | <b>2</b>        |
|   | max_depth  | 10 - 15      | <b>15</b>            | Min_samples_leaf  | 2 - 5        | <b>2</b>        |
|   | Dataset size: 2034      Tuning time: 123.36s      RMSE: 5.78 |              |                      |                   |              |                 |
| Residential buildings without occupancy profile     | N_estimators   | 100 - 150    | <b>108</b>           | min_samples_split | 2 - 5        | <b>3</b>        |
|   | max_depth  | 10 - 15      | <b>14</b>            | Min_samples_leaf  | 2 - 5        | <b>2</b>        |
|   | Dataset size: 186      Tuning time: 218.16s      RMSE: 2.19  |              |                      |                   |              |                 |
| Multifunction buildings                             | N_estimators   | 100 - 150    | <b>137</b>           | min_samples_split | 2 - 5        | <b>2</b>        |
|   | max_depth  | 10 - 15      | <b>12</b>            | Min_samples_leaf  | 2 - 5        | <b>2</b>        |
|   | Dataset size: 366      Tuning time: 234.66s      RMSE: 4.92  |              |                      |                   |              |                 |
| Non-residential buildings with occupancy profile    | N_estimators   | 100 - 150    | <b>110</b>           | min_samples_split | 2 - 5        | <b>3</b>        |
|   | max_depth  | 10 - 15      | <b>12</b>            | Min_samples_leaf  | 2 - 5        | <b>2</b>        |
|   | Dataset size: 366      Tuning time: 257.03s      RMSE: 5.58  |              |                      |                   |              |                 |
| Non-residential buildings without occupancy profile | N_estimators   | 100 - 150    | <b>146</b>           | min_samples_split | 2 - 5        | <b>3</b>        |
|   | max_depth  | 10 - 15      | <b>12</b>            | Min_samples_leaf  | 2 - 5        | <b>2</b>        |

### 4.4.3. Annual heating demand processing

In preparation for training the energy use model with ML algorithms, a sensitivity analysis was carried out using the 'feature\_importances\_' method offered by both 'LGBMRegressor' and 'RandomForestRegressor' as part of efforts to enhance the utility of ML models. This allowed for the assessment of the significance of individual input variables in the ML model training process. Through the implementation of this analysis, insignificant input variables were identified and removed in the context of training energy use models for specific building subsets. This action contributed to the improvement of model efficiency and the incorporation of the most relevant features in the model training process, ultimately resulting in more accurate and insightful outcomes.

With the optimized hyperparameters and the validation of significant variables to construct energy use models, ML models were successfully trained for each subset of buildings. As shown in Figure 4.8, both ML models exhibited exceptional performance in predicting energy intensity and heating demands at building scale for residential buildings with occupancy profile.

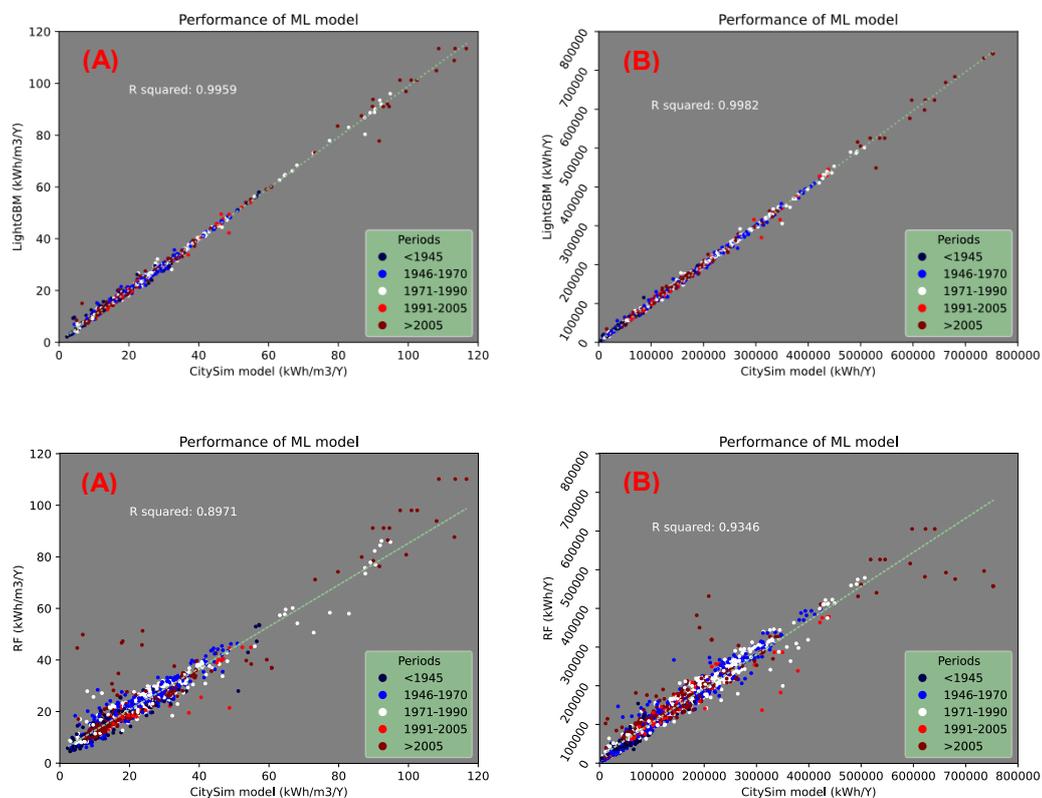


**Figure 4.8.** Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Residential Buildings with Occupancy Profile

LightGBM achieved an impressive RMSE of 1.04 for energy intensity prediction, while RF demonstrated an RMSE of 3.58. When it came to predicting heating demand, LightGBM and RF regression achieved high accuracy with R-squared values of 0.9991 and 0.9779, respectively. It's noteworthy that these outstanding performances were achieved with minimal computational time, as LightGBM required only 1.06 seconds for analysis, while RF regression took just 1.004 seconds.

The features used for training LightGBM included 'Construction period', 'S/V', 'Occupants', 'Wall thickness', 'U-Roof', 'U-Wall', 'U-Slab', 'U-Window', 'Gross heated volume', and 'Gross heated surface'. Similarly, for RF, the selected features were 'Construction period', 'S/V', 'Occupants', 'Wall thickness', 'U-Wall', 'U-Slab', 'U-Window', 'Gross heated volume', and 'Gross heated surface'.

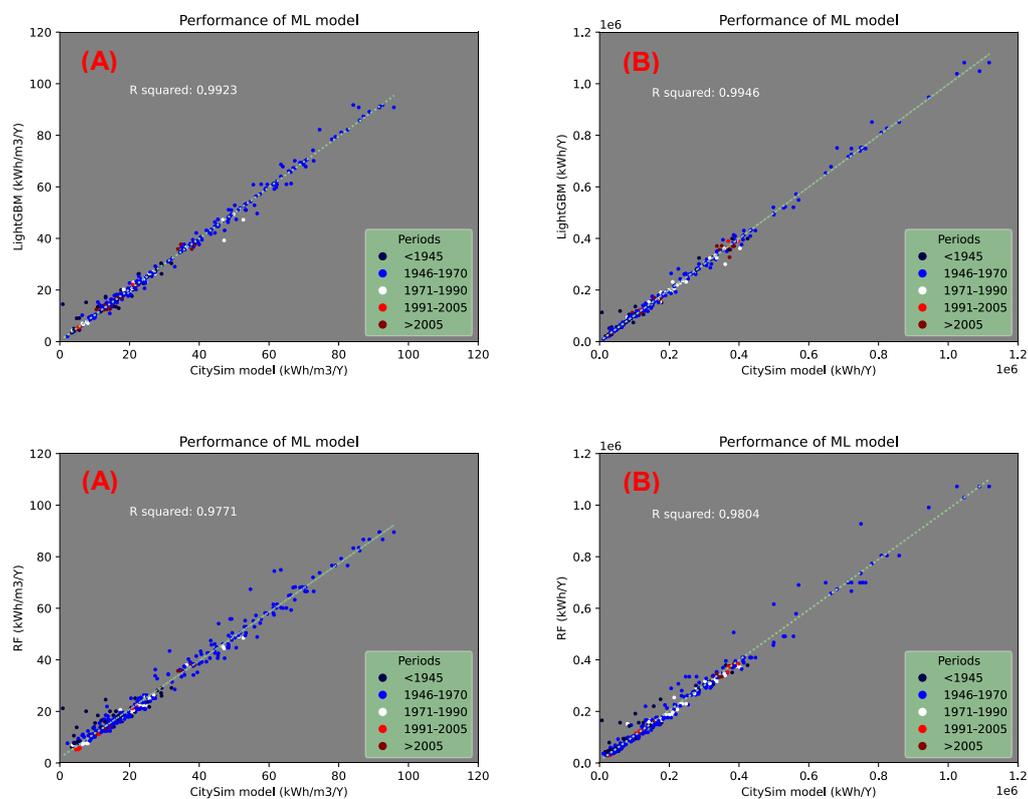
Figure 4.9, showcasing the performance of ML models in predicting energy intensity and HD for residential buildings without occupancy profiles, clearly demonstrates the substantial impact of individual variables on the performance of ML models in building energy use prediction.



**Figure 4.9.** Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Residential Buildings without Occupancy Profile

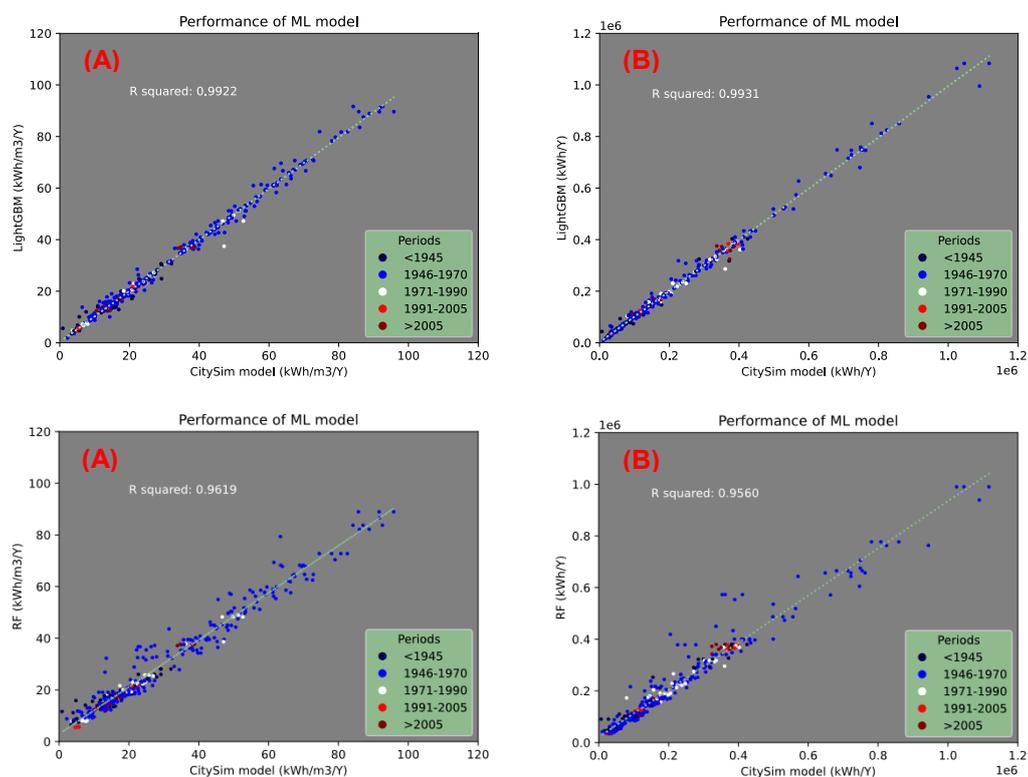
In this scenario, the input variables were consistent for both models, including 'Construction period', 'S/V', 'Wall thickness', 'U-Roof', 'U-Wall', 'U-Slab', 'U-Window', 'Gross heated volume', and 'Gross heated surface'. As shown in Figure 4.9, LightGBM maintained its impressive performance in predicting energy intensity for buildings, achieving an RMSE of 1.57, even when a significant variable was absent. However, the situation was different for RF regression. In this case, RF regression had an RMSE of 5.99 for energy intensity prediction, and there was a noticeable drop of approximately 0.5 in the R-squared precision compared to the previous model in HD prediction. Nevertheless, the results remain substantial and reliable.

Figures 4.10 and 4.11 depict the performance of LightGBM and RF regression models for nonresidential buildings, with a similar distinction as were for residential buildings, where one set of buildings lacks occupancy profile information. As anticipated, the performance was relatively better for buildings with available occupancy data.



**Figure 4.10.** Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Non-residential Buildings with Occupancy Profile

For instance, in the case of LightGBM, the RMSE for energy intensity prediction was 3.45 for buildings with occupants' information, compared to 3.42 for those without such information. For RF regression, the corresponding accuracy check was 4.87 compared to 6.09, respectively. Additionally, as evident from the figures, the accuracy in computing HD is consistently high and remarkable. For buildings with occupancy profiles, the R-squared values were 0.9946 for LightGBM and 0.9804 for RF regression. In the case of the other subset of nonresidential buildings, these values were 0.9931 and 0.9560, respectively.

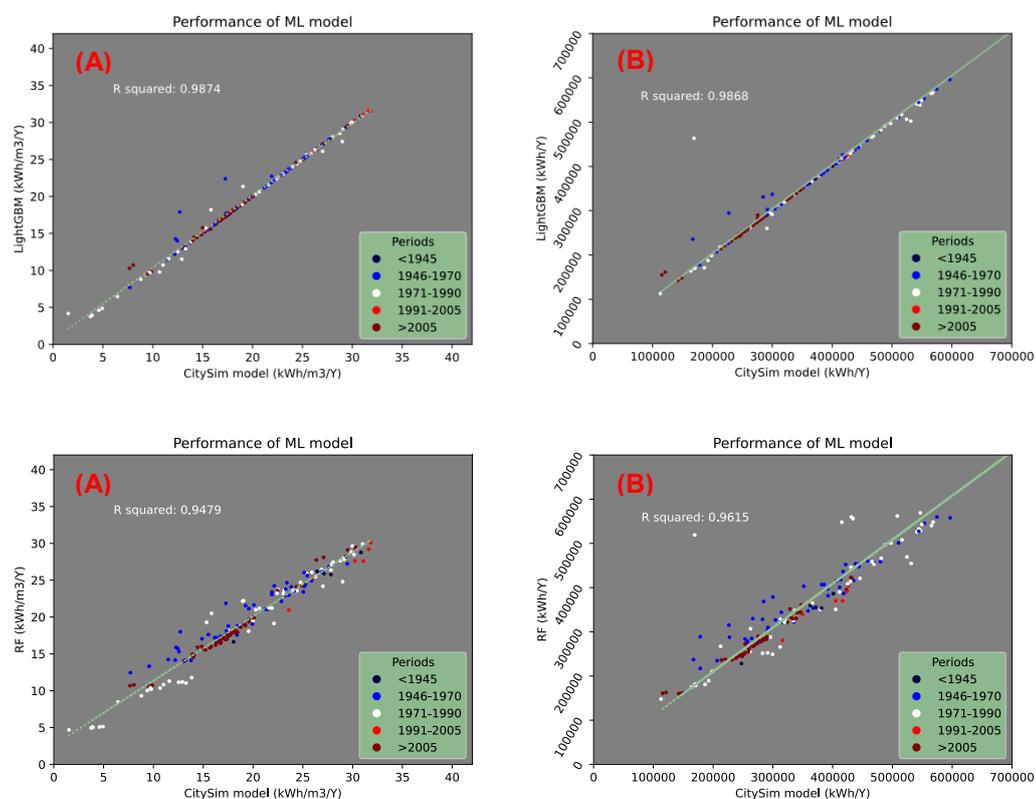


**Figure 4.11.** Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Non-residential Buildings without Occupancy Profile

For nonresidential buildings with occupancy profiles, the dependent variables for LightGBM included 'Land use', 'S/V', 'Occupants', 'Wall thickness', 'U-Roof', 'U-Wall', 'U-Slab', 'Gross heated volume', and 'Gross heated surface'. For RF regression, these variables were 'Land use', 'S/V', 'Occupants', 'Wall thickness', 'U-Slab', 'U-Window', 'Gross heated volume', and 'Gross heated surface'. In cases where nonresidential buildings lacked occupancy information, the dependent variables remained the same as the previous model, with the addition of 'Construction period' for RF regression.

It's worth noting that comparing the size of datasets used for prediction between nonresidential and residential buildings highlights the remarkable capability of ML models. The dataset size for residential buildings was 2034, whereas for nonresidential buildings, it was 366. Despite the significant gap in dataset sizes, the performance of ML models showed a relatively consistent level of accuracy.

The final group of buildings for which we trained energy use models consists of multifunctional buildings. Remarkably, despite the relatively small dataset size of just 186, as illustrated in Figure 4.12, ML algorithms once again proved highly effective in predicting their energy patterns.



**Figure 4.12.** Performance of LightGBM and RF Regression in predicting (A) Energy Intensity and (B) Heating Demands at building scale in Multifunction Buildings with Occupancy Profile

For LightGBM, the independent variables used to predict the energy use model included 'S/V,' 'Occupants,' 'Wall thickness,' 'U-Roof,' 'U-Wall,' 'U-Slab,' 'U-Window,' 'Gross heated volume,' and 'Gross heated surface.' These variables yielded an accuracy of RMSE 1.62 for energy intensity prediction and an R-squared value of 0.9869 for heating demand prediction. Meanwhile, RF regression employed 'Construction period,' 'S/V,' 'Occupants,' 'Wall thickness,' 'U-Roof,' 'U-Wall,' 'U-Slab,' 'U-Window,' 'Gross heated volume,' and 'Gross heated surface' as

independent variables, resulting in an RMSE of 2.26 for energy intensity prediction and an R-squared value of 0.9615 for heating demand prediction in multifunctional buildings.

In a comprehensive evaluation of all energy use models developed for diverse building subsets, it is clear that integrating LightGBM provided a significant advantage over RF regression. This superiority was evident both in the context of hyperparameter tuning and in the creation of energy use models. Optuna, which was used to fine-tune LightGBM's hyperparameters, demonstrated greater computational efficiency compared to the Genetic Algorithm (GA) utilized for RF regression. Consequently, LightGBM emerged as the favored algorithm at first glance.

Additionally, LightGBM consistently exhibited exceptional performance across the energy use models, consistently outperforming RF regression with higher precision. The clear superiority of LightGBM over RF regression is also apparent in table 4.5, where HD predictions are compared for different energy-saving scenarios at the urban scale. It is important to highlight that in this table, both varieties of residential and nonresidential buildings are combined. This consistent reliability and superior performance made LightGBM the unequivocal and preferred choice for subsequent analyses and model development.

**Table 4.5.** Comparing LightGBM and RF regression performance in predicting Urban-Scale heating demand with different renovation technologies

|                                   |                           | <b>Buildings not retrofitted</b> |                 |                     |                       |       |
|-----------------------------------|---------------------------|----------------------------------|-----------------|---------------------|-----------------------|-------|
| <b>Type of buildings</b>          |                           | Model                            | Whole buildings | Connected buildings | Connectable buildings |       |
| <b>Residential buildings</b>      | No.                       | --                               | 2018            | 339                 | 1679                  |       |
|                                   | Volume (Mm <sup>3</sup> ) | --                               | 4.98            | 2.02                | 2.96                  |       |
|                                   | Share of volume (%)       | --                               | 100             | 16.8                | 83.2                  |       |
|                                   | HD (GWh/Y)                | CitySim Pro                      | --              | --                  | 52                    | --    |
|                                   |                           | LightGBM                         | 138.86          | 51.86               | <b>(-0.27%)</b>       | 87    |
|                                   | RF                        | 129.97                           | 50              | <b>(-3.85%)</b>     | 79.97                 |       |
| <b>Multi-functional buildings</b> | No.                       | --                               | 39              | 31                  | 8                     |       |
|                                   | Volume (Mm <sup>3</sup> ) | ---                              | 0.89            | 0.72                | 0.17                  |       |
|                                   | Share of volume (%)       | -                                | 100             | 80.90               | 19.1                  |       |
|                                   | HD (GWh/Y)                | CitySim Pro                      | --              | --                  | 13.98                 | --    |
|                                   |                           | LightGBM                         | 17.6            | 13.96               | <b>(-0.14%)</b>       | 3.64  |
| RF                                |                           | 17.08                            | 13.57           | <b>(-2.93%)</b>     | 3.51                  |       |
| <b>Non-residential buildings</b>  | No.                       | --                               | 226             | 61                  | 165                   |       |
|                                   | Volume (Mm <sup>3</sup> ) | --                               | 1.49            | 0.54                | 0.95                  |       |
|                                   | Share of volume (%)       | --                               | 100             | 27                  | 73                    |       |
|                                   | HD (GWh/Y)                | CitySim Pro                      | --              | --                  | 14.6                  | --    |
|                                   |                           | LightGBM                         | 37.12           | 13.89               | <b>(-4.86%)</b>       | 23.23 |
| RF                                |                           | 36.9                             | 13.8            | <b>(-5.48%)</b>     | 23.1                  |       |

| <b>Buildings underwent roof insulation</b> |                           |             |                 |                     |                       |       |
|--|---------------------------|-------------|-----------------|---------------------|-----------------------|-------|
| Type of buildings                          |                           | Model       | Whole buildings | Connected buildings | Connectable buildings |       |
| <b>Residential buildings</b>               | No.                       | --          | 2018            | 339                 | 1679                  |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 4.98            | 2.02                | 2.96                  |       |
|  | Share of volume (%)       | --          | 100             | 16.8                | 83.2                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 50.15                 | --    |
|  |                           | LightGBM    |                 | 134.09              | 50.12 (-0.05%)        | 83.97 |
| RF   |                           |             | 130             | 50.05 (-0.2%)       | 79.95                 |       |
| <b>Multi-functional buildings</b>          | No.                       | --          | 39              | 31                  | 8                     |       |
|  | Volume (Mm <sup>3</sup> ) | ---         | 0.89            | 0.72                | 0.17                  |       |
|  | Share of volume (%)       | -           | 100             | 80.90               | 19.1                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 13.32                 | --    |
|  |                           | LightGBM    |                 | 16.69               | 13.31 (-0.08%)        | 3.38  |
| RF   |                           |             | 16.92           | 13.42 (+0.75%)      | 3.5                   |       |
| <b>Non-residential buildings</b>           | No.                       | --          | 226             | 61                  | 165                   |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 1.49            | 0.54                | 0.95                  |       |
|  | Share of volume (%)       | --          | 100             | 27                  | 73                    |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 13.5                  | --    |
|  |                           | LightGBM    |                 | 36.24               | 13.52 (+0.15%)        | 22.72 |
| RF   |                           |             | 36.9            | 13.8 (+2.22%)       | 23.1                  |       |
| <b>Buildings underwent slab insulation</b> |                           |             |                 |                     |                       |       |
| Type of buildings                          |                           | Model       | Whole buildings | Connected buildings | Connectable buildings |       |
| <b>Residential buildings</b>               | No.                       | --          | 2018            | 339                 | 1679                  |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 4.98            | 2.02                | 2.96                  |       |
|  | Share of volume (%)       | --          | 100             | 16.8                | 83.2                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 48.58                 | --    |
|  |                           | LightGBM    |                 | 129.36              | 48.55 (-0.06%)        | 80.81 |
| RF   |                           |             | 126.71          | 48.93 (+0.72%)      | 77.78                 |       |
| <b>Multi-functional buildings</b>          | No.                       | --          | 39              | 31                  | 8                     |       |
|  | Volume (Mm <sup>3</sup> ) | ---         | 0.89            | 0.72                | 0.17                  |       |
|  | Share of volume (%)       | -           | 100             | 80.90               | 19.1                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 12.92                 | --    |
|  |                           | LightGBM    |                 | 16.48               | 13.03 (+0.85%)        | 3.45  |
| RF   |                           |             | 16.53           | 13.09 (+1.32%)      | 3.44                  |       |
| <b>Non-residential buildings</b>           | No.                       | --          | 226             | 61                  | 165                   |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 1.49            | 0.54                | 0.95                  |       |
|  | Share of volume (%)       | --          | 100             | 27                  | 73                    |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 12.59                 | --    |
|  |                           | LightGBM    |                 | 34.11               | 12.55 (-0.32%)        | 21.56 |
| RF   |                           |             | 33.9            | 12.65 (+0.48%)      | 21.25                 |       |
| <b>Buildings underwent wall insulation</b> |                           |             |                 |                     |                       |       |
| Type of buildings                          |                           | Model       | Whole buildings | Connected buildings | Connectable buildings |       |
| <b>Residential buildings</b>               | No.                       | --          | 2018            | 339                 | 1679                  |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 4.98            | 2.02                | 2.96                  |       |
|  | Share of volume (%)       | --          | 100             | 16.8                | 83.2                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 48.61                 | --    |
|  |                           | LightGBM    |                 | 128.05              | 48.57 (-0.08%)        | 79.48 |
| RF   |                           |             | 125.66          | 49.02 (+0.84%)      | 76.64                 |       |
| <b>Multi-functional buildings</b>          | No.                       | --          | 39              | 31                  | 8                     |       |
|  | Volume (Mm <sup>3</sup> ) | ---         | 0.89            | 0.72                | 0.17                  |       |
|  | Share of volume (%)       | -           | 100             | 80.90               | 19.1                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 13.33                 | --    |
|  |                           | LightGBM    |                 | 16.71               | 13.33 (0%)            | 3.38  |
| RF   |                           |             | 16.79           | 13.33 (0%)          | 3.46                  |       |
| <b>Non-residential buildings</b>           | No.                       | --          | 226             | 61                  | 165                   |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 1.49            | 0.54                | 0.95                  |       |
|  | Share of volume (%)       | --          | 100             | 27                  | 73                    |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 13.79                 | --    |
|  |                           | LightGBM    |                 | 36.42               | 13.7 (-0.65%)         | 22.72 |

|  |                           | RF          | 36.45           | 13.77 (-0.15%)      | 22.68                 |       |
|--|---------------------------|-------------|-----------------|---------------------|-----------------------|-------|
| <b>Buildings underwent window insulation</b> |                           |             |                 |                     |                       |       |
| Type of buildings                            |                           | Model       | Whole buildings | Connected buildings | Connectable buildings |       |
| <b>Residential buildings</b>                 | No.                       | --          | 2018            | 339                 | 1679                  |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 4.98            | 2.02                | 2.96                  |       |
|  | Share of volume (%)       | --          | 100             | 16.8                | 83.2                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 42.99                 | --    |
|  |                           | LightGBM    |                 | 114.08              | 42.95 (-0.09%)        | 71.13 |
| RF   |                           |             | 117.15          | 43.29 (+0.7%)       | 73.86                 |       |
| <b>Multi-functional buildings</b>            | No.                       | --          | 39              | 31                  | 8                     |       |
|  | Volume (Mm <sup>3</sup> ) | ---         | 0.89            | 0.72                | 0.17                  |       |
|  | Share of volume (%)       | -           | 100             | 80.90               | 19.1                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 12.1                  | --    |
|  |                           | LightGBM    |                 | 15.48               | 12.29 (+1.57%)        | 3.19  |
| RF   |                           |             | 16.01           | 12.65 (+4.55%)      | 3.36                  |       |
| <b>Non-residential buildings</b>             | No.                       | --          | 226             | 61                  | 165                   |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 1.49            | 0.54                | 0.95                  |       |
|  | Share of volume (%)       | --          | 100             | 27                  | 73                    |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 13.01                 | --    |
|  |                           | LightGBM    |                 | 37.12               | 13.89 (+6.76%)        | 23.23 |
| RF   |                           |             | 35.01           | 12.98 (-0.23%)      | 22.03                 |       |
| <b>Buildings globally retrofitted</b>        |                           |             |                 |                     |                       |       |
| Type of buildings                            |                           | Model       | Whole buildings | Connected buildings | Connectable buildings |       |
| <b>Residential buildings</b>                 | No.                       | --          | 2018            | 339                 | 1679                  |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 4.98            | 2.02                | 2.96                  |       |
|  | Share of volume (%)       | --          | 100             | 16.8                | 83.2                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 34.29                 | --    |
|  |                           | LightGBM    |                 | 87.26               | 34.36 (+0.2%)         | 52.9  |
| RF   |                           |             | 87.83           | 35.65 (+3.97%)      | 52.18                 |       |
| <b>Multi-functional buildings</b>            | No.                       | --          | 39              | 31                  | 8                     |       |
|  | Volume (Mm <sup>3</sup> ) | ---         | 0.89            | 0.72                | 0.17                  |       |
|  | Share of volume (%)       | -           | 100             | 80.90               | 19.1                  |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 9.78                  | --    |
|  |                           | LightGBM    |                 | 12.99               | 10.36 (+5.93%)        | 2.63  |
| RF   |                           |             | 14.02           | 11.05 (+12.99%)     | 2.97                  |       |
| <b>Non-residential buildings</b>             | No.                       | --          | 226             | 61                  | 165                   |       |
|  | Volume (Mm <sup>3</sup> ) | --          | 1.49            | 0.54                | 0.95                  |       |
|  | Share of volume (%)       | --          | 100             | 27                  | 73                    |       |
|  | HD (GWh/Y)                | CitySim Pro | --              | --                  | 9.02                  | --    |
|  |                           | LightGBM    |                 | 25.47               | 9.24 (+2.44%)         | 16.23 |
| RF   |                           |             | 29.35           | 10.86 (+20.4%)      | 18.49                 |       |

## 4.5. DHN's optimal expansion

In the present study, an approach has been employed to establish an optimal energy-saving plan and the ideal expansion of the District Heating Network. This approach entails gaining a thorough understanding of the context and aims to chart a roadmap for maximizing network expansion and connecting new buildings over the next 30 years. In this endeavor, a wide range of factors, including the socio-economic status of the context, technical aspects of the operating District Heating Network, and the physical conditions of individual buildings, have been considered.

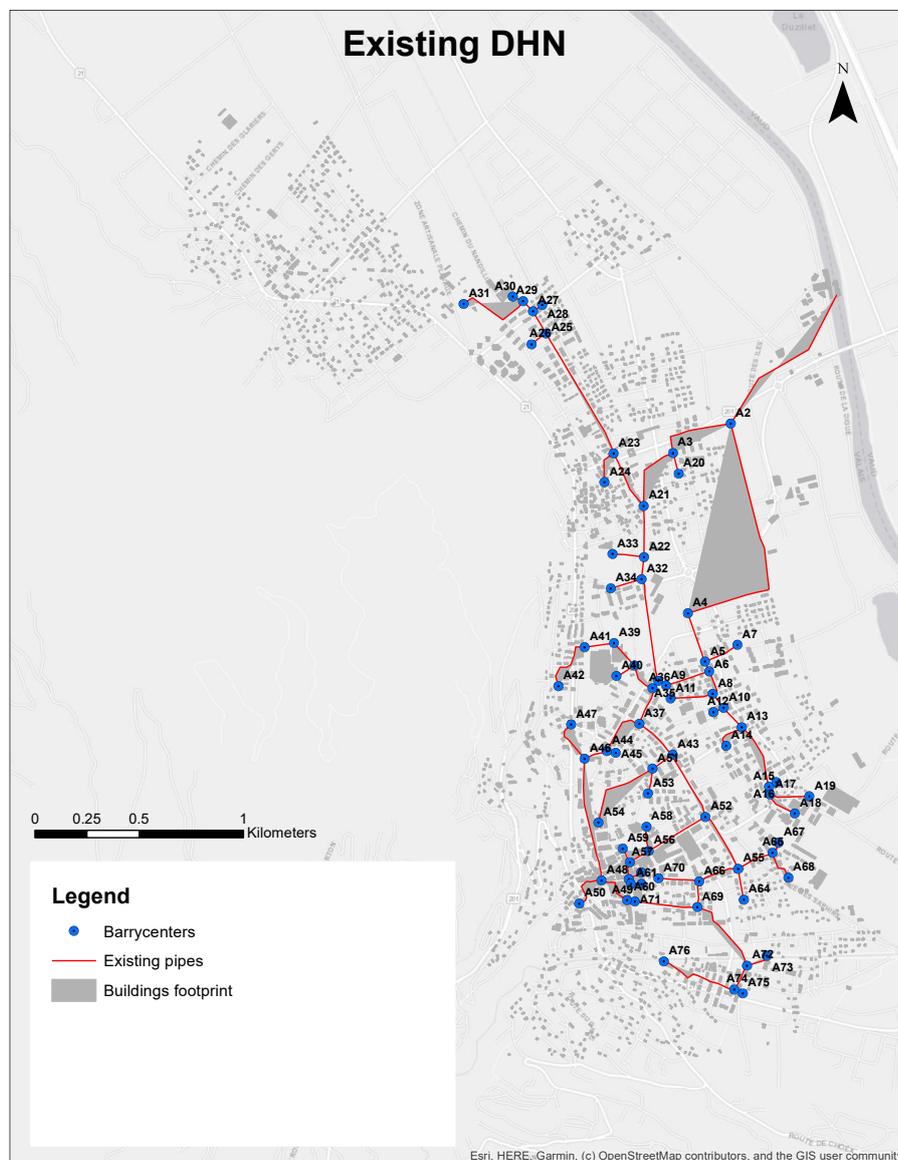
The aim is to establish a framework that can facilitate the development of an annual action plan for guiding optimal expansion. It's worth noting that still there is room for this framework to be further enhanced through the integration of more complex, physics-based network analysis.

Integral to advancing the analysis of energy conservation was the computation of peak heating demand for individual buildings within the case study. This peak demand corresponds to the highest HD during the coldest hour of the year, specifically in the targeted year of 2021. To achieve this, as previously explained, the XML file for buildings in Satom was fine-tuned based on measured data to obtain hourly HD data for buildings. As a result, through simulating the HD of connected buildings in CitySim Pro, it was established that the aggregated peak demand for buildings in Satom in 2021 occurred on January 11th at 5 a.m. Subsequently, the HD of buildings during this specific hour was utilized as a reference point for predicting the peak demand of connectable buildings. This prediction task was exclusively carried out using the LightGBM algorithm, which exhibited superior predictive performance compared to RF regression in the current research.

When using LightGBM to forecast peak HD for buildings, it involved applying all the earlier introduced preparatory measures and training procedures, such as hyperparameter tuning, sensitivity analysis for the selection of statistically significant independent variables, and the training of the energy use model. Once again, these methods were systematically implemented on different building subsets, each characterized by specific attributes that influenced the architecture of the ML model for predicting the energy use model.

The core concept behind the energy-saving scenario was to assess the maximum capacity of the Combined Heat and Power (CHP) generator and supply pipes to meet the heating demands of buildings. The central heating station of Satom SA features two 45 MW trash ovens, 5 MW of heat recovery from fumes, and dual 12.5 MW fuel-based systems for peak demands, totaling 95 MW in normal operation, with an additional 25 MW as backup capacity. Furthermore, as depicted in figure 4.13, a simplified model of the existing District Heating Network was constructed in ArcGIS to determine the maximum pipe capacity for heat transfer.

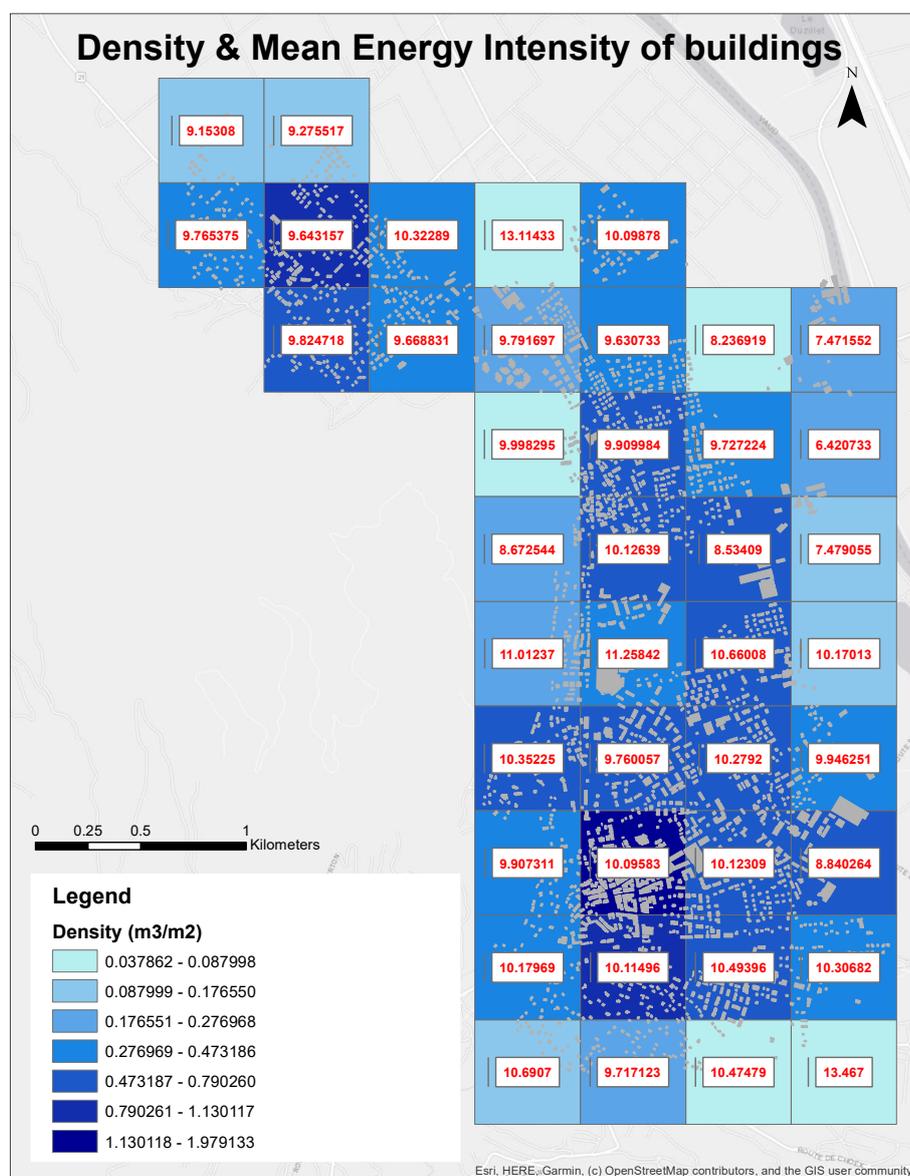
In this simplified model, the main supply pipes were retained, while secondary branches were removed. Pipes with similar diameters were grouped together to eliminate divergence points. After building this physical DHN model, the maximum load on each branch was assigned based on buildings already connected, considering all potential retrofitting scenarios. The cumulative load from outer branches towards the heating station provided the existing maximum load on the network. By subtracting this load from the maximum capacity of each branch according to its diameter, the remaining capacity of each branch to accommodate and serve additional buildings was calculated. Detailed information about the maximum load and remaining capacity of each branch can be found in Appendix E.



**Figure 4.13.** Simplified model of existing District Heating Network (DHN)

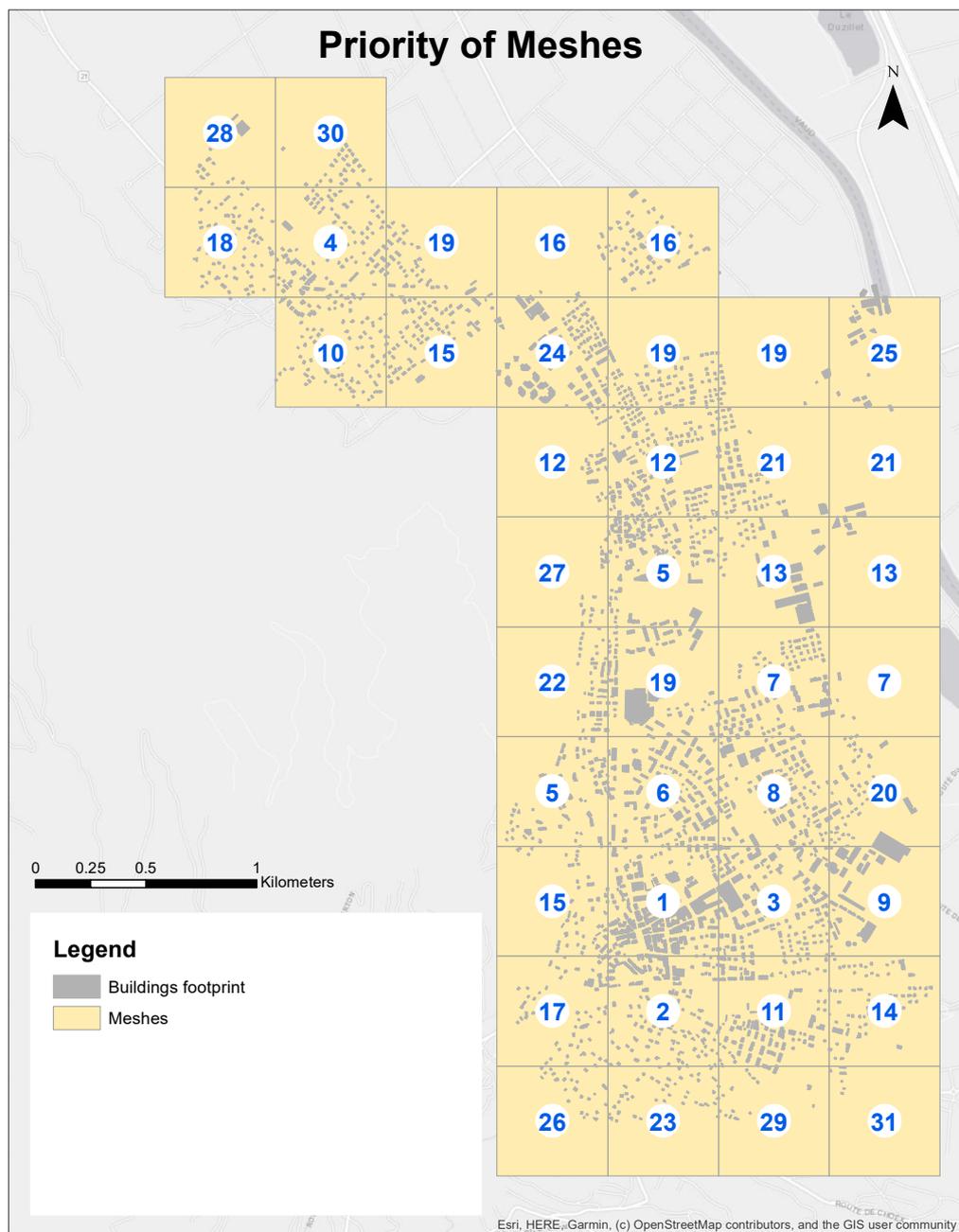
Utilizing the remaining capacity of the District Heating Network for accommodating additional buildings, a comprehensive methodology was employed to strategically select buildings for connection until the capacity was reached and could no longer support further connections. This methodology began by identifying neighborhoods that were most suitable for connection, thus allowing for the expansion of the DHN into those areas.

To achieve this, the study area was divided into a grid of meshes, each measuring 500m by 500m in dimension, as depicted in figure 4.14. Subsequently, the average energy intensity and building density within each mesh were calculated and assigned to the respective mesh.



**Figure 4.14.** Building density and energy intensity of buildings within each mesh

In the prioritization of meshes, factors such as building density and energy intensity were considered. However, building density held a greater weight in the prioritization process compared to energy intensity. This emphasis on building density stems from the general feasibility of connecting compact, high-rise buildings to the network, even though they may exhibit similar energy intensity as two-family buildings. The prioritization of neighborhoods is visually depicted in figure 4.15, and based on this prioritization, buildings within each mesh were examined sequentially.



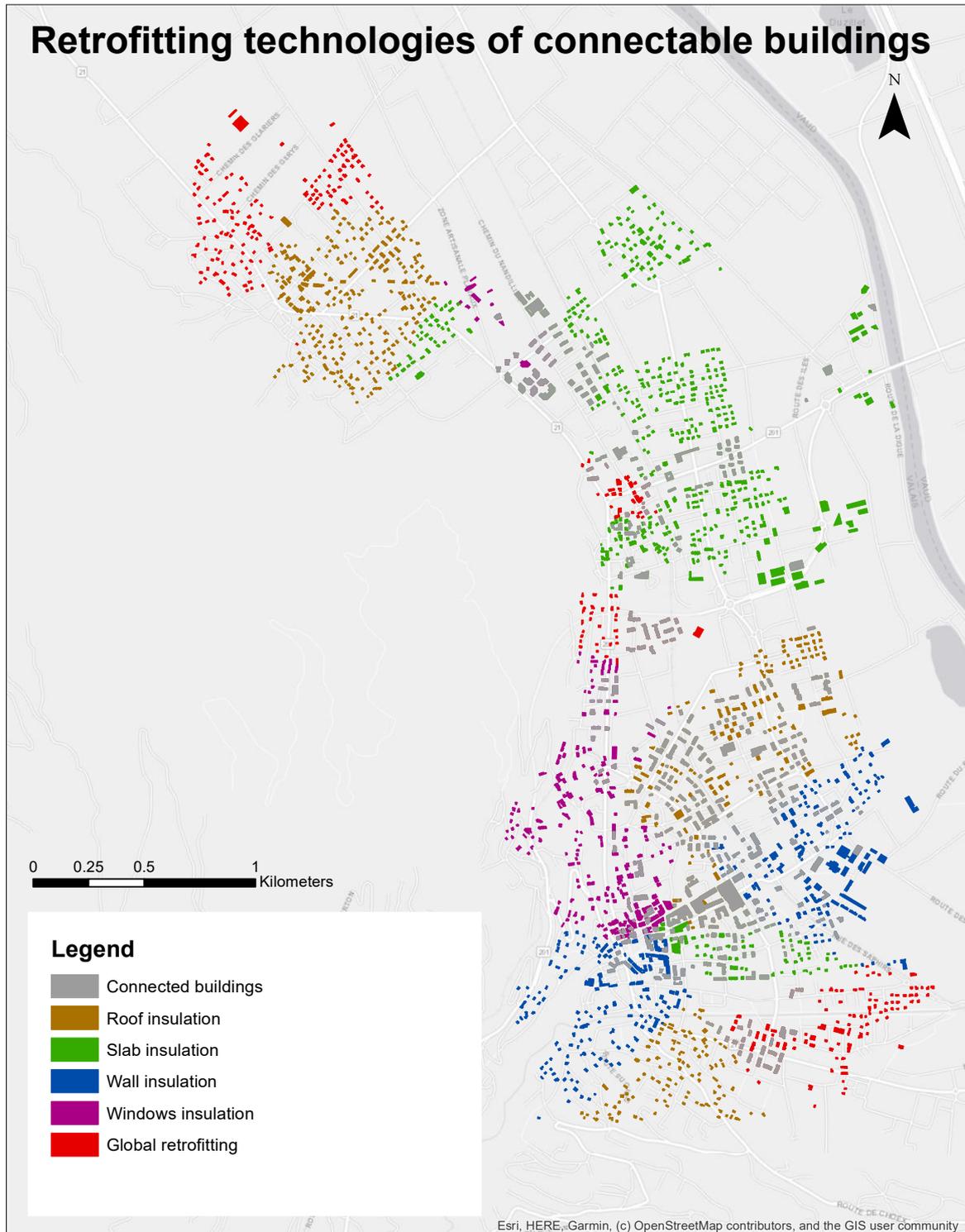
**Figure 4.15.** Mesh prioritization for optimal DHN expansion

Five scenarios were considered to explore the potential of the District Heating Network with different combinations of retrofitting technologies. These scenarios are outlined below:

- The first scenario included connected buildings without retrofitting, paired with connectable buildings that were also not retrofitted.
- In the second scenario, connected buildings without retrofitting were combined with connectable buildings that had undergone retrofitting.
- The third scenario involved retrofitting for connected buildings while the connectable buildings remained non-retrofitted.
- The fourth scenario encompassed connected buildings that were retrofitted, alongside connectable buildings that had also undergone retrofitting.
- Finally, the fifth scenario explored an extreme condition where both connected buildings and potential connectable buildings had undergone global retrofitting.

With the exception of the fifth scenario, which explores an extreme condition where both connected and connectable buildings are assumed to be globally renovated, the other scenarios consider windows substitution as the renovation technology for connected buildings. This choice is based on the understanding that connected buildings, already linked to a renewable energy resource, may have a reduced incentive for extensive renovations.

In contrast, the selection of retrofitting technology for connectable buildings is determined by their socio-economic status. To achieve this, information regarding the value of one square meter of land is collected and spatially allocated to the buildings. The range from maximum to minimum values is divided into categories corresponding to the five available renovation technologies. Consequently, each category is associated with a specific renovation technology, serving as a compelling incentive for building owners to connect their properties to the District Heating Network. Figure 4.16 provides detailed information regarding the renovation technologies assigned to each building based on this socio-economic status.



**Figure 4.16.** Distribution of retrofitting technologies for connectable buildings considering socio-economic status of context

Continuing with the prioritization of meshes for DHN expansion, connectable buildings within these meshes were randomly prioritized. Subsequently, buildings were selected for connection on an annual basis, with an annual renovation rate of 1%, based on their priorities. This iterative process was repeated for each scenario

until the network reached its maximum capacity. Table 4.6 presents the maximum number of buildings that can be connected to the DHN for each of the energy-saving scenarios. It also indicates the projected timeframe for connecting the selected buildings to the network, taking into account each of the scenarios. Figure 4.17 provides a visual representation of the spatial distribution of feasible buildings suitable for network expansion in each scenario. It's worth noting that in this figure, the buildings added in each scenario should be combined with those from previous scenarios.

**Table 4.6.** Maximum real connectable buildings considering various energy-saving scenarios

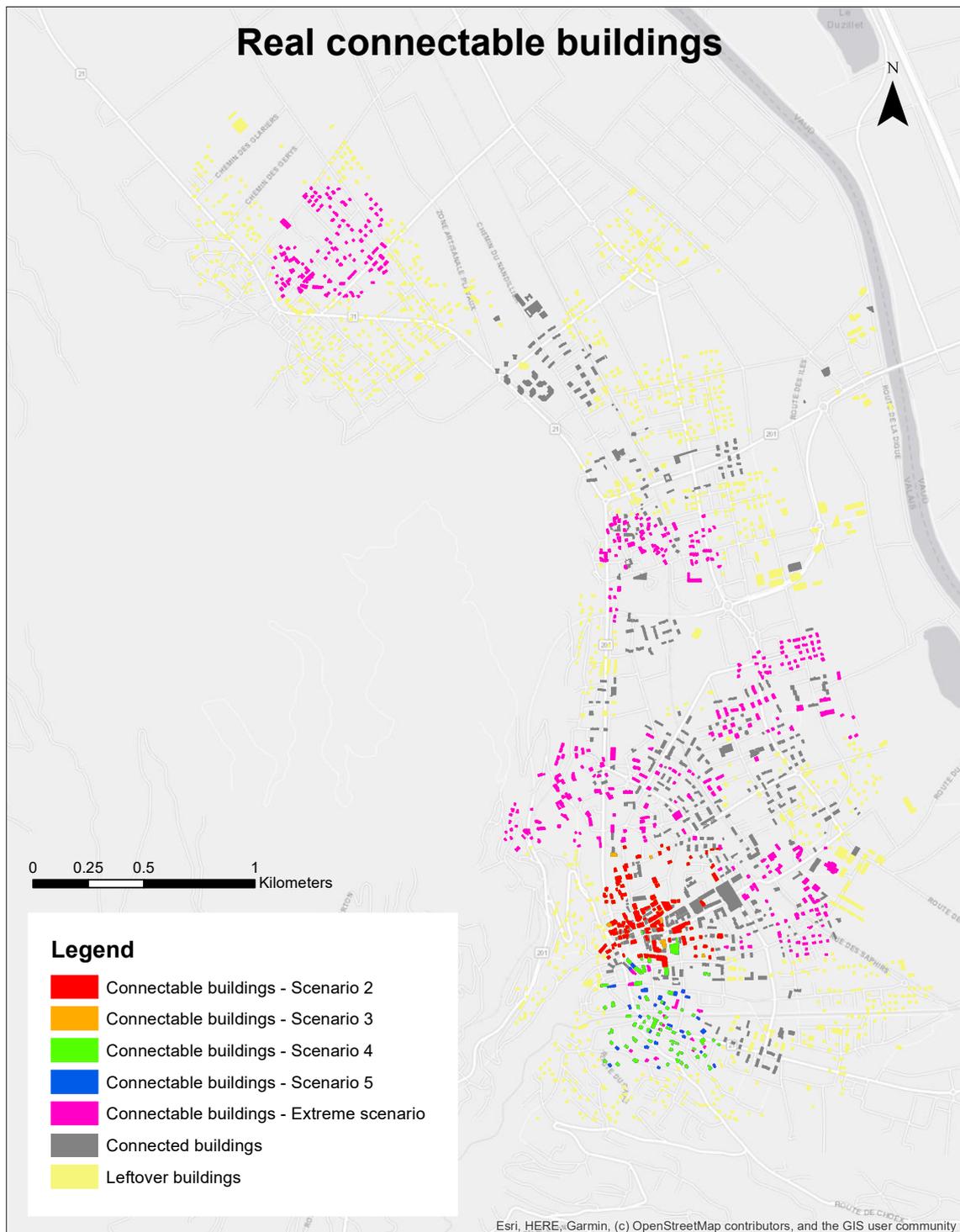
| Scenario                | Connected            | Connectable          | Real connectable buildings | Total buildings connected | Expansion timeframe (Year) |
|-------------------------|----------------------|----------------------|----------------------------|---------------------------|----------------------------|
| <b>Scenario 1</b>       | Not-retrofitted      | -                    | -                          | 431                       | -                          |
| <b>Scenario 2</b>       | Not-retrofitted      | Not-retrofitted      | 139 (7.5%)                 | 570                       | -                          |
| <b>Scenario 3</b>       | Not-retrofitted      | Retrofitted          | 146 (7.9%)                 | 577                       | 7                          |
| <b>Scenario 4</b>       | Retrofitted          | Not-retrofitted      | 209 (11.3%)                | 640                       | 20                         |
| <b>Scenario 5</b>       | Retrofitted          | Retrofitted          | 236 (12.7%)                | 667                       | 31                         |
| <b>Extreme Scenario</b> | Globally retrofitted | Globally retrofitted | 705 (38.1%)                | 1136                      | 52                         |

To assess the compatibility of the HD of connected buildings with the capacity of the DHN, the “Network simplex” algorithm was employed. The simplex algorithm is an iterative mathematical optimization technique used to solve linear programming problems. In this context, it was applied to ensure that the HD of connected buildings did not exceed the capacity of the DHN.

In pursuit of this objective, a tree graph representing the DHN was constructed using the “Networkx algorithm”. The graph consisted of edges representing pipes with their respective heat transfer capacities and nodes representing buildings with their heating demands. To establish connections between buildings and the network, a near table was generated using ArcGIS to calculate the distances of buildings from the network's barycenter's. Subsequently, individual buildings' heating demands were allocated to the nearest barycenter.

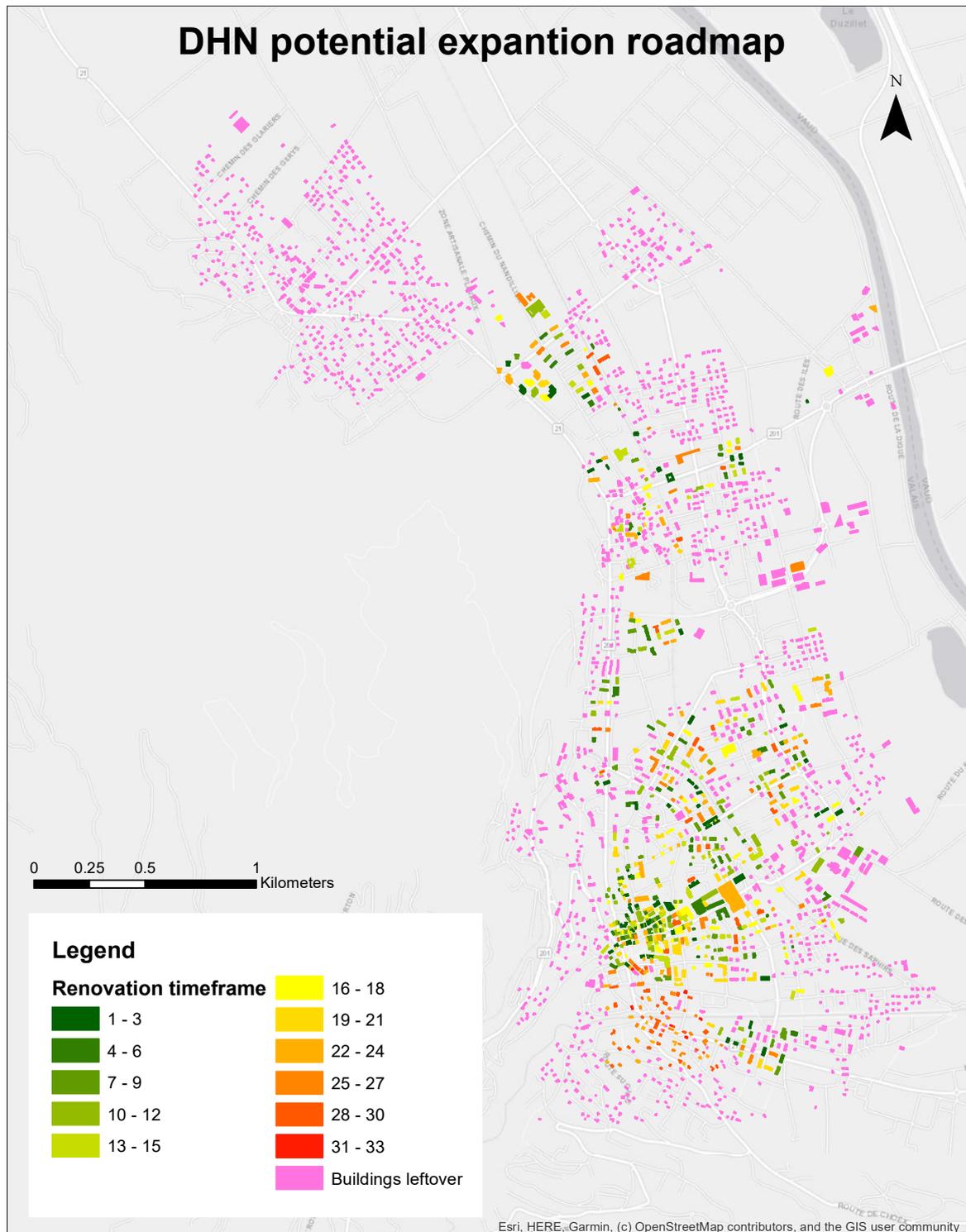
Table 4.6 makes it clear that retrofitting connected buildings, while challenging, has a significant impact on reducing peak demand and enables the connection of a substantially larger number of buildings compared to retrofitting connectable buildings. However, retrofitting connectable buildings also creates additional

opportunities to supply more buildings with renewable energy resources. Despite these facts, it's evident that the existing network lacks the capacity to accommodate a large number of buildings. In the best-case scenario, Scenario 5, only 12.7% of connectable buildings can be connected. Achieving higher capacity would require global retrofitting, which is not a feasible or realistic scenario.



**Figure 4.17.** Real connectable buildings considering different retrofitting scenarios

Scenario 5, as highlighted earlier, stands out as the most advantageous energy-saving approach for prospective DHN expansion. In this regard, Figure 4.18 provides a graphical representation of the stepwise integration of feasible buildings into the DHN over the upcoming 30 years, divided into 3-year intervals.



**Figure 4.18.** Expansion roadmap of district heating network over next 30 years in 3-year intervals

To formulate a strategy for building renovations, a random selection process was employed to determine which connected buildings would undergo the retrofitting procedure, specifically involving window substitution, as previously outlined. It is worth emphasizing that the annual renovation rate remained consistent at 1% collectively for both connected and connectable buildings, accumulating over the years.

With the presence of the existing DHN and actual connectable buildings, an algorithm was created to identify the shortest routes for connecting buildings to their nearest barycenter based on their centroid coordinates. Figure 4.19 depicts the optimized pathways for expanding the District Heating Network under the optimal energy-saving scenario.

To develop this algorithm, a networkx-based tree graph was employed to generate all potential routes between barycenter and their nearest buildings. The “Minimum Spanning Tree (MST)” technique was then applied to identify the shortest route in this graph. Additionally, the “Heapq” library was integrated to consistently select the nearest building among all possible routes from a given individual building to its neighboring structures.

While the developed algorithm yielded promising results in identifying the optimal routes for DHN expansion, there is still potential for improvement. The algorithm, as currently implemented, utilized a cloud of points as barycenter and distributed connectable buildings spatially without constraints on connections, which occasionally resulted in intersections with the existing network, leading to inefficiencies. To mitigate this issue, the optimal solution could involve mapping the networkx graph onto the urban pathway morphology and determining the shortest expansion route. However, it's essential to note that there are highly intricate algorithms designed specifically for finding the shortest paths without crossing other edges, although they can be computationally intensive.

To achieve this objective, it becomes imperative to concurrently incorporate the existing network within the graph and pursue the optimal expansion of the network towards all connectable buildings simultaneously. To accomplish this, the integration of advanced algorithms such as the “Traveling Salesman” becomes indispensable. The Traveling Salesman algorithm is a renowned optimization

technique employed to find the most efficient route that visits a set of given locations and returns to the starting point while minimizing the total distance traveled. This endeavor can significantly elevate the complexity involved in advancing the algorithm.



**Figure 4.19.** Optimal expansion pathway for District Heating Network to connect feasible buildings

As per the proposed optimal network expansion plan, which leverages a highly efficient algorithm to connect buildings through the shortest feasible routes, the expansion project will require approximately 8 kilometers of additional piping. This is to be added to the existing network, which already spans a length of 79 kilometers

Following the optimal energy-saving scenario, the annual HD and maximum peak demand of buildings are computed at the urban scale. The calculations are presented in Table 4.7.

**Table 4.7.** *The annual heating demand and peak demand of buildings connected after DHN expansion*

| Type of buildings                 |                           | Model    | Annual Demand (GWh/Y) | Peak Demand (MWh) |
|-----------------------------------|---------------------------|----------|-----------------------|-------------------|
| <b>Residential buildings</b>      | No.                       | --       |                       | 538               |
|                                   | Volume (Mm <sup>3</sup> ) | --       |                       | 2.48              |
|                                   | HD                        | LightGBM | 56.19                 | 19.9              |
| <b>Multi-functional buildings</b> | No.                       | --       |                       | 37                |
|                                   | Volume (Mm <sup>3</sup> ) | ---      |                       | 0.85              |
|                                   | HD (GWh/Y)                | LightGBM | 14.57                 | 5.41              |
| <b>Non-residential buildings</b>  | No.                       | --       |                       | 92                |
|                                   | Volume (Mm <sup>3</sup> ) | --       |                       | 0.64              |
|                                   | HD (GWh/Y)                | LightGBM | 16.42                 | 5.12              |

When considering the capacity of the CHP generator, it becomes evident that it holds substantial potential to serve a considerably larger number of buildings. However, this potential can only be fully harnessed if the heat transfer capacity of the network is substantially improved. To tackle this challenge, there are several viable scenarios worthy of consideration.

One viable strategy entails the installation of new pipes and the resizing of existing ones within the network. This approach would necessitate a significant infrastructural adjustment to ensure it can adequately handle the increased demand for heat transfer.

Alternatively, another feasible option is to seek support from a new or existing CHP plant that comes equipped with its own well-established piping infrastructure. By leveraging the capabilities of such a plant and integrating it into the existing network, the limitations of the current system could be effectively overcome.

Both of these strategies present promising avenues for expanding the network's capacity and ensuring that it can fully meet the heating demands of an expanded user base.

Furthermore, taking into account the statistical projection that over the next three decades, the frequency of buildings connected to the network will increase by 55% (equivalent to a 21% increase in the volume of buildings), the annual heating demand has only experienced a modest 8.2% increase, rising from 80.57 (GWh/Y) to 87.18 (GWh/Y). This notable achievement can be attributed to the effectiveness of energy-saving scenarios that incorporate various renovation technologies. In this light, it becomes imperative to incentivize property owners to undertake renovations and thereby tap into additional renewable energy resources.

The benefits of such a proactive approach are multifaceted. Firstly, it contributes to the reduction of greenhouse gas emissions, playing a pivotal role in combating climate change. Secondly, the expansion of the District Heating Network paves the way for a more sustainable and eco-friendlier urban environment, leading to cleaner air quality and reduced pollution. Furthermore, as the DHN grows, it offers a reliable and cost-effective solution for meeting the heating needs of a burgeoning population of connected buildings.

In conclusion, the combined efforts of expanding the DHN and promoting energy-saving renovations hold significant promise for contributing to improvements in climate conditions.

## 5. CONCLUSION

The research undertaken in this thesis, has unveiled a comprehensive exploration of advanced methodologies aimed at enhancing our understanding of urban building energy dynamics. By amalgamating the principles of Urban Building Energy Modeling (UBEM) with the potency of Machine Learning (ML) techniques, the study has made notable strides in assessing the potential for energy savings and harnessing renewable energy resources within the context of Satom.

Fundamental to the study was the development of a solid model of buildings. This was done carefully using Geographic Information System (GIS) software. This step not only made the modeling process more accurate but also made it easier to gather important information in the GIS database. By adding geographical details to the model, we got a better understanding of building features and where they are located, making our findings stronger.

The incorporation of LightGBM and Random Forest, both cutting-edge ML algorithms, has introduced a new era of Data-driven analysis. These techniques, leveraged through a bottom-up approach, have demonstrated remarkable predictive capabilities, enabling precise estimations of energy demand patterns and revealing insights into optimal utilization of renewable energy sources. The integration of state-of-the-art ML techniques has enabled a thorough assessment of building energy dynamics, leading to a more comprehensive view of energy efficiency.

The strength of the methodologies employed lies in their complementary nature. The harmonious interplay of data-driven techniques with a physically grounded GIS model has generated a comprehensive analytical approach that effectively captures the intricacies of urban building energy dynamics. The proficiency showcased by LightGBM and Random Forest, has unveiled a multifaceted perspective on heating demand patterns, potential energy savings, and the feasibility of renewable energy integration. However, in the context of this research, LightGBM outperforms RF in terms of both computational speed and accuracy, whether at the Building-scale or the Urban-scale. Consequently, the utilization of

LightGBM proved advantageous in predicting peak demand for buildings, which in turn facilitated the analysis of potential DHN expansion.

Moreover, the research has embarked on an innovative journey, constructing a forward-looking roadmap for the next 30 years within a case study focused on exploring energy-saving possibilities within buildings. This proactive initiative intricately intertwined investigations into the physical characteristics and energy efficiency of structures with a consideration of the socio-economic context of the surrounding environment. By introducing custom renovation scenarios tailored to each building and subsequently assessing post-renovation energy demands, a careful and thoughtful selection process took place, designating specific buildings for integration into the district heating network.

The true strength of this initiative lies in its iterative nature, manifesting through a recurring cycle, progressively optimizing the network's capacity to accommodate connectable buildings. This comprehensive endeavor exemplifies a pioneering approach that harmoniously balances technological innovation, environmental considerations, and infrastructural planning, ultimately paving the way towards a harmonious and sustainable urban energy ecosystem in the years to come.

To sum up, as long as cities continue to face the challenge of increasing energy needs, these approaches provide valuable instruments to shape policy creation, infrastructure design, and the sustainable growth of urban areas. They point the way toward a future marked by efficient energy use and a strong commitment to environmental well-being.

## **5.1. Limitations of the Study**

The current study faced several limitations, with the primary constraint being the scarcity of data within the context of the case study. The methodology employed heavily relies on Data-driven models, and the availability of data is a critical factor in the development of highly effective energy use models. Specifically, the lack of socio-economic information, often deemed confidential or private, posed a significant challenge, as it hindered the ability to incorporate occupants' behavior

into the Data-driven models. Furthermore, the limited richness of available physical data, which serves as the foundational data source for energy use models, presented obstacles in advancing predictive models. These data limitations underscore the need for more comprehensive and diverse datasets to enhance the accuracy and applicability of future research efforts in this domain.

Another significant constraint revolved around the researcher's limited proficiency in deploying advanced and complex machine learning (ML) algorithms. While the ML methods employed in this study exhibited considerable predictive power, it is apparent that more complex ML approaches, such as elaborate ensemble models, artificial neural networks, or deep learning methodologies, have the capacity to achieve even greater predictive accuracy. These sophisticated ML techniques often entail intricate architectures, multiple layers, and intricate mathematical computations, allowing them to capture subtle data patterns and intricate relationships, which can result in more precise predictive outcomes.

Furthermore, the building energy use models developed for predicting peak demand and evaluating the potential network expansion are constructed based on simplified assumptions regarding the physical model of the district heating network (DHN) and the willingness of property owners to adopt retrofitting technologies. Any deviations or oversimplifications in these underlying assumptions have the potential to introduce inaccuracies into the models. Real-world conditions frequently encompass a wide range of diverse factors that may not perfectly conform to the assumptions made in the models, ultimately impacting the reliability of energy demand predictions.

Last but not least, the study proposed a forward-looking roadmap for district heating network (DHN) expansion, it primarily focuses on the technical aspects of expansion. In reality, DHN expansion is a multifaceted process that involves navigating a complex landscape of regulatory requirements, financial considerations, and political dynamics. Issues related to permitting, securing funding, and gaining community and stakeholder support can pose substantial challenges. The study doesn't delve into these intricate complexities, which could impact the feasibility and timeline of implementing the proposed DHN expansion strategies.

## **5.2. Recommendation for future research**

To address the limitations of the current research, while still taking advantage of its methodological approach, future research should focus on collecting more comprehensive and diverse datasets. This could be done by partnering with other organizations that have access to relevant data, such as utility companies or government agencies. Additionally, researcher could develop new methods for extracting information from existing datasets.

In addition, future research on DHN expansion should develop a more comprehensive methodology that uses a hybrid model. This model would integrate data-driven predictive algorithms for energy use models with process-driven models. Process-driven models are based on physical principles that govern the flow of energy through a system. These models can be used to perform detailed analysis of the network without having to simplify it, however they can be computationally expensive. This would allow for more accurate and comprehensive planning of DHN expansion.

Ultimately, future research on DHN expansion should focus on developing more accurate models that take into account a wider range of factors. This could be done by conducting more detailed surveys of property owners to understand their energy usage habits, their willingness to adopt energy-saving measures, and their financial constraints. Additionally, more sophisticated Machine Learning techniques could be used to identify patterns in the data and to make predictions about future energy use. Mutually, future research should consider the regulatory, financial, and political factors that could impact DHN expansion. By understanding these factors, researcher could develop more realistic and feasible roadmaps for DHN expansion.

## 6. REFERENCES

1. Abbasabadi, N., & Ashayeri, M. (2019). Urban energy use modeling methods and tools: A review and an outlook. *Building and Environment*, 161, 106270.
2. Abolhassani, S. S., Amayri, M., Bouguila, N., & Eicker, U. (2022). A new workflow for detailed urban scale building energy modeling using spatial joining of attributes for archetype selection. *Journal of Building Engineering*, 46, 103661.
3. Ahmad, M. W., Mourshed, M., & Rezugui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and buildings*, 147, 77-89.
4. Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2018). A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy and Buildings*, 165, 301-320.
5. Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019, July). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623-2631).
6. Al Daoud, E. (2019). Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *International Journal of Computer and Information Engineering*, 13(1), 6-10.
7. Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 1192-1205.
8. Ang, B. W., Choong, W. L., & Ng, T. S. (2015). Energy security: Definitions, dimensions and indexes. *Renewable and sustainable energy reviews*, 42, 1077-1093.
9. Ang, Y. Q., Berzolla, Z. M., & Reinhart, C. F. (2020). From concept to application: A review of use cases in urban building energy modeling. *Applied Energy*, 279, 115738.
10. Ashrae, A. (2009). ASHRAE/USGBC/IES Standard 189.1 e2009, Standard for the Design of High-Performance Green Buildings. American Society of Heating, Refrigerating and Air-Conditioning Engineers.
11. Basu, S., Bale, C. S., Wehnert, T., & Topp, K. (2019). A complexity approach to defining urban energy systems. *Cities*, 95, 102358.
12. Ben, H., & Steemers, K. (2020). Modelling energy retrofit using household archetypes. *Energy and Buildings*, 224, 110224.
13. Benonysson, A., Bøhm, B., & Ravn, H. F. (1995). Operational optimization in a district heating system. *Energy conversion and management*, 36(5), 297-314.
14. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).

15. Bergstra, J., Yamins, D., & Cox, D. D. (2013, June). Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In Proceedings of the 12th Python in science conference (Vol. 13, p. 20).
16. Bjørnskov, J., Mortensen, L. K., Filonenko, K., Shaker, H. R., Jradi, M., & Veje, C. (2021). Optimization of district heating production with thermal storage using mixed-integer nonlinear programming with a new initialization approach. *Energy Informatics*, 4(2), 1-17.
17. Boghetti, R., Fantozzi, F., Kampf, J., Mutani, G., Salvadori, G., & Todeschi, V. (2020). Building energy models with Morphological urban-scale parameters: A case study in Turin. *BUILDING SIMULATION APPLICATIONS BSA...*, 2020, 131-139.
18. Bordin, C., Gordini, A., & Vigo, D. (2016). An optimization approach for district heating strategic network design. *European Journal of Operational Research*, 252(1), 296-307.
19. Breiman, L., & Cutler, R. A. (2001). Random forests machine learning [J]. *journal of clinical microbiology*, 2, 199-228.
20. Bro, R., Kjeldahl, K., Smilde, A. K., & Kiers, H. A. L. (2008). Cross-validation of component models: a critical look at current methods. *Analytical and bioanalytical chemistry*, 390, 1241-1251.
21. Brownlee, J. (2016). Master Machine Learning Algorithms Discover How They Work and Implement Them From Scratch i Master Machine Learning Algorithms. Technical report.
22. Brownlee, J. (2020). Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning. *Machine Learning Mastery*.
23. Burges, C. J. (2010). From ranknet to lambdarank to lambdamart: An overview. *Learning*, 11(23-581), 81.
24. Bruse, M., Nouvel, R., Wate, P., Kraut, V., & Coors, V. (2019). An energy-related CityGML ADE and its application for heating demand calculation. In *Architecture and Design: Breakthroughs in Research and Practice* (pp. 1306-1323). IGI Global.
25. Cerezo, C., Dogan, T., & Reinhart, C. (2014, September). Towards standardized building properties template files for early design energy model generation. In Proceedings of the ASHRAE/IBPSA-USA building simulation conference (pp. 25-32).
26. Chen, Y., Guo, M., Chen, Z., Chen, Z., & Ji, Y. (2022). Physical energy and data-driven models in building energy prediction: A review. *Energy Reports*, 8, 2656-2671.
27. Crawley, D. B., Hand, J. W., & Lawrie, L. K. (1999, September). Improving the weather information available to simulation programs. In Proceedings of Building Simulation'99 (Vol. 2, pp. 529-536).
28. Dalipi, F., Yildirim Yayilgan, S., & Gebremedhin, A. (2016). Data-driven machine-learning model in district heating system for heat load prediction: A comparison study. *Applied Computational Intelligence and Soft Computing*, 2016.
29. Delmastro, C., Mutani, G., & Corgnati, S. P. (2016). A supporting method for selecting cost-optimal energy retrofit policies for residential buildings at the urban scale. *Energy Policy*, 99, 42-56.

30. Déqué, F., Ollivier, F., & Poblador, A. (2000). Grey boxes used to represent buildings with a minimum number of geometric and thermal parameters. *Energy and buildings*, 31(1), 29-35.
31. Di Francescomarino, C., Dumas, M., Federici, M., Ghidini, C., Maggi, F. M., Rizzi, W., & Simonetto, L. (2018). Genetic algorithms for hyperparameter optimization in predictive business process monitoring. *Information Systems*, 74, 67-83.
32. Dominković, D. F., Stunjek, G., Blanco, I., Madsen, H., & Krajačić, G. (2020). Technical, economic and environmental optimization of district heating expansion in an urban agglomeration. *Energy*, 197, 117243.
33. Dou, Y., Sun, L., Fujii, M., Kikuchi, Y., Kanematsu, Y., & Ren, J. (2021). Towards a renewable-energy-driven district heating system: key technology, system design and integrated planning. *Renewable-Energy-Driven Future*, 311-332.
34. Ekundayo, I. (2020). Optuna optimization based cnn-lstm model for predicting electric power consumption (Doctoral dissertation, Dublin, National College of Ireland).
35. European Commission. Available online: [https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal\\_en](https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en) (accessed on 22 February 2023).
36. Fang, H., Xia, J., Zhu, K., Su, Y., & Jiang, Y. (2013). Industrial waste heat utilization for low temperature district heating. *Energy policy*, 62, 236-246.
37. Ferrando, M., Causone, F., Hong, T., & Chen, Y. (2020). Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches. *Sustainable Cities and Society*, 62, 102408.
38. Fouquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, 23, 272-288.
39. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
40. Gassar, A. A. A., & Cha, S. H. (2020). Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review. *Energy and Buildings*, 224, 110238.
41. Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. "O'Reilly Media, Inc."
42. Guelpa, E., Mutani, G., Todeschi, V., & Verda, V. (2017). A feasibility study on the potential expansion of the district heating network of Turin. *Energy Procedia*, 122, 847-852.
43. Guelpa, E., Mutani, G., Todeschi, V., & Verda, V. (2018). Reduction of CO2 emissions in urban areas through optimal expansion of existing district heating networks. *Journal of Cleaner Production*, 204, 117-129.
44. Guelpa, E., Toro, C., Sciacovelli, A., Melli, R., Sciubba, E., & Verda, V. (2016). Optimal operation of large district heating networks through fast fluid-dynamic simulation. *Energy*, 102, 586-595.

45. Hall, I. J., Prairie, R. R., Anderson, H. E., & Boes, E. C. (1978). Generation of a typical meteorological year (No. SAND-78-1096C; CONF-780639-1). Sandia Labs., Albuquerque, NM (USA).
46. Heidelberg, E., & Rakha, T. (2022). Inclusive urban building energy modeling through socioeconomic data: A persona-based case study for an underrepresented community. *Building and Environment*, 222, 109374.
47. Holmgren, K. (2006). Role of a district-heating network as a user of waste-heat supply from various sources—the case of Göteborg. *Applied energy*, 83(12), 1351-1367.
48. Hong, T., Piette, M. A., Chen, Y., Lee, S. H., Taylor-Lange, S. C., Zhang, R., ... & Price, P. (2015). Commercial building energy saver: An energy retrofit analysis toolkit. *Applied Energy*, 159, 298-309.
49. Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., & Modi, V. (2012). Spatial distribution of urban building energy consumption by end use. *Energy and buildings*, 45, 141-151.
50. IEA/IRENA Renewables Policies Database, 2013. (<https://www.iea.org/policies/5135-promotion-of-high-efficiency-cogeneration>).
51. IEA/IRENA Renewables Policies Database, 2019. (<https://www.iea.org/policies/890-legislative-decree-dl-2811-03032011-implementating-directive-200928ce-on-promotion-of-use-of-energy-from-res-modifying-and-repealing-directives-200177ce-and-200330ce-guarantee-fund-and-planning-for-district-heating-feed-in-tariff-targets>).
52. IEA/IRENA Renewables Policies Database, 2019. (<https://www.iea.org/policies/1355-legislative-decree-4-june-2013-dl-462013-transposition-of-the-european-directive-201031ue-on-the-energy-performance-of-buildings>).
53. IEA/IRENA Renewables Policies Database, 2020. (<https://www.iea.org/policies/606-danish-energy-agreement-for-2012-2020?s=1>).
54. Installations, E. T. F. C. (2008). Guidelines for District Heating Substations. *Euroheat & Power*, 68.
55. Jain, R. K., Smith, K. M., Culligan, P. J., & Taylor, J. E. (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, 123, 168-178.
56. Johansson, T., Olofsson, T., & Mangold, M. (2017). Development of an energy atlas for renovation of the multifamily building stock in Sweden. *Applied Energy*, 203, 723-736.
57. Johnston, D. (2003). A physically-based energy and carbon dioxide emission model of the UK housing stock (Doctoral dissertation, Leeds Metropolitan University).
58. Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia tools and applications*, 80, 8091-8126.
59. Kavgić, M., Mavrogianni, A., Mumović, D., Summerfield, A., Stevanović, Z., & Djurović-Petrović, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Building and environment*, 45(7), 1683-1697.

60. Kontokosta, C. E., & Tull, C. (2017). A data-driven predictive model of city-scale energy use in buildings. *Applied energy*, 197, 303-317.
61. Kwok, S. S., & Lee, E. W. (2011). A study of the importance of occupancy to building cooling load in prediction by intelligent approach. *Energy Conversion and Management*, 52(7), 2555-2564.
62. Kyriakides, G., & Margaritis, K. G. (2019). *Hands-On Ensemble Learning with Python: Build highly optimized ensemble machine learning models using scikit-learn and Keras*. Packt Publishing Ltd.
63. Li, P. (2012). Robust logitboost and adaptive base class (abc) logitboost. arXiv preprint arXiv:1203.3491.
64. Li, W., Zhou, Y., Cetin, K., Eom, J., Wang, Y., Chen, G., & Zhang, X. (2017). Modeling urban building energy use: A review of modeling approaches and procedures. *Energy*, 141, 2445-2457.
65. Lidberg, T., Olofsson, T., & Ödlund, L. (2019). Impact of domestic hot water systems on district heating temperatures. *Energies*, 12(24), 4694.
66. Lindenberger, D., Bruckner, T., Groscurth, H. M., & Kümmel, R. (2000). Optimization of solar district heating systems: seasonal storage, heat pumps, and cogeneration. *Energy*, 25(7), 591-608.
67. Liu, S., Zou, Y., Ji, W., Zhang, Q., Ahmed, A., Han, X., ... & Zhang, S. (2022). Energy-saving potential prediction models for large-scale building: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 156, 111992.
68. Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J. E., Hvelplund, F., & Mathiesen, B. V. (2014). 4th Generation District Heating (4Gdomestic): Integrating smart thermal grids into future sustainable energy systems. *Energy*, 68, 1-11.
69. Lund, H., Østergaard, P. A., Chang, M., Werner, S., Svendsen, S., Sorknæs, P., ... & Möller, B. (2018). The status of 4th generation district heating: Research and results. *Energy*, 164, 147-159.
70. Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2014). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865-873.
71. Madsen, H., Sejling, K., Søgaaard, H. T., & Palsson, O. P. (1994). On flow and supply temperature control in district heating systems. *Heat Recovery Systems and CHP*, 14(6), 613-620.
72. Malhotra, A., Bischof, J., Nichersu, A., Häfele, K. H., Exenberger, J., Sood, D., ... & Schweiger, G. (2022). Information modelling for urban building energy simulation—A taxonomic review. *Building and Environment*, 208, 108552.
73. Michalewicz, Z. (1992). *Genetic Algorithms+ Data Structures= Evolution Programs* Springer-Verlag, New York, NY.
74. Mills, E. (2004). Inter-comparison of North American residential energy analysis tools. *Energy and Buildings*, 36(9), 865-880.

75. MIT, (1997). Energy technology availability: review of longer-term scenarios for development and deployment of climate-friendly technologies. Cambridge, Massachusetts, USA: Massachusetts Institute of Technology Energy Laboratory.
76. Moghadam, S. T., Delmastro, C., Corgnati, S. P., & Lombardi, P. (2017). Urban energy planning procedure for sustainable development in the built environment: A review of available spatial approaches. *Journal of Cleaner Production*, 165, 811-827.
77. Moghadam, S. T., Toniolo, J., Mutani, G., & Lombardi, P. (2018). A GIS-statistical approach for assessing built environment energy use at urban scale. *Sustainable Cities and Society*, 37, 70-84.
78. Mutani, G., Fabiano, E., Garcia, D. A., & Mancini, F. (2021, November). Spatial energy modelling for the Metropolitan City of Rome. In *2021 IEEE 4th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE)* (pp. 43-48). IEEE.
79. Mutani, G., & Todeschi, V. (2017). Space heating models at urban scale for buildings in the city of Turin (Italy). *Energy Procedia*, 122, 841-846.
80. Mutani, G., & Todeschi, V. (2019). An urban energy atlas and engineering model for resilient cities. *International Journal of Heat and Technology*, 37(4), 936-947.
81. Mutani, G., & Todeschi, V. (2020). Building energy modeling at neighborhood scale. *Energy Efficiency*, 13(7), 1353-1386.
82. Mutani, G., & Todeschi, V. (2021). GIS-based urban energy modelling and energy efficiency scenarios using the energy performance certificate database. *Energy Efficiency*, 14(5), 1-28.
83. Mutani, G., Todeschi, V., Guelpa, E., & Verda, V. (2020). Building Efficiency Models and the Optimization of the District Heating Network for Low-Carbon Transition Cities. In *Improving Energy Efficiency in Commercial Buildings and Smart Communities* (pp. 217-241). Springer, Cham.
84. Mutani, G., Todeschi, V., Kämpf, J., Coors, V., & Fitzky, M. (2018, October). Building energy consumption modeling at urban scale: Three case studies in Europe for residential buildings. In *2018 IEEE International Telecommunications Energy Conference (INTELEC)* (pp. 1-8). IEEE.
85. Mutani, G., Vocale, P., & Javanroodi, K. (2023). Toward Improved Urban Building Energy Modeling Using a Place-Based Approach. *Energies*, 16(9), 3944.
86. Nagota, T., Shimoda, Y., & Mizuno, M. (2008). Verification of the energy-saving effect of the district heating and cooling system—Simulation of an electric-driven heat pump system. *Energy and Buildings*, 40(5), 732-741.
87. Nouvel, R., Mastrucci, A., Leopold, U., Baume, O., Coors, V., & Eicker, U. (2015). Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support. *Energy and Buildings*, 107, 204-212.

88. Nuorkivi, A. (2016). District heating and cooling policies worldwide. In *Advanced district heating and cooling (DHC) systems* (pp. 17-41). Woodhead Publishing.
89. Nutkiewicz, A., Yang, Z., & Jain, R. K. (2017). Data-driven Urban Energy Simulation (DUE-S): Integrating machine learning into an urban building energy simulation workflow. *Energy Procedia*, 142, 2114-2119.
90. Nutkiewicz, A., Yang, Z., & Jain, R. K. (2018). Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Applied energy*, 225, 1176-1189.
91. Olhoff, A., & Christensen, J. M. (Eds.) (2016). *The Emission Gap Report 2016: A UNEP Synthesis Report*. UNEP. <http://at.unep.org/theme/index/13#egr>
92. Perera, A. T. D., Javanroodi, K., Mauree, D., Nik, V. M., Florio, P., Hong, T., & Chen, D. (2023). Challenges resulting from urban density and climate change for the EU energy transition. *Nature Energy*, 1-16.
93. Raji, I. D., Bello-Salau, H., Umoh, I. J., Onumanyi, A. J., Adegboye, M. A., & Salawudeen, A. T. (2022). Simple deterministic selection-based genetic algorithm for hyperparameter tuning of machine learning models. *Applied Sciences*, 12(3), 1186.
94. Rao, R. B., Fung, G., & Rosales, R. (2008, April). On the dangers of cross-validation. An experimental evaluation. In *Proceedings of the 2008 SIAM international conference on data mining* (pp. 588-596). Society for Industrial and Applied Mathematics.
95. Reinhart, C. F., & Davila, C. C. (2016). Urban building energy modeling—A review of a nascent field. *Building and Environment*, 97, 196-202.
96. Reinhart, C., & Davila, C. C. (2019). Urban building energy modeling. *Building Performance Simulation for Design and Operation*, 696-722.
97. Resch, E., Bohne, R. A., Kvamsdal, T., & Lohne, J. (2016). Impact of urban density and building height on energy use in cities. *Energy Procedia*, 96, 800-814.
98. Rezaie, B., & Rosen, M. A. (2012). District heating and cooling: Review of technology and potential enhancements. *Applied energy*, 93, 2-10.
99. Richardson, M., Dominowska, E., & Ragno, R. (2007, May). Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web* (pp. 521-530).
100. Rivers, N., & Jaccard, M. (2005). Combining top-down and bottom-up approaches to energy-economy modeling using discrete choice methods. *The energy journal*, 26(1).
101. Roland, M., & Schmidt, M. (2020). Mixed-integer nonlinear optimization for district heating network expansion. *at-Automatisierungstechnik*, 68(12), 985-1000.
102. Roth A. The shockingly short payback of energy modeling; 2016. Retrieved from U.S. department of Energy's website: <https://www.energy.gov/eere/buildings/articles/shockingly-short-payback-energy-modeling>.
103. Ryan, E. M., & Sanquist, T. F. (2012). Validation of building energy modeling tools under idealized and realistic conditions. *Energy and buildings*, 47, 375-382.

104. Sailor, D. J., & Lu, L. (2004). A top–down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas. *Atmospheric environment*, 38(17), 2737-2748.
105. Salomón, M., Gómez Galindo, M. F., & Martin, A. (2014). Upgrading a biomass-based district heating plant in sweden: a technoeconomic optimization and sensitivity analysis. *Journal of Cleaner Production*.
106. Sayegh, M. A., Jadwyszczak, P., Axcell, B. P., Niemierka, E., Bryś, K., & Jouhara, H. (2018). Heat pump placement, connection and operational modes in European district heating. *Energy and Buildings*, 166, 122-144.
107. Shaikh, P. H., Nor, N. B. M., Nallagownden, P., Elamvazuthi, I., & Ibrahim, T. (2014). A review on optimized control systems for building energy and comfort management of smart sustainable buildings. *Renewable and Sustainable Energy Reviews*, 34, 409-429.
108. Shorrock, L. D., & Dunster, J. E. (1997). The physically-based model BREHOMES and its use in deriving scenarios for the energy use and carbon dioxide emissions of the UK housing stock. *Energy Policy*, 25(12), 1027-1037.
109. Sikder, S. K., Nagarajan, M., Kar, S., & Koetter, T. (2018). A geospatial approach of downscaling urban energy consumption density in mega-city Dhaka, Bangladesh. *Urban climate*, 26, 10-30.
110. Singh, K. P., Gupta, S., & Rai, P. (2013). Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmospheric Environment*, 80, 426-437.
111. Sipilä, K. (2016). Cogeneration, biomass, waste to energy and industrial waste heat for district heating. In *Advanced District Heating and Cooling (DHC) Systems* (pp. 45-73). Woodhead Publishing.
112. Sperling, K., & Möller, B. (2012). End-use energy savings and district heating expansion in a local renewable energy system—A short-term perspective. *Applied energy*, 92, 831-842.
113. Srinivas, P., & Katarya, R. (2022). hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomedical Signal Processing and Control*, 73, 103456.
114. Sola, A., Corchero, C., Salom, J., & Sanmarti, M. (2018). Simulation tools to build urban-scale energy models: A review. *Energies*, 11(12), 3269.
115. Sola, A., Corchero, C., Salom, J., & Sanmarti, M. (2020). Multi-domain urban-scale energy modelling tools: A review. *Sustainable Cities and Society*, 54, 101872.
116. Sperling, K., & Möller, B. (2012). End-use energy savings and district heating expansion in a local renewable energy system—A short-term perspective. *Applied energy*, 92, 831-842.
117. Sun, F., Fu, L., Zhang, S., & Sun, J. (2012). New waste heat district heating system with combined heat and power based on absorption heat exchange cycle in China. *Applied Thermal Engineering*, 37, 136-144.

118. Sun, J., Gong, M., Zhao, Y., Han, C., Jing, L., & Yang, P. (2022). A hybrid deep reinforcement learning ensemble optimization model for heat load energy-saving prediction. *Journal of Building Engineering*, 58, 105031.
119. Sun, Y., Haghghat, F., & Fung, B. C. (2020). A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy and Buildings*, 221, 110022.
120. Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and sustainable energy reviews*, 13(8), 1819-1835.
121. Thorsen, J. E., Christiansen, C. H., Brand, M., Olesen, P. K., & Larsen, C. T. (2011, March). Experiences on low-temperature district heating in Lystrup–Denmark. In *Proceedings of the International Conference on District Energy* (pp. 1-10).
122. Todeschi, V., Boghetti, R., Kämpf, J. H., & Mutani, G. (2021). Evaluation of urban-scale building energy-use models and tools—Application for the city of Fribourg, Switzerland. *Sustainability*, 13(4), 1595.
123. Todeschi, V., Javanroodi, K., Castello, R., Mohajeri, N., Mutani, G., & Scartezzini, J. L. (2022). Impact of the COVID-19 pandemic on the energy performance of residential neighborhoods and their occupancy behavior. *Sustainable Cities and Society*, 82, 103896.
124. Tornberg, J., & Thuvander, L. (2005). A GIS energy model for the building stock of Goteborg. In *ESRI International User Conference Proceedings*.
125. van der Zwan, S., & Pothof, I. (2020). Operational optimization of district heating systems with temperature limited sources. *Energy and Buildings*, 226, 110347.
126. Verda, V., Guelpa, E., Kona, A., & Russo, S. L. (2012). Reduction of primary energy needs in urban areas through optimal planning of district heating and heat pump installations. *Energy*, 48(1), 40-46.
127. Veza, P., Comoglio, C., Rosso, M., & Viglione, A. (2010). Low flows regionalization in north-western Italy. *Water resources management*, 24, 4049-4074.
128. Wang, L., Lee, E. W., & Yuen, R. K. (2018). Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. *Applied Energy*, 228, 1740-1753.
129. Wang, Z., & Srinivasan, R. S. (2017). A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*, 75, 796-808.
130. Wenjing, L., Yanuar, S., & Perry, P. J. (2004). GIS analysis for the climatic evaluation of 3d urban geometry: The development of GIS analytical tools for sky view factor.
131. Werner, S. (2013). *District heating and cooling*.
132. Wiltshire, R. (Ed.). (2015). *Advanced district heating and cooling (DHC) systems*. Woodhead Publishing.
133. Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6), 3586-3592.

134. Zhongming, Z., Linong, L., Xiaona, Y., Wangqiang, Z., & Wei, L. (2021). The Emissions Gap Report 2021. <https://www.unep.org/resources/emissions-gap-report-2021>
135. Østergaard, D. S., Smith, K. M., Tunzi, M., & Svendsen, S. (2022). Low-temperature operation of heating systems to enable 4th generation district heating: A review. *Energy*, 123529.
136. <http://data.europa.eu/eli/dir/2002/91/oj>
137. <http://data.europa.eu/eli/dir/2012/27/oj>
138. <http://data.europa.eu/eli/dir/2018/2002/oj>
139. [https://en.wikipedia.org/wiki/Mathematical\\_optimization](https://en.wikipedia.org/wiki/Mathematical_optimization)
140. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html#sklearn.ensemble.RandomForestRegressor>
141. [https://scikit-learn.org/stable/modules/cross\\_validation.html#cross-validation-iterators](https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation-iterators)
142. faqs.org, (2014). What are cross-validation and bootstrapping?: <http://www.faqs.org/faqs/ai-faq/neural-nets/part3/section-12.html>
143. Microsoft Corporation. (2023). <https://LightGBM.readthedocs.io/en/latest/index.html>
144. <https://www.minergie.ch/it/standard/ammodernare/ammodernamento-di-sistema/>

## APPENDIX A

The code block tunes LightGBM's hyperparameters utilizing Optuna optimizer.

```
# Importing libraries
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import LightGBM as lgb
from sklearn import linear_model
from sklearn.model_selection import train_test_split, KFold
from sklearn.metrics import mean_squared_error
import optuna
import time

# Reading the dataset
directory = r"C:\example"
df = pd.read_csv(directory + r"\Database.csv")

# Setting the value for x and y
X = df[['Age', 'Land use', 'S/V', 'Occupants', 'Wall thickness', 'U Roof', 'U Wall', 'U
Slab', 'U Window']]

y = df['EP (kWh/m3/Y)']

# Splitting the dataset into train and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42) #TODO adjust test size based on the frequency of samples

# Start timer
start = time.time()

# Define the objective function for Optuna
def objective(trial):

    param = {
        'objective': 'regression',
        'metric': 'rmse',
        'verbosity': -1,
        'boosting': 'gbdt',
        'force_col_wise': True,
        'num_leaves': trial.suggest_int('num_leaves', 50, 100),
        'max_depth': trial.suggest_int('max_depth', 1, 5),
        'min_data_in_leaf': trial.suggest_int('min_data_in_leaf', 1, 10),
        'feature_fraction': trial.suggest_float('feature_fraction', 0.4, 1.0),
        'bagging_fraction': trial.suggest_float('bagging_fraction', 0.4, 1.0),
        'bagging_freq': trial.suggest_int('bagging_freq', 1, 10),
        'lambda_l1': trial.suggest_float('lambda_l1', 0.0, 0.6),
```

```

    'lambda_l2': trial.suggest_float('lambda_l2', 0.0, 0.6),
    'max_bin': trial.suggest_int('max_bin', 500, 1000),
}

# Perform K-Fold CV
n_fold = 5 #TODO adjust No. Folds based on the frequency of samples
rmse_scores = []
kf = KFold(n_splits=n_fold, shuffle=True, random_state=42)
for train_index, valid_index in kf.split(X_train):
    train_x, valid_x = X_train.iloc[train_index], X_train.iloc[valid_index]
    train_y, valid_y = y_train.iloc[train_index], y_train.iloc[valid_index]

    train_dataset = lgb.Dataset(train_x, label=train_y)
    valid_dataset = lgb.Dataset(valid_x, label=valid_y)

    pruning_callback = optuna.integration.LightGBMPruningCallback(trial,
'rmse')

    lgb = lgb.train(param, train_dataset, valid_sets=[valid_dataset],
callbacks=[pruning_callback])
    pred = lgb.predict(valid_x)
    rmse_score = np.sqrt(mean_squared_error(valid_y, pred, squared=False))
    rmse_scores.append(rmse_score)

return np.mean(rmse_scores)

# Create LightGBM regressor
study = optuna.create_study(direction='minimize')
study.optimize(objective, n_trials=500)

# Stop timer
stop = time.time()

# Get the best parameters and best score
print('Best hyperparameters: {}'.format(study.best_params))
print('Best RMSE: {}'.format(study.best_value))

```

## APPENDIX B

The code block trains energy-use model using LightGBM and tuned hyperparameters and makes prediction for unseen data.

```
# Importing libraries
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import LightGBM as lgb
from sklearn import linear_model
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
import time

# Reading the dataset
directory = r"C:\example"
df = pd.read_csv(directory + r"\Database.csv")

# Setting the value for x and y
X = df[['Age', 'Land use', 'S/V', 'Occupants', 'Wall thickness', 'U Roof', 'U Wall', 'U
Slab', 'U Window']]

y = df['EP (kWh/m3/Y)']

# Splitting the dataset into train and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42) #TODO adjust test size based on the frequency of samples

# Define the LightGBM hyperparameters
params = {
    'objective': 'regression',
    'metric': 'rmse',
    'verbosity': -1,
    'boosting': 'gbdt',
    'force_col_wise': True,
    'num_leaves': 53,
    'max_depth': 5,
    'min_data_in_leaf': 1,
    'feature_fraction': 0.84,
    'bagging_fraction': 0.8,
    'bagging_freq': 8,
    'lambda_l1': 0.01,
    'lambda_l2': 0.1,
    'max_bin': 807
}
```

```

# Create dataset for training
train_data = lgb.Dataset(X_train, label=y_train)

# Start timer
start = time.time()

# Train the model
model = lgb.train(params, train_data, num_boost_round= 1000)

# Stop timer
stop = time.time()

# Make predictions on the testing set
y_pred = model.predict(X_test)

# Evaluate the model
RMSE = np.sqrt(mean_squared_error(y_test, y_pred))
print('RMSE - testing set: {}'.format(RMSE))

y_prediction = model.predict(X)
RMSE = np.sqrt(mean_squared_error(y, y_prediction))
print('RMSE - all dataset: {}'.format(RMSE))
print("Training time: {}s".format(stop - start))

df['Prediction_LightGBM'] = y_prediction
df.to_csv(directory + r"\Database.csv")

# Regression performance
new_y = y[:, None]

regr = linear_model.LinearRegression()
regr.fit(new_y, y_prediction)
line = regr.predict(new_y)
r2 = r2_score(new_y, y_prediction)

# Predicting unseen data
df = pd.read_csv(directory + r"\Unseen.csv")

# Setting the value for x and y
X = df[['Age', 'Land use', 'S/V', 'Occupants', 'Wall thickness', 'U Roof', 'U Wall', 'U
Slab', 'U Window']]

# Prediction
prediction = model.predict(X)
df['Prediction_lightGBM'] = prediction
df.to_csv(directory + r"\Unseen.csv")

```

## APPENDIX C

The code block tunes Random Forest (RF) regression's hyperparameters utilizing Genetic Algorithm (GA).

```
# Importing libraries
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split, KFold
from sklearn.metrics import mean_squared_error
from sklearn import linear_model
import random
import time

# Reading the dataset
directory = r"C:\example"
df = pd.read_csv(directory + r"\Database.csv")

# Setting the value for x and y
X = df[['Age', 'Land use', 'S/V', 'Occupants', 'Wall thickness', 'U Roof', 'U Wall', 'U Slab', 'U Window']]

y = df['EP (kWh/m3/Y)']

# Splitting the dataset into train and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42) #TODO adjust test size based on the frequency of samples

# Define the Genetic Algorithm parameters
population_size = 20
num_generations = 10
mutation_rate = 0.1

# Define the range of values for the hyperparameters
param_ranges = {
    'n_estimators': (100, 150),
    'max_depth': (10, 15),
    'min_samples_split': (2, 5),
    'min_samples_leaf': (2, 5)
}
```

```

# Define the fitness function
def evaluate_fitness(individual):
    # Create a Random Forest with the hyperparameters specified in the individual
    rf = RandomForestRegressor(
        n_estimators=individual[0],
        max_depth=individual[1],
        min_samples_split=individual[2],
        min_samples_leaf=individual[3],
    )

    n_fold = 5 #TODO adjust No. Folds based on the frequency of samples
    acc = []
    kf = KFold(n_splits=n_fold, shuffle=True, random_state=42)
    for train_index, valid_index in kf.split(X_train):
        train_x, valid_x = X_train.iloc[train_index], X_train.iloc[valid_index]
        train_y, valid_y = y_train.iloc[train_index], y_train.iloc[valid_index]

        rf.fit(train_x, train_y)
        accuracy = rf.score(valid_x, valid_y)
        acc.append(accuracy)

    return np.mean(acc)

# Define the genetic operators
def create_individual():
    return [np.random.randint(param_ranges[param][0], param_ranges[param][1] +
1) for param in param_ranges]

def crossover(parent1, parent2):
    child = []
    for i in range(len(parent1)):
        if np.random.random() < 0.5:
            child.append(parent1[i])
        else:
            child.append(parent2[i])
    return child

def mutate(individual):
    for i in range(len(individual)):
        if np.random.random() < mutation_rate:
            for key in param_ranges.keys():
                individual[i] = np.random.randint(param_ranges[key][0],
param_ranges[key][1] + 1)
    return individual

# Start timer
start = time.time()

```

```
# Initialize the population
population = [create_individual() for _ in range(population_size)]

# Start the evolution process
for generation in range(num_generations):
    # Evaluate the fitness of each individual
    fitness_scores = [evaluate_fitness(individual) for individual in population]

    # Select the fittest individuals for reproduction
    selected_indices = np.argsort(fitness_scores)[::-1][:int(population_size/2)]
    selected_individuals = [population[idx] for idx in selected_indices]

    # Create the next generation through crossover and mutation
    offspring = []
    while len(offspring) < population_size:
        i = random.randint(0, len(selected_individuals) - 1)
        j = random.randint(0, len(selected_individuals) - 1)
        parent1 = selected_individuals[i]
        parent2 = selected_individuals[j]
        child = crossover(parent1, parent2)
        child = mutate(child)
        offspring.append(child)

    # Replace the old population with the new generation
    population = offspring

# Select the best individual from the final population
best_individual = max(population, key=evaluate_fitness)
best_accuracy = evaluate_fitness(best_individual)

# Stop timer
stop = time.time()

print("Best Hyperparameters:", best_individual)
print("Best Accuracy:", best_accuracy)
```

## APPENDIX D

The code block trains energy-use model using Random Forest regression and tuned hyperparameters and makes prediction for unseen data.

```
# Importing libraries
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from sklearn import linear_model
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
import time

# Reading the dataset
directory = r"C:\example"
df = pd.read_csv(directory + r"\Database.csv")

# Setting the value for x and y
X = df[['Age', 'Land use', 'S/V', 'Occupants', 'Wall thickness', 'U Roof', 'U Wall', 'U
Slab', 'U Window']]

y = df['EP (kWh/m3/Y)']

# Splitting the dataset into train and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42) #TODO adjust test size based on the frequency of samples

# Define the RandomForestRegressor hyperparameters
params = {'n_estimators': 130,
          'max_depth': 12,
          'min_samples_split': 4,
          'min_samples_leaf': 2,
          'bootstrap': True,
          }

# Start timer
start = time.time()

# Train the model
model = RandomForestRegressor(**params)
model.fit(X_train, y_train)

# Stop timer
stop = time.time()
# Make predictions on the testing set
y_pred = model.predict(X_test)
```

```
# Evaluate the model
RMSE = np.sqrt(mean_squared_error(y_test, y_pred))
print('RMSE - testing set: {}'.format(RMSE))

y_prediction = model.predict(X)
RMSE = np.sqrt(mean_squared_error(y, y_prediction))
print('RMSE - all dataset: {}'.format(RMSE))
print("Training time: {}s".format(stop - start))
df['Prediction_RF'] = y_prediction
df.to_csv(directory + r"\Database.csv")

# Regression performance
new_y = y[:, None]

regr = linear_model.LinearRegression()
regr.fit(new_y, y_prediction)
line = regr.predict(new_y)
r2 = r2_score(new_y, y_prediction)

# Predicting unseen data
df = pd.read_csv(directory + r"\Unseen.csv")

# Setting the value for x and y
X = df[['Age', 'Land use', 'S/V', 'Occupants', 'Wall thickness', 'U Roof', 'U Wall', 'U
Slab', 'U Window']]

# Prediction
prediction = model.predict(X)
df['Prediction_RFR'] = prediction
df.to_csv(directory + r"\Unseen.csv")
```

## APPENDIX E

The table provides comprehensive data regarding the current network, including pipe codes, maximum capacity, and the existing load on each branch, considering various scenarios that account for connected buildings. All heating demand values in the table are expressed in the unit of watt-hours (WH).

| dn      | Starting point | Ending point | Max Capacity | Load - Not retrofitted | Load - Roof insulation | Load - Slab insulation | Load - Wall insulation | Load - Window insulation | Load - Globally retrofitted |
|---------|----------------|--------------|--------------|------------------------|------------------------|------------------------|------------------------|--------------------------|-----------------------------|
| 300/560 | A36            | A38          | 31796000     | 1166252.75             | 1109285.375            | 1099630.75             | 1115889.25             | 1003486.188              | 810728.4375                 |
| 300/560 | A32            | A35          | 31796000     | 16220395               | 15521010               | 15356562               | 15465106               | 14370029                 | 11500573                    |
| 300/560 | A3             | A21          | 31796000     | 22041590               | 21145300               | 20976290               | 21121782               | 19513334                 | 16017678                    |
| 300/560 | A3             | A20          | 31796000     | 61886.87109            | 62680.92188            | 59477.96094            | 62276.69141            | 59718.99609              | 52360.59766                 |
| 300/560 | A2             | A3           | 31796000     | 22734186               | 21832474               | 21664242               | 21813114               | 20156156                 | 16625853                    |
| 300/560 | A2             | A4           | 31796000     | 4554414                | 4345890                | 4324493                | 4393961                | 4067559                  | 3274550                     |
| 300/560 | A1             | A2           | 31796000     | 27590144               | 26478718               | 26285708               | 26505358               | 24521646                 | 20189192                    |
| 100/250 | A74            | A76          | 1975000      | 295375.125             | 282689.5625            | 279438.7813            | 286997.3438            | 253470.0625              | 209011.8438                 |
| 100/250 | A61            | A63          | 1975000      | 88464.51563            | 85114.875              | 84859.00781            | 76601.00781            | 73757.82813              | 54910.10156                 |
| 100/250 | A60            | A62          | 1975000      | 236901.6406            | 231142.6094            | 229367.8125            | 222078.375             | 212464.5313              | 162690.9219                 |
| 100/250 | A57            | A59          | 1975000      | 136632.625             | 130829.3047            | 126993.1484            | 123553.2031            | 117995.9375              | 94473.34375                 |
| 100/250 | A65            | A68          | 1975000      | 238563.2344            | 234656.6406            | 230721.7969            | 231866.875             | 201544.6563              | 178943.2813                 |
| 100/250 | A56            | A58          | 1975000      | 249133.7031            | 235580.1094            | 229266.8906            | 231753.2813            | 240738.2188              | 155197.8125                 |
| 100/250 | A51            | A53          | 1975000      | 694633.8125            | 671495                 | 667443.125             | 659671.875             | 572588.9375              | 482148.6875                 |
| 100/250 | A13            | A14          | 1975000      | 896991.0625            | 857712.4375            | 848555.125             | 858402.5               | 756693                   | 627657.6875                 |
| 100/250 | A10            | A12          | 1975000      | 60958.17969            | 52050                  | 50092.96094            | 59665.80078            | 60958.17969              | 22981.67969                 |
| 100/250 | A29            | A30          | 1975000      | 476485.75              | 429385.4063            | 414390.4063            | 469932.4063            | 476485.75                | 314416.2813                 |
| 125/280 | A69            | A71          | 3398000      | 365174.4375            | 339964.5938            | 333737.5625            | 340897.3438            | 332414.9375              | 238555.8125                 |
| 125/280 | A66            | A70          | 3398000      | 552260.0625            | 534330.375             | 542336.25              | 540714.6875            | 519174.2188              | 498252.2188                 |
| 125/280 | A55            | A64          | 3398000      | 58699.8125             | 53941.64844            | 53125.90234            | 53535.01563            | 50616.66406              | 32166.97266                 |
| 125/280 | A44            | A45          | 3398000      | 355216.5               | 333628.4063            | 329020.2813            | 338242.9063            | 305835.9375              | 232572.2969                 |
| 125/280 | A38            | A40          | 3398000      | 207690.8281            | 197475.0156            | 195093.8594            | 196905.9844            | 176971.75                | 145393.9219                 |
| 125/280 | A5             | A7           | 3398000      | 907691.125             | 894997.9375            | 888660.875             | 894870.25              | 849342                   | 755247.8125                 |
| 125/280 | A32            | A34          | 3398000      | 771653.25              | 749398.75              | 748965.75              | 744702.3125            | 665742.1875              | 591025.6875                 |
| 125/280 | A22            | A33          | 3398000      | 209017.8281            | 209820.2031            | 206128.0156            | 196956.1563            | 203122.0781              | 189496.1719                 |
| 150/315 | A48            | A49          | 6600000      | 269134.3125            | 259295.8906            | 252029.3125            | 257932.125             | 238192.9219              | 185649.1406                 |
| 150/315 | A60            | A61          | 6600000      | 88464.52344            | 85114.88281            | 84859.00781            | 76601.00781            | 73757.82813              | 54910.10156                 |
| 150/315 | A48            | A50          | 6600000      | 301270.3438            | 297929.1875            | 294448.5938            | 295883.5313            | 264335.2188              | 213332.1563                 |
| 150/315 | A55            | A65          | 6600000      | 507850.9375            | 484311.0938            | 481570.5               | 486696.4063            | 436489.75                | 339653.5625                 |
| 150/315 | A57            | A60          | 6600000      | 781076.5               | 746324                 | 732765.1875            | 722636.9375            | 693160.5625              | 512271.75                   |
| 150/315 | A56            | A57          | 6600000      | 1175935.25             | 1124909.375            | 1102937.625            | 1081852.125            | 1047230.5                | 782296.9375                 |
| 150/315 | A17            | A19          | 6600000      | 127528.1094            | 124142.8672            | 121477.1563            | 131932.2813            | 125973.0781              | 83518.55469                 |
| 150/315 | A15            | A16          | 6600000      | 296840.5625            | 269350.7813            | 266156.0313            | 254961.1875            | 251941.3594              | 156304.375                  |
| 150/315 | A46            | A48          | 6600000      | 1203022.625            | 1175764.75             | 1147148                | 1132734.25             | 1079454.25               | 783094.5                    |
| 150/315 | A46            | A47          | 6600000      | 301379.1875            | 288042.75              | 283452.5313            | 286663.9688            | 255209.8594              | 199594.3906                 |
| 150/315 | A51            | A54          | 6600000      | 204170.5781            | 204991.5469            | 184030.6875            | 194432.6875            | 198687.625               | 120207.3906                 |
| 150/315 | A37            | A44          | 6600000      | 2120867                | 2052561.5              | 2016244.5              | 2014782.25             | 1889042.75               | 1408309.75                  |

|         |     |     |          |             |             |             |             |             |             |
|---------|-----|-----|----------|-------------|-------------|-------------|-------------|-------------|-------------|
| 150/315 | A8  | A11 | 6600000  | 541888.5625 | 503093.5313 | 520577.4375 | 538533.3125 | 515634.0313 | 410586.0938 |
| 150/315 | A13 | A15 | 6600000  | 698408.375  | 648304.625  | 632334.625  | 634467.125  | 648121.875  | 407215.0625 |
| 150/315 | A6  | A9  | 6600000  | 195646.9219 | 187989.5781 | 188861.4219 | 189318.7813 | 166907.2031 | 147163.5313 |
| 150/315 | A6  | A8  | 6600000  | 3151960.5   | 2978317.5   | 2961482.75  | 3021746     | 2800969     | 2164174.25  |
| 150/315 | A5  | A6  | 6600000  | 3484348.5   | 3295296.5   | 3279821.75  | 3342279     | 3080108.25  | 2401200.5   |
| 150/315 | A41 | A42 | 6600000  | 449802.5625 | 432806.125  | 429521.2188 | 430077.1875 | 389733.75   | 322800.875  |
| 150/315 | A23 | A24 | 6600000  | 456539.8125 | 438199.8125 | 430663.625  | 438908.375  | 408939.4063 | 322627.4688 |
| 150/315 | A25 | A26 | 6600000  | 1485469     | 1485256.75  | 1484060.875 | 1459879.5   | 1313780.625 | 1288258     |
| 150/315 | A29 | A31 | 6600000  | 285347.3438 | 274755.4688 | 273476.5313 | 279612.7188 | 250311.2031 | 205693.1719 |
| 200/400 | A72 | A74 | 11146000 | 1270610.25  | 1221151.875 | 1220222     | 1234925.5   | 1100369.375 | 956448.3125 |
| 200/400 | A72 | A73 | 11146000 | 42374.80859 | 44796.94922 | 43591.52344 | 43751.26172 | 42374.80859 | 38635.67188 |
| 200/400 | A69 | A72 | 11146000 | 1435744     | 1378223.375 | 1378700.625 | 1391503.875 | 1247296.25  | 1066069     |
| 200/400 | A52 | A56 | 11146000 | 2957786.75  | 2829629.25  | 2765537.5   | 2793130     | 2700499     | 2079282.5   |
| 200/400 | A17 | A18 | 11146000 | 0           | 0           | 0           | 0           | 0           | 0           |
| 200/400 | A4  | A5  | 11146000 | 4554413.5   | 4345890     | 4324493     | 4393961     | 4067559.25  | 3274549.5   |
| 250/500 | A52 | A55 | 20224000 | 5223174     | 4991982     | 4990023     | 4971909.5   | 4623860     | 3782894.25  |
| 250/500 | A43 | A52 | 20224000 | 9487805     | 9005028     | 8934846     | 8992995     | 8456270     | 6736617.5   |
| 250/500 | A36 | A37 | 20224000 | 14686924    | 14050802    | 13896990    | 13994191    | 13058425    | 10408417    |
| 250/500 | A39 | A41 | 20224000 | 784628.6875 | 744021.3125 | 738564.5    | 752906.3125 | 679586.4375 | 538481.6875 |
| 250/500 | A23 | A25 | 20224000 | 3259388     | 3178663.5   | 3148087.5   | 3189482.75  | 2906566.75  | 2590958.75  |
| 250/500 | A27 | A28 | 20224000 | 0           | 0           | 0           | 0           | 0           | 0           |
| 250/500 | A27 | A29 | 20224000 | 847468.375  | 789776.1875 | 773502.1875 | 833118      | 798577.3125 | 589979.125  |
| 150/315 | A8  | A10 | 6600000  | 2443945     | 2316918     | 2282577.5   | 2324826     | 2148694.25  | 1638306.875 |
| 150/315 | A10 | A13 | 6600000  | 2052599.625 | 1950226.875 | 1920557.875 | 1940210.25  | 1804537.375 | 1383127.125 |
| 150/315 | A15 | A17 | 6600000  | 401567.7813 | 378953.875  | 366178.5938 | 379505.9063 | 396180.5313 | 250910.6875 |
| 150/315 | A65 | A67 | 6600000  | 155620.6406 | 142489.9531 | 145347.6406 | 151389.8438 | 140078.2813 | 94916.67969 |
| 200/400 | A55 | A66 | 11146000 | 4166336.75  | 3979936.25  | 3986792.5   | 3988433.25  | 3710025     | 3096111.25  |
| 200/400 | A66 | A69 | 11146000 | 2791018     | 2660436.5   | 2656142.5   | 2670609.5   | 2438535.5   | 1986437.25  |
| 200/400 | A74 | A75 | 11146000 | 216907.6563 | 204659.4375 | 206017.6875 | 207800.2344 | 187891.0469 | 154906.3438 |
| 150/315 | A43 | A51 | 6600000  | 1412720.75  | 1370518.75  | 1339140.125 | 1355446     | 1216321.25  | 968999.3125 |
| 250/500 | A37 | A43 | 20224000 | 11583892    | 11028679    | 10918755    | 11013884    | 10297801    | 8217429.5   |
| 150/315 | A44 | A46 | 6600000  | 1615851.5   | 1578987.75  | 1543181.125 | 1532237.75  | 1446113.125 | 1068736     |
| 250/500 | A35 | A36 | 20224000 | 16135364    | 15436001    | 15272199    | 15381835    | 14297268    | 11430124    |
| 300/560 | A21 | A22 | 31796000 | 17495440    | 16772908    | 16603920    | 16690138    | 15477736    | 12504386    |
| 250/500 | A25 | A27 | 20224000 | 1013142.188 | 955185.8125 | 937756.5625 | 993737.1875 | 939401.6875 | 724364.3125 |
| 250/500 | A21 | A23 | 20224000 | 4387922     | 4269432     | 4217854.5   | 4278476     | 3900796.75  | 3392617.25  |
| 300/560 | A22 | A32 | 31796000 | 17286422    | 16563088    | 16397792    | 16493182    | 15274614    | 12314890    |
| 300/560 | A38 | A39 | 31796000 | 860976.3125 | 820348.5    | 811947.375  | 826260      | 742945.5625 | 596137.4375 |