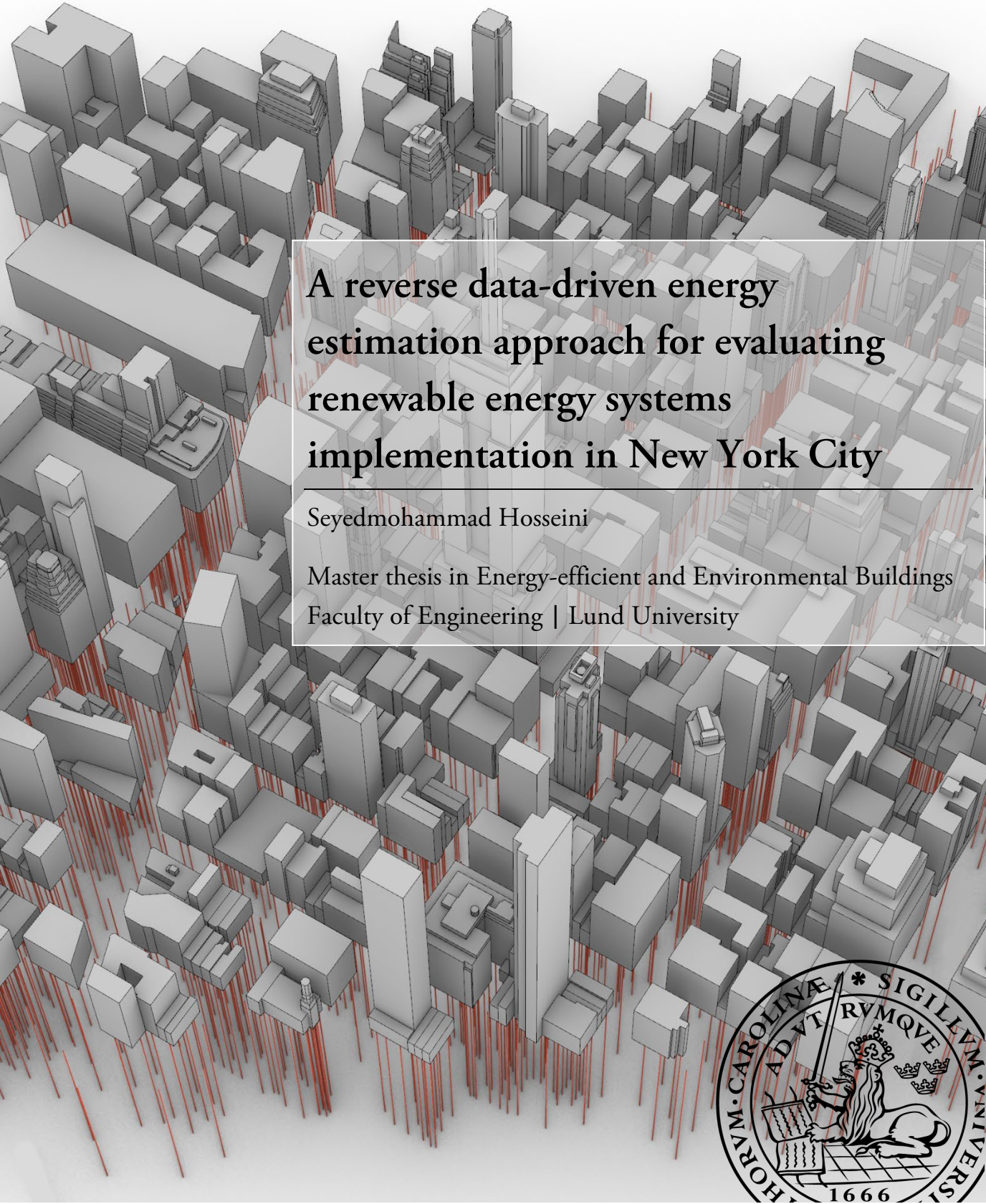


A reverse data-driven energy estimation approach for evaluating renewable energy systems implementation in New York City

Seyedmohammad Hosseini

Master thesis in Energy-efficient and Environmental Buildings
Faculty of Engineering | Lund University

Lund University

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

Master Programme in Energy-efficient and Environmental Building Design

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

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

Examiner: Dennis Johansson (Building Services)

Supervisor: Saqib Javed (Building Services), Maria Josefsson (LKAB Wassara)

Keywords: Energy use prediction, Energy performance simulation, Renewable energy source, Statistical methods, Reverse method, Regression model, Urban modelling, Energy modelling, Generative modelling, Geothermal heat pump, BTES.

Publication Year: 2021

Abstract

Urban areas are responsible for a significant part of greenhouse gas emissions. Megacities, especially, have cumulative impacts due to different anthropogenic heat sources and urban heat island phenomena. It is therefore vital to replace fossil fuel burning heating sources with renewable sources in the current building stock. Renewable energy sources can be applied through building energy retrofits, which need to be technically and economically feasible. It is, however, always a complex matter to have a fast and reasonably accurate energy use prediction for the retrofit feasibility study. New York City was chosen for this research due to the wide variety of publicly accessible data, including high-quality metered energy and high-resolution geomatics data. This study proposes a reverse method to predict hourly building energy performance according to the total metered energy use. Buildings were divided into different categories based on type and vintage. A robust relation between building categories and hourly energy use was found. Hourly energy trend was defined to describe the energy use of each category. The output was separated into cooling- and heating loads to be used in a geothermal system design program. The geothermal system was considered as an appropriate system to be established in the existing buildings. The results showed that 50% to 100% of the heating loads of the buildings could be covered by geothermal as an alternative to fossil fuels. The introduced reverse method can significantly reduce the time of energy estimation compared to the simulation-based methods. Hence, providing a retrofit proposal to implement renewable energy systems such as geothermal systems would be faster, easier, and affordable for practitioners.

Table of contents

List of figures

List of tables

Abbreviations

AMY	Actual Meteorological Year
BBL	Borough Block Lot
BIN	Building Identification Number
BTES	Borehole Thermal Energy Storage
CV (RMSE)	Coefficient of Variation of the Root Mean Square Error
DHW	Domestic Hot Water
EPS	Energy Performance Simulation
EPW	EnergyPlus Weather
EUI	Energy Use Intensity
GCM	General Circulation Model
GHG	Greenhouse Gas
GIS	Geomatics Information System
GSHP	Ground Source Heat Pump
HET	Hourly Energy Trend
HVAC	Heating Ventilation and Air Conditioning
LoD	Level of Detail
NMBE	Normalized Mean Bias Error
NYC	New York City
RC	Relative Compactness
SDGs	Sustainable Development Goals
SH	Space Heating
SPF	Seasonal Performance Factor
TMY	Typical Meteorological Year
UHI	Urban Heat Island
UWG	Urban weather Generator

1 Introduction

More than half of the world population live in urban areas. This number is expected to rise to more than two thirds by 2050 [1]. Urban areas are the primary source of greenhouse gas (GHG) emissions in different sectors [2]. Urban areas are responsible for more than 70% of CO_2 emissions [3]. Buildings are responsible for more than 40% of the total energy use, of which around 60% corresponds to residential buildings [4, 5]. Various local and international policies have been implemented to control and reduce the impacts of urbanization on the environment. United Nations Sustainable Development Goals (SDGs) have emphasized the importance of urban and energy-related challenges in Target 7: Affordable and clean energy, Target 11: Sustainable cities and communities, and Target 13: Climate action [6].

More than 7% of the entire world population live in megacities [7]. By definition, cities with more than ten million people are counted as megacities [8]. Megacities are influenced strongly by Urban Heat Island (UHI) phenomena, where the temperature in the urban areas is considerably higher than the surrounding rural areas [9]. Several studies have been performed on UHI impacts in the past few decades; however, the acceleration in urbanization has caused complexities in understanding the real effects of various urban elements. Anthropogenic modifications in the natural landscape and heat production from human activities have been shown to be the influential parameters for UHI effects [10]. Rising temperatures, a significant reduction in wind speed, changes in relative humidity, and thermal comfort disturbance are parts of UHI consequences [11]. Moreover, studies have shown that the building sector's energy demands can double due to the high cooling loads [12]. The effects of UHI cannot be mitigated by spreading the urban area to a larger land area with lower urban elements density because it is strongly related to the City's land area rather than urban elements density. Thus, UHI consequences can be observed, not only in downtowns but also in low-density urban areas [13].

1.1 Background

New York City (NYC) is one of the top 20 megacities, with an approximate population of 20 million in its metropolitan area. NYC is significantly involved with buildings' environmental impacts due to the enormous number of constructions in the 1920s and 1930s before introducing strict building standards and environmental regulations. Also, over 70% of GHG emissions in NYC are generated from the building sector [14]. Meanwhile, more than 80% of the properties use steam boilers for heating purposes, whereas 70% of the buildings exploit natural gas-driven local boilers, and 10% are connected to district steam system. At the same time, the share of electricity in supplied heat is not significant [15]. NYC has committed itself to reduce its total GHG emissions by 80%, relative to 2005 levels, by 2050. This plan is supported by Local Law 97 [16]. This law includes all buildings with an area of larger than 2300 m^2 (25,000 ft^2). In addition, the owners of the large buildings are obliged to submit their measured annual energy and water use to the city database in a benchmarking process [17]. These data are publicly disclosed from the year 2013 and later to be used openly [18].

According to Clean Energy Standard (CES) [19], achieving the ambitious energy goal requires applying different renewable energy sources [20]. NYC energy goal is a novel approach because NYC aims at existing buildings. However, similar laws that have passed worldwide are applied primarily for new constructions [21]. Considering the effects of the urban area on most renewable sources, such as sun blockage, wind obstruction, shortage of available space, and visual disturbance, prevent them from being used. Hence, the role of heat pumps and geothermal as an alternative for heating systems becomes significant. Providing the required heating demand in buildings through heat pump technology is a possible measure in reducing fossil fuel consumption and their associated GHG emissions [24, 25].

The geothermal energy source applications are among the viable solutions to reach NYC's environmental goals. Establishing a geothermal heat pump system in a building must be technically feasible and economically viable. Thus, having a relatively accurate prediction of building performance is necessary. Various uncertainties in energy use estimation influence the accuracy of the predictions and make it hard for clients to accept the feasibility of energy retrofit proposals. Generally speaking, retrofitting buildings for energy improvement is one of the most prevalent topics in the field of energy use in buildings. Several computational, statistical, and empirical methods have been developed to estimate buildings' energy performance after refurbishment, especially hourly load profile and peak powers. It is always an intricacy in feasibility studies of building retrofit to estimate energy use and potential cost savings. The energy bills after the retrofitting process are most likely different. Therefore, studies have tried to find a fast and acceptable estimation method for hourly energy demand.

Building energy performance simulation (EPS) involves various kinds of uncertainties in inputs, as well as the time of the simulation procedure. Energy simulation, however, relies not only on building performance but also on urban morphology [26-28], user behavior [29-31] and operation strategies. Therefore, a comprehensive perspective is required in different disciplines, including building typology, urban morphology, human, society, and policies, to mitigate the results' discrepancy from the actual condition. Moreover, climate change and global warming cause complexity in predicting future weather conditions. Consequently, buildings' energy performance has ambiguity in its lifetime.

1.2 Aims and Goals

This study aims to develop a customized data-driven method to estimate annual hourly heating- and cooling loads for the NYC building stock, avoiding heavy energy simulation methods. The main hypothesis is to find a robust relation between different building categories and buildings' hourly energy use profiles. The hourly load profiles are achieved through EPS to investigate their characteristic corresponding to the category. Meanwhile, the EPS' uncertainties are partly eliminated by the calibration with annual measured energy use, taking the opportunity of using publicly available data. In this research, high-resolution modellings were performed using GIS data to run EPS. The EPS' hourly results are calibrated with the measured data to achieve accurate hourly load profiles.

The approach used in this thesis is based on statistical methods performing analysis on the publicly available data from the NYC building stock and EPS results. The statistical analysis results in hourly thermal loads to assess the potential of applying the geothermal system in the existing buildings.

1.3 Thesis disposition

The following chapter reviews different articles and previous studies relevant to the study purpose. The third chapter explains the methods and tools applied in this research, including modelling, statistics, software, and delimitations. The fourth chapter presents the results of the proposed method and their discussions. The last chapter concludes the research.

2 Literature review

This chapter firstly provides a review of scholarly articles and previous studies focusing on estimating building energy performance. Then, energy simulation programs and reverse approaches for building energy estimation are presented. Thereafter, weather and climate data, UHI, geothermal energy, and NYC-related articles were reviewed.

2.1 Energy performance prediction

Estimating the energy use of buildings is one of the pivotal parts of the energy study of constructions. Regardless of the applied method and the purpose of simulation, all kinds of EPSs consist of three main components:

(1) Physical parameters refer to the physical properties of materials. They are mostly identifiable as the standard input parameters in simulation and are not influenced by designers. Examples of physical parameters include thickness, density, thermal conductivity and heat capacity of walls.

(2) Design parameters refer to the building's working conditions, which are determined entirely by the decision-makers/designers; for example, ventilation rate, heat dissipation rate by occupants, facilities, and light.

(3) Scenario parameters refer to parameters relative to the building's operation during its lifetime. These parameters are not measurable and hard to control [29]. Weather conditions, including ambient temperature, relative humidity, solar radiation, and casual heat gain, are examples of scenario parameters. Likewise, occupants' density, computers and light in use, internal and external shading coefficients, internal and external convection transfer rate, as well as infiltration through the envelope and openings, are some other examples of scenario parameters [33, 34].

One of the most recently adopted and comprehensive classification of building energy estimation introduces four approaches: (1) engineering methods, (2) statistical methods, (3) neural networks, and (4) support vector machine. These are discussed in the following.

2.1.1 Engineering methods

The calculation of thermodynamic behavior of buildings or buildings' components based on physical principles is considered as engineering methods. They have been developed extensively during the past decades [32]. In general, they can be divided into two categories: (1) comprehensive and (2) simplified methods [33]. In the comprehensive methods, known as computational techniques, all components, including outdoor and indoor condition, user behavior, equipment loads, and detailed construction, are involved. This method, however, needs more time, computation power, and a high level of expertise. The method is adequate for HVAC design or energy analysis [33]. The ISO has defined a standard for energy demand calculation for buildings [34]. A large number of programs have been developed to calculate

the detailed energy behavior of the buildings. They use different calculation methods and data library sources and provide comprehensive details [35].

Computational building EPS is a multidisciplinary, broad range method with temporal continuity and dynamic spatial features, considering the complexity of the actual situation to achieve more accurate results [36]. The accuracy of EPS relies upon model precision, building's specifications, and weather data [39, 40]. Computational methods have two crucial issues since they need the details of the building and its environment. Firstly, detailed data are not available for all buildings and places, such as envelope characteristics, room functions, user behavior, schedules, etc. Secondly, operating the software and handling so many inputs requires a high level of expertise and software skill [39].

2.1.2 Statistical methods

Prediction of the building energy use based on the measured data relies on a regression analysis of the measured and predicted values [40]. Regression analysis correlates energy use to the influencing parameters on the energy performance [39]. It is crucial to collect enough historical data to fill the statistical database or make a training dataset. In statistical methods, data accuracy, availability, and reliability are prominent [41]. Commonly, three kinds of data are available as historical data in the building sector. The first kind is real data, which includes data collected through intelligent energy meters, building management systems, and weather stations.

Moreover, utility bills, energy surveys, and energy statistics also provide real data [42]. The second type is simulated data from an energy simulation software based on modelling of an existing or non-existing building [43]. Computational simulation tools intrinsically have deviations from the prototype building and cannot represent the actual condition [43]. Simulated data, considering the limitations, are beneficial when the real data is unavailable or come with shortages. The third data type is publicly available benchmark datasets available for some urban areas and building types. This data type can be used as an inclusive source for statistical analysis and calibration [50, 51]. Apart from directly predicting energy use, several statistical methods have been developed to predict or define a range of construction specifications such as U-value, G-value, internal gains, and schedules [29, 52, 53].

Energy modelling can be calibrated and verified with the measured values. ASHRAE Guideline 14 [48], FEMP [49], and IPMVP [50] has introduced modelling and verification guideline. Mostly, all these criteria evaluate the hourly or monthly data, using Normalized Mean Bias Error (NMBE), Coefficient of Variation of the Root Mean Square Error (CV(RMSE)), and the coefficient of determination (R^2). These verification indices define the deviation of the simulated data from the measured values regression line and can be scale-dependent or scale-independent. When an index is scale-dependent, it has a unit similar to the assessed data like RMSE. Thus, it gives an understanding of the scale of discrepancy. In

contrast, scale-independent indices are usually normalized by a component of the data, such as mean or sum. Hence they are unitless like CV(RMSE) and NMBE [51].

Mean Bias Error (MBE) is the average of the errors of the sample data. It presents the relation between simulated data and the regression line of the measured data [52]. MBE can be positive or negative, where positive values show under-predicted measured data and vice versa. Nevertheless, this index has a fundamental issue related to error cancellation since positive and negative values eliminate each other in the equation [51]. NMBE is the normalization of MBE to make it scale-independent (unitless). Equation (1) and Equation (2) present the corresponding formulas to calculate these values, where *m* is measured data, *s* is simulated data, *n* is the number of values, *i* is the index in the time series, and $\overline{m}$ is the mean value of the measured data [59, 60].

$$MBE = \frac{\sum(m_i - s_i)}{n} \quad (1)$$

$$NMBE = \frac{1}{\overline{m}} \cdot \frac{\sum(m_i - s_i)}{n} \times 100 \quad (2)$$

RMSE indicates how simulated data can describe the shape of the measured data. It defines the deviations between simulation and measurement. CV(RMSE) is calculated by dividing RMSE by the mean value of the dataset (hourly or monthly energy profile) to eliminate the effect of scale. CV(RMSE) is commonly used among professionals and used in different standards [54, 59]. These two indices can be calculated from equation (3) and equation (4), respectively.

$$RMSE = \sqrt{\frac{\sum(m_i - s_i)^2}{n}} \quad (3)$$

$$CV(RMSE) = \frac{1}{\overline{m}} \cdot \sqrt{\frac{\sum(m_i - s_i)^2}{n}} \times 100 \quad (4)$$

The other commonly used index for practitioners is R-square (R^2). It describes how close is the simulation values to the regression line of the measured values. The higher index value means the closer simulated data to the regression line of the measured data. R^2 can be calculated through equation (5), and the result is in the range of 0 and 1 [55, 59], where a value close to one is desired.

$$R^2 = 1 - \frac{\sum(s_i - m_i)^2}{\sum(m_i - \overline{m})^2} \quad (5)$$

2.1.3 Neural networks

One of the most widely used artificial intelligence models in building energy prediction is the artificial neural network [55]. This method has been developed during the past 30 years. Neural network algorithms can be applied to predict the building energy performance or other parameters like internal loads, predicted mean vote, and electricity use [35]. Artificial neural networks help to solve non-linear problems [56]. Several studies have shown the better performance of neural networks than engineering methods in predicting electricity, lighting, user behavior [63].

2.1.4 Support vector machines

Applications of Support vector machines (SVM) are growing fast and becoming common in the field of building energy prediction. This method has been widely applied to solve regression problems to estimate the underlying relationship between the non-linear inputs to the continuous real-valued target [40]. The SVM used for regression is called support vector regression (SVR), which has become an important data-driven approach for predicting building energy consumption [40].

2.2 Energy simulation software

During the past decades, hundreds of building energy simulation programs have been developed. Each program provides particular capabilities based on the purpose of use. Different methods are applied in the programs, highly influenced by advanced computer technology and data availability [35]. The choice of an appropriate energy simulation software is based on the requirements of a particular purpose, the available expertise, and proper hardware [58].

The main principle of energy simulation programs is a simulation engine with assigned inputs related to building geometry, weather conditions, HVAC systems, internal loads, operating strategy, and simulation specific parameters [32]. All these inputs are inserted in a graphical user interface with different levels of capabilities. Building geometry constitutes the basic input for energy simulation. The geometry is different from the architecturally designed model and should be adopted to the simulation requirements. Internal and external loads should be considered to provide sufficient information for heat balance in a space. External loads are mainly calculated through weather data which is discussed later in this study. Internal loads and operating schedules are generally assumed based on the building type and use [32]. HVAC systems and their components are a major part of the input parameters for EPS. These systems can reflect the actual system if the energy simulation tool (graphical user interface and the engine) provides enough flexibility. Last but not least, all energy simulation tools

need specific parameters, for example, numeric convergence tolerances for the underlying mathematical model of the simulation engine. These parameters influence the numerical behavior of the simulation engine.

2.3 Reverse energy estimation

The essential demand to have a faster and accurate energy estimation have motivated practitioners to develop energy estimation methods based on data availability. ASHRAE Handbook of fundamentals [59] defines two modelling methods: forward (classic) and inverse (data-driven). In the forward approach, the workflow starts from the modelling and assigning the physical parameters, operation schedules, etc. Then it estimates the total energy use in the desired time increment. The inverse method uses measured or monitored energy use and other possible metrics as input and tries to fit them into a baseline time series [54]. Inverse models are used commonly in building retrofits [60].

Two general methods are suggested by ASHRAE Guideline 14 [48] to generate baselines. The first one is calibrated data-driven regression analysis, and the other is calibrated computer simulation tools [61]. Currently, regression techniques are the most popular inverse energy estimation approach. This method assumes that the nonlinear energy performance of the building can be appropriately captured by the regression model [54]. Calibrated simulations can provide more customized profiles with the target building stock. However, they are time-consuming and need detailed simulation inputs. Choosing between the proper methods depends on the purpose of the study and the availability of data [72, 73].

2.4 Weather

As the most important boundary condition for the building, climate conditions significantly affect EPS. It is also a source of uncertainties in the simulation due to the complexity of the forecasting. Some researchers have investigated the sensitivity of the energy simulation results to the quality of weather data. Achieving a more reliable energy model is correlated to choosing the most appropriate weather dataset [74, 75]. The effects of weather dataset accuracy have been investigated in several different studies [38]. Accuracy of weather dataset has become important since extreme climate events and different climate patterns are becoming more common due to climate change [76, 77]. Several methods, such as stochastic methods, have been developed to provide better predictions using available datasets and measurements [78, 79].

Outdoor climate condition is applied in the EPS as weather data. Weather datasets can be generated from downscaled General Circulation Models (GCMs). GCMs are computer-driven models for forecasting weather, understanding climate, and projecting climate change [70]. A GCM aims to describe the climate behavior by integrating various fluid-dynamic, chemical, or biological equations derived directly from physical laws (e.g. Newton's law) or empirical generated methods [70].

Weather data files describe a mesoclimate condition. Typical meteorological year (TMY) is one of the most commonly used weather data for energy simulations. TMY is a meteorological dataset with climatic parameters values for every hour in a year for a given geographical location. This data is captured from hourly data for a more extended period (usually ten years or more). For each month in the year, the data is selected from the year that is considered the most typical for that month [71]. Energy plus weather-file (EPW) is an adopted TMY file for EnergyPlus simulation software [72]. Other sources for weather data, such as International Weather for Energy Calculations (IWEC) and Weather Year for Energy Calculation 2 (WYEC2), can be used for energy simulations in the EPW-file structure. [72].

Energy performance is significantly influenced by surrounding objects and elements, especially in urban area. Considering the effects of the urban context, surrounding buildings, and obstacles can enhance the accuracy of EPS results. However, capturing in totality could be difficult due to several involved parameters such as material, area, size, color, and height of adjacent buildings and roads. These parameters have different impacts on sky view, solar gain, solar reflectivity, and wind speed [83, 84]. One possible solution is to use a microclimate weather file. Urban Weather Generator (UWG) is a method to make microclimate weather files for a certain urban area. UWG is based on energy conservation principles and takes input parameters that describe urban morphology, geometry, and surface materials. UWG focuses more on the heat balance equation and considers different sources of heat in the urban area based on the buildings type, vintage, and floor area [75].

UWG, among all effective parameters on the microclimate scale, focuses on site coverage, façade to site ratio, and vegetation area. Site coverage is the ratio of the covered area by buildings to the total land area, which explains the density of the urban area in the desired zone. Façade to site ratio, calculated by the façade area divided by total land area, explains the ratio of climate exposed buildings' surfaces to the site area. Finally, vegetation area is the area of the land covered by vegetation and plants in the target area [73].

2.5 Urban heat island

Urban regions usually have a higher temperature than the surrounding area due to the anthropogenic heat released from vehicles, power plants, air conditioners and other heat sources. Furthermore, huge quantities of solar radiations are mainly stored and re-radiated in urban areas due to massive construction materials [76]. This phenomenon, called Urban Heat Island (UHI), has been aggravated in connection with urbanization, industrialization, and population growth [77]. The intensity of the UHI effect can be observed by local monitoring, satellite images, and numerical model simulations [78]. However, all these methods encounter several limitations. For instance, the chosen spot for monitoring should be representative of the entire area. Moreover, monitoring points should be distributed evenly in the target zone [79]. Satellite imagery has limitations for spatial and temporal resolution and complexity due to the high technology of the required equipment. Numerical model simulations depend on

the urban morphology and regional climate conditions, which need geographical and climate data, respectively [80].

2.6 Geothermal energy

Geothermal energy, among other renewable sources, is predictable and reliable. Unlike fossil energy sources, geothermal energy is clean, sustainable, and recyclable [81]. The ground can be used as a heat sink or a heat source to provide cooling and heating. Moreover, compared to other renewable energy sources, geothermal needs less area to be established [93]. Geothermal energy is often exploited by ground source heat pumps [94].

Underground layers have a more stable temperature than the ambient. Thus, the ground temperature is lower than ambient air during cooling seasons and higher than ambient air during warming season [83]. In order to use geothermal energy more efficiently and sustainably, the ground can be charged by heat energy through a BTES. Most rocks have a relatively high volumetric heat capacity, and when it comes to underground water, it is even higher. The ground can be used as thermal storage or accumulator [84]. The application of geothermal energy in the building sector can include both the heating and cooling process. Removing heat from buildings, known as the cooling process, always involves rejecting the excess heat from the interior spaces. The cooling process happens in the warm seasons when the ground temperature is lower than the ambient temperature. In cooling mode, the ground works as a heat sink and absorbs the heat rejected from borehole heat exchangers [95]. In heating mode, the stored heat in the ground is extracted and used for heating purposes [96].

Common geothermal systems use vertical boreholes [85]. Designing the ground source heat pump relies upon the fluid temperature exiting from the borehole system. This temperature is influenced by the borehole system's thermal response in the short-term and surrounding ground's response in the long term, as well. Moreover, in multiple borehole fields, the exiting fluid temperature is affected by thermal interaction between boreholes [86]. Changes in the surrounding ground thermal response happen slowly due to the large thermal mass and thermal capacity. Consequently, the ground response is often studied in monthly or annual periods [94].

The calculation of the thermal behavior of the ground and boreholes is a complicated process. To make it possible to calculate the ground thermal response, it is necessary to define the critical parameters and neglect the rest. For instance, the ground surface temperature is influenced by the ambient temperature, and it varies day to day and year to year. Hence, in practice, the average temperature is used instead. One of the essential factors influencing the ground thermal behavior is the ground thermal conductivity. Borehole thermal resistance is another key performance characteristic of a borehole. It is thermal resistance between the fluid in the tube and the borehole wall. Lower borehole thermal resistance leads to better system performance. Then the injected heat into the ground (or extracted heat as negative values) defines the temperature in the short- and long-term [85].

The unitless step-response function "*g*", known as *g*-function, is defined to characterize the thermal response of boreholes. *g*-function relates the borehole's wall temperature and the injected heat (or extracted heat) rate to the ground thermal conductivity. Accordingly, a set of temperature response factors for different borehole arrangements can be generated. A superposition technique for heat transfer can be a beneficial approach. In this technique, the thermal response of a system is obtained by superimposing the thermal response of every phenomenon at a certain point of interest [88]. The superposition method is used in the computer software called EED to design the borehole system based on the hourly data of the heating and cooling loads of the building [100] and the corresponding g-function [101].

2.7 New York City

NYC is situated on the Atlantic Coastal Plain in the northeastern United State with the coordinate of 40.71 N and 74.00 W. The northeast region is the most economically developed, culturally diverse, and densely populated in the entire country [91]. The region accounts for 20% of the total GDP of the USA [91].

The state of New York is bounded to the north and west by Lake Erie, the Canadian provinces of Ontario and Quebec and Lake Ontario; to the east by Vermont, Massachusetts, and Connecticut; to the southeast by the Atlantic Ocean and New Jersey; and to the south by Pennsylvania. Although New York State has been known as NYC, the entire state has a broad climate and geographical diversity. The most crucial element in the state is water. Nine major rivers, more than 8000 lakes, and two shorelines of 210 km by the Atlantic and 600 km by Lake Erie and Ontario play a significant role in its climate condition. Also, Niagara Falls is a huge source of hydroelectric power in the area [92].

NYC is built mainly on the three islands of Long Island, Manhattan, and Staten Island. Additionally, the City is divided into five boroughs: Brooklyn, Queens, Manhattan, the Bronx, and Staten Island, presented in Figure 1. The Bronx is the only contiguous to upstate New York [93].

Figure 1: NYC five boroughs Adapted from https://en.wikipedia.org/wiki/Boroughs_of_New_York_City# /media/File:5_Boroughs_Labels_New_York_City_Map.svg

The geology of New York City is greatly varied, composed of areas directly underlain by bedrock with shallow overburden of unconsolidated material; for example, Manhattan and the Bronx. Other boroughs with hundreds of feet of unconsolidated material consist of multiple layers of unconfined and semi-confined aquifer systems such as Queens and Brooklyn [94]. The geology beneath the studied project in [94] is a bedrock geologic unit known as Manhattan Schist.

Geothermal ground source systems have been used in New York City and the greater New York metropolitan area since the 1940s. Early systems mainly were systems located on Long Island (including Brooklyn and Queens), which used groundwater as a source and sink, supplying water to buildings from supply wells and returning water through separate return wells located at a distance hydraulically downgradient [94].

The required parameters to design a geothermal system in the studied project in [94] are presented in the following points.

- Soil Type: Average Rock
- Thermal conductivity of the ground: 3.115 W/m·K (1.8 Btu/hr·ft·°F)
- Volumetric heat capacity of Ground: 2.343 MJ/m^3·K (34.943 Btu/°F·ft^3)
- Undisturbed ground temperature: 15.5 ℃ (60 °F)
- Borehole thermal resistance: 0.0303 m·K/W (0.0524 hr·ft·°F/Btu)

3 Method

This chapter will explain the methods applied in this research under four main headings. The first section presents the statistical methods and software used for the analysis. The second chapter explains EPS workflow. The third section presents the reverse energy estimation method identifying the details of the proposed approach. The fourth part explains the geothermal system design. Finally, delimitations of this research are introduced. The following computer programs (Figure 2) were utilized to perform this study. Rhino and Grasshopper were used mainly for the building modelling, whereas HoneyBee and LadyBug in Grasshopper were used for energy modelling. GIS data was captured from Open Street Map using the Gismo plugin in Grasshopper. DragonFly was utilized to generate microclimate weather data. The energy simulation engine was EnergyPlus with OpenStudio, while data analysis was done in Python using Pandas and SK-Learn.

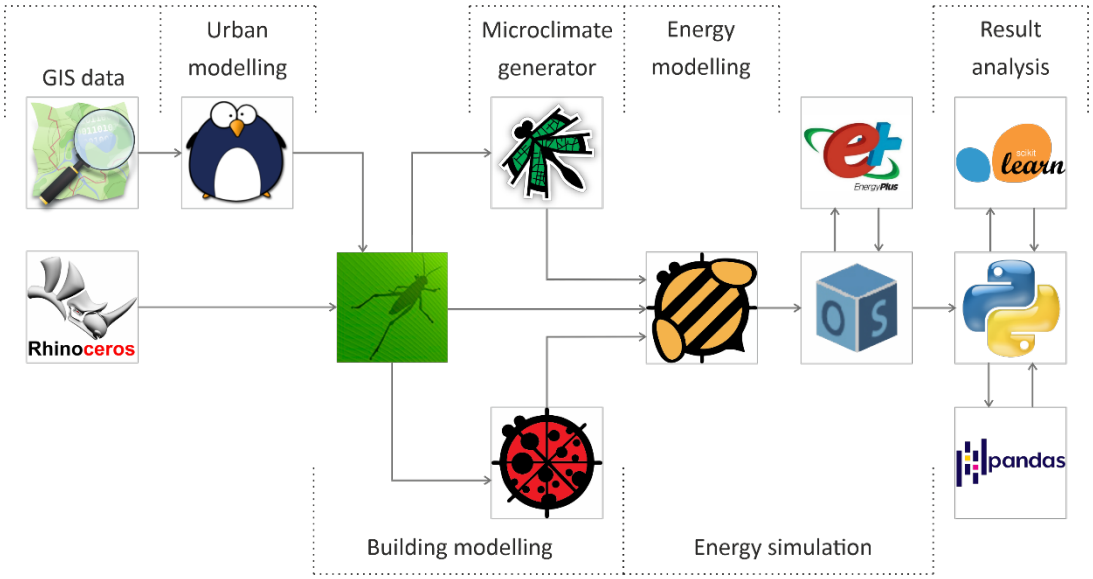


Figure 2: Utilized computer software for each stage of the entire process

3.1 Statistical analysis

Statistical methods in this research were applied in two main stages. The primitive stage worked on the benchmarking data to prepare a clean and clear dataset for the rest of the progress. Also, it gives a quantitative overview of the entire population. The conclusive stage performed to assess the results from the energy simulations.

3.1.1 Primitive stage: Benchmarking analysis

In the primitive stage, data from NYC Benchmarking [17] was used. For this study, the data from 2017 was utilized. Also, the analytical report for 2014 and 2015 benchmarking data [95], provided by the local authorities, were used as complement information. This benchmarking

data collects data for the energy and water use of buildings. The dataset includes 34,686 records of buildings in the City. Each building in the dataset has identification information, including "Borough, Block, Lot" (BBL), Building Identification Number (BIN), borough, address, postal code, and property name. Then building features are introduced consisting of the primary type, largest, second and third largest type, total floor area, area of each type, year built, and metered area (including Whole building, part of the area or specific type). Afterwards, the energy use information of the buildings is given, including source EUI, site EUI, fossil fuel sources, district systems, and electricity use. Fossil fuels are divided into fuel oil (5 types), diesel, propane, and natural gas. District systems are divided into district steam, district hot water, and district chilled water. Outliers, zeros, and empty cells can be seen in the dataset.

Table 1: Detail information of the available data for the major parameters.

	Parameter	Unit/ Description	Number of values
Identification	Order	1 to 34,686	34,686
	BBL	10-digit number	33,956
	BIN	6-digit number	14,700
	Address	Text	34,686
	Postal code	5-digit number	34,686
Building features	Gross floor area	ft^2	34,686
	Primary property type	Text	34,686
	Year built	YYYY	34,686
	Metered area	Text	31,710
Energy use	Site EUI	kBtu/ft^2	32,073
	Fuel oil (5 types)	kBtu	7,917
	Diesel	kBtu	49
	Propane	kBtu	1
	Natural gas	kBtu	13,986
	District Steam	kBtu	1,575
	District hot water	kBtu	8
	District chilled water	kBtu	31
	Electricity (grid purchase)	kBtu	13,678

Table 1 presents the units, descriptions, and the number of values of the major parameters in the dataset. The following paragraph gives an overall view of the contents of the described data. The first set of data, identification, was used to define the building and find it in the GIS. The first data column in the dataset was "Order", which was used to identify each building since it was a unique parameter for each record. The second set of data, building features, were mainly used to define the category of the building.

One of the essential parameters is "Primary building type". In total, 66 types of buildings exist in the dataset with different frequencies. A general overview of them is illustrated in Figure 3. Accordingly, the focus in this study was on "Multifamily housing" and "Office". "K-12 School" was not focused since most of them were located in mixed-type buildings, and the measured data was combined. All building types and their frequency are presented in Appendix A.

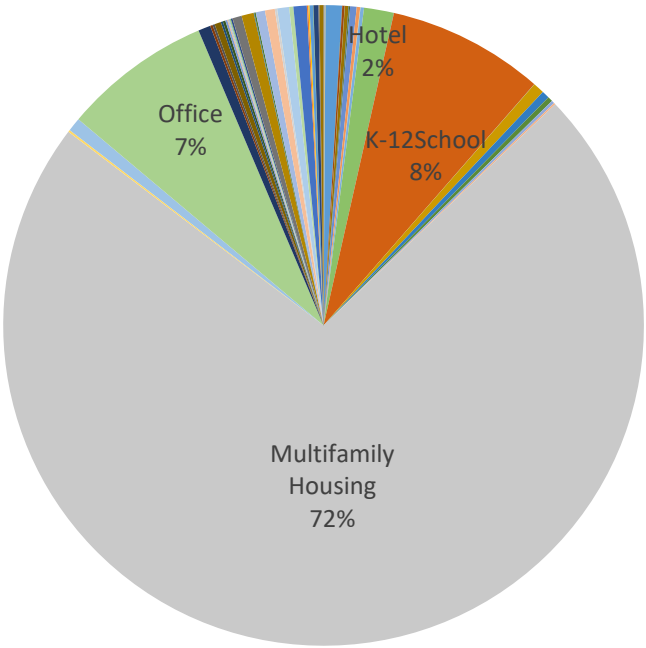


Figure 3: The variability of different building types in the dataset

This dataset is composed of a wide range of buildings total floor area. As illustrated in Figure 4, half of the buildings have an area between 4,200 m^2 (45,000 ft^2) (25^{th} percentile) and 11,200 m^2 (120,000 ft^2) (75^{th} percentile). Less than 1% of the records have an area of more than 100,000 m^2 (1,076,000 ft^2) up to around 900,000 m^2 (9,700,000 ft^2) which are not included in the histogram. "Year built" was used to define the vintage based on the later explained divisions with the illustrated distribution in Figure 5.

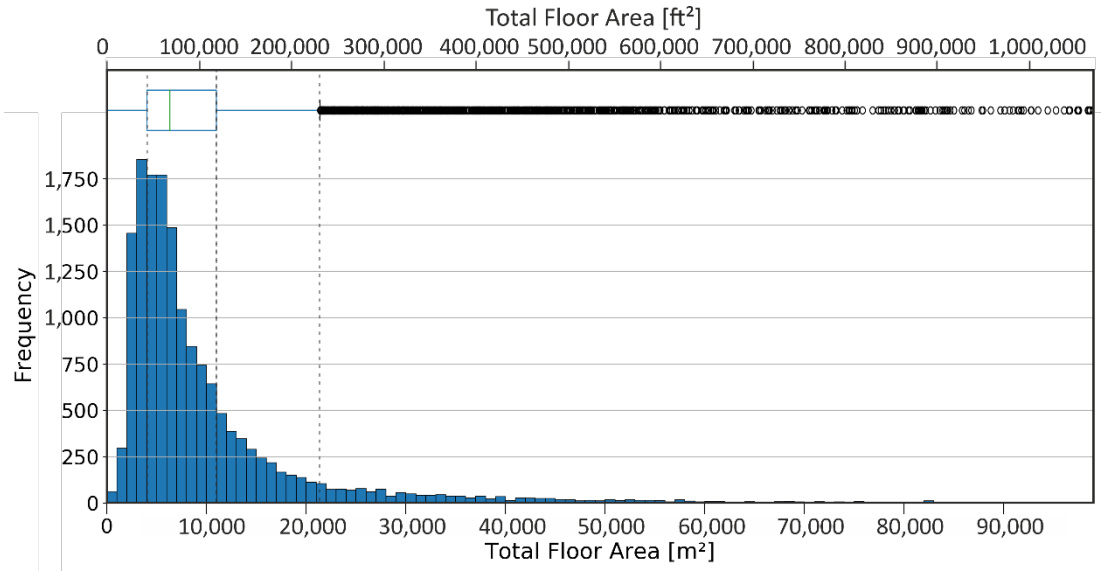


Figure 4: The frequency of the different building area

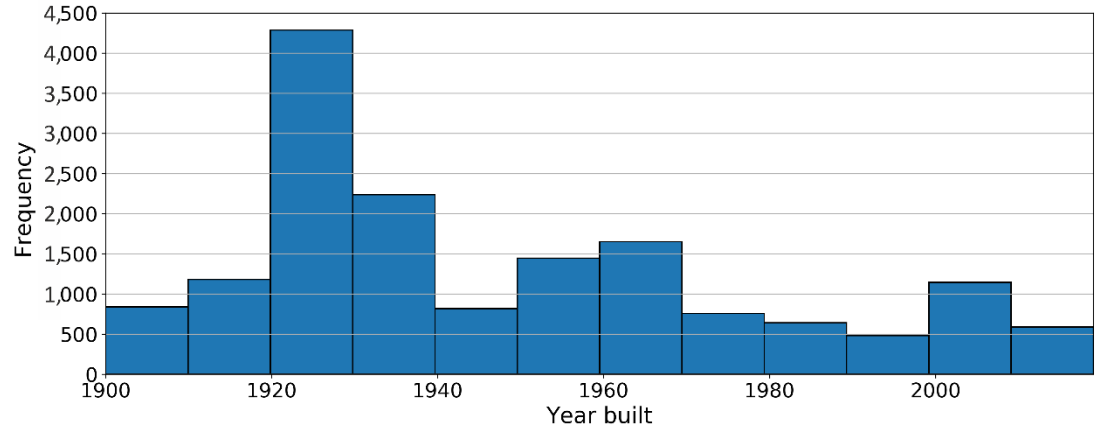


Figure 5: The frequency of building age (Year built)

The last set of data, energy use, denoted the amount of energy use for each available source. The first parameter in this set was "Site EUI", which was normalized by area. Based on the dataset explanation, Energy Use Intensity (EUI) is calculated from total energy use. Different sources of energy were presented in the dataset. Diesel, Propane, District hot water, and District chilled water had negligible frequency in the dataset. Thus, they were not considered in the statistical analysis. Except for "Electricity (grid purchase)", the rest of the sources (fossil fuels) were assumed to provide thermal energy for Space heating (SH) and Domestic Hot Water (DHW). Electricity is the provider for lighting, equipment, cooling system, pumps, fans, heat pumps, etc. These numbers were measured from the grid and not in the final place of use. Thus, the efficiency of the facilities had to be applied.

The dataset was filtered by removing the following items. The cleaned dataset included approximately 16,600 records.

a. Records with empty Site EUI
b. Records with the "metered area" column not equal to "Whole building" were removed. Meaning that the remainders had the measured energy use for the entire building and all energy sources.
c. Buildings with the "primary type area" less than 90% of the total floor area were removed. Hence, the remainders counted as a single type of building.
d. Items with year built before 1900 were removed because they were not an interest of this study, and not many specifications were available.
e. Site EUI less than 90 kWh/m^2 (30 kBtu/ft^2) and larger than 800 kWh/m^2 (250 kBtu/ft^2) were removed. The average was around 270 kWh/m^2 (80 kBtu/ft^2), then three times more and less were counted as outliers.

The previous years' analytical report [95] presented more detailed energy use in the NYC building stock. This report included an energy end-use breakdown which indicated an average percentage of separated energy demands, presented in Table 2. It should be mentioned that this breakdown is a statistical output from the dataset and does not represent necessarily the entire population. For instance, based on Figure 5, a large number of buildings were built in the 20s and 30s. They influenced the total results and caused a skewness toward older buildings' averages, like a higher heating demand due to the low-quality envelope.

Table 2: Energy end-use breakdown in Multifamily housing and Office buildings

Building type	Typically fossil fuel		Typically electric
	Space heating	DHW	Space cooling
Multifamily housing	38%	15%	8%
Office	22%	4%	11%

3.1.2 Conclusive stage: Results analysis

Results of EPS were analyzed in two stages. First, the total simulated values were controlled by the total measured data. Regarding the deviations of the EPS results from actual values, the calibration of the simulation process was not done on every single building. A linear regression model was defined to control the relations of all simulated and measured values. This calibration was done with the first 20 simulations.

The next stage was to control the hourly values of the EPS results. The hourly profile for each pair of buildings with the same category (explained in 3.3.1) had to be checked to assess the similarity of the hourly trend. To check the similarity of the hourly profiles, values were normalized by the sum, and then R^2 was calculated for each pair of buildings. In this work,

hourly profiles with $R^2 > 0.8$ were considered as a similar trend. ASHRAE recommends an $R^2 > 0.75$ for a pair of time series to be counted as similar [52].

3.2 Energy performance simulation

Several methods for calculating the energy performance of buildings have been developed during the life of construction engineering. Every EPS approach considers a part of actual situation parameters and necessarily ignores some of them. The used software and methods for modellings are explained in the following sub-chapters.

3.2.1 Building modelling

This research utilized the computer program Rhino and Grasshopper for the building modelling because of the appropriate integration with other required sources, such as GIS and Python, explained further in this report. Buildings were made by scripting in Grasshopper. The captured data from GIS, based on the building address or coordination, was converted to a 3D model with simple surfaces, divided into floors and 6×6 m^2 (20×20 ft^2) zones for residential and 10×10 m^2 (33×33 ft^2) zones for non-residential buildings. A set of native components in Grasshopper was used to define the windows' orientation and assign the percentage of the windows to the façade in Honeybee.

3.2.2 Urban modelling

The effects of the surrounding on the building energy performance are substantial. It becomes crucial when the context is a dense urban area. Open Street Map is an open-source GIS database that provides the ability to fetch data through API or direct export from the website. Gismo is a recently developed plugin for Grasshopper, which gets GIS data from Open Street Map, including geometries and types of places. Gismo interprets GIS data directly to Grasshopper objects. A sample view of a produced urban model with a subject building is shown in Figure 6.

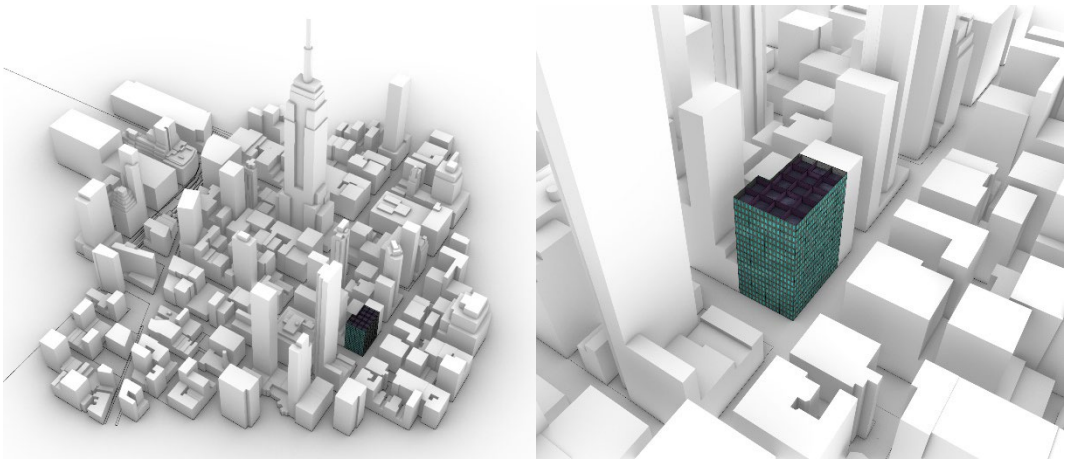


Figure 6: A sample urban model (left) and subject building (right) in the context

The identification parameters of buildings (Table 1) from benchmarking dataset were used to identify buildings in GIS data. By merging the location-related columns in the benchmarking dataset, a location string with this pattern was produced: "Address" + "postal code" + New York + U.S., which is readable by GIS software to identify the desired building.

NYC has a general format for addressing the properties, called Property Address Directory (PAD). The five boroughs in the City have a standard one-digit code: 1= Manhattan, 2= The Bronx, 3= Brooklyn, 4= Queens, 5= Staten Island. Then, each lot identified with a code made of borough code (1-digit), tax block number (5-digit), tax lot number (4-digit), which called BBL. Accordingly, each lot has a unique BBL, which refers to the lot and all buildings on it. Thereafter, buildings on the lots are assigned a Building Identification Number (BIN) [96]. BBL and BIN were used to verify the address and selected building manually by the user.

3.2.3 Weather data

The required weather data for this study were provided from two main sources: (1) Typical Meteorological Year (TMY) weather file as EPW-file format from the EnergyPlus database [72], (2) Actual Meteorological Year (AMY) from The National Centers for Environmental Information [97]. The TMY files for La Guarida Airport in NYC was used as a typical condition to have a general hourly trend. The EPW-file had an annual average temperature of 13.5 ℃ (56.3°F) with a maximum of 36.7 ℃ (98°F) and a minimum of -15 ℃ (5°F). This weather data describes the mesoclimate condition of the correlated area; thus, Urban Weather Generator (UWG) was used via the Dragonfly plugin in Grasshopper to generate the microclimate weather file.

AMY weather file for the year 2017 was used to assess the real situation since the benchmarking dataset included measured data for 2017 as well. DragonFly plugin in Grasshopper was used to make EPW-files format from downloaded AMY data. The selected stations were: La Guardia Airport, The Battery, Newark Liberty International Airport, and Bergen Point. Not all available stations on the website were convertible to the standard EPW-file; hence, those mentioned were selected among many. The utilized AMY hourly dry-bulb temperatures are presented in Figure 7. Since all four AMY weather data did not show significant differences in the weather parameters, La Guardia Airport was chosen following the TMY weather data. The maximum temperature is 37.8 ℃ (100 °F), and the minimum is -11.1 ℃ (12 °F) with an average of 14.2 ℃ (58 °F).

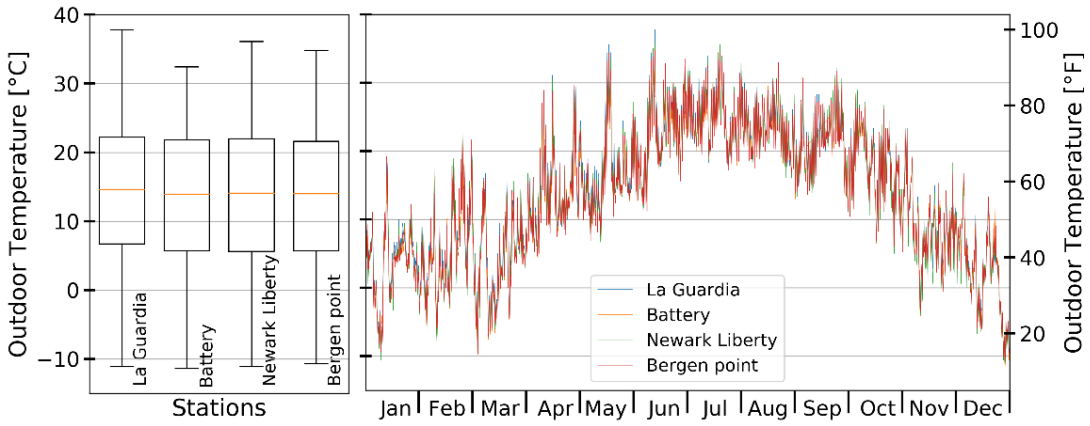


Figure 7: Hourly outdoor temperature from measured data for the four stations for 2017

3.2.4 Energy modelling

Energy modelling, in general, consists of the building geometry and its specifications, boundary conditions, and surroundings, which can be used in a simulation engine. In this study, three main components have been defined for the inputs of an EPS. These divisions are given from the needs of the simulation in this study and also inspired by ASHRAE Guideline 14 [48]: (1) Indoor condition, (2) Outdoor condition, and (3) building specifications. Based on ASHRAE Guideline 14 [48], EPS parameters can be divided into transient and steady. Accordingly, Outdoor- and Indoor conditions have been defined as transient and Building specification as steady. Moreover, indoor conditions fluctuate in a daily loop, while outdoor conditions have a yearly frequency. Hence, their effects on the simulation results appear with different scale and occurrences. Figure 8 depicts the components and their parameters.

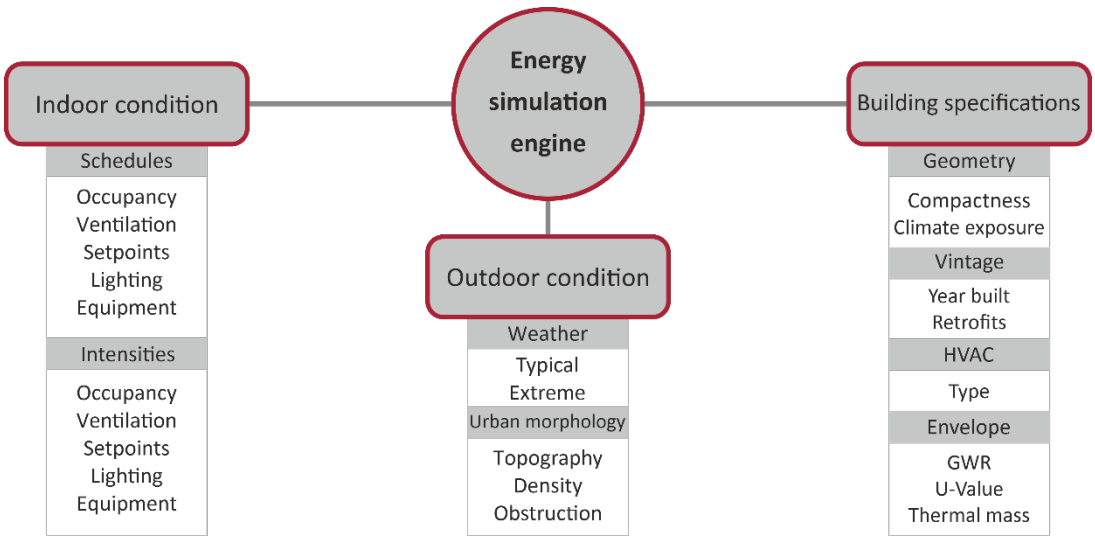


Figure 8: Three main components of EPS and their parameters.

The simulation engine for this study was OpenStudio which is connected to Honeybee plugin in Grasshopper. OpenStudio exploits EnergyPlus for energy simulation while adds the capability of applying an HVAC system in the simulation. The energy simulation parameters were given from New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs [98]. This guideline also suggested four vintages for buildings and defined different characteristics for them, denoted in Table 3 and Table 4.

Table 3: Multifamily housing buildings characteristics for EPS

Parameter	Vintage 1	Vintage 2	Vintage 3	Vintage 4
Time period	before 1940	1940-1979	1980-2006	2007-present
Wall U-value				
$[W/(m\text{^}2 \cdot K)]$	1.42	0.81	0.52	0.30
$([Btu/(hr \cdot ft\text{^}2 \cdot °F)])$	(0.25)	(0.14)	(0.09)	(0.05)
Roof U-value				
$[W/(m\text{^}2 \cdot K)]$	2.84	0.52	0.30	0.15
$([Btu/(hr \cdot ft\text{^}2 \cdot °F)])$	(0.5)	(0.09)	(0.05)	(0.02)
Windows U-value				
$[W/(m\text{^}2 \cdot K)]$	5.28	5.28	3.86	1.59
$([Btu/(hr \cdot ft\text{^}2 \cdot °F)])$	(0.93)	(0.93)	(0.68)	(0.28)
Windows G-value $[\%]$	87	87	77	49
Infiltration $[ACH]$	1	1	0.5	0.35
Heating setpoint $[℃]$ $(℉)$	22.8 (73)	21.1 (70)	21.1 (70)	21.1 (70)
Heating setback $[℃]$ $(℉)$	21.1 (70)	19.4 (67)	19.4 (67)	19.4 (67)
Cooling setpoint $[℃]$ $(℉)$	23.9 (75)	23.9 (75)	23.9 (75)	23.9 (75)
Cooling setback $[℃]$ $(℉)$	25.5 (75)	25.5 (78)	25.5 (78)	25.5 (78)
HVAC type	PTAC + hot water boiler (20% oversizing)			

Table 4: Office buildings characteristics for EPS

Parameter	Vintage 1	Vintage 2	Vintage 3	Vintage 4
Time period	before 1940	1940-1979	1980-2006	2007-present
Wall U-value				
$[W/(m\^2 \cdot K)]$	1.42	0.81	0.52	0.30
$([Btu/(hr \cdot ft\^2 \cdot °F)])$	(0.25)	(0.14)	(0.09)	(0.05)
Roof U-value				
$[W/(m\^2 \cdot K)]$	2.84	0.52	0.30	0.15
$([Btu/(hr \cdot ft\^2 \cdot °F)])$	(0.5)	(0.09)	(0.05)	(0.02)
Windows U-value				
$[W/(m\^2 \cdot K)]$	5.28	5.28	3.86	1.59
$([Btu/(hr \cdot ft\^2 \cdot °F)])$	(0.93)	(0.93)	(0.68)	(0.28)
Windows G-value $[\%]$	87	87	77	49
Infiltration $[ACH]$	1	1	0.5	0.35
Heating setpoint $[℃]$ $(℉)$	Occupied hour: 21.1 (70)		Unoccupied hour: 19.4 (67)	
Cooling setpoint $[℃]$ $(℉)$	Occupied hour: 23.9 (75)		Unoccupied hour: 25.5 (78)	
HVAC	Fan coil + DOAS (10% over sizing)			

3.2.5 Energy simulation workflow

Two sets of simulations were utilized to provide hourly values and loads ratio. The first one was called "Actual Buildings", an iterative algorithm using the selected buildings from the benchmark dataset. A list of buildings' addresses was exported from NYC benchmarking dataset according to their EUI. 25th, 50th, and 75th percentile of EUIs were captured as representatives of the low, normal, and high energy use buildings, respectively. The list of the selected buildings was used to capture data from GIS, modelling, and simulation. These simulations used AMY weather data. The HVAC system was assigned to "Actual Building" simulations to have the results of energy demands, which is the required electricity or fossil fuel to run the facilities. The energy simulation workflow is presented in Figure 9.

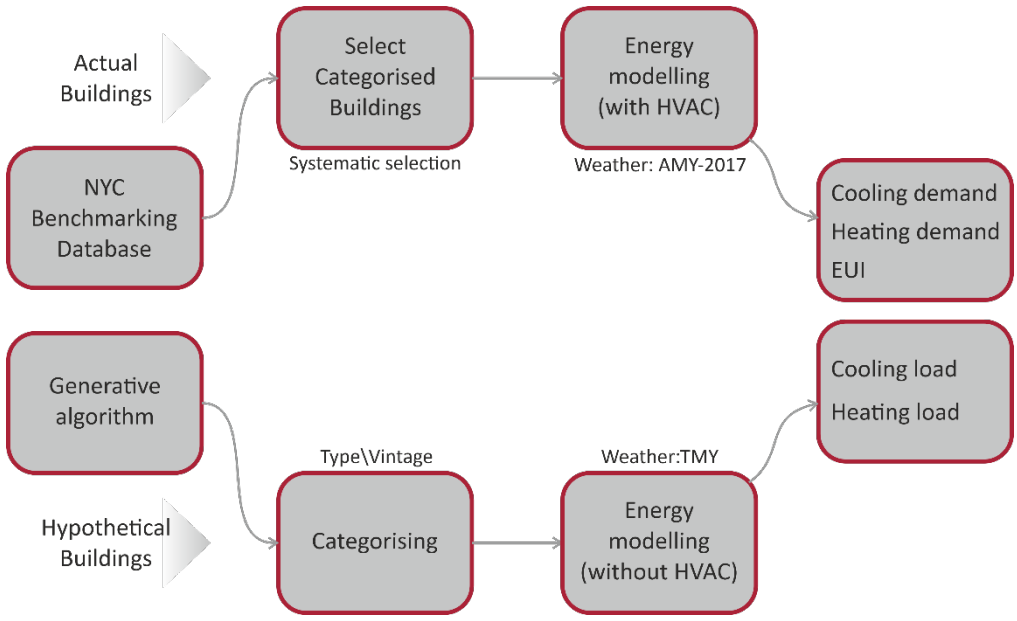


Figure 9: Energy modelling and simulation workflow

The second series of simulations was a generative algorithm to model several "Hypothetical Buildings". These models represented general forms of the buildings in the urban area with similar simulation settings. Modular models provide better results for analytical purposes due to a broad and comprehensive variety as well as a higher process speed. Generative models were made randomly from developing a cubic module of 5×5×2.7 m in the different horizontal and vertical arrangement as presented in Figure 10.

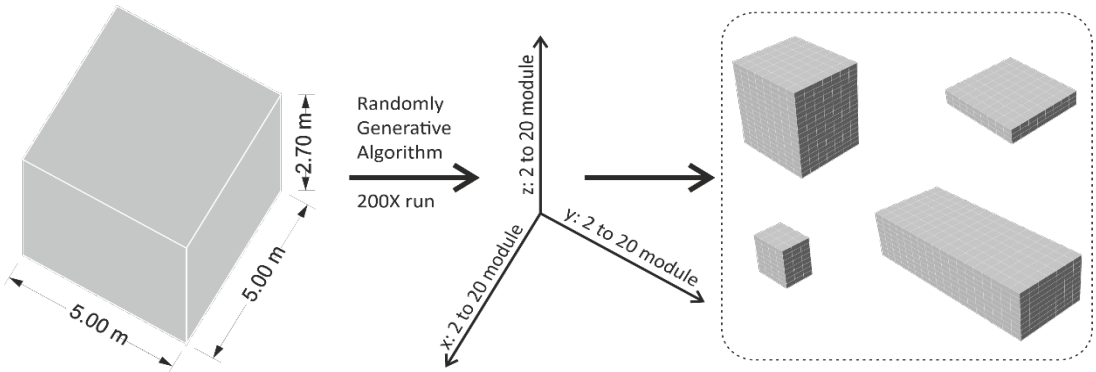


Figure 10: The generative model to produce modular geometries

This set of simulations applied the converted TMY weather file to the microclimate. "Hypothetical Buildings" simulations were run with and without an HVAC system. Simulations without assigning an HVAC system gave results as thermal loads, which is the thermal energy required to be added or removed from the building to maintain the indoor conditions. Thereafter, comparing results of energy demands and thermal loads gave the ratio

of thermal loads to energy demand which showed the efficiency or COP of the HVAC system in an hourly profile. The share of cooling demands in EUI was calculated from the total cooling demand ratio to the corresponding EUI from the same simulations. The cooling demand ratio was used to extract cooling demand from measured EUI. The cooling load was calculated by applying the COP curve to the cooling demand.

EPS in EnergyPlus includes heating and cooling demand, lighting loads, equipment loads, pumps, and fans electricity, which is EUI in total. Hence, to calculate closer to the measured data, more items needed to be added to the simulated EUI. The first one was the DHW load. Thus, to generate the DHW annual hourly profile, the DHW generator in Honeybee was used. This component worked based on the number of people per area and the number of units in the building, and it used a predefined library from ASHRAE 90.1 [99]. The second energy use to be added to the simulated EUI was the electricity demand to run elevators. According to ASHRAE 90.1, 7,000 kWh (24,000 kBtu) was added to the annual EUI for every 4,645 m^2 (50,000 ft^2) of the total floor area for each elevator. However, the defined number in ASHRAE 90.1 and the area per capita are just an approximation to add missed values for the total EUI in EPS compared to the measured EUI.

3.3 The reverse energy estimation approach

3.3.1 Backbone

The primary purpose of this study is to develop a reverse method of energy estimation based on the availability of relevant data. Regarding this purpose, a hypothesis was defined as following: Does a **group of similar buildings** have a similar **hourly energy trend**?

The following two terms should be defined to check the validity of the hypothesis. First of all, a group of similar buildings is a number of building with similar indoor conditions and building specifications. In this research, a group of similar buildings is called a category. Categories required identifiable characteristics. Three levels of categorizing were defined to separate buildings. The first level was building type, which includes 63 different types. However, "Multifamily housing" and "Office" were used in this study due to the high frequency in the dataset. The second level was vintage, divided into four groups: Pre-1940, 1940-1979, 1980-2006, and 2007-present. Each group is related to a particular development in the building codes; hence, the constructed buildings in each period had a similar specification following the regulations.

The third level was Compactness (C) as a representative parameter of the geometry. Compactness can be calculated from equation (7) [100], [101]. The ratio of the compactness of a building to the compactness of a cube with a similar volume called relative compactness (RC). Cube is the most compact geometry for a building with a specific volume. Relative compactness result from equation (8) [102]. Three levels of compactness were defined to identify buildings geometry characteristics. $0.8<RC\leq1$: compact, $0.5<RC\leq0.8$: normal, $RC\leq0.5$: exposed.

$$C = \frac{Building\ Volume}{Envelope\ Area} \quad (6)$$

$$RC = \frac{C_{cube}}{C} \quad (7)$$

The second term is the Hourly Energy Trend (HET). It is the normalized hourly loads by the sum, which is unitless and scaleless, describing the fluctuation pattern of a building's load. HET is a time series of coefficients giving the hourly percentage of the total load. For every building in different categories, HET was calculated from the simulated data. Thereafter, in each category, similar trends were kept, and the outliers were removed. To define the outliers which had significantly different HETs, R^2 was used with a threshold of 0.8. Accordingly, R^2 was calculated for every pair of buildings, and the results stored for each building. Then, buildings with an average R^2 less than 0.8 were removed. This process was performed in the "Analytic Kernel".

The remaining HETs were used to make an average HET for the category called "Reference Model". Each reference model represents the corresponding category and all buildings in it. This process was performed in "Reference Model Generator", where reference models were generated for the space heating (SH) load, cooling load and DHW load separately as 2-dimensional arrays of 8760×3.

The stored reference models were used in "Reverse Method Kernel" to estimate the total hourly loads of a building. This calculation can be done by the multiplication of the imported total loads by the reference model. The result is a time series of the hourly loads.

3.3.2 Operation

The subject building should be defined with a unique code from identification information (Table 1) to choose the building from benchmarking dataset and import it to the reverse method. Then, corresponding building features and energy use (Table 1) will be captured from the relevant columns in the benchmarking dataset. Based on the building features, the category of the building can be defined. After that, a matching reference model will be fetched from the reference model generator. Energy use information will be used to calculate separated loads based on the following points and multiply them by the correlated reference model.

a. Cooling load: $\left(\frac{\sum_1^{8760} \widehat{Cooling\ load}}{\sum_1^{8760} \widehat{EUI}}\right) \cdot EUI_{benchmark}$

b. Heating load: $(district\ steam + natural\ gas + fuel\ oils) \cdot \varepsilon$

c. SH load: $heating\ load \cdot m$

d. DHW: $heating\ load \cdot (1 - m)$

Parameters with a hat ($\widehat{\ }$) are simulated data. $\widehat{Cooling\ load}$ is from simulations without HVAC and $\widehat{EUI}$ is derived from simulations with an HVAC system. ε is the heating facilities efficiency which is considered 0.75 according to [98]. m is the ratio of SH load to the entire heating load from Table 2 which is 0.72 and 0.84 for Multifamily housing and Office, respectively. The remaining part counted as DHW load (0.28 and 0.16) [95]. The reverse method workflow is demonstrated in Figure 11. The first two boxes on the left provide data to make HET. Then Analytic kernel performs statistical analysis on HET and removes outliers. In the next stage, the Reference Model Generator calculate the median value of the remaining HETs from the previous stage. The Reverse Method Kernel allows the user to input the building and get the results based on the aforementioned process. The Python code to perform this progress is presented in Appendix B.

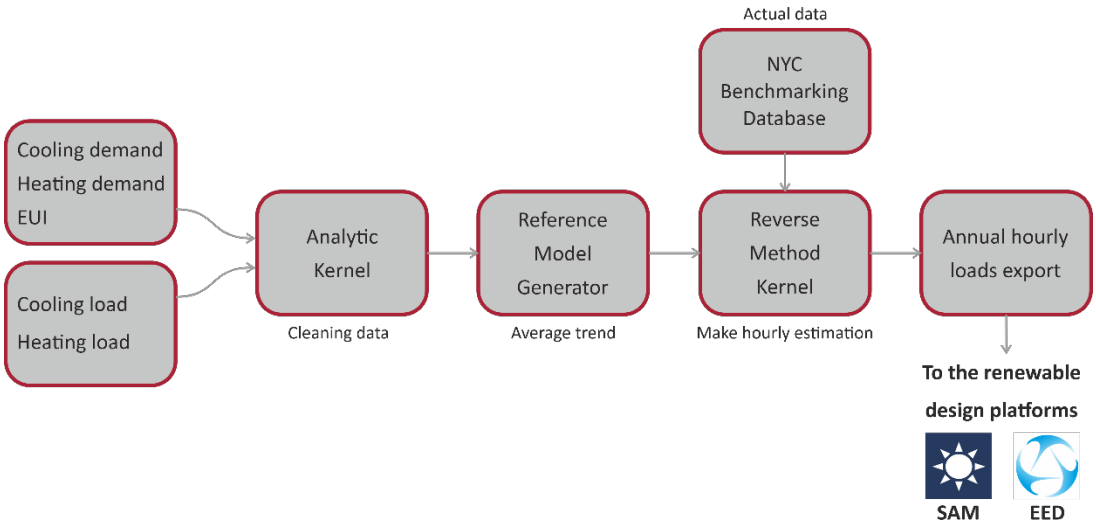


Figure 11: Reverse method workflow

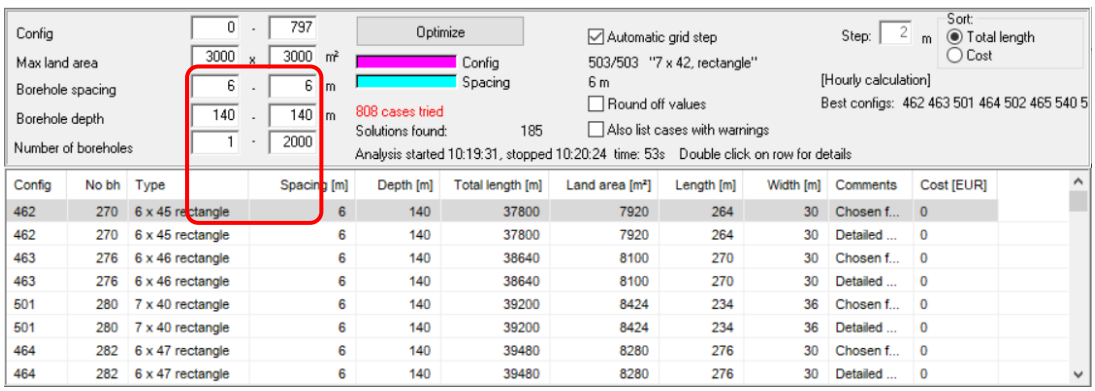
3.4 Geothermal calculation

According to the research goal and the appropriate condition of establishing geothermal energy in the urban area, a rough estimation of the BTES system was performed. Geothermal energy was selected due to the advantages compared to the other renewable sources, such as no visual disturbance, not influenced by the surrounding obstacles, and provide both heating and cooling. However, the same process as explained here can be performed for other software like System Advisor Model (SAM) from the National Renewable Energy Laboratory (NREL) to design the solar collectors and P.V. systems. These sample cases will show the validation of the outputs to be used in the EED program. Moreover, the geothermal system implementation was evaluated. Local regulations concerning drilling and excavation [103] explicitly define the drilling condition in different zones in the City. Hence, to have an overall value, a depth of 140 m (460 ft) was determined as a benchmark. Since the main target is the existing buildings, the system should be implemented in the basements or yards. As a result,

the limitation of the available area, operation restrictions such as extracted rocks, noise, and shaking will define the main arrangement and method of drilling the boreholes.

The objective of designing boreholes was to maximize the coverage of the heating load. To keep the ground balance in the long term, and avoid increasing boreholes depth, rejected heat from the cooling process was injected into the boreholes, equal to the extracted heat. Thus, a constraint was set for the cooling load to be equal to or smaller with the heating load. The excess rejected heat was assumed to be taken out by an auxiliary system. An adjustment coefficient was defined for the heating load to tune the heating peak load (system size) based on the limitations of the land. The capacity of the property to cover the thermal loads from the ground can be calculated roughly from the ratio of the possible boreholes in the property (based on the lot geometry and area) to the required boreholes (based on the calculation) for the entire loads. Thereafter, this ratio can be used to adjust the BTES system's size to cover a part of the required heating loads. The exceeded part must be covered by an auxiliary system.

Calculations in EED was performed through "Hourly Calculation", using separated hourly values for heating, cooling, and DHW. The fluid temperature was adjusted to 32°C (90°F) and 2°C (35°F), and the thermal conductivity of the ground was set to 3.08 W/m·k (1.78 Btu/h·ft·°F). The borehole thermal resistance set to 0.03 m·k/W (0.05 h·ft·°F/Btu). The volumetric ground heat capacity equals 2.5 MJ/m^3·k (37.1 Btu/ft^3·°F) with a ground surface temperature of 16.5 °C (62 °F). To consider the influenced ground loads by the heat pump, the Seasonal Performance Factor (SPF) was set to 4 and 3 for heating and cooling, respectively. The optimization tool in EED was used to calculate the required number of boreholes for the imported loads. Borehole spacing was fixed at 6 m (20 ft), and borehole depth was defined as 140 m (460 ft), based on the obtained information from a geothermal consultant in NYC. The land area and the number of boreholes were defined as limitless. Then the lowest number of required boreholes was achieved. Optimization settings are shown in Figure 12.



Config	No bh	Type	Spacing [m]	Depth [m]	Total length [m]	Land area [m²]	Length [m]	Width [m]	Comments	Cost [EUR]
462	270	6 x 45 rectangle	6	140	37800	7920	264	30	Chosen f...	0
462	270	6 x 45 rectangle	6	140	37800	7920	264	30	Detailed ...	0
463	276	6 x 46 rectangle	6	140	38640	8100	270	30	Chosen f...	0
463	276	6 x 46 rectangle	6	140	38640	8100	270	30	Detailed ...	0
501	280	7 x 40 rectangle	6	140	39200	8424	234	36	Chosen f...	0
501	280	7 x 40 rectangle	6	140	39200	8424	234	36	Detailed ...	0
464	282	6 x 47 rectangle	6	140	39480	8280	276	30	Chosen f...	0
464	282	6 x 47 rectangle	6	140	39480	8280	276	30	Detailed ...	0

Figure 12: Settings for optimization in EED

The actual bore field was defined according to the land dimensions with the same spacing and depth. Thereafter, the plausible number of boreholes was compared to the required amount

from the calculation. This comparison acquires the capacity of the land to provide the thermal load from the ground.

3.5 Delimitations

Due to the comprehensive coverage of the research subject, this study encounters several limitations. First of all, the researcher was unable to obtain high-quality temporal and spatial measured data to calibrate and verify the simulations on an hourly scale because the data was not disclosed by the authorities. Secondly, the simulation engine and the generative algorithm required a high computational hardware capacity and time. However, this study had a restricted timeframe. Adding more simulated data to the reference models can enhance the predictions, requiring more time for more simulations. Third, using the reverse method to evaluate the geothermal system needed ground characteristics that were not accessible. Finally, the applied GIS data showed a considerable number of failures to capture data in the dense urban area where the buildings density cause complexity in determining their boundary line in GIS; thus, around 20% of simulations were failed.

4 Results analysis

The results of the applied method and their corresponding analysis are presented in this chapter in four parts: (1) benchmarking data shows the analysis of the NYC benchmarking dataset contents, (2) generated microclimate weather data, (3) EPS results for actual and hypothetical buildings and DHW, and (4) results from the reverse method and sample cases.

4.1 Benchmarking data

The statistical analysis of the NYC benchmarking dataset based on the defined categories is discussed in this section.

A general overview of the buildings' construction year and the total energy use are outlined in Figure 13. The scatter plot shows a slight reduction in EUI for newer buildings. The mean value and median are close to each other, around 270 kWh/m^2 (85 kBtu/ft^2). The scatter plot also gives an understanding of the number of constructions over time. The densely scattered point around the 20s shows a large number of constructions in those years. Also, it says the older buildings with a lower construction quality have a high frequency in the City.

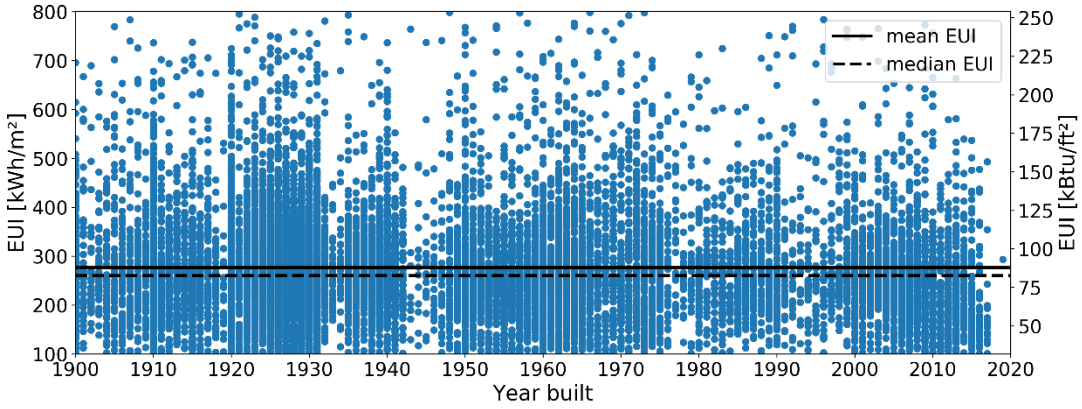


Figure 13: Scatter plot for the normalized EUI by the floor area for each year built. The solid line and dashed line show the mean and median of the EUI, respectively. This figure also visualizes the frequency of constructions every year.

The median values of the EUI for each category are illustrated in Figure 14. A decline for multifamily houses can be seen from 270 kWh/m^2 (86 kBtu/ft^2) in the first vintage to 210 kWh/m^2 (67 kBtu/ft^2) in the last vintage. This reduction can result from better building performances and applying newer building codes, especially after the global oil crisis in the 70s. Office buildings do not show a strong trend for the EUI, and they just fluctuate around 250 kWh/m^2 (80 kBtu/ft^2) over different periods.

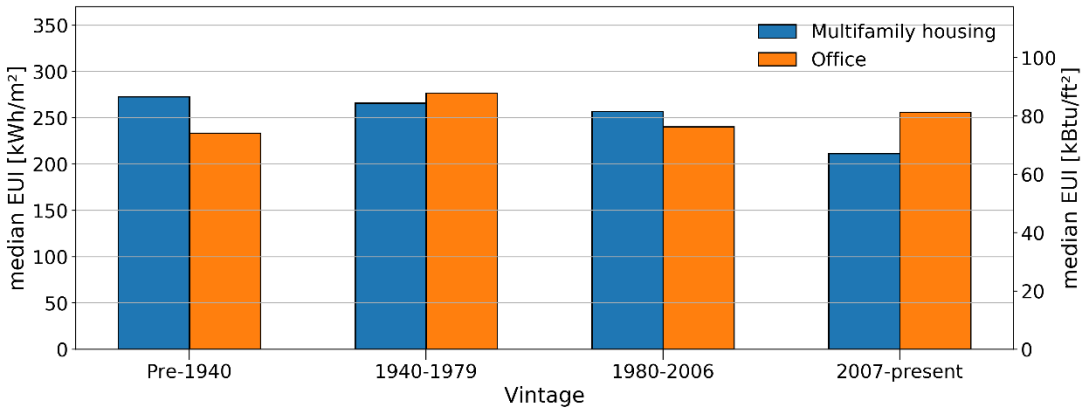


Figure 14: Median of EUI for Multifamily housing and Office divided by vintages

The frequency of the different energy sources in the dataset is illustrated in Figure 15. Electricity and natural gas are by far the largest source of energy in NYC following by fuel oils. This ratio is similar in both types over different vintages.

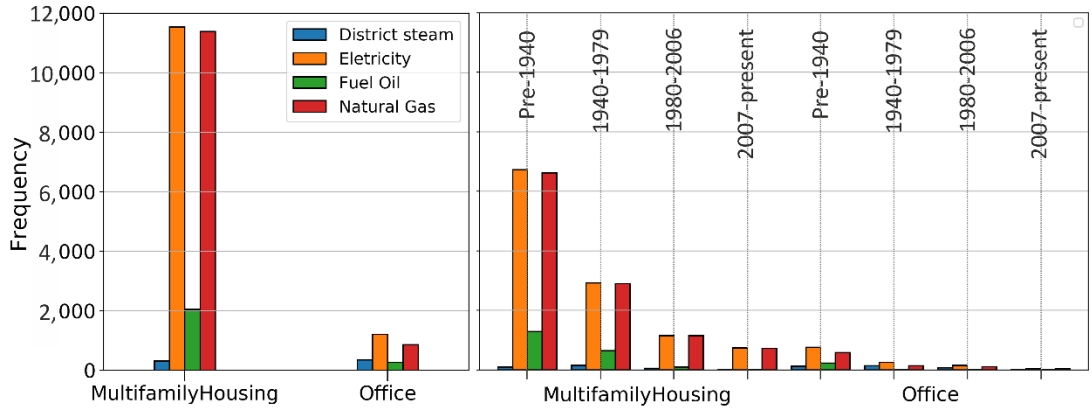


Figure 15: Frequency of different energy sources for the two main building types, overall data (left) and categorized data (right).

4.2 Microclimate

Outdoor dry-bulb temperature for the utilized EPW-file from La Guardia Airport is shown in Figure 16. The microclimate EPW-file had an annual average temperature of 14.5 °C (58.1°F) with a maximum of 37 °C (98.6°F) and a minimum yearly temperature of -14 °C (6.8°F).

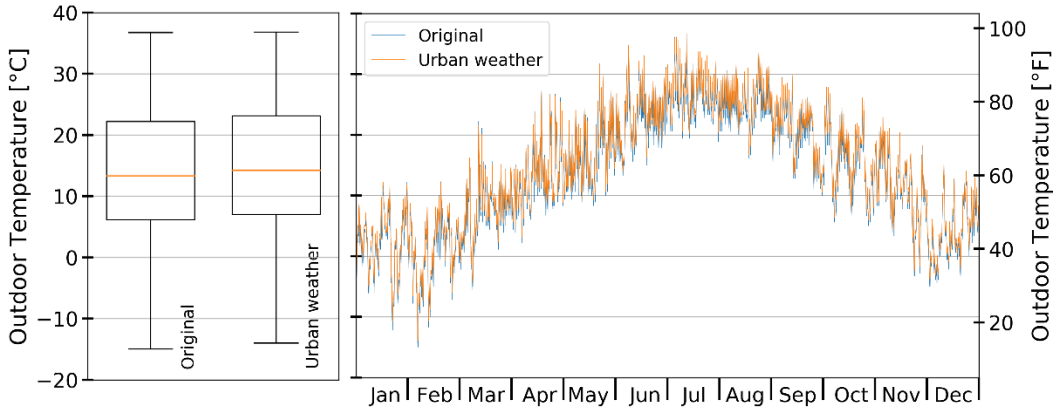


Figure 16: Dry-bulb temperature for the original and microclimate TMY weather file

4.3 Energy performance simulation

4.3.1 Actual buildings

The results for the energy simulations regarding twenty selected buildings were compared to the actual measured EUI. Simulated EUI showed a deviation of 70% to 130% from the measured one. This deviation can be a result of the different building performances since simulations used average values for the setting. Also, user behavior and operation strategies could not be considered precisely similar to the actual buildings. However, the regression line of the twenty values (blue dots in Figure 17) showed that simulated and measured values follow the regression model. In the regression line equation, the coefficient of the x-axis (Measured EUI) was almost 1, meaning that y=x+β. Hence, simulations' results could generally represent the measured data.

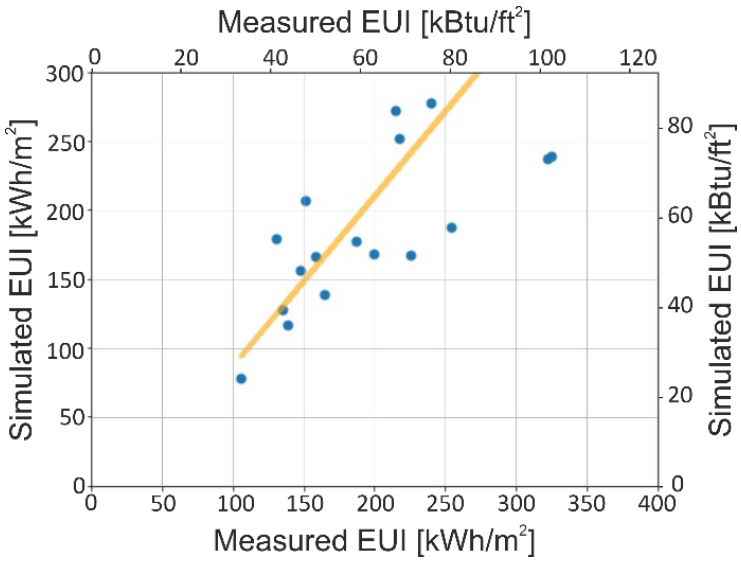


Figure 17: The comparison between simulated and measured EUI.

Randomly selected results for “Multifamily housing” with the “Order” number of 33212, 33380, 29630, and 30342 are presented in Figure 18.

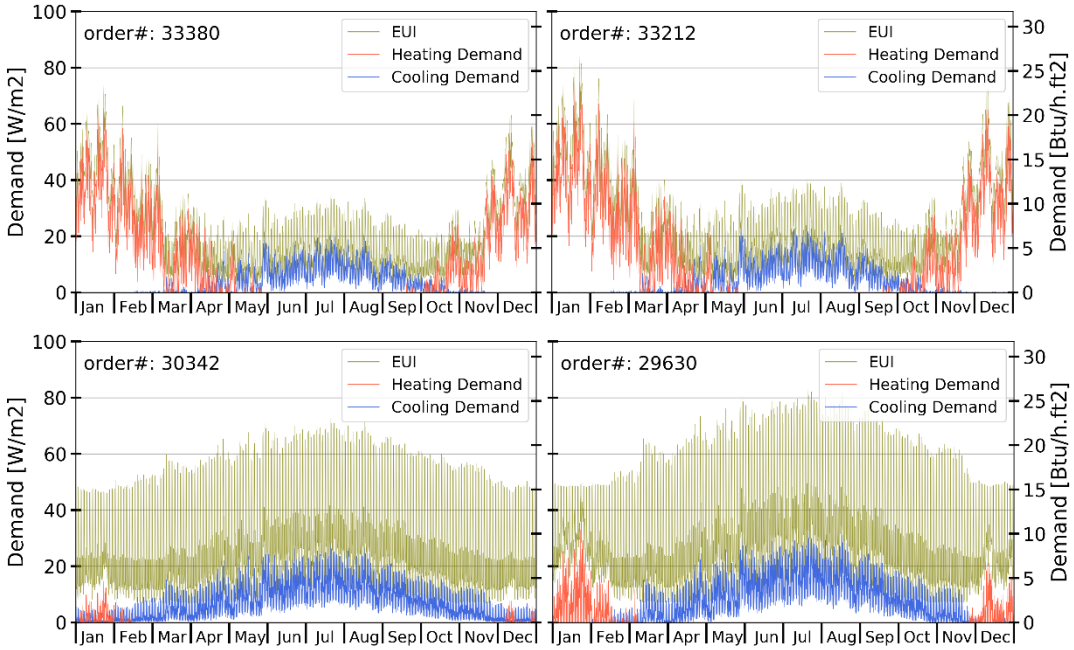


Figure 18: Annual hourly profile for EUI, Heating demand, and Cooling demand for Multifamily housing. Vintage: Pre-1940 (top) and 1940-1979 (bottom)

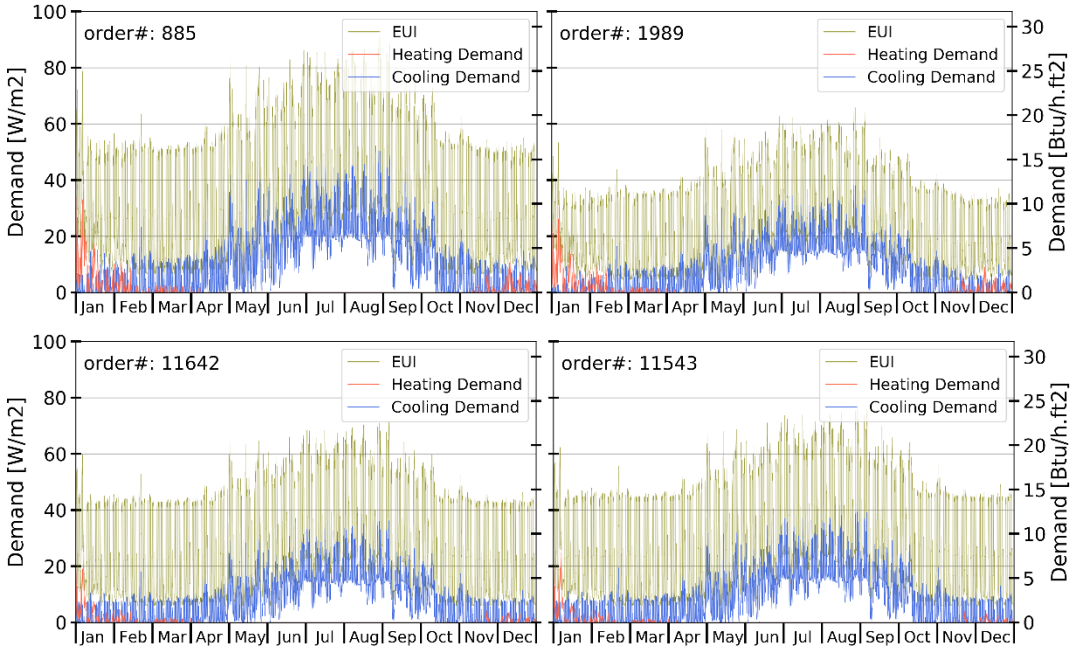


Figure 19: Annual hourly profile for EUI, Heating demand, and Cooling demand for Office. 1980-2006 (top) and 2007-present (bottom)

Results for four randomly selected offices with the "Order" numbers of 885, 1989, 11642, 11543 are also illustrated in Figure 19. The annual profile for office buildings showed an apparent weekly pattern, influenced by the indoor condition. It can be seen in the graphs as the weekends' low energy use. In comparison, residential buildings showed more variation based on the outdoor conditions in summer and winter. The results for R^2 for each pair of actual buildings with the same category are presented with boxplots in Figure 20.

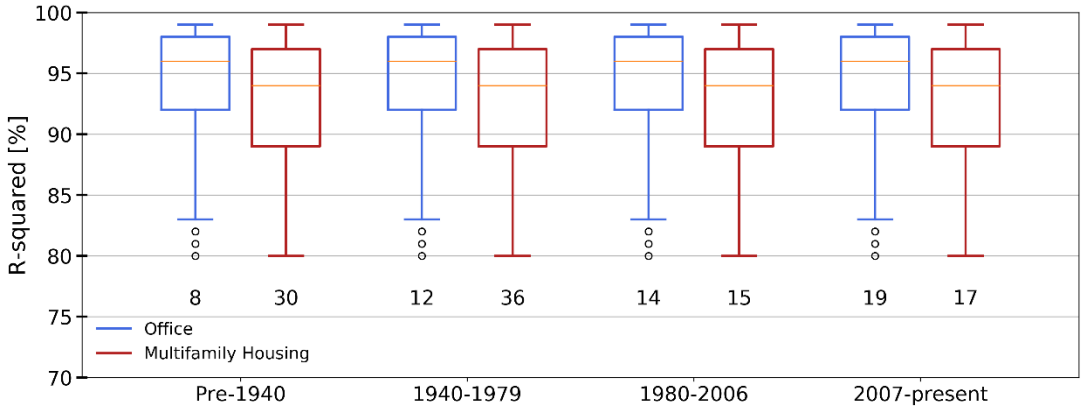


Figure 20: R^2 for each pair of actual buildings in every category.

Numbers for each box represents the number of buildings in the box.

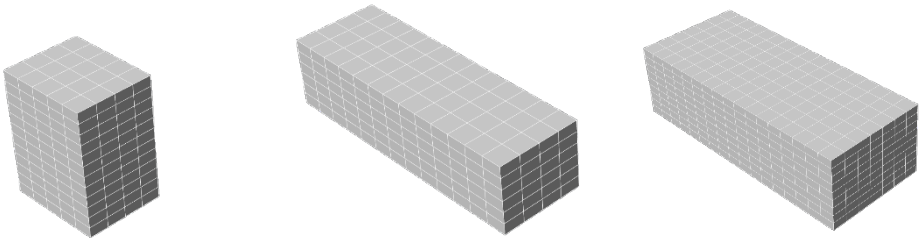
Calculated R^2 showed that the majority of the buildings are fitted with other buildings in the category. In the graph, the number of buildings in each box is presented. In total, 160 actual buildings were simulated, which 141 of them remained in the dataset, which means that 19 buildings showed an average $R^2 < 0.8$ and were removed from the process. A part of the calculated R^2 for the actual buildings is presented in Appendix C.

4.3.2 Hypothetical buildings

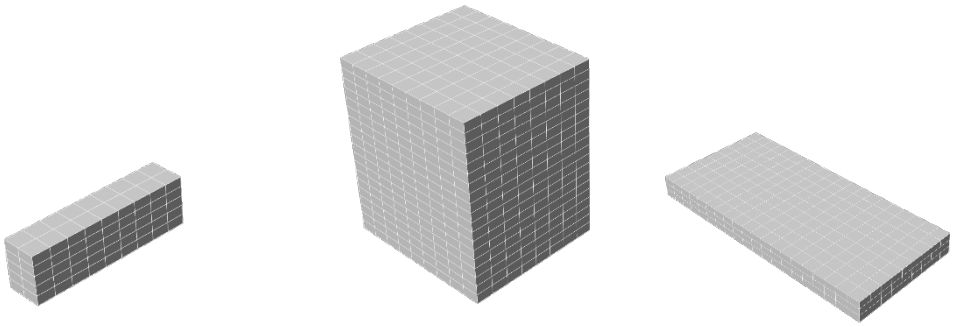
A few sample cases of randomly generated hypothetical buildings are presented in Figure 21 to show the variety of the modelling. COPs for the cooling systems were calculated as an hourly curve. The annual average of each curve is presented in Table 5, divided into different categories.

Table 5: Cooling load ratio to the EUI and average COPs

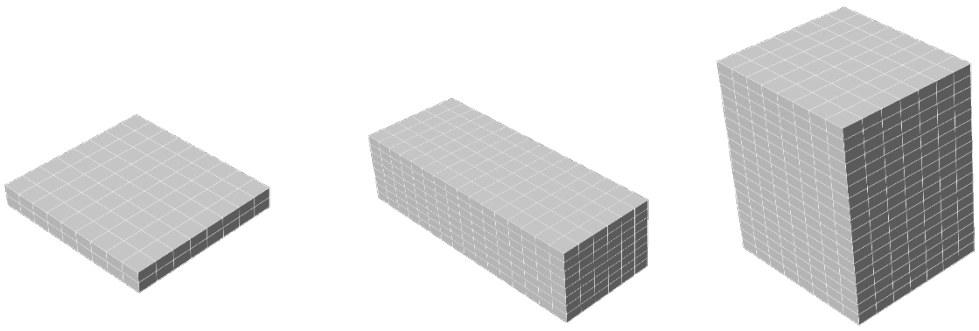
Type	Multifamily Housing				Office			
Vintage	pre-1940	1940-1979	1980-2006	2007-present	pre-1940	1940-1979	1980-2006	2007-present
Average COP	1.9	2	2	1.9	3	3.2	3.2	3



5 × 4 × 12	RC: 98%	13 × 4 × 6	RC: 85%	17 × 7 × 9	RC: 88%



2 × 9 × 5	RC: 83%	9 × 8 × 19	RC: 99%	20 × 9 × 3	RC: 58%



9 × 8 × 2	RC: 60%	16 × 6 × 8	RC: 87%	8 × 7 × 18	RC: 99%

Figure 21: Sample cases for randomly generated hypothetical buildings

The hourly trends were checked with R^2 in pairs to control their fitness. Figure 22 shows the R^2 for "Multifamily housing" and "Office" for heating and cooling load in each vintage. These R^2 were calculated in each category, and to simplify the graph, both heating and cooling loads are presented in one box. Each box contains values for 200 buildings from the cooling- and heating load for both types. Each box includes $200^2 \times 2$ values since each building was compared to all others (200^2) for heating and cooling loads ($\times 2$). R^2 is above 94% for all vintages, and as it was expected, the last vintage has higher values due to the better envelope quality and less sensitivity to the outdoor condition. Moreover, lower ranges of R^2 in office buildings means higher fitness of their hourly trend. It can be a result of less sensitivity of office buildings to the outdoor climate since office spaces are internal heat dominant.

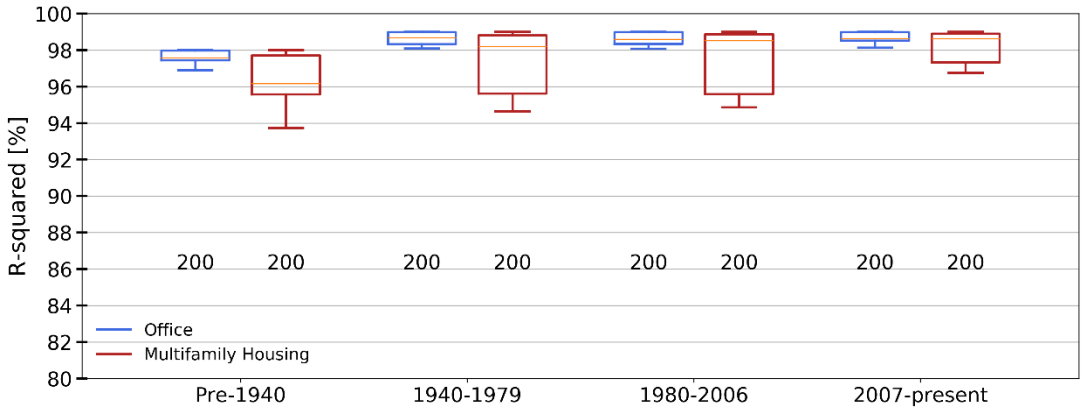


Figure 22: R^2 for each pair of hypothetical buildings in every category.

Numbers for each box represents the number of buildings in the box.

In this research, no clear relation between compactness and energy trend was found. Figure 23 shows a set of RCs for iterated actual buildings (60 cases) and hypothetical models (200 cases). In different ranges of compactness from normal to compact, R^2 showed a relatively small deviation from perfect fitness ($R^2 = 1$). As explained before, compactness and geometry as a building specification have a scaling effect on the entire hourly energy trend. Changes in geometry cause scaling in the entire data, which is eliminated in normalizing by sum. Some minor impacts on the hourly trend from the geometry of the building caused a reduction in R^2 from 1. The entire effects of the geometry were taken into account when the hourly energy trends were scaled up to the total measured energy use, which means that changing the energy use of building results in changing the total energy use with the same trend.

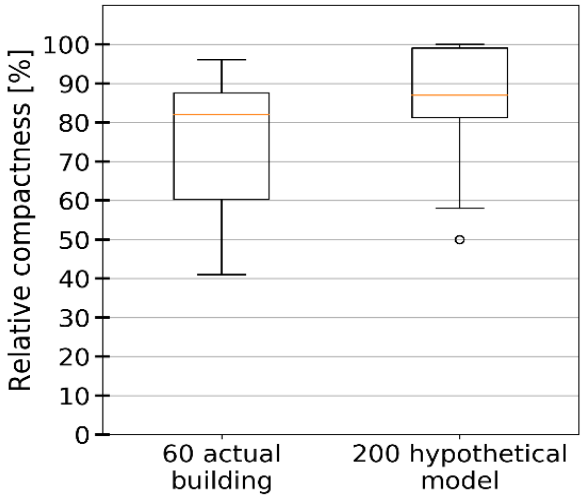


Figure 23: Compactness and RC for 60 actual buildings and 200 hypothetical models

As a result, compactness was not considered a category level because: (1) a proper classification for geometry needs a closer look and further investigation. (2) defining different geometrical parameters (e.g. orientation, exposure to the sun, neighbors blockage) in practice needs more input and collecting data which can be time-consuming and complicated contrary to the purpose of the research. (3) the current two levels of categories showed enough similarities in the hourly trend, which means more parameters cannot considerably enhance the categorization. Thus, it would be better to maintain the method with these two levels, which has an acceptable fitness in each category. Accordingly, in this research and on this level, compactness was not applied. However, it is discussed to keep the way open for further research on it, if it is required.

4.3.3 Domestic hot water

The hourly percentage of heating load for DHW is presented in Figure 24. The results are normalized by the total amount of the load. Hence, the curve shows the trend, regardless of the amount of energy. The total required heating loads for DHW were calculated based on the energy breakdown in Table 2, and the results were multiplied by the presented coefficient time series.

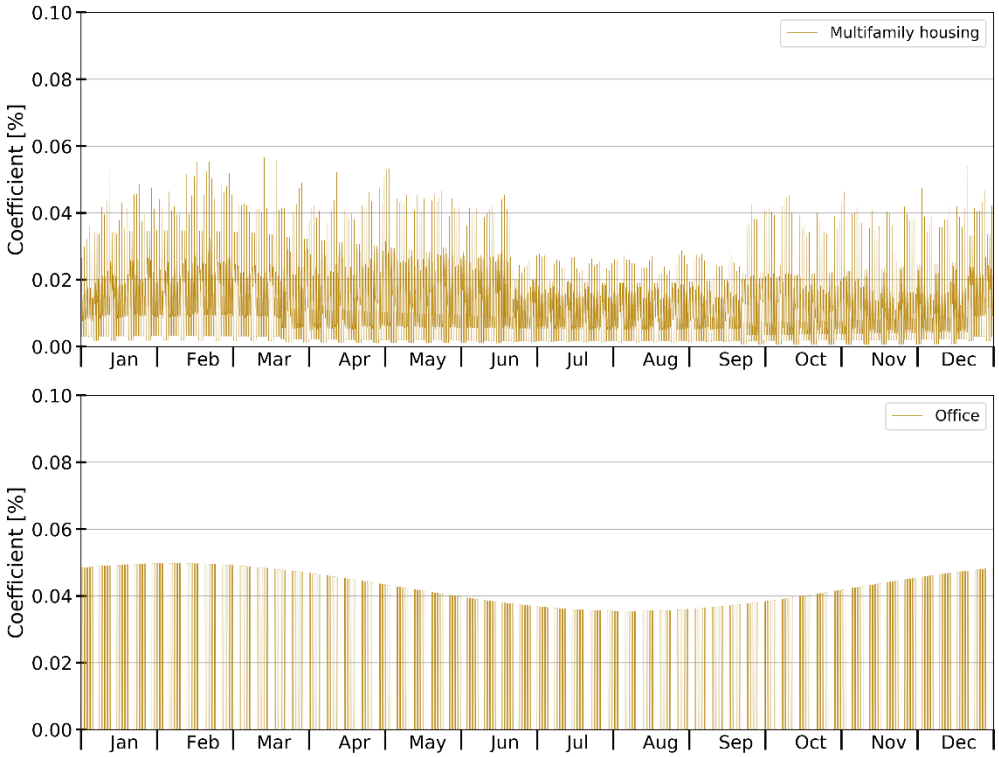


Figure 24: DHW heating load for Multifamily housing (top) and Office (bottom)

4.4 Reverse method

This section will present the generated reference models for the remaining building in the dataset based on the R^2 larger than 0.8. After that, the outputs of six sample cases are presented.

4.4.1 Reference models

The generated reference models are demonstrated in Figure 25. The presented time series shows the percentage of energy use for every single hour, and the sum of the time series is equal to one. In these graphs, a few points are vital to be considered. Heating- and cooling load are not related and not proportionally changed, and they are both defined by the total measured load. For instance, “Multifamily housing” in the “2007-present” vintage shows a high percentage of hourly heating loads (heating power); However, this percentage will be multiplied by the total measured heating load and then it is comparable to the cooling load. When the percentage of loads in the time series is zero, no loads are required. Accordingly, newer buildings show fewer hours for required loads. For example, for “Office” buildings, in “pre-1940”, heating demand is zero from almost the middle of May to late September: While in “2007-present”, it is zero from almost the middle of April to the middle of October. This difference can be a result of better building performance. As a result, the hourly energy trend

(reference model) can describe the energy performance of the building according to its category.

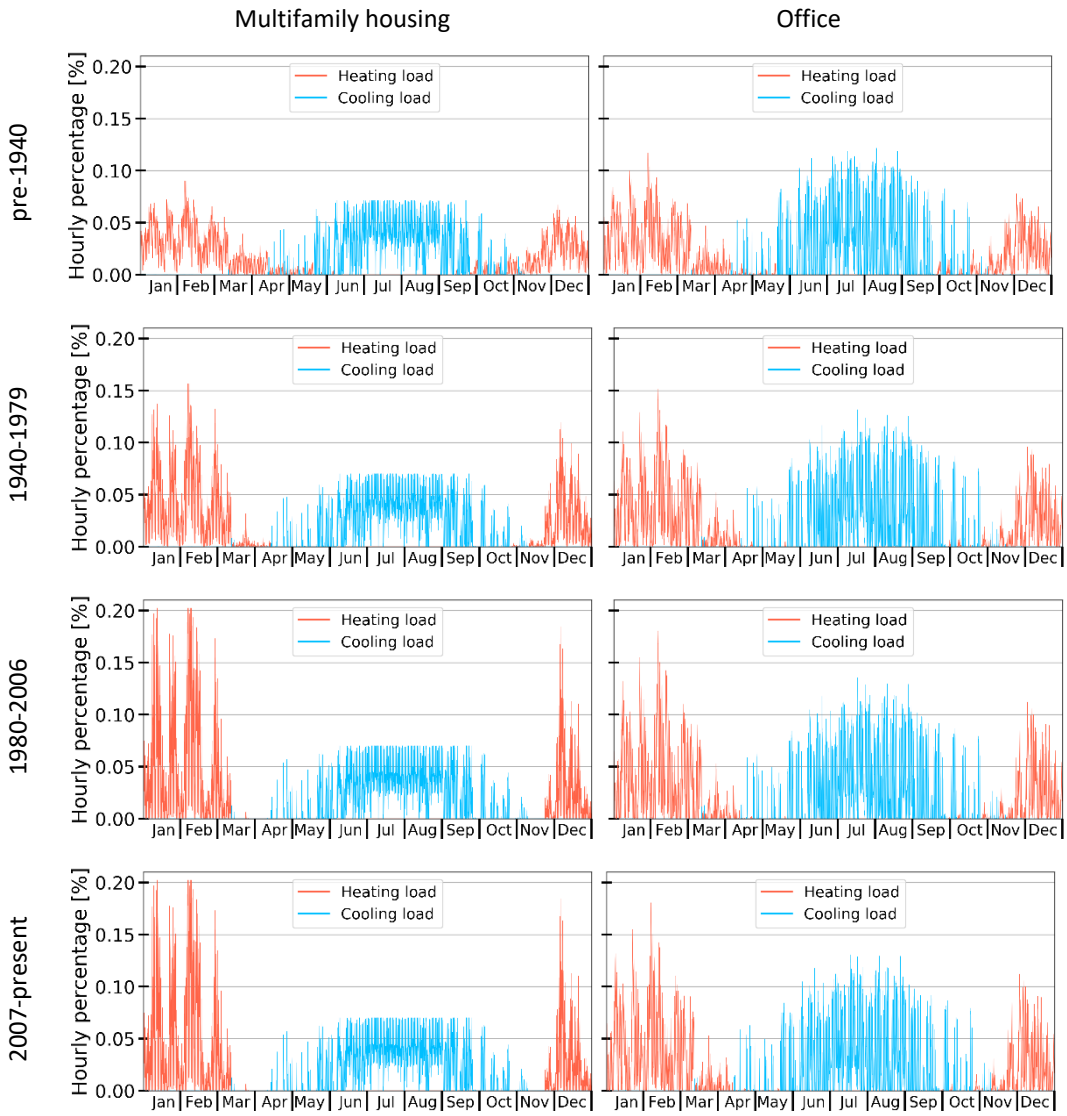


Figure 25: Generated reference models for the eight defined categories

4.4.2 Subjected buildings test

Six sample buildings are chosen from benchmarking dataset to present the results. The first part is the following text report (the grey box) in the Python interface and exported as a text file. It shows the building information to control if the correct building is picked from the dataset. Then the total demands are denoted to give a general overview of energy demands. The selected buildings are chosen among the most frequencies of the area between 5000 m^2

(54,000 ft^2) and 10,000 m^2 (108,000 ft^2). The energy use of the selected buildings is around the average of the entire data to present an overall view to the entire population.

This case shows a "Multifamily housing" for "pre-1940". Primitive information for energy use and demands are presented in the following text box, and a summary of calculations are shown in Table 6. Annual hourly loads for the building are presented in Figure 26, and the annual ground loads after applying SPF are illustrated in Figure 27. The duration diagram for ground loads and the adjusted peak powers are shown in Figure 28.

150-74th Street Brooklyn 11209 New York

---

building type: multifamily

BBL: 3059270024.0

BIN: 3148192

Area [m2]: 5298

EUI [kWh]: 1465957

Heating load [kWh]: 1062466 - 72.5%

Space heating [kWh]: 764975

Domestic hot water [kWh]: 297490

Cooling load [kWh]: 542404

---

vintage: pre-1940

reference model: multifamily_v1

Table 6: Results summary for Order number 80

Building footprint [m^2] ([ft^2])	Possible number of boreholes	Required number of boreholes	Required boreholes coverage
22×40 (72×131)	4×7	6×10	45%
Annual heating load coverage	**Peak heating power coverage**	**Annual cooling load coverage**	**Peak cooling power coverage**
87%	40%	67%	41%

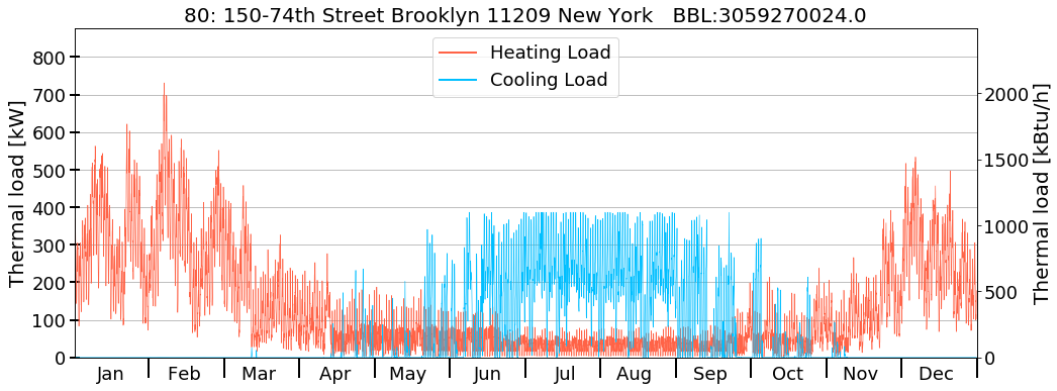


Figure 26: Predicted annual heating- and cooling load for building number 80

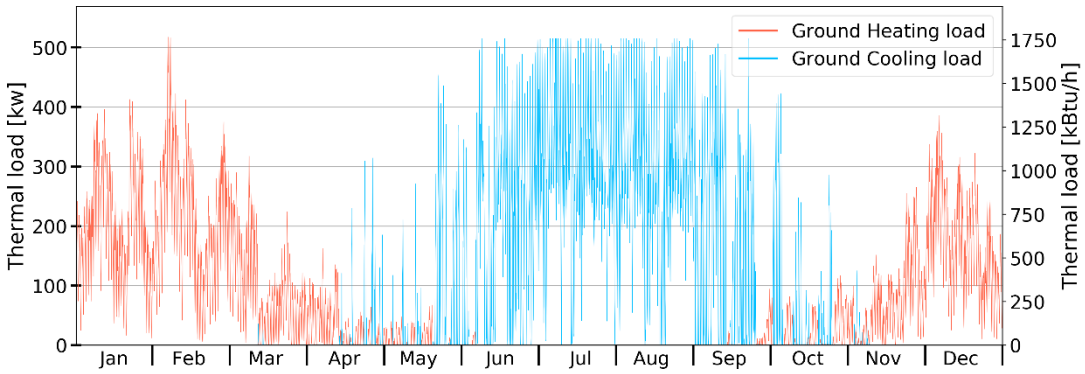


Figure 27: Ground loads for building number 80 after applying SPF

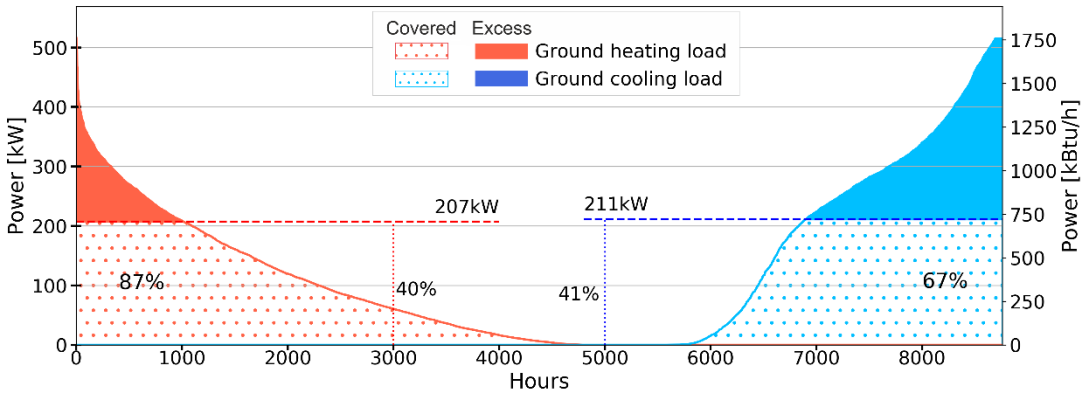


Figure 28: Duration diagram for heating and cooling loads for building number 80. The coverage percentage of loads and powers are mentioned in the graph.

The second sample building shows an “Office” for “1980-2006”. Primitive information for energy use and demands are presented in the following text box, and a summary of calculations are shown in Table 7. Annual hourly building loads are presented in Figure 29, and the annual ground loads after applying SPF are illustrated in Figure 30. The duration diagram for ground loads and the adjusted peak powers are shown in Figure 31.

110 East 55th street Manhattan 10022 New York

---

building type: office

BBL: 1013090066.0

BIN: 1036493

Area [m2]: 6689

EUI[kWh]: 1749282

Heating load [kWh]: 286594 - 16.4%

Space heating [kWh]: 243604

Domestic hot water [kWh]: 42989

Cooling load [kWh]: 822162

---

vintage: 1980-2006

reference model: office_v3

Table 7: Results summary for Order number 29983

Building footprint [m^2] ([ft^2])	Possible number of boreholes	Required number of boreholes	Required boreholes coverage
16×37 (52×121)	3×7	4×10	50%
Annual heating load coverage	**Peak heating power coverage**	**Annual cooling load coverage**	**Peak cooling power coverage**
93%	45%	14%	5%

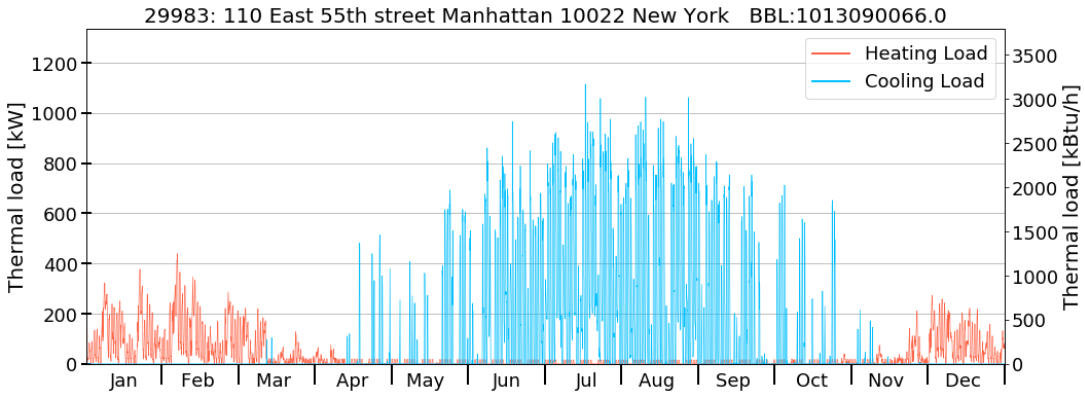


Figure 29: Predicted annual heating- and cooling load for Order number 29983

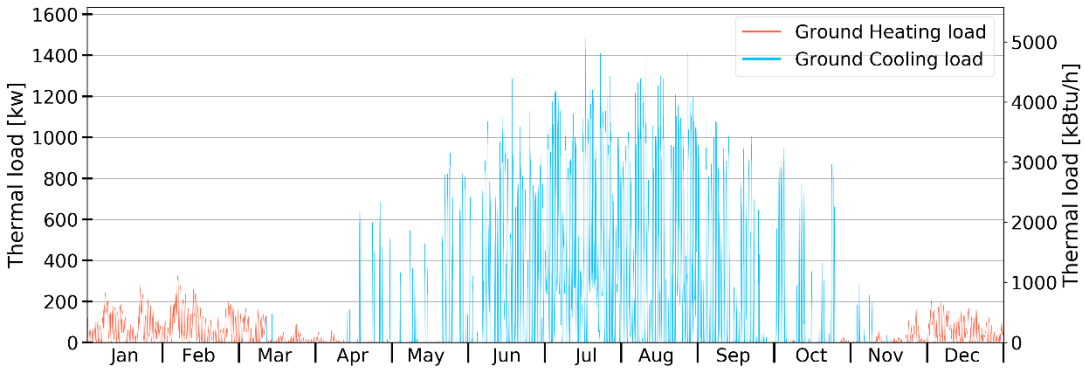


Figure 30: Ground loads for building number 29983 after applying SPF

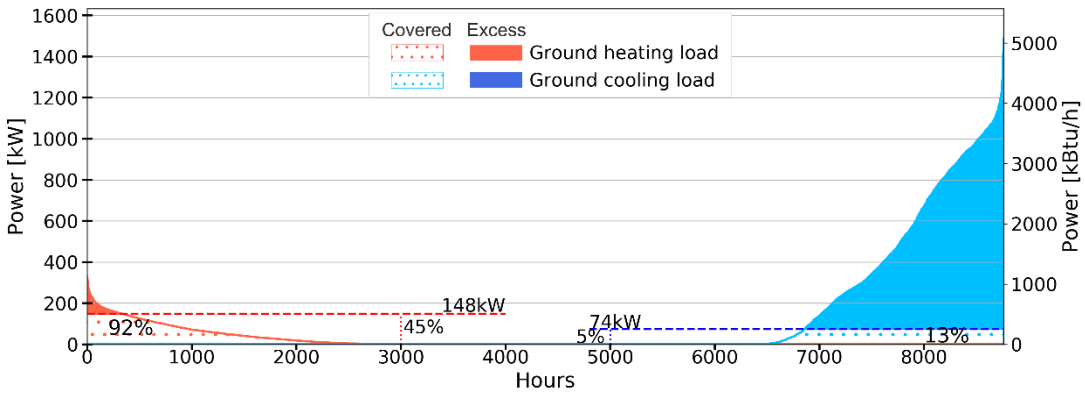


Figure 31: Duration diagram for heating and cooling loads for building number 29983. The coverage percentage of loads and powers are mentioned in the graph.

The third case is a “Multifamily housing” for “2007-present”. Primitive information for energy use and demands are presented in the following text box, and a summary of calculations are shown in Table 8. Annual hourly building loads are presented in Figure 32, and the annual ground loads after applying SPF are illustrated in Figure 33. The duration diagram for ground loads and the adjusted peak powers are shown in Figure 34.

545 Washington Avenue Brooklyn 11238 New York

---

building type: multifamily

BBL: 3020137503.0

BIN: 3397606

Area [m2]: 7369

EUI [kWh]: 1845879

Heating load [kWh]: 874463 - 47.4%

Space heating [kWh]: 629613

Domestic hot water [kWh]: 244849

Cooling load [kWh]: 719893

---

vintage: 2007-pres

reference model: multifamily_v4

Table 8: Results summary for Order number 30063

Building footprint [m^2] ([ft^2])	Possible number of boreholes	Required number of boreholes	Required boreholes coverage
21×43 (69×141)	4×7	6×17	27%
Annual heating load coverage	**Peak heating power coverage**	**Annual cooling load coverage**	**Peak cooling power coverage**
68%	30%	32%	17%

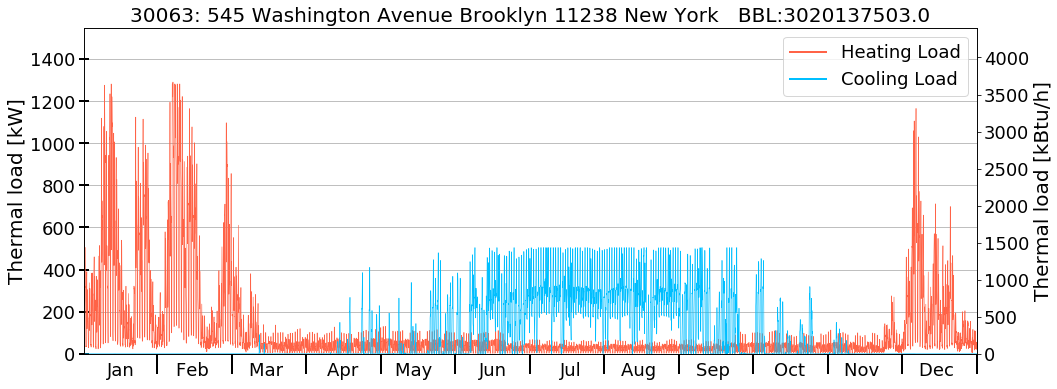


Figure 32: Predicted annual heating- and cooling load for building number 30063

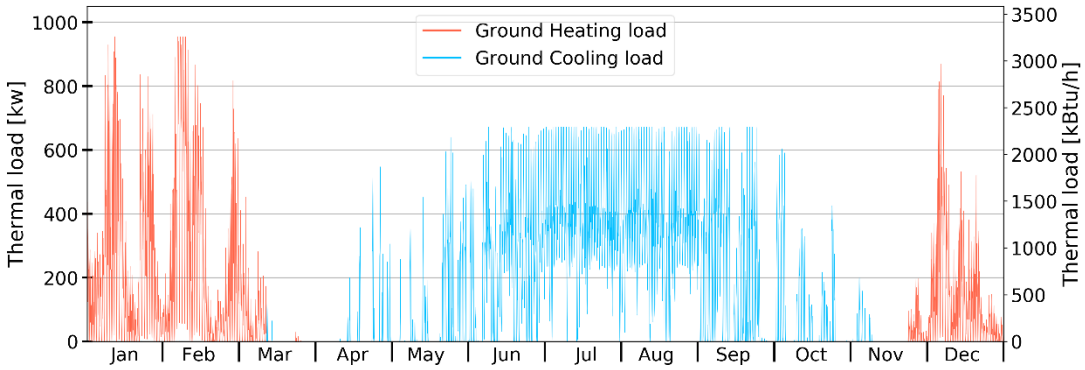


Figure 33: Ground loads for building number 30063 after applying SPF

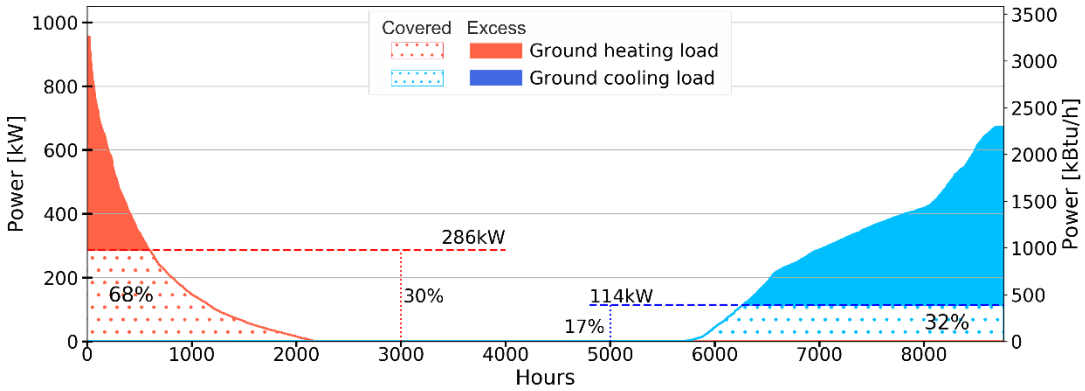


Figure 34: Duration diagram for heating and cooling loads for building number 30063. The coverage percentage of loads and powers are mentioned in the graph.

The fourth sample introduces an "Office" for "2007-present". Primitive information for energy use and demands are presented in the following text box, and a summary of calculations are shown in Table 9. Annual hourly building loads are presented in Figure 35, and the annual ground loads after applying SPF are illustrated in Figure 36. The duration diagram for ground loads and the adjusted peak powers are shown in Figure 37.

```
4055 10th Ave Manhattan 10034 New York
---------------------------------------
building type: office
BBL: 1022130001.0
BIN: 1064513
Area [m2]: 5642
---------------------------------------
EUI [kWh]: 1835085
Heating load [kWh]: 387557 - 21.1%
Space heating [kWh]: 329423
Domestic hot water [kWh]: 58133
Cooling load [kWh]: 770735
---------------------------------------
vintage: 2007-pres
reference model: office_v4
```

Table 9: Results summary for Order number 12080

Building footprint [m^2] ([ft^2])	Possible number of boreholes	Required number of boreholes	Required boreholes coverage
38×147* (125×480)	6×14	7×7	100%
Annual heating load coverage	**Peak heating power coverage**	**Annual cooling load coverage**	**Peak cooling power coverage**
100%	100%	21%	8%

* This land is a triangle with a base and height of 38×147 m^2 (125×480 ft^2). The possible grid of boreholes is considered a rectangular field. Other configurations can be used; however, the rectangular field can cover the required heating load. Thus, no more borehole is considered.

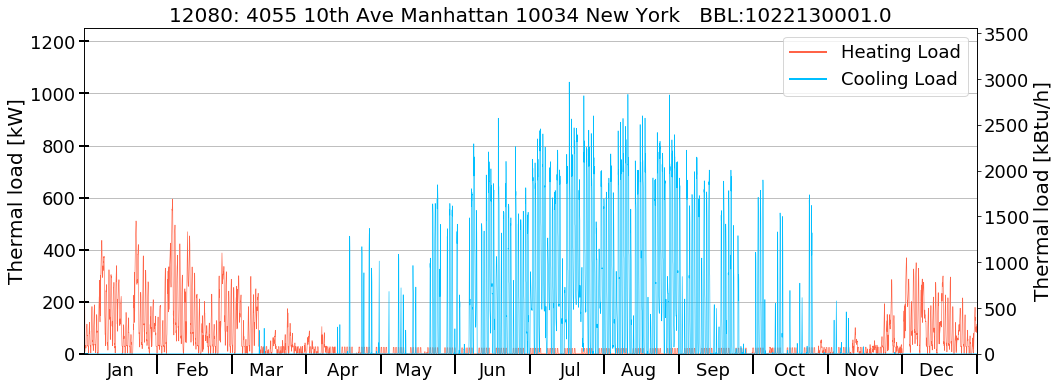


Figure 35: Predicted annual heating- and cooling load for building number 12080

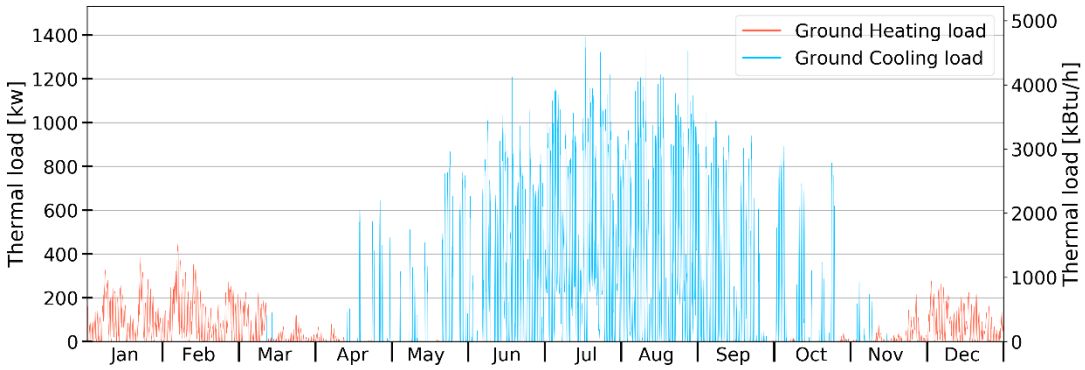


Figure 36: Ground loads for building number 12080 after applying SPF

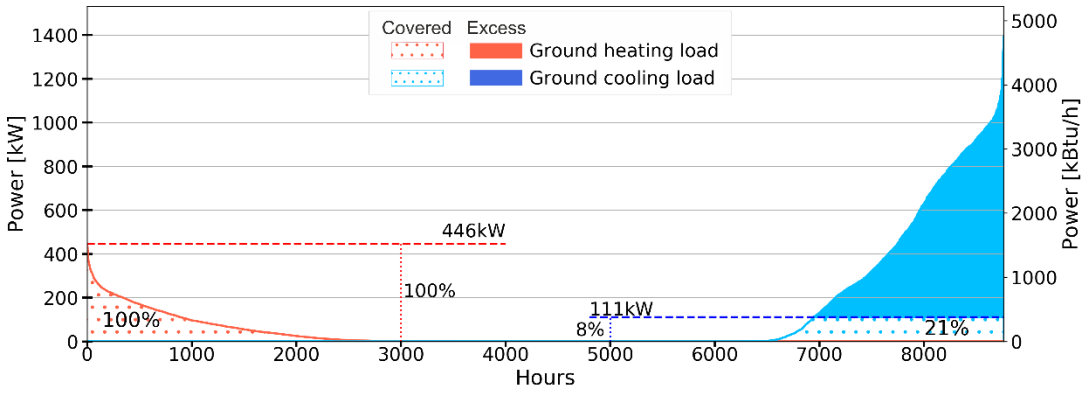


Figure 37: Duration diagram for heating and cooling loads for building number 12080. The coverage percentage of loads and powers are mentioned in the graph.

The fifth case shows an “Office” for “2007-present”. Primitive information for energy use and demands are presented in the following text box, and a summary of calculations are shown in Table 10. Annual hourly building loads are presented in Figure 38, and the annual ground loads after applying SPF are illustrated in Figure 39. The duration diagram for ground loads and the adjusted peak powers are shown in Figure 40.

16-16 Whitestone Expressway Queens 11357 New York

---

building type: office

BBL: 4041480065.0

BIN: 4099028

Area [m2]: 6489

EUI [kWh]: 2088042

Heating load [kWh]: 930568 - 44.6%

Space heating [kWh]: 790982

Domestic hot water [kWh]: 139585

Cooling load [kWh]: 876978

---

vintage: 2007-pres

reference model: office_v4

Table 10: Results summary for Order number 26965

Building footprint [m^2] ([ft^2])	Possible number of boreholes	Required number of boreholes	Required boreholes coverage
37×68 (118×223)	7×11	6×19	65%
Annual heating load coverage	**Peak heating power coverage**	**Annual cooling load coverage**	**Peak cooling power coverage**
97%	55%	47%	20%

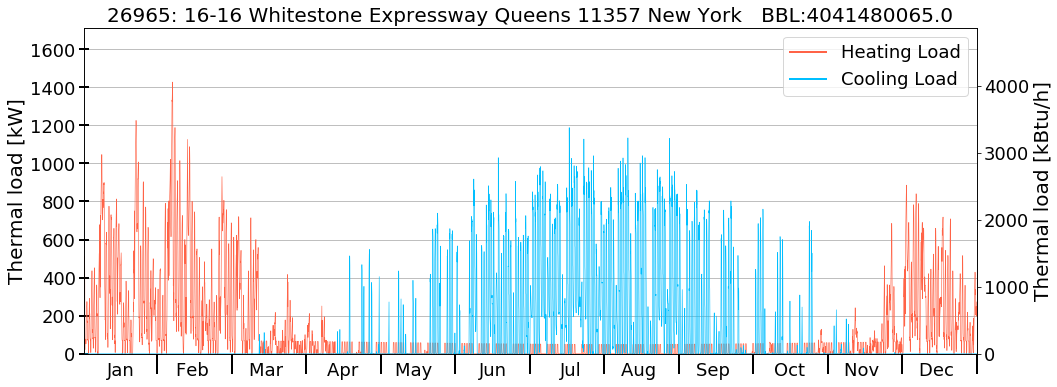


Figure 38: Predicted annual heating- and cooling load for building number 26965

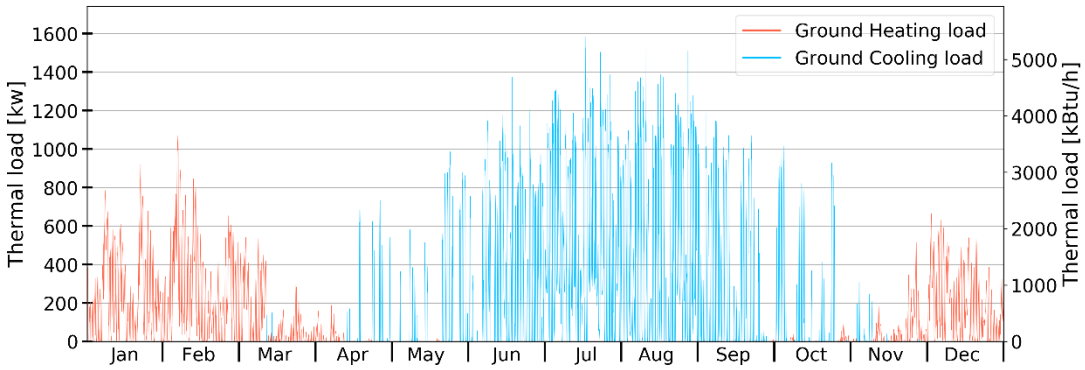


Figure 39: Ground loads for building number 26965 after applying SPF

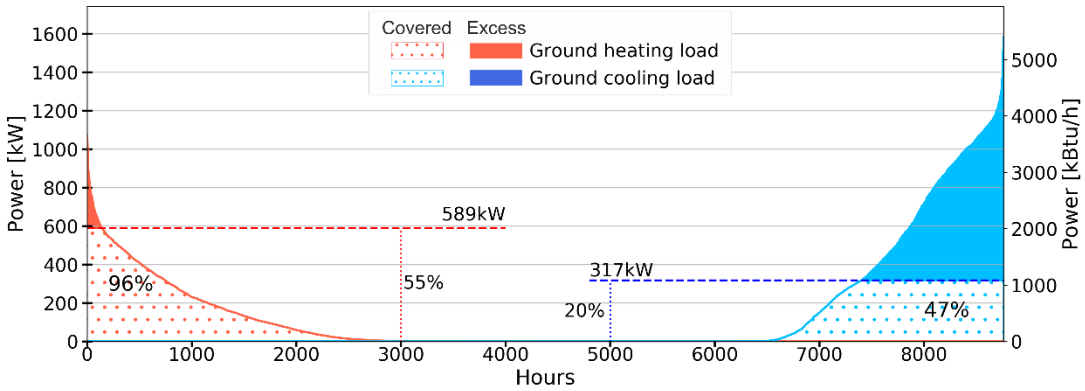


Figure 40: Duration diagram for heating and cooling loads for building number 26965. The coverage percentage of loads and powers are mentioned in the graph.

Sample case number six presents a “Multifamily housing” for “2007-present”. Primitive information for energy use and demands are presented in the following text box, and a summary of calculations are shown in Table 11. Annual hourly building loads are presented in Figure 41, and the annual ground loads after applying SPF are illustrated in Figure 42. The duration diagram for ground loads and the adjusted peak powers are shown in Figure 43.

300 West 128th Street Manhattan 10027 New York

---

building type: multifamily

BBL: 1019540036.0

BIN: 1089108

Area [m2]: 7026

EUI [kWh]: 1839861

Heating load [kWh]: 941683 - 51.2%

Space heating [kWh]: 678011

Domestic hot water [kWh]: 263671

Cooling load [kWh]: 717546

---

vintage: 2007-pres

reference model: multifamily_v4

Table 11: Results summary for Order number 4338

Building footprint [m^2] ([ft^2])	Possible number of boreholes	Required number of boreholes	Required boreholes coverage
20×40 (65×131)	4×7	6×19	25%
Annual heating load coverage	**Peak heating power coverage**	**Annual cooling load coverage**	**Peak cooling power coverage**
52%	18%	27%	14%

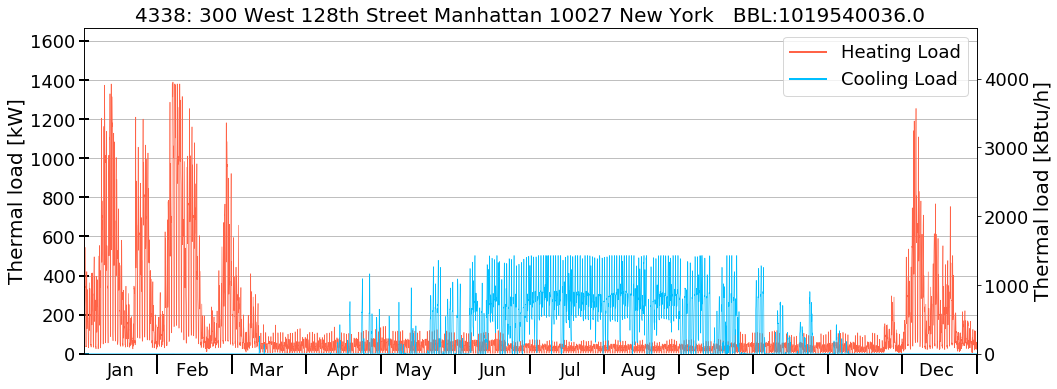


Figure 41: Predicted annual heating- and cooling load for building number 4338

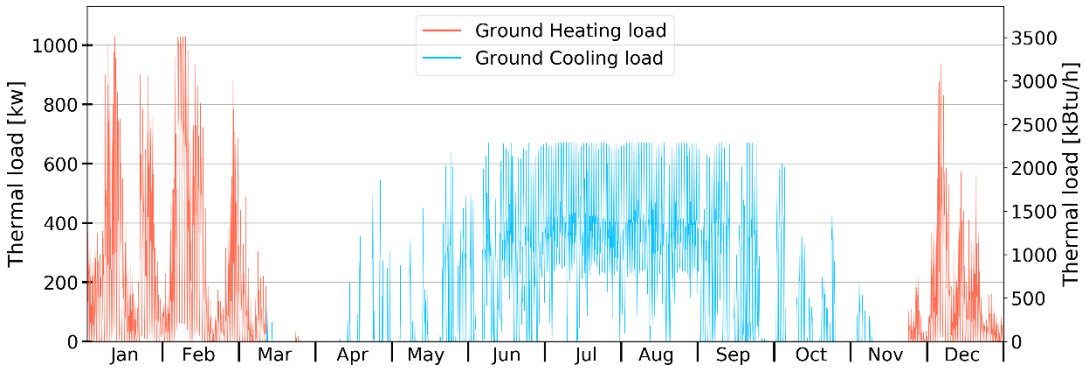


Figure 42: Ground loads for building number 4338 after applying SPF

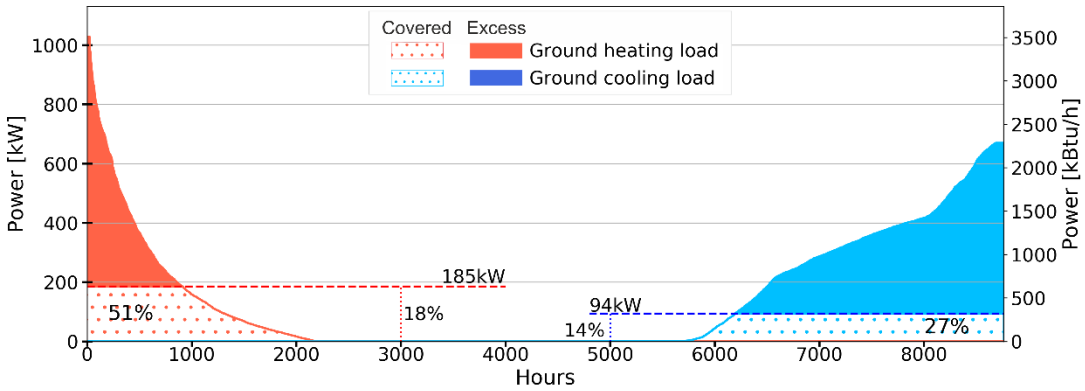


Figure 43: Duration diagram for heating and cooling loads for building number 4338. The coverage percentage of loads and powers are mentioned in the graph.

The presented samples outputs, firstly, showed the capabilities of the method in calculating the BTES system. Secondly, results showed that the BTES system could provide a considerable share of the required thermal loads of the buildings. The following text will explain the process in practice, then discuss the results.

The categorization system with two levels, consisting of the building type and vintage, does not need to collect extra input data. Both type and vintage of the buildings are provided in the NYC benchmarking dataset. Hence, having an identification number was enough to capture the building data from benchmarking dataset as the primitive information (the grey box). In the primitive information, the building category was defined, and the corresponding reference model was called from the method database. The required data was exported as CSV and TXT file with a suitable format for EED. From the input of the identification number to export the results, this process took less than a minute.

Thereafter, the process was continued in EED to design the boreholes field. The total loads were imported to EED's "hourly calculation" tool to calculate the total number of boreholes by the "optimization" tool. The possible number of boreholes in the land area was calculated based on the building footprint geometry captured from GIS. The ratio of the possible boreholes to the total required boreholes from EED showed the required boreholes coverage by land. Then the defined adjustment variable for the heating load was set to a percentage to calculate the thermal load coverage based on the possible number of boreholes in the available area in the building. The results of the adjusted heating loads were imported again to EED to check if the possible boreholes can provide the adjusted loads. The duration diagrams show the loads coverage for each case and their excess powers.

The results showed that establishing the geothermal system in the buildings can cover around 50% up to 100% heating loads. It is, however, depends mainly on the operation strategy and optimization of the system sizing, ratio of the extracted and injected heat, and peak load adjustment. For instance, in case number five, a reduction in heating peak load by 45% cover 96% of the entire annual heating load. The excess load happens in around 150 hours in the year. According to Figure 4, half of the buildings in the NYC benchmarking dataset have an area between 4,000 m^2 (43,000 ft^2) and 11,000 m^2 (118,500 ft^2). Establishing geothermal systems can cover a major part of their need for fossil fuels for heating purposes.

5 Conclusions

This study explored the effective parameters for building energy performance in NYC, including outdoor conditions, indoor conditions, and building specifications. The research was inspired by the publicly available metered energy data to perform a data-driven method. Thereafter, a reverse data-driven method was adopted to predict the annual hourly heating and cooling load from total measured use. The usual energy estimation methods are time-consuming but deviated from actual performance. Two sorts of data were used in this study, including benchmarking and simulated data.

NYC benchmarking database provides metered energy use data that can be used to calibrate energy estimations. This data characterizes the building stock based on the energy use, energy source, vintage, and type. Accordingly, combining the effective parameters on EPS and buildings characteristic from the benchmarking data divided buildings based on the type and vintage. Two sets of energy simulations were run to achieve the hourly load profile called the hourly energy trend (HET) in this research. The first one was performed on actual buildings, using GIS data, and calibrated by the annual energy use from benchmarking data. The second series of simulations were done on randomly generated hypothetical buildings with the same characteristics as the actual buildings. Outdoor condition, indoor condition, and building specifications are defined as the main components of the EPS. The outdoor condition was assigned to simulations as the weather file. While actual buildings used measured data for 2017, hypothetical models deployed TMY weather file. The rest of the settings were similar for both.

The coefficient of determination (R^2) was used to calculate how much hourly trend of buildings in each type-vintage category are similar. The results showed that buildings in each category have a similar hourly trend of thermal load. This similarity is partly related to the energy simulation settings: Indoor condition and building specifications were defined based on the standards and for different type and vintage. Hence in each category, buildings had similar settings. However, in reality, this similarity happens but to a lower degree.

The effects of geometry on the EPS were studied to assess their applicability in categorising buildings. Compactness and relative compactness were used to define the geometrical features of the buildings. Nevertheless, an investigation showed that several factors such as orientation, neighbors blockage, and horizontal or vertical compactness are involved. Thus, no clear relation was found between compactness and annual loads. The effects of the geometry are applied through calibration with the measured energy use. Defining all required inputs to identify the geometry parameters will make the process complicated and time-consuming for the user, in contrast with the research goal. Also, having complex inputs needs specific expertise, which should be avoided.

The HET of the buildings with an R^2 of larger than 0.8 on average were kept in the process. A reference model for each category was made from the median of the HET. These reference

models were created for the heating load (SH and DHW) and cooling load. Thereafter, the total measured data were multiplied by the percentage time series to generate the scaled hourly loads. These hourly loads are used for the primitive design of renewable systems such as geothermal and solar, mainly for feasibility studies.

The application of the method was tested on designing the BTES systems. The results showed that more than half of the heating loads of the buildings could be covered by the BTES system. This energy source replacement would significantly affect fossil fuel consumption as the main source of heating in NYC. This method's high speed and accuracy will help practitioners provide accurate feasibility studies for retrofit projects to establish BTES systems.

The entire workflow was processed in Python, which implied that there would be a high potential to integrate the workflow with other Python-based tools like pySAM or pygfunction. This method can be enhanced by updating reference models with more simulated- and measured data. Especially, it can be fed by energy simulations with generative or iterative algorithms. The most important limitation in this study was to access the measured data for calibrating and verifying the reference models. The required time for the simulations was also a severe problem due to the limited timeframe of this research. Future research can improve the accuracy of the predictions. More study about the geometry characteristics of the buildings, such as the compactness and orientation, can enhance the quality of the results if required.

Lastly, high-quality energy estimation is an essential step towards the implementations of renewable energy systems. Hourly energy estimation for the existing buildings is more complex than the design stage due to the complexity of building specifications measurements. In contrast, buildings under operation have metered energy data which can be helpful to estimate their energy performance on an hourly scale. BTES system as a renewable energy system shows potential in reducing fossil fuel burning as a noticeable source of GHG emissions considering the large share of fossil fuels for heating purposes in NYC.

References

[1] "World Cities Report 2020: The Value of Sustainable Urbanization | UN-Habitat." https://unhabitat.org/World 20Cities 20Report 202020 (accessed Feb. 08, 2021).

[2] Revi, A. *et al.*, "Part A: Global and Sectoral Aspects," in *Climate Change 2014: Impacts, Adaptation, and Vulnerability*, United Kingdom and New York, NY, USA: Cambridge University Press, pp. 535–612.

[3] Seto K. C. *et al.*, "Human Settlements, Infrastructure and Spatial Planning," in *Climate Change 2014: Mitigation of Climate Change*, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

[4] A. Kalua, "Urban Residential Building Energy Consumption by End-Use in Malawi," *Buildings*, vol. 10, no. 2, Art. no. 2, Feb. 2020, doi: 10.3390/buildings10020031.

[5] "World Energy Outlook 2017." IEA, 2017. Accessed: Jun. 05, 2021. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2017

[6] "THE 17 GOALS | Sustainable Development." https://sdgs.un.org/goals/goal17 (accessed May 28, 2021).

[7] United Nations, Department of Economic and Social Affairs, and Population Division, *World urbanization prospects: the 2018 revision*. 2019.

[8] K. Baier, V. Mataré, M. Liebenberg, and G. Lakemeyer, "Towards integrated intentional agent simulation and semantic geodata management in complex urban systems modeling," *Computers, Environment and Urban Systems*, vol. 51, pp. 47–58, May 2015, doi: 10.1016/j.compenvurbsys.2015.01.006.

[9] A. J. Arnfield, "Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island," *International Journal of Climatology*, vol. 23, no. 1, pp. 1–26, 2003, doi: 10.1002/joc.859.

[10] K. Lee, Y. Kim, H. C. Sung, J. Ryu, and S. W. Jeon, "Trend analysis of urban heat island intensity according to urban area change in asian mega cities," *Sustainability (Switzerland)*, vol. 12, no. 1, 2020, doi: 10.3390/su12010112.

[11] K. Javanroodi and V. Nik, "Impacts of Microclimate Conditions on the Energy Performance of Buildings in Urban Areas," *Buildings*, vol. 9, Aug. 2019, doi: 10.3390/buildings9080189.

[12] M. Santamouris, "Using cool pavements as a mitigation strategy to fight urban heat island—A review of the actual developments," *Renewable and Sustainable Energy Reviews*, vol. 26, pp. 224–240, Oct. 2013, doi: 10.1016/j.rser.2013.05.047.

[13] W. Yue, X. Liu, Y. Zhou, and Y. Liu, "Impacts of urban configuration on urban heat island: An empirical study in China mega-cities," *Science of the Total Environment*, vol. 671, pp. 1036–1046, 2019, doi: 10.1016/j.scitotenv.2019.03.421.

[14] "Inventory of New York City Greenhouse Gas Emissions in 2016." The City of New York, Dec. 2017.

[15] M. Tobias, “Steam Use in Large New York City Buildings.” https://www.ny-engineers.com/blog/steam-use-in-large-new-york-city-buildings (accessed Feb. 18, 2021).

[16] “Local laws of the city of New York for the year 2019.” NYC Council, New York, 2019.

[17] “GBEE - Greener, Greater Buildings Plan - LL84: Benchmarking.” https://www1.nyc.gov/html/gbee/html/plan/ll84.shtml (accessed Mar. 09, 2021).

[18] “NYC Energy & Water Performance Map.” http://energy.cusp.nyu.edu/ (accessed Mar. 09, 2021).

[19] “Clean Energy Standard (CES),” *NYSERDA*. https://www.nyserda.ny.gov/all-programs/programs/clean-energy-standard (accessed Feb. 24, 2021).

[20] “Renewable Generators and Developers,” *NYSERDA*. https://www.nyserda.ny.gov/All-Programs/Programs/Clean-Energy-Standard/Renewable-Generators-and-Developers (accessed Feb. 24, 2021).

[21] “To Fight Climate Change, New York City Will Push Skyscrapers To Slash Emissions,” *NPR.org*. https://www.npr.org/2019/04/23/716284808/new-york-city-lawmakers-pass-landmark-climate-measure (accessed Feb. 24, 2021).

[22] “New York City’s Roadmap to 80 x 50_Final.pdf.” Accessed: Feb. 24, 2021. [Online]. Available: https://www1.nyc.gov/assets/sustainability/downloads/pdf/publications/New%20York%20City's%20Roadmap%20to%2080%20x%2050_Final.pdf

[23] G. Mutani and V. Todeschi, “An urban energy atlas and engineering model for resilient cities,” *International Journal of Heat and Technology*, vol. 37, no. 4, pp. 936–947, 2019, doi: 10.18280/ijht.370402.

[24] A. T. D. Perera, K. Javanroodi, and V. M. Nik, “Climate resilient interconnected infrastructure: Co-optimization of energy systems and urban morphology,” *Applied Energy*, vol. 285, 2021, doi: 10.1016/j.apenergy.2020.116430.

[25] K. Javanroodi, M. Mahdavinejad, and V. M. Nik, “Impacts of urban morphology on reducing cooling load and increasing ventilation potential in hot-arid climate,” *Applied Energy*, vol. 231, pp. 714–746, 2018, doi: 10.1016/j.apenergy.2018.09.116.

[26] T. Ueno and A. Meier, “A method to generate heating and cooling schedules based on data from connected thermostats,” *Energy and Buildings*, vol. 228, 2020, doi: 10.1016/j.enbuild.2020.110423.

[27] S. Yfanti, N. Sakkas, and E. Karapidakis, “An event-driven approach for changing user behaviour towards an enhanced building’s energy efficiency,” *Buildings*, vol. 10, no. 10, pp. 1–14, 2020, doi: 10.3390/buildings10100183.

[28] P. Bacher, P. A. de Saint-Aubain, L. E. Christiansen, and H. Madsen, “Non-parametric method for separating domestic hot water heatingspikes and space heating,” *Energy and Buildings*, vol. 130, pp. 107–112, Aug. 2016, doi: http://dx.doi.org/10.1016/j.enbuild.2016.08.037.

[29] P. Huang, G. Huang, and Y. Wang, “HVAC system design under peak load prediction uncertainty using multiple-criterion decision making technique,” *Energy and Buildings*, vol. 91, pp. 26–36, 2015, doi: 10.1016/j.enbuild.2015.01.026.

[30] P. Huang, G. Huang, and G. Augenbroe, “Sizing HVAC systems under uncertainty in both load-demand and capacity-supply side from a life-cycle aspect,” *Science and Technology for the Built Environment*, vol. 23, pp. 367–381, Dec. 2016, doi: 10.1080/23744731.2016.1260409.

[31] R. Rezaee, J. Brown, J. Haymaker, and G. Augenbroe, “A novel inverse data driven modelling approach to performance-based building design during early stages,” *Advanced Engineering Informatics*, vol. 41, 2019, doi: 10.1016/j.aei.2019.100925.

[32] T. Maile, M. Fischer, and V. Bazjanac, “Building Energy Performance Simulation Tools - a Life-Cycle and Interoperable Perspective,” *Facil. Eng. (CIFE) Working Pap.*, vol. 107, Jan. 2007.

[33] F. C. McQuiston, J. D. Parker, and J. D. Spitler, *Heating, ventilating, and air conditioning: analysis and design*, 6th ed. Hoboken, N.J: John Wiley & Sons, 2005.

[34] “Energy performance of buildings — Energy needs for heating and cooling, internal temperatures and sensible and latent heat loads.” ISO, Jun. 2017.

[35] “Building Energy Modeling,” *Energy.gov, Office of Energy Efficiency & Renewable Energy*. https://www.energy.gov/eere/buildings/building-energy-modeling (accessed Apr. 28, 2021).

[36] J. L. M. Hensen and R. Lamberts, *Building Performance Simulation for Design and Operation*. Routledge, 2012.

[37] Y. Ding, Y. Shen, J. Wang, and X. Shi, “Uncertainty sources and calculation approaches for building energy simulation models,” 2015, vol. 78, pp. 2566–2571. doi: 10.1016/j.egypro.2015.11.283.

[38] S. Erba, F. Causone, and R. Armani, “The effect of weather datasets on building energy simulation outputs,” *Energy Procedia*, vol. 134, pp. 545–554, Oct. 2017, doi: 10.1016/j.egypro.2017.09.561.

[39] H. Zhao and F. Magoulès, “A review on the prediction of building energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, Aug. 2012, doi: 10.1016/j.rser.2012.02.049.

[40] Y. Wei *et al.*, “A review of data-driven approaches for prediction and classification of building energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1027–1047, Feb. 2018, doi: 10.1016/j.rser.2017.09.108.

[41] H. Do and K. S. Cetin, “Residential Building Energy Consumption: a Review of Energy Data Availability, Characteristics, and Energy Performance Prediction Methods,” *Curr Sustainable Renewable Energy Rep*, vol. 5, no. 1, pp. 76–85, Mar. 2018, doi: 10.1007/s40518-018-0099-3.

[42] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, “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*, vol. 123, pp. 168–178, Jun. 2014, doi: 10.1016/j.apenergy.2014.02.057.

[43] N. Li, Z. Yang, B. Becerik-Gerber, C. Tang, and N. Chen, "Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?," *Applied Energy*, vol. 159, pp. 196–205, Dec. 2015, doi: 10.1016/j.apenergy.2015.09.001.

[44] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1192–1205, Jan. 2018, doi: 10.1016/j.rser.2017.04.095.

[45] Q. Li, L. Gu, G. Augenbroe, C. F. J. Wu, and J. Brown, "Calibration of Dynamic Building Energy Models with Multiple Responses Using Bayesian Inference and Linear Regression Models," *Energy Procedia*, vol. 78, pp. 979–984, Nov. 2015, doi: 10.1016/j.egypro.2015.11.037.

[46] M. J. Jiménez and M. R. Heras, "Application of multi-output ARX models for estimation of the U and g values of building components in outdoor testing," *Solar Energy*, vol. 79, no. 3, pp. 302–310, Sep. 2005, doi: 10.1016/j.solener.2004.10.008.

[47] W. MacArthur, A. Mathur, and J. Zhao, "On-Line Recursive Estimation for Load Profile Prediction," *ASHRAE Transactions*, vol. 95, Jan. 1989.

[48] "ASHRAE Guideline 14-2014 Measurement of Energy, Demand and Water Savings." American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia, 2014.

[49] FEMP, *M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0*.

[50] D. Tanguay, "International Performance Measurement and Verification Protocol (IPMVP)," *Efficiency Valuation Organization (EVO)*. https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp (accessed Jan. 22, 2021).

[51] C. Fan, D. Yan, F. Xiao, A. Li, J. An, and X. Kang, "Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches," *Building Simulation*, vol. 14, no. 1, pp. 3–24, Feb. 2021, doi: 10.1007/s12273-020-0723-1.

[52] G. Ramos Ruiz and C. Fernández Bandera, "Validation of Calibrated Energy Models: Common Error," *https://www.mdpi.com/journal/energies*.

[53] J. Granderson, S. Touzani, C. Custodio, M. Sohn, S. Fernandes, and D. Jump, "Assessment of Automated Measurement and Verification (M&V) Methods," 2015.

[54] Y. Zhang, Z. O'Neill, B. Dong, and G. Augenbroe, "Comparisons of inverse modeling approaches for predicting building energy performance," *Building and Environment*, vol. 86, pp. 177–190, Apr. 2015, doi: 10.1016/j.buildenv.2014.12.023.

[55] R. Yokoyama, T. Wakui, and R. Satake, "Prediction of energy demands using neural network with model identification by global optimization," *Energy Conversion and*

Management, vol. 50, no. 2, pp. 319–327, Feb. 2009, doi: 10.1016/j.enconman.2008.09.017.

[56] A. H. Neto and F. A. S. Fiorelli, "Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption," *Energy and Buildings*, vol. 40, no. 12, pp. 2169–2176, Jan. 2008, doi: 10.1016/j.enbuild.2008.06.013.

[57] A. Kusiak, M. Li, and Z. Zhang, "A data-driven approach for steam load prediction in buildings," *Applied Energy*, vol. 87, no. 3, pp. 925–933, Mar. 2010, doi: 10.1016/j.apenergy.2009.09.004.

[58] D. B. Crawley, J. W. Hand, M. Kummert, and B. T. Griffith, "Contrasting the capabilities of building energy performance simulation programs," *Building and Environment*, vol. 43, no. 4, pp. 661–673, Apr. 2008, doi: 10.1016/j.buildenv.2006.10.027.

[59] ASHRAE, *2013 ASHRAE Handbook -- Fundamentals*, Cdr edition. Atlanta, GA.: ASHRAE, 2013.

[60] Y. Heo and V. M. Zavala, "Gaussian process modeling for measurement and verification of building energy savings," *Energy and Buildings*, vol. 53, pp. 7–18, Oct. 2012, doi: 10.1016/j.enbuild.2012.06.024.

[61] S. Thomas Ng, M. Skitmore, and K. F. Wong, "Using genetic algorithms and linear regression analysis for private housing demand forecast," *Building and Environment*, vol. 43, no. 6, pp. 1171–1184, Jun. 2008, doi: 10.1016/j.buildenv.2007.02.017.

[62] M. Krarti, J. F. Kreider, D. Cohen, and P. Curtiss, "Estimation of Energy Savings for Building Retrofits Using Neural Networks," *Journal of Solar Energy Engineering*, vol. 120, no. 3, pp. 211–216, Aug. 1998, doi: 10.1115/1.2888071.

[63] A. Dhar, T. A. Reddy, and D. E. Claridge, "Generalization of the Fourier Series Approach to Model Hourly Energy Use in Commercial Buildings," *Journal of Solar Energy Engineering*, vol. 121, no. 1, pp. 54–62, Feb. 1999, doi: 10.1115/1.2888143.

[64] P. de Wilde and D. Coley, "The implications of a changing climate for buildings," *Building and Environment*, vol. 55, pp. 1–7, Sep. 2012, doi: 10.1016/j.buildenv.2012.03.014.

[65] H. Wang and Q. Chen, "Impact of climate change heating and cooling energy use in buildings in the United States," *Energy and Buildings*, vol. 82, pp. 428–436, Oct. 2014, doi: 10.1016/j.enbuild.2014.07.034.

[66] Y. Chen, W. Moufouma-Okia, V. Masson-Delmotte, P. Zhai, and A. Pirani, "Recent Progress and Emerging Topics on Weather and Climate Extremes Since the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," *Annual Review of Environment and Resources*, vol. 43, no. 1, pp. 35–59, 2018, doi: 10.1146/annurev-environ-102017-030052.

[67] A. Baniassadi and D. Sailor, "Synergies and trade-offs between energy efficiency and resiliency to extreme heat – A case study," *Building and Environment*, vol. 132, Mar. 2018, doi: 10.1016/j.buildenv.2018.01.037.

[68] Y. Cui, D. Yan, T. Hong, C. Xiao, X. Luo, and Q. Zhang, “Comparison of typical year and multiyear building simulations using a 55-year actual weather data set from China,” *Applied Energy*, vol. 195, pp. 890–904, Jun. 2017, doi: 10.1016/j.apenergy.2017.03.113.

[69] S. Tsoka, K. Tolika, T. Theodosiou, and K. Tsikaloudaki, “Evaluation of stochastically generated weather datasets for building energy simulation,” *Energy Procedia*, vol. 122, pp. 853–858, Sep. 2017, doi: 10.1016/j.egypro.2017.07.449.

[70] “Global climate model,” *ScienceDaily*. https://www.sciencedaily.com/terms/global_climate_model.htm (accessed Mar. 15, 2021).

[71] T. E3P, “Typical Meteorological Year (TMY),” Sep. 02, 2016. https://e3p.jrc.ec.europa.eu/articles/typical-meteorological-year-tmy (accessed Mar. 15, 2021).

[72] “Weather Data Sources | EnergyPlus.” https://energyplus.net/weather/sources (accessed Mar. 15, 2021).

[73] R. Ma, B. Ren, D. Zhao, J. Chen, and Y. Lu, “Modeling urban energy dynamics under clustered urban heat island effect with local-weather extended distributed adjacency blocks,” *Sustainable Cities and Society*, vol. 56, 2020, doi: 10.1016/j.scs.2020.102099.

[74] S. Zoras, S. Veranoudis, and A. Dimoudi, “Micro- climate adaptation of whole building energy simulation in large complexes,” *Energy and Buildings*, vol. 150, pp. 81–89, 2017, doi: 10.1016/j.enbuild.2017.05.060.

[75] B. Bueno, A. Nakano, L. Norford, and C. Reinhart, “Urban Weather Generator - a Novel Workflow for Integrating Urban Heat Island Effect within Urban Design Process,” Dec. 2015.

[76] A. M. Rizwan, L. Y. C. Dennis, and C. Liu, “A review on the generation, determination and mitigation of Urban Heat Island,” *Journal of Environmental Sciences*, vol. 20, no. 1, pp. 120–128, Jan. 2008, doi: 10.1016/S1001-0742(08)60019-4.

[77] K. Zhang, R. Wang, C. Shen, and L. Da, “Temporal and spatial characteristics of the urban heat island during rapid urbanization in Shanghai, China,” *Environmental monitoring and assessment*, vol. 169, pp. 101–12, Oct. 2009, doi: 10.1007/s10661-009-1154-8.

[78] K. Oleson, G. Bonan, J. Feddema, and T. Jackson, “An examination of urban heat island characteristics in a global climate model,” *International Journal of Climatology*, vol. 31, pp. 1848–1865, Oct. 2011, doi: 10.1002/joc.2201.

[79] M. Gao, H. Shen, X. Han, H. Li, and L. Zhang, “Multiple timescale analysis of the urban heat island effect based on the Community Land Model: a case study of the city of Xi’an, China,” *Environmental Monitoring and Assessment*, vol. 190, Dec. 2017, doi: 10.1007/s10661-017-6320-9.

[80] H. Shen, L. Huang, L. Zhang, P. Wu, and C. Zeng, “Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China,” *Remote Sensing of Environment*, vol. 172, pp. 109–125, Jan. 2016, doi: 10.1016/j.rse.2015.11.005.

[81] H. Kulasekara and V. Seynulabdeen, “A Review of Geothermal Energy for Future Power Generation,” in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, Sep. 2019, pp. 223–228. doi: 10.1109/ICAEE48663.2019.8975470.

[82] M. Kheshti and X. Kang, “Geothermal Energy, a key point in Energy and Development in China,” Oct. 2015. doi: 10.1049/cp.2015.0576.

[83] H. Hemmatabady, J. Formhals, B. Welsch, D. O. Schulte, and I. Sass, “Optimized Layouts of Borehole Thermal Energy Storage Systems in 4th Generation Grids,” *Energies*, vol. 13, no. 17, Art. no. 17, Jan. 2020, doi: 10.3390/en13174405.

[84] H. Skarphagen, D. Banks, B. S. Frengstad, and H. Gether, “Design Considerations for Borehole Thermal Energy Storage (BTES): A Review with Emphasis on Convective Heat Transfer,” *Geofluids*, vol. 2019, p. e4961781, Apr. 2019, doi: 10.1155/2019/4961781.

[85] J. Claesson and P. Eskilson, “Conductive heat extraction to a deep borehole: Thermal analyses and dimensioning rules,” *Energy*, vol. 13, no. 6, pp. 509–527, Jun. 1988, doi: 10.1016/0360-5442(88)90005-9.

[86] J. Claesson and S. Javed, “An Analytical Method to Calculate Borehole Fluid Temperatures for Time-scales from Minutes to Decades (ML-11-C034),” *ASHRAE Transactions*, vol. 117, pp. 279–288, Jan. 2011.

[87] S. Javed, *Development of Modelling and Simulation tools for Geothermal Basements and Deep Foundations in Soft Clays*. 2016.

[88] C. Naranjo-Mendoza, M. Oyinlola, A. Wright, and R. Greenough, “Application of the superposition technique in conduction heat transfer for analysing arrays of shallow boreholes in ground source heat pump systems,” Sep. 2019.

[89] B. Sanner, G. Phetteplace, and G. Hellström, “Introduction to Computer Models for Geothermal Heat Pumps,” p. 9.

[90] P. Gonzalez and J. Angel, *Understanding numerically generated g-functions: A study case for a 6x6 borehole field*. 2013. Accessed: May 11, 2021. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-126017

[91] “Northeastern United States,” *Wikipedia*. Apr. 27, 2021. Accessed: Apr. 27, 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Northeastern_United_States&oldid=1020114902

[92] “New York | Capital, Map, Population, History, & Facts,” *Encyclopedia Britannica*. https://www.britannica.com/place/New-York-state (accessed Mar. 31, 2021).

[93] "New York City," *Wikipedia*. Mar. 30, 2021. Accessed: Mar. 30, 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=New_York_City&oldid=1015043116

[94] McLaren Engineering Group, "Geotechnical Feasibility Study for Beresford, 233 Central Park West, New York, NY." Jul. 2020.

[95] "UGC-Benchmarking-Report-101617-FINAL.pdf." Accessed: May 06, 2021. [Online]. Available: http://www.nyc.gov/html/gbee/downloads/pdf/UGC-Benchmarking-Report-101617-FINAL.pdf

[96] "NYC Property Address Directory (PAD)," *NYC, Department of City Planning*, May 2021, [Online]. Available: https://www1.nyc.gov/assets/planning/download/zip/data-maps/open-data/pad21b.zip

[97] N. D. A. Team, "Hourly/Sub-Hourly Observational Data Map | GIS Maps," *National Centers for Environmental Information (NCEI)*. https://www.ncdc.noaa.gov/maps/ (accessed May 04, 2021).

[98] "New York Standard Approach for Estimating Energy Savings from Energy Efficiency Programs – Residential, Multi-Family, and Commercial/Industrial Measures." New York State Joint Utilities, Apr. 2019. Accessed: May 06, 2021. [Online]. Available: https://www3.dps.ny.gov/W/PSCWeb.nsf/All/72C23DECFF52920A85257F1100671BDD

[99] S. Goel and M. I. Rosenberg, "ANSI/ASHRAE/IES Standard 90.1-2010 Performance Rating Method Reference Manual," Pacific Northwest National Lab. (PNNL), Richland, WA (United States), PNNL-25130, May 2016. doi: 10.2172/1253882.

[100] K. Javanroodi, V. M. Nik, and B. Adl-Zarrabi, "A multi-objective optimization framework for designing climate-resilient building forms in urban areas," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 588, p. 032036, Nov. 2020, doi: 10.1088/1755-1315/588/3/032036.

[101] A. Mohamed, "Understanding building compactness entity, definition and concept of assessment," *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies.*, vol. 7, no. 3, Sep. 2016, doi: DOI: 10.13140/RG.2.2.15820.82569.

[102] B. D'Amico and F. Pomponi, "A compactness measure of sustainable building forms," *Royal Society Open Science*, vol. 6, no. 6, p. 181265, doi: 10.1098/rsos.181265.

[103] "§ 57-03 Standards for Drilling and Excavation.," *American Legal Publishing Corporation*. https://codelibrary.amlegal.com/codes/newyorkcity/latest/NYCrules/0-0-0-120183 (accessed May 05, 2021).

Appendix A

Building types in the dataset and their frequencies

	Property Type	Count		Property Type	Count
1	Multifamily housing	12,579	34	Refrigerated Warehouse	28
2	Office	1,859	35	Museum	26
3	K-12School	1,386	36	Other-Recreation	26
4	College/University	426	37	Movie Theater	25
5	Hotel	385	38	Court house	22
6	Residence Hall/Dormitory	329	39	Strip Mall	20
7	Non-Refrigerated Warehouse	295	40	Prison/Incarceration	19
8	Other	152	41	Residential Care Facility	19
9	Senior Care Community	148	42	Automobile Dealership	17
10	Retail Store	144	43	Laboratory	17
11	Manufacturing/Industrial Plant	126	44	Social/Meeting Hall	17
12	Pre-school/Daycare	119	45	Adult Education	13
13	Repair Services	117	46	Financial Office	12
14	Self-Storage Facility	111	47	Wastewater Treatment Plant	11
15	Library	107	48	Food Service	9
16	Distribution Center	97	49	Bank Branch	8
17	Parking	95	50	OutpatientRehabilitation/PhysicalTherapy	8
18	Police Station	80	51	Wholesale Club/Super center	8
19	Hospital (General Medical & Surgical)	72	52	Data Center	7
20	Worship Facility	72	53	Enclosed Mall	6
21	Other-Lodging/Residential	69	54	Transportation Terminal/Station	6
22	Other-Education	60	55	Convention Center	5
23	Medical Office	58	56	Other-Services	5
24	Fire Station	55	57	Other-Utility	4
25	Supermarket/Grocery Store	52	58	Ambulatory Surgical Center	3
26	Urgent Care/Clinic/Other Outpatient	50	59	Ice/Curling Rink	2
27	Mixed Use Property	47	60	Stadium (Open)	2
28	Fitness Center/Health Club/Gym	41	61	Veterinary Office	2
29	Other-Specialty Hospital	40	62	Zoo	2
30	Other-Public Services	37	63	Convenience Store without GasStation	1
31	Other-Entertainment/Public Assembly	34	64	Mailing Center/Post Office	1
32	Other-Mall	29	65	Other-Technology/Science	1
33	Performing Arts	28	66	Restaurant	1

Appendix B

Office energy simulation settings

```
#Define variables
req_data = ['Year Built', 'Primary Property Type - Self Selected', 'Self-Reported Gross Floor Area (ft²)',
        'Site EUI (kBtu/ft²)', 'BBL - 10 digits', 'NYC Building Identification Number (BIN)', 'Borough',
        'Address 1 (self-reported)', 'Postal Code']
heating_energy = ['Fuel Oil #1 Use (kBtu)', 'Fuel Oil #2 Use (kBtu)', 'Fuel Oil #4 Use (kBtu)',
            'Fuel Oil #5 & 6 Use (kBtu)', 'Diesel #2 Use (kBtu)', 'Propane Use (kBtu)',
            'District Steam Use (kBtu)', 'District Hot Water Use (kBtu)', 'Natural Gas Use (kBtu)']
clg_ratio = [0.37, 0.45, 0.45, 0.39, 0.44, 0.47, 0.47, 0.42] #clg/EUI simulated
vintages = ['1:pre_war', '2:prior_1979', '3:1980-2006', '4:2007-pres']

#Input desired building via Order
order = int(input('Please input order:'))

#Capture coresponding data from benchmarking dataset
ind = [benchmark[benchmark['Order']==order].index][0][0]
yearbuilt = benchmark[req_data].iloc[ind][0]
building_type = benchmark[req_data].iloc[ind][1].lower().split(' ')[0]
area = benchmark[req_data].iloc[ind][2]
EUI_ft2 = benchmark[req_data].iloc[ind][3]
EUI_mes = EUI_ft2 * area
#Heating load
htg_load = int(benchmark[heating_energy].iloc[ind].sum())*.75
htg_ratio = round((htg_load/EUI_mes),3)*100
#DHW
if building_type=='office': dhw_load=(dhw_of*(htg_load*.15))
else: dhw_load=(dhw_mf*(htg_load*.28))
#Space heating
spc_htg = (htg_load - dhw_load.sum().sum())
#Vintage
if yearbuilt<1941: vintage = vintages[0]
elif yearbuilt<1980: vintage = vintages[1]
elif yearbuilt<2007: vintage = vintages[2]
else: vintage = vintages[3]
address = benchmark[req_data].iloc[ind][7]+' '+benchmark[req_data].iloc[ind][6]+\
                ' '+str(benchmark[req_data].iloc[ind][8])+' New York'
BBL = int(benchmark[req_data].iloc[ind][4])
BIN = int(benchmark[req_data].iloc[ind][5])
reg_mod = building_type+'_v'+vintage[0]
#Cooling load
```

```
if building_type=='office': clg_load=clg_ratio[4:][int(vintage[0])-1]*EUI_mes
else: clg_load=clg_ratio[:4][int(vintage[0])-1]*EUI_mes

#Output
output = address+'\n'+'-'*len(address)+'\nbuilding type: '+building_type+'\nBBL: '+str(BBL)+'\nBIN: '+str(BIN)+\
     '\n'+'-'*len(address)+'\nEUI[kBtu]: '+str(int(EUI_mes))+'\nHeating demand[kBtu]: '+str(htg_load)+' - '+\
      str(htg_ratio)+'%\nSpace heating[kBtu]: '+str(int(spc_htg))+'\nDomestic hot water[kBtu]: '+\
      str(int(dhw_load.sum().sum()))+'\nCooling load[kBtu]: '+str(int(clg_load))+'\n'+'-'*len(address)+\
     '\nvintage: '+vintage+'\nregression model: '+reg_mod
print (output)
with open('fig\\test\\'+str(order)+'.txt', 'w') as txt:
   txt.write(output)

#Call regression model
regression = pd.read_csv('reg_models\\reg_mod_'+reg_mod+'.csv').set_index('Unnamed: 0').set_index(arange(8760))
predict = regression.copy()
h = regression.columns
predict.rename(columns={h[0]:'HL', h[1]:'CL'}, inplace=True)
predict['HL'] = regression.iloc[:,0] * spc_htg + dhw_load[0]
predict['CL'] = regression.iloc[:,1] * clg_load
```

Appendix C

R^2

Multifamily housing

Pre-1940

Orders	22765	14660	33380	23651	14873	9280	9380	7115	13570	27594	5410	14164	18151	9541	15332
22765	1	0.87	0.89	0.99	0.87	0.96	0.95	0.97	0.91	0.89	0.93	0.98	0.78	0.97	0.93
14660	0.91	1	0.99	0.91	0.99	0.88	0.98	0.94	1	0.98	0.99	0.96	0.53	0.86	0.98
33380	0.92	0.99	1	0.92	0.99	0.87	0.97	0.96	0.99	1	0.98	0.97	0.56	0.86	0.96
23651	0.99	0.88	0.88	1	0.88	0.97	0.96	0.96	0.91	0.88	0.94	0.98	0.77	0.98	0.94
14873	0.91	0.99	0.99	0.91	1	0.88	0.98	0.96	0.99	0.98	0.99	0.97	0.54	0.86	0.97
9280	0.95	0.82	0.79	0.96	0.81	1	0.94	0.91	0.86	0.78	0.9	0.94	0.77	0.99	0.92
9380	0.96	0.97	0.96	0.97	0.97	0.95	1	0.98	0.98	0.95	1	0.99	0.64	0.94	0.99
7115	0.97	0.93	0.96	0.97	0.95	0.93	0.98	1	0.95	0.96	0.97	0.99	0.68	0.93	0.94
13570	0.93	1	0.99	0.93	0.99	0.91	0.99	0.96	1	0.97	0.99	0.97	0.56	0.88	0.99
27594	0.92	0.98	1	0.92	0.98	0.86	0.96	0.97	0.98	1	0.97	0.96	0.57	0.85	0.95
5410	0.95	0.99	0.98	0.95	0.99	0.93	1	0.97	0.99	0.97	1	0.99	0.61	0.91	0.99
14164	0.98	0.95	0.96	0.98	0.96	0.95	0.99	0.99	0.97	0.95	0.99	1	0.69	0.95	0.98
18151	0.73	0.22	0.24	0.72	0.23	0.75	0.5	0.57	0.31	0.26	0.41	0.58	1	0.82	0.43
9541	0.97	0.78	0.77	0.97	0.78	0.99	0.92	0.91	0.83	0.76	0.88	0.94	0.83	1	0.88
15332	0.94	0.98	0.96	0.95	0.96	0.94	0.99	0.95	0.99	0.94	0.99	0.98	0.6	0.92	1

1940-1979

Orders	9185	4429	5983	4006	20668	7298	9879	2095	4357	18581	8973
9185	1	0.71	0.8	0.87	0.67	0.96	0.56	0.73	0.52	0.8	0.64
4429	0.8	1	0.43	0.56	0.33	0.86	0.21	0.99	0.17	0.42	0.29
5983	0.76	0	1	0.98	0.94	0.72	0.87	0	0.86	0.98	0.92
4006	0.84	0.22	0.98	1	0.93	0.82	0.85	0.21	0.83	0.97	0.9
20668	0.58	0	0.94	0.93	1	0.54	0.98	0	0.98	0.96	0.99
7298	0.97	0.81	0.77	0.86	0.65	1	0.52	0.81	0.49	0.74	0.6
9879	0.47	0	0.87	0.85	0.98	0.41	1	0	1	0.92	0.99
2095	0.82	0.99	0.46	0.58	0.34	0.87	0.23	1	0.19	0.45	0.3
4357	0.42	0	0.87	0.84	0.98	0.37	1	0	1	0.9	0.99
18581	0.75	0	0.98	0.97	0.96	0.68	0.92	0	0.9	1	0.95
8973	0.56	0	0.92	0.9	0.99	0.5	0.99	0	0.99	0.95	1

1980-2006

Orders	8685	21240	1587	14417	9888	6400	27705	6528	5127	1783
8685	1	0.95	0.66	0.99	0.96	0.91	0.86	0.99	0.75	0.98
21240	0.95	1	0.78	0.94	0.99	0.97	0.97	0.96	0.9	0.97
1587	0.58	0.76	1	0.54	0.7	0.82	0.79	0.59	0.92	0.63
14417	0.99	0.95	0.61	1	0.97	0.91	0.87	1	0.73	0.99
9888	0.95	0.99	0.73	0.96	1	0.97	0.96	0.97	0.86	0.97
6400	0.88	0.97	0.82	0.89	0.96	1	0.96	0.92	0.91	0.92
27705	0.84	0.97	0.8	0.85	0.96	0.97	1	0.87	0.95	0.89
6528	0.99	0.96	0.64	1	0.97	0.93	0.88	1	0.75	0.99
5127	0.71	0.9	0.93	0.7	0.85	0.92	0.95	0.73	1	0.77
1783	0.98	0.97	0.67	0.99	0.97	0.93	0.9	0.99	0.78	1

2007-present

Orders	18771	30342	16904	25654	29630
18771	1	0.95	0.16	0.99	0.95
30342	0.96	1	0	0.97	0.96
16904	0.23	0	1	0.21	0.18
25654	0.99	0.97	0.14	1	0.98
29630	0.96	0.96	0.19	0.98	1

Office

1980-2006

Orders	10707	2805	1989	11966	20074	2177	7309	15776	6416	3275	28403	5754	885	17145
10707	1	0.93	0.99	0.96	0.96	0.66	0	0.95	0.96	0.97	0.99	0.99	0.97	0.99
2805	0.93	1	0.9	0.83	0.81	0.84	0	0.98	0.92	0.83	0.9	0.91	0.85	0.96
1989	0.99	0.9	1	0.99	0.98	0.59	0	0.92	0.94	0.99	0.99	1	0.99	0.98
11966	0.97	0.84	0.99	1	1	0.49	0	0.86	0.9	1	0.98	0.98	1	0.94
20074	0.96	0.83	0.98	1	1	0.48	0	0.85	0.9	1	0.98	0.98	1	0.94
2177	0.64	0.83	0.56	0.44	0.41	1	0.35	0.83	0.74	0.44	0.6	0.58	0.49	0.7
7309	0	0.08	0	0	0	0.56	1	0.1	0.07	0	0	0	0	0
15776	0.95	0.98	0.91	0.85	0.83	0.83	0	1	0.96	0.85	0.92	0.92	0.87	0.97
6416	0.96	0.91	0.94	0.89	0.89	0.73	0	0.96	1	0.89	0.95	0.94	0.92	0.96
3275	0.97	0.85	0.99	1	1	0.5	0	0.87	0.91	1	0.98	0.99	1	0.95
28403	0.99	0.9	0.99	0.98	0.97	0.61	0	0.92	0.96	0.98	1	0.99	0.98	0.98
5754	0.99	0.91	1	0.98	0.98	0.6	0	0.93	0.95	0.99	0.99	1	0.99	0.98
885	0.98	0.86	0.99	1	1	0.54	0	0.88	0.93	1	0.99	0.99	1	0.95
17145	0.99	0.96	0.98	0.94	0.93	0.71	0	0.97	0.96	0.95	0.98	0.98	0.95	1

LUND UNIVERSITY

Dept of Architecture and Built Environment: Division of Energy and Building Design

Dept of Building and Environmental Technology: Divisions of Building Physics and Building Services