2018

Development Of A Methodology For Fast Optimization Of Building Retrofit And Decision Making Support

Pengyuan Shen
University of Pennsylvania, penshen@upenn.edu

Development Of A Methodology For Fast Optimization Of Building Retrofit And Decision Making Support

Abstract
The condition of current building stock in the United States raises the question of whether the energy performance of existing buildings can ever be environmentally sustainable. In the United States, buildings accounted for 39% of total energy consumption and 72% of total electricity consumption (USEPA 2009). In addition, current building energy use is projected to increase by 1.7% annually until 2025 (J.D. Ryan 2004). The great potential for energy reduction in existing buildings has created opportunities in building energy retrofit projects (Noris et al. 2013). A building renovation project must not only be affordable, taking into account factors such as investor budgets, payback period, economic risks and uncertainties, but also create a thermally comfortable indoor environment and is sustainable through its lifetime. The research objective of this dissertation is to develop a novel method to optimize the performance of buildings during their post-retrofit period in the future climate. The dissertation is organized in three sections:

a) Develop a data-driven method for the hourly projection of energy use in the coming years, taking into account global climate change (GCC). Using machine learning algorithms, a validated data-driven model is used to predict the building's future hourly energy use based on simulation results generated by future extreme year weather data and it is demonstrated that GCC will change the optimal solution of future energy conservation measure (ECM) combination.

b) Develop a simplified building performance simulation tool based on a dynamic hourly simulation algorithm taking into account the thermal flux among zones. The tool named SimBldPy is tested on EnergyPlus models with DOE reference buildings. Its performance and fidelity in simulating hourly energy use with different heating and cooling set points in each zone, under various climate conditions, and with multiple ECMs being applied to the building, has been validated. This tool and modeling method could be used for rapid modeling and assessment of building energy for a variety of ECM options.

c) Use a non-dominated sorting technique to complete the multi-objective optimization task and design a schema to visualize optimization results and support the decision-making process after obtaining the multi-objective optimization results. By introducing the simplified hourly simulation model and the random forest (RF) models as a substitute for traditional energy simulation tools in objective function assessment, certain deep retrofit problem can be quickly optimized. Generated non-dominated solutions are rendered and displayed by a layered schema using agglomerative hierarchical clustering technique. The optimization method is then implemented on a Penn campus building for case study, and twenty out of a thousand retrofit plans can be recommended using the proposed decision-making method. The proposed decision making support framework is demonstrated by its robustness to the problem of deep retrofit optimization and is able to provide support for brainstorming and enumerate various possibilities during the process of making the decision.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

DEVELOPMENT OF A METHODOLOGY FOR FAST OPTIMIZATION OF BUILDING RETROFIT AND DECISION MAKING SUPPORT

Pengyuan Shen

A DISSERTATION

in

Architecture

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

_____

William W. Braham

Professor of Architecture

Graduate Group Chairperson

_____

David Leatherbarrow

Professor of Architecture

Dissertation Committee

David, Leatherbarrow, Professor & Chair of the Graduate Group in Architecture

Yun Kyu Yi, Assistant Professor, Illinois School of Architecture

Eric Eaton, Faculty in Department of Computer and Information Science

DEVELOPMENT OF A METHODOLOGY FOR FAST OPTIMIZATION OF BUILDING RETROFIT AND DECISION MAKING SUPPORT

To My Grandma in Heaven

ACKNOWLEDGMENT

I am so excited at this point. There are so many people to whom I want to express my gratitude during these years of study abroad. I still remember how eager I was to go to Penn when I opened the letter of admission and received the fellowship six years ago. Time flies by, I really cannot imagine how fast the six years at Penn have gone. The most important people that I want to thank are my family members who supported me all the way through both emotionally and financially. Without you, I would not be able to go to the United States and fulfil my dream of studying at a world-renowned university. My deepest thoughts go to my grandma who raised me and passed away shortly before my graduation. I feel so sorry for not being beside her to spend the last moment of her life with her.

I am lucky and grateful to be a student of Professor William Braham, who offered great help and advice for my thesis, shared with me his knowledge in learning and teaching, and gave me invaluable advice in every course and research project along the way. His unwavering passion for research, spirit of scholarship, visionary and innovative ideas, diligence, sense of humor, will always be the criteria to be pursued in my future professional life. I also could not be more delighted and fortunate to have Dr. Yunkyu Yi and Dr. Eric Eaton sit on my dissertation committee. Both of them gave me inspiration and suggestions for the completion of the dissertation. I wish them all the success in their field of research and study. I also want to thank Professor Ali Malkawi for bringing me to Penn at the very beginning, otherwise I will not have those wonderful memories here at all.

Moreover, I would like to thank my girlfriend, Rui Pei, for being with me in the last two years of PhD study where the most negative emotions and anxieties might occur. We shared

happiness and bitterness. Thank you for the great patience and attentiveness in proof reading my dissertation.

Finally, I want to thank all the extraordinary teachers and amazing colleagues who have taught me, given me advice and shared intellectual thoughts with me. Thank you!

ABSTRACT

DEVELOPMENT OF A METHODOLOGY FOR FAST OPTIMIZATION OF BUILDING RETROFIT AND DECISION MAKING SUPPORT

Pengyuan Shen

William W. Braham

The condition of current building stock in the United States raises the question of whether the energy performance of existing buildings can ever be environmentally sustainable. In the United States, buildings accounted for 39% of total energy consumption and 72% of total electricity consumption (USEPA 2009). In addition, current building energy use is projected to increase by 1.7% annually until 2025 (J.D. Ryan 2004). The great potential for energy reduction in existing buildings has created opportunities in building energy retrofit projects (Noris et al. 2013). A building renovation project must not only be affordable, taking into account factors such as investor budgets, payback period, economic risks and uncertainties, but also create a thermally comfortable indoor environment and is sustainable through its lifetime. The research objective of this dissertation is to develop a novel method to optimize the performance of buildings during their post-retrofit period in the future climate. The dissertation is organized in three sections:

a) Develop a data-driven method for the hourly projection of energy use in the coming years, taking into account global climate change (GCC). Using machine learning algorithms, a validated data-driven model is used to predict the building's future hourly energy use based on simulation results generated by future extreme year weather data and it is demonstrated that GCC will change the optimal solution of future energy conservation measure (ECM) combination.

b) Develop a simplified building performance simulation tool based on a dynamic hourly simulation algorithm taking into account the thermal flux among zones. The tool named SimBldPy is tested on EnergyPlus models with DOE reference buildings. Its performance and fidelity in simulating hourly energy use with different heating and cooling set points in each zone, under various climate conditions, and with multiple ECMs being applied to the building, has been validated. This tool and modeling method could be used for rapid modeling and assessment of building energy for a variety of ECM options.

c) Use a non-dominated sorting technique to complete the multi-objective optimization task and design a schema to visualize optimization results and support the decision-making process after obtaining the multi-objective optimization results. By introducing the simplified hourly simulation model and the random forest (RF) models as a substitute for traditional energy simulation tools in objective function assessment, certain deep retrofit problem can be quickly optimized. Generated non-dominated solutions are rendered and displayed by a layered schema using agglomerative hierarchical clustering technique. The optimization method is then implemented on a Penn campus building for case study, and twenty out of a thousand retrofit plans can be recommended using the proposed decision-making method. The proposed decision making support framework is demonstrated by its robustness to the problem of deep retrofit optimization and is able to provide support for brainstorming and enumerate various possibilities during the process of making the decision.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

1. Introduction and Statement of the Problem

Buildings consume a vast amount of resources in constructing, operating, maintaining, and demolishing. This process requires huge amounts of high quality resources, and human needs are steadily increasing with a higher standard of living. Buildings are the most essential and fundamental element in the formation of a city, the place where people gather and share their thoughts, beliefs, interests, and affections of their inhabitants. The networks of buildings create a multidimensional space that is essential for interpersonal relationships. Buildings have always been so important to us. However, they are also important to the functioning of the environmental system because the land we use, the trees we cut, the concrete and steel we produce, all come from the ecosystem around us. The contribution of buildings to society's collective environmental footprint is partly measurable and at the same time, unmeasurable—partly measurable because of the rapid development of engineering techniques for calculation and building performance simulation, which allows us to quantitatively assess the amounts of energy and resources used in their environmental footprint; while it is unmeasurable because a building's environmental performance system is so complex and involves so many social and cultural factors.

For the measurable or quantifiable dimension, it is reported that in the United States, buildings consume 39% of the energy and 72% of the electricity, and they generate 38% of the carbon dioxide, 49% of the sulfur dioxide and 25% of the nitrogen oxides found in the air (Ravi S. Srinivasan et al. 2012). Contemporary buildings cannot function without the energy provided by fossil fuels, while pollutants from fossils fuels pose risks to the local and global environment with the growing needs of the development of human civilization. For decades, much research attention has been paid to the energy efficiency of building, as existing buildings play an important role in building sector's energy consumption and carbon footprint. The method to find the best ECM combinations for existing building retrofit has been studied and debated over years.

Architects and engineers are concerned by the energy consumption and carbon emission associated with existing buildings. The energy performance of buildings deteriorates as it ages in an ever-changing outdoor environment. Global climate change (GCC), aging equipment and building systems, technology progress and adoption, among other factors, are challenges and opportunities for existing buildings. According to previous research by Aktas and Bilec, the average life of residential building in the United States is currently 61 years and has a linearly increasing trend (Aktas and Bilec 2011). For commercial buildings, the EIA estimates that the median service life of commercial buildings varies from 65 to 80 years depending on the type of building, based on data analysis from Commercial Building Energy Consumption Survey (CBECS) ((EIA) 2009). Approximately half of all commercial buildings were built before 1980 according to CBECS, as shown in Fig. 1, and Table 1 shows the distribution of lifetimes in different types of commercial buildings in the United States according to EIA annual energy outlook in 2011 ((EIA) 2011).

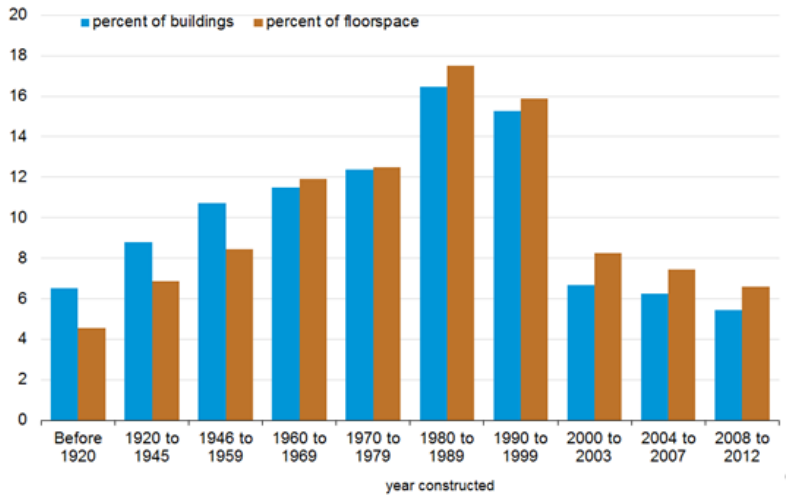


Fig. 1 Year in which commercial buildings were constructed ((EIA) 2009)

Table 1 Commercial building median lifetime in years ((EIA) 2011)

Building Type	Median	66% Survival (*)	33% Survival (*)
Assembly	55	40	75
Education	62	45	86
Food Sales	55	41	74
Food Service	50	35	71
Health Care	55	42	73
Large Office	65	46	92
Mercantile & Service	50	36	69
Small Office	58	41	82
Warehouse	58	41	82
Lodging	53	38	74
Other	60	44	81

Note: Number of years after which the building survives. For example, a third of the large office building constructed today will survive 92 years later.

As a matter of fact, an estimated 14 billion m^2 of existing buildings (about 50% of the total building stock) are expected to be renovated in the next 30 years in the United States (Zhai, Bendewald, and Hammer 2011). In 2003, only 26% of the commercial buildings in the United States were built in the last ten years (EIACBECS 2003). Approximately 86% of current building construction expenditures are spent on renovating existing buildings and the rest on new construction (Levine et al. 2014). With the low newly constructed building number and the trend in recent years, many of the existing buildings in the United States will be renovated if energy consumption and carbon emissions are expected to be significantly reduced in the future.

Regarding building retrofit, one of the most important things is to evaluate the building energy use (BEU) with different combinations of energy conservation measures (ECM). This has been done by many studies in the past under the current climatic conditions (Asadi et al. 2012a, b, Asadi et al. 2014, Chuah, Raghunathan, and Jha 2013, Lee et al. 2015, Levine et al. 2014,

Rysanek and Choudhary 2013, Yalcintas 2008). However, finding a computationally efficient way of assessing the different energy performance provided by different ECM combination in future climatic condition has not yet been fully developed. In this chapter, the background of the research subject, discussion of the state of the art, and its challenges and limitations will be discussed.

1.1 Challenges for Building Retrofit Optimization

The building retrofit process can be complicated, involving the efforts of different parties including property owners, building clients, architects, and engineers. One of the most important issues facing building owners and clients is finding the best ECM combination for the existing building. This is a problem that the building energy research community has been facing for over thirty years. Asked from different perspectives, this question also means differently by saying the "best" or the "bests" because various emphasis is often put on the renovation of buildings: sometimes, it could simply be economic benefits brought by the renovation, or the enhancement of the building indoor environment and quality, or the reduction of BEU and greenhouse gases (GHGs). Thus, the optimization function is generally multi-objective and multi-constrained (Asadi et al. 2012a, b, Asadi et al. 2014, Sun, Huang, and Huang 2015, Wang, Xia, and Zhang 2014, Lu et al. 2015). In this dissertation, the most imminent problem --- how to optimize the retrofit project in a computationally affordable manner will be addressed when decision-making factors such as indoor environment and quality, investment value, economic uncertainties are taken into account.

In recent years, the rapid development of computer technology and building energy simulation (BES) software has become a powerful tool for answering the question of how to integrate optimization approach with simulation results, which does not only give an estimation

of BEU of different ECM combinations but also its environmental impacts and building wellness. Simulation becomes a powerful tool to achieve the goal — that is, to critically assess, refute and further propose new direction of future high performance building design. It also makes it possible to study the performance of independent subsystem and their interactive behavior between them, notably by using object-oriented simulation tools. Existing BES tools like EnergyPlus, eQuest, TRNSYS, ESP-r, etc. (LBNL 2015, Hirsch 2016, Klein and al. 2010), can model and analyze building energy performance in terms of thermal dynamic features, day lighting, and human behavioral factors. Thus, various retrofit measures can be applied to virtually to buildings and simulated results inform us about the decision-making process in selecting ECMs. In general, computer-based simulation tools allow us to test different combinations of ECMs for a particular building in a relatively efficient and cost-effective way given a calibrated and accurate baseline model. However, the use of BES for optimizing ECM building combinations with parametric case studies presents several challenges:

a) First, the use of BES to evaluate building performance for the huge combinatorial problem when dealing with multiple ECMs brings an overwhelming computational cost. In a retrofit optimization problem, each ECM and its parameter are considered as a vector of design variable. Due to the non-linear nature of optimizing building retrofit problem, it is hard to construct an objective function with derivative using method of gradient descent or ascent (Wetter and Wright 2004, Michael and Jonathan 2003, Chidiac et al. 2011a). In order to find the interactions between different variables and their parameters, a huge combinatorial problem will be formed. The introduction of an additional ECM into a retrofit project will result in an exponential increase in simulation cases. Thus, using current optimization techniques to find solutions to the optimization problem would be infeasible, even with parallel computing technology. The application of gradient-free and non-derivative optimization algorithms such as pattern search and genetic algorithm can be useful in solving the combinatorial optimization

problem, but without a computationally efficient method for evaluating future year's building energy use, it is difficult to obtain optimal ECM combinations in the future.

b) Secondly, simplified and static simulation methods are unwieldy and non-generalizable because of the lack of availability of the tools developed by these methods. There are simplified BES modeling methods, such as gray models like R-C (resistance & capacitance) models and static simulation method like degree-day approach, which have been applied to building retrofit analysis. For example, Asadi et al. developed a multi-objective mathematical model to help stakeholders make decisions by seeking ECM that minimizes energy in a cost-effective manner (Asadi et al. 2012a). They used an R-C model to simultaneously evaluate the effectiveness of all available combinations of retrofit actions. Murray et al. developed a static simulation modeling process using a degree-day methodology to evaluate the gross building energy use (Murray, Rocher, and O'Sullivan 2012). The large monthly time step of these methods makes high-resolution analysis impossible, such as demand response and real-time pricing, the production of onsite renewable energy supply systems and the effect of cooling and heating setback during unoccupied hours. Hillebrand et al. developed and designed a simplified method to evaluate ECMs for office buildings in Europe by adopting a retrofit matrix method that takes into account ecological and economic efficiency (Hillebrand et al. 2014). The basis of this method is mainly predicated on standardized building types. However, each building is different from another. While the use of an established database of prototypical building to guide typical building types could assist in the preliminary phase of retrofit planning, it can not provide a detailed, reproducible and generalizable tool that addresses individual building. Moreover, each building model needs parameter optimization, model testing and calibration, and only then can a model be adopted for the prediction of energy load with different ECMs being applied. This is not easy for existing buildings where many complexities reside and where multiple zones exist. As a result, existing simplified model is generally better when applied to small-scale individual

building or individual zone, such as a single-zone building model with single occupancy schedule being applied to the building after the parameters and configurations of the model for this building are calibrated. But the modeling and calibration can be difficult and time-consuming when it is implemented on more complex building types, such as mixed-use buildings, various schedules of room temperature set point, occupancy and equipment. The advantages of the gray model are the lesser amount of inputs needed and the uncertainties of the input data are well controlled. However, the disadvantage of the existing simplified method is also salient for the purpose in this study: compared to the white models like EnergyPlus, a simulation tool based on simplified gray model that has the flexibility, the generalizability, and the reproducibility in modeling the building thermophysics including building geometry, solar heat gains and infiltration (e.g. opening, blinds, shadings) are not available and need to be developed.

c) The huge computational cost in generating a training database for the data-driven model. Data-driven method can be used to significantly reduce computational cost in solving the building retrofit optimization problems because it can be used to predict BEU of each ECM combination applied to the target building with potentially high reliability if data samples are sufficient and the model is properly trained. In a research that uses data-driven method to learn how different ECMs and their parameters will change the BEU, a huge database shall be established and used as the training database for the optimization problem (Eisenhower et al. 2012). Nevertheless, if EnergyPlus is used to simulate the BEU of different combinations of ECMs under current climatic conditions, the enormous calculation cost is expected as mentioned in point a) and also mentioned by Rysanek and Choudhary (Rysanek and Choudhary 2013), not to mention running each case for the next 30 years with hourly weather data from each year generated by the future climate scenario. In fact, without the help of data driven model, it is rather impossible to perform the optimization process in this study. However, the problem of how to

judiciously reduce the size of the training database and save computing power in database generation as well as in the training and forecasting process needs to be resolved.

1.2 Optimization Approach and Objectives

Table 2 Literature review on building retrofit optimization

Literature	Objectives	Objective Function Evaluation	Energy Use Evaluation	Optimization Algorithm
(Wang and Xia 2015)	Aggregated energy saving, internal rate of return	Weighted sum multi-objective	Model predictive control	Differential evolution
(Wang, Xia, and Zhang 2014)	Energy saving, lifecycle net present value (NPV), discounted payback period	Weighted sum multi-objective	Estimation	Differential evolution
(Jafari and Valentin 2017)	Total lifecycle cost (LCC)	Single objective optimization	eQuest & static modeling	Genetic algorithm
(Shao, Geyer, and Lang 2014)	Initial investment cost, energy consumption, global warming potential	Multi-objective Pareto front	DIN V 18599 assessment method	NSGA-II
(Asadi et al. 2012b)	Retrofit cost, energy saving in kWh, thermal comfort	Weighted Tchebycheff metric	TRNSYS	Tchebycheff programming
(Rysanek and Choudhary 2012b)	Greenhouse gas emission reduction	Single objective	TRNSYS & Matlab	Brute-force
(Eisenhower et al. 2012)	Thermal comfort, annual energy consumption	Weighted sum multi-objective	EnergyPlus	Repeated sampling and feature reduction
(Mauro et al. 2015)	Energy demand, thermal comfort, global cost	Multi-stage analysis method	EnergyPlus & Matlab	Feature reduction and brute-force
(Chidiac et al. 2011b)	Payback period	Single objective	Archetype modeling in EnergyPlus	Non-linear regression
(Asadi et al. 2012a)	Retrofit cost, energy saving in kWh	Weighted sum multi-objective	ISO 13790 RC model (monthly)	Tchebycheff programming
(Siddharth et al. 2011)	Electricity use, natural gas use	Weighted sum multi-objective	DOE 2.2	Genetic algorithm
(Roberti et al. 2017)	Energy use, thermal comfort, conservation compatibility for historic building	Multi-objective Pareto front	EnergyPlus	NSGA-II

Literature	Objectives	Objective Function Evaluation	Energy Use Evaluation	Optimization Algorithm
(Tadeu et al. 2015)	Global cost, primary energy use	Multi-objective Pareto front	EnergyPlus	Brute-force
(Son and Kim 2016)	Energy consumption, CO2 emissions, retrofit costs, and thermal comfort	Multi-objective Pareto front	EnergyPlus	NSGA-III
(Wu et al. 2017)	Annualized costs and life cycle GHGs emissions	Single obejective & Multi-objective Pareto front	EnergyPlus	Epsilon-constraint method
(Asadi et al. 2014)	Energy consumption, retrofit cost, thermal comfort	Single objective and Multi-objective Pareto front	TRNSYS	Latin-hypercube sampling, artificial neural network, MOGA (Multi-Objective Genetic Algorithm)
(Rysanek and Choudhary 2013)	Marginal abatement cost vs. GHGs emissions saved; discounted payback period vs. required capital	Primary and secondary objectives	TRNSYS & Matlab	Brute-force
(Chantrelle et al. 2011)	EnergyConsumption, investment, thermal comfort	Multi-objective Pareto front	TRNSYS	NSGA-II
(Pombo et al. 2016)	Lifecycle financial saving, energy saving	Multi-objective Pareto front	EnergyPlus	Brute-force

Table 2 lists research related to building retrofit optimization and methods. These studies dealt with different objectives and different methods for forming optimization problems. The most often used objectives in the optimization are the energy use or consumption, the economic metrics taking into account lifecycle analysis and thermal comfort. For the BES tools used in those research, EnergyPlus (LBNL 2015) is mostly used, and others include TRNSYS (Klein and al. 2010), eQuest, which uses the DOE simulation engine (Hirsch 2016). Most popular optimization algorithm is evolutionary algorithm including genetic algorithm (GA), multi-objective genetic algorithm (MOGA), non-dominated sorting genetic algorithm (NSGA). The advantages of GA are that unlike brute force, it does not exhaust the entire design space of the ECM variables included in a retrofit project. After generations of evolution, GA is able to provide an optimal solution to the given objective function, although global optima are not guaranteed. It has been very widely used in researches concerning building retrofit optimization. However, the

use of GA is usually coupled with BES tools to evaluate the energy use under different ECM bundles since energy performance always plays an indispensable role, directly or indirectly, in the objective function. As discussed, EnergyPlus is one of the most popular tools that are used together with evolutionary algorithms. White box modeling tools like EnergyPlus that involve a lot of input information and manipulate dynamic functions in building energy modeling can be time consuming to model and simulate, especially when it is faced with exponentially increasing simulations that are required in such a combinatorial optimization problem for the retrofit. How to facilitate the evaluation of objective function during optimization that may involve various factors such as building energy performance, indoor thermal comfort, investment and returns could be *Limitation One* of this research problem.

Limitation Two would be the difficulty of generalizing the result of all the ECMs to the same type of building by means of an archetypical building study (Chidiac et al. 2011b) and providing clients with better decision-making support regarding the optimization results. As we discussed in the beginning of the dissertation, it is not reasonable to treat an individual building as its archetypical representative, because modern buildings are different in terms of geometric shape, use of materials, building systems, etc. even if it is classified as the same use type. For example, the lab buildings on the University of Pennsylvania campus are so diversified in their win-wall ratio, use schedule, equipment type, and thermal capacity, even within the same type of use (laboratory), making the variance of energy use intensity large. Therefore, for the purpose of deep energy retrofitting, it would be important to develop a broadly applicable methodology for the rapid optimization of individual building retrofitting planning..

In addition, two mechanisms for optimizing the building retrofit are mainly used in the reviewed research: the deterministic method (weighted sum method is often used) and the non-dominated method (Pareto front). *Limitation Three* is that not enough support are provided for

user's decision-making in both methods. Most building retrofit optimization problems involve several objectives, making them multi-objective problems. Although giving different weights to each sub-objectives before optimization takes place would reduce the complexity of the problem by converting the multi-objective problem into a single objective one, the "a priori" nature of this method requires preferential information to solve the problem. It would be difficult for users or clients to define appropriate weight values in the final objective function, with little knowledge about how the optimization results will look like. Moreover, when implementing this deterministic method, each sub-objective function should be transformed and normalized into a uniform scale to achieve dimensionless comparison among each other. This process also requires the intervention from client or decision maker to determine the labeling criteria and transform the output of each sub-objective function into the same scale. Compared to the disadvantage of deterministic method in decision-making process, non-dominated method is able to visualize the trade-offs in retrofit planning, but the drawback could be that the optimized Pareto front curve is so widespread that it will be difficult to have an idea of where to start and which range on the Pareto curve might be interesting to look at.

Limitation Four is that it should be noted that LCA is necessary for a retrofit project because clients tend not to do retrofit too frequently in a short period of time, while the most reviewed research do not take into account future climate uncertainties. Certainly, it will not be so difficult to simulate future years' hourly energy use by using hourly downscaled future weather data (Shen , Xu et al. 2012, Chan 2011, Belcher, Hacker, and Powell 2005), but a thirty-year lifespan of hourly energy simulation by using tools like EnergyPlus for the objective function evaluation could be rather unprocurable given the immense scale in computation. A method to circumvent the huge computational cost engendered by simulation runs should be found and tested.

1.3 Future Building Energy Use

The level of greenhouse gases (GHGs) under different scenarios projected by the Intergovernmental Panel on Climate Change (IPCC) demonstrates a dramatic increase of GHGs in the future (IPCC 2007). The status quo of GCC is also updated due to changing trends in GHGs emission levels and human activities. The IPCC is continuously paying attention to the ever-changing status of GCC and the carbon emission scenarios since they were first put forward in the year of 2000. In the recently released IPCC Fifth Assessment Report (AR5), common facts and basics have been achieved and updated. It is stated that "human influence on the climate system is clear, and the atmospheric concentrations of carbon dioxide, methane, and nitrous oxide have increased to levels unprecedented in at least the last 800,000 years" (IPCC 2013). Fig. 2 shows the history of surface temperature change on both land and ocean over the last century.

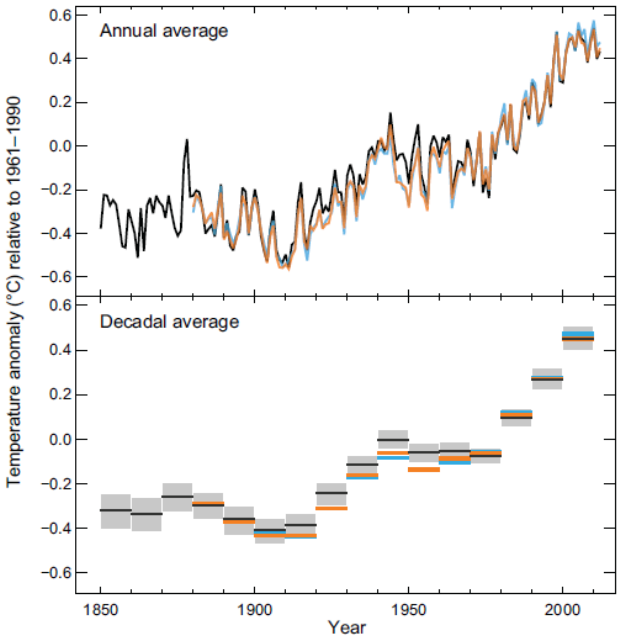


Fig. 2 Observed annually averaged combined land and ocean surface temperature anomaly 1850 – 2012 (IPCC 2013) (Top panel: annual mean values. Bottom panel: decadal mean values including the estimate of uncertainty for one dataset (black). Anomalies are relative to the mean of 1961−1990)

Many studies have shown that GCC brings great influences to BEU around the world. In Australia, Wang et al. evaluated the heating and cooling energy requirements and the corresponding carbon emissions of residential houses under different future climatic conditions(Wang, Chen, and Ren 2011), and found that the carbon emission of a 5-star house was projected to have an average increase of 30% in Darwin, 15% in Alice Springs and 19% in Sydney. Hassan Radhi assessed the potential impact of GCC on residential buildings in the United Arab Emirates with regard to CO_2 emissions (Radhi 2009). In this research, the design of building envelope and fenestrations in the future is given highlight in combating the increase of building energy consumption as well as CO_2 emission. Researchers in China have also evaluated the GCC's impact on the regional renewable energy (RE) system and found that the extreme events (such as rainstorms, frost, etc.) and the variation of climatic elements will have substantial impacts on RE systems in different province of China including Guangdong, Gansu, and Tibet will be most vulnerable to GCC in terms of installing RE system (Wang et al. 2014). Shen stated that in four representative cities in the United States, the annual energy use is expected to change from -1.64% to 14.07% for residential buildings and from -3.27% to -0.12% for office buildings under A2 scenario (a carbon emission scenarios defined by IPCC) in different regions during 2040 to 2069 (Shen 2017). Two recent studies also reveal that GCC will affect the efficiency of building onsite renewable systems: one claims that GCC will bring down the energy efficiency of the GSHP system in residential applications in the United States using TRNSYS and eQuest modeling technique as the warmer ground in the future will result in an average rise of about 2–3 °C in the inlet and outlet water temperatures of GSHP during the cooling season (Shen and Lukes 2015), while the other research indicates that for all the existing net zero energy buildings located in the ten climate zone in the United States, the proportion of decreased annual PV output is 5 out of 20, while that of wind turbines is 12 out of 20, indicating the comparatively lower stability of the renewable energy system prioritized to wind turbines than to PV prioritized RE

system (Shen and Lior 2016). The impacts of GCC on energy use in different types of buildings in the future are expected and analyzed in various research. Given the fact that the retrofitted building will last for a certain period of time during post-retrofit stage, it is essential that during the decision-making process, the situation of the possible change in building energy performance introduced by climate change as well as the retrofit measures should be studied. Therefore, it is necessary that GCC's implication to building performance should be studied since building energy retrofit will be made for the "future" building instead of buildings in current situation.

Most research on existing building retrofit evaluation and optimization method does not take into account the impacts of GCC. Meanwhile, the research on GCC's impact is generally focused on the energy performance of existing buildings. Little work has been done on the ongoing energy performance of existing buildings during the post-retrofit stage and the long-term performance of renewable energy system under climate change conditions. Recent research by Chow et al. studied the effectiveness of retrofitting existing public buildings in the face of GCC in the "hot summer and cold winter" climate region in China (Chow, Li, and Darkwa 2013). The study focuses mainly on existing buildings in Zhejiang Province. It uses the HadCM3 global climate model results to estimate the future building energy use by running the new weather file in DOE2 simulation engine. Only two retrofit measures are considered in the context of this research: a) improving the U-Values of the building enclosure; and (b) improving the domestic hot-water system by replacing the electric boiler with an air-source heat pump. Another research in 2016 by Swedish and Swiss researchers showed a research closely related to building energy retrofit measures and theirs robustness to future climate change in terms of mainly heating energy consumption using a simplified model (Nik, Mata, and Sasic Kalagasidis 2015). The research focuses on a residential stock in Stockholm to 2100 for five climate scenarios. The simulation is carried out using a lumped system in Simulink involving all 153 residential buildings. Again, only two retrofit measures are considered here: a) assuming reduction of the lighting power of the

building stock by 50%, by installing more efficient lighting equipment; b) upgrading the building envelope by adding thermal insulation. Limitations still exist in related studies, including very few ECM options, limitation of regional application, large simulation time step, and lack of an inclusive methodology that can be applied to any building type in various climate zones.

2. Development of Data-driven Model for Future Energy Projection

2.1 Parametric Study Tool

The purpose of this chapter is to look for a data driven method to replace energy simulation method to predict hourly energy use in future years. We use EnergyPlus as the baseline simulation engine for the validation of the developed data-driven method. EnergyPlus is a widely used BES model in both academic and commercial studies, which is developed by the Department of Energy (DOE). Its precedent versions are BLAST and DOE-2 and it inherited the features and strengths of both programs. EnergyPlus is able to model the whole building energy performance and has undergone numerous reliability tests (DOE 2014). One of the characters of EnergyPlus is that it uses ASCII text based weather data, building input, and simulation output files and is able to handle design and engineering alterations by modifying the textual information in the input file. This important feature allows user to process inputs and transform the user inputs into text based information for EnergyPlus to override the baseline building model. In this way, different retrofit scenarios of implementing various ECM combination for the existing building can be realized in EnergyPlus model. The text based outputs generated by EnergyPlus can be collected and analyzed later. In this study, Python is used as the packaging tool for feeding EnergyPlus with retrofit inputs by user defined retrofit options. Python is good for text file

processing, and has strong support and capability in data processing. It is chosen to program the main console and tool that helps ECM information injection and EnergyPlus output handling.

Building ECMs mainly fall into three major categories: passive, active, and renewables as shown in Fig. 3: 1) Passive retrofit is mainly about reducing the heating and cooling load of a building, including building envelope insulation, fenestration system retrofit, and natural ventilation. Passive retrofit ECMs aim to curb the heating and cooling load of the building when the building is supposed to maintain at a certain thermally comfortable condition for occupants; 2) Active retrofit refers to system efficiency, including HVAC system efficiency (heat source, pump, fan, valves, and etc.), lighting system efficiency, building control system. Active retrofit ECMs are applied to these building systems that are used to meet the energy demand caused by load as described in point 1 with decent efficiency; 3) Renewables refers primarily to the onsite renewable energy system, which provides additional or major portion of the required energy for the building in place of traditional fossil fuel energy source in order to reduce GHGs emission and total energy consumption. Renewables energies have attracted a great deal of interest in recent years, with the onsite renewable energy systems providing cleaner and less carbon-intensive energy supply alternatives.

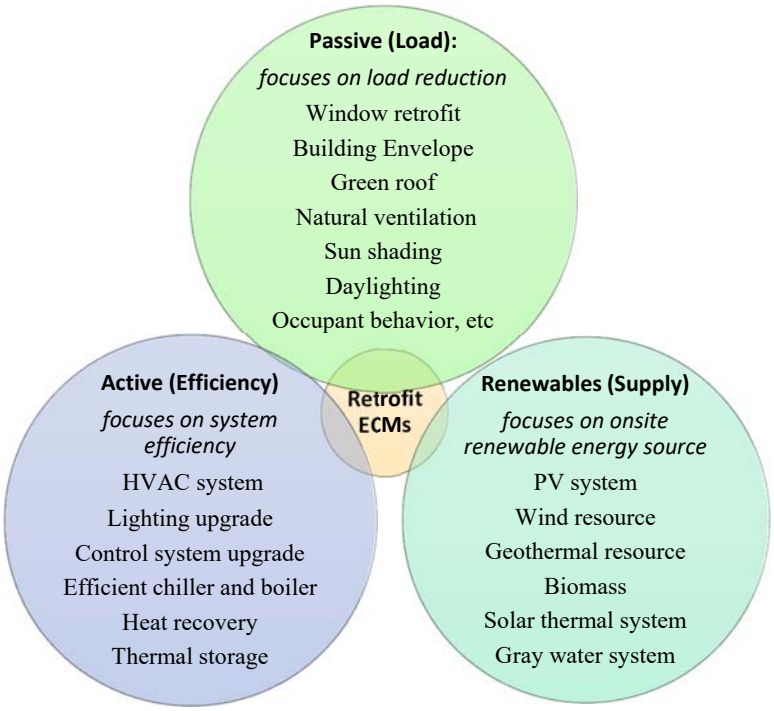


Fig. 3 Categories of building retrofit ECMs

In this study, we included eleven ECMs, but the retrofit options and their parameters are not limited to these ECMs in practice for different retrofit projects. In addition, new emerging ECMs can be added to the Python retrofit tool. The eleven pioneer ECMs and their parameters are presented in Table 3.

Table 3 ECMs and parameters

Number	ECM	Parameter Range		
1	Window SHGC	0.3	0.5	0.8
2	Window U-value (W/m^2-K)	0.4	1	2
3	Window shading	N/A	Internal Blind	External Blind
4	Wall Insulation (m^{2}K/W)	N/A	2	4
5	Air infiltration rate for residential (h^{-1})	N/A	0.5	1
	Air infiltration rate for office (h^{-1})	N/A	1	2
6	Roof Insulation (m^{2}K/W)	N/A	3	6
7	Heating efficiency	N/A	0.9	
8	Cooling COP	N/A	4.5	
9	Cooling supply air temperature (°C)	N/A	15	
10	Lighting efficiency improvement	N/A	40% (compared with current condition)	
11	Daylighting control and dimming	N/A	Applied	

Note: N/A: ECM not applied to the building; SHGC: solar heat gain coefficient.

2.2 Future Local Hourly Weather Data

In this research, the global climate model (GCM) --- HadCM3, developed at the Hadley Center in the United Kingdom, is adopted to generate future weather file. HadCM3 (Hadley Center Coupled Model Version 3), like other GCMs, is a grid point model with large grid cells (2.5^o in latitude and 3.75^o in longitude over land areas, which gives 96 * 73 grid points on the scalar grid) (Pope et al. 2000). The outputs of the HadCM3 model are monthly averages for each climatic variables of the chosen future period. This model has been used in some building energy research (Gaterell and McEvoy 2005) (Chan 2011). It is assumed that the building retrofit will last about 35 years in this research. A recent run of HadCM3 model's outputs is used (IPCC Fifth Assessment Report (AR5)). For each GCM, the simulations were performed with prescribed CO_2

concentrations reaching 421 ppm (RCP2.6), 538 ppm (RCP4.5), 670 ppm (RCP6.0), and 936 ppm (RCP 8.5) by the year 2100 (IPCC 2013). These scenarios are described in the IPCC Fifth Assessment Report (AR5) and are named representative concentration pathways (RCPs). In this dissertation, the RCP6.0 scenario is selected to predict future climate condition for the building retrofit. In RCP6.0, emissions peak around 2080, then decline. The different RCPs are plotted in Fig. 4:

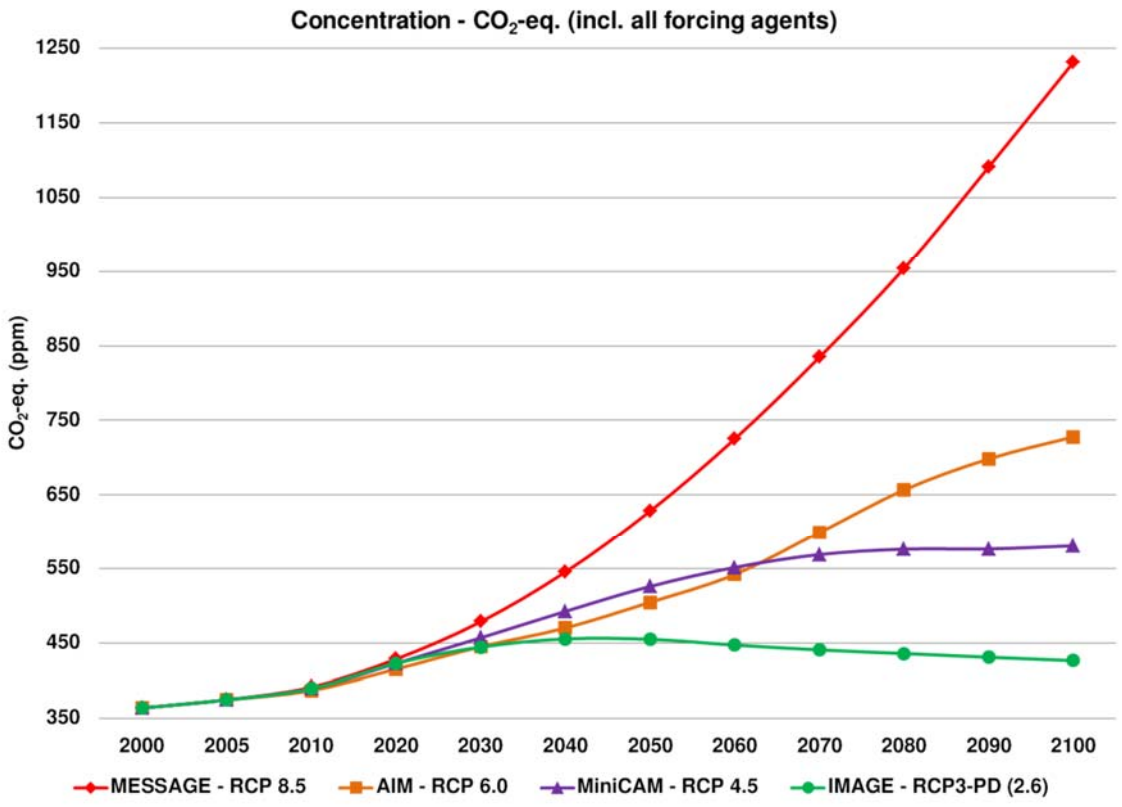


Fig. 4 All forcing agents' atmospheric CO_2-equivalent concentrations (in parts-per-million-by-volume (ppmv)) according to four RCPs (IPCC 2013)

Due to the fact that GCM operates on a coarse global grid to model and predict the future GCC scenarios using a numerical solution to Navier-Stokes equations to save the computational resources, GCC outputs can not be directly used as EnergyPlus weather data. The outputs need to

be “downscaled” to the local geographic location with finer granularity and mapped to hourly step. Thanks to the times series based morphing downscaling method proposed by Belcher et al. (Belcher, Hacker, and Powell 2005), it is computationally efficient to obtain local hourly weather data in the future. The detailed morphing algorithm and its results in 10 climate zones in the United States are conducted and obtained in (Shen and Lior 2016). In this study, San Francisco and Philadelphia are chosen to be experimented with the proposed optimization approach. The reason for choosing these two cities for this study is because San Francisco has stable temperature spread over a year with lowest temperature at 10℃ in January and high temperature at around 27℃ in summer while the variance of outdoor dry bulb temperature in Philadelphia is high throughout the year (humid subtropical climate zone that has hot summer and cold winter) (NOAA 2016), making comparison for prioritizing different ECMs for buildings in the two climate zones interesting.

2.3 Feature selection for ECMs

For the eleven proposed ECMs in Table 3, if all of their combinations are considered in the building retrofit optimization, then there will be 23328 EnergyPlus simulation cases in total, not to mention the exponential growth of simulation cases when new ECMs are added in practice. In fact, for a specific existing building, there will be ECMs that are not so effective in reducing BEU. For example, we performed the 23328 simulations for a residential building and an office building in San Francisco (see the detailed description of the model in section 3.1) respectively and found that the effects of cooling supply air temperature adjustment in residential building as well as improving the heating efficiency in office building are not obvious to BEU, while the impacts of improved wall insulation in the residential building and the improved air infiltration

rate in office building are rather significant to BEU (Fig. 5) (or any other objective function depending on the different needs of the research; other criteria may also be used, such as financial payback period, environmental impacts of greenhouse gas emission, etc.). This brings about a question: is running the combinations of all suggested ECMs required? Because it is not hard to think of the situation that some ECM's benefits for the existing building is trivial from the perspective of BEU saving. This purveys a possible way to circumvent the large combinatorial search space by fist finding the most influential ECMs to the BEU --- before starting to spend great amount of computational resources to simulate the BEU of all the possible combinations.

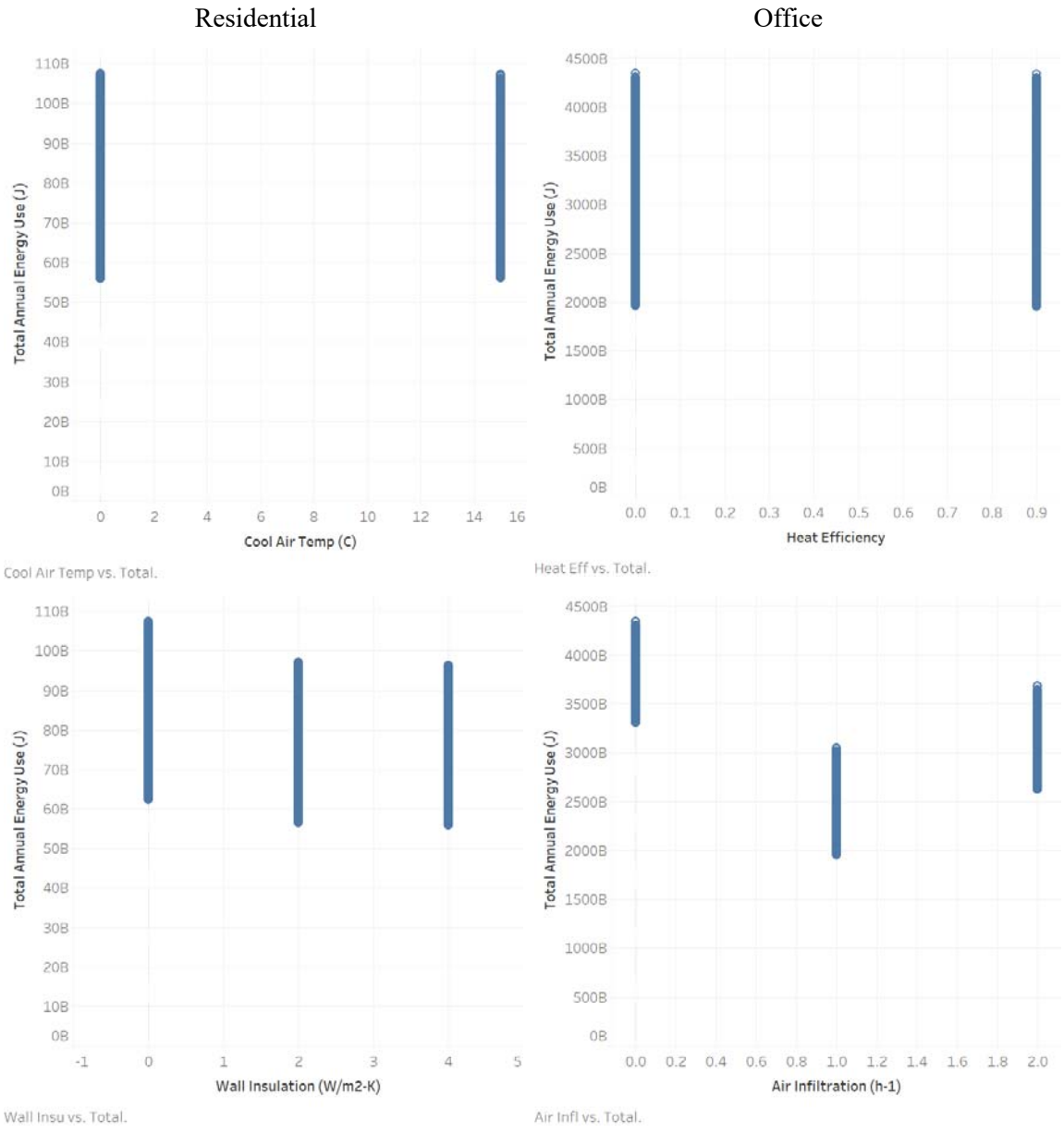


Fig. 5 Different ECM's impact to building annual energy use simulated by EnergyPlus

In this research, a feature selection algorithm based on information theory is used to search the most influential ECMs to an existing building in terms of the criterion for reducing energy use. The method is called Joint Mutual Information Maximization (JMIM), which uses joint mutual information and the "maximum of the minimum" approach that chooses the most relevant and influential features to the BEU. The algorithm is able to avoid overestimating the

importance of the features compared with conventional joint mutual information (JMI) method (Bennasar, Hicks, and Setchi 2015). To have an idea of the method, let's first start with the concept of entropy in information theory (Shannon 1948):

Equation 1: $$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x),$$

In Equation 1, H(X) is the entropy of the distribution of variable X, which concerns about the uncertainty level for its distribution and the average amount of information required to describe the random variable. For discrete variables, p(x) can be calculated by the proportion of the number of instants with value x in the total number of instants (N). Then we can calculate the conditional entropy H(X|Y) of two distributions as:

Equation 2: $$H(X|Y) = -\sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) \log p(x|y),$$

Conditional entropy basically tells the amount of uncertainty left in X after Y is seen. Then mutual information I(X;Y) between X and Y is defined as:

Equation 3: $$I(X;Y) = H(X) - H(X|Y)$$

Equation 4: $$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(xy) \log \frac{p(xy)}{p(x)p(y)}$$

The first term explains the uncertainty before Y is known, while the second term represents the uncertainty after Y is known. Mutual information can be thought of the amount of uncertainty in X that is removed by knowing Y. Then conditional mutual information can be defined as:

Equation 5: $$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = I(X;Y,Z) - I(X;Z),$$

where I(X;Y|Z) is the joint mutual information.

Lemma 1 basically says that for a featuref_i, if the m-joint mutual information is larger than that of all other features f_j, where f_i and $f_j \in F - S(i \neq j)$, then it is the most relevant feature to the target in the context of the subset of S. Here F is the full set of features, and let S be the subset of features that are selected already. JMIM employs joint mutual information and the 'maximum of the minimum' approach, which should choose the most relevant features according to Lemma 1 and is given by (Shannon 1948):

Equation 6: $$f_{jMIM} = \arg max_{f_j \in F-S} \left(min_{f_S \in S} \left(I(f_i, f_s; y) \right) \right)$$

where $I(f_i, f_s; y) = I(f_S; y) + I(f_i; y/f_S)$.

With the help of the JMIM algorithm, it is possible to find the most effective ECMs in an ECM combination. However, it is undoubted that the most effective ECM can be identified by the ranking of the features when JMIM is run on the complete annual BEU database of all possible ECM combinations, but what we need is to use fewer samples of simulation results to find out what these ECMs are. Therefore, we used Latin-hypercube sampling (LHS) technique to sample the combinatorial design space. Since LHS works only with continuous design space, the discrete retrofit design space is converted to continuous variable space, with each ECM parameter mapped to a uniform distribution. We then experimented with different sample sizes to find the minimum sample size that is sufficient to obtain a rough ranking of the feature importance. The results of the LHS sampling and random sampling are compared in this research when the method is applied to test case buildings, which will be presented in section 3.2.

2.4 Random Forest Algorithm

According to the challenge introduced in section 1.1, the BEU simulation in future years of each ECM combination by any simulation-based optimization would be implausible due to the

overwhelming computational cost. In this research, a data-driven method is used to predict the future years' BEU for each ECM combination after that the most significant ECMs are selected using the feature selection method proposed in section 2.3.

The Random forest (RF) algorithm is used as a data mining method in this study. It is an ensemble learning method based on a non-parametric supervised learning method called a decision tree algorithm that uses a graph or tree model to learn and predict the schema of the best routes or rules. When applying the decision tree algorithm, the features of the independent variables can be either categorical or continuous. For a single decision tree, the simplest consistent explanation is the best, and such bias is called inductive bias. It is the set of assumptions that the learner uses to predict outputs given inputs that it has not encountered (Mitchell 1980). The basic algorithm for top-down induction of decision trees (ID3, C4.5 by Quinlan) is as follows (Quinlan 1986):

1. A, which is the "best" decision attribute for the next node.
2. Assign A as decision attribute for node.
3. For each value of A, create a new descendent of node.
4. Sort training examples to leaf nodes.
5. If training examples are perfectly classified, stop. Else, repeat over new leaf nodes.

In choosing the "best" attribute, the concept of entropy in Equation 1 is also used in decision tree algorithm. Another important notion is Information Gain:

Equation 7: $$IG(A,S) = H(S) - \sum_{t \in T} p(t)H(t)$$

Where,

- H(S) - Entropy of set S
- T - The subsets created from splitting set S by attribute A
- p(t) - The proportion of the number of elements in t to the number of elements in set S
- H(t) - Entropy of subset t

Furthermore, if the attribute has many values, information gain will select it. Hence, in order to evade such node as date time like node, Gain Ratio should be used to evaluate the "best" attribute.

Equation 8:
$$GainRatio(X, A) = \frac{InfoGain(X, A)}{SplitInformation(X, A)}$$

Equation 9:
$$SplitInformation(X, A) = -\sum_{v \in values(A)} \frac{|X_v|}{|X|} \log_2 \frac{|X_v|}{|X|}$$

where, X_v is a subset of X for which A has value v. Using this node division criterion when choosing the next division in the node, a tree-type model will be constructed. The more the tree develops in depth, the more complex the decision rules or routes and the fitter the model will be adapted to the training data. However, a single decision tree will be biased because it will overfit the training data, which means that if the hypothesis space has many dimensions (large number of attributes), meaningless regularity in the data that is irrelevant to the true, important, distinguishing features will be established. For example, a single decision tree will be sensitive to data outliers, making it fit the outliers and decreasing its predictive power, or if there is too little training data, even a reasonable hypothesis space will overfit. To settle this problem, RF is constructed by a set of decision trees in the training process and produces the rule that is the mean prediction of the individual trees. This corrects the overfitting behavior of a single decision tree. For RF, each tree in the ensemble is built from a sample drawn with replacement of the training set. When dividing a node during the construction of the tree, the chosen division is no longer the best split among all features. Instead, the split that is picked is the best split among a random subset of the features. Using RF instead of a single decision tree will effectively lower the variance of the model and handle the overfit problem.

In this study, RF is trained by the BEU database containing the hourly simulation results of selected influential ECM combinations based on future extreme year's hourly weather data constructed as described in section 3.3. The features of the database include the weather variables including temperature, relative humidity, solar irradiation, wind speed, and building occupancy reported by EnergyPlus. Two different models are trained for electricity use and gas use of the building (if the building uses gas) for each building under certain future climate scenario (RCP6.0). Each retrofit plan will have its respective RF model for the prediction of future hourly BEU. The results and validity of the model will be discussed in section 3.3.

3. Test of the Data-driven Method and Discussions

3.1 Test residential and office building

In this research, we use two reference EnergyPlus building models in the United States that are compliant with American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) 90.1's 2004 building code and International Energy Conservation Code (IECC) 2006 as the baseline case. San Francisco and Philadelphia are two climate zones where we apply and validate the proposed data-driven workflow and methods. It is assumed that the building retrofit lasts 35 years from the year of 2020 to 2055. Detailed descriptions of the building physical and thermal characteristics are shown in Table 4. The occupancy behavior, lighting and equipment intensity, and system schedule are shared for two types of buildings in San Francisco and Philadelphia respectively.

Table 4 Major model parameter for the two buildings in San Francisco and Philadelphia

	San Francisco		Philadelphia	
	Residential	**Office**	**Residential**	**Office**
Building Area (m^2)	335	4982.2	335	4982.2
Gross Wall Area (m^2)	235.1	1977.7	235.1	1977.7
Window Area (m^2)	33.2	652.6	33.2	652.6
Window U-value (W/m^2-K)	3.695	4.913	2.273	3.045
Window SHGC	0.398	0.365	0.394	0.428
Wall U-value (W/m^2-K)	0.535	0.787	0.535	0.787
Roof U-value (W/m^2-K)	2.481	0.376	2.481	0.358
Air Infiltration rate (h^{-1})	1.5	1.5	1.5	1.5
Cooling Capacity (W)	7177.5	326921.42	14450.6	786467.1
Cooling COP	3.19	3.23	3.37	3
Heating Capacity (W)	12277.37	648888.26	22629.8	1170823.31
Heating Efficiency	0.78	0.8	0.78	0.8
Lighting (W/m^2)	2.28	10.76	2.28	10.76
Service Water Power Input (W)	11137.8	29307.11	11137.8	29307.11
Service Water Heating Efficiency	0.81	0.81	0.80	0.81

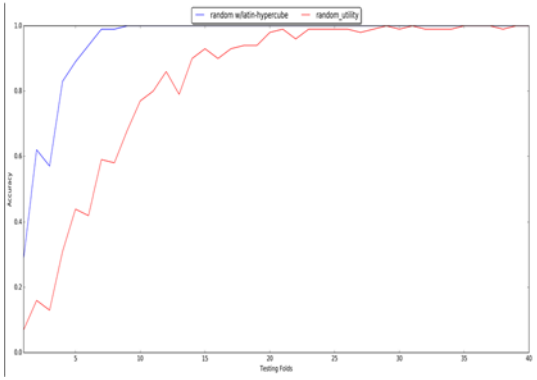
3.2 Feature selection for training database size reduction

The proposed feature selection algorithm is applied both types of building in two climate zones and we try to find the best sampling size with a decent chance to get the most influential ECMs for the target building without exhausting the whole combinatorial search space. To illustrate the validity of using the portion of the entire combinatorial space to obtain these ECMs, we validate the method in the following steps:

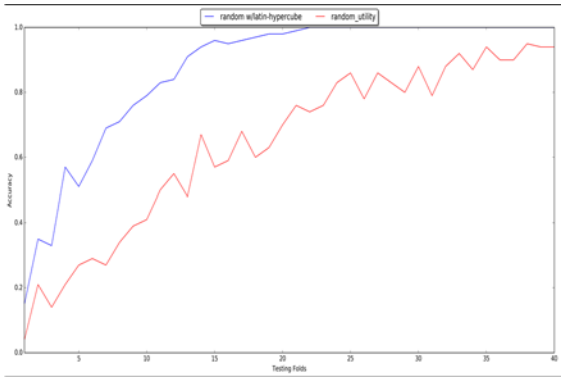
1) A “fold” concept is introduced here to describe the sample size by defining one “fold” of sample size as the number of all the ECM parameters considered in a retrofit, which means that, in this study, one fold of sampling size is thirty one samples because there are a total 31 parameters from all the ECMs.

2) LHS sampling is used to generate different number of sampling folds ranging from one to forty for the continuous design space of eleven dimensions, which has been mapped from the discrete space containing the ECM by uniform distribution between 0 and 1. We performed 200 times of feature selection for each number of folds and calculate the average accuracy of the feature importance rankings and compare it with the true feature importance rankings generated by the entire database with 23328 cases.

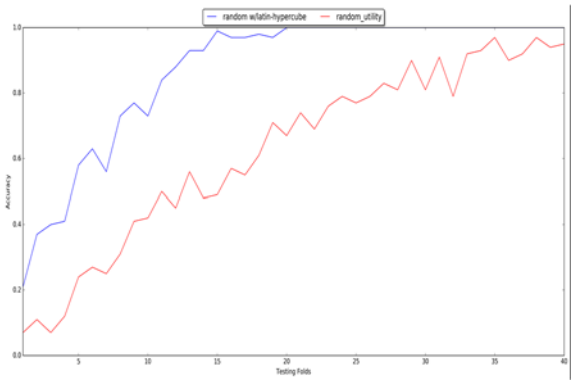
3) The same process in step 2 is performed with random sampling to compare the feature selection accuracy of LHS.



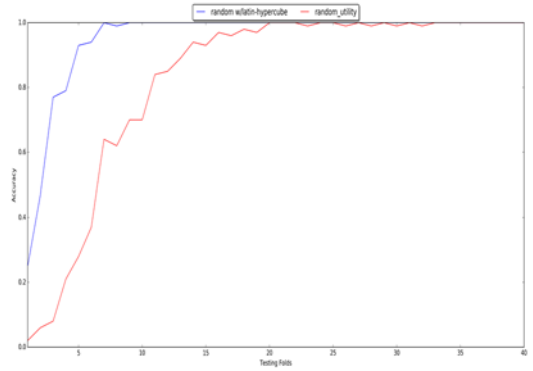
Residential Building in PHL w/ 6 chosen ECMs



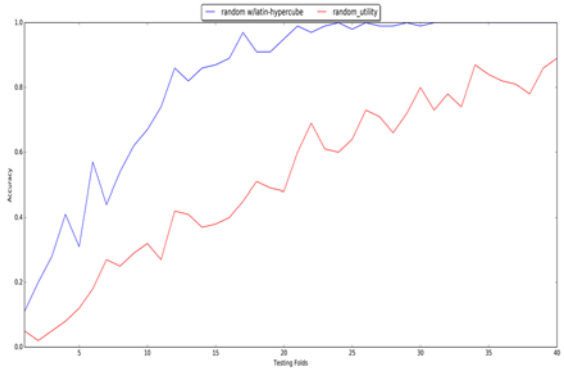
Residential Building in PHL w/ 7 chosen ECMs



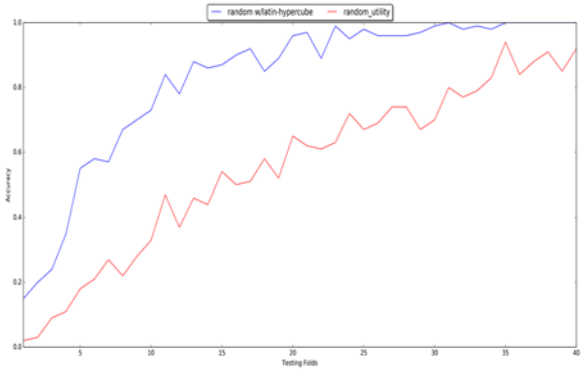
Residential Building in PHL w/ 8 chosen ECMs



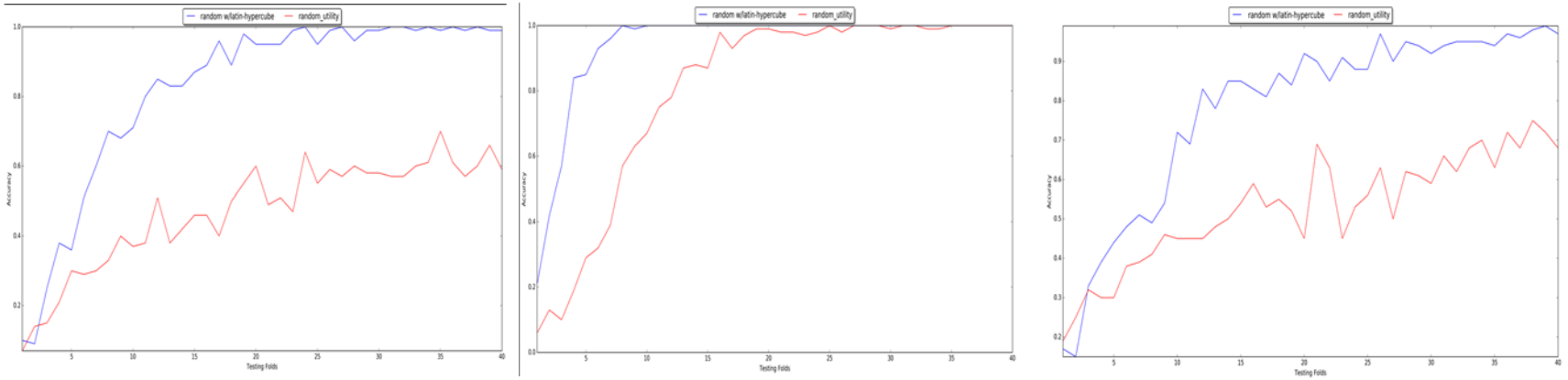
Residential Building in SF w/ 6 chosen ECMs



Residential Building in SF w/ 7 chosen ECMs



Residential Building in SF w/ 8 chosen ECMs



Office Building in PHL w/ 6 chosen ECMs | Office Building in PHL w/ 7 chosen ECMs | Office Building in PHL w/ 8 chosen ECMs

Office Building in SF w/ 6 chosen ECMs | Office Building in SF w/ 7 chosen ECMs | Office Building in SF w/ 8 chosen ECMs

Fig. 6 Average feature selection accuracy with different folds by using LHS and random sampling

The results of the feature selection test are illustrated in Fig. 6. The blue line and the red line in the figure corresponds to the latin-hypercube random sampling and uniform random sampling, respectively. For all the building types and climate zones, the use of LHS for feature selection has a great advantage over random sampling. Most importantly, it appears that using 30 folds of sample size can be sufficient to find the true feature importance ranking and the most influential ECMs can be thus obtained from the proposed eleven ECMs in Table 4. Thirty folds, 868 simulation cases, represents approximately 3.7% of the total combinatorial search space. The proposed method of "LHS feature selection with 30 folds sampling size" makes it possible to focus on the most important ECMs that affects BEU, and reduce the computational cost greatly in the simulation based parametric study.

To reduce the training database size and run simulations containing only the most important ECM combinations in EnergyPlus, the feature selection method is applied to both building types and the seven most influential ECMs are selected according to their feature importance. The ranking of the selected ECMs for each building is shown in Table 5. The objective function for feature selection here refers to the simulated annual energy use based on future "extreme year" hourly weather data. The reason, justification and method to generate the "extreme year" weather data will be introduced in section 3.3.

Table 5 Rankings of ECM importance for each building

Residential building in PHL	**Office building in PHL**	**Residential building in SF**	**Office building in SF**
Air infiltration rate (h^{-1})	Air infiltration rate (h^{-1})	Air infiltration rate (h^{-1})	Air infiltration rate (h^{-1})
Wall Insulation (m^2K/W)	Cooling COP	Wall Insulation (m^2K/W)	Lighting efficiency improvement
Window U-value (W/m^2-K)	Window U-value (W/m^2-K)	Heating efficiency	Window shading
Heating efficiency	Lighting efficiency improvement	Window shading	Cooling supply air temperature (°C)
Window SHGC	Wall Insulation (m^2K/W)	Window U-value (W/m^2-K)	Daylighting control and dimming
Window shading	Window shading	Window SHGC	Window SHGC
Lighting efficiency improvement	Window SHGC	Lighting efficiency improvement	Cooling COP

3.3 Validation of the data-driven model

As described in section 2.4, RF model is trained by the hourly BEU data generated by EnergyPlus. The features involved in the model training are: temperature, relative humidity (rh), solar irradiation, wind speed, and the building occupancy reported by EnergyPlus. Two models are generated to predict the hourly electricity use and gas use respectively. Then randomly sampled ECM combination is selected to validate the prediction results. We used TMY weather data to train the model, but it turned out that the prediction of the future BEU data is biased against the EnergyPlus simulation results in the place as shown in Fig. 7.

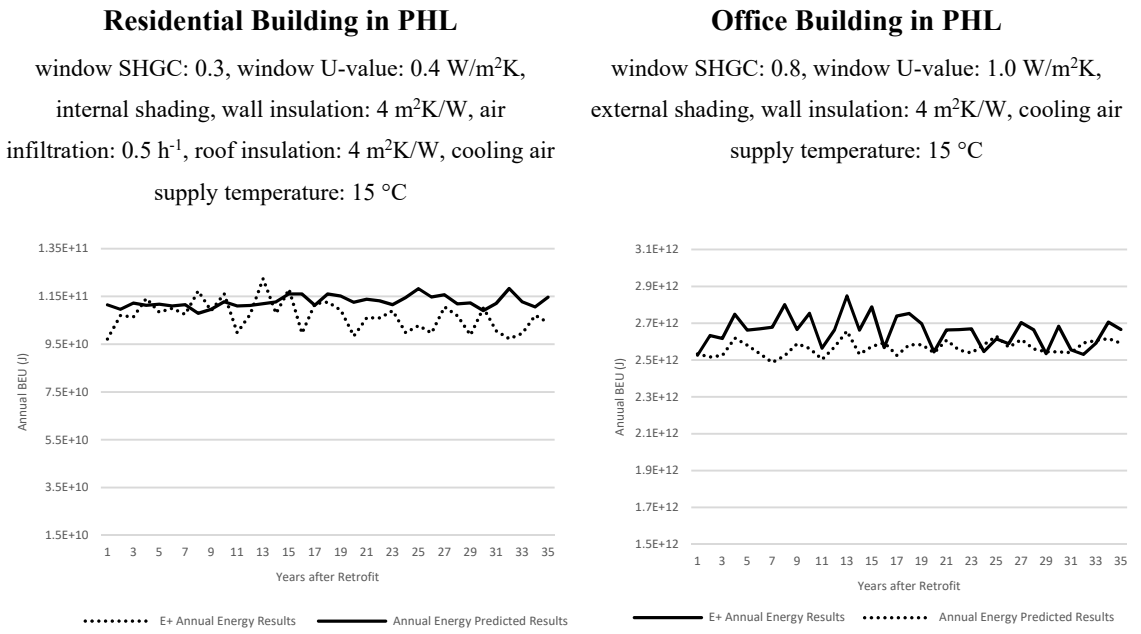


Fig. 7 Biased prediction results of the model for random ECM combination by using TMY weather data for model training

The predictive bias of the model is mainly due to the reason that the weather condition of the coming years are different from those of today. If one tries to predict the future BEU based on the model trained by TMY weather data, the results will be biased. Thus, in order to improve the model prediction for future BEU, we try to construct a year of hourly weather data that contains the extreme weather conditions in the coming years and use that weather data to train the model. Then, EnergyPlus is used to simulate the hourly BEU based on the constructed weather file, and RF is then trained by the database with various ECM combinations. The end of this approach is to allow the regression model to better understand the variance in the future hourly weather conditions so that it can predict future hourly BEU more accurately.

Since temperature and building occupancy are the main driving forces of BEU for most types of building and the latter one is assumed not to change much in the future given that

building occupancy schedule stays stable in future building operations, an “extreme year” weather data is constructed in terms of the extreme temperature in the future years. By means of concatenating half of each future weather data in the year with the most extreme winter (with the lowest temperature in 35 years) and the year of the most extreme summer (with the highest temperature in 35 years), the extreme year weather data is constructed. The construction process for Philadelphia is show in Fig. 8, where half year of the weather data in the year that has the hottest summer and the year that has the coldest winter has been concatenated to form a “pseudo year” hourly weather data. Due to that the constructed data represents a “pseudo year”, the time frame may not be continuous.

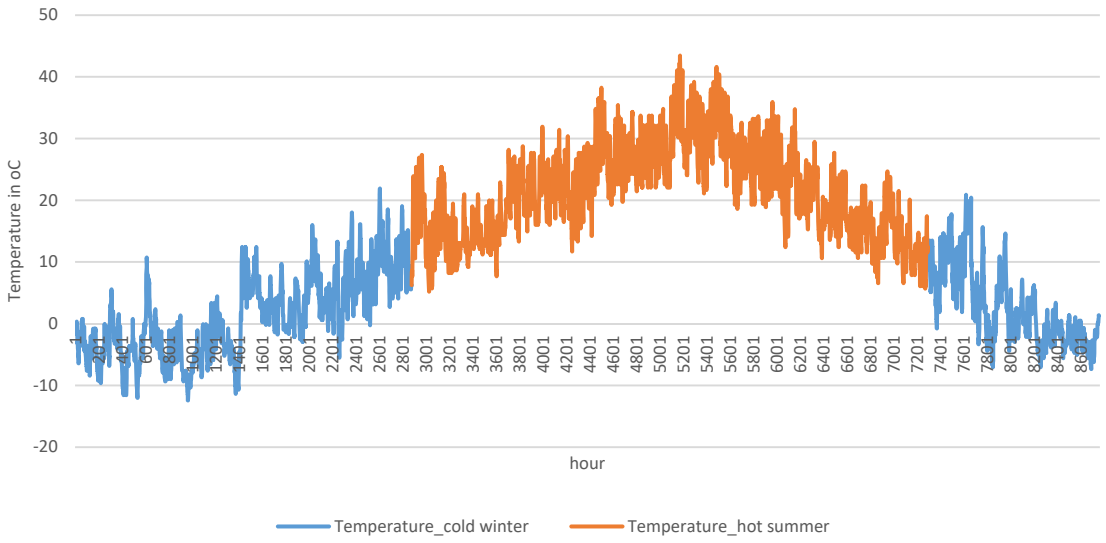
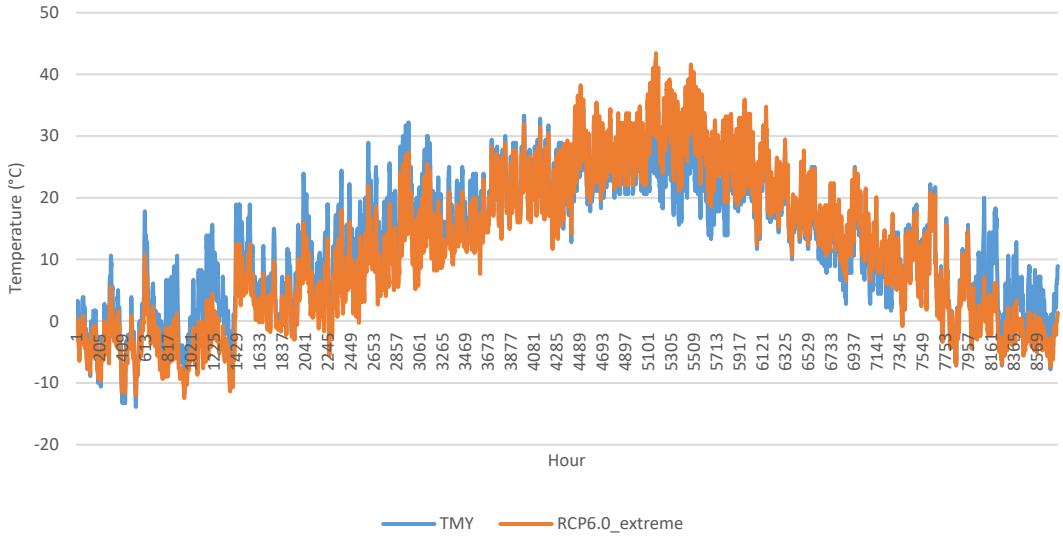


Fig. 8 The construction process of the extreme year hourly weather data (Philadelphia)

The constructed extreme year hourly temperature in Philadelphia and San Francisco are shown in Fig. 9, where the blue line is TMY weather data and the orange line is the constructed extreme year weather data.

Philadelphia



San Francisco

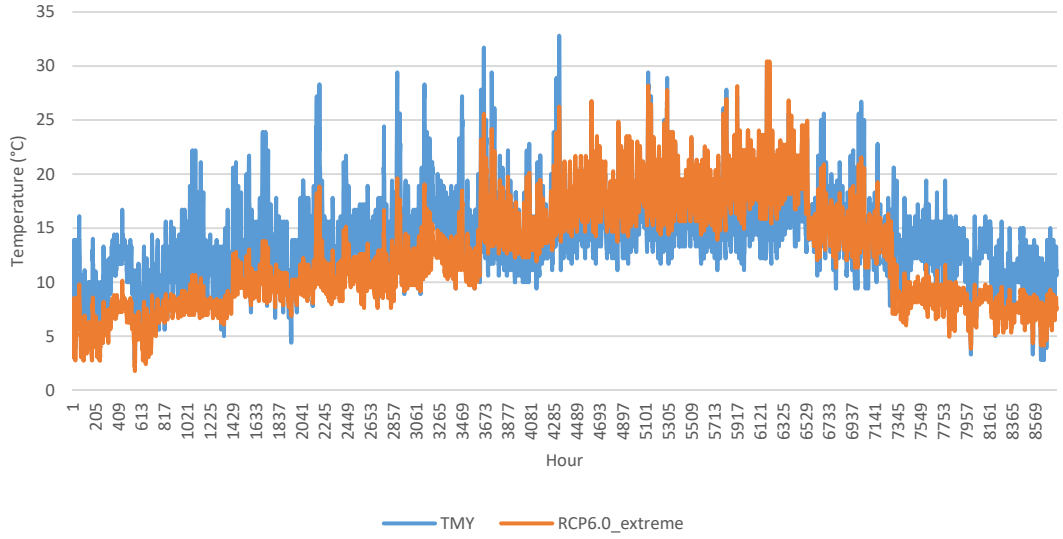


Fig. 9 The constructed hourly temperature profile for extreme weather year

Twenty cases are randomly selected among all the ECM combinations for each building to validate the prediction results by comparing the predictions with the EnergyPlus simulation results in the future years. The validation results of annual BEU during the year of 2020 to 2055 are shown in Table 6. The root mean squared error (RMSE) and coefficient of variation (CV) of the model are used to evaluate the predictive power of the model. RMSE is frequently used to

measure the difference between model predicted value and actually observed value, which can be taken as the residuals of the model, while CV is a description of the fitting of the model in terms of relative sizes of the residuals and outcome values. When CV gets lower, the smaller the residual is compared to the predicted values.

Equation 10: $$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}}$$

Equation 11 $$CV = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} / \bar{x}$$

where x_i and $\hat{x}_i$ is the true and forecasted value, $\bar{x}$ is the average of true values.

In addition, we randomly selected three ECM combinations out of the twenty from Table 5 for each building to illustrate the validation of annual BEU for each building and the results are shown in Fig. 10. According to the results in Table 6, the model trained by future extreme weather data is performing well in predicting future annual BEU. The highest CV for the annual BEU among all the cases in Table 6 is less than 3%, showing good reliability in BEU prediction.

Table 6 Future years' annual BEU validation of twenty randomly chosen ECM combinations for each building

Residential building (PHL)		**Office building (PHL)**		**Residential building (SF)**		**Office building (SF)**	
BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)	BEU_RMSE (J)	BEU_CV (%)
1.18E+09	1.00%	7.45E+10	0.95%	1.06E+09	1.61%	5.57E+10	1.49%
1.40E+09	1.19%	6.40E+10	1.06%	8.73E+08	1.15%	9.07E+09	0.34%
1.25E+09	1.01%	4.57E+10	0.94%	5.07E+08	0.68%	4.47E+10	1.12%
1.23E+09	1.24%	8.66E+10	1.07%	1.04E+09	1.16%	1.35E+10	0.56%
1.14E+09	1.09%	6.87E+10	0.84%	1.01E+09	1.24%	3.42E+10	0.89%
1.32E+09	1.29%	8.06E+10	1.60%	1.86E+09	2.12%	5.71E+10	1.48%
1.52E+09	1.07%	7.34E+10	1.52%	9.89E+08	1.58%	5.37E+10	1.72%
1.76E+09	1.28%	6.75E+10	1.08%	1.16E+09	1.79%	3.56E+10	1.43%
1.46E+09	1.01%	6.59E+10	1.11%	5.83E+08	0.94%	8.30E+10	2.05%
1.40E+09	1.07%	5.35E+10	1.11%	4.26E+08	0.64%	6.70E+10	2.75%
1.23E+09	1.37%	7.84E+10	0.93%	5.42E+08	0.72%	8.34E+10	2.81%
1.95E+09	1.20%	6.52E+10	0.84%	1.01E+09	1.35%	8.00E+09	0.28%
1.48E+09	1.00%	8.31E+10	1.03%	1.04E+09	1.37%	6.25E+10	1.78%
1.88E+09	1.58%	7.88E+10	1.26%	1.41E+09	1.55%	4.75E+10	1.55%
1.39E+09	1.03%	5.29E+10	1.15%	1.40E+09	1.89%	5.30E+10	1.87%
1.42E+09	1.27%	8.31E+10	1.06%	1.08E+09	1.70%	5.89E+10	1.99%
1.72E+09	1.10%	5.10E+10	0.99%	7.80E+08	1.01%	4.01E+10	1.72%
1.88E+09	1.23%	4.34E+10	0.87%	6.01E+08	0.95%	7.23E+10	2.34%
1.18E+09	0.97%	9.08E+10	1.15%	8.12E+08	1.25%	6.25E+10	2.33%
1.35E+09	1.25%	6.17E+10	1.28%	9.49E+08	1.28%	6.65E+10	2.05%

Note: each row's results represent the prediction accuracy of the randomly chosen ECM combination in 2020 to 2055; it should also be noted that the twenty random ECM combinations for each building differ from each other because they are randomly chosen for each building respectively.

3.4 Discussion of future ECM selection

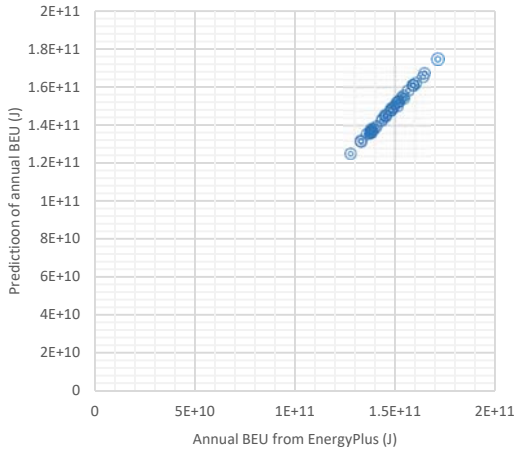
By applying the proposed method to the two typical building types in Philadelphia and San Francisco, important ECMs are selected for each building before the generation of the complete database by running the EnergyPlus simulation using the LHS feature selection with 30 folds sample size, and the future hourly BEU for the year of 2020 to 2055 of each selected ECM combination is projected by RF model. Since the proposed method is validated in section 3.2 and section 3.3, final results are generated and analyzed for each building.

3.4.1 ECM selections for each building

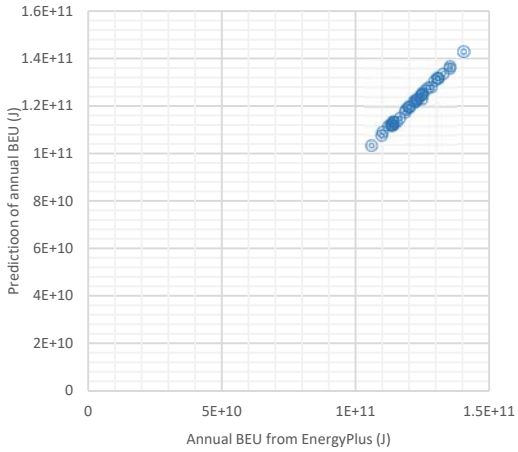
Seven important ECMs are selected using the LHS feature selection with a 30-fold sampling size for each building. The results of selected ECMs shown in Table 5 make sense for each type of the building. The proposed JMIM feature selection method works well in identifying the most important ECMs for a particular building and can be applied in future related research. In order to better illustrate the results shown in Table 5, the end use break down of BEU for each building is plotted in Fig. 11. It should be noted that the reason why heating energy use is taking most of the percentages in total BEU is that the heating site energy source is gas for both types of buildings and gas has a lower site to source energy conversion factor compared to electricity. Usually, the site to source energy conversion factor of electricity is three times higher than gas in the United States (DOE 2014).

Residential Building in PHL

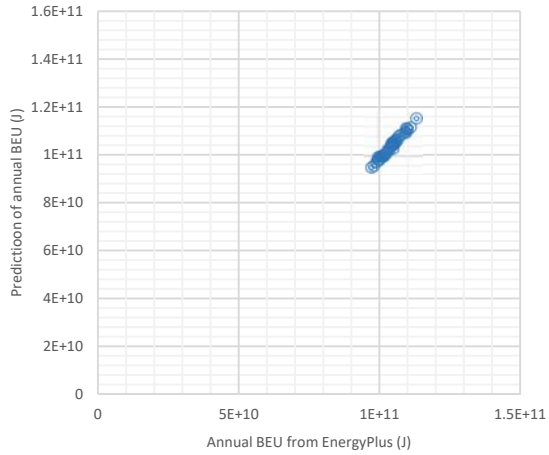
window SHGC: 0.3, window U-value: 1 W/m^{2}K, internal shading, wall insulation: 4 m^{2}K/W



window SHGC: 0.5, window U-value: 0.4 W/m^{2}K, external shading, wall insulation: 4 m^{2}K/W, heating efficiency: 0.9



window SHGC: 0.8, window U-value: 2 W/m^{2}K, wall insulation: 4 m^{2}K/W, air infiltration: 0.5 h^{-1}, heating efficiency: 0.9

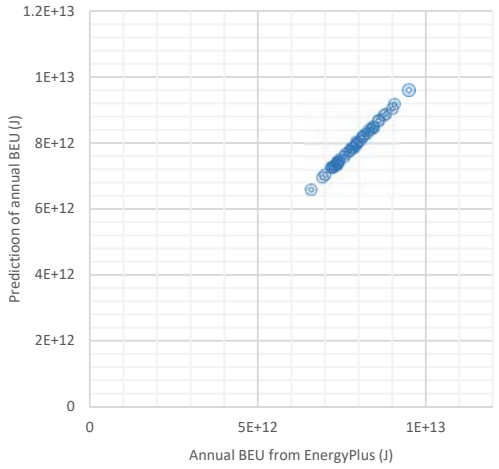


Office Building in PHL

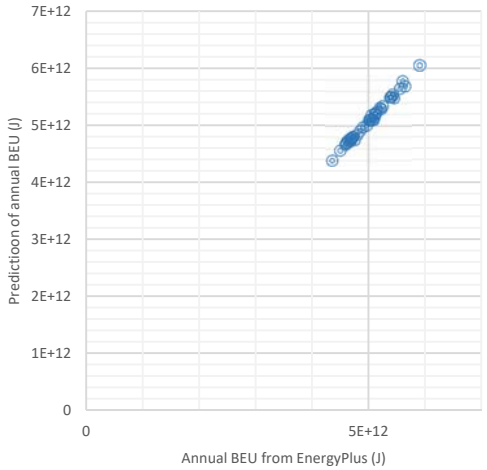
window SHGC: 0.3, window U-value: 0.4 W/m^{2}K, cooling COP: 4.5, lighting: 40%

window SHGC: 0.5, window U-value: 2 W/m^{2}K, wall insulation: 4 m^{2}K/W, air infiltration: 1.5 h^{-1}, lighting: 40%

window SHGC: 0.8, window U-value: 1 W/m^{2}K, internal shading, wall insulation: 2 m^{2}K/W, air lighting: 40%

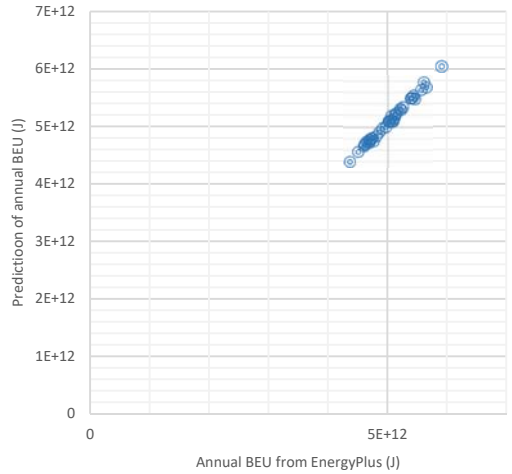


window SHGC: 0.3, window U-value: 1 W/m^{2}K, wall insulation: 4 m^{2}K/W, air infiltration: 0.5 h^{-1}

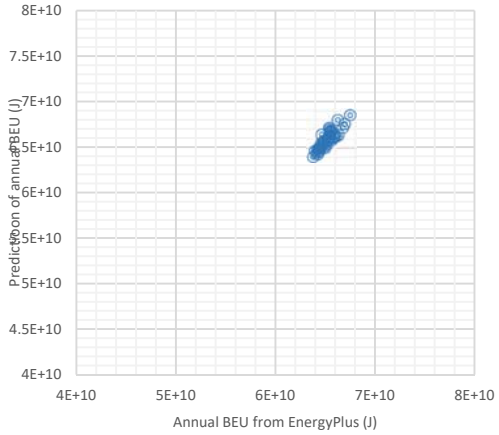


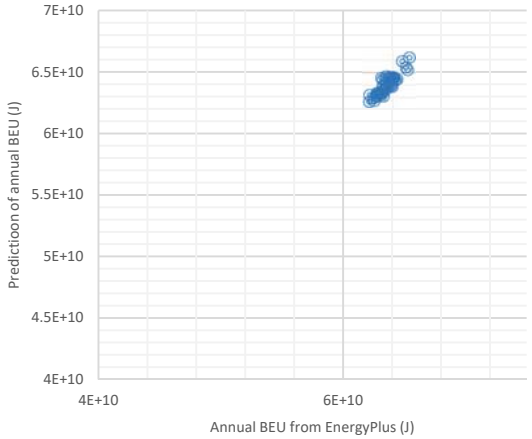
Residential Building in SF

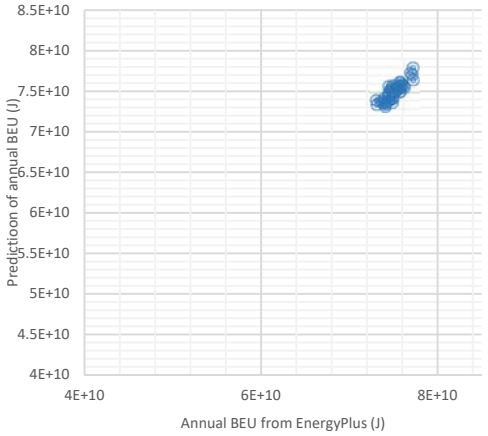
window SHGC: 0.5, window U-value: 1 W/m^{2}K, internal shading, wall insulation: 4 m^{2}K/W, air infiltration: 0.5 h^{-1}, heating efficiency: 0.9



window SHGC: 0.8, window U-value: 2 W/m^{2}K, wall insulation: 4 m^{2}K/W, air infiltration: 0.5 h^{-1},







Office Building in SF

window SHGC: 0.3, window U-value: 1 W/m^{2}K, air infiltration: 2 h^{-1}, lighting: 40%, implemented daylighting control

window SHGC: 0.5, window U-value: 1 W/m^{2}K, internal shading, air infiltration: 2 h^{-1}, cooling COP: 4.5, cooling air supply temperature: 15 °C, lighting: 40%

window SHGC: 0.8, window U-value: 1 W/m^{2}K, external shading, air infiltration: 1 h^{-1}, cooling air supply temperature: 15 °C, lighting: 40%, implemented daylighting control

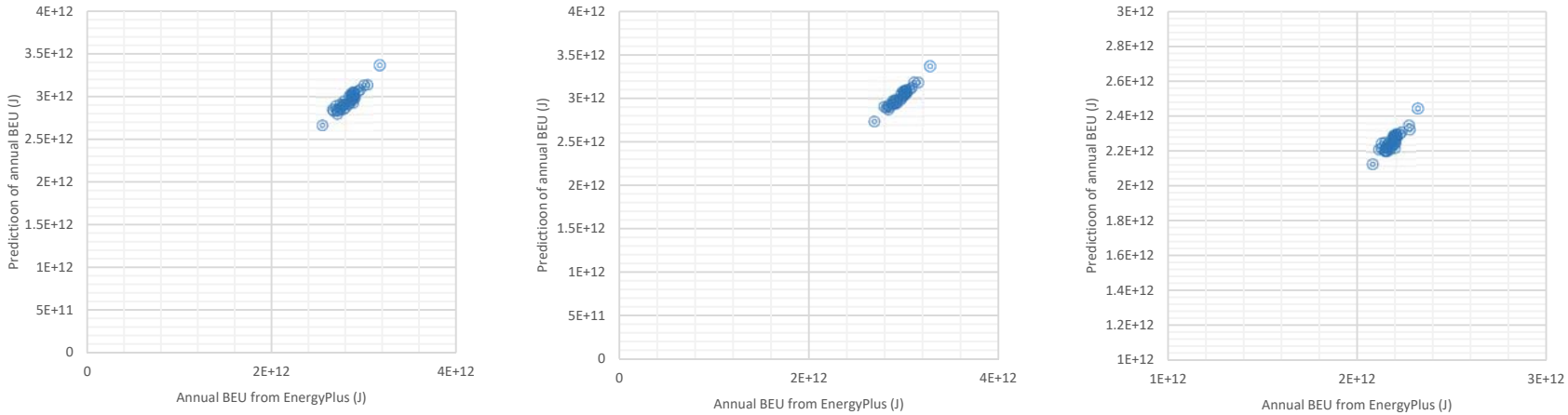


Fig. 10 Validation of annual BEU prediction of three random ECM for each building

Table 7 Rankings of top ten ECM combinations from year of 2020 to 2055

Residential building in PHL									
SHGC	win_U (W/m^{2}K)	shading	wall_insulation (m^{2}K/W)	airInfl (h^{-1})	heat_efficiency	lightings (%)	Elec (J)	Gas (J)	Total (J)
0.8	0.4	2	4	0.5	0.9	40%	1.29E+12	1.96E+12	3.24E+12
0.8	0.4	2	4	0.5	0.9	0	1.45E+12	1.9E+12	3.34E+12
0.8	0.4	2	2	0.5	0.9	40%	1.31E+12	2.05E+12	3.36E+12
0.5	0.4	2	4	0.5	0.9	40%	1.28E+12	2.11E+12	3.39E+12
0.8	1	2	4	0.5	0.9	40%	1.3E+12	2.15E+12	3.45E+12
0.8	0.4	2	2	0.5	0.9	0	1.47E+12	1.99E+12	3.45E+12
0.5	0.4	2	4	0.5	0.9	0	1.43E+12	2.04E+12	3.48E+12
0.8	0.4	2	4	0.5	0	40%	1.29E+12	2.21E+12	3.5E+12
0.3	0.4	2	4	0.5	0.9	40%	1.27E+12	2.24E+12	3.51E+12
0.5	0.4	2	2	0.5	0.9	40%	1.3E+12	2.22E+12	3.52E+12
Office building in PHL									
SHGC	win_U (W/m^{2}K)	shading	wall_insulation (m^{2}K/W)	airInfl (h^{-1})	cool_ efficiency	lightings (%)	Elec (J)	Gas (J)	Total (J)
0.8	0.4	2	4	1	4.5	40%	1.62E+14	2.17E+13	1.83E+14
0.8	0.4	2	2	1	4.5	40%	1.62E+14	2.13E+13	1.83E+14
0.8	0.4	0	4	1	4.5	40%	1.68E+14	1.74E+13	1.86E+14
0.8	1	2	4	1	4.5	40%	1.66E+14	2.14E+13	1.87E+14
0.8	0.4	0	2	1	4.5	40%	1.69E+14	1.77E+13	1.87E+14
0.8	1	2	2	1	4.5	40%	1.66E+14	2.08E+13	1.87E+14
0.5	0.4	0	4	1	4.5	40%	1.68E+14	1.98E+13	1.87E+14
0.8	0.4	2	4	1	0	40%	1.66E+14	2.15E+13	1.88E+14
SHGC	win_U (W/m^{2}K)	shading	wall_insulation (m^{2}K/W)	airInfl (h^{-1})	heat_efficiency	lightings (%)	Elec (J)	Gas (J)	Total (J)

0.5	0.4	2	4	1	4.5	40%	1.66E+14	2.11E+13	1.88E+14
0.8	0.4	1	4	1	4.5	40%	1.7E+14	1.79E+13	1.88E+14
Residential building in SF									
SHGC	win_U (W/m^2K)	shading	wall_insulation (m^2K/W)	airInfl (h^{-1})	heat_efficiency	lightings (%)	Elec (J)	Gas (J)	Total (J)
0.8	0.4	2	2	0.5	0.9	40%	1.23E+12	9.78E+11	2.2E+12
0.8	0.4	2	4	0.5	0.9	40%	1.22E+12	9.96E+11	2.21E+12
0.3	0.4	2	4	0.5	0.9	40%	1.2E+12	1.02E+12	2.22E+12
0.5	0.4	2	4	0.5	0.9	40%	1.21E+12	1.01E+12	2.22E+12
0.5	0.4	2	2	0.5	0.9	40%	1.21E+12	1.02E+12	2.23E+12
0.3	0.4	2	2	0.5	0.9	40%	1.21E+12	1.03E+12	2.23E+12
0.8	1	2	4	0.5	0.9	40%	1.22E+12	1.04E+12	2.25E+12
0.3	1	2	4	0.5	0.9	40%	1.19E+12	1.07E+12	2.26E+12
0.5	1	2	4	0.5	0.9	40%	1.2E+12	1.06E+12	2.27E+12
0.5	1	2	2	0.5	0.9	40%	1.21E+12	1.06E+12	2.27E+12
Office building in SF									
SHGC	shading	airInfl (h^{-1})	cool_efficiency	cooling air temperature (°C)	lightings (%)	daylight control	Elec (J)	Gas (J)	Total (J)
0.8	2	1	4.5	0	40%	1	7.86E+13	3.94E+12	8.26E+13
0.8	2	1	0	0	40%	1	8.04E+13	3.92E+12	8.43E+13
0.3	2	1	4.5	0	40%	1	8.04E+13	3.91E+12	8.43E+13
0.8	0	1	4.5	0	40%	1	8.14E+13	3.07E+12	8.44E+13

SHGC	shading	airInfl (h^{-1})	cool_efficiency	cooling air temperature (°C)	lightings (%)	daylight control	Elec (J)	Gas (J)	Total (J)
0.5	2	1	4.5	0	40%	1	8.1E+13	3.89E+12	8.49E+13
0.8	2	1	4.5	0	40%	0	8.12E+13	3.96E+12	8.52E+13
0.5	2	1	0	0	40%	1	8.13E+13	3.91E+12	8.52E+13
0.5	0	1	4.5	0	40%	1	8.24E+13	3.15E+12	8.55E+13
0.3	0	1	4.5	0	40%	1	8.22E+13	3.77E+12	8.6E+13
0.3	2	1	4.5	0	40%	0	8.22E+13	3.95E+12	8.62E+13

As per the results, air infiltration is the most influential factor that determines the amount of BEU, regardless of building type, indicating that a well-controlled air leakage level can greatly influence BEU performance. For residential building, wall insulation and heating efficiency are important, which implies that the reduction of heating energy use in residential building is vital. It can also be told from Fig. 11 that heating energy use in residential buildings has a higher proportion in total BEU than office buildings, which is logical due to that BEU in residential building is more vulnerable to heat loss than heat gains and that office buildings are more influenced by higher thermal capacity, greater intensity of equipment use and occupant activity. This is also validated by the fact that the improvement of cooling COP is among the top 7 for the two office buildings in Table 5.

For both types of buildings, the improvement of lighting efficiency and the control strategy of natural lighting are more important in office buildings than in residential buildings because lighting energy use in office building has a higher proportion in the end use breakdown.

Window SHGC is a factor that can never be underestimated in impacting BEU for both types of building, and the same thing happens for window shading because it is known that the heat transfer through the window is always decisive for the building heating and cooling energy use. In this study, window shading is achieved by installing blinds inside or outside the window and both installations are considered and analyzed. In addition, the heating efficiency of residential building should receive more attention than cooling, while improving the cooling efficiency of office buildings should be emphasized.

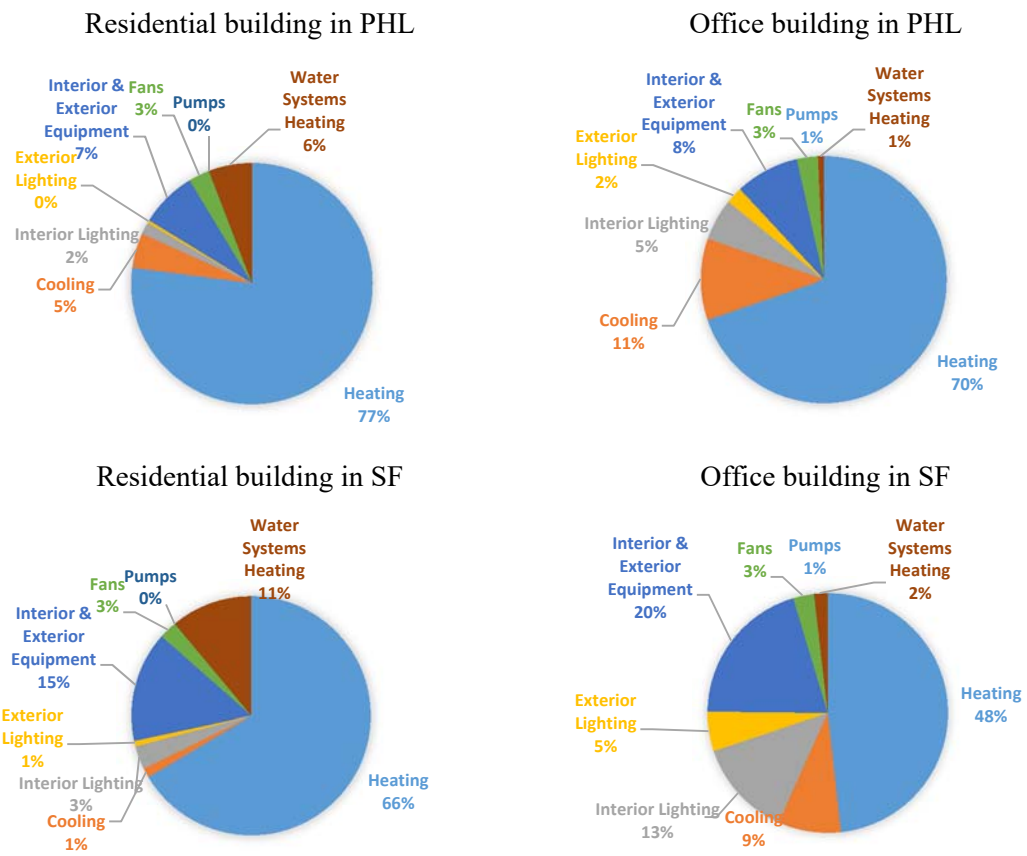


Fig. 11 End use breakdown of BEU for the four buildings

3.4.2 Future years' ranking of ECM combination

As described in section 3.3, eight RF models are trained in total, two for each building being trained by electricity use and gas use as predictands, whereas future extreme year weather data and ECM parameters as predictors. The hourly BEU for each combination of selected ECMs is projected from 2020 to 2055. Ranking by their total site energy use during the year of 2020 to 2055, the top ten ECM combinations that save the most energy in the future 35 years are presented in Table 7.

Based on the results in Table 7, the best controlled air infiltration rate (0.5 h^{-1}) saves the most energy for residential buildings in both Philadelphia and San Francisco during the next 35 years. For office buildings, better airtightness is required in San Francisco and Philadelphia (1 h^{-1}). In addition, whether in residential or office, buildings with external blind shading are always among the top listed ECMs regarding energy saving rankings. The window shading system in this research is set to work when the indoor environment is subject to high cooling load, and it is usually not turned on in heating season, so sun light can enter during most of the time in winter. This measure significantly reduces cooling load during cooling season as GCC may create extreme outdoor weather conditions during the summer over the next 35 years.

For the four buildings in Philadelphia, SHGC with a value of 0.8 dominates all other SHGC values in top-level ECM combinations. This is beyond the common expectation that buildings in Philadelphia should have a low SHGC value for the window, especially in future conditions. Even for the office building in Philadelphia that has a higher internal heat gain than the residential building, high SHGC value still saves the most energy. Low window U-value retains heating energy during the cold winter in Philadelphia, and even with the impact of, this situation is not going to change over the next 35 years.

For buildings in San Francisco, it is quite difficult to define which SHGC value is the most dominant over others for both residential and office buildings. The decision-making of the best SHGC can be clearer when other factors such as economic analysis are introduced. Low window U-value in residential building is always welcome in the future climate scenario, especially for buildings in Philadelphia.

When it comes to the renovation of building envelope, better wall insulation will save energy for buildings in Philadelphia due to its cold winter, while the renovating of wall insulation will not be among the most selected ECMs for the office building in San Francisco. The results

indicate that residential buildings in both climate regions should pay attention to wall insulation even though GCC will reduce future heating energy use for both cities.

3.4.3 Ranking change of ECMs in the future

Analyzing future top ECM combinations is not sufficient to know the impacts of GCC on the decision-making process for building retrofit. We also compared the ranking changes of each ECM combination in future 35 years against in TMY. We used EnergyPlus to run the selected ECM combinations under the TMY and compare its rankings with the results of rankings in future years for the four buildings and listed the ranking changes in Table 8. The ranking changes in the table are calculated by the following function:

Equation 12: $$RCECM_P = \frac{\sum_{i=1}^{n}(R_{i,future} - R_{i,TMY})}{n}$$

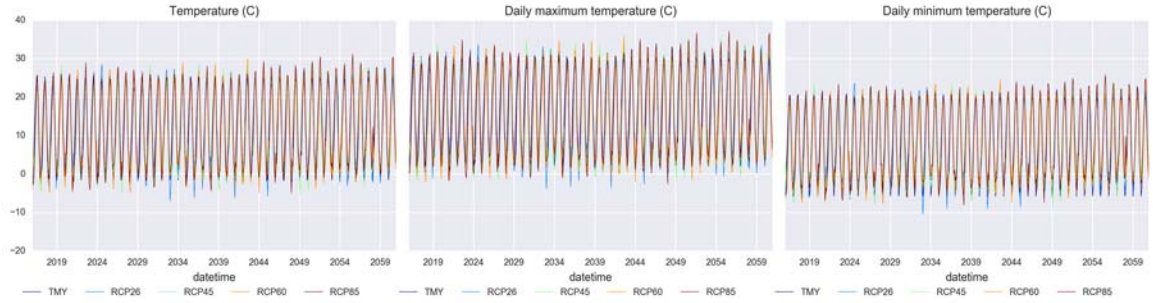
where, $RCECM_P$ is the rank change of ECM with certain parameter value P; n is the total number of ECM combinations that have the ECM parameter P; $R_{i,future}$ is the ranking of the i^{th} ECM combinations that has parameter P by total BEU in the future; $R_{i,TMY}$ is the ranking of the i^{th} ECM combinations that has parameter P by total BEU in single TMY year.

Table 8 Ranking changes of ECM combinations for four buildings compared with TMY scenario

ECM	Parameter	San Francisco		Philadelphia	
		Residential	Office	Residential	Office
Window SHGC	0.3	12.5	2.82	-5.77	-2.48
	0.5	11.47	-0.34	-8.02	-1.46
	0.8	-23.96	-2.48	13.8	3.94
Window U-value	0.4	0.47	-	-2.45	4.98
	1	2.04	-	-0.7	0.85
	2	-2.5	-	3.15	-5.84
Window shading	N/A	-23.03	-2.06	8.65	-1.88
	Internal Blind	-5.76	0.33	3.66	1.53
	External Blind	28.79	1.72	-12.3	0.35
Wall insulation	N/A	-20.85	-	5.9	2.85
	2	9.91	-	-5.54	1.3
	4	10.94	-	-0.35	-4.15
Air infiltration for residential	N/A	58.22	-	-33.2	-
	0.5	-30.51	-	5.3	-
	1	-27.71	-	27.9	-
Air infiltration for office	N/A	-	0.22	-	0
	1	-	-0.88	-	-0.01
	2	-	0.65	-	-0.01
Roof insulation	N/A	-	-	-	-
	3	-	-	-	-
	6	-	-	-	-
Heating efficiency	N/A	-8.23	-	0.21	-
	0.9	8.23	-	-0.21	-
Cooling COP	N/A	-	-	-	1.83
	4.5	-	-	-	-1.83
Cooling air temperature	N/A	-	2.67	-	-
	15	-	-2.67	-	-
Lighting efficiency	N/A	-20.4	-0.37	-4.15	5.84
	40%	20.42	0.37	4.15	-5.84
Daylighting control	N/A	-	-0.43	-	-
	Applied	-	0.43	-	-

Note: "-" means that the ECM is not selected by JMIM feature selection procedure or does not apply to the specific building type

Philadelphia



San Francisco

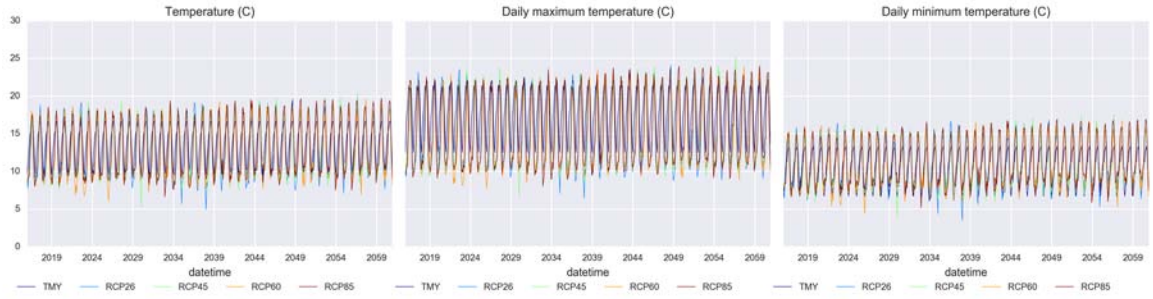


Fig. 12 Downscaled monthly mean temperature, daily maximum and minimum temperature under different RCPs (2015 to 2069) and TMY in Philadelphia and San Francisco

It is not difficult to see that residential buildings are more vulnerable to climate change since their ECM ranking changes are relatively larger than office buildings. For residential buildings in Philadelphia, a higher SHGC is more preferable in future years as shown in Fig. 12, which is due to the fact that GCC not only raises the outdoor temperature in summer, but also creates more extreme winter conditions. Thus, the decrease in heating energy due to the increase of SGHC could possibly offset the increase in cooling energy in summer. A higher window U-value also provides better insulation for the building in winter and reduces heating energy use. For the same reason, less shading and wall insulation are needed in future climate for the residential building in Philadelphia. The residential building does not need to be so airtight relative to the TMY scenario but needs to be retrofitted to maintain an air infiltration rate at 1 h^{-1}.

For the office building in Philadelphia, lower window U-value is preferred in the future than today. Better wall insulation is not so important compared to the current climate. Rankings of retrofits that involve increasing cooling COP and lighting efficiency are slightly decreasing, but according to Table 7, they are still very important in high-ranking ECM combinations because only the relative changes are reflected in the table in future climate condition.

For buildings in San Francisco, the most important finding is the big change in air infiltration rate rankings. Given that the air infiltration rate is the most important factor influencing the BEU as indicated by feature importance analysis, the magnitude in its ranking changes could alter the picture of future retrofit decision-making. As shown in Table 8, both residential and office building tend to be less air tightened in the future than in TMY condition, mainly because of the rise in outdoor temperature under future climate condition, as shown in Fig. 12. Moreover, windows with low SHGC is more preferred in the future compared to TMY condition in San Francisco and the building is needed to gain less heat from the sun. This can also be reflected by the fact that the exterior shading is also the most valued parameter in San Francisco's future climate, as it is able to best reduce solar heat gain during period of high cooling load. In the meantime, a lower U-value and better wall insulation are preferable to better insulate the building to reduce the heating energy use for residential buildings, while insulation is not an important factor in impacting the BEU in office building in San Francisco as they are not chosen by feature selection. Daylighting control and improvement in lighting efficiency are slightly more preferable in the future climate.

In conclusion, the change of preference for ECM parameters in building retrofit in San Francisco in the future against current climate condition is to make the building less airtight, to reduce the solar heat gain, and to improve thermal insulation, while for buildings in Philadelphia, more solar heat gain, less thermal insulation will be more effective to save BEU.

4. Development of Simplified Building Modeling Tool

Building energy simulation (BES) can be used to study how the building will perform under different design and engineering scenarios, such as different building thermal properties (Hillary et al. 2017), occupancy behavior (Monteiro, Fernández, and Freire 2016), changing weather conditions (Spandagos and Ng 2017), energy supply systems (Shen and Lior 2016) as well as the short-term predictive control method of the building system (Kwak and Huh 2016, Li, Wen, and Bai 2016), and etc. Most BES tools require very detailed inputs for the model because of the nature of the building performance, which usually involves many driving factors and uncertainties. Tools like EnergyPlus, TRNSYS, BLAST uses a transient method to simulate building heating and cooling loads by dynamically integrating all heat flows into calculation without simplification, making the modeling and simulation with good reliability and detailed yet complicated and heavy to use and in the meantime requiring abundant professional knowledge and modeling experience.

Another important issue is that when it comes to the computational complexity of BES tools using transient heat transfer calculation method, the tools would be expensive in terms of computation to tackle problems such as retrofit optimization. For comparative research regarding different scenarios of active and passive building systems instead of looking for very specific operational parameters of building systems, these tools can over qualify and waste unnecessary computing resources. Especially in parametric studies, for example, which aim to find the optimal combination of energy conservation measures (ECM) for an existing building, the potential combinatorial nature of the problem can have these tools consume an unaffordable time in finding the solutions (Rysanek and Choudhary 2012a). Recently, data-driven models have been used by

researchers to circumvent this problem by training models using machine learning algorithm and verifying the accuracy of data-driven methods (Eisenhower et al. 2012). However, for this method, a major challenge in huge consumption of computational resources lays in generating the training database by results of transient simulation. The same problem exists when using heuristic optimization method in finding the optimum of the objective function specified in relation to building energy use, thermal comfort level, and economic benefits. The process of evaluating the objective function usually involves transient BES tools, making the optimization process computationally expensive and unreliable since an insufficient population size and iteration steps can deteriorate the final optimum of the heuristic search and the convergence of the solution.

Research has been conducted to find lighter modeling methods in providing answers that are sufficient for comparative study in building heating and cooling load and energy use. One of them uses electrical analogue to model the thermal behavior of the building, which is better known as RC (resistance & capacitance) model. People used RC method to model the thermal reaction of the building under the synergy of indoor and outdoor conditions. Berthoua et al.'s work tried four gray box models of different complexity that have been tested and evaluated to simulate the cooling and heating needs of a multi-zone office building. The simulation results show that the two-order 6R2C semi-physical model offers the best compromise between all tested models. It is able to predict thermal needs and indoor air temperature during heating and cooling periods with an accuracy above 84% (Berthoua et al. 2014). In Terés-Zubiaga et al.'s work, a sophisticated RC gray model is developed for the dwelling using monitoring data. First, the thermal performance of an empty social housing dwelling had been monitored for 3 months. Afterwards, a gray box model development was carried out using obtained monitoring data. Model development as well as some general model results are presented and evaluated later (Terés-Zubiaga et al. 2015). Asadi et al. developed a multi-objective mathematical model to assist stakeholders make decisions on searching for energy-minimizing ECMs in a cost effective

manner (Asadi et al. 2012a). They developed an R-C model to simultaneously evaluate the effectiveness of all available combinations of retrofit actions. Liao and Dexter have developed a second-order gray physical model to simulate the dynamic behavior of the existing heating system of a multi-zone residential building (Liao and Dexter 2004). Researchers at Tongji University used 3R2C modeling method with an additional parallel structure that replaces the original serial model to describe the building internal mass. Under the help of sub metering data and starting from simplest 3R2C, a more complex RC model was formed and validated to have good prediction performance for a commercial building in Shanghai (Ji et al. 2016). In conclusion, simple low-order RC model had been verified to have a good performance in modeling the heating and cooling need of a simple building that usually has one zone or single use.

The complex RC model with high order is able to handle more sophisticated buildings with multi zones but requires more computations compared with the lower-order RC model since the root search process will be more complicated when the order of the RC model grows. Another problem is that most research using RC modeling method have not developed a simulation interface that makes the simulations of different buildings easy and feasible. The generalization of RC method to different buildings and a simple tool to use the method are undertapped.

ISO 13790 provides a monthly method and a simplified hourly method that uses 5R1C modeling method to calculate heating and cooling needs for buildings (ISO 2008). The method can be called a normative method because it adopts normative values for certain model variables and characteristics that are regressively obtained from buildings in Europe. The method was adopted to model campus building and campus level energy use by researchers at Georgia Institute of Technology (GIT) (Lee, Zhao, and Augenbroe 2013). An Excel based calculation tool was also developed by the research team in GIT (Lee, Zhao, and Augenbroe 2013), which is

called Energy Performance Standard Calculation Toolkit (EPSCT) where the monthly and simplified hourly method described in the standard are implemented. The simplified methods of ISO 13790 have been verified in many research for its performance in modeling the monthly and annual building energy use (Kokogiannakis, Strachan, and Clarke 2008, Kalema et al. 2008, Hasan 2007, Kokogiannakis 2007, Jokisalo and Kurnitski 2007). The conclusion for the monthly method is that it is generally able to give accurate results in calculating the annual energy use while the normalized gain utilization factor may lead to the failure in model accuracy for certain types of building (like light-weight building). In the research of (Kokogiannakis, Strachan, and Clarke 2008) and (Kokogiannakis 2007), the authors also used the simplified hourly method and compared the results of different parametric combinations with those of ESP-r and Energyplus, and found that the results of hourly method generally agree on the annual energy use with the reference models but in some cases they vary significantly. In reference (Hasan 2007), simplified hourly method was also tested for the modeling accuracy, which resulted in up to 25% underestimation and up to 30% overestimation from the reference results by IDA-ICE building energy software. Burhenne and Jacob (Burhenne 2008) tested the simplified hourly method to fit the model to actual measured heating energy use as well as indoor temperature. The results show that simplified hourly method is capable of modeling the annual sum of heating energy, but its performance in fitting to hourly energy use is limited (R^2 of 0.67). More importantly, the simplified hourly method in most of the research as well as in the implementation in EPSCT is merely tested for buildings that have a constant heating and cooling set point or a single zone and zone thermal interaction was not considered, which makes the application of simplified hourly method in simulating mixed-use buildings constrained. Without zone thermal coupling, the performance of the method in hourly scale will be weakened as there are buildings that have multi-purpose of use, and evaluating the adoptions of energy or cost saving measures like demand response, or onsite renewable energy systems that require certain accuracy in hourly load will not

be viable. The possibility of including the zone thermal coupling method should be explored and discussed.

It is also worth noting that an updated version of ISO 13790 --- ISO 52016, has been released in 2017, June, which uses the same RC model as the core for simplified hourly calculation of sensible energy use for heating and cooling and a new method for calculation of latent energy use for (de)humidification was added. However, it does not affect the test for zone thermal coupling process in this research. Other additional applications in the new standard include (ISO 2017):

— calculation of internal temperatures, e.g. under summer conditions without cooling or winter conditions without heating;

— calculation of design heating or cooling load.

Another difference is that the building elements are not aggregated to a few lumped parameters, but kept separate in the model, just as shown in Fig. 13 (Dick van Dijk 2016):

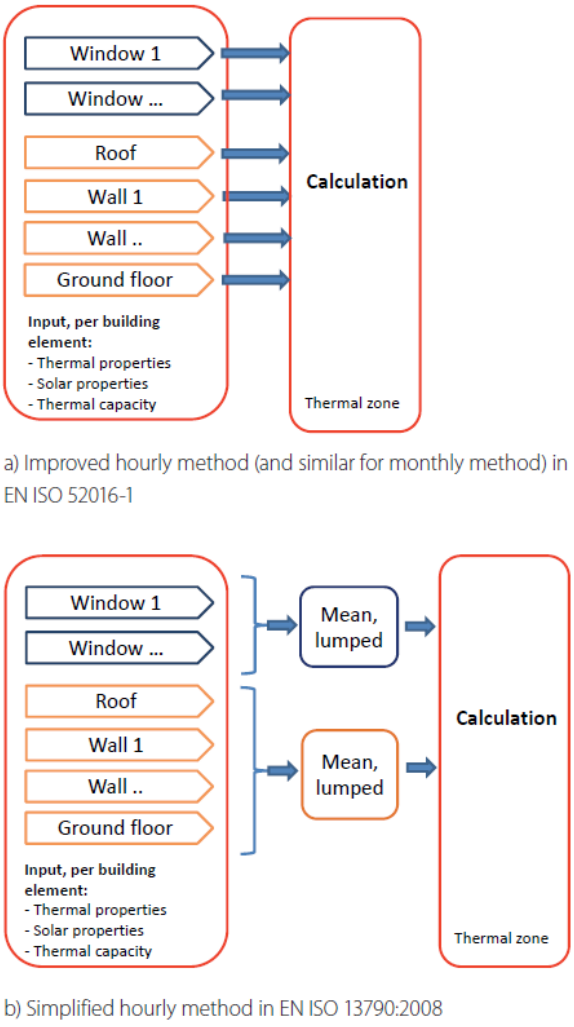


Fig. 13 Improved hourly method in EN ISO 52016-1 (b) compared to simplified method in EN ISO 13790:2008 (a).

Although the major revision in ISO 52016 is a more transparent modeling method for each component of the building envelopes instead of lumping different walls or windows respectively into a single resistor, it also create more complexity in the building information input and to the weight of the model in view of that different boundary conditions for each modeled part should be given and taken into account during simulation. Therefore, considering that the purpose of this research is to find a comparative parametric study tool that is simple to model and light in calculation, the modeling method in ISO 13790 is kept and used here.

In this chapter, a Python based simulation tool that provides modeling and calculation of building energy use based on ISO 13790 simplified hourly method in which zone thermal coupling is applied is proposed. The development of the tool aims to give researchers and professionals a platform to implement simplified, dynamic hourly method using a 5R1C thermal modeling method for modeling different buildings where multi-zones exist and various purpose of use and thermal set points are applied.

4.1 Modeling Methods

The core of the modeling tool is programmed in Python (shown in upper part of Fig. 14). The method adapts electric analogy to simulate the dynamic characteristics of the building thermal behavior. The thermal circuit follows Kirchhoff's law and is able to provide solutions for the heat flux and the temperature of each time step at three nodes.

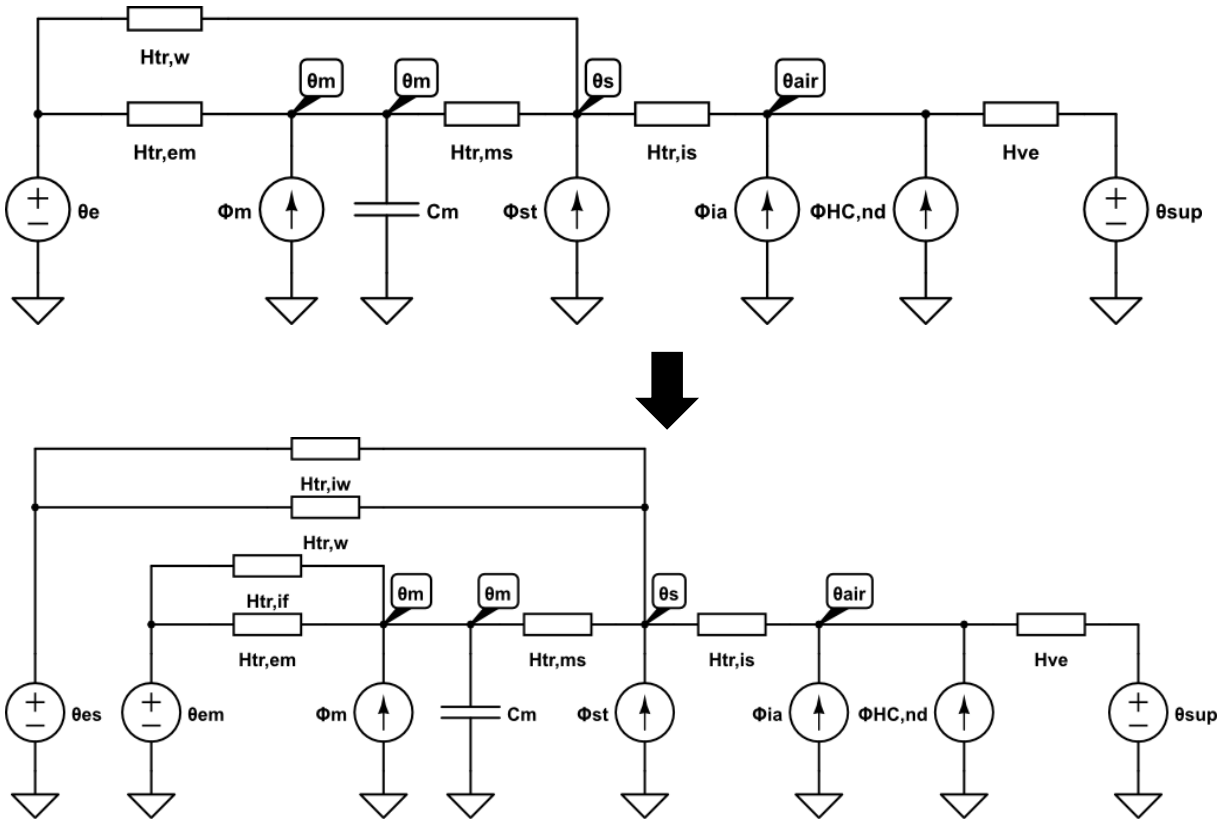


Fig. 14 5R1C model used for the simplified hourly method from non zone thermal coupling (upper) to zone thermal coupling (lower)

The main variables in the modeling method where zone thermal coupling is not considered, are C_m (internal thermal capacity per building area of the considered building, in J/K-m^2), $H_{tr,op}$ (transmission heat transfer coefficient of the opaque building elements like walls, and roofs, in W/m^2K), $H_{tr,w}$ (transmission heat transfer coefficient of windows and glazed walls in W/m^2K), $H_{tr,em}$, $H_{tr,ms}$ (transmission heat transfer coefficient of the internal structure and external structure, respectively, in W/m^2K), $H_{tr,ve}$, $H_{tr,is}$ (transmission heat transfer coefficient of ventilated air, and that between the air in the building and internal structures, respectively, in W/m^2K). $H_{tr,em}$ is calculated by $H_{tr,op}$ and $H_{tr,ms}$ in the following way:

Equation 13: $$H_{tr,em} = \frac{1}{\frac{1}{H_{tr,op}} - \frac{1}{H_{tr,ms}}}$$

The heat flux sourced from solar, building heating and cooling are $\Phi_{sol}, \Phi_{HC,nd}$, and those to the internal air node, to the central node, and to the internal mass node are named as $\Phi_{ia}, \Phi_{st}, \Phi_m$, respectively, in W. The temperature variables for the model are $\theta_e, \theta_{air}, \theta_m, \theta_s$, θ_{sup}, $\theta_{H,set}, \theta_{C,set}$, standing for outdoor temperature, internal air temperature, building thermal mass temperature, mean instantaneous temperature of internal surfaces that are in contact with internal air, supply air temperature, heating set point temperature, and cooling set point temperature, respectively, in ℃.

The three important nodes of the model are internal air node, central node, and internal mass node. For the internal air node, it is governed by the heat balance of heating and cooling load input, the heat flow from the internal air that is affected only by internal heat gain Φ_{int}, and the heat flow from ventilated air θ_{sup}*H_{ve}. The thermal electric balance equation is as follows:

Equation 14: $$\theta_{air}\left(H_{tr,is} + H_{ve}\right) = H_{tr,is}\theta_s + H_{ve}\theta_{sup} + \Phi_{HC} + \Phi_{ia}$$

For the central node, the heat flow is made up of the sum of internal mass heat flow, heat flow from internal load and solar gain, heat flow from external structure, and the combined heat flow from internal air node, which can be described by:

Equation 15:

$$H_{tr,ms}\theta_m + \Phi_{st} + \theta_e H_{tr,w} + \frac{1}{\frac{1}{H_{tr,is}} + \frac{1}{H_{ve}}} * \left(\theta_{sup} + \frac{\Phi_{ia} + \Phi_{HC}}{H_{ve}}\right) = \theta_s (H_{tr,ms} + H_{tr,w} + \frac{1}{\frac{1}{H_{tr,is}} + \frac{1}{H_{ve}}})$$

For the internal mass node, it is balanced by the heat flow from external structure, from the internal mass capacitance, from the internal mass, and the combined heat flow from central node and internal air node, which can be described by the following equation:

Equation 16:

$$C_m \frac{d\theta_m}{dt} + \left(\frac{1}{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}}+H_{tr,w}} + \frac{1}{H_{tr,ms}}} + H_{tr,em} \right) \theta_m = \Phi_m + H_{tr,em}\theta_e +$$

$$\frac{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}}+H_{tr,w}}}{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}}+H_{tr,w}} + \frac{1}{H_{tr,ms}}} \left(\Phi_{st} + H_{tr,w}\theta_e + \frac{1}{\frac{1}{H_{tr,is}}+\frac{1}{H_{ve}}} \left(\theta_{sup} + \frac{\Phi_{ia}+\Phi_{HC}}{H_{ve}}\right)\right)$$

Equation 13 to Equation 16 describe the thermal flow balance of the model without zone coupling. When zone thermal coupling is considered, the 5R1C circuit turns out to be the lower one that is presented in Fig. 14. $H_{tr,iw}$ and $H_{tr,if}$ are introduced to represent the transmission heat transfer coefficient of the internal wall and internal floor, in W/m^{2}K. It should be noted that the m^2 in the unit refers to per condition floor area of the zone, instead of per area of the material of the contact surface. Internal wall heat transfer coefficient is coupled with window heat transfer coefficient, while internal floor transfer coefficient is couple with the heat transfer coefficient of the external structure. Instead of using θ_e (outdoor air temperature), the coupled part of the circuit will be using an equivalent temperature that reflects the thermal condition on the other side of the coupled surface. For the coupled part of internal wall and window, the equivalent temperature θ_{es} would be:

Equation 17: $$\theta_{es} = \frac{\theta_e H_{tr,w} + \sum \theta_{az,i} H_{iw,i}}{H_{tr,w} + \sum H_{iw,i}}$$

Then the equivalent temperature θ_{em} of the coupled part of internal floor and external structure would be:

Equation 18: $$\theta_{em} = \frac{\theta_e H_{tr,em} + \sum \theta_{az,i} H_{if,i}}{H_{tr,em} + \sum H_{if,i}}$$

where i represents the i[th] adjacent zone that has contact with the current zone, and $\theta_{az,i}$ stands for the i[th] zone's internal air temperature.

The coupling process will not change the heat flow balance of the internal air node, but that for central node and internal mass node will be affected. After coupling internal wall and internal floor into the model, Equation 15 and Equation 16 then turn out to be:

Equation 19:

$$H_{tr,ms}\theta_m + \Phi_{st} + \theta_e H_{tr,w} + \sum \theta_{az,i} H_{iw,i} + \frac{1}{\frac{1}{H_{tr,is}} + \frac{1}{H_{ve}}} * \left(\theta_{sup} + \frac{\Phi_{ia} + \Phi_{HC}}{H_{ve}}\right) =$$

$$\theta_s\left(H_{tr,ms} + H_{tr,w} + \sum H_{iw,i} + \frac{1}{\frac{1}{H_{tr,is}} + \frac{1}{H_{ve}}}\right)$$

Equation 20:

$$C_m \frac{d\theta_m}{dt} + \left(\frac{1}{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}} + H_{tr,w}} + \frac{1}{H_{tr,ms}}} + H_{tr,em} + \sum H_{if,i} \right) \theta_m = \Phi_m + \theta_e H_{tr,em} + \sum \theta_{az,i} H_{if,i} +$$

$$\frac{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}} + H_{tr,w} + \sum H_{iw,i}}}{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}} + H_{tr,w} + \sum H_{iw,i}} + \frac{1}{H_{tr,ms}}} \left(\Phi_{st} + \theta_e H_{tr,w} + \sum \theta_{az,i} H_{iw,i} + \frac{1}{\frac{1}{H_{tr,is}} + \frac{1}{H_{ve}}} \left(\theta_{sup} + \frac{\Phi_{ia} + \Phi_{HC}}{H_{ve}}\right)\right)$$

In this coupling method, only heat transmission between zones are considered, the coupling of infiltration or air flow between zones are not considered in this model.

In this modeling method, a concept of free floating air temperature $\theta_{air,free}$ will be used to describe the indoor air temperature of the zone when heating and cooling are not provided, and the heating and cooling need of the zone space is assumed to be always satisfied, which leads to the following three situations:

1) When cooling is needed $(\theta_{air,free} > \theta_{C,set})$, HVAC system will provide enough cooling energy to make $\theta_{air} = \theta_{C,set}$
2) When heating and cooling energy is not needed ($\theta_{air,free} < \theta_{C,set}$ & $\theta_{air,free} > \theta_{H,set}$), the zone indoor air temperature will be the free floating air temperature $\theta_{air} = \theta_{air,free}$
3) When space heating is needed $(\theta_{air,free} < \theta_{H,set})$, HVAC system will provide enough heating energy to make $\theta_{air} = \theta_{H,set}$

The assumption that the indoor air temperature will always be met by the HVAC system implies maximum flexibility in HVAC system control and no dynamic factor will be taken into account in the HVAC control, which will make the HVAC system work in an ideal state. With regard to the primary energy consumption of heating and cooling, a performance curve method will be adopted in this tool. The user will be asked to provide the energy efficiency of the heating and cooling source at 20%, 40%, 60%, 80%, 100% partial load conditions. This measure is to simulate the energy performance of the heating and cooling system under different partial load conditions. After having the inputs of energy efficiency at each stage of the partial load, a linear interpolation will be made to emulate a performance curve of the system, and this processing is intended to simplify model inputs. The pump system model in this tool assumes that the pumps operate in a constant flow state, and its mass flow rate is calculated by the flow rate required for peak heating and cooling load. Thus, if ECMs that reduce the building heating or cooling load are adopted for a building, the pump energy use will also be saved if upgrading pumps are chosen as one of the ECMs.

4.2 Structure of the Tool

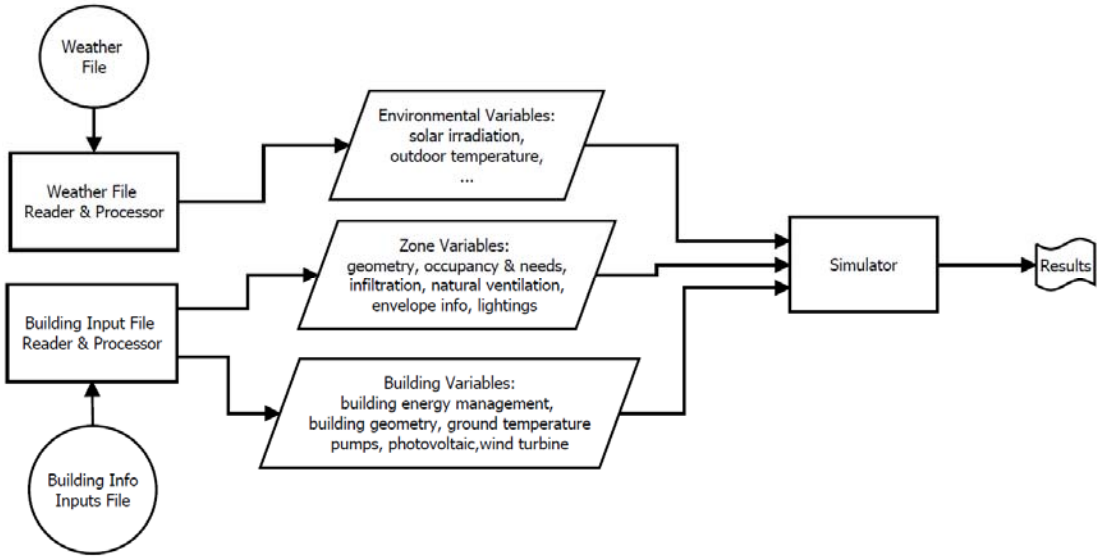


Fig. 15 Flow chart of the simulation tool

Fig. 15 shows the flowchart of how the tool is organized for detailed building simulation using simplified hourly method. The weather file reader and processor will read the designated weather file (.epw). After having useful weather variables, pre-calculations for hourly direct radiation, diffuse radiation, reflected radiation, and global insolation will be carried out for each orientation of the building. These values and other variables such as outdoor temperature, relative humidity, calculated solar azimuth and altitude degree will be used later to calculate heat gain of the building. The solar heat gain calculation will be performed in the simulator according to ISO 13790.

The building input file is text based, which ends with the extension of ".sim". The hierarchy of how the inputs are organized is described in Fig. 16:

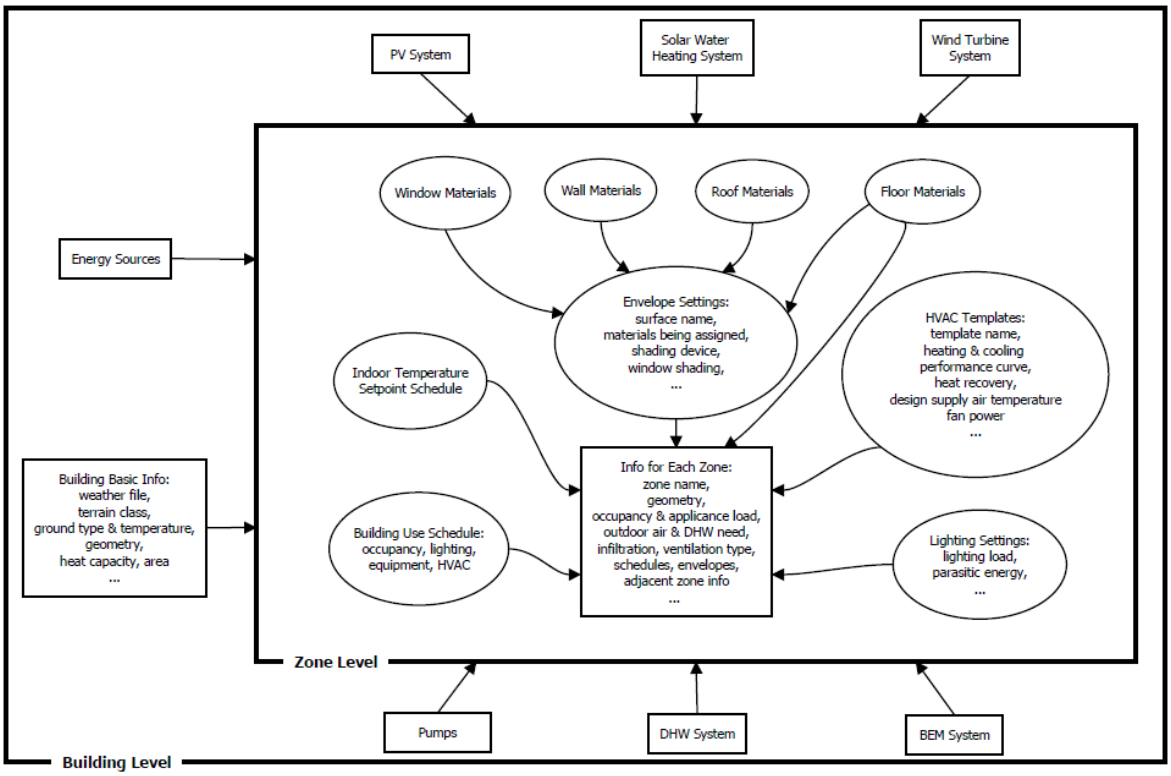


Fig. 16 Hierarchy of the building inputs information

The inputs in the circles will be read by the program as objects and these modules can be taken by other objects. For example, window material properties includes U-value, solar, absorptivity, emissivity, and the window module will be considered as an object that can be applied to the envelope setting. Then, the envelope setting containing the information including the properties of the windows, walls, roofs, and floors will be integrated into zones. Building level inputs include energy sources, renewable energy systems, building energy management (BEM) system, domestic hot water (DHW) system, pumps, and some basic information about the building.

The SimBldPy simulation results are easy to read and analyze. The hourly simulation results will be exported to a comma separated values (csv) file, including date and time, lighting energy use, equipment energy use, solar heat gain from window and opaque parts, cooling and

heating need and their energy use, DHW energy use, pump energy use, fan energy use, as well as energy production from the onsite renewable energy systems such as PV and solar water heater (SWH) system.

4.3 Built-in Parallelized Model Calibration

In building energy modeling, calibration has always been an important process in tuning the key model parameters that best fit the target energy use of a building. Previous research has validated certain methods such as Bayesian approach (Heo, Choudhary, and Augenbroe 2012), sensitivity analysis (Lomas and Eppel 1992, Tian 2013, Li et al. 2014, Enríquez, Jiménez, and Heras 2017), evolutionary algorithms (Ramos Ruiz et al. 2016), is qualified to be used in calibrating building energy model. In this tool, a differential evolution (DE) algorithm is used as a method of calibrating model parameters. This method is chosen here because that it is capable of handling non-differentiable and nonlinear cost function. In this research, the cost function of the model parameters is evaluated by running the SimBldPy simulation, which is non-differentiable and nonlinear in nature. Moreover, DE has the advantage of being parallelizable, easy to use and has good convergence properties (Storn and Price 1997). Unlike traditional genetic algorithm (GA), DE is able to deal with real number vectors as design space, which allows to handle both continuous and integer design variables (the model parameters to be calibrated here).

The mechanism of DE has been described in (Storn and Price 1997) and will not be rephrased here. The uniform crossover operator with a crossover rate of 0.9, and the Gaussian mutation operator with a mutation rate of 0.01 are used in the optimization. As a single objective optimization problem, a fitness-proportion selection method is used. The selection procedure

stochastically chooses individuals from the population with probability proportional to their fitness, which is often referred to as "roulette wheel" selection (Mitchell 1998).

The non-optional model parameters that must be calibrated for the simplified hourly model include C_m (thermal capacity of the building mass per building area), A_t (area of all surfaces facing the building zone per building area), A_m (area of internal structure per building area), and infiltration rate. Other optional calibration parameters include the U-value of exterior and interior wall, floor, and roof, solar absorptivity and emissivity of exterior wall and roof, window solar heat gain coefficient. These variables are important for the simplified hourly model and some of them such as C_m, A_m, A_t, and infiltration rate are not easy to be captured and are therefore always counted as calibration parameters. Users can choose other optional parameters to involve in the calibration process. Range of the parameters is determined by a linear space with certain assigned intervals. For example, the range of parameter C_m can be from 40kJ/K-m^2 to 100kJ/K-m^2 with an interval of 5kJ/K-m^2, and this range will be mapped into a normalized space between 0 and 1 as one input variable for the DE algorithm.

The objective function of the calibration is the sum of root mean square error (RMSE) of building heating and cooling energy use. The calculation of RMSE is defined in Equation 10.

DE will be used to minimize the objective function and choose the best combination of model parameters after iterations of 50 generations, and the total population size of each iteration is set as 20 times of the calibration parameters number.

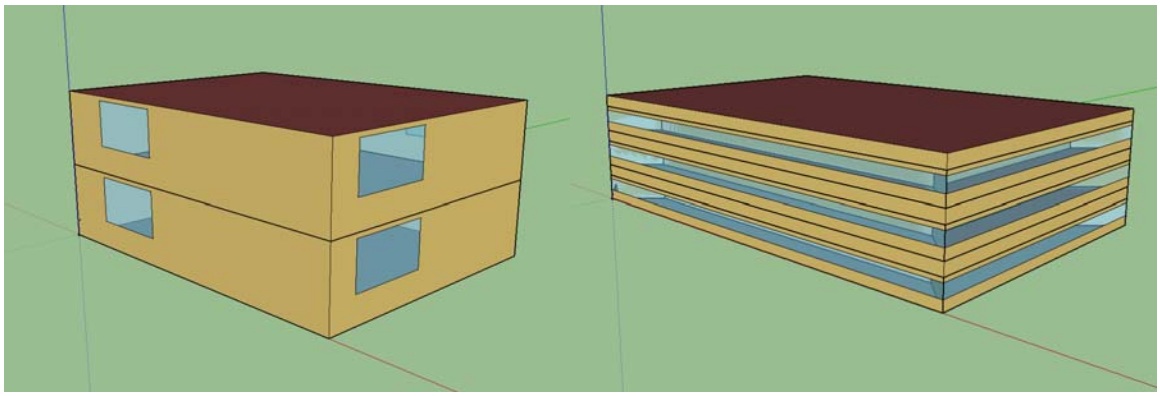
Another important feature of the calibration process is the parallelization of the objective function evaluation. The DE algorithm is adapted for parallel computing by simultaneously using all the threads of computer cores to evaluate the objective function of each individual within a generation, which means a great acceleration of the optimization. For example, in this research, a

16-core Intel Xeon v4 workstation with 32 available threads is used for parallel computing for the DE, which is 32 times faster than unparalleled algorithm. Using a simplified building model together with parallelized DE can greatly improve the computational performance of the calibration process.

5. Calibration and Validation of Simplified Modeling Tool

5.1 Reference Buildings

In this chapter, the SimBldPy tool is adopted to model two DOE reference buildings based on IECC 2006 (residential) and ASHRAE 90.1 2004 (commercial) standard. One is a residential building, and the other is a medium-sized office building. Both buildings are located in Philadelphia, PA. Typical Meteorological Year (TMY3) weather data will be used for the BES. EnergyPlus 8.5 is adopted as the reference modeling engine in this research to verify the performance of the SimBldPy model. EnergyPlus is a universally acknowledged building simulation engine that provides the most detailed modeling procedure for the transient thermal dynamics behavior of buildings (LBNL 2015), which has been validated to be accurate and reliable. The DOE (Deru 2011) and Pacific Northwest National Laboratory (Goel S 2014) carried out numerous studies in calibrating and verifying EnergyPlus models against field data and proved that it is a reliable building simulation tool. The two test buildings are illustrated in Fig. 17:



two-storey residential building three-story office building with three plenum zones

Fig. 17 Building model of the two reference buildings in EnergyPlus

Table 9 Thermal properties and system configuration of the two buildings

	Residential	Office
Building Area (m^2)	223	4982.2
Gross Wall Area (m^2)	221.2	1977.7
Window Area (m^2)	33.2	652.6
Window U-factor (W/m^2-K)	2.273	3.045
Window SHGC	0.394	0.428
Wall U-factor (W/m^2-K)	0.535	0.7
Roof U-factor (W/m^2-K)	3.0	0.358
Air Infiltration rate (h^{-1})	1	1
HVAC type	Packaged Terminal	Packaged Terminal
Nominal Cooling COP	4	4
Cooling Source Energy Type	Electricity	Electricity
Nominal Heating Efficiency	0.8	0.8
Heating Source Energy Type	Gas	Gas
Lighting (W/m^2)	2.5	10.76
Service Water Heating Efficiency	0.8	0.8

Table 10 Building occupancy schedule and use schedule

Residential Building						
time of day	wd_occ	we_occ	wd_app	we_app	wd_light	we_light
From 0 To 7	1.00	1.00	0.56	0.56	0.10	0.10
From 7 To 8	0.88	0.88	0.72	0.72	0.43	0.43
From 8 To 9	0.41	0.41	0.61	0.61	0.19	0.19
From 9 To 16	0.24	0.24	0.53	0.53	0.13	0.13
From 16 To 17	0.29	0.29	0.71	0.71	0.48	0.48
From 17 To 18	0.55	0.55	0.86	0.86	0.67	0.67
From 18 To 22	0.90	0.90	1.00	1.00	1.00	1.00
From 22 To 24	1.00	1.00	0.85	0.85	0.30	0.30
Office Building						
time of day	wd_occ	we_occ	wd_app	we_app	wd_light	we_light
From 0 To 5	0	0	0.4	0.3	0.05	0.05
From 5 To 6	0	0	0.4	0.3	0.1	0.05
From 6 To 7	0.1	0.1	0.4	0.4	0.1	0.1
From 7 To 8	0.2	0.1	0.4	0.4	0.3	0.1
From 8 To 12	0.95	0.3	0.9	0.5	0.9	0.3
From 12 To 13	0.5	0.1	0.8	0.35	0.9	0.15
From 13 To 17	0.95	0.1	0.9	0.35	0.9	0.15
From 17 To 18	0.3	0.05	0.5	0.3	0.5	0.05
From 18 To 19	0.1	0.05	0.4	0.3	0.3	0.05
From 19 To 20	0.1	0	0.4	0.3	0.3	0.05
From 20 To 22	0.1	0	0.4	0.3	0.2	0.05
From 22 To 24	0.05	0	0.4	0.3	0.05	0.05

Note: “wd”, weekday; “we”, weekend; “occ”, occupancy schedule; “app”, appliances use schedule; “light”, lighting schedule

Table 11 Building indoor air temperature set point schedule

Office Building				
time of day	wd_Tset_heat	we_Tset_heat	wd_Tset_cool	we_Tset_cool
From 0 To 6	13	16	32	28
From 6 To 7	18	16	32	28
From 7 To 21	23	16	24	28
From 21 To 24	16	16	32	28
Residential Building				
time of day	wd_Tset_heat	we_Tset_heat	wd_Tset_cool	we_Tset_cool
From 0 To 24	22.22	22.22	23.88	23.88

Note: “wd”, weekday; “we”, weekend; “Tset_heat”, heating set point; “Tset_cool”, cooling set point

Table 9 shows some of the most important modeling parameters for SimBldPy and the referenced EnergyPlus model (will be called “sim model” and “ep model” in the rest of this chapter). For the residential building, both ep and sim model the building’s two floors as two thermal zones. For the office building, for ep model, each floor is divided into five zones (four perimeter zones and one core zone) while for the sim model, each floor is divided into two zones (one perimeter zone and one core zone). This is in order to simplify the modeling process in SimBldPy, and it will be shown later when simulation results are compared, simplifying the perimeter zones in SimBldPy will not cause much difference in the results compared with EnergyPlus. In addition, for the office building in EnergyPlus model, there are three 1.22m high plenum zones between each floors and on the top of the third floor where the roof is attached to, and they are all unconditioned. These plenum spaces are modeled as unconditioned space in SimBldPy and further added uncertainties as to achieve good results from SimBldPy modeling tool since these unconditioned space can be more volatile regarding indoor air temperature because they are influenced by the synergy of heat flux from the outdoor environment and from the upper and lower zone.

Table 10 and Table 11 show the building occupancy schedule, operation schedules and indoor temperature set point schedules. The reason for showing these schedules in Table 10 and Table 11 is that these values are very important in both modeling methods and will be referenced later when different temperature set point schedules are applied to certain zones of the buildings.

5.2 Model Calibration Results

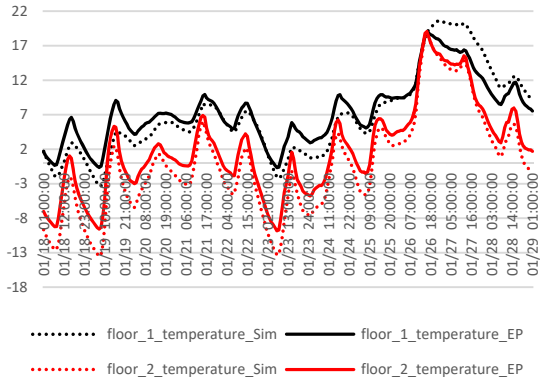
The sim models of the two buildings are calibrated using the parallelized DE algorithm described in section 4.3. The parameters that are tuned include C_m, A_m, A_t, U-values of the interior wall, interior floor, and ground floor. For this residential building, there is no internal wall, so it will not be considered in the calibration. The objective function of the DE algorithm is the sum of RMSE of the hourly heating gas and cooling electricity from the ep model simulation results. The calibration process takes about 1079 seconds and 2145 seconds for the residential building and the office buildings using parallel computing of 32 threads, respectively. The calibration results are shown in Table 12.

Table 12 Calibrated model parameters for the sim model

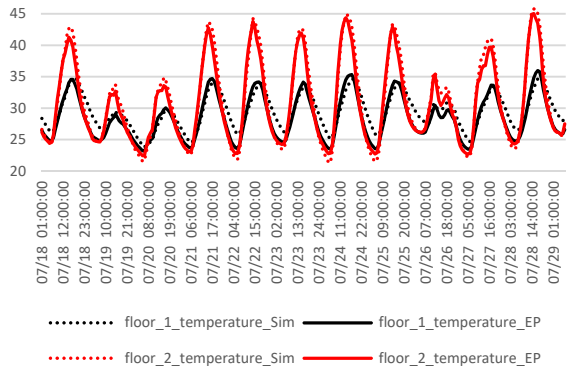
Building type	C_m (J/K-m^2)	A_m (m^2/m^2)	A_t (m^2/m^2)	Ground floor U-value (W/m^2-K)	Interior wall U-value (W/m^2-K)	Interior floor U-value (W/m^2-K)
Residential	60000	2.1	4.2	1.8	N/A	1.6
Office	155000	1.1	3.4	2.0	4.0	1.8

5.3 Indoor Air Temperature

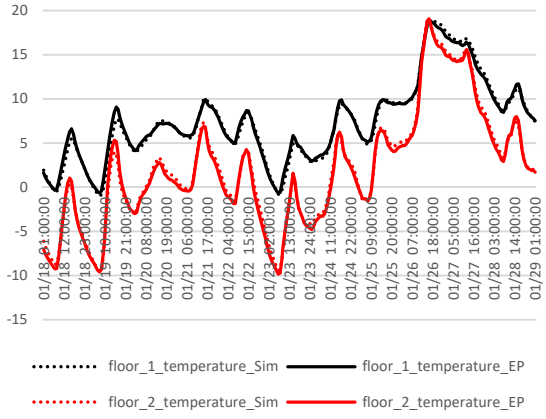
The calibrated SimBldPy residential building model using zone thermal coupling method is compared with the simplified hourly method without zone thermal coupling to show how different the internal temperature of sim model could be from ep model. The sim model and the ep model are both modified to disabled the HVAC system in this case. Thus, the indoor air temperature of the two zones in the residential building becomes free floating air temperature. The end of this is to see if the indoor air temperature of the sim models (with and without zone thermal coupling) is in agreement with the ep model when the room temperature is not controlled by the HVAC system.



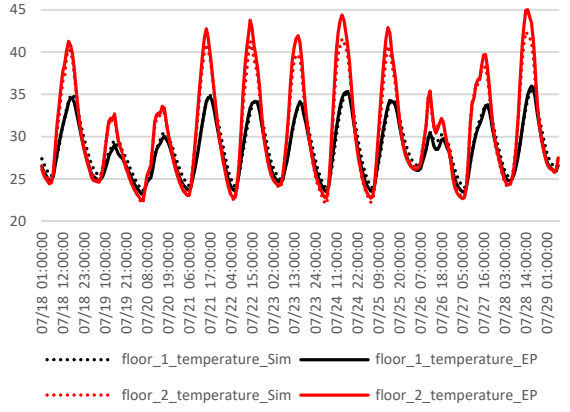
a winter week w/o thermal coupling



a summer week w/o thermal coupling



a winter week w/ thermal coupling



a summer week w/ thermal coupling

Fig. 18 Indoor air temperature of the residential building with and without zone thermal coupling

According to the indoor air temperature shown in Fig. 18, the use of zone thermal coupling will improve model accuracy because when HVAC system is not operating, the contact surface of two zones are usually assumed to have no heat flow in the model without thermal coupling, which makes the zones more susceptible to outdoor environment. This causes the model to predict the zone air temperature in the winter lower and higher in the summer. When the HVAC system in the zone is working, if the indoor air temperature set points of adjacent zones are similar, then the heat flow between them would be minimal, but a higher heat flow could be observed when the thermostat set points are distinct, causing a difference in heating and cooling loads prediction in these zones compared to the model without thermal coupling.

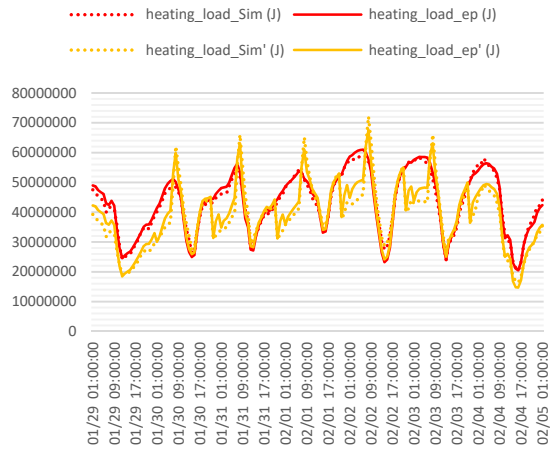
5.4 Heating and Cooling Load

In this research, it is important that if the sim model's heating and cooling load predictions are consistent with the reference ep model. The calibrated energy use in sim model is found to be a good predictor for the heating and cooling loads, which corresponds well with ep model's load predictions. The validation results of sim model have been shown in Fig. 19 and Fig. 20.

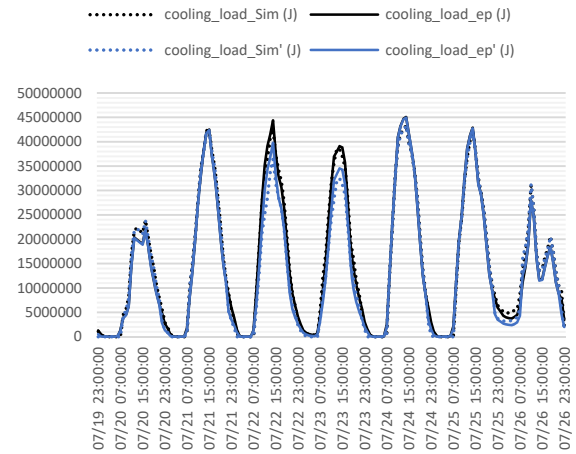
The simulation results shown in Fig. 19 and Fig. 20 indicate that the heating and cooling load prediction of the sim model agrees well with the trend and pattern of ep model results. The calculated R2 values for heating and cooling load predictions are 0.9937, 0.9914 for the residential building, and 0.9873, 0.9893 for the office building, respectively.

In order to showcase the validation of the sim model when different temperature set points are applied to certain zones in the two buildings, we first compare the heating and cooling load of the ep model and sim model by applying the same temperature set points schedules shown

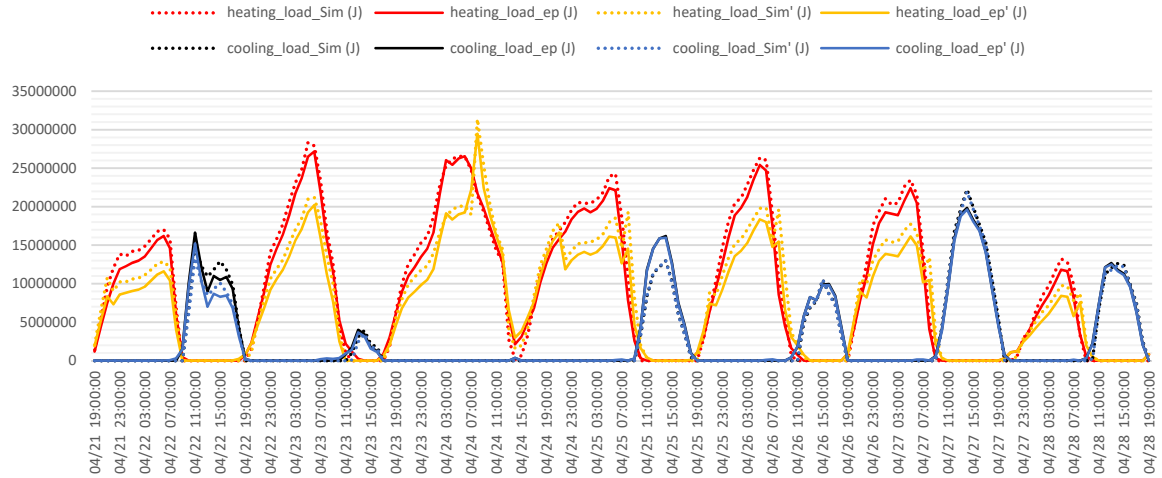
in Table 11 to all the zones in the two building. In order to validate the performance of the sim model when temperature set points are switched for certain zones as discussed in section 5.3, another test of the ep and sim model is performed by applying the set point schedule of the office building to the first floor in the residential building and the constant set point schedule of the residential building to all the core zones in the office building to see if the thermal coupled is going to give a consistent simulation results referenced to ep model.



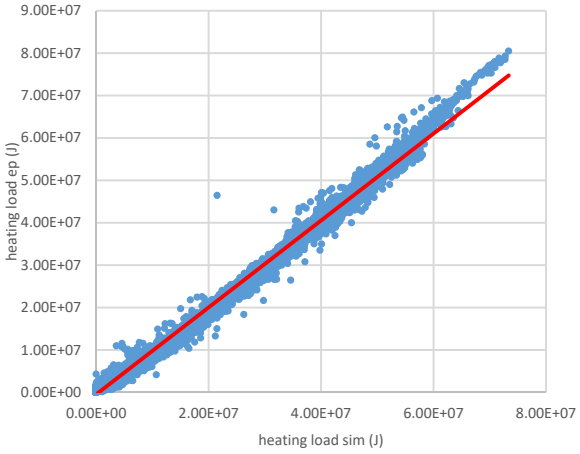
hourly heating load of a winter week



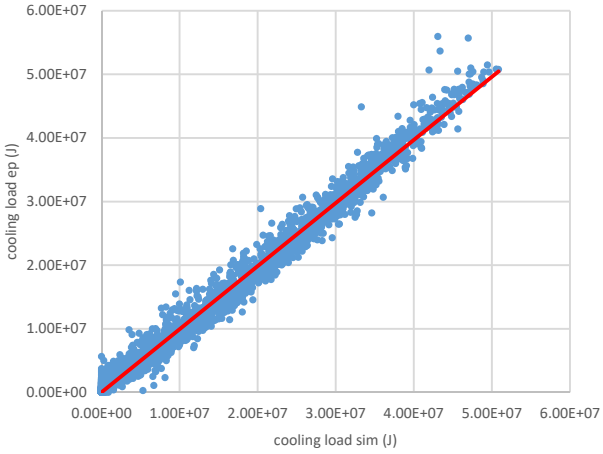
hourly cooling load of the summer week



hourly heating load of a week in swing season

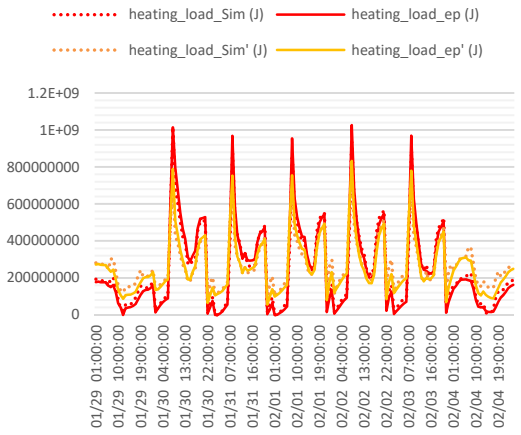


heating load prediction accuracy w/o set point switching

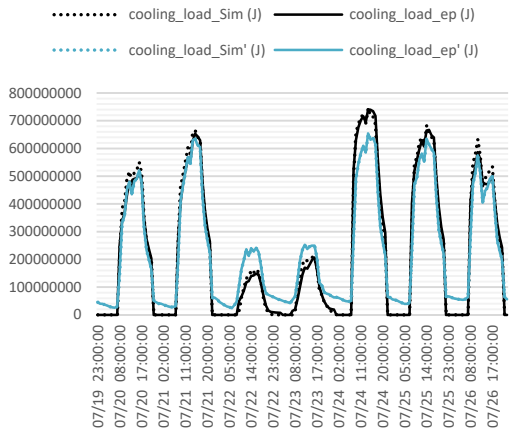


cooling load prediction accuracy w/o set point switching

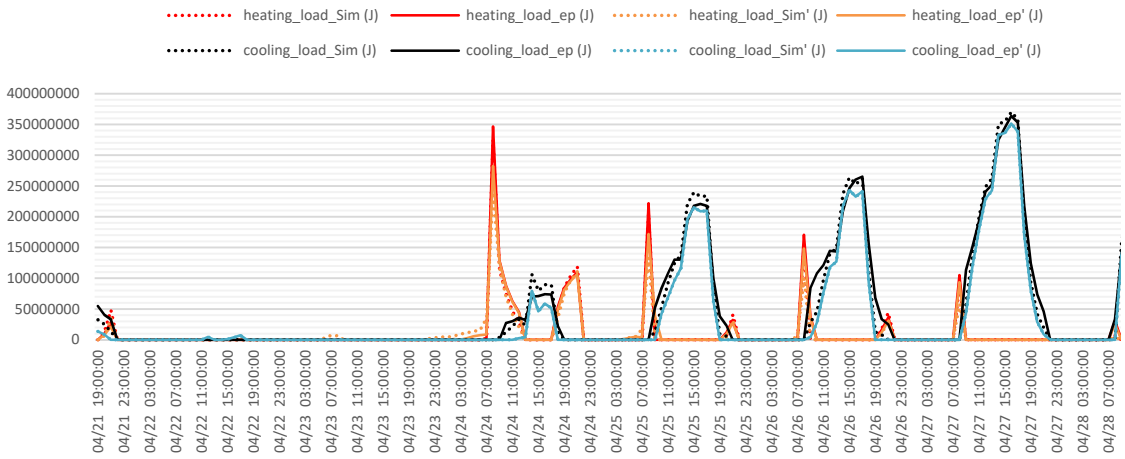
Fig. 19 Heating and cooling load comparison between ep and sim model for the residential building (line with apostrophe is the case with different indoor air temperature set points)



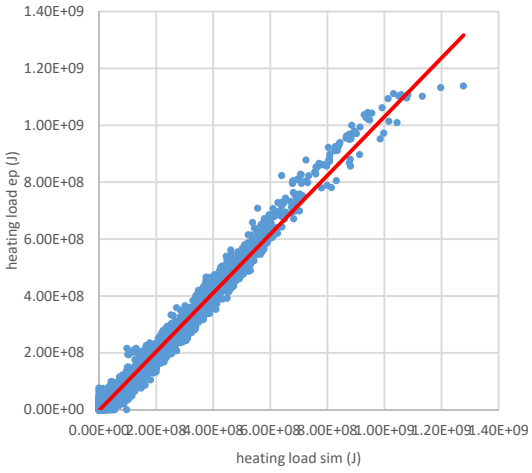
hourly heating load of a winter week



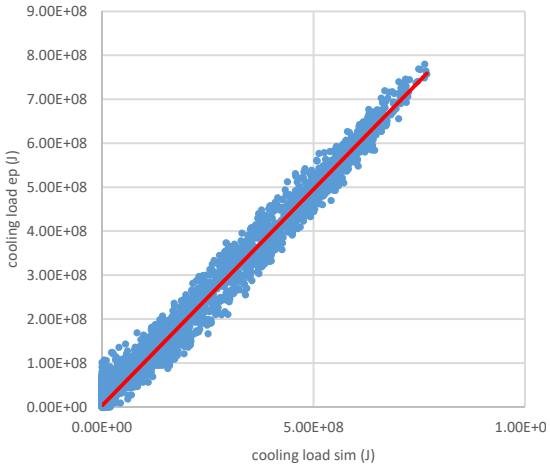
hourly cooling load of the summer week



hourly heating load of a week in swing season



heating load prediction accuracy w/o set point switching



cooling load prediction accuracy w/o set point switching

Fig. 20 Heating and cooling load comparison between ep and sim model for the office building

(line with apostrophe is the case with different indoor air temperature set points)

For the residential building, after switching the temperature set point schedule in the first floor to that of the office building, the heating load during night time falls below the baseline scenario. The heating load of the set point switching scenario also has a higher peak because the heating load will be higher when the temperature is raised from 18℃ to 23℃ as the office set point schedule states. The residential building does not need much cooling during the night, so switching the set point does not affect the cooling load pattern too much.

For the office building, on weekdays, the heating load is lower after applying the constant set point to the core zones of the office building in winter, while on weekends, due to that the office building heating set point is only 16℃, the higher residential set point in core zones raises the heating load. The same reason also applies to cooling load in summer. A lower heating load peak of the set point switching scenario makes sense because a constant set point in the core zones makes the office building more thermally "stable". The results show that the sim model has good performance in simulating heating and cooling load when different temperature set point schedules are used in different zones.

5.5 Performance of Model under Various Climates and Various Retrofit Combinations

One of the main objectives of developing the SimBldPy tool is to facilitate the parametric study of building retrofit ECM options. Since dynamic and transient BES engines are usually more computationally intensive, the development of this lightweight modeling tool provides alternatives for assessing relative impacts of ECM combinations on building load and energy use. The simulation time of sim model is about 1/30 times of the ep model, meaning 30 times faster due to the simplification. Therefore, the validity of the sim model under various weather conditions and ECM options should be guaranteed. In this chapter, the results of the sim model

will be used to compare with those of the reference model (ep model) under different climates and retrofit options to verify its performance in handling the various needs in building performance study.

5.5.1 Validation under different climate conditions

One of the concerns for the sim model is that if it is able to function well in different climate conditions for the same buildings because the sim model is calibrated with ep model results under current weather condition and the performance of the same model in various climate conditions should also be validated. We run the sim model and the ep model in five different climate zones in the United States: San Francisco in California (CA), Phoenix in Arizona (AZ), Houston in Texas (TX), Memphis in Tennessee (TN), and Burlington in Vermont (VT), located in 2A, 2B, 3A, 3C, and 6A climate zones, respectively. The annual summary of the heating and cooling energy use is presented in Table 13, where normalized root mean squared error (NRMSE) and R-square (R^2) are used as an indicator of the accuracy of the sim model's hourly prediction compared with the referenced ep model. NRMSE and R^2 are calculated in this way:

Equation 21: $$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \Big/ (x_{max} - x_{min})$$

Equation 22: $$R^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(\hat{x}_i - \bar{\hat{x}})}$$

where x_i is the true value and $\hat{x}_i$ is the model predicted value; $\bar{x}$, $\bar{\hat{x}}$, x_{max} and x_{min} are the average of true and predicted value, and the maximum and minimum of true values, respectively.

Table 13 Sim model performance in energy use prediction in different climate conditions (unit: GJ)

Residential Building								
City	heating energy ep	heating energy sim	cooling energy ep	cooling energy sim	heating energy NRMSE	cooling energy NRMSE	heating energy R^2	cooling energy R^2
Memphis	117.029	123.366	21.942	19.862	2.7%	3.8%	0.992	0.990
Phoenix	44.550	46.705	42.974	43.606	5.1%	3.3%	0.985	0.994
Houston	64.537	68.853	27.710	25.248	2.5%	3.5%	0.992	0.988
SF	114.826	121.231	4.691	4.780	3.2%	2.1%	0.990	0.987
Burlington	278.832	277.279	6.175	6.462	2.6%	1.9%	0.996	0.988
Office Building								
City	heating energy ep	heating energy sim	cooling energy ep	cooling energy sim	heating energy NRMSE	cooling energy NRMSE	heating energy R^2	cooling energy R^2
Memphis	470.208	460.225	319.897	307.896	1.6%	3.5%	0.989	0.988
Phoenix	81.202	87.467	556.582	568.304	1.5%	4.2%	0.976	0.987
Houston	167.295	161.408	408.805	398.682	1.2%	4.1%	0.984	0.988
SF	214.869	211.745	108.777	112.861	1.5%	3.9%	0.984	0.986
Burlington	1625.397	1582.362	119.930	125.549	2.1%	2.2%	0.991	0.988

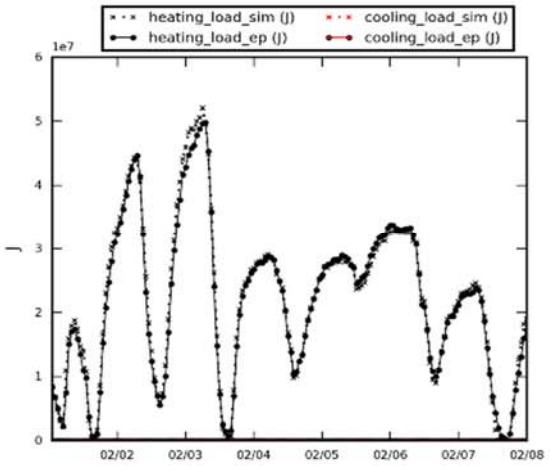
Table 13 shows that the sim model is able to accurately predict the cooling and heating energy for both test buildings. For the residential building, the highest NRMSE is 5.1% for hourly heating energy in Phoenix and 3.8% for hourly cooling energy in Memphis. The highest NRMSE for the office building is 2.1% for heating and 4.2% for cooling. The R^2 value, which indicates the correlation between the results of the sim model and the ep model, clearly shows that the performance of the sim model is good compared to the reference model for both buildings in energy use simulation.

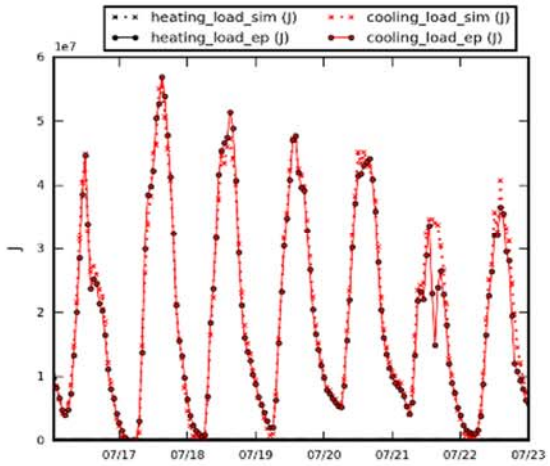
In addition, one week of results in winter, summer, and swing season, is selected respectively to demonstrate the validity of heating and cooling load prediction for both buildings and is shown in Fig. 21. Three cities: Houston, San Francisco, and Burlington, are chosen in Fig. 21 to represent distinct weather conditions where the building will be located. The predictive

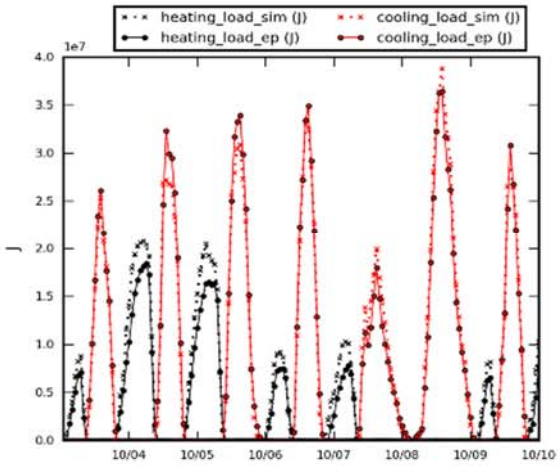
power of the sim model is comparable with that of the ep model in these cases, which shows that the sim model is able to simulate the heating and cooling load and energy use in various weather conditions.

Residential Building

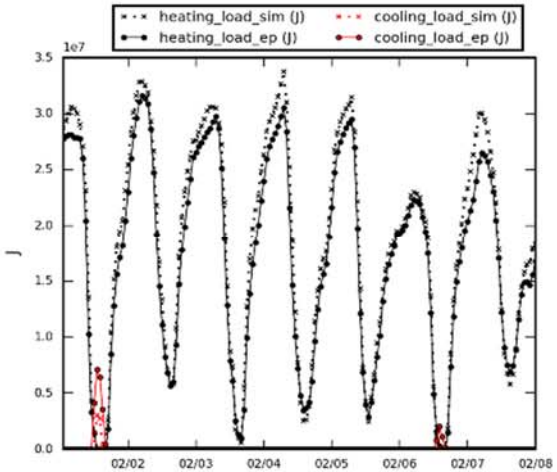
Houston

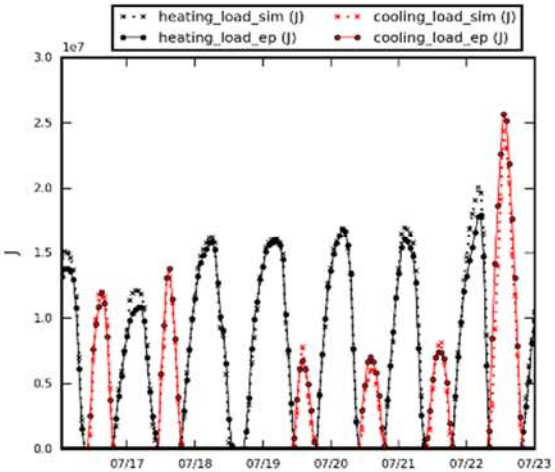


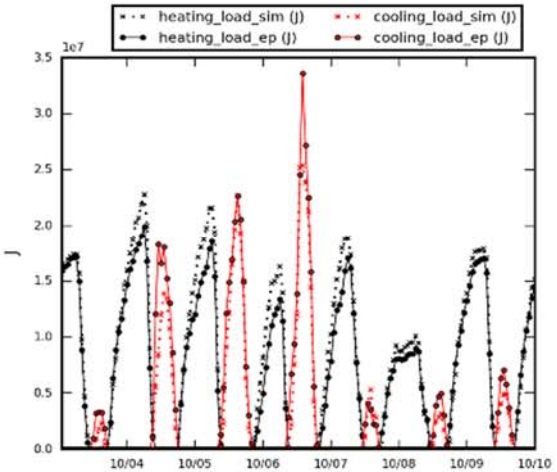




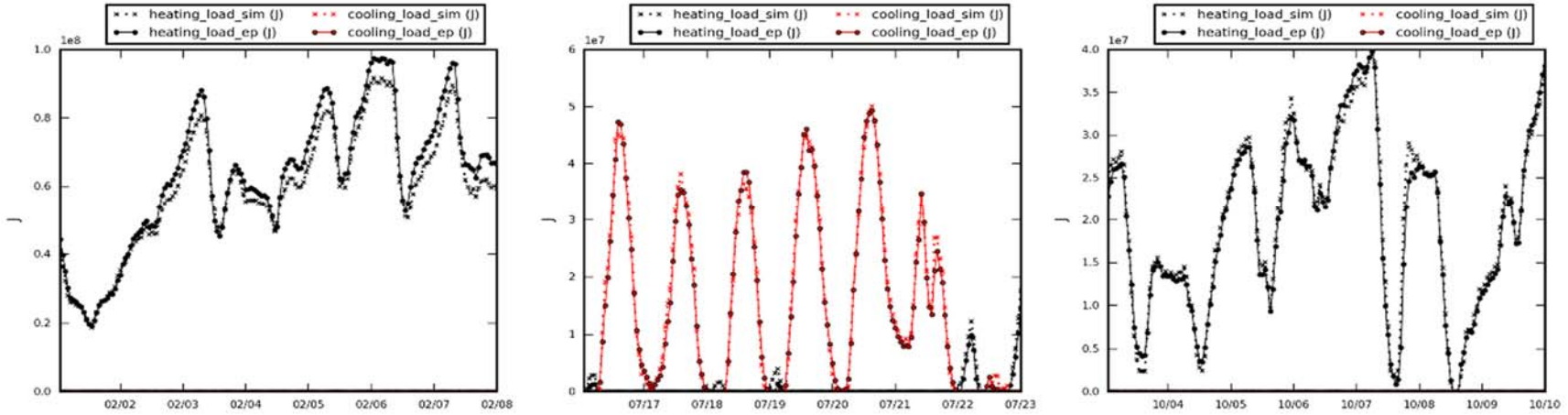
San Francisco



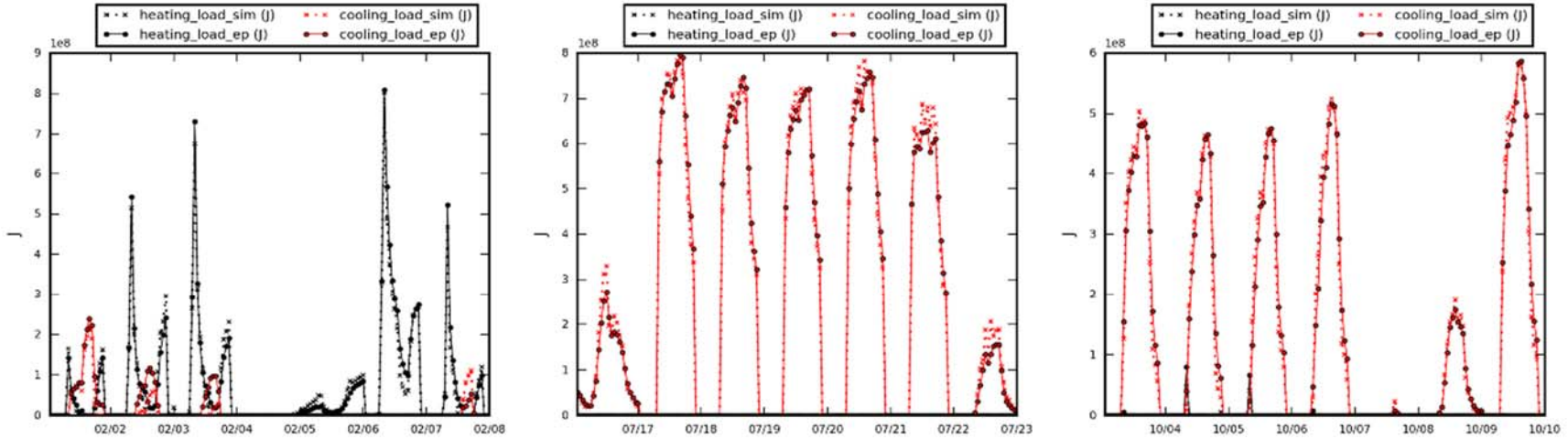




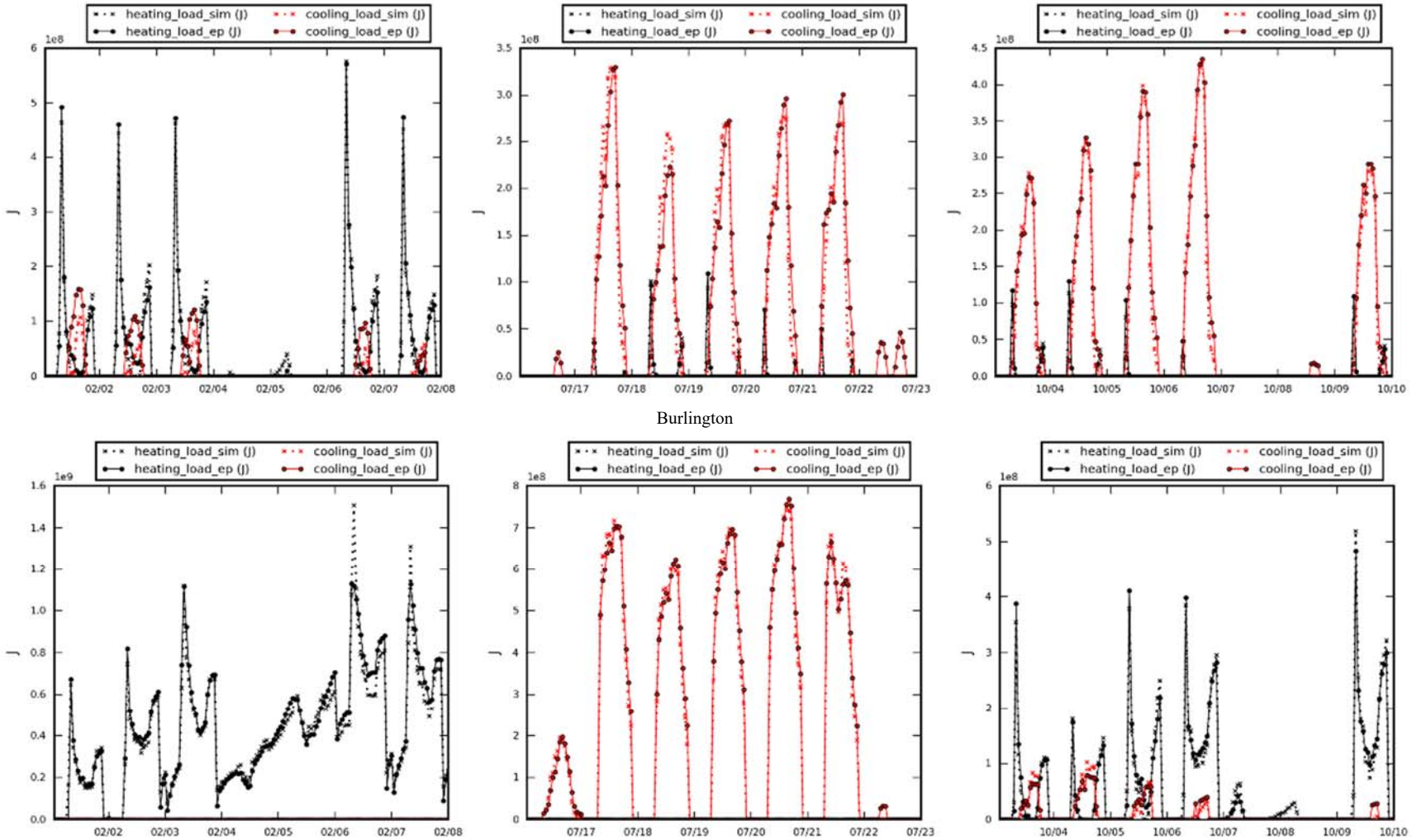
Burlington



Office Building
Houston



San Francisco



Burlington

Fig. 21 Hourly heating and cooling load validation of the buildings in other climate zones

5.5.2 Validation with ECMs

To better understand the performance of the sim models with different ECMs, an ECM parametric study function has been added to the SimBldPy tool. The ECMs currently involved in the tool are change to window U-value and SHGC (window change), wall U-value improvement (adding insulation), roof U-value improvement (adding insulation), blinds shading, air infiltration improvement, heating efficiency improvement, cooling efficiency improvement, and lighting load reduction(lighting system upgrade). The ECM parameters are listed in Table 14. It should be noted that the value 0 ((0, 0) for window retrofit) for each ECM's parameter means the retrofit option is not applied to the building in an ECM combination.

Table 14 ECMs and parameters

window (SHGC, m^2-K/W)	wall_insulation (R-value) (m^2-K/W)	roof_insulation (R-value) (m^2-K/W)	window shading	air_infl (h^{-1})	heating efficiency	cooling efficiency	lighting s
(0.0, 0.0)	0	0	0	0	0	0	0
				0.4			
(0.80, 3.6)	1.25	1.519757	internal	(0.3)	0.95	4.2	0.3
(0.75, 2.8)	1.610306	1.968504	external	0.6		4.5	0.4
(0.62, 1.6)	1.968504	2.421308		0.8			
(0.44, 1.6)	2.331002	2.873563					
(0.288, 1.05)	2.688172	3.322259					
(0.585, 0.52)	3.04878	3.773585					
(0.28, 0.33)	3.412969						
(0.63, 0.48)	3.773585						
(0.25, 0.26)							

Note: the number in () for air infiltration is the value used for office building; the lighting system upgrade coefficient in the table means how much lighting load will remain after retrofit compared with the current lighting system.

The modeling of wall and roof insulation in SimBldPy is achieved by adding an additional thermal resistance to the opaque material resistance of the building. The window shading is modeled by assigning a solar reduction factor (SRF) to the solar heat gain through the windows in each time step. In SimBldPy, the SRF for the exterior shading is assigned to 0.3, while that of interior shading is 0.7. Shading will be enabled once high zone cooling energy is observed and the same control strategy is adopted in the window shading in ep model. In SimBldPy, the activation and deactivation of window shading is determined by the cooling load of the previous time step, which is the same as the ep model.

A function programmed in the SimBldPy tool called "SimParaValidate" is designed to take any combination of ECMs from the above list as an input argument and run the simulation for both sim model and ep model with the designated ECMs. For the parametric analysis of EnergyPlus, a Python script is programmed to modify the text based EnergyPlus input file and perform parametric simulations of different ECMs. The "SimParaValidate" function is then executed in parallel to test different ECM combinations for sim and ep models. For the ECMs considered in Table 14, there could be 136,080 cases of different ECM combinations. Testing all the cases would be impossible, so we randomly pick 100 samples from all cases using latin-hypercube sampling to ensure reasonably distributed samples in the huge combinatorial design space. The statistics of model accuracy for the 100 randomly chosen samples with various ECM combinations for the residential building and office building are shown in Appendix I and Appendix II.

The average NRMSEs for office building heating load, cooling load, heating energy, and cooling energy, are 1.95%, 3.26%, 2.29%, and 3.98%, respectively. The R^2 values for them are 0.988, 0.99, 0.983, 0.983, respectively. For the residential building, the average NRMSEs for

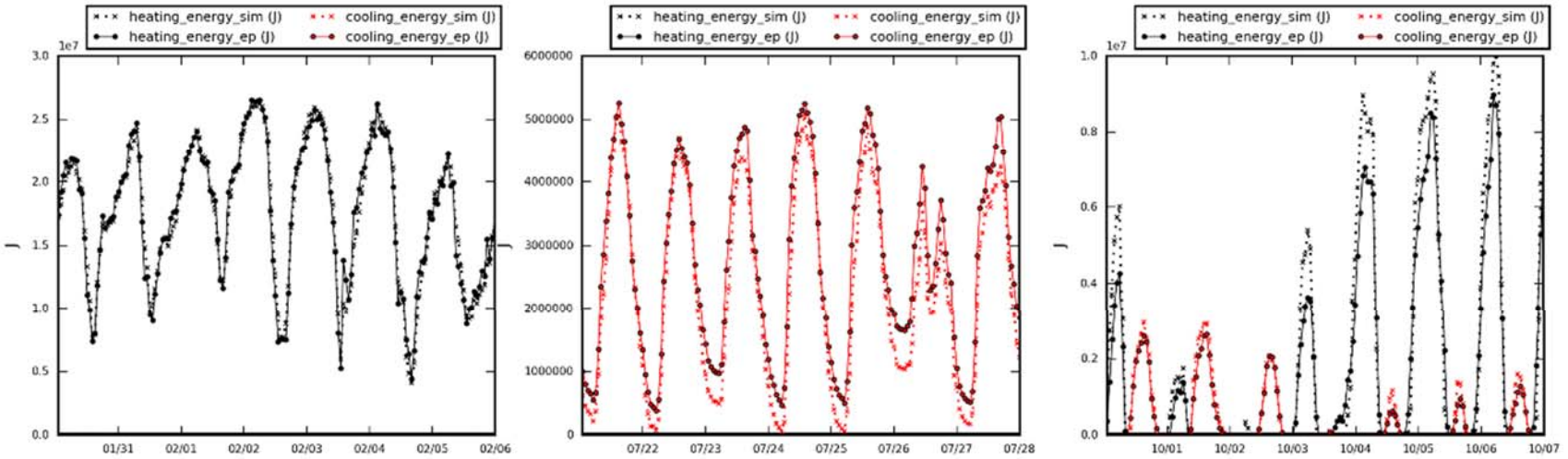
heating load, cooling load, heating energy, and cooling energy, are 2.7%, 3.61%, 2.76%, and 4.9%, respectively, while R^2 are 0.994, 0.989, 0.993, 0.979, respectively. It is proven that the sim model is efficient and accurate in simulating the hourly heating and cooling load and energy use.

For each building, the heating and cooling energy use of four random ECM combinations is plotted in Fig. 22. In the figure, the order of the ECMs shown at the top of each plot is defined as follows: window retrofit, shading position, wall insulation R-value, air infiltration level, roof insulation R-value, heating system efficiency, cooling system efficiency, and lighting load improvement.

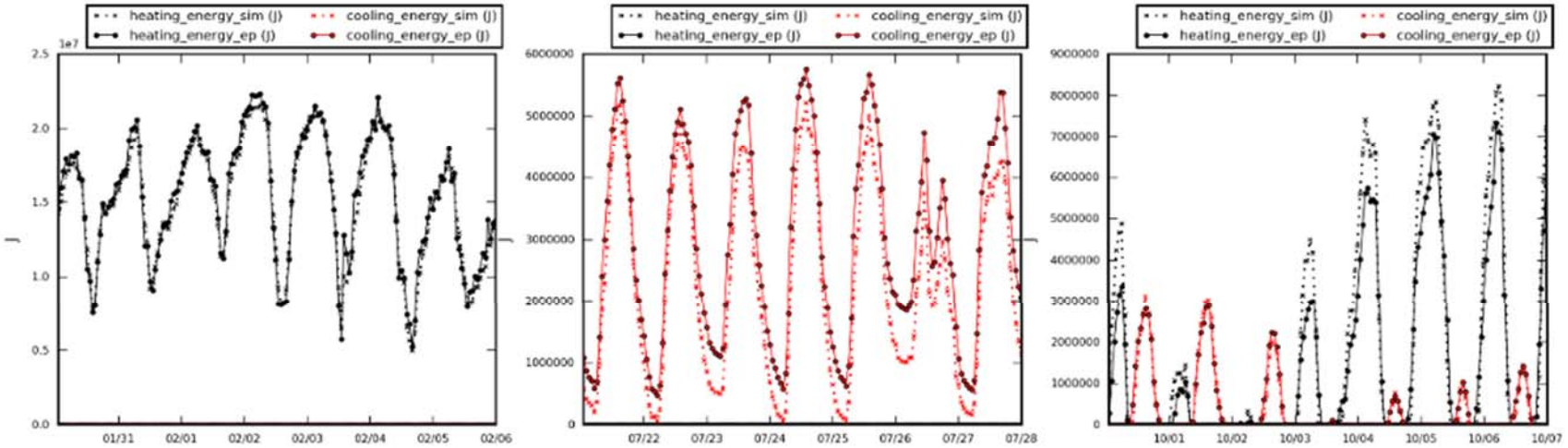
Fig. 22 shows that the sim model is able to predict hourly heating and cooling energy use with different ECM combinations. Although in some cases, the hourly results may differ from ep model results, particularly for the peak load prediction, the overall performance of the sim model is reliable when SimBldPy is used as a comparative parametric study tool to learn the impact of different ECM combinations on the heating and cooling performance of the building. Sensitivity analysis and regression analysis could be used to study the optimization and evaluation of building retrofit policy by modeling buildings in SimBldPy.

Residential Building

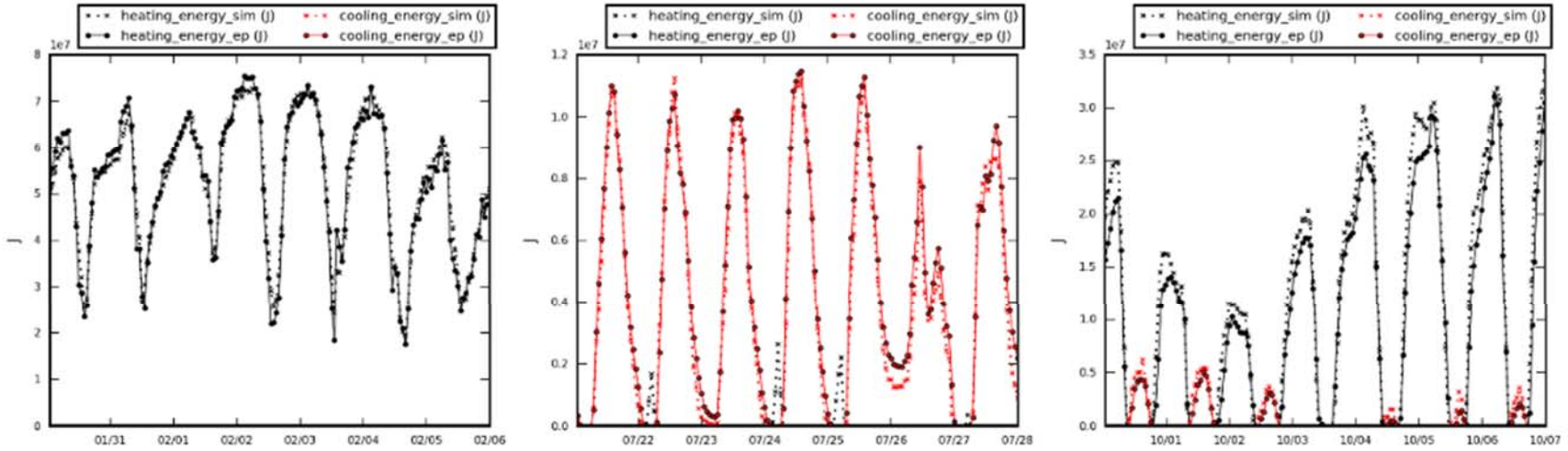
(0.0, 0.0), 1.0, 3.78, 0.4, 1.97, 0.0, 4.5, 0.0



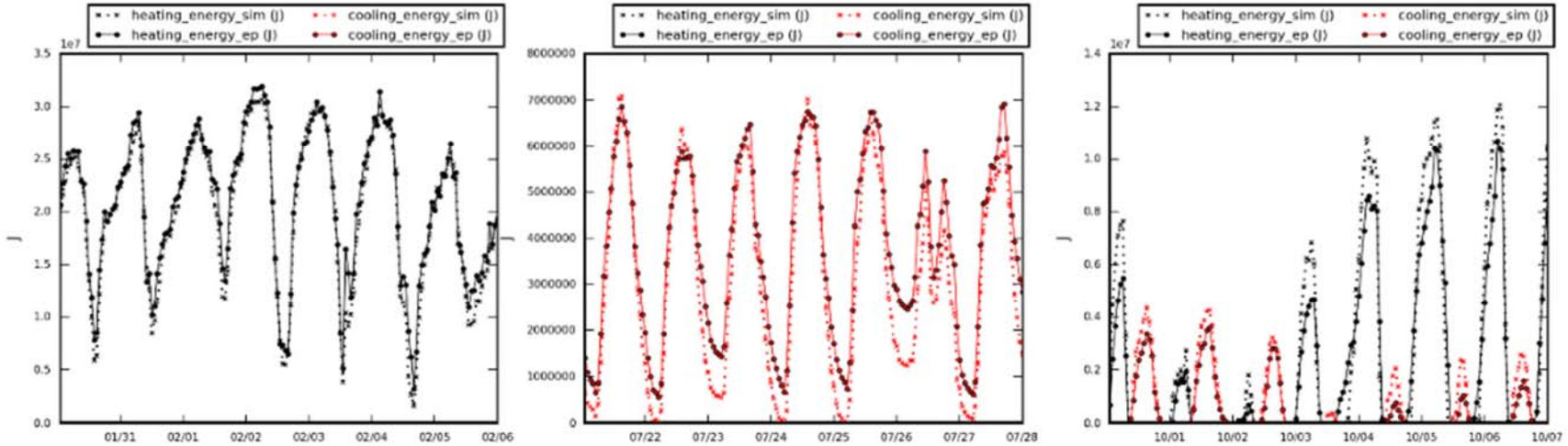
(0.28, 0.33), 0.0, 2.33, 0.6, 1.97, 0.95, 4.5, 0.4



(0.62, 1.6), 2.0, 0.0, 0.0, 0.0, 0.0, 4.5, 0.3

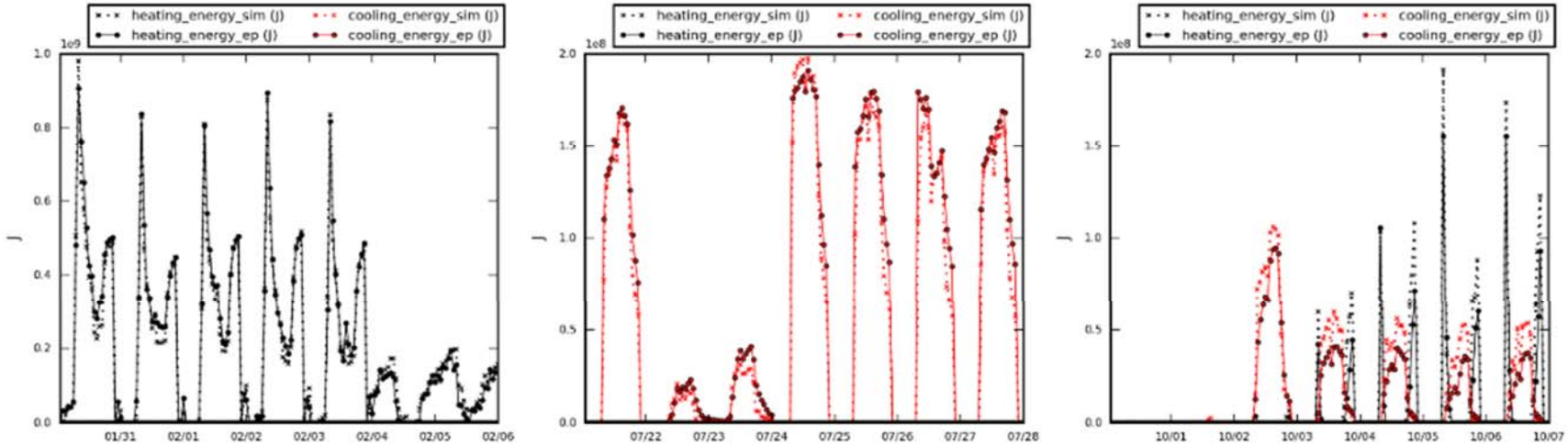


(0.585, 0.52), 1.0, 1.25, 0.8, 3.32, 0.0, 0.0, 0.0

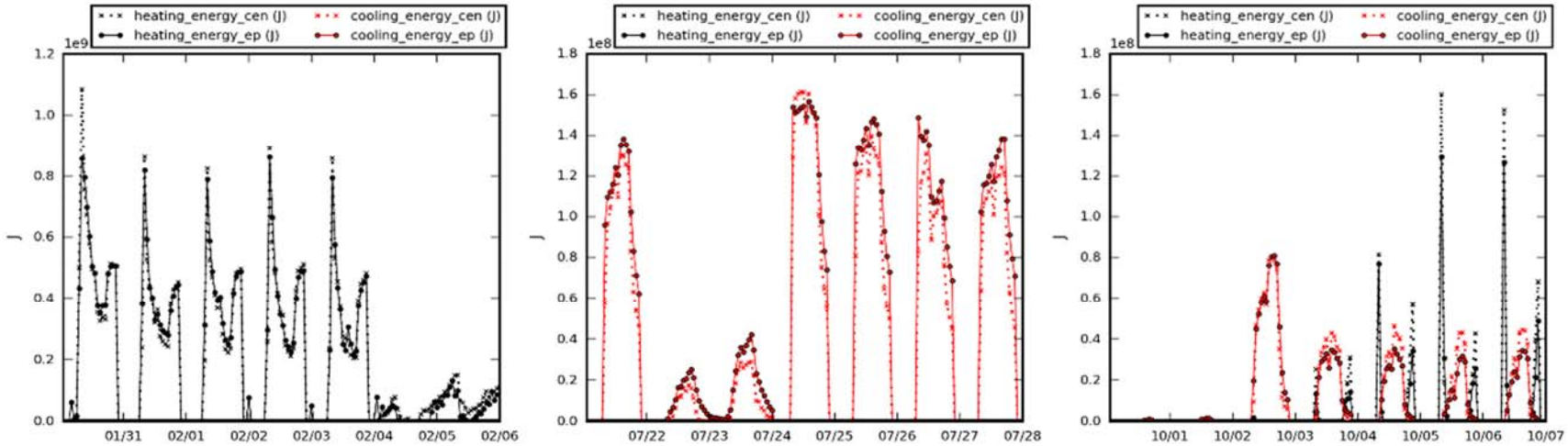


Office Building

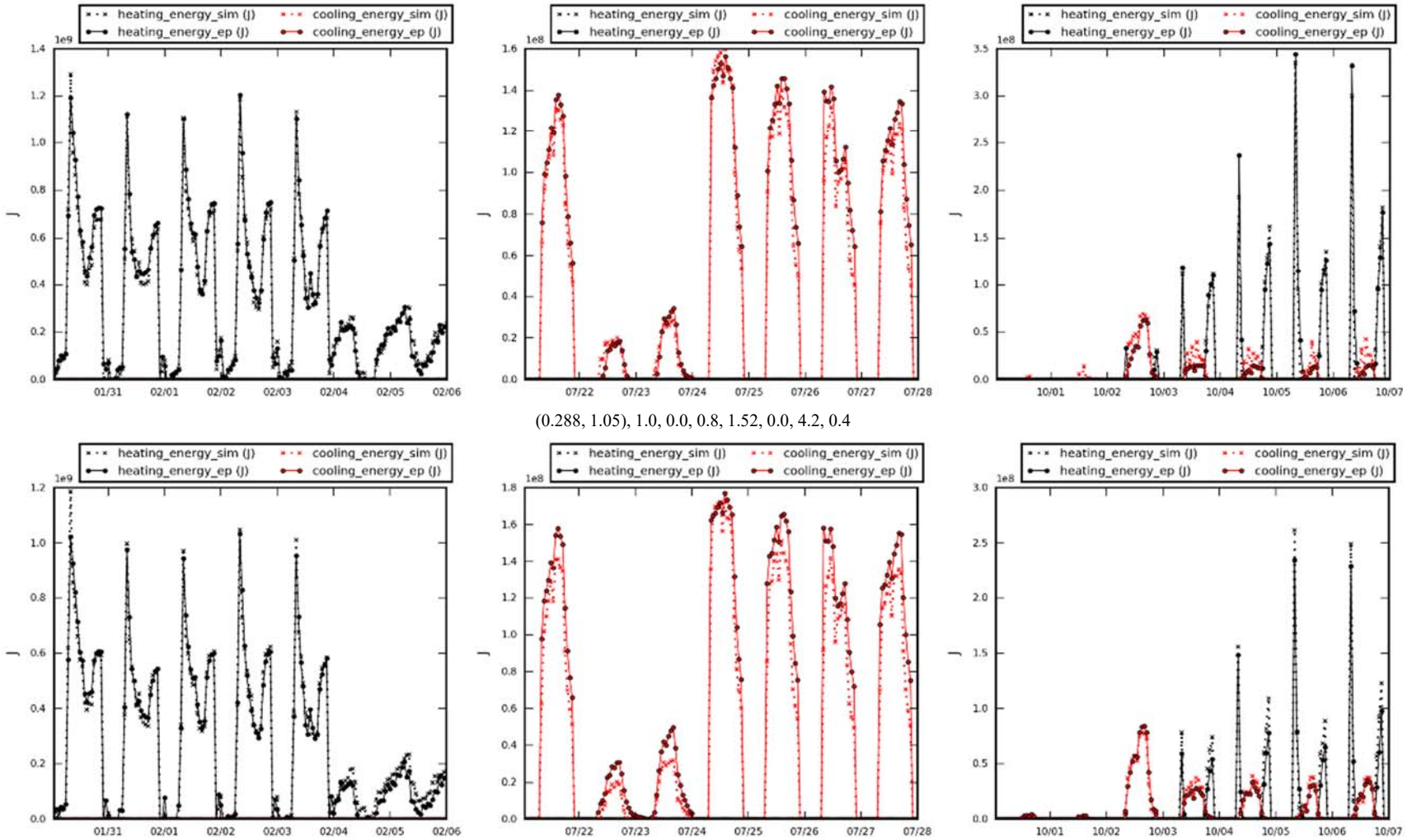
(0.0, 0.0), 2.0, 3.049, 0.8, 3.78, 0.95, 0.0. 0.0



(0.8, 3.6), 1.0, 1.25, 0.0, 3.78, 0.95, 4.2, 0.3



(0.75, 2.8), 2.0, 2.69, 0.0, 0.0, 0.0, 4.5, 0.3

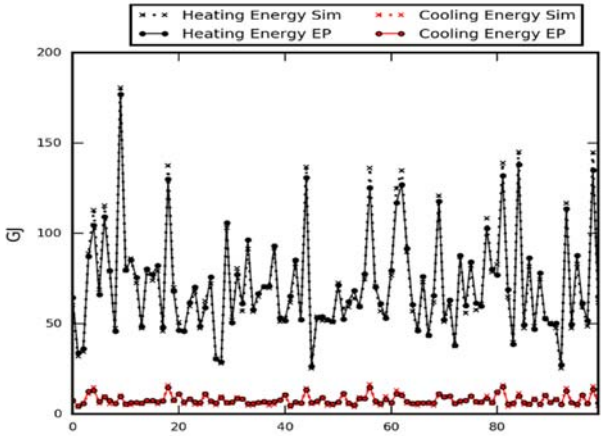


(0.288, 1.05), 1.0, 0.0, 0.8, 1.52, 0.0, 4.2, 0.4

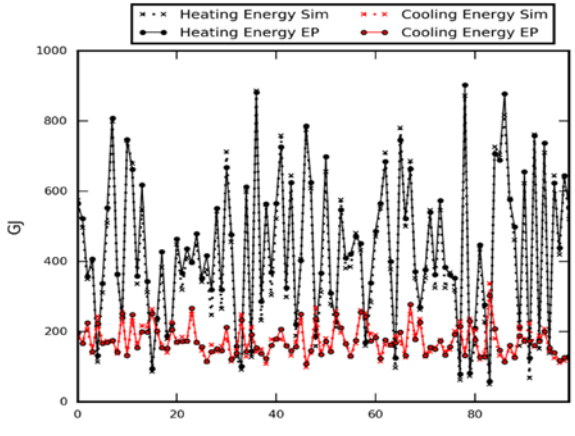
Fig. 22 Heating and cooling energy use of the two buildings when different ECM combinations are applied

The annual sums of heating and cooling load of the 100 randomly chosen ECM combination samples are plotted in Fig. 23 to visualize the comparison between the sim model and the ep model. To further study the model performance in simulating whole building energy use including pumps, fans, equipment and lighting energy use, the annual energy use of these systems are also shown in Fig. 23. A close correlation with the sim model and ep model in terms of annual heating and cooling calculation, which has an important application in ECM selection and screening at the preliminary stage of building retrofit. As shown in Fig. 23, the sim model is also able to simulate the energy use for different end use systems as well as the heating and cooling energy use with high reliability on an annual base with various ECMs being installed.

Validation of heating and cooling energy use of the buildings

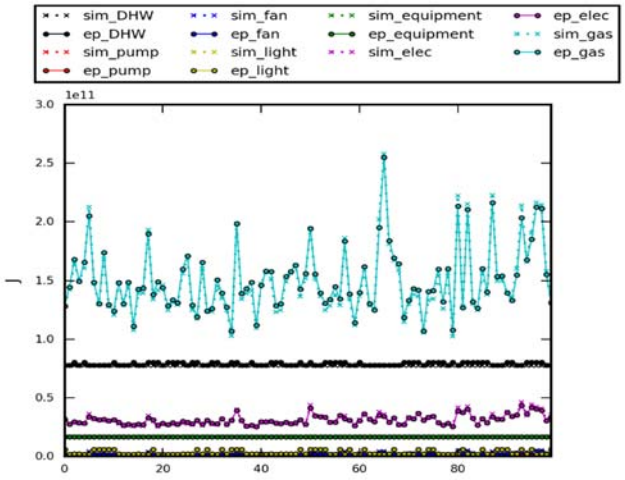


Residential Building

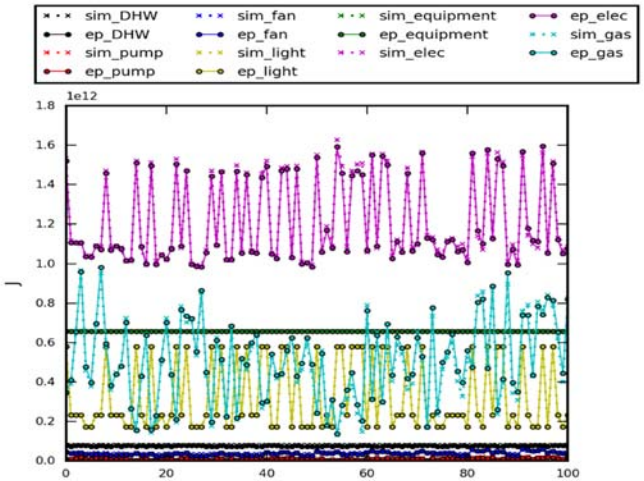


Office Building

End use validation



Residential Building



Office Building

Fig. 23 Annual heating, cooling energy use, and end use of the two buildings under different ECM combinations

5.6 Summary of SimBldPy and EnergyPlus Performance

Table 15 Comparison of computational cost and bias between EnergyPlus and SimBldPy

	E+ (residential)	SimBldPy (residential)	E+ (office)	SimBldPy (office)
Computational time (16 cores CPU)	39.6 hr	1.05 hr	60.1 hr	1.97 hr
Averaged accuracy of BEU (NRMSE)	/	4.2%	/	6.3%
PMV calculation	√	√	√	√

Table 15 shows the comparison of the EnergyPlus and SimBldPy regarding their computational performance and modeling bias. The average computational time used for EnergyPlus to simulate the residential buildings and office building out of approximately 13000 retrofit options (sampled from the entire combinatorial space, about 10% of the total design space) are 39.6 hours and 60.1 hours, respectively, while SimBldPy only uses about 1/30 to 1/40 computation time compared with EnergyPlus. The tradeoff here is the bias in simulation results, which are about 4.2% and 6.3% for residential and office building, respectively.

The results show that with affordable loss of model confidence, the developed lightweight building energy simulation tool which is dedicated for comparative parametric analysis, or for fast modeling and building energy performance evaluation, is able to greatly reduce computational complexity by about thirty to forty times. Moreover, the PMV estimation is also enabled while using SimBldPy for indoor thermal comfort analysis. The PMV calculation code is adapted from Chris Mackey's "comfort_model" Python script[1], and is optimized for

[1] https://github.com/CenterForTheBuiltEnvironment/comfort_tool/blob/master/contrib/comfort_models.py

computational efficiency by vectorizing all the variables used in PMV calculation and has been integrated with SimBldPy code.

In this chapter, it is shown that the SimBldPy can potentially be used for the calculation of BEU during the optimization procedure where BEU is an important part of the multi-objectives. In the forthcoming chapter where the optimization method is developed and where a case study will be tested with, SimBldPy will replace EnergyPlus in hourly BEU simulation based on weather data of future extreme year, which is constructed in the way described in section 3.3. The simulation results will then be used to train the RF model for future hourly BEU predictions during the post-retrofit phase.

6. Optimization Approach

It was introduced in the last chapter that a Python building simulation tool is developed based on ISO 13790 standard's simplified hourly method. The simulation tool is able to read the text based input file describing the physics of a building, including building shape and zoning, construction materials and physical materials, occupancy schedules, HVAC system type and efficiency, and etc. It will be used as the most important tool to evaluate the objective function in the optimization. The optimization work flow is shown in Fig. 24. The SimBldPy model will be calibrated based on the metered energy use. Then, a building retrofitting module will be in charge of reading ECM options entry and their respective parameters, the utility cost information in each year during the lifecycle period, as well as the information of the calibrated building model. After optimization information collection, the evolutionary optimizer will generate simulation tasks to the SimBldPy engine. This time, in order to assess the impact of climate change on building energy use (BEU) in the future years, a predictive model based RF is developed in this research

and it will make projections on the future hourly energy use. The optimizer will iterate to improve the objective function and at the mean time turn to SimBldPy to conduct energy simulation and evaluate the objective function. When the solution converges, the optimization results are exported and forwarded to a post-processor for decision-making. It is worth noting that throughout the process, parallel computation is used in the model calibration and in the evolutionary optimizer, which speeds up the entire process.

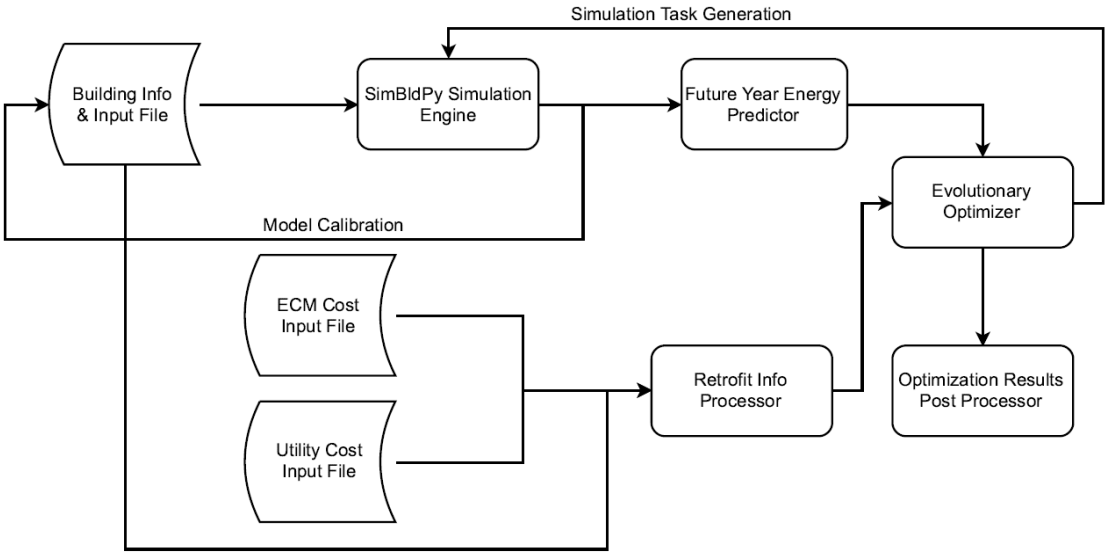


Fig. 24 Work flow of optimization approach

6.1 Decision Variables

The decision variables in this research involves various ECM options, and all the variable parameters will be normalized and scaled from 0 to 1. The ECMs will include building wall insulation, window U value and SHGC (solar heat gain coefficient), roof insulation, natural ventilation, air infiltration level, heating and cooling system efficiency, renewable energy

systems, and etc. If an ECM is not considered in the retrofit, its value will be 0. The normalization rules for different types of decision variables are as follows:

6.1.1 Discrete variable scaling

A discrete variable can be assigned to an ECM. In a non-idealized application of optimization problem in real practice, values such as the U-factor, solar transmittance, should be a specific number depending on the properties of the material for walls or windows, or, whether to adopt solar shading device for a building can be a categorical type of values including only 0 and 1. The solution of scaling such discrete variable is to map them into a continuous space between 0 and 1. The implementation will project the discrete variables into a uniformly distributed space between a specified range. Using uniform distribution here rather than other distribution is to ensure the equal chance for the selection of each parameter. For example, in a range of 0 to 1, the variable extracted between [0, 0,5) means adopting solar shading system while variable extracted between [0.5, 1) means not adopting the shading system.

6.1.2 Continuous variable scaling

For continuous variables, the scaling method works in a similar way with discrete variables. Instead of assigning a discrete value to a random number generated by uniform distribution, the continuous space is mapped into a continuous space between 0 and 1 corresponding to the designed parameter range. To simply illustrate the scaling of continuous variable, the example of air infiltration rate is used. If the design space of air infiltration is set between 0.5 h^{-1} to 3 h^{-1}, then a uniform distribution is assigned to the design space with a

sampling variable range of 0 to 1. For example, if the air infiltration is 0.8475, the projected value in the sampling space would be $\frac{0.8475-0.5}{(3-0.5)} = 0.139$ in the space of [0, 1).

6.1.3 User defined scaling

In the application of different ECM parameters, there can be both discrete and continuous variables as well as user-defined variables such as window U-factors of 0.5 W/m^2-K, 1 W/m^2-K, 4 W/m^2-K, and 10 W/m^2-K for the window retrofitting. By applying the mapping method described in section 6.1.1 and 6.1.2, all ECMs and their parameters will be projected into a uniform distributed sampling space between [0, 1).

6.2 Objective Functions

The objective functions in this research will include four parts: energy saving in Joule, energy saving in dollars, retrofit investment (including maintenance cost) and thermal comfort.

6.2.1 Building energy simulation

One of the most important parts of the objective functions is the evaluation of the energy performance of the building under different retrofit packages. We implemented the 5R1C modeling method using Python programming and created a simulation tool called SimBldPy as introduced in Chapter 4. The input file format is text based, similar to EnergyPlus and DOE2 engines and easy to manipulate by users or clients. The modified 5R1C modeling method is presented in Fig. 14.

The hourly energy use of a building with same type of utility, E, will be calculated as:

Equation 23:

$$E_u = \sum_i (E_{H,i} + E_{C,i} + E_{DHW,i} + E_{light,i} + E_{equipment,i} + E_{F,i}) + E_{pump} - E_{PV} - E_{SWH} \ (J)$$

$$E_{total} = \sum E_u F_u \ (J)$$

Where, $E_u, E_{pump}, E_{PV}, E_{SWH}$ represent hourly energy use of a same utility type, energy use of pump, energy production from solar panels, and solar water heaters, respectively, and $E_{H,i}, E_{C,i}, E_{DHW,i}, E_{light,i}, E_{equipment,i}, E_{F,i}$ is energy use of heating, cooling, domestic hot water (DHW), lighting, and equipment, respectively for i^{th} HVAC zone, in Joule. E_{total} is the total hourly energy use of all utility types and F_u is the primary energy factor for a certain type of utility. Among them, E_{PV} is electricity production from solar panel and its value of each time step can not exceed total electricity use. E_{SWH} is the thermal energy production, and its value of each time step can not exceed DHW thermal demand, in Joule. It should be noted that this equation is only for aggregating the same type of utility. If multiple utility types (electricity, gas, oil, and etc.) are used in a building, then total energy will be the primary energy in J that are transformed by primary energy factor for each type of utility.

6.2.2 Future year energy projection under climate change

In order to understand how climate change is going to affect the optimization of building retrofit, it is necessary to calculate building energy performance in future years. Using the morphing method developed by Belcher et al., future local hourly weather data can be downscaled from the global climate model (GCM) (Belcher, Hacker, and Powell 2005). The

detailed morphing algorithm and its results in 10 climate zones in the United States are obtained and presented in (Shen and Lior 2016).

After obtaining hourly weather data in the coming years, it will still be computationally heavy to calculate the hourly energy use of each year by SimBldPy. For each retrofit package (including the baseline case) in the optimization process, a data-driven RF model developed in Chapter 2 is trained by the dataset constructed by extreme year hourly weather data and the corresponding SimBldPy simulation results of each package, and the predictions of the RF model will serve as the basis for calculating energy saving in Joule and dollar .

Finally, the total energy saving in Joule of the lifespan can be simply calculated by:

Equation 24: $$E^u_{saving} = \sum_k (E^u_{post,k} - E^u_{pre,k}) \quad (J)$$

Equation 25: $$E_{saving} = \sum_u (E^u_{saving} * F_u) \quad (J)$$

Where $E_{pre,k}$ and $E_{post,k}$ is *k* year's annual energy use without retrofit (baseline case) and energy use after *k* years of retrofit for a particular type of utility, in Joule. *F* is the primary energy transforming factor. The total energy saving in Joules, E_{saving}, can be calculated by aggregating the use of different types of energy sources.

In addition, after the retrofit measures are applied to the building, the ageing of the ECMs will have an impact on the future performance of the building. In this research, by introducing retrofit maintenance during the post-retrofit period every five years which includes testing and maintaining the applied ECMs to ensure that their performance are as good as supposed to be, the ageing factor of ECMs in the future is assumed to be excluded in this research (Jaggs and Palmer 2000, Chidiac et al. 2011b, Rysanek and Choudhary 2013). The calculation of the maintenance cost will be explained in 6.2.4.

6.2.3 Thermal comfort

As an important part of the objective functions, the thermal comfort calculation is also introduced into the simulation process by assuming a constant metabolic rate at 1.1 met, a constant air velocity of 1 m/s. The objective function is defined by aggregating the absolute PMV values in all zones in each time step, making it possible to sum the thermal dissatisfaction in all overheated and underheated hours. Moreover, the future PMV values during the lifecycle will also be projected by the proposed RF model, forming the final objective function as the sum of the absolute PMV values in each zone throughout predetermined lifecycle in order to compare the results of different design vectors.

6.2.4 Financial modeling

In addition to energy saving in Joule, the objective functions also include energy saving in dollars, and retrofit investment. For the calculation of these last two sub-objectives, an LCA method is used taking into account the future increase in cost and the discount rate.

- Calculation of retrofit investment:

  Equation 26:

$$I_0 = \sum_i A_{wall,i} * C_{wall-insu,i} + \sum_j A_{roof,j} * C_{roof-insu,j} + \sum_l A_{win,l} * C_{win,l} + \sum_n A_{win,n} * C_{shade,n} + A_{floor} * C_{infl} + C_{others} \ (\$)$$

Where, $A_{wall,i}$, $A_{roof,i}$, $A_{win,l}$, $A_{win,n}$ is the area (m^2) of building wall, roof, and windows where i^{th} wall insulation, j^{th} roof insulation, l^{th} window material, n^{th} shading material

with certain cost ($), respectively. A_{floor} is the total floor area where infiltration level was improved, which has a cost of C_{infl} ($/m^2). Other costs include installing onsite renewable energy sources, upgrading building lightings, and etc.

It is assumed that all ECMs that have an initial investment will be maintained every five years, which will result in a periodic maintenance fee. Thus, the total retrofit investment can be obtained in the following equation:

$$I_{total} = I_0 + \begin{cases} \sum_k \frac{(1+\tau_m)^k * I_k}{(1+r)^k}, & k \,\%\, 5 = 0 \\ 0, & otherwise \end{cases} \quad (\$)$$

Where, τ_m (%) is the cost increase in maintenance fee of each year, r is the discount rate (%),and I_k is assumed to be proportional to the initial investment of kth ECM.

The introduction of maintenance fees every five years is to intended to ensure that the applied ECM will operate as well as in the beginning of its life. Though the maintenance cost occurs every five years, the cost increase τ_m and discount rate r will always be taken account into account in the calculation on an annual basis. For the case study in Chapter 8, the proportion of the maintenance cost of the initial investment is assumed to be 12% and is calculated every five years.

- Calculation of energy saving in dollars:

The utility cost for each different type of energy source will be assumed to increase annually, and the total energy saving in dollars in the lifecycle would be:

Equation 27: $S^u_{total} = \sum_k \frac{(1+\tau_u)^k * (C^u_{post,k} - C^u_{pre,k})}{(1+r)^k}$ ($)

Equation 28: $S_{total} = \sum_u S^u_{total}$ ($)

Where, S_{total}^{u} ($) is the total saving for type of energy source u, τ_u (%) is the cost increase for type of energy source u, and $C_{post,k}^{u}, C_{pre,k}^{u}$ is the energy cost of type u during post-retrofit phase for the building with and without retrofit. Finally, S_{total} can be obtained by aggregating the energy saving in dollars of each utility type.

6.2.5 Formulation of the optimization problem

After declaring the decision variables, and each sub-objective function, the multi-objective combinatorial optimization problem is then formed as follows:

$$\min Y_1(X) = E_{saving}(X)$$

$$\min Y_2(X) = S_{total}(X)$$

$$\min Y_3(X) = PMV_{total}(X)$$

$$\min Y_4(X) = I_{total}(X)$$

where, $X = (x_{wall-insu}, x_{roof-insu}, x_{win}, x_{infl}, x_{shade}, \dots)$

S.T.

$X \in [0,1)$

This optimization problem is a multi-objective combinatorial problem, and the possible design space could be huge. In section 6.2.6, the evolutionary algorithm that is used to solve the problem by finding the Pareto front in the solution space is described.

The reason of considering both energy saving in Joule and in dollar as sub-objectives is that the utility costs of different energy source are different and different energy source contribute

to energy use in the building incurred by different end use. For example, one ECM combination may provide propensity in reducing heating energy more than reducing cooling energy, such as sealing up the building and decrease infiltration rate while not using solar shading options during cooling season. Usually, electricity is used for cooling and gas or steam will be used for heating. Then the energy saving in Joule will not have a linear relationship with energy saving in dollars in this case. The inclusion of the two objectives is to ensure that tradeoffs incurred by such situation can be observed and analyzed in the optimization and decision-making process.

The inclusion of the sub-objective --- retrofit investment, in lieu of merging the investment and energy saving in dollar and thus taking net present value (NPV) as the objective function is that by doing this, the decision maker will be given a chance to see the total cost of each ECM combination because even though high investment can sometimes result in high returns and high net present value (NPV), the affordability of the retrofit early in the retrofit life will still be greatly favored by building owner or investor as there would be economic uncertainties in the future.

The adoption of summed PMV values as one of the objectives is to make sure that indoor thermal comfort can be taken into account as some retrofit options seem to be able to save a lot of energy, but at the same time, they can bring thermal comfort problems to the building, such as cooling and heating set point change and natural ventilation. The minimization of the PMV sub-objectives will prevent the retrofit packages that overheat and overcool the building too much from being selected. One of the concerns is that the adoption of this metric will create more complexity and tradeoffs to the multi-objective optimization problem and make the decision-making process more complicated.

In summary, other objectives can also be considered in the optimization such as greenhouse gas emission, indoor air quality (by assessing the necessary amount of fresh air

needed by a healthy indoor environment). The purpose of this dissertation is to provide a methodology in the optimization procedure, methods, and decision-making support framework. The selection of objectives in different retrofit project depends greatly on the conditions of each retrofit and can be subjective to change. In this research, the selection of the four objectives suffices for the discussion of the optimization, decision-making tradeoffs, and the need of the case study that will be introduced in Chapter 8.

6.2.6 Optimization algorithm

Traditionally a non-dominated genetic algorithm (NSGA-II) will be used to solve the problem in finding Pareto fronts (Deb et al. 2002). This non-dominated sorting algorithm has been proved to be efficient and effective in finding non-dominated solutions for multi-objective optimization problems (Chantrelle et al. 2011, Hamdy, Hasan, and Siren 2013, Shao, Geyer, and Lang 2014). However, unlike a normal optimization problem that NSGA-II confronts, the decision variables in this research have all been normalized in a continuous space between 0 and 1, so instead of GA, the differential evolution (DE) algorithm is used in handling the decision variables and the evolution for finding optimal solutions (Storn and Price 1997). The advantage of DE is that it is faster and more robust in convergence on the search for numerical optimization solution and is more likely to find the global optimum. Thus, the evolutionary algorithm is called non-dominate sorting differential evolution (NSDE) (Abbass, Sarker, and Newton 2001), where the same mutation and the same crossover strategy of DE are used while the selection criterion is adjusted by using elite non-dominated sorting as used in NSGA-II. The pseudo code for NSDE is shown in Table 16.

Table 16 Algorithm: non-dominate sorting differential evolution

Initialize population:
for i in (1, NP):
    for j in (1, D):
        $X_{i,j}$ = random_gaussian[0, 1)
    end for
end for
Do mutation, crossover, selection, and objective evaluation:
gen = 0
while (gen < Max_gen):
    for i in (1, NP):
        do a = random_gaussian[0, 1)*NP while a==i
        do b = random_gaussian[0, 1)*NP while b==i || b==a
        do c = random_gaussian[0, 1)*NP while c==i || c==a
        Perform mutation and binomial crossover for X_i and create trial vectors, $X_{t,jrand}$:
        jrand = rand*D
        for k in (1, D):
            if (rand[0,1) < CR) or k == D):
                $X_{t,jrand} = X_{c,i} + F(X_{a,i} - X_{b,i})$
            else:
                $X_{t,jrand} = X_{i,j}$
            jrand = (jrand+1)%D (get next parameter)
        end for
        Perform non-dominated sorting selection and evaluation:
        if ($X_{t,jrand}$ dominats X_i):
            add X_t to the pool P_{gen}
        else:
            add X_i to the pool P_{gen}
    end for
    F = [] (Pareto fronts)
    Perform non-dominated sorting selection and evaluation:
    for p in P_{gen}:
        for q in P_{gen}:
            Initialize np = 0, which is the number of individuals that dominate p
            Suppose $S_p = \emptyset$ which contains all the individuals being dominated by p
            if p dominates q:
                add q to the set S_p
            else:
                np += 1
        end for
        append p to F_{np} and calculate its crowding distance
    end for
    Then Sort all vectors in F by each vector's rank and crowding distance
After sorting, go through each front and add all the vectors to next generation's population pool P_{gen+1} until doing so would make len(P_{gen+1}) > NP
    m = 0
    while m <= NP:
        for p in sorted F:
            append p to P_{gen+1}
        m += 1
    end while
    gen += 1
end while

Note: NP: the population size; D: problem dimension; CR: crossover constant; F: learning rate; Max_gen: maximum generations.

- Generator

A generator is used to create the initial set of candidate solutions needed by the evolutionary computation. It is important for the convergence speed of the optimization process and the possibility of finding the global optimum. The generator in this research is using Latin hypercube sampling of Gaussian random fields that is good at generating a relatively small set of map realizations that captures most of the variability of the spatial inputs (Pebesma and Heuvelink 1999).

- Selector

The selector decides how to choose the individuals in the population who will create the offspring for the next generation. Selection has to be balanced with variation in crossover and mutation. The selector usually used for the non-dominate sorting genetic algorithm --- the tournament selection, is used in this research. The tournament selection is similar to the rank selection in terms of selection pressure, but it is more computationally efficient and more amenable to parallel implementation (Mitchell 1998). Two individuals are chosen at random from the population. A random number r is then chosen between 0 and 1. If $r < k$ (where k is a parameter ranging from 0 to 1, and 0.9 is used in this research), the fitter of the two individuals is selected to be a parent; otherwise the less fit individual is selected. Both two are then returned to the original population and can be selected again.

- Crossover

The main distinguishing feature of genetic algorithm is the use of crossover, and different crossover operator can result in different performance of the optimization (Mitchell 1998). Three different crossover operators are to be used and be compared in terms of their performance by the case study in Chapter 8. They are single point crossover, two point crossover, and uniform crossover.

For the single crossover, only one crossover position is chosen at random and the parts of two parents after the crossover position are exchanged to form two offspring. In single point crossover, the head and tail of a chromosome break up, and if both head and tail have good genetic material, then none of the offspring will get the both good features directly.

For the two point crossover, two positions are chosen at random and the segments between them are exchanged. Two point crossover is less likely to disrupt schemas with large defining lengths and can combine more schemas than single point crossover. This will allow the head and tail section of a chromosome to be accepted together in the offspring.

For the uniform crossover, each gene in offspring is created by copying it from the parent chosen according to the corresponding bit in the binary crossover mask of same length as the length of the parent chromosomes. For each element of the parents, a biased coin is flipped to determine whether the first offspring gets the "mom" or the "dad" element. In this research, the "biased coin" is set to have the same chance to adopt the element from the parents. Thus, the offspring will have a mixture of genes from both the parents.

A crossover rate of 0.8 is used in this research, which means around 80% of the offspring will be generated by crossover.

- Mutation

Mutation is basically a measure of the likeness that random elements of your chromosome will be flipped into something else. The existence of the mutation operator is to ensure the population against permanent fixation at any particular locus and thus playing more of a background role. Usually, a mutation rate between 0.005 and 0.01 is adopted (Mitchell 1998). The Gaussian mutation with a mutation rate of 0.01 is used in this research. Gaussian mutation adds a random number from a Gaussian distribution with mean zero and one as the standard

deviation to each vector entry of an individual and can be applied to float genes like the individual's genes in this research. The Gaussian distribution will be mapped to the each vector's bounding condition, which is 0 to 1 here.

- Population size

Population size defines how many chromosomes are in one generation. In this research, the population size is set to be 20 times the sum of all parameters listed in Table 20 to Table 22 for each generation. A maximum of 100 generations is used as the stopping criterion for the evolution process.

7. Decision Making Support Method

In this research, a decision-making support method is developed for the optimization results and its visualization. Traditionally, Pareto fronts archived through the optimization will be treated directly as a deliverable to the clients for decision-making (Shao, Geyer, and Lang 2014) (Roberti et al. 2017) (Tadeu et al. 2015) (Son and Kim 2016). However, the fronts could cover a wide range of solution sets in the design space, and it would still be hard for the user to target at solutions that they might be interested in by a predetermined preference, criterion, or state of mind. This situation could be aggravated with a high dimensional design space where more than three objectives are considered. Hence, a decision-making support scheme is developed here based on an unsupervised learning method --- hierarchical clustering (Johnson 1967).

A clustering problem can usually be described as follows (Hansen and Jaumard 1997):

$$\min Z = \sum_i \sum_j d_{ij}\, x_{ij}$$

S.T.

$$\sum_J x_{ij} = 1 \quad \forall i$$

$$\sum_J x_{jj} = m$$

$$x_{ij} \leq x_{jj} \quad \forall i, j$$

$$x_{ij} = 0 \; or \; 1 \quad \forall i, j$$

Where m is the designated number of clusters; d_{ij} is the dissimilarity between object i and j; x_{ij} measures if object i is assigned to certain cluster j, and it is a binary variable. The resolution to a clustering problem can be described as searching the best set of medians, which are able to assign all the points to and meanwhile minimizes the sum of the distances from all points to their respective cluster median, and one point should and only should belong to one cluster.

In this study, the hierarchical clustering technique will be used to find the group for each Pareto frontal points to which they belong. It is one of the most popular ways to assign data observations to clusters. It has been used to analyze market entry strategies (Robles 1994), design group technology manufacturing cells (Kamrani, Parsaei, and Chaudhry 1993), define employment sub centers in Los Angeles region (Giuliano and Small 1991), and most importantly, it can be used to visualize high dimensional data as other clustering techniques do (Agrawal et al. 1998, Kriegel, Kröger, and Zimek 2009, Parsons, Haque, and Liu 2004, Tadesse, Sha, and Vannucci 2005). The hierarchical clustering technique used in this research is based on agglomerative method, which starts with a cluster number of all the data points in the database. Basically, the number of all the Pareto fronts, n, is the initial clustering number. Then the algorithm will gradually merge the two most similar points into one cluster, and reduce the

number of clusters to n-1. By repeating the step, all the Pareto fronts will be agglomerated into one cluster that contains all the points, and the whole agglomeration process can be pictured by a dendrogram following a tree-like path.

The decision rule that is used to merge the clusters and form the similarity-dissimilarity matrix will be the major difference between those agglomerative methods. The decision rule that is used here is called the linkage method. The clustering method will calculate the similarity-dissimilarity matrix so as to compute the relationship between the new clusters and the remaining entities in terms of the linkage method (Blashfield 1976). There are several different linkage methods, but all of these methods can be described in the following equation to show how they compute this relationship (Müllner 2011):

Equation 29: $$d(h,k) = \alpha_i d(h,i) + \alpha_j d(h,j) + \beta d(i,j) + \delta |d(h,i) - d(h,j)|$$

where d function is the squared Euclidean distance between different entities; i and j are two clusters joined into a new cluster k = i ∪ j; h is the remaining entity. How $\alpha_i, \alpha_j, \beta, \delta$ are determined is based on different linkage method. For example, for single linkage clustering, the parameters are set as $\alpha_i = \alpha_j = 0.5, \beta = 0, \delta = -0.5$ (Sneath 1957). In this study, the ward linkage is used as proposed by Ward in 1963, which is also called the "minimum variance method" (Ward Jr 1963). The parameters used in this method are:

$$\alpha_i = \frac{n_i+n_h}{n_i+n_h+n_j}; \alpha_j = \frac{n_j+n_h}{n_i+n_h+n_j}; \beta = \frac{-n_j}{n_i+n_h+n_j}; \delta = 0$$

where n_i, n_i, n_h is defined as the number of points in cluster i, j, h, respectively.

With hierarchical clustering, a layered clustering scheme is developed to better group and visualize the Pareto fronts for decision-making support. Clustering is performed at each layer, allowing users to "zoom in" on the sub clusters of interest to them and then performing further

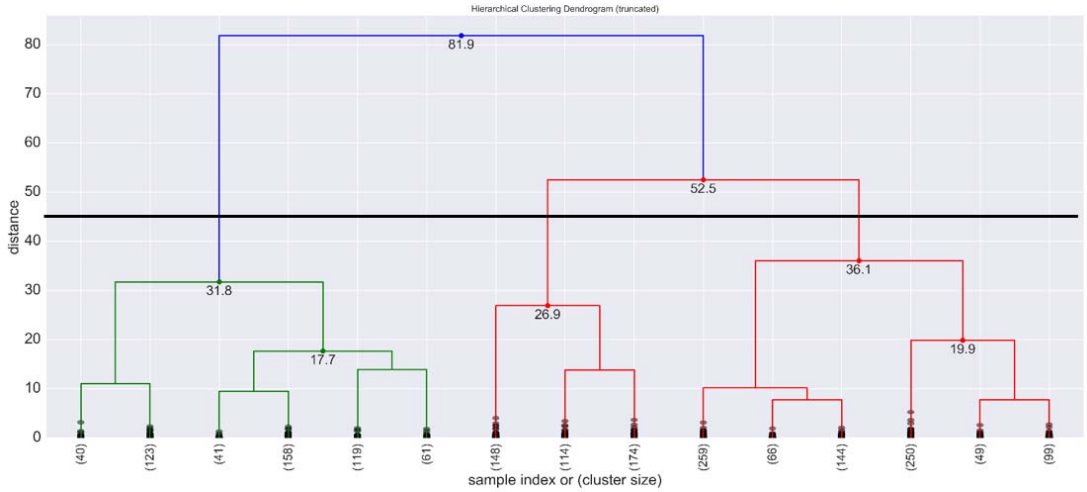
clustering on the sub clusters until the Pareto fronts in the cluster are retrieved and compared. For hierarchical clustering, this procedure can be simply conducted by examining the dendrogram and applying the linkage similarity-dissimilarity matrix to the clustering algorithm to find the certain sub clusters of the parent cluster.

However, there is still a problem: even with hierarchical clustering, the question of how many clusters to choose at each layer still exists, as other clustering techniques do (Guha, Rastogi, and Shim 1998, Halkidi, Batistakis, and Vazirgiannis 2001). Here we adopt an "elbow" method (Thorndike 1953), which attempts to find the clustering step where the biggest leap of distance growth happens in order to determine the number of clusters. That is to say, the location of a "knee" in the distance plotting for each step of agglomeration is generally considered as an indicator of the appropriate number of clusters. In this research, the proper cluster number will be determined at each layer according to the "elbow" method with a minimum number of cluster of three. This ensures that the process of layered clustering is fast (not too few clusters) and in the meantime remains visible to the users.

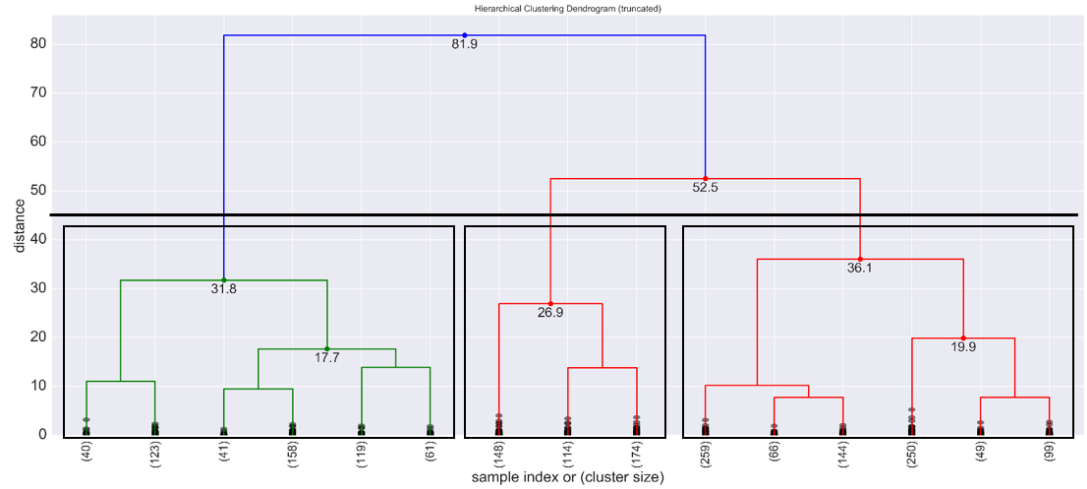
The process illustrated in Fig. 25 describes how the hierarchical clustering works in a layered framework to find the clusters of ECM combinations that are interesting to the decision maker. First, the "elbow" method described above will find the best distance of dissimilarity to determine the number of clusters in the first layer (step 1). Next, the dataset shown in Fig. 25 will be classified into three clusters (step 2). Next, choose the cluster that has the preferred sub-objective performance (step 3). Then repeat choosing the number of clusters using "elbow" method and find the sub clusters of the first layer cluster chosen in step 3 (the sub clusters are shown in step 4). Then repeat step 3 and choose the preferred sub cluster. This process can be iterated multiple times until the clusters with good overall performance are zoomed in and a

limited number of ECM combinations in the clusters are selected. It is worth noting that multiple clusters in the same layer can be chosen in the same time.

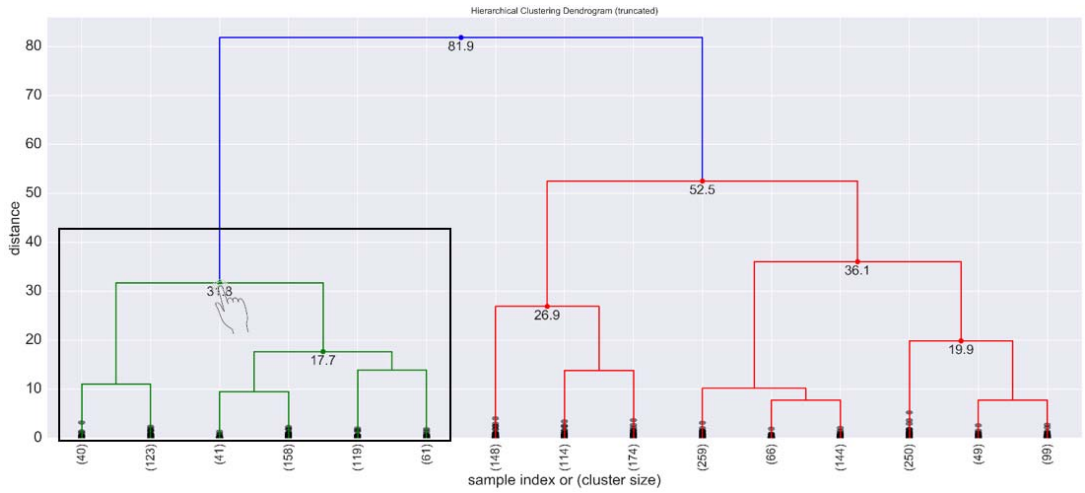
Step 1



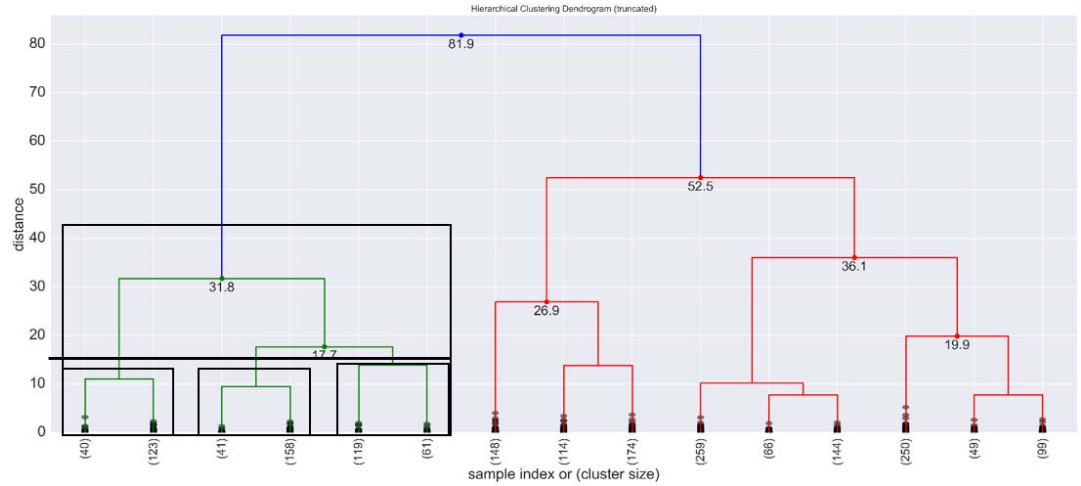
Step 2



Step 3



Step 4



Step 5

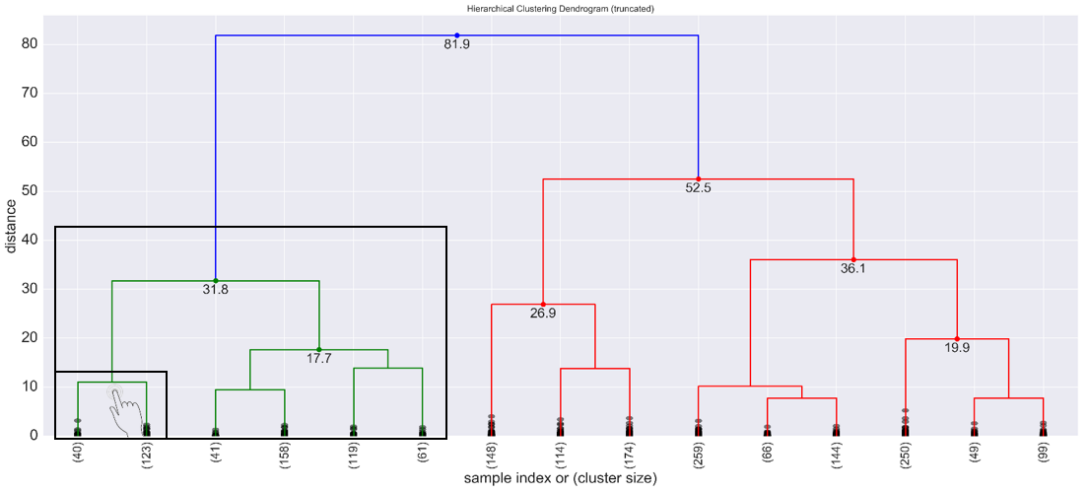


Fig. 25 Example of hierarchical clustering of a dataset in a layered framework

8. Case study and Results Discussion

8.1 Building Description and Model Calibration

The method developed in this research is implemented on one of the campus buildings at the University of Pennsylvania—the Towne building, which is designed in the manner of the English classicism of the seventeenth century. The building has 4 floors (with one basement floor) and is mainly composed of classroom and offices. The total floor area of the building is about 13000 m^2.

The simulation input for the building is collected in a text based file format including building geometric information, operation schedule, building systems, building envelopes, and etc., and then being fed to SimBldPy simulation tool. We adopted a classic "perimeter-core" modeling method to model this building. For each floor, including the basement floor, a core and a perimeter zone are modeled to make the SimBldPy model stay simple. The building envelopes are also modeled for each zone. The brief information on the building envelope is reported in Table 17.

Table 17 Towne building envelope in each orientation

Orientation	Opaque (m^2)	Window (m^2)	Below Grade Opaque (m^2)
S	1787.6	406.8	259.0
SE	0.0	0.0	0.0
E	948.4	187.3	155.7
NE	0.0	0.0	0.0
N	1127.9	256.9	218.2
NW	0.0	0.0	0.0
W	1028.9	246.5	147.3
SW	0.0	0.0	0.0
Roof	3995.9	0.0	0.0

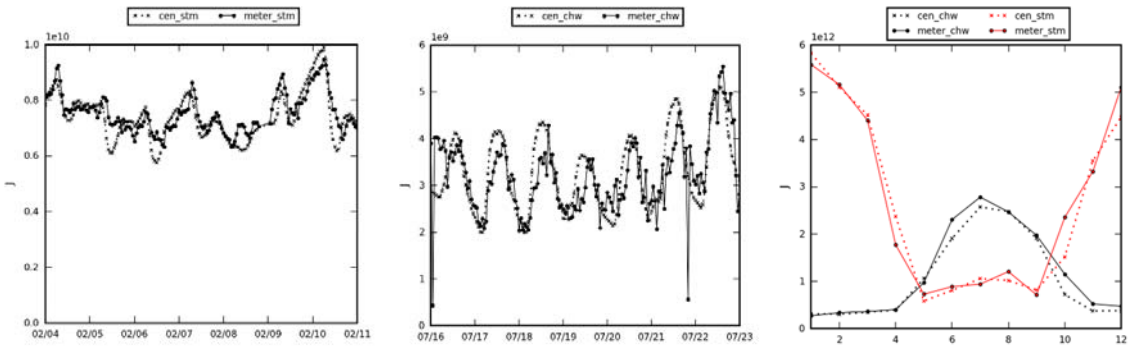
The model is calibrated with its actual energy performance in 2015 by metered hourly and monthly energy use data, which are stored and maintained by Penn Facilities and Real Estate Services (FRES). The heating and cooling set point of the building is constant, which is 21.8℃ for cooling and 22.5℃ for heating, respectively. In Table 19, the operation schedule of Towne is calibrated based on its electricity use pattern because cooling and heating energy use are separated from electricity use. The building wall section consists of outside air film, face brick, air cavity, CMU (concrete masonry unit), air cavity, veneer plaster, and inside air film. The calibrated thermophysical properties of the building envelope used in building simulation are shown in Table 18. All the campus building uses district cooling and heating source in the form of chilled water and steam. Thus, the building simulation model is calibrated with the metered chilled water and steam usage so as to get prepared for the following retrofit optimization and ensure the optimization have practical significance. The simulation results of the calibrated building model are shown in Fig. 26. The dotted line is the simulated result and the solid line is the actual use. Possible reasons for the deviation from the actual and simulated energy use of the building may be the real time use schedule, local microclimate condition, HVAC system control strategy, system fault, impact of occupant behavior, and etc.

Table 18 Calibrated thermo physical properties of building envelopes

Envelope	U-value (W/m2 °C)	Absorption coefficient (SHGC for window)
Wall	1.1	0.8
Roof	0.92	0.78
Below-grade	2.95	0.81
Window	4.16	0.69

Table 19 Towne building operation schedule

time of day	wd_occ	we_occ	wd_app	we_app	wd_light	we_light
From 0 To 7	0	0	0.72	0.51	0.72	0.51
From 7 To 8	0	0	0.73	0.55	0.73	0.55
From 8 To 9	0.3	0.4	0.78	0.58	0.78	0.58
From 9 To 16	0.5	0.5	0.82	0.62	0.82	0.62
From 16 To 17	0.95	0.5	0.82	0.62	0.82	0.62
From 17 To 18	0.7	0.5	0.83	0.63	0.83	0.63
From 18 To 22	0.95	0.5	0.84	0.64	0.84	0.64
From 22 To 24	0.7	0.3	0.75	0.65	0.75	0.65



winter week hourly heating summer week hourly cooling monthly heating & cooling

Fig. 26 Building model calibration

In order to show how the simulation results of SimBldPy model compare to the metered data, the scatter plot depicting outdoor air temperature versus cooling and heating energy use of meters, SimBldPy model and a data-driven RF model are drawn respectively in Fig. 27. The plot

shows the “signature” of the building performance in terms of heating and cooling energy use. The RF model is trained on the real measured weather data in 2015 and the occupancy schedule used in SimBldPy model. The original purpose for training this RF is to provide a self-benchmarking tool that examines how the building is performing in the years after 2015.

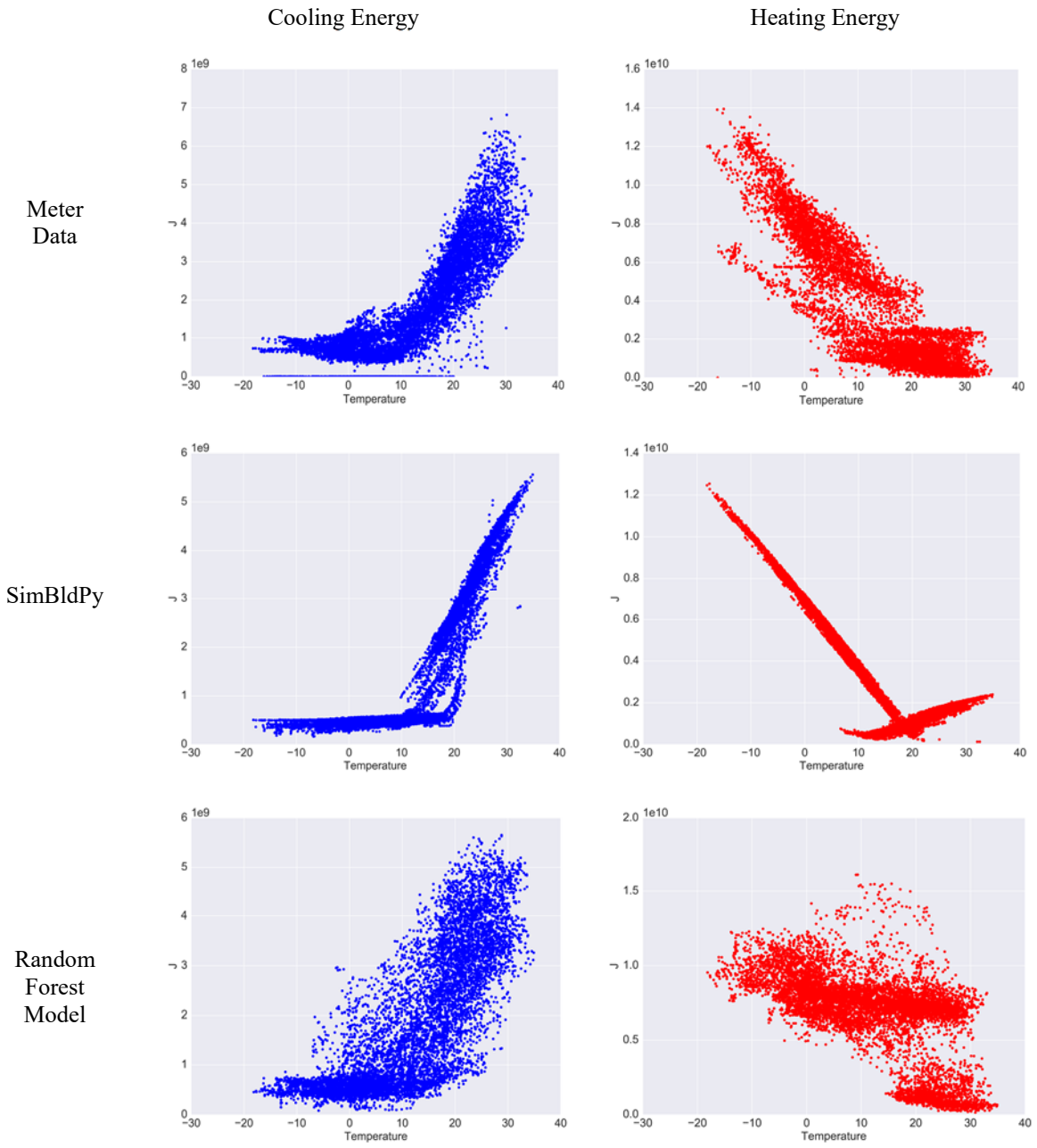


Fig. 27 “Signature” plot of cooling and heating energy from meters, SimBldPy model, and RF model for Towne Building

The difference between SimBldPy model and the metered data is mainly due to the fact that the simplified hourly model assumes that the building’s heating and cooling system is performing ideally just as described in section 4.1, while in practice, in a building like Towne, the HVAC system may suffer from degradation and ageing, which can possibly result in overheating or overcooling due to system faults or failure. Thus, the actual operation state can deviate from what the system is supposed to perform. Other reasons for the difference between the model and metered data could be attributed to the difference in actual occupancy and use, as well as the actual operation and management of the building system. It is should be noted that the “elbow” in the plots, where a watershed of different slopes of temperature versus energy use occur, reflects the different operation mode of the building system concerning heating and cooling. It is generally at that point that the building switches its system operation between cooling and heating. These points are also called “change point” (Paulus, Claridge, and Culp 2015).

The downscaled future hourly weather data is also an important input for the optimization model and is obtained by using the morphing method described in (Shen and Lior 2016). Fig. 28 and Fig. 29 show the trends of monthly mean dry bulb temperature and downwelling shortwave radiation from the year of 2017 to 2069 under different RCPs scenarios, respectively. In the following study concerning building retrofit and its optimization, RCP6.0 scenario will be used as the future climate scenario. The full set of downscaled climatic variables includes dry bulb air temperature, relative humidity, solar irradiation, and wind speed.

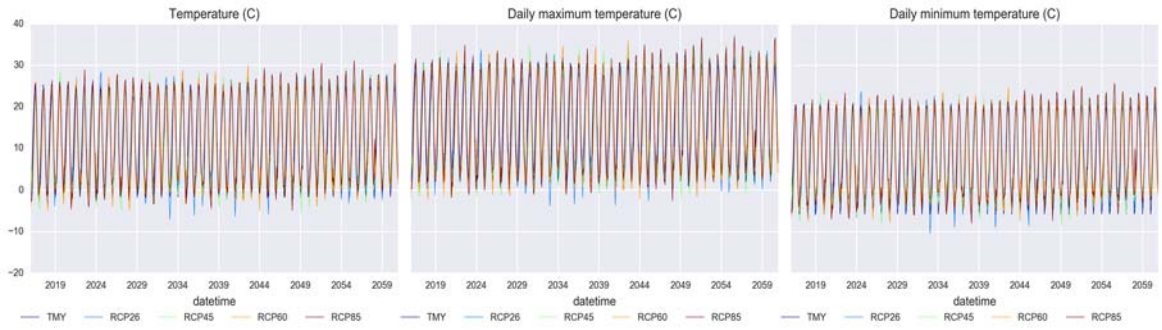


Fig. 28 Monthly mean air temperature, daily maximum and minimum temperature in different RCPs and TMY in Philadelphia

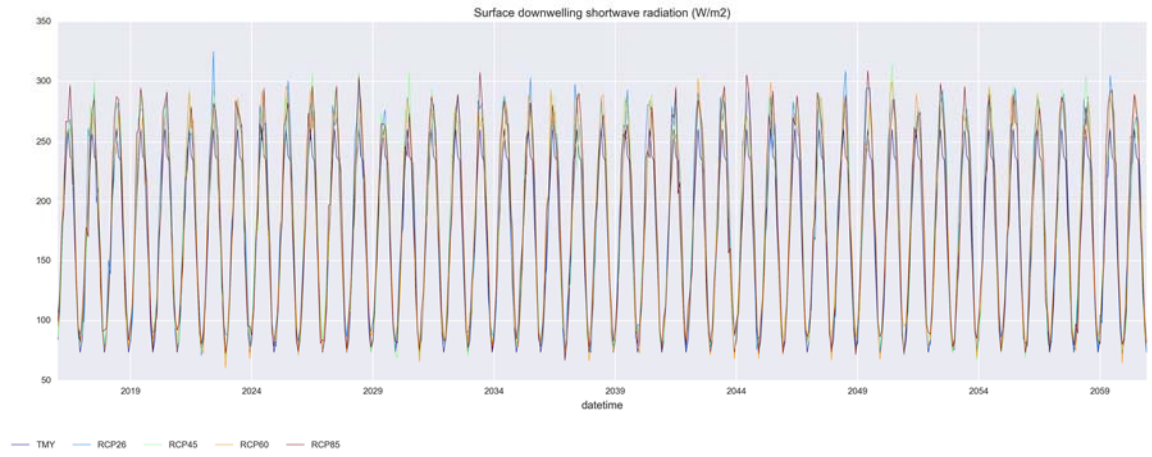


Fig. 29 Monthly mean downwelling shortwave radiation in different RCPs and TMY in Philadelphia

8.2 ECM Options and Costs

The following ECMs are considered in the retrofit: window replacement, wall insulation, roof insulation, window shading, air infiltration improvement, cooling supply air temperature, heating supply air temperature, lighting efficiency, daylighting control, natural ventilation,

cooling set point, heating set point, unoccupied hour setback, PV panels, and solar water heater (SWH). The retrofit lifecycle is assumed to be twenty years, namely, from year 2018 to 2038.

The parameter and cost of all ECMs are listed in Table 20 to Table 22 (Shao, Geyer, and Lang 2014, Wang, Xia, and Zhang 2014, Asadi et al. 2012a). For PV and SWH system, different inclination angles are also considered in the optimization. PV system is assumed to be multi crystalline silicone cell with an efficiency of 0.13kW/m^2 (Yoza et al. 2014), and the optimization parameter is the amount of available roof area used for solar power generation, while the optimization parameter of the SWH system is that if the rest of the roof area is used to install SWH collectors. The SWH system is assumed to have an overall efficiency of 50% (Yoza et al. 2014). The cost of the PV panel and the SWH is \$274.7/m^2 and \$213.4/m^2, respectively. In addition, setting the inclination angle of both PV and SWH system to non-zero will incur a frame support installation fee of \$30/m^2.

Table 20 Window replacement properties and cost

window (SHGC, U-value (W/m2 °C))	$/m2	type
(0.0, 0.0)	0.00	N/A
(0.80, 3.6)	47.0	Single glazing
(0.75, 2.8)	53.2	2bl glazing Without thermal break
(0.62, 1.6)	75.2	2bl glazing low-e window
(0.44, 1.6)	92.9	2bl glazing Window air-filled metallic frame
(0.288, 1.05)	79.2	SGSILVER
(0.585, 0.52)	98.1	SGCLIMATOP
(0.28, 0.33)	113.4	3050 SH 7/16 inch glass low-e
(0.63, 0.48)	131.7	3050 SH 7/16 inch glass
(0.25, 0.26)	183.0	3050 DH 3-7/16 insulated glass low-e krypton filled triple pane

Table 21 ECM parameters and costs (w/cost)

wall insulation (m^2 °C/W)	$/wall m^2	roof insulation (m^2 °C/W)	$/roof m^2	window shading	$/window m^2	air infiltration (h^{-1})	$/floor m^2	lighting efficiency improvement	$/floor m^2	daylight control	$/floor m^2
N/A	0	N/A	0	N/A	0	N/A	0	N/A	0	N/A	0
1.25	11.4	1.52	12.5	1	28.7	0.3	25.5	30%	3	Applied	3
1.61	12.5	1.97	16.4	2	37.2	0.5	20.2	40%	1.9		
1.97	13.5	2.42	20.1			0.7	14.4				
2.33	14.6	2.87	22.9			0.9	9.3				
2.69	15.7	3.32	26.8								
3.05	16.7	3.77	30.3								
3.41	18.5										
3.77	20.5										

Table 22 ECM parameters and costs (w/o cost)

cooling supply air temperature (℃)	heating supply air temperature (℃)	natural ventilated window ratio	cooling setpoint (℃)	heating setpoint (℃)	unoccupied hour setback
N/A	N/A	N/A	N/A	N/A	N/A
11	52	10%	22	18	Applied
12	51	20%	23	19	
13	50	30%	24	20	
14	49	40%	25	21	
	48	50%	26	22	
		60%	27	23	
		70%		24	
		80%		25	
		90%			
		100%			

8.3 Preliminary Run with Single ECM

Before the building retrofit being optimized using the procedure described in Chapter 6, a preliminary study to find how the building acts under single ECM described in Table 20, Table 21, and Table 22 is conducted. This is to find how single ECM is going to influence the building energy performance and economic return performance during the life span of post-retrofit period before the interactions and tradeoffs among ECMs are considered into the study when multiple ECMs are applied to the building simultaneously. This is conducted by looping through the parameters of each ECM while keeping other ECMs not applied to the building. Then, the parameter for each ECM with the best performance as well as its according simulation results of energy and economic performance will be recorded. As shown in Fig. 30, the simulation results are presented in a marginal abatement cost curve (MACC), which is a straightforward way to show the best effect of each single ECM to the building in terms of the NPV.

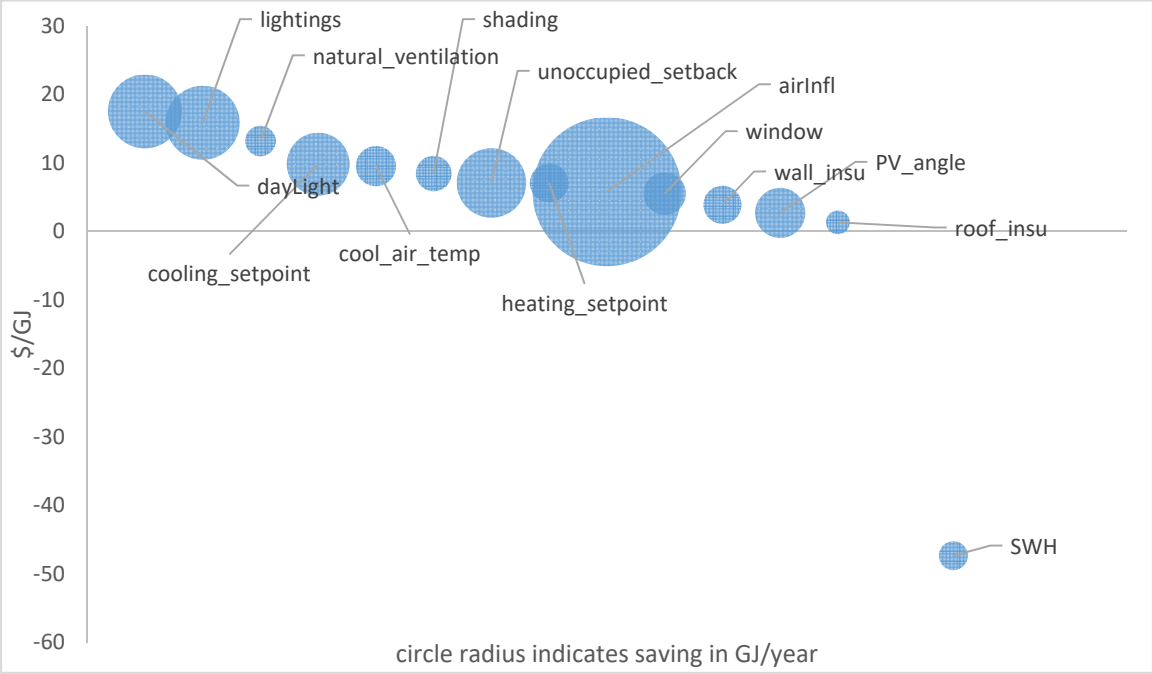


Fig. 30 Marginal abatement cost curve for Towne Building

The results shown above are obtained from an economic model with a lifespan of 20 years and 3% discount rate. The y-axis in the figure is the NPV of each ECM divided by energy saving in gigajoule (GJ) and the radius of the blue circle represents the energy saving in divided number of years in the life span. The results show that ECMs with high energy savings do not necessarily lead to a high economic return, which can be caused by high initial and maintenance costs. ECMs that are assumed to be "free" in this study (those shown in Table 22) usually show great potential in economic benefit since there is no investment involved. The best ECMs with high NPV are lighting retrofits, natural ventilation, cooling set point change. Air infiltration is a major contributor to energy saving, but the economic return is not as good as unoccupied hour set point setback because of its high test cost and retrofit investment.

This section aims to show how each single ECM is going to perform using life cycle cost analysis. When multiple ECMs are applied to the building, the situation will be more complicated. For example, when lighting power is reduced in a retrofit to save electricity, heating load will at the same time be higher and more heating energy will be consumed. In addition, utility cost for electricity and heating energy source (in this case district steam) are different, the tradeoffs and complexity of the problem will make it difficult in fathoming the optimal choice for retrofit.

8.4 Optimization Results

The optimization process takes about 23 hours for the building retrofit using parallel computation with 32 threads, which has a population size of 20 times the sum of all parameters listed in Table 20 to Table 22 for each generation. For each generation, the current non-dominated solutions will be archived and compared with last generation's archive of solutions.

The difference between the current and last generation's archived non-dominated solution will be the "newcomers" to the archive, and those "newcomers" will be used to evaluate the convergence of the solution. The reason of using this convergence-examine rule instead of using the total number of non-dominate solutions in each generation is to make sure that even with the same total number of Pareto fronts, there would be no new fronts that replace the older ones in the archive. The growth in the number of newly generated Pareto fronts in each generation are shown in Fig. 31 for the three crossover operators described in Chapter 7.

As per Fig. 31, uniform crossover outperforms the other two crossover operators. It has a better convergence performance in the problem, which can be caused by the fact that uniform crossover has no positional bias and any schemas contained at different positions in the parents can potentially be recombined in the offspring. The number of newcomers becomes stable and les than 10 for each generation. Indeed, there may be more newcomers being generated and going into the archived non-dominated solutions, but maximum generation number of one hundred is sufficient to find most of the Pareto fronts as shown in the result of convergence.

The populated Pareto fronts are displayed in Fig. 32. More than one thousand Pareto fronts are found during the optimization. The simulation time for a single year with extreme weather for each ECM combination in SimBldPy is about 1/30 to 1/40 of EnergyPlus model that has the same modeling complexity, and with the help of RF, it becomes possible for a moderate server to complete certain deep retrofit optimization task that takes into account future hourly energy projection under climate change in a fast manner.

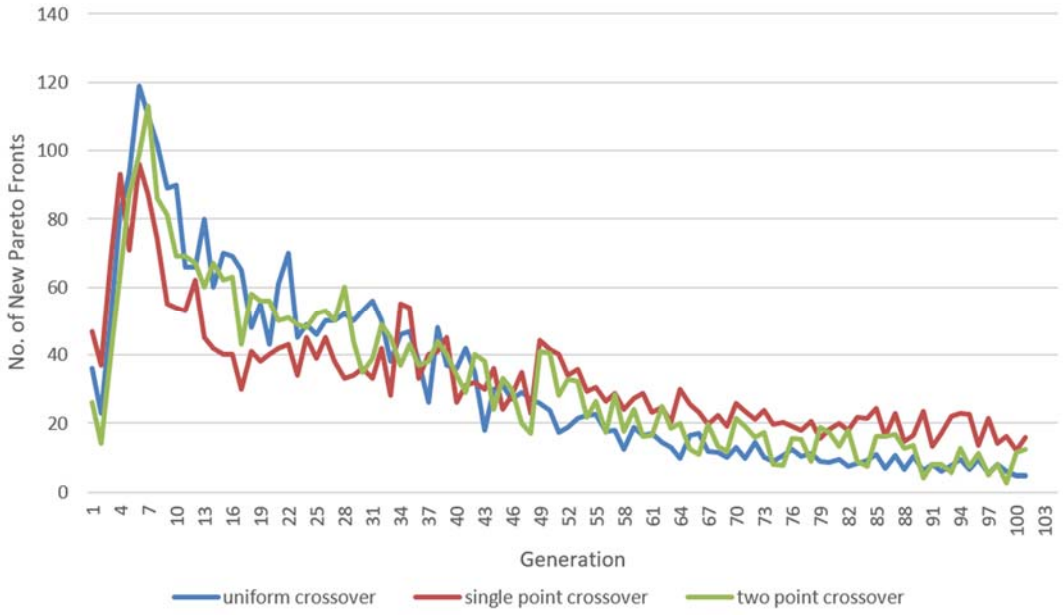
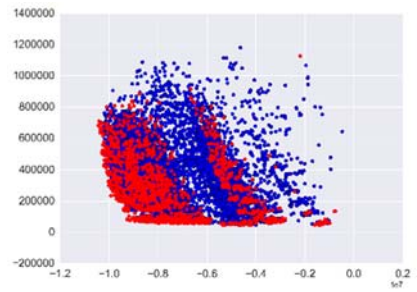
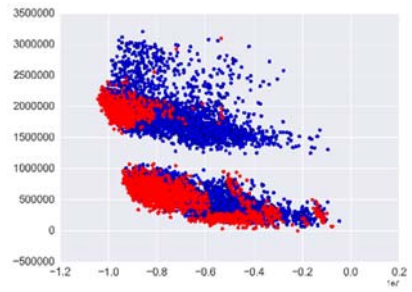


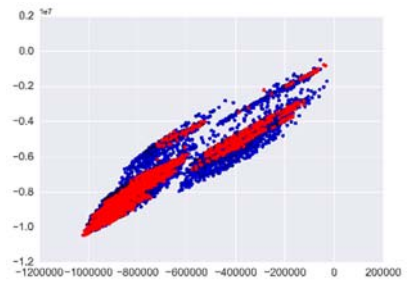
Fig. 31 Convergence of the optimization results



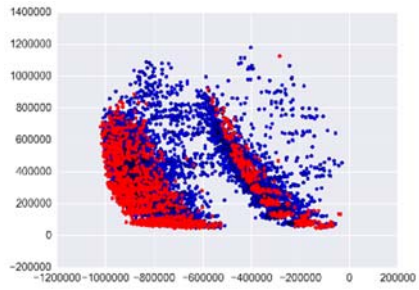
energy_saving ($) vs PMV



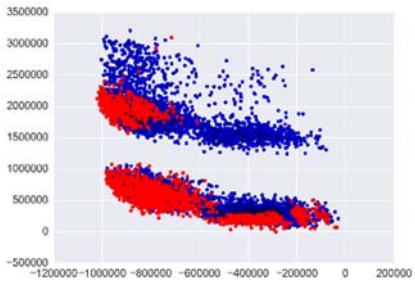
energy_saving ($) vs retrofit_cost ($)



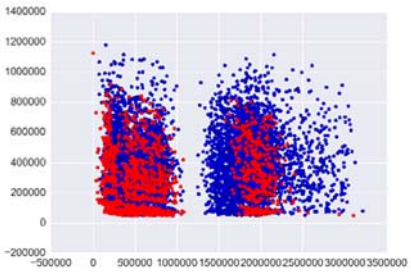
energy_saving (GJ) vs energy_saving ($)



energy_saving (GJ) vs PMV



energy_saving (GJ) vs retrofit_cost ($)



retrofit_cost ($) vs PMV

Fig. 32 2-D projection of Pareto fronts with all combinations of objective functions

As discussed in Chapter 7, it is difficult for the clients or users to fathom the optimization results with a high dimensional data structure. With the information provided in Fig. 32, it would still be difficult to make decision and have a general idea of which retrofit options to choose from. For a deep retrofit project, many ECM options as well as objective functions will be concerned as in this case study. The generation of about 1500 Pareto fronts in this example shows the difficulty in presenting the results. Thus, in the next section, the decision-making support method based on the layered hierarchical clustering proposed in Chapter 7 will be implemented to the optimization results of this project.

8.5 Implementation of Decision Making Support Method

The archived Pareto fronts dataset are first normalized by their means and standard deviations before being clustered. After applying the agglomerative hierarchical clustering method to the generated Pareto fronts at the first layer, the data is clustered into three different classes, as indicated by the suggested “elbow” method and the clustered data is shown in Fig. 33.

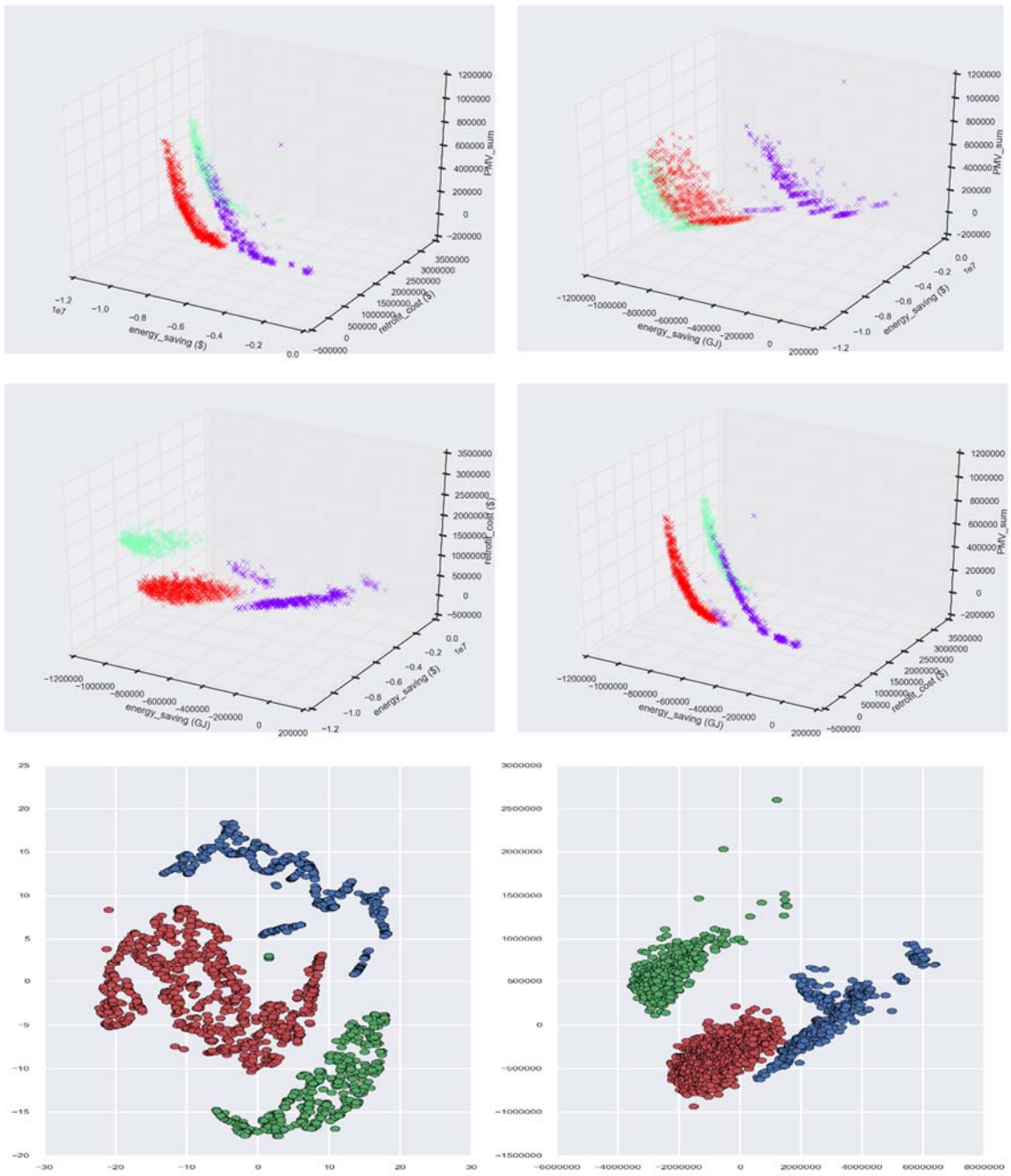


Fig. 33 3-D projections and dimensionality reduction visualization (t-sne & PCA) of the clustered Pareto fronts

According to the 3D projections on different combinations of objectives, it is shown that the hierarchical clustering in the first layer is doing a good work in classifying the data to the

right group in an unsupervised manner. For the cluster colored in blue, some data points are off the cluster, and according to the first and third plot in Fig. 32, it is inferred that the existence of these Pareto fronts could be due to their good performance in thermal comfort since some ECM combinations have the characters of low cost and high thermal comfort performance. For example, thermostat set point setback during unoccupied hours can reduce energy use and have a limited impact to indoor thermal comfort during the occupied time, while the energy use of ECM combinations having set point setback is not as much as others containing window replacement, but they will still be counted as non-dominated fronts.

The low dimension visualization in Fig. 33 further proves that the clustering result is a good representation for the nature of the data structure. With the recent development in machine learning algorithms and computational efficiency, high dimensional data can be visualized in more versatile and powerful ways. Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components, making it possible to linearly project the inherent structure of the data into low dimensional (Pearson 1901). Moreover, since non-linearity may exist in the dataset of Pareto fronts and a linear projection method such as PCA is not as sensitive to non-linearity, one of the low dimension embedding methods, also called manifold learning method, is adopted to show the 2D projection of the data points too. T-distributed stochastic neighbor embedding (t-sne) is a machine learning method that is able to reduce the dimensionality of the data to two or three in the way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points. The affinities in the original space are represented by Gaussian joint probabilities and the affinities in the embedded space are represented by Student's t-distributions (Maaten and Hinton 2008). The advantage of this algorithm is that it is able to scale each feature with different unit and dimension into one plot while avoid conglomerating them together. In both t-sne and PCA

plots in Fig. 33, it is indicated in both linear and non-linear perspectives that the chosen clusters are well-suited for this clustering problem.

When implementing the suggested clustering based decision-making support framework, the first layer clustering is extremely important because many decision vectors can be eliminated in this first stage of decision-making. Thus, it is important to visualize the data in a more intuitive way for the decision makers or users. The parallel coordinates plot together with suggestive heat map will be used as a support technique to visualize the clusters for decision-making. They are plotted in Fig. 34 and Fig. 35 (parallel coordinates figure and heat map will also be plotted at each subsequent layer for decision-making, but will not be redundantly shown here):

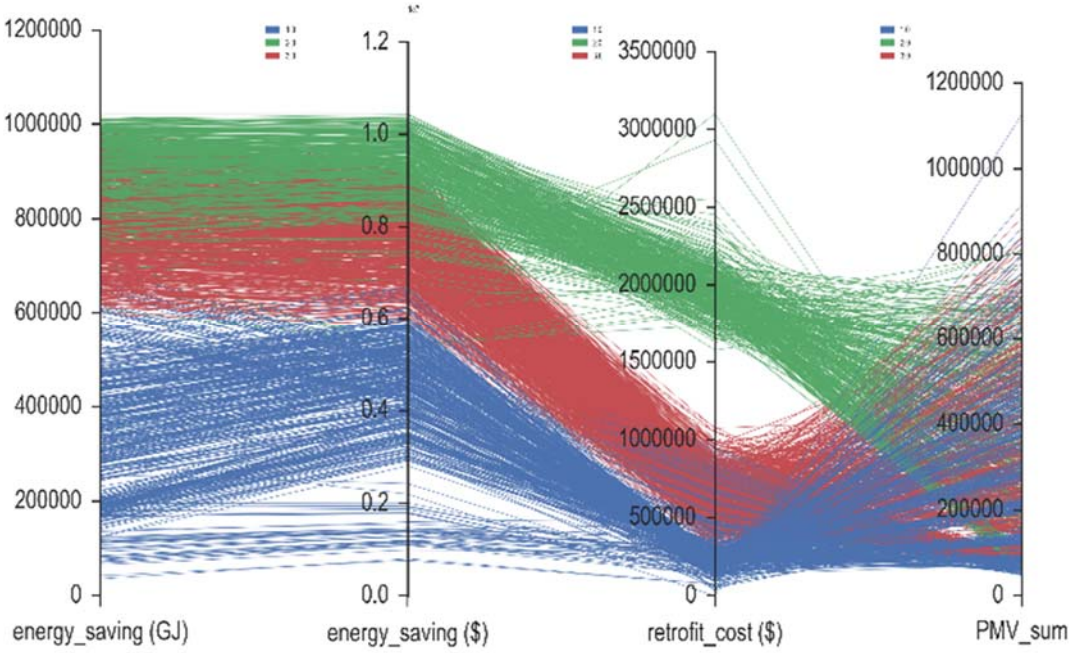


Fig. 34 Plotting of sub-objectives in the first layer clustering

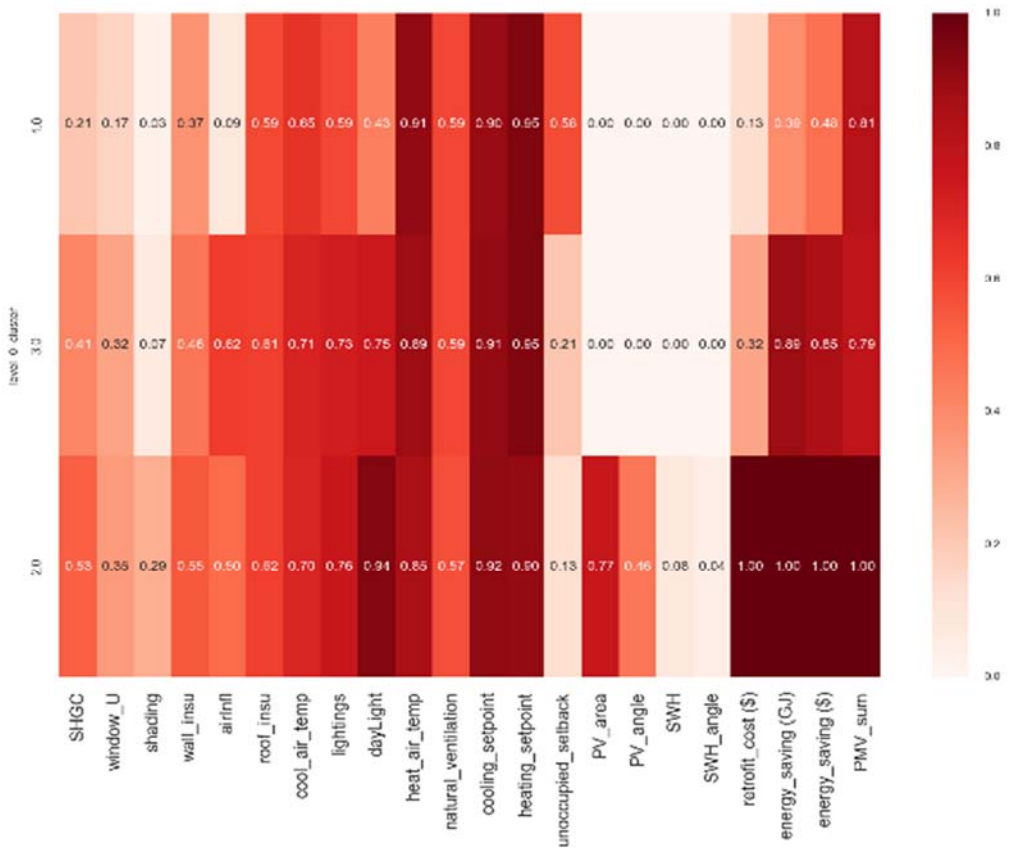


Fig. 35 Heat map of the first layer clustering

By plotting the clustered Pareto fronts in parallel coordinates, the decision-making process will become more visible as the tradeoffs among objectives can be illustrated in a straightforward way. Combined with heat map showing the average of both decision variables and objectives in one graph, it becomes intuitive for the decision makers to see what happens to the clusters. In Fig. 35, each row in the heat map represents a cluster, and each column provides a comparison of the means of each decision variables and each objectives among clusters. The clusters (rows) in the heat map are sorted by the retrofit investment of each cluster. The white-colored (or zero-valued) ECM in the heat map indicates that some ECMs are not adopted and applied in that cluster.

As analyzed at the end of section 6.2.4, including both energy savings in Joule and in dollar makes sense for the optimization problem because the visualized results in Fig. 34 show that the energy saving in Joule does not go linearly correlated with energy saving in dollar as a whole. There are ECM combinations that have both positive and negative slope rates between the two objectives, and the magnitudes of the slopes also differ from each other.

According to the parallel coordinates plot in Fig. 34, the thermal comfort levels of Pareto fronts in the three clusters are quite scattered, making it barely easy to decide which cluster to choose from, but it can be clearly said that the cluster colored in red (Cluster 3) is the most interesting cluster to look at since the unit investment produces better amount of unit energy saving. Integrated with the information provided in the heat map, the color of each ECM grows darker in Cluster 2, meaning that the certain kind of ECM is being selected more in that cluster. In addition, in this case study, ECM combinations with renewable options such as PV and SWH all belong to Cluster 2 and also have the highest investment rate, implying that renewable energy systems are major contributors to investment growth in retrofit project. In Cluster 3, with an average of 32% of the highest investment value among the Pareto fronts, about 85% of energy savings can be harvested without harming the thermal comfort on average. Therefore, retrofit options in Cluster 3 will be chosen and enter the next layer's decision-making in this case study. It should also be understood that eliminating Cluster 3 will abandon all renewable energy options thanks to the visualization provided in the heat map.

8.6 Decision Making Pathway

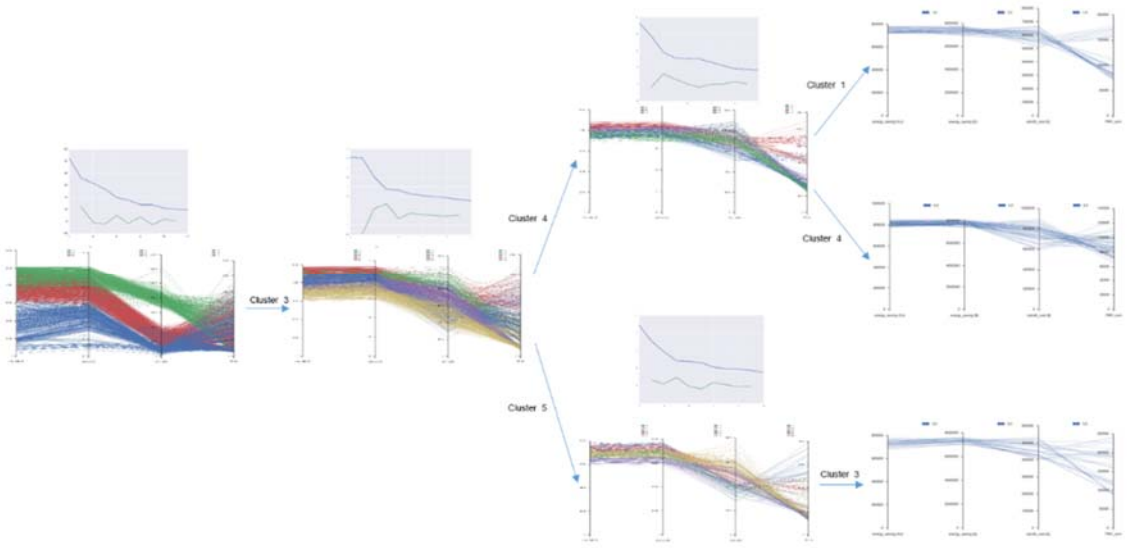


Fig. 36 Layered decision-making pathway

By repeating the process in the 1st layer clustering as described in Chapter 7, a pathway can be plotted in Fig. 36. In Fig. 36, on top of each parallel coordinates plot, the distance of each clustering step and its first-order differential curve in that layer are attached to show how the elbow method works and determines the number of clusters for each layer. For the Pareto fronts generated by the optimization process in this case study, the following clusters and sub clusters are selected as shown in Fig. 36.

The pathway clearly shows how each decision is made at each layer and what clusters are selected and zoomed. Eventually, three 4th layer clusters are chosen as the final target clusters. In this case study, the criterion in choosing the each layer's cluster is based on the principle of lower cost, higher energy saving return, and better thermal comfort level. After picking retrofit packages with a cost lower than $810,000 and sorting them by their NPV in twenty years (2018 - 2038), top twenty ECM combinations from the three chosen clusters are listed in Table 23, from which the most suitable combination that is tailor-made for the current building can be chosen.

Since renewable options have already been crossed out during the 1st layer's decision-making, they will not be shown in Table 23.

It should be noted that the final results derived from the decision-making framework and presented in this section can be subject to change depending on different criteria. For example, if the building owner or client does not care much about the initial cost and maintenance cost, it is possible that Cluster 2 would be chosen in the first layer as this cluster produces the most energy saving either in dollars and Joules and renewable options will also be included in the final results.

As already discussed in Limitation Two in the section 1.2, subjectivity will always be part of the decision-making process whereas the difference in this proposed decision-making framework is that choices are provided with increasing information to the decision maker for all possible optimized results as the layered hierarchical clustering unfolds on its pathways.

Different from the use of weighted sum or product method that requires hard-to-decide and whimsical weighting factors before the solution space is generated and visualized, this framework offers options and layered reasoning that leads to the final results; and different from traditional processing and visualization of the Pareto fronts, how the inherent structure of these high dimensional solutions are rendered, and how important it is to visualize the structure and trade-offs among a possibly large amount of Pareto fronts in a deep building retrofit project where many ECMs involved are discussed. As the case study shows, the proposed decision-making support framework is manifested to show robustness in handling deep retrofit optimization problem and is able to provide support for brainstorming and enumerating various possibilities during decision-making process.

Table 23 Top 20 selected ECM combinations using the suggested hierarchical clustering-based decision-making support framework

No.	SHGC	window_U	shading	wall insu	infl	roof insu	cool_air_temp	lights	daylight	heat_air_temp	natural vent	cooling stpt	heating stpt	unocc setback	cost ($)	ES (GJ)	ES ($)	PMV
1	0.62	1.6	2	1.97	0.3	1.97	11	0.3	1	51	30%	25	24	1	806442	-853102	-8157555	85102
2	0	0	2	1.97	0.3	2.42	11	0.3	1	51	50%	25	24	1	729747	-838175	-8039577	99441
3	0	0	1	2.69	0.3	2.42	11	0.3	1	50	50%	25	24	0	802589	-834331	-8069318	88409
4	0.75	2.8	2	1.61	0.3	1.97	11	0.3	1	51	50%	25	24	1	779231	-842379	-8039496	99160
5	0	0	2	2.33	0.3	2.42	0	0.3	1	52	40%	25	24	1	736379	-834730	-7994319	98740
6	0	0	2	3.05	0.3	2.42	11	0.3	1	50	10%	25	24	1	749549	-841017	-8003161	97708
7	0	0	2	0	0.3	1.52	11	0.3	1	48	50%	25	24	0	606774	-803426	-7856490	110277
8	0	0	1	0	0.3	1.52	11	0.3	1	50	80%	25	24	1	606774	-809337	-7854209	125219
9	0.75	2.8	1	2.33	0.3	1.97	11	0.3	1	48	20%	25	24	1	792308	-847454	-8034243	97600
10	0	0	1	2.69	0.3	1.52	0	0.3	1	52	50%	25	24	0	704289	-820412	-7930031	99152
11	0	0	0	1.97	0.3	2.87	0	0.3	1	52	60%	25	24	1	744252	-829474	-7968637	95068
12	0.29	1.05	0	0	0.3	3.77	11	0.3	1	0	60%	25	24	0	807954	-825286	-8017933	83148
13	0	0	0	1.97	0.7	1.97	11	0.3	1	50	50%	25	23	0	513410	-767228	-7720925	168888
14	0	0	0	2.33	0.3	2.87	11	0.3	1	50	60%	25	24	1	750884	-823025	-7957244	84418
15	0	0	0	1.97	0.7	3.77	11	0.3	1	52	100%	25	23	1	583642	-780945	-7785804	171453
16	0.8	3.6	0	1.25	0.3	1.97	11	0.3	1	0	50%	25	24	0	764012	-825351	-7964434	92416
17	0	0	0	2.33	0.3	3.32	11	0.3	1	49	60%	25	23	1	770427	-823478	-7958761	81261
18	0.75	2.8	0	2.69	0.3	1.97	12	0.3	1	50	10%	25	24	1	798847	-838250	-7986575	87339
19	0	0	0	1.97	0.7	2.87	11	0.3	1	49	50%	25	23	1	546388	-768497	-7727967	158759
20	0.62	1.6	0	1.61	0.3	1.97	11	0.3	1	51	70%	25	24	0	809997	-830025	-7971569	85822

8.7 Summary

In this chapter, the multi-objective optimization problem usually encountered in existing building deep retrofit project is provided with an optimization scheme and a method of decision-making support. It is described how the method in resolving the optimization problem in a rapid manner by means of applying non-dominated sorting differential evolution algorithm (NSDE) to a campus building. By introducing the SimBldPy modeling tool and random forest (RF) models as the replacer for traditional energy simulation tools in the objective function evaluation, certain deep retrofit problem can be quickly optimized. Moreover, the generated non-dominated solutions, or so called Pareto fronts, are rendered and displayed in a layered way using agglomerative hierarchical clustering technique in order to make it more intuitive and sense making in the decision-making process as well as to be better presented to the clients and decision maker.

The strength of the developed optimization procedure lies in its adaptability and generalizability to different existing buildings and retrofit problems. The use of simplified hourly calculation method in building energy simulation not only reduces the time and computation cost for objective function evaluation during optimization, but also saves time and resources for the earlier-stage building modeling and calibration. With affordable simulation bias introduced by the SimBldPy model as described in Table 15, the method still suffices for achieving the goal of comparative parametric study in building retrofit problems.

Moreover, the optimization process introduced in 6.2.6 can be used to find optimal solution for single objective problems but its major application would be in multi-objective problems that involve linearly independent sub-objectives. It is also found that the uniform crossover operator works best in finding the optimal ECM combinations in building retrofit

problem compared with the traditionally used one point or two-point crossover operator mainly because it has no positional bias and any schemas contained at different positions in the parents can potentially be recombined in the offspring during the evolutionary optimization process.

The developed layered hierarchical clustering technique for decision-making support is a novel attempt to implement unsupervised machine learning algorithm to visualize and provide information for high dimensional data structure of the optimization results. This method unveils the chance of making decision on complicated Pareto fronts space using a pathway-like procedure that zooms into clusters at each layer and progressively finding a limited amount of ECM combinations with a specific decision-making logic. As subjectivity and preference do influence decision-making, the developed method offers a tool for screening undesirable solutions with rationale and appropriate visualization, which is very important in multi-objective optimization problems because traditional methods such as weighted sum or product method force the user to arbitrarily give weights or decision strategies to sub-objectives of various dimensions, scales, and decision-making values, in which priori bias to the multi-objective optimization problem may already have been introduced.

9. Conclusions

The objective of this research is to develop a novel and generalized method to optimize the selection of building retrofits. The dissertation is organized in three procedures: 1. evaluate the feasibility and importance of taking into account global climate change (GCC) in building retrofit planning; 2. develop a simplified simulation tool for building performance assessment based on a dynamic hourly simulation algorithm taking into account the zone thermal flux and coupling, and validating the modeling method; and 3. develop an optimization approach that uses

a non-dominated sorting technique to perform the optimization task and design a scheme for visualizing the optimization results and providing support for the decision-making process after obtaining the multi-objective optimization results.

For the first procedure, a Python script was developed to perform the parametric study by running EnergyPlus for different retrofit scenarios for existing buildings. Using a latin-hypercube sample (LHS) method and a Joint Mutual Information Maximization (JMIM)-based feature selection method, the most energy efficient ECMs for a target building can be selected, greatly reducing the computational cost of model training and prediction when attempting to assess the impact of GCC on building retrofit. Then random forest (RF) model is trained with the EnergyPlus hourly building energy use (BEU) results simulated with future extreme year weather data, which is able to predict the hourly BEU in future years during the post-retrofit phase. The same procedure applies to the simulation results of hourly energy use obtained by SimBldPy for the data-driven model training in the case study. It is found that GCC does influence ECM performance in the future, and its influence varies from building to building, and location to location. Moreover, the optimal retrofit strategy for selecting the best ECM combinations under current climate condition will be subject to change in future climate condition.

For the second procedure, a light-weight BES tool --- SimBldPy programmed in Python, is proposed and developed. The modified simplified hourly method used in this dissertation adds additional thermal resistances of the zone internal floor and internal walls as well as to adjacent zone temperature for each contact surface among zones to the original 5R1C modeling method described in ISO 13790. The built-in calibration module in SimBldPy adopts a differential evolution (DE) algorithm for the optimization of major modeling parameters with computational performance boost with the aid of parallel computation. Two DOE reference buildings (one residential and one office) located in Philadelphia and San Francisco were used to verify the

validity of the tool. The simulation results of the models have been compared with that of the EnergyPlus models of the four buildings, and it is found that given the calibrated model inputs, the proposed model has good fidelity in predicting hourly energy use with different heating and cooling set points in each zone, under various climate conditions, and with multiple ECMs being applied to the building. The tool aims to provide a rapid solution for parametric study of building energy, and the results in this study show that the computational performance and precision of the simplified model compare well with EnergyPlus.

The third procedure was to develop a method for resolving the optimization problem in a fast manner by using a non-dominated sorting differential evolution algorithm (NSDE). After introducing the simplified hourly simulation model and RF models as the replacer for traditional energy simulation tools for objective function evaluation, the deep retrofit multi-objective optimization problem can be optimized in 24 hours using a moderate server. Moreover, the generated non-dominated solutions are processed and rendered by a developed schema that uses the agglomerative hierarchical clustering technique in a layered manner to provide intuitive support for the decision-making process as well as to better present the optimization results to the clients.

The suggested optimization method is applied to a retrofit project for a campus building at the University of Pennsylvania with various ECM options and costs. SimBldPy is used to model the building's energy performance, and the model is then calibrated to the metered energy use. Using the future downscaled hourly weather data and the RF model trained on extreme year weather and energy use simulated by SimBldPy, the future hourly energy use data from year 2018 to 2038 can be projected for each combination of retrofits. The optimization is run in parallel with 32 threads, and the number of generated Pareto fronts converges in about 23 hours. More than one thousand Pareto fronts are generated and analyzed according to the decision-making

framework. Twenty ECM combinations are eventually selected from those fronts using specified decision-making criterion, while the application of the framework could be versatile since different decision-making criterion can lead to various solutions.

The suggested layered hierarchical clustering method is robust in dealing with the high dimension optimization results, helping to structure and visualize the decision-making process by choosing clusters of Pareto fronts that are of interest at each layer. The method provides the opportunity to explore various solutions and enhance decision-making with more information as the process unfolds. It is very difficult to determine optimized results ahead of time or to decide which decision-making strategy should be used, especially for multi-objective optimization problems with many sub-objectives. Building retrofitting is a complex process involving different aspects and goals, so this decision method gives designers and clients a tool to help them make better design decisions.

APPENDIX

Appendix I Heating and cooling load simulation accuracy of the SimBldPy model for the residential building (unit of load: GJ)

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.25, 0.26)	2	1.610	0.8	1.520	0.95	4.2	0.3	57.721	60.026	19.688	17.156	0.023	0.050	0.996	0.989
(0.62, 1.6)	0	1.250	0.8	3.774	0.95	4.5	0.3	55.043	56.478	31.315	32.040	0.027	0.035	0.994	0.990
(0.585, 0.52)	1	3.049	0.6	3.774	0.95	4.5	0.4	32.864	32.512	22.786	23.997	0.033	0.050	0.990	0.989
(0.8, 3.6)	2	3.413	0	2.421	0.95	4.5	0.3	78.467	78.448	18.676	17.215	0.022	0.043	0.996	0.978
(0.44, 1.6)	0	2.331	0.4	1.520	0	0	0.3	38.716	42.259	28.797	27.149	0.033	0.029	0.994	0.992
(0.28, 0.33)	1	0.000	0.6	3.322	0	4.5	0.3	50.445	51.912	20.021	18.979	0.026	0.031	0.995	0.988
(0.288, 1.05)	0	0.000	0	3.322	0	4.2	0.4	81.635	80.528	20.956	21.665	0.020	0.029	0.996	0.990
(0.62, 1.6)	0	0.000	0.8	1.969	0.95	0	0.4	69.117	69.917	34.325	34.325	0.020	0.026	0.996	0.993
(0.0, 0.0)	0	2.688	0	3.322	0.95	4.2	0.4	76.280	74.644	21.860	23.239	0.020	0.028	0.997	0.994
(0.62, 1.6)	0	1.250	0	1.520	0.95	4.2	0	71.786	73.190	37.146	34.994	0.022	0.027	0.996	0.993
(0.8, 3.6)	2	3.049	0.6	1.969	0	4.2	0	54.533	56.842	20.697	18.247	0.029	0.052	0.993	0.976
(0.25, 0.26)	2	2.331	0.6	3.322	0	4.5	0.3	38.334	39.527	14.729	12.668	0.026	0.042	0.995	0.989
(0.75, 2.8)	1	0.000	0	2.421	0	0	0.4	86.714	86.645	26.882	31.134	0.019	0.040	0.997	0.990
(0.585, 0.52)	1	2.688	0.8	2.874	0	4.5	0	44.823	43.780	24.886	27.243	0.027	0.038	0.993	0.991
(0.25, 0.26)	2	3.413	0.6	1.520	0	4.5	0.4	42.244	44.026	18.950	14.485	0.023	0.053	0.996	0.988
(0.44, 1.6)	0	1.250	0	3.322	0	4.2	0	70.882	70.398	25.285	26.505	0.023	0.033	0.995	0.992
(0.28, 0.33)	0	3.049	0.6	2.421	0	4.2	0	36.890	38.120	22.984	21.628	0.023	0.027	0.996	0.993
(0.62, 1.6)	1	3.413	0.8	1.520	0	4.5	0.4	58.156	58.336	27.214	27.740	0.021	0.022	0.996	0.995
(0.8, 3.6)	0	1.610	0.8	0.000	0	0	0	124.794	130.613	53.207	53.897	0.028	0.026	0.993	0.991

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.0, 0.0)	2	3.774	0.6	0.000	0	0	0	109.896	113.452	32.485	32.411	0.025	0.017	0.994	0.994
(0.28, 0.33)	1	1.969	0.4	1.969	0	0	0.3	29.911	33.520	23.147	18.771	0.040	0.046	0.993	0.989
(0.75, 2.8)	2	3.049	0.4	2.874	0	0	0.4	37.613	39.566	17.582	16.066	0.037	0.057	0.990	0.967
(0.63, 0.48)	1	3.413	0.8	3.322	0.95	4.5	0.3	43.790	42.311	23.764	26.845	0.029	0.042	0.992	0.991
(0.585, 0.52)	1	2.688	0	0.000	0.95	0	0.4	121.663	122.663	39.236	42.848	0.024	0.025	0.995	0.994
(0.44, 1.6)	0	1.610	0.6	0.000	0.95	0	0	106.733	111.766	40.101	43.234	0.027	0.027	0.994	0.993
(0.0, 0.0)	1	1.250	0.8	3.774	0.95	4.5	0	64.196	64.998	20.350	21.331	0.024	0.034	0.995	0.989
(0.288, 1.05)	0	0.000	0	2.874	0.95	4.5	0	80.648	79.597	22.523	22.934	0.020	0.027	0.997	0.991
(0.63, 0.48)	0	2.331	0.4	1.969	0	4.5	0	23.746	26.233	40.124	38.190	0.039	0.038	0.990	0.988
(0.288, 1.05)	1	2.331	0.6	3.774	0	4.2	0.3	43.185	44.535	18.466	17.104	0.024	0.027	0.995	0.992
(0.75, 2.8)	2	3.049	0	0.000	0.95	4.5	0.4	133.618	136.107	33.621	34.476	0.024	0.019	0.995	0.993
(0.62, 1.6)	0	1.969	0.4	0.000	0	0	0.3	91.735	98.364	45.931	48.488	0.033	0.029	0.992	0.991
(0.63, 0.48)	2	3.774	0	3.774	0.95	0	0	53.237	49.886	16.959	16.337	0.029	0.046	0.993	0.976
(0.0, 0.0)	1	3.413	0	3.322	0.95	4.2	0.3	75.988	73.670	18.631	19.769	0.020	0.024	0.997	0.995
(0.63, 0.48)	0	1.969	0.4	1.520	0	4.2	0.3	27.964	31.404	40.267	37.000	0.041	0.038	0.991	0.988
(0.63, 0.48)	0	2.331	0.4	1.520	0	4.2	0.3	27.394	30.497	40.174	36.980	0.039	0.038	0.992	0.989
(0.28, 0.33)	0	2.331	0.6	1.969	0.95	4.5	0	39.384	41.504	24.512	22.268	0.026	0.030	0.995	0.992
(0.585, 0.52)	2	3.774	0.6	3.322	0.95	4.2	0.4	33.326	32.356	16.199	14.988	0.034	0.055	0.989	0.971
(0.288, 1.05)	2	2.331	0.8	2.421	0.95	0	0	56.452	56.908	17.744	15.924	0.019	0.042	0.997	0.990
(0.0, 0.0)	0	1.250	0.6	1.969	0	0	0	55.227	58.493	26.519	25.931	0.028	0.028	0.995	0.992
(0.75, 2.8)	0	2.331	0	1.520	0.95	0	0.4	77.146	78.668	41.275	37.648	0.021	0.032	0.996	0.991
(0.62, 1.6)	1	0.000	0	1.969	0	4.5	0.3	82.399	81.630	26.385	28.589	0.019	0.032	0.997	0.990
(0.8, 3.6)	1	1.610	0	0.000	0.95	0	0.4	139.464	143.967	41.245	46.555	0.025	0.030	0.994	0.993
(0.585, 0.52)	0	3.413	0.6	2.421	0.95	0	0.4	34.206	34.793	32.762	32.873	0.027	0.034	0.993	0.991

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.44, 1.6)	2	0.000	0.4	1.969	0.95	0	0.4	48.818	51.601	19.511	16.752	0.028	0.055	0.995	0.973
(0.8, 3.6)	1	1.610	0.4	2.874	0.95	4.5	0	42.393	46.435	27.164	32.142	0.040	0.057	0.990	0.987
(0.25, 0.26)	0	2.688	0.8	2.874	0.95	4.5	0	48.826	48.791	20.837	20.117	0.020	0.024	0.996	0.994
(0.25, 0.26)	1	3.413	0.4	1.969	0	0	0	26.278	28.621	23.182	19.091	0.032	0.047	0.994	0.989
(0.0, 0.0)	0	0.000	0.8	2.421	0	0	0.3	78.649	79.001	24.034	25.326	0.020	0.030	0.996	0.991
(0.63, 0.48)	1	1.969	0.4	2.874	0	0	0.3	24.289	26.002	26.436	28.704	0.045	0.047	0.985	0.988
(0.288, 1.05)	0	1.250	0.4	3.774	0.95	4.2	0	31.319	34.970	22.090	21.533	0.042	0.038	0.992	0.988
(0.63, 0.48)	2	1.250	0.8	3.774	0	4.2	0.4	46.920	47.170	17.230	16.022	0.032	0.051	0.991	0.971
(0.8, 3.6)	1	1.250	0	2.421	0	4.2	0	79.656	81.607	27.865	31.992	0.023	0.042	0.995	0.991
(0.62, 1.6)	2	2.688	0.8	3.774	0.95	0	0.3	53.368	52.481	16.059	15.048	0.026	0.046	0.994	0.976
(0.44, 1.6)	2	2.331	0.8	2.421	0.95	4.5	0	58.261	57.978	18.624	15.925	0.022	0.044	0.995	0.987
(0.8, 3.6)	0	1.969	0.4	3.322	0	0	0	39.791	44.176	42.808	40.190	0.039	0.042	0.991	0.985
(0.288, 1.05)	1	2.688	0.6	1.520	0.95	0	0.4	48.501	51.101	23.628	19.287	0.024	0.043	0.996	0.991
(0.44, 1.6)	2	1.610	0.6	1.969	0	0	0.3	50.574	52.886	19.072	17.336	0.028	0.052	0.994	0.985
(0.25, 0.26)	0	3.774	0.4	2.874	0	4.5	0.4	24.690	26.552	20.124	19.259	0.032	0.028	0.994	0.992
(0.0, 0.0)	2	1.969	0	0.000	0	4.2	0.3	139.586	142.015	32.750	32.552	0.023	0.016	0.995	0.995
(0.8, 3.6)	2	3.774	0.6	0.000	0.95	4.5	0	110.733	115.705	33.750	34.684	0.029	0.020	0.993	0.992
(0.44, 1.6)	0	3.774	0.4	1.520	0	4.2	0.3	37.067	39.655	28.354	26.901	0.029	0.028	0.995	0.992
(0.585, 0.52)	2	2.688	0	0.000	0.95	0	0.3	122.089	123.184	33.136	33.948	0.024	0.018	0.995	0.993
(0.28, 0.33)	0	0.000	0.8	1.969	0.95	4.5	0	64.386	65.289	24.656	23.665	0.021	0.028	0.996	0.991
(0.28, 0.33)	2	2.331	0.8	2.874	0.95	4.2	0.4	51.196	51.357	15.804	13.550	0.021	0.040	0.996	0.989
(0.288, 1.05)	2	1.610	0.6	3.774	0	0	0	42.870	44.658	16.234	14.116	0.028	0.041	0.994	0.988
(0.0, 0.0)	1	1.610	0.4	3.774	0	0	0.3	40.164	43.185	18.949	19.854	0.036	0.039	0.993	0.987
(0.25, 0.26)	1	1.610	0	2.874	0	4.2	0.3	65.605	65.480	19.779	17.380	0.020	0.030	0.996	0.992

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.585, 0.52)	1	1.969	0.6	0.000	0.95	4.2	0.3	97.801	102.278	39.343	42.824	0.029	0.027	0.993	0.993
(0.75, 2.8)	1	3.774	0.4	3.774	0.95	4.2	0.4	35.111	36.515	23.931	29.432	0.036	0.062	0.990	0.988
(0.25, 0.26)	0	2.688	0.8	1.520	0	4.5	0.3	55.526	56.926	23.496	20.452	0.020	0.033	0.997	0.992
(0.75, 2.8)	2	1.250	0.6	1.520	0.95	4.5	0.3	56.644	60.915	22.261	18.380	0.033	0.054	0.993	0.977
(0.25, 0.26)	1	1.969	0.8	2.874	0	4.2	0.4	51.797	52.484	19.700	17.229	0.021	0.032	0.996	0.992
(0.8, 3.6)	2	1.969	0.8	2.874	0	4.2	0.4	66.879	68.793	18.327	17.097	0.026	0.046	0.994	0.974
(0.75, 2.8)	0	0.000	0	3.322	0.95	4.5	0	82.552	82.336	38.036	38.237	0.019	0.025	0.997	0.993
(0.28, 0.33)	2	2.331	0	2.421	0	4.5	0.3	65.219	64.479	16.668	14.068	0.019	0.040	0.997	0.990
(0.585, 0.52)	2	1.610	0.6	3.774	0.95	4.5	0.4	35.555	36.277	16.749	15.391	0.036	0.054	0.989	0.970
(0.44, 1.6)	2	3.413	0.4	2.874	0	4.2	0	31.572	32.694	17.451	15.019	0.031	0.053	0.993	0.982
(0.63, 0.48)	1	3.413	0.6	0.000	0.95	4.5	0	93.192	96.461	40.950	44.945	0.028	0.027	0.993	0.993
(0.0, 0.0)	0	3.774	0.4	2.421	0.95	0	0.4	39.047	41.157	23.365	23.833	0.027	0.029	0.995	0.993
(0.63, 0.48)	1	1.610	0	2.421	0	4.5	0.4	60.116	59.497	26.167	27.862	0.025	0.033	0.994	0.992
(0.8, 3.6)	0	2.688	0	3.322	0.95	4.2	0	74.134	74.794	39.835	38.268	0.021	0.032	0.996	0.990
(0.28, 0.33)	2	3.049	0.4	0.000	0.95	4.2	0.3	87.593	93.311	31.016	30.948	0.031	0.017	0.993	0.994
(0.0, 0.0)	1	3.774	0.6	1.520	0.95	0	0.3	55.572	57.162	22.894	21.313	0.021	0.024	0.997	0.994
(0.28, 0.33)	2	1.250	0.8	2.421	0	4.2	0.4	54.587	56.367	17.468	15.447	0.025	0.044	0.995	0.988
(0.63, 0.48)	0	0.000	0.4	2.421	0	4.2	0	33.583	35.931	37.620	38.056	0.031	0.035	0.993	0.990
(0.62, 1.6)	2	2.688	0	3.322	0.95	4.2	0	64.256	62.185	17.841	16.684	0.024	0.042	0.995	0.981
(0.28, 0.33)	2	1.250	0.8	1.969	0	4.2	0.4	56.063	58.244	18.654	16.965	0.025	0.047	0.995	0.988
(0.585, 0.52)	1	3.413	0.8	2.421	0.95	0	0.4	46.659	45.475	24.399	26.242	0.026	0.032	0.994	0.993
(0.25, 0.26)	2	1.969	0.4	3.774	0.95	4.2	0.4	26.010	28.465	14.828	12.714	0.041	0.047	0.992	0.986
(0.288, 1.05)	2	2.688	0.6	2.421	0.95	0	0.4	45.052	46.577	16.523	14.559	0.023	0.047	0.996	0.989
(0.44, 1.6)	1	1.969	0.6	3.774	0	4.5	0.3	45.700	46.678	19.626	21.002	0.028	0.037	0.994	0.989

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.75, 2.8)	1	3.049	0.8	1.520	0	4.2	0.3	65.171	66.595	28.630	30.537	0.023	0.026	0.995	0.994
(0.288, 1.05)	1	2.688	0.8	2.874	0.95	4.5	0.4	56.394	56.535	19.572	17.736	0.019	0.027	0.997	0.994
(0.44, 1.6)	1	3.774	0	0.000	0	4.5	0.4	131.718	132.235	36.331	38.755	0.022	0.021	0.996	0.995
(0.585, 0.52)	2	3.049	0	1.969	0.95	0	0.3	61.330	59.403	19.158	18.529	0.023	0.049	0.995	0.982
(0.288, 1.05)	1	3.413	0.4	3.322	0.95	4.5	0.3	30.164	32.093	18.852	17.141	0.030	0.031	0.995	0.992
(0.75, 2.8)	0	3.049	0.8	2.874	0.95	4.5	0.3	59.419	60.536	37.645	36.219	0.023	0.033	0.995	0.990
(0.62, 1.6)	1	3.774	0.4	1.969	0	4.2	0.4	32.362	33.895	26.037	27.710	0.033	0.031	0.992	0.993
(0.75, 2.8)	1	3.049	0.6	1.520	0	0	0	51.542	53.740	30.059	32.029	0.028	0.028	0.994	0.994
(0.62, 1.6)	0	3.049	0.4	3.322	0	4.5	0	27.442	29.315	34.477	34.941	0.035	0.040	0.992	0.989

Appendix II Heating and cooling load simulation accuracy of the SimBldPy model for the office building (unit of load: GJ)

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load C sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.585, 0.52)	1	1.610	0.8	1.520	0.95	0	0	228.30	227.01	789.52	755.43	0.019	0.032	0.986	0.992
(0.25, 0.26)	2	2.688	0.8	3.322	0.95	4.5	0.4	334.06	356.49	404.93	388.36	0.022	0.026	0.992	0.993
(0.25, 0.26)	0	3.049	0.3	3.322	0	4.5	0	37.19	46.94	875.99	816.25	0.023	0.037	0.973	0.991
(0.585, 0.52)	0	2.331	0.8	2.421	0.95	4.2	0	195.49	211.51	931.15	848.93	0.020	0.037	0.986	0.989
(0.44, 1.6)	0	3.774	0	1.520	0.95	4.5	0	419.51	424.80	695.96	670.33	0.016	0.029	0.992	0.992
(0.0, 0.0)	2	1.610	0.6	0.000	0.95	4.2	0.4	444.44	446.42	423.95	407.87	0.015	0.026	0.994	0.993
(0.28, 0.33)	1	3.049	0.3	1.969	0	4.2	0.3	86.48	101.86	631.59	530.92	0.028	0.043	0.981	0.989
(0.63, 0.48)	2	3.413	0.3	3.774	0	4.2	0.3	78.31	68.39	521.46	529.13	0.029	0.044	0.952	0.982
(0.62, 1.6)	1	1.969	0.3	2.874	0.95	0	0	86.69	85.19	906.59	877.57	0.018	0.034	0.978	0.991
(0.28, 0.33)	2	0.000	0.3	0.000	0	4.5	0.4	172.50	181.45	461.03	476.21	0.026	0.030	0.982	0.992
(0.62, 1.6)	2	1.250	0	1.969	0	0	0.4	505.58	511.13	434.51	411.94	0.016	0.033	0.993	0.988
(0.28, 0.33)	1	3.049	0.3	3.774	0	4.5	0	40.14	47.52	887.42	785.97	0.021	0.039	0.974	0.991
(0.63, 0.48)	1	1.969	0.3	3.322	0.95	4.5	0.4	64.60	67.16	793.87	753.02	0.022	0.038	0.975	0.988
(0.0, 0.0)	1	2.688	0.6	1.520	0	4.2	0.4	401.91	417.86	532.50	499.23	0.016	0.025	0.994	0.994
(0.63, 0.48)	2	2.331	0	3.774	0.95	4.5	0	312.91	294.88	572.35	578.32	0.020	0.035	0.986	0.988
(0.75, 2.8)	2	1.969	0.3	3.322	0	4.5	0.4	194.11	192.81	509.53	487.43	0.022	0.040	0.983	0.985
(0.288, 1.05)	1	2.688	0.8	3.774	0.95	4.5	0	300.63	317.43	641.81	593.86	0.017	0.026	0.992	0.994
(0.288, 1.05)	2	2.688	0.8	3.774	0	0	0.3	413.52	436.33	376.82	357.28	0.020	0.025	0.993	0.993
(0.75, 2.8)	1	0.000	0.6	2.421	0	4.2	0.3	414.66	430.51	583.31	579.07	0.016	0.026	0.993	0.993
(0.28, 0.33)	0	1.610	0.6	2.421	0	0	0	153.37	178.65	759.45	699.22	0.024	0.033	0.985	0.992
(0.44, 1.6)	1	3.049	0.8	2.874	0	4.2	0	323.19	322.55	656.65	643.53	0.016	0.028	0.991	0.993
(0.0, 0.0)	2	1.610	0.6	3.774	0.95	4.2	0.3	421.64	445.30	412.83	378.20	0.018	0.031	0.993	0.991

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load C sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.44, 1.6)	1	0.000	0.8	1.520	0	0	0.3	503.50	529.40	484.79	470.19	0.017	0.025	0.994	0.993
(0.288, 1.05)	1	2.688	0.3	2.421	0	4.5	0.3	135.04	153.78	574.74	496.86	0.026	0.037	0.986	0.991
(0.8, 3.6)	2	2.688	0	2.874	0.95	4.5	0.3	624.91	635.75	399.89	384.87	0.015	0.032	0.994	0.988
(0.25, 0.26)	1	2.688	0.6	3.774	0.95	0	0.3	227.17	257.91	502.48	442.48	0.027	0.032	0.990	0.992
(0.75, 2.8)	1	1.250	0	1.969	0.95	4.5	0.3	582.14	599.09	550.91	536.44	0.014	0.022	0.995	0.995
(0.585, 0.52)	0	2.331	0.6	0.000	0.95	0	0	138.53	141.53	970.17	916.70	0.017	0.035	0.985	0.991
(0.8, 3.6)	0	3.413	0.3	0.000	0.95	4.2	0	166.44	172.04	1175.28	993.51	0.015	0.050	0.988	0.987
(0.0, 0.0)	1	0.000	0.6	1.520	0	0	0.3	490.66	518.94	503.25	480.95	0.017	0.024	0.994	0.994
(0.585, 0.52)	2	0.000	0.3	2.421	0.95	4.2	0.3	140.55	142.09	500.20	499.06	0.029	0.040	0.972	0.985
(0.63, 0.48)	1	3.413	0.6	1.969	0	4.2	0.4	179.41	177.83	676.12	650.11	0.021	0.033	0.986	0.991
(0.28, 0.33)	0	2.688	0	3.774	0	4.5	0.4	427.78	466.24	484.31	461.20	0.020	0.025	0.993	0.994
(0.8, 3.6)	1	1.969	0	1.969	0.95	4.5	0	521.82	520.07	711.84	704.71	0.014	0.023	0.993	0.995
(0.288, 1.05)	1	2.688	0.8	2.421	0	4.2	0.4	390.50	420.20	490.26	438.09	0.019	0.027	0.993	0.994
(0.28, 0.33)	1	2.331	0.6	0.000	0.95	4.5	0	180.12	182.90	712.01	664.31	0.018	0.027	0.988	0.994
(0.28, 0.33)	2	3.774	0	1.969	0.95	4.2	0.3	462.49	482.71	374.30	353.70	0.019	0.026	0.994	0.993
(0.288, 1.05)	0	0.000	0	1.969	0.95	4.2	0	478.38	509.92	623.29	606.11	0.017	0.026	0.992	0.994
(0.75, 2.8)	1	1.969	0.3	1.520	0.95	4.2	0.3	203.35	205.50	676.43	639.06	0.018	0.030	0.989	0.992
(0.62, 1.6)	2	0.000	0.3	3.322	0.95	4.5	0.4	192.05	193.53	498.64	495.25	0.024	0.037	0.981	0.987
(0.63, 0.48)	2	1.250	0.8	3.322	0.95	0	0.4	300.82	301.59	456.72	450.33	0.022	0.036	0.987	0.987
(0.44, 1.6)	2	1.610	0.6	2.421	0	4.2	0	242.01	243.70	613.02	577.55	0.019	0.033	0.987	0.990
(0.288, 1.05)	1	1.250	0	1.520	0	0	0.3	543.37	586.39	464.14	403.52	0.018	0.029	0.994	0.994
(0.25, 0.26)	2	1.610	0.3	1.969	0.95	0	0.3	104.38	119.79	494.88	449.56	0.030	0.036	0.978	0.990
(0.25, 0.26)	0	2.331	0.3	1.520	0.95	4.2	0	43.75	55.85	888.68	808.52	0.023	0.039	0.972	0.990

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load C sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.25, 0.26)	0	2.331	0.8	1.520	0	4.5	0.4	325.77	365.88	513.84	472.15	0.024	0.028	0.991	0.993
(0.585, 0.52)	2	1.969	0	3.322	0.95	0	0.3	423.97	422.56	401.99	398.53	0.019	0.031	0.992	0.989
(0.0, 0.0)	1	3.049	0.3	3.774	0.95	4.5	0	166.52	168.79	761.75	733.18	0.017	0.030	0.987	0.993
(0.62, 1.6)	1	1.250	0.8	2.874	0.95	0	0.3	396.78	412.48	547.96	534.71	0.017	0.027	0.994	0.993
(0.63, 0.48)	0	1.969	0	2.874	0.95	4.5	0.4	344.07	384.29	745.88	661.94	0.021	0.037	0.992	0.989
(0.8, 3.6)	2	3.774	0.3	0.000	0.95	0	0.3	269.64	254.50	446.99	453.63	0.020	0.036	0.987	0.986
(0.0, 0.0)	2	3.413	0.6	1.969	0	4.5	0.3	414.25	428.70	416.78	378.33	0.017	0.032	0.994	0.991
(0.63, 0.48)	1	1.969	0.6	2.874	0.95	4.5	0	130.37	127.58	872.91	840.04	0.020	0.034	0.980	0.991
(0.75, 2.8)	2	3.049	0.8	0.000	0	0	0.3	494.76	475.99	397.65	406.19	0.016	0.031	0.993	0.988
(0.44, 1.6)	1	3.774	0	1.520	0	0	0.4	525.26	531.94	484.27	467.62	0.015	0.023	0.995	0.994
(0.62, 1.6)	2	3.774	0.6	2.874	0.95	4.5	0.4	272.22	261.63	455.29	447.46	0.021	0.035	0.986	0.987
(0.8, 3.6)	0	3.413	0	2.874	0	0	0	469.30	499.22	981.04	802.45	0.016	0.045	0.991	0.989
(0.62, 1.6)	1	1.969	0	3.774	0	0	0.4	482.03	488.83	533.19	535.52	0.015	0.026	0.994	0.994
(0.25, 0.26)	2	3.049	0.8	0.000	0.95	4.5	0.3	384.50	387.94	366.49	371.41	0.020	0.022	0.992	0.994
(0.62, 1.6)	0	3.774	0.3	2.421	0	4.5	0.3	108.31	123.91	867.59	793.39	0.023	0.047	0.985	0.986
(0.288, 1.05)	2	3.413	0.8	1.969	0	0	0.4	397.65	415.87	409.99	381.45	0.019	0.027	0.993	0.993
(0.288, 1.05)	1	3.413	0.3	0.000	0	4.2	0.4	151.17	152.29	565.77	531.39	0.020	0.028	0.986	0.993
(0.0, 0.0)	2	3.774	0.6	3.322	0.95	0	0.3	408.74	421.33	404.39	376.44	0.017	0.029	0.993	0.992
(0.585, 0.52)	0	3.049	0.8	1.520	0	4.2	0	195.07	208.06	938.79	851.67	0.020	0.037	0.986	0.990
(0.8, 3.6)	1	3.774	0.8	2.421	0	4.5	0.3	502.00	508.39	552.80	550.32	0.014	0.023	0.994	0.995
(0.8, 3.6)	0	2.331	0	0.000	0.95	4.2	0.3	608.69	646.04	780.39	631.60	0.016	0.044	0.993	0.989
(0.62, 1.6)	0	3.774	0.6	1.969	0.95	0	0.3	245.45	272.82	754.28	654.00	0.021	0.040	0.990	0.988
(0.585, 0.52)	1	1.610	0.6	0.000	0	4.5	0	165.78	152.87	806.80	812.60	0.019	0.033	0.983	0.992
(0.0, 0.0)	0	2.688	0	3.774	0	0	0.4	620.45	646.27	552.10	512.90	0.015	0.026	0.995	0.993

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load C sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.288, 1.05)	0	2.331	0.8	2.421	0	4.2	0.4	387.99	424.02	505.85	470.34	0.020	0.027	0.993	0.993
(0.585, 0.52)	0	3.049	0	1.969	0.95	4.2	0	283.00	297.62	871.42	802.86	0.018	0.035	0.989	0.990
(0.0, 0.0)	0	1.250	0.3	1.520	0	4.2	0	177.17	195.91	900.38	796.73	0.019	0.038	0.986	0.990
(0.75, 2.8)	0	3.413	0.3	1.520	0	0	0	112.57	123.24	1218.88	1160.17	0.018	0.053	0.985	0.987
(0.28, 0.33)	2	1.250	0.8	3.774	0	4.5	0.3	361.00	395.39	393.78	370.76	0.023	0.027	0.992	0.993
(0.75, 2.8)	1	3.774	0.3	2.421	0	4.2	0	131.25	122.29	892.07	873.10	0.017	0.031	0.982	0.993
(0.585, 0.52)	0	1.250	0.3	3.322	0	0	0.3	63.41	86.99	915.03	887.88	0.031	0.049	0.979	0.984
(0.75, 2.8)	0	1.250	0	2.874	0.95	0	0.3	535.91	593.79	771.78	623.65	0.019	0.045	0.992	0.988
(0.8, 3.6)	1	1.610	0	3.322	0	0	0.4	611.59	626.31	558.28	556.74	0.014	0.022	0.995	0.995
(0.44, 1.6)	1	3.049	0.6	2.421	0.95	0	0.4	297.59	306.64	533.41	511.84	0.018	0.028	0.992	0.993
(0.75, 2.8)	2	3.413	0.8	3.322	0	4.2	0.4	442.00	439.48	429.76	421.90	0.017	0.033	0.992	0.987
(0.44, 1.6)	0	2.331	0.6	2.874	0	4.2	0	216.51	228.34	781.73	744.14	0.018	0.033	0.988	0.991
(0.62, 1.6)	1	1.610	0.6	0.000	0.95	4.5	0.4	305.52	296.51	584.05	595.54	0.017	0.029	0.991	0.993
(0.28, 0.33)	0	0.000	0.8	0.000	0	4.5	0.4	419.37	453.10	493.74	495.78	0.020	0.023	0.993	0.995
(0.25, 0.26)	2	2.331	0.8	1.969	0.95	0	0.3	356.04	383.54	394.76	377.10	0.023	0.028	0.992	0.992
(0.62, 1.6)	2	1.969	0	3.322	0.95	4.2	0.4	494.58	493.00	417.59	408.83	0.016	0.031	0.993	0.989
(0.63, 0.48)	1	0.000	0.3	2.874	0.95	0	0.3	115.64	127.73	713.46	686.59	0.025	0.039	0.981	0.988
(0.8, 3.6)	0	1.250	0.6	2.421	0.95	4.5	0.4	351.67	407.12	920.72	801.23	0.022	0.057	0.989	0.985
(0.25, 0.26)	2	3.049	0.6	0.000	0	4.2	0	186.19	178.65	578.85	589.93	0.020	0.029	0.985	0.993
(0.63, 0.48)	0	2.688	0.6	2.421	0	4.5	0	105.52	118.86	1081.95	956.33	0.021	0.043	0.982	0.988
(0.63, 0.48)	2	3.413	0	2.874	0.95	4.2	0.4	387.87	375.84	424.76	428.76	0.020	0.034	0.990	0.987
(0.44, 1.6)	0	3.413	0	2.874	0.95	4.5	0	416.68	423.04	683.62	668.47	0.016	0.029	0.993	0.992
(0.44, 1.6)	0	2.688	0	1.969	0.95	0	0.3	533.62	558.58	519.65	493.72	0.016	0.027	0.994	0.992

(Window_SHGC & Window_U)	Shading position	Wall_Insulation	Air Insulation	Roof Insulation	Heating Efficiency	Cooling COP	Lighting Load Reduction	Heating Load EP	Heating Load sim	Cooling Load EP	Cooling Load C sim	Heating Load NRMSE	Cooling Load NRMSE	Heating Load R2	Cooling Load R2
(0.8, 3.6)	2	0.000	0.8	3.774	0	0	0.3	581.63	601.10	408.23	404.34	0.017	0.032	0.993	0.987
(0.44, 1.6)	2	3.774	0.6	1.520	0	4.5	0	233.28	225.65	613.38	580.07	0.018	0.034	0.987	0.990
(0.0, 0.0)	1	1.610	0.3	2.874	0	4.2	0.4	239.34	255.72	585.73	545.15	0.019	0.029	0.991	0.992
(0.25, 0.26)	0	2.331	0.8	3.322	0	4.2	0	239.94	264.42	664.54	635.44	0.021	0.029	0.989	0.993
(0.288, 1.05)	0	0.000	0.8	3.322	0.95	4.5	0	363.12	397.00	643.15	629.46	0.019	0.027	0.991	0.994
(0.28, 0.33)	0	1.610	0.6	1.520	0.95	0	0.4	221.05	262.51	582.33	518.92	0.028	0.033	0.988	0.991
(0.585, 0.52)	2	1.250	0	2.421	0.95	4.2	0.4	418.02	421.23	437.15	423.74	0.019	0.032	0.991	0.988
(0.75, 2.8)	0	1.969	0.8	3.774	0	4.2	0.4	399.88	446.27	825.86	771.12	0.020	0.047	0.992	0.987

BIBLIOGRAPHIES

(EIA), Energy Information Administration. 2009. Commercial Sector Demand Module of the National Energy Modeling System: Model Documentation 2009. Washington D.C.: U.S. Department of Energy.

(EIA), Energy Information Administration. 2011. Assumptions for the Annual Energy Outlook 2011. Washington, DC: Department of Energy (DOE).

Abbass, H. A., R. Sarker, and C. Newton. 2001. PDE: a Pareto-frontier differential evolution approach for multi-objective optimization problems. Paper read at Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546), 2001.

Agrawal, Rakesh, Johannes Gehrke, Dimitrios Gunopulos, and Prabhakar Raghavan. 1998. *Automatic subspace clustering of high dimensional data for data mining applications.* Vol. 27: ACM.

Aktas, Can B, and M Bilec. 2011. "Impact of lifetime on US residential building LCA results." *The International Journal of Life Cycle Assessment* no. 17 (3):337-349.

Asadi, Ehsan, Manuel Gameiro da Silva, Carlos Henggeler Antunes, and Luís Dias. 2012a. "Multi-objective optimization for building retrofit strategies: A model and an application." *Energy and Buildings* no. 44 (0):81-87. doi: http://dx.doi.org/10.1016/j.enbuild.2011.10.016.

Asadi, Ehsan, Manuel Gameiro da Silva, Carlos Henggeler Antunes, and Luís Dias. 2012b. "A multi-objective optimization model for building retrofit strategies using TRNSYS simulations, GenOpt and MATLAB." *Building and Environment* no. 56:370-378. doi: http://dx.doi.org/10.1016/j.buildenv.2012.04.005.

Asadi, Ehsan, Manuel Gameiro da Silva, Carlos Henggeler Antunes, Luís Dias, and Leon Glicksman. 2014. "Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application." *Energy and Buildings* no. 81:444-456. doi: http://dx.doi.org/10.1016/j.enbuild.2014.06.009.

Belcher, S.E, J.N Hacker, and D.S Powell. 2005. "Constructing design weather data for future climates." *BUILDING SERV ENG RES TECHNOL* no. 26 (1):49-61. doi: 10.1191/0143624405bt112oa.

Bennasar, Mohamed, Yulia Hicks, and Rossitza Setchi. 2015. "Feature selection using Joint Mutual Information Maximisation." *Expert Systems with Applications* no. 42 (22):8520-8532. doi: http://dx.doi.org/10.1016/j.eswa.2015.07.007.

Berthoua, Thomas, Pascal Stabata, Raphael Salvazetb, and Dominique Marchioa. 2014. "Development and validation of a gray box model to predict thermal behavior of occupied office buildings." *Energy and Buildings* no. 74:91-100.

Blashfield, Roger K. 1976. "Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods." *Psychological Bulletin* no. 83 (3):377.

Burhenne, S., Jacob, D. 2008. Simulation models to optimize the energy consumption of buildings. In *The International Conference for Enhanced Building Operations 2008*. Berlin, Germany.

Chan, A. L. S. 2011. "Developing future hourly weather files for studying the impact of climate change on building energy performance in Hong Kong." *Energy and Buildings* no. 43 (10):2860-2868. doi: 10.1016/j.enbuild.2011.07.003.

Chantrelle, Fanny Pernodet, Hicham Lahmidi, Werner Keilholz, Mohamed El Mankibi, and Pierre Michel. 2011. "Development of a multicriteria tool for optimizing the renovation of buildings." *Applied Energy* no. 88 (4):1386-1394. doi: http://dx.doi.org/10.1016/j.apenergy.2010.10.002.

Chidiac, S. E., E. J. C. Catania, E. Morofsky, and S. Foo. 2011a. "Effectiveness of single and multiple energy retrofit measures on the energy consumption of office buildings." *Energy* no. 36 (8):5037-5052. doi: http://dx.doi.org/10.1016/j.energy.2011.05.050.

Chidiac, S. E., E. J. C. Catania, E. Morofsky, and S. Foo. 2011b. "A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings." *Energy and Buildings* no. 43 (2):614-620. doi: http://dx.doi.org/10.1016/j.enbuild.2010.11.002.

Chow, D. H. C., Zhilei Li, and J. Darkwa. 2013. "The effectiveness of retrofitting existing public buildings in face of future climate change in the hot summer cold winter region of

China." *Energy and Buildings* no. 57:176-186. doi: http://dx.doi.org/10.1016/j.enbuild.2012.11.012.

Chuah, Jun Wei, Anand Raghunathan, and Niraj K. Jha. 2013. "ROBESim: A retrofit-oriented building energy simulator based on EnergyPlus." *Energy and Buildings* no. 66:88-103. doi: http://dx.doi.org/10.1016/j.enbuild.2013.07.020.

Deb, Kalyanmoy, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. 2002. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation* no. 6 (2):182-197.

Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., Liu, B., Halverson, M., Winiarski, D., Rosenberg, M., Yazdanian, M., Huang, J., Crawley, D. 2011. U.S. Department of Energy commercial reference building models of the national building stock. http://digitalscholarship.unlv.edu/renew_pubs/44: DOE.

Dick van Dijk, Marleen Spiekman, Linda Hoes. 2016. "EPB standard EN ISO 52016: Calculation of the building's energy needs for heating and cooling, internal temperatures and heating and cooling load." *REHVA* (03/2016).

The Encyclopedic Reference to EnergyPlus Input and Output 8.0. EnergyPlus Development Team.

EIACBECS. 2003. U.S. Energy Information Administration (EIA): Commercial BuildingsEnergy Consumption Survey. http://www.eia.gov/consumption/commercial/index.cfm.

Eisenhower, Bryan, Zheng O'Neill, Satish Narayanan, Vladimir A. Fonoberov, and Igor Mezić. 2012. "A methodology for meta-model based optimization in building energy models." *Energy and Buildings* no. 47:292-301. doi: http://dx.doi.org/10.1016/j.enbuild.2011.12.001.

Enríquez, R., M. J. Jiménez, and M. R. Heras. 2017. "Towards non-intrusive thermal load Monitoring of buildings: BES calibration." *Applied Energy* no. 191:44-54. doi: http://dx.doi.org/10.1016/j.apenergy.2017.01.050.

Gaterell, M. R., and M. E. McEvoy. 2005. "The impact of climate change uncertainties on the performance of energy efficiency measures applied to dwellings." *Energy and Buildings* no. 37 (9):982-995. doi: http://dx.doi.org/10.1016/j.enbuild.2004.12.015.

Giuliano, Genevieve, and Kenneth A Small. 1991. "Subcenters in the Los Angeles region." *Regional science and urban economics* no. 21 (2):163-182.

Goel S, et al. 2014. Enhancements to ASHRAE standard 90.1 prototype building models. https://www.energycodes.gov/development/commercial/prototype_models: Pacific Northwest National Laboratory.

Guha, Sudipto, Rajeev Rastogi, and Kyuseok Shim. 1998. CURE: an efficient clustering algorithm for large databases. Paper read at ACM Sigmod Record.

Halkidi, Maria, Yannis Batistakis, and Michalis Vazirgiannis. 2001. "On clustering validation techniques." *Journal of intelligent information systems* no. 17 (2):107-145.

Hamdy, Mohamed, Ala Hasan, and Kai Siren. 2013. "A multi-stage optimization method for cost-optimal and nearly-zero-energy building solutions in line with the EPBD-recast 2010." *Energy and Buildings* no. 56:189-203. doi: http://dx.doi.org/10.1016/j.enbuild.2012.08.023.

Hansen, Pierre, and Brigitte Jaumard. 1997. "Cluster analysis and mathematical programming." *Mathematical Programming* no. 79 (1):191-215. doi: 10.1007/bf02614317.

Hasan, Kai Sirén and Ala. 2007. Comparison of two calculation methods used to estimate cooling energy demand and indoor summer temperatures. Paper read at Proceedings of Clima 2007 WellBeing Indoors.

Heo, Y., R. Choudhary, and G. A. Augenbroe. 2012. "Calibration of building energy models for retrofit analysis under uncertainty." *Energy and Buildings* no. 47:550-560. doi: http://dx.doi.org/10.1016/j.enbuild.2011.12.029.

Hillary, Jason, Ed Walsh, Amip Shah, Rongliang Zhou, and Pat Walsh. 2017. "Guidelines for developing efficient thermal conduction and storage models within building energy simulations." *Energy* no. 125:211-222. doi: http://dx.doi.org/10.1016/j.energy.2017.02.127.

Hillebrand, G., G. Arends, R. Streblow, R. Madlener, and D. Müller. 2014. "Development and design of a retrofit matrix for office buildings." *Energy and Buildings* no. 70 (0):516-522. doi: http://dx.doi.org/10.1016/j.enbuild.2013.10.029.

Hirsch, James J. 2016. eQuest. edited by Lawrence Berkeley National Laboratory. U.S.

IPCC. 2000. IPCC special report on emissions scenarios (SRES): summary for policymakers-- a special report of IPCC Working Group III Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC.

IPCC. 2007. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007 http://www.ipcc.ch/pdf/assessment-report/ar4/wg3/ar4-wg3-chapter1.pdf.

IPCC. 2013. Climate change 2013-- the physical science basis, fifth assessment report of the IPCC. edited by Thomas F.Stocker and Dahe Qin. New York.

ISO. 2008. ISO 13790: Energy performance of buildings - Calculation of energy use for space heating and cooling.

ISO. 2017. ISO 52061-1: Energy performance of buildings — Energy needs for heating and cooling, internal temperatures and sensible and latent heat loads - Part I. Geneva, Switzerland: ISO.

J.D. Ryan, A. Nicholls. 2004. Commercial building R&D program multi-year planning: opportunities and challenges. Paper read at ACEEE Summer Study on Energy Efficiency in Buildings.

Jafari, Amirhosein, and Vanessa Valentin. 2017. "An optimization framework for building energy retrofits decision-making." *Building and Environment* no. 115:118-129. doi: http://dx.doi.org/10.1016/j.buildenv.2017.01.020.

Jaggs, Michael, and John Palmer. 2000. "Energy performance indoor environmental quality retrofit — a European diagnosis and decision making method for building refurbishment." *Energy and Buildings* no. 31 (2):97-101. doi: https://doi.org/10.1016/S0378-7788(99)00023-7.

Ji, Ying, Peng Xu, Pengfei Duan, and Xing Lu. 2016. "Estimating hourly cooling load in commercial buildings using a thermal network model and electricity submetering data." *Applied Energy* no. 169:309-323. doi: http://dx.doi.org/10.1016/j.apenergy.2016.02.036.

Johnson, Stephen C. 1967. "Hierarchical clustering schemes." *Psychometrika* no. 32 (3):241-254.

Jokisalo, Juha, and Jarek Kurnitski. 2007. "Performance of EN ISO 13790 utilisation factor heat demand calculation method in a cold climate." *Energy and Buildings* no. 39 (2):236-247. doi: http://dx.doi.org/10.1016/j.enbuild.2006.06.007.

Kalema, Timo, Gudni Jóhannesson, Petri Pylsy, and Per Hagengran. 2008. "Accuracy of Energy Analysis of Buildings: A Comparison of a Monthly Energy Balance Method and Simulation Methods in Calculating the Energy Consumption and the Effect of Thermal

Mass." *Journal of Building Physics* no. 32 (2):101-130. doi: 10.1177/1744259108093920.

Kamrani, Ali K, Hamid R Parsaei, and Mahfooz A Chaudhry. 1993. "A survey of design methods for manufacturing cells." *Computers & industrial engineering* no. 25 (1-4):487-490.

TRNSYS 17: A Transient System Simulation Program. Solar Energy Laboratory, University of Wisconsin, Madison, USA.

Kokogiannakis, G., Clarke, J. & Strachan, P. 2007. "Impact of using different models in practice-a case study with the simplified methods of ISO 13790 standard and detailed modelling programs." *International Building Performance Simulation Association (IBPSA)*:39-46.

Kokogiannakis, Georgios, Paul Strachan, and Joe Clarke. 2008. "Comparison of the simplified methods of the ISO 13790 standard and detailed modelling programs in a regulatory context." *Journal of Building Performance Simulation* no. 1 (4):209-219. doi: 10.1080/19401490802509388.

Kriegel, Hans-Peter, Peer Kröger, and Arthur Zimek. 2009. "Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering." *ACM Transactions on Knowledge Discovery from Data (TKDD)* no. 3 (1):1.

Kwak, Younghoon, and Jung-Ho Huh. 2016. "Development of a method of real-time building energy simulation for efficient predictive control." *Energy Conversion and Management* no. 113:220-229. doi: http://dx.doi.org/10.1016/j.enconman.2016.01.060.

LBNL. *EnergyPlus (version 8.5.0) [Software]*. Lawrence Berkeley National Laboratory 2015. Available from http://www.eere.energy.gov/buildings/energyplus/.

Lee, Sang Hoon, Tianzhen Hong, Mary Ann Piette, Geof Sawaya, Yixing Chen, and Sarah C. Taylor-Lange. 2015. "Accelerating the energy retrofit of commercial buildings using a database of energy efficiency performance." *Energy* no. 90, Part 1:738-747. doi: http://dx.doi.org/10.1016/j.energy.2015.07.107.

Lee, Sang Hoon, Fei Zhao, and Godfried Augenbroe. 2013. "The use of normative energy calculation beyond building performance rating." *Journal of Building Performance Simulation* no. 6 (4):282-292. doi: 10.1080/19401493.2012.720712.

Levine, Mark D., Wei Feng, Ke Jing, Tianzhen Hong, and Nan Zhou. 2014. A Retrofit Tool for Improving Energy Efficiency of Commercial Buildings. edited by Mark D. Levine. Berkeley, CA: Lawrence Berkeley National Laboratory.

Li, S., M. Zhou, K. Deng, Z. Song, and Y. Lu. 2014. Sensitivity analysis for building energy simulation model calibration via automatic differentiation. Paper read at Proceedings of the 11th IEEE International Conference on Networking, Sensing and Control, 7-9 April 2014.

Li, Xiwang, Jin Wen, and Er-Wei Bai. 2016. "Developing a whole building cooling energy forecasting model for on-line operation optimization using proactive system identification." *Applied Energy* no. 164:69-88. doi: http://dx.doi.org/10.1016/j.apenergy.2015.12.002.

Liao, Z., and A. L. Dexter. 2004. "A simplified physical model for estimating the average air temperature in multi-zone heating systems." *Building and Environment* no. 39 (9):1013-1022. doi: http://dx.doi.org/10.1016/j.buildenv.2004.01.034.

Lomas, Kevin J., and Herbert Eppel. 1992. "Sensitivity analysis techniques for building thermal simulation programs." *Energy and Buildings* no. 19 (1):21-44. doi: http://dx.doi.org/10.1016/0378-7788(92)90033-D.

Lu, Yuehong, Shengwei Wang, Yang Zhao, and Chengchu Yan. 2015. "Renewable energy system optimization of low/zero energy buildings using single-objective and multi-objective optimization methods." *Energy and Buildings* no. 89:61-75. doi: http://dx.doi.org/10.1016/j.enbuild.2014.12.032.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing data using t-SNE." *Journal of Machine Learning Research* no. 9 (Nov):2579-2605.

Mauro, Gerardo Maria, Mohamed Hamdy, Giuseppe Peter Vanoli, Nicola Bianco, and Jan L. M. Hensen. 2015. "A new methodology for investigating the cost-optimality of energy retrofitting a building category." *Energy and Buildings* no. 107:456-478. doi: http://dx.doi.org/10.1016/j.enbuild.2015.08.044.

Michael, Wetter, and A. Wright Jonathan. 2003. Comparison of a generalized pattern search and a genetic algorithm optimization method. In *Proc. of the 8th IBPSA Conference*. Eindhoven, Netherlands.

Mitchell, Melanie. 1998. *An introduction to genetic algorithms*.

Mitchell, T. M. 1980. *The need for biases in learning generalizations*. New Brunswick, New Jersey, USA.: CBM-TR 5-110, Rutgers University.

Monteiro, Helena, John E. Fernández, and Fausto Freire. 2016. "Comparative life-cycle energy analysis of a new and an existing house: The significance of occupant's habits, building systems and embodied energy." *Sustainable Cities and Society* no. 26:507-518. doi: http://dx.doi.org/10.1016/j.scs.2016.06.002.

Müllner, Daniel. 2011. "Modern hierarchical, agglomerative clustering algorithms." *arXiv preprint arXiv:1109.2378*.

Murray, Sean N., Benoit Rocher, and D. T. J. O'Sullivan. 2012. "Static Simulation: A sufficient modelling technique for retrofit analysis." *Energy and Buildings* no. 47 (0):113-121. doi: http://dx.doi.org/10.1016/j.enbuild.2011.11.034.

Nik, Vahid M., Érika Mata, and Angela Sasic Kalagasidis. 2015. "A statistical method for assessing retrofitting measures of buildings and ranking their robustness against climate change." *Energy and Buildings* no. 88:262-275. doi: http://dx.doi.org/10.1016/j.enbuild.2014.11.015.

NOAA. 2016. NowData – Online Weather Data. National Oceanic and Atmospheric Administration.

Noris, Federico, Gary Adamkiewicz, William W. Delp, Toshifumi Hotchi, Marion Russell, Brett C. Singer, Michael Spears, Kimberly Vermeer, and William J. Fisk. 2013. "Indoor environmental quality benefits of apartment energy retrofits." *Building and Environment* no. 68:170-178. doi: http://dx.doi.org/10.1016/j.buildenv.2013.07.003.

Parsons, Lance, Ehtesham Haque, and Huan Liu. 2004. "Subspace clustering for high dimensional data: a review." *Acm Sigkdd Explorations Newsletter* no. 6 (1):90-105.

Paulus, Mitchell T, David E Claridge, and Charles Culp. 2015. "Algorithm for automating the selection of a temperature dependent change point model." *Energy and Buildings* no. 87:95-104.

Pearson, Karl. 1901. "LIII. On lines and planes of closest fit to systems of points in space." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* no. 2 (11):559-572.

Pebesma, Edzer J, and Gerard BM Heuvelink. 1999. "Latin hypercube sampling of Gaussian random fields." *Technometrics* no. 41 (4):303-312.

Pombo, Olatz, Karen Allacker, Beatriz Rivela, and Javier Neila. 2016. "Sustainability assessment of energy saving measures: A multi-criteria approach for residential buildings

retrofitting—A case study of the Spanish housing stock." *Energy and Buildings* no. 116:384-394. doi: http://dx.doi.org/10.1016/j.enbuild.2016.01.019.

Pope, V.D., M.L. Gallani, P.R. Rowntree, and R.A. Stratton. 2000. "The impact of new physical parameterizations in the Hadley Centre climate model - HadAM3." *Climate Dynamics* no. 16 (2-3):123-146.

Quinlan, J. R. 1986. "Induction of Decision Trees." *Machine Learning* no. 1 (1):81 - 106.

Radhi, Hassan. 2009. "Evaluating the potential impact of global warming on the UAE residential buildings – A contribution to reduce the CO2 emissions." *Building and Environment* no. 44 (12):2451-2462. doi: 10.1016/j.buildenv.2009.04.006.

Ramos Ruiz, Germán, Carlos Fernández Bandera, Tomás Gómez-Acebo Temes, and Ana Sánchez-Ostiz Gutierrez. 2016. "Genetic algorithm for building envelope calibration." *Applied Energy* no. 168:691-705. doi: http://dx.doi.org/10.1016/j.apenergy.2016.01.075.

Ravi S. Srinivasan, William W. Braham, Daniel E. Campbell, and Charlie D. Curcija. 2012. "Re(De)fining Net Zero Energy: Renewable Emergy Balance in environmental building design." *Building and Environment* no. 47:300- 315.

Roberti, Francesca, Ulrich Filippi Oberegger, Elena Lucchi, and Alexandra Troi. 2017. "Energy retrofit and conservation of a historic building using multi-objective optimization and an analytic hierarchy process." *Energy and Buildings* no. 138:1-10. doi: http://dx.doi.org/10.1016/j.enbuild.2016.12.028.

Robles, Fernando. 1994. "International market entry strategies and performance of United States catalog firms." *Journal of Direct Marketing* no. 8 (1):59-70.

Rysanek, A. M., and R. Choudhary. 2012a. "A decoupled whole-building simulation engine for rapid exhaustive search of low-carbon and low-energy building refurbishment options." *Building and Environment* no. 50 (0):21-33. doi: http://dx.doi.org/10.1016/j.buildenv.2011.09.024.

Rysanek, A. M., and R. Choudhary. 2013. "Optimum building energy retrofits under technical and economic uncertainty." *Energy and Buildings* no. 57:324-337. doi: http://dx.doi.org/10.1016/j.enbuild.2012.10.027.

Rysanek, A.M., and R. Choudhary. 2012b. "A decoupled whole-building simulation engine for rapid exhaustive search of low-carbon and low-energy building refurbishment options." *Building and Environment* no. 50:21-33.

Shannon, C.E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* no. 27:379 - 423.

Shao, Yunming, Philipp Geyer, and Werner Lang. 2014. "Integrating requirement analysis and multi-objective optimization for office building energy retrofit strategies." *Energy and Buildings* no. 82:356-368. doi: http://dx.doi.org/10.1016/j.enbuild.2014.07.030.

Shen, Pengyuan. "Impacts of climate change on U.S. building energy use by using downscaled hourly future weather data." *Energy and Buildings*. doi: http://dx.doi.org/10.1016/j.enbuild.2016.09.028.

Shen, Pengyuan. 2017. "Impacts of climate change on U.S. building energy use by using downscaled hourly future weather data." *Energy and Buildings* no. 134:61-70. doi: http://dx.doi.org/10.1016/j.enbuild.2016.09.028.

Shen, Pengyuan, and Noam Lior. 2016. "Vulnerability to climate change impacts of present renewable energy systems designed for achieving net-zero energy buildings." *Energy* no. 114:1288-1305. doi: http://dx.doi.org/10.1016/j.energy.2016.07.078.

Shen, Pengyuan, and Jennifer R. Lukes. 2015. "Impact of global warming on performance of ground source heat pumps in US climate zones." *Energy Conversion and Management* no. 101:632-643. doi: http://dx.doi.org/10.1016/j.enconman.2015.06.027.

Siddharth, V., P. V. Ramakrishna, T. Geetha, and Anand Sivasubramaniam. 2011. "Automatic generation of energy conservation measures in buildings using genetic algorithms." *Energy and Buildings* no. 43 (10):2718-2726. doi: http://dx.doi.org/10.1016/j.enbuild.2011.06.028.

Sneath, Peter HA. 1957. "The application of computers to taxonomy." *Microbiology* no. 17 (1):201-226.

Son, Hyojoo, and Changwan Kim. 2016. "Evolutionary Multi-objective Optimization in Building Retrofit Planning Problem." *Procedia Engineering* no. 145:565-570. doi: http://dx.doi.org/10.1016/j.proeng.2016.04.045.

Spandagos, Constantinos, and Tze Ling Ng. 2017. "Equivalent full-load hours for assessing climate change impact on building cooling and heating energy consumption in large Asian cities." *Applied Energy* no. 189:352-368. doi: http://dx.doi.org/10.1016/j.apenergy.2016.12.039.

Storn, Rainer, and Kenneth Price. 1997. "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces." *Journal of Global Optimization* no. 11 (4):341-359. doi: 10.1023/A:1008202821328.

Sun, Yongjun, Pei Huang, and Gongsheng Huang. 2015. "A multi-criteria system design optimization for net zero energy buildings under uncertainties." *Energy and Buildings* no. 97:196-204. doi: http://dx.doi.org/10.1016/j.enbuild.2015.04.008.

Tadesse, Mahlet G, Naijun Sha, and Marina Vannucci. 2005. "Bayesian variable selection in clustering high-dimensional data." *Journal of the American Statistical Association* no. 100 (470):602-617.

Tadeu, S., C. Rodrigues, A. Tadeu, F. Freire, and N. Simões. 2015. "Energy retrofit of historic buildings: Environmental assessment of cost-optimal solutions." *Journal of Building Engineering* no. 4:167-176. doi: http://dx.doi.org/10.1016/j.jobe.2015.09.009.

Terés-Zubiaga, J., C. Escudero, C. García-Gafaro, and J. M. Sala. 2015. "Methodology for evaluating the energy renovation effects on the thermal performance of social housing buildings: Monitoring study and grey box model development." *Energy and Buildings* no. 102:390-405. doi: http://dx.doi.org/10.1016/j.enbuild.2015.06.010.

Thorndike, Robert L. 1953. "Who belongs in the family?" *Psychometrika* no. 18 (4):267-276.

Tian, Wei. 2013. "A review of sensitivity analysis methods in building energy analysis." *Renewable and Sustainable Energy Reviews* no. 20:411-419. doi: http://dx.doi.org/10.1016/j.rser.2012.12.014.

USEPA. 2009. Building and Their Impact on the Environment: a Statistical Summary.

Wang, Bing, Ruo-Yu Ke, Xiao-Chen Yuan, and Yi-Ming Wei. 2014. "China′s regional assessment of renewable energy vulnerability to climate change." *Renewable and Sustainable Energy Reviews* no. 40:185-195. doi: http://dx.doi.org/10.1016/j.rser.2014.07.154.

Wang, Bo, and Xiaohua Xia. 2015. "Optimal maintenance planning for building energy efficiency retrofitting from optimization and control system perspectives." *Energy and Buildings* no. 96:299-308. doi: http://dx.doi.org/10.1016/j.enbuild.2015.03.032.

Wang, Bo, Xiaohua Xia, and Jiangfeng Zhang. 2014. "A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings." *Energy and Buildings* no. 77:227-235. doi: http://dx.doi.org/10.1016/j.enbuild.2014.03.025.

Wang, Xiaoming, Dong Chen, and Zhengen Ren. 2011. "Global warming and its implication to emission reduction strategies for residential buildings." *Building and Environment* no. 46 (4):871-883. doi: 10.1016/j.buildenv.2010.10.016.

Ward Jr, Joe H. 1963. "Hierarchical grouping to optimize an objective function." *Journal of the American statistical association* no. 58 (301):236-244.

Wetter, Michael, and Jonathan Wright. 2004. "A comparison of deterministic and probabilistic optimization algorithms for nonsmooth simulation-based optimization." *Building and Environment* no. 39 (8):989-999. doi: http://dx.doi.org/10.1016/j.buildenv.2004.01.022.

Wu, Raphael, Georgios Mavromatidis, Kristina Orehounig, and Jan Carmeliet. 2017. "Multiobjective optimisation of energy systems and building envelope retrofit in a residential community." *Applied Energy* no. 190:634-649. doi: http://dx.doi.org/10.1016/j.apenergy.2016.12.161.

Xu, Peng, Yu Joe Huang, Norman Miller, Nicole Schlegel, and Pengyuan Shen. 2012. "Impacts of climate change on building heating and cooling energy patterns in California." *Energy* no. 44 (1):792-804. doi: 10.1016/j.energy.2012.05.013.

Yalcintas, Melek. 2008. "Energy-savings predictions for building-equipment retrofits." *Energy and Buildings* no. 40:2111-2120.

Yoza, Akihiro, Atsushi Yona, Tomonobu Senjyu, and Toshihisa Funabashi. 2014. "Optimal capacity and expansion planning methodology of PV and battery in smart house." *Renewable Energy* no. 69:25-33. doi: http://dx.doi.org/10.1016/j.renene.2014.03.030.

Zhai, John, Michael Bendewald, and Nicole Hammer. 2011. "Deep Energy Retrofit of Commercial Buildings: A Key Pathway Toward Low-Carbon Cities." *Carbon Management* no. 2 (4):425 - 430.