Energy 172 (2019) 892–912

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Rapid multi-objective optimization with multi-year future weather condition and decision-making support for building retrofit

Pengyuan Shen [a, b, *], William Braham [c], Yunkyu Yi [d], Eric Eaton [e]

[a] School of Architecture, Harbin Institute of Technology, Shenzhen, 518055, China
[b] Urban Smart Energy Group, Shenzhen Key Laboratory of Urban Planning and Decision Making, Harbin Institute of Technology, Shenzhen, 518055, China
[c] Department of Architecture, University of Pennsylvania, PA, 19104, United States
[d] School of Architecture, University of Illinois at Urbana–Champaign, 61821, United States
[e] Department of Computer and Information Science, University of Pennsylvania, PA, 19104, United States

ARTICLE INFO

Article history:
Received 4 December 2018
Received in revised form
19 January 2019
Accepted 30 January 2019
Available online 5 February 2019

Keywords:
Building retrofit
Optimization
Heuristic method
Pareto fronts
Hierarchical clustering
Climate change

ABSTRACT

A method of fast multi-objective optimization and decision-making support for building retrofit planning is developed, and lifecycle cost analysis method taking into account of future climate condition is used in evaluating the retrofit performance. In order to resolve the optimization problem in a fast manner with recourse to non-dominate sorting differential evolution algorithm, the simplified hourly dynamic simulation modeling tool SimBldPy is used as the simulator for objective function evaluation. Moreover, the generated non-dominated solutions are treated and rendered by a layered scheme using agglomerative hierarchical clustering technique to make it more intuitive and sense making during the decision-making process as well as to be better presented.

The suggested optimization method is implemented to the retrofit planning of a campus building in UPenn with various energy conservation measures (ECM) and costs, and more than one thousand Pareto fronts are obtained and being analyzed according to the proposed decision-making framework. Twenty ECM combinations are eventually selected from all generated Pareto fronts. It is manifested that the developed decision-making support scheme shows robustness in dealing with retrofit optimization problem and is able to provide support for brainstorming and enumerating various possibilities during the decision-making process.

1. Introduction

The function and use of building have evolved greatly after the industrial revolution as they now can be more efficiently and rapidly constructed in vast scale, which facilitated the process of urbanization and the diversification of building's functionality. Buildings are different in terms of their types of use, geometrical shape, material use, building systems, and etc., so individual building may not be treated equally in the process of planning and decision-making in design, construction, and retrofitting. 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 [1]. For commercial buildings, about half of all commercial buildings were constructed before 1980 according to Commercial Building Energy Consumption Survey (CBECS) [2]. 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 [3]. Approximately 86% of current building construction expenditures are spent on renovating existing buildings and the rest on new construction [4]. With the trend of slow increase in newly constructed buildings in recent years, many of the existing buildings in the United States should be renovated if energy consumption and carbon emissions are expected to be significantly reduced in the future.

On the other hand, regarding sustainable building design, there could be great deal of tailor-made design variables to be considered for each individual building. One of the most important factors in sustainable building design is the energy performance of a building, and this involves many factors including site location, onsite climate condition, building envelope, passive design, energy

* Corresponding author. School of Architecture, Harbin Institute of Technology, Shenzhen, 518055, China.
E-mail address: Pengyuan_pub@163.com (P. Shen).

https://doi.org/10.1016/j.energy.2019.01.164
0360-5442/

Nomenclature

A_{floor}	Total floor area (m^2)
$A_{roof,j}$	Area of building wall with jth wall insulation, m^2
$A_{wall,i}$	Area of building roof with ith wall insulation, m^2
$A_{win,l}$	Area of window with lth wall insulation, m^2
$A_{win,n}$	Area of window with nth shading material, m^2
C_{infl}	Retrofit cost for air tightness improvement, $/$m^2$
C_{others}	Other costs, $
$C_{roof-insu,j}$	Roof insulation cost, $/$m^2$
$C_{shade,n}$	Shading material cost, $/$m^2$
$C_{wall-insu,i}$	Wall insulation cost, $/$m^2$
$C_{win,l}$	Window replacement cost, $/$m^2$
$E_{C,\,i}$	Hourly cooling energy use, J
$E_{DHW,\,i}$	Hourly domestic hot water heating energy use, J
$E_{equipment,\,i}$	Hourly energy use of equipment, J
$E_{H,\,i}$	Hourly heating energy use, J
$E_{light,\,i}$	Hourly lighting energy use, J
$E_{post,k}$	k year's annual energy use after retrofit, J
$E_{pre,k}$	k year's annual energy use without retrofit, J
E_{pump}	Hourly pump energy use, J
E_{PV}	Hourly solar panel electricity production, J
E_{saving}	Aggregated energy saving during lifecycle, J
E_{SWH}	Hourly solar water heater energy production, J
E_u	Hourly energy use of a utility, J
F_u	Primary energy conversion factor
I_k	Proportion of the maintenance cost of the initial investment, %
I_{total}	Total investment cost, $
n_{pv}	Number of PV panels
I_{pv}	Solar irradiation hitting on the PV panel surface, W/m^2
I_{sh}	Solar irradiation hitting on the solar water heater, W/m^2
P_{pv}	Power output of PV system, W
Q_{sh}	Thermal energy output of the solar water heater, W
S^u_{total}	total saving for type of energy source u, $
S_{pv}	Array area of each PV panel, m^2
t_a	Outside air temperature of PV panel, $^\circ C$
τ_m	Cost increase in maintenance fee of each year, %
τ_u	Cost increase for type of energy source u, %
μ_{pv}	Efficiency of PV panel, %
μ_{sh}	Efficiency of the solar water heater, %
μ_{ex}	Efficiency of the heat exchanger of the solar water heating system, %
u	Utility type

Acronyms

ANN	Artificial neural network
AR5	IPCC Fifth Assessment Report
ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers
BES	Building energy simulation
BEU	Building energy use
CBECS	Commercial Building Energy Consumption Survey
CMU	Concrete masonry unit
CV	Coefficient of variation
DE	Differential evolution
DHW	Domestic hot water
DOE	Department of Energy
ECM	Energy conservation measure
FRES	Penn Facilities and Real Estate Services
HVAC	Heating, Ventilating, and Air Conditioning
IPCC	Intergovernmental Panel on Climate Change
LCA	Lifecycle cost analysis
GA	Genetic algorithm
GCM	Global climate model
GHG	Greenhouse gas
MOGA	Multi-objective genetic algorithm
NPV	Net present value
NSDE	Non-dominated sorting differential evolution
NSGA	Non-dominated sorting genetic algorithm
PMV	Predicted mean vote
PV	Solar panel
RAM	Random access memory
RC	Resistance-capacitance
RCP	Representative concentration pathway
RF	Random forest
RMSE	Root mean squared error
SHGC	Solar heat gain coefficient
SWH	Solar water heater
TMY	Typical meteorological year
UPenn	University of Pennsylvania

system, electric appliances, occupant comfort, etc. 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 [5]. Moreover, it has been projected that the energy consumption of current buildings is going to grow annually by 1.7%–2025 [6]. Mills et al. [7,8] has shown that improving existing buildings will yield median energy savings of 16% in the United States. The great potential in energy reduction in existing building has created opportunities in building energy retrofit projects [9]. A successful building energy retrofit project should also be affordable, which takes into account factors like investment budget, payback period, economic risks and uncertainties, and etc.

Building retrofit usually involves updating building device and system, in which a series of choices – energy conservation measures (ECM), can be applied. It is not an easy project considering that building is a complex system with a decision-making process aiming to meet the goal of high energy performance and some constraints related to project costs, cultural and social needs, expectations from building owners, and technology limitations. The answer to the problem of searching the optimal ECMs is to look for an admissible set of values of the command variables (alternatives) compatible with the constraints (inner environment), that maximize the utility function (interface connecting inner and outer environment) for the given parameters of design variables (outer environment). The difficulty in solving such a problem lies in that the objective function of the retrofit project usually possesses nonlinearity as well as multicollinearity with respect to the different design variables. One ECM in a retrofit bundle may have influence on other ECMs and the options for choosing them. To solve this problem, building simulation is often a method adopted to study the influences and interactions among ECMs and the trade-offs

according to selected objectives. However, this process can be complicating considering the collection of existing building information and retrofit options, the complexity in building modeling and simulation, the optimization process, and the decision-making phase, in which not only expert panels such as architects and engineers are involved, but clients should also actively participate. There are many on-going researches in the field looking for the method of finding optimal solutions for the building energy retrofit, and they are discussed in the following literature review.

2. Literature review

A comprehensive literature study concerning building retrofit optimization research have been listed in Table 1.

Table 1 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 the thermal comfort. For the BES tools used in those research, EnergyPlus [30] is mostly used, and others include TRNSYS [31], eQuest, which uses the DOE simulation engine [32]. 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 in dynamic way and has undergone numerous reliability tests [33] and is validated to be well within the accuracy needed for building design [34]. Tools like EnergyPlus and TRNSYS are mostly used for energy diagnoses and scientific investigations because of their modeling and simulation accuracy compared with steady or semi-steady state simulation tools.

The 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 simulate, especially when it is faced with exponentially increasing simulation number that is 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: Evaluation of Objective Functions* for the research problem.

Limitation Two: Generalization of Archetypical Building. This limitation 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 [18] and providing clients with better decision-making support. As we have discussed in the beginning of the paper, 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 (UPenn) 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 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 (where weighted sum method is often used) and the non-dominated method (Pareto front). Then here is *Limitation Three: Lack of Decision-making* Support. Not enough support is 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-objective before optimization 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 the determination of the weighting factors. It would be difficult for users or clients to define appropriate weight values in the final objective function, with little knowledge on 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: Necessity of Using LCA. It should be noted that lifecycle cost analysis (LCA) is necessary for a retrofit project because clients tend not to do retrofit too frequently in a short period of time, while most LCA methods in the reviewed research do not take into account future climate uncertainties. In literature [29], Ascione et al. used NSGA-II algorithm to optimize the energy use and the global cost under different global warming scenarios. In this research, it is assumed that the future air temperature increment is discretized in 32 years and an 8-year period is used to perform the building simulation in EnergyPlus. It is not difficult to simulate future years' hourly energy use by using hourly downscaled future weather data [35–38], 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 are going to be found and tested in this research.

Confronted with the four limitations, in this research, a method of fast multi-objective optimization and decision-making support for building retrofit planning is developed and introduced, and LCA method taking into account of climate change will be used in the objective function. A retrofit optimization of an on-campus building in University of Pennsylvania to which the developed method and framework is implemented will be studied and discussed. The highlights of this research corresponding to each stated limitation, which can be potential solutions to the limitations are listed as follows:

a) Limitation One: Evaluation of Objective Functions

Table 1
Literature review on building retrofit optimization.

Literature	Objectives	Objective Function Evaluation	Energy Use Evaluation	Optimization Algorithm	Authors	Year of Publication	Type of Building
[10]	Aggregated energy saving, internal rate of return	Weighted sum multi-objective	Model predictive control	Differential evolution	Bo Wang, Xiaohua Xia	2015	Office building built in 1960s
[11]	Energy saving, lifecycle NPV, discounted payback period	Weighted sum multi-objective	Estimation	Differential evolution	Bo Wang, Xiaohua Xia, Jiangfeng Zhang	2014	Office building built in 1960s
[12]	Total lifecycle cost (LCC)	Single objective optimization	eQuest & static modeling	Genetic algorithm	Amirhosein Jafari, Vanessa Valentin	2017	House built in 1964
[13]	Initial investment cost, energy consumption, global warming potential	Multi-objective Pareto front	DIN V 18599 assessment method	NSGA-II	Yunming Shao, Philipp Geyer, Werner Lang	2014	Office building built in 1900
[14]	Retrofit cost, energy saving in kWh, thermal comfort	Weighted Tchebycheff metric	TRNSYS	Tchebycheff programming	Ehsan Asadi, Manuel Gameiro da Silva, Carlos Henggeler Antunes, Luís Dias	2012	House built in 1945
[15]	Greenhouse gas emission reduction	Single objective	TRNSYS & Matlab	Brute-force	A.M. Rysanek, R. Choudhary	2012	Office: built in 1960s; School: late 19th century
[16]	Thermal comfort, annual energy consumption	Weighted sum multi-objective	EnergyPlus	Repeated sampling and feature reduction	Bryan Eisenhower, Zheng O'Neill, Satish Narayanan, Vladimir A. Fonoberov, Igor Mezi'c	2012	Office and gym building Built in 1910s
[17]	Energy demand, thermal comfort, global cost	Multi-stage analysis method	EnergyPlus & Matlab	Feature reduction and brute-force	Gerardo Maria Mauro, Mohamed Hamdy, Giuseppe Peter Vanoli, Nicola Bianco, Jan L.M. Hensenf	2015	Building stock erected between 1920 and 1970
[18]	Payback period	Single objective	Archetype modeling in EnergyPlus	Non-linear regression	S.E. Chidiac, E.J.C. Catania, E. Morofsky, S. Foo	2011	Archetype #1 – built prior to 1950, Archetype #2 – built between 1950 and 1975, Archetype #3 – built post 1975
[19]	Retrofit cost, energy saving in kWh	Weighted sum multi-objective	ISO 13790 RC model (monthly)	Tchebycheff programming	Ehsan Asadi, Manuel Gameiro da Silva, Carlos Henggeler Antunes, Luís Dias	2010	Built in 1945
[20]	Electricity use, natural gas use	Weighted sum multi-objective	DOE 2.2	Genetic algorithm	V. Siddharth, P.V. Ramakrishna, T. Geetha, Anand Sivasubramaniam	2011	Existing single-zone office building
[21]	Energy use, thermal comfort, conservation compatibility for historic building	Multi-objective Pareto front	EnergyPlus	NSGA-II	Francesca Roberti, Ulrich Filippi Oberegger, Elena Lucchi, Alexandra Troi	2017	Built before 1780s
[22]	Global cost, primary energy use	Multi-objective Pareto front	EnergyPlus	Brute-force	S. Tadeu, C.Rodrigues, A.Tadeu, F.Freire, N.Simões	2015	Historic building
[23]	Energy consumption, CO2 emissions, retrofit costs, and thermal comfort	Multi-objective Pareto front	EnergyPlus	NSGA-III	Hyojoo Son, Changwan Kim	2016	Existing public school building
[24]	Annualized costs and life cycle GHG emissions	Single obejective & Multi-objective Pareto front	EnergyPlus	Epsilon-constraint method	Raphael Wu, Georgios Mavromatidis, Kristina Orehounig, Jan Carmeliet	2017	Existing residential building
[25]	Energy consumption, retrofit cost, thermal comfort	Single objective and Multi-objective Pareto front	TRNSYS	Latin-hypercube sampling, artificial neural network (ANN), MOGA (Multi-Objective Genetic Algorithm)	Ehsan Asadi, Manuel Gameiro da Silva, Carlos Henggeler Antunesc,d,Luís Diasc,e, Leon Glicksman	2014	School built in 1983
[26]	Marginal abatement cost vs. GHG emissions saved; discounted payback period vs. required capital	Primary and secondary objectives	TRNSYS & Matlab	Brute-force	A.M. Rysanek, R. Choudhary	2013	Office building built in 1960s
[27]	Energy consumption, investment, thermal comfort	Multi-objective Pareto front	TRNSYS	NSGA-II	Fanny Pernodet Chantrelle, Hicham Lahmidi, Werner Keilholz, Mohamed El Mankibi, Pierre Michel	2011	Existing office building
[28]	Lifecycle financial saving, energy saving	Multi-objective Pareto front	EnergyPlus	Brute-force	Olatz Pombo, Karen Allacker, Beatriz Rivela, Javier Neila	2016	Residential block built in 1960s
[29]	Thermal energy demand, global cost	Multi-objective Pareto front	EnergyPlus & Matlab	NSGA II	Fabrizio Ascionea, Nicola Biancoa, Rosa Francesca De Masib, Gerardo Maria Mauroa, Giuseppe Peter Vanoli	2017	Civil building built prior to 1900

The proposed lightweight simulation engine, SimBldPy, and the validated regression technique based on random forest algorithm greatly reduces the computational cost in objective function evaluation during the optimization.

b) Limitation Two: Generalization of Archetypical Building

Since each building is unique in terms of their geometry, thermal properties, operation schedules, and etc., the developed SimBldPy simulation tool, which offers thermal coupling between neighboring building zones, is capable of the rapid modeling of most geometrically unsophisticated existing mixed-use buildings and performing rapid optimization.

c) Limitation Three: Lack of Decision-making Support

The proposed decision-making support framework based on layered hierarchical clustering technique is manifested to show robustness in handling retrofit optimization problem and is able to provide support for brainstorming and enumerating various possibilities and solutions during decision-making process.

d) Limitation Four: Necessity of Using LCA

The use of LCA method with future climate condition cooperated into the evaluation of the energy and economic performance of the building retrofit, provides robustness to retrofit planning against climate uncertainty during the lifecycle.

3. Methodology

3.1. Optimization framework

In this research, we developed a Python building simulation tool called SimBldPy, which uses a simplified hourly dynamic method based on resistance-capacitance (RC) gray modeling technique. 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 building simulation tool to evaluate the objective function in the optimization. The SimBldPy is a tool designed and developed mainly for the purpose of the fast-parametric study for building simulation optimization problem, and in this case, it is used for the optimization for the design space of building retrofit problem.

The optimization work flow is shown in Fig. 1. The SimBldPy model will first be calibrated based on the metered energy use. Then, a building retrofitting module will be in charge of reading ECM entries 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 is collected, the evolutionary optimizer will generate simulation tasks to the SimBldPy engine. In order to assess the impact of climate change on building energy use (BEU) in the future years, a predictive model based random forest (RF) is developed in this research and it will conduct projections on the future hourly energy use, which will be introduced in section 3.1.2.2 in details. The optimizer will iterate to improve the objective function while turning 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.

3.1.1. 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, U value and SHGC (solar heat gain coefficient) of transparent part of the building envelope, 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:

3.1.1.1. 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

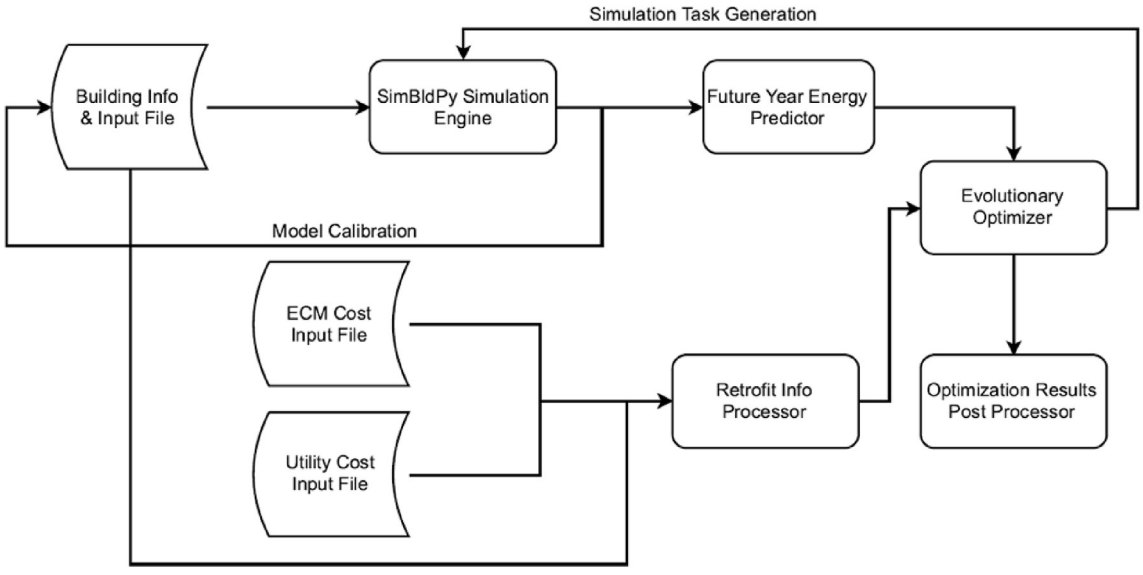

Fig. 1. Work flow of optimization approach.

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 0–1, 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.

3.1.1.2. 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 1 W/m^2-K, 2 W/m^2-K, and 4 W/m^2-K for the window retrofitting. By applying the mapping method described in section 3.1.1.1, all ECMs and their parameters will be projected into a uniform distributed sampling space between [0, 1).

3.1.2. 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.

3.1.2.1. 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. As discussed in Literature Review, *Limitation One* points out the significance of reducing the computational cost when dealing with building energy performance evaluation. Since in building retrofitting optimization problem, the relative comparison among different retrofit bundles is the priority, a lightweight energy simulation engine could be of great help in reducing computational complexity. One of those modeling methods is using electrical analogue for modeling the thermal behavior of the building, which is more commonly known as RC model. People has been using RC method to model building's thermal reaction under the synergy of indoor and outdoor conditions [39], and it has been validated that RC model is able to handle hourly building energy simulation requirement for simple building that usually has one zone or single type of use. However, RC modeling has its limitation of application in more complicated building types. ISO 13790 provides a monthly method and a simplified hourly dynamic method that uses 5R1C modeling method for the calculation of building heating and cooling need [40]. The monthly method is the one that has been used in Ref. [19] for building retrofit analysis, whereas in this research, the simplified dynamic hourly method is used. Moreover, zone thermal coupling method is also adopted to extend its use to more complex mixed-use building type. The modified 5R1C modeling method is presented in Fig. 2. The detailed introduction of the thermal modeling process in SimBldPy is included in Appendix A. Thermal modeling in SimBldPy.

We implemented the 5R1C modeling method using Python programming and created a simulation tool called SimBldPy. The input file format is text based, which resembles EnergyPlus, DOE2 engines and makes it easy to input and manipulate by users or clients. The detailed description of the modeling method and the structure of the modeling tool is introduced in Ref. [41]. The validity of the modeling tool and its responsiveness under various climate conditions and the synergy of different ECMs have been proved by EnergyPlus results with two DOE reference buildings in Ref. [41].

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

$$E_u = \sum_i \left(E_{H,i} + E_{C,i} + E_{DHW,i} + E_{light,i} + E_{equipment,i}\right) + E_{pump} - E_{PV} - E_{SWH}\ (J) \quad (1)$$

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

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 $E_{H,\,i}, E_{C,\,i}, E_{DHW,\,i}$, $E_{light,\,i}, E_{equipment,\,i}$ is energy use of heating, cooling, domestic hot water (DHW), lighting, and equipment, respectively for ith HVAC zone, in Joule. E_{total} is the total hourly energy use of all utility types and F_u is the primary energy conversion factor for a certain type of utility. Among them, E_{PV} is electricity production from solar panel and its value of each time step cannot 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.

3.1.2.2. Energy projection in future years. It is known that climate change is going to change the building retrofit optimization results if future decade's climate conditions are considered in the optimization process [42]. It is concluded by Shen et al. that the retrofit strategy of selecting the best ECM combinations is subject to potential change brought by the changes in climate condition [42]. To understand how climate change is going to affect the building retrofit optimization, future year's building energy performance needs to be calculated. Using Belcher et al.'s morphing method,

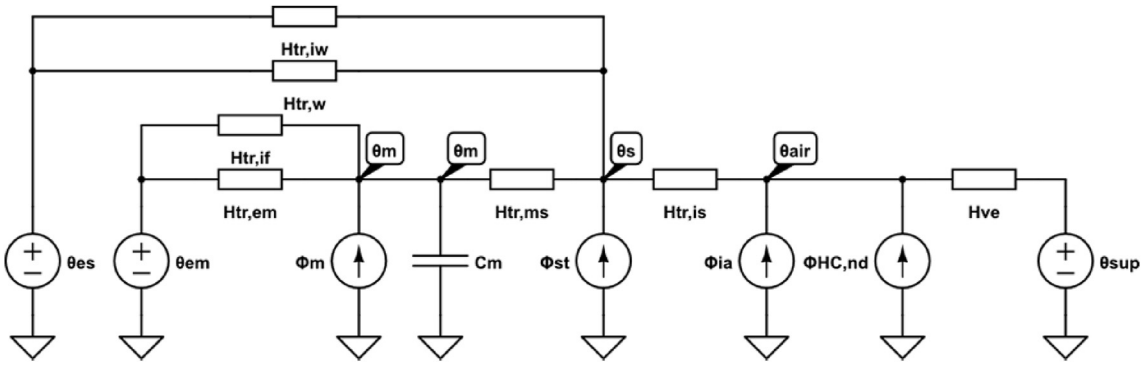

Fig. 2. 5R1C model with zone thermal coupling.

future local hourly weather data can be downscaled from global climate model (GCM) [38]. 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 [43]. These scenarios are described in the IPCC Fifth Assessment Report (AR5) and are named representative concentration pathways (RCPs). In this paper, the RCP6.0 scenario is selected to predict future climate condition for the building retrofit as an attempt to taking into account the potential impacts of climate change on building retrofit. The detailed introduction of morphing future hourly weather data is given in Appendix B. Morphing of future hourly weather data.

Even with the help of SimBldPy, it will still be computationally heavy to calculate each year's hourly energy use. In Refs. [42,44], a method of projecting future years' energy performance has been developed and validated by EnergyPlus results. Random forest (RF) algorithm is used as the data driven model training method for predicting future year's energy use. RF is an ensemble learning method based on a non-parametric supervised learning method called decision tree algorithm that makes use of a tree-like graph or model to learn and predict the pattern of the best routes or rules to specific goals. When applying the decision tree algorithm, the features for the independent variables can be either categorical or continuous. For a regular decision tree, the deeper the tree grows, the more complex the decision rules or routes and the fitter the model will be with the training data. However, a single decision tree will be biased because it overfits 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 included in the model training. To settle this problem, RF is created by a sets of decision trees in the training process and output the rule that is the mean prediction of the individual trees. This corrects the overfitting behavior of a single decision tree. Detailed application of the algorithm in future hourly energy use predictions can be referred to Refs. [42,44].

Since training a data driven model based on typical meteorological year (TMY) data will generate biased energy use results, RF is used to train the dataset constructed by extreme year hourly weather data and the corresponding simulation results of energy use from SimBldPy for each single ECM combination (including the baseline case without retrofit) in the optimization process, and the predictions of the RF model will serve as the basis for calculating energy saving in Joule and dollar.

Finally, total energy saving in kth year after the retrofit is simply given by:

$$E^{u}_{saving} = \sum_{k}\left(E^{u}_{post,k} - E^{u}_{pre,k}\right)\ (J) \tag{2}$$

$$E_{saving} = \sum_{u}\left(E^{u}_{saving} * F_u\right)\ (J) \tag{3}$$

where, $E_{pre,k}$ and $E_{post,k}$ is kth year's annual energy use without retrofit (baseline case) and energy use after kth 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 retrofit measures being 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 [18,26,45]. The calculation of the maintenance cost will be explained in 3.1.2.4.

3.1.2.3. 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 0.5 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.

3.1.2.4. 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:

$$\begin{aligned} I_0 = &\sum_{i} A_{wall,i} * C_{wall-insu,i} + \sum_{j} A_{roof,j} * C_{roof-insu,j} \\ &+ \sum_{l} A_{win,l} * C_{win,l} + \sum_{n} A_{win,n} * C_{shade,n} + A_{floor} * C_{infl} \\ &+ C_{others}\ (\$) \end{aligned} \tag{4}$$

where, $A_{wall,i}$, $A_{roof,j}$, $A_{win,l}$, $A_{win,n}$ is the area (m^2) of building wall, roof, and windows where ith wall insulation, jth roof insulation, lth window material, nth shading material with certain cost ($/$m^2$), respectively. A_{floor} is the total floor area where infiltration level was improved, which has a cost of C_{infl} ($/$m^2$). 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:

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

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 each 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 that will be discussed later, 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:

$$S^{u}_{total} = \sum_{k} \frac{(1+\tau_u)^k * \left(C^{u}_{post,\,k} - C^{u}_{pre,\,k} \right)}{(1+r)^k} \quad (\$) \tag{5}$$

$$S_{total} = \sum_{u} S^{u}_{total} \quad (\$) \tag{6}$$

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

Then, the NPV can be obtained by:

$$NPV = S_{total} - .I_{total} \tag{7}$$

3.1.2.5. 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, $X = (x_{wall-insu}, x_{roof-insu}, x_{win}, x_{infl}, x_{shade}, \ldots)$

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 3.1.3, 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 is 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 taking net present value (NPV) as the objective function is that by doing this, the decision maker will be given a chance to look at 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 stage will still be an important concern of the building owner or investor.

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 selection of retrofit packages that overheat and overcool the building too much. 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 and etc. The purpose of this paper 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 is subjective to individual building. 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.

3.1.3. Optimization algorithm

Traditionally a non-dominated genetic algorithm (NSGA-II) will be used to solve the problem in finding Pareto fronts [46]. This non-dominated sorting algorithm has been proved to be efficient and effective in finding non-dominated solutions for multi-objective optimization problems [13,27,47]. 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 [48]. 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) [49], 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 2.

- 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, which is good at generating a relatively small set of map realizations that captures most of the variability of the spatial inputs [50].

- 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 [51]. 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

Table 2
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.

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 [51]. Three different crossover operators are to be used and be compared in terms of their performance by the case study in this research. 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

Table 3
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

Table 4
Calibrated thermo physical properties of building envelopes.

Envelope	U-value (W/m2 °C)	Absorption coefficient for opaque envelope/SHGC for window
Wall	1.1	0.63
Roof	0.97	0.76
Below-grade	1.35	N/A
Window	4.5	0.72

Table 5
Description of ECMs.

Window replacement	change the window to another window type with different properties in thermal conductivity and solar heat gain coefficient
Wall insulation	adding insulation to the external walls of the building
Roof insulation	adding insulation to the external roofs of the building
Lighting efficiency upgrade	use energy efficient lighting equipment for all the zones in the building
Natural ventilation	open parts (%) of the windows in the building during swing seasons and during summer nights to flush redundant heat in the building
Window shading	use blinds shading for the windows when the building zone is overheated and needs cooling internally or externally
Cooling (heating) supply air temperature	change the cooling (heating) supply air temperature to certain value
Air tightness improvement	tighten up the building and improve the air change rate between indoor and outdoor environment

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 the 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 [51]. 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 each vector's bounding condition, which is 0–1 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 Tables 6–8 for each generation. A maximum of 100 generations is used as the stopping criterion for the evolution process.

3.2. 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 [21]. 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 [52].

A clustering problem can usually be described as follows [53]:

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

S.T.

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

Table 6
Window replacement properties and cost [14,19,69].

Window (SHGC, U-value (W/m2 °C))	$/m2	type
(0.0, 0.0)	0.00	N/A
(0.80, 5.6)	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 1.11 cm glass low-e
(0.63, 0.48)	131.7	3050 SH 1.11 cm glass
(0.25, 0.26)	183.0	3050 DH 3–7/16 insulated glass low-e krypton filled triple pane

Table 7
ECM parameters and costs (w/cost) [11,13,70–72].

wall insulation (m^2 °C/W)	$/wall m^2	roof insulation (m^2 °C/W)	$/roof m^2	windowshading	$/window m^2	air infiltration (h^{-1})	$/$m^2$	lighting efficiency improvement	$/$m^2$	daylight control	$/$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 8
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	32	10%	22	18	Applied
12	31	20%	23	19	
13	30	30%	24	20	
14	29	40%	25	21	
	28	50%	26	22	
		60%	27	23	
		70%		24	
		80%		25	
		90%			
		100%			

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

$$x_{ij} \le x_{jj} \ \forall i,j$$

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

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 [54], design group technology manufacturing cells [55], define employment sub centers in Los Angeles region [56], and most importantly, it can be used to visualize high dimensional data as other clustering techniques do [57–60]. 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 in this research 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 [61]. 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 [62]:

$$d(h,k) = \alpha_i d(h,i) + \alpha_j d(h,j) + \beta d(i,j) + \delta |d(h,i) - d(h,j)| \quad (8)$$

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 α_i, α_j, β, δ are determined is based on different linkage method. For example, for single linkage clustering, the parameters are set as $\alpha_i = \alpha_j = 0.5$, $\beta = 0$, $\delta = -0.5$ [63]. In this study, the ward linkage is used as proposed by Ward in 1963, which is also called the "minimum variance method" [64]. The parameters used in this method are:

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

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 [65,66]. Here we adopt an "elbow" method [67], 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 clusters 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. 3 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. 3 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.

4. Implementation and results discussion

4.1. Building model

The method developed in this research is implemented on one of the campus buildings in UPenn—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 edifice of the building is shown in Fig. 4.

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

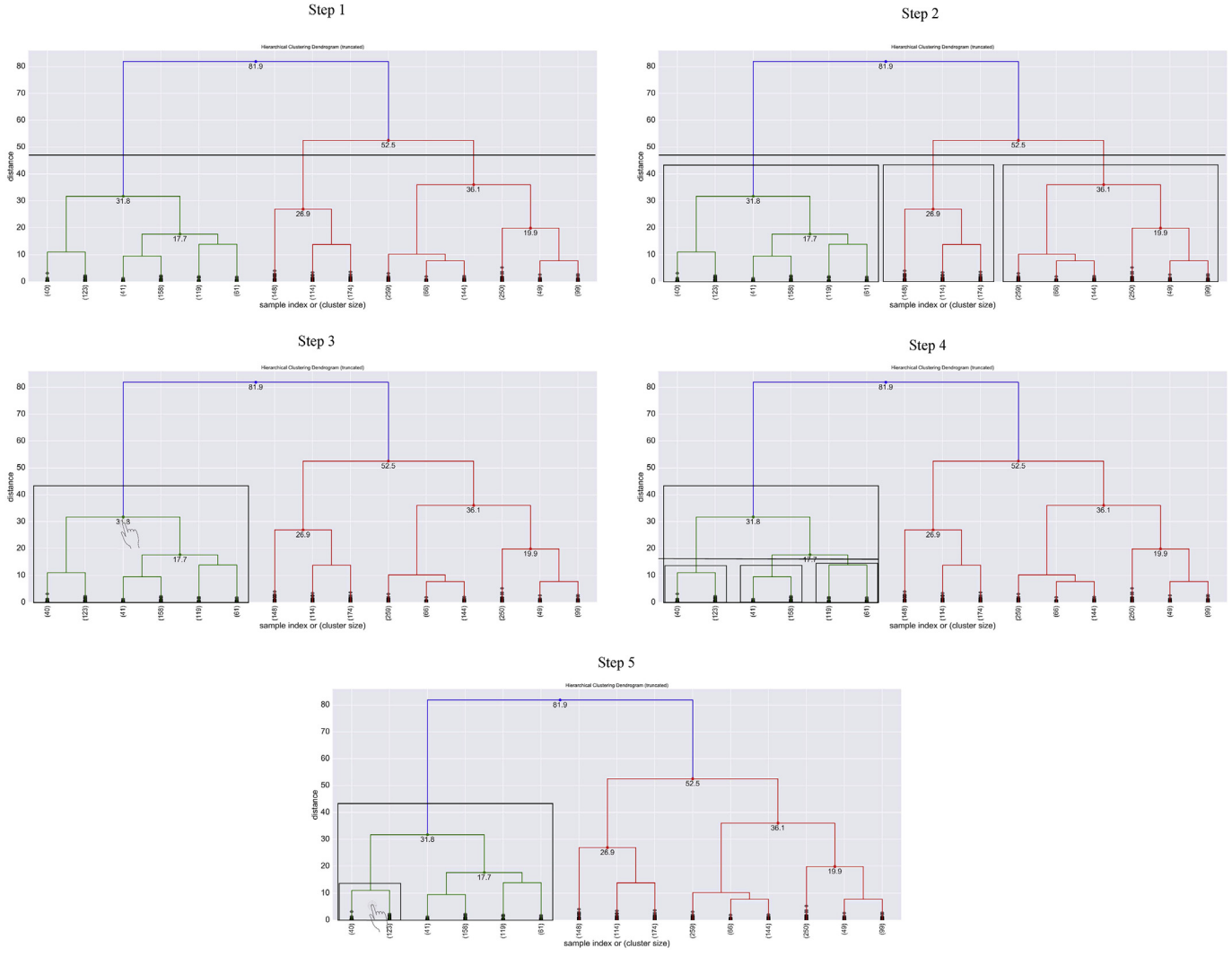

Fig. 3. Example of hierarchical clustering of a dataset in a layered framework.

Fig. 4. Towne building in UPenn.

modeled to make the SimBldPy model stay simple. The building envelopes are also modeled for each zone. The brief information of the building envelope is reported in Table 3.

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 calibrated thermophysical properties of the building envelope used in building simulation are shown in Table 4. 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 model is calibrated with its actual energy performance in 2015 by metered hourly and monthly energy use data, which are stored and maintained by FRES. The heating and cooling set point of the building is constant, which is 21.8°C for cooling and 22.5°C for heating, respectively. The floor height above the ground is 5.2 m and that of the basement is 4.4 m. 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 4. 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 detailed modeling and calibration of the building in SimBldPy can be found in Ref. [41]. The root mean squared error (RMSE) and coefficient of variation (CV) of the model are used to evaluate the predictive power of the calibrated model, and they are calculated in the following way:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \widehat{x}_i)^2}{n}} \tag{9}$$

$$CV = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \widehat{x}_i)^2}{n}} \Big/ \overline{x} \tag{10}$$

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

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 Ref. [68]. Fig. 5 shows the trends of monthly mean dry bulb temperature from the year of 2017–2069 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.

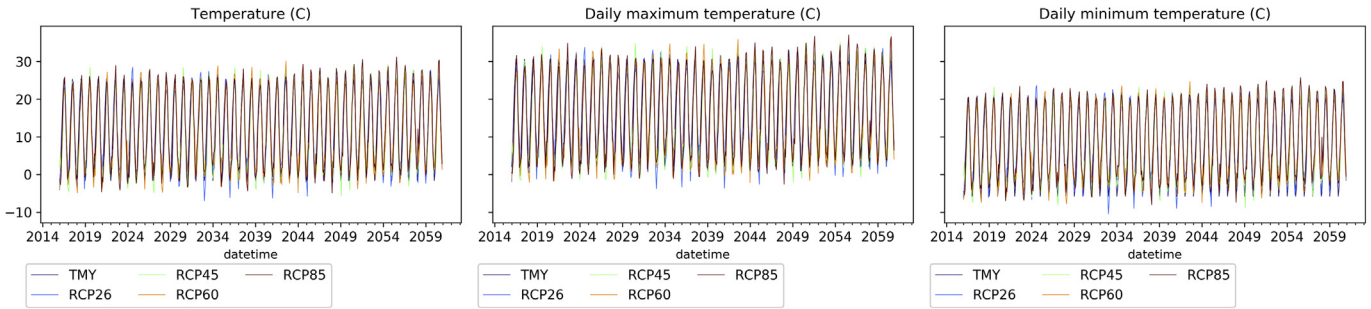

Fig. 5. Monthly mean air temperature, daily maximum and minimum temperature in different RCPs and TMY in Philadelphia.

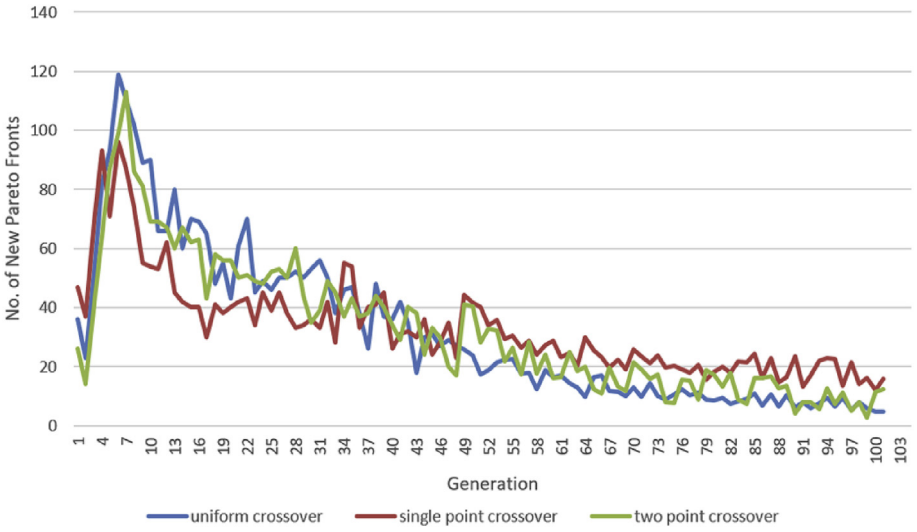

Fig. 6. Convergence of the optimization results.

4.2. ECM options and costs

The following ECMs are considered in the retrofit: window replacement (without window frames), wall insulation, roof insulation, window shading, air tightness 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). For PV and SWH system, different inclination angles are also considered in the optimization. The retrofit lifecycle is assumed to be twenty years, namely, from year 2018–2038. The parameter and cost of all ECMs are listed in Tables 6–8. The detailed description of the ECMs are provided in Table 5. For PV and SWH system, different inclination angles are also considered in the optimization.

The power output of the PV system is calculated by a method proposed by Erbs et al. [73]:

$$P_{pv} = n_{pv}\mu_{pv}S_{pv}\, I_{pv}(1 - 0.005(t_a - 25))\ [W]$$

where, n_{pv} is the number of panels, S_{pv} is the array area (2 m^2), μ_{pv} is the conversion efficiency of the solar cell used for the array. The

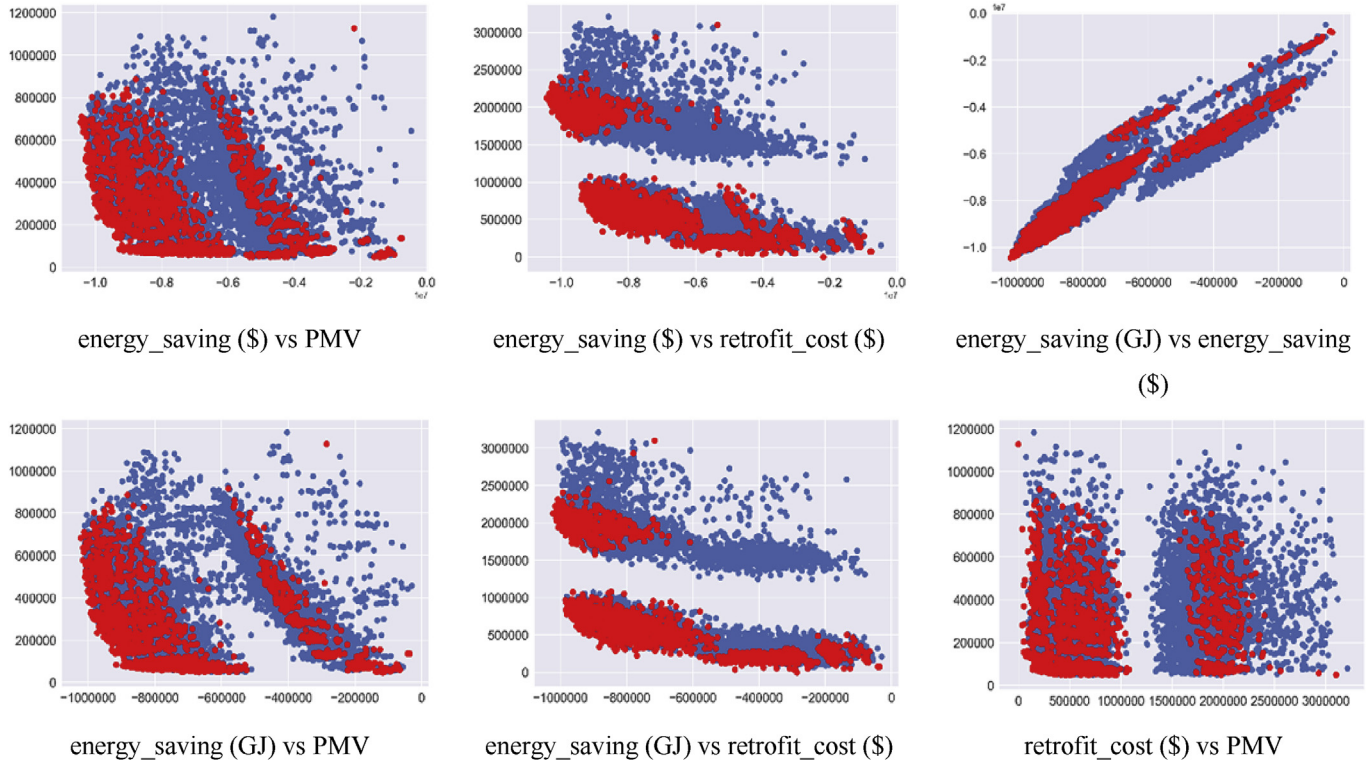

Fig. 7. 2-D projection of Pareto fronts with all combinations of objective functions.

solar cell used here is poly-crystalline silicon (p-Si) which has an efficiency of 14% [74]. I_{pv} is the solar irradiation hitting on the panel surface (W/m^2), t_a is the outside air temperature.

The solar thermal energy collected by the system can be calculated by:

$$Q_{sh} = \mu_{sh} S_{sh} I_{sh} \mu_{ex} [W]$$

where, μ_{sh} is the heat collection efficiency, which is 60% [75]; S_{sh} is the total heat collection area (2 m^2 per panel), and I_{sh} is the solar irradiation hitting on the thermal collector system surface (W/m^2), μ_{ex} is the efficiency of the heat exchanger set as 50%.

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.

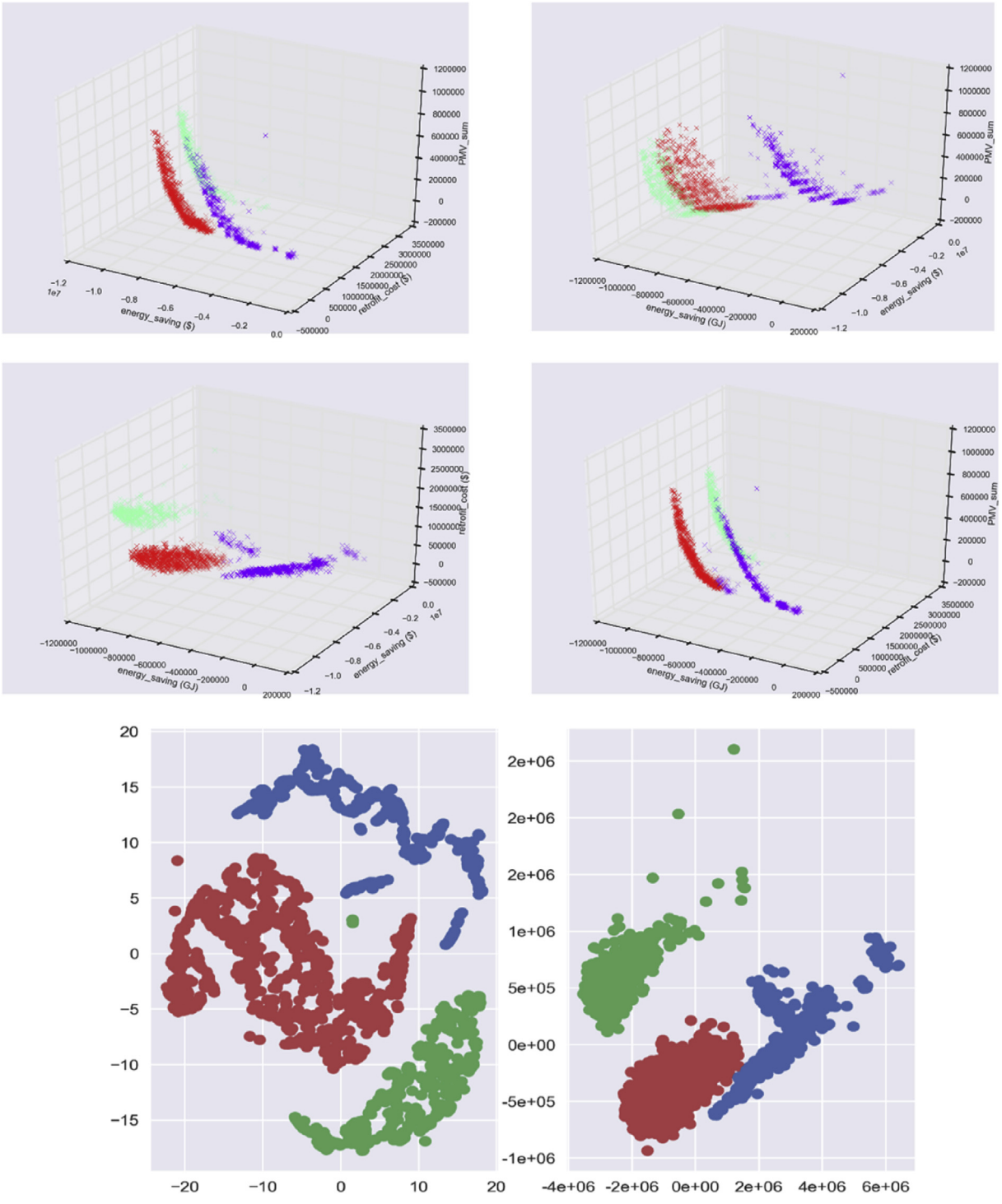

Fig. 8. 3-D projections and dimensionality reduction visualization (t-sne & PCA) of the clustered Pareto fronts.

4.3. Optimization results

The optimization and simulation are conducted on an Intel(R) Xeon(R) AWS-1100 v4 2.40 GHz server, which has a RAM of 32G. The optimization process takes about 23 h 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 Tables 6–8 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 the number of 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. 6 for the three crossover operators described in section 3.1.3.

As per Fig. 6, 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 less 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. 7. 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 retrofit optimization task that takes into account future hourly energy projection under climate change in a fast manner.

As discussed in section 2, 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. 7, it would still be difficult to make decision and have a general idea of which retrofit options to choose from. For a retrofit project, many ECM options as well as objective functions will be concerned. 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 section 3.2 will be implemented to the optimization results of this project.

4.4. 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. 8.

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. 7, 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. 8 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 [76]. 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 [77]. 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. 8, 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. 9 and Fig. 10 (parallel coordinates figure and heat map will also be plotted at each subsequent layer for decision-making, but will not be redundantly shown here):

By plotting the clustered Pareto fronts in parallel coordinates, the decision-making process will become more visible as the

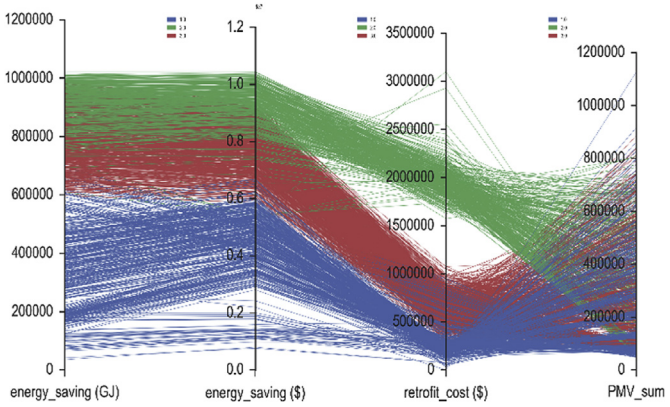

Fig. 9. Plotting of sub-objectives in the first layer clustering.

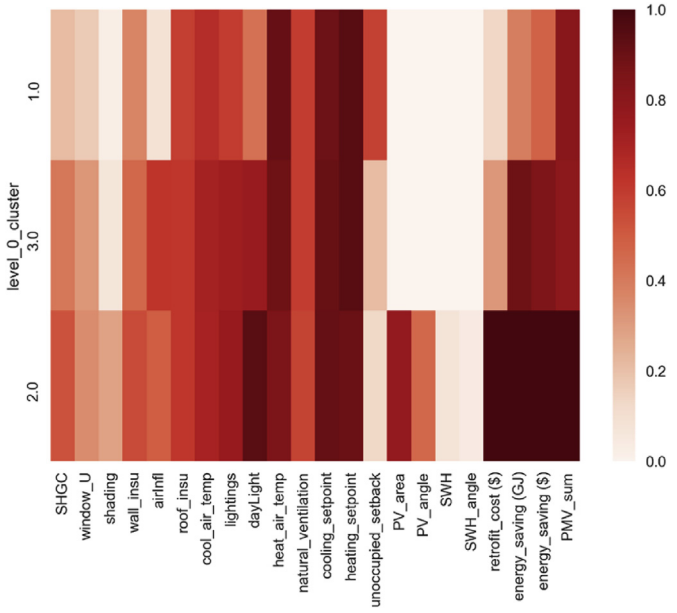

Fig. 10. Heat map of the first layer clustering.

tradeoffs among objectives can be illustrated in a straightforward way. Combined with the 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. 10, each row in the heat map represents a cluster, and each column provides a comparison of the means of each decision variables and each objective 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 3.1.2.5, including both energy savings in Joule and in dollar makes sense for the optimization problem because the visualized results in Fig. 9 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. 9, 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 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.

4.5. Decision-making pathway

By repeating the process in the 1st layer clustering as described in section 3.2, a pathway can be plotted in Fig. 11. In Fig. 11, 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. 11.

The pathway clearly shows how each decision is made at each

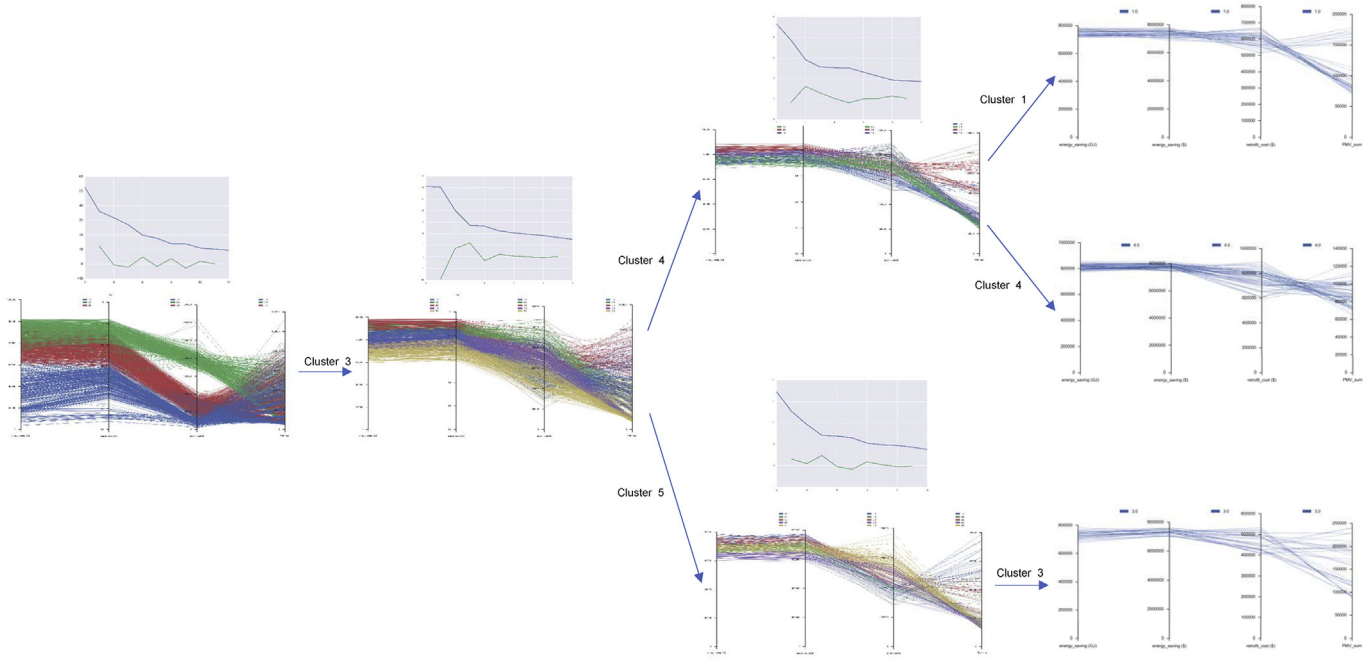

Fig. 11. Layered decision-making pathway.

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 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 9, 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 9.

It can be concluded from the results shown in Table 9 that the insulation of the building envelope in the climate of Philadelphia is important. For most of the bundles, the insulation of wall and roof is adopted except for Bundle No. 7, 8, and 12, indicating that the future winter climate in Philadelphia still demands high quality in envelope insulation. For more than half of the bundles, the setback of the thermostat during unoccupied hours is adopted. The upgrade of the lightings is also cost effective during the lifecycle of the retrofit and the lightings having 70% less heat rejection are selected even with higher initial cost. Daylighting control is favorable among the ECMs and has good economic performance. Higher percentage of ventilated window during swing season and summer is preferred when the insulation level becomes higher, which is reasonable considering that during the season with high cooling demand, the night flush and natural ventilation helps removing redundant heat from the building. Concerning the high cost related with window replacement, more than half of the bundles do not include window refurbishment in the retrofit, and windows with high insulation and lower SHGC are not so welcomed because of their economic and thermal performance in the future climate of Philadelphia. Shading is considered important during cooling season, which may be ascribed to the rise of air temperature during cooling season in the future. The optimal cooling and the heating setpoint for Towne Building are 25°C and 24°C, respectively.

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 section 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. 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 in the Pareto fronts for a building retrofit project where many ECMs are involved, are discussed. As the case study shows, the proposed decision-making support framework is manifested to show robustness in handling retrofit optimization problem and is able to provide support for brainstorming and enumerating

Table 9
Top 20 selected ECM combinations using the suggested hierarchical clustering-based decision-making support framework.

No.	SHGC	window_U (W/m²K)	shading	wall insu (m²K/W)	Infl (h-1)	roof insu (m²K/W)	cool_air_temp (°C)	lights daylight	heat_air_temp (°C)	natural vent (%)	cooling stpt (°C)	heating stpt (°C)	unocc setback	cost ($/m²)	ES (GJ)	ES ($)	Sum of PMV
1	0.62	1.6	2	1.97	0.3	1.97	11	0.3 1	31	40%	25	24	1	57.6	−853102	−8157555	85102
2	0	0	2	1.97	0.3	2.42	11	0.3 1	31	50%	25	24	1	52.1	−838175	−8039577	99441
3	0	0	1	2.69	0.3	2.42	11	0.3 1	30	50%	25	24	0	57.3	−834331	−8069318	88409
4	0.75	2.8	2	1.61	0.3	1.97	11	0.3 1	31	50%	25	24	1	55.7	−842379	−8039496	99160
5	0	0	2	2.33	0.3	2.42	0	0.3 1	32	40%	25	24	1	52.6	−834730	−7994319	98740
6	0	0	2	3.05	0.3	2.42	11	0.3 1	30	100%	25	24	1	53.5	−841017	−8003161	97708
7	0	0	2	0	0.3	1.52	11	0.3 1	28	40%	25	24	0	43.3	−803426	−7856490	110277
8	0	0	1	0	0.3	1.52	11	0.3 1	30	30%	25	24	1	43.3	−809337	−7854209	125219
9	0.75	2.8	1	2.33	0.3	1.97	11	0.3 1	28	60%	25	24	1	56.6	−847454	−8034243	97600
10	0	0	1	2.69	0.3	1.52	0	0.3 1	32	50%	25	24	0	50.3	−820412	−7930031	99152
11	0	0	0	1.97	0.3	2.87	0	0.3 1	32	60%	25	24	1	53.2	−829474	−7968637	95068
12	0.29	1.05	0	0	0.3	3.77	11	0.3 1	0	30%	25	24	0	57.7	−825286	−8017933	83148
13	0	0	0	1.97	0.7	1.97	11	0.3 1	30	50%	25	23	0	36.7	−767228	−7720925	168888
14	0	0	0	2.33	0.3	2.87	11	0.3 1	30	60%	25	24	1	53.6	−823025	−7957244	84418
15	0	0	0	1.97	0.7	3.77	11	0.3 1	32	100%	25	23	1	41.7	−780945	−7785804	171453
16	0.75	2.8	0	1.25	0.3	1.97	11	0.3 1	0	50%	25	24	0	54.6	−825351	−7964434	92416
17	0	0	0	2.33	0.3	3.32	11	0.3 1	29	60%	25	23	1	55.0	−823478	−7958761	81261
18	0.75	2.8	0	2.69	0.3	1.97	12	0.3 1	30	100%	25	24	1	57.1	−838250	−7986575	87339
19	0	0	0	1.97	0.7	2.87	11	0.3 1	29	50%	25	23	1	39.0	−768497	−7727967	158759
20	0.62	1.6	0	1.61	0.3	1.97	11	0.3 1	31	70%	25	24	0	57.9	−830025	−7971569	85822

various possibilities during decision-making process.

5. Conclusions

In this paper, the multi-objective optimization problem usually encountered in existing building 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 non-dominated sorting differential evolution algorithm (NSDE) is implemented 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 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. The method suffices for achieving the goal of comparative parametric study in building retrofit problems.

Moreover, the optimization process used in this research can be used to find optimal solution 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. The suggested decision-making support method is successfully applied to the case study building in UPenn and generated 20 sets of ECMs that can be potentially of interest to the future retrofitting.

Appendix A. Thermal modeling in SimBldPy

The main variables in the modeling method include 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^{2}K), $H_{tr,w}$ (transmission heat transfer coefficient of windows and glazed walls in W/m^{2}K), $H_{tr,em}$, $H_{tr,ms}$ (transmission heat transfer coefficient of the internal structure and external structure, respectively, in W/m^{2}K), $H_{tr,ve}$, $H_{tr,is}$ (transmission heat transfer coefficient of ventilated air, and that between the air in the building and internal structures, respectively, in W/m^{2}K). 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:

$$\theta_{air}\left(H_{tr,is}+H_{ve}\right)=H_{tr,is}\theta_s+\ H_{ve}\theta_{sup}+\ \Phi_{HC}+\Phi_{ia} \tag{11}$$

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 $\Phi_{ia,}$ $\Phi_{st,}$ Φ_m, respectively, in W. The temperature variables for the model are $\theta_{e,}$ $\theta_{air,}$ $\theta_{m,}$ θ_s $\theta_{sup,}$ $\theta_{H,set,}$ $\theta_{C,\ set}$, 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 $^\circ C$.

The zone thermal 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, the governing equation for central node and internal mass node are:

$$C_m\frac{d\theta_m}{dt}+\left(\frac{1}{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}}+H_{tr,w}}+\frac{1}{H_{tr,ms}}}+H_{tr,em}+\sum H_{if,\ i}\right)\theta_m=\Phi_m+\theta_eH_{tr,em}+\sum\theta_{az,i}H_{if,\ i}+\frac{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}}+H_{tr,w}+\sum H_{iw,\ i}}}{\frac{1}{\frac{H_{ve}H_{tr,is}}{H_{ve}+H_{tr,is}}+H_{tr,w}+\sum H_{iw,\ i}}+\frac{1}{H_{tr,ms}}}\left(\Phi_{st}+\theta_eH_{tr,w}+\sum\theta_{az,i}H_{iw,\ i}+\frac{1}{\frac{1}{H_{tr,is}}+\frac{1}{H_{ve}}}\left(\theta_{sup}+\frac{\Phi_{ia}+\Phi_{HC}}{H_{ve}}\right)\right) \tag{13}$$

$$H_{tr,ms}\theta_m + \Phi_{st} + \theta_e H_{tr,w} + \sum \theta_{az,i} H_{iw,\,i} + \frac{1}{\frac{1}{H_{tr,is}} + \frac{1}{H_{ve}}} * \left(\theta_{sup} + \frac{\Phi_{ia} + \Phi_{HC}}{H_{ve}}\right) = \theta_s \left(H_{tr,ms} + H_{tr,w} + \sum H_{iw,\,i} + \frac{1}{\frac{1}{H_{tr,is}} + \frac{1}{H_{ve}}}\right) \quad (12)$$

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. $H_{tr,iw}$ 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.

In this modeling method, a concept of free-floating air temperature $\theta_{air,\,free}$ is introduced 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.

Appendix B. Morphing of future hourly weather data.

In the morphing method, there is a "baseline" where the "morphing" starts from and the baseline refers to the TMY3 weather data. The first step of this method is to calculate the mean value for each climate variable of each month m for the baseline scenario, the baseline climate value of x_0 for month m is defined to be:

$$\left\langle x_0 \right\rangle_m = \frac{1}{24 \times d_m} \sum_{month\ m} x_0 \quad (14)$$

where d_m is the number of days in month m and the 24 comes from averaging the hourly measurements over the 24 h of each day.

The morphing method adopted here includes three operations, which can be described as: 1) a shift; 2) a linear stretch (scaling factor); 3) a shift and a stretch, the following equations demonstrates the three operations:

$$x = x_0 + \Delta x_m \quad (15)$$

$$x = \alpha_m x_0 \quad (16)$$

$$x = x_0 + \Delta x_m + \alpha_m (x_0 - \langle x_0 \rangle_m) \quad (17)$$

where xo is the existing hourly climate variable, Δx_m is the absolute change in monthly mean climate variable for month m (which is obtained from GCM outcome), α_m is the fractional change in monthly mean climate variable for month m, and x_{0m} is the climate variable xo average over month m.

The adding of absolute change in monthly mean climate variable for certain month is called shift as described by Equation (15). It indicates the mean value in baseline scenario experiences an absolute change, like the change in atmospheric pressure. A stretch is used when the change of certain variable is embodied as fractional change rather than absolute increment, like solar radiation which can never be a positive number at night (Equation (16)). The combination of stretch and shift can be applied to those variables like dry-bulb temperature, where both fractional diurnal change and absolute increment take place, especially when taking into account the changes in both maximum and minimum daily temperatures (Equation (17)). After applying the stretch and/or shift changes to specific climate variables, the future hourly weather data can be morphed on the basis of current TMY hourly weather data.

References

[1] Aktas CB, Bilec M. Impact of lifetime on US residential building LCA results. Int J Life Cycle Assess 2011;17(3):337–49.
[2] (EIA) EIA. Commercial sector demand module of the national energy modeling system: model documentation 2009. Washington D.C.: U.S. Department of Energy; 2009.
[3] Zhai J, Bendewald M, Hammer N. Deep energy retrofit of commercial buildings: a key pathway toward low-carbon cities. Carbon Manag 2011;2(4):425–30.
[4] Levine MD, Feng Wei, Ke Jing, Hong Tianzhen, Zhou N. A retrofit tool for improving energy efficiency of commercial buildings. In: Levine MD, editor. Berkeley, CA: Lawrence Berkeley National Laboratory; 2014.
[5] USEPA. Building and their impact on the environment: a statistical summary. 2009.
[6] J.D. Ryan AN. Commercial building R&D program multi-year planning: opportunities and challenges. Conference Commercial building R&D program multi-year planning: opportunities and challenges. p. 307-319.
[7] Mills E. Building commissioning: a golden opportunity for reducing energy costs and greenhouse gas emissions in the United States. Energy Efficiency 2011;4(2):145–73.
[8] Mills HF E, Powell T, Claridge D, Haasl T, Piette MA. The costeffectiveness of commercial-buildings commissioning: a meta-analysis of energy and non-energy impacts in existing buildings and new construction in the United States. Lawrence Berkeley National Laboratory; 2004.
[9] Noris F, Adamkiewicz G, Delp WW, Hotchi T, Russell M, Singer BC, et al. Indoor environmental quality benefits of apartment energy retrofits. Build Environ 2013;68:170–8.
[10] Wang B, Xia X. Optimal maintenance planning for building energy efficiency retrofitting from optimization and control system perspectives. Energy Build 2015;96:299–308.
[11] Wang B, Xia X, Zhang J. A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. Energy Build 2014;77:227–35.

[12] Jafari A, Valentin V. An optimization framework for building energy retrofits decision-making. Build Environ 2017;115:118–29.
[13] Shao Y, Geyer P, Lang W. Integrating requirement analysis and multi-objective optimization for office building energy retrofit strategies. Energy Build 2014;82:356–68.
[14] Asadi E, da Silva MG, Antunes CH, Dias L. A multi-objective optimization model for building retrofit strategies using TRNSYS simulations, GenOpt and MATLAB. Build Environ 2012;56:370–8.
[15] Rysanek AM, Choudhary R. A decoupled whole-building simulation engine for rapid exhaustive search of low-carbon and low-energy building refurbishment options. Build Environ 2012;50:21–33.
[16] Eisenhower B, O'Neill Z, Narayanan S, Fonoberov VA, Mezić I. A methodology for meta-model based optimization in building energy models. Energy Build 2012;47:292–301.
[17] Mauro GM, Hamdy M, Vanoli GP, Bianco N, Hensen JLM. A new methodology for investigating the cost-optimality of energy retrofitting a building category. Energy Build 2015;107:456–78.
[18] Chidiac SE, Catania EJC, Morofsky E, Foo S. A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings. Energy Build 2011;43(2):614–20.
[19] Asadi E, da Silva MG, Antunes CH, Dias L. Multi-objective optimization for building retrofit strategies: a model and an application. Energy Build 2012;44(0):81–7.
[20] Siddharth V, Ramakrishna PV, Geetha T, Sivasubramaniam A. Automatic generation of energy conservation measures in buildings using genetic algorithms. Energy Build 2011;43(10):2718–26.
[21] Roberti F, Oberegger UF, Lucchi E, Troi A. Energy retrofit and conservation of a historic building using multi-objective optimization and an analytic hierarchy process. Energy Build 2017;138:1–10.
[22] Tadeu S, Rodrigues C, Tadeu A, Freire F, Simões N. Energy retrofit of historic buildings: environmental assessment of cost-optimal solutions. Journal of Building Engineering 2015;4:167–76.
[23] Son H, Kim C. Evolutionary multi-objective optimization in building retrofit planning problem. Procedia Eng 2016;145:565–70.
[24] Wu R, Mavromatidis G, Orehounig K, Carmeliet J. Multiobjective optimisation of energy systems and building envelope retrofit in a residential community. Appl Energy 2017;190:634–49.
[25] Asadi E, Silva MGD, Antunes CH, Dias L, Glicksman L. Multi-objective optimization for building retrofit: a model using genetic algorithm and artificial neural network and an application. Energy Build 2014;81:444–56.
[26] Rysanek AM, Choudhary R. Optimum building energy retrofits under technical and economic uncertainty. Energy Build 2013;57:324–37.
[27] Chantrelle FP, Lahmidi H, Keilholz W, Mankibi ME, Michel P. Development of a multicriteria tool for optimizing the renovation of buildings. Appl Energy 2011;88(4):1386–94.
[28] Pombo O, Allacker K, Rivela B, Neila J. Sustainability assessment of energy saving measures: a multi-criteria approach for residential buildings retrofitting—a case study of the Spanish housing stock. Energy Build 2016;116:384–94.
[29] Ascione F, Bianco N, De Masi RF, Mauro GM, Vanoli GP. Resilience of robust cost-optimal energy retrofit of buildings to global warming: a multi-stage, multi-objective approach. Energy Build 2017;153:150–67.
[30] LBNL. Energyplus (version 8.5.0) [software]. Lawrence Berkeley National Laboratory; 2015. http://www.eere.energy.gov/buildings/energyplus/.
[31] Klein SA. TRNSYS 17: a transient system simulation program. University of Wisconsin. Madison, USA: Solar Energy Laboratory; 2010. al. e.
[32] Hirsch JJ. eQuest. In: Laboratory LBN, editor. U.S; 2016.
[33] DOE. The encyclopedic reference to EnergyPlus input and output, 8.0. EnergyPlus Development Team; 2014.
[34] Olsen EL, Chen Q. Energy consumption and comfort analysis for different low-energy cooling systems in a mild climate. Energy Build 2003;35(6):560–71.
[35] Shen P. Impacts of climate change on U.S. building energy use by using downscaled hourly future weather data. Energy Build.
[36] Xu P, Huang YJ, Miller N, Schlegel N, Shen P. Impacts of climate change on building heating and cooling energy patterns in California. Energy 2012;44(1):792–804.
[37] Chan ALS. Developing future hourly weather files for studying the impact of climate change on building energy performance in Hong Kong. Energy Build 2011;43(10):2860–8.
[38] Belcher SE, Hacker JN, Powell DS. Constructing design weather data for future climates. Build Serv Eng Res Technol 2005;26(1):49–61.
[39] Berthoua T, Stabata P, Salvazetb R, Marchioa D. Development and validation of a gray box model to predict thermal behavior of occupied office buildings. Energy Build 2014;74:91–100.
[40] ISO. ISO 13790. Energy performance of buildings - calculation of energy use for space heating and cooling. 2008.
[41] Shen P, Braham W, Yi Y. Development of a lightweight building simulation tool using simplified zone thermal coupling for fast parametric study. Appl Energy 2018;223:188–214.
[42] Shen P, Braham W, Yi Y. The feasibility and importance of considering climate change impacts in building retrofit analysis. Appl Energy 2019;233–234:254–70.
[43] IPCC. Climate change 2013– the physical science basis, fifth assessment report of the IPCC. In: Stocker T F, Qin D, editors; 2013. New York.
[44] Shen P. TThe feasibility and importance of considering the impacts of climate change on the energy use in building retrofit analysis. Department of Architecture, University of Pennsylvania; 2018.
[45] Jaggs M, Palmer J. Energy performance indoor environmental quality retrofit — a European diagnosis and decision making method for building refurbishment. Energy Build 2000;31(2):97–101.
[46] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 2002;6(2):182–97.
[47] Hamdy M, Hasan A, Siren K. A multi-stage optimization method for cost-optimal and nearly-zero-energy building solutions in line with the EPBD-recast 2010. Energy Build 2013;56:189–203.
[48] Storn R, Price K. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. J Global Optim 1997;11(4):341–59.
[49] Abbass HA, Sarker R, Newton C. PDE: a Pareto-frontier differential evolution approach for multi-objective optimization problems. Conference PDE: a Pareto-frontier differential evolution approach for multi-objective optimization problems, vol. vol. 2. p. 971-978 vol. 2.
[50] Pebesma EJ, Heuvelink GB. Latin hypercube sampling of Gaussian random fields. Technometrics 1999;41(4):303–12.
[51] Mitchell M. An introduction to genetic algorithms1998.
[52] Johnson SC. Hierarchical clustering schemes. Psychometrika 1967;32(3):241–54.
[53] Hansen P, Jaumard B. Cluster analysis and mathematical programming. Math Program 1997;79(1):191–215.
[54] Robles F. International market entry strategies and performance of United States catalog firms. J Direct Mark 1994;8(1):59–70.
[55] Kamrani AK, Parsaei HR, Chaudhry MA. A survey of design methods for manufacturing cells. Comput Ind Eng 1993;25(1–4):487–90.
[56] Giuliano G, Small KA. Subcenters in the Los Angeles region. Reg Sci Urban Econ 1991;21(2):163–82.
[57] Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. ACM; 1998.
[58] Kriegel H-P, Kröger P, Zimek A. Clustering high-dimensional data: a survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Trans Knowl Discov Data 2009;3(1):1.
[59] Parsons L, Haque E, Liu H. Subspace clustering for high dimensional data: a review. Acm Sigkdd Explorations Newsletter 2004;6(1):90–105.
[60] Tadesse MG, Sha N, Vannucci M. Bayesian variable selection in clustering high-dimensional data. J Am Stat Assoc 2005;100(470):602–17.
[61] Blashfield RK. Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods. Psychol Bull 1976;83(3):377.
[62] Müllner D. Modern hierarchical, agglomerative clustering algorithms. 2011. arXiv preprint arXiv:11092378.
[63] Sneath PH. The application of computers to taxonomy. Microbiology 1957;17(1):201–26.
[64] Ward Jr JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963;58(301):236–44.
[65] Guha S, Rastogi R, Shim K. CURE: an efficient clustering algorithm for large databases. Conference CURE: an efficient clustering algorithm for large databases, vol. vol. 27. ACM, p. 73-84.
[66] Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. J Intell Inf Syst 2001;17(2):107–45.
[67] Thorndike RL. Who belongs in the family? Psychometrika 1953;18(4):267–76.
[68] Shen P, Lior N. Vulnerability to climate change impacts of present renewable energy systems designed for achieving net-zero energy buildings. Energy 2016;114:1288–305.
[69] DOE. Building component cost community. DOE; 2015.
[70] Laboratory NRE. National residential efficiency measures database. In: Laboratory NRE, editor. Golden, CO2013.
[71] Monteiro H, Fernández JE, Freire F. Comparative life-cycle energy analysis of a new and an existing house: the significance of occupant's habits, building systems and embodied energy. Sustain Citie Soc 2016;26:507–18.
[72] Kokogiannakis G, Clarke J, Strachan P. Impact of using different models in practice-a case study with the simplified methods of ISO 13790 standard and detailed modelling programs. Int Build Perfo Simul Assoc (IBPSA) 2007:39–46.
[73] Erbs DG, Klein SA, Duffie JA. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Sol Energy 1982;28(4):293–302.
[74] Agrawal B, Tiwari GN. Life cycle cost assessment of building integrated photovoltaic thermal (BIPVT) systems. Energy Build 2010;42:1472–81.
[75] Yoza A, Yona A, Senjyu T, Funabashi T. Optimal capacity and expansion planning methodology of PV and battery in smart house. Renew Energy 2014;69:25–33.
[76] Pearson K. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magaz J Sci 1901;2(11):559–72.
[77] Lvd Maaten, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9(Nov):2579–605.