Exploring data-driven building energy-efficient design of envelopes based on their quantified impacts

Zhichao Tian, Xing Shi, Sung-Min Hong

PII: S2352-7102(21)00876-7

DOI: https://doi.org/10.1016/j.jobe.2021.103018

Reference: JOBE 103018

To appear in: *Journal of Building Engineering*

Received Date: 24 January 2021

Revised Date: 17 July 2021

Accepted Date: 23 July 2021

Zhichao Tian: Conceptualization; Formal analysis; Validation; Visualization; Writing- original draft.
Xing Shi: Data curation; Funding acquisition; Project administration; Resources; Supervision.
Sung-Min Hong: Supervision; Writing- review & editing.

Exploring data-driven building energy-efficient design of envelopes based on their quantified impacts

Abstract

Building performance design plays a key role in reducing the energy consumption of buildings. However, the widely used simulation-based design is facing several challenges, such as the labor-intensive modeling process and the performance gaps between design stage estimations and operational energy use. For these reasons, artificial intelligent methods are expected by designers to improve the efficiency and reliability of building energy-efficient design. To date, there has not been a practical data-driven design method of envelopes. This study aimed at exploring data-driven building energy-efficient design of envelopes based on their quantified impacts. A feature selection method and a game-theoretic method were applied to quantify the impacts of envelopes on space heating and cooling energy, which were performed on two building datasets, one of which is from the U.S. and the other from China. Random forest classifiers were developed to conduct the study. Based on discovered energy patterns and quantified impacts of envelopes on energy consumption, a rectified linear design method of envelopes was proposed with the idea of improving the performance of high-impact envelopes. Besides, a validation study was conducted on two office buildings in the hot-summer cold-winter region. To design the envelopes of a building, the data-driven analysis was driven by its similar buildings other than the whole dataset. Moreover, a detailed energy simulation was conducted to evaluate the energy performance of different design solutions. The results showed that compared with baseline design solutions, new strategies could save 1.05% to 21.2% energy for space heating and cooling for these two case buildings. The proposed method is a general building envelope design approach and allows designers to easily find an energy-efficient configuration of envelopes. This study demonstrated the feasibility and effectiveness of the data-driven energy-efficient design of building envelopes.

Keywords:

Building envelopes, data-driven, energy-efficient design.

Abbreviations and Nomenclature

AIRTGT	The airtight level of fenestration
AREA	Building area
BEDID	Building Energy Design Information Dataset
CBECS	Commercial Building Energy Consumption Survey
CELV	Cooling energy-efficient level
CLLD	Cooling load
CLMT	Climate zone
DDBED	Data-driven building energy-efficient design
GLSSPC	Percent exterior glass
HDD65	Heating degree days
HELV	Heating energy-efficient level
HSCW	Hot-summer-cold-winter
HTLD	Heating load
NFLOOR	Number of floors
NFLOOR	Number of floors
PFCN	Principle function
PUBCLIM	Climate region
RFCNS	Roof construction
SC	Solar coefficient
SC	Solar coefficient
SHAP	SHapley Additive exPlanations
SHGC	solar heat gain coefficient
SQFT	Square footage
URF	U-value of roofs
UWIN	U-value of windows
UWLL	U-value of exterior walls
WINTYP	Window glass type
WLCNS	Wall construction
WWRE	WWR east orientation
WWRN	WWR north orientation
WWRS	WWR south orientation
WWRW	WWR west orientation

1. Introduction

For decades, the world has recognized the necessity of improving building energy efficiency to achieve sustainable development. Up to now, green building rating programs have been adopted to construct green buildings [1]. The energy efficiency of buildings has been emphasized in those programs by weighting more scores on energy-related measures and setting baseline performance for the passive components [2]. In the early design stage, building envelopes, as the main part of passive systems, determine the heating and cooling demands of a building, and affect the construction cost and building operation energy consumption. Thus, stricter energy standards and regulations have been formulated, especially on envelopes. For instance, the latest national energy efficiency standard for public buildings in China poses higher requirements than the last generation [3]. Besides, researchers have developed several building energy-efficient design methods. The heating- and cooling-loads-based design and the simulation-based design are two primary methods in this field [4]. The former focuses on reducing the heating and cooling loads rather than the actual operational energy use. While for the latter, the building energy modeling process is time-consuming and labor-intensive and has a steep learning curve [5, 6]. In addition, these performance gaps involving discrepancies between design stage estimations and operational energy use jeopardize the reliability of the simulation-based design [7-9]. Although researchers are dedicating to remedying the shortcomings of this method [10, 11], there still exists a strong demand for effective and smart methods.

Recently, data-driven building energy-efficient design (DDBED) is gaining much attention with the establishment of many building energy databases, such as the building performance database [12, 13] and the Commercial Building Energy Consumption Survey (CBECS) [14]. These databases make it possible to analyze building energy from the perspectives of big-data and data-mining, rather than traditional energy simulation programs, in which it is difficult to include actual operation circumstances, especially human behaviors [10, 11, 15]. As for building envelopes, their quantified impacts can conduce to intuitive knowledge for designers before conducting energy-efficient design. To the authors' knowledge, the impacts of envelopes on energy use have not been quantified with the data-driven methods based on a large amount of building energy data till now.

Data-driven methods can be applied to select heating, ventilation, and air-conditioning systems (HVAC) [16], [17]. When building data-driven models for heating energy-efficient design, Tian et al. [18] found that supervised learning can mainly be applied in the design of important designable features. Although envelope features have been utilized as input parameters in several data-driven studies, envelopes performed a little function in those models. In other words, building envelopes cannot be designed by traditional machine learning models. There are still no practical data-driven building energy-efficient design methods of building envelopes. Therefore, this study aims to explore new design approaches of envelopes with data-driven methods.

The remaining parts of this paper are organized as follows. A focused literature review is given in Section 2 to elucidate the research status of building envelope design and DDBED. Section 3 presents the methodology. The results are presented in Section 4. Discussions and Conclusions are addressed in the final two sections.

2. Literature review

2.1. Impacts of building envelopes

The impacts of envelopes on energy consumption should always be quantified before conducting energy-efficient designs. Previously, parametric or even optimization techniques have been performed to investigate the impacts of building envelopes on energy consumption [19-21]. However, most of these experiments were conducted only on a single building [19]. In other words, these studies may not be instructive to other buildings.

Data-driven analysis on a large amount of building energy use data can unveil important energy patterns [22-24]. Brom et al. [22] analyzed the energy saving effects of thermal retrofits with data of almost 90,000 retrofitted dwells in the Netherlands. They found that the energy-savings of deep renovations were lower than expectations, but could still reach the highest average values. Scofield and Doane [23] compared the energy consumption between LEED-certified school buildings and conventional school buildings in Chicago. Their findings indicated that LEED-certified buildings consumed 17% more source energy than other buildings.

Even used in several data-driven studies, however, building envelopes only exert a scarce impact on the machine learning models [18, 25]. With the 219,000 airtightness measurements mainly from residential buildings, Mélois et al. [26] analyzed the impacts of insulation, ventilation systems, and main building materials on building leakage measurement results. Their study identified influential factors of air leaks for both single-family and multi-family houses. To identify the most significant retrofit strategy, Pistore et al. [27] proposed a stepwise approach to analyze the impacts of different building envelopes, including walls, roofs, and windows. Tian et al. [18] ranked the impacts of building features on heating energy consumption for office buildings in the cold region. The results showed that heating equipment has a tremendous impact on heating energy. Bartusch et al. [28] estimated the impacts of household features on electricity with t-tests and analyses of variance. They found that the boiler and heat pump were determinant features of energy consumption. In a nutshell, the impacts of building envelopes have not been quantified specifically with data-driven methods in any study.

2.2. Design methods of envelopes

Before introducing the methodology, this section attempts to summarize existing design methods of envelopes to figure out whether data-driven methods have been adopted to design building envelopes. The design of envelopes has been explored in many studies based on numeric equations [4, 29], load-calculations [30, 31], or energy simulation programs [32]. Yu et al. [4] explored the optimum insulation thickness of the external walls for buildings with numeric heat transfer equations. Yong et al. [30] investigated the impacts of envelope design factors on heating and cooling loads of the reference building in different climate zones of the U.S. They concluded that among ten variables related to the envelope design, solar heat gain coefficient (SHGC) and window-wall-ration (WWR) are two determinant features of cooling loads under different climate zones of the U.S. Koo et al. [31] proposed a finite element model for heating and cooling demand prediction of a residential building. This study did not, however, prove that the model can apply to other residential buildings. To explore the impacts of envelopes, Košir et al. [33] conducted a series of energy simulations on a fictional building.

3. Methodology

To quantify the impacts of envelopes on energy consumption and figure out a new design method with data-driven approaches, a four-step scheme is designed, as shown in **Fig. 1**. The general design procedure of building envelopes for a design building is depicted in the dotted box on the top of Fig. 1. The first step is to select suitable datasets that contain available energy data and building envelope information. Two datasets, one from the U.S. and the other from China, are adopted to carry out the following study. To mine underline wisdom, using a group of similar buildings to drive the analysis is a brilliant strategy [18]. Therefore, only similar buildings are adopted to train and test the supervised learning models. The second step involves quantifying the impacts of building envelopes on space heating and cooling energy by explaining random forest models. Subsequently, an innovative rectified linear method is proposed by distributing more resources on high-impact features. Finally, a case study on two office buildings is conducted with the proposed method. Detained energy simulation is adopted to evaluate the effectiveness of the new design solutions, with details depicted in the following several sub-sections. On the ground that building function has a profound influence on energy consumption. This study only focuses on office buildings that account for the largest proportion of these two datasets. Due to the limitation of the features included in existing datasets, only such major components of the building envelopes as windows, walls, and roofs would be analyzed in this study.

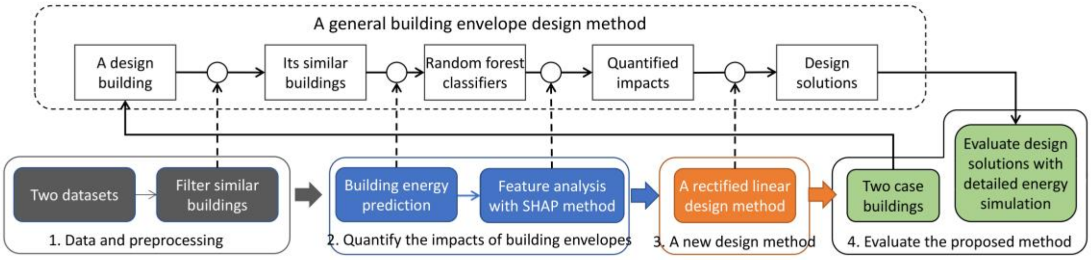
A general building envelope design method
A design building
Its similar buildings
Random forest classifiers
Quantified impacts
Design solutions
Two datasets
Filter similar buildings
1. Data and preprocessing
Building energy prediction
Feature analysis with SHAP method
2. Quantify the impacts of building envelopes
A rectified linear design method
3. A new design method
Two case buildings
Evaluate design solutions with detailed energy simulation
4. Evaluate the proposed method

Fig. 1 The scheme of data-driven design of envelopes based on their quantified impacts

3.1. Data and preprocessing

3.1.1. CBECS dataset

To investigate the energy uses of 560 million commercial buildings, the U.S. Energy Information Administration conducted large-scale surveys [14]. The CBECS dataset 2012 was generated from a survey finished in 2012. Because it has a large number of buildings and features, many of which relate to envelopes, it was used to quantify the impacts of building envelopes on space heating and cooling energy. Table 1 presents the main features used in this study. Because it does not contain the detailed performance of envelopes, such as the U-values of exterior walls, the exact values of envelopes cannot be given by traditional supervised learning models.

Table 1 Main features of CBECS used in this study

Abbreviation	Explanation	Abbreviation	Explanation
GLSSPC	Percent exterior glass	SQFT	Square footage
HDD65	Heating degree days	WINTYP	Window glass type
NFLOOR	Number of floors	WLCNS	Wall construction
PUBCLIM	Climate region	HELV	Heating energy-efficient level
RFCNS	Roof construction	CELV	Cooling energy-efficient level

3.1.2. BEDID dataset

The Building Energy Design Information Dataset (BEDID) is a customized dataset summarized from the design energy reports of about 2500 buildings in Jiangsu, China. Because those reports only recorded the design details in the design stage, the BEDID dataset does not contain on-site energy consumption but heating and cooling loads. Table 2 presents the main features of the BEDID dataset. Since building envelopes are strongly related to cooling and heating loads, this dataset provides excellent foundations for exploring their impacts on energy demands. Besides, this dataset has advantages over a detailed description of building envelopes, especially WWRs in four orientations.

Table 2 Main features of BEDID

Abbreviation	Explanation	Abbreviation	Explanation
PFCN	Principle function	WWRN	WWR north orientation
CLMT	Climate zone	UWIN	U-value of windows, W/(m^{2}K)

AREA	Building area	UWLL	U-value of exterior walls, W/(m^{2}K)
NFLOOR	Number of floors	URF	U-value of roofs, W/(m^{2}K)
SC	Solar coefficient	AIRTGT	The airtight level of fenestrations
WWRE	WWR east orientation	CLLD	Cooling Load, W/m^2
WWRW	WWR west orientation	HTLD	Heating Load, W/m^2
WWRS	WWR south orientation		

3.1.3. Data preprocessing

To conduct successful data-driven analysis, raw data need cleaning, integration, and selection. As a common practice, missing values, which refer to non-application in these datasets, were filled with 0. Based on the reasonable ranges of envelope parameters, outliers were detected and excluded. According to China's thermal design code for civil buildings, the HDD65 of HSCW ranges from 1260°F/day to 3600°F/day. Because this study only focused on office buildings in the HSCW region, buildings that satisfied this requirement were picked out.

3.1.4. Similar buildings

Data-driven building energy-efficient design aims at finding energy-efficient design solutions for buildings with a large amount of building data. In other words, data-driven algorithms can evacuate underlying design wisdom from existing buildings. Thus, buildings used to drive the analysis shall be similar to the design building. Similar buildings could be sorted out with the Euclidian distance algorithm [16]. However, the threshold of similarity of this method is difficult to define. In building fields, energy benchmarking involves selecting a group of buildings based on their function, area, year of construction, climate, and so forth [34]. Similarly, a similarity analysis was conducted by restricting the ranges of key features, including building area, climate zone, and function. Table 3 presents features and their ranges for filtering similar buildings for a design building.

Table 3 Features and their ranges for filtering similar buildings

Feature	Range
Building area, m^2	10000, 100000
Function	Office
Climate zone	HSCW
HDD65, °F/day	1260, 3600

In the early design stage, only several features have already been determined, such as climate zone, building function, and area. Thus, only these basic features and envelope features would be used in the following study, as shown in Tables 1 and 2.

3.2. Random forest classifiers

This study adopts random forest classifiers to predict the energy categories of the design buildings. As an ensemble learning, random forest combines many simple decision tree models to output an average prediction. When training, Random Forest can correct the overfitting of a decision tree. Previous studies indicate that ensemble learning algorithms, like Random Forest, outperform traditional machine learning algorithms, such as decision trees [35-38]. Although attracted huge attention, deep learning is not a good choice. For one reason, deep learning is always applied to image recognition, natural language processing, and speech recognition. For another one, compared with deep learning, ensemble learning can achieve better results under the condition that features have actual meanings [39]. In this study, Sklearn [40], a python machine learning algorithm library, was employed to construct the random forest classifiers.

To increase models' reliability, cross-validation was adopted to minimize the overfitting due to the random division of training and testing datasets [41]. K-fold cross-validation cuts the original data randomly into k equal-size parts, known as folds. Generally, k would be selected as either 5 or 10. K-fold cross-validation is popular for its effectiveness in minimizing the imperfect effect of partitioning data. In this study, K was set to 4 in the feature selection process. ROC_AUC is a magical strategy that can remedy the drawback of arbitrary positive score thresholds used by many classification algorithms. In this study, the ROC_AUC was applied as the criterion to assess the performance of the random forest classifiers.

3.3. Quantifying the impacts of building envelopes

Unearthing determinant features is a common but successful data-driven analysis in existing studies. Just as depicted in the literature review section, two types of methods, i.e., statistical and supervised learning, can be used to quantify the impacts of envelopes on energy consumption. Because statistical methods, such as Pearson Correlation Coefficient, cannot consider the combined effects of features on energy consumption, supervised learning algorithms, such as decision tree [25, 42-45] and Random Forest [27, 46], have been adopted in many studies to unveil the impacts of features on energy consumption. Feature analysis usually serves as a preprocessing or a by-product of supervised learning [44, 47]. In this study, classification learning was adopted. For one thing, it can produce high accuracy outcomes; for another, for building energy design, we want to design high-performance buildings, other than predicting the energy consumption of a building. In this study, office buildings were classified into three categories, i.e., high-efficient, medium-efficient, and low-efficient buildings. Previously, buildings were classified into 2, 3, and 5 categories in data-driven studies on building energy, in which 2 categories achieved the highest classification accuracy [47]. In this study, 2 parts of the data, i.e., high- and low-efficient ones, were adopted to train the models.

3.3.1. Feature selection

To select a subset of relevant features, feature selection is an indispensable process for supervised learning. Feature selection can help to improve prediction accuracy, build cost-effective models, and gain a thorough understanding of the data [48]. In this study, step forward feature selection was adopted to improve the prediction accuracy and select high-impact features. In each step, a feature that results in the highest prediction accuracy is picked out. This method can be employed to rank features based on the order to be selected.

3.3.2. SHAP methods

Visualization is an intuitive method to show the impacts of each feature. The decision tree allows designers to visualize the structures that are top-down generated by the splitting metric [25]. For other supervised learning algorithms that are hard to achieve visualization, Lundberg and Lee [49] proposed the SHAP (SHapley Additive exPlanations) method, a game-theoretic approach to explain the outcome of any machine learning model. SHAP provides several methods to visualize the impacts of features on the output. The SHAP method can remedy the shortcoming of feature selection that cannot quantify the impacts of features on energy consumption. This method was adopted to quantify the impacts of different envelopes on energy use. Furthermore, design solutions could be formulated based on the impacts, which would be described in the following section in detail.

3.4. A rectified linear design method of envelopes

Design solutions should also reflect the impacts of envelopes on energy consumption except complying with design standards or regulations. A solution that satisfied the basic requirements of energy standards is treated as the baseline. The performance of envelopes can be enhanced based on their impacts. Therefore, we proposed a rectified linear design method of envelopes. The core idea is to increase the performance of high-impact features linearly according to their impacts, as shown in Fig. 2. The turning point is defined by the average impacts of different envelopes. For those envelopes whose impacts are smaller than the average value, their performances can just satisfy the baseline requirements. For envelopes whose impacts exceed the average value, their performance should be improved linearly.

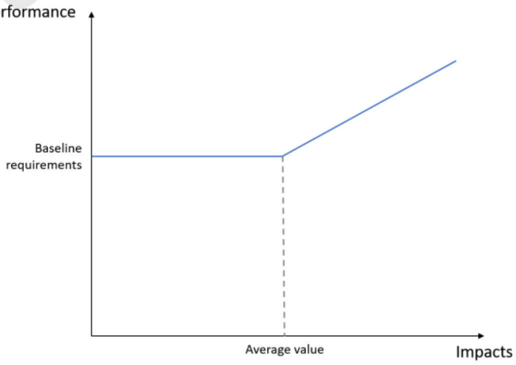

Fig. 2 A schematic diagram of the rectified linear design method

3.5. Evaluation with detailed simulation

3.5.1. Case buildings

In this study, two office buildings were selected to evaluate the effectiveness of the new design method. If the size of the exterior window system is treated as the criterion to classify office buildings, the Green Office represents a typical office building in the HSCW region. As a symbol of modern high-class buildings, the Internet High Rising has a large curtain wall system. The basic parameters of these two buildings are presented in Table 4. Since the linear distance between these two buildings is about 100 km, they have similar climate conditions. In the early design stage, the design team only has limited information about the building, such as building area, main function, and location. From this point, these two buildings are akin to each other. This study used same criteria to select similar buildings for both buildings as shown in Table 3.

Table 4 Main features of the two case buildings

Feature	Green Office	Internet High Rising
Function	Office	Office
Area, m^2	25,500	30,000
Height, m	42.9	96.8
Climate zone	HSCW	HSCW
City	Changzhou	Nanjing
Number of floors	9	22
WWRE	0.025	0.630
WWRW	0.025	0.680
WWRN	0.246	0.650
WWRS	0.234	0.660

3.5.2. Detailed energy simulation

To evaluate the energy performance of new and conventional design solutions, the EnergyPlus was utilized as the energy simulation engine. Developed by the National Renewable Energy Laboratory and several other laboratories in the U.S., EnergyPlus is a dynamic building energy simulation software for modeling various building components, including envelopes, lights, people, heating ventilation, and air-conditioning systems [50]. The 3D geometry of the model was established with the Openstudio SketchUp plugin, as shown in Fig. 3. The idea of a standard floor was employed to decrease the calculation time, as shown in Fig. 3.

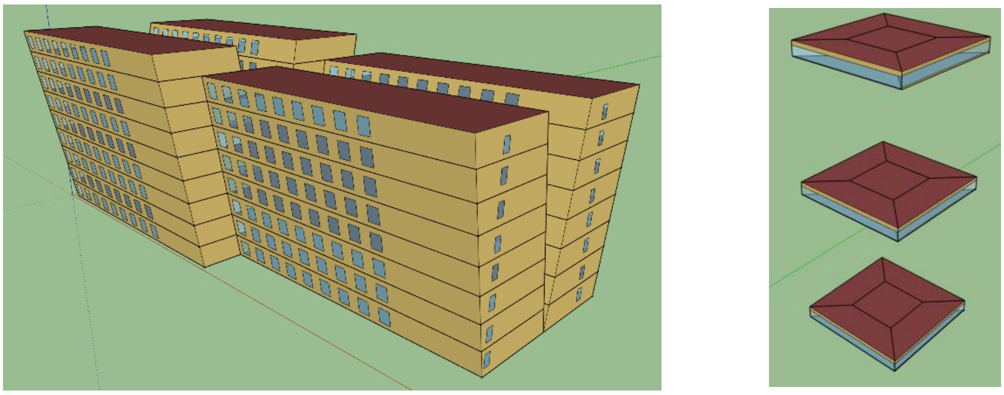

Fig. 3 EnergyPlus models of Green office (left) and Internet-High-Rising)

In the early design stage, when the design team wants to select suitable building geometry and envelopes, the HVAC systems have not been defined yet. Therefore, these two buildings are supposed to be conditioned with packaged rooftop heat pumps (PTHP). Heat pump is one of the most commonly used air-conditioning systems in office buildings in China. Other input parameters, including people, lights, equipment, and their schedules, are tuned to reflect the actual building operation status according to the on-site surveys. For each building, several detailed energy models are built to model the energy consumption of different solutions. For each building, except for the envelope features, all other parameters are kept the same.

4. Results

4.1. Impacts of envelopes

4.1.1. By the feature selection

Table 5 presents the feature selection results when predicting HLLV on two groups of similar buildings of those two case buildings. Steps 1, 2, and 3 represent the prediction of the HLLV with 1, 2, and 3 features, respectively. Both models got the highest prediction accuracy within three steps. With the CBECS dataset, SQFT is the feature with the highest impact. With the BEDID dataset, UWIN was found to have the highest impact on HLLV.

Table 5 Feature selection results on predicting HLLV

	On similar buildings from CBECS		On similar buildings from BEDID	
Step	Features	ROC_AUC	Features	ROC_AUC
1	SQFT	0.58	UWIN	0.62
2	WINTYP	0.71	UROOF	0.66
3	NFLOOR	0.74	AIRTGT	0.67

Table 6 shows the feature selection results when predicting CLLV on two groups of similar buildings from two datasets. Among many envelope parameters, GLSSPC was found to be a predominant feature by the CBECS. Whether a building is energy-efficient in cooling can almost be predicted only with the parameter of GLSSPC. While, as for the BEDID, solar coefficient is a high-impact feature. The results also indicate that the prediction accuracy on CBECS is much higher than that on BEDID.

Table 6 Feature selection results on predicting CLLV

	On similar buildings from CBECS		On similar buildings from BEDID	
Step	Features	ROC_AUC	Features	ROC_AUC
1	GLSSPC	0.83	SC	0.68
2	NFLOOR	0.86	AIRTGT	0.75
3	PUBCLIM	0.88		
4	WLCNS	0.90		

4.1.2. By the SHAP method

In this section, all the features were employed to build classifiers. The SHAP algorithm interprets those random forest models to obtain the impact of each feature on the outcome. Fig. 4 and Fig. 5 show the impacts of each feature on heating and cooling energy respectively. Each point in these figures represents a building. In these figures, features were ranked by the summation of their SHAP values. As for heating, the U-values of exterior walls have the highest impacts. As for cooling, WWR and airtightness are the high-impact features. Fig. 6 and Fig. 7 show the mean SHAP values for each feature on energy demand for space heating and cooling, respectively. Fig. 8 shows that those envelopes have weak impacts on heating energy consumption. Referring to previous studies by Tian et al.[18], the main heating equipment type has the highest impact on heating energy consumption among usual passive and active components and devices. However, results from both datasets verify that window size is a determinant feature of cooling energy consumption.

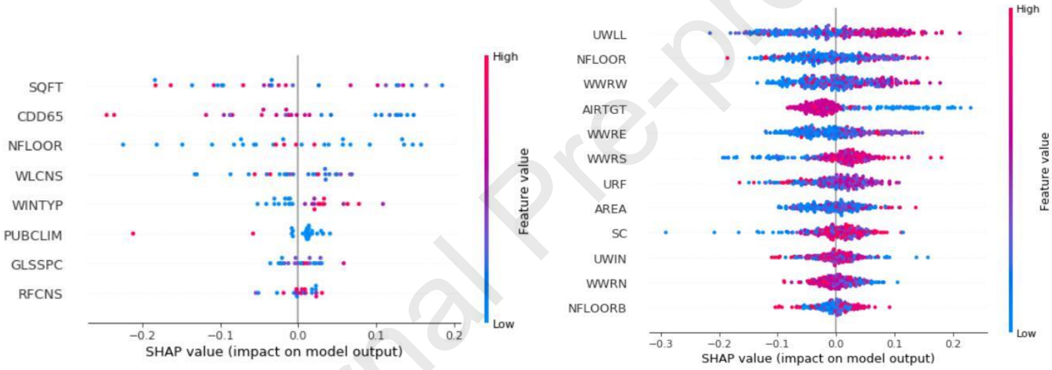

Fig. 4 Impacts of features on heating energy levels with CBECS (left) and BEDID (right)

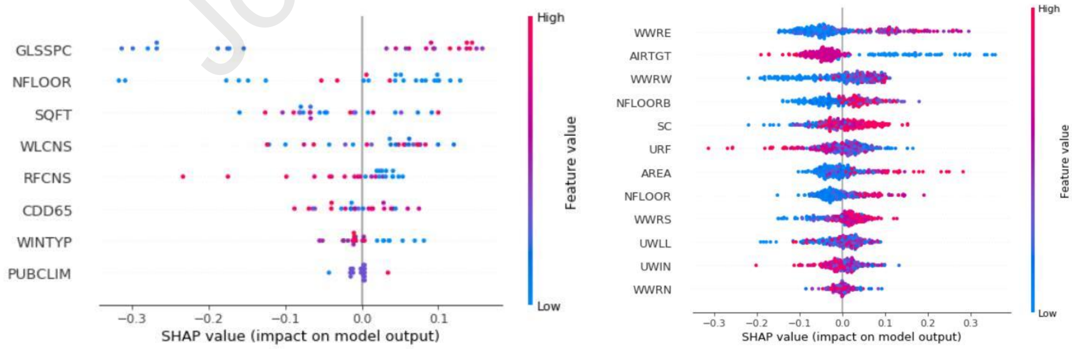

Fig. 5 Impacts of envelopes on cooling energy levels with CBECS (left), and BEDID (right)

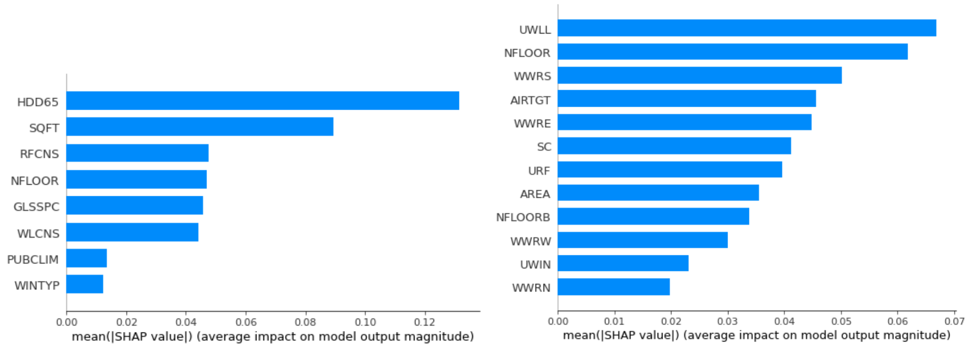

Fig. 6 Mean SHAP values of each feature for heating on CBECS (left) and BEDID (right)

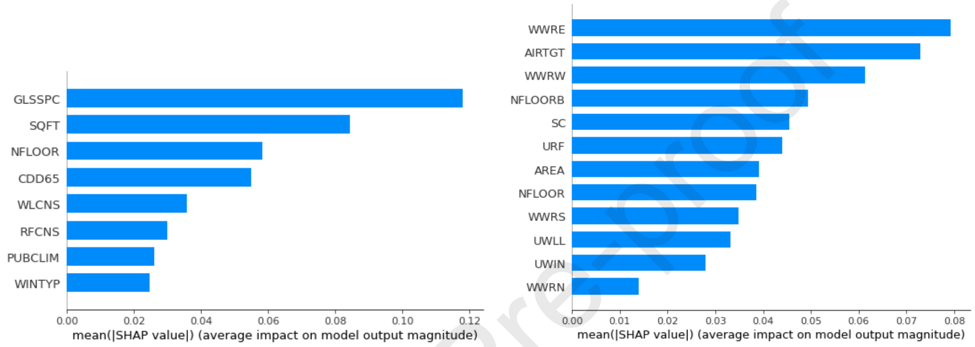

Fig. 7 Mean SHAP values of each feature for cooling on CBECS (left) and BEDID (right)

4.2. Envelope design strategies

Table 7 and Table 8 present the average SHAP impacts of building envelopes. The results indicate that the WWRs especially in the east and west orientations have tremendous impacts on energy consumption. Besides, the impacts of wall constructions are a little higher than the mean value. Thus, the design strategies should focus on the AIRTGT, UWLL, and especially WWRs.

Table 7 Impacts of envelopes on energy consumption with the CBECS dataset

Features	Heating	Cooling	Mean
GLSSPC	0.045	0.192	0.119
WLCNS	0.043	0.036	0.040
WINTYP	0.012	0.024	0.018
RFCNS	0.050	0.025	0.038
Mean	0.037	0.069	0.053

Table 8 Impacts of envelopes on heating or cooling loads with the BEDID dataset

Features	Heating	Cooling	Mean
WWRE	0.044	0.078	0.061
WWRW	0.030	0.062	0.046

WWRS	0.049	0.035	0.042
WWRN	0.019	0.013	0.016
UWIN	0.022	0.028	0.025
UWLL	0.065	0.034	0.060
URF	0.039	0.043	0.041
AIRTGT	0.045	0.072	0.059
Mean	0.039	0.046	0.042

The new design strategies conduce to improving the performance of windows and walls and keeping the performance of other envelopes on the baseline level. As for the Green Office, since it has small WWRs in four orientations, the design solution can only improve the insulation of walls. Table 9 lists the design solutions for the Green Office. Fig. 8 shows the distribution of WWRs for office buildings in the BEDID dataset. As for the Internet High Rising, because it has much larger WWRs, the design solutions can decrease the WWRs or increase the performance of windows. Thus, the first design solution is to decrease the WWRs in east and west orientations. The second design solution is to alter the window construction to the triple-panel with interior blinds window systems for windows in the east and west orientation. Table 11 shows the new design solutions for the Internet High Rising.

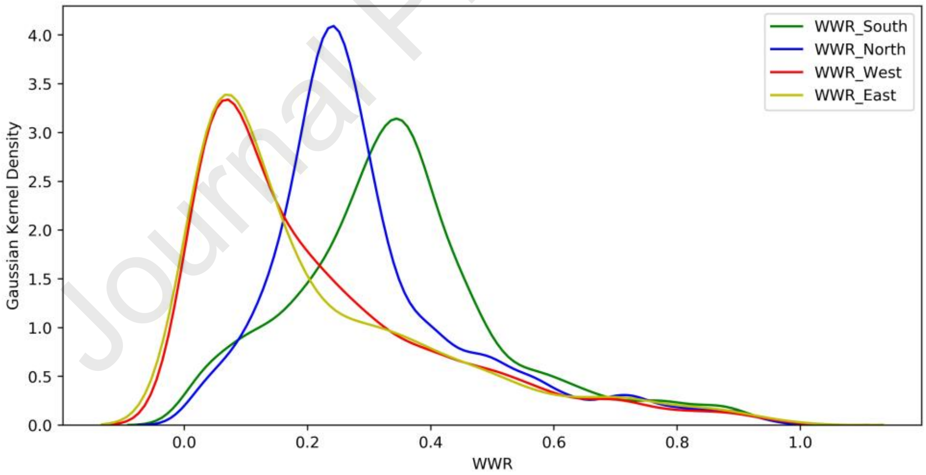

Fig. 8 Distribution of WWRs for office buildings in BEDID dataset

Table 9 Design solutions of building envelopes for the Green Office

	Baseline	Conventional	New
UWLL, W/(m^{2}K)	1.00	0.80	0.84
URF, W/(m^{2}K)	0.80	0.50	0.80
UWIN, W/(m^{2}K)	2.60	2.60	2.60
SHGC	0.497	0.497	0.497

Table 10 Design solutions of building envelopes for the Internet High Rising

	Baseline	Conventional	New 1	New 2

UWLL, W/(m^{2}K)	1.00	0.80	0.84	0.84
URF, W/(m^{2}K)	0.80	0.50	0.80	0.80
UWIN, W/(m^{2}K)	2.60	2.60	2.60	1.273
SHGC	0.704	0.704	0.704	0.525
WWRE	0.630	0.630	0.150	0.630
WWRW	0.680	0.680	0.150	0.680
WWRN	0.650	0.650	0.650	0.650
WWRS	0.660	0.660	0.660	0.660

Table 11 Window constructions of these design solutions

New solution	Conventional and baseline
6mmLow-e + 13mmAir+ 6mmClear+ 50mm (Blind inside) +6mmClear	6mmClear + 13mmAir + 6mmClear

4.3. Validation

To evaluate the reliability of the proposed design method, detailed energy simulation was adopted to evaluate the energy performances of different solutions of two case buildings. Table 12 shows the simulation results of different envelope design solutions. As for the Green Office, compared with the baseline solution, the energy saved by conventional and new solutions is less than 2%. Although installed with better wall and roof constructions, the conventional solution only saves less than 1% than the new design solution. As for the Internet High Rising, two new design solutions save 10.6% and 21.2% energy, respectively, for maintaining a comfortable indoor climate. While the conventional solution only saves less than 1%. These results indicate that for a building if the performances of these high-impact envelopes are low, the energy-saving potential will be large. Besides, the improvement of low-impact envelopes will not bring much energy-saving.

Table 12 Annual energy consumption for space heating and cooling of two case buildings

	Green Office		Internet High Rising	
	Energy, kWh/m^2	Savings	Energy, kWh/m^2	Savings
Baseline	56.4	--	84.9	--
Conventional	55.4	-1.86%	85.5	0.75%
New 1	55.8	-1.05%	76.4	-10.6%
New 2			66.9	-21.2%

5. Discussion

This study focused on exploring the impacts of different envelopes on building energy consumption and proposing a data-driven design method for building envelopes. To quantify the impacts of building envelopes, a feature selection and the SHAP approach were adopted to interpret the random forest models. Each data-driven analysis has been conducted on a group of similar buildings other than the whole dataset. As for energy prediction, data-driven models have commonly been built only with few features [37, 51]. The impacts of envelopes may be neglected because that they were not determinant features of energy consumption or the dataset does not contain envelope features [37, 46, 52, 53]. This study ranked envelope features based on their impacts. The feature selection results indicate that, with the CBECS dataset, WINTYP and GLSSPC are high-impact features for space heating and cooling energy. With the BEDID dataset, transparent features and U-values of the roof are favorable indicators for predicting heating or cooling energy consumption.

However, feature selection is weak in quantifying the contributions of each feature in predicting the energy. The results of the SHAP analysis show that envelopes exert little impact on the heating energy consumption as the mean SHAP values are smaller than 0.05. However, features related to transparent envelopes, especially WWRs, have high mean SHAP values, which indicates that they have high impacts on cooling energy. In the HSCW region, cooling energy is typically higher than heating energy and residents are less tolerant of the hot season than that of the cold season [54].

Based on the quantified impacts of envelopes on energy consumption, a data-driven design method of building envelopes was proposed. Compared with existing methods, it emphasizes the design of high-impact features. Design solutions are formulated for two case buildings. Previously, parametric analyses, even optimization techniques have been applied to find out a favorable configuration of envelopes [19]. However, the computational burden is one of those main obstacles to adopt building energy simulation and optimization methods [55]. The energy simulation results demonstrate that the improvements in low-impact features would not save much energy. By comparing the energy savings between different design solutions, it can be concluded that the improvement of high-impact features will bring much energy savings. The results also manifest the effectiveness of the proposed design method.

This study implies that a large amount of realistic building energy data can provide insights into the energy patterns of various building features. From the perspective of a policymaker, to increase the energy efficiency of buildings in a specific region, it is not necessary to pose strict requirements for all kinds of envelopes. A large scale of building data makes it possible to conduct similar analyses such as the design of building heating and cooling systems.

Before concluding, it is also necessary to expound the limitation of the proposed method. First, initial construction cost is one of the main factors considered by building owners, but it is omitted on the ground that it outreaches the scope of this study. The objective of this study is to explore the energy patterns of building envelopes with data-driven methods based on big on-site building data of office buildings, which would arouse the reconsideration of possible solutions from real-world practices, other than only with physical functions. Second, other key features, such as the heating equipment and opening of windows, are not taken into account.

The major reason is that these datasets do not contain some of the fundamental features. Future research is expected to include more building features.

6. Conclusions

Designers desire smart and effective design methods of building envelopes, instead of the time-consuming and labor-intensive simulation-based design. This study quantified the impacts of envelopes on energy use for office buildings with two data-driven methods, i.e., the feature selection method and the SHAP method. Two datasets, one from China and the other from the U.S., were employed to perform the analyses. For each building, random forest classifiers were built on a group of its similar buildings. A rectified linear design method of envelopes was proposed based on quantified impacts of envelopes on energy use. Based on the impacts of different envelopes, new design strategies can be generated. Finally, this study evaluated the performances of new design solutions with detailed energy simulation for two case buildings.

Several remarkable findings stem from the results. Quantifying the impacts of envelopes can provide designers with an intuitive understanding of the contributions of each envelope. A new design method of envelopes was proposed and successfully applied in two case buildings. It helps designers to figure out energy-efficient design solutions instead of the trial-and-error process of the simulation-based design. The proposed method is able to generate specific values of the performance of envelopes for an individual building. The results indicate that no outstanding envelope can exert a major influence on heating energy for office buildings in the HSCW region. As per the analysis of office buildings in the HSCW region as an instance, the proposed method can be applied to any building in the design stage when a number of similar buildings can be found to perform the analysis. The informative findings provide guidelines for designers and useful references for policymakers and standard-setters. To sum up, the proposed method exhibits the effectiveness, feasibility, and practicability of data-driven building energy-efficient design in the early design stage.

Acknowledgements

This paper is financially supported by the China Postdoctoral Science Foundation (Project number: 2021M692426).

Reference

[1]. Doan, D.T., A. Ghaffarianhoseini, N. Naismith, T. Zhang, A. Ghaffarianhoseini, and J. Tookey, *A critical comparison of green building rating systems.* Building and Environment, 2017. **123**: 243-260.

[2]. Chen, X., H. Yang, and L. Lu, *A comprehensive review on passive design approaches in green building rating tools.* Renewable & Sustainable Energy Reviews, 2015. **50**: 1425-1436.

[3]. China, M.o.H.a.U.-R.C.o.t.P.s.R.o., *Design standard for energy efficiency of public buildings*. 2015, China Architecture& Building Press. 2.

[4]. Yu, J., C. Yang, L. Tian, and D. Liao, *A study on optimum insulation thicknesses of external walls in hot summer and cold winter zone of China.* Applied Energy, 2009. **86**(11): 2520-2529.

[5]. Heo, Y., R. Choudhary, and G. Augenbroe, *Calibration of building energy models for retrofit analysis under uncertainty.* Energy and Buildings, 2012. **47**: 550-560.

[6]. Granadeiro, V., J.P. Duarte, J.R. Correia, and V.M. Leal, *Building envelope shape design in early stages of the design process: Integrating architectural design systems and energy simulation.* Automation in Construction, 2013. **32**: 196-209.

[7]. Turner, C. and M. Frankel, *Energy performance of LEED for new construction buildings.* New Buildings Institute, 2008. **4**: 1-42.

[8]. Calì, D., T. Osterhage, R. Streblow, and D. Müller, *Energy performance gap in refurbished German dwellings: Lesson learned from a field test.* Energy and Buildings, 2016. **127**: 1146-1158.

[9]. van den Brom, P., A. Meijer, and H. Visscher, *Performance gaps in energy consumption: household groups and building characteristics.* Building Research & Information, 2018. **46**(1): 54-70.

[10]. Zou, P.X., D. Wagle, and M. Alam, *Strategies for minimizing building energy performance gaps between the design intend and the reality.* Energy and Buildings, 2019. **191**: 31-41.

[11]. Yan, D., W. O'Brien, T.Z. Hong, X.H. Feng, H.B. Gunay, F. Tahmasebi, and A. Mahdavi, *Occupant behavior modeling for building performance simulation: Current state and future challenges.* Energy and Buildings, 2015. **107**: 264-278.

[12]. LBL. *Building Performance Database*. 2019 [cited 2019 21-Apr.]; Available from: https://bpd.lbl.gov/#explore.

[13]. Mathew, P.A., L.N. Dunn, M.D. Sohn, A. Mercado, C. Custudio, and T. Walter, *Big-data for building energy performance: Lessons from assembling a very large national database of building energy use.* Applied Energy, 2015. **140**: 85-93.

[14]. EIA. *COMMERCIAL BUILDINGS ENERGY CONSUMPTION SURVEY (CBECS)*. 2019 [cited 2019 21-Apr.]; Available from: https://www.eia.gov/consumption/commercial/data/2012/index.php?view=microdata.

[15]. Niu, S., W. Pan, and Y. Zhao, *A virtual reality integrated design approach to improving occupancy information integrity for closing the building energy performance gap.* Sustainability cites and society, 2016. **27**: 275-286.

[16]. Tian, Z., B. Si, X. Shi, and Z. Fang, *An application of Bayesian Network approach for selecting energy efficient HVAC systems.* Journal of Building Engineering, 2019. **25**: 100796.

[17]. Yamaguchi, Y., Y. Miyachi, and Y. Shimoda, *Stock modelling of HVAC systems in Japanese commercial building sector using logistic regression.* Energy and Buildings, 2017. **152**: 458-471.

[18]. Tian, Z., S. Wei, and X. Shi, *Developing data-driven models for energy-efficient heating design in office buildings.* Journal of Building Engineering, 2020. **32**: 101778.

[19]. Shi, X.J.E., *Design optimization of insulation usage and space conditioning load using energy simulation and genetic algorithm.* 2011. **36**(3): 1659-1667.

[20]. Chvatal, K.M.S. and H.J.J.o.B.P.S. Corvacho, *The impact of increasing the building envelope insulation upon the risk of overheating in summer and an increased energy consumption.* 2009. **2**(4): 267-282.

[21]. Wright, J.A., H.A. Loosemore, and R. Farmani, *Optimization of building thermal design and control by multi-criterion genetic algorithm.* Energy and Buildings, 2002. **34**(9): 959-972.

[22]. van den Brom, P., A. Meijer, and H. Visscher, *Actual energy saving effects of thermal renovations in dwellings-longitudinal data analysis including building and occupant characteristics.* Energy and Buildings, 2019. **182**: 251-263.

[23]. Scofield, J.H. and J. Doane, *Energy performance of LEED-certified buildings from 2015 Chicago benchmarking data.* Energy and Buildings, 2018. **174**: 402-413.

[24]. Streicher, K.N., P. Padey, D. Parra, M.C. Burer, and M.K. Patel, *Assessment of the current thermal performance level of the Swiss residential building stock: Statistical analysis of energy performance certificates.* Energy and Buildings, 2018. **178**: 360-378.

[25]. Lin, M., A. Afshari, and E. Azar, *A data-driven analysis of building energy use with emphasis on operation and maintenance: A case study from the UAE.* Journal of Cleaner Production, 2018. **192**: 169-178.

[26]. Melois, A.B., B. Moujalled, G. Guyot, and V. Leprince, *Improving building envelope knowledge from analysis of 219,000 certified on-site air leakage measurements in France.* Building and Environment, 2019. **159**: 106145.

[27]. Pistore, L., G. Pernigotto, F. Cappelletti, A. Gasparella, and P. Romagnoni, *A stepwise approach integrating feature selection, regression techniques and cluster analysis to identify primary retrofit interventions on large stocks of buildings.* Sustainable Cities and Society, 2019. **47**(101438).

[28]. Bartusch, C., M. Odlare, F. Wallin, and L. Wester, *Exploring variance in residential electricity consumption: Household features and building properties.* Applied Energy, 2012. **92**: 637-643.

[29]. Lin, Y.-H., K.-T. Tsai, M.-D. Lin, and M.-D. Yang, *Design optimization of office building envelope configurations for energy conservation.* Applied Energy, 2016. **171**: 336-346.

[30]. Yong, S.-G., J.-H. Kim, Y. Gim, J. Kim, J. Cho, H. Hong, Y.-J. Baik, and J. Koo, *Impacts of building envelope design factors upon energy loads and their optimization in US standard climate zones using experimental design.* Energy and Buildings, 2017. **141**: 1-15.

[31]. Koo, C., S. Park, T. Hong, and H.S. Park, *An estimation model for the heating and cooling demand of a residential building with a different envelope design using the finite element method.* Applied Energy, 2014. **115**: 205-215.

[32]. Shi, X., *Design optimization of insulation usage and space conditioning load using energy simulation and genetic algorithm.* Energy, 2011. **36**(3): 1659-1667.

[33]. Košir, M., T. Gostiša, and Ž. Kristl, *Influence of architectural building envelope characteristics on energy performance in Central European climatic conditions.* Journal of Building Engineering, 2018. **15**: 278-288.

[34]. Perez-Lombard, L., J. Ortiz, R. Gonzalez, and I.R. Maestre, *A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes.* Energy and Buildings, 2009. **41**(3): 272-278.

[35]. Hsu, D., *Identifying key variables and interactions in statistical models of building energy consumption using regularization.* Energy, 2015. **83**: 144-155.

[36]. Papadopoulos, S. and C.E. Kontokosta, *Grading buildings on energy performance using city benchmarking data.* Applied Energy, 2019. **233**: 244-253.

[37]. Deb, C., S.E. Lee, and M. Santamouris, *Using artificial neural networks to assess HVAC related energy saving in retrofitted office buildings.* Solar Energy, 2018. **163**: 32-44.

[38]. Deng, H.F., D. Fannon, and M.J. Eckelman, *Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata.* Energy and Buildings, 2018. **163**: 34-43.

[39]. Chen, T. and C. Guestrin. *Xgboost: A scalable tree boosting system*. in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.

[40]. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, *Scikit-learn: Machine learning in Python.* Journal of machine learning research, 2011. **12**(Oct): 2825-2830.

[41]. Cawley, G.C. and N.L.C. Talbot, *On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation.* Journal of Machine Learning Research, 2010. **11**: 2079-2107.

[42]. Yoon, Y.R. and H.J. Moon, *Energy consumption model with energy use factors of tenants in commercial buildings using Gaussian process regression.* Energy and Buildings, 2018. **168**: 215-224.

[43]. Zhou, X., D. Yan, T.Z. Hong, and X.X. Ren, *Data analysis and stochastic modeling of lighting energy use in large office buildings in China.* Energy and Buildings, 2015. **86**: 275-287.

[44]. Yu, Z., F. Haghighat, B.C.M. Fung, and H. Yoshino, *A decision tree method for building energy demand modeling.* Energy and Buildings, 2010. **42**(10): 1637-1646.

[45]. Martinez, A. and J.H. Choi, *Exploring the potential use of building facade information to estimate energy performance.* Sustainable Cities and Society, 2017. **35**: 511-521.

[46]. Ma, J. and J.C.P. Cheng, *Identifying the influential features on the regional energy use intensity of residential buildings based on Random Forests.* Applied Energy, 2016. **183**: 193-201.

[47]. Kuo, C.F.J., C.H. Lin, and M.H. Lee, *Analyze the the energy consumption characteristics and affecting factors of Taiwan's convenience stores-using the big data mining approach.* Energy and Buildings, 2018. **168**: 120-136.

[48]. Guyon, I. and A. Elisseeff, *An introduction to variable and feature selection.* Journal of machine learning research, 2003. **3**(Mar): 1157-1182.

[49]. Lundberg, S.M. and S.-I. Lee. *A unified approach to interpreting model predictions*. in *Advances in neural information processing systems*. 2017.

[50]. NREL. *EnergyPlus 9.2.0 Introduction* 2020 Sep.-27-2019 [cited 2020 Apr.-29].

[51]. Huebner, G., D. Shipworth, I. Hamilton, Z. Chalabi, and T. Oreszczyn, *Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes.* Applied Energy, 2016. **177**: 692-702.

[52]. Deb, C. and S.E. Lee, *Determining key variables influencing energy consumption in office buildings through cluster analysis of pre-and post-retrofit building data.* Energy and Buildings, 2018. **159**: 228-245.

[53]. Esmaeilimoakher, P., T. Urmee, T. Pryor, and G. Baverstock, *Identifying the determinants of residential electricity consumption for social housing in Perth, Western Australia.* Energy and Buildings, 2016. **133**: 403-413.

[54]. Xiong, Y., J. Liu, and J. Kim, *Understanding differences in thermal comfort between urban and rural residents in hot summer and cold winter climate.* Building and Environment, 2019. **165**: 106393.

[55]. Tian, Z., X. Zhang, X. Jin, X. Zhou, B. Si, and X. Shi, *Towards adoption of building energy simulation and optimization for passive building design: A survey and a review.* Energy and Buildings, 2018. **158**: 1306-1316.

- Smart and effective design methods of building envelopes are needed.
- The impacts of building envelopes on energy consumption are quantified with a data-mining method.
- A rectified linear design method is proposed based on quantified impacts of envelopes.
- The proposed method can be applied to any building in the design stage.
- This study exhibits the practicability of data-driven building energy efficient design of envelopes.

None