# Energy-saving potential prediction models for large-scale building: A state-of-the-art review

Yang, X, Liu, S, Zou, Y, Ji, W, Zhang, Q, Ahmed, A, Han, X, Shen, Y & Zhang, S



**Original citation & hyperlink:**
Yang, X, Liu, S, Zou, Y, Ji, W, Zhang, Q, Ahmed, A, Han, X, Shen, Y & Zhang, S 2022, 'Energy-saving potential prediction models for large-scale building: A state-of-the-art review', Renewable and Sustainable Energy Reviews, vol. 156, 111992.
https://doi.org/10.1016/j.rser.2021.111992

DOI 10.1016/j.rser.2021.111992
ISSN 1364-0321

Publisher: Elsevier

# Energy-saving Potential Prediction Models for Large-Scale Building: A state-of-the-art review

Xiu'eYang[ab]ShuliLiu[a]YuliangZou[c]WenjieJi[a]QunliZhang[de]AbdullahiAhmed[f]XiaojingHan[a]YongliangShen[a]ShaoliangZhang[a]

[a]School of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081, China
[b]School of Civil Engineering, Tangshan University, Tangshan, 063000, China
[c]Campus Planning and Asset Management Division; Beihang University, China
[d]Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China
[e]Beijing Key Lab of Heating, Gas Supply, Ventilating and Air Conditioning Engineering, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China
[f]Institute for Future Transport and Cities, Coventry University, CV1 5FB, UK

**Abstract:** Energy consumption prediction models for large-scale building reveal energy use pattern, which play an irreplaceable role in formulating energy policy and developing building industry. The existing review papers on energy consumption models for large-scale building mainly focus on approaches for building energy consumption prediction, but rarely summarize models used to predict energy-saving effect of large-scale buildings retrofit. . Therefore, this paper reviews the energy consumption models for predicting energy saving potential of building in large-scale. The advantages, disadvantages and accuracy of the models have been analyzed, and future research direction of these models have been discussed. The results reveals three types of approaches, including data-driven, physics-based, and hybrid approaches which can be used for predicting energy saving potential in large-scale buildings. Some problems have been solved in the existing models, such as building physical parameters not included in data-driven models, lack of geometric parameter and time consuming requirement for input parameters in physical-based model. However, there are research gaps in the prediction accuracy and prediction range which requires attention to be given to : 1) model verification; 2) model rebound effect; 3) dynamic prediction of energy saving potential; 4) residents' willingness to retrofit in the model. This paper promotes the development of models for predicting energy-saving potential for large-scale buildings, and help to formulate appropriate strategies for the retrofit of existing buildings.

**Keywords:** Large-scale building models Energy-saving effect Physical-based Data-driven Building retrofit

## 1 Introduction

Energy and environmental issues are important obstacles hindering the sustainable development of society. As the largest energy consuming sector, buildings consume over 1/3 of the overall energy consumption every year, and about the same proportion of associated greenhouse gas emissions.[1]. Meanwhile, with the continuous growth of economy, urbanization and population, building energy consumption will continue to increase. For example, it is expected to increase by two to three times by 2050 in the BRIC countries (Brazil, Russia, India and China) [2].

Therefore, building energy efficiency has become the key for achieving energy saving and low-carbon development globally. Several countries have set up the building energy saving goals. For example, the UK has set a target to reducecurrent carbon emissions levels by up to 66% by 2050

[3]. The Swiss Energy Strategy 2050 projects 64% reduction in heating energy demand [4]. American government has proposed greenhouse gas emissions reduction of 26~28% by 2025 compared to 2005 emissions level [5]. Chinese government has promised carbon dioxide emissions up to peak in approximately 2030 [6]. However, the influence of new buildings on building sector energy consumption is restricted due to a large number of existing buildings with high energy consumption [7]. For example, in the EU new buildings built after 2009 are consuming 30% to 60% less than buildings built before 1990 [8]. Therefore, it is very necessary to retrofit existing buildings for improved energy performance, which not only reduces the energy-intensity of the building, but also improves the comfort of the occupants.

Energy retrofit of the building stocks have been considered as one of the main approaches with relatively low cost and high rate of return for achieving sustainable development [9]. Many countries or regional governments， such as Danish and New York City .etc. [10,11], have begun to promote energy retrofit of building in large-scale. The building in large-scale here refers to buildings of regional scale, including neighborhood, district and cityetc. Identifying retrofit measures is one major phase of retrofit process. In this phase, reliable estimation and quantification energy saving potential are essential by using appropriate prediction model [12]. For a single building, developing physical model of building and quantifying the energy saving potential of retrofit measures using simulation engine, such as EnergyPlus, Design Builder, IESVE etc is common approach. However, the number of buildings in large-scale has been upgraded from single buildings to hundreds of thousands, or even building stocks at the national level. Therefore, it is complex to develop physical models of all the buildings in the study area. This makes prediction of energy saving potential for existing large-scale buildings a big challenge.

Currently, many energy consumption prediction models on large-scale buildings have been developed. At the same time, some papers have comprehensively reviewed the existing prediction models, as shown in Table 1. These review papers mainly focus on prediction approaches of energy consumption for buildings in large-scale, but rarely summarize models used to predict energy-saving effect of buildings retrofit in large-scale. Some models for predicting energy consumption cannot predict the energy-saving effect of building retrofit due to certain building characteristic parameters not being included. Therefore, this paper will review the energy consumption models for predicting energy saving potential of large-scale building retrofit. . The advantages, disadvantages and accuracy of the models have been analyzed, and future research direction of these models has been discussed. This paper has a role in promoting the development of prediction models for large-scale building energy retrofit, to boost the retrofit for the existing buildings. Prediction models for large-scale buildings retrofit will be introduced in the second section,, and in the third section, we will discuss characteristics, prediction accuracy of models and the influence of residents' willingness to renovate in the model, conclusions are given in the last section.

Table1 Summary of energy prediction models review

| Publication year | Research scale | Research content |
|---|---|---|
| 2008[13] | Residential sector | Reviewing modelling techniques |
| 2010[14] | Existing building stock | Comparing several bottom-up models for building stocks with the respect to their purpose, strengths and shortcomings |
| 2011[15] | --- | Reviewing multiple energy prediction modelling approaches, including time series, regression, autoregressive integrated moving average, fuzzy logic, genetic algorithm, and neural networks. |
| 2012[16] | Building | Reviewing recently developed models including physical-based, |

| | | |
|---|---|---|
| | | statistical and artificial intelligence methods. |
| 2013[17] | --- | A detailed review and discussion of simulation methods including physical statistical and hybrid method. |
| 2015[18] | Urban | Reviewing the simulation method and workflows of bottom-up building energy modeling. |
| 2017 [5] | Urban | Reviewing the basic workflow and applications of physics-based urban building energy use models. |
| 2018[19] | Building stock | Summarizing the characteristics of models for prediction energy-saving effect using archetype |
| 2018[20] | Urban and rural-level | Highlighting large-scale energy demand and data-driven prediction models |
| 2019[21] | Building | Reviewing data-driven building energy prediction models |
| 2020 [22] | Large-scale | Reviewing building energy prediction techniques for large-scale buildings |

## 2 Energy-savings prediction for building retrofit in large-scale

To estimate the energy-saving effect of buildings in large-scale, models need to have the ability to estimate the energy consumption at the pre and post retrofit phases. Energy consumption of pre retrofit phase can be estimated using multiple approaches, such as multiple linear regression, archetypes approaches, whereas estimating post retrofit building energy consumption requires that impact of new technologies and retrofit measures on building energy consumption should be evaluated. Therefore, not all prediction models of energy consumption can be used to predict energy-saving effect. Generally speaking, approaches estimating the energy-saving effect of buildings in large-scale can be broadly classified into three categories: data-driven, physics-based, and hybrid approaches[23], as shown in Figure 1.

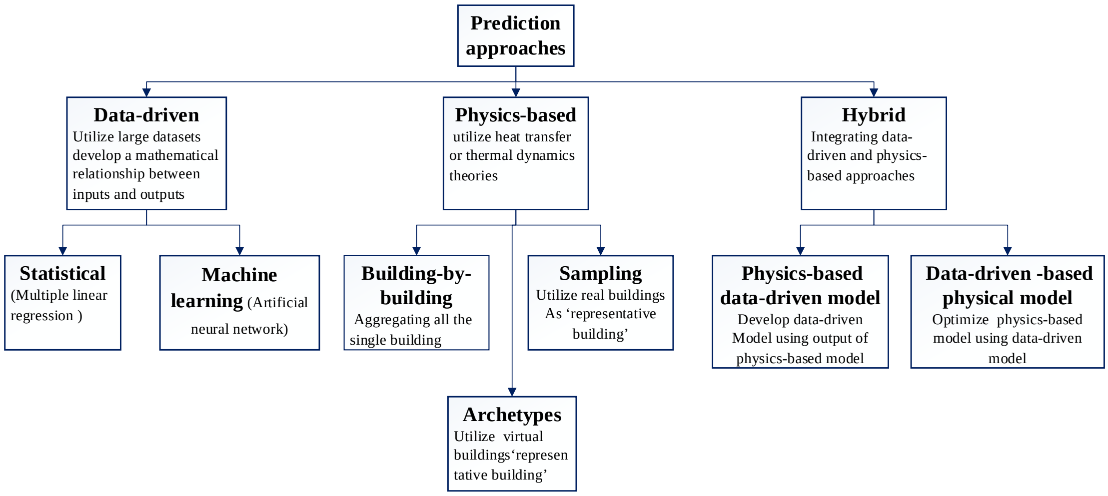


Fig.1 Approaches for estimating the energy-saving effect of building in large-scale

### 2.1 **Data-driven approaches**

Data-driven approaches are also called black-box based approaches[22]. Assuming a mathematical relationship exists between inputs (e.g. heat transfer coefficient of envelope) and outputs (e.g. space heating energy consumption), the approaches utilize large datasets, such as Building Performance Database of the United States or energy performance certificates[24], provided by public authorities (utilities and energy companies. etc.) to developed energy models. The models mainly correlate energy consumption (outputs) and influencing variables (inputs). To predict energy saving potential for large-scale buildings using data-driven approaches, buildings in a region are often divided into several types, and the total energy-saving effect for building in large-scale is obtained by aggregating the energy saving of all types of buildings in a region. Data-driven models of a type building for retrofit analysis based on large datasets can be presented as shown in Figure 2. In upper side of Figure 2, process of estimating energy consumption for buildings in large-scale is depicted. The approach which are depicted in the bottom side of Figure 2, is used to assess the impact on energy consumption due to adopting retrofit measure or renewable/alternative energy technology.

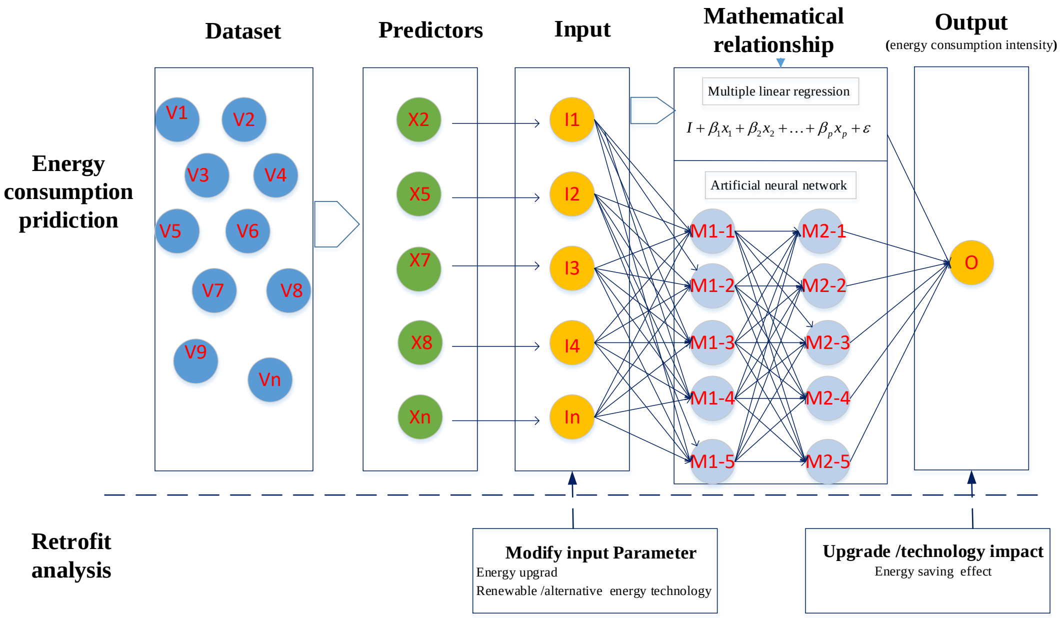


Fig.2. The steps involved in developing data-driven model for buildings retrofit

A common feature of data-driven models is to identify predictors. However, each model contains different predictors, as it was developed based on different datasets including different information. This means that some variables affecting energy consumption were not considered as predictors in the model, because relevant information is not or rarely included in the database, which lead to the model prediction inaccuracy. And for this reason, it is meaningless to compare the models developed by different studies. Unfortunately, none of the studies has solved how will the prediction results change when models ignore some factors affecting energy consumption of buildings. Therefore, increasing availability of building data and identifying predictors is the key to developing data-driven models, which directly determines the accuracy of the model.

In addition, there is limitation for data-driven models due to models trained and tested using datasets from the same sample space. The model can run the risk of being inaccurate when new inputs is out of the parameter space of the training and test inputs. This is a general limitation of building energy consumption prediction at pre-retrofit phase, whereas there is a grave limitation for post-retrofit buildings due to the absence on post-retrofit buildings performance data in the study area. Using post-retrofit datasets to train models and learn from representations of pre- and post-retrofit data can provide a solution to solve this. However, the lack of scalable measurement techniques limits the development of such data [25].

The most immediate measure for model validation is the actual energy consumption. Energy consumption of pre retrofit buildings can be verified using actual energy consumption from dataset, whereas the validation mainly focuses on an aggregated level (the average energy use intensity level for each type of building), which makes it ignore gap differences between buildings. Validation of energy consumption for post retrofit buildings is difficult due to lack of related data. This is also a major challenge for all types of modelling approaches.

The greatest advantage for data-driven models is that the occupant behavior that has a significant impact on building energy consumption can be considered [26]. Its simple application is also considered a great advantage, and is highly rated for this reason [27].

For application, not all data-driven models can predict energy saving potential of buildings, except taking characteristics parameters of buildings as predictors. In other words, data-driven model can predict the energy-saving effect of parameters that are considered as predictors. Therefore, prediction ability of data-driven models for energy saving potential is limited. With the development of large databases in recent years (data on building energy consumption, building systems, physical characteristics, equipment systems, personnel behavior, and socio-economic factors are available) the availability of building characteristics parameters and energy databases has become more and more widespread, which means that data-driven models will be an effective method to estimate the energy saving potential for building retrofit. There have been significant effort to develop multiple linear regression and artificial neural network models to predict the energy-saving effect for large scale buildings [28]. The characteristics and applications of each methods will be described as follows. Current research have also shown data-driven approache**s** have the high potential for the analysis building retrofit. In the absence of post-retrofit data and the retrofit measures implemented, some studies focused on benchmarks developed based on data-driven approache**s** using the collected data, which provide guidelines for regional energy retrofit policies. By comparing benchmarking buildings within the district and nationally, , it helps to identify suitable candidates for renovation.

**2.1.1 Multiple linear regression**

Multiple linear regression was first introduced by Galton in 1886 [22], which was used to explain the linear relationship between multiple independent variables (contribution inputs) and dependent variables (output). The model can be expressed as:**:**

$$y = I + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \tag{1}$$

Where $y$ represent dependent variable (energy consumption); $I$ represent constant; $x_i$ refers to the input variables (predictors, $i$=1,2,3,….p ); $\beta_i$ represent the regression coefficient of the input variables; $\varepsilon$ represent the random error (to measure the random difference between the y variables for all buildings and the corresponding prediction for a specific building) and remaining errors [29]. The coefficients of predictors for multiple linear regression models reflect the contribution of predictors on energy use intensity. However, there are different coefficients for the same predictor in different models due to the different predictors included, which makes it meaningless to compare the models developed by different studies. In addition, the correlation of the predictors in the model will lead to the confusion effects of the model. It reduces prediction accuracy of energy saving potential. Another drawback of the multiple linear regression model is overfitting, which leads to the prediction inaccuracy of the model. Many shrinkage methods for multiple linear regression have been developed to improve the least- squares estimator by adding constraints on the value of coefficient.

The peer group including 926 commercial buildings based on the U.S. Department of Energy Building Performance Database (870,000 buildings performance data) was determined to develop a multivariate linear regression model [30]. Occupant density, operating hours, year of construction and building typology etc. have been identified as predictors. The predicted results are shown in Figure 3. Energy saving potential of two retrofit measures have been estimated using this model**:** change the window walls to concrete and change the windows from single to double glazing. The model provides information for the stakeholders of retrofit to make decisions on retrofit investment. However, the authors also point out that energy saving potential of some measures are obviously inaccurate, such as from T12 to T8 fluorescent lights, for which negative saving appear. Lack of data is the main reason. The U.S. Department of Energy, Building Performance Database only represents less than 1% of the building stocks, and only about 5% of building contain building equipment information. In the future, with the rapid increase in the availability of building data, it is possible to obtain high-quality data to provide reliable and accurate estimation for energy saving potential using multiple linear regression model.

In the reference, the authors tried to explain the reasons for determining the predictors: considering physical intuition, variables correlation, and data availability. For example, it is intuitive that climate will impact building's energy use intensity. Operating hours is correlated with energy use intensity. Therefore, climate and operating hours have been considered as predictors. In addition, some variables, such as the number of air conditioners, temperature setting. etc. are not identified as predictors due to not being included in the database. In addition, the model also solves the problem of confounding effects of multiple parameters by increasing the number of predictors.

Although the authors tries to verify the model with the actual measured building energy intensity, the model fits well in the overall average energy use intensity. However, in terms of the individual building level, it overestimates the energy consumption of buildings with low energy use intensity and underestimates the energy consumption of buildings with high energy use intensity. An important improvement that needs to be made to verify the energy saving potential of pre and post building retrofit using detailed building information and energy intensity.

A GIS-based multiple linear regression model developed by Mastrucci et al. was introduced to estimate energy saving potential of building stocks at city scale [31]. The energy saving potential of retrofit measures, including window replacement, envelope insulation, heating and ventilation system upgrade) have been analyzed for Rotterdam. The main advantage of this model is for simple and quick energy consumption prediction in large-scale. Compared with physical-based model, lots of data input or assumptions are not required. However, the model paid no attention to the change of occupants' behaviour and the state of the building retrofit.

Streicher et al. have estimated the space heating energy consumption for Swiss dwellings built before 2000 by the Swiss residential energy model. Great energy saving potential for envelopes retrofit in large-scale has been discussed and been confirmed theoretically, such as heating energy consumption reduction up to 84% in theory through building envelope retrofit. However, actual energy-saving effect will be lower than expected from the Swiss residential energy model due to technical or social constraints [32].

Howard et al. [11] estimated the energy intensity of the New York City through a GIS-based multiple linear regression model. The building function was selected as variable in the model, ignoring the impact of building form and age. The heating, cooling and hot water energy intensity estimated can be viewed intuitively on the GIS map. Therefore, it has been identified to be an effective tool formulating energy plans for city managers.

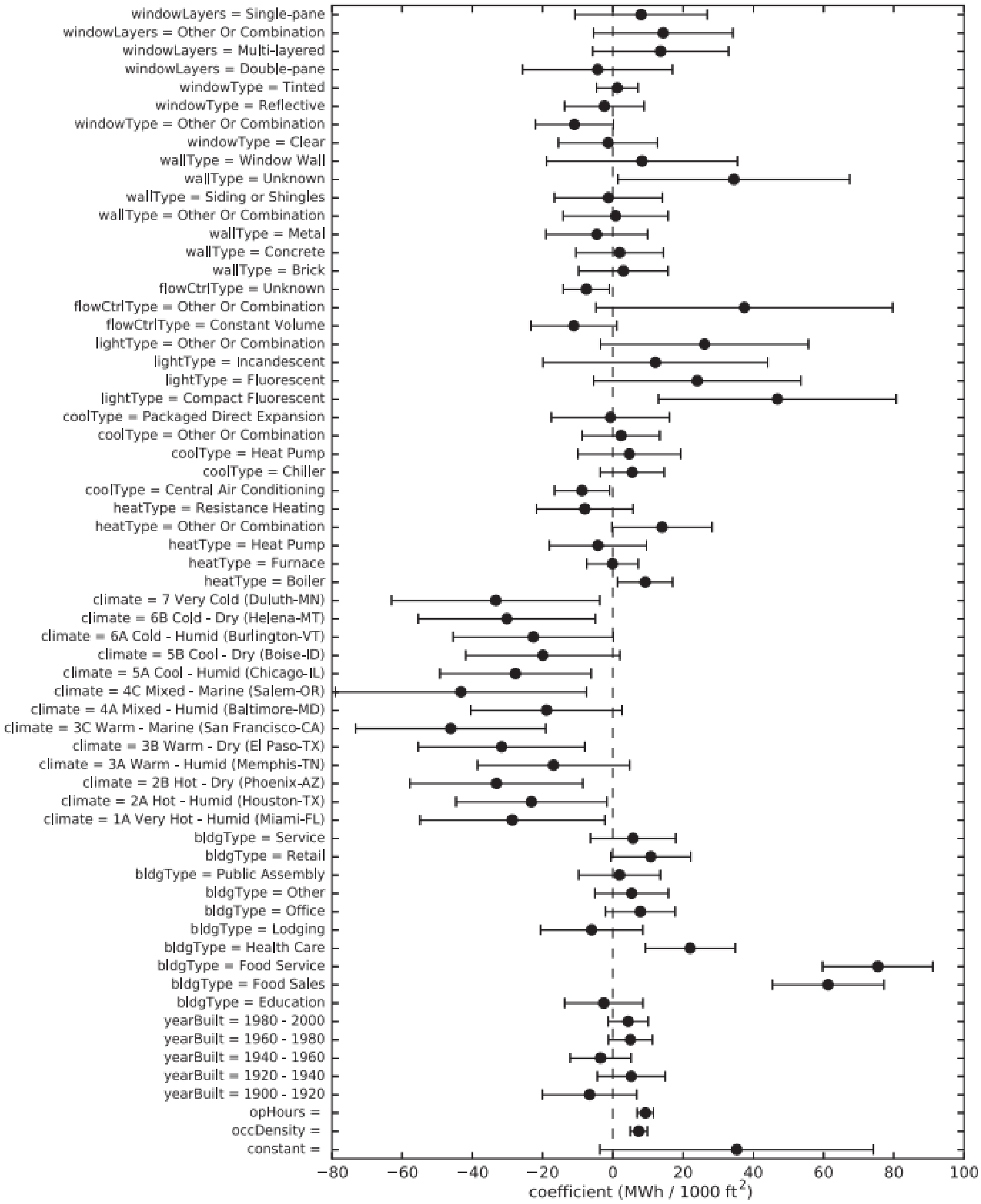


Fig.3. Coefficients of regression model [30]

2.1.2 Artificial neural network

Artificial neural network ( ANN) is an algorithm mathematical model， which use a structure similar to the connection of brain synapses to process information. The processing unit, also called neuron, is basic element of artificial neural network. Figure 4 shows a typical artificial neural network multilayer feedforward network. It consists of three layers: input layer, hidden layer, and output layer [33]. Many processing units are arranged on each layer, and every processing unit represents a specific output function, called activation function. The output of the previous layer happens to be the input of the next layer. The following Eq. (2) (3) describe function of each process units.

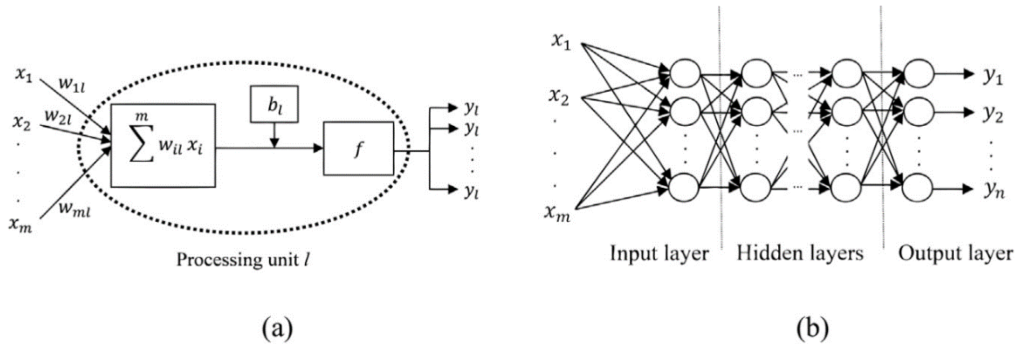


Fig.4. Schematic of artificial neural network. (a) A single process unit ; (b) Artificial neural networks. [33]

$$y_i = f\left(\sum_{i=1}^{m} w_{il} x_i + b_l\right) \tag{2}$$

$$f(x) = \frac{1}{1+e^{-x}} \tag{3}$$

Where $y$ represents output signal of process unit, $w_{il}$ represents weight coefficient, $l$ is each process unit, $b_l$ represents a bias, $f$ represents the activation function, $x_i$ represents the input signal of process unit, it is also used as input to the next layer of the artificial neural network.

The main advantage for artificial neural network model is that it uses less data to obtain reliable data, and it consumes less time to predict energy consumption. Compared with multivariate linear regression model, the artificial neural network approach has a greater advantage in predicting nonlinear relationships due to scaling and activation functions applied during the modeling process [34]. At the same time, it has been considered to successfully predict end-use in large-scale considering socioeconomic factors. However, although each processing unit has bias and coefficients, which can be seen from Eq.3, these have no physical meaning. Therefore, determining the contribution of a variable for the total energy consumption is difficult through artificial neural network approach. Several studies have used artificial neural network to predict energy consumption: from the first studies involved in the prediction of energy loads in 1990s, until more recent research on the energy saving potential of building retrofit [35]。

Using data from the Building Energy Certification open database（encompassing data about buildings energy labelling）provided by Regione Lombardia , F. Re Cecconi classified school buildings into seven categories according to the year of construction: before 1930 ,1931-1945,1946-1960,1961-1976,1977-1992,1993-2006, after 2006, and eight artificial neural network (seven for each type of building and one artificial neural network on the whole cleaned Building Energy Certification open database) have been trained to estimate the energy saving potential of building retrofit [36]. 11 parameters were selected as predictors, including 1. winter degree days; 2. year of construction; 3. gross surface area; 4. gross volume; 5. dispersant surface; 6. ratio between glazing surface and dispersant surface; 7. ratio between opaque surface and dispersant surface; 8. average U-value of walls; 9. average U-value of roof; 10. average U-value of windows; 11. average U-value of basement. The three retrofit scenarios are predicted respectively using artificial neural network. The conclusions have shown that as the number of buildings renovated increases, the energy saving potential decreases due to lowering the thresholds of retrofit parameters (walls, roof and windows

transmittance).
In the reference, Database encompassing data about buildings energy labelling in Regione Lombardia have been used to verify the model. The result shows that the difference is small. However, verification can only be done at the overall level. In addition, the influence of the trend demographic change have not been considered in the model, which is also the final factor of energy consumption. this may lead to the rebound effect. The authors have solved the problem of building surface area calculated to support the cost assessment of different retrofit measures. The energy saving potential of different retrofit scenarios in the Lombardy region can be visualized in the spatial map by the combination of artificial neural network and GIS, as shown in Figure 5, which provides visualized energy saving retrofit tool for regional energy policy makers.

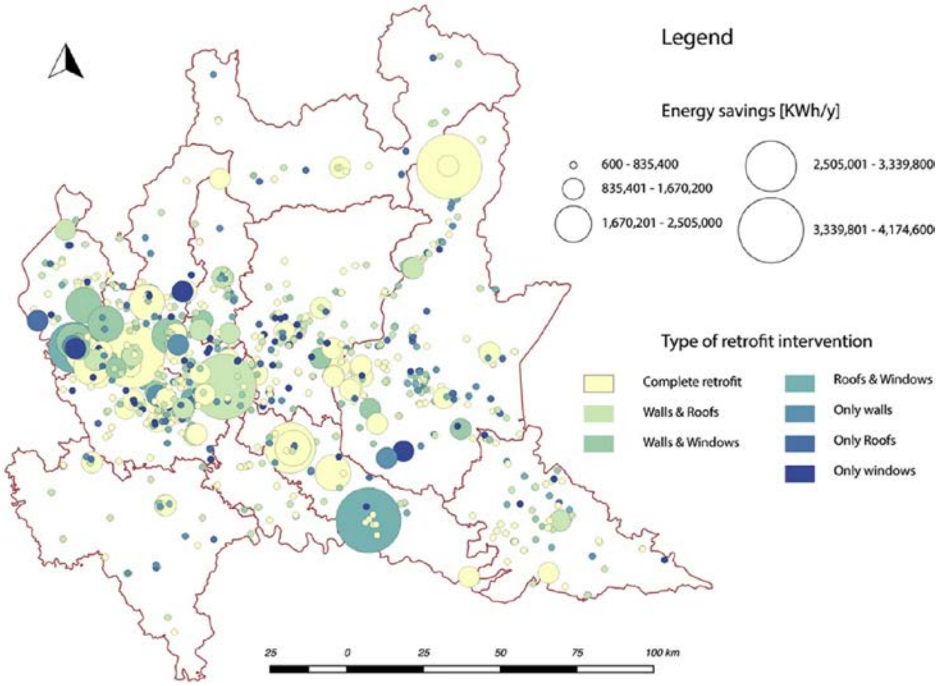


Fig.5. Retrofit scenarios visualization [36]

**2.2 Physics-based approach**

Physics-based approaches are called engineering approaches, also called as white-box approaches [28], which calculate energy consumption in the entire building or sub-level components based on heat transfer or thermal dynamics theories [16]. Figure 6 shows the steps involved in developing physics-based model for building retrofit. These approaches can be used to test the energy saving potential of any building system or technology retrofit due to energy consumption calculated based on physical principles. For example, EnergyPlus has been recently used to compare Phase Change Material (PCM) applications for building retrofit [37]. These approaches have the most flexibility and ability for modeling retrofit measures and new technologies without historical energy consumption data.

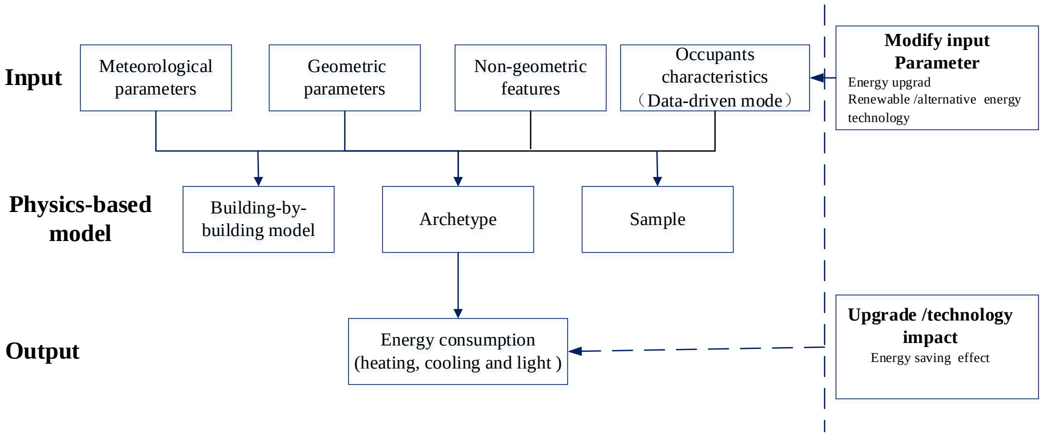


Figure.6. The steps involved in developing physics-based model for retrofitting buildings

Developing physical models are a common feature for physics-based approaches. Many energy simulation software such as Equest, EnergyPlus, Dest, Doe-2 .etc. can be used to build physical model, the regional building simulation software, such as Umi, CityBES, and CitySim [38] can also be directly used. Meteorological parameters, building geometric parameters (building form, area, and height), non-geometric features parameters (heating systems, air conditioning systems, equipment, etc.) and occupants characteristics parameters need to be input when developing physical models, which leads to two challenges. First of all, collecting building information is difficult. Dozens, hundreds, or even thousands of buildings are included in larger-scale, more information need to be collected. Not only does it consume time, field investigation to obtain detailed information of all buildings becomes unrealistic [42]. Although GIS and remote sensing techniques can provide the possibility to obtain detailed building geometric information and to visualize distribution of buildings in large-scale [43], it is still not possible to input occupants characteristics parameters. . Secondly, there is the performance gap of energy saving potential due to lots of incomplete information in most cases, such as user behaviour and infiltration in post-retrofit buildings. Therefore, some default values in the standard or the simulation engine are often used. In addition, a timetable approach is often used to define occupant behavior. However，the actual schedule for residential behavior, such as opening window, hours of occupation, set-point temperatures etc., are often random and uncertain. The factors, such as occupants’ thermal comfort demand, climatic, building type, state of occupant, economical parameters, cultural, traditional factors et al. affect occupant’s energy usage [26]. These lead to performance gap between actual and theoretical energy consumption. [39] identified that occupant behavior is one of important factors for thermal demand changes at regional level. The gap was also noted by Morten Brøgger [10] and HugoHens [40].etc.

In additional to occupant behavior, Meteorological parameters is another important reason for gap in energy saving potential. Currently, there are many forms of meteorological data used in building models, such as Test Reference Year. These databases are obtained by averaging historical meteorological data. However, the microclimate environment has greater impact on regional building energy consumption [41]. The factors, such as changes of community greening, the enhancement of heat island effect and global warmingetc. will cause microclimate changes. Therefore, using the same meteorological parameters to predict the energy saving potential of buildings for pre and post retrofit will cause the prediction deviation [42], and introducing microclimate model into energy models will become an important research direction.

Three methods exist for estimating the energy-saving potential of buildings in large-scale: building-by-building, archetypes and sample aggregation approaches.

### 2.2.1 Building-by-building aggregation approaches

Building-by-building aggregation approaches refer to energy saving potential of buildings obtained by aggregating the difference of energy consumption between pre and post retrofit of all the single building in the study area. A physical model of each building needs to be built and analyzed. The major advantage of this approach is that the characteristics of each building are included. Shading between buildings can also be considered. The prediction accuracy of energy consumption is relatively high. One major drawback is that lots of building-related information need to be provided as input, which is time consuming. GIS and remote sensing techniques provides solution for obtaining detailed building geometric information [43]. However, GIS datasets are not available to public in many countries.

Chen et al. presented urban building energy model-- the City Building Energy Saver to quantify the energy saving potential of buildings in large-scale [38]. Figure 7 show the workflow of this model. Using San Francisco as a case, the energy saving potential and cost for five retrofit measures （upgrade heating ,cooling system, replace windows, add air-economizer, replace lighting with LED） for 940 offices have been analyzed using this tool. The results show that the combination of five retrofit measures could significantly reduce energy consumption of buildings, but upgrading the air-conditioning system and replacing windows in this area were not economical due to low cooling and heating loads in San Francisco. The major contribution for this model is that it can automatically generate building physical models based on buildings datasets and GIS of cities. The tool solves the problem of inputting a large amount of physical information to build physical models, and saves time. However, GIS consolidated a city’s multiple datasets is key barrier when using the method in other countries. In addition, the City Building Energy Saver provides visualization by color-coding the 3D-view of the buildings based on performance metrics (Fig.8). The authors tried to introduce meteorological data closer to the study area through GIS positioning, but the documents introduced were still typical meteorological data of American cities. These data are not representative of the microclimate of cities or regions. Despite attempts to verify the model with monthly electricity and natural gas energy consumption, the model can only be calibrated on a regional level. It is difficult to obtain the monthly electricity and water bill data for each individual.

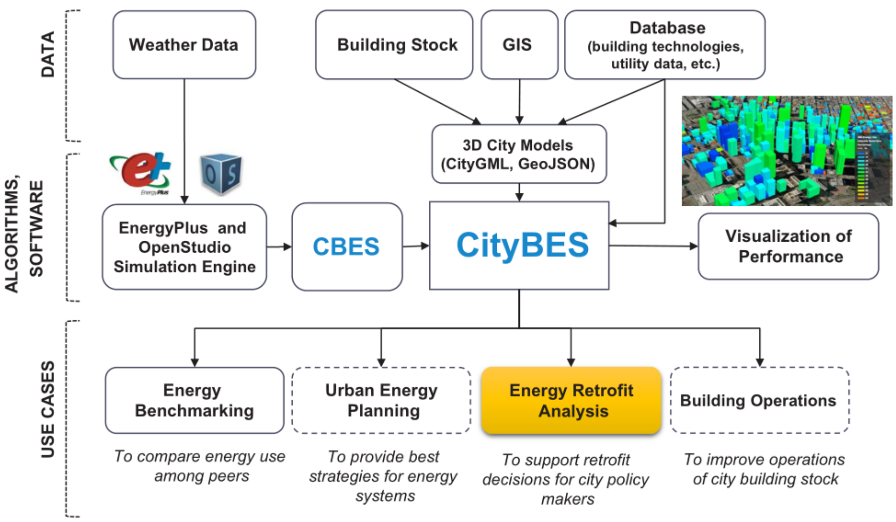


Figure.7. The workflow of the City Building Energy Saver

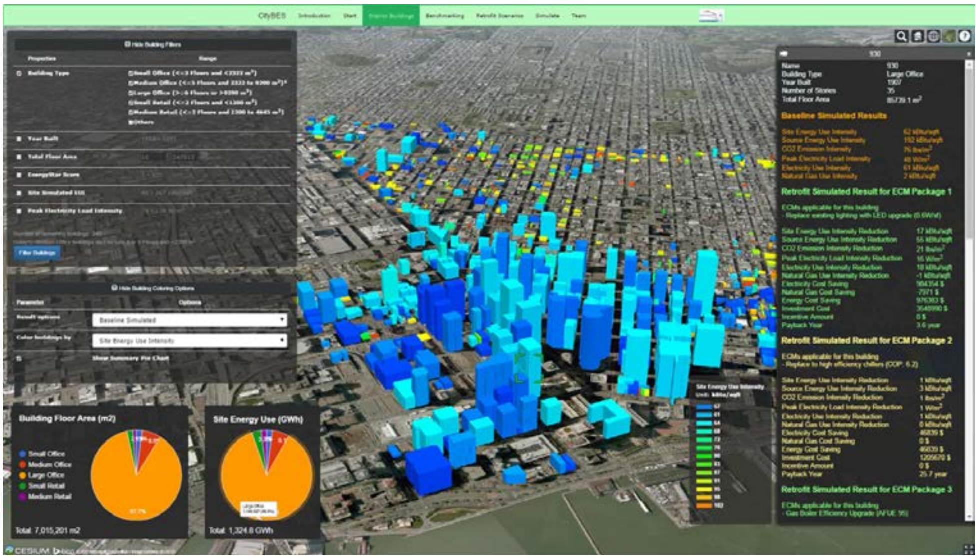


Fig.8. The 3Dview of the buildings based on selected performance Metrics [38]

Wang et al. developed named 'Combined Energy Simulation and Retrofitting' (CESAR) model [1], consist of two models: energy demand model and the retrofit model. The building future energy performance and carbon dioxide emissions of three different residential districts under two retrofit strategies ("Business as Usual" and the "New Energy Policy" based on the Swiss Energy Strategy 2050) were analyzed using CESAR model. The energy demand model employed the simulation tool EnergyPlus to calculate current building energy demand at all three scales. The retrofit model predicted the building future energy consumption according to different retrofit strategies. It was concluded that more buildings could reach the desired energy-saving goal under the 'New Energy Policy' scenario compared to 'Business as Usual'. Since the shading and occlusion of adjacent buildings can be considered in this model. The accuracy of the model is relatively high.

### 2.2.2 Archetypes aggregation approach

Using virtual buildings as archetypes to represent the building group in study area according to the building characteristics and performance, archetypes aggregation approach obtains energy saving potential of buildings in large-scale by aggregating the difference of energy consumption between pre and post retrofit of the archetypes building. This method has been widely used to predict the energy saving potential of buildings in large-scale due to fewer building models.

Defining archetype buildings is necessary for this approach. However, there are different concepts for each archetype in the existing studies. In other word, archetype buildings include different characteristics in different studies. Building structure (e.g. envelope types, number of floors, floor area, and building age etc.), building system performance (e.g., ventilation system, status of refurbishment; heating source, energy use intensity, and total energy consumption etc.), spatial factor (e.g. geographical location and climate zone etc.) and socio-economic factor (e.g. population and household income etc.) are used as a reference to identify archetypes. For example, Mata et al. have classified buildings into several archetypes according to building type, construction year, main heating system, climate zone for four European countries [44]. Dall'O' et al. have established archetypes based on the building types (detached, semi-detached ,up to 6 apartments and over 6 apartments) [45]. Pasichnyi et al. established archetypes based on buildings age, types and type heating source [24]. Unfortunately, none of the studies really addresses the questions of what makes a building representative or why the " archetype " in a given study is representative building in large-scale. Therefore, it is meaningless to attempt to compare the models in the different studies. Brøgger et al. have questioned the lack of support for the process of defining and selecting typical archetype buildings [19]. To overcome the shortcomings of the archetypes method, a new “hourglass” model has been proposed, which combines the archetype models with the diversification process to achieve the diversity of the archetypes [46].

Verification of the model is the main challenge of the archetype buildings. Archetype buildings are usually virtual. The average value of each type of buildings were often utilized as the main characteristic parameters of the archetypes. This makes it difficult to verify the validity of the proposed models at single building level. It can only be done at the aggregate level due to lack of actual energy consumption data of individual archetype.

‘Trade-offs’ between speed and accuracy of the models need to be identified using archetypes aggregation approaches. The models with more archetypes are accurate, but are complex. In contrast, models with few archetypes are simple. In addition, the limited data availability makes the archetype building unable to represent a larger scale of building stock, which limits the application scope of the model.

Using data from the survey, simulate the software's default data and Data published in ASHRAE and CIBSE guidelines, Hui Ben etc. developed five archetypes: active spenders, conscious occupiers, average users, conservers and inactive users based on comfort, behavior, energy use and household characteristics, to represent the building stocks in Cambridge [47]. The energy saving potential of archetypes building for 8 types of retrofit measure were simulated using energy consumption simulation software IES-VE. Simultaneously, the 8 retrofit measures were applied to the urban building scale of Cambridge, and the results showed the energy saving potential of five archetypes is obviously different with the same retrofit measure.

In develop archetypes, the author took human behavior as the basis to classify archetypes buildings for the first time. This solves retrofit shortfalls that building energy model often uses standardized assumptions about a set of comfort conditions, such as heating settings remaining the same among heterogeneous household. It provides a practical method for tailoring retrofit scheme for specific target groups. However, the model only compares the changes in human behavior horizontally, without considering the longitudinal comparison. In other words, the energy use behavior of different archetypes families is different, but it is assumed no change in a period of time (pro and post retrofit). However, there is some change for energy use behavior in pro and post retrofit for the same archetype families. For example, after retrofit, the time of turning on heating equipment may be shortened due to the improvement of indoor comfort. This leads to gap between predicted and actual energy consumption.

The Canadian Residential Energy End-Use Model (CREEM) proposed by Farahbakhsh et al. is one of the most widely known Canadian building stock modeling [48]. In the model, the classification of 8,767 houses based on types and age of building, heating source fuels, location resulted in 16 archetypes to simply determine all building stock energy consumption. Two retrofit schemes, including upgrade to R-2000 standards and the National Energy Code for Housing, were compared using this model, and the energy-saving effect for the building stocks were discussed. It was found that both schemes had obvious energy-saving effects, but the schemes for R-2000 standards had more energy-saving than the National Energy Code for Housing. The CREEM has been considered to be effective for evaluating the energy-saving and carbon dioxide emissions reduction brought by various retrofit measures for the Canadian building stocks.

Dascalaki et al. treated archetype buildings as a tool to predict energy saving potential for Hellenic building stocks [49]. The building stocks were classified into 24 archetypes based on the construction year (pre-1980, 1981–2001 and 2002–2010), building type and the four climates (601–1100 heating degree day (HDD), 1101–1600 HDD, 1601–2200 HDD, and 2201–2620 HDD). The Software- TEE-KENAK [50] was employed to develop model for analyzing energy saving potential and investment return period of residential buildings retrofit in two different scenarios: a “standard” and an “ambitious” scenario. The results showed that both of these scenarios were important in improving building performance, and the energy saving potential for “standard” scenario is higher than “ambitious”. In addition, it has better energy-saving effect for buildings with poor initial performance. Simultaneously, percentage of building retrofit built in different periods were formulated to achieve the overall national indicative target of 9% energy-saving by 2016 based on the simulation results [49]**.**

Moghadam et al. proposed the visualization approach evaluating the energy saving potential for the building stocks in Settimo Torinese [51]. The archetypes composed of 87 buildings have been created for Settimo Torinese. Two tools CitySim and GIS have been coupled to visualize energy consumption. The energy simulation software CitySim was employed to develop a 3D retrofit model. The ESP for two scenarios, which were determined in the light of the Tabula and the Minergie-P label standards, were predicted. Thanks to the 3D model, the microclimatic conditions based on the real city urban geometrical could be considered when analyzing the energy saving potential at the city scale. Building energy consumption with different retrofit scenarios was visible by color range GIS maps. This proposed method offered a user-friendly interface for architects or city manager to visualize the impact of the retrofit strategies.

Several other studies on the energy saving potential of large-scale building have been investigated by using archetype buildings based on different building attributes. For example, Dall’O’ et al. have analyzed energy saving potential of the building stocks retrofitted for member states using archetype buildings, which were identified by the building types (detached, semi-detached ,up to 6 apartments and over 6 apartments) [45]. The possible energy demand reductions were investigated in [52]. The building stocks in Norwegian were grouped into forty archetypes based on heating carrier share, sector and energy classification.

### 2.2.3 Sampling approaches

Using some real buildings as sample to represent the buildings in the study area, the energy consumption of each sample building is calculated by physical-based models, and then the consumption for large-scale building is aggregated using escalation factors, such as number of buildings, floor surfaces, etc. The approaches are called as the sample approach.

As archetypes aggregation approaches, determining number of sample buildings is a major challenge of the sample aggregation approaches. With the increase of the number of samples, the models become more complex and accurate. On the contrary, for the simple models the accuracy decreases. A trade-off between precision and complexity should be made. In addition, how to select the sample buildings is an important challenge for the sample aggregation approaches. Which buildings can be used as samples to represent the buildings in the study area and why these buildings can represent the buildings in the study area have not been really solved by the study, obviously, for the same study area, different samples may result in different results.

One advantage of this approach is that the model can be validated at the level of a single building. Since sample buildings are extracted from the actual building, the model can be verified using actual measured energy consumption at the level of a single building.

The Energy, Carbon and Cost Assessment for Building Stocks (ECCABS) model has been presented by Érika Mata [53]. This model was developed with the Matlab and Simulink tools according to the single-zone hourly heat balance principle. 1400 buildings have been determined as sample buildings representing Swedish residential building stock [54]. The energy saving potential of 12 measures have been estimated. The result showed energy consumption of the Swedish residential stock will be reduced by 53% using all 12 retrofit measures compared with benchmark year. Despite effort to verify the model, the authors only used models that have been developed, such as HAM-tools, which may have rebound effect on the prediction result. It is not compared with the actual energy consumption data of the building stock, and the verification of individual building energy consumption is not considered.

Shimoda et al. have determined 23 household forms and 20 residential forms to represent all residential building for Osaka City. The total energy consumption of residential buildings in city-scale was summed up though multiplying the representative building energy consumption by corresponding building numbers [55]. The energy saving potential of wall and air conditioner upgrade were evaluated. The way of classifying buildings based on household type and dwelling type offered the distinct advantage that the impact of occupant behaviour change on energy consumption can be considered.

**2. 3 Hybrid approaches**

Hybrid approaches combined data-driven and physics-based approaches [21].This approaches were first proposed by Swan et al. to take full advantage of physical and statistical models [56]. Hybrid models have been divided into two categories according to the embedding manner of data-driven and physics-based approaches: 1) Physics-based data-driven model (Fig. 9), 2) Data-driven -based physical model (Fig. 10).

To developed physics-based model, meteorological parameters, building geometric parameters, non-geometric features and occupants characteristics still need to be input into the hybrid model. Therefore, it has flexibility and ability to model energy saving potential of building due to building parameters included. However, compared to physics-based approaches, uncertain factors in the physical models, such as occupant behavior (turn-on schedule for appliances or lighting etc.) are determined through the data-driven models. Therefore, only bounds on physical parameters are required, and a limited number of data need to be considered [17]. The prediction accuracy of energy consumption for the model is clearly improved. However, for data-driven -based physical model, the prediction accuracy of energy saving potential does not change greatly except for the adoption of different residential behavior model pre and post the retrofit. In addition to occupant behavior, dynamic meteorological parameter models can also be developed, and used it as input in a building-physics based model. However, none of studies have embeded dynamic meteorological models into physical models. As integrating the advantages of data-driven and physics-based approaches and overcoming a part of their shortcomings, hybrid models has become a research hotspot in predicting energy consumption for large scale building in recent years [57].

2.3.1 Physics-based data-driven model

Physics-based data-driven models refers to data-driven model, such as multiple linear regression, artificial neural networks and support vector machines using the output of the physical model. Figure 9 presents the steps involved in developing data-driven -based physical models for retrofit analysis. The output results of various scenarios for physics-based models are used as training data.

The data-driven models are developed based on these training data, and the energy saving potential was predicted through the data-driven model.
The great advantage of this model is that there is no need defining occupant behaviour explicitly. It can be accounted for by the data-driven model. It provides a solution for building energy consumption datasets lacking by using physical models to generate datasets.

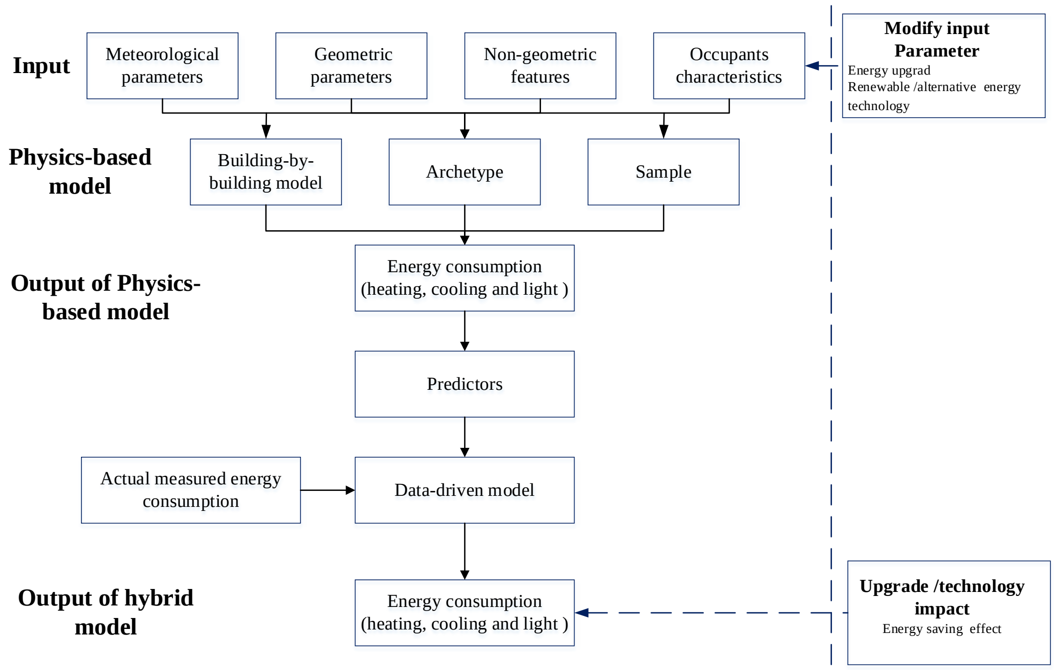


Figure. 9. The workflow of physics-based data-driven model for retrofit analysis.

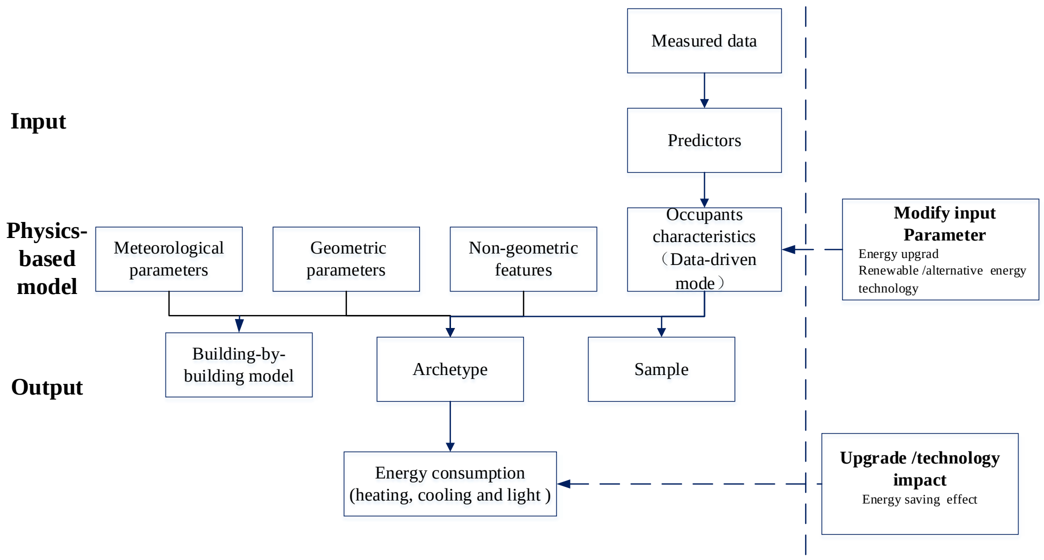


Figure. 10. The workflow of Data-driven -based physical model for retrofit analysis.

Using available 134,065 buildings data from EPC (Danish Energy Performance Certificate) and BBR (Danish Building and Dwelling Register) [58], Morten Brøgger developed a regression-based hybrid model to predict energy-saving potential for the Danish building stock. Figure 11 depicts flowchart of a regression-based hybrid model. The unique physics-based model for each building was built using the European standard ISO 13970. The energy consumption calculated based on the physical-based model is used as the predictor of the statistical model in the hybrid model as follow.

$$Q_{reg,i} = \beta_0 + \beta_1 \cdot Q_{calc,i} + \beta_2 \cdot A_{floor,i} + \beta_3 \cdot Q_{calc,i} \cdot A_{floor,i} + \psi \quad (4)$$

In the hybrid model, output result of the physics-based model is taken as the input, and the building physical parameters can be included in model. This makes it capable of estimating the energy saving potential for any building retrofit and upgrade. The distinct advantage for the hybrid model was that socio-technical factors had been taken into account by taking heating area as a predictor of the statistical model, but not as input parameters. It provides a simple way to correct errors that arise from uncertainty in the physics-based parts of the model, which makes it more accurate for estimating energy consumption for invisible attributes in building samples. However, this simple approach does not allow the identification of the source of the uncertain relationship. Therefore, it is impossible to detect which energy-saving measures cause the rebound effect.

Pasichnyi et al. estimated energy saving potential of seven retrofitting packages using hybrid model for 5,532 buildings in Stockholm City [24]. In this model, three virtual archetypes are determined to represent all buildings. Energy consumption simulation software DesignBuilder was employed to create physical models of virtual archetypes, and the results were calibrated using data-driven model based on measured data of 6,732 heating networks. It was concluded that the energy saving potential for the combined retrofitting packages is sometimes lower than the same single measure. The energy saving potential of building is also closely related to type of archetypes.

Nouvel et al. predicted energy saving potential for the buildings in Bospolder using a multi-scale framework hybrid model, which integrate by GIS-based data-driven and physical-based model [59]. In this hybrid model, the statistical model was used to predict energy consumption of building in the city-scale at first, and the retrofitting neighborhood areas were selected based on energy consumption distribution. The physical models of retrofit building for the neighborhood were created based on the 3D city model, and the result is calibrated by data-driven model. The energy saving potential of several retrofit measures, including building envelope insulations, energy-efficient windows replacement and thermal bridges treatment, were analyzed through the physical model. The result showed 59% reduction in average heating energy consumption post retrofit.

Ascione et al. presented a hybrid model combining artificial neural network and physics-based approaches to predict energy saving potential for office buildings built during 1920 to 1970 in South Italy in different retrofit scenarios [7]. Simulation software EnergyPlus was employed to develop energy consumption dataset. The artificial neural network model was trained and tested in Matlab using energy dataset. In this research, the proposed artificial neural network replaced standard building performance simulation tools to analyze energy saving potential. Therefore, computational effort and time were obviously decreased.

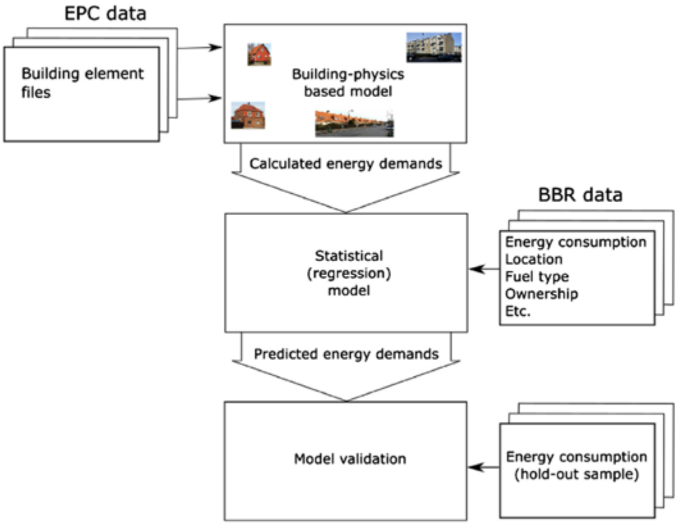


Figure. 11. The flowchart of a regression-based hybrid model

### 3.3.2 Data-driven -based physical model

Data-driven -based physics model is used tomeasure data to develop data-driven models modelling energy use of appliances, such as DHW, and lighting, and used this as input in building- physics based model. Figure 10 shows process of data-driven -based physical model. Using the Canadian Single-Detached and Double/Row House Database, including thermal envelope, equipment, occupancy, and air infiltration data from 16,952 residential buildings, a hybrid energy model (the Canadian Hybrid Residential End-Use Energy and GHG Emissions Model [CHREM)) integrating the neural network and physics-based approaches for building stock in Canada developed by Lukas G. Swan [60]. Figure 12 depicts the composition of the model CHREM. The neural network is used for modelling appliance and lighting energy consumption and domestic hot water volume draw. And this is used as input in a physical model. Energy saving potential of building stocks caused by retrofit and new technology can be evaluated through the right part of Figure 12. There is a distinct advantage for this way of combining data-driven and physics -based model that occupant behavior need not be assumed. However, several other parameters, such as indoor setting temperatures and air change rates. etc. remain uncertain in energy model of building stocks. Therefore, models that can account for all uncertain parameters while offering physical description of the system for buildings are required.

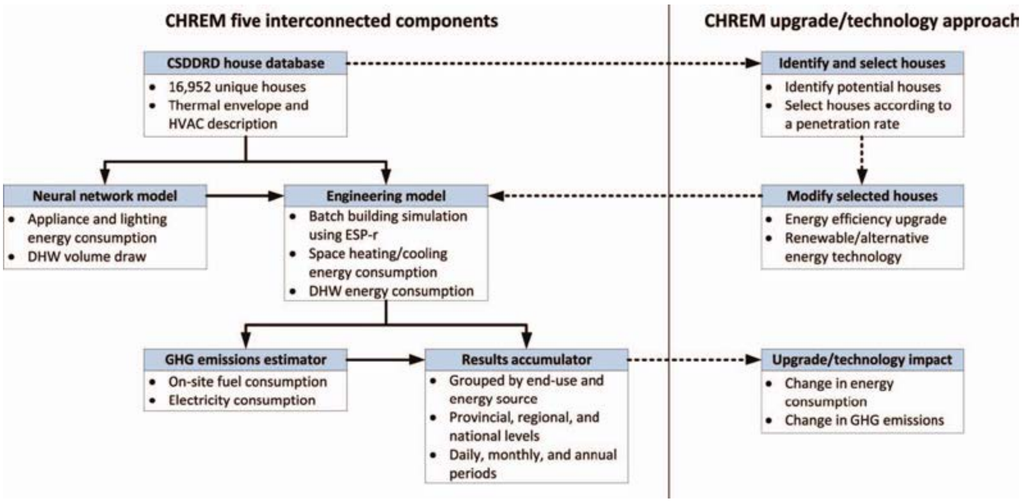


Fig. 12. The composition of the model CHREM [60]

## 3 Discussion

### 3.1 Model characteristics

All of the three approaches (data-driven，physical-based and hybrid approaches) can be used to predict the energy saving potential of building upgrade in large-scale. However, compared with the physical-based models, both data-driven and hybrid approaches developed late due to restrictions on the development of building data, and corresponding application are fewer.

Physical-based models require a simulation engine to build model, and detailed information such as climate parameters, building physical information, occupant behavior, and equipment systems need to be inputted during the modelling process. Therefore, energy performance of building retrofit and equipment update can be predicted and assessed. It can provide technical support for decision-makers to make reasonable energy saving retrofit strategy decisions. However, it takes a long time to input these detailed information, and professional expertise are required to build up the model, which limits the application of the model. Although some efforts have been made to reduce input information requirement, such as GIS [61], for automatic data generation city energy model (CITEBES)[38]. However, these are only in terms of building geometry information. There are lots of information for equipment systems and occupant behaviour parameters that need to be inputed. In addition, the simulation process also consumes a lot of time to calculate the potential savings at the regional level due to thousands of buildings. Despite archetypes approaches simplifying input data requirement, this leads to problem of accuracy.

Compared with the physical-based models, data-driven models do not require inputting so much information, but the relevant building information and historical energy consumption data need to be collected to predict the energy saving potential. It will be a great challenge for some countries. Hybrid models combining data-driven and physical-based model consider both the uncertainty of the building and relevant building information, which is significant to simulate the improvement of the energy saving potential for building retrofit, but also has time-consuming defects in data acquisition and information input. The future main research direction can develop fast and accurate regional energy consumption models to provide technical support for decision makers by establishing the data availability of building energy consumption, automatic models of importing building information, and simplified processes of modelling.

### 3.2 Model accuracy

#### 3.2.1 Influence of human behavior

Occupants’ behavior is one of the main factors influencing building energy consumption. Delzendeh et al. summarized the factors affecting occupant’s energy usage, including occupants’ thermal comfort demand, climatic, building type, state of occupant, economical parameters, etc [26]. Additionally, some cultural, traditional factors and Race/Ethnic factors also affect occupant’s energy usage. It was reported that heating energy use intensity was low for the median income, in contrast, it was high for poor households and Race/Ethnic Minority Family [62]. The total behavioral energy consumption, including district lighting consumption, heating load, cooling, hot water supply, is not simply the accumulation of behavioral energy consumption. For example, Gilani et al. has proved that the influence of occupant behavior on energy consumption would be diminished as the number of offices increases [63]. When changing from one building to several buildings, even districts, cities or building stocks, the total energy use of the occupants will have an aggregation and smoothing effect, and uncertainties caused by occupants’ behavior would be overlapped.

Resident behaviour is complex and uncertain, which contribute to the gap between prediction and actual energy performance. Current energy consumption simulation software handles presence (passive effects) and actions (active effects) of occupant in fixed or scheduled manner, which cannot reflect real energy consumption [64,65]. Data-driven models and hybrid models can consider human behavior, but these regressions are based on historical data. Behavior changes (opening window, adjusting the set temperature of HVAC, using solar shading and blinds, etc.) caused by development of socio-economic (such as income increase) and climate warming are not involved. In other words, these models reflect static energy saving potential, little consider future energy savings prediction post retrofit. In order to address the uncertainty for district energy simulations, some scholars have begun to study residential behavior models. For example, Baetens et al. developed stochastic residential occupant behavior model [66]. The model combines multiple models according to activity prerequisites and occupancy. An J et al. proposed a random human behavior models, which parameters are determined according to typical human behavior pattern and probability distribution based on a large number of questionnaire [67]. This method took into account the differences in pattern and random for different households' behavior in detail.

### 3.2.2 Influence of climate change

Climate is an important factor affecting buildings performance， its change directly affects the building energy consumption [68,69]. Due to the long service life of the building, considering climate change is necessary when analyzing the energy saving potential for buildings in large scale [25]. According to the fourth assessment report of the Intergovernmental Panel on Climate Change [69], climate changes induce temperature rise, climate variability and extreme events. These means that there will be less heating and more cooling demand due to climate warming in the future. For example, a region specific simulation for a residence in Argentina showed that for each 1◦C of increment in monthly mean outdoor temperature in summer (January)[70], an increase of cooling energy demand of about 2.2 kWh/$m^2$ per month is predicted. Similarly, for each 1◦C of increment in monthly mean outdoor temperature in winter (July), a decrease of 3.0 kWh/$m^2$ per month is predicted [71]. However, climate change is rarely considered in prediction models for building retrofit in large-scale.

Restricted by the availability of data on building characteristic parameters, data-driven models have been developed to predict the energy saving potential for buildings in large-scale in recent years. None of the models developed take climate parameters as predictors. Although meteorological parameters is input for physical-based model, these databases, such as Test Reference Year, Typical Metrological Year weather database, International Weather Files, ASHRAE IWEC2 datasets, WATSUN data, and European Test Reference Year weather data [4], are obtained by averaging historical meteorological data. The future weather files in the simulation software should consider not only typical but also extreme conditions. It should be determined based on dynamic climate models. Using different weather data estimates of energy consumption of pre and post building retrofit in large-scale to predict energy saving potential. The authors of [71] have shown that global climate change will alter the optimal scenario for future portfolios energy saving measure.

### 3.2.3 Influence of willingness

All three categories of models rarely consider people's willingness to retrofit. Building retrofit is characterized by subjectivity and diversity due to the different energy saving consciousness and objective conditions of the subjects. A variety of factors, such as age [72], household income and education levels[73] etc. affect owners' willingness for building retrofit. The probability of retrofit implementation is also different for different owners, which is ambiguous and uncertain. Jia JJ et al.[74] investigated household's willingness to accept six technical and behavioral energy-saving measures in Beijing. The results showed that socio-economic variables had stronger influence on willingness to accept technical energy-saving measures than behavioral measures. Hrovatin N et al.[72] stated that higher age of homeowners was an important barrier for energy-saving renovation of the building envelope in Slovenia, and increasing demand for living comfort and thermal comfort could drive residents' willingness for building envelope renovation. However, the previous research on the prediction model of building retrofit neglected the influence of willingness, and the prediction result will deviate from the actual situation to a certain extent. Therefore, it is necessary to introduce the willingness model into the prediction model of building retrofit.

## 4. Conclusion

Large-scale building retrofit models can provide theoretical support for formulating energy saving policies and developing strategies for buildings' retrofit, and it is of great significance for sustainable development of social. This paper summarizes models recently developed for large-scale building retrofit, including physics-based models, data-driven models and hybrid models. The modelling processes of these models have been introduced in detail, the advantages and limitations have been analyzed, and some examples of model applications have been demonstrated. In summary, the limitations and development trends are as followings:

1). Physics-based models consume a lot of time, and prone to overestimate, and the statistical models are simple, but the data acquisition is a challenge. The hybrid models have higher accuracy, but consume more time in data acquisition and information input. It is important to research fast and accurate regional energy consumption models to provide technical support for decision makers.

2). Uncertainty is one of main reasons causing the rebound effect of prediction accuracy for energy consumption modelling. Studying on how to introduce uncertain factors (climate change, microclimate and human behavior changes, etc.) into the model is an important direction to enhance the prediction accuracy.

3) Occupant's willingness to retrofit was ignored in all three categories of models, which can lead to the prediction result deviate from the actual situation in a certain extent. Therefore, it is necessary to introduce the willingness model into the energy consumption prediction model for energy retrofit.


## Acknowledgements

This study was supported by the China Beijing Municipal education Commission. The work is part of the research content of Beijing advanced innovation center for future urban design (grant number 3030036121906).


## References


[1] Ascione F, Bianco N, De Stasio C, Mauro GM, Vanoli GP. CASA, cost-optimal analysis by multi-objective optimisation and artificial neural networks: A new framework for the robust assessment of cost-optimal energy retrofit, feasible for any building. Energy Build 2017;146:200–19.

[2] Berardi U. A cross-country comparison of the building energy consumptions and their trends. Resour Conserv Recycl 2017;123:230–41.

[3] Cabrera Serrenho A, Drewniok M, Dunant C, Allwood JM. Testing the greenhouse gas emissions reduction potential of alternative strategies for the english housing stock. Resour

Conserv Recycl 2019;144:267–75.
[4] Asaee SR, Nikoofard S, Ugursal VI, Beausoleil-Morrison I. Techno-economic assessment of photovoltaic (PV) and building integrated photovoltaic/thermal (BIPV/T) system retrofits in the Canadian housing stock. Energy Build 2017;152:667–79.
[5] Li W, Zhou Y, Cetin K, Eom J, Wang Y, Chen G, et al. Modeling urban building energy use: A review of modeling approaches and procedures. Energy 2017;141:2445–57.
[6] Li B, Han S, Wang Y, Li J, Wang Y. Feasibility assessment of the carbon emissions peak in China's construction industry: Factor decomposition and peak forecast. Sci Total Environ 2020;706.
[7] Ascione F, Bianco N, De Stasio C, Mauro GM, Vanoli GP. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. Energy 2017;118:999–1017.
[8] Caputo P, Pasetti G. Overcoming the inertia of building energy retrofit at municipal level: The Italian challenge. Sustain Cities Soc 2015;15:120–34.
[9] Ma Z, Cooper P, Daly D, Ledo L. Existing building retrofits: Methodology and state-of-the-art. Energy Build 2012;55:889–902.
[10] Brøgger M, Bacher P, Madsen H, Wittchen KB. Estimating the influence of rebound effects on the energy-saving potential in building stocks. Energy Build 2018;181:62–74.
[11] Howard B, Parshall L, Thompson J, Hammer S, Dickinson J, Modi V. Spatial distribution of urban building energy consumption by end use. Energy Build 2012;45:141–51.
[12] Ma Z, Cooper P, Daly D, Ledo L. Existing building retrofits: Methodology and state-of-the-art. Energy Build 2012;55:889–902.
[13] Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renew Sustain Energy Rev 2009;13:1819–35.
[14] Kavgic M, Mavrogianni A, Mumovic D, Summerfield A, Stevanovic Z, Djurovic-Petrovic M. A review of bottom-up building stock models for energy consumption in the residential sector. Build Environ 2010;45:1683–97.
[15] Suganthi L, Samuel AA. Energy models for demand forecasting - A review. Renew Sustain Energy Rev 2012;16:1223–40.
[16] Zhao HX, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16:3586–92.
[17] Foucquier A, Robert S, Suard F, Stéphan L, Jay A. State of the art in building modelling and energy performances prediction: A review. Renew Sustain Energy Rev 2013;23:272–88.
[18] Reinhart CF, Cerezo Davila C. Urban building energy modeling - A review of a nascent field. Build Environ 2016;97:196–202.
[19] Brøgger M, Wittchen KB. Estimating the energy-saving potential in national building stocks – A methodology review. Renew Sustain Energy Rev 2018;82:1489–96.
[20] Ahmad T, Chen H, Guo Y, Wang J. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. Energy Build 2018;165:301–20.
[21] Bourdeau M, Zhai X qiang, Nefzaoui E, Guo X, Chatellier P. Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustain Cities Soc 2019;48:101533.
[22] Gassar AAA, Cha SH. Energy prediction techniques for large-scale buildings towards a

sustainable built environment: A review. Energy Build 2020;224:110238.
[23] Dong B, Li Z, Rahman SMM, Vega R. A hybrid model approach for forecasting future residential electricity consumption. Energy Build 2016;117:341–51.
[24] Pasichnyi O, Wallin J, Kordas O. Data-driven building archetypes for urban building energy modelling. Energy 2019;181:360–77.
[25] Deb C, Schlueter A. Review of data-driven energy modelling techniques for building retrofit. Renew Sustain Energy Rev 2021;144:110990.
[26] Delzendeh E, Wu S, Lee A, Zhou Y. The impact of occupants' behaviours on building energy analysis: A research review. Renew Sustain Energy Rev 2017;80:1061–71.
[27] Bassani M, Catani L, Cirillo C, Mutani G. Night-time and daytime operating speed distribution in urban arterials. Transp Res Part F Traffic Psychol Behav 2016;42:56–69.
[28] Tardioli G, Kerrigan R, Oates M, O'Donnell J, Finn D. Data driven approaches for prediction of building energy consumption at urban level. Energy Procedia 2015;78:3378–83.
[29] Torabi Moghadam S, Toniolo J, Mutani G, Lombardi P. A GIS-statistical approach for assessing built environment energy use at urban scale. Sustain Cities Soc 2018;37:70–84.
[30] Walter T, Sohn MD. A regression-based approach to estimating retrofit savings using the Building Performance Database. Appl Energy 2016;179:996–1005.
[31] Mastrucci A, Baume O, Stazi F, Leopold U. Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam. Energy Build 2014;75:358–67.
[32] Nino Streicher K, Parra D, Buerer MC, Patel MK. Techno-economic potential of large-scale energy retrofit in the Swiss residential building stock. Energy Procedia 2017;122:121–6.
[33] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, et al. A review of data-driven approaches for prediction and classification of building energy consumption. Renew Sustain Energy Rev 2018;82:1027–47.
[34] Aydinalp M, Ismet Ugursal V, Fung AS. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. Appl Energy 2002;71:87–110.
[35] Hawkins D, Hong SM, Raslan R, Mumovic D, Hanna S. Determinants of energy use in UK higher education buildings using statistical and artificial neural network methods. Int J Sustain Built Environ 2012;1:50–63.
[36] Re Cecconi F, Moretti N, Tagliabue LC. Application of artificial neutral network and geographic information system to evaluate retrofit potential in public school buildings. Renew Sustain Energy Rev 2019;110:266–77.
[37] Park JH, Jeon J, Lee J, Wi S, Yun BY, Kim S. Comparative analysis of the PCM application according to the building type as retrofit system. Build Environ 2019;151:291–302.
[38] Chen Y, Hong T, Piette MA. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. Appl Energy 2017;205:323–35.
[39] Kazas G, Fabrizio E, Perino M. Energy demand profile generation with detailed time resolution at an urban district scale: A reference building approach and case study. Appl Energy 2017;193:243–62.
[40] Hens H, Parijs W, Deurinck M. Energy consumption for heating and rebound effects. Energy Build 2010;42:105–10.

[41] Allegrini J, Dorer V, Carmeliet J. Influence of the urban microclimate in street canyons on the energy demand for space cooling and heating of buildings. Energy Build 2012;55:823–32.
[42] Santamouris M, Papanikolaou N, Livada I, Koronakis I, Georgakis C, Argiriou A, et al. On the impact of urban climate on the energy consuption of building. Sol Energy 2001;70:201–16.
[43] Alhamwi A, Medjroubi W, Vogt T, Agert C. GIS-based urban energy systems models and tools: Introducing a model for the optimisation of flexibilisation technologies in urban areas. Appl Energy 2017;191:1–9.
[44] Mata É, Sasic Kalagasidis A, Johnsson F. Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. Build Environ 2014;81:270–82.
[45] Dall'O' G, Galante A, Pasetti G. A methodology for evaluating the potential energy savings of retrofitting residential building stocks. Sustain Cities Soc 2012;4:12–21.
[46] Jaffal I, Inard C, Ghiaus C. Fast method to predict building heating demand based on the design of experiments. Energy Build 2009;41:669–77.
[47] Ben H, Steemers K. Modelling energy retrofit using household archetypes. Energy Build 2020;224:110224.
[48] Farahbakhsh H, Ugursal VI, Fung AS. A residential end-use energy consumption model for Canada. Int J Energy Res 1998;22:1133–43.
[49] Dascalaki EG, Droutsa KG, Balaras CA, Kontoyiannidis S. Building typologies as a tool for assessing the energy performance of residential buildings - A case study for the Hellenic building stock. Energy Build 2011;43:3400–9.
[50] V.1.28.1.73 T-K software. No Title 2011.
[51] Torabi Moghadam S, Coccolo S, Mutani G, Lombardi P, Scartezzini JL, Mauree D. A new clustering and visualization method to evaluate urban heat energy planning scenarios. Cities 2019;88:19–36.
[52] Sartori I, Wachenfeldt BJ, Hestnes AG. Energy demand in the Norwegian building stock: Scenarios on potential reduction. Energy Policy 2009;37:1614–27.
[53] Mata É, Kalagasidis AS, Johnsson F. A modelling strategy for energy, carbon, and cost assessments of building stocks. Energy Build 2013;56:100–8.
[54] Mata É, Sasic Kalagasidis A, Johnsson F. Energy usage and technical potential for energy saving measures in the Swedish residential building stock. Energy Policy 2013;55:404–14.
[55] Shimoda Y, Fujii T, Morikawa T, Mizuno M. Residential end-use energy simulation at city scale. Build Environ 2004;39:959–67.
[56] Wang S, Yan C, Xiao F. Quantitative energy performance assessment methods for existing buildings. Energy Build 2012;55:873–88.
[57] Abbasabadi N, Ashayeri M, Azari R, Stephens B, Heidarinejad M. An integrated data-driven framework for urban energy use modeling (UEUM). Appl Energy 2019;253:113550.
[58] Brøgger M, Bacher P, Wittchen KB. A hybrid modelling method for improving estimates of the average energy-saving potential of a building stock. Energy Build 2019;199:287–96.
[59] Nouvel R, Mastrucci A, Leopold U, Baume O, Coors V, Eicker U. Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support. Energy Build 2015;107:204–12.
[60] Swan LG, Ugursal VI, Beausoleil-Morrison I. Hybrid residential end-use energy and greenhouse gas emissions model - development and verification for Canada. J Build Perform Simul 2012;6:1–23.

[61] Cerezo Davila C, Reinhart CF, Bemis JL. Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. Energy 2016;117:237–50.

[62] Reames TG. Targeting energy justice: Exploring spatial, racial/ethnic and socioeconomic disparities in urban residential heating energy efficiency. Energy Policy 2016;97:549–58.

[63] Gilani S, O'Brien W, Gunay HB. Simulating occupants' impact on building energy performance at different spatial scales. Build Environ 2018;132:327–37.

[64] Fabi V, Andersen RV, Corgnati SP, Olesen BW. A methodology for modelling energy-related human behaviour: Application to window opening behaviour in residential buildings. Build Simul 2013;6:415–27. https://doi.org/1227301301196.

[65] Martinaitis V, Zavadskas EK, Motuziene V, Vilutiene T. Importance of occupancy information when simulating energy demand of energy efficient house: A case study. Energy Build 2015;101:64–75.

[66] Baetens R, Saelens D. Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour. J Build Perform Simul 2016;9:431–47.

[67] An J, Yan D, Hong T, Sun K. A novel stochastic modeling method to simulate cooling loads in residential districts. Appl Energy 2017;206:134–49.

[68] Daly D, Cooper P, Ma Z. Implications of global warming for commercial building retrofitting in Australian cities. Build Environ 2014;74:86–95.

[69] Nik VM, Mata É, Sasic Kalagasidis A. A statistical method for assessing retrofitting measures of buildings and ranking their robustness against climate change. Energy Build 2015;88:262–75.

[70] 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.

[71] Flores-Larsen S, Filippín C, Barea G. Impact of climate change on energy use and bioclimatic design of residential buildings in the 21st century in Argentina. Energy Build 2019;184:216–29.

[72] Hrovatin N, Zorić J. Determinants of energy-efficient home retrofits in Slovenia: The role of information sources. Energy Build 2018;180:42–50.

[73] Gamtessa SF. An explanation of residential energy-efficiency retrofit behavior in Canada. Energy Build 2013;57:155–64.

[74] Jia JJ, Xu JH, Fan Y, Ji Q. Willingness to accept energy-saving measures and adoption barriers in the residential sector: An empirical analysis in Beijing, China. Renew Sustain Energy Rev 2018;95:56–73.

[75] Hsu D. Identifying key variables and interactions in statistical models ofbuilding energy consumption using regularization. Energy 2015;83:144–55.