Renewable and Sustainable Energy Reviews 199 (2024) 114478

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Archetype building energy modeling approaches and applications: A review

Pengyuan Shen [a,*], Huilong Wang [b,**]

[a] Harbin Institute of Technology, Shenzhen, Guangdong Province, China
[b] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, Guangdong Province, China

ARTICLE INFO

Keywords:
Building archetypes
Bottom-up modeling
Building stock
Building simulation
Machine learning

ABSTRACT

Archetype Building Energy Models (ABEMs) are representations of buildings in a certain region that enable the assessment of energy use across building stocks in a bottom-up manner, playing an important role in building energy policy making, energy efficiency measure evaluation, and sustainable urban planning. However, the selection of the suitable modeling approach can be up to various factors due to the diversity of available methods and the specific requirements of application scenarios. This review aims to address this issue by analyzing and comparing different approaches in ABEM, namely the building codes-based approach, data-driven approach, and hybrid approach, and present the strengths, weaknesses, and real-world applicability of the modeling approaches. The fitness of each method to different research purposes and contexts is explored. This study sheds light on key factors influencing the choice of the ABEM method, including the objectives of the research, available data quality, computational resources, and model accuracy. By gathering and synthesizing available information from the state of art studies, an overview and guideline for researchers and decision-makers who intend to leverage ABEM for various purposes is provided. It not only helps in the better understanding of existing modeling methods but also identifies challenges faced with ABEMs and the potential in future improvement. This work also identifies opportunities in future ABEM and its flexibility in various application scenarios, and it is anticipated that ABEMs will continue to play important roles in informing engineering design, influencing regulations, optimizing energy systems, guiding policy decisions for sustainable building development.

1. Introduction

Building sector plays an important role in energy consumption and greenhouse gas emissions, accounting for approximately 36 % of the total energy consumption and 40 % of CO_2 emissions worldwide [1]. The importance of developing accurate and reliable building energy models (BEMs) cannot be overstated, as these models are essential tools for understanding building energy consumption (BEC) patterns, identifying energy-saving potentials, and informing decision-making in energy-saving measures [2]. Archetype building energy models (ABEMs) are a set of BEMs that aim to represent a group of buildings with similar characteristics, which provides a generalized or representative understanding of energy consumption for the built environment [3]. This work aims to present a comprehensive analysis of various methods and approaches for constructing engineering-based archetype building energy models and their applications.

The idea of ABEMs emerged from the need to simplify and generalize the analysis of large building stocks [4]. Analyzing each individual building within a region or city would be computationally expensive, particularly when considering the numerous variations in building characteristics, usage patterns, and construction details. ABEMs address this issue by representing a group of buildings that share common features, such as age, construction type, size, or BEC characteristics [5]. The primary objective of ABEMs is to provide a simplified yet representative model of BEC, which can be used for purposes such as estimation of the overall energy performance of building stock, identification of potential energy-saving measures, and assessment of energy-saving strategies [6]. ABEMs are increasingly being recognized as an essential tool for urban energy planning, energy-saving measures and policy evaluation, and building stock analysis, which offers several advantages over building-per-building energy modeling, such as reduced computational complexity, increased scalability, and ease of integration with other urban planning tools. ABEMs have been applied to a wide range of purposes, including urban energy planning [7,8], climate change mitigation [9] and strategies [10], policy evaluation [11,12], and building

* Corresponding author. Harbin Institute of Technology, Shenzhen, China.
** Corresponding author.
E-mail addresses: Pengyuan_pub@163.com (P. Shen), wanghuilong@szu.edu.cn (H. Wang).

https://doi.org/10.1016/j.rser.2024.114478
Received 9 August 2023; Received in revised form 25 March 2024; Accepted 25 April 2024
Available online 3 May 2024
1364-0321/

Abbreviations

ABEM	Archetype Building Energy Models
AIC	Akaike Information Criterion
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
BEC	Building Energy Consumption
BEM	Building Energy Model
BETSI	Building Energy, Technical Status and Indoor Environment
BIC	Bayesian Information Criterion
BIM	Building Information Modeling
CBECS	Commercial Building Energy Consumption Survey
CCI	Climate Change Impact
CEN	European Committee for Standardizations
CIM	City Information Modeling
CRECS	China Residential Energy Consumption Survey
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DOE	Department of Energy, U.S
ECM	Energy Conservation Measure
EU	European Union
EUI	Energy Use Intensity
HVAC	Heating, Ventilation, and Air Conditioning
IECC	International Energy Conservation Code
ISO	International Organization for Standardization
GIS	Geographical Information System
NECB	National Energy Code of Canada for Buildings
PCA	Principal Component Analysis
RC	Thermal Resistance and Capacitance
RECS	Residential Energy Consumption Survey
TABULA	Typology Approach for Building Stock Energy Assessment
UHI	Urban Heat Island
UK	United Kingdom
WWR	Window to Wall Ratio
ZERH	NREL's Zero Energy Ready Home

stock retrofit analysis [9,13–16]. They can be used to assess the potential impact of energy conservation measures (ECMs) on different building typologies [17,18] and to evaluate the energy-saving potential of various technologies, such as building envelope improvements and renewable energy systems under a changing climate [19–21].

However, there is no universally recognized standard for obtaining building archetypes for thermal and energy use modeling. In different contexts, different terms are coined and used, such as reference building, prototype building, representative building, and typical building, and they may even refer to different implications as some of the building archetypes are developed based on building codes while others are extracted through survey or national level database to represent the status quo or reality of building stocks. This work provides a comprehensive review of the state-of-the-art methods and approaches involved in constructing such archetype buildings for various research purposes and attempts to offer critical perspectives on challenges and limitations of different methods as well as guidelines in how to make better choices and use of ABEMs with different contexts and research purposes. Our analysis provides insights into the global implications of adopting diverse ABEMs for industry stakeholders, policymakers, and researchers across various regions and countries, underscoring the practical relevance of ABEMs in driving sustainable urban development and shaping international energy policies and standards.

2. Contribution of this review

While several review papers have been published in the field of building energy modeling, our review aims to provide a more comprehensive and up-to-date analysis of methods and approaches for constructing ABEMs. Previous reviews have focused on specific aspects of building energy modeling, such as bottom-up building stock models [22,23], urban-level building energy use modeling [7,24,25], residential end-use energy consumption modeling techniques [3,5,26], and the application of machine learning techniques in ABEM generation [27–29]. However, a review that encompasses the methodologies for constructing physics-based ABEMs for various building types other than residential buildings and compares their advantages, limitations, and applications is lacking in the studies [3,5,7,22–26,28]. To show the most recent progress in the related fields, pertinent literature is selected in this review and most of the research works were conducted in the past decade as shown in Fig. 1.

The goal of this review is to offer a more comprehensive and in-depth examination of the various methods and approaches for constructing ABEMs based on engineering methods, comparing different approaches including building codes-based approaches, data-driven approaches, and hybrid approaches. This review provides a detailed comparison of these methods, considering factors that influence the choice of method, practical considerations, and applications. It shall be noted that the scope of this review is restrained to the building engineering model,

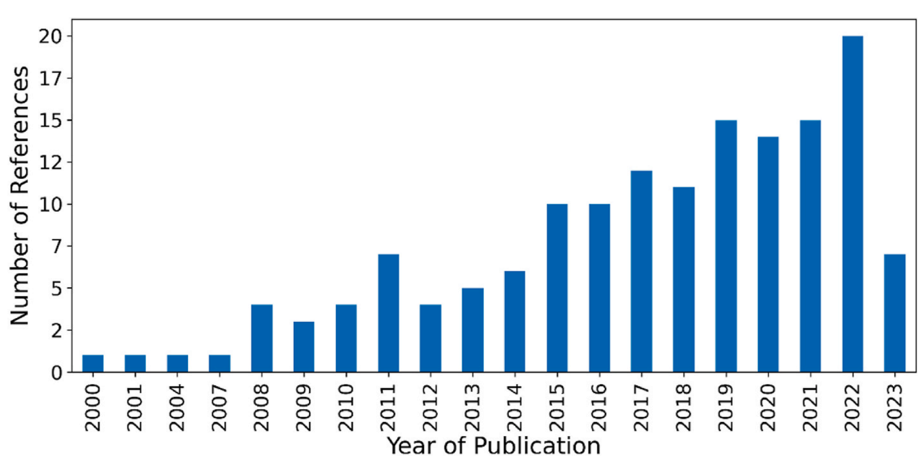

Fig. 1. Distribution of publication years of the reviewed papers.

which is usually adopted in bottom-up modeling. The challenges and future research directions in the field are also identified. This review aims to provide researchers, practitioners, and policymakers with a better understanding of the different methods and approaches available, their strengths and weaknesses, and their applicability to various energy analysis and policy evaluations.

3. Engineering-based archetype building energy models

3.1. Overview of the engineering method

The engineering method for constructing archetype building energy models (ABEMs) is based on a detailed analysis of individual building characteristics, such as building geometry, envelope properties, HVAC systems, and occupancy patterns [30]. This approach involves the use of building energy simulation tools to create representative models of building typologies that share similar BEC characteristics. The physics-based engineering archetype modeling method is widely adopted in building energy modeling, as it provides a high level of detail and accuracy, allowing for a better understanding of the underlying factors influencing BEC [31].

The process of constructing ABEMs using the engineering method typically involves several steps. First, a building stock analysis is conducted to identify the most common building typologies, based on factors including vintage, construction type, size, and geometry. Next, representative buildings for each identified typology are selected or developed, and their characteristics are either gathered from available data sources, such as building codes, energy audit reports, statistical analysis, or data-driven methods. BEM tools [32] are then employed to create detailed energy models of the representative buildings considering various model input parameters. The engineering method can provide detailed and accurate results with hourly granularity, allowing for a comprehensive understanding of energy performance in different building types [18,33]. However, the approach can sometimes be time-consuming and computationally intensive [24], as it requires the modeling of individual building characteristics and the simulation of building energy performance. This can be a factor that leads to difficulties and challenges when dealing with large building stocks or when data availability is under challenge. Despite these challenges, the engineering method remains an essential approach to provide reliable and detailed insights into BEC as different ABEM approaches can be employed to facilitate this modeling process.

3.2. Simulation tools for ABEM

BEM enables researchers, designers, and engineers to simulate building energy performance [34]. There is a wide range of BEM tools available, with differing levels of capabilities and complexity. Some of the most widely used tools in the field include EnergyPlus [35], TRNSYS [36], ESP-r [37], and DeST [38]. These tools are capable of providing hourly dynamic thermal simulation for ABEMs and are widely used in ABEM [38–44]. Besides dynamic simulation tools, some studies developed or used steady-state simulation tools for ABEM. Li et al. adopted a reduced-order RC (thermal resistance and capacitance) model to simulate DOE commercial reference buildings [45] and further considered urban heat island (UHI) impacts on ABEM simulation [46]. Dall'O et al. reported that they used a lightweight steady-state simulation tool called CENED + programmed based on European standard (CEN) to create statistical relationships for reference buildings constructed in various years [47]. Bagle et al. developed a gray-box RC model to facilitate the simulation of archetype models of Norwegian apartment blocks [48]. Steady-state ABEMs are faster in computation but may lead to a certain extent of loss of model confidence.

The choice of a BEM tool depends on various factors, including the required level of detail, available data, computational resources, and user expertise. Each tool has its strengths and weaknesses, and researchers and practitioners must carefully consider their specific needs and requirements when selecting an appropriate tool for ABEM [32]. Recently, more interest has been invested in integrating emerging tools with other platforms, such as Building Information Modeling (BIM) and Geographical Information Systems (GIS), to enable comprehensive and efficient analysis of building energy performance at the urban scale [44,49–51]. These integrated approaches can help overcome some of the limitations of traditional BEM tools and facilitate the construction of ABEMs for various energy analysis and policy evaluation tasks.

3.3. Modeling parameters considered in ABEMs

In developing ABEMs, various factors need to be considered to ensure that the models accurately represent the target building population. These factors include but are not limited to, building design and configuration, construction materials, energy systems, operational and management practices, and occupant behaviors [52]. Simply put, the important factors considered in ABEMs can be categorized into geometric-related and non-geometric parameters, and the modeling parameters of ABEMs are summarized in Fig. 2.

Factors of building design and configuration are critical factors, in which geometric-related information is considered. It includes the building's shape, size, orientation, and layout, which have a notable impact on the building energy performance [53,54]. To generate representative geometric configurations of buildings, GIS and City Information Modeling (CIM) are gaining popularity in being used to create 3D building models based on automated generation methods that vary in the level of detail [44,49,51,55,56]. However, not all countries and regions share high-quality GIS or CIM data publicly [57]. In some studies, geometric parameters of archetype buildings are generated based on personal expertise and assumptions [58,59].

Thermal property of construction material is one of the most important factors considered in non-geometric factors, which includes the type and quality of the building envelope, thermal insulation, window, and other envelope components. The thermal properties of these materials directly affect heat transfer rates, thereby impacting heating and cooling loads [60]. Other important factors include configurations of building energy systems, including HVAC, lighting, and appliances, which contribute significantly to BEC. The efficiency, control strategy, and maintenance of these systems can vary widely, leading to significant differences in energy use [6]. Operational and management practices, such as HVAC indoor setpoints, lighting control strategies, and maintenance practices, can also greatly influence building energy performance, and archetypes of these aspects have recently been studied as well [61].

Lastly, occupant behaviors, which include actions that affect building energy use, such as window opening, appliance use, and thermostat settings, are increasingly recognized as a crucial factor [62–64]. Recent research has demonstrated the importance of incorporating realistic occupant behavior models into ABEMs to improve their accuracy [62]. Studies have shown that variations in these practices can lead to differences in energy use of up to 50 % [63]. In recent years, there has been a growing interest in incorporating more of these factors into ABEMs to better capture the diversity and complexity of real-world buildings. Various methods including classic statistical methods and machine learning methods have been applied to the characterization of these factors for ABEM, which will be further discussed in detail in subsequent sections.

3.4. Thermal zoning in ABEMs

Thermal zoning is a critical aspect as it allows for the representation of variations in thermal conditions within a building [65]. The process of defining thermal zones in ABEMs involves the division of the building's interior space into distinct areas that share similar thermal characteristics, such as temperature, humidity, and solar heat gains. Proper

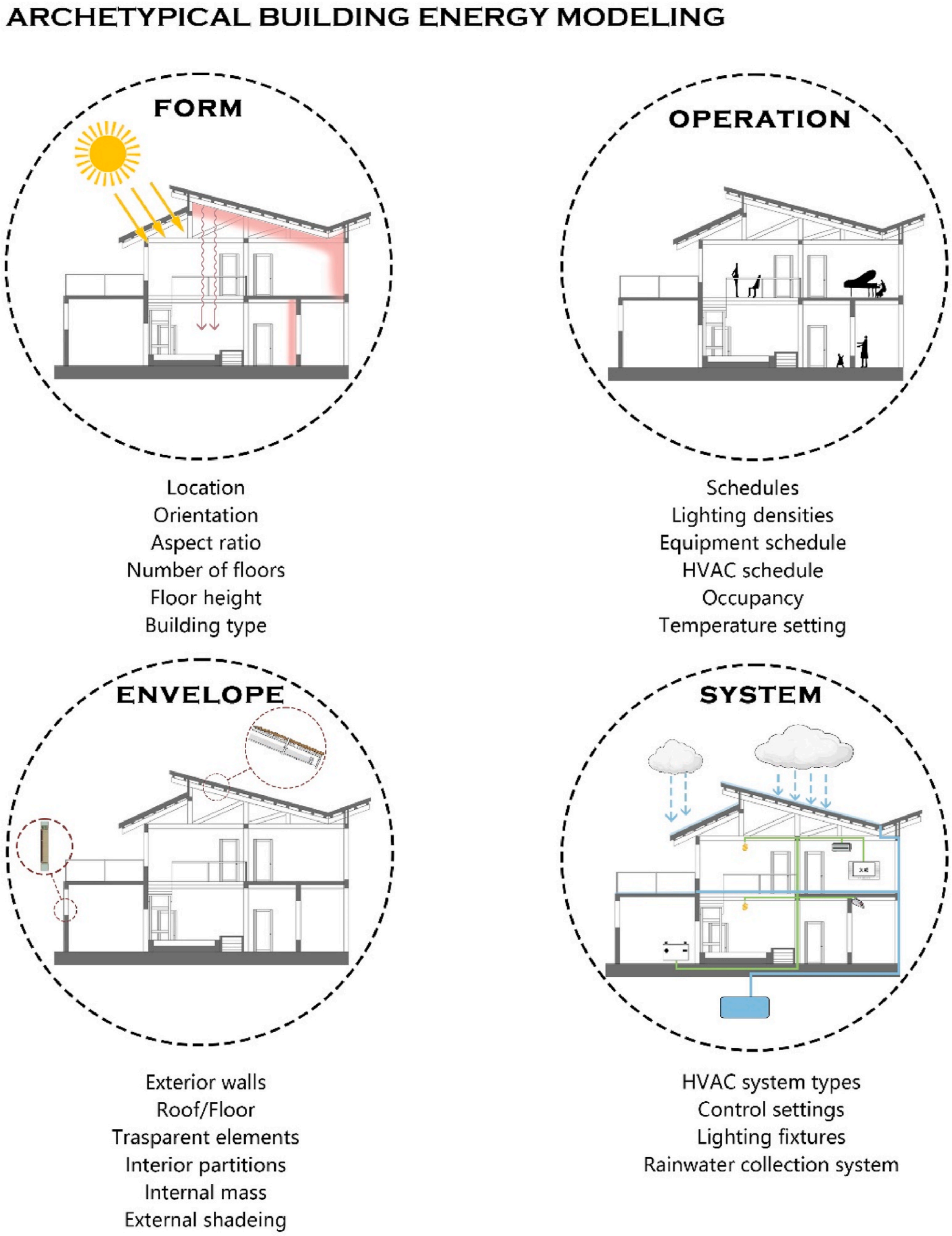

Fig. 2. Modeling parameters of ABEMs.

thermal zoning is essential for capturing the complex interactions between zones having various functions.

The level of detail in thermal zoning can vary depending on the objectives of the energy analysis and the available knowledge. In low-detail ABEMs, buildings may be represented as a single thermal zone. In urban building energy modeling, single thermal zone building models have been widely used in recent years. According to Cerezo et al., they used one thermal zone per floor method to reduce simulation time and complexity in modeling urban-level building energy use in Boston and acknowledged that a more detailed zoning scheme (core-perimeter zoning) would have increased accuracy but also simulation time [66]. Another urban-level building energy practice conducted by Deng et al. also used a single thermal zone model in which each floor was considered as a single thermal zone and annual BEC was calculated by EnergyPlus [44]. Some studies on energy consumption simulation of building stocks in the European Union (EU) also used the single thermal zone method for building archetypes [42,43,67]. While this approach can provide a quick estimate of BEC, it may not accurately capture the variations in thermal conditions within the building or the impact of specific energy-saving measures.

In medium- and high-detail ABEMs, buildings are divided into multiple thermal zones based on factors such as occupancy patterns, building envelope properties, and HVAC system configuration [65]. This approach allows for a more accurate representation of the thermal conditions within the building and can provide valuable insights into the impact of various ECMs and policy interventions. For medium-detail ABEMs, the core-perimeter thermal zoning strategy is usually adopted as proposed by ASHRAE 90.1–2016 appendix G [68] under the circumstances that the detailed HVAC system design has not yet been done or is not available. For high-detail ABEM, thermal zones are generated more complicatedly than the core-perimeter rule for a certain floor [69,70]. Research has shown that there can be discrepancies between simulation results of the same archetype buildings using different thermal zoning methods. Usually, low-detail thermal zoning leads to underestimated thermal loads and energy use as supported by several studies [70,71].

However, the process of defining medium and high-detail thermal zones can be time-consuming and may require a significant amount of data and user expertise. Recently, the research community has begun to explore automatic techniques for defining thermal zoning in ABEMs, leveraging advancements in computational algorithms and increasing availability of high-resolution building data [69,70,72–74]. Faure et al. designed a core-perimeter thermal zoning method by iterating over perimeter zone depth starting from 3 m and simulated urban-level building energy use [74]. Using closed boundary representations as inputs, Dogan et al. [73] proposed an "autozoner" algorithm that can automatically create multi-zone energy models that comply with ASHRAE 90.1 Appendix G [68]. The algorithm can simulate annual load profiles faster and with similar accuracy as multi-zone thermal models that follow ASHRAE90.1 Appendix G guidelines for different types of perimeter and core floorplans [72]. Chen and Hong developed a new pixel-based algorithm that splits any polygon into perimeter and core zones, following the ASHRAE 90.1 standards, which uses a method similar to discrete element method that is usually implemented in fluid dynamic simulations [70]. They named this algorithm "Autozone" and conducted EnergyPlus simulation on DOE commercial reference buildings [45], and found that the AutoZone method and the Prototype method differ in source energy use by −12.1 %–19.0 %, and more so in thermal loads and equipment capacities [70]. Another research has developed a more computationally heavy automated thermal zoning method called the grid/cluster method and conducted simulation comparing the results calculated by one zone and core-perimeter zone methods [69].

The recent advent of automatic techniques for defining thermal zoning in ABEMs has led to new horizons in the realm of energy modeling. The development of these techniques, which make use of cutting-edge computational algorithms and rich building datasets, marks a significant shift from traditional manual processes that are time-consuming and expertise dependent. With different methodologies such as the core-perimeter thermal zoning and the recently proposed methods such as the autozoner algorithm, autozone method, and grid/cluster method, previous research has demonstrated the capability to create multi-zone energy models with higher efficiency and better accuracy to conventional multi-zone models in ABEM.

4. Building codes-based approach

4.1. Building codes-based approach and its application

The building codes-based approach, or the so-called reference building approach, refers to the method of constructing ABEMs that relies on the specifications and requirements set forth in national or regional building codes and standards. Building archetypes constructed based on this method are called reference buildings or prototype buildings. This approach adopts standardized guidelines for building geometry, envelope properties, and mechanical systems. The building codes-based approach also incorporates occupant behavior archetypes to better represent the actual BEC. Occupant behavior archetypes can be developed based on field surveys, time-use studies, or expert opinions and then integrated into the building codes-based ABEMs [75], which is an indispensable component in ABEMs [76]. Many studies have demonstrated the application of the building codes-based approach for constructing ABEMs to support energy-saving measure analysis [67], climate change impact (CCI) [9], and urban building energy modeling [16]. Some building codes-based archetype models have been adopted in studies around the world, such as ASHRAE 90.1 [68], International Energy Conservation Code (IECC) [77], National Energy Code of Canada for Buildings (NECB) [78], and models conforming to national or EU building codes [59,79].

The DOE reference model represents a set of archetype building models that conform to the commercial building energy consumption survey (CBECS) data and residential energy consumption survey (RECS) data [80,81], which have been widely used for evaluating ECMs in residential and commercial buildings [82]. ASHRAE Standard 90.1 [68] and the IECC [83] are integrated into DOE reference models. The two sets of codes have served as a benchmark for energy-efficient building design in the United States. They define the baseline energy performance, including specifications for building envelopes, HVAC systems, lighting, and other energy-related equipment. The adoption of this set of ABEMs has been widely applied to, but not limited to, the purpose of the assessment of CCIs on regional residential [20,84] and commercial energy consumption [85,86], green roof impact on building performance [87], CCIs on building renewable systems [19,88], building sector retrofit scenarios [16,89], and urban level energy use simulation [16,90,91]. Moreover, DOE reference buildings are also adopted in other places when reference buildings are not available in specific places for research purposes including energy efficiency optimization integrated with renewable energy systems in Hong Kong, China [92], heat island modeling in Shenzhen, China [93], model predictive control opportunities in commercial buildings in Canada [94], sensitivity analysis of office building cooling demand in the UK [95], etc.

Building codes-based ABEMs predicated on other national building codes also have been developed and used to fulfill various research purposes. The NECB is a standard that provides minimum energy efficiency requirements for new and retrofit buildings in Canada. It is used by Girgis-McEwen and Ullah to model a ten-story Large Office archetype with a 13,380 m^2 floor area. This research grouped measures into tiers for future NECB editions (2020, 2025, and 2030). Without cross effects, these measures could save 50 % of energy compared to NECB 2015 [96]. Another research evaluated the CCI on a single family archetype building meeting the NCEB codes and revealed that the archetype houses consume less energy in all zones [97]. NREL's Zero Energy Ready Home (ZERH) is a building code that prescribes extra insulation, air sealing, windows, and efficient HVAC equipment and ducts in conditioned space in order to reach the goal of zero energy. Munankarmi et al. modeled a residential community based on their developed archetype house conforming to ZERH and assessed its responsiveness to ECMs and demand management strategies [98]. d

On ABEMs developed based on local building codes, studies have also been carried out around the world. Chen et al. developed prototype buildings based on California Title 24 to simulate urban-level energy use in San Francisco and energy-saving potential by applying various ECMs [99]. Another case study by Ballarini et al. investigated the energy performance of Italian residential buildings by using a set of archetypal buildings constructed based on the national building code. The study analyzed the impact of different ECMs on BEC and identified the most cost-effective strategies for building retrofit [59].

In Asia, research recently conducted by Alhamlawi et al. applied Dubai's green building system - Al Sa'fat to statistically generated building archetypes to understand the energy-saving effectiveness of the building code in various building types [100]. In China, Xiong et al. developed four typical archetypes of residential buildings in China that comply with the national building code GB 50096–1999 and evaluated the CCI on them in China's five representative climate zones by DeST [20]. Deng et al. developed an automated prototype building generator called AutoBPS-OSS given that a building geometry is provided based on non-geometric modeling information including envelope property, building system, program, and schedule stipulated in the Energy Efficiency of Residential Buildings in Hot Summer and Cold Winter Zone (JGJ 134–2001, 2010), and Design Standard for Energy Efficiency of Public Buildings (GB 50189–2005, 2015). It can then be used for urban-level building simulation and energy policy evaluation [16]. In Saudi Arabia, Alardhi et al. developed a residential archetype based on the Saudi energy conservation code (SBC-602) to test its energy reduction impact in three climate zones of Saudi Arabia and found that decreasing wall U-value is the most efficient way to reduce total electricity consumption [101]. All these studies demonstrate the utility of the building codes-based approach for constructing ABEMs and evaluating energy efficiency strategies. With the help of the established building codes-based ABEMs, people are allowed to provide insights into the potential energy savings associated with code compliance, as well as identify opportunities for further efficiency improvements.

4.2. Advantages and limitations

Using building codes as a basis for constructing ABEMs has several benefits. It is capable of providing a consistent and standard framework, which ensures that the developed models meet regulatory standards and minimum performance benchmarks. This uniformity aids in comparing energy performance across various buildings, regions, climates, and scenarios, which purveys a standardized modeling process, making the modeling and simulation more efficient and time-saving. This simplicity is especially beneficial when working with large quantities of buildings or when data resources for model development are scarce [3].

However, building codes-based ABEM also comes with certain drawbacks. One significant downside is the potential failure to capture the real-world diversity and complexity of buildings, as models constructed based solely on this method depend only on model assumptions and standardized parameters. Consequently, the energy performance of ABEMs based on building codes might not accurately depict the real performance of buildings, which can vary due to differences in building vintage, construction quality, maintenance, and occupant behavior [102]. Another limitation is that building codes might not always keep up with the latest energy-efficient design technologies and best practices, which could lead to underestimating potential energy savings. Also, this approach might not be suitable for modeling older or unique buildings that don't comply with current codes or that deviate from standard assumptions [103].

5. Data-driven approach

The data-driven approach in ABEM makes use of extensive datasets to analyze building energy performance, aiming to improve model precision and real-world representation through empirical data utilization. These methodologies rely on extracting insights and predictions from real-world data. The data-driven approach aims to identify representative building archetypes that capture the main features of the building stock and accurately estimate its energy performance [104]. It usually involves a five-stage workflow, which includes data collection, data segmentation, characterization, scaling, and modeling and simulation [27]. There is a growing trend in recent research that during the characterization stage, the application of machine learning algorithm is conducted. Within this framework, traditional statistical techniques refer to a subset of data-driven methods, focusing on regression analyses to discern patterns and correlations in data. This approach is particularly effective for structured data, offering clear insights into the energy usage and characteristics of different building types. Machine learning methods, also under the data-driven methods umbrella, stand for emerging algorithms which utilize supervised and unsupervised methods, including clustering algorithms, neural networks, decision trees, to process and learn from data. Classified under machine learning, clustering methods can be employed to handle unstructured or complex datasets, adeptly identifying intricate, non-linear relationships of dataset. Together, these methods underpin the data-driven approaches in ABEM, ranging from traditional statistical analysis to advanced machine learning modeling.

5.1. Statistical data approach and its application

The statistical data approach for developing ABEM involves the analysis of statistical and empirical data related to building stock, including building geometry, construction materials, building equipment, and BEC. Building archetypes constructed based on this method are also called representative buildings or typical buildings. In this review, the scope of the statistical data approach only refers to those who help in the process of constructing ABEMs, and those studies purely focusing on statistical or empirical analysis of building stocks (though they might be referred to by other studies for ABEM development in) are not considered within the scope of this review.

The statistical data approach can be extended to include occupant behavior archetypes. Field surveys and questionnaires can be adopted to collect data on occupant behavior, which can then be statistically analyzed to identify patterns and create representative behavior archetypes [64]. These occupant behavior archetypes can be integrated into the statistical data-based ABEMs, providing a more comprehensive representation of the building stock's energy performance by accounting for variations in occupant behavior [76]. The key advantage of the statistical data approach is its ability to provide a detailed representation of the building stock based on real-world data, which is particularly helpful when assessing the impact of ECMs and policies as well as evaluating BEC and carbon emission goals on a regional scale building sector [105,106].

5.1.1. Statistical data-based residential ABEMs

The statistical data approach usually involves an analysis of large-scale building datasets. Among those many attempts of developing statistical-based ABEMs around the world, one of the most well-known series of ABEMs is the *Typology Approach for Building Stock Energy Assessment* (*TABULA*) project in Europe [59]. The TABULA project (2009–2012) aimed to create a common structure for building typologies in Europe to estimate and improve the energy performance of residential buildings at the national level. It has developed a set of "exemplary buildings" to represent each building type and a "common calculation method" to compare their energy demand and potential savings. The project involved 13 countries initially and classified buildings by climate zone, building size and shape, building vintage and other parameters by the year 2013 [59]. Now that there are 21 European countries in total have finished their residential building archetypes development in TABULA [104]. Due to the heterogeneity of the large national-level building stocks, buildings are grouped into different categories, eventually forming a building typology matrix for each country as shown in Fig. 3.

Most of the counties participating in the TABULA project developed their building archetypes using a synthetic average building approach. This approach generates a "virtual" building that represents the common features (average values) of envelope thermophysical properties and building system energy efficiency of a group of buildings in the stock based on statistical analysis [107]. Ongoing studies have adopted the building archetypes in TABULA for regional and national energy consumption analysis and evaluation of energy-saving strategies in the building sector, i.e. Czech Republic [108], Italy [59], Greece [109]. In these studies, the estimation of the aggregated performance of the residential building sector is usually realized by statistically enlarging the energy performance of representative subsets in the building archetype matrix based on how often the certain building type appears (frequency) in the area. However, the energy use calculation or simulation of the TABULA building archetypes is usually achieved by a series of steady-state, quasi-steady-state, or simplified dynamic calculation methods that use monthly or seasonal time steps [59], which makes the calculation procedure computationally efficient while may lead to potential loss of fidelity.

Considering the fact that the TABULA building archetype may not be able to capture the characteristics of the local building stock, many studies have developed their statistical data-based residential building archetypes. Research conducted by Attia et al. seeks to create an energy performance dataset and benchmark archetype model for near-zero-energy homes in Brussels, focusing on a post-2010 renovated terraced house by walkthrough audit, in-situ measurements, and utility bill

Fig. 3. Residential building archetype portfolio of individual countries in TABULA [104].

[110]. Near-zero-energy terraced homes are the focus of this study, in which their actual performance was evaluated. These homes have been renovated and have an average energy use intensity (EUI) of 29 kWh/m^2/year, which serves as the benchmark for this type of housing in Brussels. Caputo et al. constructed four typical residential buildings namely semi-detached, line block, tower block, and central patio buildings based on GIS building database and formed an archetype building matrix of 56 buildings after taking building vintage into account in Milan [11], and then used EnergyPlus to evaluate the building retrofit strategies on residential building stock for the city. In Sweden, research has applied 1400 statistically representative single family and multifamily dwellings from the BETSI program [111] to assess the effectiveness of different energy-saving measures on residential building stock, and the buildings are treated as one thermal zone when conducting BEM using ISO 13790 calculation method [67]. Mata et al. also collected 593 archetype buildings covering both residential and non-residential building types from France, Germany, Spain, and the UK, and validated their choice of archetype building by comparing their simulation results to various sources of statistics in the four countries [42]. In Greece, Theodoridou developed four typical multifamily residential buildings according to statistical classifications in four climate zones and compared the EnergyPlus simulation results with actual measurements [112]. In Switzerland, a study developed an archetype database called "SwissRes", which contains 54 residential archetypes generated based on a database of over 25000 buildings. They have discerned that the largest contributor to national heating energy consumption is those buildings constructed before 1980, and the model can further be harnessed for assessing national-scale energy retrofit strategies [14]. In France, Portella and Ribas developed 54 residential archetypes for single family dwellings, and private and public multi-family dwellings based on statistical data from various international and national databases, and they further used them to calculate national building stock energy consumption [113]. In Sweden, Pasichnyi et al. generated residential archetypes buildings for EnergyPlus simulation based on statistical models that segmented and characterized the building stocks in Stockholm, and estimated energy-saving potential from seven retrofitting packages [28].

In places other than Europe, Shen et al. developed five residential archetypes according to the RECS database in the United States for the state of Texas [9] and New York [89]. The five archetype buildings were fine-tuned based on the survey data in the two states in terms of building sizes, energy utility type, and HVAC system end-use types, and the models were further calibrated on the actual per capita energy use data. The validated ABEMs in the two states were then utilized for regional energy consumption projection under various scenarios including future climate conditions, urbanization and population growth, and adoption of ECMs. In Chongqing, China, Li et al. used statistical approach to derive the socioeconomic, building vintage, and shape factors from the 2012 census data of Chongqing Statistics Bureau, and generated 27 residential archetypes covering three different construction ages [114].

5.1.2. *Statistical data-based ABEMs for non-residential buildings*

There are studies and examples demonstrating the effectiveness of the statistical data approach in constructing ABEMs for various building sectors and regions [115]. Table 1 lists studies related to statistical data-based non-residential ABEMs in different places around the world. Some of those studies produced ABEMs at urban or regional scale [116, 117], while others at a national level [42,53,113,118–124]. The most frequently categorized features for those developed ABEMs are building shape-related parameters, followed by building vintage, envelope thermal properties, and operation schedule. Other factors are ad-hoc to specific building types such as retail, hotel, and educational building. The major purposes for developing the non-residential ABEMs are to understand the energy consumption patterns of building stocks, to figure out the most sensitive parameters to BEC of certain building type, and to

Table 1
Summary of ABEMs of non-residential buildings developed around the world.

Building Type	Reference	Year	Country	Database	Number of Archetypes	Data Analyzed	Purpose
Office	[119]	2013	Italy	65000 offices from ENEA (Italian National Agency for New Technologies Energy and Sustainable Economic Expansion) database	3 (Type A: statistical data-based ABEM)	Building form, envelope and building system features, climate zone	Compare simulation results among ABEMs constructed based on different methods
Office	[118]	2021	Kuwait	463 buildings in Kuwait	3	Built-up area, number of floors, vintage, AC type, building schedule, EUI, thermostat setpoints	Provide recommendations for Kuwait's addressing educational, technological, and policy challenges of the building sector
Office, hotel, mall, mixed-use	[116]	2019	China	200 commercial buildings	6	Building shape and zoning, Window-to-Wall Ratio (WWR), envelope thermal properties	ABEM development and energy use analysis
Educational	[120]	2021	England	9551 educational buildings	168 (5 seed models)	area of building footprint, number of floors, average WWR, vintage	Assessment of national fossil-thermal energy consumption of primary school and natural ventilation performance
Educational	[53]	2021	Brazil	298 schools	7	Building shape, vintage, area of floor plan, number of students	Point out that building shape can be important when generating bottom-up benchmarking models
Retail	[117]	2013	Japan	5869 retail facilities	14	Building stock category, sales area, gross floor area, number of stories, store hours, lighting and equipment load	Validate the energy simulation results of models with surveyed mean EUI, and calculate the total BEC of the retail sector in Keihanshin area
Mix-use hotel	[124]	2017	Italy	Tabula and Hellenic Tourism Organization (referred from Ref. [125])	1	Building size, vintage, and hotel opening period	Analyzing the energy use patterns of the developed archetype building
Commercial, leisure, office, sports, culture and leisure (SCL), and other	[42]	2014	France, Spain, UK	209 non-residential archetype buildings	45 (France), 80 (Spain), 84 (UK)	Building type, vintage, heating system, climate zone	Compare the simulated results of archetypes with statistical data from various sources and evaluate the sensitivity of modeling parameters on energy demand
commercial, office, sports and leisure, and others	[121]	2008	Spain	Instituto Nacional de Estadística, Ministerio de Fomento, IDEA, Eurostat, GAINS, IEA	80	climate zone, building vintage, thermal properties of envelope, HVAC, building program	Cross-database comparison and national BEC calculation
Retail, accommodation, restaurant, private office, public office, hospital, university, school	[122]	2017	Korea	Survey of the Ministry of Trade, Industry and Energy (MOTIE) and the Ministry of Land, Infrastructure and Transport (MOLIT) in South Korea	8	Floor area, number of floors, aspect ratio, WWR, EUI	Develop reference buildings for national building stock
Restaurant, school, hospital, office, shop	[123]	2020	Switzerland	31662 non-residential buildings	45 in total (9 for each building type)	Floor area, building height, and building period of all buildings	Understanding BEC profiles of the Swiss building stock and the effect of retrofitting strategies on the stock
Office, commercial, health, education	[113]	2012	France	Multiple international and national database	45	Floor area, ventilation rate, envelope U-value, etc.	Development of residential and non-residential building archetypes for national building stock energy use calculation

evaluate the effect of ECMs on building stocks.

The studies around the world further demonstrate the versatility and applicability of the statistical data approach in constructing ABEMs for different building sectors and at various scales. By using large-scale building data sets to identify representative building archetypes, this approach better informs decision-makers of the BEC profiles of various building typologies, and supports energy policy evaluation and decision-making processes, allowing for the development of targeted ECMs and strategies in the non-residential building sector.

5.2. Applications of machine learning methods

The application of machine learning techniques has emerged as promising approaches for generating ABEMs based on statistical data due to their ability to handle large-scale, complex, and diverse datasets. These techniques are particularly useful for identifying patterns and relationships in data that may not be easily discernible through traditional statistical methods [52] and are usually implemented during the characterization stage of statistical data approach when generating ABEM. It should be noted that studies involving the application of machine learning algorithms to building stocks while didn't take into account the engineering-based building archetypes are not included in the scope of this review. Generally, machine learning methods applied in archetype generation can be broadly categorized into supervised and unsupervised learning techniques.

Supervised learning techniques, such as regression and decision trees [126], require labeled data to train models that can predict outcomes based on input features. In the context of ABEMs, supervised learning

Table 2
Studies applying machine learning methods in ABEMs.

Author (s)	Country	Year	Method Used	Target Parameters	Application and Purpose
[127]	Belgium	2016	Agglomerative hierarchical clustering	Total loss area, building vintage, heated floor area, footprint, WWR, number of floors, loss-to-floor ratio, compactness, building typology, window area	Develop ABEMs for urban energy modeling
[132]	Italy	2014	Hierarchical Clustering	Building form, vintage, heated gross volume, net floor area, envelope thermal properties, energy demand	Development of residential building archetypes
[128]	Switzerland	2018	PCA, k-Means Clustering, Partitioning Around Medoids, Hierarchical Clustering	building typology information, vintage, building usage	Identification and characterization of representative buildings in urban datasets
[133]	United States and Denmark	2012	Gaussian Mixture Models with Expectation-Maximization	Building vintage and EUI	Produce synthetic building energy data
[93]	China	2021	Gaussian Mixture Models with PCA and BIC	11 parameters related to buildings geometry and number of floors	Urban climate modeling using GIS data
[131]	Greece	2010	Clustering and Probabilistic Techniques	Heated surface, building vintage, building insulation level, number of classrooms, number of students, operating hours, heating system vintage	Identification of representative school buildings
[27]	Ireland	2019	k-Means Clustering	Construction material, usage patterns, and building systems	Comparing modeling results of multi-scale building archetype at building and urban levels
[54]	Italy	2015	k-Means Clustering	12 features including buildings shape and size, WWR, envelope thermal properties, heating system capacity	Identification of representative school buildings from a dataset of 60 buildings
[130]	Brazil	2016	Hierarchical and k-Means Clustering	16 features related to floor area, information related to space function, number of floors	Obtaining building archetypes for low-income housing stock
[134]	Austria	2017	Hierarchical Clustering, k-Means Clustering, and Model-based Clustering	17 variables related to building geometry, solar gain, envelope thermophysical properties, and usage patterns	Reductive urban energy modeling
[44]	China	2022	Clustering, Random Forest, Convolutional Neural Network	Building shape, building type, vintage	Urban building energy modeling
[135]	China	2022	Stratified Sampling, k-Means Clustering	Building vintage, number of floors, building shape, floor areas, WWR	provide building geometric information and characteristic-based evaluations for ABEM generation and further analysis on building performance
[126]	China	2023	k-Means Clustering, CART Decision Tree	Occupant behavior pattern, socio-demographic characteristics	Reduce the performance gap between the simulation and actual residential buildings' energy consumption
[136]	China	2022	k-Means Clustering and Random Forest Classifier	Building vintage, typology, floor area, number of floors, building volume, surface area, shape factor, envelope thermal properties	Modeling and simulation of urban building energy performance
[57]	China	2018	k-Means Clustering and Partitioning around medoids	Building shape	Validate with district-level EUI
[137]	United States	2017	k-Modes Clustering and Probabilistic Neural Network	Occupant behavior and schedule	Show the difference in energy use of the generated archetype against that with the ASHRAE standard occupancy schedule
[138, 139]	United Kingdom	2019	k-Modes Clustering	Occupancy schedule	Showcase the difference in heating demand between ordinary building archetypes with uniform occupancy schedules and archetypes with occupancy-integrated schedule
[29]	Italy	2022	k-Means Clustering	Occupancy schedule	Develop data-driven schedules for ABEMs in urban energy modeling
[140]	Japan	2019	Logistic Regression	HVAC system	Improve the reliability of urban building energy modeling and evaluate energy efficiency technologies on HVAC
[141]	China	2021	k-Means clustering	Building footprints and primary building type	Urban energy modeling
[123]	Switzerland	2020	k-Medoids Clustering	Floor area, height of the building, and building operation	Understanding BEC profiles of Swiss buildings and retrofitting effects

techniques have been used to identify the parameters that are worth being featured in unsupervised learning [44,127]. In De Jaeger et al.'s research, they conducted a regression analysis and Akaike Information Criterion (AIC) to identify the most influential parameters for ABEM [127]. In Deng et al.'s research, Random Forest and Convolutional Neural Network were utilized to determine the building typologies and building vintage based on GIS and historical satellite image data [44].

However, the majority of the work done in building archetype generation is predicated on unsupervised learning. Unsupervised learning techniques, such as clustering algorithms, do not require labeled data and are used to group similar objects based on their attributes, which are gaining popularity in recent studies for the characterization of the building stock and building archetype identification. Clustering techniques like k-Means, hierarchical clustering, and k-Modes clustering have been employed to identify representative building archetypes from large datasets. Since many studies adopted the clustering methods in ABEM, they have been listed and summarized in Table 2. For clustering problems, the identification of the proper number of clusters can be achieved by algorithms such as AIC, Bayesian Information Criterion (BIC), and Partitioning around medoids [57,93,127–129]. The normalization and dimensional reduction algorithms for data-preprocessing that are generally involved before conducting clustering include Z-score [57,130], Principal Component Analysis (PCA) [128,131], and Box-Cox [128].

Machine learning techniques also offer the potential to capture occupant behavior patterns more accurately and efficiently than traditional methods [142,137]. By analyzing large datasets collected from field surveys, smart meters, or sensor networks, machine learning algorithms can identify occupant behavior archetypes and their impact on BEC [64,143]. By incorporating occupant behavior archetypes in building codes-based, statistical data-based, and machine learning methods, ABEMs can provide a more accurate and comprehensive representation of the energy performance of building stocks. Field surveys and other data collection techniques can be employed to construct occupant behavior archetypes, which can be integrated into the respective modeling approaches [29,62]. This additional layer of detail can lead to improved energy analysis and policy evaluation, ultimately supporting more effective and targeted ECMs.

5.3. Advantages and limitations

The intention behind using a data-driven approach is to facilitate the identification and analysis of trends in energy consumption across various building sectors and regions. By analyzing large-scale building data sets, researchers can identify patterns and correlations between different building characteristics and their energy performance. This can help decision-makers to develop targeted strategies for improving energy efficiency in the building sector. Moreover, the data-driven approach can contribute to the development of more accurate and reliable benchmarking tools for comparing the energy performance of different building types. This can be particularly valuable for energy service companies and utilities, which often require reliable and detailed building energy performance data to support their energy efficiency programs and initiatives.

While machine learning techniques offer several advantages for generating ABEMs, some limitations need to be considered. Ensuring data quality and availability, addressing the complexity and interpretability of models, managing computational resources, and avoiding overfitting are key challenges that need to be addressed when applying machine learning techniques to ABEMs. The data-driven approach for constructing ABEMs has several advantages and limitations that should be considered when evaluating its applicability in specific contexts, which is summarized in Table 3.

6. Hybrid approach

6.1. Overview of the hybrid method

The hybrid method combines the strengths of building codes based approach and data-driven approach to create more accurate and representative ABEMs. This integrated approach leverages the advantages of each method to overcome its limitations, resulting in more comprehensive and reliable ABEMs for various applications. One of the implementations can be starting with the building codes-based approach, which provides a solid foundation based on established building regulations and standards for non-geometric modeling parameters. The resulting archetypes serve as a starting point for further refinement using statistical data-based and machine-learning techniques. Then the data-driven approach can be applied to adjust building geometry, construction materials, building occupancy and load, and system efficiencies based on real-world building stock data. This process allows for a better representation of actual building stock characteristics and the identification of trends in energy consumption patterns meanwhile providing convenience in ABEM without heavy reliance on data sources of existing building stocks. Vice versa, researchers can also choose to start their development of ABEMs using a data-driven

Table 3
Advantages and limitations of data-driven approach.

Advantage		Limitation	
Real-world data representation	Rely on actual data, which provides a more accurate representation of the existing building stock's energy performance compared to other modeling methods based on assumptions or idealized building characteristics; It ensures ABEMs mirror the actual scenario and models are grounded in reality	Data availability and quality	Rely heavily on the availability and quality of building data, which can be limited or inconsistent across different regions and building sectors. This can affect the reliability and accuracy of the ABEMs developed using this approach
Scalability and adaptability	Suitable for modeling the energy profile of entire cities or regions and be applied to diverse building types and datasets, allowing for a more comprehensive understanding of the building stock	Model complexity	Most of the data-driven ABEMs are modeled by a single thermal zone, which can be computationally light whilst lose real-world representation and model predictability
Identification of trends and patterns	Enable the identification of trends, patterns, and correlations between building characteristics and energy performance; results and data can be analyzed across ABEMs to uncover overarching energy usage insights and inform policy under "what-if" scenarios	Model generalizability	May not always produce generalizable results, as the ABEMs developed are specific to the analyzed building stock, limiting their applicability to other contexts or regions without additional calibration and validation efforts
Support for benchmarking	Contribute to the development of accurate and reliable benchmarking tools for comparing the energy performance of different building types, which is essential for energy efficiency initiatives and programs	Computationally intensive and overfitting	Some machine learning techniques (i.e. hierarchical clustering), can be computationally intensive, requiring significant processing power and time to train and evaluate models; some algorithms may sometimes overfit the data, capturing noise rather than the underlying patterns, which can lead to poor generalization to new, unseen data

approach, characterizing the building archetype based on the database in the first place. This ensures that factors such as building forms and envelope materials of the developed archetype conform to the local features of the building stocks. Then building codes-based approach can be used to fill out those necessary modeling parameters that are missing or difficult to acquire from the existing database. Either of these approaches have been implemented in the existing studies when generating ABEMs in various places around the world.

Moreover, in the hybrid method, machine learning techniques can be incorporated to further refine the ABEMs by identifying the characteristics of building attributes and energy consumption that may not be apparent in the building codes or statistical data. Clustering and regression techniques can be used to categorize buildings with similar characteristics and model the complex interactions between modeling parameters and energy performance. This step enhances the accuracy and granularity of the baseline building codes-based ABEMs, providing insights for energy consumption pattern recognition, targeted energy efficiency interventions, and policy development of building stocks.

6.2. Application of the hybrid method

Several studies have demonstrated the effectiveness of the hybrid method in generating ABEMs. For instance, in Wuhan, China, Ding and Zhou used the geometries of DOE reference buildings together with the occupancy schedule data extracted from China Residential Energy Consumption Survey (CRECS) to construct the residential and office building archetypes and validated the EnergyPlus simulation results against surveyed mean energy consumption data in CRECS [18]. In Changsha, China, Deng et al. utilized machine learning algorithms on segmenting building type, vintage, and shape based on GIS dataset and generated 22 building archetypes to represent 59332 buildings in the city. Later, ASHRAE 90.1 standards are applied to these archetypes and used for urban energy use simulation by EnergyPlus [44]. In Chongqing, China, Li et al. made use of the clustering technique on categorizing buildings based on their geometric information including height, compactness ratio, and aspect ratio. Then building code of China's Residential Buildings Design Standards JGJ 134–2001 is referred to in assigning non-geometric modeling parameters to the EnergyPlus model, such as envelope thermophysical properties, HVAC system coefficient of performance and HVAC set points, and internal load gains [57]. Moreover, in another research, when developing residential ABEM, the same team also attempted to use statistical methods on the census data from the 2012 Chongqing Statistics Bureau to determine the socio-economic, building vintage and shape factors, and further integrated the JGJ 134–2001 on the extracted archetypes for retrofit measure evaluation [114].

Research conducted in England developed 168 school archetypes based on national survey data concerning building footprint area, number of stories, average WWR, and vintage. After that, they applied building codes (National Calculation Methodology and Building Bulletin 101 in England) to model parameters such as internal loads, ventilation, etc. EnergyPlus is used to simulate building performance regarding BEC, indoor air quality, and natural ventilation. In Korea, Kim et al. developed 11 reference buildings for national building stock covering residential and non-residential archetypes [122] based on national survey data concerning floor area, number of floors, aspect ratio, WWR, and EUI. Then they applied building codes to non-geometric parameters such as envelope U-value, infiltration, ventilation, internal loads, domestic hot water, and arbitrary parameters including HVAC system efficiency, and thermostat setpoint [122]. In Italy, a reference hotel building is developed based on a hybrid approach integrated with statistical method (building form and building system) and building codes method (building operation from EU 15232 and DOE reference building) [124].

The hybrid method, which combines building codes based, statistical data-based, and machine learning approaches, offers a powerful and flexible framework for generating ABEMs. By leveraging the strengths of each approach, the construction of ABEMs in places with a variety of imparities in data availability and quality can be achieved since one of the most prominent advantages of the hybrid approach is its flexibility when dealing with limited data sources and availability. Nevertheless, the selection of which approach to be adopted is subject to the various research purposes and factors.

6.3. Factors influencing the choice of method

The selection of an appropriate approach for generating ABEMs is a critical decision that is shaped by the interplay of several key factors. Each approach has its strengths and limitations, which can affect its suitability for a given research or policy context. Understanding these considerations can help make informed decisions that align with their objectives, data availability, and computational complexity.

Research Purpose and Desired Model Accuracy: The objectives driving the ABEM analysis and the precision required from the models are closely linked, guiding the selection of the modeling approach. Projects aimed at informing policy development or evaluating energy efficiency measures demand high accuracy and detailed representation of building energy behaviors. In such cases, hybrid methods, which combine the building code specified assumptions for less sensitive model parameters in the chosen region with the empirical information provided by data-driven methods, are particularly valuable. They provide a comprehensive understanding, suitable for forming the basis of policy or strategic energy interventions. Conversely, studies with broader, more exploratory, and goal-driven aims, such as future building code development compared to the current building code, might prioritize a wider overview over granular accuracy. For these purposes, simpler approaches such as building codes-based methods might suffice, offering general insights without delving into the complexities of existing building performance. Essentially, the choice between a detailed, high-accuracy model and a more generalized overview depends on the research's intended impact.

Data Availability and Quality: The type, quality, and quantity of available data can directly impact the choice of the ABEM approach. Building codes-based methods are particularly useful when comprehensive building codes exist, while data-driven methods are more effective when large, high-quality datasets on building characteristics and energy performance are available.

Computational Complexity and Resources: The computational demands of different methods vary. More sophisticated methods, like some of the machine learning approaches, may require more advanced computing resources and technical expertise. Thus, the selection of a method should consider the available computational resources and the balance between the complexity of the method and the desired model accuracy.

It can be a concern that while the hybrid method aims to synergize the strengths of both building codes-based and statistical data approaches, it also risks compounding their individual limitations. Recent studies suggest that a judicious combination of methods, grounded in a deep understanding of their inherent biases and constraints [144] while leverage emerging techniques such as GIS technology and smart meter for more precise representation of socioeconomic details, can help mitigate these risks [143], producing models that more faithfully represent actual building stocks without necessarily amplifying the limitations of the respective approaches. This balance underscores the importance of methodological diversity in the ABEM development process.

7. Summary of modeling approaches

The summary provided here in this chapter aims to offer a comprehensive overview that enhances understanding and delineates their applications and potential of these methodologies.

7.1. Building codes-based approach

This method stands out for its use of established regulatory frameworks as the foundation of modeling, ensuring ABEMs adhere to minimum or required performance standards. Its key advantage lies in its ability to facilitate standardized comparisons of energy performance, offering an efficient modeling process that is ideal for large-scale applications. However, its significant limitation is the potential to overlook the possible diversity and complexity of the status quo of real-world building stocks.

7.2. Data-driven approach

The data-driven approach in ABEM generation leverages extensive datasets to illuminate building energy performance trends, enabling targeted energy efficiency strategies and policy development. This approach can be further categorized into classic statistical methods and machine learning techniques, each bringing distinct advantages and confronting unique challenges.

7.3. Classic statistical methods

Rooted in traditional statistical analysis, this segment of the data-driven approach focuses on regression analyses and other statistical tools to identify correlations and patterns within building data. Its strength lies in providing clear, interpretable models that offer insights into energy usage and the characteristics of buildings. Classic statistical methods are particularly effective for structured datasets and are valued for their straightforward analysis and ease of understanding. However, they may not be able to fully capture the complexity of energy consumption behaviors, especially in the presence of non-linear relationships or when dealing with large, unstructured datasets.

7.4. Application of machine learning techniques

Under the umbrella of data-driven approaches, machine learning techniques employ algorithms with various complexities to process and learn from data. This includes both supervised and unsupervised learning methods, such as decision trees, neural networks, and clustering algorithms. They are good at handling unstructured or complex datasets, uncovering intricate, non-linear relationships that classical statistical approaches might overlook. They can offer enhanced model accuracy and the ability to predict energy performance with greater details. The challenges associated with machine learning techniques involve ensuring data quality and availability, computational resources, model interpretation, and mitigating the risk of overfitting.

7.5. Hybrid approach

Aiming to combine the strengths of building codes-based and data-driven approaches (including both classic statistical methods and machine learning techniques), the hybrid method has the potential to offer increased modeling flexibility and efficiency. It can represent actual building stocks more accurately by mitigating individual method limitations. However, researchers should be cautious with the effectiveness of this approach since it depends on balancing the strengths and weaknesses of different methods and has a case-by-case nature, which is highly contingent upon specific research purposes.

Considering that each method offers unique advantages suited to different research purposes, data availabilities, and computational resources, the selection among these approaches should be informed by a careful consideration of these factors, ensuring that the chosen method aligns with the specific objectives and constraints of the project at hand and catering to the varied needs of analysis and policymaking in pursuit of sustainable urban environments using ABEM.

8. Challenges and future research

8.1. Data availability and model accuracy

One of the primary challenges in the development of ABEMs is the availability and quality of data used in the modeling process. Data availability and uncertainty are closely intertwined with model accuracy, as the quality of input data directly influences the accuracy and reliability of the resulting ABEMs [145]. The data required for ABEMs typically include information on building geometry, construction materials, energy systems, occupancy patterns, and local climate conditions, among others. However, obtaining accurate and representative data for these parameters can be challenging, particularly for large and diverse building stocks or in cases where detailed data are protected due to privacy concerns [27]. The use of public databases, national surveys, and regional or local building registries can help address data availability issues. However, the quality and granularity of such data sources can vary significantly, potentially leading to inaccuracies or uncertainties in the resulting ABEMs [146]. To mitigate this, researchers have developed various approaches to estimate or impute missing or uncertain data, including statistical methods, expert judgments, and machine learning techniques such as DBSCAN and autoencoders for data noise elimination [147].

Furthermore, researchers should validate and calibrate their ABEMs against measured BEC data or other established benchmarks to ensure the accuracy and credibility of their models, and many efforts have been made [33,40,51,145,148]. This is extremely important in identifying potential errors or inconsistencies in the input data or modeling assumptions, enabling the refinement and improvement of the ABEMs. Hence, model calibration and validation are essential steps in ensuring the accuracy and credibility of ABEMs. However, these processes can be challenging due to various factors, including data limitations, computational complexity, modeling uncertainties, and the complexity of building energy systems.

One of the main challenges in model calibration is the availability of high-quality, measured BEC data to compare against the model's predictions [149]. In many cases, measured data may be unavailable, incomplete, or protected due to privacy concerns [150,151]. Obtaining accurate and representative data is crucial for a successful calibration, as discrepancies between the model's output and the actual BEC can help identify areas for improvement in the model. Another challenge is the inherent uncertainty in building energy modeling. ABEMs rely on numerous input parameters and assumptions, such as building geometry, construction materials, energy systems, occupancy patterns, and climate conditions. Uncertainties in these inputs can propagate through the model, leading to inaccuracies in the predicted BEC [152]. Researchers should conduct sensitivity analyses to identify the most influential parameters and prioritize efforts to improve the accuracy of these inputs [153,154]. The validity of a calibrated model depends on the representativeness of the calibration dataset. Future research could focus more on the development of standardized data protocols, which would schematize the model creation process, improve the accuracy and reliability of models, and facilitate model comparison and validation. The standardized data protocols in this case refer to the creation of a uniform set of rules and procedures for data collection, treatment, and use of survey structure for ABEMs. The focus of this standardization is on the assurance of consistency, replicability, and comparability among datasets and models from different sources. Data acquired in that manner can be consistent as they employ the same methodologies of definition and units across various projects. The repeatability of data processing will allow researchers and engineers to use similar processing and cleansing techniques, which makes it possible to compare models built on various datasets. Moreover, the advent of smart meters and IoT devices in buildings could pave the way for more precise data collection [149], and future research can explore efficient ways to integrate this detailed and real-time data into ABEMs.

8.2. Computational complexity and scalability

The energy dynamics of cities and communities are influenced by the interaction between numerous buildings, infrastructure systems, and environmental factors. Recently, there is a growing number of studies focusing on the topic of urban energy modeling [23,24,152]. Future research will continue advancing and applying ABEMs at the urban scale, enhancing our understanding of urban-level energy use patterns and identifying opportunities for energy savings and resilience at community-level urban planning. This approach will support policy decisions with quantifiable, spatial-temporal data on energy consumption across the building sector in the future. In this context, computational complexity and scalability are critical aspects to consider when developing and applying ABEMS for this purpose, especially when dealing with large building stocks or high-resolution simulations. These challenges arise due to the increasing amount of data, the complexity of building energy systems, and the need for accurate and reliable results in a reasonable time frame [155]. Model calibration can also be computationally demanding, particularly when dealing with large building stocks or complex energy systems [66]. Iteratively adjusting model parameters to minimize the discrepancy between model predictions and measured data can require significant computational resources and time.

Advanced optimization algorithms and machine learning techniques can help with the calibration process and improve the overall efficiency of the modeling workflow (i.e. artificial neural network [156]). The computational complexity of ABEMs is influenced by several factors, such as the number of modeled buildings, the level of detail and resolution, the simulation timestep, and the choice of modeling techniques [155]. For example, detailed BEM tools, such as EnergyPlus, TRNSYS, or IES VE, require significant computational resources and time to simulate large building stocks, which can become a limiting factor for their practical application in urban planning and policy analysis. To address the computational complexity issue, previous research has proposed various strategies, such as using simplified or surrogate models, parallel computing, cloud-based simulation services, and machine learning techniques [49,155]. Simplified models, such as engineering equations or regression-based methods, can provide faster simulations at the expense of reduced accuracy. Parallel computing and cloud-based services can help distribute the computational workload across multiple processors or servers, significantly reducing the simulation time. Machine learning techniques, such as artificial neural networks and support vector machines, can provide fast and reasonably accurate predictions, but they may require extensive training datasets and fine-tuning of model parameters [157]. As building stocks continue to grow, and the demand for detailed building energy analysis increases, the need for scalable modeling solutions becomes more critical. Scalability can be achieved by developing modular and hierarchical modeling approaches, automating the generation of input data, and adopting standards for data exchange and interoperability [56,157].

8.3. Advances in modeling techniques and demands

Recent advances in simulation tools and modeling techniques have the potential to address some of the challenges faced in ABEM and improve the accuracy, efficiency, and usability of these models. One of the developments is the integration of BIM and GIS into BEM tools, which can streamline the data exchange process and enhance the ability to model complex urban environments [44,49,143,158]. BIM provides detailed building data and geometry, while GIS offers spatial context and environmental data, such as urban texture, terrain, and sky view factor. Combining these tools can help create better quality and comprehensive ABEMs.

As recent research integrates machine learning techniques into the ABEM more frequently, the integration of machine learning and artificial intelligence into ABEMs has the potential to greatly enhance their capabilities. They could be used for clustering building characteristics, automating the model creation process, optimizing the selection of model parameters, and even learning from previous simulations to improve future ones. There is also potential for exploring hybrid modeling approaches that combine the strengths of building codes-based and data-driven methods. Such approaches could provide a balance between model accuracy, data availability, and research purposes and adapt to the specific needs of different applications.

Looking forward, with the rise of renewable energy technologies, buildings are increasingly not merely consumers of resources, but also producers. In the future, the integration of renewable energy systems and the changing grid dynamics may pose new challenges and opportunities for ABEMs [159]. Buildings are increasingly becoming active participants in the energy grid, with the ability to generate and store energy, as well as adjust demand in response to grid conditions. This requires ABEMs to not only model the BEC and production within the building but also interact with the grid. This could add another layer of complexity to the modeling process but also opens up new avenues for energy savings and grid resilience.

9. Conclusion

This review has provided a comprehensive examination of the various methods employed in the generation of engineering-based archetype building energy models (ABEMs), detailing their merits and limitations, and exploring how different approaches are better suited to particular scenarios and objectives. This research investigated the various factors that influence the choice of method for generating ABEMs, including the specific objectives of the research, data availability, the required model accuracy, and computational resources. Each method brings its unique advantages and constraints, and the choice of method should thus be strategically made after a careful evaluation of these factors. Future research in ABEMs will undoubtedly deal with these factors, but as technological advancements continue to improve our capabilities for data collection and processing, further developments within the field of ABEMs are anticipated. To summarize, the main findings of this review are as follows.

- The building codes-based approach offers a standardized modeling framework aligned with regulatory standards, facilitating energy performance comparisons given different external scenarios (e.g. climate change impact, population and building stock changes). However, it is not able to fully capture the diversity of real-world buildings.
- The data-driven approach, including both classical statistical methods and machine learning techniques, allows for the detailed representation of real-world building stocks. They are good at identifying complex patterns and trends in energy consumption but require high-quality and extensive datasets.
- The hybrid approach can combine the strengths of building codes-based and data-driven approaches to improve model accuracy and reliability, but users should also be aware of the aggregation of their limitations. The hybrid approach can offer balanced and comprehensive modeling solution.
- The technological advancements in leveraging smart metering and IoT devices are anticipated to vastly improve the quality and granularity of data for ABEMs. This evolution could lead to models that more accurately reflect real-time energy dynamics and occupant behavior, enhancing model precision and applicability.
- There's a growing emphasis on addressing computational complexity and scalability, particularly for urban-scale ABEM applications. Future strategies may include the development of more efficient modeling algorithms, the adoption of cloud-based simulation platforms, and the exploration of more advanced model calibration methods, aiming to balance accuracy with computational feasibility.

- The integration of Building Information Modeling (BIM) and Geographic Information Systems (GIS) with ABEMs is expected to advance the capability to model complex urban energy systems, which provides better spatial-temporal analysis and improve the understanding of building energy flows within urban contexts.
- The expansion of ABEMs to incorporate renewable energy systems and grid interaction reflects the evolving role of buildings from mere energy consumers to active energy participants. This shift necessitates models that can deal with the complexities in onsite energy production, storage, and demand response, offering new pathways for enhancing energy efficiency and grid resilience.

As ABEMs evolve, their potential to contribute to sustainable urban development and to inform energy policies will continue to grow. The ability to generate accurate, representative ABEMs is growing with the ongoing improvements in data availability and computational resources. ABEMs will continue to play important roles in informing engineering design, influencing regulations, optimizing energy systems, guiding policy decisions, and contributing to finance and environmental, social, and governance strategies by providing a quantitative basis for decision-making and strategy development.

It should also be noted that this review's scope can be limited by its dependence on available literature and existing datasets, which may not encompass the latest developments in ABEMs or the full diversity of building stocks worldwide. Such constraints could potentially affect the comprehensiveness of our analysis and the applicability of our findings across different regions and building types. Furthermore, the dynamic nature of energy policies, technological advancements, and sustainability practices necessitates ongoing research to ensure the relevance and accuracy of our conclusions. Therefore, future work would benefit from incorporating emerging studies and data to maintain its timeliness and validity in guiding energy-efficient building design and policy.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is partly supported by the National Natural Science Foundation of China (NSFC) (Grant No. 52008132) and Shenzhen Science and Technology Program (Grant No. 20220810173620003).

References

[1] International Energy Agency. Global status report for buildings and construction 2019. UN Enviroment programme 2019;224.
[2] Li X, Wen J. Review of building energy modeling for control and operation. Renew Sustain Energy Rev 2014;37:517–37.
[3] 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.
[4] Torcelini P, Deru M, Griffith B, Benne K, Halverson M, Winiarski D, et al. DOE commercial building benchmark models. Golden, CO (United States): National Renewable Energy Lab.(NREL); 2008.
[5] 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.
[6] Fonseca JA, Nguyen T-A, Schlueter A, Marechal F. City Energy Analyst (CEA): integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts. Energy Build 2016;113:202–26.
[7] Abbasabadi N, Ashayeri M. Urban energy use modeling methods and tools: a review and an outlook. Build Environ 2019;161:106270.
[8] Keirstead J, Jennings M, Sivakumar A. A review of urban energy system models: approaches, challenges and opportunities. Renew Sustain Energy Rev 2012;16: 3847–66.
[9] Shen P, Yang B. Projecting Texas energy use for residential sector under future climate and urbanization scenarios: a bottom-up method based on twenty-year regional energy use data. Energy 2020;193:116694.
[10] Oraiopoulos A, Hsieh S, Schlueter A. Energy futures of representative Swiss communities under the influence of urban development, building retrofit, and climate change. Sustain Cities Soc 2023;91:104437.
[11] Caputo P, Costa G, Ferrari S. A supporting method for defining energy strategies in the building sector at urban scale. Energy Pol 2013;55:261–70.
[12] Cerezo C, Sokol J, AlKhaled S, Reinhart C, Al-Mumin A, Hajiah A. Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): a residential case study in Kuwait City. Energy Build 2017; 154:321–34.
[13] Ramokone A, Popoola O, Awelewa A, Temitope A. A review on behavioural propensity for building load and energy profile development – model inadequacy and improved approach. Sustain Energy Technol Assessments 2021;45:101235.
[14] Streicher KN, Padey P, Parra D, Bürer MC, Schneider S, Patel MK. Analysis of space heating demand in the Swiss residential building stock: element-based bottom-up model of archetype buildings. Energy Build 2019;184:300–22.
[15] Filippi Oberegger U, Pernetti R, Lollini R. Bottom-up building stock retrofit based on levelized cost of saved energy. Energy Build 2020;210:109757.
[16] Deng Z, Chen Y, Yang J, Causone F. AutoBPS: a tool for urban building energy modeling to support energy efficiency improvement at city-scale. Energy Build 2023;282:112794.
[17] Abd Alla S, Bianco V, Tagliafico LA, Scarpa F. Life-cycle approach to the estimation of energy efficiency measures in the buildings sector. Appl Energy 2020;264:114745.
[18] Ding C, Zhou N. Using residential and office building archetypes for energy efficiency building solutions in an urban scale: a China case study. Energies 2020; 13:3210.
[19] Shen P, Lior N. Vulnerability to climate change impacts of present renewable energy systems designed for achieving net-zero energy buildings. Energy 2016; 114:1288–305.
[20] Xiong J, Guo S, Wu Y, Yan D, Xiao C, Lu X. Predicting the response of heating and cooling demands of residential buildings with various thermal performances in China to climate change. Energy 2023;269:126789.
[21] Reyna JL, Chester MV. Energy efficiency to reduce residential electricity and natural gas use under climate change. Nat Commun 2017;8:14916.
[22] Brøgger M, Wittchen KB. Estimating the energy-saving potential in national building stocks – a methodology review. Renew Sustain Energy Rev 2018;82: 1489–96.
[23] Ferrando M, Causone F, Hong T, Chen Y. Urban building energy modeling (UBEM) tools: a state-of-the-art review of bottom-up physics-based approaches. Sustain Cities Soc 2020;62:102408.
[24] Reinhart CF, Cerezo Davila C. Urban building energy modeling – a review of a nascent field. Build Environ 2016;97:196–202.
[25] Dahlström L, Broström T, Widén J. Advancing urban building energy modelling through new model components and applications: a review. Energy Build 2022; 266:112099.
[26] Nayak BK, Sansaniwal SK, Mathur J, Chandra T, Garg V, Gupta R. A review of residential building archetypes and their applications to study building energy consumption. Architect Sci Rev 2023:1–14.
[27] Ali U, Shamsi MH, Hoare C, Mangina E, O'Donnell J. A data-driven approach for multi-scale building archetypes development. Energy Build 2019;202:109364.
[28] Pasichnyi O, Wallin J, Kordas O. Data-driven building archetypes for urban building energy modelling. Energy 2019;181:360–77.
[29] Ferrando M, Ferroni S, Pelle M, Tatti A, Erba S, Shi X, et al. UBEM's archetypes improvement via data-driven occupant-related schedules randomly distributed and their impact assessment. Sustain Cities Soc 2022;87:104164.
[30] Clarke JA. Energy simulation in building design. Routledge; 2001.
[31] Harish VSKV, Kumar A. A review on modeling and simulation of building energy systems. Renew Sustain Energy Rev 2016;56:1272–92.
[32] Crawley DB, Hand JW, Kummert M, Griffith BT. Contrasting the capabilities of building energy performance simulation programs. Build Environ 2008;43: 661–73.
[33] Mohammadiziazi R, Copeland S, Bilec MM. Urban building energy model: database development, validation, and application for commercial building stock. Energy Build 2021;248:111175.
[34] Hong T, Chou SK, Bong TY. Building simulation: an overview of developments and information sources. Build Environ 2000;35:347–61.
[35] Henninger RH, Witte MJ, Crawley DB. Analytical and comparative testing of EnergyPlus using IEA HVAC BESTEST E100–E200 test suite. Energy Build 2004; 36:855–63.
[36] Klein SA, al e. Trnsys 17: a transient system simulation program. Madison, USA: University of Wisconsin; 2010. Solar Energy Laboratory.
[37] Strachan PA, Kokogiannakis G, Macdonald IA. History and development of validation with the ESP-r simulation program. Build Environ 2008;43:601–9.
[38] Yan D, Zhou X, An J, Kang X, Bu F, Chen Y, et al. DeST 3.0: a new-generation building performance simulation platform. Build Simulat 2022;15:1849–68.
[39] Heiple S, Sailor DJ. Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles. Energy Build 2008;40:1426–36.

[40] Ascione F, De Masi RF, de Rossi F, Fistola R, Sasso M, Vanoli GP. Analysis and diagnosis of the energy performance of buildings and districts: methodology, validation and development of Urban Energy Maps. Cities 2013;35:270–83.
[41] Robinson D, Haldi F, Leroux P, Perez D, Rasheed A, Wilke U. CitySim: comprehensive micro-simulation of resource flows for sustainable urban planning. In: Proceedings of the eleventh international IBPSA conference; 2009. p. 1083–90.
[42] 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.
[43] Fonseca JA, Schlueter A. Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. Appl Energy 2015;142:247–65.
[44] Deng Z, Chen Y, Yang J, Chen Z. Archetype identification and urban building energy modeling for city-scale buildings based on GIS datasets. Build Simulat 2022;15:1547–59.
[45] Deru M, Field K, Studer D, Benne K, Griffith B, Torcellini P, et al. US Department of Energy commercial reference building models of the national building stock. 2011.
[46] Qi L, Steven JQ, Godfried A, Jason B, Perry P-JY. Building Energy Modelling At Urban Scale: Integration of Reduced Order Energy Model with Geographical Information. In: Proceedings of Building Simulation 2015: 14th Conference of IBPSA. IBPSA; 2015. p. 190–9.
[47] DallO G, Galante A, Torri M. A methodology for the energy performance classification of residential building stock on an urban scale. Energy Build 2012; 48:211–9.
[48] Marius EB, Philip M, Harald TW, Igor S. Identifying grey-box models from archetypes of apartment block buildings. In: Proceedings of Building Simulation 2021: 17th Conference of IBPSA. IBPSA; 2021. p. 1091–8.
[49] 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.
[50] Agugiaro G, Benner J, Cipriano P, Nouvel R. The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations. Open Geospatial Data, Software and Standards 2018;3:2.
[51] Yang X, Hu M, Heeren N, Zhang C, Verhagen T, Tukker A, et al. A combined GIS-archetype approach to model residential space heating energy: a case study for The Netherlands including validation. Appl Energy 2020;280:115953.
[52] 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.
[53] Geraldi MS, Gnecco VM, Barzan Neto A, Martins BAdM, Ghisi E, Fossati M, et al. Evaluating the impact of the shape of school reference buildings on bottom-up energy benchmarking. J Build Eng 2021;43:103142.
[54] Arambula Lara R, Pernigotto G, Cappelletti F, Romagnoni P, Gasparella A. Energy audit of schools by means of cluster analysis. Energy Build 2015;95:160–71.
[55] Rosser JF, Long G, Zakhary S, Boyd DS, Mao Y, Robinson D. Modelling urban housing stocks for building energy simulation using CityGML EnergyADE. ISPRS Int J Geo-Inf 2019;8:163.
[56] Ali U, Shamsi MH, Hoare C, O'Donnell J. GIS-based residential building energy modeling at district scale. In: BSO 2018: 4th IBPSA-england conference on building simulation and optimization. Cambridge, United Kingdom: International Building Performance Simulation Association; 2018. 11-12 September 2018.
[57] Li X, Yao R, Liu M, Costanzo V, Yu W, Wang W, et al. Developing urban residential reference buildings using clustering analysis of satellite images. Energy Build 2018;169:417–29.
[58] Kragh J, Wittchen KB. Development of two Danish building typologies for residential buildings. Energy Build 2014;68:79–86.
[59] Ballarini I, Corgnati SP, Corrado V. Use of reference buildings to assess the energy saving potentials of the residential building stock: the experience of TABULA project. Energy Pol 2014;68:273–84.
[60] Augiseau V, Kim E. Spatial characterization of construction material stocks: the case of the Paris region. Resour Conserv Recycl 2021;170:105512.
[61] Candanedo JA, Vallianos C, Delcroix B, Date J, Saberi Derakhtenjani A, Morovat N, et al. Control-oriented archetypes: a pathway for the systematic application of advanced controls in buildings. Journal of Building Performance Simulation 2022;15:433–44.
[62] Yan D, O'Brien W, Hong T, Feng X, Burak Gunay H, Tahmasebi F, et al. Occupant behavior modeling for building performance simulation: current state and future challenges. Energy Build 2015;107:264–78.
[63] Hong T, Taylor-Lange SC, D'Oca S, Yan D, Corgnati SP. Advances in research and applications of energy-related occupant behavior in buildings. Energy Build 2016; 116:694–702.
[64] Happle G, Fonseca JA, Schlueter A. A review on occupant behavior in urban building energy models. Energy Build 2018;174:276–92.
[65] Shin M, Haberl JS. Thermal zoning for building HVAC design and energy simulation: a literature review. Energy Build 2019;203:109429.
[66] 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.
[67] Mata É, Sasic Kalagasidis A, Johnsson F. Energy usage and technical potential for energy saving measures in the Swedish residential building stock. Energy Pol 2013;55:404–14.
[68] Ashrae A. ASHRAE/IES standard 90.1-2016: energy standard for buildings except low-rise residential buildings. 2016. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2016.
[69] Shin M, Haberl JS. A procedure for automating thermal zoning for building energy simulation. J Build Eng 2022;46:103780.
[70] Chen Y, Hong T. Impacts of building geometry modeling methods on the simulation results of urban building energy models. Appl Energy 2018;215: 717–35.
[71] Fiorentini M, Gomis LL, Chen D, Cooper P. On the impact of internal gains and comfort band on the effectiveness of building thermal zoning. Energy Build 2020; 225:110320.
[72] Dogan T, Reinhart C. Shoeboxer: an algorithm for abstracted rapid multi-zone urban building energy model generation and simulation. Energy Build 2017;140: 140–53.
[73] Dogan T, Reinhart C, Michalatos P. Autozoner: an algorithm for automatic thermal zoning of buildings with unknown interior space definitions. Journal of Building Performance Simulation 2016;9:176–89.
[74] Faure X, Johansson T, Pasichnyi O. The impact of detail, shadowing and thermal zoning levels on urban building energy modelling (UBEM) on a district scale. Energies 2022;15:1525.
[75] Livingston OV, Cole P, Elliot D, Bartlett R. Building energy codes program: national benefits assessment (1992-2040). Richland, WA (United States): Pacific Northwest National Lab.(PNNL); 2014.
[76] Malik J, Bardhan R, Hong T, Piette MA. Developing occupant archetypes within urban low-income housing: a case study in Mumbai, India. Build Simulat 2022; 15:1661–83.
[77] Magenes L, Dinh HQ, Klein A. How new energy codes impact electrical substation designs: an overview of the changing regulations. IEEE Ind Appl Mag 2020;26: 21–8.
[78] Baba FM, Ge H. Effect of climate change on the energy performance and thermal comfort of high-rise residential buildings in cold climates. In: MATEC web of conferences. EDP Sciences; 2019, 02066.
[79] Buckley N, Mills G, Reinhart C, Berzolla ZM. Using urban building energy modelling (UBEM) to support the new European Union's Green Deal: case study of Dublin Ireland. Energy Build 2021;247:111115.
[80] Department of Energy, U.S.. Commercial buildings energy consumption survey (CBECS). Washington, DC, USA: US Department of Energy; 2015.
[81] Administration EI. Residential Energy Consumption Survey. http://www.eia.gov/consumption/residential/data/2009/index.cfm?view=microdata2009.
[82] Field K, Deru M, Studer D. Using DOE Commercial Reference Buildings for Simulation Studies: Preprint. 2010. United States: N. Web.
[83] International Energy Conservation Code (IECC). IECC code and commentary. 2012. Illinois, US2012.
[84] Shen P. Understanding Texas energy use for residential sector under future climate and urbanization scenarios. In: 4th international conference on fossil and renewable energy. Houston: Oral presentation; 2020. Texas.
[85] Berardi U, Jafarpur P. Assessing the impact of climate change on building heating and cooling energy demand in Canada. Renew Sustain Energy Rev 2020;121: 109681.
[86] Shen P. Impacts of climate change on U.S. building energy use by using downscaled hourly future weather data. Energy Build 2017;134:61–70.
[87] Yaghoobian N, Srebric J. Influence of plant coverage on the total green roof energy balance and building energy consumption. Energy Build 2015;103:1–13.
[88] Shen P, Lukes JR. Impact of global warming on performance of ground source heat pumps in US climate zones. Energy Convers Manag 2015;101:632–43.
[89] Shen P, Wang Z, Ji Y. Exploring potential for residential energy saving in New York using developed lightweight prototypical building models based on survey data in the past decades. Sustain Cities Soc 2021;66:102659.
[90] Ang YQ, Berzolla ZM, Letellier-Duchesne S, Jusiega V, Reinhart C. UBEM.io: a web-based framework to rapidly generate urban building energy models for carbon reduction technology pathways. Sustain Cities Soc 2022;77:103534.
[91] Abolhassani SS, Amayri M, Bouguila N, Eicker U. A new workflow for detailed urban scale building energy modeling using spatial joining of attributes for archetype selection. J Build Eng 2022;46:103661.
[92] Chen X, Yang H, Peng J. Energy optimization of high-rise commercial buildings integrated with photovoltaic facades in urban context. Energy 2019;172:1–17.
[93] Shen P, Liu J, Wang M. Fast generation of microclimate weather data for building simulation under heat island using map capturing and clustering technique. Sustain Cities Soc 2021;71:102954.
[94] Kavgic M, Hilliard T, Swan L. Opportunities for implementation of MPC in commercial buildings. Energy Proc 2015;78:2148–53.
[95] Zeferina V, Wood FR, Edwards R, Tian W. Sensitivity analysis of cooling demand applied to a large office building. Energy Build 2021;235:110703.
[96] Girgis-McEwen E, Ullah R. Energy efficiency measures for net-zero energy ready building codes. In: Proceedings of the eSim; 2018.
[97] Mutasim Baba F, Ge H. Effect of climate change on the annual energy consumption of a single family house in British Columbia. MATEC Web Conf. 2018;251:03018.
[98] Munankarmi P, Maguire J, Balamurugan SP, Blonsky M, Roberts D, Jin X. Community-scale interaction of energy efficiency and demand flexibility in residential buildings. Appl Energy 2021;298:117149.
[99] 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.
[100] Alhamlawi F, Alaifan B, Azar E. A comprehensive assessment of Dubai's green building rating system: Al Sa'fat. Energy Pol 2021;157:112503.
[101] Alardhi A, S Alaboodi A, Almasri R. Impact of the new Saudi energy conservation code on Saudi Arabia residential buildings. Aust J Mech Eng 2022;20:1392–406.

[102] Foroushani S, Bernhardt R, Bernhardt M. On the use of the reference building approach in modern building energy codes. Energy Build 2022;256:111726.
[103] Ascione F, Cheche N, Masi RFD, Minichiello F, Vanoli GP. Design the refurbishment of historic buildings with the cost-optimal methodology: the case study of a XV century Italian building. Energy Build 2015;99:162–76.
[104] Loga T, Stein B, Diefenbach N. TABULA building typologies in 20 European countries—making energy-related features of residential building stocks comparable. Energy Build 2016;132:4–12.
[105] Zhang N, Luo Z, Liu Y, Feng W, Zhou N, Yang L. Towards low-carbon cities through building-stock-level carbon emission analysis: a calculating and mapping method. Sustain Cities Soc 2022;78:103633.
[106] Heisel F, McGranahan J, Ferdinando J, Dogan T. High-resolution combined building stock and building energy modeling to evaluate whole-life carbon emissions and saving potentials at the building and urban scale. Resour Conserv Recycl 2022;177:106000.
[107] Sartori I, Wachenfeldt BJ, Hestnes AG. Energy demand in the Norwegian building stock: scenarios on potential reduction. Energy Pol 2009;37:1614–27.
[108] Vimmr T, Loga T, Diefenbach N, Stein B, Bachová L. Tabula—residential building typologies in 12 European countries—good practice example from the Czech Republic. Central Europe towards sustainable building. 2013.
[109] 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.
[110] Attia S, Canonge T, Popineau M, Cuchet M. Developing a benchmark model for renovated, nearly zero-energy, terraced dwellings. Appl Energy 2022;306: 118128.
[111] Boverket. Energi i bebyggelsen—tekniska egenskaper och beräkningar—resultat från projektet BETSI. Boverket Karlskrona; 2010.
[112] Theodoridou I, Papadopoulos AM, Hegger M. A typological classification of the Greek residential building stock. Energy Build 2011;43:2779–87.
[113] Portella J, Riibas M. Bottom-up description of the French building stock, including archetype buildings and energy demand. 2012.
[114] Li X, Yao R, Yu W, Meng X, Liu M, Short A, et al. Low carbon heating and cooling of residential buildings in cities in the hot summer and cold winter zone - a bottom-up engineering stock modeling approach. J Clean Prod 2019;220:271–88.
[115] Gaglia AG, Balaras CA, Mirasgedis S, Georgopoulou E, Sarafidis Y, Lalas DP. Empirical assessment of the Hellenic non-residential building stock, energy consumption, emissions and potential energy savings. Energy Convers Manag 2007;48:1160–75.
[116] Lv Y, Peng H, He M, Huang Y, Wang J. Definition of typical commercial building for South China's Pearl River Delta: local data statistics and model development. Energy Build 2019;190:119–31.
[117] Matsuoka A, Yohei Y, Yusuke S, Yoshiyuki S. Urban scale modelling of energy demand of retail facilities. In: Proceedings of BS2013: 13th conference of international building performance simulation association; 2013. p. 1288–95.
[118] Azar E, Alaifan B, Lin M, Trepci E, El Asmar M. Drivers of energy consumption in Kuwaiti buildings: insights from a hybrid statistical and building performance simulation approach. Energy Pol 2021;150:112154.
[119] Corgnati SP, Fabrizio E, Filippi M, Monetti V. Reference buildings for cost optimal analysis: method of definition and application. Appl Energy 2013;102:983–93.
[120] Schwartz Y, Godoy-Shimizu D, Korolija I, Dong J, Hong SM, Mavrogianni A, et al. Developing a Data-driven school building stock energy and indoor environmental quality modelling method. Energy Build 2021;249:111249.
[121] Medina Benejam G. Bottom-up characterisation of the Spanish building stock–Archetype buildings and energy demand. 2011.
[122] Kim DW, Kim YM, Lee SH, Park WY, Bok YJ, Ha SK, et al. Development of reference building energy models for South Korea. In: Proceedings of the 15th IBPSA conference; 2017. p. 7–9. San Francisco, CA, USA.
[123] Yilmaza S, Murrayb P, Bollingerb A, Guidatic G. Hourly demand profiles for space heating and electricity. ETH Zurich; 2020.
[124] Buso T, Corgnati SP. A customized modelling approach for multi-functional buildings – application to an Italian Reference Hotel. Appl Energy 2017;190: 1302–15.
[125] Boemi SN, Slini T, Papadopoulos AM, Mihalakakou Y. A statistical approach to the prediction of the energy performance of hotel stock. Int J Vent 2011;10: 163–72.
[126] Liu X, Hu S, Yan D. A statistical quantitative analysis of the correlations between socio-demographic characteristics and household occupancy patterns in residential buildings in China. Energy Build 2023;284:112842.
[127] De Jaeger I, Reynders G, Callebaut C, Saelens D. A building clustering approach for urban energy simulations. Energy Build 2020;208:109671.
[128] Tardioli G, Kerrigan R, Oates M, O'Donnell J, Finn DP. Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach. Build Environ 2018;140:90–106.
[129] Hsu D. Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. Appl Energy 2015;160:153–63.
[130] Schaefer A, Ghisi E. Method for obtaining reference buildings. Energy Build 2016; 128:660–72.
[131] Gaitani N, Lehmann C, Santamouris M, Mihalakakou G, Patargias P. Using principal component and cluster analysis in the heating evaluation of the school building sector. Appl Energy 2010;87:2079–86.
[132] Ballarini I, Corgnati SP, Corrado V, Talà N. Improving energy modeling of large building stock through the development of archetype buildings. In: Proceedings of the 12th conference of the international building performance simulation association. Sydney, Australia: IBPSA); 2011. p. 14–6.
[133] Han M, Wang Z, Zhang X. An approach to data acquisition for urban building energy modeling using a Gaussian mixture model and expectation-maximization algorithm. Buildings 2021;11:30.
[134] Ghiassi N, Mahdavi A. Reductive bottom-up urban energy computing supported by multivariate cluster analysis. Energy Build 2017;144:372–86.
[135] Ma Y, Deng W, Xie J, Heath T, Xiang Y, Hong Y. Generating prototypical residential building geometry models using a new hybrid approach. Build Simulat 2022;15:17–28.
[136] Wang C, Ferrando M, Causone F, Jin X, Zhou X, Shi X. An innovative method to predict the thermal parameters of construction assemblies for urban building energy models. Build Environ 2022;224:109541.
[137] Diao L, Sun Y, Chen Z, Chen J. Modeling energy consumption in residential buildings: a bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation. Energy Build 2017;147:47–66.
[138] Buttitta G, Turner WJN, Neu O, Finn DP. Development of occupancy-integrated archetypes: use of data mining clustering techniques to embed occupant behaviour profiles in archetypes. Energy Build 2019;198:84–99.
[139] Buttitta G, Turner W, Finn D. Clustering of household occupancy profiles for archetype building models. Energy Proc 2017;111:161–70.
[140] Kim B, Yamaguchi Y, Kimura S, Ko Y, Ikeda K, Shimoda Y. Urban building energy modeling considering the heterogeneity of HVAC system stock: a case study on Japanese office building stock. Energy Build 2019;199:547–61.
[141] Deng Z, Chen Y, Pan X, Peng Z, Yang J. Integrating GIS-based point of interest and community boundary datasets for urban building energy modeling. Energies 2021;14:1049.
[142] Aerts D, Minnen J, Glorieux I, Wouters I, Descamps F. A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. Build Environ 2014; 75:67–78.
[143] Ang YQ, Berzolla Z, Reinhart C. Smart meter-based archetypes for socioeconomically sensitive urban building energy modeling. Build Environ 2023; 246:110991.
[144] Borges P, Travesset-Baro O, Pages-Ramon A. Hybrid approach to representative building archetypes development for urban models – a case study in Andorra. Build Environ 2022;215:108958.
[145] Sokol J, Cerezo Davila C, Reinhart CF. Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. Energy Build 2017;134:11–24.
[146] Gholami M, Torreggiani D, Tassinari P, Barbaresi A. Narrowing uncertainties in forecasting urban building energy demand through an optimal archetyping method. Renew Sustain Energy Rev 2021;148:111312.
[147] Krayem A, Al Bitar A, Ahmad A, Faour G, Gastellu-Etchegorry J-P, Lakkis I, et al. Urban energy modeling and calibration of a coastal Mediterranean city: the case of Beirut. Energy Build 2019;199:223–34.
[148] Kristensen MH, Hedegaard RE, Petersen S. Hierarchical calibration of archetypes for urban building energy modeling. Energy Build 2018;175:219–34.
[149] Wang C, Ferrando M, Causone F, Jin X, Zhou X, Shi X. Data acquisition for urban building energy modeling: a review. Build Environ 2022;217:109056.
[150] Wang C-K, Tindemans S, Miller C, Agugiaro G, Stoter J. Bayesian calibration at the urban scale: a case study on a large residential heating demand application in Amsterdam. Journal of Building Performance Simulation 2020;13:347–61.
[151] Cerezo Davila C, Jones N, Al-Mumin A, Hajiah A, Reinhart C. Implementation of a calibrated Urban Building Energy Model (UBEM) for the evaluation of energy efficiency scenarios in a Kuwaiti residential neighborhood. Build Simulat 2017: 745–54. IBPSA; 2017.
[152] Johari F, Peronato G, Sadeghian P, Zhao X, Widén J. Urban building energy modeling: state of the art and future prospects. Renew Sustain Energy Rev 2020; 128:109902.
[153] Fennell P, Ruyssevelt P, Mata É, Jakob M. A review of the status of uncertainty and sensitivity analysis in building-stock energy models. Build Simulat 2019. 2019;16:3353–60.
[154] Prataviera E, Vivian J, Lombardo G, Zarrella A. Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis. Appl Energy 2022;311:118691.
[155] Hong T, Chen Y, Luo X, Luo N, Lee SH. Ten questions on urban building energy modeling. Build Environ 2020;168:106508.
[156] Nutkiewicz A, Yang Z, Jain RK. Data-driven Urban Energy Simulation (DUE-S): integrating machine learning into an urban building energy simulation workflow. Energy Proc 2017;142:2114–9.
[157] Thrampoulidis E, Hug G, Orehounig K. Approximating optimal building retrofit solutions for large-scale retrofit analysis. Appl Energy 2023;333:120566.
[158] Crosbie T, Dawood N, Dawood S. Improving the energy performance of the built environment: the potential of virtual collaborative life cycle tools. Autom ConStruct 2011;20:205–16.
[159] Deng S, Wang RZ, Dai YJ. How to evaluate performance of net zero energy building – a literature research. Energy 2014;71:1–16.