Title	Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis
Authors(s)	Ali, Usman, Shamsi, Mohammad Haris, Hoare, Cathal, Mangina, Eleni, O'Donnell, James
Publication date	2021-09-01
Publication information	Ali, Usman, Mohammad Haris Shamsi, Cathal Hoare, Eleni Mangina, and James O'Donnell. "Review of Urban Building Energy Modeling (UBEM) Approaches, Methods and Tools Using Qualitative and Quantitative Analysis." Elsevier, September 1, 2021. https://doi.org/10.1016/j.enbuild.2021.111073.
Publisher	Elsevier
Item record/more information	http://hdl.handle.net/10197/26158
Publisher's version (DOI)	10.1016/j.enbuild.2021.111073

Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis

Usman Ali[a], Mohammad Haris Shamsi[a], Cathal Hoare[a], Eleni Mangina[b], James O'Donnell[a]

[a]*School of Mechanical and Materials Engineering and UCD Energy Institute, UCD, Dublin, Ireland*
[b]*School of Computer Science and UCD Energy Institute, UCD, Dublin, Ireland*

Abstract

The world has witnessed a significant population shift to urban areas over the past few decades. Urban areas account for about two-thirds of the world's total primary energy consumption, of which the urban building sector constitutes a significant proportion approximately 40%. Stakeholders such as urban planners and policy makers face substantial challenges when targeting sustainable energy and climate goals related to the buildings' sector, i.e. to reduce energy use and associated emissions. Urban energy modeling is one possible solution that leverages limited resources to estimate building energy use and support appropriate policy formation. Over the past few years, there have been only a few review studies on urban building energy modeling approaches. These studies lack an in-depth discussion of the challenges and future research opportunities related to data-driven, reduced-order, and simulation-based modeling methods. This paper proposes Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis of approaches, methods and tools used for urban building energy modeling. Furthermore, this paper proposes a generalized framework based on existing literature for different urban energy modeling methods. The aim of this study is to assist urban planners and energy policymakers when choosing appropriate methods to develop and implement in-depth sustainable building energy planning and analysis projects based on limited available resources.

Keywords: urban building energy modeling, top-down, bottom-up, data-driven, energy modeling, UBEM, energy efficiency

---

Nomenclature

ADE	Application Domain Extensions
BEM	Building Energy Modeling

BPD	Building Performance Database
CA	Classification Accuracy
CART	Classification and Regression Tree
CBECS	Commercial Buildings Energy Consumption Survey
CDR	Conditional Demand Analysis
CE	Classification Error
CEA	City Energy Analyst
CEA	City Energy Analyst
CityBES	City Building Energy Saver
CityGML	City Geography Markup Language
DBSCAN	Density-based Spatial Clustering of Applications with Noise
DOE	Department of Energy
DT	Decision Tree
EPBD	European Union Energy Performance of Buildings Directive
EPC	Energy Performance Certificate
GBT	Gradient Boosted Trees
GIS	Geographic Information System
GLM	Generalized Linear Model
HVAC	Heating Ventilation, and Air Conditioning
IESVE	Integrated Environmental Solutions Virtual Environment
ISO	International Organization for Standardization
KML	Keyhole Markup Language
LOD	Level of Detail
LOF	Local Outlier Factor
LRD	Local Reachability Density
MAPE	Mean Absolute Percentage Error

$MARS$	Multivariate Adaptive Regression Splines
MLR	Multiple Linear Regression
$RMSE$	Root Mean Squared Error
SVM	Support Vector Machines
SVR	Support Vector Regression
$SWOT$	Strengths, Weaknesses, Opportunities, and Threats
$TEASER$	Tool for Energy Analysis and Simulation for Efficient Retrofit
TMY	Typical Meteorological Year
$UBEM$	Urban Building Energy Modeling
$UEUM$	Urban Energy Use Modeling
UMI	Urban Modeling Interface

1. Introduction

More than half (55%) of the global population lives in urban areas, and this proportion is projected to increase to more than two-thirds of the world population (68%) by 2050 [1]. Levels of urbanization vary across different regions of the world. In 2018, Europe's urbanization level was 74% and is projected to increase to 83.7% by 2050 [1]. Urban areas consume more than two-thirds of global energy and account for more than 70% of global CO_2 emissions [2].

The urban building sector is one of the highest contributors of CO_2 emissions worldwide [3]. Buildings account for approximately 39% and 36% of CO_2 emissions in the United States and European Union respectively [4, 5]. Commercial and residential buildings consume more than 70% of the United States overall electricity consumption [5]. A similar trend has been observed in Europe. The building sector forms the core of urban energy supply and demand, and hence, a considerable potential exists to reduce energy consumption and associated emissions. These reductions could be facilitated through the implementation of energy efficiency policies and awareness programs. For instance, The European Union has set energy-efficient targets to achieve an efficient and decarbonized stock by 2050 [6].

Urban planners and policymakers at the urban scale face significant issues when planning and implementing large scale sustainable and energy-efficient scenarios [7]. Building energy models can help optimize and analyze various design scenarios to improve a building's energy performance. Numerous energy models have been devised for multiple energy-driven applications, for instance, energy planning, energy supply-demand calculation, retrofit analysis, forecasting, renewable energy impact, emission reduction, and optimization [8, 9, 10, 11, 12].

Considering the scale of implementation, energy modeling could be performed at the individual building level, Building Energy Modeling (BEM), or at the urban level, Urban Building Energy Modeling (UBEM). BEM approaches mainly comprise three categories, white-box models (forward modeling or physics-based), black-box models (empirical or statistical-based), and grey-box models (both physics and statistical-based) [13, 14]. Urban energy models further comprise two categories, namely, top-down and bottom-up models [15]. The top-down models employ macroeconomic variables and statistical data to determine energy predictions. Bottom-up models consider and formulate building clusters with similar building characteristics (geometric and non-geometric parameters) at different modeling scales. While top-down approaches are suitable for a broad and aggregated level large-scale analysis, bottom-up approaches are instrumental in the identification of plausible efficiency improvements for the urban building sector. Based on existing studies [10, 15], the bottom-up approach has been proven ideal for in-depth urban scale building analyses. In general, urban building energy modeling requires data sources, including geometric and non-geometric information related to any building [16]. Geometric data includes building shape, dwelling type, building envelope, number of floors, walls, and windows. In contrast, non-geometric building data includes envelope U-values, construction assemblies, and HVAC systems. However, building stock data collection is difficult at an urban scale due to a lack of data and users' privacy issues [8]. An initial query in the World of Science database indicates that approximately 63% of the publication topics (title, abstract and keywords) focus on BEM and the remaining 37% actually relate to UBEM. An initial query refers to topic keywords (Urban (U), Building (B), and Energy Modeling (EM)) that are used for quantitative analysis of urban building energy modeling literature studies (Table 2). Although the trend of BEM and UBEM publications has experienced a gradual increase between 2011-2019, UBEM still requires more consideration(Fig 1).

Urban building energy modeling has great potential to improve building energy performance at a large scale in a manner that supports energy efficiency, sustainability, and management of cities [8]. The key practice areas of urban energy system models include technology design, building design, urban climate, systems design, and policy assessment [17]. UBEM also helps in urban and regional building energy planning approaches for sustainable development in the built environment [10]. However, UBEM is quite challenging due to the urban energy system's complexity, required resources, time, and effort to achieve accurate results [18]. Furthermore, urban energy modeling (for instance, building stock occupant behavior) and existing spatial data are insufficient to represent the entire building stock [15].

Although BEM has been extensively reviewed in the building simulation domain, only a few studies specifically focus on UBEM approaches and techniques [19, 20]. The majority of the UBEM studies employ methods and tools that implement bottom-up simulation-based approaches [19, 21]. A recent review of UBEM by Reinhart and Cerezo Davila [8] discussed different aspects of urban scale modeling that include urban simulation data inputs, thermal model generation, the associated tools, and results validation. Li et al. [22] published a review of the broad categories of UBEM, which describes the basic workflow, and strengths and weaknesses of physics-based, bottom-up models, and their applications to simulating

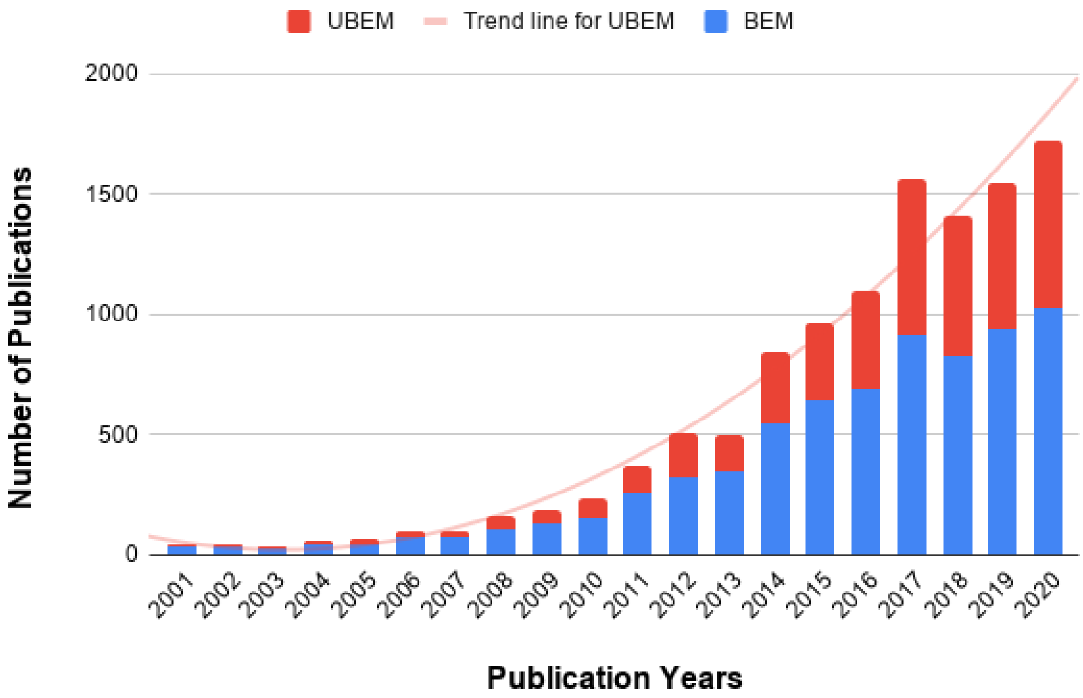

Figure 1: Building energy modeling and urban building energy modeling publication trend from World of Science academic search platform.

urban-scale building energy use. Sola et al. [23] reviewed simulation tools for urban-scale energy modeling and further classified UBEM based on different applications, such as building energy demand prediction, land use, transportation, environment interaction, multi-domain integrated simulation, and multi-domain co-simulation tools. A comprehensive review by Ferrari et al. [24] compared 17 different tools that evaluate several energy services, sources, and technologies targeted at an urban/district scale. A vast majority of the available UBEM tools focus on a single component of urban building energy use rather than overall energy use [18]. Therefore, an integrated urban energy modeling approach that facilitates the identification of accurate, time-efficient, and feasible solutions remains a significant challenge for researchers and stakeholders. Furthermore, current UBEM studies lack an in-depth discussion of the modeling challenges and research opportunities associated with the entire UBEM spectrum. These studies mostly focus on bottom-up UBEM. This paper reviews the top-down as well as bottom-up (data-driven, reduced-order, and simulation-based) methods

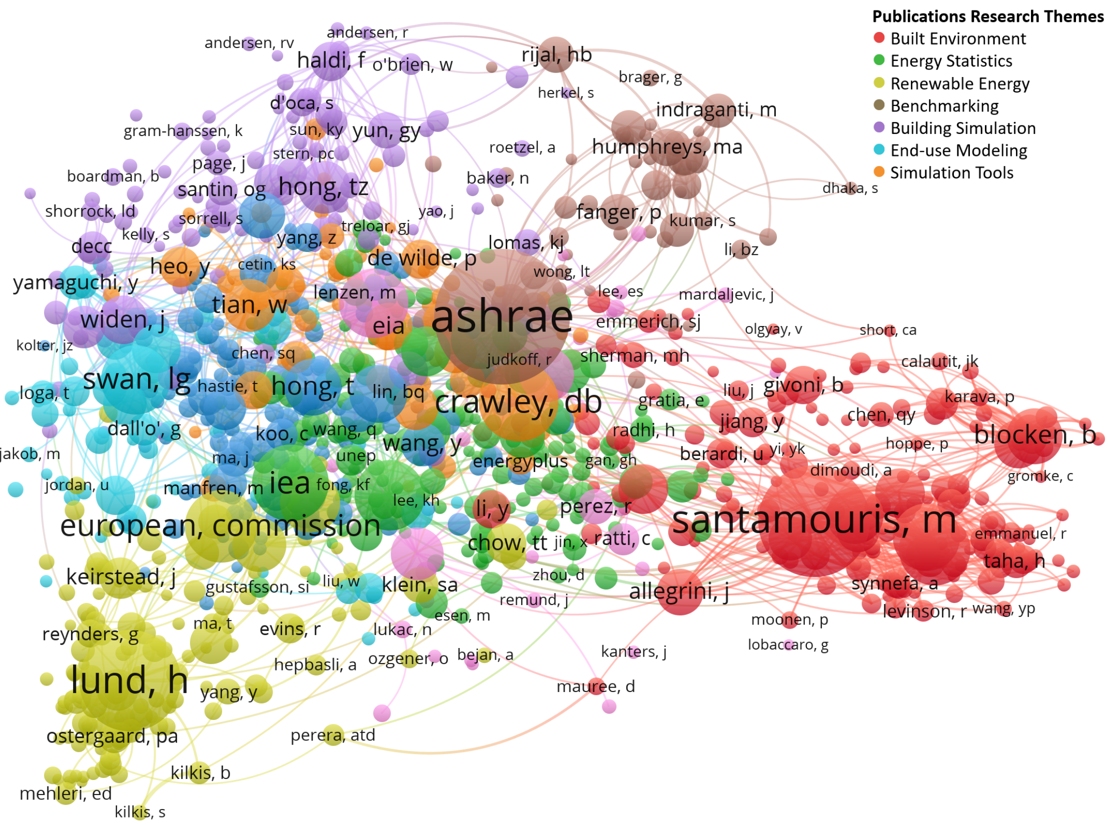

Figure 2: UBEM publication co-citation network categorized under six clusters based on closely related research themes and obtained from World of Science academic search platform using VOSViewer tool [25].

The aim of this research paper is to provide a detailed overview of the strengths, weaknesses, opportunities, and threats of the approaches, methods and tools used for building energy modeling at the urban scale. This paper also includes a comprehensive review of all urban energy modeling categories including top-down and bottom-up (data-driven, reduced-order, and simulation-based) methods. Furthermore, the research proposes generalized modeling frameworks based on the existing literature for different urban energy modeling methods. This review paper uses World of Science academic search platform to perform systematic review analysis. The co-citation network map, formulated using the VOSViewer tool, illustrates the frequently cited papers in the UBEM field categorized under six clusters that are based on closely related publications research themes (Fig. 2) [25]. The search criteria based on the initial query refer to topic keywords (Urban (U), Building (B), and Energy Modeling (EM)) used for co-citation network map development using the World of Science database (Table 2). The developed co-citation network then identifies the top-cited articles in the UBEM field (Table 1).

Table 1: List of the top relevant articles having a strong citation in the field of UBEM

Authors	Title	Citation
Crawley et al. [26]	EnergyPlus: creating a new-generation building energy simulation program	1900+
Swan et al. [27]	Modeling of end-use energy consumption in the residential sector: A review of modeling techniques	1700+
Zhao et al. [28]	A review on the prediction of building energy consumption	1200+
Lund et al. [7]	Renewable energy strategies for sustainable development	1100+
Santamouris et al. [3]	Energy and climate in the urban built environment	900+
Suganthi et al. [11]	Energy models for demand forecasting—A review	950+
Jebaraj et al. [12]	A review of energy models	800+
Kavgic et al. [29]	A review of bottom-up building stock models for energy consumption in the residential sector	700+
Keirstead et al. [17]	A review of urban energy system models: Approaches, challenges and opportunities	400+
Hong et al. [21]	Building simulation: an overview of developments and information sources	400+
Reinhart et al. [8]	Urban building energy modeling–A review of a nascent field	400+

This paper is structured as follows: Section 2 describes the review methodology and the associated steps; Section 3 illustrates a detailed overview of UBEM approaches with qualitative and quantitative analysis; Section 4 includes a detailed discussion of the challenges associated with modeling methods. Section 5 concludes this research study and discusses future work.

2. Review Methodology

The review methodology used in this work consists of four steps: 1) approaches, 2) methods, 3) quantitative analysis, and 4) qualitative analysis (Fig 3). The process starts with an overview of different categories (and sub-categories) of urban energy modeling approaches. The second step involves a detailed discussion of the existing projects and tools that implement urban energy modeling methods and further illustrates a generalized methodology based on the current literature for each urban energy modeling approach.

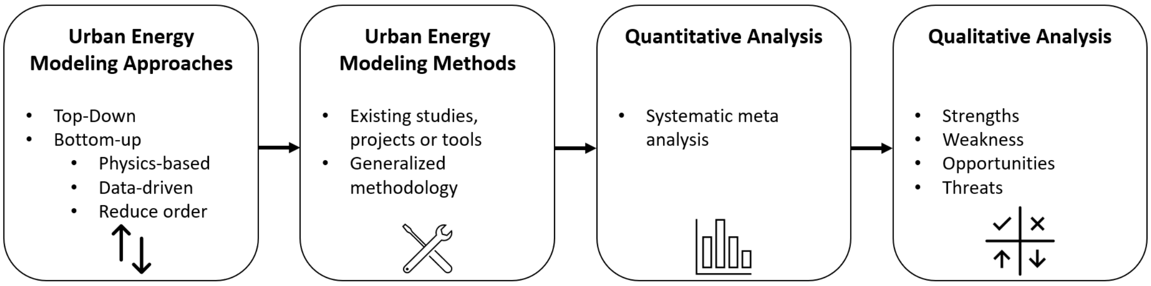

Figure 3: Methodology for the review of urban building energy modeling approaches.

The third step includes a systematic review based on quantitative analysis to identify the trends in the implementation of modeling approaches. The systematic review uses a meta-analysis approach to determine the statistics of studies conducted in the urban energy modeling domain. As meta-analysis combines the results of multiple scientific studies, this approach is suitable to compare various UBEM studies. These statistics are identified through the academic search engine known as Web of Knowledge, a well-recognized database for academic studies and publications [30]. Furthermore, the search engine performs multiple results' analysis using publication years, types, and sources. We used the advanced search feature of the academic search engine with the Keyword Topic (KT) structure (Eq. 1) and combinations (Table 2). Topic A encodes the different urban energy modeling keyword "Approaches" such as top-down, bottom-up (their sub-categories), which might be referred differently in various studies. To remove any inconsistencies, Topic A keywords search inside the paper title. Topic U defines "Urban" and similar keywords. Topic B focuses mainly on "Buildings" and related keywords. Topic EM includes the "Energy Modeling" keyword, and Topic P comprises the Application keyword. Topics U, B, EM, and P search inside the paper title, abstract, and keyword of the paper.

$$KT = A + U + B + EM + P \tag{1}$$

Finally, the fourth and final step performs a qualitative analysis (SWOT analysis) of different UBEM approaches and methods. The process involves the identification of strengths (S) and weaknesses (W) of the approaches as well as the opportunities (O) for future research and threats (T) presented by the existing studies [31]. More notably, strengths and weaknesses help the researchers choose accurate methods to develop and implement sustainable building energy planning and analysis projects. Opportunities and threats contribute to an in-depth discussion of challenges and future research opportunities across different modeling methods.

Table 2: Keyword topics used for quantitative analysis of urban building energy modeling literature studies using advanced search academic engine Web of Knowledge.

Topic	Keywords
Approaches (A)	1) top-down, econometric, technological, 2) bottom-up, physic-based, simulation, engineering, 3) data-driven, machine learning, artificial intelligence, statistic*, 4) reduced-order, lumped parameter, R–C, grey-box, gray-box
Urban (U)	urban, city, district, cities, large, region
Building (B)	building, house, residential, non-residential, commercial, office, school, institution
Energy Modeling (EM)	energy model*
Application (P)	forecast*, benchmark*, load profile*, retrofit, predict*, classification, mapping, planning, energy analysis

* is used to search for keywords singular, plural and similar forms.

3. Literature Review

Urban building energy models can be broadly categorized as top-down and bottom-up approaches as listed below [27].

- Top-down modeling approach
- Bottom-up modeling approach
  - Physics-based
  - Data-driven
  - Reduced-order

The top-down models consider the entire building sector a single energy entity in order to estimate energy consumption at scale [27]. On the other hand, bottom-up modeling approaches focus on individual buildings and end-uses and estimate energy consumption of individual buildings' or groups of buildings [22]. The bottom-up models work at a disaggregated level can further be classified as physics-based, data-driven and reduced-order methods [15].

This review paper analyzes each modeling approach using a common structure. The review initiates with an analysis of the modeling approach at the broad level that defines the different kinds of models. The review further analyzes the methods that provide an indication of the various implementations of the modeling approach. These approaches are then analyzed quantitatively and qualitatively. While the quantitative analysis determines the number of publications published under the considered modeling domain, the qualitative analysis involves a SWOT analysis that defines the strengths, weaknesses, opportunities and threats of the considered modeling approach.

3.1. Top-down Urban Building Energy Modeling

3.1.1. Top-down Approaches

Top-down modeling primarily focuses on statistical energy use, historical data, and socio-economic factors such as population, weather conditions, and fuel prices [29]. Typically, these approaches investigate the connections between the energy sector and the economic outputs to fit historical time-series datasets of energy consumption or CO_2 emissions [29]. Top-down modeling approaches comprise econometric and technological top-down models [27]. Econometric models identify the relationship between the energy sector and variables such as price (fuel) and income to represent the economic outputs [29]. Technological models consider factors, such as saturation, technological growth, and structural reforms that influence the energy use [29].

3.1.2. Top-down Methods

Top-down methods are relatively easier to implement and have been widely used to calculate urban energy consumption. For instance, Meha et al. [32] proposed a generalized top-down model to calculate the total thermal energy consumed by residential buildings. The model estimated the overall space heat demand using the population distribution dataset and yearly energy balances. The study extracted the population distribution from the national census database and yearly energy balances from the global end-use energy consumption data from the International Energy Agency. Summerfield et al. [33] developed a model to identify the total annual household energy consumption trajectory of households in the United Kingdom. The Annual Delivered Energy Price and Temperature (ADEPT) model implemented regression models to analyze the relationship between household energy consumption, heating season temperature, and energy price [33].

Shkurti et al. [34] proposed a top-down macro-econometric model investigating the relationship between energy consumption and GDP growth of 6 western Balkan countries (Albania, Bosnia and Herzegovina, Croatia, Serbia, Montenegro, and Macedonia). The data were collected from World Development Indicators, considering electricity use per capita and oil price variables. Adam et al. [35] proposed top-down models to calculate the energy savings for the residential sector. The models are regarded as thermal energy and electricity (household appliances, lighting) consumption to determine the household sector's energy savings.

3.1.3. Top-down Quantitative Analysis

The top-down quantitative analysis uses the advanced search function of Web of Knowledge engine using Eq.1 and Table 2. The keywords search is based on any paper's initial content, such as title, abstract, and keywords used in a paper. The main keywords include top-down, econometric, and technological. A total of 101 articles are found that contain these keywords. The research publication trend between 2007-2020 indicates that the number of papers published has experienced a gradual increase after 2013 and fewer papers were published in 2018 compared to 2017 (Fig. 4). The year 2018 experienced a decline in the number of top-down UBEM publications. This might suggest the growing shift towards bottom-up approaches (Fig. 8). The main applications of the top-down models include building energy analysis (33%), energy planning (18%), and optimization (14%) respectively (Fig. 4). The top publication sources on the top-down approach topics are Applied Energy and Energy and Buildings journals: 83% of these papers were published as articles, and the 13% were published as conference proceedings. (Fig. 5).

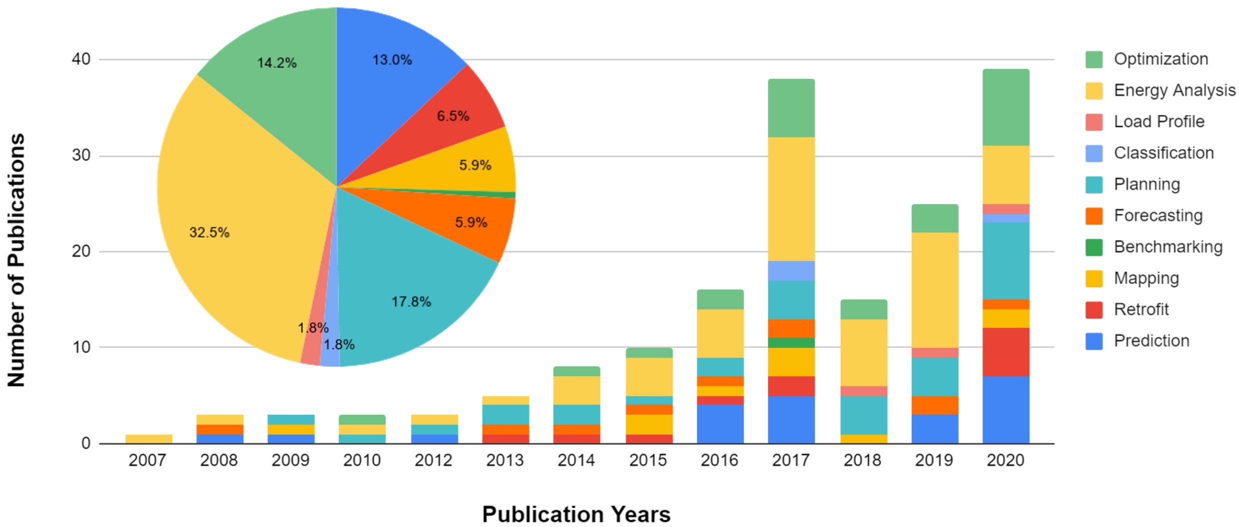

Figure 4: Top-down urban building energy modeling publication trend with intended area of implementation from the World of Science academic search platform.

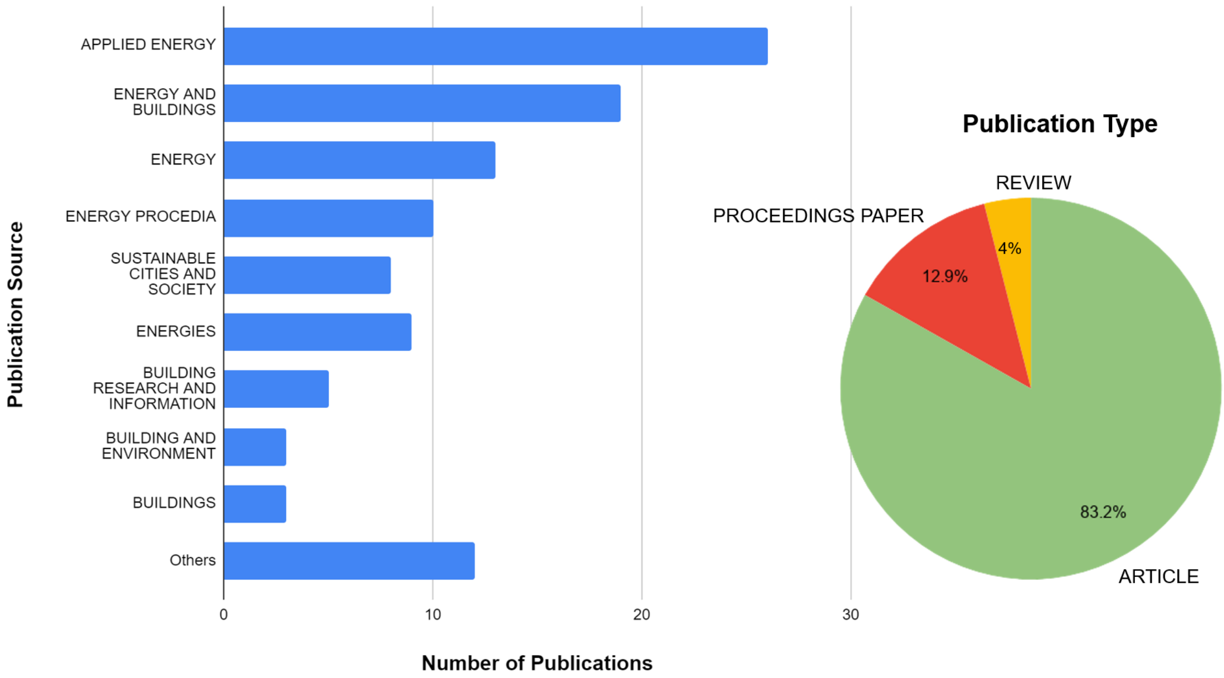

Figure 5: Top-down urban building energy modeling publication distribution on the basis of publication sources and publication type from the World of Science academic search platform.

3.1.4. Top-down Qualitative Analysis

The qualitative analysis of top-down models determines the strengths and weaknesses associated with these models' past implementation. Furthermore, this analysis identifies the threats and challenges of top-down model suitability and scope, which would eventually evaluate various opportunities to enhance these models (Fig. 6).

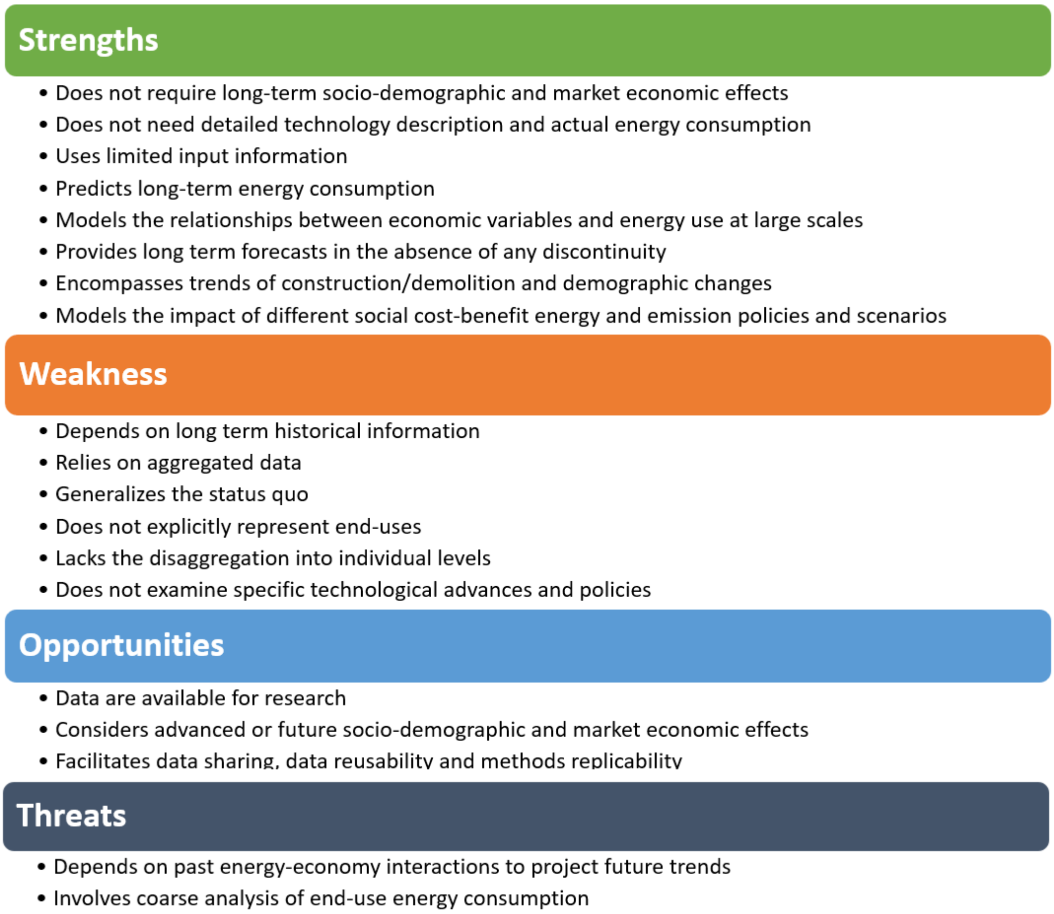

Figure 6: Strengths, weaknesses, opportunities, and threats analysis of top-down modeling approach for qualitative analysis.

The top-down modeling approach provides an integrated overview of the urban energy workflow that includes long-term socio-demographic as well as market economic effects [22]. As top-down models solely rely on available historical aggregated energy data, these models can help predict long-term energy consumption behavior in the absence of any discontinuity [27]. Moreover, top-down models' formulation requires limited information (mostly aggregated energy-economic data), thereby avoiding the need for detailed technological descriptions or actual energy consumption data [22]. This approach captures the interaction between economic indicators and energy use at a large scale to reflect the impact of social-economic policies [36]. Furthermore, this approach could easily model the effects of different social cost-benefit energy strategies, emission policies, and scenarios [29].

As top-down models use long term historical data to estimate energy consumption, one significant concern regarding the approach relates to the generalization of the existing conditions [27]. Hence, this type of model is less suitable for examining urban energy supply and demand [18]. Since the models rely on historical data, the underlying approach would not be ideal for exploring specific technological advancements in energy policies [29]. Furthermore, these models do not provide an explicit representation and disaggregation of individual component levels to analyze the energy consumption [36].

Data availability remains a significant challenge to implement these models, mainly due to the associated privacy concerns and limited resources. One considerable opportunity would be to provide easy access to the national historical database for detailed long-term analysis to facilitate the modeling process. Furthermore, researchers should consider advanced or future socio-demographic and market economic effects, such as the social, economic, and environmental influences of renewable energy.

The use of past energy-economy interactions to project future trends could pose a significant threat to top-down models [29]. Furthermore, as these models do not explicitly consider each building's different physical characteristics, top-down models provide a coarse analysis of end-use energy consumption [36]. Thus, top-down building stock models eliminate the possibility of identifying areas of improvement in building energy use, building physical design, and operational specifications [36].

3.2. Physics-based Bottom-up Urban Building Energy Modeling

3.2.1. Physics-based Approaches

The physics-based method, also known as the engineering or simulation method, uses simulation techniques, building characteristics, construction, climatic, and system data to calculate end-use energy consumption. Bottom-up building physics-based models provide useful insights to policymakers for enhancing end-user building efficiencies [29].

The majority of the bottom-up engineering models are based on distribution, sample, or archetype-based approaches [27]. The distribution approach determines the end-use energy consumption by analyzing the regional or national distribution of building energy use. The sample modeling approach refers to the use of actual sample building data as the model input information. This method requires an extensive database to represent the group of buildings. The archetype approach classifies the building stock according to dwelling types, size, climate, and construction year [37]. Each building archetype is modeled in the simulation engine to estimate energy consumption. These estimates are further scaled up to represent the regional or national building stock through aggregation [27].

The archetype bottom-up approach has been extensively used to understand the cumulative impact of energy performance strategies and new technologies for the regional or national building stock [8]. These approaches rely on quantitative data of building physics. The inputs required for these methods include thermal properties (U-values) of building elements (walls, windows, roof, floor, doors), internal or external temperatures, patterns of space heating systems, ventilation rates, number of appliances, occupancy factor, schedules, and internal loads [38, 36]. These models require several assumptions to determine occupant behavior and a significant amount of technical data to calculate energy consumption.

3.2.2. Physics-based Methods

The past decades has experienced tremendous growth in physics-based energy modeling studies in the United States and Europe. Some well-known studies in this field include

CityBES [39], UMI [40], CitySim [41], UrbanOPT [42], and Boston-UBEM [43].

Table 3: Summary of physics-based bottom-up urban building energy modeling studies [15, 19, 44].

Projects/ Paper	Simulation Engine	Stakeholders	Time scale	Availability	Developer	Reference
CityBES	EnergyPlus	Urban planners, policy makers	Sub-hourly	Research	LBNL	[39]
UMI	EnergyPlus	District energy managers	Hourly	Free	MIT	[40]
CitySim	CitySim Solver	Urban planners, policy makers	Hourly	Free	EPFL	[41]
UrbanOPT	EnergyPlus, OpenStudio	District energy managers	Hourly	Free, Open-source	NREL	[42]
Nageler et al.	IDA-ICE	Urban planner	Hourly	Research	IWT, TU Graz	[45]
Quan al.	Urban-EPC Engine	Urban energy policy maker	Hourly	Research	Georgia Tech	[46]
MIT's UBEM	EnergyPlus	Urban planner	Hourly	Research	MIT	[43]
SEMANCO	Custom simulation engine	Urban planners, policy makers	Yearly	Free	FUNITEC	[47]

One of the most prominent projects, City Building Energy Saver (CityBES) [48], provides a platform to analyze different retrofit scenarios when integrating building technology with performance and cost data. CityBES integrates the EnergyPlus simulation engine with CityGML and GeoJSON data formats to model a district or city-scale buildings. Another project, Urban Modeling Interface (UMI) [40], integrates Radiance/Daysim and EnergyPlus simulation engines to model urban building operational energy, daylighting, and walkability. CitySim project involves a decision support tool that aids energy planners and stakeholders to minimize energy usage and emissions and further integrate various optimization and retrofit analyses [49, 41]. URBANopt (Urban Renewable Building And Neighborhood optimization) [42] provides the EnergyPlus and OpenStudio-based simulation platform that facilitates district-level modeling and campus-scale thermal and electrical analyses. MIT's UBEM [43] platform uses the EnergyPlus simulation engine to model approximately 83,541 buildings by integrating official GIS datasets and a custom building archetype library.

The aforementioned studies use a similar modeling structure although specific to the developed cased study. It is therefore crucial to devise a framework that is generalized and henceforth could be used to model different use cases. This paper proposes a generalized bottom-up physics-based modeling framework based on the existing literature [8]. The generalized methodology for urban-scale bottom-up physics-based modeling requires several steps that include model inputs, followed by a simulation engine to perform energy simulations, validation, and finally, visualization and analysis (Fig. 7).

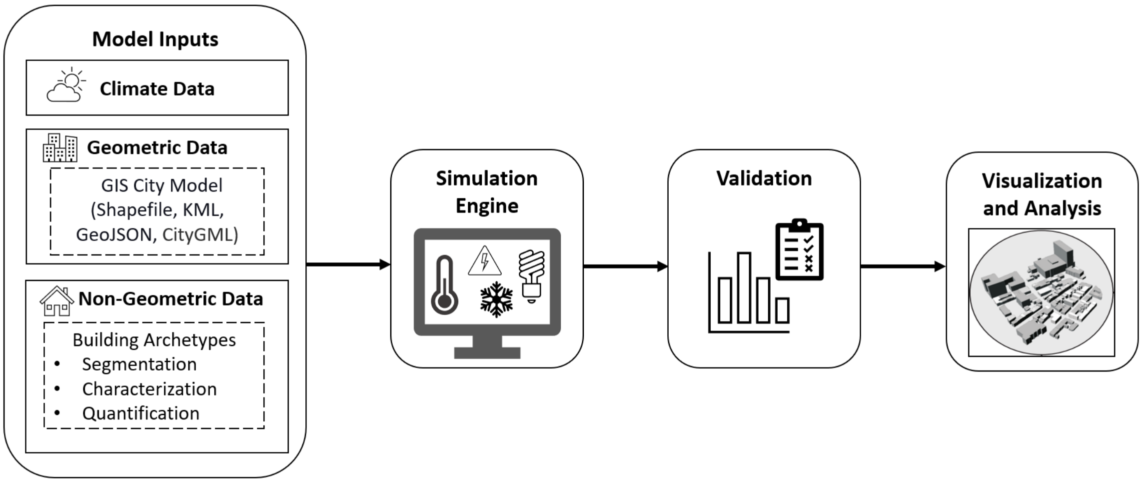

Figure 7: Generalized methodology for physics-based bottom-up building energy modeling at urban scale based on the existing literature [8].

Model Inputs: Bottom-up physics-based models require several data inputs, including climate, building geometry, and non-geometry data [8]. Weather data sets are needed for building thermal simulations. Typically, hourly climate datasets, referred to as Typical Meteorological Year datasets (TMY), have been collected for several years. However, existing data sets are based on historical data, which do not consider climate change scenarios, uncertainties, future weather files or extreme conditions [50].

The geometry input data required by the bottom-up approach consists of building envelope data, shapes, geometry, and geospatial positions [50]. Typically, geometric building data are collected through Geographic Information Systems (GIS) city model database. The building geometric data exists in various formats, and no standard building identifier is used to link the different sets of data. The most common data model formats are Shapefile, KML, GeoJSON, and CityGML [20]. However, the Shapefile, KML, and GeoJSON formats do not provide a schema to describe the building characteristics, leading to property inconsistencies between various datasets [15]. Recently, CityGML has grown popular; it is a comprehensive data format within the geospatial domain based on the Open Geospatial Consortium (OGC) standard [51]. CityGML can be represented in five different levels of details (LOD 0 to LOD 4) [52]. CityGML presents several Application Domain Extensions (ADEs) used to enrich the 3D city model and develop user-defined objects and attributes for bottom-up building energy modeling. The Energy ADE for CityGML is currently under development and focuses on integrating the building geometric and non-geometric for energy simulations at an urban scale using bottom-up approach [39, 53]. For instance, CityBES [39] uses building stock data in both CityGML and GeoJSON formats as model input along with weather data. The MIT UBEM [43] model uses a TMY file for weather data, GIS shapefiles for geometric data, and characterized non-geometric input data for modeling archetypes. The UMI model [40] uses a variety of geographic and urban fabric-specific input parameters such as building typology, construction type, and vegetation. URBANopt [42] can use different formats (e.g., GeoJSON, CityGML) that input features such as location, floor area, number of stories, building type, cooling source.

In addition, bottom-up modeling requires non-geometric building properties, including envelope construction, user occupancy, usage pattern, infiltration rates, equipment loads, and HVAC system type [50]. One of the significant challenges associated with physics-based models relates to the aforementioned information's availability to model the entire urban area. Typically, non-geometric building data are gathered through the building archetypes approach using existing available building stock data. Archetype development requires three major steps, namely, segmentation, characterization, and quantification [16, 54]. The segmentation process involves building stock classification according to building type, age, and climate. The characterization process defines a complete set of thermal and building physics characteristics, including construction materials, usage patterns, and building systems for each archetype [55]. The quantification process determines the number of buildings under each building archetype using national census statistics [16].

Simulation Engine:

Physics-based modeling approaches use a simulation engine to capture the full dynamic of high-resolution building performance [15]. Simulations are performed using the climate, geometric and other inputs available for each developed building archetype model [56]. At the urban scale, energy modeling is a highly complex and time-consuming process. Generally, simulation engines, such as EnergyPlus [26], IDA-Indoor Climate Energy (IDA-ICE) [57], DOE2 [58], IES VE [59] and TRNSYS [60], are used to calculate the thermal loads of buildings. These simulation engines have ben broadly used for dynamic simulation models. In dynamic building energy simulations, building envelope phenomena are captured for annual and hourly or even sub-hourly granular level [20]. The most commonly used simulation engine is EnergyPlus, as this engine allows for an in-depth analysis of complex urban building systems [16].

Validation:

Validation of the results is of paramount importance in the bottom-up modeling process. The resulting models comprise several simplifications as energy performance assessment uses aggregated building properties obtained from archetypes/reference models [20]. It is crucial to produce reliable estimates to support stakeholders in the implementation of energy-efficient policy decisions. Individual building modeling results may differ significantly from measured results due to assumptions in inputs such as those relating to occupant behavior. However, when aggregating multiple buildings' energy estimates, the modeling inaccuracies tend to average out with errors ranging between 7% and 21% for thermal load estimation and 1% and 19% for total Energy Use Intensity (EUI) [8, 61, 62]. The modeling results of each archetype are aggregated at the urban level and can be compared with national energy use consumption statistics [16] or against the established validation criteria [63]. Previous studies often perform results validation at an aggregated level [8]. The reliability of these results highly depend on model validation against measured data as archetypes only partially represent the actual diversity of buildings including usage patterns. For instance, the MIT UBEM model [43] validates the hourly and annual consumption of Boston city against the measured energy use and fuel consumption. The resulting deviations lie between 5 and 20%

for individual zip codes. Another study by Quan et al. [46] performed a similar annual energy use validation for 1680 buildings in Manhattan. For 80% of the buildings, the estimated energy use lied within the range of 0.5–2 times of the measured energy use, suggesting an overall good fit. Tools such as CityBES also perform validation using test cases from standards such as ASHRAE Standard 140. According to ASHRAE guidelines, the computer model shall have an NMBE of 5% and a CV(RMSE) of 15% relative to monthly aggregated calibration energy consumption data. If hourly calibration data are used, these requirements shall be 10% and 30%, respectively [64].

Visualization and Analysis:

Stakeholders could eventually use the modeling results to perform building performance analysis at an urban scale. The model outputs often comprise large datasets with a high resolution both spatially and temporally. Therefore, computationally efficient solutions are required to manage the results for visualization and further processing [19]. The best way to present extensive scale data in 3D or 2D format is to use GIS data models, namely, shapefile, KML, or CityGML [65]. CityGML is a popular choice to store the building geometric data and present data in 3D format [66]. GIS can help visualize building energy performance results with socio-economic, demographic, and other correlated data [50].

3.2.3. Physics-based Quantitative Analysis

The bottom-up physics-based quantitative analysis uses the Web of Knowledge search engine's advanced search function using Eq.1 and Table.2. The keywords search is based on any paper's initial content, such as title, abstract, and keywords used in a paper. The main keywords include physics-based, simulation, and engineering. A total of more than 2368 articles are found to contain these keywords. The research publication trend between 2001-2020 indicates that the number of papers published has experienced a significant increase after 2012 (Fig. 8). This is indicative of the popularity of the bottom-up approach in building energy modeling. The main applications of bottom-up models include building energy analysis (29%), optimization (21%) and energy predictions (20%) (Fig. 8). The top publication sources on the physics-based approach topics are Energy and Buildings and Applied Energy journals: 83% of the papers were published as articles, and the 15% were published as conference proceedings (Fig. 9).

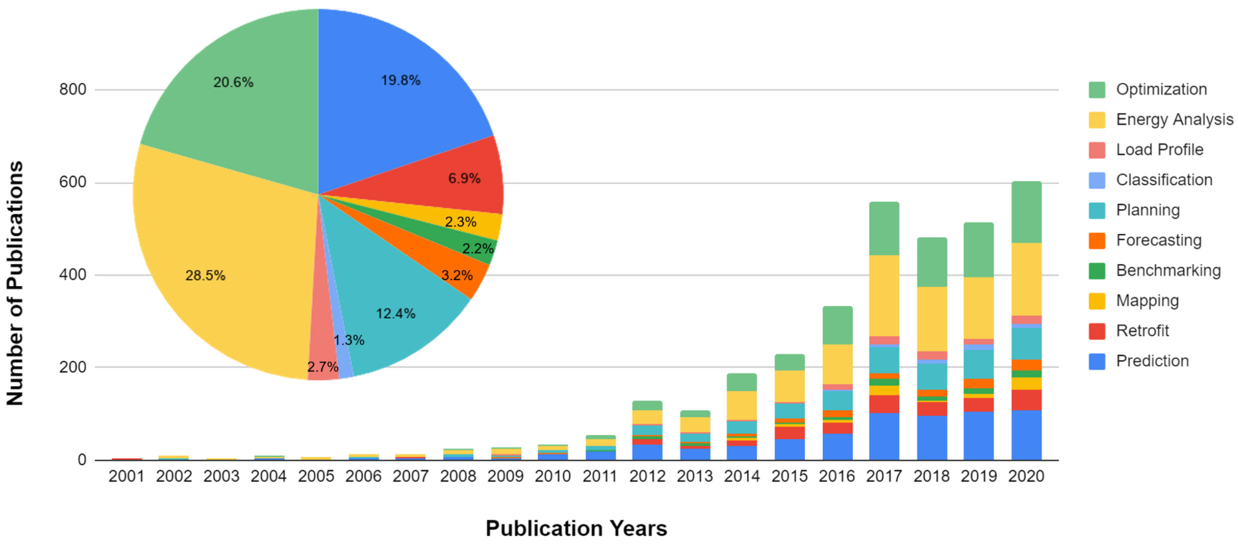

Figure 8: Physics-based urban building energy modeling publication trend with intended area of implementation from the World of Science academic search platform.

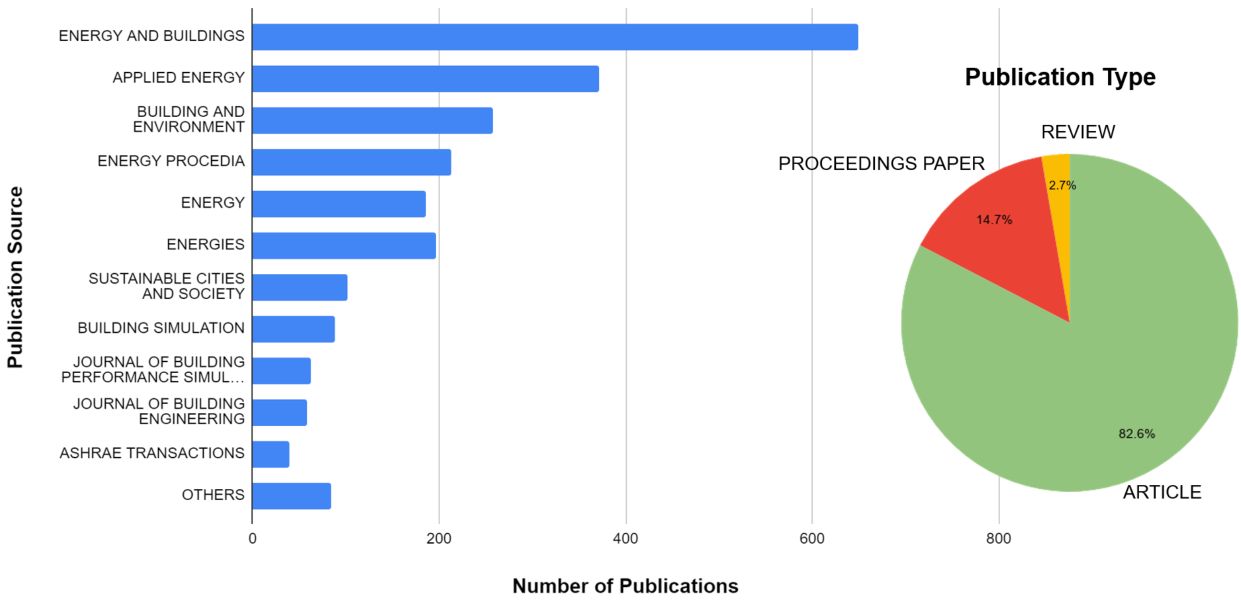

Figure 9: Physics-based urban building energy modeling publication distribution according to publication source and publication type from World of Science academic search platform.

3.2.4. Physics-based Qualitative Analysis

The qualitative analysis of bottom-up models determines the strengths and weaknesses associated with these models' past implementation. Furthermore, this analysis identifies the threats and challenges of bottom-up model suitability and scope, which would eventually evaluate various opportunities to enhance these models (Fig. 10).

The bottom-up approach provides a two fold advantage. By facilitating the evaluation of current and prospective technologies. These models further enable the assessment and quantification of the effects of combining different technologies on the delivered building energy [36]. For instance, this approach could model the on-site building energy usage or generation using renewable energy technologies [18]. The bottom-up models determine each building's end-use energy consumption by type and rating using measurable data [27]. The bottom-up physics-based approach does not require detailed historical energy consumption and socio-economic factors [22].

The most apparent drawback of the bottom-up physics-based approach relates to the availability of detailed physical and technical building information [36]. Bottom-up models incorporate simplifications and assumptions regarding human-related factors for energy modeling [18]. The modeling process further neglects and fails to capture the relationships between energy use and macroeconomic activities. The bottom-up approach uses a simulation engine, and due to the complexity of the urban systems, that is computationally intensive and time-consuming [27]. Furthermore, socio-economic factors often capture the change in social policies and human behavior that requires historical records to formulate the dependencies. As physics-based models simulate building energy use based on the physical characteristics of the building and use single inputs, these models do not consider the long term interactions between energy use and socio-economic factors [22]. As these methods often entail numerous parameters that might be uncertain, further calibration might be required to account for input uncertainties. Furthermore, the existing calibration methods are more generalized, and there is still a need to further enhance the archetype development process, especially in classification and characterization methodologies.

Figure 10: Strengths, weaknesses, opportunities, and threats analysis of bottom-up physics-based modeling for qualitative analysis.

The biggest challenge in the bottom-up physics-based modeling is the granular level building archetype development. There are still opportunities to further improve the archetype development process, particularly in segmentation and characterization steps. Machine learning and data mining techniques can help to improve the data quality required for building archetype development. Parallel and cloud computing-based simulation are possible solutions to improve computation efficiency. These technologies offer faster simulations, even for more complex building energy models [67]. Physics-based thermal modeling relies on climate data; therefore, the integration of full micro-climate models can improve modeling results [68]. Furthermore, there is a need for realistic modeling of human activity at the building level because of limited consideration of occupant-driven factors [69, 20, 70].

The most plausible threat to a bottom-up physics-based approach is occupants' behavioral assumptions as the effect of occupant behavior can significantly impact building energy consumption [27]. Furthermore, due to the lack of detailed building data at the urban scale, this approach can introduce significant uncertainty in building energy estimates [36]. The simulated energy-use data using physics-based methods require measured data for validation. However, the measured data is usually not available for individual buildings, this poses a huge risk when considering the accuracy of the developed models.

3.3. Data-driven Bottom-up Urban Building Energy Modeling

3.3.1. Data-driven Approaches

The bottom-up data-driven method is used for urban energy modeling to predict and estimate building energy consumption, including the basic knowledge of the buildings' features [71, 72]. These methods rely on existing available data, such as building stock datasets, billing data (e.g., electricity, gas), survey data, and socio-economic variables. Generally, data-driven modeling comprise statistical and Artificial Intelligence (AI) approaches (Fig 11) [18]. The statistical approach involves regression methods to determine the inverse mathematical models corresponding to the building design or operational parameter data. These models regress aggregate building energy consumption onto parameters or combinations of parameters that influence the energy consumption [27]. The most commonly used algorithms include Linear Regression (LR), Multiple Linear Regression (MLR), Non-linear Regression (NR), and Conditional Demand Analysis (CDR) [18, 22].

An artificial intelligence approach is mainly based on Machine Learning (ML) techniques that model the energy use of urban buildings by automatically learning the data patterns. The model learns and trains with the historical dataset to find the mathematical association between building energy use and influential factors such as building features, urban properties, and occupancy features [18].

Generally, machine learning techniques are grouped into two main categories, namely, supervised and unsupervised learning (Fig 11) [73, 74]. Supervised learning predicts the output through the formulation of a sophisticated relationship between multiple features. This approach further comprises regression and classification algorithms [75, 76]. A classification algorithm learns to predict output data when the outcome is a label, such as energy rating and building type [73]. Regression algorithms learn from the input data and predict real output values, such as energy consumption [73]. Unsupervised learning algorithms learn from the input data and discover inherent structure, associations, or unknown patterns in the data, for example, energy consumption patterns [75].

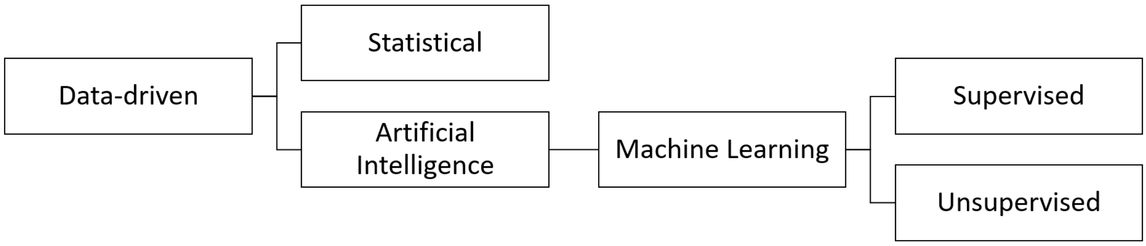

Figure 11: Data-driven modeling techniques for modeling building energy at urban scale.

3.3.2. Data-driven Methods

In recent urban energy studies, artificial intelligence models using machine learning algorithms have been widely used compared to traditional statistical techniques. Statistical models are best for determining associations between variables, while machine learning models identify the most accurate predictions possible. Furthermore, recent advancements in meter and open data initiatives have accelerated the development of new data-driven machine learning models that have produced computational intelligence to discover hidden patterns of urban building energy use [77]. The data-driven models have multiple applications such as prediction, forecasting, benchmarking, mapping, and classification [78, 11, 12].

For instance, Mastrucci et al. [79] proposed the bottom-up statistical methodology based on a Geographical Information System (GIS) to estimate residential urban area building energy consumption. Pedersen et al. [80] suggested that the load prediction method estimates the heat and electricity demand profiles of different building categories to plan hybrid energy distribution systems. Dall'O' et al. [61] proposed a methodology for building energy performance classification of urban residential stock using a statistical approach. This methodology's data are based on the existing available building stock data (cartographic documentation, thematic maps, geometric data, and others) from energy audits on sample buildings. Williams et al. [81] established reliable statistical learning models for predicting future monthly residential energy consumption using building features and climate data.

Recent studies have extensively used machine learning-based modeling to predict future trends (Table 4). For instance, Rahman et al. [82] developed and optimized models to predict medium to long-term electricity consumption for commercial and residential buildings using deep recurrent neural networks. Robinson et al. [83] proposed a methodology to determine commercial building energy consumption by training machine learning models (gradient boosting model, linear regression, SVR, Random forest) using national data from the Commercial Buildings Energy Consumption Survey (CBECS). Abbasabadi et al. [84] proposed an integrated framework for Urban Energy Use Modeling (UEUM) that uses a machine learning approach to model urban building and transportation energy. Another framework, Data-driven Urban Energy Simulation (DUE-S) [77] uses synthesis data by combining physics-based simulation and machine learning methods to model multi-scale urban energy. Data-driven Urban Energy Benchmarking (DUE-B) is a data-driven methodology based on recursive partitioning (Classification and Regression Tree (CART)) and stochastic frontier analysis for benchmarking urban building energy consumption. The DUE-B [85] evaluated over 10,000 buildings' energy performance in New York City using real energy and building data. Kontokosta et al. [86] proposed statistical models (linear regression, random forest, and support vector regression) to estimate the electricity and natural gas energy use of 1.1 million buildings in New York City with building physical, spatial, and energy use characteristics datasets. GREEN grading [87] system framework implements the XGBoost machine learning algorithm to assess building energy performance using existing energy use and building data. The framework is applied to 7,500 residential buildings in New York City for urban energy efficiency and carbon reduction policies. Ali et al. [88] proposed a generic methodology to optimize energy retrofit decisions for urban scale residential buildings using machine learning algorithms.

Table 4: Summary of data-driven bottom-up urban building energy modeling studies.

Projects/ Paper	Model	Application	Target Users	Developer	Reference
UrbanFootprint	Statistics	Mapping	Urban planners, policy makers	LBNL	[89]
DUE-B	Classification and Regression Tree	Benchmarking	Policy makers	Urban Informatics Lab	[85]
CoBAM	Statistics	Energy Analysis	Policymakers, practitioners and local communities	EPFL	[90]
Energy Proforma	Statistics	Energy Analysis	Urban planners, policy makers	NREL	[91]
DUE-S	Residual network model	Energy Analysis	Policy makers	Urban Informatics Lab	[77]
UEUM	k-NN, and ANNs	Energy Analysis	Designers, planners, and policymakers	Illinois Institute of Technology	[84]
GREEN grading system	XGBoost	Benchmarking	Policymakers	New York University	[87]
Dall'O' et al	Regression	Classification	Local administrators	BEST Politecnico di Milano	[61]
Pedersen et al	Regression	Load prediction	Urban planners	Norway, NTNU	[80]
Mastrucci et al	Multiple linear regression	Urban planning	Urban planners	CRTE	[79]
CRECM	Statistical method	Energy Analysis	City policy makers	SCUT	[92]
Rahman et al	Recurrent neural network	Forecasting	Planning	SSESLab, University of Utah	[82]
Robinson et al	Gradient boosting model, Linear regression, SVM	Energy Analysis	City planners and policy makers	Georgia Tech	[83]
Kontokosta et al	Linear regression, random forest, and support vector regression	Energy Analysis	Policymakers	New York University	[86]
Williams et al	Linear regression, regression trees, and MARS	Energy Analysis	Policymakers	TSERI	[81]
Ali et al	Deep Learning	Retrofit analysis	Urban planners and energy policymakers	University College Dublin	[88]

This paper proposes generalized bottom-up data-driven modeling frameworks based on the existing literature [16]. Generally, a data-driven artificial intelligence approach using machine learning algorithms requires several steps for urban building energy modeling. The generalized methodology starts with data collection, data pre-processing followed by feature selection, data splitting for training and testing purposes, machine learning model development, and finally, model performance evaluation (Fig. 12).

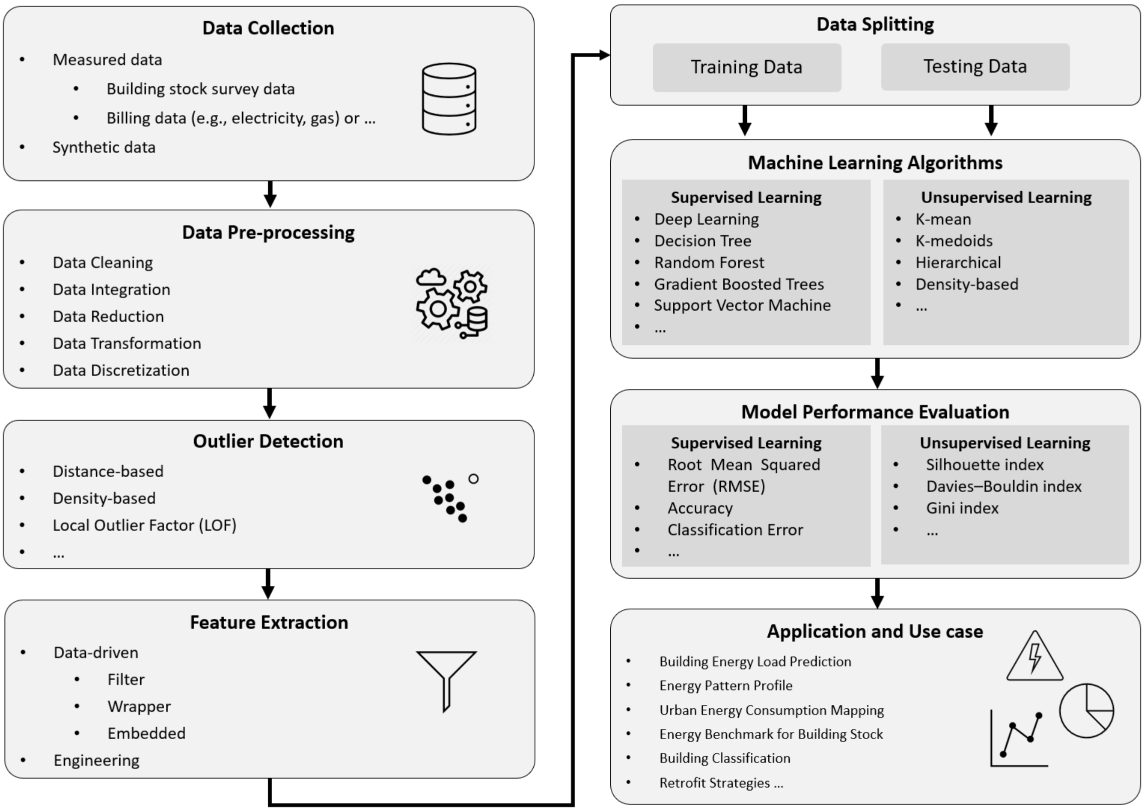

Figure 12: Generalized methodology to model urban building stock using data-driven bottom-up building energy modeling [16].

The collection process collects the datasets needed for data-driven modeling. Datasets are classified into synthetic or measured data. Synthetic data can be generated by using a physics-based simulation platform. Measured datasets include building stock databases, billing data (e.g., electricity, gas), surveys, and benchmark data. At the urban scale, buildings measured data can be collected from existing building stock databases such as the building Energy Performance Certificates (EPC) database that contains building characteristics information. The information includes shape, type, fabric, construction assemblies, and Heating Ventilation and Air Conditioning (HVAC) system properties. For instance, the United States' Department of Energy (DOE) [93] manages the nation's largest dataset of information about commercial and residential buildings' energy-related characteristics (Building Performance Database (BPD)).

Similarly, each European Union member state maintains its own EPC database containing primary information about building stock with energy performance [94]. The energy consumption dataset often includes information regarding building energy use (electricity and gas) and is collected from billing data. Furthermore, quantification data are needed to quantify the number of buildings present in an urban area from national statistics or census databases.

The data collected from surveys and measurement methods for modeling are often incomplete or lack several essential variables [16]. In the pre-processing stage, the data passes through several operations such as data cleaning, data integration, data reduction, data transformation, and data discretization [95]. Simialry, outlier detection or removal is a necessary step before implementing a machine learning algorithm. Outlier detection has been widely used for abnormal building energy consumption detection, load profile discord identification, and fault detection analysis in buildings [96, 97, 98].

Feature extraction eliminates unnecessary or irrelevant variables and determines the essential features for developing the data-driven model. This method helps to reduce the model's complexity, computational load, and improves model performance [99]. Feature extraction can be performed using data-driven and engineering methods. Data-driven feature extraction can be done by filter, wrapper, and embedded methods [100, 101, 102]. Engineering methods are manual processes in which existing literature and expert opinion judgment are considered [103, 104, 105]. Feature extraction can also be used for learning models and retrofit decision optimization. For instance, Seyedzadeh et al. [106] proposed optimizing data-driven machine learning models for both heating and cooling loads features by multi-objective optimization. Ali et al. [88] proposed a novel generalized approach to optimize urban scale residential building features for energy retrofit decisions using data-driven strategies.

Before implementing machine learning algorithms, data splitting is distributing the dataset across two subsets; a training dataset (a subset to train a model) and a test dataset (a subset to test the trained model) [107]. Commonly used supervised learning algorithms include Gradient Boosted Trees (GBT), Logistic Regression (LG), Deep learning (DL), Artificial Neural Networks (ANN), Decision Tree (DT), Random Forest (RF), Generalized Linear Model (GLM), Support Vector Machine (SVM), and Support Vector Regression (SVR) [108, 75, 109, 73]. Over the past two decades, the deep learning or neural network algorithm performed well compared with other supervised learning algorithms to model the building energy use at a large scale [110]. For instance, Ali et al. [111] proposed a data-driven approach for multi-scale GIS-based building energy modeling using deep learning algorithms with an accuracy of 88%. Nutkiewicz et al. [77] proposed a novel Data-driven Urban Energy Simulation (DUE-S) framework that can accurately predict urban scale energy consumption at hourly, daily and monthly intervals using a residual network (ResNet) machine learning model. Deng et al. [112] compared ANN, SVR, RF, and GBT for prediction of US commercial building energy use and the individual energy end-uses of HVAC, lighting, and plug loads using 2012 Commercial Building Energy Consumption Survey (CBECS) microdata. Abbasabadi et al. [84] proposed the framework for data-driven urban energy use modeling using MLR, RF, ANN, Nonlinear Regression (NLR), Classification and Regression Trees (C&RT) and k-Nearest Neighbors (k-NN) algorithms. Robinson et al. [83] used RF, SVR, and GBT to estimate New York City commercial building energy consumption using data from the Commercial Buildings Energy Consumption Survey (CBECS). Beccali et al. [113] used the ANN decision support tool to assess 151 commercial and educational buildings' energy performance in South Italy's four regions.

The standard unsupervised learning algorithms include k-mean, k-medoids, hierarchical, and density-based [75, 74, 114]. The unsupervised learning algorithms widely used clustering analysis at the urban level for categorizing buildings' functionality, characteristics, and consumption patterns, to model urban building efficient energy performance. For instance, Jaeger et al. [115] established a building clustering approach for modeling residential buildings of 1230 single-family dwellings using a k-mean algorithm. Tardioli et al. [116] proposed a methodology for calibration of building energy models of 2646 buildings from the city of Geneva, Switzerland, using clustering and surrogate techniques. Ali et al. [114] performed that detailed comparative analysis of clustering algorithms for building archetypes development in urban building energy modeling. The results show that the k-mean algorithm performs the best in terms of cluster formation. Papadopoulos et al. [117] implemented k-means clustering to identify buildings with similar temporal energy performance patterns using commercial and residential buildings energy time-series data.

To examine the effectiveness of supervised learning models evaluates using standard performance measures [76]. Generally, the Classification Accuracy (CA), Classification Error (CE), precision and recall are used for classification models [88]. Regression model performance can be evaluated using the Coefficient of Variation (CV), Root Mean Squared Error (RMSE), Normalized Mean Bias Error (NMBE) and Mean Absolute Percentage Error (MAPE) [118, 108]. Furthermore, NMBE and RMSE have been commonly used as evaluation criteria for building energy models by international standards, such as ASHRAE [108].

The unsupervised models based on clustering algorithms can be evaluated using internal validity indices that determine the resulting clusters' properties, such as compactness, separation, and roundness. The most common performance indices are the silhouette index, davies Bouldin index, Gini index, and cophenetic correlation coefficient [114].

Data-driven approaches are widely applied to building energy load prediction, energy pattern profiling, urban energy consumption mapping, energy benchmarking for building stock, building classification, and identification of retrofit strategies [76].

3.3.3. Data-driven Quantitative Analysis

Similar to the top-down analysis, the bottom-up data-driven quantitative analysis uses the Web of Knowledge search engine's advanced search function using Eq.1 and Table.2. The keywords search is based on any paper's initial content, such as title, abstract, and keywords used in a paper. The main keywords include data-driven, machine learning, artificial intelligence, and statistic. A total of more than 589 articles are found that contain these keywords. The research publication trend between 2001-2020 indicates that the number of published papers have experienced a steep increase after 2012 (Fig. 13). This is indicative of the increasing implementation of different data-driven models in the building simulation domain. The keywords analysis based on modeling methods indicates that 33% and 67% of the published papers contain machine learning and statistical keywords (Fig. 15).

Implementing machine learning methods in the building simulation domain offers tremendous potential for identifying pattern in existing energy consumption profiles. The main applications of data-driven approaches include building energy analysis (26%) and energy prediction (25%). These applications do not commonly address energy benchmarking and retrofit analysis (Fig. 13). Commonly used algorithms for data-driven modeling include regression, ANN, and SVM (Fig. 15). The top publication sources on the data-driven approach topics are Energy and Buildings and Applied Energy journals: 88% of the papers were published as articles, and 10% were published as conference proceedings (Fig. 14).

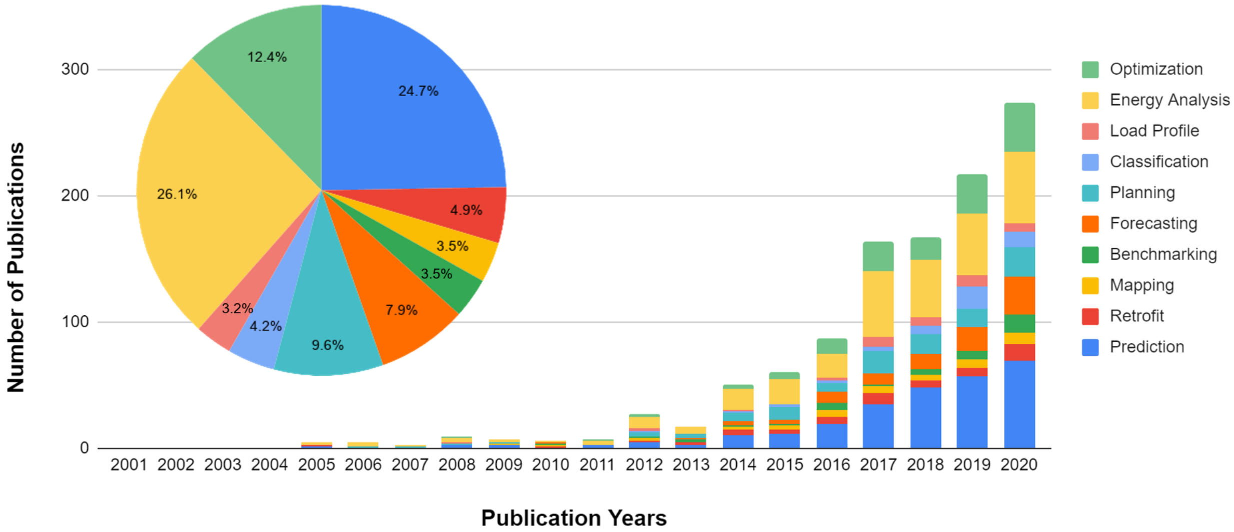

Figure 13: Data-driven urban building energy modeling publication trend with intended area of implementation from the World of Science academic search platform.

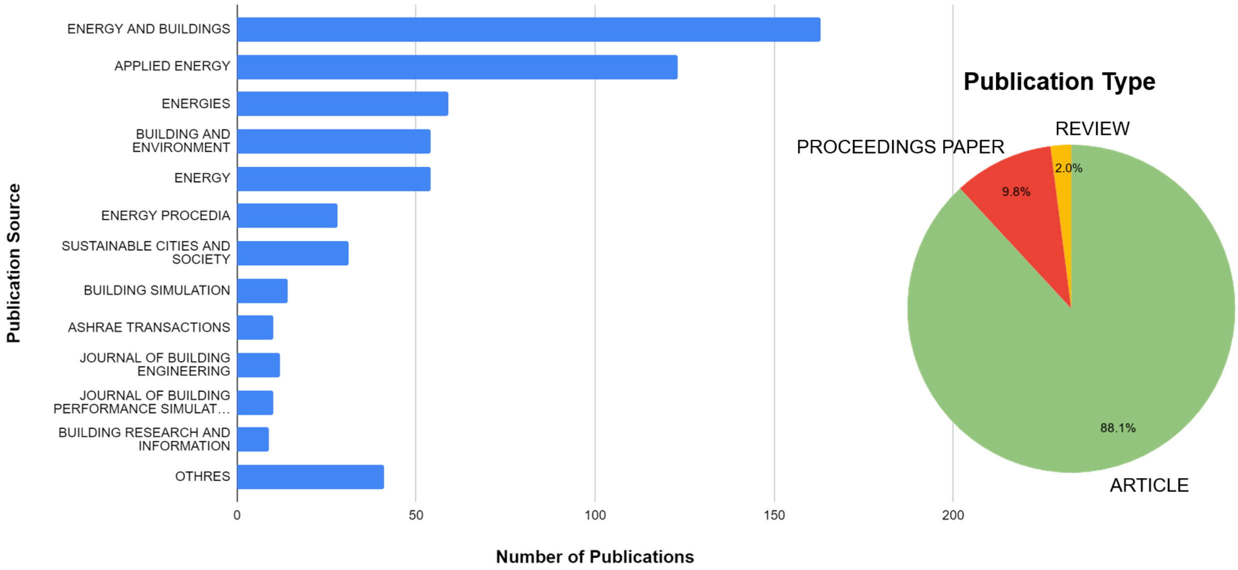

Figure 14: Data-driven UBEM publication distribution according to source and type from World of Science academic search platform.

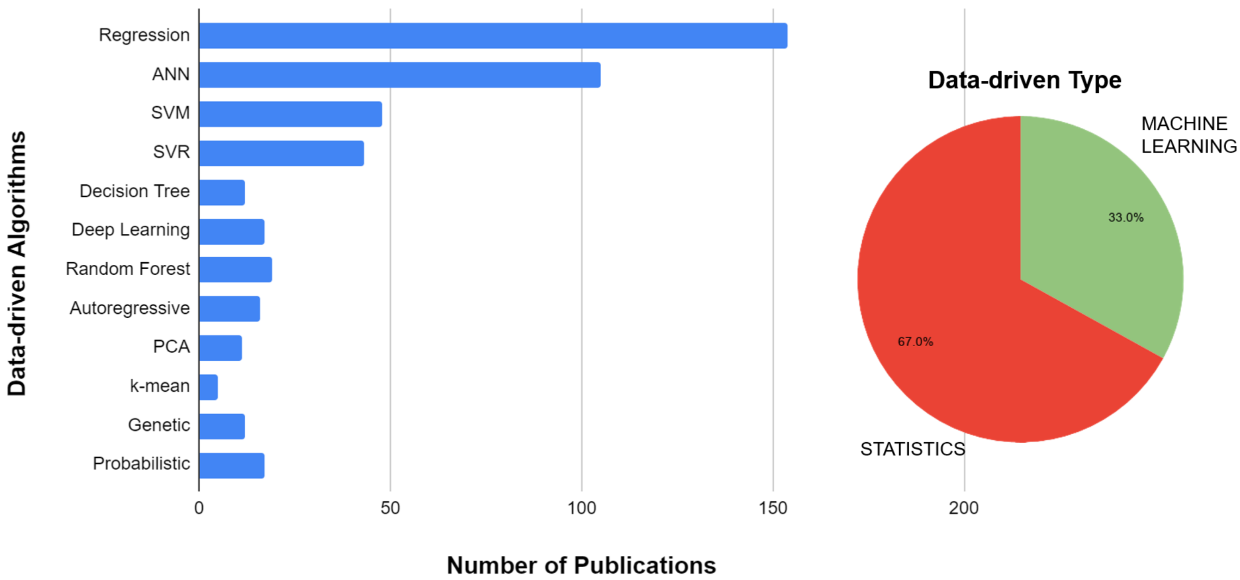

Figure 15: Data-driven urban building energy modeling publication distribution according to algorithms, data-driven model type from World of Science academic search platform.

3.3.4. Data-driven Qualitative Analysis

The qualitative analysis of data-driven models determines the strengths and weaknesses associated with these models' past implementation. Furthermore, this analysis identifies the threats and challenges of data-driven model suitability and scope, which would eventually evaluate various opportunities to enhance these models (Fig 16).

Figure 16: Strengths, weaknesses, opportunities, and threats analysis of bottom-up data-driven modeling approaches.

Data-driven approaches are relatively easier to formulate compared to other approaches. These approaches provide an meaningful representation of urban energy where existing data are limited [18]. The data-driven modeling process does not require detailed building physics data (geometric and non-geometric) when compared to a simulation-based modeling approach. Furthermore, data-driven models can effectively estimate the typical end-user building energy consumption by incorporating occupancy and socio-economic factors [36].

These models can also further learn any correlations between input variables and output variables when the data are of sufficient quality [75].

As data-driven approaches rely on historical data, these approaches offer limited modeling capability when assessing the impact of retrofit or new technologies with limited data [29]. It is often required to conduct extensive surveys or collect large-scale data samples when the data are not available [18]. The data-driven models offer limited flexibility for design modifications, and hence, the application of the developed models is restricted to a particular location and building type [15]. Furthermore, data-driven modeling could be computationally intensive and time-consuming, especially when implementing machine learning algorithms.

The advancement in smart meters and advanced metering infrastructure admits new methods of collection. Easy access to the smart-metering data would open novel opportunities for different analysis types to enhance the current understanding of building energy performance [118]. As data-driven modeling algorithms are trained using existing building stock datasets, these algorithms might eliminate essential variables to evaluate building energy performance, such as buildings physical properties, operation strategies, weather conditions, and occupant behavior [76]. Therefore, these algorithms must further determine crucial features to enhance the suitability of a data-driven model [108]. Furthermore, to reduce computational complexity, the model formulation process could be integrated with a cloud-based resources such as Google Cloud Platform or Amazon Web Services. These models would eventually provide opportunities to examine the integration of building stock energy performance and social science-based research (such as occupancy behavior and demand patterns) [72].

One of the significant threats to data-driven modeling concerns the quality of the building stock dataset. As the data are acquired through surveys, they might contain numerous human errors, such as duplicate data entries, incomplete records, and broken formats. Low-quality data does not yield accurate results, especially when training a machine learning model. Furthermore, survey-based data for modeling often creates a massive gap between predicted and actual energy use (or measured energy use). Moreover, privacy concerns associated with the use of consumer energy data would pose a significant threat to the implementation of data-driven modeling.

3.4. Reduced-order Bottom-up Urban Building Energy Modeling

3.4.1. Reduced-order Approaches

Reduced-order approaches are being widely adopted to perform a quick assessment of building energy performance. These approaches require fewer inputs than the physics-based and data-driven approaches [15]. One of the processes specific to this approach is the determination of model parameter values, which could be estimated using various calculation standards developed by the European Committee for Standardization (CEN) and the International Organization for Standardization (ISO). These standard procedures define the calculation method through a set of normative statements that contain the physical building parameters and the associated systems for different building types. These calculation methods have been implemented to calculate the energy performance rating in the EU

[119]. ASHRAE's standard thermal network model describes the heat transfer and thermal dynamics through the building envelope and the subsequent effect on indoor temperature [120].

Table 5: Summary of reduced-order bottom-up urban building energy modeling studies [15]

Projects	Method	Target Users	Time scale	Availability	Developer	Reference
SimStadt	ISO/CEN standards based reduced-order model	Urban planners, policy makers	Monthly	Research	Hochschule für Technik Stuttgart	[121]
LakeSIM	ISO/CEN standards based reduced-order model	Urban planners, policy makers	sub-hourly	Open	ANL	[122]
OpenIDEAS	Modelica based reduced order model	District energy managers	Hourly	Free, open-source	KU Leuven	[123]
TEASER	Modelica based reduced order model	District energy managers	Hourly	Free, open-source	RWTH Aachen University	[124]
Energy Atlas	ISO/CEN standards based reduced-order model	Urban planners, policy makers	-	-	Technische Universität München	[125]
CEA	Tool specific calculation modules	Urban planners, policy makers	Hourly	Free, open-source	ETH Zurich	[126]
Tool by Georgia Institue of Technology	ISO/CEN standards based reduced-order model	Urban planners, policy makers	-	-	Georgia Institute of Technology	[127]

3.4.2. Reduced-order Methods

Commonly used tools that employ reduced-order methods include SimStadt [121], City Energy Analyst (CEA) [126], TEASER [124], OpenIDEAS [123] using the Modelica Buildings library [128] (Table 5).

One of the prominent projects include SimStadt [121] that provides an energy modeling and energy simulation platform for monthly and hourly analysis of city districts using CityGML city models. City Energy Analyst (CEA) [126] is an integrated open-source tool for analyzing and optimizing urban building energy systems. CEA offers energy demand and supplies analysis for buildings to support energy efficiency planning. Remmen et al. [124] developed a TEASER tool to address the challenge of data acquisition required for dynamic modeling of buildings at an urban scale. The tool uses a data enrichment process for the identification of the network model for various building archetypes. However, the tool implements standard grey box networks that only represent a certain level of complexity (maximum network order of 4). Also, the grey-box networks in TEASER have been developed according to German standards. Hence, they still need to be integrated with international statistical building stock data to extend its applicability to other countries. Another urban scale project OpenIDEAS [129] formulated an open framework for integrated building and district energy simulations based on Modelica based reduced-order models. The framework allows simultaneous transient simulation of thermal and electrical systems at both building and feeder levels. The Modelica buildings library [130] is an open-source library with dynamic simulation models for rapid prototyping of new building and district energy and control systems.

This paper proposes generalized reduced-order modeling frameworks based on the existing literature [121, 124, 126, 129]. These studies follow a similar urban-scale bottom-up reduced-order modeling structure that involves several steps namely model input identification, followed by reduced-order model formulation, simulation, validation and finally, model implementation (Fig. 17).

Bottom-up reduced-order models require building meta-information; the requirements depend on the complexity of the model [131]. When considering simplified models (single-zone), any prior information of the building cluster is usually not required besides the buildings' location as described in [13]. TMY datasets provide the hourly climate profiles of any particular location. Remmen et al. introduced a parameter enrichment procedure to help identify initial guesses and further validate estimated parameters [124] when considering enhanced reduced-order models. The enrichment procedure uses information regarding the building size, window area, and building envelope. Furthermore, the position of available zone measurements is very useful for determining the zoning strategies for multi-zone models [132]. The formulation of these models requires the knowledge of existing HVAC systems inside the building to adapt the model structure as per the installed system.

Guang et al. replaced the meta-information with a brief analysis of the available data [133]. The operational dataset should contain a minimal set of measured variables to identify the model parameters, including ambient and indoor temperature profiles and heating or cooling loads at hourly (or lower) intervals. Previous studies further recommend the availability of electricity consumption variables to enhance model accuracy [121]. When modeling the HVAC systems, the modeling process also requires the energy use measurements of different components. Occupancy factors could be modeled indirectly using internal heat gains, relative humidity, or electric plug consumption profiles.

As identified from the literature, reduced-order model formulation at the urban level comprises two crucial procedures, namely, model identification and network parameterization [124]. Bacher et al. devised a model identification procedure that employs a forward selection strategy or statistical method to identify the simplest model (first order single-zone model) with fewer free parameters [131]. The method further adds additional parameters to increase the model complexity until the model accuracy could not be improved further. Previous studies either implements a manual trial and error identification procedure or an automated one where the user identifies a set of models and specifies the validation tests in the model workflow [134].

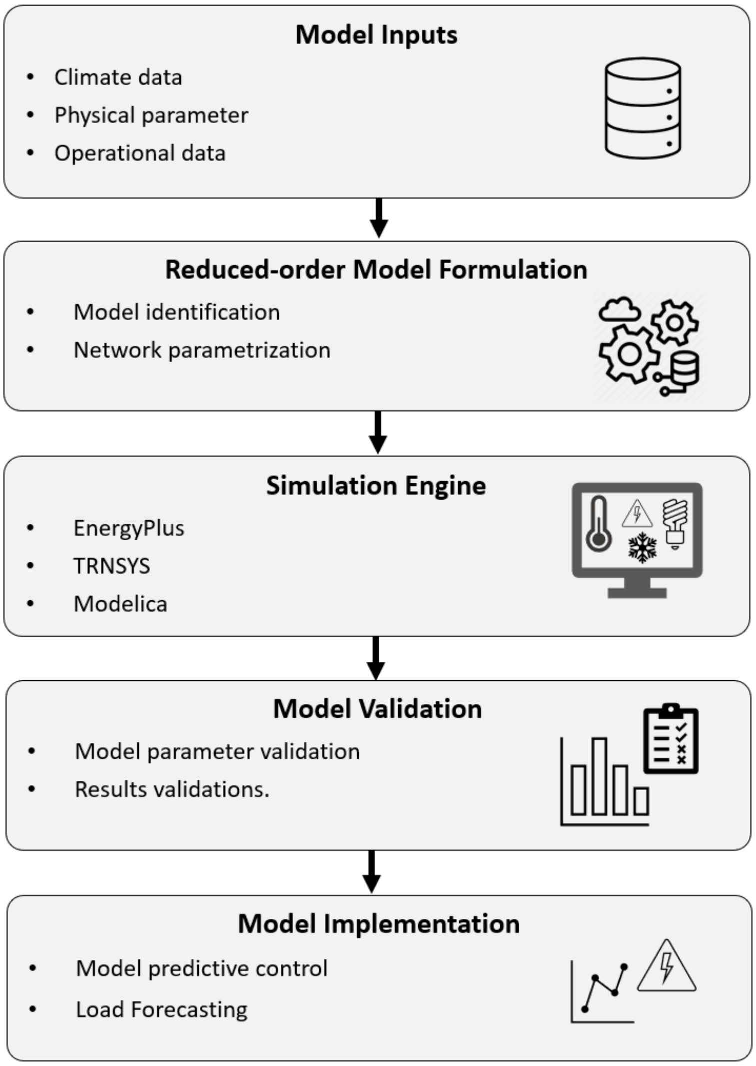

Figure 17: Generalized methodology for to model urban building stock using reduced-order bottom-up building energy modeling.

The network parameterization procedure (or model calibration) involves the estimation of reduced-order model parameters. A vast majority of urban reduced-order modeling studies use simulated data to estimate the model parameters of building clusters [135, 134]. While these studies provide reliable estimates, the uncertainties in these parameters are not usually accounted for, especially when considering the actual building operation [133]. Commonly used statistical estimation procedures include the maximum likelihood technique, Monte Carlo sampling, genetic algorithm, and interior-point algorithm [135, 131, 134].

Urban scale reduced-order models significantly reduce the simulation time when considering building clusters [124]. To examine the effectiveness of the models, studies usually implement two validation criteria: model parameter validation and results validations [136]. This procedure provides a means to check whether the model satisfies the assumptions and identifies reliable parameter estimates from a physical perspective. Berthou et al. implemented a cross-validation criterion, which simulates the building performance on a subset of dataset not used for identification [137]. Furthermore, when validating the model, special attention is given to calculating the confidence intervals of the parameters. It is crucial to ensure that the confidence intervals are small to retain the model's physical validity.

Reduced-order modeling approaches are considered a robust framework for creating low-order models for analysis and control of monitored buildings. Stakeholders could eventually use these models to test various demand management scenarios or generate load profiles of existing building clusters.

3.4.3. Reduced-order Quantitative Analysis

The bottom-up reduced-order model quantitative analysis uses the advanced search function of the Web of Knowledge search engine using Eq.1 and Table.2. The keywords search is based on the paper's initial content, such as title, abstract, and keywords used in a paper. The main keywords include reduced-order, lumped parameter, R–C, grey-box, and gray-box. A total of more than 72 articles are found that contain these keywords. The research publication trend between 2008-2020 indicates that the number of published papers have experienced a steep increase after 2014 (Fig. 18). This is indicative of the increasing implementation of different reduced-order models in the building simulation domain. The main applications of reduced-order approaches include building energy prediction (32%) and energy analysis (25%). Classification and mapping are at the lower end of reduced-order modeling applications (Fig. 18). Top publication sources on the reduced-order approach topics are Energy and Buildings and Applied Energy journals: 81% of the papers were published as articles, and 17% papers were published as conference proceedings (Fig. 19).

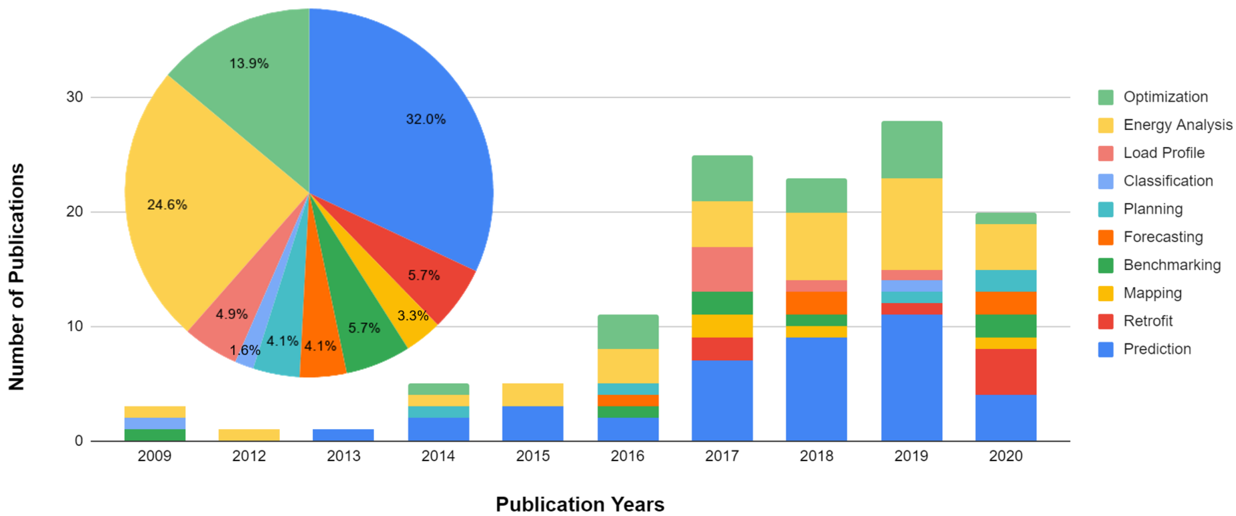

Figure 18: Reduced-order urban building energy modeling publication trend with intended area of implementation from the World of Science academic search platform.

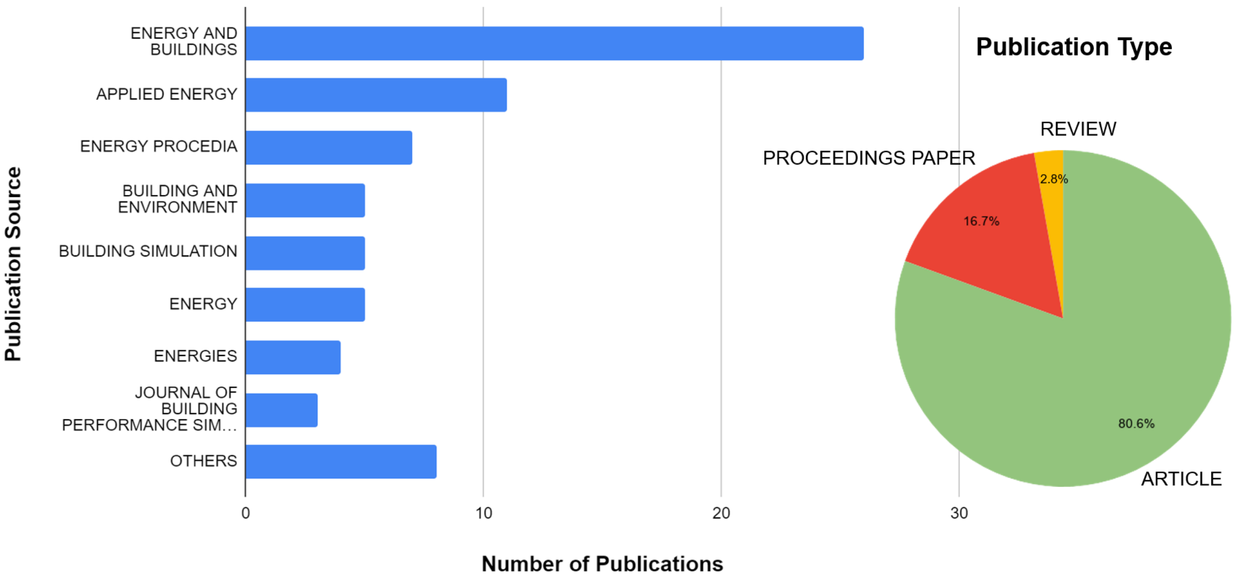

Figure 19: Reduced-order urban building energy modeling publication distribution according to publication source and type from World of Science academic search platform.

3.4.4. Reduced-order Qualitative Analysis

The qualitative analysis of reduced-order models determines the strengths and weaknesses associated with these models' past implementation. Furthermore, this analysis identifies the threats and challenges of reduced-order model suitability and scope, which would eventually evaluate various opportunities to enhance these models (Fig. 20).

The reduced-order modeling approach combines the best of both approaches (physics-based and data-driven). As these models contain the physical representation of the building while maintaining optimal computational efficiency. Furthermore, these models require fewer inputs for the model formulation process. Reduced-order models integrate the building's simplified physical model with a model identification process to identify the design parameters. Using simplified physical models reduces the requirement of training data sets and hence, the calculation time. The grey box approach uses trained simplified physical models, for instance, RC networks, to represent the building physics. Reduced-order energy models provide more flexibility (compared to black-box models) when optimizing the building operation when implementing different design scenarios. This includes determining the effects of variations in building geometry, right through to a change in material parameters. Moreover, a reduced-order model could formulate a data-driven model that could be used as a design tool itself.

Figure 20: Strengths, weaknesses, opportunities, and threats analysis of bottom-up reduced-order modeling approaches.

One of the significant weaknesses concerns the assumption that non-linear parameters or processes are often represented through linear relationships. Furthermore, areas of physical detail are often simplified by averaging the localized properties. While reduced-order building energy models have been widely implemented, these models' applicability has often been specific to particular applications and stakeholders. Another major drawback of reduced-order models is that the network order limits these models' scalability. The network order defines the level of complexity incorporated in the model. Furthermore, reduced-order methods often require extensive calibration using measured data especially when considering uncertainties in model parameters

The reduced-order modeling approach enables the rapid and accurate creation of building energy models at various modeling levels. These models offer opportunities to assess the pre and post-consumption patterns of the existing building stock.

One significant threat associated with reduced-order models occurs when addressing the overfitting issue of model parameters. A good fit of model predictions to the measured data are often considered for validating the model. However, it should be noted that if a model is flexible enough to fit several design scenarios, the identification of a good fit becomes less meaningful. With enhanced building dynamics, the associated reduced-order model needs to be compensated with additional complexity at the same level. Reduced-order models also require additional analysis to identify and optimize the trade-offs between model complexity and the desired accuracy. As measured building energy consumption is often unavailable at a large scale, validation of the simulated consumption poses a significant threat.

4. Discussion

The urban building energy modeling domain has experienced tremendous growth over the past few decades and offers numerous innovative opportunities. Stakeholders and researchers currently face various challenges while implementing energy modeling at this scale. These significant challenges include data availability, data inconsistencies, scalability, integration, detailed geo-spatial analysis, privacy issues, and computational resources.

- **Data Availability**

  Data availability for building stock modeling remains a key challenge as existing building data, especially building physical properties, are often unavailable for an entire city or a district. Similarly, the prediction accuracy of algorithms used in the data-driven approaches relies heavily on availability of the high-quality building input data [18]. In the past few years, the Internet of Things (IoT) and other advanced technologies help gather data from different resources [138]. Most developed countries' cities already have a significant amount of energy-related information available, but not all are easily accessible to the public and researchers [139]. Therefore, there should be national and global strategies to overcome these data availability issues in the face of future research needs.

- **Data Inconsistencies**

  The available data for urban energy modeling, such as building stock, census, land use, and building footprints, have different sources and formats that cause data inconsistencies. Therefore, due to inconsistencies and lack of standard urban scale data, available data are a significant ongoing barrier to implementing accurate energy modeling [140].

- **Data Quality**

  Urban building energy modeling requires building-related data, which is usually acquired through surveys [15]. Data contains errors, inaccuracies, redundancies, and anomalies. For instance, building stock data are collected and assembled manually [139]. Although every country measures data quality and mandates to follow defined standards, the process is prone to human error [141]. There are lots of duplication or extra information that are not crucial for building energy performance calculation.

- **Scalability**

  Stakeholders face scalability issues for detailed level analysis because of the lack of scalable building energy mapping approaches. Therefore a gap persists between building energy modeling and traditional planning methods [88]. Multi-scale modeling requires building, neighborhood, district to city-scale to quantify and analyze the influence of the building's characteristics on the building's energy performance. However, the biggest challenge is the complex nature of the multi-scale modeling that requires access to building-related data and computational resources.

- **Integration**

  Buildings play an essential role in determining energy demand calculations. However, buildings are not unique contributors; there are other urban environment elements, such as urban climate, energy generators, mobility, and socio-economic factors (income and number of people) [18, 20]. Furthermore, developing large scale models requires data from different sources such as building stock databases, geographical databases, surveys, and national census databases. Therefore, integrating different climate models, energy generators, and mobility would be a significant challenge [20]. Similarly, data integration of various sparse, inconsistent, and heterogeneous databases for urban scale energy modeling would require a better future research solution.

- **Geo-spatial Analysis**

  Stakeholders (urban planners, energy policymakers, and local authorities) use geospatial analysis for large scale planning. GIS techniques commonly use energy modeling, especially for bottom-up approaches, to increase the utility of building energy modeling results from the individual building level to urban, regional, or national levels [22]. However, the available traditional geospatial data face high-resolution limitations and provide incomplete or restricted coverage. Similarly, most of the survey data used for modeling are not geocoded for GIS mapping [139]. Furthermore, to help stakeholders for decision-making on many modeling results requires a 3D GIS mapping integrated visualization workflow [15].

- **Privacy issues**

  Data privacy, protection, and security are one of the most significant issues when implementing urban energy modeling; this especially due to human occupancy factors or other sensitive building data [15, 141]. Detailed building properties and actual energy use data are often required to estimate the accurate and granular level modeling results.

- **BIM Integration**

  Future work would benefit from highlighting links to BIM (Building Information Modeling). Data constraints refer to the lack of detailed building properties. The use of BIM furnishes these gaps (detailed inventories of materials, insulation, orientation are growingly available) provided that it was mandatory for the information to be made available by building owners. The BIM sector is experiencing tremendous growth these days. While some of the identified data limitations could be resolved in the coming years, there would be merits in acknowledging the integration of BIM as a future line of research work in building energy modeling and its application within industry practices.

- **Computation**

  Modeling all urban buildings requires significant computational resources. For example, one million buildings' energy performance simulation can be an exascale computing problem that requires next-generation supercomputers [15]. Parallel and cloud computing are the best options for accelerating the more sophisticated urban energy simulation procedure [20]. Similarly, urban energy modeling requires a massive amount of data, especially for a data-driven approach. Implementing processes such as data pre-processing, outlier deduction, feature extraction, and training models requires significant computational power. Integration with cloud-based or big data approaches using Google Cloud Platform or Amazon Web Services services adds substantial value to future research. Big data and artificial intelligence integration can also optimize complex geo-spatial analysis in future urban planning [142].

5. Conclusions and future research

Urban building energy modeling offers a robust framework that can be used for energy planning, retrofits, and city development to achieve optimal urban stock building performance. Over the last decade, a wide variety of research has been conducted in the urban building energy modeling domain. However, existing review papers focus on single approaches and fail to overview this domain's research trends. This paper fills this gap and attempts to provide a comprehensive literature review of urban building energy modeling. The reviews of state-of-the-art studies indicate that these models provide significant opportunities to capture or study complex urban system variations for enhancing system accuracy.

Top-down approaches are widely used at the national scale for long-term projections of energy demand. On the other hand, bottom-up techniques focus on a detailed, granular level building energy demand. Furthermore, a significant amount of research in this domain implement bottom-up physics-based modeling. Advanced data-driven approaches using artificial intelligence algorithms or hybrid models can be considered promising techniques for modeling the existing or new urban building stock.

Bottom-up data-driven building energy modeling approaches have been attracting significant research attention for future applications. Different tools and methodologies exist in the literature, which is formulated for different purposes, scopes, and scenarios using case-specific datasets and features. The data-driven models have their pros and cons and perform differently in different situations. Therefore, an application-specific model is needed to achieve the desired research goal. Future research should focus on developing a generalized solution in these fields that works with different data and scenarios.

The bottom-up physics-based modeling approach is most widely used in the building energy modeling field. Numerous research articles have been published in this area, and the domain has seen tremendous growth in the simulation of future technologies such as retrofit and renewable studies. However, due to the inherent uncertainties associated with an urban-scale energy analysis of buildings, the integration of spatio-temporal human activity patterns and socio-technical factors would improve modeling results.

Although the bottom-up reduced-order modeling approach is not quite popular in urban energy modeling domain, the implementation is gaining momentum as this approach combines the advantages of both approaches (physics-based and data-driven). For instance, the use of a physical structure enhances the interpretability of the problem. Moreover, building characteristics can be determined by optimization techniques such as genetic algorithms and hence, detailed building data are not often required.

Nowadays, big data and cloud-based computing are being widely used in public health and safety, social tourism, and real estate sectors. Limited work has been done to integrate cloud-based computing with the urban energy modeling domain. Future work could implement these advanced techniques for data acquisition, preprocessing, and simulation of large-scale energy modeling data, especially when implementing data-driven and physics-based bottom-up approaches.

6. Acknowledgements

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under the SFI Strategic Partnership Programme Grant number SFI/15/SPP/E3125. The opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the SFI. The authors would also like to acknowledge IBPSA Project 1, an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA).

References

[1] United Nation, `https://www.un.org/`, [Online; accessed 02-Mar-2020].

[2] C40 Cities, `https://www.c40.org`, [Online; accessed 02-Mar-2020].

[3] M. Santamouris, Energy and climate in the urban built environment, Routledge, 2013.

[4] Energy for Europe by European Commission, `https://ec.europa.eu/energy/en`, [Online; accessed 02-Mar-2020].

[5] Building and Climate Change by Environmental and Energy Study Institute, `http://www.eesi.org/`, [Online; accessed 02-Mar-2020].

[6] EU, Directive (EU) 2018/844 of the European Parliament and of the Council of 30 may 2018 amending Directive 2010/31/eu on the energy performance of buildings and Directive 2012/27/eu on energy efficiency, Official Journal of the European Union 61.

[7] H. Lund, Renewable energy strategies for sustainable development, Energy 32 (6) (2007) 912–919.

[8] C. F. Reinhart, C. C. Davila, Urban building energy modeling–a review of a nascent field, Building and Environment 97 (2016) 196–202.

[9] T. Hong, M. A. Piette, Y. Chen, S. H. Lee, S. C. Taylor-Lange, R. Zhang, K. Sun, P. Price, Commercial building energy saver: an energy retrofit analysis toolkit, Applied Energy 159 (2015) 298–309.

[10] S. T. Moghadam, C. Delmastro, S. P. Corgnati, P. Lombardi, Urban energy planning procedure for sustainable development in the built environment: A review of available spatial approaches, Journal of Cleaner Production 165 (2017) 811–827.
[11] L. Suganthi, A. A. Samuel, Energy models for demand forecasting—a review, Renewable and sustainable energy reviews 16 (2) (2012) 1223–1240.
[12] S. Jebaraj, S. Iniyan, A review of energy models, Renewable and sustainable energy reviews 10 (4) (2006) 281–311.
[13] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: A review, Renewable and Sustainable Energy Reviews 23 (2013) 272–288.
[14] X. Li, J. Wen, Review of building energy modeling for control and operation, Renewable and Sustainable Energy Reviews 37 (2014) 517–537.
[15] T. Hong, Y. Chen, X. Luo, N. Luo, S. H. Lee, Ten questions on urban building energy modeling, Building and Environment 168 (2020) 106508.
[16] U. Ali, M. H. Shamsi, C. Hoare, E. Mangina, J. O'Donnell, A data-driven approach for multi-scale building archetypes development, Energy and Buildings 202 (2019) 109364.
[17] J. Keirstead, M. Jennings, A. Sivakumar, A review of urban energy system models: Approaches, challenges and opportunities, Renewable and Sustainable Energy Reviews 16 (6) (2012) 3847–3866.
[18] N. Abbasabadi, M. Ashayeri, Urban energy use modeling methods and tools; a review and an outlook for future tools, Building and Environment (2019) 106270.
[19] M. Ferrando, F. Causone, T. Hong, Y. Chen, Urban building energy modeling (ubem) tools: A state-of-the-art review of bottom-up physics-based approaches, Sustainable Cities and Society (2020) 102408.
[20] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, J. Widén, Urban building energy modeling: State of the art and future prospects, Renewable and Sustainable Energy Reviews 128 (2020) 109902.
[21] T. Hong, S. Chou, T. Bong, Building simulation: an overview of developments and information sources, Building and environment 35 (4) (2000) 347–361.
[22] W. Li, Y. Zhou, K. Cetin, J. Eom, Y. Wang, G. Chen, X. Zhang, Modeling urban building energy use: A review of modeling approaches and procedures, Energy 141 (2017) 2445–2457.
[23] A. Sola, C. Corchero, J. Salom, M. Sanmarti, Multi-domain urban-scale energy modelling tools: A review, Sustainable Cities and Society 54 (2020) 101872.
[24] S. Ferrari, F. Zagarella, P. Caputo, M. Bonomolo, Assessment of tools for urban energy planning, Energy 176 (2019) 544–551.
[25] N. J. Van Eck, L. Waltman, Vosviewer manual, Leiden: Univeristeit Leiden 1 (1) (2013) 1–53.
[26] D. B. Crawley, L. K. Lawrie, C. O. Pedersen, F. C. Winkelmann, Energy plus: energy simulation program, ASHRAE journal 42 (4) (2000) 49–56.
[27] L. G. Swan, V. I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and sustainable energy reviews 13 (8) (2009) 1819–1835.
[28] H.-x. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews 16 (6) (2012) 3586–3592.
[29] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Building and environment 45 (7) (2010) 1683–1697.
[30] T. Hong, Z. Wang, X. Luo, W. Zhang, State-of-the-art on research and applications of machine learning in the building life cycle, Energy and Buildings 212 (2020) 109831.
[31] P. Bonato, R. Fedrizzi, M. D'Antoni, M. Meir, A. Kostro, A. Hafner, A. V. Souto, B. Arregi, B. Copertaro, B. Wilkinson, et al., State-of-the-art and swot analysis of building integrated solar envelope systems: deliverables a. 1 and a. 2.
[32] D. Meha, T. Novosel, N. Duic, Bottom-up and top-down heat demand mapping methods for small municipalities, case gllogoc, Energy (2020) 117429.
[33] A. Summerfield, R. Lowe, T. Oreszczyn, Two models for benchmarking uk domestic delivered energy,

Building Research & Information 38 (1) (2010) 12–24.
[34] A. Shkurti, Energy consumption modeling in the western balkan countries using a top-down approach, Academic Journal of Interdisciplinary Studies 7 (3) (2018) 35.
[35] A. A. Adam, A. Badea, Top-down model for the calculation of energy savings, in: 2017 International Conference on ENERGY and ENVIRONMENT (CIEM), IEEE, 2017, pp. 211–215.
[36] H. Lim, Z. J. Zhai, Review on stochastic modeling methods for building stock energy prediction, in: Building Simulation, Vol. 10, Springer, 2017, pp. 607–624.
[37] M. Lavagna, C. Baldassarri, A. Campioli, S. Giorgi, A. Dalla Valle, V. Castellani, S. Sala, Benchmarks for environmental impact of housing in europe: Definition of archetypes and lca of the residential building stock, Building and Environment 145 (2018) 260–275.
[38] D. Johnston, A physically-based energy and carbon dioxide emission model of the uk housing stock, Ph.D. thesis, Leeds Metropolitan University (2003).
[39] T. Hong, Y. Chen, S. H. Lee, M. A. Piette, Citybes: A web-based platform to support city-scale building energy efficiency, Urban Computing 14.
[40] C. Reinhart, T. Dogan, J. A. Jakubiec, T. Rakha, A. Sang, Umi-an urban simulation environment for building energy use, daylighting and walkability, in: 13th Conference of International Building Performance Simulation Association, Chambery, France, 2013.
[41] T. Vermeulen, J. H. Kämpf, B. Beckers, Urban form optimization for the energy performance of buildings using citysim, Tech. rep., EPFL Solar Energy and Building Physics Laboratory (LESO-PB) (2013).
[42] B. Polly, C. Kutscher, D. Macumber, M. Schott, S. Pless, B. Livingood, O. Van Geet, From zero energy buildings to zero energy districts, No. NREL/CP-5500-66292 (National Renewable Energy Laboratory, NREL, Golden, CO, USA, 2016).
[43] C. C. Davila, C. F. Reinhart, J. L. Bemis, Modeling boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, Energy 117 (2016) 237–250.
[44] M. Ferrando, F. Causone, An overview of urban building energy modelling (ubem) tools, in: Building Simulation 2019, Vol. 16, 2020, pp. 3452–3459.
[45] P. Nageler, G. Zahrer, R. Heimrath, T. Mach, F. Mauthner, I. Leusbrock, H. Schranzhofer, C. Hochenauer, Novel validated method for gis based automated dynamic urban building energy simulations, Energy 139 (2017) 142–154.
[46] S. J. Quan, Q. Li, G. Augenbroe, J. Brown, P. P.-J. Yang, Urban data and building energy modeling: A gis-based urban building energy modeling system using the urban-epc engine, in: Planning Support Systems and Smart Cities, Springer, 2015, pp. 447–469.
[47] L. Madrazo, A. Sicilia, G. Gamboa, Semanco: Semantic tools for carbon reduction in urban planning, in: Proceedings of the 9th European Conference on Product and Process Modelling, 2012.
[48] Y. Chen, T. Hong, M. A. Piette, Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis, Applied Energy 205 (2017) 323–335.
[49] E. Walter, J. H. Kämpf, A verification of citysim results using the bestest and monitored consumption values, in: Proceedings of the 2nd Building Simulation Applications conference, no. CONF, Bozen-Bolzano University Press, 2015, pp. 215–222.
[50] U. Ali, M. H. Shamsi, C. Hoare, J. O'Donnell, Gis-based residential building energy modeling at district scale, in: BSO 2018: 4th IBPSA-England Conference on Building Simulation and Optimization, Cambridge, United Kingdom, 11-12 September 2018, International Building Performance Simulation Association, 2018.
[51] G. Agugiaro, J. Benner, P. Cipriano, R. Nouvel, The energy application domain extension for citygml: enhancing interoperability for urban energy simulations, Open Geospatial Data, Software and Standards 3 (1) (2018) 2.
[52] G. Gröger, L. Plümer, Citygml–interoperable semantic 3d city models, ISPRS Journal of Photogrammetry and Remote Sensing 71 (2012) 12–33.
[53] R. Nouvel, J.-M. Bahu, R. Kaden, J. Kaempf, P. Cipriano, M. Lauster, K.-H. Häfele, E. Munoz,

O. Tournaire, E. Casper, Development of the citygml application domain extension energy for urban energy simulation, in: Building Simulation 2015-14th Conference of the International Building Performance Simulation Association, 2015, pp. 559–564.

[54] É. Mata, A. S. Kalagasidis, F. Johnsson, Building-stock aggregation through archetype buildings: France, germany, spain and the uk, Building and Environment 81 (2014) 270–282.

[55] C. Cerezo, J. Sokol, C. Reinhart, A. Al-Mumin, Three methods for characterizing building archetypes in urban energy simulation. a case study in kuwait city, in: Proceedings of BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India, 2015.

[56] Z. Li, B. Lin, S. Zheng, Y. Liu, Z. Wang, J. Dai, A review of operational energy consumption calculation method for urban buildings, Building Simulation.
URL `https://doi.org/10.1007/s12273-020-0619-0`

[57] I. ICE, Ida indoor climate and energy (2018).

[58] J. Hirsch, Doe2. 2 whole building energy simulation analysis using equest version 3.65 (2018).

[59] I. E. Solutions, Ies ve software (2020).

[60] S. A. Klein, Trnsys-a transient system simulation program., University of Wisconsin-Madison, Engineering Experiment Station Report (1988) 38–12.

[61] A. Galante, M. Torri, et al., A methodology for the energy performance classification of residential building stock on an urban scale, Energy and Buildings 48 (2012) 211–219.

[62] J. A. Fonseca, A. Schlueter, Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts, Applied Energy 142 (2015) 247–265.

[63] R. Judkoff, D. Wortman, B. O'Doherty, J. Burch, A methodology for validating building energy analysis simulations, Tech. Rep. April (2008).
URL `http://www.osti.gov/bridgehttp://www.stanford.edu/group/narratives/classes/08-09/CEE215/ReferenceLibrary/BIMandBuildingSimulationResearch/AMethodologyforValidatingBuildingEnergyAnalysisSimulations.pdf`

[64] ASHRAE, Standard 140-2017: Standard method of test for the evaluation of building energy analysis computer programs. (2017).

[65] P. Wate, V. Coors, 3d data models for urban energy simulation, Energy Procedia 78 (2015) 3372–3377.

[66] J. F. Rosser, G. Long, S. Zakhary, D. S. Boyd, Y. Mao, D. Robinson, Modelling urban housing stocks for building energy simulation using citygml energyade, ISPRS International Journal of Geo-Information 8 (4) (2019) 163.

[67] Z. Zheng, J. Chen, X. Luo, Parallel computational building-chain model for rapid urban-scale energy simulation, Energy and Buildings 201 (2019) 37–52.

[68] N. Lauzet, A. Rodler, M. Musy, M.-H. Azam, S. Guernouti, D. Mauree, T. Colinart, How building energy models take the local climate into account in an urban context–a review, Renewable and Sustainable Energy Reviews 116 (2019) 109390.

[69] G. Happle, J. A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy and Buildings 174 (2018) 276–292.

[70] S. Hu, D. Yan, E. Azar, F. Guo, A systematic review of occupant behavior in building energy policy, Building and Environment (2020) 106807.

[71] G. Tardioli, R. Kerrigan, M. Oates, O. James, D. Finn, Data driven approaches for prediction of building energy consumption at urban level, Energy Procedia 78 (2015) 3378–3383.

[72] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, P. Chatellier, Modeling and forecasting building energy consumption: A review of data-driven techniques, Sustainable Cities and Society 48 (2019) 101533.

[73] U. Ali, M. H. Shamsi, C. Hoare, F. Alshehri, E. Mangina, J. O'Donnell, Application of intelligent algorithms for residential building energy performance rating prediction, in: 16th IBPSA International Conference and Exhibition Building Simulation 2019, International Building Performance Simulation Association, 2019.

[74] G. Tardioli, R. Kerrigan, M. Oates, J. O'Donnell, D. P. Finn, Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and

predictive modelling approach, Building and Environment 140 (2018) 90–106.
[75] Y. Zhao, C. Zhang, Y. Zhang, Z. Wang, J. Li, A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis, Energy and Built Environment 1 (2) (2020) 149–164.
[76] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, X. Zhao, A review of data-driven approaches for prediction and classification of building energy consumption, Renewable and Sustainable Energy Reviews 82 (2018) 1027–1047.
[77] A. Nutkiewicz, Z. Yang, R. K. Jain, Data-driven urban energy simulation (due-s): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow, Applied energy 225 (2018) 1176–1189.
[78] T. Ahmad, H. Chen, Y. Guo, J. Wang, A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review, Energy and Buildings 165 (2018) 301–320.
[79] A. Mastrucci, O. Baume, F. Stazi, U. Leopold, Estimating energy savings for the residential building stock of an entire city: A gis-based statistical downscaling approach applied to rotterdam, Energy and Buildings 75 (2014) 358–367.
[80] L. Pedersen, J. Stang, R. Ulseth, Load prediction method for heat and electricity demand in buildings for the purpose of planning for mixed energy distribution systems, Energy and Buildings 40 (7) (2008) 1124–1134.
[81] K. T. Williams, J. D. Gomez, Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach, Energy and Buildings 128 (2016) 1–11.
[82] A. Rahman, V. Srikumar, A. D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Applied energy 212 (2018) 372–385.
[83] C. Robinson, B. Dilkina, J. Hubbs, W. Zhang, S. Guhathakurta, M. A. Brown, R. M. Pendyala, Machine learning approaches for estimating commercial building energy consumption, Applied energy 208 (2017) 889–904.
[84] N. Abbasabadi, M. Ashayeri, R. Azari, B. Stephens, M. Heidarinejad, An integrated data-driven framework for urban energy use modeling (ueum), Applied Energy 253 (2019) 113550.
[85] Z. Yang, J. Roth, R. K. Jain, Due-b: Data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis, Energy and Buildings 163 (2018) 58–69.
[86] C. E. Kontokosta, C. Tull, A data-driven predictive model of city-scale energy use in buildings, Applied energy 197 (2017) 303–317.
[87] S. Papadopoulos, C. E. Kontokosta, Grading buildings on energy performance using city benchmarking data, Applied Energy 233 (2019) 244–253.
[88] U. Ali, M. H. Shamsi, M. Bohacek, C. Hoare, K. Purcell, E. Mangina, J. O'Donnell, A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings, Applied Energy 267 (2020) 114861.
[89] C. Analytics, The ultimate technical guide to urbanfootprint.
[90] F. Zhao, I. J. Martinez-Moyano, G. Augenbroe, Agent-based modeling of commercial building stocks for policy support, in: Building Simulation, 2011, pp. 2385–92.
[91] M. Cascadden, The energy proforma.
[92] Y. Wang, T. Wu, H. Li, M. Skitmore, B. Su, A statistics-based method to quantify residential energy consumption and stock at the city level in china: The case of the guangdong-hong kong-macao greater bay area cities, Journal of Cleaner Production 251 (2020) 119637.
[93] Building Performance Database by Department of Energy, `https://bpd.lbl.gov/`, [Online; accessed 02-Mar-2020].
[94] Energy Performance Certificates across europe from design to implementation, BPIE, `http://www.buildup.eu/en/practices/publications/energy-performance-certificates-across-europe-design-implementation`, [Online; accessed 02-Mar-2020] (2010).

[95] U. Ali, C. Buccella, C. Cecati, Households electricity consumption analysis with data mining techniques, in: Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE, IEEE, 2016, pp. 3966–3971.

[96] C. Xu, H. Chen, Abnormal energy consumption detection for gshp system based on ensemble deep learning and statistical modeling method, International Journal of Refrigeration.

[97] J. Y. Park, E. Wilson, A. Parker, Z. Nagy, The good, the bad, and the ugly: Data-driven load profile discord identification in a large building portfolio, Energy and Buildings 215 (2020) 109892.

[98] A. Capozzoli, F. Lauro, I. Khan, Fault detection analysis using data mining techniques for a cluster of smart office buildings, Expert Systems with Applications 42 (9) (2015) 4324–4338.

[99] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, R. J. Millar, Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems, IEEE Access 7 (2019) 91463–91475.

[100] X. Gao, A. Malkawi, A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm, Energy and Buildings 84 (2014) 607–616.

[101] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, Applied Energy 195 (2017) 222–233.

[102] C. Feng, M. Cui, B.-M. Hodge, J. Zhang, A data-driven multi-model methodology with deep feature selection for short-term wind forecasting, Applied Energy 190 (2017) 1245–1257.

[103] A. A. Famuyibo, A. Duffy, P. Strachan, Developing archetypes for domestic dwellings—an irish case study, Energy and Buildings 50 (2012) 150–157.

[104] J. Egan, D. Finn, P. H. D. Soares, V. A. R. Baumann, R. Aghamolaei, P. Beagon, O. Neu, F. Pallonetto, J. O'Donnell, Definition of a useful minimal-set of accurately-specified input data for building energy performance simulation, Energy and Buildings 165 (2018) 172–183.

[105] D.-S. Kapetanakis, E. Mangina, E. H. Ridouane, K. Kouramas, D. Finn, Selection of input variables for a thermal load prediction model, Energy Procedia 78 (2015) 3001–3006.

[106] S. Seyedzadeh, F. P. Rahimian, S. Oliver, I. Glesk, B. Kumar, Data driven model improved by multi-objective optimisation for prediction of building energy loads, Automation in Construction 116 (2020) 103188.

[107] R. R. Picard, K. N. Berk, Data splitting, The American Statistician 44 (2) (1990) 140–147.

[108] Y. Sun, F. Haghighat, B. C. Fung, A review of the-state-of-the-art in data-driven approaches for building energy prediction, Energy and Buildings (2020) 110022.

[109] V. Marinakis, Big data for energy management and energy-efficient buildings, Energies 13 (7) (2020) 1555.

[110] S. Fathi, R. Srinivasan, A. Fenner, S. Fathi, Machine learning applications in urban building energy performance forecasting: A systematic review, Renewable and Sustainable Energy Reviews 133 (2020) 110287.

[111] U. Ali, M. H. Shamsi, M. Bohacek, K. Purcell, C. Hoare, E. Mangina, J. O'Donnell, A data-driven approach for multi-scale gis-based building energy modeling for analysis, planning and support decision making, Applied Energy 279 (2020) 115834.

[112] H. Deng, D. Fannon, M. J. Eckelman, Predictive modeling for us commercial building energy use: A comparison of existing statistical and machine learning algorithms using cbecs microdata, Energy and Buildings 163 (2018) 34–43.

[113] M. Beccali, G. Ciulla, V. L. Brano, A. Galatioto, M. Bonomolo, Artificial neural network decision support tool for assessment of the energy performance and the refurbishment actions for the non-residential building stock in southern italy, Energy 137 (2017) 1201–1218.

[114] U. Ali, M. H. Shamsi, F. Alshehri, E. Mangina, J. O'Donnell, Comparative analysis of machine learning algorithms for building archetypes development inurban building energy modeling, in: Building Performance Modeling Conference and SimBuild, 2018.

[115] I. De Jaeger, G. Reynders, C. Callebaut, D. Saelens, A building clustering approach for urban energy simulations, Energy and Buildings 208 (2020) 109671.

[116] G. Tardioli, A. Narayan, R. Kerrigan, M. Oates, J. O'Donnell, D. P. Finn, A methodology for cali-

bration of building energy models at district scale using clustering and surrogate techniques, Energy and Buildings 226 (2020) 110309.

[117] S. Papadopoulos, B. Bonczak, C. E. Kontokosta, Pattern recognition in building energy performance over time using energy benchmarking data, Applied Energy 221 (2018) 576–586.

[118] K. Amasyali, N. M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renewable and Sustainable Energy Reviews 81 (2018) 1192–1205.

[119] S. H. Lee, F. Zhao, G. Augenbroe, The use of normative energy calculation beyond building performance rating, Journal of Building performance simulation 6 (4) (2013) 282–292.

[120] A. Handbook, Fundamentals, ashrae–american society of heating, Ventilating and Air-Conditioning Engineers.

[121] R. Nouvel, K.-H. Brassel, M. Bruse, E. Duminil, V. Coors, U. Eicker, D. ROBINSON, Simstadt, a new workflow-driven urban energy simulation platform for citygml city models, in: Proceedings of International Conference CISBAT 2015 Future Buildings and Districts Sustainability from Nano to Urban Scale, no. CONF, LESO-PB, EPFL, 2015, pp. 889–894.

[122] J. Bergerson, R. T. Muehleisen, W. B. Rodda, J. A. Auld, L. B. Guzowski, J. Ozik, N. Collier, Designing future cities: Lakesim integrated design tool for assessing short and long term impacts of urban scale conceptual designs, ISOCARP Review 11.

[123] R. Baetens, R. De Coninck, F. Jorissen, D. Picard, L. Helsen, D. Saelens, Openideas-an open framework for integrated district energy simulations, in: Building simulation, 2015.

[124] P. Remmen, M. Lauster, M. Mans, M. Fuchs, T. Osterhage, D. Müller, Teaser: an open tool for urban energy modelling of building stocks, Journal of Building Performance Simulation 11 (1) (2018) 84–98.

[125] R. Kaden, T. H. Kolbe, City-wide total energy demand estimation of buildings using semantic 3d city models and statistical data, in: Proc. of the 8th International 3D GeoInfo Conference, 2013.

[126] J. A. Fonseca, T.-A. Nguyen, A. Schlueter, F. Marechal, City energy analyst (cea): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts, Energy and Buildings 113 (2016) 202–226.

[127] Q. Li, S. J. Quan, G. Augenbroe, P. P.-J. Yang, J. Brown, Building energy modelling at urban scale: integration of reduced order energy model with geographical information, in: Building Simulation, 2015.

[128] M. Wetter, C. van Treeck, Iea ebc annex 60: New generation computing tools for building and community energy systems, The Regents of the University of California and RWTH Aachen University.

[129] R. Baetens, R. De Coninck, J. Van Roy, B. Verbruggen, J. Driesen, L. Helsen, D. Saelens, Assessing electrical bottlenecks at feeder level for residential net zero-energy buildings by integrated system simulation, Applied Energy 96 (2012) 74–83.

[130] M. Wetter, W. Zuo, T. S. Nouidui, X. Pang, Modelica buildings library, Journal of Building Performance Simulation 7 (4) (2014) 253–270.

[131] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, Energy and Buildings 43 (7) (2011) 1511–1522. doi:10.1016/j.enbuild.2011.02.005.
URL https://www.sciencedirect.com/science/article/pii/S0378778811000491

[132] M. H. Shamsi, U. Ali, E. Mangina, J. O'Donnell, A framework for uncertainty quantification in building heat demand simulations using reduced-order grey-box energy models, Applied Energy 275 (2020) 115141. doi:10.1016/j.apenergy.2020.115141.

[133] G. Yang, Z. Li, G. Augenbroe, Development of prototypical buildings for urban scale building energy modeling: A reduced order energy model approach, Science and Technology for the Built Environment 24 (1) (2018) 33–42. doi:10.1080/23744731.2017.1328943.

[134] M. Heidarinejad, N. Mattise, M. Dahlhausen, K. Sharma, K. Benne, D. Macumber, L. Brackney, J. Srebric, Demonstration of reduced-order urban scale building energy models, Energy and Buildings 156 (2017) 17–28. doi:10.1016/j.enbuild.2017.08.086.

[135] M. H. Shamsi, U. Ali, F. Alshehri, J. O. Donnell, A Framework To Assess The Interoperability Of Commercial Buildings At A District Scale, in: Building Simulation and Optimization, Vol. 4, International Building Simulation Association England, 2018, pp. 389–395.

[136] M. Manfren, B. Nastasi, Parametric performance analysis and energy model calibration workflow integration - A scalable approach for buildings, Energies 13 (3) (2020) 621. doi:10.3390/en13030621.
URL https://www.mdpi.com/1996-1073/13/3/621

[137] T. Berthou, P. Stabat, R. Salvazet, D. Marchio, Development and validation of a gray box model to predict thermal behavior of occupied office buildings, Energy and Buildings 74 (2014) 91–100. doi:10.1016/j.enbuild.2014.01.038.
URL https://www.sciencedirect.com/science/article/pii/S0378778814000760

[138] M. M. Rathore, A. Ahmad, A. Paul, S. Rho, Urban planning and building smart cities based on the internet of things using big data analytics, Computer Networks 101 (2016) 63–80.

[139] O. Pasichnyi, J. Wallin, F. Levihn, H. Shahrokni, O. Kordas, Energy performance certificates—new opportunities for data-enabled urban energy policy instruments?, Energy policy 127 (2019) 486–499.

[140] Y. Chen, T. Hong, X. Luo, B. Hooper, Development of city buildings dataset for urban building energy modeling, Energy and Buildings 183 (2019) 252–265.

[141] Y. Li, S. Kubicki, A. Guerriero, Y. Rezgui, Review of building energy performance certification schemes towards future improvement, Renewable and Sustainable Energy Reviews 113 (2019) 109244.

[142] Y. Yamagata, P. P. Yang, S. Chang, M. B. Tobey, R. B. Binder, P. J. Fourie, P. Jittrapirom, T. Kobashi, T. Yoshida, J. Aleksejeva, Chapter 2 - urban systems and the role of big data, in: Y. Yamagata, P. P. Yang (Eds.), Urban Systems Design, Elsevier, 2020, pp. 23 – 58. doi:https://doi.org/10.1016/B978-0-12-816055-8.00002-6.