Energy & Buildings 257 (2022) 111809

Contents lists available at ScienceDirect

Energy & Buildings

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

Occupant-centric urban building energy modeling: Approaches, inputs, and data sources - A review

Sanam Dabirian [a,*], Karthik Panchabikesan [a], Ursula Eicker [a,b]

[a] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec H3G 1M8, Canada
[b] Canada Excellence Research Chair on Smart, Sustainable and Resilient Cities and Communities, Canada

ARTICLE INFO

Article history:
Received 13 September 2021
Revised 2 December 2021
Accepted 20 December 2021
Available online 24 December 2021

Keywords:
Urban building energy modeling
Occupancy modeling approaches
Occupant-related inputs
Urban data
Level of details
Co-simulation

ABSTRACT

Occupant-related inputs are significant parameters that influence energy simulation accuracy at both the building and urban levels. In most previous research works, fixed occupant schedules were used in urban-scale building energy modeling. The main reason is the lack of data availability to model the dynamic occupancy schedules. In recent years, urban data sets and modern estimation and detection techniques were introduced to increase the availability of occupant-related data sets. Yet, using these data to model the detailed occupancy at the urban scale is challenging and not much explored. Also, it is unclear how detailed the input regarding the occupancy should be. Addressing these research questions is the main objective of this study. This paper presents a comprehensive review of the occupant behavior (OB) modeling approaches, occupant-related input parameters with particular focus on the occupancy schedule, lighting, appliances use schedule, temperature set-point schedule, and domestic hot water usage for urban building energy modeling (UBEM). Strategies to consider the occupancy sub-models as co-simulation connecting to urban building energy simulation are discussed. Some potential datasets that could be used to derive occupant-related inputs at the urban scale are presented and highlighted. Further, the correlation between occupants' activity level, plug, and lighting loads is discussed in detail. Finally, the limitations and challenges of the occupant-related data connecting to building energy modeling are discussed, along with the research gaps and future directions of occupant-centric UBEM.

Contents

Abbreviations: ABM, Agent-based modeling; ARIMA, Autoregressive integrated moving average; ASHRAE, American Society of Heating, Refrigerating and Air-Conditioning Engineers; ATUS, American time use survey; BEMS, Building energy management system; BPS, Building performance simulation; CEA, City energy analyst; CESAR, Combined energy simulation and retrofitting; CityBES, City building energy saver; DNAS, Drivers, needs, actions, and systems; DOE, Department of energy; DYD, Donate your data; EnergyADE, Energy application domain extension; GIS, Geographical information system; HUES, Holistic urban energy simulation; HVAC, Heating, ventilation, air-conditioning; ISTAT, Italian National Institute of Statistics; k-NN, k-nearest neighbors; LBS, Location-based services; MuMo, Multiple Modules; NECB, National energy code of Canada for buildings; OB, Occupant behavior; obFMU, Occupant behavior functional mockup unit; obXML, Occupant behavior XML; PIR, Passive infra-red; RNN, Recurrent neural network; SHEU, Survey of household energy use; SIA, Swiss Society of Engineers and Architects; SOB, Stochastic occupant behavior; StROBe, Stochastic residential occupancy behavior; SVM, Support Vector Machine; TEASER, Tool for energy analysis and simulation for efficient retrofit; UBEM, Urban building energy modeling; UMI, Urban modeling interface; UML, Unified modeling language.

* Corresponding author.
E-mail address: sanam.dabirian@mail.concordia.ca (S. Dabirian).

https://doi.org/10.1016/j.enbuild.2021.111809
0378-7788/

1. Introduction

1.1. Background and motivation

More than half of the world's population lives in the cities and is responsible for 70% of the global energy consumption and environmental impacts [1]. Hence, energy planners and stakeholders are interested in sustainability, resilience, and enhancing the buildings' energy efficiency in the urban context [2]. In this regard, researchers and building engineers apply urban building energy modeling (UBEM) to estimate the buildings' energy demand at the urban scale. UBEM refers to physics-based computational modeling to simulate a group of buildings' performance considering the buildings' dynamic and inter-connection [3]. Although UBEM can be an integrated planning tool to estimate the building energy demand and assess the buildings' energy efficiency, a reliable, automated modeling approach is lacking to simulate the urban buildings' energy demand [4].

In general, detailed data and realistic modeling assumptions are required so that a UBEM could estimate the representative energy demand of buildings [5]. In specific, UBEM requires input related to geographical information system (GIS) [6], weather data [7], building characteristics including materials [8], year built [9], heating, ventilation, and air conditioning (HVAC) types, occupancy, and equipment usage schedules [10]. The most critical challenge in UBEM is providing appropriate inputs for several parameters and accordingly reducing the uncertainties in the estimated energy demand, which generally arises from the simplified/default assumptions made while performing the UBEM [11]. Similar to the weather and building-related parameters (building envelopes, energy system configurations, operation, and maintenance), occupant related input parameters also have significant influence over the energy consumption in buildings and are one of the important sources for the uncertainties in UBEM [3,12–16]. It is reported in the literature that the occupant-related uncertainty is relatively higher compared to inputs related to infiltration rate and building envelope materials [17], accounting for up to 30% of the estimated building energy demand variation at the urban scale [18]. Romero [19] mentioned that the main reasons for the uncertainties in UBEM are the limited knowledge about the input parameters (epistemic uncertainty) and the system components' variability (stochastic uncertainty).

Epistemic uncertainty has been slowly addressed in recent years by utilizing data collected from smart energy meters or building energy management systems (BEMS) [19]. Several researchers are interested in developing novel data mining/data-driven, stochastic models representing occupancy and its associated inputs in energy simulations, recommending occupant-centric controls in buildings, and providing user-specific energy-saving advisories. Though many studies on occupancy modeling are available in the literature, most of them focused on the buildings level, and detailed research on occupant-centric UBEM is scarce. Most of the UBEM tools focus on the effects of weather conditions, building physics, and construction features rather than the impact of occupants [20]. Therefore, OB modeling and considering distinct occupant patterns in UBEM can provide representative energy consumption data, which helps decision-makers provide more efficient measures to control the urban building systems. It is still unclear how to represent the variation associated with the occupancy (at the urban scale) concerning different building archetypes, periods of the day, week, month, and even seasons. In this context, this review aims to (1) consolidate research works available in the literature on occupant-centric urban building energy modeling, (2) different occupancy modeling approaches, (3) provide insights on occupant-related inputs required for UBEM, and (4) finally reviews the challenges in accommodating the variations in occupancy at the urban scale. Besides, some urban data sources available for understanding and extracting occupant-related parameters in buildings are highlighted.

1.2. Occupancy and occupant behavior modeling

Occupancy refers to occupant presence and absence schedule in the given space, whereas OB refers to occupant interactions with appliances, lighting, thermostats, and other energy equipment. Several studies on occupancy and OB modeling have advanced the building energy performance analysis and enriched the modeling inputs in the past decade. Annex 53 described the influence of occupant activity and behavior quantitatively on the residential and office buildings' energy use [21]. Annex 66 established a framework to simulate the OB model and integrate it with the building energy performance tools [15]. Following Annex 66, Annex 79 focused on occupant-centric building design and operation [22].

Several reviews have been conducted to summarize the work of research studies that majorly focused on understanding the impact of OB in energy modeling at building scale and reviewing the OB modeling approaches, implementation, and representation of OB models in BPS. Hong et al. [23] reviewed the implementation of the OB models in building performance simulation (BPS). They provided insight into BPS programs' OB implementation approaches, representing the OB model inputs and the approaches' pros and cons. Li et al. [24] conducted a review focused on three critical issues regarding the uncertainties arising from OB in BPS, selecting an appropriate OB model, and the requirements to improve OB models. Since OB influences building energy consumption, the driving factors of the energy-related OB should be identified. Zhang et al. [25] conducted a review on rethinking OB in the BPS. They categorized the occupant energy-related behavior into interaction with window, lighting, heating, and cooling behavior. Regarding the occupancy modeling approaches, Jia et al. [20] reviewed state-of-the-art occupancy modeling methodologies and compared them based on their advantages and disadvantages. They concluded that the agent-based modeling integrates appropriately with the building energy simulation tools, while stochastic-based and data mining techniques are used to track long-term OB patterns to estimate the energy demand. The above-mentioned studies investigated the occupancy at the building scale.

The proliferation of occupant-centric big data such as IoT, sensor-based, and mobility data has paved the way to model the occupant behavior at a neighborhood, district, or city scale. However, there is limited study regarding OB modeling for urban building design and operation. Hou et al. [26] proposed a novel framework to model the interaction between the building occupants and urban energy systems using Wi-Fi infrastructure to extract the occupancy profiles. The location-based services data such as Google Maps or Facebook is used to create context-specific, data-driven occupancy schedules in 13 different U.S. cities [27]. In another study, the occupancy profiles were derived from high-resolution mobile positioning data in 900 buildings in downtown San Antonio, Texas [28]. In [28], using the captured empirical occupancy profiles shows a significant reduction in thermal demand compared with the DOE occupancy profiles. With extracted urban scale building occupancy profiles from mobile positioning data, Barbour et al. [29] and Kang et al. [30] calculated the energy demand of different building types at the urban level. Parker et al. [31] developed a framework to create the occupancy schedules using personal location metadata to input dynamic building simulation that could be extended for different building types. Miller and Meggers [32] proposed a two-step framework to characterize the building model-based and pattern-based behavior using temporal feature extraction of 507 non-residential buildings.

In addition to developing the OB models based on the available urban data, some studies have investigated the impact of different occupant-related input parameters on energy management, urban building energy simulation, and energy system sizing. Xu et al. [33] proposed an energy-saving alignment strategy to achieve maximum energy efficiency, focusing on occupants' thermostat preferences and apartments' operative temperatures in a public housing project in New York City. Seasonal effects of the input parameters on urban building energy simulation (e.g., operating parameters and internal loads) have been investigated utilizing a sensitivity analysis in a mixed-used district in central Zurich, Switzerland [34]. Mosteiro-Romero and Schlueter [12] analyzed the effects of the variations of the occupants and climate in electrical and thermal demands caused by the sizing of the supply energy systems. They compared three different occupancy modeling approaches on the district planned systems, such as deterministic, stochastic, and population-based methods. Baetens and Saelens [35] quantified the occupant-related uncertainties arising from the stochastic nature of OB in residential building districts using the StROBe stochastic simulation tool. The review papers focused on OB models at the urban scale, and the uniqueness of the present review is highlighted in section 1.3.

1.3. Novelty of the review

Several review papers have been published at the urban scale in recent years on various aspects of UBEM, such as bottom-up physics-based UBEM tools, including required input, output, and workflows [36,37]. However, there exist only few reviews focused on the occupancy modeling for UBEM [38–40]. For example, Happle et al. [38] conducted a study focusing mainly on OB modeling approaches for the UBEM. They assessed the diversity of OB in different building types. Salim et al. [39] discussed various urban data sources that enable the researchers to model the human-building interactions and mobility behavior at the urban scale. They categorized the occupant-centric urban data into six categories: occupancy schedules, mobility data, building performance, operational data, environmental, and survey data. In another review, Dong et al. [40] investigated the gap between occupancy modeling and data sources in building science and other fields to identify the challenges, modeling requirements, and policy at the urban scale. The above-mentioned reviews were focused on the categorization of OB modeling in UBEM and discussed the advantages, disadvantages of each occupancy model. To the author's knowledge, these review papers have not discussed the required occupant-related input parameters for UBEMs in detail except occupancy presence. The current paper provides a consolidated overview and discussion on the choice of occupancy modeling in urban contexts, the correlations existing between the occupant-related input parameters (occupancy schedule, lighting, plug load schedule, setpoint temperature schedule), potential data sources that shall be used to extract occupant related inputs and co-simulation options available in the literature to represent OB in UBEM.

1.4. Review methodology

In this study, a literature review was conducted to consolidate the past research on occupancy modeling, detailing the occupant-related inputs for building energy modeling at the neighborhood, district, and city scale, challenges in assigning realistic occupancy in the models, and finally providing an overview of available data sources to model occupancy for UBEM purposes. In this context, a text mining analysis was performed to find the journal articles and conference proceedings since 2000. Before 2003, no study was found that investigated UBEM (Fig. 1). The keywords included neighborhood, district, urban, or city combining with building energy modeling or simulation, occupancy modeling, and occupant-related input parameters such as lighting, appliance usage, temperature setpoints, and domestic hot water.

The publications were screened considering two main criteria: (1) the energy simulation must be implemented on the urban scale; (2) it contains the occupancy modeling parameters and approaches. Due to the first keyword's tremendous search results, another search was done, adding the second keyword to limit the output. If the references passed both the above-mentioned screening criteria, they are included in this review as well. All the collected papers have been examined to define the UBEM primary application, occupancy sub-models, source of occupancy profiles, and used energy simulation engine (Fig. 2).

2. Urban building energy modeling (UBEM) - an overview

To model the building energy requirements and improve energy sustainability at the city scale, UBEM is a powerful tool that

Fig. 1. Reviewed papers publication year.

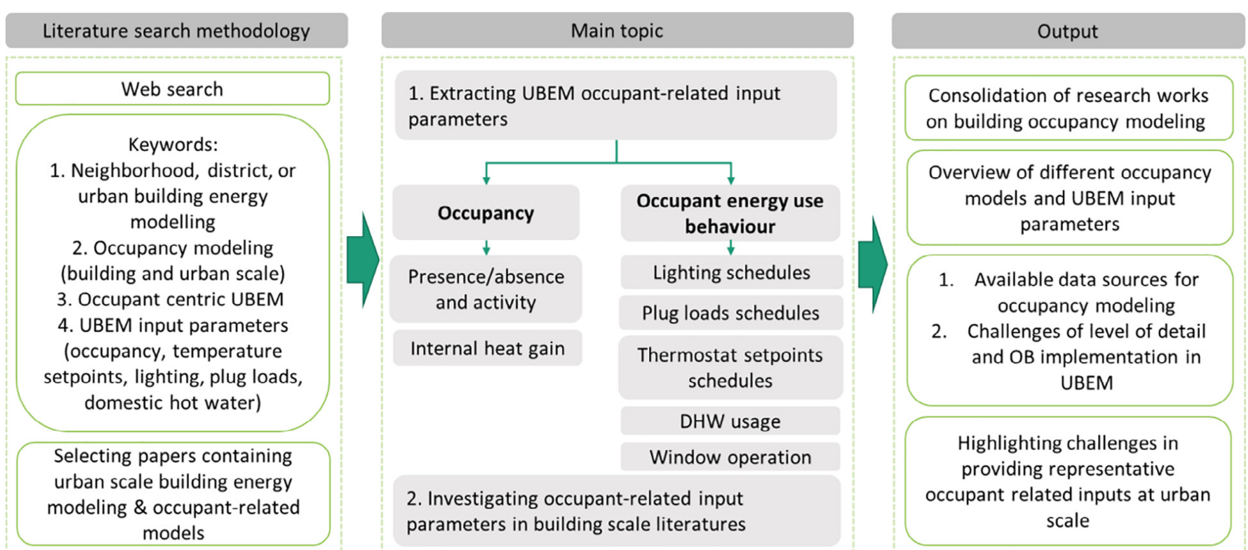

Fig. 2. Review methodology.

combines urban data with energy simulation engines. The major applications of UBEM are evaluating the system performance [41], identifying the potential for building retrofit [42], energy-driven planning, and urban decision-making [43]. UBEM approaches are classified as top-down and bottom-up models [44,45]. The top-down model as a data-driven approach predicts long-term energy consumption using macroeconomic data and statistical information at the urban and regional levels [41]. This approach does not deal with individual end-uses. The top-down model's major drivers include gross domestic product, population, household size, energy price, and climate [46].

On the other hand, the bottom-up model that is the most applied method to model urban building energy examines the individual's energy use or a group of buildings [11]. The bottom-up approach is applied when the energy demand analysis assesses each building's contribution to urban energy consumption [47]. The bottom-up model's procedure consists of model development, calibration, validation, and simulation of urban energy demand. It represents the UBEM more accurately using extensive data to simulate the energy demand. The bottom-up method's strengths are the capability to use a high level of details, simulation of energy use at different temporal scales, and reliable calculation [48].

In the literature, several UBEM platforms have been proposed, such as SUNtool [49], CitySim [50], UMI [51], TEASER [52], CityBES [53], and CESAR [54]. UBEM tools like CESAR were proposed to model the energy demand and retrofitting based on an automated bottom-up approach in different spatial scales [54]. The model was built by considering geometry information and enriched using census data, building characteristics, and OB data.

Table 1 summarizes the overview of 24 studies focused on UBEM platforms and workflows published between 2003 and 2018, with a specific focus on occupancy sub-models in UBEM. The occupancy sub-models' characteristics include the occupancy modeling approach, occupancy data type, data source, and building type. Additionally, the UBEM simulation engine considered in the research studies is highlighted in the table. It is also important to mention that Table 1 indicates the UBEM models that have exploited more than one occupant-related input parameter in their simulation.

Figure 3 indicates the percentage of each occupant-related input parameter considered in the UBEM studies mentioned in Table 1. Fig. 3 is drawn as the result of the *meta*-analysis of reviewed papers mentioned in Table 1. Occupants' presence is the primary input parameter for 92% of the UBEMs. 79% and 75%

Table 1
Urban building energy modeling workflows and platforms selected for review with a focus on occupancy models.

Authors [Ref]	Platform/tool	UBEM objective	Occupancy modeling approach	Occupancy-related data type	Data source	Building type	Simulation engine
Yamaguchi et al., [55]2003	–	Evaluating options for district energy systems design	Stochastic	Presence, appliance usage, activity schedules, HVAC	Customized	Office	HASP
Robinson et al., [49]2007	SUNtool	Describing the factors which influence dynamic demand	Stochastic	Presence, appliances, windows, lighting & shading, HVAC	Literature	Office	Customized
Shimoda et al., [56]2007	–	Simulating heating and cooling demand, operation of appliances & occupants' behavior	Stochastic	Presence, lighting, appliance, HVAC use, DHW, activity schedules	Customized	Residential	Customized R-C model
Heiple & Sailor, [57]2008	–	Estimating hourly & seasonal energy consumption profile	Deterministic	Presence, lighting, appliance, HVAC use	ASHRAE, NREL	Commercial & Residential	eQuest
Robinson et al., [50]2009	CitySim	Simulating building's energy flows	Deterministic & stochastic	Presence, activity schedules, windows	Customized	Residential	Customized R-C model
Caputo et al., [58]2013	–	Assessing different energy strategies by characterizing energy performance	Deterministic	Presence, ventilation, internal loads	Customized	Commercial & Residential	EnergyPlus
Reinhart et al., [51]2013	(Urban Modeling Interface)UMI	Analyzing operational energy	Deterministic	Presence, appliance, lighting, HVAC use	ASHRAE	All types	EnergyPlusRadiance, Daysim
Orehounig et al., [59]2014	–	Integrating decentralized energy systems	Deterministic	Appliance, lighting	SIA 2024	Residential, hotel & office	EnergyPlus
Remmen et al., [52]2015	(Tool for Energy Analysis and Simulation for Efficient Retrofit) TEASER	Simulating spatiotemporal energy demand	Deterministic& stochastic	Presence, appliance, lighting	Richardson lighting use model (TUS)	Residential	Modelica, AixLib
Bollinger & Evins, [60]2015	(Holistic Urban Energy Simulation) HUES	Simulating urban multi-energy systems	Stochastic	Presence, appliance, activity schedules	Richardson activity and appliance use model (TUS)	Residential	EnergyPlus
Nouvel et al. [61], 2015	SIMSTADT	Analyzing city districts energy	Deterministic	Presence, appliance, lighting, HVAC	The Association of *German* Engineers (VID)	All types	ISO/CEN standards-based reduced-order model
Fonseca et al., [62]2016	(City Energy Analyst) CEA	Retrofitting and designing district energy systems	Deterministic	Presence, ventilation, lighting, appliance, HVAC, DHW	SIA	All types	R-C model
Baetens & Saelens, [35]2016	StROBe	Accounting uncertainty for district energy simulations	Stochastic	Presence, lighting, activity schedules, appliances, DHW	TUS, questionnaire survey	Residential	Modelica, IDEAS
Ellis, [63]2016	Params-NZP	Generating building energy models for parametric analysis	Deterministic	Presence, lighting, appliance, HVAC, DHW	DOE	Army, commercial & residential	EnergyPlus
Schiefelbein et al., [6465]2016	–	Thermal simulation & load flow calculation	Stochastic	Presence, appliance & lighting	Richardson model (TUS)	Residential	TEASER (Modelica)
Cerezo Davila et al. [7]	MIT UBEM Tool	Calculating hourly energy demand load	Deterministic	Presence, lighting, appliance, HVAC, DHW	ASHRAE	All types	EnergyPlus
Hong et al., [531]2017	City Building Energy Saver (CityBES)	Energy benchmarking, urban energy planning, Retrofit analysis	Deterministic	Presence, lighting, HVAC, appliances	DOE	office & commercial buildings	EnergyPlus
Ahmed et al., [66]2017	–	Forecasting city-scale energy demand (end-use profiles)	Deterministic	Presence, lighting, HVAC, appliances, activity schedules	DOE	All types	EnergyPlus
Sokol et al., [67]2017	–	UBEM calibrationProviding usage profiles	Stochastic	Presence, appliance, lighting	Customized	All types	EnergyPlus
Nageler et al., [68]2017	–	Simulating annual heating & DHW energy consumption	Deterministic	Presence, window ventilation, equipment, lighting	IDA ICE occupancy model, SIA	Residential, office, commercial, school	IDA ICE
An et al., [69]2017	Stochastic Occupant Behavior (SOB)	Simulating cooling load	Stochastic	Presence, lighting, appliances, HVAC control, window	Questionnaire survey	Residential	DeST
Wang et al., [54]2018	(Combined Energy Simulation and Retrofitting) CESAR	Simulating energy demand & retrofitting modeling	Stochastic	Presence, activity schedules	SIA 2024	Residential	EnergyPlus
Torabi Moghadam et al., [70]2018	–	Estimating energy consumption	Deterministic	Presence (Building occupation ratio)	ISTAT national census	Residential	Customized
El Kontar & Rakha, [71]2018	–	Modeling occupancy & consequent energy loads	Deterministic	Lighting, equipment, appliances, presence	Energy measured data	Residential	EnergyPlus

of the UBEMs considered the appliance loads and lighting profiles in their energy simulation. In 21% of the UBEM models, domestic hot water (DHW) schedules were used. However, it is still unknown how much representative DHW usage could influence building energy simulation [72].

Figure 4 illustrates the data source of occupancy sub-models in the UBEM presented in Table 1. The inference from Fig. 4 is that 22% of the UBEMs used the customized co-simulation for occupant-behavior inputs. 18%,17%, and 13% of the UBEMs used the Swiss Society of Engineers and Architects normative (SIA), Time use survey (TUS) dataset, and American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) schedules to model occupancy. The other sources such as Department of Energy (DOE) occupant-related profiles, Italian National Institute of Statistics (ISTAT) national census, questionnaire surveys, Association of German Engineers (Verein Deutscher Ingenieure), and energy measured data have been used less than 10%, respectively to model the occupancy in the UBEMs. The distribution of the building types considered in UBEMs' occupancy sub-models is shown in Fig. 4. In 25% of the UBEMs, all type of buildings has been modeled. Also, the percentage is shown in Figs. 3 and 4 need not necessarily be equal to 100% as the same origin, and building types might have applied for different UBEMs' occupancy sub-models. Fig. 5 indicates the number of UBEM models that support different building types based on Table 1.

3. Occupancy modeling approaches

OB modeling is fundamentally determined by simulating the occupant's presence, movement, and interaction with the building systems [26]. The term 'occupancy' describes the occupants' presence and absence in each building's usage zone. Occupant energy use behavior describes occupants' interaction with electrical appliances, lighting systems, thermostat setpoint settings, and other building energy systems [73,74]. In the previous studies, various parameters such as occupant presence/absence state [75], lighting [76], plug loads [77], and thermostat setpoints [78] have been exploited to model the occupancy and occupant energy use behavior.

The modeling of realistic occupancy accounting for its inherent stochastic characteristics is a challenge. For this reason, most of the preceding research works applied static and deterministic occupancy schedules for UBEM, which might not necessarily capture the variation associated with occupancy. For instance, UMI [51], CEA [62], and CityBES [53] simulation tools use standard or deterministic schedules for occupancy modeling. Recent studies focused on stochastic/probabilistic occupancy modeling to account for the occupant behavior's temporal variations. The TEASER tool used occupancy presence and electrical load to generate the occupancy schedules using the Richardson presence model in the district's residential building stocks [52]. In the CESAR platform, Wang et al. [54] model the occupancy schedules based on the occupant's presence and activity profiles to predict the district energy demand and model the retrofitting. The modeling approaches of occupancy (Fig. 6) in both building and urban scale generally include deterministic, data-driven, stochastic/probabilistic, and agent-based approaches [24,79].

3.1. Deterministic model

The deterministic approach, the typically used input in the UBEM models, refers to the fixed schedules (low-level of complexity) extracted from standards (such as ASHRAE [80], DOE [38] and National Energy Code of Canada for Buildings (NECB) [81]) or a set of certain rules [38,82]. The deterministic schedules provide a simple representation of occupant behavior for different building types composed of the hourly values to form a daily profile [83]. In the literature [38,84], "deterministic" model" is mentioned as 'deterministic schedules, deterministic rules/profiles' or rule based models. In general, both rule-based model and deterministic model is a modeling approach that are conceived from a set of rules or data source meaning that the model is a repeated manifestation of limited number of patterns without any randomness. Deterministic schedules or rules can be translated into Markov-chains to introduce the variations in the schedules. Though technically, deterministic and rule-based models represent the same meaning, in many literatures [7,38–40,50,62,79] such model is commonly termed as deterministic models. Hence, in this review both rule-based, and deterministic models are commonly termed as 'deterministic model'. Fonseca et al. [62] in the CEA used standard time-series data to represent occupant schedules, ventilation rates, and temperature and humidity setpoints for all buildings embedded in the archetype database. Robinson et al. [50], in the CitySim tool, considered the deterministic profiles of presence, windows and blinds opening, lights, appliances to simulate buildings' energy flows. Fig. 7 is an example of deterministic schedules for occupancy, lighting and appliances, and domestic hot water consumption, respectively considered in the CEA tool for a school building in Switzerland.

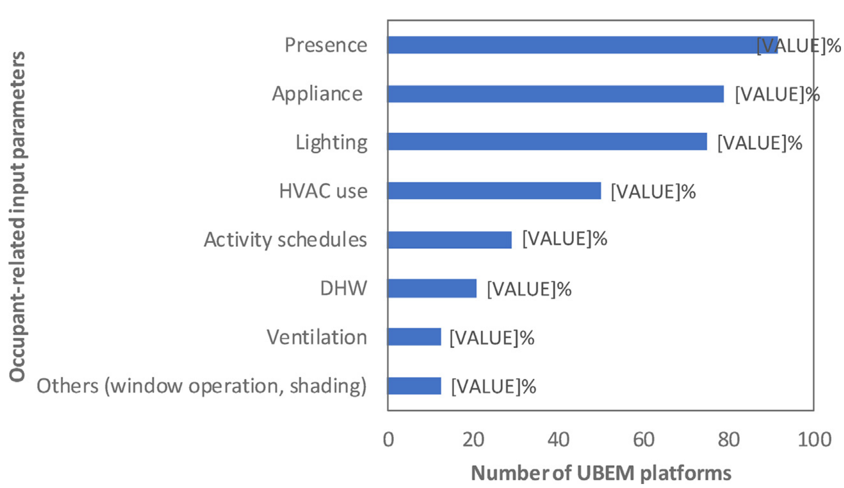

Fig. 3. Occupant-related input parameters considered in the UBEM based on Table 1.

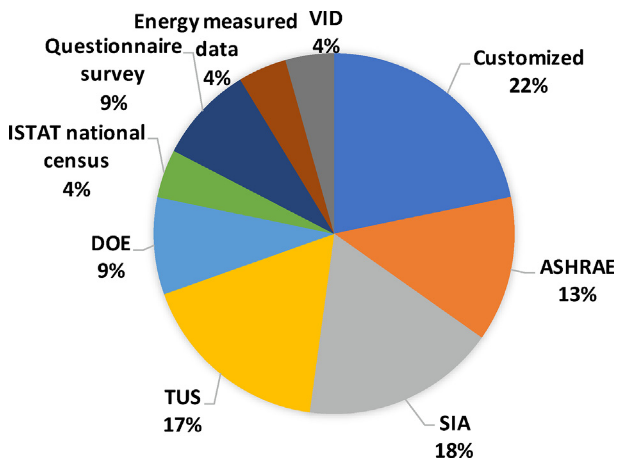

Fig. 4. The data source for occupancy sub-models in the UBEMs considered in Table 1.

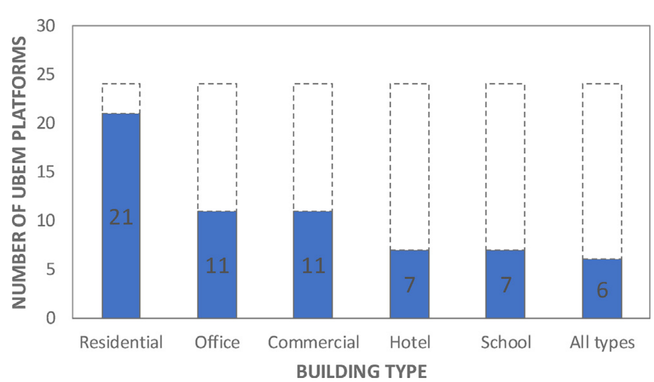

Fig. 5. Number of UBEM models that support different building types based on Table 1.

3.2. Data-driven modeling

On the other hand, occupant-related inputs derived using the data-driven approach in a way are fixed but are derived using the robust data collected from buildings using different sources. In recent years, the increment of smart energy monitoring devices (such as BEMS, smart thermostats) has made it possible to create publicly available databases, including building characteristics, occupant-related data, and energy consumption. This made it possible to develop data-driven models with the integration of engineering simulation to create a more accurate building energy simulation model. The data-driven approach mostly uses data mining techniques (most commonly k-means [86,87], k-Shape clustering [88,89], decision tree [90,91], associate rule mining techniques [92,93]), machine learning models [94,95], and statistical methods [96]. Fig. 8 is an example of daily occupancy presence probability in a three-bedroom apartment derived using a data-driven approach. The figure is drawn using the occupancy presence/absence data monitored (at one-minute resolution) using passive infra-red (PIR) motion sensors installed in various locations of an apartment located in Lyon, France. The data relating to occupancy (presence/absence state), plug load, and lighting energy consumption was collected from 32 apartments of a high-performance mixed-use building for one year, and the data collected from one of the apartments was used for illustration. The data-driven framework for deriving occupant schedules using the data collected from several apartments of the same building considered in this study is explained in detail [89]. The readers can refer to the following papers [89] and [93] for more details on the data collection, sensor details, and dataset description. The purpose of Fig. 8 is to emphasize the variation in occupancy schedule between each day of the week in the same apartment. By defining the same occupant schedule for all weekdays, there is a high chance of repeated schedules and the variations in the occupancy with respect to time of the day, day of the week, seasons are missed out. Therefore, though a rule-based occupancy schedule is a simple and basic solution for defining occupancy in UBEM, stochastic approaches are suggested in the literature to incorporate the occupants' dynamic nature [69,97].

Researchers investigated the relation between electric energy consumption and occupant presence using big data analysis [98,99]. However, few studies have been implemented to recognize the occupant-related patterns using data analysis in residential buildings due to privacy issues. Consequently, Time use surveys (TUS) as occupant's behavior schedules are applied to model the energy use in the residential buildings [100]. Annex 79 is investigating to model occupancy (e.g., presence and activities) using data-driven models [101]. Bianchi et al. [102] proposed a novel methodology to model parametric occupancy-related schedules using real 15-min resolution electric metered data in the diverse building stock. Replacing fixed schedules with data-driven occupancy schedules for urban building energy simulation may impact occupant-related applications' annual energy demand. Besides, it might affect predicting the peak power, leading to different supply systems sizing [27].

Data mining and machine learning help step forward to the model development after data preprocessing in data-driven mod-

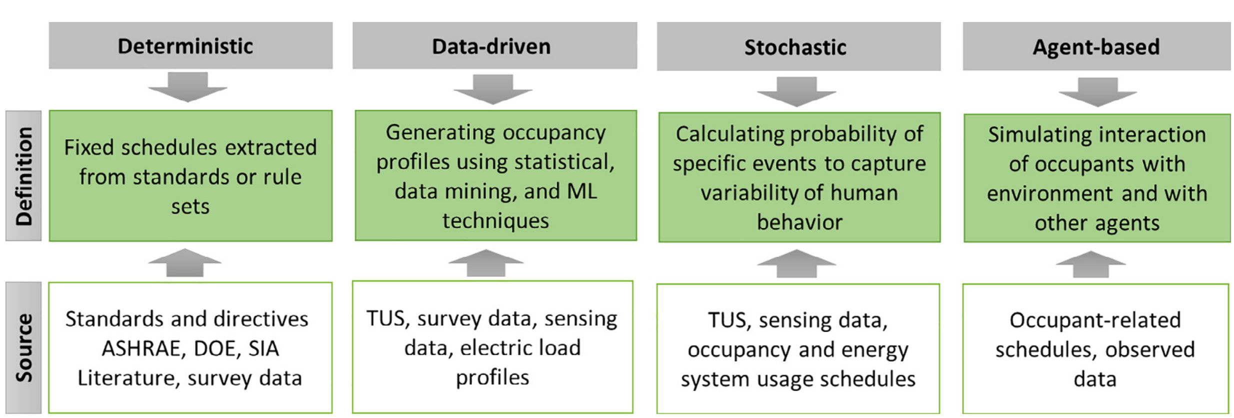

Fig. 6. Occupancy modeling approaches.

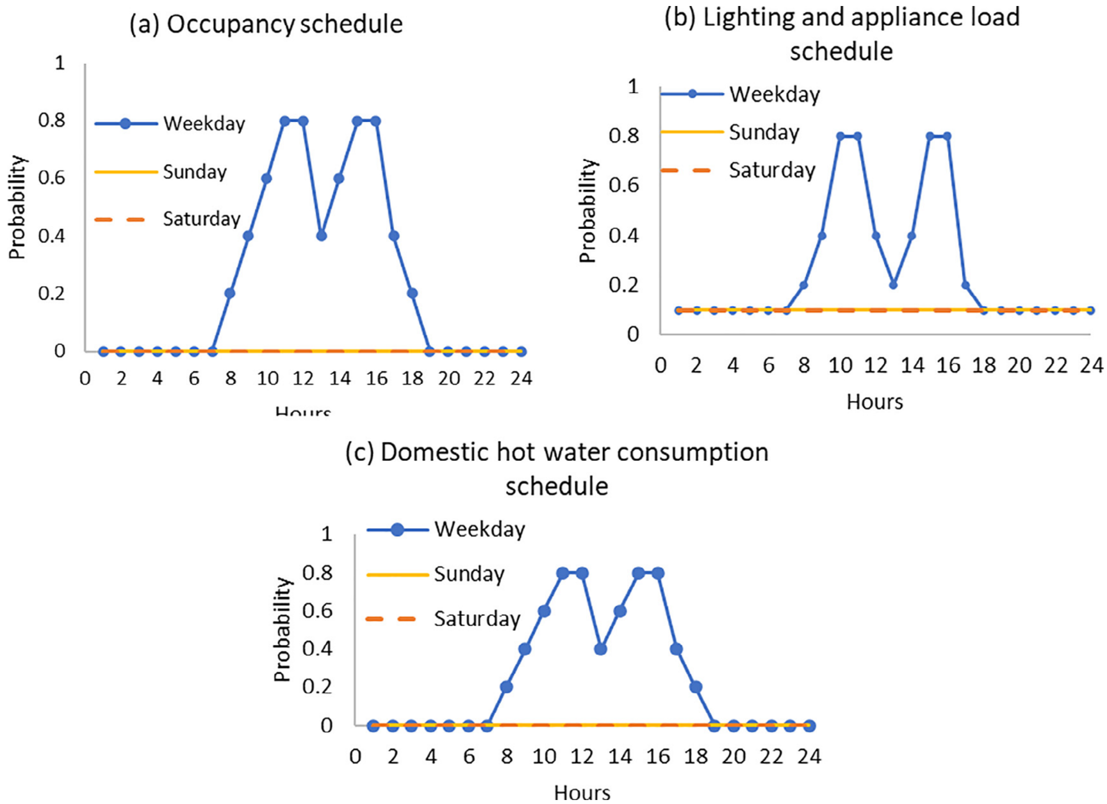

Fig. 7. Example of deterministic schedules for (a) occupancy, (b) lighting and appliances, and (c) domestic hot water consumption considered in CEA tool for a school building in Switzerland, modified from [85].

eling procedures. The commonly used ML techniques in OB pattern recognition and prediction are clustering, decision tree, association rule mining, and Bayesian network [24]. It is found that the predicted OB patterns using ML techniques can represent the stochastic nature of the OB in building energy simulation [24]. In addition to using one ML technique in occupancy modeling, more than one ML technique was applied to benefit the strength of various methods in some research works [103]. Huchuk et al. [94] compared various ML models, including traditional classification, logistic regression, Markov models, Random forest, and recurrent neural network (RNN), for occupancy prediction in residential buildings to evaluate their performance in terms of the accuracy level and computationally viability. Noticeably, they used connected thermostat data as input of the ML models. The ML methods are mostly applied to estimate the occupancy without exactly tracing them [20]. Currently, there are few studies of occupancy data-driven models that are used for UBEM. Happle et al. [27] proposed a data-driven occupancy model to generate the urban building occupant presence model based on Location-Based Services (LBS) using the CEA tool for energy simulation. They concluded that the occupancy profiles are sensitive parameters in urban building energy prediction. Besides, they found that the standard schedules significantly overestimate the building occupancy. Another purpose of using ML models is to calibrate the UBEM models based on OB patterns. El kontar and Rakha [71] calibrated the UBEM model of a residential neighborhood in Austin, Texas, utilizing clustering analysis of building energy use variables such as lighting, equipment, cooling, and heating.

3.3. Stochastic/probabilistic modeling

Studies at urban scale illustrate that stochastic modeling is preferred to predict realistic energy loads with a high temporal resolution [69]. Generating occupancy profiles using stochastic models increases the reliability of building energy simulation [103]. Developing the stochastic occupancy models is mostly based on the Markov Chain processes, built upon TUS or sensor-based data [104]. Page et al. [74] simulated the occupant's presence and absence of a single office based on time-inhomogeneous Markov chain considering occasional periods of prolonged absence. They used the weekly presence probability statistics to simulate the occupant presence and absence event.

Furthermore, probabilistic models are used to develop stochastic occupancy models in which the state of the OB follows a specific probability distribution [24]. Wang et al. [105] proposed a simple mathematical formula to represent the OB's stochastic characteristics to the environmental conditions. Cecconi et al. [106] proposed an automated and tested probabilistic occupancy model to reduce epistemic and stochastic uncertainties using a multi-layered supervised feedforward Artificial neural network (ANN). Virote and Neves-Silva [107] presented an energy consumption prediction model, considering the stochastic nature of occupant presence. They modeled the OB based on the lighting system's usage. However, they acknowledged that the other energy system could be replaced to represent the occupant's presence's randomness. Additionally, it is reported that using a dynamic occupancy model provides a more accurate result than a static algorithm [108]. Gilani et al. [109] quantified lighting energy use prediction uncertainty in offices of different sizes. They concluded that the occupant impact on the annual lighting energy consumption decreases by increasing the office size depending on the occupancy modeling approach. This study illustrates that stochastic occupancy modeling would not be beneficial if the objective is to simulate the average energy use prediction of large office buildings. The studies mentioned above proposed the stochastic occupancy models at

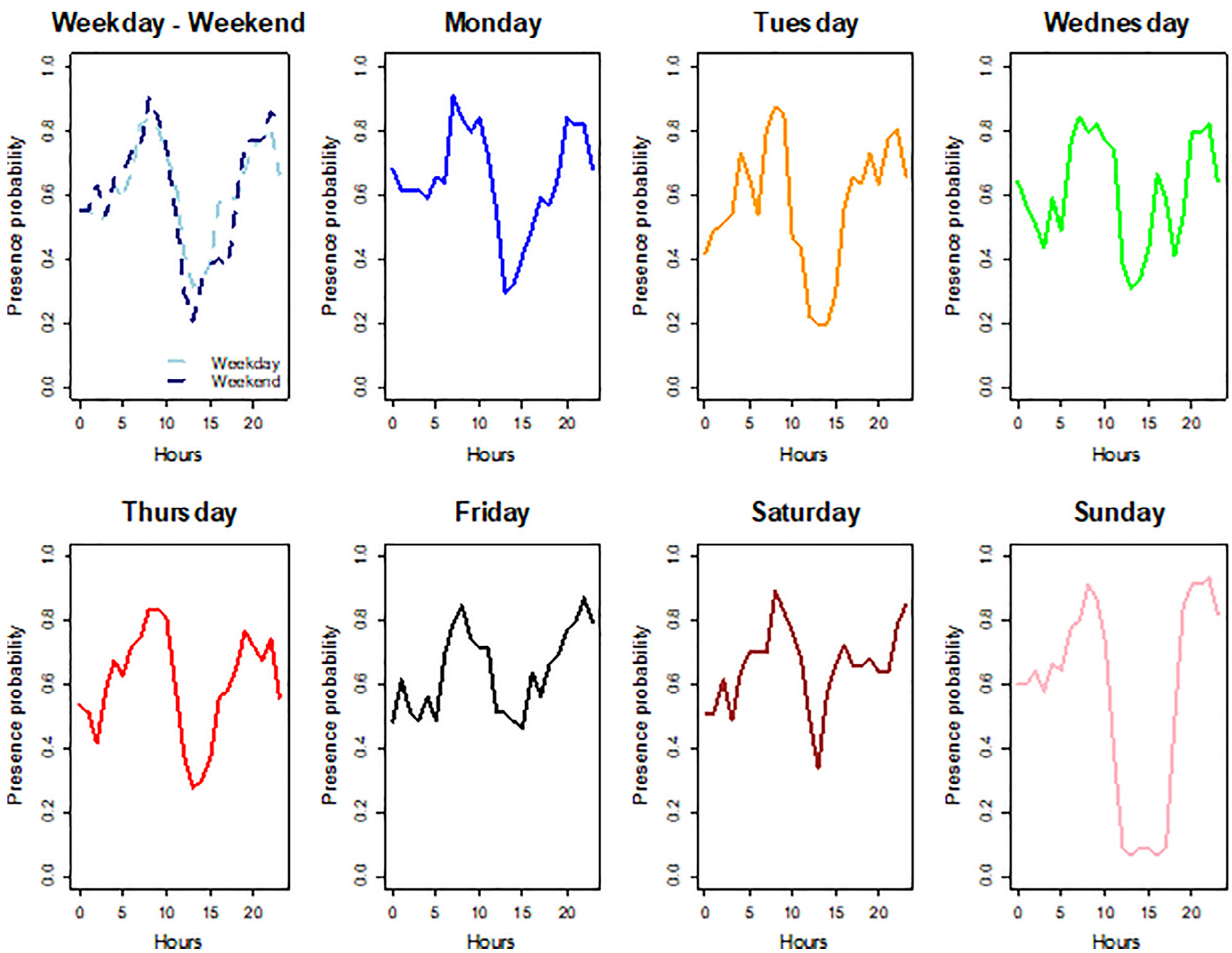

Fig. 8. Variation in daily occupant schedule observed in an apartment.

the building level. Several UBEM models considered the stochasticity of the occupant-related behavior using the stochastic occupancy models in connection with their platforms. In TEASER tool, Schiefelbein et al. [65] generated the annual occupancy, appliance, and lighting profiles using the Richardson lighting use model [110] to simulate district buildings' thermal demand. An et al. [69] proposed a stochastic modeling method with the OB model reflecting different spatiotemporal diversity and stochastic sampling processes to predict the district cooling loads. They concluded that using oversimplified OB schedules could lead to overestimated energy use and cooling loads. He et al. [97] developed a stochastic occupancy model to generate neighborhood hourly thermal demand profiles. It is reported that the stochastic model could generate more representative hourly thermal demand profiles. A significant difference was observed between the profiles generated using the stochastic occupancy model and the standard heating hours.

3.4. Agent-based modeling

The agent-based modeling (ABM) considers the aggregated actions through occupants' interactions and relationships [111]. In ABM, the individuals are considered agents, and their interactions in the built environment are modeled. ABM is used to model occupant movement in buildings, their energy-related behavior, and adaptability in terms of thermal comfort [79]. Liao et al. [112] proposed a novel Multiple Modules (MuMo) model. One of the pioneering studies explored the advantages and disadvantages of using the ABM approach for occupancy modeling. They simulated the multi-occupant single-zone scenario and multi-occupant multi-zone scenario. Liao and Barooah [113] used the agent-based approach to model a real-time estimation of occupants' numbers in a commercial building. They used Monte Carlo simulation to extract a reduced-order statistical model from the ABM. In their review, Li et al. [24] reported that developing ABM models require detailed insights on the agent's specific behavior and sequences of different behavior. Also, ABM requires high computational time, which challenges its application in large-scale simulations. Though ABM can model the heterogeneity associated with the occupants, detailed data about each agent is required, further hindering their application at the urban scale. Because of the complexity, high computational time in modeling the agents and their interactions with multiple zones in the built environment, other modeling approaches are often preferred in the previous research.

4. Occupant-related inputs

The main and widely represented occupant-related input parameters in the energy simulation are occupant presence/absence state, appliance use, lighting, setpoint temperature schedules, DHW usage, and window operation [50]. Insights of each occupant's energy-related input at both building and urban scale are presented in the following sections.

4.1. Occupancy

Kjærgaard et al. [114] categorized the occupancy into occupancy detection and occupancy estimation. Occupancy detection refers to the binary inference of the presence and absence of the occupants, and occupancy estimation is meant to determine the number of occupants in the space. In addition to the above, one more input related to occupancy is occupant activity level, which is useful to estimate the internal gains due to occupants in buildings. Quan et al. [16] reported that the main difference between the energy modeling performed at building and urban scale is the interactions of various systems (for example, interactions between buildings and their surroundings, interactions between occupants and buildings) in the urban context. Happle et al. [38] stated that interactions between occupants with multiple buildings in the city/district are often sidelined in UBEM. For demonstration, the authors [38] gave an example of an occupant being absent in the office during lunchtime would eventually be present in the restaurant or home, and these interactions are not often considered in UBEM. The main reason for oversimplifying the assumptions related to occupancy at the urban scale is the lack of occupant-related data availability in building types. It must be mentioned that the impact of such detailed occupancy depends on the objective of the UBEM. If the aim is to estimate the yearly energy demand of buildings, then the interactions of individual occupants with multiple buildings might get diminished in the urban context.

Some studies estimated the occupants' presence based on operational control, such as the occupants' interaction with thermostats [94], lighting and shading systems [115]. The occupants' arrival and departure and the absence duration could influence the building energy consumption [116]. Energy use during an occupant's absence has a considerable impact on energy-saving potential [117]. Data-driven based approach is a promising method to extract the occupancy schedules and predict the number of occupants [118]. Panchabikesan et al. [89] developed a data-driven framework to extract the occupant activity and presence schedule from several apartments using shape-based clustering and change-point detection techniques. D'Oca and Hong [119] used data mining techniques (k-means clustering and decision tree) to extract the occupant schedules from 16 office buildings. It is suggested that the extracted occupant schedules shall be used to study the effect of occupant presence schedule on system operation and energy use in office buildings.

4.2. Occupant energy-related inputs

In most UBEM models, the occupant's presence was considered to calculate the associated internal heat gain. As mentioned earlier in this section, plug load, lighting load, setpoint schedules, and DHW usage are the other occupant-related inputs that need to be defined while performing the energy simulation. Assigning the appliances and lighting load schedules by occupancy is crucial as they are highly correlated [118].

4.2.1. Plug load schedules

Plug load schedules are related to the electrical equipment usage behavior of the occupants. In general, the plug load' schedule highly depends on the occupancy schedule. In residential buildings, these two possess a strong correlation with each other [86]. Fig. 9 depicts the plug load and occupant activity schedule (represented by the hourly average number of movements detected on the specific day of the week) in four apartments. Figs. 8 and 9 are drawn using the data collected from the high-performance mixed-use buildings located in Lyon, France. The data in Fig. 9 illustrates the average daily profile of occupant activity schedule and plug load schedule (starting from Monday) derived from the data collected for one year from four apartments. At the building scale, it is possible to understand and derive the plug load schedules for the specific building and assign the representative occupancy and appliances schedules (as shown in Fig. 9) while performing energy simulation. However, it is challenging to assume a distinct/representative plug load schedule for each building at the urban scale. Therefore, in most UBEM models, plug load schedules derived from deterministic models are used to simulate the building energy demand. Assigning similar profiles for different building functions would misrepresent the actual energy demand and peak loads [19]. Carlucci et al. [82] reported that most of the studies in the literature had used data-driven models to extract the appliance use schedules and related occupant activities. Clustering algorithm coupling with event detection algorithms has also been established to describe the state of the appliance use [120]. Nevertheless, the appliance use patterns could be stochastically described using data-driven techniques. Only a few stochastic models predict the appliance use patterns in specific buildings, not in a mixed-used district [38]. Spiegel [121] proposed an approach to identify the appliance use schedules considering the electric power consumption change using Bayesian networks.

The internal heat gains generated from electrical appliances are highly correlated with the occupants' activities and interaction with the appliances [122]. Richardson et al. [123] presented a high-resolution time-correlated appliance use model based on occupants' TUS data to estimate the dwellings' annual electricity demand. The appliance activity (e.g., cooking) and the number of active occupants has been considered to determine the probability of appliance switching on at each step.

4.2.2. Lighting schedules

Lighting schedules are one of the parameters used to predict occupancy and assess building energy consumption. While lighting load contributes to a considerable amount of energy consumption, monitoring real-time lighting use could optimize the buildings' energy efficiency [124]. Buildings' occupants meet their need for visual comfort by interacting with the lighting system. Therefore, a correlation between lighting energy use and occupants' indoor activities must be considered to model the lighting use schedules. As an example, the lighting load variation with respect to occupant activity schedule in residential buildings is shown in Fig. 10. The figure shows that the lighting load follows the occupant activity pattern and is different for each building. Assigning a fixed/default schedule for several buildings leads to uncertainty in UBEM results. Also, Zhou et al. [125] found that the lighting energy use is mainly driven by the occupancy schedules. Reinhart [116] proposed the Lightswitch-2002 to model the lighting energy consumption. The annual occupancy schedules and illuminance use patterns are the input of the Lightswitch-2002 model. Another approach to model the occupant action related to lighting system use is implementing probabilistic modeling and different machine learning algorithms such as Support Vector Machine (SVM) and k-nearest neighbors (k-NN) [82]. A high-resolution domestic lighting demand model was proposed by Richardson et al. [126] that considers the natural lighting coupled with occupants' activities. They generated the residential buildings' lighting electricity demand profiles using a time-series active occupancy schedule.

4.2.3. Temperature setpoint schedules

Temperature setpoint schedules are a crucial input to simulate the HVAC system performance in buildings. The occupant space-heating/cooling operation is influenced by several drivers, including environmental factors, building systems, and occupant-related factors [127]. The NECB's assumptions associated with the living spaces' heating and cooling setpoints are 21 °C and

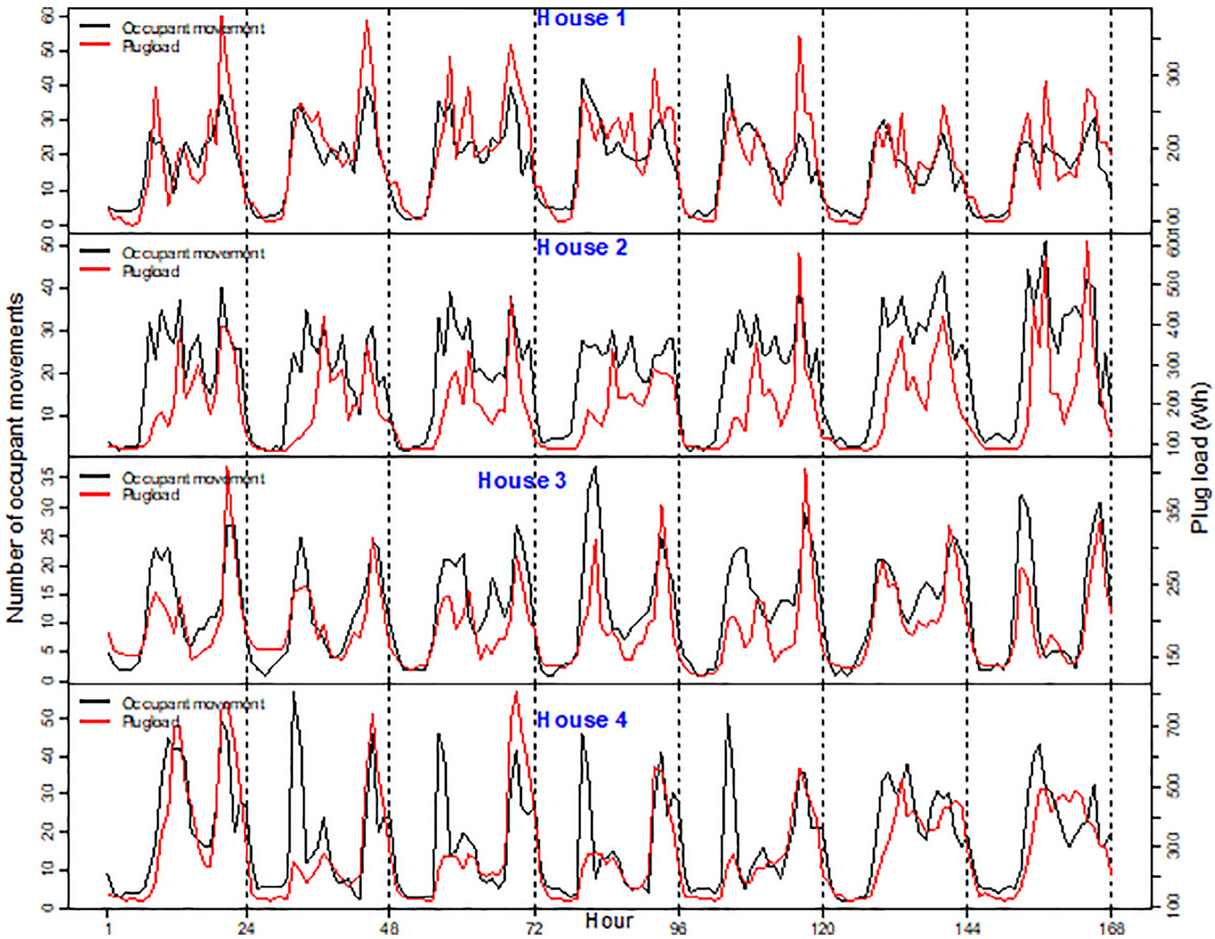

Fig. 9. Variation in daily plug load schedule (starting from Monday) with respect to occupant activities.

25 °C, respectively [81]. Generally, widening the setpoint temperature could conserve the energy demand, especially during the night and when the building is unoccupied [10]. Besides, in the design stage, considering a higher setpoint temperature in the summer and a lower setpoint in the winter could lead to a more efficient HVAC system. In building energy simulation, typical setpoint temperature schedules are applied. Since the occupants occasionally adjust the setpoint temperature schedules, applying default profiles in building energy demand simulation leads to a gap between simulated and actual energy consumption [82].

Chuck et al. [128] reported that variation in input related to a setpoint temperature of even 0.5 ℃ in building energy simulations could impact the heating and cooling demand predictions up to 10%. The setpoints temperature schedules could be extracted from building management systems data [129] and thermostat data [130]. A logistic regression model can approximate the frequency of thermostat interactions while the indoor temperature is considered the predictor variable [131]. Fabi et al. [132] proposed a probabilistic model to simulate the occupants' interaction with building controls using the measurement of thermostat radiator valves' setpoint. In another study, Peng et al. [133] applied a neural network algorithm to learn the time-dependent HVAC setpoints to provide the possibility that the HVAC system automatically adapts to the occupant temperature preferences in commercial and office buildings. They concluded that using dynamic temperature setpoints leads to an energy-saving between 4% and 25% compared to the fixed schedules at the low values of the preferred temperature range. Ren et al. [134] used data mining techniques to derive the distinct room temperature behavioral patterns using temperature data collected from 62 apartments. Their results indicated different room temperature profiles among the considered apartments and that the space heating system's operations (cycling frequency) were found to be more frequent than expected because of the tight range of the thermostat settings. Panchabikesan et al. [130] used the time-series clustering method to extract distinct patterns of average daily heating and cooling setpoint temperatures in residential buildings, which can be used in energy simulations as the replacement of the fixed schedules from standards. For the analysis, data collected from around 13,000 residential buildings across Canada was used. In addition to the clustering analysis, the authors used random forest model to determine the relative importance of different attributes that influence average thermostat setpoint preferences.

4.2.4. Domestic hot water usage

Domestic hot water (DHW) usage is another significant driver in different building types energy consumption. Extracting the detailed DHW schedules allows assessing the building systems' energy efficiency reliably and optimizing the control operating systems. The influential parameters on water consumption include climatic conditions, seasonality, socio-economic factors, and the building use type [72]. Like other usage profiles, the DHW profiles applied in the building energy simulation are extracted from technical standards, representing the identical water consumption of

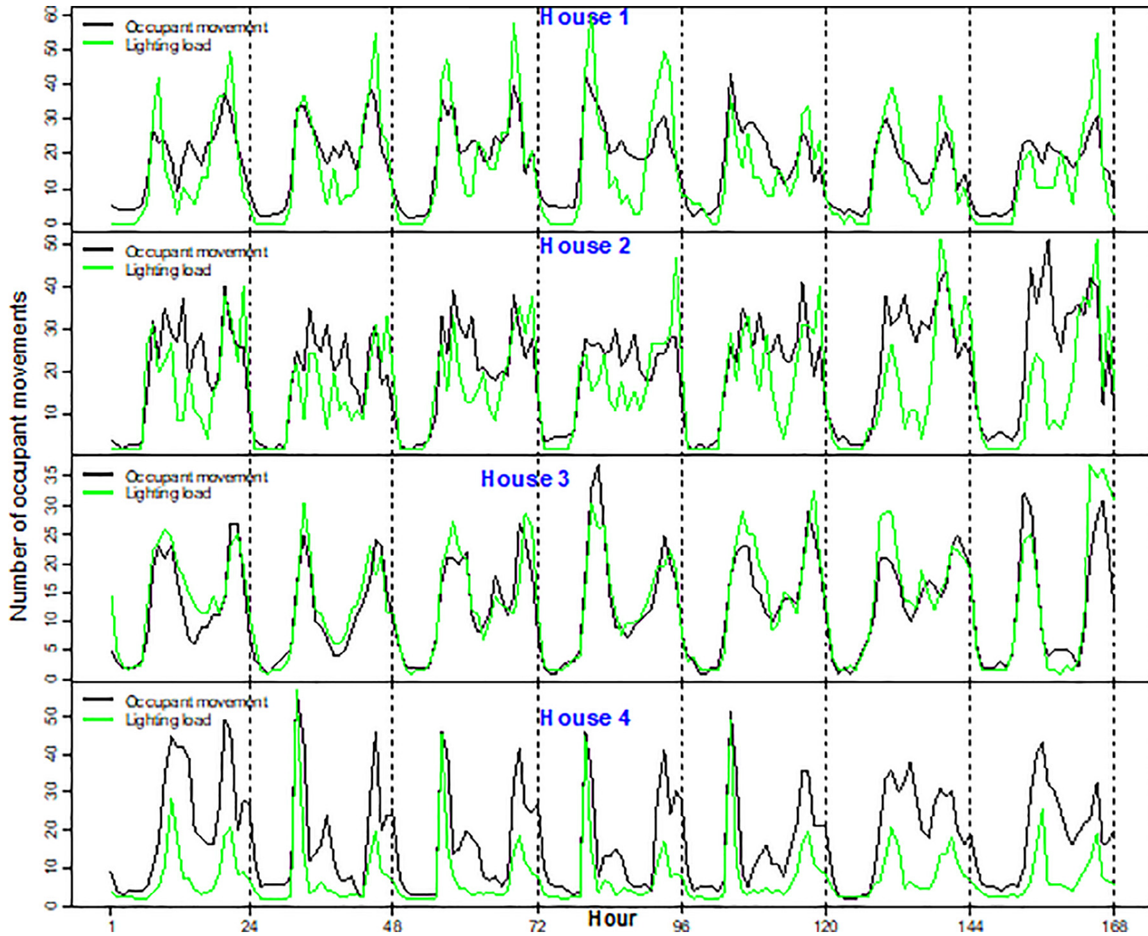

Fig. 10. Variation in lighting load schedule concerning occupant activities in residential buildings (apartments).

each day of the year. Several methods have been proposed to generate the DHW profiles, including stochastic models [135], forecasting the DHW consumption of individual residential houses using estimated exponential smoothing, autoregressive integrated moving average (ARIMA), and seasonal decomposition models [136], machine learning-based models such as ANN [137], statistical approaches, and long-term filed recording DHW consumption data. Since the DHW patterns are geographically dependent, the local measurement must tune the DHW models [138].

According to Table 1, 21% of the UBEM platforms have considered the DHW schedules as an input of building energy simulation. For instance, Fonseca et al. [62] in CEA computed the hourly demand of the DHW with the spatiotemporal demand model developed in [139]. The required data to calculate the DHW was gathered from local building standards and literature. Cerezo Davila et al. [7] extracted the average energy use intensities of DHW and other parameters such as lighting and appliance use from the Energy Information Agency surveys for different building types. The hot water energy use model was proposed by Shimoda et al. [56] to simulate the energy consumption of the residential sector at the city level. The amount of hot water uses of living activities, hot water temperature, and activities frequency were taken into consideration.

4.2.5. Window operation

The window opening/closing behavior influences the indoor air quality and the room temperature and influences the building energy consumption. Therefore, along with the above-mentioned occupant energy-related behavior, this study provided an overview of some studies that considered window operation as an urban building energy simulation input in the UBEM platforms.

Robinson et al. [49], in their urban energy modeling platform (SUNtool), developed a window opening/closing behavioral model considering the probability of interaction and the consequences of this interaction. Then, the result of this model is parsed to the whole building thermal model. The window opening schedules induce a deterministically calculated ventilation rate. Additionally, they considered the appliance usage, lighting, shading to calculate urban buildings' dynamic demand. In another GIS-based UBEM platform, Nageler et al. [68] calculated the urban building energy demand using the IDA ICE model instead of occupancy profiles considering the action pertaining to windows and its impact on ventilation rate. To simulate the cooling load of a residential district, An et al. [69] considered five types of occupant behavior, including occupancy presence, lighting, HVAC control, cooling temperature setpoint, and window operation. Considering the indoor thermal conditions, they determined the HVAC and windows state using the probability functions to simulate the thermal demand. CitySim [50] is another UBEM model that takes windows opening and corresponding ventilation exchanges into account to model the buildings' energy flow and interaction of the buildings on the urban scale. Recently, the option for representing window opening/closings as level of detail 3 (LoD3) is made available in City Geographical Markup Language (CityGML) [140]. For more details on different LoD's of CityGML, the readers can refer to [140–142].

5. Implementation of OB in UBEM

5.1. Some occupant-centric datasets for extracting occupant related inputs at the urban scale

In general, monitoring the OB is complicated due to its spatial and temporal stochastic nature and privacy issues [143]. Therefore, OB is often expressed based on the probability of the actions carried out by the occupants. In the past, the occupancy presence data were collected using a questionnaire as a simple method considering the daily pattern of presence and activities [144]. One example of the questionnaire dataset is the American time use survey (ATUS) dataset containing information collected from 210,000 interviews from households across America from 2003 to 2019 [145]. ATUS is a robust public dataset containing the details on the amount of time people spend doing life's activities, including housework, working, and leisure activities [145]. This dataset was widely used to identify the occupancy schedules [146,147]. The UBEM platforms, including [53,61,64], used the TUS data indirectly by applying Richardson occupancy models connecting to urban building energy simulation. The 2015 Survey of household energy use (SHEU) is another nationwide data collected from several Canadian households, including self-reported heating and cooling setpoints [10]. The SHEU result does not differentiate the cooling setpoints for homes with and without air-conditioning and dwelling ownership status. Although SHEU surveyed the temperature setpoints in several thousand households across Canada, the simulation result is limited for OB modeling and energy management purposes due to not providing high-resolution data.

In recent years, urban occupant-centric data sets are being collected through several heterogeneous sources such as sensor-based and IoT data, mobility, city image-based data, building energy, and occupancy data. Sensor-based technologies are another available approach to detect and estimate occupancy [148]. However, using sensors to collect occupancy data has limitations such as installation and maintenance costs and probable errors due to the installation orientation [82]. The sensor-based data enables the energy modelers to effectively improve the active occupancy schedules to obtain the occupant's stochastic nature [149]. Regarding the sensor-based dataset available for modeling the occupancy at the urban scale, ecobee's 'Donate Your Data (DYD)' program [150] is a potential data source to understand occupant's preference on thermostat setpoint settings and to derive occupant schedules in residential buildings. The data is collected from different countries; however, most of the data is collected from residential buildings located in North America [94]. The thermostats have PIR motion detection sensors which could be used for occupancy modeling. Huchuk et al. [94] evaluated various machine learning models to predict occupancy motion state using ecobee's Donate Your Data dataset collected by modern connected thermostats in residential buildings. They found that in order to collect the occupancy data effectively, the number of sensors and their placement, the number of people at home, and people's general mobility pattern would be significant.

Pecan Street institute provides high-resolution data of approximately 1,000 homes in Texas, Colorado, California, and New York in the US [151,152]. The dataset consists of electricity in 1-second, 1-minute, and 15-minutes resolution, indoor temperature, water data, gas data. Hoon Yoon et al. [153] used Pecan Street data to design the internal gain and occupancy schedules to analyze the HVAC power consumption of residential buildings. The Building Data Genome Project 2 is an open-source dataset collected to represent the energy-related data from 1,636 non-residential buildings at an hourly frequency for two years across North America and Europe [154]. This dataset includes building electricity, water, steam, solar, and irrigation data. The metadata of each building contains the use type, area, weather, and year built. Miller and Meggers [32] presented a framework to characterize the model-based and pattern-based behavior by feature extraction using the Building Data Genome Project dataset.

Development of the mobile positioning data and IoT as modern technology has provided the possibility of deriving more realistic occupancy profiles to analyze its impact on the urban building performance [28]. Wi-Fi connection is a data source that occupancy and plug loads schedules could derive [155]. Hou et al. [26] proposed a novel framework to model the interaction between the building occupants and urban energy systems using Wi-Fi infrastructure to extract the occupancy profiles. The real-time data extracted from LBS such as Google Maps and Facebook's could be used to determine the number of occupants in a specific building in the structure of typical 24-hour profiles of each day of the week [27]. The process of applying the LBS data comprises the data collection, categorization of the places based on the building types, and extracting the occupancy schedules. Happle et al. [27] pointed out that collecting enough LBS data to run meaningful statistical analysis is feasible in large cities. However, several challenges exist in LBS data collection, data processing, and data validation. For instance, probable biases in data collection may happen once the data is gathered in larger cities or in a particular group with more cell phone ownership. The studies indicate that the extracted occupancy schedules from social networking services follow a similar trend with a slight difference [156].

The novel image-based technology is another source that provides the positioning and orientation information for the occupants in real-time. Digital image processing and three-dimensional reconstruction have been developed to detect the human pose that is not uniformly distributed in the buildings [157]. Meng et al. [158] developed a real-time estimation method to detect the number of indoor occupants to control the HVAC system by integrating image information. In another study, Petersen et al. [159] proposed an image-based method to detect the occupants anonymously by the ceiling-mounted cameras. They acknowledged that their method has a 99% accuracy in a three weeklong room occupancy test. Choi et al. [154] tested the applicability of vision-based occupancy counting in office buildings with deep learning algorithms. Also, they evaluated the performance of the method on the occupant-centric control strategies to the HVAC system. Although most of the studies which used image-based data are in the building scale, this occupant-centric data source could potentially be applied to detect the occupancy at the urban level.

5.2. Unified modeling language (UML) diagram representing occupant related inputs in UBEM

Urban building energy simulation work with a variety of data source of different size and format [160]. To organize the required input data of the UBEM models, assembling all the datasets extracted from surveys, city datasets, and public records into a single standardized database is essential [3]. CityGML is a data model that several UBEM models use to represent the 3D city models and the semantics to predict the district building energy demand [161]. CityBES served urban data model to store data from various sources and provides the input parameters of the modeling, simulation, and visualization [53]. In this platform, CityGML as an XML-based open data model is exploited to provide a standardized geometry model and exchange the data between the building energy model and other urban analysis platforms. The Energy Application Domain Extension (ADE) extends the CityGML standard by energy-related features required to simulate a standalone building's energy or city scale. TEASER utilized CityGML EnergyADE to export the enriched urban building information [52]. As one of the modules of the EnergyADE of CityGML, the OB module enables the building OB model [161]. The OB module consists of the classes representing the occupancy and occupant-related parameters required for an energy simulation. The operating schedules such as heating, cooling, and ventilation schedules could specify the indoor climate condition. The occupant-related classes are associated with the building unit and the usage and thermal zone classes (building-related objects) in the upper level. Fig. 11 shows the occupant behavior UML diagram that could connect to UBEM models providing the occupant-related inputs.

5.3. Level of details and co-simulation required to represent OB in UBEM

This section discusses the challenges, the level of detail, and co-simulation required to represent occupant-related input parameters in UBEM. Although several UBEM occupant input parameters were considered, there are still discussions on how detailed the input data should be. Do the simplified deterministic schedules support UBEM as an appropriate tool to estimate the energy demand? Or the complex probabilistic models and ABM are required to model the occupancy in the buildings accurately. In most UBEMs, the occupancy and OB models defined at the building scale are generalized as the urban scale's occupancy model [28]. Literature indicated that developing occupancy models that dynamically simulate the occupant schedules could provide relatively accurate simulation results and pave the way to identify energy-efficient strategies. Nevertheless, Romero [19] reported that although the stochastic-based model proposed by Page et al. [74] can represent the occupant's temporal variation, stochastic schedules on average will converge towards similar schedules used in the deterministic model. This is because of consideration of the occupant schedules from the deterministic approach for generating the stochastic schedules. The same authors provided the concluding remark to the question of which model should be used for UBEM. They stated that simple occupancy models (deterministic schedules) are good enough to assess the building's energy demand on a yearly or sub-yearly scale. The effect of accounting for the temporal variations of the occupant is less when assessing the building's yearly demand, especially in space-heating-dominated countries [79]. However, occupants' dynamic nature plays a significant role in assessing the hourly energy demand and predicting the peak demands. In this case, more detailed occupancy models such as probabilistic models and ABM could be considered so that the uncertainty in the simulation results shall be reduced [13,37,38,79,139].

While previous studies indicate that stochastic occupancy modeling reduces the uncertainty related to the variation of the occupant behavior, according to Table 1, among the 24 studies reviewed at the urban scale, 50% of the studies used the stochastic occupancy modeling approach. Further, it is observed that the urban-scale stochastic occupancy models are used for specific use-types (80% and 20% for residential and office buildings, respectively), not for a mixed-use district. One reason for this could be that developing the stochastic occupancy models for urban scale modeling is computationally intensive and requires large behavior datasets that are not commonly available for all the building types [7]. One of the significant challenges of UBEM is providing appropriate inputs for several parameters and reducing the uncertainties in the estimated energy demand. The studies indicated that the main reason for the challenges mentioned above is the lack of publicly available occupant-related data for different building types [27]. Besides, the available datasets, especially electricity loads, are aggregated data, and disaggregation is challenging.

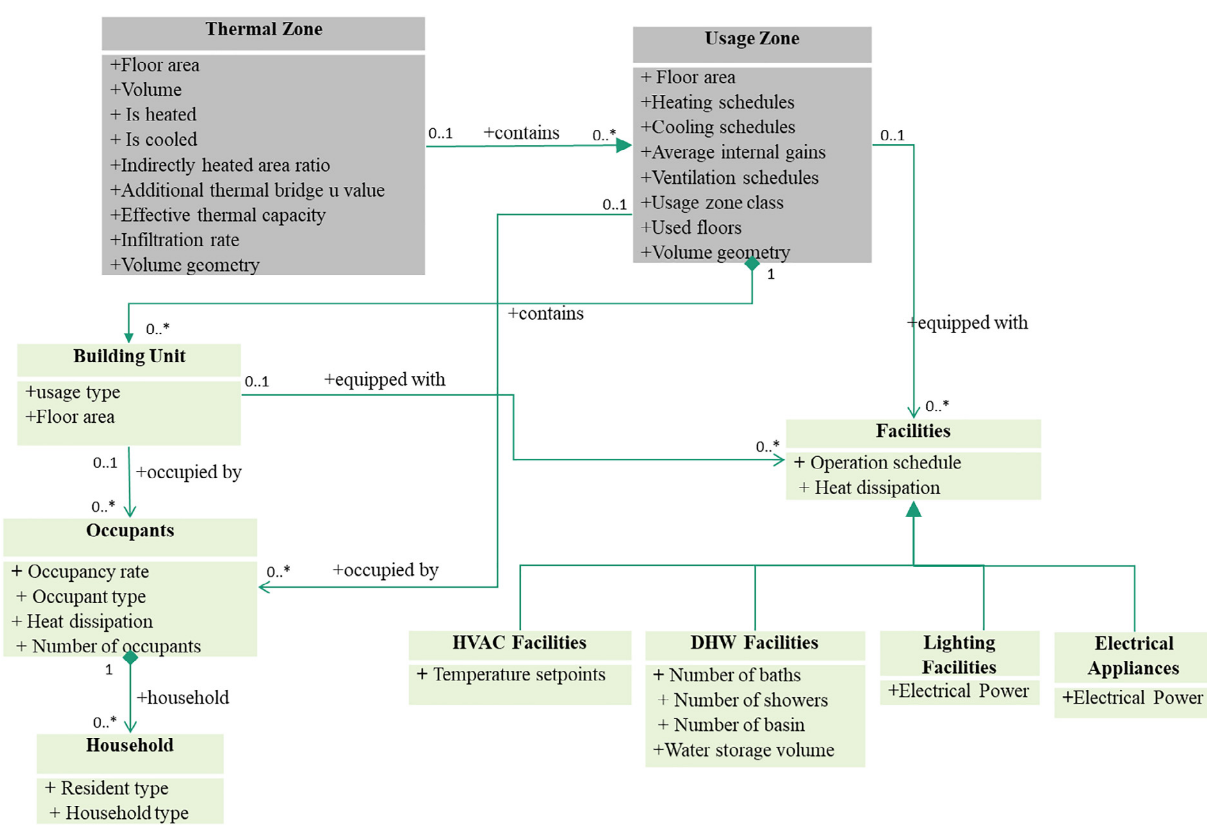

Fig. 11. Occupant behavior UML diagram.

In addition to the level of details regarding the occupancy, it is also important to discuss the OB-related data input methods to the simulation engines. The representation and implementation approach of the OB models to the energy simulation engine includes direct input, built-in OB models, user function or custom code, and co-simulation, which should be properly integrated with building performance simulation programs [23]. The definition of co-simulation, examples for co-simulation in BPS, workflow for co-simulation approach, and implementation of OB models in BPS are reviewed in detail in [23]. If the deterministic models are applied, the schedules and related parameters could be given as direct input. However, linking the stochastic OB models to UBEM needs a co-simulation to generate the occupancy and occupant-related schedules. The co-simulation approach allows the occupant-related schedules to be simulated by different simulation tools integrating with the BPS program [162]. For example, Hong et al. [163] developed an OB functional mockup unit (obFMU) to simulate the OB. In another study, Hong et al. [164] developed an occupant behavior XML (obXML) schema to implement the drivers, needs, actions, and systems (DNAS) [165] framework to represent OB in building simulation engines. In the study, the operation of the HVAC system and window blinds were provided as two examples to demonstrate the functionality of the developed obXML schema. Chen et al. [166] proposed an approach to analyze and visualize the OB in the office buildings using the obFMU through co-simulation with EnergyPlus. Another example of co-simulation is the StROBe, a residential human behavioral model that exploited the TUS data and household budget survey to model households' heating load, plug loads, DHW, and internal heat gains [35]. Occupancy chains and household members' activities are generated using survival analysis. The epistemic and aleatory uncertainties of the district energy simulations regarding non-deterministic OB have been addressed in the StROBe model [35].

6. Conclusion and discussions

The present study aimed at investigating the research works to extract the occupant-centric parameters for UBEM. Initially, an overview of UBEM is presented, and later, the focus is shifted towards summarizing the different occupancy modeling approaches, addressing their advantages and disadvantages, highlighting the correlation between occupancy and their energy use behaviors. The importance of occupant-centric UBEM is emphasized. Further, some potential urban datasets that can be used to develop detailed occupancy models are suggested. A brief discussion on occupant behavior UML diagram represents how to implement different occupant-related inputs in the UBEM is made. Finally, a discussion is made on the level of details and co-simulation required to represent OB models in UBEM.

OB modeling is a multi-faceted subject that requires an understanding of occupancy, occupant energy use behavior, and data from different building use types in temporal and spatial aspects. Previously, exploring the uncertainty in UBEM due to occupant-related inputs is side-lined in most research works because of the model complexity and lack of data availability. Subsequently, simple/fixed schedules were assigned for multiple buildings, resulting in a significant difference between the simulation and the measured data. In recent times, the installation of smart energy meters, thermostats, and BEMS has gained momentum and collected different data in buildings, including occupancy. Accordingly, research on occupant energy use behavior, developing occupancy models using different approaches, and implementing the model outputs in BPS have increased substantially. Specifically, if the objective of UBEM is to perform the peak load analysis and explore energy-efficient strategies, probabilistic occupancy models and ABM in BPS are recommended in the literature. However, most probabilistic models were developed for residential and office buildings. The models focused on other types of buildings are scarce. The other unspoken topic in the literature emphasizes the importance of defining harmonized schedules for occupant presence and occupant energy usage in energy simulation. Since the receptacle, lighting, thermostat setpoint temperature schedules, and DHW usage are highly correlated with occupant presence schedule (especially in residential buildings), assigning a coordinated input for occupancy and occupant energy use behavior is crucial BPS.

The literature review's main understanding is that each occupancy modeling approach has its advantages/disadvantages, and domain knowledge is crucial for adopting a suitable modeling approach based on the requirements. For instance, a simple deterministic modeling approach is sufficient to estimate the buildings' energy demand on a sub-yearly or yearly scale and design the energy systems' appropriate size. If the goal is to estimate the hourly energy demand and implement energy management strategies and demand response programs at the urban scale, more detailed occupancy models such as probabilistic approaches or ABM must be followed. However, very few studies explored the outcomes of deterministic, probabilistic stochastic occupancy models at an urban scale and compared the uncertainties in the simulation results. In this context, future research must be considered for UBEM results' uncertainties by employing different occupancy modeling approaches. This would provide guidelines for choosing suitable occupancy models or developing hybrid occupancy models by making a trade-off between the model complexity, accuracy, and data requirements. The other understanding from the literature is that the lack of sufficient occupant-related datasets is the major hurdle in developing a detailed occupancy model for UBEM. In this regard, the use of data collected from different sensing techniques and TUS data is important. Developing a generic data-driven model that systematically extracts the dynamic occupancy schedules from the collected data (from different building use types) is the other scope for future research in this field. These distinct schedules could be used later to develop stochastic-based occupancy models.

Declaration of Competing Interest

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

Acknowledgment

This research was undertaken, in part, thanks to funding from the Canada Excellence Research Chairs Program with grant number CERC-2018-00005.

References

[1] 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, Appl. Energy 205 (August) (2017) 323–335, https://doi.org/10.1016/j.apenergy.2017.07.128.

[2] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (8) (2009) 1819–1835, https://doi.org/10.1016/j.rser.2008.09.033.

[3] T. Hong, Y. Chen, X. Luo, N. Luo, S.H. Lee, Ten questions on urban building energy modeling, *Build. Environ.* 168(August 2019) 2020 106508, doi: 10.1016/j.buildenv.2019.106508.

[4] 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, Appl. Energy 225 (January) (2018) 1176–1189, https://doi.org/10.1016/j.apenergy.2018.05.023.

[5] C.S. Monteiro, C. Costa, A. Pina, M.Y. Santos, P. Ferrão, An urban building database (UBD) supporting a smart city information system, Energy Build. 158 (2018) 244–260, https://doi.org/10.1016/j.enbuild.2017.10.009.
[6] A. Alhamwi, W. Medjroubi, T. Vogt, C. Agert, GIS-based urban energy systems models and tools: introducing a model for the optimisation of flexibilisation technologies in urban areas, Appl. Energy 191 (2017) 1–9, https://doi.org/10.1016/j.apenergy.2017.01.048.
[7] C. Cerezo 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, https://doi.org/10.1016/j.energy.2016.10.057.
[8] O. Pasichnyi, J. Wallin, O. Kordas, Data-driven building archetypes for urban building energy modelling, Energy 181 (2019) 360–377, https://doi.org/10.1016/j.energy.2019.04.197.
[9] M. Zirak, V. Weiler, M. Hein, U. Eicker, Urban models enrichment for energy applications: challenges in energy simulation using different data sources for building age information, Energy 190 (2020) 116292, https://doi.org/10.1016/j.energy.2019.116292.
[10] A. Abdeen, W. O'Brien, B. Gunay, G. Newsham, H. Knudsen, Comparative review of occupant-related energy aspects of the National Building Code of Canada, Build. Environ. 183 (2020) 107136, https://doi.org/10.1016/j.buildenv.2020.107136.
[11] J. Keirstead, M. Jennings, A. Sivakumar, A review of urban energy system models: approaches, challenges and opportunities, Renew. Sustain. Energy Rev. 16 (6) (2012) 3847–3866, https://doi.org/10.1016/j.rser.2012.02.047.
[12] M. Mosteiro-Romero, A. Schlueter, Effects of occupants and local air temperatures as sources of stochastic uncertainty in district energy system modeling, *Energies* 14(8) 2021, doi: 10.3390/en14082295.
[13] J. An, D. Yan, T. Hong, Clustering and statistical analyses of air-conditioning intensity and use patterns in residential buildings, Energy Build. 174 (2018) 214–227, https://doi.org/10.1016/j.enbuild.2018.06.035.
[14] D. Yan, W. O'Brien, T. Hong, X. Feng, H.B. Gunay, F. Tahmasebi, A. Mahdavi, Occupant behavior modeling for building performance simulation: current state and future challenges, Energy Build. 107 (2015) 264–278, https://doi.org/10.1016/j.enbuild.2015.08.032.
[15] D. Yan, T. Hong, B. Dong, A. Mahdavi, S. D'Oca, I. Gaetani, X. Feng, IEA EBC Annex 66: definition and simulation of occupant behavior in buildings, Energy Build. 156 (2017) 258–270, https://doi.org/10.1016/j.enbuild.2017.09.084.
[16] S.J. Quan, Q. Li, G. Augenbroe, J. Brown, Urban data and building energy modeling: a GIS-based urban building energy modeling system using the urban-EPC engine, (September) 2015, doi: 10.1007/978-3-319-18368-8.
[17] V. Fabi, R.V. Andersen, S.P. Corgnati, B.W. Olesen, A methodology for modelling energy-related human behaviour: application to window opening behaviour in residential buildings, 2013 415–427, doi: 10.1007/s12273-013-0119-6.
[18] W. Tian, Y. Heo, P. de Wilde, Z. Li, D. Yan, C.S. Park, X. Feng, G. Augenbroe, A review of uncertainty analysis in building energy assessment, Renew. Sustain. Energy Rev. 93 (2018) 285–301, https://doi.org/10.1016/j.rser.2018.05.029.
[19] M.M. Romero, Stochastic sources of uncertainty in urban energy systems: occupancy and microclimate, PhD Thesis, ETH Zürich, 2020, doi: 10.3929/ethz-b-000438893.
[20] M. Jia, R.S. Srinivasan, A.A. Raheem, From occupancy to occupant behavior: an analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency, *Renew. Sustain. Energy Rev.* 68(June 2016) 2017 525–540, doi: 10.1016/j.rser.2016.10.011.
[21] H. Yoshino, T. Hong, N. Nord, IEA EBC annex 53: Total energy use in buildings — Analysis and evaluation methods, *Energy Build.* 152(March 2013) 2017 124–136, doi: 10.1016/j.enbuild.2017.07.038.
[22] W. O'Brien et al., Introducing IEA EBC annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation, Build. Environ. 178 (May) (2020), https://doi.org/10.1016/j.buildenv.2020.106738 106738.
[23] T. Hong, Y. Chen, Z. Belafi, S. D'Oca, Occupant behavior models: a critical review of implementation and representation approaches in building performance simulation programs, Build. Simul. 11 (1) (2018) 1–14, https://doi.org/10.1007/s12273-017-0396-6.
[24] J. Li, Z. Jerry, F. Haghighat, G. Zhang, Development and improvement of occupant behavior models towards realistic building performance simulation: a review, Sustain. Cities Soc. 50 (June) (2019), https://doi.org/10.1016/j.scs.2019.101685 101685.
[25] Y. Zhang, X. Bai, F.P. Mills, J.C.V. Pezzey, Rethinking the role of occupant behavior in building energy performance: a review, Energy Build. 172 (2018) 279–294, https://doi.org/10.1016/j.enbuild.2018.05.017.
[26] H. Hou, J. Pawlak, A. Sivakumar, B. Howard, J. Polak, An approach for building occupancy modelling considering the urban context, Build. Environ. 183 (2020) 107126, https://doi.org/10.1016/j.buildenv.2020.107126.
[27] G. Happle, J.A. Fonseca, A. Schlueter, Context-specific urban occupancy modeling using location-based services data, Build. Environ. 175 (2020) 106803, https://doi.org/10.1016/j.buildenv.2020.106803.
[28] W. Wu, B. Dong, Q. Wang, M. Kong, D. Yan, J. An, Y. Liu, A novel mobility-based approach to derive urban-scale building occupant profiles and analyze impacts on building energy consumption, Appl. Energy 278 (2020) 115656, https://doi.org/10.1016/j.apenergy.2020.115656.
[29] E. Barbour, C.C. Davila, S. Gupta, C. Reinhart, J. Kaur, M.C. González, Planning for sustainable cities by estimating building occupancy with mobile phones, *Nat. Commun.* 10(1) 2019, doi: 10.1038/s41467-019-11685-w.
[30] X. Kang, D. Yan, H. Sun, Y. Jin, P. Xu, An approach for obtaining and extracting occupancy patterns in buildings based on mobile positioning data, Build. Simul. Conf. Proc. 4 (2019) 2364–2370, https://doi.org/10.26868/25222708.2019.211151.
[31] J. Parker, A. Hardy, D. Glew, C. Gorse, A methodology for creating building energy model occupancy schedules using personal location metadata, Energy Build. 150 (2017) 211–223, https://doi.org/10.1016/j.enbuild.2017.06.014.
[32] C. Miller, F. Meggers, Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings, Energy Build. 156 (2017) 360–373, https://doi.org/10.1016/j.enbuild.2017.09.056.
[33] X. Xu, P.J. Culligan, J.E. Taylor, Energy Saving Alignment Strategy: achieving energy efficiency in urban buildings by matching occupant temperature preferences with a building ' s indoor thermal environment, Appl. Energy 123 (2014) 209–219, https://doi.org/10.1016/j.apenergy.2014.02.039.
[34] M. Mosteiro-Romero, J.A. Fonseca, A. Schlueter, Seasonal effects of input parameters in urban-scale building energy simulation, Energy Procedia 122 (2017) 433–438, https://doi.org/10.1016/j.egypro.2017.07.459.
[35] R. Baetens, D. Saelens, Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour, J. Build. Perform. Simul. 9 (4) (2016) 431–447, https://doi.org/10.1080/19401493.2015.1070203.
[36] U. Ali, M.H. Shamsi, C. Hoare, E. Mangina, J. O'Donnell, Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis, Energy Build. 246 (2021) 111073, https://doi.org/10.1016/j.enbuild.2021.111073.
[37] 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, *Sustain. Cities Soc.* 62(March) 2020 102408, 2020, doi: 10.1016/j.scs.2020.102408.
[38] G. Happle, J.A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy Build. 174 (2018) 276–292, https://doi.org/10.1016/j.enbuild.2018.06.030.
[39] F. Salim et al., Modelling urban-scale occupant behaviour, mobility, and energy in buildings: a survey, *Build. Environ.* 2020 106964, doi: 10.1016/j.buildenv.2020.106964.
[40] B. Dong, E. Azar, Occupant behavior modeling methods for resilient building design, operation and policy at urban scale: a review, *J. Appl. Energy (In Press.*, 293(March) 2021 116856, doi: 10.1016/j.apenergy.2021.116856.
[41] J. Allegrini, K. Orehounig, G. Mavromatidis, F. Ruesch, V. Dorer, R. Evins, A review of modelling approaches and tools for the simulation of district-scale energy systems, Renew. Sustain. Energy Rev. 52 (2015) 1391–1404, https://doi.org/10.1016/j.rser.2015.07.123.
[42] M.H. Kristensen, Urban building energy modelling for retrofit analysis under uncertainty, PhD Thesis, Aarhus University, Denmark, 2018, DOI: 10.7146/aul.301.208.
[43] S. Ferrari, F. Zagarella, P. Caputo, M. Bonomolo, Assessment of tools for urban energy planning, Energy 176 (2019) 544–551, https://doi.org/10.1016/j.energy.2019.04.054.
[44] 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, Build. Environ. 45 (7) (2010) 1683–1697, https://doi.org/10.1016/j.buildenv.2010.01.021.
[45] C.F. Reinhart, C. Cerezo Davila, Urban building energy modeling - A review of a nascent field, Build. Environ. 97 (2016) 196–202, https://doi.org/10.1016/j.buildenv.2015.12.001.
[46] W. Li, Y. Zhou, K. Cetin, J. Eom, Y.u. Wang, G. Chen, X. Zhang, Modeling urban building energy use: a review of modeling approaches and procedures, Energy 141 (2017) 2445–2457, https://doi.org/10.1016/j.energy.2017.11.071.
[47] G. Buttitta, W.J.N. Turner, O. Neu, D.P. Finn, Development of occupancy-integrated archetypes: use of data mining clustering techniques to embed occupant behaviour profiles in archetypes, Energy Build. 198 (2019) 84–99, https://doi.org/10.1016/j.enbuild.2019.05.056.
[48] S. Ferrari, F. Zagarella, P. Caputo, A. D'Amico, Results of a literature review on methods for estimating buildings energy demand at district level, Energy 175 (2019) 1130–1137, https://doi.org/10.1016/j.energy.2019.03.172.
[49] D. Robinson, N. Campbell, W. Gaiser, K. Kabel, A. Le-Mouel, N. Morel, J. Page, S. Stankovic, A. Stone, SUNtool - A new modelling paradigm for simulating and optimising urban sustainability, Sol. Energy 81 (9) (2007) 1196–1211, https://doi.org/10.1016/j.solener.2007.06.002.
[50] D. Robinson et al., Citysim: comprehensive micro-simulation of resource flows for sustainable urban planning, IBPSA 2009 – Int. Build. Perform. Simul. Assoc. 2009 (2009) 1083–1090.
[51] C.F. Reinhart, T. Dogan, J.A. Jakubiec, T. Rakha, A. Sang. UMI - an urban simulation environment for building energy use, daylighting, and walkability, Massachusetts Institute of Technology Department of Architecture, pp. 476–483, 2013.
[52] P. Remmen, M. Lauster, M. Mans, M. Fuchs, T. Osterhage, D. Müller, TEASER: an open tool for urban energy modelling of building stocks, J. Build. Perform. Simul. 11 (1) (2018) 84–98, https://doi.org/10.1080/19401493.2017.1283539.

[53] T. Hong, Y. Chen, S.H. Lee, M.P. Piette, Y. Chen, M.P. Piette, CityBES: A web-based platform to support city-scale building energy efficiency, 5th Int. Urban Comput. Work. San Fr. (August) (2016) 10.
[54] D. Wang, J. Landolt, G. Mavromatidis, K. Orehounig, J. Carmeliet, CESAR: A bottom-up building stock modelling tool for Switzerland to address sustainable energy transformation strategies, Energy Build. 169 (2018) 9–26, https://doi.org/10.1016/j.enbuild.2018.03.020.
[55] Y. Yamaguchi, Y. Shimoda, M. Mizuno, Development of district energy system simulation model based. Department of Environmental Engineering, Graduate School of Engineering, Osaka University System district, *Eighth Int. IBPSA Conf. Eindhoven, Netherlands*, pp. 1443–1450, 2003.
[56] Y. Shimoda, T. Asahi, A. Taniguchi, M. Mizuno, Evaluation of city-scale impact of residential energy conservation measures using the detailed end-use simulation model, Energy 32 (9) (2007) 1617–1633, https://doi.org/10.1016/j.energy.2007.01.007.
[57] S. Heiple, D.J. Sailor, Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles, Energy Build. 40 (8) (2008) 1426–1436, https://doi.org/10.1016/j.enbuild.2008.01.005.
[58] P. Caputo, G. Costa, S. Ferrari, A supporting method for defining energy strategies in the building sector at urban scale, Energy Policy 55 (2013) 261–270, https://doi.org/10.1016/j.enpol.2012.12.006.
[59] K. Orehounig, G. Mavromatidis, R. Evins, V. Dorer, J. Carmeliet, Towards an energy sustainable community: an energy system analysis for a village in Switzerland, Energy Build. 84 (2014) 277–286, https://doi.org/10.1016/j.enbuild.2014.08.012.
[60] L.A. Bollinger, R. Evins, Hues: a Holistic Urban Energy Simulation Platform for Effective Model Integration, *Cisbat*, pp. 841–846, 2015, [Online]. Available: https://infoscience.epfl.ch/record/213425/files/9_BOLLINGER.pdf.
[61] R. Nouvel, A. Mastrucci, U. Leopold, O. Baume, V. Coors, U. Eicker, Combining GIS-based statistical and engineering urban heat consumption models: towards a new framework for multi-scale policy support, Energy Build. 107 (2015) 204–212, https://doi.org/10.1016/j.enbuild.2015.08.021.
[62] 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 Build. 113 (2016) 202–226, https://doi.org/10.1016/j.enbuild.2015.11.055.
[63] P. Ellis, "A Parametric Tool for Community-Scale Modeling," *SimBuild 2016*, pp. 118–124, 2016, [Online]. Available: http://www.ibpsa-usa.org/index.php/ibpusa/article/download/348/334.
[64] J. Schiefelbein *et al.*, "Development of a to ol chain for complex city district energy system modeling and simulation," *14th Int. Conf. IBPSA - Build. Simul. 2015, BS 2015, Conf. Proc.*, no. December, pp. 1774–1781, 2015.
[65] J. Schiefelbein, A. Javadi, M. Lauster, P. Remmen, R. Streblow, "Development of a City Information Model to support Data Management and Analysis of Building Energy Systems within Complex City," *Proc. CISBAT 2015 - Futur. Build. Dist. - Sustain. from Nano to Urban Scale*, pp. 949–954, 2015, doi: 10.5075/epfl-cisbat2015-949-954.
[66] K. Ahmed, L.E. Ortiz, J.E. Gonzalez, On the spatio-temporal end-user energy demands of a dense urban environment, J. Sol. Energy Eng. Trans. ASME 139 (4) (2017) 1–11, https://doi.org/10.1115/1.4036545.
[67] J. Sokol, C. Cerezo Davila, C.F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, Energy Build. 134 (2017) 11–24, https://doi.org/10.1016/j.enbuild.2016.10.050.
[68] 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, https://doi.org/10.1016/j.energy.2017.07.151.
[69] J. An, D. Yan, T. Hong, K. Sun, A novel stochastic modeling method to simulate cooling loads in residential districts, Appl. Energy 206 (September) (2017) 134–149, https://doi.org/10.1016/j.apenergy.2017.08.038.
[70] S. Torabi Moghadam, J. Toniolo, G. Mutani, P. Lombardi, A GIS-statistical approach for assessing built environment energy use at urban scale, *Sustain. Cities Soc.* 37(March 2017) 2018 70–84, doi: 10.1016/j.scs.2017.10.002.
[71] R. El Kontar, T. Rakha, Profiling occupancy patterns to calibrate urban building energy models (UBEMs) using measured data clustering, Technol. Archit. Des. 2 (2) (2018) 206–217, https://doi.org/10.1080/24751448.2018.1497369.
[72] E. Fuentes, L. Arce, J. Salom, A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis, *Renew. Sustain. Energy Rev.* 81(February 2017) 2018 1530–1547, doi: 10.1016/j.rser.2017.05.229.
[73] T. Hong, J. Langevin, K. Sun, Building simulation: ten challenges, 2018 871–898.
[74] J. Page, D. Robinson, N. Morel, J.L. Scartezzini, A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2) (2008) 83–98, https://doi.org/10.1016/j.enbuild.2007.01.018.
[75] J. Page, D. Robinson, J.L. Scartezzini, Stochastic simulation of occupant presence and behaviour in buildings, *Solar Energy and Building Physics Laboratory*, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, pp. 757–764, 2007.
[76] J. Wang, W. Zuo, S. Huang, D. Vrabie, Data-driven prediction of occupant presence and lighting power-A case data-driven prediction of occupant presence and lighting power: a case study for small commercial buildings, (March) 2020.
[77] J. Zhao, B. Lasternas, K.P. Lam, R. Yun, V. Loftness, Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining, Energy Build. 82 (2014) 341–355, https://doi.org/10.1016/j.enbuild.2014.07.033.
[78] B. Huchuk, S. Sanner, W.O. Brien, Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data, *Build. Environ.* 160(March) 2019 106177, doi: 10.1016/j.buildenv.2019.106177.
[79] M. Mosteiro-Romero, I. Hischier, J.A. Fonseca, A. Schlueter, A novel population-based occupancy modeling approach for district-scale simulations compared to standard-based methods, Build. Environ. 181 (2020) 107084, https://doi.org/10.1016/j.buildenv.2020.107084.
[80] ASHRAE, "90.1 User's Manual ANSI/ASHRAE/IESNA Standard 90.1 - 2004," *ASHRAE Stand.*, p. 392, 2004.
[81] "National Energy Code of Canada for Buildings," Ottawa, Canada, 2017.
[82] S. Carlucci, M. De Simone, S.K. Firth, M.B. Kjærgaard, R. Markovic, M.S. Rahaman, M.K. Annaqeeb, S. Biandrate, A. Das, J.W. Dziedzic, G. Fajilla, M. Favero, M. Ferrando, J. Hahn, M. Han, Y. Peng, F. Salim, A. Schlüter, C. van Treeck, Modeling occupant behavior in buildings, Build. Environ. 174 (2020) 106768, https://doi.org/10.1016/j.buildenv.2020.106768.
[83] T. Hong, D. Yan, S. D'Oca, C. fei Chen, Ten questions concerning occupant behavior in buildings: the big picture, Build. Environ. 114 (2017) 518–530, https://doi.org/10.1016/j.buildenv.2016.12.006.
[84] J. Widén, A. Molin, K. Ellegård, Models of domestic occupancy, activities and energy use based on time-use data: deterministic and stochastic approaches with application to various building-related simulations, J. Build. Perform. Simul. 5 (1) (2012) 27–44, https://doi.org/10.1080/19401493.2010.532569.
[85] "Modeling building occupants in CEA – Part 1: Deterministic occupant model." https://cityenergyanalyst.com/blog/2019/7/11/modeling-building-occupants-in-the-cea-part-1-deterministic-schedules.
[86] J. Li, K. Panchabikesan, Z. Yu, F. Haghighat, M. El, D. Corgier, Systematic data mining-based framework to discover potential energy waste patterns in residential buildings, Energy Build. 199 (2019) 562–578, https://doi.org/10.1016/j.enbuild.2019.07.032.
[87] F. Xiao, C. Fan, Data mining in building automation system for improving building operational performance, Energy Build. 75 (2014) 109–118, https://doi.org/10.1016/j.enbuild.2014.02.005.
[88] J. Yang, C. Ning, C. Deb, F. Zhang, D. Cheong, S.E. Lee, C. Sekhar, K.W. Tham, k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement, Energy Build. 146 (2017) 27–37, https://doi.org/10.1016/j.enbuild.2017.03.071.
[89] K. Panchabikesan, F. Haghighat, M.E. Mankibi, Data driven occupancy information for energy simulation and energy use assessment in residential buildings, Energy 218 (2021) 119539, https://doi.org/10.1016/j.energy.2020.119539.
[90] Z. Jerry, F. Haghighat, B.C.M. Fung, E. Morofsky, H. Yoshino, A methodology for identifying and improving occupant behavior in residential buildings, Energy 36 (11) (2011) 6596–6608, https://doi.org/10.1016/j.energy.2011.09.002.
[91] S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408, https://doi.org/10.1016/j.enbuild.2014.11.065.
[92] Z. Jerry, F. Haghighat, B.C.M. Fung, L. Zhou, A novel methodology for knowledge discovery through mining associations between building operational data, Energy Build. 47 (2012) 430–440, https://doi.org/10.1016/j.enbuild.2011.12.018.
[93] J. Li, K. Panchabikesan, Z. Yu, F. Haghighat, M. El Mankibi, D. Corgier, Systematic data mining-based framework to discover potential energy waste patterns in residential buildings, Energy Build. 199 (2019) 562–578, https://doi.org/10.1016/j.enbuild.2019.07.032.
[94] B. Huchuk, S. Sanner, W. O'Brien, Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data, Build. Environ. 160 (2019) 106177, https://doi.org/10.1016/j.buildenv.2019.106177.
[95] B. Yang, F. Haghighat, B.C.M. Fung, K. Panchabikesan, Season-based occupancy prediction in residential buildings using machine learning models, *e-Prime*, 2021 100003, doi: 10.1016/j.prime.2021.100003.
[96] P. Anand, D. Cheong, C. Sekhar, M. Santamouris, S. Kondepudi, Energy saving estimation for plug and lighting load using occupancy analysis, Renew. Energy 143 (2019) 1143–1161, https://doi.org/10.1016/j.renene.2019.05.089.
[97] M. He, T. Lee, S. Taylor, S.K. Firth, K. Lomas, "Coupling a stochastic occupancy model to EnergyPlus to predict hourly thermal demand of a neighbourhood," School of Civil and Building Engineering , Loughborough University , Loughborough , UK," *Proc. Build. Simul. 2015*, pp. 2101–2108, 2015.
[98] S.H. Kim, H.J. Moon, Y.R. Yoon, "Improved occupancy detection accuracy using PIR and door sensors for a smart thermostat Department of Architectural Engineering , Dankook University, Yongin, Republic of Korea Abstract," pp. 2753–2758, 2017.
[99] N. Khalil, D. Benhaddou, O. Gnawali, J. Subhlok, Nonintrusive ultrasonic-based occupant identification for energy efficient smart building applications, *Appl. Energy* 220(August 2017) 2018 814–828, doi: 10.1016/j.apenergy.2018.03.018.
[100] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, F. Descamps, A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison, Build. Environ. 75 (2014) 67–78, https://doi.org/10.1016/j.buildenv.2014.01.021.

[101] M.B. Kjærgaard, B. Dong, S. Carlucci, F.D. Salim, C.J. Andrews, Poster abstract: data-driven occupant modeling strategies and digital tools enabled by IEA EBC Annex 79, (2), 188–189.
[102] C. Bianchi, L. Zhang, D. Goldwasser, A. Parker, H. Horsey, Modeling occupancy-driven building loads for large and diversified building stocks through the use of parametric schedules, *Appl. Energy* 276(March) 2020 115470, doi: 10.1016/j.apenergy.2020.115470.
[103] F. Causone, S. Carlucci, M. Ferrando, A. Marchenko, S. Erba, A data-driven procedure to model occupancy and occupant-related electric load profiles in residential buildings for energy simulation, Energy Build. 202 (2019), https://doi.org/10.1016/j.enbuild.2019.109342 109342.
[104] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, Build. Simul. 4 (2) (2011) 149–167, https://doi.org/10.1007/s12273-011-0044-5.
[105] C. Wang, D. Yan, H. Sun, Y. Jiang, A generalized probabilistic formula relating occupant behavior to environmental conditions, Build. Environ. 95 (2016) 53–62, https://doi.org/10.1016/j.buildenv.2015.09.004.
[106] F.R. Cecconi, M. Manfren, L.C. Tagliabue, A.L.C. Ciribini, E. De Angelis, Probabilistic behavioral modeling in building performance simulation: a Monte Carlo approach, Energy Build. 148 (2017) 128–141, https://doi.org/10.1016/j.enbuild.2017.05.013.
[107] J. Virote, R. Neves-Silva, Stochastic models for building energy prediction based on occupant behavior assessment, Energy Build. 53 (2012) 183–193, https://doi.org/10.1016/j.enbuild.2012.06.001.
[108] D. Yan, W. O'Brien, T. Hong, X. Feng, H. Burak Gunay, F. Tahmasebi, A. Mahdavi, Occupant behavior modeling for building performance simulation: current state and future challenges, Energy Build. 107 (2015) 264–278, https://doi.org/10.1016/j.enbuild.2015.08.032.
[109] S. Gilani, W. O'Brien, H.B. Gunay, Simulating occupants' impact on building energy performance at different spatial scales, Build. Environ. 132 (January) (2018) 327–337, https://doi.org/10.1016/j.buildenv.2018.01.040.
[110] I. Richardson, M. Thomson, D. Infield, A high-resolution domestic building occupancy model for energy demand simulations, Energy Build. 40 (8) (2008) 1560–1566, https://doi.org/10.1016/j.enbuild.2008.02.006.
[111] Y.S. Lee, A.M. Malkawi, Simulating multiple occupant behaviors in buildings: an agent-based modeling approach, Energy Build. 69 (2014) 407–416, https://doi.org/10.1016/j.enbuild.2013.11.020.
[112] C. Liao, Y. Lin, P. Barooah, Agent-based and graphical modelling of building occupancy, J. Build. Perform. Simul. 5 (1) (2012) 5–25, https://doi.org/10.1080/19401493.2010.531143.
[113] C. Liao, P. Barooah, An integrated approach to occupancy modeling and estimation in commercial buildings, *Proc. 2010 Am. Control Conf. ACC 2010*, pp. 3130–3135, 2010, doi: 10.1109/acc.2010.5531035.
[114] M.B. Kjærgaard *et al.*, Current practices and infrastructure for open data based research on occupant-centric design and operation of buildings, *Build. Environ.* 177(March) 2020 106848, doi: 10.1016/j.buildenv.2020.106848.
[115] A. Mahdavi, A. Mohammadi, E. Kabir, Occupants ' operation of lighting and shading systems in office buildings, 1493, 2008, doi: 10.1080/19401490801906502.
[116] C.F. Reinhart, Lightswitch-2002: a model for manual and automated control of electric lighting and blinds, 77 2004 15–28, doi: 10.1016/j.solener.2004.04.003.
[117] C.M. Stoppel, F. Leite, Integrating probabilistic methods for describing occupant presence with building energy simulation models, Energy Build. 68 (PARTA) (2014) 99–107, https://doi.org/10.1016/j.enbuild.2013.08.042.
[118] X. Liang, T. Hong, G. Qiping, Occupancy data analytics and prediction: a case study, Build. Environ. 102 (2016) 179–192, https://doi.org/10.1016/j.buildenv.2016.03.027.
[119] S.D. Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408, https://doi.org/10.1016/j.enbuild.2014.11.065.
[120] D.M. Franco-Duran, G. Mejia A, Construction research congress 2016 2039, *Proc. Constr. Res. Congr. 2016* 2016 2039–2049, doi: 10.1061/9780784479827.203.
[121] S. Spiegel, "Optimization of In-House Energy Demand," Chapter 10, no. January, 2015.
[122] G. Buttitta, W.J.N. Turner, O. Neu, D.P. Finn, Development of occupancy-integrated archetypes : Use of data mining clustering techniques to emb e d occupant b ehaviour profiles in archetypes, Energy Build. 198 (2019) 84–99, https://doi.org/10.1016/j.enbuild.2019.05.056.
[123] I. Richardson, M. Thomson, D. Infield, C. Clifford, Domestic electricity use: a high-resolution energy demand model, Energy Build. 42 (10) (2010) 1878–1887, https://doi.org/10.1016/j.enbuild.2010.05.023.
[124] M. Haq, M.Y. Hassan, H. Abdullah, H.A. Rahman, M.P. Abdullah, F. Hussin, D. M. Said, A review on lighting control technologies in commercial buildings, their performance and affecting factors, Renew. Sustain. Energy Rev. 33 (2014) 268–279, https://doi.org/10.1016/j.rser.2014.01.090.
[125] X. Zhou, D. Yan, T. Hong, X. Ren, Data analysis and stochastic modeling of lighting energy use in large office buildings in China, Energy Build. 86 (2015) 275–287, https://doi.org/10.1016/j.enbuild.2014.09.071.
[126] I. Richardson, M. Thomson, D. Infield, A. Delahunty, Domestic lighting: a high-resolution energy demand model, Energy Build. 41 (7) (2009) 781–789, https://doi.org/10.1016/j.enbuild.2009.02.010.
[127] S. Wei, R. Jones, P. De Wilde, Driving factors for occupant-controlled space heating in residential buildings, Energy Build. 70 (2014) 36–44, https://doi.org/10.1016/j.enbuild.2013.11.001.
[128] C. Booten *et al.*, Residential Indoor Temperature Study Residential Indoor Temperature Study, Technical report, NREL/TP-5500-68019, April 2017.
[129] H.B. Gunay, W. Shen, G. Newsham, A. Ashouri, Modelling and analysis of unsolicited temperature setpoint change requests in office buildings, *Build. Environ.* 133(December 2017) 2018 203–212, doi: 10.1016/j.buildenv.2018.02.025.
[130] K. Panchabikesan, M. Ouf, U. Eicker, Investigating thermostat setpoint preferences in Canadian households, *Proc. Bs2021, IBPSA*, 2021.
[131] H.B. Gunay *et al.*, Development and implementation of a thermostat learning algorithm Development and implementation of a thermostat learning algorithm, 4731, 2018, doi: 10.1080/23744731.2017.1328956.
[132] V. Fabi, R.V. Andersen, S.P. Corgnati, Influence of occupant's heating set-point preferences on indoor environmental quality and heating demand in residential buildings, HVAC R Res. 19 (5) (2013) 635–645, https://doi.org/10.1080/10789669.2013.789372.
[133] Y. Peng, Z. Nagy, A. Schlüter, Temperature-preference learning with neural networks for occupant-centric building indoor climate controls, Build. Environ. 154 (January) (2019) 296–308, https://doi.org/10.1016/j.buildenv.2019.01.036.
[134] X. Ren, D. Yan, T. Hong, Data Mining of Space Heating System Performance in Affordable Housing, pp. 1–30.
[135] D. Fischer, T. Wolf, J. Scherer, B. Wille-Haussmann, A stochastic bottom-up model for space heating and domestic hot water load profiles for German households, Energy Build. 124 (2016) 120–128, https://doi.org/10.1016/j.enbuild.2016.04.069.
[136] L. Gelažanskas, K.A.A. Gamage, Forecasting hot water consumption in residential houses, Energies 8 (11) (2015) 12702–12717, https://doi.org/10.3390/en81112336.
[137] V. Bakker, A. Molderink, J.L. Hurink, G.J.M. Smit, Domestic heat demand prediction using neural networks, *Proc. 19th Int. Conf. Syst. Eng. ICSEng 2008*, pp. 189–194, 2008, doi: 10.1109/ICSEng.2008.51.
[138] J. de Santiago, O. Rodriguez-Vialon, B. Sicre, The generation of domestic hot water load profiles in Swiss residential buildings through statistical predictions, Energy Build. 141 (2017) 341–348, https://doi.org/10.1016/j.enbuild.2017.02.045.
[139] J.A. Fonseca, A. Schlueter, Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts, Appl. Energy 142 (2015) 247–265, https://doi.org/10.1016/j.apenergy.2014.12.068.
[140] A. Malhotra, J. Bischof, A. Nichersu, K.-H. Häfele, J. Exenberger, D. Sood, J. Allan, J. Frisch, C. van Treeck, J. O'Donnell, G. Schweiger, Information modelling for urban building energy simulation—A taxonomic review, Build. Environ. (2021) 108552, https://doi.org/10.1016/j.buildenv.2021.108552.
[141] "CityGML Energy Application Domain Extension," Guideline, 2016.
[142] R. Nouvel, R. Kaden, J. Bahu, Genesis of the CityGML Energy ADE, (September) 2015, doi: 10.5075/epfl-cisbat2015-931-936.
[143] J. Yang, M. Santamouris, S.E. Lee, Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings, Energy Build. 121 (2016) 344–349, https://doi.org/10.1016/j.enbuild.2015.12.019.
[144] M.S. Gul, S. Patidar, Understanding the energy consumption and occupancy of a multi-purpose academic building, Energy Build. 87 (2015) 155–165, https://doi.org/10.1016/j.enbuild.2014.11.027.
[145] ATUS, "American Time Use Survey," *United States Dep. Labor. Bur. Labor Stat.*, 2015.
[146] L. Diao, Y. Sun, Z. Chen, J. Chen, Modeling energy consumption in residential buildings: a bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation, Energy Build. 147 (2017) 47–66, https://doi.org/10.1016/j.enbuild.2017.04.072.
[147] D. Mitra, N. Steinmetz, Y. Chu, K.S. Cetin, Energy & Buildings Typical occupancy profiles and behaviors in residential buildings in the United States, Energy Build. 210 (2020) 109713, https://doi.org/10.1016/j.enbuild.2019.109713.
[148] A. Wagner, B. Dong, W.O. Brien, *Exploring Occupant Behavior in Buildings, Methods and Challenges*, 2018. doi: 10.1007/978-3-319-61464-9.
[149] X. Feng, D. Yan, T. Hong, Simulation of occupancy in buildings, Energy Build. 87 (2015) 348–359, https://doi.org/10.1016/j.enbuild.2014.11.067.
[150] "Ecobee Inc 'Donate Your Data,'" 2018. https://www.ecobee.com/donateyourdata/.
[151] "P.S. Dataport, Pecan street dataport," 2016. https://dataport.pecanstreet.org/.
[152] A.R.O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt, Dataport, nilmtk, A building data set designed for non-intrusive load monitoring, in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015, pp. 210–214. 58.
[153] J.H. Yoon, R. Bladick, A. Novoselac, Demand response for residential buildings based on dynamic price of electricity, Energy Build. 80 (2014) 531–541, https://doi.org/10.1016/j.enbuild.2014.05.002.
[154] C. Miller, B. Picchetti, P. Arjunan, J.Y. Park, The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition, (August), 2020.

[155] A. Chong, G. Augenbroe, D. Yan, Occupancy data at different spatial resolutions: building energy performance and model calibration, (August), 2020.
[156] X. Lu, F. Feng, Z. Pang, T. Yang, Extracting typical occupancy schedules from social media (TOSSM) and its integration with building energy modeling, 2019.
[157] X.L.H. Wang, G. Wang, Image-based occupancy positioning system using pose-estimation model for demand-oriented ventilation, 39(January) 2021, doi: 10.1016/j.jobe.2021.102220.
[158] Y. Meng, T. Li, G. Liu, S. Xu, T. Ji, Real-time dynamic estimation of occupancy load and an air-conditioning predictive control method based on image information fusion, *Build. Environ.* 173(October 2019) 2020 106741, doi: 10.1016/j.buildenv.2020.106741.
[159] S. Petersen, T.H. Pedersen, K.U. Nielsen, M.D. Knudsen, Establishing an image-based ground truth for validation of sensor data-based room occupancy detection, Energy Build. 130 (2016) 787–793, https://doi.org/10.1016/j.enbuild.2016.09.009.
[160] G. Langlois *et al.*, "A data model for simulation models relying on spatio-temporal urban data To cite this version: HAL Id: hal-01521335 A data model for simulation models relying on spatio-temporal urban data," 2017.
[161] G. Agugiaro, J. Benner, P. Cipriano, R. Nouvel, The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations, *Open Geospatial Data Softw. Stand.* 3(1) 2018, doi: 10.1186/s40965-018-0042-y.
[162] T. Nouidui, M. Wetter, W. Zuo, Functional mock-up unit for co-simulation import in EnergyPlus, 1493 2014, doi: 10.1080/19401493.2013.808265.
[163] T. Hong, H. Sun, Y. Chen, S.C. Taylor-Lange, D. Yan, An occupant behavior modeling tool for co-simulation, Energy Build. 117 (2016) 272–281, https://doi.org/10.1016/j.enbuild.2015.10.033.
[164] T. Hong, S.D. Oca, S.C. Taylor-lange, W.J.N. Turner, Y. Chen, S.P. Corgnati, An ontology to represent energy-related occupant behavior in buildings. Part II: implementation of the DNAS framework using an XML schema, Build. Environ. 94 (2015) 196–205, https://doi.org/10.1016/j.buildenv.2015.08.006.
[165] T. Hong, S.D. Oca, W.J.N. Turner, S.C. Taylor-lange, An ontology to represent energy-related occupant behavior in buildings. Part I: introduction to the DNAs framework, Build. Environ. 92 (2015) 764–777, https://doi.org/10.1016/j.buildenv.2015.02.019.
[166] Y. Chen, X. Liang, T. Hong, X. Luo, Simulation and visualization of energy-related occupant behavior in office buildings, Build. Simul. 10 (6) (2017) 785–798, https://doi.org/10.1007/s12273-017-0355-2.