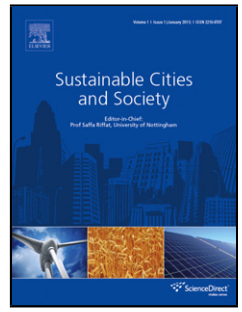
Accepted Manuscript



Title: Development and improvement of occupant behavior models towards realistic building performance simulation: A review

Authors: Jun Li, Zhun (Jerry) Yu, Fariborz Haghighat, Guoqiang Zhang

PII: S2210-6707(19)30377-4
DOI: https://doi.org/10.1016/j.scs.2019.101685
Article Number: 101685

Reference: SCS 101685

To appear in:

Received date: 9 February 2019
Revised date: 20 June 2019
Accepted date: 25 June 2019

Please cite this article as: Li J, Yu Z(Jerry), Haghighat F, Zhang G, Development and improvement of occupant behavior models towards realistic building performance simulation: A review, *Sustainable Cities and Society* (2019), https://doi.org/10.1016/j.scs.2019.101685

Development and improvement of occupant behavior models towards realistic building performance simulation: A review

Jun Li [a, b], Zhun (Jerry) Yu [a, b, *], Fariborz Haghighat [c], Guoqiang Zhang [a, b]

[a] College of Civil Engineering, National Center for International Research Collaboration in Building Safety and Environment, Hunan University, Changsha, Hunan 410082, China

[b] Collaborative Innovation Center of Building Energy Conservation & Environmental Control, Hunan 412007, China

[c] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec, H3G 1M8, Canada

Highlights:

- Uncertainty about impacts of occupant behavior on various spatial scales and temporal granularities.
- Modeling methods for occupant behavior and respective selection issues.
- Model input and output selection for occupant behavior model improvement.
- Research gaps and future directions for occupant behavior modeling.

Abstract

With the rise of concern about newly-designed or retrofitted buildings to have robust performance under different realistic scenarios, it is of vital importance to providing reliable energy predictions for building design and planning. Occupant behavior (OB), as one source of the significant uncertainties, is generally oversimplified as static schedules or predetermined inputs, which could cause a significant gap between the simulated and measured one. To bridge such gap, growing interests have been raised to understand the role of OB on building energy performance and develop OB models which can be integrated into building simulation tools. This paper aims to provide a systematic review with the focus on three important issues: a) the impact uncertainty caused by OB in building performance simulation and their differences in various spatial scales and temporal granularities; b) main criteria for the comparison and selection of modeling methods; c) requisite considerations to improve the performance of OB models. Based on this review, a framework was proposed towards improving the predictive performance of future OB models. Existing research gaps and key challenges for OB modeling are identified and future directions in this area are highlighted.

Keywords: Occupant behavior; Model; Building energy demand; Simulation; Uncertainty

1. Introduction

1.1 Background

The building sector possesses huge potentials for efficiency gains and greenhouse gas emission reduction, so as to positively contribute to global climate change. To ensure the high energy-efficiency and decarbonization of buildings, various innovative solutions have been recently proposed such as nearly zero-energy buildings (nZEBs) [1, 2], deep renovation of existing buildings [3] and smart buildings[4, 5]. For example, in EU, all new buildings are required to be nZEBs by 2020, under the Energy Performance of Buildings Directive (EPBD) [6]. According to previous studies, the building sector is still far from the expectation to be high energy-performance though the use of energy-conservation and low-carbon technologies [7]. For instance, the energy consumption of 196 apartments in two similar high-performance buildings varies greatly and some of them is higher than the desired performance due to occupants' behavior [8]. To foster building's energy efficiency, it is critical for buildings being designed or retrofitted to have robust performance under real and variable scenarios, particularly considering the large variability of occupant behavior.

Building performance simulation (BPS) tools are powerful techniques to predict building energy performance. An unsolved performance gap issue (i.e. discrepancy between predicted and actual performance), however, creates significant barriers in the effectiveness and reliability of BPS in producing accurate prediction results. One possible reason is the deterministic predictions of energy performance by using BPS tools in current practices. To be more specific, many input parameters inherent in BPS are assumed imprecisely or oversimplified and the variety of uncertainties is often ignored. Occupant behavior (OB), therefore, is represented by standardized schedules or predefined inputs which are oversimplified versions of complex reality. Such a deterministic simulation results in considerable difficulties for practitioners and stakeholders to make more rational decisions on selecting energy saving technologies especially in cost-optimal analysis and life cycle analysis [9, 10]. Accordingly, a paradigm shift in current deterministic simulation practices to a probabilistic form which can address the impact of various uncertainties (especially OB) in BPS are attracting much attention in recent years [11, 12].

Due to the increasing importance of OB, research community has placed substantial efforts to model the process of energy-related occupant behavior which can be viewed as solutions to address the uncertainty caused by the simplification of OB in BPS. For instance, the International Energy Agency (IEA) proposed a project Annex 66-'Definition and Simulation of Occupant Behavior in Buildings', aiming at accurately quantifying occupant behavior in a standard way [13, 14]. During the last two decades, various OB models have been developed to mimic the random nature of OB and generate stochastic and high-resolution OB profiles. In spite of significant improvements, the wide implementation of such OB models in BPS is limited as the robustness of these models is still under investigation.

As the most emerging expectations are to predict building energy performance in a reliable manner, understanding the role of OB and creating more robust OB models might remain a major issue for the following years. Therefore, more comprehensive and critical research and review works are required in order to address the existing challenges in modeling OB and to consider their dynamic nature during the building/district energy prediction and simulation.

1.2 Existing reviews and research gaps

Several reviews relevant to OB and its modeling have been conducted in recent years. For example, Yan et al. [15] conducted an inclusive review on occupant monitoring and data collection, model development, model evaluation and implementation. Gunay et al. [16] critically reviewed observational studies, modeling and simulation methods for adaptive behaviors. These studies mainly focus on statistical and stochastic methods while data mining and agent-based models were rarely included. Wei et al. [17] identified key drivers for space heating behavior in residential buildings and discussed methods of modeling space heating behavior in building simulation tools. Jia et al. [18] reviewed the studies from a unified view of 'sensing, modeling and coupling' and discussed the advantages and limitations of modeling methods from the view of building scale. Hong et al. [19] specifically focused on the approaches to represent OB in major building performance simulation tools. Stazi et al. [20] focused on the identification of important environmental factors and time-related events for six categories of behavior, and reviewed the commonly used variables in existing OB models. Gilani and O'Brien [21] critically reviewed the existing monitoring approaches and their future possibilities to facilitate OB modeling. Zhang et al. [22] conducted a review with particular focus on the understanding of OB, data collection methods, quantitative modeling methods and respective energy-saving potentials. Happle et al. [23] presented an overview of the existing modeling strategies of OB in urban building energy models and suggested a multi-agent approach for urban OB modeling.

Despite the importance of these efforts, there still remains several research gaps that are needed to be addressed. First, with the substantial interests in district and urban energy modeling, understanding the extent to which OB can affect building energy performance at different spatial scales is imperative for the development of OB models. Existing reviews were primarily carried out from the perspective in a single zone or building, while their impacts in large-scale simulation (e.g. building stocks, districts and cities) have not been well-understood. Second, different types of OB models have been established and each has its own strengths and limitations. A major problem faced by modelers and practitioners is to select suitable modeling methods with the highest potential for meeting their own requirements in practice. However, there is not yet a set of uniform and common criteria available for supporting such selection. In response, identifying model selection criteria that are of a rational and systematic basis is urgently needed based on a thorough understanding of existing modeling methods. Third, some technical details in OB model construction profoundly affect the model performance while they have seldom been fully addressed in previous reviews yet. An effort is necessary to understand and discuss the considerations in terms of these technical details that are directly/indirectly involved in the process of OB model development.

1.3 Scope of this review

Occupants generally influence their built environment through their impacts of presence (e.g. heat release and carbon dioxide emissions) and their interactions with buildings (e.g. building envelopes, systems and appliances) [24]. In this view, Schweiker et al. [24] sorted all energy-related behavior into four categories: 1) physiological adjustments regarding occupants' unconscious controls such as sweating and shivering, 2) individual adjustments such as clothing changes, 3) environmental adjustments broadly including all possible behavior influencing indoor environment conditions and use of systems and appliances, and 4) occupancy involves occupants' presence and spatial movement in buildings. Note that occupancy, particularly occupants' presence, can be viewed as a prerequisite of any behavior. Hence, the scope of this paper is environmental and individual adjustments for OB modeling purposes, while studies on psychological adjustments and occupancy models are not included.

1.4 Objective

To address the abovementioned gaps, this paper critically reviewed current research efforts in OB modeling with an emphasis on the following topics: a) impact quantification of OB on building energy performance in terms of different spatial and temporal scales and difficulties of addressing OB-related uncertainties for large-scale simulation (Section 2); b) establishment of a set of criteria for the rational selection from existing OB models (Section 3 and Section 5); c) technical issues and key challenges of model input and output selection, consideration of occupant diversity and comprehensive model evaluation, so as to the improvements of OB models (Section 4). Moreover, an integrated framework is proposed to help understand the whole process of OB model development so as to improve model performance. Such a review aims to assist modelers to develop more robust OB models with proper selection of modeling methods and requisite considerations in terms of the technical details. It can also help the building research community and building engineers to rethink the role of sophisticated OB models in building energy simulation.

2. Review methodology

In order to comprehensively review up-to-date research efforts to address the above said research gaps, a three-step procedure was taken in this review. At first, an extensive keyword-based search of relevant studies with regard to occupant behavior modeling in buildings was carried out using the following academic databases: Web of Science, ScienceDirect, Springer, Scopus and Google Scholar. In this step, examples of keywords used in this study are: occupant behavior, user behavior, modeling, uncertainty, impact, building performance simulation. Secondly, journals with an impact factor above 1.5 or highly endorsed by experts were selected. Through these two steps, more than 200 journals and conference papers were found. Since this research is not intended to review all the topics related with occupant behavior in buildings, as some of them have been addressed in prior reviews. Instead, we primarily focus on the abovementioned research gaps to critically review the impact of occupant behavior, existing models and approaches to improve model performance. Hence, at the third step, irrelevant articles were eliminated after going through the abstract using domain knowledge.

3. Impact of occupant behavior on building energy performance

Previously, many in-situ monitoring studies investigated the energy use in similar buildings with (nearly) identical structure and environment. For example, a monitoring case in high-performance social housing buildings observed that energy use between different dwellings varies between 54 and 273 kWh/m^2 [25]. Gill et al. [26] stated that OB contributes 51%, 37% and 11% of variance in heating, electricity and water consumption. These results confirmed the large deviation of building energy performance caused by OB in real buildings which implies the importance of considering the uncertainty of OB and quantify such uncertainty on BPS. With the rapid advancement of computation capability, conducting a large number of parametric simulations becomes feasible within an acceptable computational burden. Hence, uncertainty analysis (UA) and sensitivity analysis (SA) have been widely used in BPS to address different sources of uncertain input parameters (e.g. weather conditions, thermal properties of building envelope and OB). In this section, the impact quantification of OB in building energy performance as well as the differences of such impacts on different spatial and temporal scales are reviewed and some suggestions are provided.

3.1 Approaches for impact quantification

UA is a method of analyzing the response of simulation outputs along with the possible variation of input parameters. SA, however, primarily focuses on identifying the order of most influential input parameters in terms of their contribution to the variation of simulation results. Though these two methods originate from different disciplines, SA is normally performed in combination UA where the output of UA can be directly used as the input for SA. In building performance analysis, UA and SA have been deployed to find potentially economic solutions for energy efficiency and thermal environment improvement. For instance, Belleri et al [27] addressed the key design parameters of natural ventilation systems in the early-design stage for avoiding overheating risk. Winkler and Munk [28] examined the sensitivity of indoor humidity to changes in OB and air-conditioning control and found that a reduction of cooling supply airflow rates and a cooling blower-off delay would be a good solution rather than a simply lowering of cooling set points. To facilitate the usage of UA and SA in building energy assessment, Mavromatidis et al. [29] proposed a general procedure incorporated with UA and SA, as shown in Fig.1.

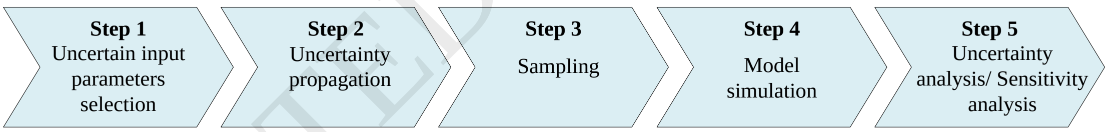


Fig. 1. A general procedure incorporated with UA and SA in building energy assessment [29]

In prior studies, weather, building and occupancy-related parameters (e.g. internal gains and presence) were the main focus, while OB- related parameters (e.g. blind, window and light control) were either ignored or assumed to be fixed scenarios [30]. Until now, few research works accounted for OB when applying UA and SA in BPS. In particular, their solutions for the investigation of OB impacts are to analyze the changes in performance metrics (e.g. annual total and peak heating load) along with the alteration of OB-related parameters (e.g. *met* value [31] and schedules [32, 33]) by: 1) assuming a number of given conditions with different value settings or a probability distribution to each parameter; 2) using generated profiles from OB models [34]. These studies indicated that OB resulted in significant variations in energy demand and indoor environment, indoor environment and thermal comfort (e.g. positive and negative thermal sensation [31]), ranging from 23.6% to 65% [31, 35].. Table 1 summarizes some studies that investigated the impact of OB on BPS. It should be noted that, though there are many studies that use UA and SA in building energy analysis, the studies that primarily consider the impact of OB were summarized in Table 1.

Table 1 Reviewed studies focusing on the impact quantification of OB on building energy performance

Ref.	Type of OB	Use of OB model	UA *	SA **	Performance metrics		
					Energy	IAQ/TC ***	Others
[36]	Use of blinds and lighting Occupancy Internal gain	Y	UA1	S2.1	Annual heating and cooling energy demand; Primary energy use;	PMV; Max. and min. room temperature	-
[37]	Window use Shade use	Y	-	-	Heating and cooling demand	PMV	-
[27]	Window use, Internal gains	-	UA1.1	SA2.1	-	Air change rate per hour	window opening factor
[31]	Blinds Lights Window, Temperature set point Fan and clothing	-	-	SA1	Heating and total energy demand	Thermal sensation	-
[38]	HVAC use Occupancy	Y	-	-	Total energy use	-	-
[39]	Occupancy Window shade use Lighting use	Y	-	-	Lighting use	-	-

[40]	Occupancy, Internal gains		UA1.1	-	Heating and cooling load	-	-
[32]	Window; Light; equipment	-	UA1.1	SA2.1	Energy consumption for air conditioning	-	Degree-hours for heating and cooling
[41]	Thermostat level, Ventilation behavior Metabolic rate Clothing Presence	-	UA1.1	SA2.3	Annual heating energy consumption	PMV	-
[33]	Thermostat set points Occupancy, Light use	-	UA1.1	SA2.1& SA2.2	Annual and peak facility electricity consumption, annual and peak HVAC electricity consumption	-	-
[28]	Cooling set point Air-conditioner configuration, controls	-	-	SA1		Indoor humidity	
[42]	Occupancy Window behavior, DHW Electrical appliances use Light use	Y	-	-	Heating and cooling demand	Temperature and CO2	Power mismatch factor
[43]	Occupancy Cooling temperature set point Lighting control Window use HVAC control	Y	UA1.1	-	Total and peak cooling load Load distribution	-	-
[44]	DHW use schedule		UA1.1	SA2.1& SA2.2	Peak/annual whole building water consumptions, DHW system water consumption, DHW system gas consumptions, and DHW system electricity consumptions	-	-

[45]	Cooling set point; Internal gain	-	-	SA1	Peak total cooling and dehumidification loads	-	-
[46]	Occupancy, use of lights, window shades, operable windows, plug-in equipment, and thermostats	Y	-	-	Total natural gas, peak heating load, total electricity use, and peak cooling load	-	-

Note: * There are two categories of UA methods, i.e. probabilistic (UA1) and non-probabilistic s (UA2). The probabilistic methods can further divided into sampling-based (UA1.1) and non-sampling based (UA1.2).

** Local sensitivity analysis (SA1) and global sensitivity analysis are two main types of SA methods. The global sensitivity analysis includes regression-based (SA2.1), screening-based method (SA2.2), variance-based method (SA2.3); meta-model based method (SA2.4). Details with respect to categories of UA and SA can be found in [29, 47, 48].

*** IAQ and TC are the abbreviations of indoor air quality and thermal comfort respectively.

**** 'Y' is the abbreviation of yes which indicates if the OB models is used in uncertainty and sensitivity analysis.

In Table 1, UA employed in most studies are probabilistic and sampling-based approaches in which the uncertainty was represented by a single probability distribution, consequently combining both natural variability (i.e. aleatory uncertainty) and the uncertainty deriving from the lack of knowledge (i.e. epistemic uncertainty). However, some uncertainties are due to the inappropriate assumptions for OB parameters which is epistemic instead of aleatory. For instance, O'Neill and Niu [33], Pang and O'Neill [44] argued that OB-related parameters employed in these studies are time-independent and thus the temporal variation of OB cannot be reflected. To address this issue, some research efforts [33, 44] attempted to apply a Karhunen-Loève expansion sampling method to consider the temporal variation of OB on residential building energy usage and domestic hot water usage respectively. These two studies indicated that the uncertainties in terms of annual and peak energy consumption associated with temporal variations of OB are required to be carefully addressed. In addition, it can be seen from Table 1 that annual and peak energy performance metrics were analyzed in most existing studies. However, the extent to which OB influences load distribution remains unclear. Such an influence needs to be further examined as it is critical to evaluate the adoption of energy-efficient technologies and control strategies [43].

3.2 Impact on different spatial and temporal scales

Accurate prediction of building energy demand at different scales (building, district and city) is essential for sizing and managing different-scale energy systems and for decision-makers to plan energy-saving strategies. Considering the importance of OB in predicting energy demand, the diversity among different occupants at district or larger scale would result in distinct results with respect to their impact on building energy performance. Accordingly, it is of high importance to identify the impact of OB on different spatial scales. Table 2 summarizes research studies found in the literature related to investigate the impacts of OB on either one particular scale (e.g. small and large) or compared the differences of such impact on different spatial scales. It can be seen from Table 2 that most of existing studies were conducted in a single zone or building [31-33, 35, 49]. So far, very few studies have been conducted to examine the uncertainty of OB on different spatial scales or in large-scale simulation [50]. For instance, the effects of OB on different number of rooms within a building were investigated. Based on the comparison of aggregated total lighting energy use, it is found that the impacts of OB (window shade use, light use and occupancy) on lighting consumption might be reduced as the number of offices increases, for instance, from 1 to 100 [34]. When moving from small scales to large scales (building stocks, districts, and cities), aggregating and smoothing effects could possibly exist and uncertainties caused by OB would be overlapped [51, 52]. For two buildings with diverse load profiles (i.e. peak load occurs at different periods), the total peak load would be smoothed. On the contrary, for two buildings with similar load profiles (i.e. peak load occurs at the same time), the total peak load will be maximized. Baetens and Saelens [50] reported that the variation in annual total electricity load and hot water consumption caused by OB (plug-in appliance use and hot water use) varies between 0.81 and 1.6 times the standard value (i.e. the mean consumption) for more than 10 houses, while the variation is between 0.88 and 1.3 times for more than 20 houses. Such results imply that the uncertainty effect of OB on annual total energy consumption tends to be reduced as the number of buildings increases. However, An et al. [43] recently investigated the effects of OB on district cooling systems. Results showed that the oversimplification of OB could result in significant overestimation of total and peak cooling loads, as well as load distribution. Hence, more research works are needed to identify the degree of uncertainty introduced by OB on large-scale simulation and understand the aggregating and smoothing effects caused by OB as well.

Table 2. Research studies on analyzing the impacts of OB at different spatial scales

Ref.	Spatial scales				Building type
	Room/Zone	Building	District	City	
[36]	√				Residential building (single family house)
[53]		√			Office building (private offices)
[35]		√			Office building
[32]		√			Residential building (low-income house)
[31]	√				Not mentioned
[54]		√			Residential building (single house)
[50]			√		Residential buildings
[39]	√				Office building
[33]		√			Residential building
[43]			√		Residential buildings
[34]	√	√			Office building

The impact of temporal granularities on BPS is also of interests as many OB models were developed with different time intervals. Temporal granularities involve both temporal resolution (also called time intervals) and time length of the simulation. With regard to temporal resolution, it directly determine how long the behavior needs to be modelled (e.g. 10min, half-hour, hourly) and therefore has a great impact on capturing the dynamic variation of OB [15, 55] and also reducing computational costs for simulation [56]. Generally, low-resolution models tend to have a relatively poor performance in capturing the temporal variation of OB, on the other hand a fine resolution would significantly increase the size of the required database (i.e. data needed for constructing OB models) [17, 23, 25, 26]. These studies mainly involved in qualitative analyses of the impact while a quantitative analysis is necessary to choose suitable time intervals. Recently a simulation-based analysis was conducted by Feng et al. [56] to compare the predicted cooling energy demand based on different time intervals of stochastic OB models. The authors reported that the simulated energy consumption distribution by using time intervals 5, 10 and 15 min were similar and shows a relatively small deviation. While the simulation with time intervals at 30 and 60 min presents a relatively large deviation. Based on their findings, 15 min was suggested for situations where computational time is restricted. However, this study was limited to cooling and window behavior. In this view, more studies on the test of other behavior are still needed in the future. With respect to the time length of simulation, existing simulation studies were usually performed on a yearly basis. In such a short period, possible changes in OB may be rather small and thus can be negligible for short-term prediction. Nevertheless, occupant's preferences and behavior might change over a long period since there could exist evolution to more energy-efficient appliances and possibility to modify behavior by energy policies and education [50]. Such change is important for assessing medium-to-long term energy performance and energy saving potentials in building sectors. Hence, future studies need to be carried out to consider OB and varied demographic compositions for long-term prediction [50].

3.3 Summary

In summary, UA and SA have been applied in addressing the impact of OB in BPS in recent years. The results obtained from these studies hints the important role of OB and on the BPS. However, existing UA and SA studies that consider the

OB-related parameters are limited in current building energy analysis. The representation of OB in these studies are simple and static while the temporal variation of OB has not been well addressed. Ongoing research works should take such variation into consideration when quantifying the impact of OB in BPS. Moreover, few studies investigated the use of advanced OB models (stochastic and data-driven behavioral models) in UA and SA studies [36]. Despite the importance for supporting effective building design and decision-making, consistency has not been reached about how sophisticated model would be beneficial to BPS. The uncertainty of an OB model itself which is ignored in current studies needs to be addressed in the future.

In addition, the impact of OB on different spatial scales and temporal granularities has not been fully investigated in the previous research works and further studies are still needed. In particular, procedures and guidelines for the selection of temporal granularities on different scales are necessary. In addition, some factors in large-scale simulation, such as the interaction between buildings and urban heat island effects, are rather complex. The influence of these factors still remains unclear and therefore their combined effects on OB also need to be identified. Filling such research gaps will provide a deep understanding of the role of OB models in building energy simulation, and facilitate research community to make a rational decision on how to choose suitable modeling methods for different simulation purposes..

4. Approaches to model occupant behavior

A lack of knowledge in terms of how occupants interact with buildings (e.g. use of windows and lights) in real buildings (i.e. epistemic uncertainty) possesses significant challenges in improving BPS. It is urgently needed to understand the process of occupants' energy-related behavior in reality and to develop effective methods to model such process. To date, various mathematical models have been proposed and Jia et al. [18], Zhang et al. [22] classify them into four broad categories: statistical models, stochastic/probabilistic models, data mining models and agent-based models (ABM). Note that, there exists some overlaps in these four categories which are not mutually exclusive. For instance, an agent-based model is a computational model with bottom-up structure that enables to model the behavior of each occupant independently and produce respective behavior patterns. In line with this definition, some statistical models can be considered as agent-based style as they simulated individuals' behavior independently through developing separate models developed for each occupant. However, other statistical models consider the average behavior of multiple occupants or simulate the behavior from the whole-building level and consequently, these models should be distinguished from agent-based models. To get a clear view of current OB models, up-to-date modeling methods are reviewed and their strengths and weaknesses are discussed based on to the abovementioned categories. Such a comprehensive review is also a crucial step towards the establishment of a set of criteria for supporting rational selection of OB models, which will be further discussed in Section 5. Additionally, important technical details and information (e.g. data source, data resolution and possible application for large-scale simulation) of all reviewed models are summarized in Table 3-7, respectively.

4.1 Statistical models

Statistical models generally use traditional regression methods (e.g. linear regression) or generalized linear methods (e.g. logistic regression) to quantitatively describe the relationship between influencing factors to the behavioral metrics.. Logistic regression methods have gained the greatest popularity due to their capability to deal with binary variables (such as behavior states) [57] and to allow for non-normal distribution [58]. Based on one or several environmental factors, it has been successfully applied for predicting the various adaptive behavior [57]. For example, most regression models for window opening and closing behavior were correlated with indoor/outdoor temperature or both. The early studies were conducted in office buildings while recently several field studies have been extended to residential buildings [57, 59-62] and school buildings [63]. In these later studies, the effects of other environmental stimuli such as indoor CO_2 concentration [57, 59, 61] and outdoor $PM_{2.5}$ concentration [61, 62] were also explored. It is inferred from the above studies that indoor CO_2 concentration is an important predictor of window opening behavior. However, inconsistency was found in [63] that the correlation of indoor CO_2 concentration to window status in classrooms was relatively small. In addition, it is noticed that a few studies considered time-related contextual factors (e.g. time of day, season) by developing sub-models under different contexts [64].

Table 3. Summary of statistical models of OB

Methods	Behavior type	Building Type	Input data needed	ABM	Resolution(duration of measurement and measure intervals)	Sample size	Ref.
Quadratic equation	Window use	Office	Occupancy, outdoor temperature	N	13 months, 1 min	21 offices	[65]
Logistic regression	Window use	Office	Outdoor/indoor temperature	N	Up to 3 months, four times per day	15 buildings	[66]
Logistic regression	Window use	Office	Indoor temperature, time of day	Y	3 months, 1 h	6 offices (5 single-occupant offices, 1 two-occupant office)	[64]
Logistic regression	Window use	Office	Outdoor/indoor temperature, or both	N	7 years, 5 min	14 offices	[67]
Logistic regression	Window use	Office	Outdoor temperature	N	8 weeks (only weekdays)	Office room	[68]
Logistic regression	Window use	Residential	Indoor CO_2 concentration; outdoor	N	8 months; 10 min	10 rented apartments, 5 privately	[59]

			temperature			owned houses	
Logistic regression	Window use	Residential	Time of day, CO_2 concentration, daily average outdoor temperature	N	1 year; 1 min	60 apartments	[57]
Logistic regression	Window use	Residential	indoor and outdoor air temperature and RH, wind speed, solar radiation, and rainfall, time of day and season	N	370 days 10 min	7 flats, 3 houses	[60]
Linear and logistic regression	Window use	School	Outdoor /indoor temperature	N	25 days 1-15min	a single classroom	[63]
Logistic regression	Blinds use	Office	Average illuminance of window, vertical solar radiation at the window	N	Not specified	25	[69]
Logistic regression	Shading devices use	Office	Local stimuli on the work plane	N	5 years and 3 months	14 celluar offices	[70]
Logistic regression	Air-conditioning use	Dormitory	Mean outdoor air temperature	N	10 weeks; 2 min	39 dormitory rooms	[71]

Statistical models allow understanding the influence of many independent variables on OB, particularly adaptive behavior that is predominantly driven by physical environmental stimuli. However, the success of statistical model development heavily relies on the identification of important influencing factors that are complex and sometimes inter-correlated. In addition, some statistical models only simulate various states of a building component (e.g. windows and blinds) [63, 65-68, 72, 73] rather than occupant's actions (opening a window) [57, 60, 64, 70]. Thus, they are quasi-static and not able to reveal the dynamic variation of OB (i.e. time-varying probability of a behavior to be performed). A possible solution is to directly relate the state transition (e.g. from closed to opened) with its influencing factors. Furthermore, OB is sometimes completely habitual or psychologically driven, and it is difficult for statistical models to interpret such randomness. For example, statistical models using the combination of outdoor and indoor operative temperature can only predict 22% of the total variance of clothing levels [74].

4.2 Stochastic/probabilistic models

4.2.1 Markov chain models

Markov chain models are essentially one type of stochastic models adopted to predict OB. The basic assumption behind this approach is that the future states only depend on the current state while being independent of all previous states (i.e. Markov property). The Markov property is described by the state transition probabilities, i.e. the conditional probability of being state '*i*'given that the current state is '*j*' [75]. Thus, Markov chain can be used to directly predict the behavioral state sequence since the model output has a one-to-one correspondence to a state of behavior [76].

Existing Markov chain models were mainly developed based on two categories of datasets: Time-use Survey (TUS) data and sensor-measured data. In the TUS data, household demographics and occupants' daily activities were collected nationwide through self-report diaries. Correspondingly, the synthetic domestic activities of a group of occupants rather than a single occupant could be modeled [77-79]. In particular, the transition probabilities are directly estimated [77] from a population by dividing the number of state transitions with the total amount of transitions. On-site sensors (e.g. cameras and contact sensors) are used to collect data in selected buildings. Thus, it can provide more information on the environmental factors and energy use. Markov chain models based on sensor-measured datasets have been applied in predicting adaptive behavior (mainly window operation and air-conditioning usage) [80, 81]. As behavioral information and possible influencing factors are both available in sensor-measured data, the transition probabilities can be estimated by direct calculation or using logistic regression analysis [82].

Table 4. Summary of Markov chain models of OB

Data source	Behavior Type	Building Type	Input data needed	Resolution	Application*	Ref.
TUS	Domestic activities	Residential	Nine activities (e.g. sleeping, and cooking)	1 min	Small-scale distributed power generation; building simulations; demand side management	[77]
TUS	Domestic activities**	Residential	Nine activities (e.g. sleeping, and cooking)	5 min	Indoor climate simulations; load management; load matching	[78]
TUS	Air-conditioning use	Office	17 domestic activities (e.g. cooking, and cleaning) Day of week (weekdays, Saturday, and Sunday); Type of person (e.g. working male/female)	15 min	NA	[79]
Sensor -measured	Window use	Residential	Window angle (classified into six classes) Outdoor air temperature	30 min	NA	[80]
Sensor -measured	Air-conditioning use	Office	Window state (open and closed) Outdoor air temperature	1 h	NA	[81]
Sensor -measured	Adaptive behavior (e.g. blind and fan use)	Office	Adaptive action (on and off) PMV; illuminance level; noise level; CO_2 concentration; wind speed; day of week; floor number; window operability; permission of night cooling strategy	1 h	NA	[83]

* Application in the table presents whether particular application purposes are highlighted in the reviewed papers such as small- and large-scale simulation. ** Note that, domestic activities refer to residents' daily activities such as doing laundry and preparing food which are generally obtained by time-use survey. Domestic activities correlate with appliance use (e.g. when preparing food kitchen appliances are used) and consequently, they can be

considered as indicators for infer residential appliance use. However, it require clear relationship between activities with possible used appliances.

An advantage of Markov chain models lies in its ability to simulate the transition between different states rather than predicting the probability of one state and therefore it can realistically represent the behavior of occupants and predict the evolvement of OB. With regard to model accuracy, McKenna et al. [84] reported that over a large number of runs (e.g. 50 runs), the synthetic OB profiles generated by Markov chain models have aggregated statistical properties has good agreement with original data. It should be noted that, model accuracy of an OB model can be represented in terms of different performance metrics (such as the probability of a behavior and respective duration) which are further discussed in Section 5.3. Moreover, this model is eminently suitable for long-term OB schedule prediction (e.g. months or whole year) with consideration of both computational efficiency and prediction accuracy [85]. The model accuracy is also substantially dependent on the calculation of transition matrix (i.e. a matrix of state transition probabilities) [78]. However, since the transition probabilities are independent of the time when an activity was started and do not consider the duration for a particular behavior, they show potential deficiencies in simulating the duration distribution of occupants' activities [86]. Meanwhile, the amount of states directly increases the difficulties to estimate the exact probability of each state transition, especially for the transitions that occurs rarely. In addition, time steps in existing Markov chain models are pre-defined and the selection is often subjective. A coarse interval might fail to capture the variations that occur between two successive time steps and lead to redundant calculation [67]. So far, only limited literature reported the relationship between the time step and model prediction accuracy [87].

4.2.2 Other probabilistic models

Other probabilistic models assume that the state of OB follows specific probabilistic distribution (e.g. normal distribution and exponential distribution). In this study, existing probabilistic models are categorized into two groups: discrete-time models and discrete-event models. In discrete-time models, the variation of states is represented as the progress of time and the states only change at fixed discrete time intervals. Moreover, the state is independent at each time step, which can be deemed as a Bernoulli process and satisfies the memoryless property. Different from discrete-time models, OB in discrete-event is modeled as a discrete and ordered sequence of events and each event occurs at a specific point in time. Hence, the frequency and duration of states describe each behavior. Such duration can be obtained from empirical distribution or survival analysis [67].

Table 5. Summary of probabilistic models of OB

Model type	Distribution type	Behavior type	Building type	Variables	Application	Ref.
Discrete-event	Weibull	Domestic activities	Residential	Individual characteristics (e.g. age and gender)	Building and urban energy simulation	[86]
Discrete-time	Weibull	Air-conditioning use	Residential	Indoor temperature	NA	[55]
Discrete-time	Exponential	Light use	Office	Psychological magnitude	NA	[88]
Discrete-time	Poisson, exponential, and normal	Light use	Office	Time of day	NA	[89]
Discrete-event	Empirical	Domestic activities	Residential	Starting time, number of starts, and duration	NA	[90]
Discrete-event	Gaussian and Uniform	Domestic activities	Residential	Day of week (weekdays, Saturday and Sunday), People attributes (e.g. working male/female	NA	[79]

				and housewife)		
Discrete-event	Cumulative probability	Domestic activities	Residential	Occupants' attributes, day type	Community-/ urban-scale energy demand modeling	[91]

Probabilistic models can partially reflect the randomness of OB by characterizing its time-dependency probability. Discrete-event probabilistic models have been proved to be efficient in simulating the duration of OB since the probability is directly correlated with a particular behavior as well as its starting time. Examples of OB correlated with its starting time include light use [88], domestic activities [86] and air-conditioning use [55]. Meanwhile, discrete-event probabilistic models are computationally efficient since they do not need to simulate at each time step [15]. Though discrete-time probabilistic models require a fixed time step, they can directly incorporate various time steps for more flexible simulations [55, 88], which is different from discrete-time Markov chain models. A key to the success of probabilistic model development is to obtain suitable probabilistic distribution curves fitting the original datasets (such as behavioral state and influencing factors). To get the distribution curve, the data associated with each model input variables need to be discretized. The statistical performance of such curve fit is easily influenced by the number of data points in each discretized group (i.e. continuously numeric factors divided into several categories with different ranges). A narrow discretization might lead to less or even no points in some groups and thereby poor fitting results will be made because of the large jumping up and down from one group to another. On the other hand, if this discretization is coarse, it is also not beneficial to curve fitting, as it would be subject to over-fitting problems. Therefore, a proper discretization of datasets for each model parameter is essential for the model performance [55].

4.3 Data mining models

Data mining (DM) techniques are powerful methods to extract hidden patterns and knowledge from large datasets and have also been applied to develop OB models [92-94]. Among various DM techniques, decision tree, Bayesian network, cluster analysis, and association rule mining methods are commonly used for OB patterns prediction and recognition. It is reported that the application of these techniques in OB modeling has shown promising results identifying typical behavior patterns as well as predicting OB [95, 96] and their performance has also been compared. Zhao and Lasternas et al. [97] compared the decision tree method, Bayes method and support vector machine method to investigate the office appliance usage pattern and its relationship with power consumption. The authors found that among the above said methods; the decision tree method has significantly better performance for individual OB prediction. The predicted OB patterns can represent the stochastic nature of OB and be further used in BPS.

In addition to the usage of a single DM technique, some researchers [66, 67] combined several DM techniques to take the advantage of the strengths associated with each technique. For example, Ren et al. [98] firstly used cluster analysis to find typical room temperature setting behavior patterns representing occupants' diverse comfort demand. Then decision tree was employed to predict heating energy consumption based on the identified temperature patterns and system operation. D'Oca et al. [99] used cluster analysis to obtain window opening and closing behavior patterns based on distinct datasets, and association rule mining was then conducted to discover the frequent patterns that concurrently existed.

Table 6. Summary of data mining models of OB

Methods	Behavior type	Building type	Variables	Ref.
Cluster analysis	Appliance use	Residential	Twelve factors (e.g. annual mean air temperature, annual mean RH, and annual mean wind speed,)	[95]
Association rule mining	Light use	Educational	Season, time, weekends, occupancy, light, waste, events, day of week	[96]
Decision tree	Air-conditioning use	Residential	Temperature, RH, behavior data (on/off)	[100]
Decision tree	Appliance use	Office	Behavior data (occupancy, computer logging time), computer consumption data	[97]
Decision tree	Appliance use	Residential	Hour, day, season, month	[101]
Cluster analysis and decision tree	Space heating	Residential	Hourly average temperature on weekday, hourly standard deviation on weekday during the studied period (for cluster analysis); Room temperature clusters, heating system operation class, heating energy consumption (for decision tree)	[98]
Cluster analysis and association rule mining	Window use	Office	Environment-related variables, weather data, behavior data, building and system-related variables	[99]
Nearest Neighbor Model	Domestic activities	Residential	Thirty occupant activities (cooking, clothes washing, ironing etc.) Household composition, dwelling type (detached house or apartment)	[102]
Bayesian network	Window use	Residential	Indoor and outdoor temperature, time of the day, indoor CO_2 concentration, indoor RH	[103]
Bayesian approach	Shading devices use	Office	Work plane illuminance; vertical illuminance; shade position; electric light level; outside view; visual privacy	[104]
Deep learning method	Window use	Office	Outdoor/ indoor environmental-related variables, time-related contextual factors, etc.	[105]

DM techniques can extract and exploit valuable information from large amounts of data and discover different behavior patterns among the data. This information and patterns can easily be interpreted and are understandable, as well as provide deep insights into the way occupants behave. In addition, DM techniques can analyze both numerical and categorical attributes. This provides possibilities to study the influence of socioeconomic, psychological and physiological factors on OB since these factors are often categorical. The main limitation of these models is that they are static and incapable of simulating the dynamic and stochastic characteristics of OB. The successful application of DM techniques for describing OB, particularly cluster analysis and association rule mining, heavily relies on the size of databases (i.e. databases include data of potential influencing factors and behavior parameters) which are usually restricted by cost and effort in practice. Furthermore, the users' prior expertise directly impacts the effectiveness of model construction and knowledge extraction, which adds difficulties in model development. In addition, some DM techniques might not be suitable for time series data [100]. For example, Zhou et al. [100] reported that decision tree has relatively poor performance in mining air-conditioning patterns from time-dependent parameters (temperature and relative humidity). It should be noted that, not all the data mining techniques can be directly integrated with building simulation tool as some techniques such as clustering analysis and association rule mining are descriptive which are generally used to extract useful profiles and understand the occupant behavior. Nevertheless, the discovered typical profiles and rules can be further implemented in the simulation tool [99].

4.4 Agent-based models

Agent-based model (ABM) is a computational method that allows researchers to create, analyze and manipulate the models that constitute agents interacting with a specified environment [106]. In this method, each agent is described as real-world individuals with various features, capabilities, and interactions with other agents and the built environment. In response, each agent can evaluate their situation and reach decisions on changing their behavior. In this review, the agent-based model in this section specifies the models that involve the social science and consider the interactions between different agents.

Table 7. Summary of ABMs of OB

Behavior Type	Building Type	Variables	Theoretical framework	Simulation tools	Ref.
Light use	Commercial	Occupant characteristics (e.g. location in building and light sensitivity); Design options and available controls (e.g. shading devices, illumination devices and target illumination levels); Environment conditions (e.g. sunrise and sunset)	Belief-Desire-Intention (BDI) framework	graphical user interface (GUI); RADIANCE	[107]
Appliance use	Residential	Perceptual: beliefs, psychological (cognitive) factors, social influence, domestic context: inhabitant, appliances, physical location	BDI framework	Business redesign agent-based holistic modeling system (Brahms)	[108]
Appliance use	Office	User agents identified by working time: early birds, timetable compliers, and flexible workers User agents identified by energy saving awareness: environment champion, energy saver, regular user, and big user	N/A	Anylogic	[109]
Not specified	Office	High energy consumers, medium energy consumers, and low energy consumers	N/A	Anylogic	[110]
Not specified	Dormitory	Social network: relationship with their peers; initialized energy consumption	N/A	N/A	[111]
Blinds, door, fan, heater, and window use clothing adjustments	Commercial	Behavioral belief, control belief, normative belief	N/A	Matlab, Building Controls Virtual Test Bed (BCVTB) and MLE+; EnergyPlus	[112]
Not specified	N/A	Social network types:	overview, design concepts, and details (ODD)	Java using Repast J 3.0	[113]
Clothing	Office	Occupants' attributes (e.g. commuting	(ODD)	Human and	[114]

adjustment	method and personal traits)	Building
Personal fan	Thermal acceptability and comfort,	Interaction
on/off	behavioral preference and sequencing,	Toolkit (HABIT),
Thermostat	behavioral constraints, the role of	MATLAB,
up/middle/down	clothing and activity level	BCVTB;
Windows	Local thermal environment (e.g.	EnergyPlus
open/close	ambient indoor/outdoor temperature)	

The ABM is capable of directly simulating the agent-to-agent and has flexibility in being further extended to different levels such as a system level, building level. Such flexibility also involves the addition of more agents and provides a possible solution for simulating multiple behaviors. Moreover, this model enables various factors (especially social and psychological factors [108]) to be easily considered, and thus it is well-suited for understanding the social and psychological influence on OB. On the other hand, due to the addition of the granularity of occupant diversity in ABM, it would significantly increase the computational time, which in turn increase the challenges to a large-scale simulation. Also, the development of ABM requires special expertise in detailed settings on each behavior and sequences of different behavior. The settings, however, are empirical and its applicability on other building contexts is questionable. Note that even minor changes in behavioral rules may have considerable impacts on model output. Besides, the real-time communication between the ABM and building energy simulation programs increases the difficulty in its application since these programs are normally unique in coding languages and protocols [115].

5. OB model performance improvement

A robust OB model is expected to have acceptable predictive accuracy and high generalization capability (i.e. the model's performance when it is applied to buildings or occupants other than the sample data used for the model development). The robustness of a model refers to the feasibility to predict OB for different contexts (such as two similar buildings in different climates). With respect to the predictive accuracy of OB models, the selection of model input and output has a direct impact and has been investigated in different studies. The model input are parameters used in a model to describe the variation of OB and the output are OB model output parameters to represent occupants' behavior. With respect to the generalization capability of OB models, accounting for occupant diversity and evaluating models comprehensively pose two major challenges to researchers. Hence, model input and output selection, occupant diversity and performance metrics for model evaluation are reviewed in Section 5.1 to 5.3, respectively.

5.1 Model input/output selection

Existing models discussed in Section 3 are intended to predict and interpret OB with consideration of their influencing factors in an explicitly quantitative way. However, OB is influenced by a wide range of factors which can be broadly classified into five categories: physical environmental (e.g. indoor environmental conditions), contextual (e.g. heating system type), physiological (e.g. age and gender), psychological (e.g. habits and attitudes), and social factors (e.g. household composition) [116]. Due to the diversity and complexity of these factors, the selection of them in OB models is a non-trivial task. Moreover, the impact of some factors (e.g. rainfall, outdoor CO_2 concentration and ambient $PM_{2.5}$ concentration etc.) remains unclear and these factors are unavailable in current BPS tools. It is highly desirable that the optimal factors can be identified and used for model construction not only to improve model's predictive accuracy but also to reduce model's complexity. Ideally, a model with optimal factors has high predictive accuracy and relatively low model complexity. It should be noted that, the issues about data collection is out of scope of this review, and model input and output selection issues are discussed based the assumptions that sufficient data needed for a specific OB model are available in the dataset.

5.1.1 Evaluation of the importance of influencing factors

The importance of various influencing factors of OB was commonly evaluated based on statistical methods and significance metrics. In general, a factor with higher significance is more likely to be considered as important factors. Table 8 summarizes the analysis methods and significance metrics employed in previous studies.

Statistical analysis methods have been effectively used to process physical factors for their importance evaluation. However, a major problem with this method is that categorical factors (e.g. psychological factors) are difficult to be processed. Compared to the numerical factors (mainly physical environmental variables), usually the amount of data points of the categorical factors is relatively small. Such data scarcity might result in a much lower significance value of these factors when calculated by statistical analysis methods. As a result, the importance of these factors is normally underestimated particularly in the understanding of general patterns in large datasets [99]. To recognize the real effect of these categorical factors, constructing controlled experiments in which only the investigated factor changes was considered as an effective solution [117].However, O'Brien et al. [117] claimed that this is not practically feasible due to the difficulties in implementing such experiments in real life. Alternatively, information entropy-based measures used in DM techniques such as information gain ratio [118] might provide a possible approach to quantify such effects.

Table 8**.** Summary of significance analysis methods and indicators of OB influencing factors

Behavior type	Influencing factors	Significance analysis method	Significance metrics	Ref.
Lighting use	Outdoor illuminance level	Qualitative	N/A	[89]
Small appliances use	Attitude, subjective norms, perceived behavioral control, habits	Stepwise multiple regression analysis	F statistics and p-value	[119]
Window use	Outdoor and indoor air temperature, outdoor and indoor RH, indoor CO_2,	Multi-factor variance analysis	F statistics and p-value	[68]

	outdoor wind speed			
Window use(position)	Outdoor air temperature, non-environmental factors: seasonal, effects, change of daylight saving time, occupant absence in subsequent days, window orientation, floor level, gender, personal preference	Logistic regression analysis	Wald statistics and p-value	[120]
Window use	Outdoor air temperature, wind speed, rainfall, sunshine hours, solar radiation, RH	Correlation analysis	Pearson correlation and p-value	[121]
	time of day, occupancy patterns, window orientation, seasonal effects	qualitative analysis	N/A	
Window use	Outdoor temperature, indoor temperature	Logistic regression analysis	r^2	[67]
Window use	indoor temperature, outdoor temperature	Regression analysis	Pearson correlation	[65]
	seasonal effects, time of day, building occupancy	Qualitative analysis	N/A	
Blind use	Temperature, wind speed, rainfall, sunshine hours, RH, solar altitude, solar radiation	Correlation analysis	Pearson correlation	[122]
Manual solar shades	Outdoor temperature, solar radiation	Regression analysis	Wald statistics and p-value	[123]

Significance metrics can aid researchers in adopting the uniform standards to judge which factor is important. Despite the widespread usage of significance metrics (as shown in Table 8), they do not necessarily represent the causation of behavior [67]. More importantly, if the usage of these metrics lacks a consistency in different studies, the significance found in one study may not be true in another study. Thus, further research should be undertaken to develop suitable procedures for quantifying the significance of both categorical and numeric factors based on consistent metrics.

5.1.2 Input variable selection methods

Selection of optimal variables for model input can further improve model performance, model generalization, and computational effectiveness. For instance, Pan et al [124] reported that if suitable input variables were selected, the accuracy of a

Gaussian distribution model to predict window behavior was approximately 9% higher than that of a commonly used Logistic regression model with the same input parameter, i.e. outdoor temperature. Based on different evaluation criteria, the existing variable selection methods can be classified into two categories, namely filter methods and wrapper methods [125].

Filter methods assess the significance of each factor of OB, and those factors with a higher metric value than a pre-determined threshold are selected as model inputs. Filter methods are relatively simple and computationally efficient. Its major limitation lies in the fact that the selection criterion might not be suitable for the modeling since it does not directly correlate with the model performance evaluation. In addition, the method cannot distinguish highly correlated factors and thus might result in the selection of redundant factors. It should be noted that, there exists a wide range of criteria and metrics such as distance measure and mutual information that can be applied for filter methods and significance analysis methods (e.g. correlation analysis and logistic regression) mentioned in section 5.1.1 can also be applied. Wrapper methods evaluate the model performance of all possible subsets of factors, and the subsets with the best performance are normally considered as optimal variables. The subsets can be found by different searching strategies (e.g., forward selection and backward selection [125]). Unlike the filter methods, the wrapper methods can quantify the possible inter-relationship between different factors. However, the iteration of model evaluation based on the subsets results in more computational time. Also, the wrapper method has a high risk of over-fitting problems. The research works that applied the filter method and wrapper method in behavior modeling are summarized in Table 9.

Table 9. Summary of research studies that adopted filter and wrapper methods in behavioral modeling

Behavior type	OB model type	Optimal model input variable	Variable selection method	Ref.

Window use	Univariate and multivariate logistic regression models; Markov chain models	Univariate: outdoor temperature; Multivariate: outdoor and indoor temperature	Filter and wrapper method	[67]
Window use	Bayesian network	time of the day, indoor CO_2 concentration, solar radiation, indoor air temperature, indoor RH, and outdoor air temperature	Filter method	[103]
Window use	Logistic regression model	Outdoor temperature	Filter method	[68]
Window use	Gaussian distribution model	Outdoor temperature	Filter method	[124]
Blinds use	Logistic regression model and probabilistic model	local stimuli on the work plane	Wrapper method	[70]
Adaptive actions	Logistic regression model	Indoor and outdoor temperature	Wrapper method	[126]
Blinds control	Logistic regression model	Air temperature	Wrapper method (backward elimination method)	[69]
Shading devices and light use	Bayesian model	Work plane illuminance, vertical illuminance, shade position, electric lighting level, lighting condition preference, outside view need, visual privacy need	Wrapper method (forward selection)	[58]

As can be seen in Table 5, both the filter methods and wrapper methods have been widely adopted in existing studies, particularly when developing statistical and data mining models. These studies illustrated the advantages of variable selection methods in identifying optimal factors for OB models. Their results implies that only few variables are needed for OB models, for example, outdoor temperature and indoor temperature were frequently identified as important variables for window use models [68, 124]. Indeed, a possible way to take advantage of their strengths is the combined use of both methods. For example, Haldi and Robinson [67] used a combination of them to develop the multivariate logistic regression models for modeling the window behavior. However, very few research works applied both filter and wrapper methods when constructing OB models. Additionally, some data mining models, such as the decision tree model, are able to determine the optimal variables by using the entropy-based measure [118]. Nevertheless, removing irrelevant and redundant variables from an initial set of measured data is still required and is critical for these models to reduce data dimensionality, thereby enhancing the model's computational efficiency. Such efforts would be beneficial to provide useful insights in the identification of important factors so that reduce unnecessary measurement sensors for future studies and real-world applications.

5.1.3 Model output selection

OB model output is the parameters used to represent OB and are commonly identified in terms of research purposes. Existing OB model output can be grouped into two categories, i.e. states and events/actions [127]. The use of states can directly indicate the status of building components or systems caused by occupants' specific actions (e.g. window status-opened/closed). A transition event, i.e. an action, indicates the changes of status which can be considered as a proximity of behavior. Both states and actions can be applied for aggregation and individual comparison. Note that, however, the former one is only indicative of occupant's preferred states rather than an actual characterization of OB. Thus, it could not be able to identify the triggers that cause occupants to perform an action and OB models as well. The latter overcomes such limitations and it can directly reflect the triggering factors.

An important issue needs to be addressed is the proper identification of model output variables because an inappropriate use might hinder the future application. For example, shade movement rates, i.e. the average percentage of shades that are moved at different times, is a useful indicator describing how occupants operate the shades devices [128]. Yet this indicator is not applicable to OB models in visual simulation, as it does not specify the direction of shade position change. Meanwhile, when multi-class states (e.g. small, medium and large window openings) are adopted as OB model output, appropriate discretization is essential and there has to be a trade-off between accuracy requirements and technical difficulty [16, 128]. This is due to the fact that most experimental instruments (e.g. contact sensor for measuring window status) monitor only the discontinuous parameters and it is not practical for the researchers to use a continuously varied metric to represent the state of OB. This type of OB mainly includes window opening behavior, blind turning up and down, and dim lights. Moreover, such discretization needs to be considered based on model application purposes. For example, discretization of shade positions plays a vital role in visual comfort and daylighting simulation than in a whole-building simulation.

5.2 Occupant diversity characterization

5.2.1 Approaches for occupant diversity characterization

The diversity of occupants and its impact on energy-related OB prediction have raised increasing interests among research community in recent years [129]. Basically, occupant diversity can be defined as differences in occupants and their responses and actions to indoor environment and energy use in buildings. Several studies reported that the diversity of different occupants, if properly addressed, can improve the capability of OB models used in BPS [129]. In this review, the occupant diversity characterized in existing models were categorized into two main dimensions: explicit-implicit and individual-group, as illustrated in Fig. 2.

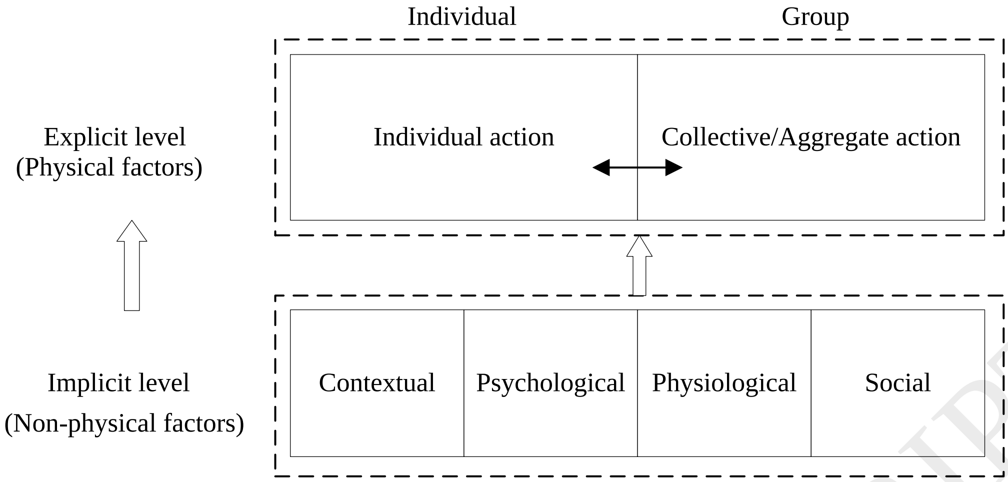


Fig. 2. Two-dimension framework of occupant diversity characterization with consideration of different influencing factors

The former dimension represents the methods by which OB is expressed (i.e. which factors are used to predict the behavior). The latter dimension represents occupants' actions observed in individuals and groups (e.g. zone/building). Based on these two dimensions, four strategies can be developed to characterize occupant diversity, including implicit-group, implicit-individual, explicit-individual, and explicit-group. Except for agent-based models, most existing OB models adopted the explicit-group strategies. In particular, by using 'average'/'typical' occupants, they tended to characterize occupant actions in an aggregate level with physical factors. For these modeling methods, an independent model can be developed for each occupant (i.e. explicit-individual) which can partially reflect the occupant diversity. For example, in [129], seven logistic regression models of window operation were developed for seven occupants in an open-plan office where the different coefficients can be viewed as indicators of occupant diversity. The limitation of explicit-group or explicit-individual characterization lies in the fact that it neglects the relationship of occupants' energy-related behavior with non-physical factors such as the social context [130]. Consequently, most OB models developed are yet case-specific and their generalization capability is highly questioned. However, one fundamental challenge is that non-physical factors are qualitative and quite diverse, thereby increasing the difficulties in being directly incorporated into OB models [14, 131].

Different from the other models, agent-based models enable non-physical factors to be directly added to represent intra-individual diversity among different occupants in addition to physical factors (i.e. implicit-individual). For example, in [114] personal traits were added in order to assess individual's adaptive behavior. Furthermore, this model can easily and flexibly reflect the occupant diversity in the implicit-group level by adding more agents and defining relationship among different agents. For example, Kashif et al. [108] considered perceptual beliefs, psychological factors, and social influence among different agents through a theoretical framework originated from the social domain (i.e. Belief-Desire-Intension framework). In implicit-individual and implicit-group characterization, however, one crucial issue is that strong domain expertise is required in the selection and determination of these non-physical factors.

To address the issue of occupant diversity, some studies proposed an improved version of 'average' occupants based on simple classifications (e.g. active and passive occupants) [109, 132] or occupant typologies obtained by cluster analysis. Although this method can improve the model accuracy, such a division may have been still overly simplistic and tends to underestimate the diversity of occupants and their behavior. Also, it is argued that the occupant traits should be described by a continuous function rather than a discrete typology [133]. A recent advancement was made by Haldi et al. [134] which applied a generalized linear mixed model by adding a built-in probabilistic term to capture the random effects of occupant diversity. Results indicated that such expression could help to differentiate the uncertainty of behavior caused by environmental stimuli and individual diversities.

5.2.2 Occupant diversity characterization in large-scale simulation

As discussed in previous sections, the uncertainty of building energy performance introduced by OB on different scale simulation remains unclear. One main reason is due to the simplification of occupant diversity in previous studies. Hence, how and to which extent occupant diversity needs to be presented in OB models plays important role in understanding such uncertainty (particularly for large-scale simulation [133, 134]). Since occupants in different buildings may behave in a very dissimilar manner. For example, the peak load in district heating and cooling simulation would be largely smoothed if the preferences of thermostat setting points are rather diverse. Baetens and Saelens [50] reported that the local disaggregation of demographic statistics (an indicator of the occupant diversity), which assumed as homogeneous in OB models) might create epistemic uncertainties in the simulation at a district level. To address this issue, several studies attempted to incorporate occupant's intra- and inter- diversity in OB models [133]. They reported that, compared with large scale simulation, modeling occupant diversity in small scale simulation is much more important due to the smoothing effects of energy demand. Note that only 16 occupants were investigated in their study and more evidence is still needed to support such argument. An et al. [43] used stochastic sampling methods to distribute typical patterns of each behavior to individual apartments for district cooling energy simulation. Despite the promising results, it requires detailed and extensive questionnaire surveys to extract occupants' information (e.g. setting point preferences).

To explore the occupant diversity in large-scale simulation, a real challenge is to establish a database with sufficient sample size so that the selected samples can be properly characterized to represent aggregate behavior among various occupants. It is reported that epistemic uncertainty would increase if behavior and occupant related data is not properly collected [12]. Currently, there is no clear guidelines for selecting appropriate samples for developing OB models so as to consider the impact of occupant diversity. However, O'Brien et al. [133] investigated the diversity issues for modeling occupancy pattern in office buildings through in-situ measurements which would give implications for OB modeling. Results of [133] indicated that the sample size is of great importance to monitor occupant diversity in a population. They suggested that hundreds of occupant samples were more appropriate rather than 10 to 15 samples which were used in most existing studies. Aside from the appropriate data collection, some advanced machine learning models (such as deep learning models and Bayesian approaches) shows some advantages to generalize its predictive performance to a large number of occupants and automatically recognize their individual behavior patterns [105].

5.3 Comprehensive evaluation of OB models

A proper model evaluation is an important step in OB model development and future implementation in building energy simulation. Internal evaluation and external evaluation are two common methods to evaluate OB models. The former method is to assess the performance of OB models by checking their fitness with the measured data from the same dataset. Although it is relatively simple and effective, its applicability to other datasets is highly questionable. In this view, model evaluation based on independent datasets is desirable and correspondingly, external evaluation procedures have been recently highlighted. These procedures test the predictive performance of OB models based on external datasets from a different but 'slightly related' population with respect to different locations, other built environment or occupants [135, 136]. The issues of OB model evaluation have been thoroughly summarized in the main report of Annex 66 and details can found in [127].

To understand the whole process of model construction, a fundamental challenge in both internal and external procedures [137] is the selection and usage of suitable performance metrics. In this study, existing performance metrics are sorted into two groups: behavior-oriented and application-oriented. The behavior-oriented metrics refer to indicators associated with the accuracy of behavior prediction. Several indicators were used according to model output's category (i.e. numeric output such as the probability of an action, or categorical output such as the states of a behavior, as mentioned in Section 4.1.3). If the outcome is numeric, root mean squared error (RMSE) and coefficient of determination (R^2) are frequently used. Otherwise, confusion matrixes i.e., a matrix consisting of predicted classes being classified into four groups: truly positive (TP), falsely positive (FP), truly negative (TN) and falsely negative (FN)), are normally used to evaluate the overall accuracy, precision, recall and F1 [105, 118, 125]. Overall accuracy indicates the proportion of correct prediction outcome, which is defined as (TP+TN)/ (TP+FP+TN+FN). Precision refers to ratio of positively predicted outcome and all positively predicted results, i.e. TP/(TP+FP). Recall refers to the proportion of truly predicted positive results and all true positive results (i.e., TP/TP+FN). F1 is defined as 2TP/(2TP+FP+TN+FN), which is the harmonic mean of precision and recall. The application-oriented metrics are used to assess the simulation performance of energy demand models integrated with OB models. Besides, these application-oriented metrics can evaluate the combined effects when simultaneously employing multiple OB models. For example, Andersen et al. [136] estimated the mixed predictive power for indoor environmental parameters (i.e. temperature, relative humidity and CO_2 concentration) by applying two stochastic models of the window opening and thermostat set-point adjustments. As most performance metrics are numeric, RMSE is commonly used. Different performance metrics employed in previous studies are summarized in Table 10.

Table 10. Summary of performance metrics used in OB models evaluation

Type	Performance metrics	Statistic	Category	Units
Behavior-oriented	The fraction of a building component's state	Overall probability/ daily change rate/average daily change rate	Numerical	%
	The probability of state transitions or behavior			%
	Number of behavioral state	–		–

		transitions			
		Duration of a behavioral state			hours
		Two-class behavioral states		Categorical	
		Multi-class behavioral states			
		Total energy consumption	1. Annual/ monthly/ daily/ hourly 2. during occupied/ unoccupied hours	Numerical	kWh kJ
		Heating energy consumption			
		Cooling energy consumption			
		Lighting energy consumption			
	Energy	Energy consumption of plug loads			
		Total peak demand			kW
		Heating peak demand			
		Cooling peak demand			
		Lighting peak demand			
		Load distribution	–	-	–
Application -oriented	Electricity grid interaction [a]	Mismatch hours			hours
	Indoor environment quality	Indoor air temperature	Mean/ minimum/ maximum/ standard deviation	Numerical	ºC
		Indoor operative temperature			ºC
		Indoor relative humidity			%
		Count of hours [b]			hours
		CO_2 concentration			ppm
		Transmitted solar radiation			kWh
	Indoor thermal comfort	Predicted mean vote (PMV)	–		–
		Predicted percentage of dissatisfied (PPD)			%

[a] Electricity grid interaction is used to assess the effect of intermittent power generation from the building (e.g. photovoltaics systems) on electricity grid [42].

[b] Count of hours refers to numbers of occurrence of temperatures above a given threshold/cut-off temperature [138].

Behavior-oriented metrics are normally used as stand-alone criteria and they only describe how well the OB models can predict the data, while the influence of these models on building energy simulation remains unknown. For instance, a transition event (e.g. blinds from closed to opened) causes a change in the indoor climate. Failing to predict these transition events leads to prediction errors in energy demand models, resulting in a discrepancy between the actual and simulated indoor environmental conditions and energy consumption. In this aspect, application-oriented metrics should also be considered if an OB model is selected for fit-for-purpose. However, lacking standardized metrics and respective statistics presents a real challenge to modelers and engineers, and very few OB models have been assessed based on both behavior- and application-oriented metrics. Currently, the performance metrics are scattered and often used by researcher based on their own experience. Future efforts in this direction are thus of utmost importance. Extensive and cross-sectional observational studies are also suggested to promote more representative evaluation. In addition, special focus should be placed on models developed and used for large-scale simulation purposes.

6. Tools to integrate OB models with BPS

The integration of OB models into building energy demand simulation plays an important role in achieving the goal of model accuracy improvement [139]. For current BPS tools, existing integration methods were classified into four groups: direct input or control, built-in OB models, user function or custom code, and co-simulation [19]. Among these methods, co-simulation has received considerable interests due to its flexibility in offering a co-operative way of different programs. So far, two different methods have been further developed for realizing co-simulation: middleware coupling methods and standardized coupling methods.

The middleware coupling method uses a middleware (e.g. building controls virtual test bed (BCTVB) and MLE+ Toolbox) as a master to manage data exchange between OB models and simulation programs. This method has been used to integrate agent-based models with EnergyPlus [112, 114]. It enables different programs to communicate with the middleware while each program needs to be implemented in a specific interface defined by the middleware. This implies that researchers must be familiar with different data coding format (e.g. R-code, C-code) to pre-define the exchange parameters.

Instead of using a specific interface, the standardized coupling method provides a uniform interface for information and data exchange. In this method, functional mock-up interface (FMI) has been widely adopted due to its popularity and widespread application in many simulation environments [140]. Different from the middleware coupling method, this method allows for direct link between the simulation programs without using a middle data exchange tool. Thus, it is more efficient and less complex than the middleware coupling method. Hong et al [140] developed the obFMU through FMI framework and tested several OB models with EnergyPlus and ESP-r. However, other simulation programs have not been tested yet and the performance of co-simulation have been rarely reported. Also, future studies need to be done to create a set of OB functional mock-up units for different categories of behavior.

7. Discussion

7.1 Selection of modeling methods

The use of OB models in building energy simulation plays a vital role in improving the accuracy of energy demand prediction. As existing modeling methods have different advantages and constraints, the selection of a suitable method from them poses a real challenge as a random selection might result in ineffectiveness or even failure in OB model development and applications. For example, Haldi et al. [37] claimed that stochastic models are not the necessities for the total energy performance simulation while they are desirable if the distribution of peak demand is expected.

To facilitate such a selection, detailed comparisons of modeling methods are prerequisites to identify the model’s capabilities (e.g. prediction accuracy) and requirements (e.g. computational time) versus different application purposes (e.g. building system control). Existing comparative studies of different modeling approaches were scarcely found and also limited to window behavior prediction.

Haldi et al. [67] developed window behavior prediction models based on the logistic regression method, the Markov chain method and a combination of both. A comparison between them indicated that the predicted fraction of windows being opened varied significantly. Moreover, compared with the Markov chain model, the hybrid model obtained relatively small percentage of opening and closing durations. Indeed, consensus has not been reached about which approaches are preferred for different simulation purposes. This is possibly because OB model's intended application scenarios and objectives are rather diverse and specific rules cannot be provided (and might be even ultimately unachievable). On the other hand, a lack of sound documentation and validation of OB models can also be attributed to the insufficiency of comparative studies among existing models. To address such a research gap, a variety of general simulation scenarios in contexts of different spatial scales and temporal granularities of simulation (as discussed in Section 2), application domains (e.g. heating and cooling demand) should be studied in the future. Recommendations and insights need to be gained from the level of difference introduced by different OB models in which such difference relies on the assessment of different performance indicators (e.g. annual total energy demand v.s. hourly).

A few studies have been carried out with the focus on the establishment of model selection criteria and framework. Isabella Gaetani et al. [141] and Yan et al. [14] suggested a fit-for-purpose (FFP) strategy framework with a focus on the influence of model complexity (predictability of different aspects of OB) in selecting the modeling methods. Nevertheless, the criteria involved in prior studies are scattered and there is not yet a set of uniform and common criteria for model selection.

To address this issue, this study attempts to suggest a set of general selection criteria concerning the comprehensive performance of OB models. The comprehensive performance calls for the trade-off in prediction accuracy under different scenarios, requirements and contexts. Based on a thorough review of existing modeling methods, the following six criteria were proposed, and for future reference, the performance of each modeling method regarding these criteria is compared in Table 11. It should be noted that, this table presents an empirical comparison among the four main categories of modeling methods according to the proposed criteria based on the comprehensive review of recent works.

• *Model complexity*. This can be mainly explained by model structure and resolution (i.e. the number of parameters and their granularity [141]). It is an important criterion for selecting a modeling method due to the principle of parsimony [141]. Typically, a model with a simple structure and high accuracy suited for its attended application is often desirable. For example, Akaike information criterion (AIC), commonly used in current statistical models, is a criterion to measure the relative model performance of a set of models with different input data (i.e. the complexity level) and therefore, it can balance the trade-off between model complexity and prediction performance [60].

• *Computational efficiency*. This refers to the computational efforts and time required for modeling OB and its further integration into building energy models. Its importance is particularly stressed when OB models are used for the different application purposes (e.g. building system control).

• *Variable selection*. This refers to a model's ability to automatically select suitable input parameters that straightly influence the model accuracy and complexity. Such parameters can be factors measured through experiments/fields and/or available in simulation tools. The questions concern to variable selection were discussed in Section 5.

• *Flexibility*. This relates to a model's applicability to different modeling tasks such as prediction and pattern recognition (usually needed for sensor-measured data). This criterion considers the implication of model application but has seldom been highlighted in previous studies.

• *Integration capability*. This criterion explores the capability of OB models to be further integrated into BPS tools. This criterion enables the linkage between OB and building energy model performance for future application and thus is non-trivial.

• *Expertise requirements*. This criterion describes the grade/degree of expertise required in OB (e.g. how detailed input information are needed in an advanced OB model) and mathematical background. This criterion deserves deliberate consideration since the extent of which the impacts of specific level of modeler's knowledge might imposes uncertainty for the building energy analysis. Such impact was demonstrated for occupancy variables [142] but remains unclear for the usage of occupants' action models.

Table 11. Comparison of different models with respect to six criteria

Model type	Model complexity	Computational time [a]	Modeler/researchers Flexibility			Integration capability	Expertise requirements
			Variable selection	Function	Behavior type		
Statistical	Low	Low	N	Prediction	Adaptive behavior	Low	Low
Stochastic	Medium	Medium	N	Prediction	Both adaptive behavior and other behavior	Medium	Medium
Data mining	Medium	Medium	Y (decision tree)/N	Prediction Pattern recognition	Both adaptive behavior and other behavior	High	High
Agent-based	High	High	N	Prediction	Adaptive behavior	Medium	High

[a] Computational efficiency: computational effort for model application is considered while the effort for model development is not included.

7.2 Proposed framework: towards improving the overall performance of OB models

The usefulness of OB models in facilitating realistic simulations of building performance has been recognized by academics. However, it tends to be neglected by practitioners and the availability of behavioral models in existing BPS tools is scarce. Necessarily, efforts are needed to enable the OB models understandable and easy to use by different building stakeholders (e.g. engineers and policy makers) who are not familiar with the concepts of human-building interaction and relevant details. Based on the literature reviewed in previous sections, the critical issue to enhance the capability and generalizability of OB models is to address the practical problems associated with necessary steps during the model developments (i.e. model input and output selection, occupant diversity characterization and model evaluation), indexed to different spatial scales (e.g. building zone to city) and application scenarios. In this view, a detailed framework was proposed based on those important technical aspects, as shown in Fig. 3. This framework makes explicit the interactive relationship inherent to different steps and also defines the options for modeling methods selection, occupant diversity characterization and performance metrics selection.

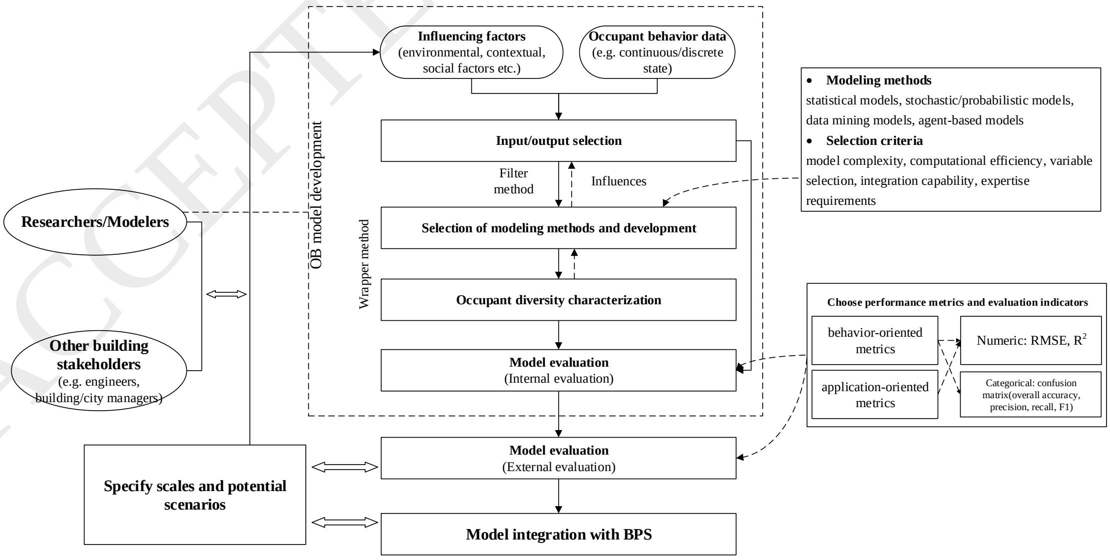


Fig. 3 A detailed framework of OB model development

The first step in the proposed framework is to specify spatial scales and possible application scenarios. Due to the difficulties to collect occupant related data from different data sources, such information is critical to other researchers who intend to conduct comparable studies in the future and different building stakeholders who might use such models. As stated in Section 3, as the aggregation and smoothing effects, the uncertain effects of OB on energy demand of buildings in a district or city scale could possibly be reduced. Hence, modeling approaches with simple structure (e.g., less required input information) seem to be a general trend. In line with this, statistical models and stochastic models would be preferable for large-scale simulation because of their relatively lower complexity level (as compared in Table 11). On the other hand, Happle et al. [23] argued that agent-based models could be also options for urban-scale building energy simulation due to its high flexibility for different levels. Because of the computation concerns for large-scale simulation, a trade-off needs to be made when applying such models with consideration of the computational efficiency. Additionally, once the scales and scenarios are specified, it would also support the selection of suitable performance metrics in assessing OB models (which has been addressed in Section 5.3). The second step consists of conduct input and output selection based on the data collected according to defined scales and scenarios. This step is highlighted in this framework since it is not only important for reducing model complexity as well as model generalizability (both are important features for large-scale simulation). This framework suggests that the input variables can be selected based on a combination of the filter method and wrapper method. Note that, for the wrapper selection method, the choice of performance metrics also has a direct impact on the results of the selected input variables. The results obtained from such steps can provide useful evidence for determining important factors that are necessarily monitored in the future. The third step is to select modeling methods based on a set of general selection criteria proposed in this review. Occupant diversity, a critical aspect for OB models' generalization capability, needs to be considered in the fourth step. As discussed in Section 5.1, the second and fourth step are simultaneously influenced by modeling methods selected. In the fifth step, internal evaluation is conducted to assess model performance by comparing its predicted results with measured data from the same population. It should be mentioned that, in this step, the evaluation indictors need to be carefully chosen and the usage of a single overall accuracy indicator may be insufficient as it cannot reflect the fact whether different behavioral states are accurately predicted. This is particularly important when a certain behavioral state dominates the database in which a high overall accuracy can be easily achieved. After the internal evaluation, the model can be optionally tested with independent datasets (if available) to further assess its performance under the potential application scenarios. With the increasing data released as open datasets (e.g. [143]), this would allow for further support of OB development and assessment. Finally, the sixth step is to integrate OB models with BPS tools and apply developed models. As discussed in Section 6, advanced OB models (such as agent-based models) often require the co-simulation when implementing the developed OB models in real cases. These scenarios can be specified by application domains, simulation purposes in different stages of building projects (e.g. early-stage design and retrofitting analysis). In general, modeling approaches for a wide array of different applications should be determined in line with the availability of data and application domains. For instance, in terms of a building-scale early-stage design evaluating the natural ventilation performance and total annual energy use, sophisticated and advanced models are not necessarily needed as limited information related occupant's window operation behavior in this stage. On the other hand, such models would provide more realistic estimations of total and peak energy demand for a building-scale retrofitting analysis with the possibility of investigating necessary information. The proposed framework would be useful to give directions for new studies in the area of OB modeling for fully-documented and models that can be transferrable and comparative performance. It could be vital for the systematic use of OB related information and models in the context of simulating building performance in a realistic way at different scales.

8. Conclusions and future work

The uncertainty and stochastic nature of OB is a complex, multi-faceted issue that has a substantial impact on the variation of building energy consumption. In previous studies, such impacts on building energy consumption and indoor environment were explored by applying UA and SA in BPS tools. However, the temporal variation of OB has not been well-understood. Hence, OB models, including statistical models, stochastic/probabilistic models, data mining models and agent-based models, and their application in building energy demand modeling and simulation have been widely investigated in recent years. The advancement of developing different OB modeling approaches (especially stochastic methods) integrated with BPS would facilitate a paradigm shift from existing deterministic simulation practices to a probabilistic form. It is found that OB models could also be used in the uncertainty analysis procedure to understand the uncertain effects of OB on BPS and improve the simulation performance, while its integration can also introduce the uncertainties if inappropriate modeling approach is used. At present, OB needs to be considered in building energy simulation at different spatial scales (zone, building, district, and city). Considering the model complexity at different scales, the simplification (especially for occupant diversity) and modeling granularities associated with the OB should be recognized for various spatial scales and also computational costs for temporal granularities. However, existing OB models are diverse and each has its own objectives, constraints, and technical capabilities. To successfully develop and apply OB models, modeling methods should be elaborately selected based on a set of criteria with the trade-off in prediction accuracy under different simulation scenarios and requirements. Moreover, as the model performance and generalizability are directly influenced by the selection of model input and output, a combination of internal and external evaluation with standardized metrics (i.e. behavior-oriented and application-oriented metrics) is necessary.

Despite the considerable researches carried out on OB modeling, yet there are several important challenges need to be addressed before implementing its wide application in building energy simulation. The following future work is therefore suggested:

1) Only few research works explored the simulation uncertainty associated with OB and the outcomes of OB models in different spatial and temporal scales. The understanding of OB in large-scale simulation are still vague which is imperative for the increasing need for energy simulation at multi-domain and multi-scale. Future research is required to conduct sensitivity and uncertainty analysis of OB under commercial districts and mix-districts. It would help to form guidelines for developing and choosing OB models that are suitable for using on different simulation scales.

2) While a multitude of OB models have been developed, their effectiveness in behavioral prediction and building energy simulation still needs rigorous evaluation. To improve model performance, a promising way is to establish hybrid models making the best of each method. At the same time, seeking new methods to address the increasing amount of data is necessary with thorough documentation of the details in model development in terms of development process, purpose and validation to avoid unnecessary duplication of research effort. Additionally, as consensus has not been reached about the model chosen preferences, a systematic framework with standardized criteria and metrics is necessary to support optimal model selection.

3) Interdisciplinary studies that incorporate physical, social, and psychological

science to reveal the intrinsic cause of OB (especially social and psychological factors) are essential for future focus. Well-established methods to quantify significance of such factors on OB are necessary for model performance improvement. Moreover, existing approaches to represent occupant diversity should be tested extensively which requires suitable sample size and careful selection of samples.

4) Many of the research issues discussed in previous sections can be ascribed to the lack of sufficient large-scale data due to the fairly large cost and efforts. Nevertheless, the advent of smart sensors, metering technologies and on-line distrusting questionnaires could support large-scale surveys and long-term data measurement. To consolidate the international monitoring campaigns, holistic methods and roadmaps to determine appropriate sample size, sensor deployment and length of measurement are of high-priority. Besides, Time-use surveys (TUSs) data shows significant potentials for developing detailed stochastic models but has been only conducted in several developed countries (Sweden, Japan, German France, the United States, etc.). Therefore, more extensive TUSs are needed for other countries.

Acknowledgment

The authors would like to express their gratitude to the National Natural Science Foundation of China (Project No. 51408205), the Fundamental Research Funds for the Central Universities, and Hunan Collaborative Innovation Center of Building Energy Conservation & Environmental Control for supporting this research project. In addition, the authors would like to thank the CSC (China Scholarship Council) to provide financial support for the study in Canada.

References

[1] I. Sartori, A. Napolitano, K. Voss, Net zero energy buildings: A consistent definition framework, Energy and Buildings. 48 (2012) 220-232.

[2] E. Annunziata, M. Frey, F. Rizzi, Towards nearly zero-energy buildings: The state-of-art of national regulations in Europe, Energy. 57 (2013) 125-133.

[3] A. Fotopoulou, G. Semprini, E. Cattani, Y. Schihin, J. Weyer, R. Gulli, A. Ferrante, Deep renovation in existing residential buildings through façade additions: A case study in a typical residential building of the 70s, Energy and Buildings. 166 (2018) 258-270.

[4] T. Weng, Y. Agarwal, From Buildings to Smart Buildings—Sensing and Actuation to Improve Energy Efficiency, IEEE Design & Test of Computers. 29 (2012) 36-44.

[5] M. De Groote, M. Fabbri, Smart buildings in a decarbonised energy system, Buildings Performance Institute Europe, BPIE, Brussels, Belgium (2016).

[6] E. Recast, Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), Official Journal of the European Union. 18 (2010) 2010.

[7] F. Ascione, N. Bianco, O. Böttcher, R. Kaltenbrunner, G.P. Vanoli, Net zero-energy buildings in Germany: Design, model calibration and lessons learned from a case-study in Berlin, Energy and Buildings. 133 (2016) 688-710.

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

[9] L.C. Tagliabue, M. Manfren, A.L.C. Ciribini, E. De Angelis, Probabilistic behavioural modeling in building performance simulation—The Brescia eLUX lab, Energy and Buildings. 128 (2016) 119-131.

[10] T. Buso, V. Fabi, R.K. Andersen, S.P. Corgnati, Occupant behaviour and robustness of building design, Building and Environment. 94 (2015) 694-703.

[11] L. Van Gelder, H. Janssen, S. Roels, Probabilistic design and analysis of building performances: Methodology and application example, Energy and Buildings. 79 (2014) 202-211.

[12] S. Lee, I. Bilionis, P. Karava, A. Tzempelikos, A Bayesian approach for probabilistic classification and inference of occupant thermal preferences in office buildings, Building and Environment. 118 (2017) 323-343.

[13] I. EBC, Annex 66: definition and simulation of occupant behavior in buildings (2013).

[14] 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.

[15] 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.

[16] H.B. Gunay, W. O'Brien, I. Beausoleil-Morrison, A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices, Build. Environ. 70 (2013) 31-47.

[17] S. Wei, R. Jones, P. de Wilde, Driving factors for occupant-controlled space heating in residential buildings, Energy and Buildings. 70 (2014) 36-44.

[18] 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, Renewable and Sustainable Energy Reviews. 68, Part 1 (2017) 525-540.

[19] 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 (2017) 1-14.

[20] F. Stazi, F. Naspi, M. D'Orazio, A literature review on driving factors and contextual events influencing occupants' behaviours in buildings, Build. Environ. 118 (2017) 40-66.

[21] S. Gilani, W. O'Brien, Review of current methods, opportunities, and challenges for in-situ monitoring to support occupant modelling in office spaces, Journal of Building Performance Simulation. 10 (2017) 444-470.

[22] 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.

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

[24] M. Schweiker, S. Carlucci, R.K. Andersen, B. Dong, W. O'Brien, Occupancy and Occupants' Actions, in: Exploring Occupant Behavior in Buildings, Springer, 2018, pp. 7-38.

[25] J. Rouleau, L. Gosselin, P. Blanchet, Understanding energy consumption in high-performance social housing buildings: A case study from Canada, Energy. 145 (2018) 677-690.

[26] Z.M. Gill, M.J. Tierney, I.M. Pegg, N. Allan, Low-energy dwellings: the contribution of behaviours to actual performance, Build. Res. Inf. 38 (2010) 491-508.

[27] A. Belleri, R. Lollini, S.M. Dutton, Natural ventilation design: An analysis of predicted and measured performance, Building and Environment. 81 (2014) 123-138.

[28] J. Winkler, J. Munk, J. Woods, Effect of occupant behavior and air-conditioner controls on humidity in typical and high-efficiency homes, Energy and Buildings. 165 (2018) 364-378.

[29] G. Mavromatidis, K. Orehounig, J. Carmeliet, A review of uncertainty characterisation approaches for the optimal design of distributed energy systems, Renewable and Sustainable Energy Reviews. 88 (2018) 258-277.

[30] C.J. Hopfe, J.L.M. Hensen, Uncertainty analysis in building performance simulation for design support, Energy and Buildings. 43 (2011) 2798-2805.

[31] M. Bonte, F. Thellier, B. Lartigue, Impact of occupant's actions on energy building performance and thermal sensation, Energy and Buildings. 76 (2014) 219-227.

[32] A.S. Silva, E. Ghisi, Uncertainty analysis of user behaviour and physical parameters in residential building performance simulation, Energy and Buildings. 76 (2014) 381-391.

[33] Z. O'Neill, F. Niu, Uncertainty and sensitivity analysis of spatio-temporal occupant behaviors on residential building energy usage utilizing Karhunen-Loève expansion, Building and Environment. 115 (2017) 157-172.

[34] S. Gilani, W. O'Brien, H.B. Gunay, Simulating occupants' impact on building energy performance at different spatial scales, Building and Environment. 132 (2018) 327-337.

[35] E. Azar, C.C. Menassa, A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings, Energy and Buildings. 55 (2012) 841-853.

[36] P. Hoes, J.L.M. Hensen, M.G.L.C. Loomans, B. de Vries, D. Bourgeois, User behavior in whole building simulation, Energy and Buildings. 41 (2009) 295-302.

[37] F. Haldi, D. Robinson, The impact of occupants' behaviour on building energy demand, Journal of Building Performance Simulation. 4 (2011) 323-338.

[38] K. Sun, T. Hong, A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures, Energy and Buildings. 146 (2017) 383-396.

[39] S. Gilani, W. O'Brien, H.B. Gunay, J.S. Carrizo, Use of dynamic occupant behavior models in the building design and code compliance processes, Energy and Buildings. 117 (2016) 260-271.

[40] W. Belazi, S. Ouldboukhitine, A. Chateauneuf, A. Bouchair, Uncertainty analysis of occupant behavior and building envelope materials in office building performance simulation, Journal of Building Engineering. 19 (2018) 434-448.

[41] A. Ioannou, L.C.M. Itard, Energy performance and comfort in residential buildings: Sensitivity for building parameters and occupancy, Energy and Buildings. 92 (2015) 216-233.

[42] N. Nord, T. Tereshchenko, L.H. Qvistgaard, I.S. Tryggestad, Influence of occupant behavior and operation on performance of a residential Zero Emission Building in Norway, Energy and Buildings. 159 (2018) 75-88.

[43] J. An, D. Yan, T. Hong, K. Sun, A novel stochastic modeling method to simulate cooling loads in residential districts, Applied Energy. 206 (2017) 134-149.

[44] Z. Pang, Z. O'Neill, Uncertainty quantification and sensitivity analysis of the domestic hot water usage in hotels, Applied Energy. 232 (2018) 424-442.

[45] J. Munk, J. Winkler, Effect of occupant behavior on peak cooling and dehumidification loads in typical and high-efficiency homes, Energy and Buildings. 184 (2019) 122-140.

[46] S. Gilani, W. O'Brien, Exploring the impact of office building users' modeling approaches on energy use across Canadian climates, Energy and Buildings (2019).

[47] W. Tian, A review of sensitivity analysis methods in building energy analysis, Renewable and Sustainable Energy Reviews. 20 (2013) 411-419.

[48] 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, Renewable and Sustainable Energy Reviews. 93 (2018) 285-301.

[49] I. Gaetani, P. Hoes, J.L.M. Hensen, On the sensitivity to different aspects of occupant behaviour for selecting the appropriate modelling complexity in building performance predictions, Journal of Building Performance Simulation. 10 (2017) 601-611.

[50] R. Baetens, D. Saelens, Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour, Journal of Building Performance Simulation. 9 (2016) 431-447.

[51] B. Talebi, F. Haghighat, P.A. Mirzaei, Simplified model to predict the thermal demand profile of districts, Energy and Buildings. 145 (2017) 213-225.

[52] B. Talebi, F. Haghighat, P. Tuohy, P.A. Mirzaei, Validation of a community district energy system model using field measured data, Energy. 144 (2018) 694-706.

[53] W. Parys, D. Saelens, H. Hens, Coupling of dynamic building simulation with stochastic modelling of occupant behaviour in offices – a review-based integrated methodology, Journal of Building Performance Simulation. 4 (2011) 339-358.

[54] R. Evins, K. Orehounig, V. Dorer, Variability between domestic buildings: the impact on energy use, Journal of Building Performance Simulation. 9 (2016) 162-175.

[55] X. Ren, D. Yan, C. Wang, Air-conditioning usage conditional probability model for residential buildings, Build. Environ. 81 (2014) 172-182.

[56] X. Feng, D. Yan, C. Wang, On the simulation repetition and temporal discretization of stochastic occupant behaviour models in building performance simulation, Journal of Building Performance Simulation. 10 (2017) 612-624.

[57] D. Calì, R.K. Andersen, D. Müller, B.W. Olesen, Analysis of occupants' behavior related to the use of windows in German households, Building and Environment. 103 (2016) 54-69.

[58] S.A. Sadeghi, N.M. Awalgaonkar, P. Karava, I. Bilionis, A Bayesian modeling approach of human interactions with shading and electric lighting systems in private offices, Energy and Buildings. 134 (2017) 185-201.

[59] R. Andersen, V. Fabi, J. Toftum, S.P. Corgnati, B.W. Olesen, Window opening behaviour modelled from measurements in Danish dwellings, Building and Environment. 69 (2013) 101-113.

[60] R.V. Jones, A. Fuertes, E. Gregori, A. Giretti, Stochastic behavioural models of occupants' main bedroom window operation for UK residential buildings, Building and Environment. 118 (2017) 144-158.

[61] M. Yao, B. Zhao, Window opening behavior of occupants in residential buildings in Beijing, Building and Environment. 124 (2017) 441-449.

[62] S. Shi, B. Zhao, Occupants' interactions with windows in 8 residential apartments in Beijing and Nanjing, China. 9 (2016) 221-231.

[63] F. Stazi, F. Naspi, M. D'Orazio, Modelling window status in school classrooms. Results from a case study in Italy, Building and Environment. 111 (2017) 24-32.

[64] G.Y. Yun, K. Steemers, Time-dependent occupant behaviour models of window control in summer, Building and Environment. 43 (2008) 1471-1482.

[65] S. Herkel, U. Knapp, J. Pfafferott, Towards a model of user behaviour regarding the manual control of windows in office buildings, Building and Environment. 43 (2008) 588-600.

[66] H.B. Rijal, P. Tuohy, M.A. Humphreys, J.F. Nicol, A. Samuel, J. Clarke, Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings, Energy and Buildings. 39 (2007) 823-836.

[67] F. Haldi, D. Robinson, Interactions with window openings by office occupants, Build. Environ. 44 (2009) 2378-2395.

[68] N. Li, J. Li, R. Fan, H. Jia, Probability of occupant operation of windows during transition seasons in office buildings, Renewable Energy. 73 (2015) 84-91.

[69] V. Inkarojrit, Monitoring and modelling of manually-controlled Venetian blinds in private offices: a pilot study, Journal of Building Performance Simulation. 1 (2008) 75-89.

[70] F. Haldi, D. Robinson, Adaptive actions on shading devices in response to local visual stimuli, Journal of Building Performance Simulation. 3 (2010) 135-153.

[71] M. Schweiker, M. Shukuya, Comparison of theoretical and statistical models of air-conditioning-unit usage behaviour in a residential setting under Japanese climatic conditions, Building and Environment. 44 (2009) 2137-2149.

[72] V. Inkarojrit, G. Paliaga, Indoor climatic influences on the operation of windows in a naturally ventilated building (2004) 19-22.

[73] J.F. Nicol, Characterising occupant behaviour in buildings: towards a stochastic model of occupant use of windows, lights, blinds, heaters and fans. 2 (2001) 1073-1078.

[74] S. Schiavon, K.H. Lee, Dynamic predictive clothing insulation models based on outdoor air and indoor operative temperatures, Build. Environ. 59 (2013) 250-260.

[75] G.A. Fink, Markov models for pattern recognition: from theory to applications, Springer Science & Business Media, 2014.

[76] W. Ching, X. Huang, M.K. Ng, T.K. Siu, Markov Chains: Models, Algorithms and Applications, 2nd ed., Springer Publishing Company, Incorporated, 2013.

[77] J. Widén, E. Wäckelgård, A high-resolution stochastic model of domestic activity patterns and electricity demand, Applied Energy. 87 (2010) 1880-1892.

[78] 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, Journal of Building Performance Simulation. 5 (2012) 27-44.

[79] J. Tanimoto, A. Hagishima, H. Sagara, A methodology for peak energy requirement considering actual variation of occupants' behavior schedules, Building and Environment. 43 (2008) 610-619.

[80] R. Fritsch, A. Kohler, M. Nygård-Ferguson, J.-. Scartezzini, A stochastic model of user behaviour regarding ventilation, Building and Environment. 25 (1990) 173-181.

[81] G.Y. Yun, P. Tuohy, K. Steemers, Thermal performance of a naturally ventilated building using a combined algorithm of probabilistic occupant behaviour and deterministic heat and mass balance models, Energy and Buildings. 41 (2009) 489-499.

[82] J. Tanimoto, A. Hagishima, State transition probability for the Markov Model dealing with on/off cooling schedule in dwellings, Energy and Buildings. 37 (2005) 181-187.

[83] S. Wei, R. Buswell, D. Loveday, Probabilistic modelling of human adaptive behaviour in non-airconditioned buildings (2010).

[84] E. McKenna, S. Higginson, P. Grunewald, S.J. Darby, Simulating residential demand response: Improving socio-technical assumptions in activity-based models of energy demand, Energy Efficiency (2017) 1-15.

[85] E. McKenna, M. Thomson, High-resolution stochastic integrated thermal–electrical domestic demand model, Applied Energy. 165 (2016) 445-461.

[86] U. Wilke, F. Haldi, J. Scartezzini, D. Robinson, A bottom-up stochastic model to predict building occupants' time-dependent activities, Building and Environment. 60 (2013) 254-264.

[87] Y. Zhou, Z. Yu, J. Li, Y. Huang, G. Zhang, The Effect of Temporal Resolution on the Accuracy of Predicting Building Occupant Behaviour based on Markov Chain Models, Procedia Engineering. 205 (2017) 1698-1704.

[88] C. Wang, D. Yan, H. Sun, Y. Jiang, A generalized probabilistic formula relating occupant behavior to environmental conditions, Build. Environ. 95 (2016) 53-62.

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

[90] D. Fischer, A. Härtl, B. Wille-Haussmann, Model for electric load profiles with high time resolution for German households, Energy and Buildings. 92 (2015) 170-179.

[91] Y. Yamaguchi, Y. Shimoda, A stochastic model to predict occupants' activities at home for community-/urban-scale energy demand modelling, Journal of Building Performance Simulation. 10 (2017) 565-581.

[92] Z. Yu, F. Haghighat, B.C.M. Fung, Advances and challenges in building engineering and data mining applications for energy-efficient communities, Sustainable Cities and Society. 25 (2016) 33-38.

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

[94] Z. Yu, F. Haghighat, B.C.M. Fung, E. Morofsky, H. Yoshino, A methodology for identifying and improving occupant behavior in residential buildings, Energy. 36 (2011) 6596-6608.

[95] Z. Yu, B.C.M. Fung, F. Haghighat, H. Yoshino, E. Morofsky, A systematic procedure to study the influence of occupant behavior on building energy consumption, Energy Build. 43 (2011) 1409-1417.

[96] D.F. Motta Cabrera, H. Zareipour, Data association mining for identifying lighting energy waste patterns in educational institutes, Energy Build. 62 (2013) 210-216.

[97] 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.

[98] X. Ren, D. Yan, T. Hong, Data mining of space heating system performance in affordable housing, Build. Environ. 89 (2015) 1-13.

[99] S. D'Oca, T. Hong, A data-mining approach to discover patterns of window opening and closing behavior in offices, Build. Environ. 82 (2014) 726-739.

[100] H. Zhou, L. Qiao, Y. Jiang, H. Sun, Q. Chen, Recognition of air-conditioner operation from indoor air temperature and relative humidity by a data mining approach, Energy Build. 111 (2016) 233-241.

[101] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, S. Ploix, A prediction system for home appliance usage, Energy Build. 67 (2013) 668-679.

[102] M. Baptista, A. Fang, H. Prendinger, R. Prada, Y. Yamaguchi, Accurate Household Occupant Behavior Modeling Based on Data Mining Techniques. (2014) 1164-1170.

[103] V.M. Barthelmes, Y. Heo, V. Fabi, S.P. Corgnati, Exploration of the Bayesian Network framework for modelling window control behaviour, Build. Environ. 126 (2017) 318-330.

[104] S.A. Sadeghi, N.M. Awalgaonkar, P. Karava, I. Bilionis, A Bayesian modeling approach of human interactions with shading and electric lighting systems in private offices, Energy Build. 134 (2017) 185-201.

[105] R. Markovic, E. Grintal, D. Wölki, J. Frisch, C. van Treeck, Window opening model using deep learning methods, Building and Environment. 145 (2018) 319-329.

[106] N. Gilbert, Agent-based models, Sage, 2008.

[107] C.J. Andrews, D. Yi, U. Krogmann, J.A. Senick, R.E. Wener, Designing buildings for real occupants: An agent-based approach, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans. 41 (2011) 1077-1091.

[108] A. Kashif, X.H.B. Le, J. Dugdale, S. Ploix, Agent based Framework to Simulate Inhabitants' Behaviour in Domestic Settings for Energy Management. (2011) 190-199.

[109] T. Zhang, P. Siebers, U. Aickelin, Modelling electricity consumption in office buildings: An agent based approach, Energy and Buildings. 43 (2011) 2882-2892.

[110] E. Azar, C.C. Menassa, Agent-based modeling of occupants and their impact on energy use in commercial buildings, J. Comput. Civ. Eng. 26 (2011) 506-518.

[111] J. Chen, J.E. Taylor, H. Wei, Modeling building occupant network energy consumption decision-making: The interplay between network structure and conservation, Energy and Buildings. 47 (2012) 515-524.

[112] Y.S. Lee, A.M. Malkawi, Simulating multiple occupant behaviors in buildings: An agent-based modeling approach, Energy and Buildings. 69 (2014) 407-416.

[113] K. Anderson, S. Lee, C. Menassa, Impact of social network type and structure on modeling normative energy use behavior interventions, J. Comput. Civ. Eng. 28 (2013) 30-39.

[114] J. Langevin, J. Wen, P.L. Gurian, Simulating the human-building interaction: Development and validation of an agent-based model of office occupant behaviors, Building and Environment. 88 (2015) 27-45.

[115] S. Papadopoulos, E. Azar, Integrating building performance simulation in agent-based modeling using regression surrogate models: A novel human-in-the-loop energy modeling approach, Energy and Buildings. 128 (2016) 214-223.

[116] V. Fabi, R.V. Andersen, S. Corgnati, B.W. Olesen, Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models, Building and Environment. 58 (2012) 188-198.

[117] W. O'Brien, H.B. Gunay, The contextual factors contributing to occupants' adaptive comfort behaviors in offices – A review and proposed modeling framework, Build. Environ. 77 (2014) 77-87.

[118] J. Han, J. Pei, M. Kamber, Data mining: concepts and techniques, Elsevier, 2011.

[119] R.M. Tetlow, C. van Dronkelaar, C.P. Beaman, A.A. Elmualim, K. Couling, Identifying behavioural predictors of small power electricity consumption in office buildings, Building and Environment. 92 (2015) 75-85.

[120] S. Wei, R. Buswell, D. Loveday, Factors affecting 'end-of-day' window position in a non-air-conditioned office building, Energy and Buildings. 62 (2013) 87-96.

[121] Y. Zhang, P. Barrett, Factors influencing the occupants' window opening behaviour in a naturally ventilated office building, Building and Environment. 50 (2012) 125-134.

[122] Y. Zhang, P. Barrett, Factors influencing occupants' blind-control behaviour in a naturally ventilated office building, Building and Environment. 54 (2012) 137-147.

[123] J. Yao, Determining the energy performance of manually controlled solar shades: A stochastic model based co-simulation analysis, Applied Energy. 127 (2014) 64-80.

[124] S. Pan, Y. Han, S. Wei, Y. Wei, L. Xia, L. Xie, X. Kong, W. Yu, A model based on Gauss Distribution for predicting window behavior in building, Building and Environment. 149 (2019) 210-219.

[125] M. Kuhn, K. Johnson, Applied predictive modeling, Springer, 2013.

[126] F. Haldi, D. Robinson, On the behaviour and adaptation of office occupants, Building and Environment. 43 (2008) 2163-2177.

[127] T.H. Da Yan, Annex 66 Final ReportL Definition and Simulation of Occupant Behavior in Buildings . ISBN 978-0-9996964-7-7 (2018).

[128] W. O'Brien, K. Kapsis, A.K. Athienitis, Manually-operated window shade patterns in office buildings: A critical review, Building and Environment. 60 (2013) 319-338.

[129] F. Tahmasebi, A. Mahdavi, On the utility of occupants' behavioural diversity information for building performance simulation: An exploratory case study, Energy and Buildings. 176 (2018) 380-389.

[130] K.B. Janda, Building communities and social potential: Between and beyond organizations and individuals in commercial properties, Energy Policy. 67 (2014) 48-55.

[131] T. Hong, D. Yan, S. D'Oca, C. Chen, Ten questions concerning occupant behavior in buildings: The big picture, Build. Environ. 114 (2017) 518-530.

[132] E. Azar, C. Menassa, An agent-based approach to model the effect of occupants' energy use characteristics in commercial buildings, in: Computing in Civil Engineering (2011), 2011, pp. 536-543.

[133] W. O'Brien, H.B. Gunay, F. Tahmasebi, A. Mahdavi, A preliminary study of representing the inter-occupant diversity in occupant modelling, Journal of Building Performance Simulation. 10 (2017) 509-526.

[134] F. Haldi, D. Calì, R.K. Andersen, M. Wesseling, D. Müller, Modelling diversity in building occupant behaviour: a novel statistical approach, Journal of Building Performance Simulation. 10 (2017) 527-544.

[135] M. Schweiker, F. Haldi, M. Shukuya, D. Robinson, Verification of stochastic models of window opening behaviour for residential buildings, Journal of Building Performance Simulation. 5 (2012) 55-74.

[136] R.K. Andersen, V. Fabi, S.P. Corgnati, Predicted and actual indoor environmental quality: Verification of occupants' behaviour models in residential buildings, Energy and Buildings. 127 (2016) 105-115.

[137] A. Mahdavi, F. Tahmasebi, On the quality evaluation of behavioural models for building performance applications, Journal of Building Performance Simulation. 10 (2017) 554-564.

[138] A. Mavrogianni, M. Davies, J. Taylor, Z. Chalabi, P. Biddulph, E. Oikonomou, P. Das, B. Jones, The impact of occupancy patterns, occupant-controlled ventilation and shading on indoor overheating risk in domestic environments, Building and Environment. 78 (2014) 183-198.

[139] H.B. Gunay, W. O'Brien, I. Beausoleil-Morrison, Implementation and comparison of existing occupant behaviour models in EnergyPlus, Journal of Building Performance Simulation. 9 (2016) 567-588.

[140] T. Hong, H. Sun, Y. Chen, S.C. Taylor-Lange, D. Yan, An occupant behavior modeling tool for co-simulation, Energy and Buildings. 117 (2016) 272-281.

[141] I. Gaetani, P. Hoes, J.L.M. Hensen, Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy, Energy and Buildings. 121 (2016) 188-204.

[142] Q. Wang, G. Augenbroe, J. Kim, L. Gu, Meta-modeling of occupancy variables and analysis of their impact on energy outcomes of office buildings, Applied Energy. 174 (2016) 166-180.

[143] B.W. Hobson, D. Lowcay, H.B. Gunay, A. Ashouri, G.R. Newsham, Opportunistic occupancy-count estimation using sensor fusion: A case study, Building and Environment. 159 (2019) 106154.