Building and Environment 219 (2022) 109195

Contents lists available at ScienceDirect

Building and Environment

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

A guideline to document occupant behavior models for advanced building controls

Bing Dong [a,*], Romana Markovic [b], Salvatore Carlucci [c], Yapan Liu [a], Andreas Wagner [b], Antonio Liguori [d], Christoph van Treeck [d], Dmitry Oleynikov [e], Elie Azar [f], Gianmarco Fajilla [g], Ján Drgoňa [h], Joyce Kim [i], Marika Vellei [j], Marilena De Simone [g], Masood Shamsaiee [k], Mateus Bavaresco [l], Matteo Favero [m], Mikkel Kjaergaard [n], Mohamed Osman [o], Moritz Frahm [p], Sanam Dabirian [o], Da Yan [q], Xuyuan Kang [q]

[a] Department of Mechanical & Aerospace Engineering, Syracuse University, 263 Link Hall, Syracuse, NY, 13244, United States
[b] Building Science Group, Karlsruhe Institute of Technology, Englerstr. 7, 76131, Karlsruhe, Germany
[c] Energy, Environment and Water Research Center, The Cyprus Institute, Nicosia, Cyprus
[d] E3D - Institute of Energy Efficiency and Sustainable Building, RWTH Aachen University, Mathieustr. 30, 52074, Aachen, Germany
[e] Katernberger Straße 107, Essen, Nordrhein-Westfalen, 45327, Germany
[f] Department of Industrial and Systems Engineering, Khalifa University, Abu Dhabi, United Arab Emirates
[g] Department of Environmental Engineering, University of Calabria, P. Bucci 44A, 87036, Rende, CS, Italy
[h] Pacific Northwest National Laboratory, Richland, WA, USA
[i] Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada
[j] LaSIE (UMR CNRS 7356), La Rochelle University, La Rochelle, France
[k] School of Mechanical Engineering, Iran University of Science and Technology, Tehran, Iran
[l] Laboratory of Energy Efficiency in Buildings, Department of Civil Engineering, Federal University of Santa Catarina, Brazil
[m] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway
[n] The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark
[o] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, H3G 1M8, Canada
[p] Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, 76344, Eggenstein-Leopoldshafen, Germany
[q] Building Energy Research Center, School of Architecture, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Guideline documentation
Occupant behavior model
Building control
Model Predictive control (MPC)

ABSTRACT

The availability of computational power, and a wealth of data from sensors have boosted the development of model-based predictive control for smart and effective control of advanced buildings in the last decade. More recently occupant-behavior models have been developed for including people in the building control loops. However, while important objectives of scientific research are reproducibility and replicability of results, not all information is available from published documents. Therefore, the aim of this paper is to propose a guideline for a thorough and standardized occupant-behavior model documentation. For that purpose, the literature screening for the existing occupant behavior models in building control was conducted, and the occupant behavior modeling processes were studied to extract practices and gaps for each of the following phases: problem statement, data collection, and preprocessing, model development, model evaluation, and model implementation. The literature screening pointed out that the current state-of-the-art on model documentation shows little unification, which poses a particular burden for the model application and replication in field studies. In addition to the standardized model documentation, this work presented a model-evaluation schema that enabled benchmarking of different models in field settings as well as the recommendations on how OB models are integrated with the building system.

* Corresponding author.
E-mail address: bidong@syr.edu (B. Dong).

https://doi.org/10.1016/j.buildenv.2022.109195
Received 1 April 2022; Received in revised form 6 May 2022; Accepted 11 May 2022
Available online 20 May 2022
0360-1323/

1. Introduction

Building energy consumption has been proven to be a systematic procedure comprehensively influenced by not only engineering technologies but also cultural concepts, occupant behavior, social equity, etc. Occupant behavior (OB), discussed in this paper, refers to occupancy presence and the number of people in the spaces of a building, and human building interactions, such as window and blind operations, turning on/off lighting, as well as thermostat adjustment and use of electric appliances. As, occupants are one of the major factors that influence energy consumption [[1], p. 79], depending on the level of building automation, the inclusion of the occupant-behavior modeling in building controls could lead to optimized building operation and reduced energy consumption [2,3]. Furthermore, the inclusion of human-building interaction [4] or OB in the control loop [5] could lead to a higher thermal comfort level and a general increase in occupants' satisfaction with the indoor environment.

However, OB models are rarely included in building controls, despite the vast scientific evidence that considering OB in building energy management could lead to optimized building performance [6]. Prior research studies show that the reason for the limited field applications of OB models could be the lack of OB model standardization and clear documentation [7,8], which results in models' limited replicability. The development of OB model standardization can enable easy integration and compatibility with existing or new building automation system (BAS). For instance, the inputs and outputs of OB models can be mapped with sensors and objects in BAS to enable occupant-centric building controls. Additionally, this standardization will ensure that the functionalities and requirements of an OB model are in alignment with building control requirement.

1.1. Existing reviews

The state-of-the-art, as well as an overview of related reviews that focus on OB modeling, are presented by Refs. [9–12], while the human dimension of energy consumption is reviewed by Ref. [13]. As concluded in the work by Carlucci et al. the predictive OB models are emerging, and this trend is evolving in parallel with the rise in the number of data-driven OB models. In this place, such predictive nature of data-driven OB models makes them promise for the application in advanced building controls such as model-based and model predictive controls. For the detailed revision of OB in the context of building control the reader is referred to Refs. [7,14–16]. Furthermore [17], reviewed occupant-centric control strategies, while the OB modeling was not in the particular focus of the latter work.

Complementary to the reviews of general OB modeling, the OB in the context of building simulation and in the context of building control has also been the focus of several recent studies [18–22].

1.2. Contribution of this paper

This work aims to fill the gaps required for the inclusion of OBs into building control, by proposing a guideline for model documentation and evaluation based on a comprehensive review of the scientific evidence and current state of technology. Since the building control and OB modeling were researched separately during the past years, the literature evidence did not provide a clear set of OB models that are developed and implemented in building control. For instance, OB models were commonly developed with a mentioned practical application for HVAC control, general building automation of smart buildings. However, the existing literature evidence does not provide clear recommendations, on which OB models can be used in building control and how to document these models for their real-world deployment. In order to bridge the gap between the two communities, we relied on our best domain expert knowledge and considered the OB models that are applicable to the building control.

From the control side, we put the spotlight on the OB models for the application in rule-based and more advanced control such as model based predictive controls. Further adaptive control paradigms that could include, but are not restricted to reinforcement learning, are not considered in the scope of this work. In this place, comprehensive and unified model documentation is required for model standardization and wide applicability. This model documentation also includes the guideline for suitable model performance evaluation, which is of crucial importance for the realistic presentation of the model's capabilities. In summary, this study aims to: (1) standardize OB model documentation to promote transparency through clear communication among researchers, reproducibility of experiments, (2) help researchers to select and adopt suitable models to fit their research needs, and (3) help researchers to understand the prerequisite, performance, application of the models they intend to use. In order to fulfill these goals, this work focus on the following research questions:

(1) How are occupant models for real-time/predictive controls currently documented in the scientific literature?
(2) How should occupant models be documented and implemented?
(3) What are the evaluation metrics for different occupant behavior models?
(4) What are the software platforms for future researchers to evaluate/validate their models?

2. Methodology

2.1. Guideline development

This section documents the development of the proposed guideline. As Fig. 1 shows, four major parts are included in this guideline: 1) Model description and applications which describes information representation, model inputs and output, and domain of applicability; 2) Model development detailed out data preparation, modeling formalism, and gaps in current model development documentation; 3) Model evaluation provides guidelines of selecting performance metrics which include absolute metrics, domain metrics, and indirect performance metrics; 4) Finally, in model implementation, we discussed the computational environment, computational time, experiment setup, and integration into MPC.

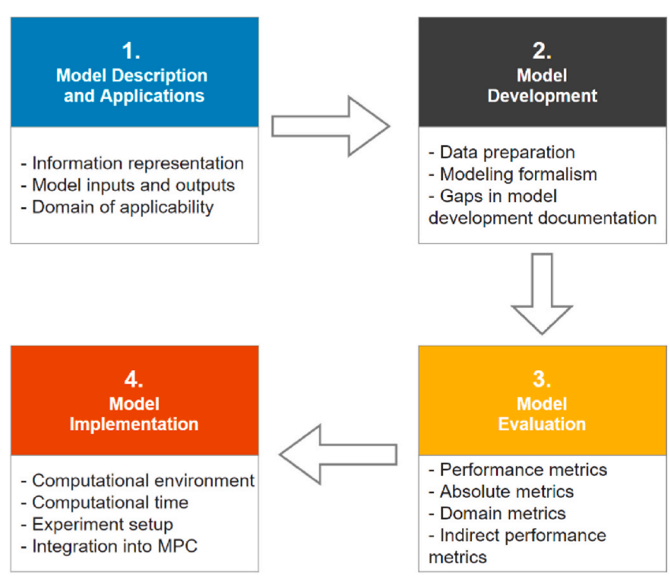

Fig. 1. Overview of the guideline development.

2.2. Review approaches and structure of the article

Based on an in-depth literature review process, this article aims to provide a guideline for a thorough and standardized occupant-behavior model documentation. The review focuses on six different categories of OBs, including Appliance Use, Lighting Operation, Occupancy Estimation and Prediction, Shading Operation, Thermostat Adjustment, and Window Operation. The literature search was conducted in Google Scholar with "Building" plus the aforementioned categories as keywords. Following the pre-defined categories, all the related literature was selected. Among those literature, OB models were reviewed from different perspectives, such as model description and applications (Section 3), model development (Section 4), model evaluation (Section 5), and model implementation (Section 6). We discussed our findings and future challenges in Section 7, and Section 8 concludes this paper.

3. Model description and applications

In order to answer the first research question, we have conducted a review on how current OB models are presented. The description of an occupant model typically includes three parts: information representation, model inputs, and model outputs. We will discuss those elements in the following sections. In addition, the domain of applicability is reviewed as well.

3.1. Information representation

Several ontologies and schemas, such as Industry Foundation Class (IFC), Green Building XML (gbXML), BPD Ontology, Brick, ASHRAE 201,. Have been developed to organize knowledge and structure data by describing both static (e.g. building geometry) and dynamic building data (e.g. time series temperature data) about building technical system, equipment, sensors, and corresponding relationships [23]. Each ontology or schema has its own focus area. For example, gbXML has been used to represent mostly energy performance simulation models with detailed material properties and geometries. BPD Ontology and Brick focus more on building operation data, which is typically measured by physical sensors and has a relationship with its location and measurement type. All ontologies or schemas aim to describe data and their relationship with building's devices. However, it is concluded that there is a lack of detailed documentation on existing data sets and models mostly due to the lack of guidelines as described in the discussion section. In addition, there is no metadata schema or ontology that can represent the full spectrum of occupant behavior models. For example, the occupant presence potentially could be represented by IFC as the "Timeseries" and attached to the "IfcOccupant" Class. However, the description of other occupant behaviors is very limited. Another necessary part of the occupant behavior modeling is a systematic description of input variables and parameters, prediction horizon or time interval. According to the review done by Na [8], eight out of 24 selected data tools can represent indoor and outdoor environmental data; however, none of the existing tools can store occupant behavior model parameters, unfortunately.

Furthermore, the terminology is an essential part of information representation. Na [8] concludes that only three data tools have defined terminology for occupant behavior-related data, which are ADI [24], Brick [25], and Project Haystack [26]. Within those three different metadata schemas, they have different naming for the same building component. The current lack of standardization in the names of sensors in commercial buildings creates challenges not only for occupant behavior modeling but also for building data integration and interoperability in general. There is a need to adapt different naming from the schemas and ontologies to have a unified naming guideline when documenting occupant behavior models.

3.2. Model inputs and outputs

The current section details the inputs (independent variables) and outputs (dependent variables) used in the OB models reviewed in the current work. The information is presented for the six main categories of behaviors considered in the models: (1) Appliance use, (2) Lighting operation, (3) Occupancy estimation and prediction, (4) Thermostat adjustment, (5) Shading operation, and (6) Window operation.

Fig. 2 presents a count of the inputs and outputs used when modeling appliance use (left side of the figure) and lighting operation (right side of the figure). Starting with the former, the most commonly studied outputs are predicting the multi-state of appliances, or their energy consumption levels. The most frequent inputs used to predict the mentioned output are mostly the plug-load energy (i.e., historical data used to predict future use), followed by the space's occupancy status (i.e., using occupancy presence/absence information to predict appliance usage). As for lighting operation (Fig. 2, right), the most frequent outputs are the state (either binary or multi-state) and the operation time. Illuminance levels, occupancy status, and power consumption of other systems (e.g., plug-loads) are the main inputs used to model lighting operation.

Fig. 3 summarizes the inputs and outputs used in occupancy estimation and prediction (left side) and thermostat adjustment (right side). Starting with occupancy models, the two most dominant outputs are the presence status (binary) and the number of occupants. Unlike the previous target behaviors, a wide variety of inputs is used to predict occupancy, including historical occupancy patterns, motion detection, power usage, and indoor environmental measures (e.g., illuminance, temperature, relative humidity, CO2, and VOC levels). As for thermostat adjustment (Fig. 3, right), the temperature setpoint setting is the most frequent target variable. Other outputs include indoor temperature, the probability of adjusting thermostat settings, and energy consumption. Here again, various parameters are used as predictors for these models, such as indoor/outdoor temperatures and humidity, solar radiation, CO2 levels, hour of the day, and electricity load and price.

Fig. 4 presents the count of inputs and outputs used in models of shading operation (left side) and window operation (right side). For shading operation, the listed outputs are all well represented in the reviewed models. They include the shading state (binary or multi-state), the probability of having blinds up or down, and the portion of blinds up or down. The predictors of the stated outputs are primarily environmental in this case, namely indoor/outdoor temperature, illuminance, and solar radiation. Moving to the right side of Fig. 4, the most considered output of window operation is the probability of window state, followed by the probability of taking action (e.g., opening/closing a window) and the portion of a window open, respectively. Here again, the inputs to such models are mostly environmental, namely indoor/outdoor temperature and humidity, wind speed and direction, solar radiation, rainfall, and concentrations of CO2 and particulate matter.

3.3. Domain of applicability

This section details how the spatio-temporal domain is documented in the OB models reviewed in the current work. The temporal dimension is represented by time granularity, prediction horizon, and control horizon. While the spatial dimension is represented by the space (e.g., room, floor, building level) the OB model is addressing.

The time granularity is the time-step or shortest time window operation from which the information regarding the occupant's behavior is used for prediction (e.g., presence model in 15-min resolution). Fig. 5 shows the distribution of the availability of this information in the reviewed papers according to each target behavior. Models of lighting operation are the least documented in terms of time granularity. For the other target behaviors, about 30% of the papers do not report the used time discretization information. Sometimes, the time-step is not explicitly documented because the authors imply that the time granularity of the model is the same as that of the sensed data. The

Appliance Use Inputs↓ Outputs→	Multi-State	Binary-State	Energy Consumption	Prob. of On/Off	Presence	Schedule
Visual Data	1	0	1	1	2	0
Acoustic data	2	0	0	0	0	0
Plug Load	16	0	11	0	1	1
Price	1	0	0	0	0	0
Wireless Signal	1	0	1	0	2	0
Tin	1	0	0	0	0	0
Tout	1	0	1	0	0	0
Door Operation	1	0	0	0	0	0
Presence	2	0	7	0	1	2
RH(in)	0	0	1	0	2	0
CO2	1	0	0	0	2	0
Wind Speed	0	0	1	0	0	0

Lighting Operation Inputs↓ Outputs→	Binary	Multi-State	Portion of Light On	Visual Comfort	Probab. of Switching	Power Consumption	Operation Time
Tin	0	0	0	0	0	1	0
Tout	1	0	0	1	0	1	0
RH(in)	0	0	0	0	0	1	0
RH(out)	0	0	0	1	0	1	0
Wind Speed	1	0	0	1	0	0	0
Illuminance	11	21	1	1	1	3	6
Solar Radiation	1	1	1	1	0	1	1
Occupancy	5	5	1	2	0	1	11
Power	2	4	2	1	1	2	10
Color	0	1	0	0	0	0	0
Visual Comfort	0	2	1	0	1	1	0

Fig. 2. Count of inputs and outputs used in models of appliance use (left) and lighting operation (right) behaviors.

Occupancy Estimation and Prediction Inputs↓ Outputs→	Presence (Binary)	Occupant Number	Percent Occupied	Position
Occupant Entry/Leave	0	2	0	0
Occupant Number	0	3	0	0
Presence	6	2	1	0
Motion	3	5	0	0
Illuminance(in)	3	3	0	0
Tin	4	2	0	0
RH(in)	4	2	0	0
LED Reading (Volt)	0	1	0	0
CO2 concentration	5	7	0	0
VOC	2	0	0	0
Power Usage	6	1	0	0
echo intensity	1	0	0	0
Acoustic Level	1	3	0	0
Telephone State	1	0	0	0
Keyboard/Mouse State	0	1	0	0
Chair State	0	1	0	0
AC State	1	0	0	0
Door State	1	1	0	0
Light State	0	1	0	0
Window State	1	0	0	0
Signal Strength	0	0	0	2
Number of Device on	0	1	0	0
MAC address	0	3	0	0
Device Location	0	1	0	0

Thermostat Adjustment Inputs↓ Outputs→	Multi-State	Binary-State	Energy Consumption	Cycle Idle Time	Prob. of Adjustment	COP	Tin	Tsetpoint	Portion of Adjustment	Tsupply
Tin	1	1	5	0	5	1	4	12	2	0
Tout	1	1	3	0	4	1	5	10	3	1
Presence	0	1	2	0	0	0	2	6	0	1
RH(in)	0	1	0	0	2	0	1	6	0	0
RH(out)	0	1	0	0	1	1	1	4	0	0
Solar Radiation	0	1	2	0	1	0	2	3	0	1
Wind Speed	0	0	1	0	1	0	0	1	0	0
CO2	1	0	1	0	2	0	1	3	0	0
Tsetpoint	0	0	1	0	0	0	0	0	0	0
WinOpen	0	0	1	0	1	0	0	1	0	0
ShadeOpen	0	0	0	0	1	0	0	0	0	0
Hour	0	0	0	0	0	0	0	3	0	0
Electricity Load	0	0	1	0	0	0	0	4	0	0
Electricity Price	0	0	0	0	0	0	0	3	0	0
Air Movement	0	0	0	0	0	0	0	1	0	0
Occupant thermal	0	0	0	0	0	0	1	1	0	1
Clothing thermal	0	0	0	0	0	0	0	1	0	0

Fig. 3. Count of inputs and outputs used in models of occupancy (left) and thermostat adjustment (right) behaviors.

documented time granularities cover a broad range from less than 1 min to hours (Fig. 6). This depends on several factors, including the granularity of the sensed data available, the temporal range in which the change of behavior in question occurs, and the envisaged predictive horizon. A time resolution between 10 and 19 min is the most frequently adopted.

The predictive horizon is the time horizon over which the OB is modeled. The predictive horizon is much less documented than the time granularity (Fig. 7) and covers a wide range of values, from less than 1–24 h (Fig. 8). This is to be expected since it is strongly dependent on the controlled variable. For example, the predictive horizon for building predictive HVAC control is related to the type of heating and cooling system and can be relatively long for radiant floor heating systems compared to air-based ones, the response time of which is very fast.

The control horizon is the time horizon over which the control variable is modeled. The control horizon is commonly equal to or longer

Shading Operation Inputs↓ Outputs→	Shading (Binary-State)	Shading (Multi-State)	Prob. of Blind Up/down	Portion of Blind Up/down
Tin	4	4	4	3
RH(in)	1	2	4	2
Illuminance(in)	3	6	3	2
VOC(in)	0	0	2	1
CO2	0	0	1	0
Air Velocity(in)	1	0	1	2
Heating/Cooling	1	0	0	1
Lighting Load	1	0	0	1
Tout	3	5	5	3
RH(out)	0	3	2	0
Illuminance(out)	1	0	0	1
Solar radiation	3	5	4	4
Solar	0	2	1	0
Wind Speed(out)	0	1	2	0
Heating/Cooling	1	0	0	1
Lighting Load	1	0	0	1

Window Operation Inputs↓ Outputs→	Multi-State	Prob. of window state	Prob. Pf Action	Time of Opening	Portion of window open
Tin	2	20	15	1	10
Tout	2	20	15	0	10
RH(in)	1	10	9	1	3
RH(out)	1	13	8	0	4
Air Velocity	0	0	0	0	1
Wind Speed	1	12	8	0	5
Wind Direction	1	9	2	0	3
Illuminance (in)	0	0	0	0	0
Illuminance (out)	1	2	1	1	1
Solar Radiation (out)	1	5	6	0	3
Sunshine Hour/ Date/ Time	0	4	1	0	2
Rainfall	0	6	2	0	1
CO2	0	12	8	0	5
PM2.5/PM10 (out)	0	6	2	0	2
VOC	0	0	1	0	1
Image	0	1	0	0	0
Position sun protection	0	1	0	0	0
Occupancy	1	2	1	0	1
Person charactersitic	1	1	1	0	1
Room charactersitic	1	1	1	0	1
Time of day	1	1	1	0	1

Fig. 4. Count of inputs and outputs used in models of shading operation (left) and window operation (right) behaviors.

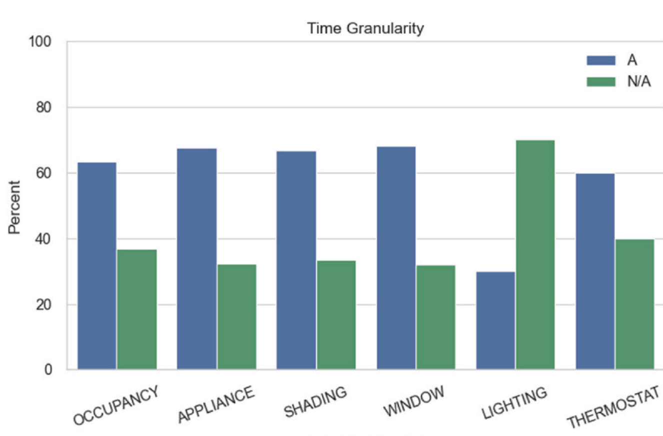

Fig. 5. Availability of the time granularity according to the target behavior.

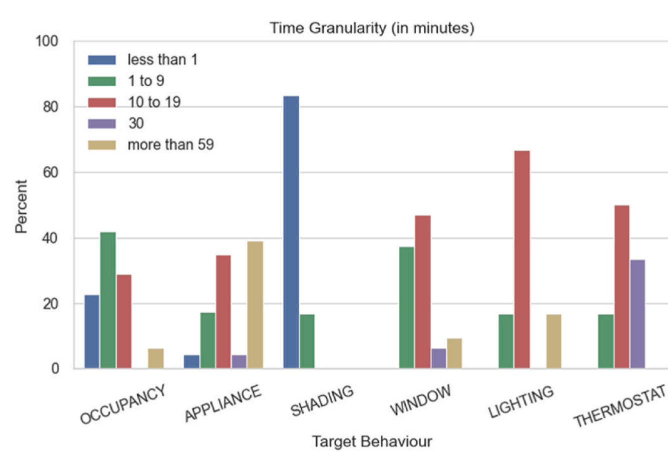

Fig. 6. Distribution of the time granularity in minutes according to the target behavior. (If a model was developed with more than one time-step; only the smallest was considered in the figure).

than the predictive horizon. The control horizon is explicitly documented in only 5% of the reviewed papers. The lack of this information can be explained by the fact that the great majority of the reviewed papers propose OB models as an input for model predictive control (MPC) but do not actually apply it in predictive control. However, the practical implementation of these OB models in the field is currently lacking.

Regarding the space granularity, i.e. the space (room, floor, building level) the OB model is addressing, a majority of occupancy, shading, and lighting models have been developed only at room level (Fig. 9). Instead, appliance use, window operation, and thermostat adjustment models have been mostly addressing the building level that, for the residential case, corresponds to an entire house. Only a minority of OB models have been addressing a lab-based installation, such as a test cubicle.

4. Model development

To develop an occupant model, one needs to prepare raw data into a format that can be used for modeling data (4.1) and identify a modeling method or algorithm that is applicable and practical for a particular

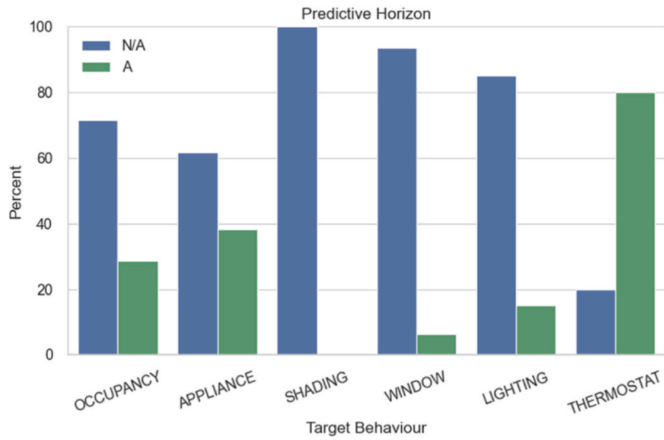

Fig. 7. Availability of the predictive horizon according to the target behavior.

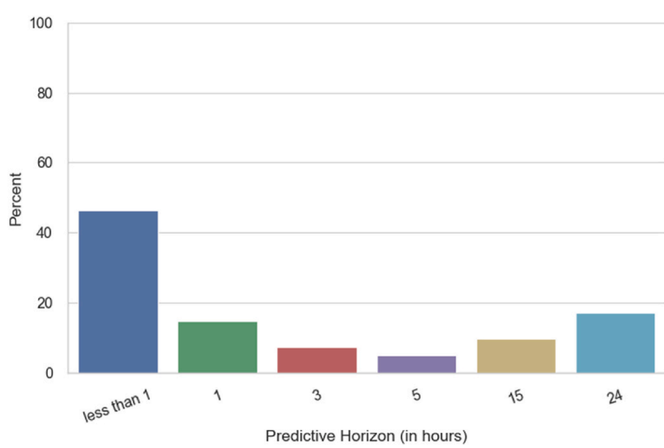

Fig. 8. Distribution of the predictive horizon in hours according to the target behavior. (If a model was developed with more than 1 predictive horizon; only the smallest value was considered in the figure).

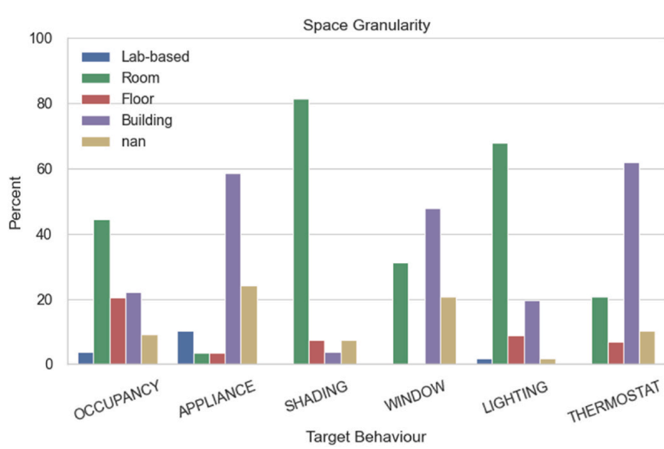

Fig. 9. Distribution of the space granularity for occupancy models.

problem (4.2). In this section, we will review diverse approaches and techniques employed in occupant models and discuss gaps in current research and development in this field (4.3).

4.1. Data preparation

Data used in occupant modeling is often collected in different structures, granularities, and volumes as described in section 2. Hence, it is important to prepare or preprocess the raw data into a format that is suitable for intended analysis and modeling [27]. Preprocessing can include, but is not limited to, the following steps:

1. Cleaning and imputing the missing and corrupt data, outliers by discarding or replacing them with inferenced values (e.g., moving average, mean, or median) to be easily parsed by machine;
2. Reducing data dimensions using row-wise for data sample reduction or column-wise for data variable reduction and random sampling methods;
3. Data scaling using min-max normalization [28], distribution-based standardization [29], or structure-based techniques [30] to scale the data into a consistent range;
4. Feature creation to construct new variables of existing features for data analysis;
5. Data partitioning using supervised (e.g., decision tree [31]) and unsupervised techniques (e.g., k-means [32], Gaussian Mixture Models [33]) to divide the dataset into the test and training subsets to evaluate the trained model based on the test set;
6. Merging data from different sources with various time intervals within time-series data.

Preparing data is often the most time-consuming portion of data-driven modeling. Yet, only a few studies in the reviewed literature describe the data preprocessing methods used for occupant models. The examples are as follows. Jin et al. [34] used the confusion matrix to evaluate the quality of PIR sensor data and remove inaccurate occupancy states. Q-test was used to identify outliers [10]. To handle the missing values, Yu et al. [29] used the moving average method to fill in missing entries. Also, they calculated the SHapley Additive exPlanation (SHAP) values of each feature to reduce the HVAC operation data dimensions. Ashouri et al. [28] employed min-max normalization to standardize energy consumption data. K-means clustering was used to recognize distinct air handling unit operation patterns and group the BAS data accordingly [29]. In another study, hierarchical clustering was conducted to extract the occupancy patterns in the building and create new features for occupancy models [35]. Given that methods and assumptions used in the preprocessing stage can affect data analysis and prediction outcome, there needs to be a concerted effort to document detailed preprocessing steps in future studies.

4.2. Modeling formalism

In this section, the models' category distribution is presented for each target behavior. The modeling categories were defined based on the state of the research: deterministic rule-based models, statistical/stochastic, and data-driven [36]. Rule-based models are the deductive models that use an a priori set of rules for describing occupant behaviors in building models, including deterministic models and schedules. The statistical/stochastic is stochastically modeled the OB to represent the various behaviors among the population [9], potential change over time [18], and complexity [10]. These models are commonly represented by statistical models such as a-priori probability density functions [37]. The third modeling formalism is the data-driven modeling, where the focus was put on computational intelligence or machine learning without an explicit aim to explain the relationship between the input variables and the OB [38]. It includes the ML models and ABM.

The reviewed OB models for building control are screened for the used modeling formalism, and the results are presented in Fig. 10.

The data-driven methods are the most used. The second most implemented category is represented by statistical models that have been applied especially to model windows operation (80%). Rule-based models' category has been applied mainly for shading and has not been tested for appliance use and window operation behavior. Some documents did not provide information about the implemented models. It

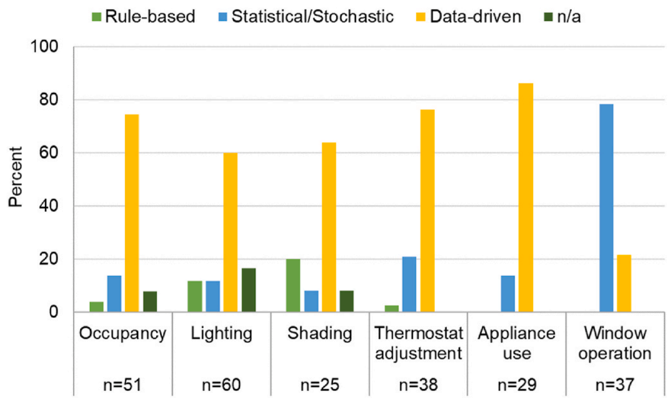

Fig. 10. Distribution of the model categories according to the target behavior.

happened for occupancy, lighting, and shading behavior in a percentage of 8%, 17%, and 8%, respectively.

Fig. 11 deepens the most adopted models in the target behavior, considering those with a percentage higher than 10%. Regarding the rule-based category, most of the reviewed studies do not provide detailed information on the models used, merely defining their belonging category. Schedules have been implemented in around 40% of the cases related to lighting usage.

About the statistical models, the Markov chain appeared the most adopted for occupancy and thermostat adjustment; Markov chain Monte Carlo models have been often implemented for modeling lighting use. Regression models have been mainly found in the case of window and shading operations with a percentage of around 74% and 100%, respectively. Often, information on the adopted model was not provided, as in the case of appliance usage (67%).

Regarding the data-driven models, neural networks (NN) have been the most common. Control logic and fuzzy logic were utilized with the same percentage of neural networks (around 20%) for shading operation. Similarly, also in the case of appliance use, clustering and long short term memory (LSTM) were implemented with the same percentage of neural networks.

Fig. 12 provides the distribution of models' categories in six building types. Rule-based modeling has been mainly applied in commercial buildings for occupancy, lighting, and shading operation. Furthermore, they have been implemented in residential buildings for occupancy, in educational buildings for lighting, and in institutional edifices to model shading operation. The statistical approach has been principally used in residential buildings and prevailed in the case of thermostat, appliance use, and window operation; the application in commercial buildings was diffuse in case of occupancy, shading, and window behavior. In the case of lighting, the implementation of statistical models was equally distributed in residential and commercial buildings. Moreover, statistical models are diffuse to predict occupancy and window operation in

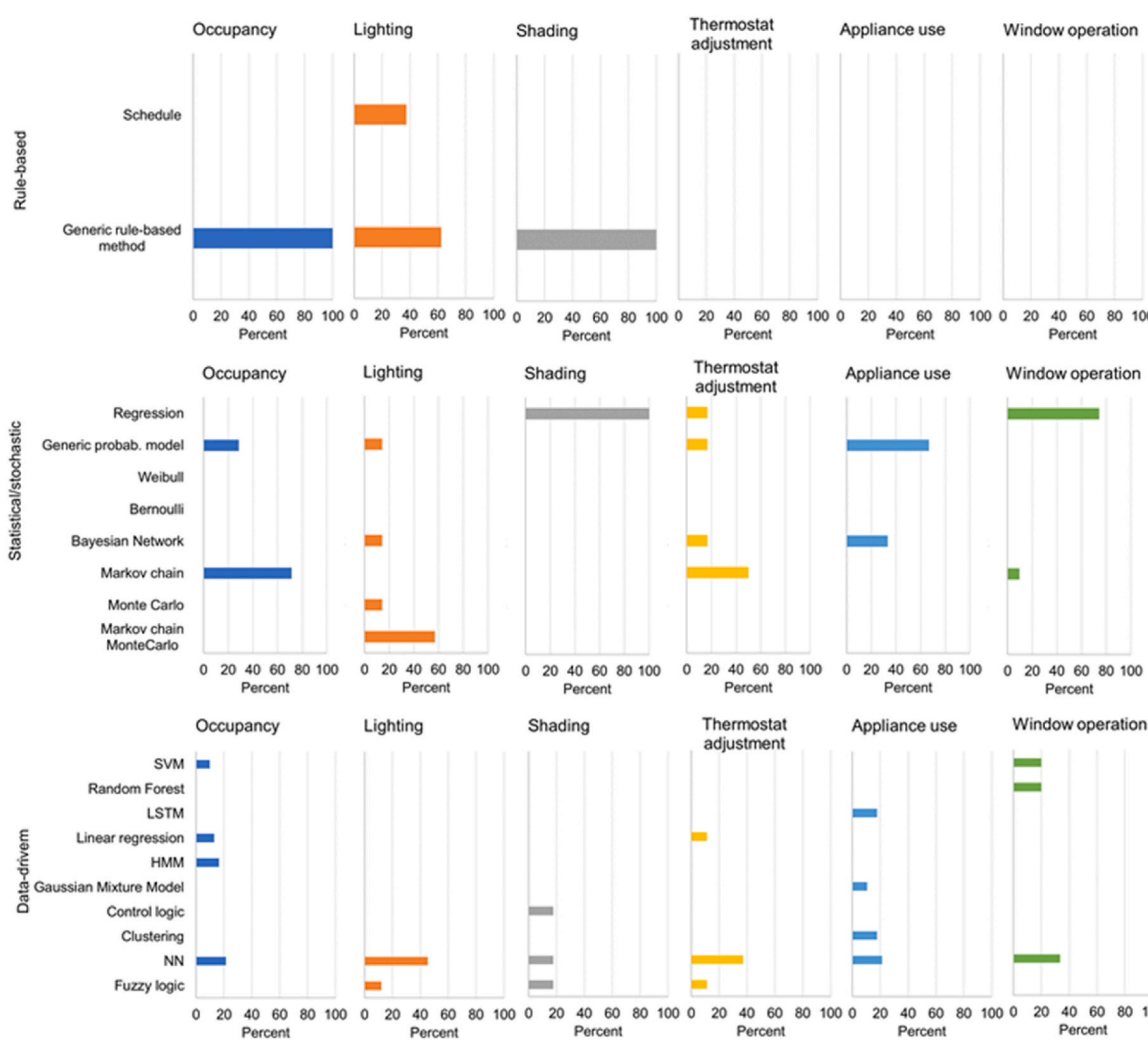

Fig. 11. Most adopted models according to the target behavior and the model categories (rule-based, statistical/stochastic, and data-driven).

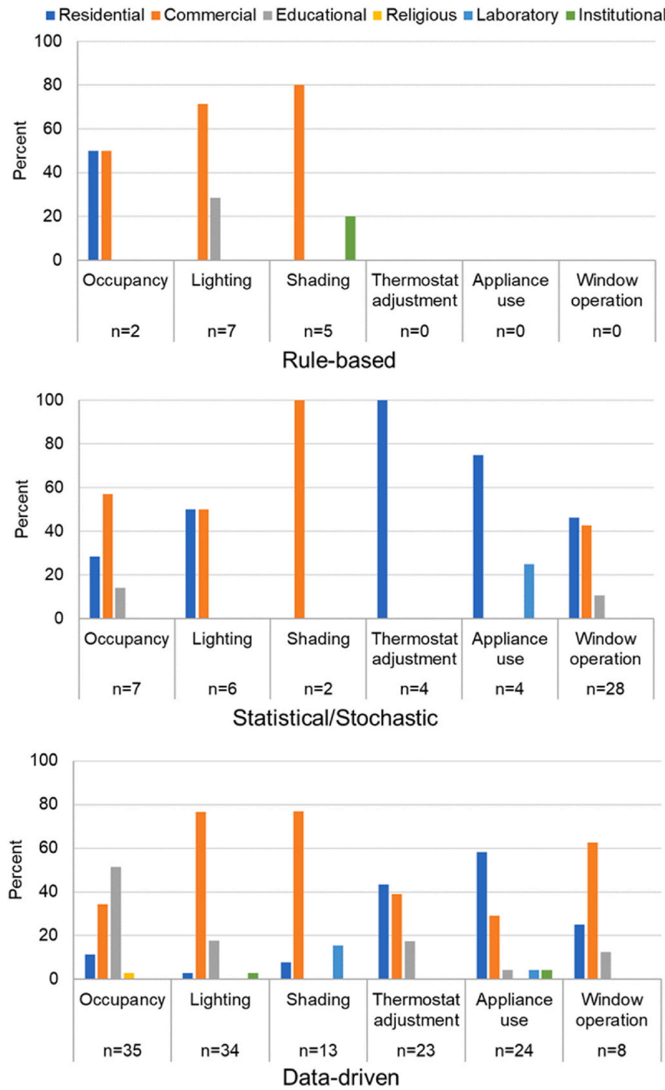

Fig. 12. Model categories (rule-based, statistical/stochastic, and data-driven) according to the target behavior and building typology.

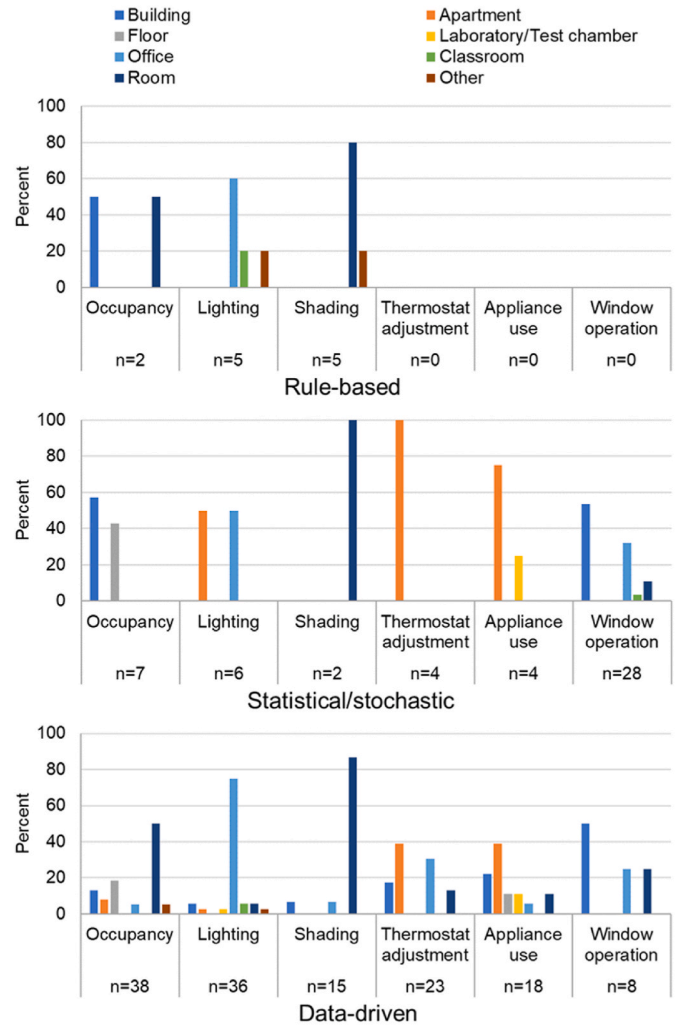

Fig. 13. Spatial granularity in model categories (rule-based, statistical/stochastic, and data-driven) according to the target behavior.

educational edifices, whereas appliances use in laboratories.

The data-driven category found notable implementations in commercial buildings for lighting, shading, and window operation in both residential and commercial buildings for thermostat adjustment and appliance use. Educational buildings prevail for occupancy modeling, and laboratory edifices for shading operation. Other buildings typologies, such as religious and institutional buildings, have been rarely investigated for occupancy, lighting, and appliance use.

Fig. 13 presents the model categories distribution considering the space granularity. In general, shading and occupancy behavior have been investigated at room level and lighting in offices.

On the other hand, thermostat adjustment and appliance use have been modeled in apartments as they are principally studied in residential buildings.

Occupancy is also frequently detected in buildings, whereas lighting operation has been modeled considering more varied space typologies: controlled environments, such as laboratories and test chambers, classrooms, and offices.

Fig. 14 presents the model categories distribution considering the time granularity. Generally, rule-based models were developed collecting occupancy data with a time step minor than 10 min, whereas the temporal time step was prolonged and longer than 60 min in case of shading behavior. In statistical models, occupancy, thermostat, and window were commonly detected with a time step minor than 20 min. Lighting operation behavior was mainly modeled by data collected with a temporal step minor than 10 min. In the data-driven category, the time step used for appliances usage detection was generally equal to 45 min. For the other behaviors, brief time steps, less than 20 min, were also utilized.

4.3. Gaps in model development documentation

Scientific research should be ‘reproducible and replicable’. Reproducibility ‘means obtaining consistent results using the same input data, code, computational steps, and conditions’ while replicability ‘means obtaining consistent results across studies aimed at answering the same scientific questions using different data (https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science). Reproducibility is challenging to attain because it involves sharing the data, which may be nowadays limited by, for instance, personal data protection needs and privacy issues. Replicability could be easily achievable compared with reproducibility. However, it demands documenting the steps undertaken in the development of the model in a transparent and detailed way. Therefore, the critical aspect is to detail the entire process – the complete workflow – rather than a specific part of it (e.g., the results).

With these premises, the model development should start with the explicit formulation of its problem; that is the modeling purpose. This implies that if a model is developed for control purposes, it should be

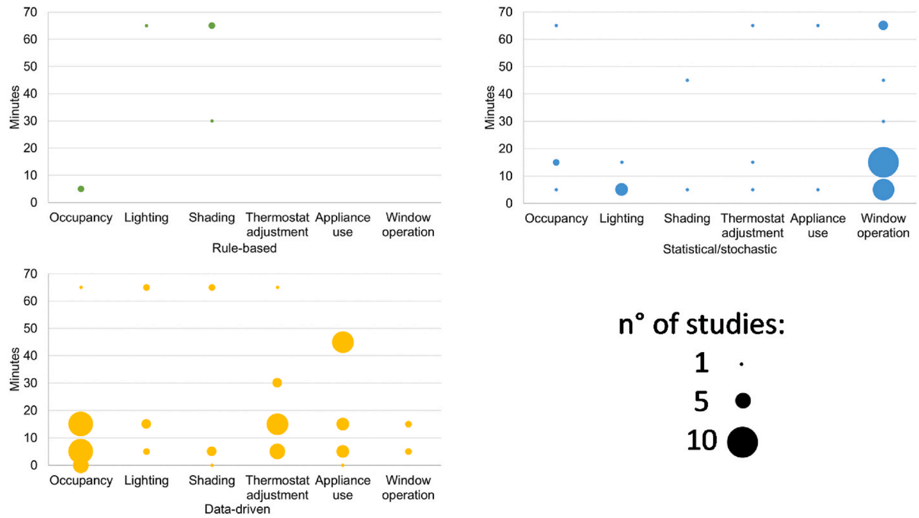

Fig. 14. Time granularity in model categories (rule-based, statistical/stochastic, and data-driven) according to the target behavior.

ready to be directly implemented in control logic. Consequently, this aspect should be already addressed in the model development. For example, for application in model based control or MPC, the control horizon depends on the kind of heating and cooling system used: it can be rather long for radiant floor heating systems compared to air-based systems. Subsequently, the modeling formalism that provides the OB sequence during the whole predictive horizon should be chosen and described. Reasons for the selection of a particular method should be given from both a practical and theoretical point of view. It should also explain how the predictors included in the model were chosen. This includes stating whether feature selection was used to reduce the data dimension and which approach was used. Also, it is important that the variables included in the model are commonly monitored in buildings. For instance, if RH is included in the model as a predictor but is not measured in the building, the model is not applicable in practice.

From the reviewed literature, only 35% of the papers state that the aim was building/zone/HVAC control application, and only 19% offer a formal integration of the model into a control logic. Therefore, the majority of the available occupant models are not designed with control purposes in mind, which directly impacts the modeling formalism used. This kind of model aimed to represent what the behaviors were based on the data collected and not what the behavior will be. In models for control purposes, time becomes a critical factor that a model should directly account for. This translates to recognizing and considering the model's input variables as a function of time.

Another issue regards the integration of real-time data into the model to be updated when new information is collected. This would also require measures of the dependent variable, that is, the behavior that the model aims to predict. The ability to self-update and adapt to real-time data strongly affects the modeling strategy's choice.

Furthermore, most of the behaviors are modeled independently from each other. However, in reality, this is rarely the case. For instance, the shading operation can affect both lighting operation and thermostat adjustment behavior. Moreover, the necessary condition for most of the occupant behavior is occupancy estimation and prediction. This implies being able to measure it or predict it. In the latter case, this will result in ulterior uncertainty in predicting behavior. For example, if window operation behavior is the dependent variable to predict, its prediction will be affected by the model error plus the prediction error for the occupancy status.

5. Model evaluation

Mode evaluation is a logic and necessary next step after model development. This section focuses on the model evaluation and documenting the model's performance. It consists of the reviewed literature evidence on the OB model evaluation and a proposed guideline for the standardized documentation of model performance. For that purpose, the performance metrics are structured into absolute, domain-specific, and indirect metrics, and their purpose is briefly elaborated. Finally, the sensitivity analysis is introduced, as an additional tool to quantify the model performance and document the uncertainties.

5.1. Performance metrics

The model evaluation [39] is structured into absolute evaluation, domain metrics, and indirect metrics as showing in Fig. 15. The absolute metrics relate to the performance indicators used for general statistical or data-driven modeling. Here, we quantify how often an OB model provides a correct prediction, or we use the absolute metrics to evaluate the performance in case of data imbalance. The domain metrics are defined using the OB and buildings physics knowledge. For instance, in the case of window operation modeling, we are not only interested in what percentage of window states is modeled correctly, but also how often a window operation occurs or what the median duration of sequences with open windows is. Lastly, the indirect metrics quantify the impact of the modeled OB on the data modeling objectives of building control: does the use of the window operation model lead to improved thermal comfort, or how does it affect the HVAC system, such as the resulting impact on energy consumption or thermal comfort.

5.1.1. Absolute metrics

The absolute metrics is based on the definition proposed by Ref. [40], namely "the metrics that are based on the absolute error calculation". The main goal is to assess the goodness of the model for fulfilling a particular task, to compare alternative approaches, and to quantify if the design updates made on an OB model led to the model's improvement. The initial step in the selection of the evaluation metrics is the assessment of the nature of the modeling objective; for example, whether the target variable is categorical or continuous.

Since the absolute metrics may result in bias in the interpretation of the results [41], these should be selected based on the nature of the modeled data and target function formulation. In this regard, the target function formulations considered for the OB modeling are continuous, categorical and the special case of binary categorical variables. The resulting absolute evaluation metrics should be defined based on the target function formulation and the particular challenge in each OB model.

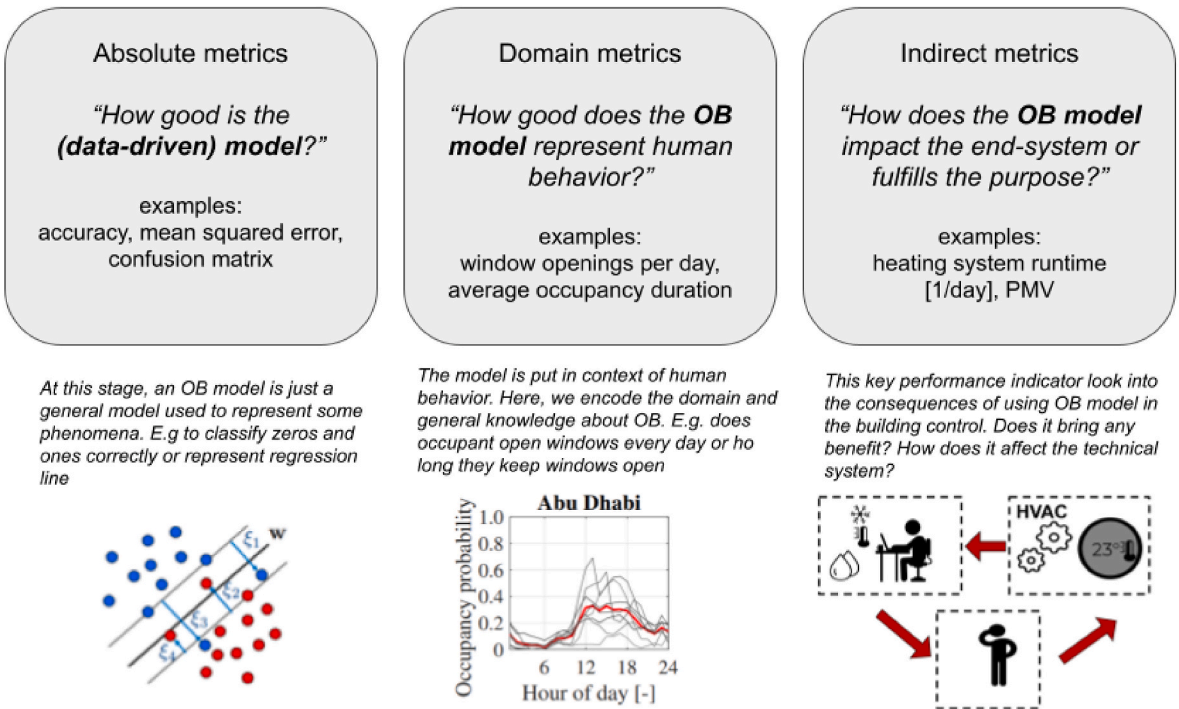

Fig. 15. Model evaluation metrics for OB in building control.

The literature screening showed that there is a significant portion of the modeling studies in which the validation and testing performance was not performed. Furthermore, the conducted literature review pointed out the lack of standardized model evaluation metrics. However, similarly to the generic data-driven or statistical modeling, the performance is commonly reported using mean average error (MAE), mean squared error (MSE). Additionally, the precision, recall, and F1 scores were used by some of the existing studies that focused on classification tasks.

Based on the empirical evidence on the target variable formulation and the nature of OB data, the minimal requirements on the set of absolute metrics for each OB are summarized in Table 1. In summary, most of the OB modeling should be treated as classification problems. The multi-class categorical target functions include the shading operation, while the window operation, occupancy estimation and prediction, lighting operation, and appliance use should be modeled as the binary classification. At the same time, the occupancy count can be modeled as a regression problem.

We use modeling window status as an example. The fundamental issue to be addressed is the imbalanced prior probabilities of the window states. Therefore, the model evaluation should include the MAE, confusion matrix, and F1 scores. Similar to the window states, the occupancy estimation and prediction, appliance use, and lighting operation are also commonly formulated as binary classification problems, and the model's performance can be quantified using the same metrics. To this end, the model goodness quantified using confusion matrix and F1 score could also be reported using precision and recall. The models that represent shading operation can be evaluated using MAE, confusion matrix, and Fβ score. The reader is referred to Li et al. [42] for further elaboration on the choice of the evaluation metrics.

In the case of the thermostat setpoint modeling, there are very limited studies on model validation. However, we argue that the setpoint modeling should be treated as a continuous problem since the setpoint changes could be treated as rare events. By treating the thermostat set point modeling as a regression problem, the relative value of the thermostat set point is to be modeled, while the setpoint changes would be addressed in an implicit fashion.

5.1.2. Domain metrics

The domain metrics are defined as a fit-for-purpose metric that evaluates the competence of the model in representing a certain OB considering the stochasticity of results. The intention of developing these metrics is to provide comparable means for assessing how well do OB models represent particular forms of human behavior. Moreover, considering the main purpose of building as a comfortable and productive space for people [43,44], these metrics standardize building performance from the perspective of its occupants. In the existing work, Tahmasebi and Mahdavi [45] presented domain metrics for comparing

Table 1
Recommended target formulation and minimal set of evaluation metrics for each modeled OB.

	Data type			Absolute metrics								
	Continuous	Categorical	Binary	ACC	Balanced accuracy	Confusion matrix	F1	Fβ	MAE	MSE	RMSE	*N*-RMSE
Window Operation		X	X	X		X	X		X			
Thermostat Adjustment	X								X	X	X	X
Occupancy Estimation and Prediction	X	X	X	X		X	X		X	X	X	
Shading Operation		X		X		X		X	X			
Lighting Operation		X	X	X		X	X		X			
Appliance Use		X	X	X	X	X	X		X			

the performance of OB models in building simulation. However, the related literature in the context of building control is sparse and there is a need for developing consistent domain metrics [46–48].

The domain metrics are categorized based on types into aggregated and interval-by-interval groups as summarized in Table 2. The aggregated domain metrics stand out as proper evaluation criteria when the tested OB model is used for long-term purposes such as (estimating total energy saving, benchmarking building performance, etc.). On the other hand, the interval-by-interval metrics are preferable for evaluating OB models in short-term applications such as (building control, MPC, demand response studies, etc.). Table 1 summarizes the metrics defined for each distinct behavior, and it highlights the research gap in developing novel interval-to-interval metrics for thermostat domains. To fill this gap, we argue that some of the developed metrics could be transferred between different domains with minor adjustments. For example, the occupancy state matching error is the percentage of false state predictions which indicate the mismatch between actual and predicted occupancy; this metric could be adjusted for evaluating appliance use models. Chong et al. [46] used the coverage width-based criterion that comprehensively evaluates the quality of prediction interval to evaluate the performance of occupancy estimation and prediction models which can easily be adjusted to evaluate the performance of OB thermostat adjustment models.

Table 2
Domain metrics for each OB type.

	Purpose	Domain	Metric	Refs
Aggregated	Long-term purposes (estimate energy saving, model building performance, etc.)	Lighting Operation	Typical lighting operation profile Frequency of switching-on actions	[48]
		Window Operation	Overall fraction of open state [%] Mean number of actions per day averaged over the observation time Open state durations' median and interquartile range [hour] Closed state durations' median and interquartile range [hour].	[47]
		Occupancy Estimation and Prediction	Occupancy State Matching (SM) error Occupancy Duration (OD) error [h] Number of Transitions (NT) error	[49]
		Appliance Use	Appliance's daily turn on times Appliance's average use duration Accumulated on-state duration	[50]
		Thermostat Adjustment	N. A	
Interval-by-interval	Short-term purposes (demand response, MPC, etc.)	Lighting Operation	The stepwise energy use	[48]
		Window Operation	N. A	
		Occupancy Estimation and Prediction	First Arrival time (FA) error [h] Last Departure time (LD) error [h]	[49]
			Prediction interval (PI) Coverage width-based criterion (CWC)	[46]
		Appliance Use	N. A	
		Thermostat Adjustment	Prediction interval (PI)[a] Coverage width-based criterion (CWC)[a]	[49]

[a] Metrics are suitable to be shifted from different domain.

5.1.3. Indirect performance metrics

Indirect performance metrics evaluate to which extent the OB model contributes to fulfilling the control goal, such as energy consumption reduction or improving thermal comfort. For example, the integration of occupancy estimation and prediction into temperature control can minimize the heating demand. In that case, the used energy could be defined as the control metric together with absolute and direct metrics in the comprehensive model evaluation. One of the examples of jointly used domain and indirect evaluation metrics is presented by Peng et al. [51] the occupancy model was evaluated using both domain metrics (probability and duration of room occupancy) and indirect metrics (total consumed energy consumption).

A summary of literature about control metrics applied for the different OB models can be obtained from Appendix I. The literature screening pointed out that there is only limited evidence of the documented indirect control metrics for OB models, which could be a result of the rare availability of control use cases. When included, control metrics are frequently used to compare the performance of control including OB modeling versus one without it. The control metrics can be absolute or relative, for example, the energy consumption or saving in kWh (absolute) or the energy reduction in percent (relative).

In the literature, where control metrics are available, the main focus of most control algorithms is to minimize energy consumption while maintaining comfort constraints. Naylor et al. [17] reviewed occupant-centric building control strategies in regard to their energy reduction and obtained between 20 and 50% reduction in most cases. Despite the energy reduction, the comfort should usually remain in a certain range, e.g. an indoor temperature between 20 and 23 °C. The most dominant comfort control metric obtained from literature is thermal comfort, as most cases of including OB (occupancy, thermostat, windows, shading) into control are related to HVAC systems [52,53]. Shading operation is not only relevant for thermal but also visual comfort metrics. Indeed, previous work has combined parameters from both domains in an algorithm to control window blinds [54]. Lightning is only related to visual comfort, mainly by guaranteeing appropriate illuminance levels at workstations. By doing so, visual comfort and energy efficiency metrics may be combined - even during non-office hours, using algorithms able to minimize illuminance targets for unoccupied workstations [55]. For thermal comfort, most authors use the control metrics indoor air temperature, the predicted percentage of dissatisfied (PPD), or the predicted mean vote (PMV); for visual comfort, illuminance, or false-off frequency are typically used.

6. Model implementation

This section aims to provide a guideline on the best practices for implementing OB models in a software environment. This part assumes that the creator has completed the model development phase for the OB models so that these models are in the form of a stand-alone application. In this context, we refer to a stand-alone application as "a set of a USER's information processing requirements" [56]. In that context, this section provides recommendations regarding the model's computational environment, runtime analysis and scalability analysis for the real-time capabilities, and the experimental hardware settings.

Primarily, we focus on the OB models' implementation as the outcome of the academic or general research activities. The aim of this section is to propose the documentation that enables model's reproducibility in building control systems that may have different software architectures [57].

6.1. Computational environment

The documentation on the computational environment in the sense of OB models in building control should include requirements for specific operating systems (OS), programming languages, and library or other software dependencies. Beyond this information, the used versions should also be reported. The details required are relevant both for the model's reproducibility in form of the application and due to the copyright requirements of each dependency in case of the (potentially commercial) field deployment. Furthermore, the operating system and programming language should be documented in the context of the runtime evaluation that is a crucial component of the OB model documentation and that is a programming language and OS-specific.

When selecting a suitable computational environment for model development, the evidence regarding the widely adopted environments could be beneficial, and this information is summarized in Table 3. Among others, R and Matlab/Simulink are the most commonly utilized programming language and software packages that can be employed for almost every type of OB. Oppositely, in terms of programming languages C/C++ and VBA and for software packages IBM SPSS and Weka are less used for developing already existing OB models. In practice, one can observe a large variability for different platforms and the computational efficiency of OB algorithms [57]. Nevertheless, information related to the utilized operating system is sparsely documented in existing studies.

6.2. Computation time

Commonly, the OB models are developed with the aim to be included in the end-systems such as building control that typically operates in real-time. Since the computation resources within building control and related end-systems are limited, an estimate of the required resources is required to assure the real-time operation. In this place, we refer to the computation resources of the OB models, which are defined as stand-alone modules that can be coupled with an end-system in various distributed manners. We focus on the runtime of the developed final models where the executed steps include the data-preprocessing and computing the value of the OB target variable. In the case of the machine learning-based models, this would correspond to a model test, while the model training and validation are considered to be previously completed. The hosting of these models is taking place within the building control, using a cloud-edge solution or on a remote cloud. In case of any of the listed computation settings, the following information should be provided:

- In which computation environment is the runtime analysis conducted?
- What is OB model inference runtime?
- Inference memory requirements?
- Optional: total required training runtime
- How does the OB model scale in space and time with the number of modeled OB instances?

The runtimes should be expressed either in core hours or the clock time, given the standard setting. The runtime should be documented together with the used hardware model. Since the majority of the OB models were created in the scope of academic research efforts, there is limited literature evidence on the runtime documentation. Among the others, the reader is referred to Refs. [57–59], and [60] for some best practices.

Additionally, the model's scalability is of particular importance and should be documented. Namely, OB models can be applied to a large number of occupants within a building and therefore the model's scalability in space and time (footnote: for further information regarding the space and time complexity, the reader is referred to Ref. [61] should be documented and expressed using "big O" notation with respect either to the number of occupants, rooms or buildings (further information about the "big O″ notation is summarized by Ref. [62]. Additionally, in case an O model is intended to be used for varied temporal resolution and predictive horizon, the time complexity should also be documented with respect to these two parameters.

6.3. Experiment setup

With the computational environment guidelines discussed above, this section focuses on presenting the experiment setup by summarizing the findings from the literature review. Sensor choices and implementation location will be discussed in the following subsections. The discussions are based on six main categories of occupant behavior models, as followed throughout this paper. The sensor choices subsection offers information of sensors that have been deployed among different studies, implementation locations subsection presents different locations of deployed sensors in different research experiments. This section aims to provide guidelines for future occupant behavior researchers to deploy sensors and set up experiments.

6.3.1. Sensor choices

From the reviewed literature, in total 85% of the studies have explicitly provided information about sensors that have been deployed. Table 4 summarizes the commonly used types of sensors and the aggregated frequency that they have been picked in the literature. The color scales in the table represent how often the specific sensors were adopted. It can be observed that, for "Appliance Use" studies, current/power sensors and smart meters are very commonly used; for "Light Operation" studies, lighting sensors and PIR sensors are primarily adopted; for "Occupancy Estimation and Prediction" studies, PIR sensor and CO2 sensor are commonly used; for "Shading Operation" studies, lighting sensors and indoor temperature sensors are commonly used; for "Thermostat Adjustment", indoor temperature sensors, sound sensors, and airspeed sensors are primarily used; for "Window Operation" studies, indoor temperature sensors and window state sensors are commonly used. Apart from the aforementioned most commonly used sensors, other sensors are also summarized in the table.

Table 3
The most common computational environments for each OB model.

Domain	Programming languages					Software packages/tools					
	R	Python	C/C++	Java	VBA	IBM SPSS	Modelica/Dymola	Matlab/Simulink	Weka	LabVIEW	RapidMiner
Window Operation	X	X				X	X	X			X
Thermostat Adjustment	X		X					X			
Occupancy Estimation and Prediction	X	X		X			X	X		X	
Shading Operation	X							X			
Lighting Operation	X			X	X			X			
Appliance Use								X	X	X	X

Table 4
Sensor Choices of the Reviewed Studies.

Sensor Type	Appliance Use	Lighting Operation	Occupancy Estimation and Prediction	Shading Operation	Thermostat Adjustment	Window Operation
AC State	2	0	1	0	1	0
Air Pressure	0	0	1	0	5	2
Air Speed	0	1	2	0	10	8
Airflow Rate	0	1	3	0	5	0
Bluetooth Beacon	1	0	1	0	0	0
Camera	4	3	11	0	2	2
Chair	0	0	1	0	0	0
CO	0	0	1	0	1	1
CO2	1	0	18	0	7	16
Current/Power	10	2	5	1	5	0
Door State	1	0	4	0	0	1
GPS Location	0	0	1	0	0	0
Keyboard&Mouse	0	0	1	0	0	0
LED	0	1	1	0	0	0
Light	2	36	8	11	1	5
Light Switch	0	5	1	0	0	0
Motion (Unspecified)	2	6	4	1	1	0
PIR	3	11	19	2	1	6
RF	0	1	1	0	0	0
RH	1	1	12	1	9	18
Smart Meter	9	2	5	0	1	0
Smart Plug	4	0	0	0	0	0
Solar Irradiance	0	3	2	5	10	7
Sound (Acoustic)	0	0	6	0	0	0
Sound (Echo-Based)	1	0	2	0	0	0
Sound recording device	1	0	0	0	0	0
Telephone(State)	0	0	1	0	0	0
Temperature (Indoor)	0	3	13	6	16	33
VOC	0	0	2	0	0	1
WiFi Connection/Probe	1	3	5	0	0	0
Window State	0	2	0	0	2	32

6.3.2. Implementation location

Table 5 summarized the common locations of sensors deployed from the literature. The locations are categorized into two levels: Space Level and Building Level. Under each level of locations, detailed locations and related sensor types are provided. Based on the review work, this table provides a guideline for future researchers to refer to when deploying sensors in an experimental set up.

6.4. Integration into MPC

OB models can be used in MPC for setpoint/reference scheduling or for including measurable and predictable disturbances and for shaping the constraints. An illustrative example for the consideration of OB in MPC is presented in Fig. 16. In terms of disturbances, the OB should be considered as a cause of thermal gains from people, thermal losses from appliances or ventilation gain, and losses during window operations. This includes information about the number of occupants, time of use, and possibly information about used equipment. As measuring a direct heat dissipation is difficult, the forecasted behavior has to be used to infer the information about internal heat gains. Additionally, the OB could be considered by the constraints, such as by setting different upper and lower indoor air temperature bounds during occupancy hours. In case a specific setpoint instead of bounds is desired, the setpoint can also be explicitly defined in the cost function.

As reviewed in subsection 4.2.3, the indirect metrics evaluate the control outputs comfort and energy consumption. For the integration of OB models into MPC, we focus on the most relevant and most common use case in literature, HVAC control. The relevance results from the significant energy savings potential. Additionally, most OB models could be meaningfully coupled with HVAC control (such as Appliance Use, Lighting Operation, Occupancy Estimation and Prediction, Shading Operation, Thermostat Adjustment, Window Operation models). Most relevantly, the thermostat adjustment and attendance profiles shape the occupants' demand for thermal satisfaction by HVAC. Knowledge about absences and reduced thermal demands can significantly reduce the energy demand. The other OB models, shading and windows operation, have an impact on the thermal energy balance.

There are several requirements that need to be documented with OB

Table 5
The locations of sensors deployed from the reviewed studies.

Level of Locations	Locations	Sensor Type
Space Level	Ceiling	● CO2 Sensor
		● CO Sensor
		● Light Sensor
	Chair	● Chair Sensor
	Desk	● Keyboard & Mouse
		● Telephone (State)
	Door Frame	● Door State Sensor
	Wall	● Air Pressure Sensor
		● Air Speed Sensor
		● Bluetooth Beacon
		● GPS Location
		● LED Sensor
		● Light Switch
		● Motion (Unspecified)
		● Smart Plug
		● Sound Sensor (Acoustic)
		● Sound Sensor (Echo-Based)
		● Sound Recording Device
		● PIR Sensor
		● RF Sensor
		● RH Sensor
		● Temperature Sensor (Indoor)
		● VOC Sensor
		● WiFi Connection/Probe
	Window Frame	● Window State Sensor
Building Level	Electrical Panel	● Current/Power Sensor
		● Smart Meter
	HVAC Equipment	● AC State Sensor
		● Airflow Rate Sensor
	Main Entrance	● Camera
	Rooftop	● Solar Irradiance Sensor

model to ensure the use in advanced optimal control methods such as MPC. Firstly, the models must provide a forecast of the occupancy behavior over the length of the prediction horizon of the optimization problem, which is typically between 1 and 24 h. The quality of the presence and OB forecasts also depends on the type of measurement sensors used to gather the occupancy data [63]. Most accurate predictions are obtained from occupancy dedicated-sensor data such as PIR and cameras. However, as pointed out in Ref. [64], other sensors such as CO_2 [65] or plug power can also provide sufficiently accurate data for control-oriented occupancy models.

7. Discussion and future challenges

Based on the previous review, to define a standard guideline to document occupant behavior for building controls poses the following challenges:

7.1. Model description and formal representation

The current model description varies among different schemas and formalization methods in terms of naming schema, description structures, and presentation of OB models. In addition, neither the OB model nor the building control model lacks a standard representation for model inputs, outputs, and model description. Hence, it creates a gap between OB and building control models. This results in a customized working process for every OB-driven building control study in the literature. In addition, such a process is not consistent and creates a very different performance (e.g. energy savings) even using the same type of OB model. As a matter of fact, the various inputs for the same OB model reflect this inconsistency. Prior researchers were using different sensors and instruments to develop different mathematical models to model and simulate the same behavior over decades. There is a need to standardize the model description and representation based on one formal language. Recently, a review paper [8] on data tools for building information and performance also concludes that ontologies or schemas represent the need to be developed. An effort to extend the current Brick schema to represent the OB model is ongoing.

7.2. Model development

Currently, the model development is not fully described in the scientific articles. Information is missing on preprocessing procedures and model selection limiting the reproducibility and even replicability of the results of the studies. To overcome these challenges in model development documentation, the authors should clearly state the model purpose, and the practical and theoretical arguments supporting the choice of a given modeling technique. Also, to foster transparency and clarity, they should explicitly document the adopted cleaning and imputing procedures for the missing and corrupt data, the outliers treatment chosen, the data dimensional reduction process implemented, the data scaling method used, the techniques used for feature creation, the approach adopted for partitioning the original dataset for the definition of the test and training subsets, and the anonymization techniques used, if any. Furthermore, an important challenge is to develop newer multi-domain models, which can consider the multi-exposure of occupants to indoor environmental conditions and a multitude of controlling opportunities for a better and tailored adjustment of the building devices.

7.3. Model evaluation

The standardized and comprehensive model evaluation is crucial for OB model deployment in real-world scenarios of building control products. Therefore, a standardized evaluation schema is proposed for each distinct form of OB. With the ambition to provide a comprehensive evaluation, this work proposes joint use of three sets of evaluation metrics, namely absolute metrics, domain metrics, and indirect metrics. The absolute metrics were derived based on the vast literature evidence on the evaluation of general data-driven methods and their existing applications for OB modeling. Furthermore, we argue that the OB model evaluation has to include specific domain metrics. These are based on domain expertise in OB modeling and are supported by the existing literature evidence.

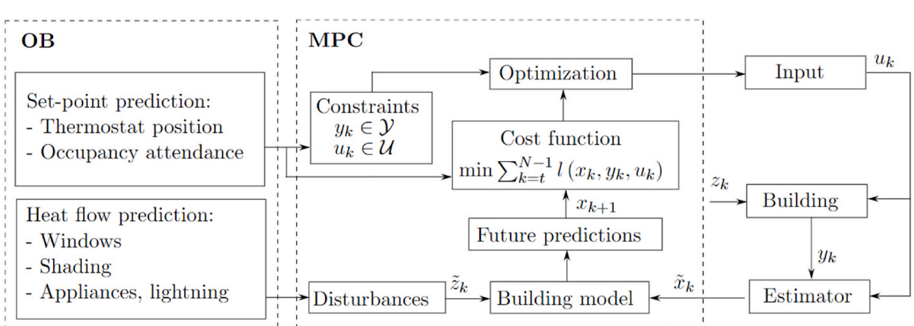

Fig. 16. An overview of the MPC structure with the proposed inclusions of the OB.

As human behavior is highly diverse and sensitive to unpredictable events, sensitivity analysis could be used as a supportive tool to assess the uncertainties related to the used OB model. Sensitivity analysis (SA) is a statistical technique that assesses the effects that changes in input or design variables have on the model output variables [16,66]. There are two approaches to SA that could be applied when a model can be also used for building control purposes. The simpler is the local sensitivity analysis, where the impact of an input variable's variation on a model response is estimated while keeping the values of the other input factors constant. The global sensitivity analysis, on the opposite, tests simultaneously all the input variables and enables assessing the impact of both individual input variables and their interactions on the model output. However, the existing applications of the sensitivity analysis for OB in building control are sparse, further adoption would provide useful insights on testing control algorithm robustness against noise or uncertainty in input variables and parameters.

7.4. Benefits of the inclusion of OB models in building control

Finally, the indirect metrics quantify how well the model contributes to fulfilling the higher goal of building control, such as maintaining comfort or optimizing energy consumption. Up to date, the impact of the OB model on the end-system has been limitedly explored. Namely, most of the existing OB models were not tested in field studies, and therefore, the relationship between whole system performance and the OB model is rarely analyzed. In order to come one step closer to filling this gap, this work proposed a set of indirect metrics for evaluating the impact of OB models for inclusion in HVAC control. The future challenge includes assessing the suitability of proposed metrics in field studies. Furthermore, the indirect metrics for alternative systems, such as shadings should be explored.

7.5. Model implementation

The documented model implementation should include the information about the used computational environment in which the OB model is tested, the experimental setting, and the recommendations for the intended application in the building control. Here, a particular challenge is that the buildings are commonly a one-time product. As there is limited literature evidence on documenting the model implementation, future research should focus on how to standardize information related to implementation in different buildings or HVAC systems.

Furthermore, the future model documentation should include the estimated OB model inference time. As highlighted in Ref. [67], state-of-the-art OB models are too computationally expensive to be included in real-time control applications, such as MPC. In order to obtain stable and reliable results, the computation of the next control signal should be indeed completed before the start of the actual period of observation. Based on the previous literature review, the time related to a single forward pass of the proposed data-driven models, i.e. inference time, is however rarely documented.

7.6. Model integration into advanced building control

One of the challenges of leveraging occupancy estimation and prediction models is their integration into advanced building control algorithms, e.g. MPC. These controllers are typically based on HVAC and building envelope thermal dynamics models and consider future system dynamics and future control inputs or constraints. Special care needs to be given to properly couple the occupancy estimation and prediction models into these dynamic equations. The OB models can serve as additional control variables (setpoint, constraint, or disturbances) for the building dynamics equations. These dynamic equations are usually represented by a set of first-order differential equations. As a common practice in the control engineering field, these equations can be reformulated into a state space model and into discrete time [68,69].

8. Conclusion

In this paper, we evaluated current documentation of OB models for advanced building controls from four different perspectives: model description and formal representation, model development and evaluation, inclusion of OB model into building controls, and modeling implementation. During the literature review, we found that the building control and OB modeling were mostly researched distinctly. Most of the OB models were developed as stand-alone models. In that context, there is only a spoonful of publications that proposed a formal integration of the OB models into building control (e.g. Refs. [70,71]). Based on a comprehensive review and analysis on current documentation of OB model for advanced building controls, it can be concluded that: 1) There is no standard representation of various OB models; 2) no unified guidelines of OB model development; 3) a standardized evaluation schema is proposed for each distinct form of OB models; 4) a set of indirect metrics for evaluating the impact of OB models for the inclusion in HVAC control is defined; and 5) a systematic documentation of indented model implementation is proposed; and 6) OB models can be integrated into MPC for HVAC as predicted setpoints, constraints, or disturbances.

Given the current review and discussions, this paper also provides following future research opportunities: a) A formal representation of OB models based on the same schema and semantics. While there is an on-going effort in the Brick schema [25], such presentation can be further enriched with more common data sets [72]; b) open sourcing a library with OB model documentation that follows this guideline, c) deployment of existing OB models in building control studies.

Limitation of the study: 1) In this paper, the "occupant" is referred to as office workers in general. The review does not cover other occupant types such as elderly, who has different interactions in response to thermal stimuli; 2) The review study found very limited or no papers considering how to integrate sensor drifting into controls. Although it is an important issue for the control implementation, the paper focuses mainly on documenting occupant behavior. Future studies could further explore this topic; and 3) The guideline paper does not cover occupant behavior of personalized cooling and heating systems. This could be included in the future studies.

CRediT authorship contribution statement

Bing Dong: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Romana Markovic:** Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Salvatore Carlucci:** Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Conceptualization. **Yapan Liu:** Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing, Conceptualization. **Andreas Wagner:** Writing – original draft, Resources, Methodology, Investigation, Conceptualization. **Antonio Liguori:** Conceptualization, Investigation, Methodology, Writing – original draft. **Christoph van Treeck:** Writing – original draft, Resources, Investigation, Conceptualization. **Dmitry Oleynikov:** Conceptualization, Investigation, Methodology, Writing – original draft. **Elie Azar:** Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Gianmarco Fajilla:** Conceptualization, Investigation, Methodology, Writing – original draft. **Ján Drgoňa:** Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Joyce Kim:** Conceptualization, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Marika Vellei:** Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Marilena De Simone:** Visualization, Writing – original draft, Writing – review & editing, Conceptualization, Investigation, Methodology. **Masood Shamsaiee:** Writing – original draft, Methodology, Investigation, Conceptualization. **Mateus Bavaresco:** Conceptualization, Investigation, Methodology, Writing – original draft. **Matteo Favero:** Writing – original draft, Methodology, Investigation, Conceptualization. **Mikkel Kjaergaard:** Conceptualization, Investigation, Methodology, Writing – original draft. **Mohamed Osman:** Writing – original draft, Methodology, Investigation, Conceptualization. **Moritz Frahm:** Conceptualization, Investigation, Methodology, Visualization, Writing – original draft. **Sanam Dabirian:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Da Yan:** Conceptualization, Investigation, Methodology, Writing – original draft. **Xuyuan Kang:** Writing – original draft, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Acknowledgment

Bing Dong and Yapan Liu would like to thank the support from the U. S. National Science Foundation (Award No. 1949372). Romana Markovic and Andreas Wagner gratefully acknowledge the financial support by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), promotional reference 03EN1002A. Dmitry Oleynikov gratefully acknowledges the financial support by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), promotional reference 03EN1002B. Moritz Frahm gratefully acknowledges the financial support from the German Federal Ministry for Economic Affairs and Climate Action (BMWK), promotional reference 03ETW016A. Gianmarco Fajilla would like to thank the Calabria Region Government for his Ph.D. scholarship (POR Calabria FSE/FESR 2014- 2020, Identification Project Number H21G18000170006). Antonio Liguori and Christoph van Treeck were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – TR 892/4-1. This paper benefited greatly from discussions with members of IEA EBC Annex 79.

Appendix

Appendix I. Summary of the indirect metrics for each OB modeling objective

Model	Comfort-related metrics	Energy-related metrics
Occupancy Estimation and Prediction	Thermal comfort and indoor air quality: hours of setpoints not met	energy consumption/saving, start/stop time, in most cases related to HVAC
Thermostat Adjustment	Thermal comfort: indoor temperature, PMV, PPD	energy consumption/saving, monetary savings, related to HVAC, duration of unnecessary heating [h], peak load change (energy shifting for DR), energy use during peak, setpoint reduction, HVAC coefficient of performance
Window Operation	–	energy consumption/saving, related to HVAC
Shading Operation	Thermal comfort: air temperature, PPD, overheated hours; Visual comfort: illuminance	energy consumption/saving, related to HVAC and lighting, optimal dimming
Lighting Operation	Visual comfort: illuminance, false off frequency, discomfort probability	energy consumption/saving, optimal dimming (% of lighting power used needed on daylight availability), peak power, illuminance reduction in unoccupied workstations
Appliance Use	–	energy consumption/saving, related to HVAC and the appliances

References

[1] W. O'Brien, et al., Introducing IEA EBC annex 79: key challenges and opportunities in the field of occupant-centric building design and operation, Build. Environ. 178 (Jul. 2020), 106738, https://doi.org/10.1016/j.buildenv.2020.106738.

[2] M. Kong, B. Dong, R. Zhang, Z. O'Neill, HVAC energy savings, thermal comfort and air quality for occupant-centric control through a side-by-side experimental study, Appl. Energy 306 (Jan. 2022), 117987, https://doi.org/10.1016/j.apenergy.2021.117987.

[3] Z. Pang, Y. Chen, J. Zhang, Z. O'Neill, H. Cheng, B. Dong, Nationwide HVAC energy-saving potential quantification for office buildings with occupant-centric controls in various climates, Appl. Energy 279 (Dec. 2020), 115727, https://doi.org/10.1016/j.apenergy.2020.115727.

[4] D. Li, C.C. Menassa, V.R. Kamat, Personalized human comfort in indoor building environments under diverse conditioning modes, Build. Environ. 126 (Dec. 2017) 304–317, https://doi.org/10.1016/j.buildenv.2017.10.004.

[5] J. Xie, H. Li, C. Li, J. Zhang, M. Luo, Review on occupant-centric thermal comfort sensing, predicting, and controlling, Energy Build. 226 (Nov. 2020), 110392, https://doi.org/10.1016/j.enbuild.2020.110392.

[6] J.Y. Park, et al., A critical review of field implementations of occupant-centric building controls, Build. Environ. 165 (Nov. 2019), 106351, https://doi.org/10.1016/j.buildenv.2019.106351.

[7] B. Dong, D. Yan, Z. Li, Y. Jin, X. Feng, H. Fontenot, Modeling occupancy and behavior for better building design and operation—a critical review, Build. Simulat. 11 (5) (Oct. 2018) 899–921, https://doi.org/10.1007/s12273-018-0452-x.

[8] N. Luo, M. Pritoni, T. Hong, An overview of data tools for representing and managing building information and performance data, Renew. Sustain. Energy Rev. 147 (Sep. 2021), 111224, https://doi.org/10.1016/j.rser.2021.111224.

[9] S. Carlucci, et al., Modeling Occupant Behavior in Buildings, 174, Building and Environment, May 2020, 106768, https://doi.org/10.1016/j.buildenv.2020.106768.

[10] A. Wagner, W. O'Brien, B. Dong (Eds.), Exploring Occupant Behavior in Buildings, Springer International Publishing, Cham, 2018, https://doi.org/10.1007/978-3-319-61464-9.

[11] Y. Cardinale, Occupant Activity Detection in Smart Buildings: A Review," Social Science Research Network, Jun. 2020, https://doi.org/10.2139/ssrn.3671533. Rochester, NY, SSRN Scholarly Paper ID 3671533.

[12] S. Norouziasl, A. Jafari, Y. Zhu, Modeling and simulation of energy-related human-building interaction: a systematic review, J. Build. Eng. 44 (Dec. 2021), 102928, https://doi.org/10.1016/j.jobe.2021.102928.

[13] S. D'Oca, T. Hong, J. Langevin, The human dimensions of energy use in buildings: a review, Renew. Sustain. Energy Rev. 81 (Jan. 2018) 731–742, https://doi.org/10.1016/j.rser.2017.08.019.

[14] M. Han, J. Zhao, X. Zhang, J. Shen, Y. Li, The reinforcement learning method for occupant behavior in building control: a review, Energy. Built. Env. 2 (2) (Apr. 2021) 137–148, https://doi.org/10.1016/j.enbenv.2020.08.005.

[15] E.T. Maddalena, Y. Lian, C.N. Jones, Data-driven methods for building control — a review and promising future directions, Control Eng. Pract. 95 (Feb. 2020), 104211, https://doi.org/10.1016/j.conengprac.2019.104211.

[16] A. Mirakhorli, B. Dong, Occupancy behavior based model predictive control for building indoor climate—a critical review, Energy Build. 129 (Oct. 2016) 499–513, https://doi.org/10.1016/j.enbuild.2016.07.036.

[17] S. Naylor, M. Gillott, T. Lau, A review of occupant-centric building control strategies to reduce building energy use, Renew. Sustain. Energy Rev. 96 (Nov. 2018) 1–10, https://doi.org/10.1016/j.rser.2018.07.019.

[18] D. Yan, et al., Occupant behavior modeling for building performance simulation: current state and future challenges, Energy Build. 107 (Nov. 2015) 264–278, https://doi.org/10.1016/j.enbuild.2015.08.032.

[19] G. Happle, J.A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy Build. 174 (Sep. 2018) 276–292, https://doi.org/10.1016/j.enbuild.2018.06.030.

[20] T. Hong, J. Langevin, K. Sun, Building simulation: ten challenges, Build. Simulat. 11 (5) (Oct. 2018) 871–898, https://doi.org/10.1007/s12273-018-0444-x.

[21] E. Azar, et al., Simulation-aided occupant-centric building design: a critical review of tools, methods, and applications, Energy Build. 224 (Oct. 2020), 110292, https://doi.org/10.1016/j.enbuild.2020.110292.
[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. 172 (Aug. 2018) 279–294, https://doi.org/10.1016/j.enbuild.2018.05.017.
[23] M.B. Kjærgaard, et al., Current practices and infrastructure for open data based research on occupant-centric design and operation of buildings, Build. Environ. 177 (Jun. 2020), 106848, https://doi.org/10.1016/j.buildenv.2020.106848.
[24] ARM research facility." https://www.arm.gov/(accessed Mar. 18, 2022).
[25] B. Balaji, et al., Brick: towards a unified metadata schema for buildings, in: Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, Nov. 2016, pp. 41–50, https://doi.org/10.1145/2993422.2993577. Palo Alto CA USA.
[26] Home – Project Haystack." https://project-haystack.org/(accessed Mar. 18, 2022).
[27] J. Kim, S. Schiavon, G. Brager, Personal comfort models – a new paradigm in thermal comfort for occupant-centric environmental control, Build. Environ. 132 (Mar. 2018) 114–124, https://doi.org/10.1016/j.buildenv.2018.01.023.
[28] M. Ashouri, B.C.M. Fung, F. Haghighat, H. Yoshino, Systematic approach to provide building occupants with feedback to reduce energy consumption, Energy 194 (Mar. 2020), 116813, https://doi.org/10.1016/j.energy.2019.116813.
[29] X. Yu, S. Ergan, G. Dedemen, A data-driven approach to extract operational signatures of HVAC systems and analyze impact on electricity consumption, Appl. Energy 253 (Nov. 2019), 113497, https://doi.org/10.1016/j.apenergy.2019.113497.
[30] S. D'Oca, T. Hong, A data-mining approach to discover patterns of window opening and closing behavior in offices, Build. Environ. 82 (Dec. 2014) 726–739, https://doi.org/10.1016/j.buildenv.2014.10.021.
[31] S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (Feb. 2015) 395–408, https://doi.org/10.1016/j.enbuild.2014.11.065.
[32] O. Van Cutsem, G. Lilis, M. Kayal, Automatic multi-state load profile identification with application to energy disaggregation, in: 2017 *22nd* IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Sep. 2017, pp. 1–8, https://doi.org/10.1109/ETFA.2017.8247684.
[33] B. Dong, K.P. Lam, A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting, Build. Simulat. 7 (1) (Feb. 2014) 89–106, https://doi.org/10.1007/s12273-013-0142-7.
[34] Y. Jin, D. Yan, X. Zhang, J. An, M. Han, A data-driven model predictive control for lighting system based on historical occupancy in an office building: Methodology development, Build. Simulat. 14 (1) (Feb. 2021) 219–235, https://doi.org/10.1007/s12273-020-0638-x.
[35] D. Mora, G. Fajilla, M.C. Austin, M. De Simone, Occupancy patterns obtained by heuristic approaches: cluster analysis and logical flowcharts. A case study in a university office, Energy Build. 186 (Mar. 2019) 147–168, https://doi.org/10.1016/j.enbuild.2019.01.023.
[36] S. Dabirian, K. Panchabikesan, U. Eicker, Occupant-centric urban building energy modeling: approaches, inputs, and data sources - a review, Energy Build. 257 (Feb. 2022), 111809, https://doi.org/10.1016/j.enbuild.2021.111809.
[37] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, Build. Simulat. 4 (2) (Jun. 2011) 149–167, https://doi.org/10.1007/s12273-011-0044-5.
[38] M.B. Kjærgaard, M. Werner, F.C. Sangogboye, K. Arendt, DCount - a probabilistic algorithm for accurately disaggregating building occupant counts into room counts, in: 2018 *19th* IEEE International Conference on Mobile Data Management (MDM),, Jun. 2018, pp. 46–55, https://doi.org/10.1109/MDM.2018.00021.
[39] R. Markovic, E. Azar, M.K. Annaqeeb, J. Frisch, C. van Treeck, Day-ahead prediction of plug-in loads using a long short-term memory neural network, Energy Build. 234 (Mar. 2021), 110667, https://doi.org/10.1016/j.enbuild.2020.110667.
[40] M. Shcherbakov, A. Brebels, N.L. Shcherbakova, A. Tyukov, T.A. Janovsky, V. A. Kamaev, A survey of forecast error measures, World Appl. Sci. J. 24 (Jan. 2013) 171–176, https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032.
[41] M.Z. Naser, A.H. Alavi, Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences, Archit. Struct. Constr., Nov. (2021), https://doi.org/10.1007/s44150-021-00015-8.
[42] Z. Li, H. Zhu, B. Dong, X. Xu, Development of a systematic procedure to establish customized shading behavior identification model, Energy Build. 239 (May 2021), 110793, https://doi.org/10.1016/j.enbuild.2021.110793.
[43] W. Kampel, S. Carlucci, B. Aas, A. Bruland, A proposal of energy performance indicators for a reliable benchmark of swimming facilities, Energy Build. 129 (Oct. 2016) 186–198, https://doi.org/10.1016/j.enbuild.2016.07.033.
[44] W. O'Brien, I. Gaetani, S. Carlucci, P.-J. Hoes, J.L.M. Hensen, On occupant-centric building performance metrics, Build. Environ. 122 (Sep. 2017) 373–385, https://doi.org/10.1016/j.buildenv.2017.06.028.
[45] F. Tahmasebi, A. Mahdavi, The sensitivity of building performance simulation results to the choice of occupants' presence models: a case study, J. Build. Perform. Simul. 10 (5–6) (Nov. 2017) 625–635, https://doi.org/10.1080/19401493.2015.1117528.
[46] A. Chong, G. Augenbroe, D. Yan, Occupancy data at different spatial resolutions: building energy performance and model calibration, Appl. Energy 286 (Mar. 2021), 116492, https://doi.org/10.1016/j.apenergy.2021.116492.
[47] A. Mahdavi, F. Tahmasebi, On the quality evaluation of behavioural models for building performance applications, J. Build. Perform. Simul. 10 (5–6) (Nov. 2017) 554–564, https://doi.org/10.1080/19401493.2016.1230148.
[48] D. Yan, X. Feng, Y. Jin, C. Wang, The evaluation of stochastic occupant behavior models from an application-oriented perspective: using the lighting behavior model as a case study, Energy Build. 176 (Oct. 2018) 151–162, https://doi.org/10.1016/j.enbuild.2018.07.037.
[49] A. Mahdavi, F. Tahmasebi, Predicting people's presence in buildings: an empirically based model performance analysis, Energy Build. 86 (Jan. 2015) 349–355, https://doi.org/10.1016/j.enbuild.2014.10.027.
[50] Y. Jin, et al., Appliance use behavior modelling and evaluation in residential buildings: a case study of television energy use, Build. Simulat. 13 (4) (Aug. 2020) 787–801, https://doi.org/10.1007/s12273-020-0648-8.
[51] C. Peng, D. Yan, R. Wu, C. Wang, X. Zhou, Y. Jiang, Quantitative description and simulation of human behavior in residential buildings, Build. Simulat. 5 (2) (Jun. 2012) 85–94, https://doi.org/10.1007/s12273-011-0049-0.
[52] I.T. Michailidis, et al., Energy-efficient HVAC management using cooperative, self-trained, control agents: a real-life German building case study, Appl. Energy 211 (Feb. 2018) 113–125, https://doi.org/10.1016/j.apenergy.2017.11.046.
[53] K.X. Perez, M. Baldea, T.F. Edgar, Integrated smart appliance scheduling and HVAC control for peak residential load management, in: 2016 American Control Conference (ACC), Jul. 2016, pp. 1458–1463, https://doi.org/10.1109/ACC.2016.7525122.
[54] T. Sanjeev Kumar, C. Kurian, S. Shetty, A data-driven approach for the control of a daylight–artificial light integrated scheme, Light. Res. Technol. 52 (2) (Apr. 2020) 292–313, https://doi.org/10.1177/1477153519841104.
[55] Z. Wang, Y.K. Tan, Illumination control of LED systems based on neural network model and energy optimization algorithm, Energy Build. 62 (Jul. 2013) 514–521, https://doi.org/10.1016/j.enbuild.2013.03.029.
[56] ANSI/ASHRAE Standard 135-2016," p. 48.
[57] M. Aleksy, P. Bauer, Ensuring the practical applicability of algorithms for user behavior modeling through integration into building automation systems, in: Advanced Information Networking and Applications, 2020, pp. 715–721, https://doi.org/10.1007/978-3-030-44041-1_63. Cham.
[58] R. Markovic, Generic Occupant Behavior Modeling for Commercial Buildings, 2020.
[59] X. Zhang, et al., IEHouse: a non-intrusive household appliance state recognition system, in: 2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/*UIC/ATC/CBDCom/IOP*/SCI), Aug. 2017, pp. 1–8, https://doi.org/10.1109/UIC-ATC.2017.8397510.
[60] D. Egarter, V.P. Bhuvana, W. Elmenreich, PALDi: online load disaggregation via particle filtering, IEEE Trans. Instrum. Meas. 64 (2) (Feb. 2015) 467–477, https://doi.org/10.1109/TIM.2014.2344373.
[61] W. Kuo, M.J. Zuo, Optimal Reliability Modeling: Principles and Applications, John Wiley & Sons, 2003.
[62] P.E. Black, 'big-O notation', in dictionary of algorithms and data structures. https://xlinux.nist.gov/dads/HTML/bigOnotation.html, Sep. 2019 accessed Mar. 24, 2022.
[63] A. Capozzoli, M.S. Piscitelli, A. Gorrino, I. Ballarini, V. Corrado, Data analytics for occupancy pattern learning to reduce the energy consumption of HVAC systems in office buildings, Sustain. Cities Soc. 35 (Nov. 2017) 191–208, https://doi.org/10.1016/j.scs.2017.07.016.
[64] F.C. Sangogboye, K. Arendt, A. Singh, C.T. Veje, M.B. Kjærgaard, B.N. Jørgensen, Performance comparison of occupancy count estimation and prediction with common versus dedicated sensors for building model predictive control, Build. Simulat. 10 (6) (Dec. 2017) 829–843, https://doi.org/10.1007/s12273-017-0397-5.
[65] F. Jorissen, W. Boydens, L. Helsen, Simulation-based occupancy estimation in office buildings using CO2 sensors, in: Proceedings of the 15th IBPSA Conference, 2017, pp. 1073–1082. Accessed: Mar. 24, 2022. [Online]. Available: http://hdl.handle.net/1854/LU-8580815.
[66] S. Goyal, H.A. Ingley, P. Barooah, Effect of various uncertainties on the performance of occupancy-based optimal control of HVAC zones, in: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), Dec. 2012, pp. 7565–7570, https://doi.org/10.1109/CDC.2012.6426111.
[67] J. Drgoňa, et al., All you need to know about model predictive control for buildings, Annu. Rev. Control 50 (Jan. 2020) 190–232, https://doi.org/10.1016/j.arcontrol.2020.09.001.
[68] I. Hazyuk, C. Ghiaus, D. Penhouet, Optimal temperature control of intermittently heated buildings using Model Predictive Control: Part I – building modeling, Build. Environ. 51 (May 2012) 379–387, https://doi.org/10.1016/j.buildenv.2011.11.009.
[69] I. Hazyuk, C. Ghiaus, D. Penhouet, Optimal temperature control of intermittently heated buildings using Model Predictive Control: Part II – control algorithm, Build. Environ. 51 (May 2012) 388–394, https://doi.org/10.1016/j.buildenv.2011.11.008.
[70] W. Turner, T. Hong, A Technical Framework to Describe Occupant Behavior for Building Energy Simulations, LBNL-6671E, Lawrence Berkeley National Lab.

(LBNL), Berkeley, CA (United States), Dec. 2013. Accessed: Mar. 24, 2022. [Online]. Available: https://www.osti.gov/biblio/1164208.

[71] B. Dong, K. Lam, C. Neuman, Integrated building control based on occupant behavior pattern detection and local weather forecasting, in: In Twelfth International IBPSA Conference, Nov. 2011, pp. 14–17. Accessed: Mar. 24, 2022. [Online]. Available: https://www.semanticscholar.org/paper/Integrated-building-control-based-on-occupant-and-Dong-Lam/d9c366355f9fef0678e47f448fc50014ad77814b.

[72] G. Fierro, et al., Formalizing tag-based metadata with the Brick ontology, Frontiers in Built Environment 6 (2020). Accessed: Mar. 28, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fbuil.2020.558034.