Energy & Buildings 251 (2021) 111345

Contents lists available at ScienceDirect

Energy & Buildings

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

Building occupancy forecasting: A systematical and critical review

Yuan Jin [a], Da Yan [a,*], Adrian Chong [b], Bing Dong [c], Jingjing An [d]

[a] Building Energy Research Center, School of Architecture, Tsinghua University, Beijing 100084, China
[b] Department of Building, School of Design and Environment, National University of Singapore, Singapore 117566, Singapore
[c] Department of Mechanical and Aerospace Engineering, Syracuse University, Syracuse, NY 13244, USA
[d] School of Environment and Energy Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

ARTICLE INFO

Article history:
Received 22 December 2020
Revised 23 July 2021
Accepted 10 August 2021
Available online 14 August 2021

Keywords:
Occupancy prediction
Forecast
Occupant behavior
Building
Operation
Energy conservation

ABSTRACT

Indoor environment construction for occupants has high energy consumption; as such, occupancy plays a noteworthy role in the complete life cycle phase of buildings, including design, operation, and retrofitting. In the past few years, building occupancy, which is considered the basis of occupant behavior, has attracted increasing attention from researchers. There are increasing requirements for buildings to be both comfortable and energy efficient; with the development of detection methods and analyzing algorithms, occupancy prediction has become a topic of interest for building automation and energy conservation. Therefore, this article reviews the literature regarding future building occupancy predictions (forecasting). This review is distinguished from occupancy simulation and detection research and focuses on the research purpose, physical routine, and complete methodology of occupancy forecasting. First, the research purposes, including the application field and detailed requirements for occupancy forecasting, are summarized and analyzed. Next, an overall methodology of occupancy forecasting, including data acquisition, modeling techniques, and evaluation, is discussed in terms of issues affecting prediction performance. Finally, the current challenges and perspectives of occupancy forecasting are highlighted, considering the insights of natural characteristics, on-site implementation, valid dataset sharing, and research techniques. Overall, accurate and robust future occupancy predictions will help to improve building system operations and energy conservation.

Contents

* Corresponding author at: Building Energy Research Center, Tsinghua University, Beijing 100084, China.
E-mail address: yanda@tsinghua.edu.cn (D. Yan).

https://doi.org/10.1016/j.enbuild.2021.111345
0378-7788/

1. Introduction

1.1. Background

Buildings intend to provide occupants with a comfortable environment of relatively steady temperature, sufficient illuminance, quietness, and fresh air by installing heating, ventilation, air conditioning (HVAC), lighting, and other systems and equipment. These systems cause buildings to constitute a large proportion of global energy consumption. In 2017, the total consumption of residential, commercial, and public services was 2.85 billion toe, accounting for 29.3% of worldwide energy consumption [1]. Occupants are one of the main service targets of buildings, meaning that the design, construction, and operation of buildings should carefully consider occupants [2]. Energy consumption is affected by occupant behavior under occupant-related building designs [3] and occupant control actions or occupant-centric operations [4]. For this reason, research and policy have been increasingly focused on occupant behavior in buildings over the past few years [5,6]. To date, the influential cooperating research projects on occupant behavior are the International Energy Agency Energy in Buildings and Communities programme Annex 66 [7] and Annex 79 [8], which have achieved significant research outcomes.

Within the scope of occupant behavior research, occupancy is one of the most important factors during different phases, including building design, operation, and renovation. Therefore, this section summarizes the definition, physical properties, and scope of related research, which require a review of occupancy forecasting research.

(1) Definition and importance of building occupancy
Among the factors related to occupants, building occupancy is the most crucial and is the basis of occupant behavior research. Fig. 1 illustrates the definition and impact of occupancy in buildings. Building occupancy includes the presence state, occupant number, and trajectory, which indicate whether the occupant is in a certain space in the building, the number of occupants, and how the occupants are moving. Irregular and partial occupancy affects the indoor environment and energy consumption [9]. Therefore, occupancy is carefully considered during the life cycle of building research and practice, including design, operation,

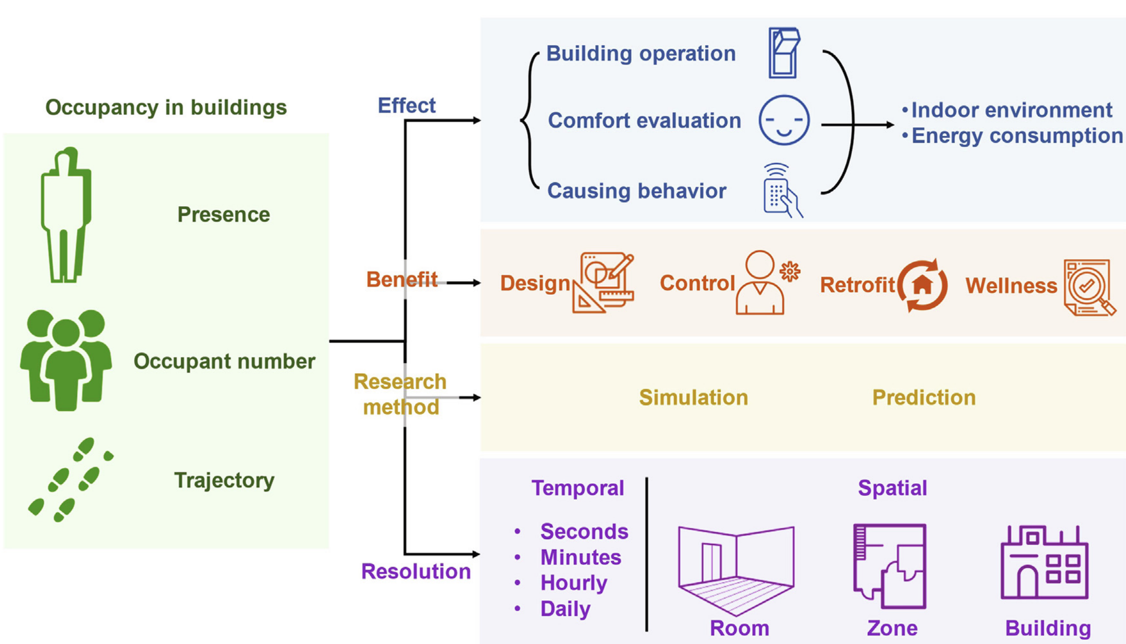


Fig. 1. Definition and concept of building occupancy.

retrofitting, and evaluation. The resolutions of occupancy studies vary for different issues. Furthermore, temporal resolution varies from seconds to days, while spatial resolution varies from rooms to the entire building. The difference between this review article and previously published reviews is that this article focuses on occupancy forecasting, a branch of occupancy prediction research. The research method of simulation and prediction is described in Section (3). However, for both research targets, the entire methodology loop includes data measurement, modeling, and evaluation.

(2) Physical properties of building occupancy
Research on building occupancy is difficult because of its physical nature, which is the basis of this article. In recent years, the notion that building occupancy possesses the properties of stochasticity [10–12] and variety has been widely analyzed and recognized [13–15]. The stochastic nature of occupancy leads to different situations compared with the static settings included in the standards according to an international review of occupant-related aspects of building energy codes [16]. From one on-site measurement of eight homes, only half of the rooms were used for up to 60% of the time when the home was occupied [17]. Meanwhile, occupancy comprises various occupancy and vacancy time distributions [18] and bimodal distributions [19]. Various studies have considered profile extraction and applications in building performance simulations [20,21]. Furthermore, occupancy behavior was examined at both the spatial [22] and temporal scales [23]. Spatial scale refers to the transfer relationship from one space to another; temporal scale refers to the autocorrelation among occupancy data with different time steps [24] or the correlation between the occupancy and ambient parameters at consecutive time steps [25]. The location or activity of one occupant also depends on the duration the occupant already spends in the current location [26,27]. There are also other types of contextual temporal information that are frequently discussed, including the time of day, weekdays or weekends, and season [28–30]. Hou et al. [31] investigated the spatial and temporal dependencies between building occupancy and the conditions of urban surroundings. Several studies have concentrated on the predictability of occupancy routines. Song et al. [32] used the concept of entropy, which is likely the most fundamental quantity for determining the degree of predictability and characterizing a time series, to depict the predictability of the occupancy state at a district scale. Ahn et al. [33] found that according to different building types and features of zones or areas in buildings, the characteristics of the presence of occupants may differ. The normalized cumulative periodogram method is used to test the random walking pattern of the presence of occupants in rooms, levels, or an entire building; furthermore, this method precisely demonstrates that the presence pattern is unpredictable. This indicates that more work on occupancy models should be performed based on different types of buildings. Gunay et al. [34] discussed the predictability of recurring occupancy patterns in offices. They found that occupancy is a nonstationary process that changes with time and space. Based on an autocorrelation analysis, the occupancy in their case study had a weak correlation with the arrival and departure times. It was also asserted that the positions of individuals are almost predictable in office buildings or institutions with a fixed timetable [35].

(3) Research on building occupancy: occupancy simulations and occupancy prediction
For the past few years, several studies have been conducted within the area of occupancy in an attempt to demonstrate physical routine, mathematical depiction, and research applications. The research methods employed are commonly denoted as “occupancy simulation” and “occupancy prediction” (comparison of model development is shown in Fig. 2). A simulation denotes the methods that represent the occupancy profiles of a building for a certain period. Modeling methods of occupancy simulation research include refined and attentive analyses, a typical embodiment as a Markov chain model, and a logistic regression model [36]. Occupancy simulation methods have been effectively integrated into energy consumption simulations and building design and retrofit evaluations [37]. Meanwhile, a prediction denotes the methods used for estimating or forecasting the occupancy state of a certain space scale in a building based on historical occupancy data or ambient parameter data from various sensors. Research on occupancy simulations and predictions has certain common features and distinctions. A physical routine and pattern are crucial common features, and a basic mathematical model can be proposed to achieve both objectives. As for the distinctions, research on occupancy simulation usually trains the model using historical data inputs, and there is no need to update the model during simulation implementation. Instead, the results are evaluated according to the distribution of certain metrics during a period, such as the accumulated occupancy

	Occupancy simulation	Occupancy prediction
Commons of model establishment	• Based on physical routine analysis • Mathematical model	
Model update	NO model update with new data during simulation	Model update with new data during prediction
Training & testing split	NO need for training & testing set split during evaluation	Split training & testing sets during evaluation
Data evaluation	Evaluate data in a certain PERIOD	Evaluate data for each TIME STEP

Fig. 2. Comparison of occupancy simulation and prediction research regarding model development.

duration, arrival time, and departure time. Additionally, there is no need to split the training and testing datasets for model verification. Conversely, research on occupancy prediction normally contributes to on-site system operation, requiring precise occupancy prediction for each time step and the estimation of real-time errors, such as root-mean-square error (RMSE), coefficient of variation of RMSE (CVRMSE), and accuracy. To overcome the variety and fluctuation of occupancy patterns, the model must be updated regularly with new historical data. Because the evaluation should be conducted at each time step, a split between the training and testing sets is also recommended. To demonstrate the limitation of occupancy simulations using the Markov chain model, Mahdavi et al. [24] proposed an MT model (simple nonprobablistic model). The model generates daily binary occupancy profiles based on aggregated past–presence data with concise threshold rules in order to show the distinction between simulation and prediction research.

Research on occupancy prediction has the potential to enhance building operation performance via occupancy-based control. According to review articles of occupancy prediction research, there are two categories of occupancy prediction: "occupancy detection/estimation" and "occupancy forecast." The prediction window should be distinguished well, as shown in Fig. 3. "Occupancy detection" refers to predicting the occupancy for the current time step, whereas "occupancy forecast" refers to predicting the occupancy for a future time step. The two research avenues differ in terms of their purpose, data acquisition, and input data. To optimize building operation strategies, such as model predictive control (MPC), it is crucial to forecast occupancy [38]. Therefore, this review mainly focuses on reviewing occupancy prediction, also named occupancy forecasting, to analyze the detailed modeling techniques and potential challenges.

Despite several efforts that have been made towards occupancy forecasting, the performance and on-site application conditions are less satisfactory than those of occupancy simulations. Thus, improving occupancy detection devices, data mining algorithms, and building automation technology will provide occupancy forecasting research with immense potential.

1.2. Objectives

To date, several articles have reviewed occupancy simulations and detection research [39–42]. However, there are no specific review articles that focus on a systematic summary and detailed technical analysis of occupancy forecasting. For instance, Dai et al. [43] reviewed machine learning models for occupancy and window-opening behavior. However, they mainly focused on occupancy detection and estimation, and their review of occupancy prediction in future time steps could be improved. Moreover, occupancy-related research, including simulations, detections or estimations, and forecasting, has not been well distinguished in many research articles. Therefore, in this review, we focus on future occupancy forecasting for buildings by incorporating the knowledge of occupancy essence and the potential of occupancy forecasting. The objective of this article was to review the literature on occupancy forecasting, including previous research purposes and overall research methodologies. Based on this comprehensive review, interpretation, and critical evaluation, we ultimately identify and narrate the challenges and perspectives of occupancy forecasting research.

In Section 2, the methodology and framework for this review are proposed, and a summary of the basic information is provided. In Section 3, we present the review results of the research purpose for occupancy forecasting and the detailed requirements for realizing this purpose. In Sections 4-6, the overall methodology of occupancy forecasting, including data acquisition, modeling, and evaluation, is reviewed and critically analyzed. In Section 7, the challenges and perspectives of the study are discussed. Finally, in Section 8, the conclusions of this study are summarized.

2. Methodology

2.1. Methodology

This article conducted a review using the three steps shown in Fig. 4.

Step 1: Literature search

The literature was searched using two databases: Web of Science (Core Collection) and Scopus. After a preliminary search, the articles were refined based on their abstracts. The literature was also supplemented by two means: related articles and one more search before article submission.

Step 2: Systematic and critical evaluation

According to the research methodology, the systematic and critical evaluation of occupancy forecast-related articles can be divided into four parts: research purposes, data acquisition, modeling techniques, and evaluation methods (Section 2.3).

Step 3: Challenges and perspective

	Occupancy detection	Occupancy forecast
Prediction window	• Real-time estimation	• Future horizon
Research purpose	• Occupancy pattern analysis • Lighting system control • Ventilation control	• For system predictive control (HVAC) • Demand response analysis • Delivery route/space optimization
Data acquisition	• Sensors	• Sensors • Results of occupancy detection
Input data	• Ambient parameters • Contextual information	• Historical data • Ambient parameters • Contextual information

Fig. 3. Distinctions between "occupancy detection" and "occupancy forecast"

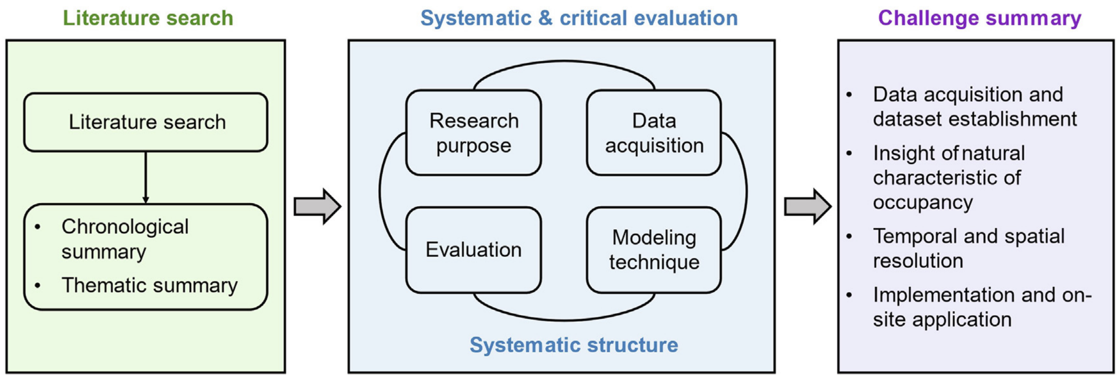


Fig. 4. Overall review methodology.

After an overall review and technical analysis, the challenges and perspectives of occupancy forecasting research were proposed.

2.2. Literature search

Currently, occupancy forecast-related research mostly uses the term "prediction" Therefore, the key phrase "occupancy prediction in buildings" was used to search the Web of Science and Scopus databases, with the publication year limited to after 2009. The search operators for the two databases were as follows:

$$TS = ((\text{occupan}* \ \text{NEAR}/5\,\text{predict}*)\ \text{AND}\ \text{building}*), \tag{1}$$

$$(\text{occupan}*\ \text{W}/5\,\text{predict}*)\ \text{AND}\ \text{building}* \tag{2}$$

where, (1) was the search operator used for Web of Science, and (2) was the search operator used for Scopus.

Based on the literature search, related article supplement, and final round of searching on April 25, 2021, 258 articles mentioning occupancy prediction research were found. Among these articles, 88 articles mainly focused on future occupancy prediction. We conducted a preliminary analysis based on these articles to identify the basic information of the search results.

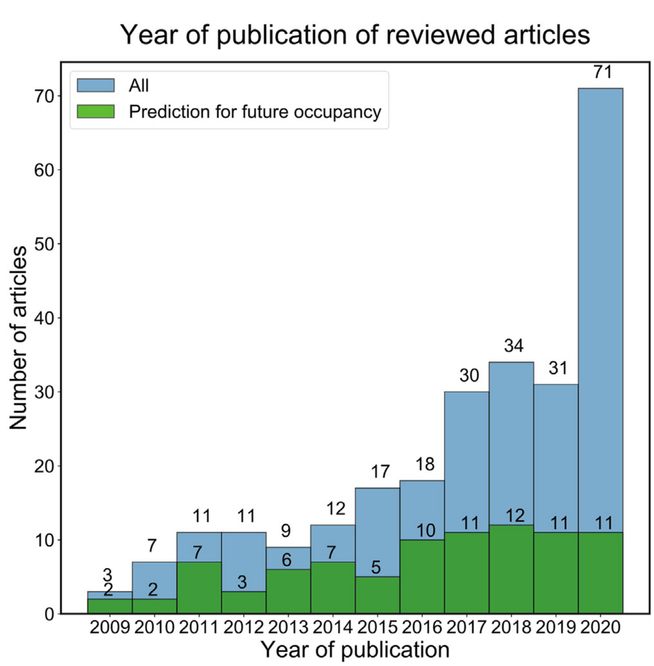


Fig. 5. Year of publication of reviewed articles.

(1) Chronological summary
Since 2009, occupancy prediction has been an attractive research topic, and the number of publications has continued to increase gradually (Fig. 5). A detailed review was conducted to ensure that the research focused on future occupancy prediction. To avoid misleading information, the data for 2021 are not presented because of their incompleteness.
(2) Thematic summary
For the thematic analysis, we summarized the type and publisher of all related articles (Fig. 6). Overall, 59% of the articles were from journals (153 articles), and there were 48 more journal articles than conference articles (105 in total).

According to the article numbers for each publisher, the main thematic research areas of occupancy forecasting were electronic engineering, building technology, and computing machinery. Research in electronic engineering focuses on occupancy detection technology and related control realization. For building technology, the research mainly focuses on the impact of occupancy on building design and operations to conserve energy consumption. For computing machinery, the occupancy detection and forecasting algorithm was analyzed.

2.3. Systematic and critical evaluation

Occupancy forecasting research was critically analyzed from the aspects of research purpose and complete methodology, which includes data acquisition, modeling techniques, and evaluation. As such, the main part of this article is composed of four aspects (Fig. 7).

All research articles related to occupancy forecasting were categorized and critically evaluated according to these four aspects. The research purpose consists of the application areas and detailed requirements for occupancy forecasting. Data acquisition considers the building type, prediction object (e.g., occupancy state or occupant number), occupancy detection method, and the size of the dataset. The modeling technique includes input, temporal and spatial resolution, lagged variable (the time steps of historical data used for prediction), prediction window (the time steps of the predicted occupancy data, e.g., next hour or next day), and a prediction algorithm. Finally, evaluation comprises a split of the training/validation/testing set, evaluation methods, and evaluation results.

There is a high correlation among these four aspects that comprehensively instruct occupancy forecasting research. For example, the application area determines the type of occupancy data to be acquired. Different applications are related to different temporal and spatial resolutions for occupancy prediction research, as well

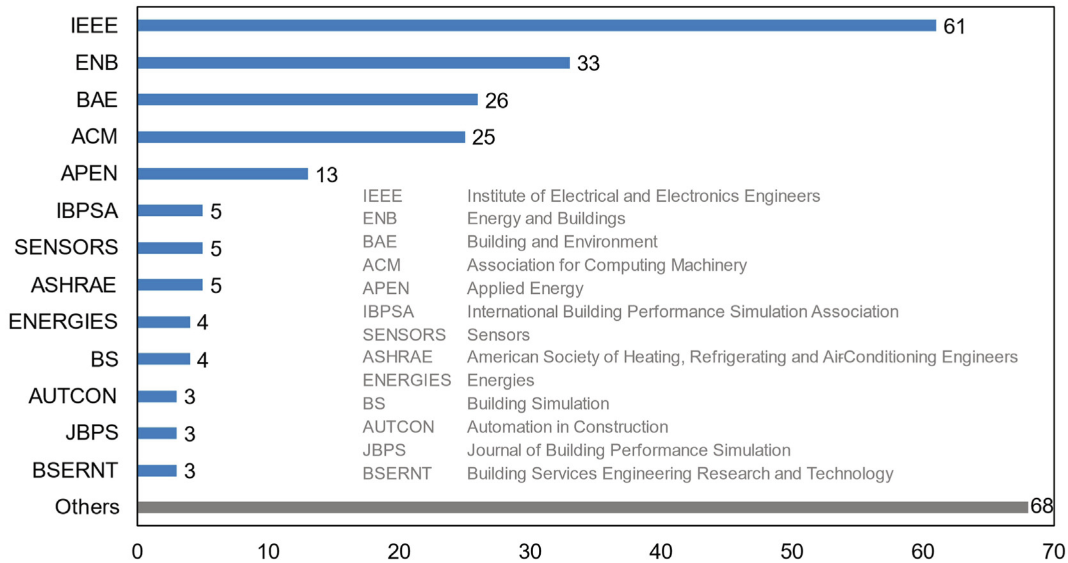


Fig. 6. Thematic summary of reviewed articles.

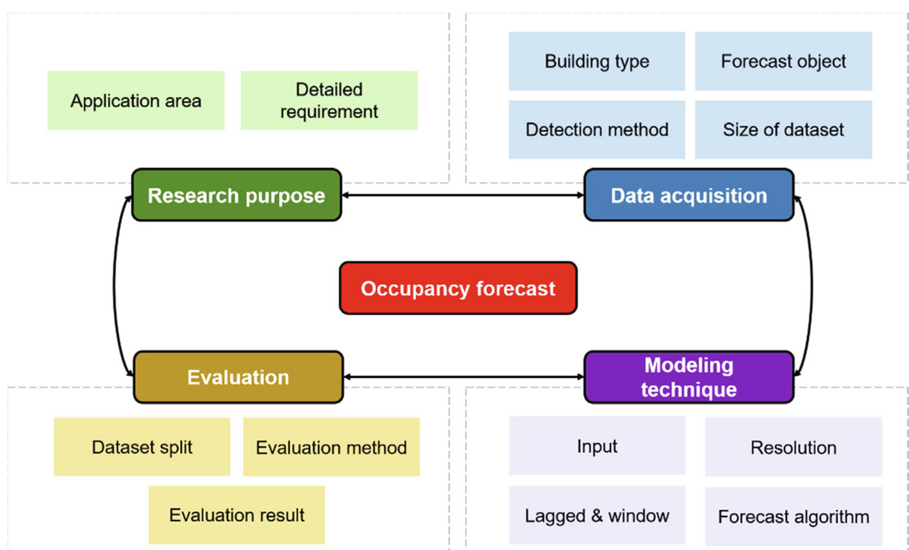


Fig. 7. Review framework on future occupancy prediction.

as different prediction windows. Furthermore, for application for occupancy forecasting, the results should generally be evaluated using the occupancy forecast performance and fit-for-purpose metrics.

3. Research purpose

3.1. Application areas

As we identified many similar conference or journal articles, there were a total of 70 articles with relatively independent research on occupancy forecasts. From the results of the literature review, 68.5% of the articles did not only cover occupancy forecasting but also mentioned potential applications. With the development of occupant-centric designs and controls in buildings [8,9], future occupancy prediction has been widely explored, as it can benefit various application fields. Specifically, occupancy forecasting can facilitate building management planning and supervision of building operations while considering electricity market needs [44].

The statistical results of the articles that mention different application areas are shown in Fig. 8. Among the application areas, 43 of the 48 articles were related to control strategies, including HVAC, lighting, blinds, and appliances; these factors are major contributors to occupancy forecasting in buildings. Furthermore, predictive strategies show better energy-saving performance and quality of controlling services than reactive strategies [44]. Thus, the predictive control method could enhance the energy-saving performance and temperature regulation [45] and could be combined with a demand–response strategy [46]. D'Oca et al. [18] found that occupancy profiles can help optimize appliances, plug loads, lighting use, HVAC control systems, fresh air requirements, internal heat gain, and building design plans in office buildings.

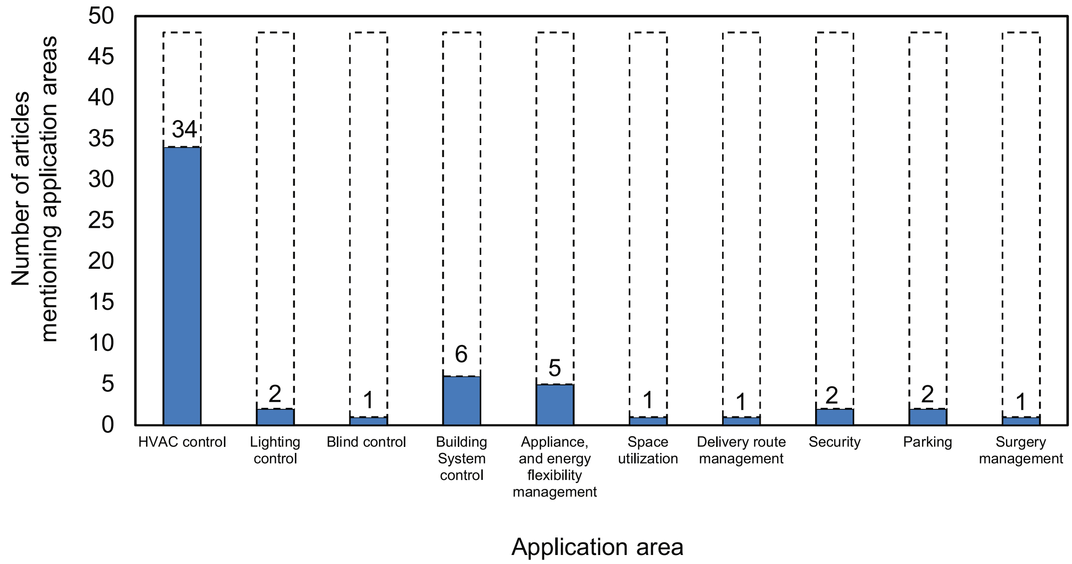


Fig. 8. Number of articles mentioning different application areas (45 articles mention the application for occupancy forecast research).

Moreover, other studies have demonstrated the significance of occupancy forecasts for building operations, which guarantee both energy conservation and occupant comfort. Erickson [15,22,47,48] proposed a series of occupancy forecast methods for usage-based demand control strategies for air conditioning in order to reduce energy consumption. Oldewurtel et al. [49] found that an occupancy forecast could be integrated with an MPC framework for room automation, such as HVAC, lighting, and blind control. Dong [50–52] focused efforts on the MPC of building heating and cooling systems. The occupancy state and number can be predicted to improve energy and comfort management, and the number of occupants can guide ventilation strategies [52].

In residential buildings, energy consumption is affected by occupancy behavior [53]. Many studies have been conducted to improve heating and cooling control in homes using occupancy state forecasts. Preheating ensures a comfortable thermal environment and simultaneously decreases energy consumption [54,55]. Specifically, a home HVAC system may be automatically turned on/off, and a deep setback temperature may be selected according to the occupancy state forecast of the home [17]. Furthermore, the predictive control strategy of smart homes also considers the requirements of the entire power grid to some extent [56].

At different scales, occupancy forecasts can be adapted for various applications. At the building level, they can be used for automated building energy management, real-time evaluation of building energy flexibility, and accurate operation strategies. Demand-response events can be scheduled by predicting the occupancy state in commercial buildings at different resolutions [57]. At the grid level, occupancy forecasts can promote demand–response analyses, dynamic pricing, and demand-side management [44]. Using occupant behavior cognition, real-time smart grid energy management can be realized to reduce the peak load and conserve energy by allocating renewable resources, such as solar sources [58].

Moreover, a few studies have applied occupancy forecasting in buildings to domains outside of operation control, which increases the number of occupancy forecast applications. Das et al. [59] found that space usage patterns can be identified to enhance space efficiency by predicting the occupancy presence in commercial buildings. Ohsugi et al. [60] optimized a package delivery route by predicting the occupancy state in residential buildings based on electricity usage. Sama et al. [61] identified security issues in smart homes, and Ali et al. [62] predicted the occupancy of parking places to provide advanced guidance to drivers regarding the location and daily and hourly occupation data of parking lots. Atif et al. [63] analyzed a method to predict the availability of parking spaces and recommended a route to minimize the journey duration for selecting a parking lot, ultimately relieving traffic congestion.

3.2. Detailed forecast characteristics

Gaetani et al. [64] emphasized the importance of fit-for-purpose occupant behavior modeling for buildings, which is the same as occupancy forecasting for buildings. The detailed requirements of occupancy forecast research can vary with the aims of different applications. Determining the application aim and comprehensively analyzing the specific requirements and characteristics of occupancy forecasts are crucial for conducting validated research and occupancy forecast-based applications.

Because of the transient nature of heat transfer in building fabrics, as for demand–response HVAC control, real-time occupancy detection is not adequate to control HVAC systems, thereby emphasizing the importance of occupancy prediction in the future [34,65]. The results of occupancy forecasting must be accurate, reliable, and able to determine occupancy changes and repetitive patterns in real time [18,22]; these factors are considered to be the center of occupancy-based building energy management [15]. For current HVAC operations, schedules tend to consider a room to be fully occupied when it contains the maximum number of occupants [47].

However, it is sometimes insufficient to simply predict the presence state without the number of occupants during HVAC control [23]. Because the effect of HVAC possesses hysteresis, an accurate occupancy forecast will help guide precooling or preheating.

Furthermore, occupancy forecasting should ensure system maintenance. If the occupancy forecast is too aggressive, home systems may be frequently turned on/off; the resulting recurring equipment cycling wastes energy and shortens the lifetime of the system [17].

Meanwhile, the stochasticity, variety, repetitive, nonstationary, and drift characteristics of occupancy present significant challenges to its application performance. Owing to the nonstationary features of occupancy patterns in buildings, an offline occupancy model is not recommended for on-site applications [34].

4. Data acquisition

4.1. Building type and forecast object

Due to accuracy and privacy issues, data acquisition is one of the challenges of occupancy-related studies [66]. The limited applicability and acceptability of occupancy detection means that studies tend to be limited to academic buildings (laboratories or offices in university/research institutes), which affects the amount of valid collected data. In this section, the issues of occupancy data acquisition are reviewed and discussed.

Based on the literature review, 39.7% of the buildings studied in the occupancy forecast field were academic buildings. Other building types included residential (27.4%), office (19.2%), and other buildings (13.7%), such as mosques [67], airports [68], parking lots [62,63], hospitals [69], and self-established test beds [25] (Fig. 9). Note that some articles mention more than one building type, causing the sum of all the building types of the occupancy forecasts to be more than 70.

The object of occupancy forecasting and data acquisition varies for different applications and aims according to detailed requirements. The lighting control of a room or HVAC control at the zone level is based on occupant presence detection [70–72], while the fresh air ventilation control or airflow rate control at the zone level requires the occupant number as an input [29,73]. Research has also been conducted to adjust the temperature setpoint according to the predicted number of occupants in a room [10]. To realize the control of HVAC systems, several studies predicted the duration of presence [50] or that for the remainder of the day [74] to guide turning off the HVAC system or setback temperature control ahead of the leaving time. The setback temperature and thermostat control can also be based on the forecast of arrival and departure times, which are considered another form of occupancy state [34,75,76]. There is also research predicting the occupancy stage (such as empty, low, medium, and full), which differentiate between the presence and detailed occupant number [44,77]. For more elaborate research, the occupant location is predicted to maximize the energy efficiency [78] or security guard [61]. The number of occupied beds in hospitals is forecasted for the specific application of surgery management in hospitals, which is similar to occupant number forecasting [69].

Meanwhile, for residential buildings, HVAC system controls may also consider different temperature setpoints in the active or inactive (sleeping) stages [17]. The occupancy state in a household is instructive for smart home operations [79] and grid requirements [56]. For space utilization and optimization, the forecast object turns to agents to determine the future presence location in a certain commercial building [59]. For delivery route optimizations, future forecasting of occupancy states in households may overcome this issue [60].

A summary of the predicted occupancy objects is presented in Table 1. For occupancy prediction in the future, nearly half of the research focused on the occupancy state forecast, while over a quarter of the research predicted the detailed number of occupants in the building. Meanwhile, only four articles mentioned location forecasts for smart home operations and smart grid energy management [58,59,61,78]. Thus, the dominant forecast objects are presence and occupant number, not occupancy trajectory, which is mainly the research target of occupancy detection for security issues and target of occupancy simulation for layout design.

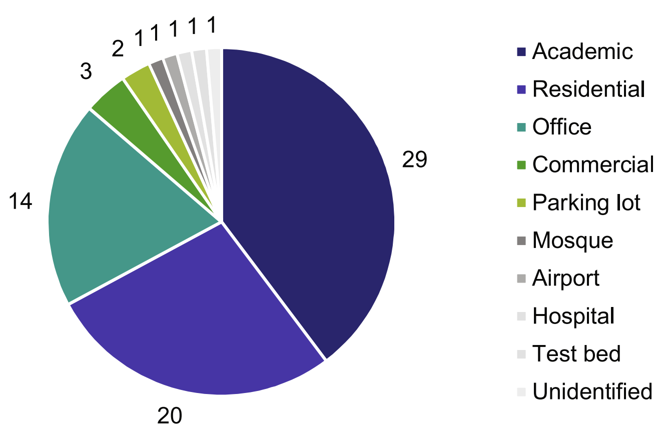


Fig. 9. Building types examined in occupancy forecast studies.

4.2. Collection method and size of dataset

Occupant-related data acquisition faces the challenges of data accuracy and personal security, which require manual labor and affect the data quality and size of the occupant behavior dataset [73]. With the innovation and development of occupancy data collection, research on occupancy forecasting has been inundated with valid ground truth data rather than occupancy data from simulation methods [80].

For occupancy data collection, traditional research methods tracked the presence and absence state of rooms using typical occupancy and vacancy duration questionnaires [49], motion sensors [24,26,34,57,61,71,74,76,81–92], radio frequency identification (RFID) sensors [55], door sensors [17], Bluetooth beacons [93–95] and mobile phone records [54,75]. For occupant number acquisition, detection methods include motion sensors

Table 1
Articles with different forecast objects.

Forecast object		Number of articles
Presence-related	Presence	34
	Presence and duration of presence	1
	Presence and time for presence	1
	Presence and status	1
	Time for presence	2
Occupant number -related	Occupant number	21
	Occupancy level (low, medium, high)	1
	Occupant number pattern type	1
Location-related	Location	3
Combined	Presence and location	1
	Presence and occupant number	1
	Presence and occupancy level (low, medium, high)	1
	Duration of presence and occupant number	1
	Presence of occupant and occupant number	1

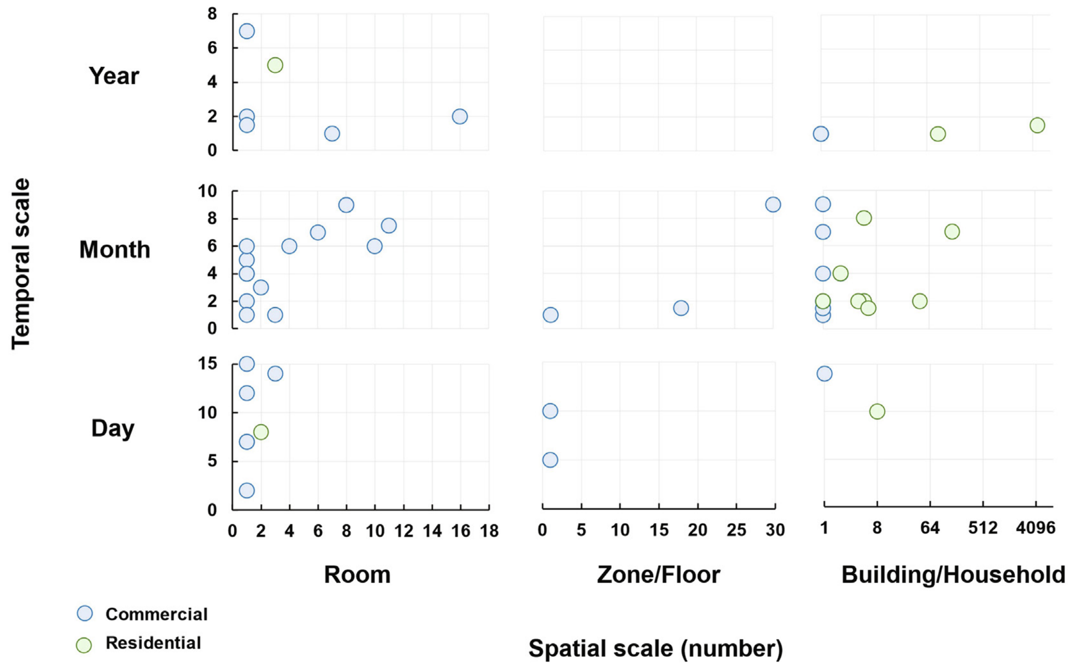


Fig. 10. Summary plot of occupancy dataset size for different studies.

installed at each exit [96], sensor network capturing and processing image data [22,67,72,97], depth-image/video cameras [44,59,65,83,98,99], probing signals from Wi-Fi networks [11,58,68,73,100,101], manual logs using a mobile (Android) app [29], geo-fencing apps on mobile phones [102], and manual counting via on-site surveys [77,102] or video records [103]. The accuracy of image/video data processing to distinguish occupant numbers ranged from 80% to 95% in different studies. Some studies have estimated the number of occupants by analyzing the ambient parameters or appliance electrical loads, in which the temperature, relative humidity, acoustics, CO_2 data [25,50,104], and electrical load [79] may be collected by sensors. Additionally, several studies have directly developed occupancy forecast models based on publicly available datasets [30,60,78,105,106].

Based on our literature review, we found several articles focusing on occupancy detection and estimation in buildings that conflated the expression of "detection" with "forecast" [107–111]. These studies mainly estimated real-time occupancy and did not determine future situations. Because the application areas and detailed modeling techniques differ, it was suggested that a specific term for "detection and forecast" should be used.

Owing to the distinction of the data acquisition method, the size of the occupancy dataset also varies among different studies, which influences the stability and generalizability of the occupancy forecast models. Fig. 10 shows a summary of the results of the amount of valid data based on the findings of the literature review. The dots are separated between commercial (blue) and residential buildings (green). Regarding the spatial scale, "building" is the expression used for commercial, and "household" is used for residential buildings. As seven articles did not mention the dataset size and several articles only stated the size on the temporal or spatial scale, these articles were not included in the figure. More detailed information is listed in Table A.1 in Appendix A. Overall, most of the articles mentioned the size of the valid dataset used for their occupancy forecast analyses. The scale of the datasets includes rooms or buildings. For the spatial scale of the room, the valid duration was concentrated from two months to eight months, whereas that of an entire building/household tended to be less than two months. Razavi et al. [106] proposed an occupancy forecast model based on a large dataset from a smart meter dataset that contained 5000 households and spanned for 1.5 years. Such large-scale datasets are valuable for model establishment and validation.

In this review, we mainly analyzed the temporal and spatial scales to evaluate the size of the occupancy datasets. We found that few studies predicting occupant positioning collected data that were mainly quantified by tracking the number of occupants [78] and are thus not discussed in detail in Table A.1.

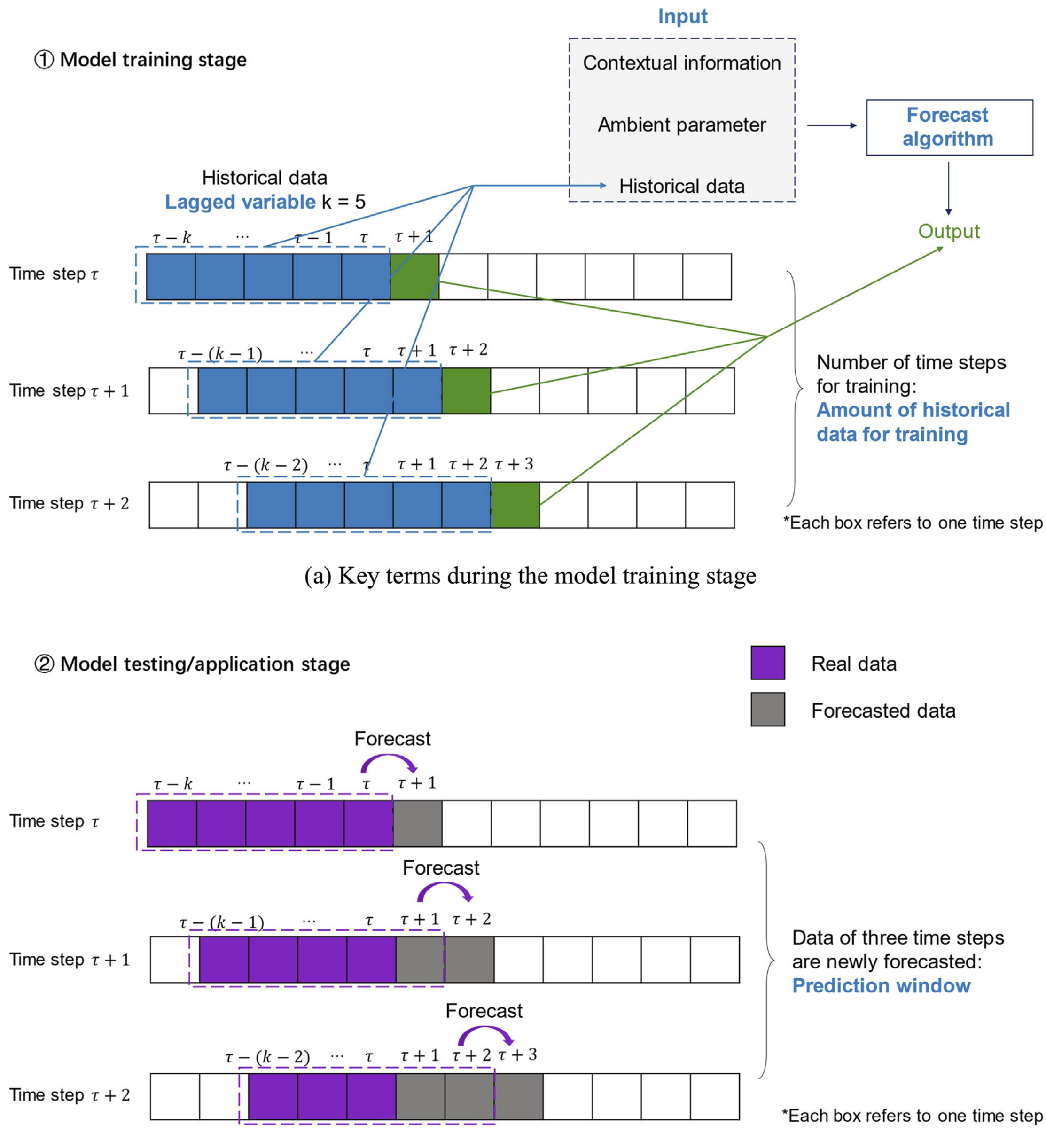
5. Forecast methods and techniques

Several key terms are used for occupancy forecasting, most of which are discussed in this section. The terms have been written in blue, as shown in Fig. 11, and include the input, lagged variable, and forecast algorithm. These terms are reviewed and analyzed in the following subsections.

5.1. Inputs

This subsection analyzes the forecasting models and techniques that are crucial for performing occupancy forecasts. In an occupancy forecast study, it is important to correlate the study with the physical nature and characteristics of occupancy, which have the potential to improve robustness and generalizability.

The input of a forecast model is important because it reflects a deep analysis of the inner physical features and characteristics of occupancy. As improper inputs may cause bias during model establishment and training, inputs greatly affect the accuracy of occupancy forecasting. In general, the inputs for occupancy forecast methods can be divided into three categories: (1) historical occupancy data; (2) contextual information [18], such as season, hour of the day, day of the week, and special holidays; and (3) ambient parameters [50,60,101,106,112], including temperature, CO_2, plug loads, motion sensor data, and acoustics. Position-based occupancy forecasts use information from mobile phones



(a) Key terms during the model training stage

(b) Key terms during the model testing or application stage

Fig. 11. Key terms of occupancy forecast research.

Table 2
Summary of inputs for occupancy forecasting.

Historical data	Contextual information	Ambient parameter	Reference
√			[10,11,22,24,26,34,49,54–58,61–63,65,70,71,74–76,79,82–84,87–90,94,95,97,98,100,104,105,113]
	√		[59,67–69]
		√	[50,112,114]
√	√		[18,29,30,44,73,78,80,81,86,91,93,96,102,103,115,116]
√		√	[25,99,101]
	√	√	[17,28,60,72,77,106]

as inputs, such as cellular information, Wi-Fi fingerprints, and location coordinates from Wi-Fi positioning systems (WPSs) and global positioning systems (GPSs) [75]. Inputs from more than one category are sometimes introduced into the forecast model [17,25,44]. An input summary is listed in Table 2, and detailed information is provided in Table B.1, B.2, and B.3.

For historical occupancy data input, the researcher normally utilizes probability analyses to predict occupancy or temporal sequential characteristic analyses such as the autoregressive integrated moving average method [65]. For contextual information input, the periodicity of occupancy is considered with a period of one day or one week. Meanwhile, holidays are usually listed as an important factor in occupancy patterns. In specific places, contextual information can be supplemented by flight schedules for airports [68] or meeting schedules for conference rooms [28]. For the ambient parameters, researchers determine the forecast based on the assumption that the occupants in buildings will affect certain environmental parameters such as CO_2 concentration, triggering motion, or acoustic sensors. However, only the environmental parameters of past time steps can be utilized to predict occupancy in future time steps. This is different from occupancy detection research, in which the occupancy of the current time step is inferred using the parameters of the current time step as input.

It is essential to comprehensively determine the purpose and requirements of the forecast while determining the inputs. For example, occupancy and ventilation control can both affect the CO_2 concentration in a room; thus, if the purpose of the study is to optimize the ventilation control strategy using an occupancy forecast, the CO_2 concentration should not be listed as the input variable.

5.2. Temporal and spatial resolution, lagged variable, and prediction window

The temporal and spatial resolutions of occupancy forecasts vary for different applications and scales [36]. The temporal resolution for an occupancy forecast varies from second to minute and hour, while the spatial resolution ranges from a single room to a zone, level, or even an entire building (Fig. 12). In total, 88.6% of the studies reported both temporal and spatial resolutions, while 98.6% stated spatial resolution. For studies controlling the HVAC system based on an occupancy forecast, the temporal resolution is often set at 10–20 min. Meanwhile, the resolution for heating (1 h) is normally longer than that for cooling (15 min) [50]. Several studies have also analyzed the impact of temporal resolution on forecast performance [65,82,83,95,97,100,102]. The results show that a higher resolution normally achieves higher accuracy. Most (61.4%) of the studies focused on the room scale for spatial resolution, while 27.1% focused on the

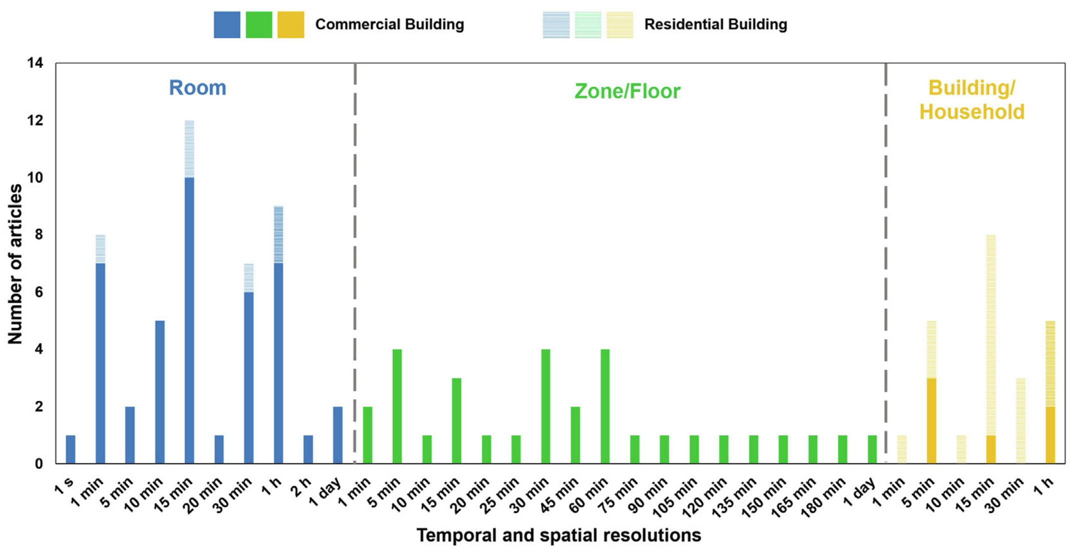


Fig. 12. Temporal and spatial resolutions of reviewed studies.

Table 3
Summary of lagged variable in reviewed articles.

Lagged time step	Lagged duration	Reference
1	1 s	[22]
1	10 min	[18]
28 (historical data with the same time point)	28 days	[24]
5	75 min	[54]
7 (historical data with the same time point)	7 days	
All historical data of current partial days start from 14	/	[55]
Last prayer event	/	[67]
24	1 day	[73]
Artificial neural network – 3, 11, 12, 15	45 min, 5.5 h, 12 h, 30 h	[65]
Support vector machine – 3, 5, 12, 18	45 min, 2.5 h, 12 h, 36 h	
60 spatial coordinate data points	/	[59]
24	12 h	[30]
12	3 h increments from 3 h to 36 h	[97]
Change points detection for the moving windows	/	[82]
6	30 min	[26]
6 (historical data with the same day of week and time point)	6 weeks	[94]
5	5 h	[80]
3–12 symbols/activities	/	[61]
7 for k-nearest neighbors method	70 min	[57]
4 for preheat method	40 min	
24	24 h	[99]
7 (clustering day type)	7 days	[101]
8	8 min	[114]
20	20 days	[69]

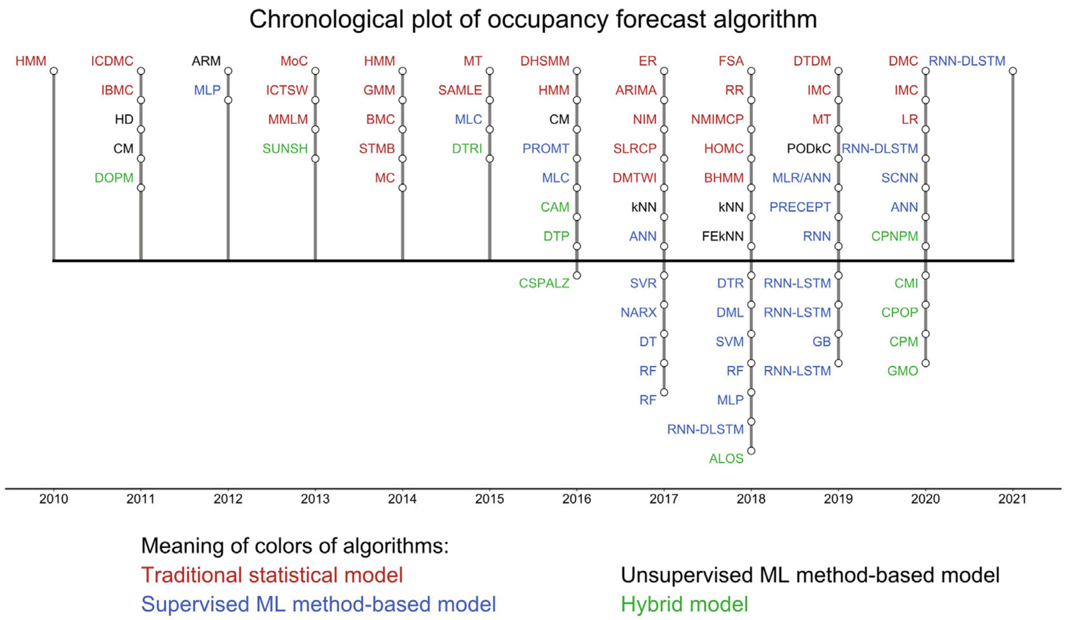


Fig. 13. Chronological plot of occupancy forecast algorithms (As for the abbreviations of different algorithms in the figure, please refer to Table 4 below).

building scale (including the presence forecast in a single household or residential building).

A lagged variable refers to the order of historical data used to correlate or conduct a regression analysis at each time step. For example, the Markov chain model assumes that the occupancy state is related to the state of the last time step, making the lagged variable equal to one time step. This reflects the cognition of the study when considering occupancy patterns. If the lagged variable is too short, the occupancy pattern will not be well reflected, causing a deficiency in the historical data for the occupancy forecast. Conversely, if the lagged variable is too long, the forecast performance is compromised because of the introduction of excessive noise. In the current research, the lagged variable is not well discussed, and only 30.0% of the reviewed articles use the term. Chen et al. [65] optimized the lagged variable for different temporal resolutions, revealing that the optimized lagged variable decreases at higher temporal resolutions. Thus, temporal resolutions of 15 min, 30 min, 1 h, and 2 h had optimal lagged variables of 3, 11, 12, and 15 time steps, respectively. The details are summarized in Table 3.

The prediction window, also called the prediction horizon, denotes the duration of the occupancy forecast in the following time steps and is highly related to on-site applications. In an HVAC system control study [50], the prediction window for heating was 24 h, while the prediction window for cooling was 3 h. In a study on residential building preheating control [55], the prediction window was evaluated from 15 to 180 min (3 h) and finally set to 3 h. Moreover [54], for home heating control, a prediction window is required to guarantee the temperature of home to return to the comfort zone before people return home; this window is determined and calculated using the current indoor air temperature, the target comfort temperature, and the forecast for the outside temperature. Adamopoulou et al. [44] considered the frequency of forecasting and found that different combinations of prediction windows and frequencies correspond to different applications. For example, 8 h – 8 h (prediction window – frequency) is related to operational and energy resource planning, and 15 min – 15 min is related to demand responses or real-time predictive control. Note that a longer prediction window induces higher errors with more uncertainty in the unknown future period [29,60,65,82,83,99,106].

For forecast research, the target variable depends on a certain hidden and unknown context. A changing context may lead to variations in forecasting targets. With deeper insight into the "concept drift" in building occupancy forecasting, which is used to depict the routine/characteristics of occupancy change with time [117], the amount of historical data used for model training or generating the state transition probability matrix is also important. Hence, if the amount of historical data is too large, the trained model may encounter too much chaotic information, thereby decreasing the accuracy of the forecast. By contrast, the amount of historical data may be too limited to grasp the overall routine pattern for the forecast. Consequently, it is crucial to determine the valid amount of data for occupancy forecasting. This amount can be highly influenced by pattern modifications; a sensitivity analysis is the optimal solution to overcome concept drift. Furthermore, regularly updating the model could help improve the forecast performance and overcome the concept drift of the occupancy pattern. However, selecting an appropriate update interval is difficult. Yuan et al. [116] retrained the forecast model every day using all historical data before the forecast day and achieved a steady improvement in the forecast performance. Currently, only 32.9% of the reviewed articles mention the amount of training data used. In one study [30], the historical data for model training was collected over eight weeks. Although discourse on this topic is currently limited, Tahmasebi et al. [89] analyzed the influence of data utilization options with three alternatives: 5, 10, and 20 days. Salimi et al. [95] conducted a sensitivity analysis to investigate the effect of the data collection period on the accuracy of the occupant number forecast model. Turley et al. [102] changed the collection period of training data from 1 week to 4 weeks to optimize the forecast model. They also discussed the efficacy of a moving mode for model training over time.

5.3. Forecast algorithm

Over the past decade, forecast algorithms have significantly improved the performance of occupant forecasts. Fig. 13 shows the evolution of methods for occupancy forecasting (models for benchmark and comparison are omitted in the figure). The complete descriptions of the forecast algorithm abbreviations are listed in Table 4.

In general, forecast methods can be classified into four categories: (1) traditional statistical methods, (2) unsupervised machine learning methods, (3) supervised machine learning methods, and (4) hybrid models. With chronological evolution, the supervised machine learning method-based model began to dominate, replacing traditional statistical models. Currently, research concerning hybrid models is relatively limited; however, these models have the potential to improve calculation precision and consider the natural characteristics and routines of occupancy.

A major traditional statistical model is the Markov chain-based model [17,22,50,63,70,77,79,104,112,113], which is based on transition probability and has been developed and innovated over time. Another type of traditional model predicts the occupancy state based on the historical occupied ratio [24,87] and the occupant number based on historical data at the same time point [29]. Other researchers have established forecast models based on other statistical algorithms or distributions, such as Gaussian distribution [50], Gamma distribution [68], maximum likelihood estimation [34], Smith-Waterman alignment algorithm [75], logistic regression [105], finite state automata [10], and temporal sequential analysis [65].

Machine learning methods have greatly contributed to data analytics in buildings, including occupancy research [118]. For unsupervised machine learning method-based models, a predominant method predicts the occupancy state or number using the clustering results of the data from several past data points. These results are considered to represent different occupancy patterns and distinguish weekdays from weekends, even on different days of the week. Scott et al. [55] utilized the Hamming distance to determine the distinctions between the patterns. In addition, several researchers have introduced clustering algorithms for further occupancy forecasting, such as k-means [56] and k-nearest neighbor techniques [60,84]. Vazquez et al. [90] evaluated the performance of different clustering forecast algorithms, including

Table 4
Abbreviations of forecast algorithms in the reviewed articles.

Abbreviation	Forecast algorithm	Ref.	Abbreviation	Forecast algorithm	Ref.
ALOS	Automatic learning of an occupancy schedule	[76]	ANN	Artificial neural network	[65,69]
ARIMA	Autoregressive integrated moving average	[65]	ARM	Association rule mining	[78]
BHMM	Backward hidden Markov model	[77]	BMC	Blended Markov chain model	[70]
CAM	Context-aware method based on the spatiotemporal analysis from (semi-)Markov model	[44]	CM	Clustering-based model	[90,115]
CMI	Clustering and motif identification-based approach	[101]	CPM	Contextual probabilistic model	[102]
CPNPM	Clustering probability based nonparametric modeling	[91]	CPOP	Clustering-based probabilistic occupancy prediction model	[95]
CSPALZ	Compression-based sequential prediction methods, based on the active LeZi algorithm	[61]	DHSMM	Dynamic hidden semi-Markov model	[112]
DMC	Discrete Markov chain model	[63]	DML	Distributed machine learning	[58]
DMTWI	Dynamic Markov time-window inference approach	[11]	DOPM	Occupancy prediction model built by using decision guidance query language (DGQL) framework	[28]
DT	Decision tree	[103]	DTDM	Dwell time distribution-based mathematical model	[68]
DTP	Decision tree with pattern	[96]	DTR	Decision tree-based model with routine	[29]
DTRI	Decision tree with rule induction	[18]	ER	Event-based regression model	[67]
FEkNN	Feature extraction-based k-nearest neighbor (KNN)	[84]	FSA	A novel finite state automata	[10]
GB	Gradient boosting	[106]	GMM	Gaussian mixture models	[50]
GMO	Graph mining-based optimization	[119]	HD	Hamming distance-based model	[55]
HMM	Hidden Markov model	[17,25,50]	HOMC	Higher order Markov chain occupancy model	[104]
IBMC	Inhomogeneous blended Markov chain model	[22]	ICDMC	Inhomogeneous closest distance Markov chain model	[22]
ICTSW	Inter-cell transition - Smith-Waterman alignment algorithm	[75]	IMC	Inhomogeneous Markov chain model	[113,116]
KNN	k-nearest neighbors	[60,74]	LR	Linear regression model	[116]
MC	Markov chain model	[79]	MLC	Multilabel classification	[81,86]
MLP	Multilayer perceptron	[60,72]	MLR/ANN	Model based on combination of linear regression and artificial neural networks	[73]
MMLM	Mixtures of multilag Markov chains	[26]	MoC	Monte Carlo-based model	[49]
MT	Daily binary occupancy profiles based on aggregated past–presence data	[24,87]	NARX	Nonlinear autoregressive with eXternal input neural network	[93]
NIM	New inhomogeneous Markov model	[82]	NMIMCP	New moving window inhomogeneous Markov model based on change point analysis	[83]
PODkC	Proper orthogonal decomposition-based k-means clustering for occupancy prediction model	[56]	PRECEPT	A variant of recurrent neural network known as gated recurrent unit (GRU) network	[59]
PROMT	Predicting occupancy presence in multiple resolution with time-shift agnostic classification	[57]	RF	Random forest	[60,98,103]
RNN	Recurrent neural network	[30]	RNN-DLSTM	Recurrent neural network with deep long short-term memory units	[62,100,114]
RNN-LSTM	Recurrent neural network with long short-term memory units	[94,97,99]	RR	Rule-based model with routine	[29]
SAMLE	Self-adaptive occupancy learning control algorithm based on maximum likelihood estimation	[34]	SCNN	Sequential & contextual neural network	[116]
SLRCP	Novel statistical model based on logistic regression model with change points	[105]	STMB	Self-tuning Markov occupancy model with on-line Bayesian training	[71]
SUNSH	Improved sensor-utility-network (SUN) algorithm with incorporation of scheduling and adaptive historical data	[80]	SVM	Support vector machine	[60]
SVR	Support vector regression	[65]			

self-organizing maps, eXclusive self-organizing maps, fuzzy C-means clustering, k-means, k-means with repeated bisection, graph clustering, and support vector clustering.

For supervised machine learning method-based models, studies have analyzed the forecast performance of multilayer perceptrons [72] and multilabel classifications [81]. With the increasing complexity of models, more complicated neural networks (nonlinear autoregressive with eXternal input neural network [93], recurrent neural network with long short-term memory units [62,94,97,99], gradient boosting [106]), and other models, such as support vector regression [65], decision tree [103], and random forest [60,98], must be introduced to forecast research.

There are also models that introduce hybrid concepts. Liang et al. [96] combined a clustering method for occupancy patterns

Table 5
Occupancy forecast evaluation metrics.

Evaluation metrics	Equation	Reference
Average static occupancy duration	/	[22]
Jensen-Shannon divergence (JSD) for room occupancy distribution	$KL(P_1\|P_2) = \sum_{x\in X} P_1(x) log \frac{P_1(x)}{P_2(x)}$ $JS(P_1\|P_2) = \frac{1}{2}KL\left(P_1\|\frac{P_1+P_2}{2}\right) + \frac{1}{2}KL\left(P_2\|\frac{P_1+P_2}{2}\right)$	[22]
Duration between entrances and exits of a room	/	[22]
Percentage of forecasts with errors below specific thresholds for five statistics	FA: First arrival time error; LD: Last departure time error OD: Occupancy duration error SM: Occupancy state matching error NT: Number of transitions error	[24,89]
Confusion matrix and related metrics	$\text{CM} = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}$; $\text{PPV} = \frac{TP}{TP+FP}\text{TPR} = \frac{TP}{TP+FN}$; $\text{ACC} = \frac{TP+TN}{TP+TN+FP+FN}\text{inACC} = 1 - \text{ACC}$; $\text{F} = \frac{(1+\beta^2)*PPV*TPR}{\beta^2*PPV+TPR}LL = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(TPR)$ $+(1-y_i)\log(1-TPR))MMC = \frac{TP\times TN - FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$ CM: confusion matrixP: positive; N: negativeT: true; F: falsePPV: precision; TPR: recallACC: accuracy; inACC: inaccuracy or error rateLL: logarithmic loss; MCC: Matthews correlation coefficient	[17,18,26,54,55,57,60,72,76,81–83,86–88,90,91,102,112,114]
Confusion matrix-based receiver operating characteristic curve	/	[26,54,55,82]
Accuracy for occupant number or location forecast	$I\left[Y_i = \widehat{Y}_i\right] = 1, if Y_i = \widehat{Y}_i, else 0 \text{ACC} = \frac{\sum_{i=1}^{N} I\left[Y_i = \widehat{Y}_i\right]}{N}$	[10,25,30,50,61,78,113]
Accuracy with x-number tolerance	$I\left[\left\|Y_i - \widehat{Y}_i\right\|, x\right] = 1, if \left\|Y_i - \widehat{Y}_i\right\| \le x, else\ 0$ $\text{ACC} = \frac{\sum_{i=1}^{N} I\left[\left\|Y_i - \widehat{Y}_i\right\|, x\right]}{N}$	[11]
Proportion of different occu_num errors	/	[104]
ME (mean error)	$\text{ME} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)$	[65]
medE (median error)	$\text{medE} = \text{median}\left(Y_i - \widehat{Y}_i\right)$	[96]
MdAE (median absolute error)	$\text{MdAE} = \text{median}\left(\left\|Y_i - \widehat{Y}_i\right\|\right)$	[62]
MAE (mean absolute error)	$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left\|Y_i - \widehat{Y}_i\right\|$	[29,59,62,83,96,97,105,114,116]
Normalized MAE	$\text{NormalizedMAE} = \frac{MAE}{\overline{Y}}$	[29]
MAPE (mean absolute percentage error)	$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{Y_i - \widehat{Y}_i}{Y_i}\right\|$	[58,80,97]
MSE (mean squared error)	$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)^2$	[59,62]
MSLE (mean squared log error)	$\text{MSLE} = \frac{1}{N}\sum_{i=1}^{N}\left(\log(Y_i+1) - \log\left(\widehat{Y}_i+1\right)\right)^2$	[62]
RMSE (root-mean-squared error)	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{Y}_i - Y_i\right)^2}$	[10,29,58,59,62,65,69,71,73,83,94,96,97,100,114,116]
CVRMSE (coefficient of variation RMSE)	$\text{CVRMSE} = \frac{RMSE}{\overline{Y}}$	[29,67,73,95,98,99]
Accuracy based on CVRMSE	$\text{ACC} = 100 \times (1 - \text{CVRMSE})$	[73]
Total average NRMSE (normalized root-mean-squared error)	$RMSE\_period_t = \sqrt{\frac{\sum_{k=1}^{n}\left(\widehat{occ}_{jk} - occ_{jk}\right)^2}{n}}NRMSE\_period_t$ $= \frac{RMSE\_period_t}{\text{max}oxx - \text{min}occ}AverageNRMSE\_zone_{ij}$ $= \frac{\sum_{t=1}^{tp} NRMSE\_period_t}{tp}AverageNRMSE\_day_i$ $= \frac{\sum_{j=1}^{z} AverageNRMSE\_zone_{ij}}{z}TotalAverageNRMSE$ $= \frac{\sum_{i=1}^{TestD} AverageNRMSE_day_i}{TestD}$j: zone; k: time step; t: period; n: total number of compared values maxoxx, minocc: maximum and minimum occupancy value tp: total number of time periods; z: total number of zones TestD: total number of testing days	[44]
R-squared value	$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(\widehat{Y}_i - Y_i\right)^2}{\sum_{i=1}^{N}\left(\overline{Y} - Y_i\right)^2}$	[67,68,73,95,113,114]
SD (standard deviation)	$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \overline{Y}\right\|^2}$	[114]

with a decision tree algorithm to predict the number of occupants in a building. Nacer et al. [76] proposed the automatic learning of an occupancy schedule method to classify arrivals and departures using a clustering method and calculated the detailed duration using the expectation maximization algorithm. Ryan et al. [80] improved a linear state-space model by introducing a schedule term in the optimization function and setting up a mechanism to automatically update the historical data. Sama et al. [61] used the Active LeZi algorithm as a sequential compression method to generate an order-k Markov model.

However, several gaps in the research need to be carefully defined and addressed. Although occupant behavior varies in different buildings and rooms, summarizing common features or similar routines is the key to generalizing a prediction method. As the previously discussed concept drift, determining how to overcome the uncertain changes in the occupancy pattern over time is important for on-site applications. To date, efforts have been made to establish additional open-source datasets concerning occupancy states in buildings and to improve the generalizability and performance of methods and techniques.

6. Evaluation

6.1. Training/validation/testing set

The forecast method should be evaluated with sufficient historical data from either a test bed or on-site measurement to show the average performance of the model and guarantee its stability and robustness. A forecast evaluation should be a complete and normative process. In the initial step, the dataset is separated into training, validation, and testing sets to avoid potential overfitting, especially in a complex model [120].

The goal of an occupancy simulation is to represent the features and characteristics of the objective. Thus, during the parameter acquisition of the simulation model, it is not strictly required to separate the training and testing sets. However, dataset splitting is crucial for occupancy forecasting. A training set is used for model training, which is an optimization process employed to minimize the loss function of the model. A validation set is more popular in the area of machine learning studies and is used to guide hyperparameter optimizations for the model. Finally, a testing set indicates occupancy data occurring in the future and provides standard data for the comparison of different models.

According to the literature review, 43% of the articles mentioned separating training and testing sets or cross-validation, while only a few (6%) mentioned a validation set. The proportions of training and testing sets are also important for overfitting or underfitting issues. Among the review articles, no study has discussed the impact of set proportions on forecast performance, which potentially needs further exploration.

6.2. Evaluation metrics and results

To demonstrate the overall performance, occupancy forecast-related studies usually establish a set of evaluation metrics. These metrics are composed of two parts: the evaluation of the occupancy forecast results and related application performance.

For the occupancy forecast metrics (occupant number), the RMSE and normalized RMSE are most commonly used. Erickson et al. [22] evaluated how long room occupancy remains static in order to measure occupancy variability. They also proposed the Jensen-Shannon divergence for the room occupancy distribution. To predict the presence state in buildings, the confusion matrix and receiver operating characteristic curve were utilized to evaluate forecast performance. Mahdavi et al. [24] focused on the first arrival time, last departure time, occupancy duration error, occupancy state matching error, and number of transitions. A previous study also analyzed [91] the uneven frequency of the occupancy state, which refers to the phenomenon in which people generally spend daytime in the office. Therefore, certain metrics, such as the Matthews correlation coefficient, are used to avoid an imbalance in the evaluation.

Table 5 summarizes the detailed evaluation metrics for future occupancy forecasting. In total, 40% of the articles reviewed proposed accuracy to evaluate the final performance of the occupancy forecast (Fig. 14). Note that the definition of accuracy in some studies contains tolerance (refer to "Accuracy with x-number tolerance" in Table 5). In one study [11], accuracy was expressed as x-accuracy, in which the × occupant error was allowed. Therefore, the green columns in Fig. 14 refer to accuracy metrics with relative tolerance.

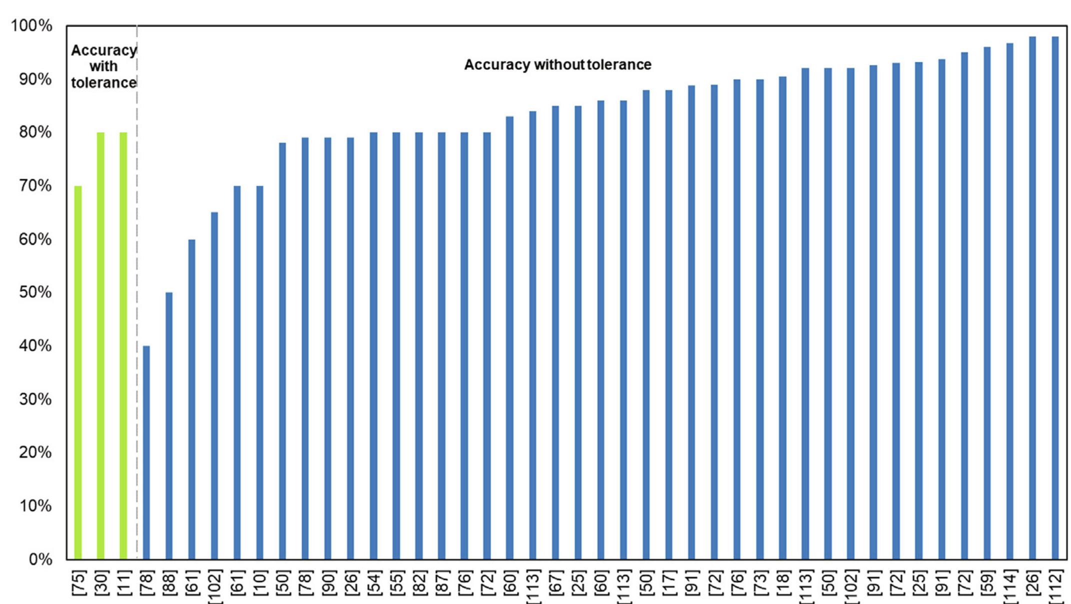


Fig. 14. Accuracy results in review articles.

Of the occupancy forecast-related applications based on the metrics, the most common is energy conservation with the help of occupancy-forecast-based control. In total, 16 articles mentioned energy conservation performance. Of these, 72% are simulation case studies, for which energy conservation ranges from 9 to 42%. Meanwhile, the energy conservation from on-site applications and measurements ranges from 18 to 30%. A detailed comparison of the baseline values is presented in Table 6.

Occupant comfort is an important indicator in research. In a study of residential heating and cooling control [17], a miss time was introduced to represent the control results; miss time is defined as the total duration when the home is occupied but the temperature does not meet the setpoint temperature. Kleiminger et al. [54] introduced discomfort degree hours as a measure of comfort loss. This term is defined as the average sum of the hourly differences between the actual indoor air temperature and the comfort temperature for all occupied intervals. In one study [67], the traditional predicted mean vote and predicted percentage dissatisfied metrics were used as indicators of occupant comfort. Dobbs et al. [71] considered comfort loss as a part of the loss function of model predictive control. They found that comfort loss increases with increasing deviation between the actual zone, the comfort setpoint temperature, and the occupancy level. In another study [90], the average difference between the real and desired temperatures when there are people at home, time in comfort during the occupied state, and the accumulated time to reach the comfort temperature during the occupied state were utilized to depict the control performance.

For other applications, the metrics are also fit-for-purpose. For instance, for parking lot guidance, the failure rate of the available parking space and the time consumed before entering a parking space are used to evaluate the control algorithm [63].

7. Challenges and perspectives

The structure of the summary of occupancy forecast research is shown in Fig. 15. Based on the literature review and analysis, there are still several challenges and perspectives for researchers to consider, including the following: (1) data acquisition and dataset establishment, (2) consideration of the natural characteristics of occupancy in buildings, (3) applicability and unification of temporal and spatial resolutions, and (4) on-site application and occupant behavior combined analyses.

Table 6
Energy conservation performance of occupancy forecast-based control.

Type of case study	Comparison baseline case	Proportion of energy conservation	Reference
Simulation	Static temperature setting	9%	[102]
Simulation	International Organization for Standard (ISO) 13790-standard heating model	12%	[54]
Simulation	Static nighttime temperature setback strategy	13%	[34]
Simulation	Base case without occupancy forecast	15%	[79]
Simulation	Conventional scheduled control	19%	[71]
Simulation	Static schedule (model predictive control with occupant number forecast)	20%	[116]
Simulation	Basic control with no occupancy info	20%	[10]
Simulation	Reactive heating system (On/off based on occupancy detection)	22%	[76]
Simulation	Typical heating system	28%	[17]
Simulation	Static schedule (adaptive model predictive HVAC controller)	30%	[116]
Simulation	Case of the corresponding day of the previous week with traditional control	31%	[67]
Simulation	Base case without pre-heating and pre-cooling	39%	[119]
Simulation	Typical HVAC control strategy assuming maximum occupancy for ventilation and conditions all rooms from 7:00–22:00	42%	[22]
On-site	Conventional scheduled temperature set-points (cooling)	18%	[50]
On-site	Existing single-zoned thermostat case	20%	[88]
On-site	standard sensible cooling control strategy embedded in the Building Management System (BMS) after the normalization against outdoor climate and room occupancy	20%	[74]
On-site	Conventionally scheduled cooling systems	30%	[84]
On-site	Conventional scheduled temperature set-points (heating)	30%	[50]

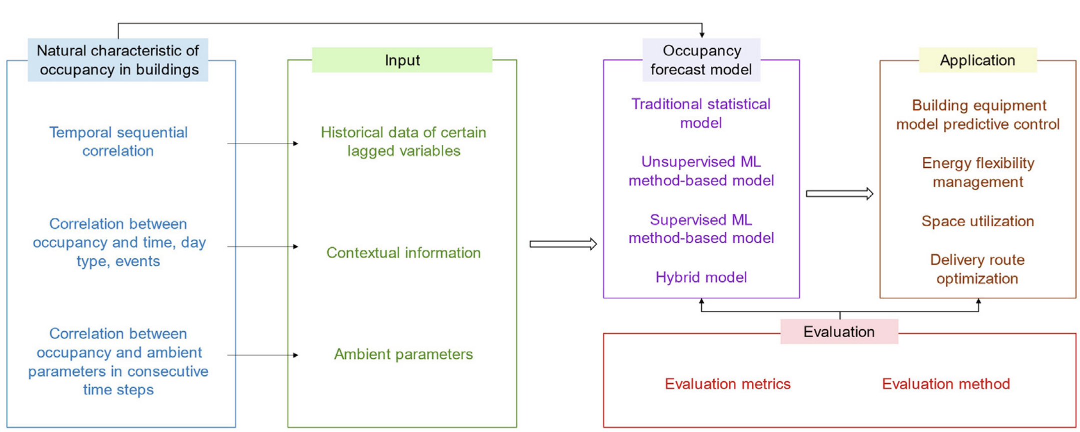
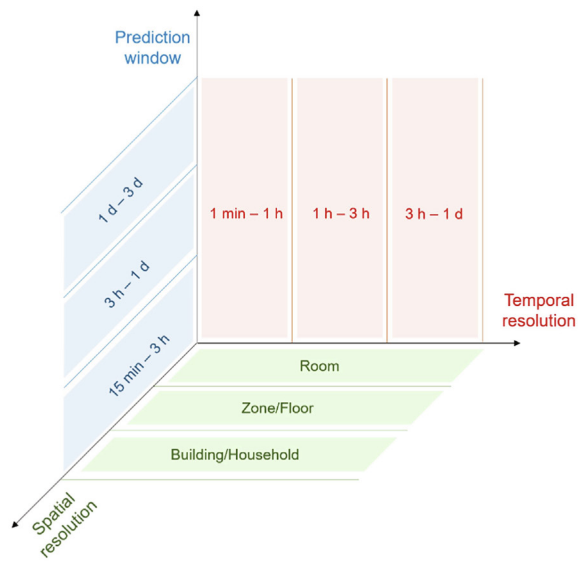
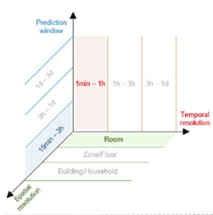
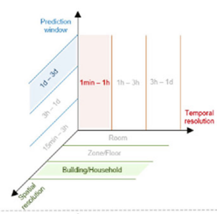
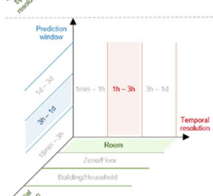
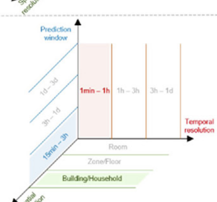
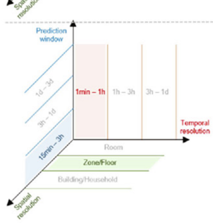
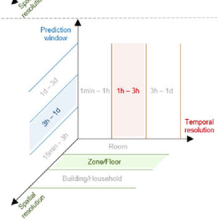
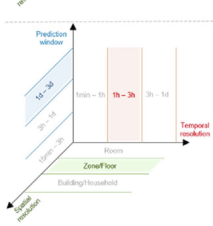


Fig. 15. Logical structure of occupancy forecast research.



(a) Structure of temporal and spatial resolutions and prediction window for occupancy forecasts

Resolution figure	Application	Resolution figure	Application
	Lighting control Cooling control		Energy flexibility management
	Heating control		Delivery route management
	Air handling unit (AHU) control Fresh air supply control		Parking
	Space utilization		

(b) Temporal and spatial resolutions and prediction window under different application scenarios

Fig. 16. Summary of application scenarios relating to temporal and spatial resolutions and prediction window.

7.1. Data acquisition and dataset establishment

Occupancy-related data are challenging to obtain in buildings because of accuracy and privacy issues. However, with the development of the Internet of Things, in addition to RFID and Wi-Fi solutions for occupancy detection, big data collection methods and analytics could potentially improve occupancy-related studies using GPS and social media data. Thus, the new challenge is how to carefully manage these data, from data cleaning to data mining. Meanwhile, occupancy information with enlarged and detailed spatial tags will help to deepen routine information and potentially improve occupancy-based research.

Additionally, according to the literature review, building occupancy datasets are relatively sparse, making it difficult to validate research results. Therefore, a shared and authentic dataset is required. One project sponsored by the American Society of Heating, Refrigerating, and Air-Conditioning Engineers collects valid global occupant behavior databases from researchers worldwide to help promote the standardization of occupancy forecast research.

7.2. Natural characteristics of occupancy

A basic physical routine analysis of occupancy is introduced in Section 1. However, deeper analyses and accurate occupancy forecasting require additional information regarding physical occupancy routines and nature analyses. Features such as periodicity, temporal sequential characteristics, and predictability would contribute to forecast studies and guide parameter selection and online model updates. Based on the reviewed modeling techniques, supervised machine learning methods are beginning to dominate the field. However, hybrid models, which could be a potential solution to combine the machine learning method with insight into the natural characteristics of occupancy routines, are relatively limited. Additionally, the diversity of occupancy patterns should be specified and distinguished under different circumstances, and sensitivity analyses may help identify the key influential factors for occupancy forecasts.

7.3. Temporal and spatial resolution

The temporal and spatial resolutions were highly correlated with the application scenario (Fig. 16). For different operation objects, the spatial resolution may be at the room, floor, or building level. Currently, research mainly focuses on predictive controls at the room scale, controlling the on/off operation state or temperature setpoint of terminals. Occupancy forecast results are important for larger spatial scales, and unifying the results for different resolutions is important.

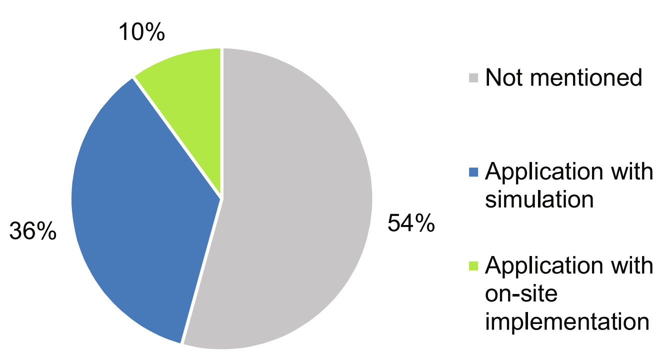


Fig. 17. Proportion of occupancy-based simulation and on-site implementation applications in reviewed articles.

Moreover, it remains a challenge to systematically propose a framework of occupancy in buildings using quantification on the temporal and spatial scales.

7.4. Implementation and on-site application

Based on the literature review results, a major application area for future occupancy predictions is to guide and optimize building operations. However, 80% of the application studies reviewed were based on simulation test beds (Fig. 17). Further, the results of the review indicate that it is necessary to improve the on-site implementation of occupancy forecast-based applications.

Meanwhile, the gaps in the knowledge corresponding to on-site implementation include: (1) designing an effective and more accurate occupancy information detection method, (2) developing a robust method for occupancy forecasting, and (3) determining valid model-updating rules for on-site applications.

Moreover, occupant behaviors with different operable appliances and systems could also be predicted in future studies. To accomplish this, feedback from occupants should be recorded and used to guide control strategy improvements. These results would provide additional information during the operation phase. The performance of predictive and feedback-based controls is promising compared with traditional control methods in terms of the ambient state and energy consumption.

8. Conclusions

Occupancy is the basis for energy demand analyses and comfort-centric operations. Occupancy prediction is a key issue in building occupant behavior studies, which have recently undergone significant improvements. This article provides a systematic and critical review of the literature concerning future occupancy predictions (occupancy forecasting). This review summarizes a complete and systematic framework for occupancy forecasting. The related articles were analyzed by application area, data acquisition, modeling techniques, and evaluation metrics. For each aspect, the detailed requirements and influential parameters are analyzed and summarized as follows:

(1) Most occupancy forecast studies were employed for operation optimization. At different spatial scales, the detailed applications may vary from HVAC terminal control to building energy management. Thus, the forecast model should capture both changes and repetitive patterns of occupancy for future predictions.
(2) Nearly one-third of the occupancy forecasting research was conducted in academic buildings, and the forecast objects mostly consisted of occupancy presence and occupant number. Based on the size of the collected datasets, the two most frequently used spatial scales were room and building (household). Furthermore, occupancy data were usually collected during a measurement period of several months.
(3) Most studies on forecasting methods are based on the input of historical data. However, several studies have considered contextual information, such as the hour of the day and the day of the week. The temporal resolution of the occupancy forecast was mostly distributed from 15 min to 1 h. Moreover, the lagged variable and the prediction window (horizon) are two key factors that must be carefully considered in this field. Previous studies have discussed the influence of lagged variables on forecast performance. Nevertheless, the impact of the prediction window remains an issue for further study. Moreover, the concept drift of the occupancy pattern is crucial for model generalizability and robustness.

Regarding the forecast algorithm, chronological evolution shows that the dominance of methods changed from traditional statistical models to supervised machine learning and hybrid methods.

(4) Because occupancy forecasts are faced with on-site control applications, evaluating the accuracy of each time step is crucial. Normally, separating the training and testing sets is required. However, to date, few studies have discussed the impact of the amount of training data on forecast performance.

Furthermore, the following challenges for occupancy forecasting must be solved:

(1) Data acquisition and dataset establishment. In addition to residential and academic buildings, the occupancy data for different types of buildings should be analyzed. However, it is necessary to design methods that can obtain occupancy data with high accuracy and without intruding on privacy.
(2) Insight of the natural characteristics of occupancy in buildings. According to a review of the occupancy forecast algorithms, the inputs of forecast methods are mainly continuous historical occupancy data and contextual information, such as hours, weekdays, and holidays. The combination of natural characteristics, such as temporal and spatial features, with machine learning methods may improve forecast accuracy. Furthermore, it is also essential to consider different event-related drivers, such as the timetables of airlines or trains for airports or railway stations, and the information of meetings and seminars for academic buildings to achieve accuracy improvements.
(3) Generalizability and robustness of model. It is important to enhance the generalizability of the model in buildings with different temporal and spatial resolutions for different application scenarios. Currently, concept drift during occupancy forecasting has not been sufficiently discussed with regard to model robustness. Thus, the realization of self-learning and updating remains a challenge for occupancy forecasting models.
(4) On-site implementation and occupant behavior combined analysis. The gap between occupancy forecasting and on-site application to the operation must be addressed by considering the relationship between occupancy information and system operation variables. Building energy conservation via operation may be greatly promoted by generalizing occupancy forecast studies and related applications.

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.

Acknowledgments

This research was supported by the National Key R&D (Research and Development) Program of China "Research and Integrated Demonstration on Suitable Technology of Net Zero Energy Building" (Grant No. 2019YFE0100300), the National Natural Science Foundation of China (Grant Number: 51778321), and the quantitative description and simulation methodology of occupant behavior in buildings. This research was also supported by the Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture.

Appendix A. Summary of sizes of occupancy datasets

Table A1
Summary of sizes of occupancy datasets.

Ref.	Temporal scale				Spatial scale			
	Day	Month	Year	Not mentioned	Room	Zone/ Floor	Building/Household	Not mentioned
[22]	5					1		
[49]			5					×
[18]			2		16			
[50]	14						1	
[24]		9			8			
[54]		2					45	
[17]	10						8	
[55]		2					5	
[44]	14				3			
[67]		1					1	
[28]		1						×
[73]		1.5					1	
[115]		5			1			
[70]				×				×
[34]			1		7			
[65]		4			1			
[59]		1				1		
[71]		2			1			
[10]		6			1			
[29]		1			3			
[112]	8				2			
[58]				×	1		2	
[68]		2					1	
[30]			1				100	
[79]				×				×
[81]		2						×

(*continued on next page*)

Table A1 (*continued*)

Ref.	Temporal scale				Spatial scale			
	Day	Month	Year	Not mentioned	Room	Zone/ Floor	Building/Household	Not mentioned
[56]		1						×
[97]	10					1		
[75]		2						×
[82]		2					4	
[83]		6			4			
[83]		4			1			
[96]			1				1	
[72]				×	2			
[26]		6			10			
[76]		4					2	
[93]				×				×
[60]		8					5	
[104]				×			4	
[74]		7			6			
[84]		7.5			11			
[94]		7					161	
[100]		1.5				18		
[106]			1.5				5000	
[78]		1.5						×
[78]		1.5						×
[80]				×				×
[25]	7				1			
[113]		1			1			
[61]				×				×
[57]		4					2	
[103]				×	2			×
[86]		3.5						×
[98]		3			2			
[87]				×	4			
[105]		2					1	
[88]				×				×
[77]	12				1			
[89]		9			8			
[90]			5		3			
[11]	2				1			
[11]	7				1			
[99]		4					1	
[99]			1				1	
[62]		9				30		
[63]				×				×
[91]			2		1			
[91]		6			1			
[101]		7					1	
[95]			1.5		1			
[102]		1.5					6	
[116]		9					1	
[119]		1			1			
[114]	15				1			
[69]			7		1			

Appendix B. Input summary of occupancy forecast model

Table B1
Input summary of historical data.

Historical occupancy state	Historical occupant number	Historical position	Clustering result from historical data	Historical forecast error	Historical probability of time and duration	Ref.
√						[18,24,26,30,34,49,54–58,62,70,71,74,76,79,81,82,84,86–88,90,94,102,104,105,113]
	√					[10,11,22,25,44,63,65,73,80,93,96–100,103,116]
		√				[59,61,75,78]
			√			[101]
					√	[89]
√	√					[83]
√			√			[91,95,115]
	√			√		[29]

Table B2
Input summary of contextual information.

Time of day	Day of week	Day of year	Clustering day type	Month of year	Holiday	Season	Schedule	Ref.
√								[17,60,72]
	√							[78,91,102,115]
							√	[28,68,80]
√	√							[29,30,116]
√	√	√						[93]
√	√		√					[77]
√	√			√	√			[106]
√	√				√			[73]
√	√					√		[18,81]
	√				√	√		[86,103]
	√				√		√	[67]
	√					√		[44,96]
	√						√	[69]

Table B3
Input summary of ambient parameters.

Temp.	Humid.	CO_2	Acoustic	Motion	Lighting	Pressure	Electricity	Wi-Fi count	Ref.
√	√	√	√	√					[72]
√	√	√		√	√				[114]
		√	√	√	√				[50]
		√					√		[25]
				√					[17,28,77]
				√		√			[112]
					√		√		[101]
					√		√	√	[99]
							√		[60,106]

References

[1] IEA, World Energy Balances 2019. 2019.
[2] S. D'Oca, T. Hong, J. Langevin, The human dimensions of energy use in buildings: a review, Renew. Sustain. Energy Rev. 81 (2018) 731–742.
[3] E. Delzendeh, S. Wu, A. Lee, Y. Zhou, The impact of occupants' behaviours on building energy analysis: a research review, Renew. Sustain. Energy Rev. 80 (2017) 1061–1071.
[4] K.-U. Ahn, C.-S. Park, Correlation between occupants and energy consumption, Energy Build. 116 (2016) 420–433.
[5] M. Jia, R.S. Srinivasan, A.A. Raheem, From occupancy to occupant behavior: an analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency, Renew. Sustain. Energy Rev. 68 (2017) 525–540.
[6] S. Hu, D. Yan, E. Azar, et al., A systematic review of occupant behavior in building energy policy, Build. Environ. 175 (2020).
[7] D.a. Yan, T. Hong, B. Dong, A. Mahdavi, S. D'Oca, I. Gaetani, X. Feng, IEA EBC Annex 66: definition and simulation of occupant behavior in buildings, Energy Build. 156 (2017) 258–270.
[8] W. O'Brien, A. Wagner, M. Schweiker, et al. Introducing IEA EBC Annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation. 2020: p. 106738.
[9] J.Y. Park, M.M. Ouf, B. Gunay, et al., A critical review of field implementations of occupant-centric building controls, Build. Environ. 165 (2019) 106351.
[10] J. Dong, C. Winstead, J. Nutaro, et al., Occupancy-based HVAC control with short-term occupancy prediction algorithms for energy-efficient buildings, Energies 11 (9) (2018).
[11] W. Wang, J. Chen, X. Song, Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference approach, Build. Environ. 124 (2017) 130–142.
[12] T. Osaragi, T. Hoshino, Predicting Spatiotemporal Distribution of Transient Occupants in Urban Areas, in: J. Gensel, D. Josselin, D. Vandenbroucke (Eds.), Bridging the Geographic Information Sciences, Springer, Berlin, Heidelberg, 2012, pp. 307–325.
[13] V.M. Barthelmes, R. Li, R.K. Andersen, W. Bahnfleth, S.P. Corgnati, C. Rode, Profiling occupant behaviour in Danish dwellings using time use survey data, Energy Build. 177 (2018) 329–340.
[14] Y. Wu, Y. Li, C. Wang, et al., Hourly occupant density prediction in commercial buildings for urban energy simulation, IOP Conference Series: Earth and Environmental Science, 2019.
[15] V.L. Erickson, Y. Lin, A. Kamthe, et al., Energy efficient building environment control strategies using real-time occupancy measurements, Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, 2009.
[16] W. O'Brien, F. Tahmasebi, R.K. Andersen et al. An international review of occupant-related aspects of building energy codes and standards. 2020: p. 106906.
[17] J. Lu, T. Sookoor, V. Srinivasan, et al., The smart thermostat: using occupancy sensors to save energy in homes, Proceedings of the 8th ACM conference on embedded networked sensor systems, 2010.
[18] S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408.
[19] Z. Wang, Y. Ding, An occupant-based energy consumption prediction model for office equipment, Energy Build. 109 (2015) 12–22.
[20] A. Muroni, I. Gaetani, P.-J. Hoes, J.L.M. Hensen, Occupant behavior in identical residential buildings: a case study for occupancy profiles extraction and application to building performance simulation, Build. Simul. 12 (6) (2019) 1047–1061.
[21] E. Cuerda, O. Guerra-Santin, J.J. Sendra, F.J. Neila González, Comparing the impact of presence patterns on energy demand in residential buildings using measured data and simulation models, Build. Simul. 12 (6) (2019) 985–998.
[22] V.L. Erickson, M.A. Carreira-Perpinan, and A.E. Cerpa. OBSERVE: Occupancy-based system for efficient reduction of HVAC energy. in: Proceedings 2011 10th International Conference on Information Processing in Sensor Networks. 2011.
[23] V.L. Erickson, M.A. Carreira-Perpinan, A.E. Cerpa, Occupancy modeling and prediction for building energy management, ACM Trans. Sens. Netw. 10 (3) (2014).
[24] Ardeshir Mahdavi, Farhang Tahmasebi, Predicting people's presence in buildings: an empirically based model performance analysis, Energy Build. 86 (2015) 349–355.
[25] Seung Ho Ryu, Hyeun Jun Moon, Development of an occupancy prediction model using indoor environmental data based on machine learning techniques, Build. Environ. 107 (2016) 1–9.
[26] C. Manna, D. Fay, K.N. Brown, et al., Learning occupancy in single person offices with mixtures of multi-lag markov chains, Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2013.
[27] Vincent Tabak, Bauke de Vries, Methods for the prediction of intermediate activities by office occupants, Build. Environ. 45 (6) (2010) 1366–1372.
[28] A. Alrazgan, A. Nagarajan, A. Brodsky, et al., Learning occupancy prediction models with decision-guidance query language, Proceedings of the Annual Hawaii International Conference on System Sciences, 2011.
[29] C.R. Garaza, P. Hespanhol, Y. Mintz, et al., Impact of Occupancy Modeling and Horizon Length on HVAC Controller Efficiency, in: 2018 European Control Conference, ECC, 2018, p. 2018.
[30] B. Huchuk, S. Sanner, W. O'Brien, Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data, Build. Environ. 160 (2019) 106177.

[31] H. Hou, J. Pawlak, A. Sivakumar, et al., An approach for building occupancy modelling considering the urban context, Build. Environ. 183 (2020).
[32] C. Song, Z. Qu, N. Blumm, A.-L. Barabasi, Limits of predictability in human mobility, Science 327 (5968) (2010) 1018–1021.
[33] Ki-Uhn Ahn, Deuk-Woo Kim, Cheol-Soo Park, Pieter de Wilde, Predictability of occupant presence and performance gap in building energy simulation, Appl. Energy 208 (2017) 1639–1652.
[34] H.B. Gunay, W. O'Brien, I. Beausoleil-Morrison, Development of an occupancy learning algorithm for terminal heating and cooling units, Build. Environ. 93 (P2) (2015) 71–85.
[35] S. Gopika, Agent based HVAC optimization model for building energy efficiency. in: Proceedings of 2015 IEEE 9th International Conference on Intelligent Systems and Control, ISCO 2015. 2015.
[36] Bing Dong, Da Yan, Zhaoxuan Li, Yuan Jin, Xiaohang Feng, Hannah Fontenot, Modeling occupancy and behavior for better building design and operation—a critical review, Build. Simul. 11 (5) (2018) 899–921.
[37] Xiaohang Feng, Da Yan, Tianzhen Hong, Simulation of occupancy in buildings, Energy Build. 87 (2015) 348–359.
[38] Amin Mirakhorli, Bing Dong, Occupancy behavior based model predictive control for building indoor climate—a critical review, Energy Build. 129 (2016) 499–513.
[39] Sylvia T. Kouyoumdjieva, Peter Danielis, Gunnar Karlsson, Survey of non-image-based approaches for counting people, IEEE Commun. Surv. Tutorials 22 (2) (2020) 1305–1336.
[40] K. Ngamakeur, S. Yongchareon, J. Yu, et al., A survey on device-free indoor localization and tracking in the multi-resident environment, ACM Comput. Surv. 53 (4) (2020).
[41] L. Rueda, K. Agbossou, A. Cardenas, et al., A comprehensive review of approaches to building occupancy detection, Build. Environ. 180 (2020).
[42] K. Sun, Q. Zhao, J. Zou, A review of building occupancy measurement systems, Energy Build. 216 (2020).
[43] X. Dai, J. Liu, X. Zhang, A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings, Energy Build. 223 (2020).
[44] Anna A. Adamopoulou, Athanasios M. Tryferidis, Dimitrios K. Tzovaras, A context-aware method for building occupancy prediction, Energy Build. 110 (2016) 229–244.
[45] Y. Ma, G. Anderson, F. Borrelli, A distributed predictive control approach to building temperature regulation, 2011 American Control Conference, 2011.
[46] Sama Aghniaey, Thomas M Lawrence, Javad Mohammadpour, WenZhan Song, Richard T Watson, Marie C Boudreau, Optimizing thermal comfort considerations with electrical demand response program implementation, Build. Serv. Eng. Res. Technol. 39 (2) (2018) 219–231.
[47] V.L. Erickson, A.E. Cerpa, Occupancy based demand response HVAC control strategy, Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, 2010.
[48] V.L. Erickson, S. Achleitner, and A.E. Cerpa. POEM: Power-efficient occupancy-based energy management system. in: IPSN 2013 - Proceedings of the 12th International Conference on Information Processing in Sensor Networks, Part of CPSWeek 2013. 2013.
[49] Frauke Oldewurtel, David Sturzenegger, Manfred Morari, Importance of occupancy information for building climate control, Appl. Energy 101 (2013) 521–532.
[50] Bing Dong, Khee Poh 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. Simul. 7 (1) (2014) 89–106.
[51] B. Dong, B. Andrews, Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings, Proceedings of building simulation, 2009.
[52] Bing Dong, Khee Poh Lam, Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network, J. Build. Perform. Simul. 4 (4) (2011) 359–369.
[53] S. Hu, D. Yan, S. Guo, et al., A survey on energy consumption and energy usage behavior of households and residential building in urban China. 2017. 148: p. 366-378.
[54] Wilhelm Kleiminger, Friedemann Mattern, Silvia Santini, Predicting household occupancy for smart heating control: a comparative performance analysis of state-of-the-art approaches, Energy Build. 85 (2014) 493–505.
[55] J. Scott, A. Bernheim Brush, J. Krumm, et al., PreHeat: controlling home heating using occupancy prediction, Proceedings of the 13th international conference on Ubiquitous computing, 2011.
[56] M. Killian, M. Kozek, Short-term occupancy prediction and occupancy based constraints for MPC of smart homes, IFAC-PapersOnLine 52 (4) (2019) 377–382.
[57] F.C. Sangogboye, M.B. Kjærgaard, PROMT: predicting occupancy presence in multiple resolution with time-shift agnostic classification, Comput. Sci. – Res. Dev. 33 (1–2) (2018) 105–115.
[58] H. Huang, H. Xu, Y. Cai, et al., Distributed machine learning on smart-gateway network toward real-time smart-grid energy management with behavior cognition, ACM Trans. Des. Autom. Electron. Syst. 23 (5) (2018).
[59] A. Das, M.B. Kjærgaard, Precept: Occupancy presence prediction inside a commercial building, Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, 2019.
[60] S. Ohsugi, N. Koshizuka, Delivery route optimization through occupancy prediction from electricity usage, Proceedings - International Computer Software and Applications Conference, 2018.
[61] S.K. Sama, M. Rahnamay-Naeini, A study on compression-based sequential prediction methods for occupancy prediction in smart homes, 2016 IEEE 7th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2016, 2016.
[62] Ghulam Ali, Tariq Ali, Muhammad Irfan, Umar Draz, Muhammad Sohail, Adam Glowacz, Maciej Sulowicz, Ryszard Mielnik, Zaid Bin Faheem, Claudia Martis, IoT based smart parking system using deep long short memory network, Electronics 9 (10) (2020) 1696, https://doi.org/10.3390/electronics9101696.
[63] Y. Atif, S. Kharrazi, D. Jianguo, et al., Internet of Things data analytics for parking availability prediction and guidance, Trans. Emerg. Telecommun. Technol. 31 (5) (2020).
[64] Isabella Gaetani, Pieter-Jan Hoes, Jan L.M. Hensen, Occupant behavior in building energy simulation: towards a fit-for-purpose modeling strategy, Energy Build. 121 (2016) 188–204.
[65] Z. Chen, Y.C. Soh, Comparing occupancy models and data mining approaches for regular occupancy prediction in commercial buildings, J. Build. Perform. Simul. 10 (5–6) (2017) 545–553.
[66] N. Haidar, N. Tamani, F. Nienaber, et al., Data collection period and sensor selection method for smart building occupancy prediction, IEEE Vehicular Technology Conference, 2019.
[67] Muhammad Aftab, Chien Chen, Chi-Kin Chau, Talal Rahwan, Automatic HVAC control with real-time occupancy recognition and simulation-guided model predictive control in low-cost embedded system, Energy Build. 154 (2017) 141–156.
[68] W. Huang, Y. Lin, B. Lin, et al., Modeling and predicting the occupancy in a China hub airport terminal using Wi-Fi data, Energy Build. 203 (2019).
[69] J. Schiele, T. Koperna, J.O. Brunner, Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks, Nav. Res. Logist. (2020).
[70] A. Beltran, A.E. Cerpa, Optimal HVAC building control with occupancy prediction, BuildSys 2014 - Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, 2014.
[71] Justin R. Dobbs, Brandon M. Hencey, Model predictive HVAC control with online occupancy model, Energy Build. 82 (2014) 675–684.
[72] S. Mamidi, Y.H. Chang, R. Maheswaran, Adaptive learning agents for sustainable building energy management, AAAI Spring Symposium - Technical Report, 2012.
[73] A. Ashouri, G.R. Newsham, Z. Shi, et al., Day-ahead Prediction of Building Occupancy using WiFi Signals, IEEE International Conference on Automation Science and Engineering, 2019.
[74] Yuzhen Peng, Adam Rysanek, Zoltán Nagy, Arno Schlüter, Occupancy learning-based demand-driven cooling control for office spaces, Build. Environ. 122 (2017) 145–160.
[75] Seungwoo Lee, Yohan Chon, Yunjong Kim, Rhan Ha, Hojung Cha, Occupancy prediction algorithms for thermostat control systems using mobile devices, IEEE Trans. Smart Grid 4 (3) (2013) 1332–1340.
[76] Amel Nacer, Bruno Marhic, Laurent Delahoche, Jean-baptiste Masson, ALOS: automatic learning of an occupancy schedule based on a new prediction model for a smart heating management system, Build. Environ. 142 (2018) 484–501.
[77] M Soudari, V Kaparin, S Srinivasan, S Seshadhri, Ü Kotta, Predictive smart thermostat controller for heating, ventilation, and air-conditioning systems, Proc. Est. Acad. Sci. 67 (3) (2018) 291, https://doi.org/10.3176/proc.2018.3.11.
[78] C. Ryan, K.N. Brown, Occupant location prediction using association rule mining, CEUR Workshop Proceedings, 2012.
[79] J. Hyung-Chul, L. Jaehee, J. Sung-Kwan, Scheduling of air-conditioner using occupancy prediction in a smart home/building environment, in: 2014 IEEE International Conference on Consumer Electronics, 2014, pp. 298–299.
[80] Tim Ryan, Jeffrey S. Vipperman, Incorporation of scheduling and adaptive historical data in the Sensor-Utility-Network method for occupancy estimation, Energy Build. 61 (2013) 88–92.
[81] K. Imamovic, F.C. Sangogboye, M.B. Kjærgaard, Poster abstract: improving occupancy presence prediction via multi-label classification, BuildSys 2015 - Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built, 2015.
[82] Zhaoxuan Li, Bing Dong, A new modeling approach for short-term prediction of occupancy in residential buildings, Build. Environ. 121 (2017) 277–290.
[83] Zhaoxuan Li, Bing Dong, Short term predictions of occupancy in commercial buildings-performance analysis for stochastic models and machine learning approaches, Energy Build. 158 (2018) 268–281.
[84] Yuzhen Peng, Adam Rysanek, Zoltán Nagy, Arno Schlüter, Using machine learning techniques for occupancy-prediction-based cooling control in office buildings, Appl. Energy 211 (2018) 1343–1358.
[85] F.C. Sangoboye, M.B. Kjærgaard, Poster abstract: Predicting occupancy presence in multiple resolutions for commercial buildings, Proceedings of the 3rd ACM Conference on Systems for Energy-Efficient Built Environments, BuildSys 2016, 2016.
[86] F.C. Sangogboye, K. Imamovic, M.B. Kjærgaard, Improving occupancy presence prediction via multi-label classification, 2016 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom Workshops, 2016.

[87] U. Saralegui, M. Angel Anton, O. Arbelaitz, et al., Smart meeting room usage information and prediction by modelling occupancy profiles, Sensors 19 (2) (2019).
[88] T. Sookoor, B. Holben, K. Whitehouse, Feasibility of retrofitting centralized HVAC systems for room-level zoning, 2012 International Green Computing Conference IGCC, 2012, 2012.
[89] Farhang Tahmasebi, Ardeshir Mahdavi, Stochastic models of occupants' presence in the context building systems control, Adv. Build. Energy Res. 10 (1) (2016) 1–9.
[90] F.I. Vazquez, W. Kastner, Clustering methods for occupancy prediction in smart home control, in: 2011 IEEE 20th International Symposium on Industrial Electronics, 2011, pp. 1321–1328.
[91] Y. De Bock, A. Auquilla, A. Nowé, et al., Nonparametric user activity modelling and prediction, User Model. User-Adap. Inter. (2020).
[92] Y. Jin, D. Yan, X. Zhang, et al., A data-driven model predictive control for lighting system based on historical occupancy in an office building: methodology development, Build. Simul. (2020).
[93] S. Naylor, M. Gillott, and G. Herries. The development of occupancy monitoring for removing uncertainty within building energy management systems. in: 2017 International Conference on Localization and GNSS. 2017 of Conference.
[94] S. Pesic, M. Tosic, O. Ikovic, et al., BLEMAT: data analytics and machine learning for smart building occupancy detection and prediction, Int. J. Artif. Intell. Tools 28 (6) (2019).
[95] S. Salimi, A. Hammad, Sensitivity analysis of probabilistic occupancy prediction model using big data, Build. Environ. 172 (2020).
[96] Xin Liang, Tianzhen Hong, Geoffrey Qiping Shen, Occupancy data analytics and prediction: a case study, Build. Environ. 102 (2016) 179–192.
[97] Seonghyeon Kim, Seokwoo Kang, Kwang Ryel Ryu, Giltae Song, Real-time occupancy prediction in a large exhibition hall using deep learning approach, Energy Build. 199 (2019) 216–222.
[98] F.C. Sangogboye, M.B. Kjærgaard, Poster abstract: occupancy count prediction for model predictive control in buildings, BuildSys 2017 - Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments, 2017.
[99] Zhe Wang, Tianzhen Hong, Mary Ann Piette, Data fusion in predicting internal heat gains for office buildings through a deep learning approach, Appl. Energy 240 (2019) 386–398.
[100] B. Qolomany, A. Al-Fuqaha, D. Benhaddou, et al. Role of Deep LSTM Neural Networks and Wi-Fi Networks in Support of Occupancy Prediction in Smart Buildings. in: Proceedings - 2017 IEEE 19th Intl Conference on High Performance Computing and Communications, HPCC 2017, 2017 IEEE 15th Intl Conference on Smart City, SmartCity 2017 and 2017 IEEE 3rd Intl Conference on Data Science and Systems, DSS 2017. 2018.
[101] B.W. Hobson, H.B. Gunay, A. Ashouri, et al., Clustering and motif identification for occupancy-centric control of an air handling unit, Energy Build. 223 (2020).
[102] C. Turley, M. Jacoby, G. Pavlak, et al., Development and evaluation of occupancy-aware HVAC control for residential building energy efficiency and occupant comfort, Energies 13 (20) (2020).
[103] Fisayo Caleb Sangogboye, Krzysztof Arendt, Ashok Singh, Christian T. Veje, Mikkel Baun Kjærgaard, Bo Nørregaard Jørgensen, Performance comparison of occupancy count estimation and prediction with common versus dedicated sensors for building model predictive control, Build. Simul. 10 (6) (2017) 829–843.
[104] Theis Heidmann Pedersen, Steffen Petersen, Investigating the performance of scenario-based model predictive control of space heating in residential buildings, J. Build. Perform. Simul. 11 (4) (2018) 485–498.
[105] Jie Shi, Nanpeng Yu, Weixin Yao, Energy efficient building HVAC control algorithm with real-time occupancy prediction, Energy Procedia 111 (2017) 267–276.
[106] Rouzbeh Razavi, Amin Gharipour, Martin Fleury, Ikpe Justice Akpan, Occupancy detection of residential buildings using smart meter data: a large-scale study, Energy Build. 183 (2019) 195–208.
[107] I.B. Arief-Ang, M. Hamilton, F.D. Salim, A scalable room occupancy prediction with transferable time series decomposition of CO2 sensor data, ACM Trans. Sens. Netw. 14 (3–4) (2018).
[108] Irvan B. Arief-Ang, Margaret Hamilton, Flora D. Salim, RUP: large room utilisation prediction with carbon dioxide sensor, Pervasive Mob. Comput. 46 (2018) 49–72.
[109] U. Habib, G. Zucker, Automatic occupancy prediction using unsupervised learning in buildings data, IEEE International Symposium on Industrial Electronics, 2017.
[110] Yixuan Wei, Liang Xia, Song Pan, Jinshun Wu, Xingxing Zhang, Mengjie Han, Weiya Zhang, Jingchao Xie, Qingping Li, Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks, Appl. Energy 240 (2019) 276–294.
[111] Wei Wang, Jiayu Chen, Tianzhen Hong, Na Zhu, Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology, Build. Environ. 138 (2018) 160–170.
[112] J.L. Gomez Ortega, H. Liangxiu, and N. Bowring, A Novel Dynamic Hidden Semi-Markov Model (D-HSMM) for Occupancy Pattern Detection from Sensor Data Stream. 2016 8th IFIP International Conference on New Technologies, Mobility and Security. 2016. 5 pp.-5 pp.
[113] Shide Salimi, Zheng Liu, Amin Hammad, Occupancy prediction model for open-plan offices using real-time location system and inhomogeneous Markov chain, Build. Environ. 152 (2019) 1–16.
[114] E. Hitimana, G. Bajpai, R. Musabe, et al., Implementation of IoT framework with data analysis using deep learning methods for occupancy prediction in a building, Future Internet 13 (3) (2021).
[115] A. Auquilla, Y. De Bock, A. Nowe, et al., Combining occupancy user profiles in a multi-user environment: An academic office case study, Proceedings - 12th International Conference on Intelligent Environments, IE 2016, 2016.
[116] Y. Yuan, K.S. Liu, S. Munir, et al., Leveraging fine-grained occupancy estimation patterns for effective HVAC control, Proceedings - 5th ACM/IEEE Conference on Internet of Things Design and Implementation, IoTDI 2020, 2020.
[117] Gerhard Widmer, Miroslav Kubat, Learning in the presence of concept drift and hidden contexts, Machine Learning 23 (1) (1996) 69–101.
[118] C. Fan, D. Yan, F. Xiao, et al., Advanced data analytics for enhancing building performances: from data-driven to big data-driven approaches, Build. Simul. (2020) 1–22.
[119] N. Haidar, N. Tamani, Y. Ghamri-Doudane, et al. Occupant Behavior Prediction and Real-Time Correction-based Smart Building Energy Optimization. in: 2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings. 2020.
[120] R. Wang, S. Lu, W. Feng, A novel improved model for building energy consumption prediction based on model integration, Appl. Energy 262 (2020) 114561.