BUILD SIMUL (2022) 15: 333–356
https://doi.org/10.1007/s12273-021-0813-8

Review on occupancy detection and prediction in building simulation

Yan Ding[1], Shuxue Han[1], Zhe Tian[1] (✉), Jian Yao[2], Wanyue Chen[1], Qiang Zhang[1]

1. School of Environmental Science and Engineering, Tianjin Key Laboratory of Built Environment and Energy Application, Tianjin University, Tianjin, 300072, China

2. Department of Architecture, Ningbo University, Ningbo, China

Review Article

Architecture and Human Behavior

Abstract

Energy simulation results for buildings have significantly deviated from actual consumption because of the uncertainty and randomness of occupant behavior. Such differences are mainly caused by the inaccurate estimation of occupancy in buildings. Therefore, the error between reality and prediction could be largely reduced by improving the accuracy level of occupancy prediction. Although various studies on occupancy have been conducted, there are still many differences in the approaches to detection, prediction, and validation. Reports published within this domain are reviewed in this article to discover the advantages and limitations of previous studies, and gaps in the research are identified for future investigation. Six methods of monitoring and their combinations are analyzed to provide effective guidance in choosing and applying a method. The advantages of deterministic schedules, stochastic schedules, and machine-learning methods for occupancy prediction are summarized and discussed to improve prediction accuracy in future work. Moreover, three applications of occupancy models—improving building simulation software, facilitating building operation control, and managing building energy use—are examined. This review provides theoretical guidance for building design and makes contributions to building energy conservation and thermal comfort through the implementation of intelligent control strategies based on occupancy monitoring and prediction.

Keywords

occupancy rate monitoring;
occupancy model;
occupancy prediction;
occupancy verification

Article History

Received: 10 January 2021
Revised: 9 May 2021
Accepted: 13 May 2021

1 Introduction

Addressing climate change is a global challenge and requires responses from all sectors and on all aspects, including in terms of building energy use, which is an important contributor to the increase in CO_2 emissions (Guo et al. 2020). The building construction consumed 36% of the global final energy and nearly 40% of total CO_2 emissions in 2018. Direct and indirect emissions from building electricity and commercial heat have risen to 10 $GtCO_2$ in 2019, which have recorded the highest level. (Sun et al. 2020). In addition, the energy consumption of buildings in the world is still growing rapidly, and reducing building energy consumption is critical to achieving the necessary reduction in overall energy use and carbon emissions. However, there are still many buildings that have not achieved the necessary energy-saving effect in the actual operation stage (Cross et al. 2017), and the design performance of buildings is significantly different from their actual performance. Energy consumption prediction during the design stage is realized by building energy simulation software, but it is not completely reliable. The energy-saving during operation is significantly different from the expected savings (Ahmad and Culp 2006). Prediction of energy consumption made in the simulation stage may be less than one-fifth of the reality (Menezes et al. 2012). There is no doubt that narrowing the gap between prediction and reality can better improve the energy supply and demand management system in buildings (Zou et al. 2018). The six most critical parameters that influence energy consumption of buildings were indicated in the IEA annex 53: (1) weather, (2) envelope, (3) building energy and service systems, (4) indoor design standards, (5) building operation and maintenance, and (6) occupant behavior (Yoshino et al. 2017). Therefore, inaccurate estimation of the occupancy

E-mail: tianzhe@tju.edu.cn

in a building could cause a large disparity between the prediction and operation results. The importance of occupant presence information is beginning to attract the attention of researchers.

The occupancy has an important significance in energy saving for buildings. Occupant behavior and lifestyle choices are closely related to energy usage (Oikonomou et al. 2009). Shen et al. (2017) argued that the key to improving the operating efficiency of heating, ventilation, and air conditioning (HVAC) systems is to provide heating and ventilation according to the number of controlled objects in areas that need to be controlled. The beginning and end of the ventilation system control usually depend on the time when occupants enter and leave a room because the number of occupants determines the flow volume of fresh air. Moreover, control of HVAC and lighting systems based on occupancy has shown great energy-saving potential. Based on using Wi-Fi and smartphone systems to obtain occupancy data in each room, Balaji et al. (2013) carried out a small-scale experiment in a university building and established a sensor network to predict the number of occupants. The results showed that the proposed control strategy for the HVAC system control strategy can save electricity by 17.8%. Brooks et al. (2014) developed a measured occupancy-based setback (MOBS) controller. Compared with the previous model predictive control (MPC)-based HVAC control, the implementation scale of the MOBS controller is not limited to specific buildings or areas, and its implementation scale can be extended to multiple areas of office buildings. The establishment of an occupancy model can not only provide a load basis for building design and HVAC-system-related equipment selection, but it also provides theoretical guidance for humanized management of building systems and optimization of HVAC and lighting system control.

At present, research on occupancy is still inadequate. In the actual building design and system control application, a fixed schedule was mainly used, resulting in a huge gap between prediction and operation. Later, a stochastic model and machine learning model were developed to simulate the building energy consumption better. Although there are various models, the lack of uniform evaluation criteria makes the model selection difficult. First, the method of comparing prediction accuracy is quite different. Chang and Hong (2013) used a logistic regression method to analyze five types of occupancy in a room. The logistic regression model is simple to construct, but the prediction accuracy is not known. D'Oca and Hong (2015) applied a three-step data-mining framework to explore the occupancy in office buildings. The construction of the model is complex, and the parameter acquisition depends on a long period of data monitoring. Despite this fact, compared with the measured data, the prediction accuracy can reach more than 90%. Yang and Becerik-Gerber (2014) used four models to simulate occupancy, of which the time series models had the highest accuracy (97.3%). However, the predicted results were compared with a fixed schedule rather than measured data. In addition, the occupancy model lacks universality. In another study (Hailemariam et al. 2011), although the simulation results showed that the prediction accuracy could reach 98.4%, the model was limited to a small range of office areas and could not be applied to the entire building level. The prediction effect would be significantly reduced if the model were extended to other types of buildings.

In the past decade, many studies focused on sensor technology for occupant detection. However, the monitoring method does not have a uniform guided selection mechanism. In addition, the lack of clear standards for different types of building sensor combinations may lead to unsatisfactory data collection with higher experimental installation costs. Passive infrared (PIR) has been widely used because of its low cost and easy of operation. However, in a previous study (Wang et al. 2005), the occupancy model could not be established because the delay of PIR leads to inaccurate estimation. Even if the modeling methods are consistent, the data collection methods vary. In other works (Page et al. 2008; Salimi et al. 2019), an inhomogeneous Markov chain was used for modeling. The former study used PIR to collect data, and the accuracy was not reported. The latter study used Bluetooth and smartphones to collect data with an accuracy of 84%. In recent studies (Khan et al. 2021), a new occupancy sensing technique whereby a chair-based temperature sensor array was introduced, which has high accuracy in occupant detection. However, it was just well suited for office rooms with seating-related tasks. In another work (Mikkilineni et al. 2019), researchers suggested a novel occupancy sensing method that was based on long-wave infra-red focal-plane arrays. It can be coupled with radio frequency and ultrasonic-based radar to improve accuracy. There are many combinations of sensors in the same type of building. In two studies (Hailemariam et al. 2011; Wang et al. 2017), data were collected in office buildings. The former used a combination of Wi-Fi and cameras, while the latter used a combination of PIR and environmental sensors. The researchers lack consistency in the choice of monitoring methods. Therefore, the current review papers still have some limitations: (1) There is no unified selection mechanism for monitoring methods in the current review paper, which cannot provide clear sensor combination selection for reader's data collection. (2) There have appeared some new sensor technologies and sensor combination schemes in recent years, but the current review papers about the sensor types for data collection were not comprehensive and novel enough. (3) The descriptions about the specific classification, prediction accuracy, and practical application of the occupancy model in the current review paper were less, and the connections between data collection, occupancy modeling, and model application have not been comprehensively analyzed.

At present, there are many research articles, but there still lacks of a systematic summary of occupancy research about data collection and model application. Meanwhile, the classification of models is not detailed enough and the description of monitoring tools needs to be improved. In view of the limitation of the above research, the latest sensor types will be introduced in this review. The use of different sensor combinations will be defined after classification, and the prediction accuracy of different occupancy models will be compared. Combined with the practical application, the model selection methods will be proposed. In terms of monitoring technology, PIR, radio frequency (RF) signal technology, cameras, environmental sensors, Wi-Fi, and Bluetooth are analyzed. The monitoring technologies are classified, and each type of sensor is given a corresponding application scope and field. Then, the advantages and disadvantages are analyzed, and two commonly used combinations of sensors are summarized. The theoretical guidance for the selection of monitoring methods is provided. In terms of modeling methods, the deterministic schedule, stochastic model, and machine-learning model are summarized, and the establishment and prediction accuracies of the models are compared. In addition, three aspects of the application of the occupancy model are highlighted: writing in energy consumption simulation software or making it a functional model unit, HVAC and lighting control, and a basis of load and energy consumption prediction model.

In this article, occupancy detection refers to the placement of physical tools inside the building to obtain the actual number of occupants and the movement of people in different time periods, which is the basis of the occupancy prediction model. Prediction refers to the use of physical tools or questionnaires and other ways to obtain the real changes of occupant number in the building, then establish an occupancy model after analyzing the variation rule of occupancy. There is no definite relationship between the two. However, the occupancy detection will provide the data basis for prediction.

This article is organized as follows. In Section 1, the background and significance of this review are discussed. In Section 2, the research trends of the related literature are surveyed and analyzed. In Section 3, the data collection methods for occupancy are summarized. Section 4 provides a comprehensive review of the current occupancy models. In Section 5, the application of the model is analyzed. Sensors, models, and simulations are summarized and discussed in Section 6, and the problems and future research directions in the occupancy modeling process are described. Finally, Section 7 summarizes this study. The structure of the review is shown in Figure 1.

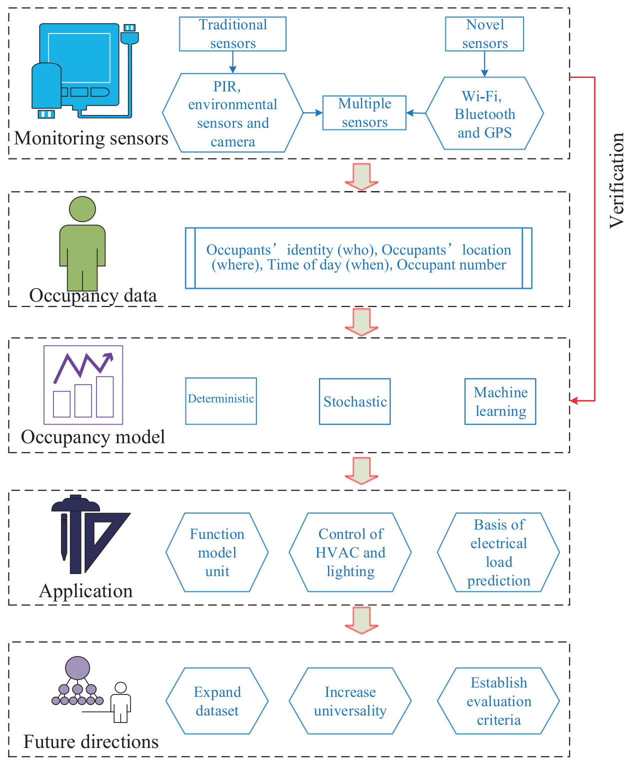

Fig. 1 Review framework of building occupancy

2 Publication collection and summary

The aim of this review is to identify current important topics and predict the future research trends of occupancy-related fields. Detailed knowledge networks can be established by comparing published literature using bibliometrics (Daim et al. 2006). Therefore, bibliometrics was applied to the study of building occupancy issues by using Scopus as a publication search engine. The keywords for searching were "occupancy" and "building" and "energy". A total of 2142 reports until March 2021 were collected. Although the keywords were included in all these articles, some were irrelevant to the purpose of this review. By deleting 373 articles, 1799 papers were finally included in this review.

As shown in Figure 2, the number of publications on building occupancy has increased from 1981 to 2021 and can be divided into two stages. The first stage was from 1981 to 2003. The number of publications was fewer than 20, which means that few studies paid attention to the occupant behavior in buildings during that time period. The second stage began in 2004, and the number of publications grew rapidly. Since then, the key role of occupants

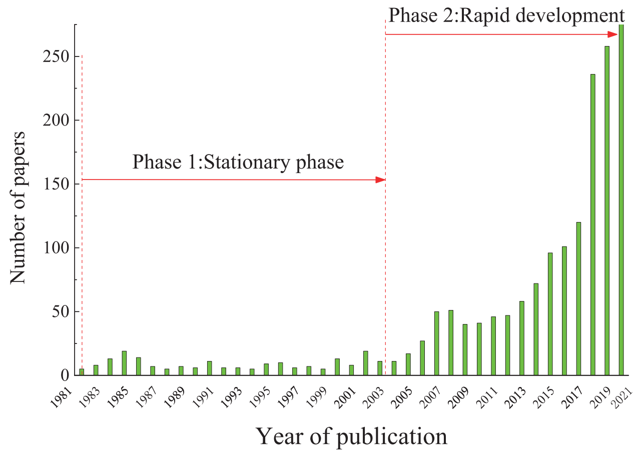

Fig. 2 Annual number of publications from 1981 through 2021

in building energy consumption became universally valued and increasingly investigated.

Through keyword mining of publication titles and abstracts, this review used clustering technology to identify topics that are currently being widely researched. The specific steps were as follows: (1) extract keywords from the title and abstract of the publication and only collect words that appear more than 10 times and (2) use VOSviewer, a free bibliometric software (Park et al. 2019; Park and Nagy 2018), to cluster the keywords that appear at the same time.

As shown in Figure 3, keywords are distinguished by different solid circles with a specific size and color. The larger the circle, the more frequently the keyword appears in the literature. The circles are connected to each other by a curve. The closer the two circles, the more frequently two keywords are connected. According to the simultaneous occurrence of keywords, four categories are obtained, as listed in Table 1, which are represented by four colors. For example, the red category represents words that appear simultaneously with "energy performance."

Through clustering, four research topics related to building occupancy can be obtained (as shown in Table 1). Red indicates that the keyword is related to building energy consumption simulation. It focuses on the application of the occupancy model and whether the simulation accuracy

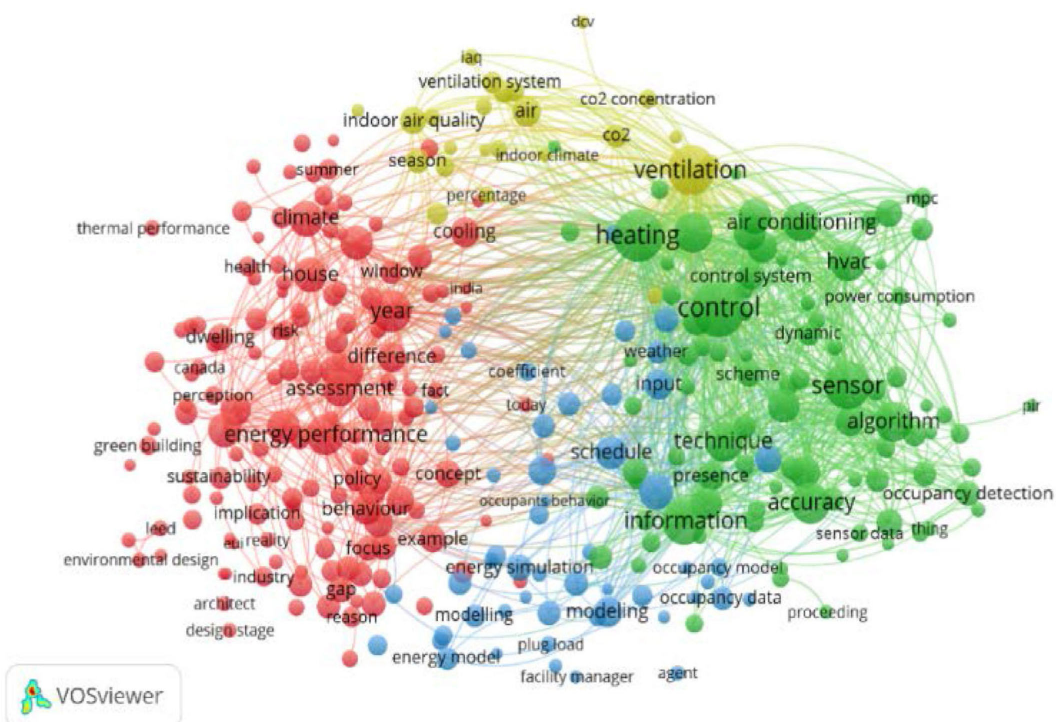

Fig. 3 Clustering results of occupancy related keywords

Table 1 Keyword clustering results of publications through 2021

Cluster color	Research topic	Observed keyword	Number of keywords
Red	Building simulation	Energy performance	120
Green	Building control	Control, HVAC, sensor	163
Blue	Occupancy model	Modeling, simulation	68
Yellow	Ventilation	Ventilation, indoor air quality	91

of building energy consumption is improved. Green indicates occupancy detection and HVAC and lighting system control based on the occupancy model. Blue indicates occupancy data acquisition and model establishment. It focuses on the process of building the occupancy model, including the generally used methods of building the model and the predicted results of the occupancy. Yellow indicates occupancy-based ventilation control to improve indoor air quality. Therefore, in this review, the application of building occupancy models is discussed based on the aforementioned research topics.

3 Occupancy monitoring method

Occupancy data are mainly obtained through two methods: one is the interviewee's record, questionnaire, or interview, which does not require a large amount of investment. The presence and activity status of occupants can be reflected through several questions. It is a relatively simple and basic data collection method, but the accuracy relies heavily on the memory of the surveyed occupant. The other relies on installing monitoring tools in buildings. This method requires the installation of sensors and meters, as well as the participation of professionals in the data processing. Although this objective monitoring method can provide more detailed and high-precision data, it is more suitable for small-scale research.

3.1 Questionnaire survey

Because of sensor security, sample size, and occupant privacy problems, questionnaires are often used for data collection in residential buildings. Richardson et al. (2008) used the United Kingdom's time use survey (TUS) data to establish an occupancy model. Hourly transition probability matrices were generated from the data. And Markov chain models were used to simulate occupancy curves. Similarly, the U.S. Bureau of Labor Statistics conducted a detailed survey of occupants in residential buildings, and this survey was called the American TUS (ATUS). The purpose was to record the time and place of the daily activities of respondents. Diao et al. (2017) identified and classified occupants' behavior with direct energy consumption outcomes and energy time use data through unsupervised clustering. Based on the ATUS, the proposed approach integrates k-modes clustering and demographic-based probability neural networks and identifies 10 distinctive behavior patterns. The results showed that the proposed behavior model offers a more accurate and reliable prediction than the ASHRAE standard schedule. To study the presence of urban residential occupants in China, Hu et al. (2019) conducted a large-scale survey of households in four cities in China and received 3424 valid samples. Based on the occupancy data, rooms are divided into four types: bedroom, study, living room, and kitchen. Each hour is divided into two parts for the recording of half-hour steps. The marked time indicates "present"; otherwise, "absent." The data analysis results showed that, from a time perspective, the most commonly used room (i.e., the master bedroom) was not occupied for more than half of the testing time. From a spatial perspective, when the statistical unit was a family, the occupancy was only 40%–50%. However, when the statistical unit was a single room, the occupancy was only 20%. Four typical Chinese urban household living patterns were extracted from K-means clustering analysis, and a large gap in the distribution of occupancy was found in different cities. The questionnaire survey obtained the occupancy mode in a simple way. By recording the occupant's position information 24 h a day, the statistics of the questionnaire can be used to classify the occupancy mode or to generate a time-shift probability for simulation.

The main purpose of the questionnaire design is to obtain the time distribution of the survey objects in different areas within a day, to obtain the room occupancy and the transfer probability in different time periods. However, the time step interval division of 24 h may cause inaccurate records in a subhour period of time if the number of occupants changes significantly. For example, it is more flexible when occupants leave the office to eat and then return to the office at noon. The time period from 11:00 a.m. to 1:00 p.m. can be divided into shorter time intervals, such as 10 min, which may improve the accuracy of the occupancy data. For shorter time intervals, however, either the questionnaire needs to be more frequent, or the subject needs to be able to accurately recall everything that happened in the past.

3.2 Monitoring tools

In this section, an overview of the monitoring tools used in the literature to collect occupancy data is provided. Figure 4 summarizes the sensor types applied in recent years, and Table 2 compares the advantages and weaknesses of different monitoring tools in the literature.

A PIR sensor is an electronic device that works by detecting energy emitted by other objects. It has a detection accuracy range of 80% to 90%. Because of its easy of implementation and low cost, the PIR sensor is widely used in detecting occupant movements and has become the most commonly used sensor (see Figure 4). Combined with PIR, the most commonly used model for predicting occupancy is the statistical model (Castanedo et al. 2011). Dodier et al. (2006) deployed a PIR sensor network in a space system

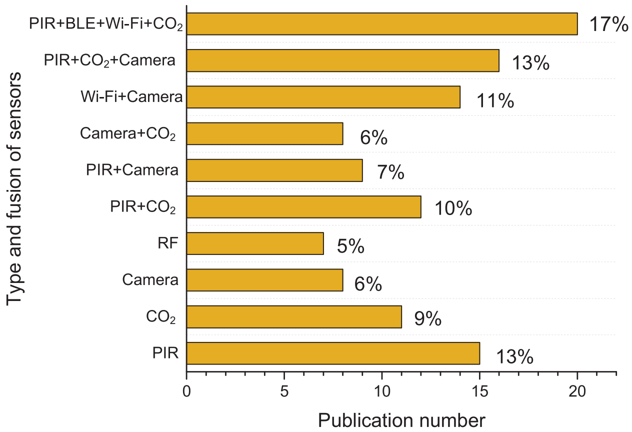

Fig. 4 Number and proportion of different sensor types and combinations in the literature

to obtain occupant presence information and predicted the existence of occupants in a building through Bayesian probability theory based on the statistics of occupancy. However, PIR could not precisely detect occupant numbers in the case of multiple people entering a room together. To solve this problem, a new method combining a PIR sensor and gateway device was proposed by Liu et al. (2016b). The accuracy of occupancy detection was significantly improved when the signal from the PIR sensor was processed by a filter and amplifier before backing to the gateway. To avoid false-negative detects caused by the stationary occupants, Wu and Wang (2019) first developed a Lavet motor PIR (LAMPIR) sensor for true presence detection with 97% accuracy. Then they (Wu et al. 2020) developed a synchronized low-energy electronically chopped PIR which uses a liquid crystal shutter to detect stationary occupants with 100% accuracy. Liu et al. (2017b) developed a chopped PIR (C-PIR) presence sensor to detect stationary occupants with 100% accuracy. In addition, pyroelectric infrared sensors come into use for detection (Fang et al. 2006; Hao et al. 2009; Liu et al. 2016a; Lu et al. 2017).

RF technology is based on electromagnetic signal detection. It has a detection accuracy range of 85% to 95%. The detection system applying RF technology is flexible in deployment and has a large communication range. Li et al. (2012) developed an occupant presence prediction system based on RF identification (RFID) and applied the k-nearest neighbor algorithm to locate the RFID tag of each person. In research conducted by Manzoor et al. (2012), passive person RFID was used to replace PIR to transform the indoor occupant detection mechanism. However, the application of RF signal systems must be combined with a large number of reference marks. The error increases when the target is too far from the occupants' label or the number of reference marks is insufficient.

Compared with the PIR and RF sensors, a camera has higher accuracy. It has a detection accuracy range of 95% to 100%. With a camera installed, not only can the specific

Table 2 Comparison of different monitoring tools and studies

Monitoring tool	Advantage	Weakness	Literature
PIR	Easy to deploy and low cost	Unable to detect static state, multioccupant situation, and limited detection range	Wang et al. 2005; Dodier et al. 2006; Delaney et al. 2009; Lu et al. 2010; Davis and Nutter 2010; Castanedo et al. 2011; Wahl et al. 2012; Duarte et al. 2013; Hong et al. 2016b; Raykov et al. 2016; Liu et al. 2016b
RF	Easy to deploy and wide communication range	Subject to indoor electro-magnetic conditions	Ni et al. 2004; Zhen et al. 2008; Khoury and Kamat 2009; Tesoriero et al. 2010; Li et al. 2012; Manzoor et al. 2012; Conte et al. 2014; Yang et al. 2016; Ahmad et al. 2021
Camera	High accuracy	High cost, privacy problems	Stancil et al. 2008; Erickson et al. 2009; Erickson and Cerpa 2010; Erickson et al. 2011; Benezeth et al. 2011; Erickson et al. 2013; Liu et al. 2013; Petersen et al. 2016; Flament et al. 2016; Chen et al. 2018c
CO_2 sensor	Easy to deploy	Delay problems	Meyn et al. 2009; Dong et al. 2010; Calì et al. 2015; Weekly et al. 2015; Ebadat et al., 2015a, 2015b; Szczurek et al. 2016; Candanedo et al. 2017; Alam et al. 2017; Zuraimi et al. 2017; Jin et al. 2018; Li et al. 2019
Combination sensor	High accuracy	High cost	Ito and Nishi 2012; Zhang et al. 2012; Ebadat et al. 2013; Javed et al. 2015; Masood et al. 2015; Candanedo and Feldheim 2016; Jiang et al. 2016; Liu et al. 2016a; Chen et al. 2016; Amayri et al. 2016; Szczurek et al. 2017; Masood et al. 2017; Zhu et al. 2017; Javed et al. 2017b; Liu et al. 2017a; Chen et al. 2017b; Wang et al. 2018a; Hobson et al. 2019; Wang et al. 2019b
GPS, Wi-Fi, Bluetooth	Efficient and convenient	The join point does not match the number of occupants	Xi et al. 2014; Corna et al. 2015; Zhou et al. 2016; Zou et al. 2017; Wang et al. 2017; Stazi et al. 2017; Wang and Shao 2017; Wang et al. 2019a; Wang et al. 2019e

number of occupants at a certain moment be obtained, but the location of each person can be determined as well. It is beneficial to collect information of multiple occupants when they share one room. Liu et al. (2013) developed a video detection-based occupant detection system at the entrance and indoors of buildings. They used the dynamic Bayesian network method to combine the detection results from multiple visual sensors at the entrance and indoors to establish a more accurate prediction of occupancy. An image-based method was developed by Petersen et al. (2016) to realize a real-time unsupervised function for counting occupants when people move into and out of a room. An accuracy of more than 98% was demonstrated using the camera Kinect V2 for Windows combined with various embedded sensors and different image-processing techniques. However, the method based on the Kinect camera and a PC can only be used in research on small buildings with fewer entrances because of its complicated computational technique and high cost; otherwise, the method leads to a rather large investment particularly in cases where entrance space needs to be detected at the same time.

Complications arise when using acoustics because of the influence of sound in adjacent room bays. Other parameters (temperature, humidity, light, etc.) have a poor correlation with the number of occupants. The presence of occupants directly increases the CO_2 concentration in the space. The CO_2 sensor has been proved to be the most effective environmental sensor for estimating the presence of occupants in a room (Candanedo et al. 2017). It has a detection accuracy range of 85% to 95%. Szczurek et al. (2016) proposed a simple method to estimate the number of occupants using statistical pattern matching. First, the authors measured and recorded the CO_2 concentration using a time sequence method, and then used pattern matching to explore the connection between the internal structure of the time series and the occupancy. Weekly et al. (2015) developed a data-driven partial-differential-equation–ordinary-differential-equation model to describe the change of CO_2 concentration in conference rooms to estimate the occupancy. The accuracy of the equation was verified by experiments. However, different types of occupants need to be distinguished according to the current activity, diet, and body size.

In recent studies, a sensor network composed of a variety of sensors (light, sound, temperature, humidity, passive infrared, CO_2, and so on) has become popular. It has a detection accuracy range of 95% to 100%. Meyn et al. (2009) proposed a new method of occupancy estimation based on various sensor data: the sensor utility network (SUN). Compared with single-type sensor data, occupancy error can be reduced from 70% to 11%. In the use of multiple sensor combinations, Ebadat et al. (2013) used CO_2, temperature, and wind speed sensor data to develop a dynamic occupancy model. They found that temperature parameters had the least impact on modeling. Dong (2010) also used a combination of sensors (temperature, humidity, CO_2, and light) to investigate the presence of occupants in a room and found that the CO_2 sensor data had the greatest impact on the prediction results. The two most common combinations of sensors are shown in Figure 4. The first is a PIR sensor, a CO_2 sensor, and a camera. A CO_2 sensor is arranged in the room to obtain information on the number of occupants. A PIR sensor is arranged on the seat of the occupant to detect the presence status. The camera records the actual movement of the occupant. This combination is suitable for the prediction of the presence of multiple occupants in an office building. Combined with the traditional Markov chain method, the general accuracy requirements of lighting and HVAC system control can be achieved. The second is PIR sensors, environmental sensors, and wireless networks, which are suitable for professional research laboratories. The number of sensors, placement location, and coverage should be precisely calculated. It is designed to monitor real-time occupant positions in minutes. However, because of its high cost, the second type of sensor network has not been applied in many buildings thus far.

With the rapid development of information technology and the spread of smartphones, the latest technologies, such as the global positioning system (GPS), wireless local-area networks, radar sensor, and Bluetooth, have been widely used in occupant detection. It has a detection accuracy range of 95% to 100%. Lu et al. (2016) proposed a room-level presence estimation system based on Wi-Fi. The system obtains temporary occupancy information by matching the received signal strength measurements with the anchor point measurements in different areas. Depatla et al. (2015) calculated the total number of occupants based on the wireless received signal strength indicator measurement between a pair of fixed transmitters/receivers. However, Wi-Fi communication always may cause significant detection uncertainties. Wang et al. (2019c) proposed an event-triggered updating method to cater for the discontinuity in Wi-Fi communication of smartphones which can improve the detection accuracy from 77.3% to 96.8%. Santra et al. (2018) proposed a new signal processing technology for radar sensors to detect human motions even minor movements, which can achieve 100% accuracy in estimating accuracy.

In addition, many researchers have proposed a sensor network monitoring system. Naser et al. (2021) showed that the classification approach based on Thermal sensor array (TSA) using the adaptive boosting algorithm is an accurate approach that has achieves an accuracy of 98.43% and 100% from vertical and overhead sensor locations respectively. Qu et al. (2019) applied a thermopile sensor to localize and track indoor multiple human targets. Chen et al. (2018a) developed an unobtrusive sensor (ARGUS) to detect occupants in the room. The average accuracy reaches 89.1%, 95.3%, and 95.1% at the distance of 0.6 m, 1.2 m, and 1.8 m respectively. Tyndall et al. (2016) proposed an occupancy estimation sensor system based on low-pixel count sensor arrays.

According to different physical properties, the classification of sensors is also different (Table 3). For different resolutions of occupancy, sensors can be selected according to the following principles: On/off (PIR), low price, high accuracy for indoor occupants monitoring; Count (seat pressure sensor, seat thermocouple sensor, and camera). In recent years, with the improvement of sensor technology, the monitoring accuracy has significantly improved, which also caused the higher cost. Distribution and tracking (RF, Wi-Fi, and BLE), can effectively track the occupant movement at different time points to get the hourly distribution of occupancy in the whole building.

When only concerning the prediction accuracy, the combination of detection technologies and occupancy models can be ordered from high to low as PIR and deterministic schedule model; RF, camera, CO_2 sensor, and stochastic schedule model; GPS, Wi-Fi, Bluetooth, and machine learning model. When only concerning the sensor price, the combination of detection technologies and occupancy models can be ordered from high to low as PIR and deterministic schedule model; CO_2 sensor and stochastic schedule model; RF, camera, GPS, Wi-Fi, Bluetooth, and machine learning model. When both concerning the prediction accuracy and the sensor price, the combination of detection technologies and occupancy models can be chosen as: If the building has deployed a wireless network signal, the optimal combination is GPS, Wi-Fi, Bluetooth, and machine learning model. Otherwise, the CO_2 sensor and stochastic model can also achieve relatively ideal prediction accuracy at a lower cost.

For different types of buildings, sensors can be selected according to the following principles: For residence buildings, PIR sensor is commonly used, which is selected for its low cost and privacy-preserving. Recently, with the popularity of wireless networks for most families, the best combination of sensors is PIR, Wi-Fi and Bluetooth; For commercial buildings and industrial buildings, occupant counters can be set at the entrance and exit. When the privacy problem is not considered, the camera can be applied for occupancy detection; For the station and shopping mall with a large population density, the installation of cameras and wireless transmitting equipment can effectively obtain the building occupancy information. The best combination of sensors and models is camera, Wi-Fi and Bluetooth. The hourly occupancy information can be used to control the air conditioning system; For office buildings, the first selection is the combination of PIR/seat pressure sensor and the camera. The data obtained by the camera can be used to calibrate the PIR and pressure sensor error and verify the accuracy of the model. Because the wireless network in office building is spread, the second selection is to set up LAN to monitor the distribution of smart phones and indirectly obtain occupant information.

4 Establishment of occupancy models

Because of the randomness and complexity of occupancy, it significantly increases the difficulty of energy consumption simulation. Therefore, it is necessary to quantify the occupancy through a mathematical model. By using the occupancy model as the input of the building energy simulation software, the gap between simulation and the actual energy consumption is considerably reduced. In this section, the occupancy models in the existing studies are summarized. They can be divided into the following three models: deterministic schedule (fixed model), random

Table 3 Classification of sensors according to different physical properties

	Property 1		Property 2		Property 3		Property 4		Property5		
	Active	Passive	Wearable	Remote	Privacy-reserve	Privacy-invasive	Single occupants	Multiple occupants	Single sensor node	Sensor fusion	Sensor network
PIR		√	√		√			√	√		
RF	√			√		√		√			√
Camera	√		√			√		√		√	
CO_2 sensor		√	√		√			√	√		
GPS, Wi-Fi, Bluetooth	√			√	√		√				√

model (probability), and machine learning. The occupancy models are classified and compared in Table 4.

4.1 Deterministic schedule method

The deterministic schedule (fixed model) represents the lowest level of complexity and is commonly used in building energy simulation software (Gaetani et al. 2016). The data collection of the deterministic schedule method is based on the long-term monitoring and statistics of the actual occupancy in the building. Probability distributions are obtained from the statistical results by analyzing the frequency of certain occupant behaviors. The probability distribution can help estimate the possibility of concentrated group behavior and establish a probabilistic model for occupant behaviors. Commonly used statistical methods include linear regression models, logistic regression models, time series models, and Bayesian estimation.

Wang et al. (2005) studied a single-person office in a large office building and proposed a predictive probability model for the occupancy of a single-person office. The model was a nonhomogeneous Poisson model with two different

Table 4 Occupancy model classification and literature summary

Model	Classification	Advantage	Weakness	Reference
Deterministic schedule mode	Poisson distribution	Simple model construction	It cannot accurately describe the randomness of single-person movement and the prediction accuracy is low	Wang et al. 2005
	Ordinal logistic model			Haldi and Robinson 2011
	Time series			Feng et al. 2015
	Logistic regression			Haldi and Robinson 2008; Tabak and de Vries 2010; Chang and Hong 2013
	Bayesian probability			Harris and Cahill, 2005; Dodier et al. 2006; Zhao et al. 2015
	Support vector regression (SVR)			Wang et al. 2017
Stochastic schedule model	Markov model	It can reflect the randomness and autocorrelation of occupant movement and has high prediction accuracy, which can be used for HVAC system control	The parameter setting of the model is complicated and the applicability of the model is poor	Page et al. 2008; Widén et al. 2009; Erickson and Cerpa 2010; Erickson et al. 2011; Wang et al. 2011; Chiou et al. 2011; Virote and Neves-Silva 2012; Han et al. 2012; Dobbs and Hencey 2014; Hong et al. 2015; Chen et al. 2015
	MCMC			Bartley and Plewis 2002; Dong and Lam 2014; Wang and Ding 2015; Chen et al. 2017a
	Hidden Markov model (HMM)			Dong et al. 2010; Hong et al., 2015
	Autoregressive HMM (ARHMM)			Ai et al. 2014
	Layered HMM (LHMM)			Milenkovic and Amft 2013
	Dynamic Markov time-window inference			Wang et al. 2017
Machine learning	RNN	High prediction accuracy, can achieve real-time occupant prediction	Data quality and quantity requirements are high, and model training takes a long time	Javed et al. 2017a; Wang et al. 2018b
	SVM			Lin et al. 2001; Shih 2014; Zhen et al. 2008
	Radial basis function (RBF) neural network			Yang et al. 2012
	K-means			Harris and Cahill 2005; Zhao et al., 2017
	Polynomial regression			Wang and Ding 2015
	Decision tree			Nesa and Banerjee 2017; Hailemariam et al. 2011
	Genetic programming (GP)			Yu 2010; Newsham et al. 2017
	Presence sense(PS)			Jin et al. 2014
	Adaptive neuro-fuzzy inference system (ANFIS)			Ekwevugbe et al. 2012

exponential distributions, which can be used to reflect the statistical characteristics in office buildings. The model assumed that the present time was not related to the time of day, and the office vacancy interval was exponentially distributed. In addition, the coefficient of the single office vacancy distribution can be considered to be a constant within a day. However, the occupancy changes with time and is more complex than the vacancy interval distribution because a motion sensor records all the information irrespective of the number of people in a room. This method is simple and easy to implement for a single office, but it is severely limited when describing the occupancy of different areas in a building because of the movement of occupants.

4.2 Stochastic schedule method

A deterministic schedule cannot accurately describe and predict the stochastic behavior of occupants. Based on the uncertainty of occupant behaviors, researchers have explored the correlation between occupant behavior and environmental stimulus or specific events. They used sample statistics to calculate the probability of an event (movement) (Happle et al. 2018) and then developed a stochastic model to describe the behavioral characteristics of occupants in buildings. The stochastic process considers the room occupancy as a random variable, and the probability of occupancy state at each time point is determined according to the previous state (Yang et al. 2016).

4.2.1 Markov chain model

A Markov chain is a time series model. In the time series, all states of a system can be directly observed. The Markov chain technique is used for the analysis aiming to develop the probabilistic occupancy profiles. Since the occupants' movement among the zones inside and outside rooms creates the occupancy profile, random mobility between different work states is assumed. Therefore, the next work state of the occupant only depends on his/her present state and some rules about the work states. Knowing that the future state of the occupant depends on his/her current state, the transitions of states are defined in Markov matrices. The future state depends only on the current state and not on all past states (Zhang et al. 2018). Page et al. (2008) proposed a nonhomogeneous two-state Markov chain model that described the presence of an occupant in a single-person office. The Markov chain has been the most frequently used occupancy model in recent years. The room occupancy can be considered as nonhomogeneous Markov chains interrupted by occasional periods of long absence. The transition probabilities between the three states of entering, leaving, and staying can be estimated based on the presence probability in each time step. However, the Markov model has two main disadvantages. (1) Because of the strong dependence on the time of occupant movements and the presence in the room, it is difficult to obtain detailed timetables in surveys and tests. (2) The Markov model cannot describe the movement of occupants from one area to another, which affects the accuracy of the occupancy prediction.

To simulate the movement process of multiple occupants, Wang et al. (2011) developed an event-driven occupancy prediction method on the basis of Markov chains. The occupancy is treated as a direct result of the movement of occupants inside and outside the building. However, there are two problems with this method. One is that the definition of the event is unclear. The state changes of some events are ambiguous. The other is that, in the case study, the authors only showed simulated data without actual measured data. The model could not be calibrated and verified. To simulate the occupancy in a multiperson single-room and multiperson multiroom situation accurately, Chen et al. (2015) proposed two nonhomogeneous Markov chain models based on previous research (Chen and Soh 2014). The transition probability matrix was a key parameter of the Markov chain model. By defining the state as a vector, the matrix calculation was significantly simplified.

4.2.2 Monte Carlo model

Monte Carlo models are also widely used for the stochastic process. According to the probability distribution of the occupancy in each room, the building occupancy can be estimated. This is a simple method to study the uncertainty of occupancy. It does not consider the interrelationships and effects between occupancy at different times in one room, nor does it consider the interrelationships and effects between different rooms at the same time.

MCMC is based on an assumption that the current state only depends on the most recent past state. Generally, the state is defined as the number of occupants or the individual location. The states can transit between the time steps based on the transition probability matrix. In general, the Markov model is only the representation of the occupancy, and the Monte Carlo method can be used to traversal the process of moving inside and outside the building to get random results, which can simplify the computational complexity of the transfer matrix.

The earlier published stochastic occupancy model was proposed by Richardson et al. (2008). The author extracted information from a large survey dataset (Bartley and Plewis 2002) conducted in the United Kingdom in 2000 and developed models using two-state nonhomogeneous Markov chain Monte Carlo (MCMC) technology. Similarly, Widén et al. (2009) adopted MCMC technology and developed a three-state occupancy model. Later, the model was extended to nine states by adding six activities that could cause power consumption, and a lighting control model was proposed based on power demand (Widén and Wäckelgård 2010). Aerts et al. (2014) used a hierarchical clustering method to identify seven significantly different indoor patterns from time-use data and developed a three-state indoor probability model. The model complexity was reduced by storing state transition probabilities in a matrix. Mckenna et al. (2015) developed the two-state indoor activity model established by Richardson et al. (2008) into a four-state model.

4.2.3 Hidden Markov model

The hidden Markov model (HMM) assumes that the possible states of the system are connected in a Markov chain, but the state of the system is not directly observed. Instead, each system state is associated with a set of observable parameters through a probability distribution.

Dong et al. (2010) deployed a complex environmental sensor network, including a wireless environmental sensing system, wired carbon dioxide sensing system, and indoor air quality wired sensing system. The wired cameras were used to obtain real occupancy data. The result showed that there is a significant correlation between environmental and living conditions. During the test, the average accuracy of the HMM for room occupancy prediction was 73%. Later, Dong and Lam (2011) developed an HMM is based on a Gaussian mixture model. The average prediction accuracy of the number of occupants was raised to 83%.

The classical HMM assumes that the observed variables are independent. However, in actual measurement, environmental parameters, such as CO_2 concentration, are not only related to the current number of occupants, but also to the concentration at the previous moment. In addition, because these parameters are related to occupancy, they are also related to each other. Under the circumstance that these environmental parameters are closely interrelated, the prediction accuracy of the HMM model decreases. To solve this problem, researchers proposed an autoregressive HMM (ARHMM) model, which combined autoregressive time series and HMMs. The autoregressive structure recognizes the existence of dependencies between the time series and hidden Markov. The HMM can capture the probabilistic characteristics of transitions between underlying states. Therefore, the model has the ability to establish the relationship between environmental observation variables. Han et al. (2012) developed an ARHMM and compared the results with the support vector machine (SVM) and the HMM. The results showed that the ARHMM performs best, with a simulation accuracy of 80.78%.

4.3 Machine-learning method

Using approximate models guarantees the robustness of a system to the associated uncertainties. A robust system could react to uncertainties and accordingly tune itself. However, accurate models with low rates of prediction error are required to maximize the system's efficiency. Statistical methods, when used alone, cannot ensure robustness; however, both efficiency and robustness can be achieved when statistical methods are combined with approximate models. In recent years, more and more researchers have developed occupancy models with data-mining methods (Ren et al. 2015). Data mining requires a large database and a large amount of data storage, and the occupant presence pattern can be inferred from the large data flow. Data-mining methods can effectively identify the behavior of occupants in buildings and provide energy-saving modes of occupant behavior, which is conducive to reducing the building energy consumption (Yu et al. 2010, 2011, 2012). By combining statistical and stochastic methods to process historical data, data mining can obtain rules from which to predict future development directions. It usually includes artificial neural networks (ANNs), decision trees, SVMs, polynomial regression, and Bayesian networks (Tsanas and Xifara 2012).

Dong et al. (2010) analyzed the correlation between environmental conditions and the number of indoor occupants by deploying a complex sensor network. Because of a large number of open offices, CO_2 and acoustic parameters have the greatest correlation with the number of occupants. During the test, three machine-learning methods (HMM, ANN, and SVM) were used to estimate the schedule of the occupancy on a typical day, and it was found that the HMM model more accurately described the overview of the staff. However, Han et al. (2012) thought that all three machine-learning algorithms have difficulty capturing the correlation between different environmental parameter measurements. By considering the correlation, it is possible to improve the estimation accuracy. Han assumed that occupants affect the indoor environment by emitting carbon dioxide, heat, moisture, and sound, and measurements should be conducted by combining those indoor environmental variables. However, no clear conclusion can be given to comparing the prediction accuracy of the three models.

Machine-learning models have a high degree of complexity and strict requirements for the number of sample data, and they require large databases and large amounts of data storage. In the studies that used machine learning to predict occupancy, most data were collected through a combination of sensor networks. Single-sensor data cannot meet the model requirements. The prediction accuracy is improved, but the cost of data collection is also increased. After obtaining a large number of data based on combined sensors, finding a suitable method to process the data to develop a model is crucial for occupancy prediction.

4.4 Simulation accuracy of occupancy models

Occupancy means the ratio of the actual number of occupants in an area or building to the maximum designed number in a given period of time. Simulation of occupancy models means the application of the model. Occupancy schedule obtained from model is used for building air-conditioning system control or energy consumption prediction. The lack of standardization and consistency in the evaluation of occupancy model accuracy has led to difficulties in comparing different studies. This is mainly reflected in two ways. First, the approach for verifying simulation results is different. It can mainly be divided into three types: comparison with the previous model, comparison with fixed schedule models, and comparison with the measured data. Second, the forms of simulation results are inconsistent. The simulation results for verification can be roughly divided into three types: presence or absence, number of occupants, and energy consumption. In addition, there was still no clear conclusion after simulation in some studies, in which the results were only displayed in charts with the form of simulation results and actual measurement results. Table 5 classifies the occupancy model, data collection method, building type, and comparison of simulation results.

With the statistical method, researchers simply described the change of the occupancy, while the accuracy of the simulation results was not investigated in detail. To obtain room-occupancy curves for different time periods for different rooms, Duarte et al. (2013) deployed 629 PIR sensors in large-scale office buildings to obtain occupant presence data, including private offices, multiperson offices, and other types of room. The analysis results showed that the measured occupancy is significantly lower than the recommended value of ASHRAE 90.1 2004, and the peak occupancy curves for multiperson and private offices decreased by approximately 46% and 12%, respectively. Although the variation tendency

Table 5 Comparison of occupancy model, data collection method, building type, and simulation results

	Literature	Modeling method	Sensor	Building type			Absence state	Number	Energy consumption	Comparison type		
Machine learning	Zhen et al. 2008	SVM	RFID	—	—	—	—	93%	—	—	—	—
	Meyn et al. 2009	SUN	Camera, PIR, and CO_2	—	Public office	—	—	89%	—	—	With the measurement	—
	Yu 2010	Genetic algorithm	PIR	Single office	—	—	—	80%–83%	—	—	With the measurement	—
	Shih 2014	SVM	PTZ camera network	—	Public office	—	—	—	—	—	—	—
	Hailemariam et al. 2011	Decision tree	PIR, CO_2, sound, and light	—	Public office	—	—	98.4%	—	—	With the measurement	—
	Yang et al. 2012	RBF neural network	Temperature, humidity, CO_2, light, sound, and PIR	—	—	—	—	87.62%	—	—	With the measurement	—
Statistical methods	Wang et al. 2005	Exponential function	PIR	Single office	—	—	—	—	—	—	—	—
	Dodier et al. 2006	Bayes probability theory	PIR	Single office	—	—	—	—	—	—	—	—
	Wang and Ding 2015	Cubic polynomial regression	Camera	—	Public office	—	—	—	Less than 5%	—	With the measurement	—
	Duarte et al. 2013	Statistical methods	PIR, camera	Single office	Public office	—	—	46% difference with ASHRAE	—	—	—	With deterministic models

Table 5 Comparison of occupancy model, data collection method, building type, and simulation results (Continued)

	Literature	Modeling method	Sensor	Building type			Absence state	Number	Energy consumption	Comparison type		
Stochastic model	Richardson et al. 2008	MCMC	TUS data	—	—	Residential building	—	—	—	—	With the measurement	—
	Chiou et al. 2011	MCMC	ATUS data	—	—	Residential building	—	—	—	—	With the measurement	—
	Widén et al. 2009	MCMC	TU-SCB-1996	—	—	Residential building	—	—	—	—	With the measurement	—
	Dong et al. 2010	Hidden Markov	Environmental sensor networks, and cameras	—	Public office	—	—	73%	—	—	With the measurement	—
	Han et al. 2012	ARHMM	PIR, CO_2, RH	—	Public office	—	—	80.78%	—	With the former	With the measurement	—
	Ai et al. 2014	ARHMM	PIR, CO_2, RH, temperature, and wind speed	—	Public office	—	—	Higher accuracy than HMM	—	With the former	With the measurement	—
	Calì et al. 2015	CO_2 concentration mass balance equation algorithm	CO_2	—	Public office	Residential building	√	80.6%	—	—	With the measurement	—
	Page et al. 2008	Inhomogeneous Markov chain	PIR	—	Public office	—	—	—	—	—	—	—
	Salimi et al. 2019	Inhomogeneous Markov chain	Bluetooth and smartphone	—	Public office	—	—	84%	—	—	With the measurement	—
	Chen et al. 2015	Inhomogeneous Markov chain	Camera	Single office	Public office	—	—	—	—	—	With the measurement	—
	Erickson et al. 2011	BMC (blend Markov chain)	SCOPES system	—	Public office	—	—	—	42% energy consumption reduction of HVAC	—	—	With deterministic models ASHRAE
	Erickson and Cerpa 2010	MWMC	SCOPES system	—	Public office	—	—	—	20% energy consumption reduction of HVAC	—	—	—
	Yang and Becerik-Gerber 2014	ARMA, ANN, MC, and logistic regression	Motion, sound, indoor temperature, humidity, CO_2, light, and PIR	Single office	Public office	—	—	97.3%	—	—	—	With deterministic models
	Wang et al. 2017	DMTWI	Wi-Fi and camera	—	Public office	—	—	80%	—	—	With the measurement	—

of occupancy is consistent with reality, there is no comparison between the predicted and actual values.

The most widely used stochastic model is the Markov chain. Salimi et al. (2019) proposed a nonhomogeneous Markov chain adaptive probability indoor prediction model based on a real-time positioning system. The prediction accuracy of the single-person room occupancy reached 84%. Erickson and Cerpa (2010) proposed a control strategy for an HVAC system based on the Markov chain occupancy model. HVAC energy consumption can be reduced by 20%. Wang et al. (2017) proposed a dynamic Markov time-window inference (DMTWI) approach based on Wi-Fi signal detection and compared it with autoregressive moving average (ARMA) and SVR models. The result showed that DMTWI has the highest accuracy (80%) in the prediction of the occupant number.

Machine learning requires a large number of datasets, which provides a higher simulation accuracy than that of the two methods described above. Yu et al. (2010) explored the presence mode of a single-person office using the genetic programming (GP) algorithm. Testing in five different offices resulted in the accuracy of the prediction model reaching 80%. Hailemariam et al. (2011) used a decision tree with multiple sensor types for real-time occupant presence detection. The results showed that the combination of data from multiple motion sensors can improve the accuracy to 98.4%. Calì et al. (2015) proposed an algorithm based on indoor CO_2 concentration. It was verified and evaluated in different scenarios. The results showed that the prediction accuracy of occupant presence in the room and the number of occupants can reach 95.8% and 80.6% respectively.

5 Practical application of occupancy models

5.1 Application to FMU

There is a big gap between the building energy consumption simulation and the actual building operation results. One of the main reasons is the inaccurate estimation of the occupancy. In the traditional energy consumption simulation program, the description of the occupancy and the interaction correlation between the occupants and the building system may be too simple. Most simulation results were presented in the form of graphs, but it was still difficult to show the influence of occupant behavior on the building energy efficiency. As a matter of fact, the occupancy mode varies greatly due to different building types and occupant densities, so the key to improve the accuracy of building simulation is to get a more realistic random occupancy schedule. In addition, to model building occupancy needs a lot of real data through monitoring, which is related to the quality of building simulation results. In recent years, with the development of computer technology, the use of cosimulation to integrate separate occupant behavior software modules with building energy modeling has been increasing, but the coupling mechanism between behavior models and energy consumption simulation is still in its early stages.

Obtaining a schedule through the occupancy model and inserting it into energy simulation software is the most commonly used method. Yang and Becerik-Gerber (2014) used regression models, time series models, pattern recognition models, and stochastic processes to model the occupancy and found that the ARMA time series model has the highest accuracy. The occupancy information of different models was simulated in OpenStudio.

There are still some more complicated methods, including customized code, customized tools, and cosimulation, that can represent the occupant behaviors through collaborative simulation. Gunay et al. (2014) used the discrete event system specification building energy consumption model to couple building, HVAC, and occupant data. Langevin et al. (2015) developed an agent-based model (ABM) of occupant behavior in office buildings and conducted energy consumption simulation through the Building Controls Virtual Test Bed, which combines the ABM in MATLAB with EnergyPlus. Hong et al. (2015) proposed a DNAs (Drivers, Needs, Actions, and systems) framework that describes the occupancy model as a standardized structure. The framework can be integrated into current simulation programs or functional model units to support model exchange and collaborative simulation of dynamic models. On the basis of a previous study (Hong et al. 2015), Hong et al. (2016a) developed a new occupant behavior modeling tool with a behavior functional mock-up unit. In further research, Chen et al. proposed an agent-based random occupancy simulator. Users can download simulation results in the CSV and EnergyPlus IDF file format for joint simulation of other building energy simulation programs (Luo et al. 2017; Chen et al. 2017a, 2018b; Hong et al. 2018).

5.2 Control of energy system

The key to improving the operating efficiency of the air conditioning system is to provide the right amount of heat and fresh air into the area, the volume of which will be directly determined by the number of occupants. And the system control requires high-precision occupancy information in the room. The start and end time of the control system depend on the arriving and leaving time of occupants, and the number of occupants determines the air volume. Many studies have shown that the control of HVAC and lighting systems based on occupancy has great energy-saving potential. Therefore, obtaining a random occupancy schedule through occupancy prediction can improve the accuracy of building simulation and provide a better HVAC control strategy.

The occupancy model is widely used in practical HVAC and lighting control. Agarwal et al. (2010) divided the building area according to the real-time presence status of occupants, thus saving building energy consumption by holding indoor temperature set points individually. Erickson et al. (2011) defined the temperature control algorithm in an HVAC system based on a hybrid Markov chain model and then substituted it into EnergyPlus software to simulate energy consumption. The results showed that, compared with the existing control strategy based on ASHRAE, 40% energy consumption can be saved every year. Goyal et al. (2012) analyzed the impact of uncertainty in the implementation of model predictive control (MPC) on energy consumption, thermal comfort, and indoor air quality. The result showed that occupancy prediction and model error have the greatest impact on energy consumption. Frequent changes in occupancy lead to a significant decrease in MPC prediction performance, increasing the energy consumption of the HVAC system by 25%.

In open office spaces, there is no consistency in the time and number of people arriving and leaving the office (Kwok et al. 2011). The flexibility of working styles increases the diversity of working hours and working positions. Therefore, lighting control based on occupancy has become more challenging. Haq et al. (2014) summarized the early occupant detection technologies that can be used for lighting control, and Guo et al. (2010) outlined the energy-saving effects of lighting that result from the implementation of occupancy detection sensors. However, control through presence detection has defects. The sensor has a certain delay, thus increasing error when judging there are no people in the room. To overcome the shortcomings of the above sensors, more advanced lighting controls with adaptive and learning capabilities have been developed. In one study (Hughes and Dhannu 2008), adding an illuminance sensor based on occupancy sensor was found to save 65% of energy consumption compared with the continuous lighting mode in the entire day. Nagy et al. (2015) proposed an adaptive control strategy by analyzing the PIR detection data and the illuminance threshold of occupant interaction with a lamp. The results showed that, in terms of lighting control, the optimal delay time setting of PIR should be able to guarantee 95% detection accuracy, and the optimal delay value of different types of room can differ by five times (4–20 min).

5.3 Basis of energy model

The occupancy model is often used as the basis of a power load forecasting model. More accurate acquisition of occupancy information is a key factor for improving the simulation accuracy of the entire load forecasting model.

Wang and Ding (2015) used an MCMC method to build the equipment input power model. Based on the indoor occupancy model and the equipment input power model, they developed the occupant-based energy consumption prediction model through time accumulation. The results showed that there is a 5% error lower than the actual energy consumption record. Wang et al. (Wang et al. 2019d) used camera sensors to detect the number of occupants, obtained plug load data from the energy consumption monitoring system, and found that there is a linear relationship between the two during the working day. The plug load during the nonworking day is almost unchanged and has nothing to do with occupancy changes. An et al. (2017) proposed a stochastic occupant behavior method (SOB) to simulate regional cold load. They used Markov chains to simulate the location of each person and the presence of occupants in the area and to store the output results in an SQLite database. Then they used the Designer's Simulation Toolkit software to simulate the cooling load by combining the air conditioning, lighting, and window control schedules with the occupant. The results showed that the peak load error is within 9% of the actual error.

Buttitta et al. (2019) developed a comprehensive indoor model to predict the heating demand of residential buildings. They integrated the occupancy, electrical appliances, and lighting heat gain models to obtain the thermal load prediction model of the residential building. Compared with the model that does not consider the occupant in the room, the difference was as much as 30%. Later, they used the MCMC method to develop occupancy models (Buttitta and Finn 2020). The simulated occupancy profile was used as the input of a dynamic building equivalent resistance–capacitance model to obtain high-precision heating load forecast curves for a day, month, and year.

6 Discussion

Researchers need to consider the accuracy of data collection, the complexity of the model, and the accuracy of the prediction when choosing an occupancy model. Because the types of sensors are complex and diverse, such problems as deployment costs, monitoring effects, and privacy must be considered during data collection. If the selected sensor is part of the existing infrastructure of the building, such as Wi-Fi, the cost is not generated during the data collection stage. Otherwise, cost becomes a major problem that researchers must consider in practical applications. Whether the monitoring effect can reach the accuracy requirement is related to the prediction result of the model, which determines the performance of the entire system in practice. In addition, the privacy problem is also the key to smooth data collection. In the data collection process, the will of the indoor occupant must be fully considered, especially when the occupancy is monitored by means of images, such as by the use of cameras. Accurate statistical data can essentially improve the accuracy of the occupancy model, and the choice of different models and the diversity of the number and types of model input parameters also lead to differences in the speed of modeling, calculations, and model prediction accuracy. The accuracy of model prediction is the most concerning issue, and it usually acts as the criterion to evaluate the quality of the model. Therefore, the difficulty of data collection, cost level, and calculation speed must be considered during the establishment and verification of the occupancy model. In the following, the data collection sensors are compared, and the advantages and disadvantages of the occupancy model and the model prediction accuracy are discussed separately.

6.1 Application of occupancy monitoring sensors

The data collection of the early occupancy model relied on PIR sensors, environmental sensors (CO_2, temperature, humidity, and so on), and camera networks, which are traditional sensors. PIR and environmental sensors are inexpensive and easy to deploy. However, PIR and environmental sensors always cause occupant detection errors and uncertainties in the data collection because of their limitations. It is recommended that this type of sensor be used to reduce the difficulty of monitoring in situations where the accuracy requirements are not very high and the approximate range of occupancy in the room must be estimated. For example, lighting control based on the status of presence only needs to determine whether the occupant is in the room. PIR and environmental sensors are recommended for lighting control, because their setup and maintenance costs are much lower than those of cameras, which require larger storage space and higher image-processing capabilities. For determining the number of occupants in a room, a camera network system has high detection accuracy. Cameras are usually deployed in professional research laboratories or public places, where test subjects can accept any movements being monitored and recorded. However, cameras cannot be used in offices or residential buildings in which occupant privacy is a concern.

With the development of wireless network communication technology, the rapid application of new wireless-technology-based Wi-Fi and Bluetooth detection systems show great advantages compared with traditional sensors. On the one hand, they use the existing wireless network of the building. Therefore, the launch facility can basically be deployed with no cost. On the other hand, they can effectively distinguish occupant identity and real-time positioning, providing a fairly accurate occupant movement trajectory and long-term data collection. However, using the number of Wi-Fi and Bluetooth connections to determine indoor occupant numbers often causes errors in the case of weak signals. The use of GPS for monitoring can make up for the shortage of Wi-Fi and Bluetooth, but it requires high-precision GPS positioning equipment that is not restricted by environmental conditions. The deployment cost is high. At the same time, the monitored occupant should carry GPS or turn on GPS positioning, which causes a privacy problem to a certain extent. Therefore, when there is a high requirement for the prediction accuracy of the occupancy and it is necessary to determine the identity of the person and the time-based location information, the use of the existing network environment in the building is recommended. Within their scope, traditional sensors and GPS positioning equipment can be added as assistance and verification to establish a database of occupant presence information consistent with the actual situation.

In terms of occupant detection, each sensor has specific advantages and disadvantages. Combining multiple sensors is a good choice to improve the quality of data collection. For example, camera detection is greatly affected by light, and a CO_2 sensor can make up for this. The camera is installed at the entrance of the building. When the light conditions are good during the day, the occupancy is calculated by recording the data of occupants entering and leaving. At night, there are fewer occupants and poor lighting conditions. If CO_2 sensors are installed along with cameras to record data, the recording accuracy is high.

6.2 Advantages and limitations of occupancy models

The establishment of the statistical model is based on a large amount of monitoring of statistical data. The probability distribution of the occupant within a fixed time can be obtained by analyzing the data of the occupant in the room, and the model has low complexity. In principle, when the historical data are sufficient and accurate, the established probability model can reproduce the close-to-actual distribution of occupants in the building. However, this type of model does not consider the inherent driving factors of occupant movement and does not distinguish between different categories of occupants. What is obtained is only the general statistical law, such as the logistic regression method in statistical models. It depends on the relationship between the researchers' state in the room and time or environmental parameters. When using logistic regression, the coefficients obtained from the analysis are used to identify the main influencing variables for occupant movements. The model still relies on the original dataset inferred from historical data. Therefore, the statistical model has the following problems. (1) It can only roughly describe the entire movement of the person in the building, and it cannot reflect the random movement process of individuals or a group of occupants. (2) It is inconsistent with the actual data because the time and space connections of occupants are not considered. Therefore, for the indoor environment of a building where the working mode and movement type of indoor occupants are similar, or where the movement of occupants is largely affected by environmental parameters and the accuracy demand of the occupancy prediction is not high, it is recommended to adopt the deterministic method. It can reduce the difficulty of model establishment and the speed of calculation.

Compared with the deterministic model, the stochastic model starts from the changes in the status of the occupants in the room. The core embodiment is the randomness of individuals moving to different areas during different time periods, maintaining the correlation between the adjacent moments in the same room and the number of occupants at the same moment in different rooms. There are many types of stochastic models. In the early days, simulation was a purposeless movement and the model accuracy was low. Later, event-based methods fundamentally described the movement mechanism, which was widely used in the modeling of office buildings with higher accuracy. However, because of the complicated parameter setting during model construction, the parameter acquisition depends on long-term monitoring, so the investigation is complicated and the cost is high. The data-processing problem is difficult for event-based methods, and there are still some limitations. In addition, the model is not universal, and the same types of buildings need to be modeled separately. When using the Markov model to predict the presence of occupants in an office building, the regular movement of occupants at a fixed time point can be better described. However, there are certain limitations to the time point where the movement of occupants is stochastic. Including the Monte Carlo method can better solve this problem. However, for commercial buildings, such as medical buildings and transportation buildings, where the movement of occupants is not obvious and the density of occupants is high, the stochastic model cannot accurately describe the building occupancy because the model parameters are difficult to obtain, and the probability of the transfer of a single person is difficult to calculate. In this case, statistical methods often have a higher prediction accuracy.

At present, this type of model is not widely used because of the complexity of machine-learning algorithms and the slow calculation speed. The accuracy of machine-learning algorithms is high, and they overcome the problem of fixed simulation results in the traditional fixed work schedule and stochastic model methods. The accuracy of the model can be continuously improved through automatic correction and adjustment. With the development of sensor networks and communication technologies, more data should be available for learning. Machine learning can analyze and process big data without manual intervention. It is a promising tool for building occupancy models in the future. It has a good application prospect in some areas of indoor environment predictive control based on high-precision occupancy.

6.3 Future prospects of occupancy models

It is often necessary to verify the model to estimate whether the established prediction model can truly reflect the presence of occupants in a building. Models based on statistical methods are less complex. Researchers mainly compare simulation results with fixed schedules but do not precisely quantify the simulation accuracy. Compared with the statistical method, the Markov chain model greatly improves the simulation accuracy and can reach an accuracy of more than 80% in single-person and multiperson rooms. The Markov chain model is an explicit model. Its data come from direct monitoring of the occupant movement and room conditions. In the HMM, the occupancy is hidden, and various types of environmental sensors are installed to record changes in environmental parameters. The ARHMM is based on the assumption that the observed variables are independent in the HMM. The relationship between environmental parameters can be considered. The temperature, humidity, and CO_2 concentration are combined and used to reflect the occupancy characteristics. This method solves the problem of feature extraction in the HMM and further improves the simulation progress, which can achieve an accuracy of 90%. Multiple types of historical data are the keys to improve the accuracy of the simulation in machine learning. A complex sensor network is the basis for applying various intelligent algorithms. Compared with the above model, the largest difference in data collection using the machine-learning model is adding a large number of wireless network real-time positioning systems, including Bluetooth, Wi-Fi, GPS, and other communication technologies, which can distinguish and track occupants in the building. This model has the highest accuracy. It can reach an accuracy of more than 95%. The overall prediction accuracy of the above occupancy models is within 80%–95%, which is sufficient for the use of building automation control systems. However, the prediction accuracy obtained here is in the context of professional research. If it is used for different behavioral feature types, the accuracy is reduced. The accuracy requirement depends on the system control application. For example, in the automatic control of lighting systems, even if the accuracy is 90% it may still cause occupant dissatisfaction. This is because, if there is an error, the lighting is turned off directly. Therefore, high precision is necessary. In HVAC control, a large prediction error is acceptable when an incorrect prediction does not cause obvious discomfort. However, the application of various intelligent algorithms of machine learning for occupancy modeling is still immature, such as unquantified data sets, the lack of the comparison of application effects between different algorithms and the set standards of parameters, etc., which still need to be further explored in future research.

In recent years, researchers have done much work on data collection and modeling methods. There are many types of sensor and occupancy models. For different types of buildings, choosing appropriate sensors and modeling methods is conducive to improving prediction accuracy. However, the current research on occupancy models still has limitations. Some models do not specifically classify occupant preferences and consider the interaction between occupants too much. This has led to a reduction in prediction accuracy. Moreover, data coverage for different types of buildings is not sufficiently comprehensive. Most of them are concentrated in residential and office buildings, while complete monitoring data and appropriate prediction models are not available for stations and airports with a large population density. Furthermore, there is no uniform standard for the length of data collection, data quality, and processing methods in the literature. Researchers often use a single building as an example for model verification; thus, applicability to other buildings can be lacking. At the same time, the verification results do not have consistent evaluation standards. This is not conducive to making comparisons between different models. The above limitations should be taken into consideration in future research.

Also, Building Performance Simulation is often used to quantify the impact of occupant presence and behavior on simulation predictions, the energy-saving potentials of different energy conservation measures, and the performance of various building systems. Occupant presence and behavior in Building Performance simulations are commonly represented with diversity factors in the form of standardized hourly schedules or profiles. When actual schedules are unknown, default diversity factors will be typically obtained from codes and standards. A limitation of these standardized profiles is that they are not specific to the buildings being modeled, instead, they have been verified to differ substantially from actual diversity profiles. By integrating the occupancy model into IoT for the control of air-conditioning and lighting systems, the predictive control and energy consumption simulation of intelligent buildings will be realized.

The number of occupants in the building group is large, and the change rule of the number of occupants in different buildings varies a lot, such as hospital buildings and campus buildings. The monitoring of buildings with large numbers of occupants should be carried out from two aspects. One is to monitor individuals. Wi-Fi signals in existing buildings and new buildings are generally used to monitor the smartphones held by users. The second is to install a counter at the entrance and exit of each building and take the building as the unit to count the number of occupants at different time points. The occupancy model is easy to establish and the number of occupants controlled by the air conditioning system can be easily obtained, therefore, the statistical method may be more suitable for the prediction of occupancy.

7 Conclusions

This article provides a comprehensive review of previous occupancy data collection methods and modeling techniques, and it provides a theoretical basis for their better application in different building design and operation stages. The conclusions are as follows.

(1) Through the summary and analysis of the occupancy data collection methods, six common sensor types were obtained: PIR, RF, camera, CO_2 sensor, radio technology, and sensor combination. Based on evaluation criteria (cost, detection accuracy, and privacy violation), the different types of sensors were compared and discussed, and the applicability of different sensors was determined. Among sensor types, traditional sensors, such as PIR sensors, CO_2 sensors, and cameras, are widely used but have lower accuracy. However, such new sensors as Wi-Fi, Bluetooth, and GPS have great advantages in terms of deployment cost and detection accuracy, and they are predicted to be the main development trend in the future. In terms of data collection quality, a combination of sensors works best. There are two common methods. The first is PIR sensors, CO_2 sensors, and cameras, which are suitable for office building detection. The second is PIR sensors, environmental sensors, and wireless networks, which are suitable for professional research laboratories and have high prediction accuracy. In future data collection, the number of data should be expanded to establish a standard database suitable for various types of building modeling and verification.

(2) Three main methods were obtained by reviewing the occupancy model: the deterministic schedule method, the stochastic model, and machine learning. The deterministic schedule model is simple, and it is only suitable for situations where the prediction accuracy is not high. The stochastic model can basically reproduce the situation of occupants in a room, and the prediction accuracy can reach approximately 90%. It is suitable for occupancy prediction in office buildings and residential buildings. It can achieve good energy-saving effects for lighting and HVAC control. However, the model establishment is complex, and parameter acquisition depends on long-term monitoring. Machine learning uses various intelligent algorithms, and the model prediction accuracy can be improved to 95% through automatic

correction adjustment. In the future, the establishment of the occupancy model should focus on improving applicability for multiple buildings, overcoming the limitations of individual modeling of a single building, and establishing an evaluation standard system to quantitatively describe the simulation results at the same time.

(3) The application of the occupancy model mainly has three aspects: writing in energy consumption simulation software or making it a functional model unit, HVAC and lighting control, and the basis of a load and energy consumption prediction model.

Acknowledgements

This work is supported by the Nature Science Foundation of Tianjin (No.19JCQNJC07000) and the National Nature Science Foundation of China (No. 51678396).

References

Aerts D, Minnen J, Glorieux I, et al. (2014). A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. *Building and Environment*, 75: 67–78.

Agarwal Y, Balaji B, Gupta R, et al. (2010). Occupancy-driven energy management for smart building automation. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, Zurich, Switzerland.

Ahmad M, Culp C (2006). Uncalibrated building energy simulation modeling results. *HVAC&R Research*, 12: 1141–1155.

Ahmad J, Larijani H, Emmanuel R, et al. (2021). Occupancy detection in non-residential buildings—A survey and novel privacy preserved occupancy monitoring solution. *Applied Computing and Informatics*, 17: 279–295.

Ai B, Fan Z, Gao RX (2014). Occupancy estimation for smart buildings by an auto-regressive hidden Markov model. In: Proceedings of the American Control Conference, Portland, OR, USA.

Alam AG, Rahman H, Kim JK, et al. (2017). Uncertainties in neural network model based on carbon dioxide concentration for occupancy estimation. *Journal of Mechanical Science and Technology*, 31: 2573–2580.

Amayri M, Arora A, Ploix S, et al. (2016). Estimating occupancy in heterogeneous sensor environment. *Energy and Buildings*, 129: 46–58.

An J, Yan D, Hong T, et al. (2017). A novel stochastic modeling method to simulate cooling loads in residential districts. *Applied Energy*, 206: 134–149.

Balaji B, Xu J, Nwokafor A, et al. (2013). Sentinel: Occupancy based HVAC actuation using existing Wi-Fi infrastructure within commercial buildings. In: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, Roma, Italy.

Bartley M, Plewis I (2002). Accumulated labour market disadvantage and limiting long-term illness: Data from the 1971–1991 Office for National Statistics' Longitudinal Study. *International Journal of Epidemiology*, 31: 336–341.

Benezeth Y, Laurent H, Emile B, et al. (2011). Towards a sensor for detecting human presence and characterizing activity. *Energy and Buildings*, 43: 305–314.

Brooks J, Goyal S, Subramany R, et al. (2014). An experimental investigation of occupancy-based energy-efficient control of commercial building indoor climate. In: Proceedings of the 53rd IEEE Conference on Decision and Control, Los Angeles, CA, USA.

Buttitta G, Turner WJN, Neu O, et al. (2019). Development of occupancy-integrated archetypes: Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes. *Energy and Buildings*, 198: 84–99.

Buttitta G, Finn DP (2020). A high-temporal resolution residential building occupancy model to generate high-temporal resolution heating load profiles of occupancy-integrated archetypes. *Energy and Buildings*, 206: 109577.

Calì D, Matthes P, Huchtemann K, et al. (2015). CO_2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. *Building and Environment*, 86: 39–49.

Candanedo LM, Feldheim V (2016). Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models. *Energy and Buildings*, 112: 28–39.

Candanedo LM, Feldheim V, Deramaix D (2017). A methodology based on hidden Markov models for occupancy detection and a case study in a low energy residential building. *Energy and Buildings*, 148: 327–341.

Castanedo F, López-De-ipiña D, Aghajan H, et al. (2011). Building an occupancy model from sensor networks in office environments. In: Proceedings of the 2011 5th ACM/IEEE International Conference on Distributed Smart Cameras, Ghent, Belgium.

Chang WK, Hong T (2013). Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data. *Building Simulation*, 6: 23–32.

Chen Z, Soh YC (2014). Modeling building occupancy using a novel inhomogeneous Markov chain approach. In: Proceedings of the IEEE International Conference on Automation Science and Engineering, New Taipei, Taiwan, China.

Chen Z, Xu J, Soh YC (2015). Modeling regular occupancy in commercial buildings using stochastic models. *Energy and Buildings*, 103: 216–223.

Chen Z, Masood MK, Soh YC (2016). A fusion framework for occupancy estimation in office buildings based on environmental sensor data. *Energy and Buildings*, 133: 790–798.

Chen Y, Liang X, Hong T, et al. (2017a). Simulation and visualization of energy-related occupant behavior in office buildings. *Building Simulation*, 10: 785–798.

Chen Z, Zhao R, Zhu Q, et al. (2017b). Building occupancy estimation with environmental sensors via CDBLSTM. *IEEE Transactions on Industrial Electronics*, 64: 9549–9559.

Chen Z, Wang Y, Liu H (2018a). Unobtrusive sensor-based occupancy facing direction detection and tracking using advanced machine learning algorithms. *IEEE Sensors Journal*, 18: 6360–6368.

Chen Y, Hong T, Luo X (2018b). An agent-based stochastic Occupancy Simulator. *Building Simulation*, 11: 37–49.

Chen Z, Jiang C, Xie L (2018c). Building occupancy estimation and detection: A review. *Energy and Buildings*, 169: 260–270.

Chiou Y, Carley KM, Davidson CI, et al. (2011). A high spatial resolution residential energy model based on American Time Use Survey data and the bootstrap sampling method. *Energy and Buildings*, 43: 3528–3538.

Conte G, de Marchi M, Nacci AA, et al. (2014). BlueSentinel: A first approach using iBeacon for an energy efficient occupancy detection system. In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, Memphis, TN, USA.

Corna A, Fontana L, Nacci AA, et al. (2015). Occupancy detection via iBeacon on Android devices for smart building management. In: Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition, Grenoble, France.

Cross JE, Shelley TO, Mayer AP (2017). Putting the green into corrections: Improving energy conservation, building function, safety and occupant well-being in an American correctional facility. *Energy Research & Social Science*, 32: 149–163.

D'Oca S, Hong T (2015). Occupancy schedules learning process through a data mining framework. *Energy and Buildings*, 88: 395–408.

Daim TU, Rueda G, Martin H, et al. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, 73: 981–1012.

Davis JA III, Nutter DW (2010). Occupancy diversity factors for common university building types. *Energy and Buildings*, 42: 1543–1551.

Delaney DT, O'Hare GMP, Ruzzelli AG (2009). Evaluation of energy-efficiency in lighting systems using sensor networks. In: Proceedings of the 1st ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, Berkeley, CA, USA.

Depatla S, Muralidharan A, Mostofi Y (2015). Occupancy estimation using only WiFi power measurements. *IEEE Journal on Selected Areas in Communications*, 33: 1381–1393.

Diao L, Sun Y, Chen Z, et al. (2017). Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation. *Energy and Buildings*, 147: 47–66.

Dobbs JR, Hencey BM (2014). Predictive HVAC control using a Markov occupancy model. In: Proceedings of the 2014 American Control Conference, Portland, OR, USA.

Dodier RH, Henze GP, Tiller DK, et al. (2006). Building occupancy detection through sensor belief networks. *Energy and Buildings*, 38: 1033–1043.

Dong B, Andrews B, Lam KP, et al. (2010). An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network. *Energy and Buildings*, 42: 1038–1046.

Dong B, Lam KP (2011). Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network. *Journal of Building Performance Simulation*, 4: 359–369.

Dong B, Lam KP (2014). A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. *Building Simulation*, 7: 89–106.

Duarte C, van den Wymelenberg K, Rieger C (2013). Revealing occupancy patterns in an office building through the use of occupancy sensor data. *Energy and Buildings*, 67: 587–595.

Ebadat A, Bottegal G, Varagnolo D, et al. (2013). Estimation of building occupancy levels through environmental signals deconvolution. In: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, Roma, Italy.

Ebadat A, Bottegal G, Molinari M, et al. (2015a). Multi-room occupancy estimation through adaptive gray-box models. In: Proceedings of the 2015 54th IEEE Conference on Decision and Control, Osaka, Japan.

Ebadat A, Bottegal G, Varagnolo D, et al. (2015b). Regularized deconvolution-based approaches for estimating room occupancies. *IEEE Transactions on Automation Science and Engineering*, 12: 1157–1168.

Ekwevugbe T, Brown N, Fan D (2012). A design model for building occupancy detection using sensor fusion. In: Proceedings of the 2012 6th IEEE International Conference on Digital Ecosystems and Technologies, Campione d'Italia, Italy.

Erickson VL, Lin Y, Kamthe A, et al. (2009). Energy efficient building environment control strategies using real-time occupancy measurements. In: Proceedings of the 1st ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, Berkeley, CA, USA.

Erickson VL, Cerpa AE (2010). Occupancy based demand response HVAC control strategy. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, Zurich, Switzerland.

Erickson VL, Carreira-Perpiñán MÁ, Cerpa AE (2011). OBSERVE: Occupancy-based system for efficient reduction of HVAC energy. In: Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks, Chicago, IL, USA.

Erickson VL, Achleitner S, Cerpa AE (2013). POEM: Power-efficient occupancy-based energy management system. In: Proceedings of the 12th International Conference on Information Processing in Sensor Networks, Philadelphia, PA, USA.

Fang JS, Hao Q, Brady DJ, et al. (2006). Real-time human identification using a pyroelectric infrared detector array and hidden Markov models. *Optics Express*, 14: 6643–6658.

Feng X, Yan D, Hong T (2015). Simulation of occupancy in buildings. *Energy and Buildings*, 87: 348–359.

Flament, M., Unbehaun, M., Zhao, H., Wang, Q., Shi, K., Shen, C., … Qian, Y. (2016). Performance Evaluation of wireless Optical communication for BAN scenario with blocking Effects. Proc of the ICRA 2010 Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation, 16(December), 319-324. https://doi.org/10.1109/MNET.2016.1500236RP

Gaetani I, Hoes PJ, Hensen JLM (2016). Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy. *Energy and Buildings*, 121: 188–204.

Goyal S, Ingley HA, Barooah P (2012). Effect of various uncertainties on the performance of occupancy-based optimal control of HVAC zones. In: Proceedings of the 2012 IEEE 51st IEEE Conference on Decision and Control, Maui, HI, USA.

Gunay HB, O'Brien W, Beausoleil-Morrison I, et al. (2014). Coupling stochastic occupant models to building performance simulation using the discrete event system specification formalism. *Journal of Building Performance Simulation*, 7: 457–478.

Guo X, Tiller DK, Henze GP, et al. (2010). The performance of occupancy-based lighting control systems: A review. *Lighting Research & Technology*, 42: 415–431.

Guo S, Yan D, Hu S, et al. (2020). Global comparison of building energy use data within the context of climate change. *Energy and Buildings*, 226: 110362.

Hailemariam E., Goldstein R, Attar R., et al. (2011). Real-time occupancy detection using decision trees with multiple sensor types. In: Proceedings of the 2011 Symposium on Simulation for Architecture and Urban Design, Toronto, ON, Canada.

Haldi F, Robinson D (2008). On the behaviour and adaptation of office occupants. *Building and Environment*, 43: 2163–2177.

Haldi F, Robinson D (2011). Modelling occupants' personal characteristics for thermal comfort prediction. *International Journal of Biometeorology*, 55: 681–694.

Han Z, Gao RX, Fan Z (2012). Occupancy and indoor environment quality sensing for smart buildings. In: Proceedings of the 2012 IEEE International Instrumentation and Measurement Technology Conference, Graz, Austria.

Hao Q, Hu F, Xiao Y (2009). Multiple human tracking and identification with wireless distributed pyroelectric sensor systems. *IEEE Systems Journal*, 3: 428–439.

Happle G, Fonseca JA, Schlueter A (2018). A review on occupant behavior in urban building energy models. *Energy and Buildings*, 174: 276–292.

Haq MAU, Hassan MY, Abdullah H, et al. (2014). A review on lighting control technologies in commercial buildings, their performance and affecting factors. *Renewable and Sustainable Energy Reviews*, 33: 268–279.

Harris C, Cahill V (2005). Exploiting user behaviour for context-aware power management. In: Proceedings of the WiMob'2005, IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Montreal, QC, Canada.

Hobson BW, Lowcay D, Gunay HB, et al. (2019). Opportunistic occupancy-count estimation using sensor fusion: A case study. *Building and Environment*, 159: 106154.

Hong T, D'Oca S, Turner WJN, et al. (2015). An ontology to represent energy-related occupant behavior in buildings. Part I: Introduction to the DNAs framework. *Building and Environment*, 92: 764–777.

Hong T, Sun H, Chen Y, et al. (2016a). An occupant behavior modeling tool for co-simulation. *Energy and Buildings*, 117: 272–281.

Hong T, Taylor-Lange SC, D'Oca S, et al. (2016b). Advances in research and applications of energy-related occupant behavior in buildings. *Energy and Buildings*, 116: 694–702.

Hong T, Chen Y, Belafi Z, et al. (2018). Occupant behavior models: A critical review of implementation and representation approaches in building performance simulation programs. *Building Simulation*, 11: 1–14.

Hu S, Yan D, An J, et al. (2019). Investigation and analysis of Chinese residential building occupancy with large-scale questionnaire surveys. *Energy and Buildings*, 193: 289–304.

Hughes RF, Dhannu SS (2008). Substantial energy savings through adaptive lighting. In: Proceedings of the 2008 IEEE Canada Electric Power Conference, Vancouver, BC, Canada.

Ito S, Nishi H (2012). Estimation of the number of people under controlled ventilation using a CO_2 concentration sensor. In: Proceedings of the IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, Montreal, QC, Canada.

Javed A, Larijani H, Ahmadinia A, et al. (2015). Experimental testing of a random neural network smart controller using a single zone test chamber. *IET Networks*, 4: 350–358.

Javed A, Larijani H, Ahmadinia A, et al. (2017a). Design and implementation of a cloud enabled random neural network-based decentralized smart controller with intelligent sensor nodes for HVAC. *IEEE Internet of Things Journal*, 4: 393–403.

Javed A, Larijani H, Ahmadinia A, et al. (2017b). Smart random neural network controller for HVAC using cloud computing technology. *IEEE Transactions on Industrial Informatics*, 13: 351–360.

Jiang C, Masood MK, Soh YC, et al. (2016). Indoor occupancy estimation from carbon dioxide concentration. *Energy and Buildings*, 131: 132–141.

Jin M, Jia R, Kang Z, et al. (2014). PresenceSense: Zero-training algorithm for individual presence detection based on power monitoring. In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, Memphis, Tennessee, USA.

Jin M, Bekiaris-Liberis N, Weekly K, et al. (2018). Occupancy detection via environmental sensing. *IEEE Transactions on Automation Science and Engineering*, 15: 443–455.

Khan DS, Kolarik J, Anker Hviid C, et al. (2021). Method for long-term mapping of occupancy patterns in open-plan and single office spaces by using passive-infrared (PIR) sensors mounted below desks. *Energy and Buildings*, 230: 110534.

Khoury HM, Kamat VR (2009). Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, 18: 444–457.

Kwok SSK, Yuen RKK, Lee EWM (2011). An intelligent approach to assessing the effect of building occupancy on building cooling load prediction. *Building and Environment*, 46: 1681–1690.

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

Li N, Calis G, Becerik-Gerber B (2012). Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations. *Automation in Construction*, 24: 89–99.

Li H, Hong T, Sofos M (2019). An inverse approach to solving zone air infiltration rate and people count using indoor environmental sensor data. *Energy and Buildings*, 198: 228–242.

Lin SF, Chen JY, Chao HX (2001). Estimation of number of people in crowded scenes using perspective transformation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 31: 645–654.

Liu D, Guan X, Du Y, et al. (2013). Measuring indoor occupancy in intelligent buildings using the fusion of vision sensors. *Measurement Science and Technology*, 24: 074023.

Liu T, Li Y, Bai Z, et al. (2016a). Two-stage structured learning approach for stable occupancy detection. In: Proceedings of the 2016 International Joint Conference on Neural Networks, Vancouver, BC, Canada.

Liu P, Nguang SK, Partridge A (2016b). Occupancy inference using pyroelectric infrared sensors through hidden Markov models. *IEEE Sensors Journal*, 16: 1062–1068.

Liu Z, Zhang J, Geng L (2017a). An intelligent building occupancy detection system based on sparse auto-encoder. In: Proceedings of the 2017 IEEE Winter Applications of Computer Vision Workshops, Santa Rosa, CA, USA.

Liu H, Wang Y, Wang K, et al. (2017b). Turning a pyroelectric infrared motion sensor into a high-accuracy presence detector by using a narrow semi-transparent chopper. *Applied Physics Letters*, 111: 243901.

Lu J, Sookoor T, Srinivasan V, et al. (2010). The smart thermostat: Using occupancy sensors to save energy in homes. In: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, Zurich, Switzerland.

Lu X, Wen H, Zou H, et al. (2016). Robust occupancy inference with commodity WiFi. In: Proceedings of the 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications, New York, USA.

Lu J, Zhang T, Hu F, et al. (2017). Preprocessing design in pyroelectric infrared sensor-based human-tracking system: On sensor selection and calibration. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47: 263–275.

Luo X, Lam KP, Chen Y, et al. (2017). Performance evaluation of an agent-based occupancy simulation model. *Building and Environment*, 115: 42–53.

Manzoor F, Linton D, Loughlin M (2012). Occupancy monitoring using passive RFID technology for efficient building lighting control. In: Proceedings of the 2012 4th International EURASIP Workshop on RFID Technology, Turin, Italy.

Masood MK, Soh YC, Chang VWC (2015). Real-time occupancy estimation using environmental parameters. In: Proceedings of the 2015 International Joint Conference on Neural Networks, Killarney, Ireland.

Masood MK, Soh YC, Jiang C (2017). Occupancy estimation from environmental parameters using wrapper and hybrid feature selection. *Applied Soft Computing*, 60: 482–494.

McKenna E, Krawczynski M, Thomson M (2015). Four-state domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 96: 30–39.

Menezes AC, Cripps A, Bouchlaghem D, et al. (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97: 355–364.

Meyn S, Surana A, Lin Y, et al. (2009). A sensor-utility-network method for estimation of occupancy in buildings. In: Proceedings of the 48h IEEE Conference on Decision and Control, Shanghai, China.

Mikkilineni AK, Dong J, Kuruganti T, et al. (2019). A novel occupancy detection solution using low-power IR-FPA based wireless occupancy sensor. *Energy and Buildings*, 192: 63–74.

Milenkovic M, Amft O (2013). Recognizing energy-related activities using sensors commonly installed in office buildings. *Procedia Computer Science*, 19: 669–677.

Nagy Z, Yong FY, Frei M, et al. (2015). Occupant centered lighting control for comfort and energy efficient building operation. *Energy and Buildings*, 94: 100–108.

Naser A, Lotfi A, Zhong J (2021). Adaptive thermal sensor array placement for human segmentation and occupancy estimation. *IEEE Sensors Journal*, 21: 1993–2002.

Nesa N, Banerjee I (2017). IoT-based sensor data fusion for occupancy sensing using dempster–shafer evidence theory for smart buildings. *IEEE Internet of Things Journal*, 4: 1563–1570.

Newsham GR, Xue H, Arsenault C, et al. (2017). Testing the accuracy of low-cost data streams for determining single-person office occupancy and their use for energy reduction of building services. *Energy and Buildings*, 135: 137–147.

Ni LM, Liu Y, Lau YC, et al. (2004). LANDMARC: indoor location sensing using active RFID. *Wireless Networks*, 10: 701–710.

Oikonomou V, Becchis F, Steg L, et al. (2009). Energy saving and energy efficiency concepts for policy making. *Energy Policy*, 37: 4787–4796.

Page J, Robinson D, Morel N, et al. (2008). A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings*, 40: 83–98.

Park JY, Nagy Z (2018). Comprehensive analysis of the relationship between thermal comfort and building control research—A data-driven literature review. *Renewable and Sustainable Energy Reviews*, 82: 2664–2679.

Park JY, Ouf MM, Gunay B, et al. (2019). A critical review of field implementations of occupant-centric building controls. *Building and Environment*, 165: 106351.

Petersen S, Pedersen TH, Nielsen KU, et al. (2016). Establishing an image-based ground truth for validation of sensor data-based room occupancy detection. *Energy and Buildings*, 130: 787–793.

Qu D, Yang B, Gu N (2019). Indoor multiple human targets localization and tracking using thermopile sensor. *Infrared Physics & Technology*, 97: 349–359.

Raykov YP, Ozer E, Dasika G, et al. (2016). Predicting room occupancy with a single passive infrared (PIR) sensor through behavior extraction. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, USA.

Ren X, Yan D, Hong T (2015). Data mining of space heating system performance in affordable housing. *Building and Environment*, 89: 1–13.

Richardson I, Thomson M, Infield D (2008). A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 40: 1560–1566.

Salimi S, Liu Z, Hammad A (2019). Occupancy prediction model for open-plan offices using real-time location system and inhomogeneous Markov chain. *Building and Environment*, 152: 1–16.

Shen W, Newsham G, Gunay B (2017). Leveraging existing occupancy-related data for optimal control of commercial office buildings: A review. *Advanced Engineering Informatics*, 33: 230–242.

Shih HC (2014). A robust occupancy detection and tracking algorithm for the automatic monitoring and commissioning of a building. *Energy and Buildings*, 77: 270–280.

Stancil BA, Zhang C, Chen T (2008). Active multicamera networks: From rendering to surveillance. *IEEE Journal of Selected Topics in Signal Processing*, 2: 597–605.

Stazi F, Naspi F, D'Orazio M (2017). A literature review on driving factors and contextual events influencing occupants' behaviours in buildings. *Building and Environment*, 118: 40–66.

Sun Y, Haghighat F, Fung BCM (2020). A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy and Buildings*, 221: 110022.

Szczurek A, Maciejewska M, Wyłomańska A, et al. (2016). Detection of occupancy profile based on carbon dioxide concentration pattern matching. *Measurement*, 93: 265–271.

Szczurek A, Maciejewska M, Pietrucha T (2017). Occupancy determination based on time series of CO_2 concentration, temperature and relative humidity. *Energy and Buildings*, 147: 142–154.

Tabak V, de Vries B (2010). Methods for the prediction of intermediate activities by office occupants. *Building and Environment*, 45: 1366–1372.

Tesoriero R, Tebar R, Gallud JA, et al. (2010). Improving location awareness in indoor spaces using RFID technology. *Expert Systems with Applications*, 37: 894–898.

Tsanas A, Xifara A (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49: 560–567.

Tyndall A, Cardell-Oliver R, Keating A (2016). Occupancy estimation using a low-pixel count thermal imager. *IEEE Sensors Journal*, 16: 3784–3791.

Virote J, Neves-Silva R (2012). Stochastic models for building energy prediction based on occupant behavior assessment. *Energy and Buildings*, 53: 183–193.

Wahl F, Milenkovic M, Amft O (2012). A distributed PIR-based approach for estimating people count in office environments. In: Proceedings of the 2012 IEEE 15th International Conference on Computational Science and Engineering, Paphos, Cyprus.

Wang D, Federspiel CC, Rubinstein F (2005). Modeling occupancy in single person offices. *Energy and Buildings*, 37: 121–126.

Wang C, Yan D, Jiang Y (2011). A novel approach for building occupancy simulation. *Building Simulation*, 4: 149–167.

Wang Z, Ding Y (2015). An occupant-based energy consumption prediction model for office equipment. *Energy and Buildings*, 109: 12–22.

Wang Y, Shao L (2017). Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning. *Building and Environment*, 114: 106–117.

Wang W, Chen J, Song X (2017). Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference approach. *Building and Environment*, 124: 130–142.

Wang W, Chen J, Hong T (2018a). Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings. *Automation in Construction*, 94: 233–243.

Wang W, Chen J, Hong T, et al. (2018b). Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology. *Building and Environment*, 138: 160–170.

Wang W, Hong T, Li N, et al. (2019a). Linking energy-cyber-physical systems with occupancy prediction and interpretation through WiFi probe-based ensemble classification. *Applied Energy*, 236: 55–69.

Wang W, Hong T, Xu N, et al. (2019b). Cross-source sensing data fusion for building occupancy prediction with adaptive lasso feature filtering. *Building and Environment*, 162: 106280.

Wang J, Tse NCF, Chan JYC (2019c). Wi-Fi based occupancy detection in a complex indoor space under discontinuous wireless communication: A robust filtering based on event-triggered updating. *Building and Environment*, 151: 228–239.

Wang Z, Hong T, Piette MA (2019d). Predicting plug loads with occupant count data through a deep learning approach. *Energy*, 181: 29–42.

Wang Z, Hong T, Piette MA, et al. (2019e). Inferring occupant counts from Wi-Fi data in buildings through machine learning. *Building and Environment*, 158: 281–294.

Weekly K, Bekiaris-Liberis N, Jin M, et al. (2015). Modeling and estimation of the humans' effect on the CO_2 dynamics inside a conference room. *IEEE Transactions on Control Systems Technology*, 23: 1770–1781.

Widén J, Nilsson AM, Wäckelgård E (2009). A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand. *Energy and Buildings*, 41: 1001–1012.

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

Wu L, Wang Y (2019). A low-power electric-mechanical driving approach for true occupancy detection using a shuttered passive infrared sensor. *IEEE Sensors Journal*, 19: 47–57.

Wu L, Gou F, Wu ST, et al. (2020). SLEEPIR: Synchronized low-energy electronically chopped PIR sensor for true presence detection. *IEEE Sensors Letters*, 4: 1–4.

Xi W, Zhao J, Li X, et al. (2014). Electronic frog eye: Counting crowd using WiFi. In: Proceedings of the IEEE INFOCOM 2014 - IEEE Conference on Computer Communications, Toronto, ON, Canada.

Yang Z, Li N, Becerik-Gerber B, et al. (2012). A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations. *Simulation Series*, 44: 100–107.

Yang Z, Becerik-Gerber B (2014). Modeling personalized occupancy profiles for representing long term patterns by using ambient context. *Building and Environment*, 78: 23–35.

Yang J, Santamouris M, Lee SE (2016). Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. *Energy and Buildings*, 121: 344–349.

Yoshino H, Hong T, Nord N (2017). IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods. *Energy and Buildings*, 152: 124–136.

Yu Z, Haghighat F, Fung BCM, et al. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, 42: 1637–1646.

Yu T (2010). Modeling occupancy behavior for energy efficiency and occupants comfort management in intelligent buildings. In: Proceedings of the 2010 9th International Conference on Machine Learning and Applications, Washington D.C. USA.

Yu Z, Fung BCM, Haghighat F, et al. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings*, 43: 1409–1417.

Yu Z, Haghighat F, Fung BCM, et al. (2012). A novel methodology for knowledge discovery through mining associations between building operational data. *Energy and Buildings*, 47: 430–440.

Zhang R, Lam KP, Chiou Y, et al. (2012). Information-theoretic environment features selection for occupancy detection in open office spaces. *Building Simulation*, 5: 179–188.

Zhang Y, Bai X, Mills FP, et al. (2018). Rethinking the role of occupant behavior in building energy performance: A review. *Energy and Buildings*, 172: 279–294.

Zhao Y, Zeiler W, Boxem G, et al. (2015). Virtual occupancy sensors for real-time occupancy information in buildings. *Building and Environment*, 93: 9–20.

Zhao Z, Kuendig S, Carrera J, et al. (2017). Indoor location for smart environments with wireless sensor and actuator networks. In: Proceedings of the 2017 IEEE 42nd Conference on Local Computer Networks, Singapore.

Zhen Z, Jia Q, Song C, et al. (2008). An indoor localization algorithm for lighting control using RFID. In: Proceedings of the 2008 IEEE Energy 2030 Conference, Atlanta, GA, USA.

Zhou H, Qiao L, Jiang Y, et al. (2016). Recognition of air-conditioner operation from indoor air temperature and relative humidity by a data mining approach. *Energy and Buildings*, 111: 233–241.

Zhu Q, Chen Z, Masood MK, et al. (2017). Occupancy estimation with environmental sensing via non-iterative LRF feature learning in time and frequency domains. *Energy and Buildings*, 141: 125–133.

Zou J, Zhao Q, Yang W, et al. (2017). Occupancy detection in the office by analyzing surveillance videos and its application to building energy conservation. *Energy and Buildings*, 152: 385–398.

Zou PXW, Xu X, Sanjayan J, et al. (2018). Review of 10 years research on building energy performance gap: Life-cycle and stakeholder perspectives. *Energy and Buildings*, 178: 165–181.

Zuraimi MS, Pantazaras A, Chaturvedi KA, et al. (2017). Predicting occupancy counts using physical and statistical Co2-based modeling methodologies. *Building and Environment*, 123: 517–528.