Title

Linking energy-cyber-physical systems with occupancy prediction and interpretation through WiFi probe-based ensemble classification

Authors

Wang, Wei
Hong, Tianzhen
Li, Nan
et al.

Publication Date

2019-02-01

DOI

10.1016/j.apenergy.2018.11.079

Peer reviewed

Linking energy-cyber-physical systems with occupancy predication and interpretation through WiFi probe-based ensemble classification

Wei Wang [a], Tianzhen Hong [b], Nan Li [c], Ryan Qi Wang [d], and Jiayu Chen [e1]

[a] *School of Architecture, Southeast University, 210018, Nanjing, Jiangsu, China*

[b] *Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, Cyclotron Road, Berkeley, CA 94720, USA*

[c] *Department of Construction Management, Tsinghua University, 100084, Beijing, China*

[d] *Department of Civil and Environment Engineering, Northeastern University, 433 SN, 360 Huntington Avenue, Boston, MA 02115, USA*

[e] *Department of Architecture and Civil Engineering, City University of Hong Kong, Y6621, AC1, Tat Chee Ave, Kowloon, Hong Kong*

Abstract:

With rapid advances in sensing and digital technologies, cyber-physical systems are regarded as the most prominent platforms to improve building design and management. Researchers investigated the possibility of integrating energy management system with cyber-physical systems as energy-cyber-physical systems to promote building energy management. However, minimizing energy consumption while fulfilling building functions for energy-cyber-physical systems is challenging due to the dynamics of building occupants. As occupant behavior is one major source of uncertainties for energy management, ignoring it often results in energy wastes caused by overheating and overcooling as well as discomfort due to insufficient thermal and ventilation services. To mitigate such uncertainties, this study proposed an occupancy linked energy-cyber-physical system that incorporates WiFi probe-based occupancy detection. The proposed framework utilized ensemble classification algorithms to extract three types of occupancy information. It creates a data interface to link energy management system and cyber-physical systems and allows automated occupancy detection and interpretation through assembling multiple weak classifiers for WiFi signals. A validation experiment in a large office room was conducted to

---

[1]*Corresponding author. tel.: +852 3442 4696; fax: +852 3442 0427.

e-mail addresses: jiaychen@cityu.edu.hk (Jiayu Chen)

examine the performance of the proposed occupancy linked energy-cyber-physical systems. The experiment and simulation results suggest that, with a proper classifier and occupancy type, the proposed model can potentially save about 26.4% of energy consumption from the cooling and ventilation demands.

Keywords: Energy-Cyber-Physical Systems, Building occupancy, Wi-Fi probe technology, ensemble algorithm

Nomenclatures

TPM	Transition probability matrix of one occupant x_k	G_{other}	Load from other potential sources
x_k^{i-o}, x_k^{i-i}	Probability that occupancy status transfers from “in” to “in” or “out”	Q_r	Load of room r
N_{i-i}, N_{i-o}	Frequency that occupancy status transfers from “in” to “in” or “out”	E_r	Energy cost to satisfy the cooling load at room r
x_k^{Mac}	MAC address of occupancy x_k	m_r	Total supply air flow rate
$X(t)$	Input feature vector at time *t*	T_s	Supply air temperature
Y	Actual occupancy vector	$m_{OA,r}$	Outdoor air flow rate of room r
$F(x)$	Ensemble occupancy algorithm function	R_p	Outdoor air requirement for each occupant
$f_m(x)$	Meta occupancy algorithm function *m*	P_r	Total number of occupants
w_m	Weight value of function *m*	R_a	Outdoor air requirement for per area
L	Loss function	A_r	Total floor area of room r
Q_{nor}	Non-occupant-related load	$E_{ven,r}$	Energy use for ventilation of room r
$Q_{¿}$	Occupant-related load	$Q_{vent,r}$	Ventilation load of room r
$Q_{inf,r}$	Heat gains from infiltration of room r	h_{OA}, h	Enthalpy value of outdoor and room air
$Q_{surf,r}$	Heat gains from surface of room r	$p_{pred.A}^r$	Prediction value of occupancy type A
$m_{inf,r}$	Flow rate of the infiltration air	t_0	Time resolution of the occupancy
C_p	Specific heat capacity of air	T	Length of the averaging time window
$T_{¿,r}$	Temperature of room r	TP	Number of true positives
T_{air}	Temperature of outdoor air	TN	Number of true negatives
$A_{surf,r}$	Surface area of room r	FP	Number of false positives
K_{surf}	Heat transfer coefficient of surface	FN	Number of false negatives
G_p	H eat gain from per occupant	BM	Baseline model
G_{eq}	Load from equipment	OLEM	Occupancy-linked e-CPS model

1. INTRODUCTION

Buildings consume more than 40% of primary energy among all energy-consuming sectors [1] and energy bills become the largest overhead in building maintenance and operation budget. An increasing number of building owners and decision makers recognize promoting building energy efficiency as the most cost-effective approach for conservation. In modern buildings, the majority of energy is consumed by the mechanical/facility systems, which consists of heating, ventilation, air-conditioning (HVAC), lighting, water, safety, and similar allied subsystems. However, promoting energy efficiency of these facility systems is extremely challenging, as they usually have to comply with complicated working conditions, comfort requirements, and dynamic energy demand. In recent years, researchers propose to integrate both physical building systems with engineered cyber models so that building systems can be monitored, coordinated, controlled, optimized with a computing and communication core [2]. The integrated system is able to model, visualize, and operate complex building systems with various computing tools, and such systems are called cyber-physical systems (CPSs). With advances in the sensors, sensor networks, and embedded computing systems, CPSs unlocked the potential of optimizing building energy systems, such consolidated system is called energy-cyber-physical systems (e-CPSs) [3]. The ideal e-CPSs are designed to reduce the power demand though computational optimization so that the demand can be satisfied by the available power with minimum waste [3]. In this context, strategies were developed to optimize building facility operation through frequency control, voltage control, or sleep state scheduling [4]. However, dynamic demand caused by occupants and distributed operation cause poor system coordination in the centralized control system [5]. The physical facility systems require the computational outcomes from cyber model to optimize their operation, but the biggest challenge is the unreliable and incorrect demand estimation, which often results in energy wastes or unsatisfied thermal comfort.

Therefore, a well-integrated e-CPSs should ensure reliability of demand information, which is usually captured by the physical system. With accurate and meaningful data inputs, the cyber model can provide effective operation suggestions. However, a building's energy demand is mainly generated by occupants' thermal, lighting, and functional requirements, which are extremely dynamic and difficult to be captured by the physical building system. Conventional e-CPSs can synchronize physical mechanical and energy management systems with digital models, but they lack the ability to response to uncertain demand of occupants. Due to this constraint, in practice, conventional e-CPSs usually are rigid and static systems that based on certain assumed operation schedules. To fill this research gap, this study proposes to implement structured occupancy information to bridge the cyber and physical systems and form a new occupancy linked e-CPSs. Such system incorporates WiFi probe technology and interpreters that are based on ensemble Wi-Fi signals classifiers. The WiFi probe infrastructure on the physical model side and the ensemble signal classifiers on the cyber model side can be integrated and bridged by the accurate and reliable occupancy estimation. With such occupancy information, accurate demand can be estimated and the facility operation can be optimized for the energy saving purpose.

The rest of the paper is organized as follows. Section 2 reviews related works, including energy-cyber-physical systems (e-CPSs) studies and buildings. Section 3 introduces the framework and quantitative occupancy linked e-CPSs. Section 4 describes the validation experiment. Section 5 presents the results of experiment and simulation. Section 6 discusses the implication and limitation of this study, and Section 7 concludes this study.

2. BACKGROUND

2.1 Energy management and cyber-physical system

With the increased capability and decreased cost of wireless sensors, CPSs are capable of capture various building information through efficient networks and abundant computing powers. Thus, researches proposed to develop CPSs for building energy management systems in future smart buildings [6]. Kleissl and Agarwal looked at modern smart buildings entirely as a cyber-physical energy systems and examined the opportunities with joint optimization of energy use by occupants and information processing equipment [7]. Balaji et al. explored two case studies on smart buildings and electric vehicles to examine the feasibility of implementation of CPSs for energy management [8]. Zhao et al. developed a conceptual scheme for CPSs based energy management in buildings that combines the building energy information system, net-zero energy system, and demand-driven system [9]. Paridari et al. proposed a cyber-physical-security framework that also includes building energy management system (BEMS) with resilient policy and security analytics [10]. Based on upon these efforts, researchers concluded that e-CPSs is one of most prominent platforms in promoting building efficiency by introducing energy management into the cyber-physical interaction loop.

Current research on e-CPSs mainly focuses on framework design and data-driven control. For the framework design studies, researchers integrate building information models (BIM) [11] and energy simulation programs [12], such as Modelica [13] or EnergyPlus [14], with physical sensor networks. For example, Delwati et al. compared the design features of the demand-controlled-ventilation methods with Modelica and proposed guidelines for building ventilation designers [15]. Hong et al simulated variable refrigerant flow systems with EnergyPlus and tested the model with typical houses in California [16]. Grigore et al. studied a case of deploying an e-CPSs for thermal optimization through electrical load monitoring, forecasting, HVAC control, and smart grid integration [17]. Behl et al. proposed an open source e-CPSs, DR-Advisor, which also allows data-driven modeling and control with rule-based algorithms. Based on a comparison with DOE commercial reference buildings, their system showed a 17% energy saving [18]. For the data-driven thermal control studies, researchers focus on converting physically captured data to system operation schedule and settings. For example, Ferreira et al. utilized neural network to implement predictict control to imporve thermal comfort in public buildings [19]. Costanzo et al.employed reinforcement learning tool to develop data-driven control for heating systems [20].

As the premise of effective e-CPSs is to ensure human-centric services (e.g. thermal comfort, visual comfort) while saving as much as possible energy, researchers recognized that occupancy information played a central role to guarantee the e-CPSs' performance in smart buildings [21]. Latest studies suggest that accurate occupancy information not only links the physical building systems and cyber models but also mitigates the discrepancies between the designed/simulated and the actual building operation performance [22]. Menezes et al. conducted a comprehensive study on the non-domestic buildings and concluded that occupancy information is significant to building energy and occupancy comfort benchmarking [23]. Liang et al. also stated occupancy data should be included to improve accuracy of building energy use predicting since occupancy is highly correlated with energy use and thermal comfort [24]. Wang et al. applied neural networks and WiFi technology to predict occupancy and integrate it to efficient building HVAC control and save 20% energy through avoiding overheating and overcooling [25]. Barbeito et al. assessed occupant thermal comfort and energy efficiency in buildings using statistical quality control (SQC) with integrated big data web energy platform [26]. Zhang et al optimized ventilation systems to satisfy occupant thermal comfort and saved 7.8% of total energy consumption [27]. Korkas et al. proposed a study of matching energy generation and consumption with occupant behavior to guarantee occupant thermal comfort and developing demand response in microgrids with renewable energy sources [28]. Chen et al. applied occupant feedback based model predictive control (MPC) for thermal comfort and energy optimization and proposed a novel dynamic thermal sensation model, saving 25% of energy use while maintaining thermal comfort level [29]. Lim et al. discussed occupant visual comfort in office spaces based on occupants' behaviors and reported 33.39% of lighting energy saving [30]. Shen et al. integrated lighting control strategies with occupancy state to guarantee visual comfort and resulted in a 48.8% saving [31].

2.2 e-CPS and occupancy information

Usable and efficient building cyber models require a good understanding of occupants' energy demand and meaningful inputs from physical building systems [32,33]. Many studies suggested that the actual energy consumption of physical buildings severely deviates from the estimations of cyber models due to incorrect estimation of occupancy behavior [34]. Significant discrepancies between actual and estimated energy performance have been observed due to the complicated interrelationship between occupancy and building facility operation and the uncertainty of human behavior [35]. Oldewirtel et al. investigates the potential of using occupancy information to realize a more energy efficient building climate control and in the simulations with alternating occupancy, the savings are in the range of 50% of the savings with homogeneous occupancy [36]. Hong et al. discussed ten questions concerning occupant behavior and building energy performance [37]. The International Energy Agency (IEA) Energy in Building and Community (EBC) Programme Annex 66 also highlighted and concluded that occupancy and occupants' behaviors are the most significant role for various research of enhancing building performance and human-centric services [38]. However, both physical building and cyber model are seldom changed in CPSs after the building has been built and the system uncertainties mainly arise from dynamic occupants' behavior and weather conditions. Many studies concluded that the occupancy information is one of the most significant considerations in energy conservation or low energy building design [39,40]. Therefore, as occupancy is the most critical data sources in energy demand estimation, e-CPSs should allow accurate and reliable occupancy information exchange between the physical system and cyber model.

Real opportunities for improving current e-CPSs exist where sensors, Information and Communication Technology (ICT), and data analytics can provide real-time occupant-related energy demand to guide building operation. Due to the complicated interrelationship of the energy consumption in building facilities and occupant behaviors [35,36,41], implementing occupancy information to improve building energy efficiency has been proven a feasible and cost-effective approach. For example, Kim et al. employed occupancy in simulation models and significantly reduced the deviated plug-load estimation [42]. Yang et al. investigated energy consumption of three institutional building in Singapore with the variability of daily occupancy and additional occupancy due to visitors [43]. Yang and Becerik-Gerber reported in their studies that the occupancy profiles-based operation schedule and room assignment can reduce 8% of HVAC energy use [44]. Pisello et al. suggested human-based energy retrofits can effectively promote energy efficiency in residential buildings with simulated post-occupancy information [45]. Chen et al. utilized occupancy information to visualize and validate the impact of occupants' behavior on commercial buildings [46].

To acquire occupancy information, researchers have proposed various methods. Jin et al. detected occupancy information through environmental sensing based on proxy measurements, such as temperature and CO_2 concentrations, and achieved 0.6044 mean squared error and 55% ventilation cost reduction [47]. Other researchers focused on using smart meters to infer occupancy presence when no data or limited data is available and reported a detection accuracy of 93% for residences and 90% for offices, respectively [48]. On the other hand, Radio frequency identification (RFID) can be applied for indoor occupant positioning, e.g. Weekly applied RFID based sampling importance resampling particle filtering algorithm for occupant positioning in a real office and achieved an accuracy of 50% estimates within 3 m range and 90% estimates within 5 m range [49]. WiFi networks are the most preferable infrastructure in existing buildings, since they are efficient, affordable, and convenient [50]. In addition, WiFi access points are usually pre-installed in most modern buildings and multiple networks can cross-reference each other. The occupants' smartphones can serve as signal receivers or tags by measuring the signal strength indicators (RSSI) and hardware addresses. Thus, with these considerations, researchers developed various WiFi-based occupancy approaches to optimize HVAC operation [51]. For example, Chen et al. showed the number of Wi-Fi connections have a positive relationship with building energy consumption [52]. Balaji utilized WiFi networks and smartphones to adjust HVAC operation setting and achieved a 17.8% electricity saving [53]. Jin et al. proposed a PresenceSense research with data collection through multiple sensing sources, including ultrasonic sensors, acceleration sensors, and WiFi [54]. Zou et al. proposed a non-intrusive occupancy sensing system, called WinOSS, to count WiFi-enabled mobile devices, which can achieve 98.85% occupancy detection accuracy when occupants stay stationary [55]. Zou et al. claimed implementing Internet of Things (IoT) technologies the counting accuracy can be as high as 99.1% [56].

Although many researchers recognized that the key of e-CPSs to promote building energy efficiency is integrating occupancy information, the interface to bridge sensing outcomes and e-CPS platform remains unfeasible. Inspired by previous researches, this study intends to develop a quantitative framework to interpret dynamic WiFi signals as useful occupancy schedules and profiles for cyber energy models. To achieve this goal, this study proposed an occupancy linked e-CPSs model (OLEM) to take advantage of existing Wi-Fi infrastructure in buildings and to incorporate ensemble classification algorithm for occupancy detection and predication. The proposed OLEM utilized three occupancy data formats as interface and WiFi probe technology toolset to bridge energy management system and CPSs.

3. METHODOLOGY

3.1 Occupancy linked e-CPSs

A fundamental e-CPSs framework includes at least a physical building system and cyber model for energy management and optimization. The physical building model reflects the actual conditions and performance of a building while the cyber model is a digital twin that can be used for various computational processes. The physical buildings usually have sensors and sensor network installed which allows acquiring the various types of environmental information, such as temperatures, CO2 concentration, and relative humidity (RH), and system operation information, such as supply/outdoor air flow rate and temperature, pump efficiency, and instantaneous energy load. The building information model is the key to associate both components and to create a dependable digital twin for the actual building. The building information model contains static features and dynamic operation settings. The static features include building materials, geometry, location, system type, and etc., while the dynamic operation settings include the operation schedule, efficiency, and settings of HVAC, lighting, and security systems.

To extend conventional e-CPSs, this study proposes to integrate dynamic occupancy information to enable data exchange between the physical building and cyber model. As the physical infrastructure of the building system, Wi-Fi networks were utilized to obtain the signal strength of occupants' device/tag. The obtained occupancy information serves as the inputs for a cyber model for data analysis and system optimization. To connect both components of e-CPSs, this study also developed an occupancy interpreter based on ensemble algorithms to convert Wi-Fi signal strengths to occupant number and schedule. Once detailed occupancy information is captured, the cyber model can conduct energy simulation with the building information model and suggest proper operational settings for the facility/mechanical systems. The Figure 1 shows the structure of the proposed occupancy linked e-CPSs.

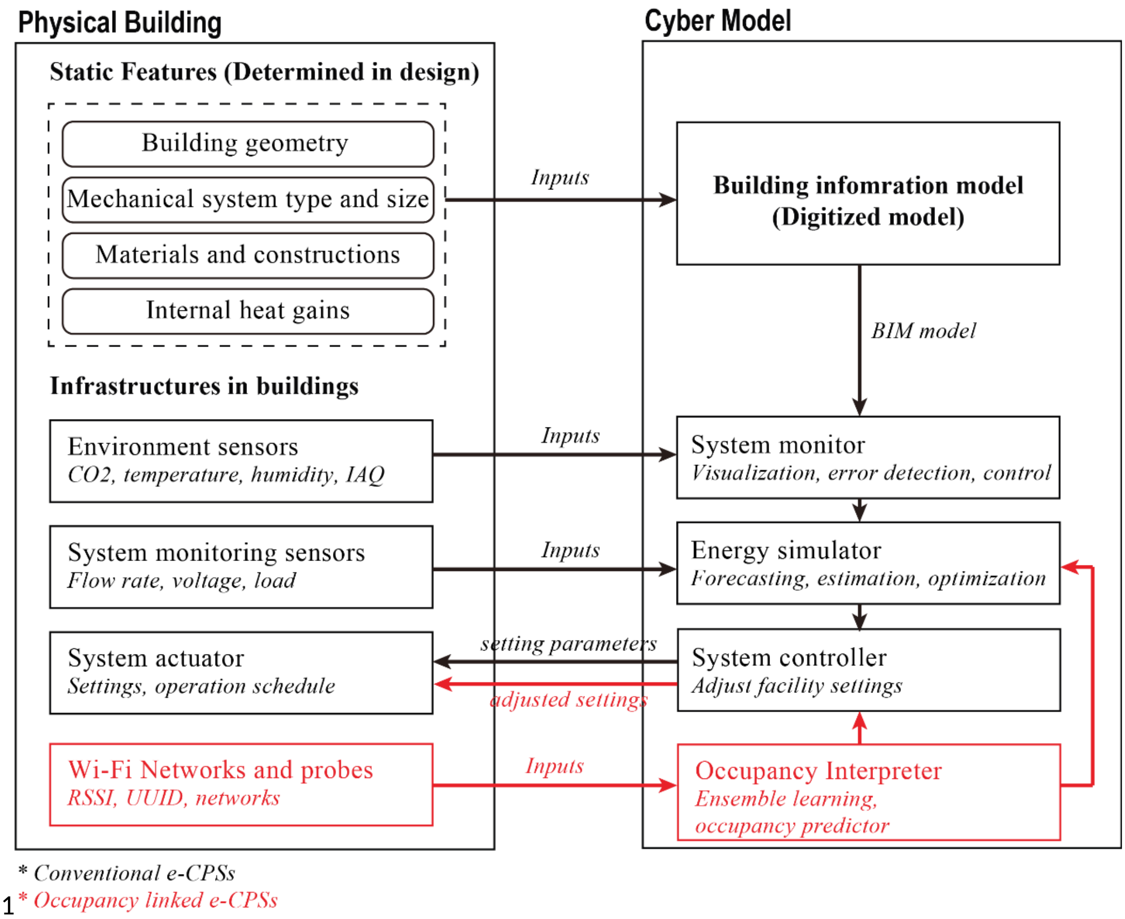

Fig. 1. The scheme of the occupancy linked e-CPSs.

3.2 Wi-Fi Probe-based ensemble learning algorithm for occupancy prediction

This study proposes to utilize Wi-Fi probes as the active detector for occupants (occupants are assumed to have a smartphone or tag with the capacity of Wi-Fi connection) and the proposed prediction algorithm implements a set of ensemble algorithms. The algorithm serves as the occupancy interpreter to convert received Wi-Fi signal strengths to the number and residency patterns of occupants and send the results as the inputs for energy simulator. The process of data interpretation includes three steps: (1) Feature extraction; (2) Ensemble learning; and (3) Occupancy pattern matching. Figure 2 shows a simplified process of the proposed algorithm.

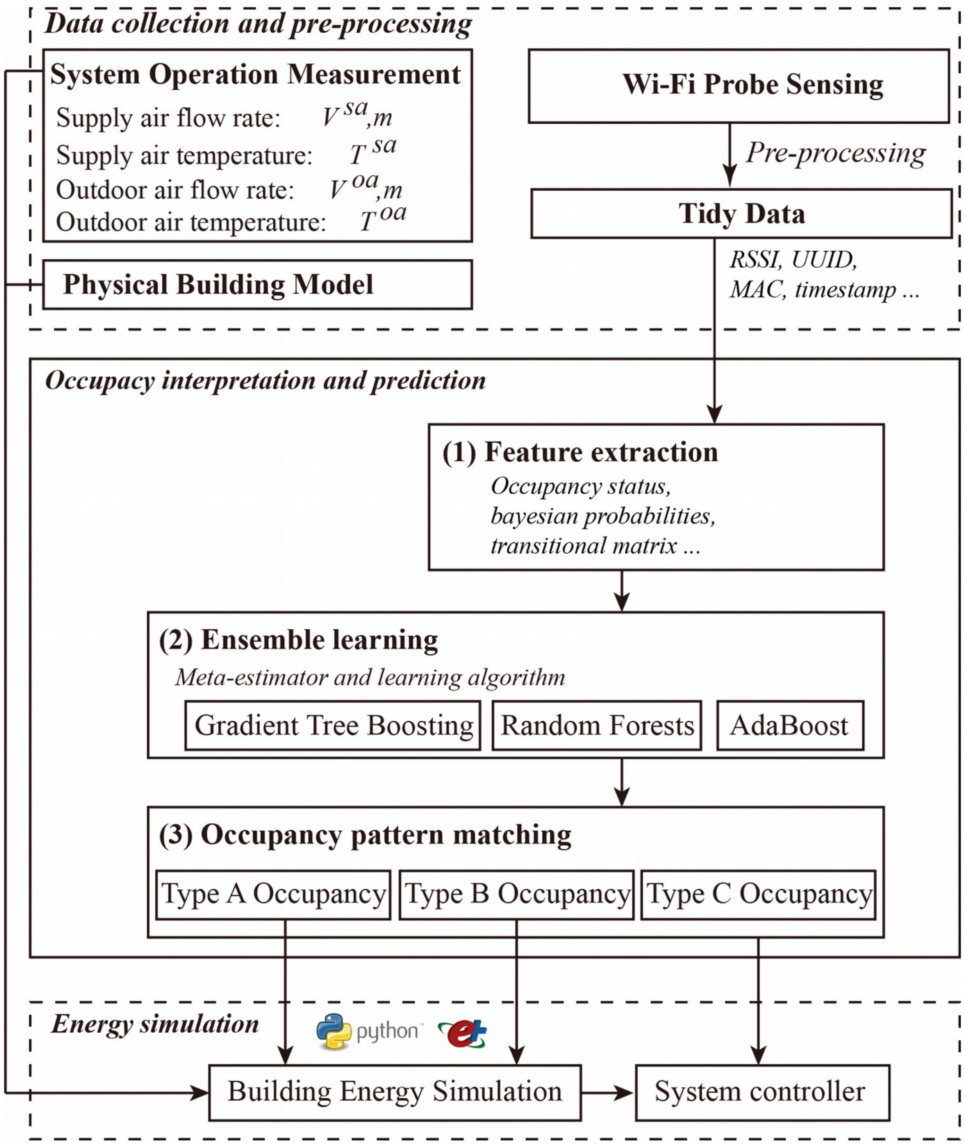

Fig. 2. The process of the Wi-Fi Probe-based ensemble learning for occupancy.

3.2.1 Feature extraction

The appearance of occupants in a building space shows a strong stochastic characteristic [57], thus, the occupancy prediction is usually modeled as a Markov process [58,59], in which current occupancy status depends on previous occupancy status. For example, the probability of an occupant leaves a space only feasible when he/she is already in the space. Therefore, the feature extraction step models an occupant status in a given space as "in" or "out" and the transfer probability and transition matrix of the Markov process can be modeled as

$$TPM|_{x_k} = \begin{bmatrix} x_k^{i-o} & x_k^{i-i} \\ x_k^{o-o} & x_k^{o-i} \end{bmatrix} \tag{1}$$

Where $TPM|_{x_k}$ represents the transition probability matrix of one occupant x_k . In the transition matrix, x_k^{i-o} and x_k^{i-i} denote the observed probability that one occupant whose status is "in" at the current time would be "out" or still "in" at the next time. x_k^{o-o} and x_k^{o-i} denote the observed probability that one occupant whose status is "out" at the current time would be "out" or "in" in the next time interval. The probability can be computed with an observed conditional probability based on Bayesian models.

$$x_k^{i-i} = P\left(observed\ state = i \middle| observed\ state = i \text{¿}\right. \tag{2}$$

Therefore, the occupied probability of one media access control (MAC) address is

$$x_k^{i-i} = \frac{\sum N_{i-i}}{\sum N_{i-i} + \sum N_{i-o}}\ x_k^{o-o} = \frac{\sum N_{o-o}}{\sum N_{o-o} + \sum N_{o-i}} \tag{3}$$

Where N_{i-i} is the frequency in which the occupancy status transfers from "in" to "in". N_{i-o} is the frequency in which the occupancy status transfers from "in" to "out". Similarly, N_{o-o} and N_{o-i} represent the frequencies in which the occupancy status transitioned from "out" to "out" and from "out" to "in", respectively. With an assigned probability for MAC addresses in the room. Each MAC address is formatted as

$$x_k = \left\{x_k^{Mac}, x_k^{o-i}, x_k^{i-i}\right\} \tag{4}$$

Then, suppose there are n occupants at one time spot t, then input feature vector at time can be as

$$X(t) = \left\{x_1^{Mac}, x_1^{o-i}, x_1^{i-i}, \ldots, x_k^{Mac}, x_k^{o-i}, x_k^{i-i}, \ldots, x_n^{Mac}, x_n^{o-i}, x_n^{i-i},\right\} \tag{5}$$

3.2.2 Ensemble learning algorithms

There are main families of ensemble methods. The first method is averaging, which builds several estimators independently and then average predictions through minimizing their prediction variance, such as Bagging methods and Forests of

Randomized Trees. The second method is boosting, which builds sequential estimators to reduce the bias by combining several weak models, such as AdaBoost and Gradient Tree Boosting. The ensembled learning algorithm in this study integrates multiple meta-estimators through boosting method.

Through feature extraction, raw data can be interpreted as an input vector of ($X(t), Y$¿ . Where Y is actual occupancy (label) as the learning object and $X(t)$ are extracted features in previous section. The ensemble learning is built upon numbers of multiple meta-estimators, which are usually simple and weak models, such as a decision tree. Decision tree uses a tree structure to create a model that predicts the value of a target variable based on several input variables. The tree can be learned by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset until the splitting no longer adds value to the predicting model. Figure 3 shows the structure of the ensembled learning for occupancy prediction. $X=\{x_1, x_2, \ldots, x_N\}$ is defined as a set of N observations of Wi-Fi dataset inputs with associated output $Y=\{y_1, y_2, \ldots, y_N\}$.

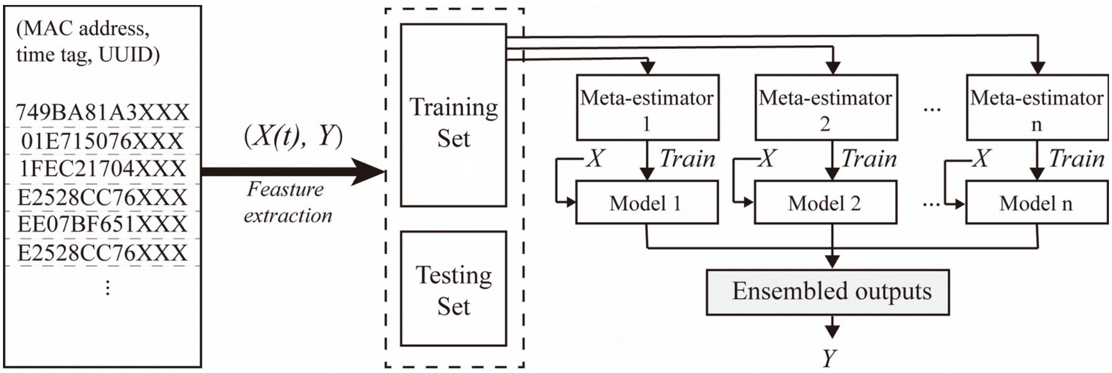

Fig. 3. The ensemble learning algorithm for occupancy prediction.

Suppose the ensembled outputs can be estimated from the aggregated results from multiple meta-estimators as:

$$F(x)=\sum_{m=1}^{M} w_m f_m(x) \tag{6}$$

Where $f_m(x)$ are the basis functions of meta-estimators. n is the index of meta-estimators and w_m is the weight parameter assigned to one meta-estimator. The iterative form of above equation can be represented as:

$$F_m(x)=F_{m-1}(x)+w_m f_m(x) \tag{7}$$

w_m is the weight of the estimators. In each iteration, the decision tree $f_m(x)$ is chosen to minimize the loss function L given the current model $F_{m-1}(x_i)$.

$$F_m(x)=F_{m-1}(x)+\arg\min_{f}\sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i)+f(x)\right) \tag{8}$$

Other than the regular decision tree, the meta-estimators can be substituted with other more complicated classifiers. This study also embedded three other ensemble algorithms (Gradient Tree Boosting classifier, Radom Forest classifier, and Adaptive Boosting classifier) in the occupancy prediction model.

(1) Gradient Tree Boosting (GTB)

Gradient Tree Boosting (GTB) classifier is a generalization of boosting to arbitrary differentiable loss functions. GTB classifier can easily handle the mixed type of data and is robust to outliers with improved loss functions. GTB attempts to solve the minimization problem numerically via steepest descent, the direction of which is the negative gradient of the loss function.

The GTB algorithm generates a model, which combines multiple simple trees in sequence. The minimum error is achieved by searching the best split of trees. The simple process of GTB can be illustrated as:

- Initial predicted value is assumed for all observation in the datasets. Error is calculated using the assumed predictions and actual datasets.
- A decision tree model is created using the errors. Split the tree branches to search the minimal error.
- Model should be updated and be used to generate new predictions. New errors can be calculated with new predictions and actual datasets.
- Repeat this process till maximum number of iterations is reached or error converges.

(2) Random Forests (RF)

Random Forests (RF) is another ensemble machine learning algorithm that follows the bagging technique. The base estimators in random forest are decision trees. Unlike bagging meta estimator, RF classifier randomly selects a set of features which are used to decide the best split from the training set. By doing this, the sample bias can be eliminated and the best split among trees can be selected. With averaging, the variance of meta-estimators can be minimized, hence yielding a better model.

The RF model create multiple trees for subsets of the whole dataset. Each tree is much smaller than that of GTB. The final classification is the aggregated results based on all trees. The minimum error is achieved by properly selecting trees for subsets. The process of a random forest algorithm can be summarized as:

- Random subsets are created from the original dataset (as bootstrapping).
- Formulate decision trees for subsets. At each node in the decision tree, only a random set of features are considered for the best split.
- An optimized decision tree model is fitted for each subset for all features.
- The final predictions of the outputs are averaged from the predictions of all decision trees.

(3) Adaptive Boosting (AdaBoost)

Adaptive Boosting (AdaBoost) classifier, one of the simplest boosting algorithms, implements multiple sequential rules (weak classifiers) on the meta-estimators. The predictions from all of the estimators are combined through a weighted majority vote (or sum) to produce the final prediction. For each successive iteration, the weights are individually modified and the learning algorithm is reapplied to the reweighted data.

The AdaBoost uses rules to classify the inputs, and the final classification is the aggregated results based on all rules. Different from RF, AdaBoost assigns unequal weights to subsets. The minimum error is achieved by properly selecting rules and subset weights. Below is a brief summary of the process of performing the AdaBoost algorithm:

- Assign equal weights to all observations in the dataset.

- Rule models are built for subsets and compute the predictions for the whole data set.
- Compute errors by comparing the predictions and actual data. Update the rule models and assign higher weights for incorrectly predicted observations.
- Repeat above steps until errors are minimized.

3.2.3 Occupancy pattern matching

Buildings consume energy to ensure the thermal comfort and indoor air quality for occupants. The energy load of a building can be categorized as non-occupant-related load ($Q_{nor}¿$ and occupant-related load $\frac{Q}{(¿¿¿)}¿$. The non-occupant-related load comes from the heat transfer across the building envelope and outside environment, which highly depends on weather conditions. The total energy load can be roughly estimated as

$$Q_{nor,r} = Q_{inf,r} + Q_{surf,r} \tag{9}$$

$$Q_{inf,r} = m_{inf,r} * C_p * \left(T_{¿,r} - T_{air}\right) \tag{10}$$

$$Q_{surf,r} = A_{surf,r} * K_{surf} * \left(T_{¿,r} - T_{air}\right) \tag{11}$$

Where $Q_{inf,r}$, $Q_{surf,r}$ are the heat gains from infiltration and surface, respectively. $m_{inf,r}$ is the flow rate of the infiltration air; C_p is the specific heat capacity of air; $T_{¿,r}$ and T_{air} are the temperature of a room and outdoor air, respectively; $A_{surf,r}$ is the surface area of a room; K_{surf} is the heat transfer coefficient.

The occupant-related load includes internal gain from occupants and equipment operated by occupants.

$$Q_{¿,r} = \sum_{P_r} G_p + \sum_{p_{eq}} G_{eq} + \sum G_{other} \tag{12}$$

$$Q_r = Q_{nor,r} + Q_{¿,r} \tag{13}$$

Where P_r is the number of occupants and G_p is the heat gain from per occupant. G_{eq} contains the load from computers, water heaters, lights etc.; p_{eq}

is the index of equipment; Q_r is the total cooling load of a room. At room level, the ventilation and air conditioning system should provide enough conditioned air to maintain proper indoor temperature and the air handling system should supply sufficient fresh air.

$$E_r = Q_r = m_r * C_p * \left(T_{¿,r} - T_{s,r}\right) \tag{14}$$

Where E_r is the energy cost to satisfy the cooling load at room level. m_r is the total supply air flow rate. T_s is the supply air temperature.

In practice, American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) standards recommends minimum ventilation approach, which requires a rough estimation on the number of occupants. The suggested ventilation amount includes both a people component (to dilute contaminants from people and their activities) and an area component (to dilute contaminants from non-occupant-related sources that are more related to floor area than occupants) [60]. Outdoor airflow required in the breathing zone of the occupied space or spaces in a zone should be computed first.

$$m_{OA,r} \geq R_p * P_r + R_a * A_r \tag{15}$$

Then,

$$E_{ven,r} = Q_{vent,r} = m_{OA,r} * \left(h_{OA} - h_{¿}\right) = m_{OA,r} * \left(f\left(T_{air}, H_{air}\right) - f\left(T_{¿,r}, H_{¿,r}\right)\right) \tag{16}$$

Where $m_{OA,r}$ is the outdoor air flow rate of a room. R_p is the outdoor air flow rate requirement for each occupant. R_a and A_r are the outdoor air flow rate requirement for per area and the total floor area of room, respectively. $E_{ven,r}$ and $Q_{vent,r}$ are the energy consumption for cooling of ventilation. h_{OA} and $h_{¿}$ are the enthalpy value of outdoor air and room air, respectively. H_{air} and $H_{¿,r}$ are the humidity of outdoor air and indoor air, respectively.

Based on above itemized energy loads, to match the system operation and energy simulation model, this study utilized three operation schedules based on different occupancy types. Figure 4 illustrates a typical occupancy schedule of each occupancy type. In the baseline simulation model, all other system operation settings, such as the supply air flow rate and outdoor air flow rate, are either set by facility managers or captured by sensors.

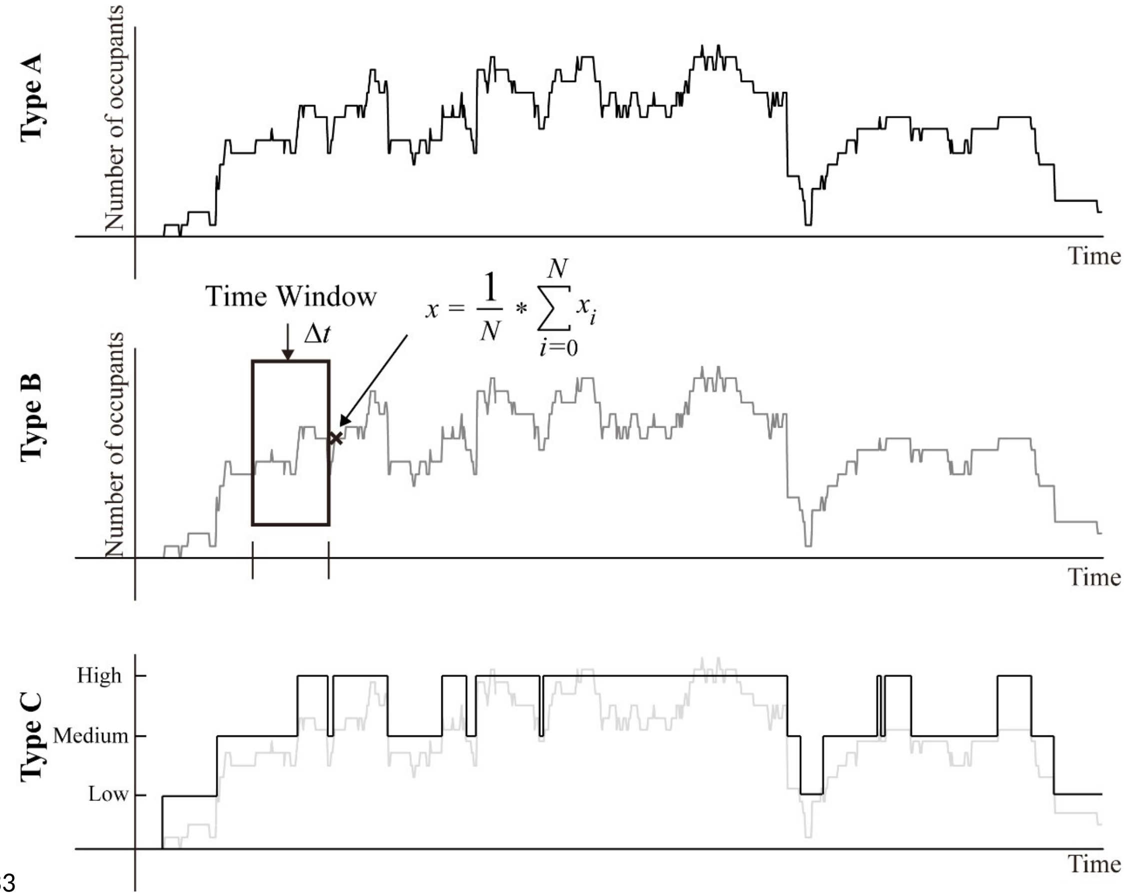

Fig. 4. Sample occupancy schedules for three occupancy types.

(1) Type A occupancy

Type A occupancy reports the continuous and exact occupancy information (number of occupants in a space) that estimated by the ensemble algorithm. The operative temperature and relative humidity settings are computed with ASHRAE standard 62.1-2013 recommended thermal comfort based on the number of occupants. Then the minimum outdoor air flow rate can be computed accordingly.

$$m_{OA} = m_{pred.min}^{OA} = R_p * p_{pred.A}^{r} + R_a * A_r \tag{17}$$

Where $T_{¿,r} = T_{setting}$ and $H_{¿,r} = H_{setting}$ are the temperature and humidity settings. $p_{pred.A}^{r}$ is the predicted results of type A occupancy. $m_{pred.min}^{OA}$ is the minimum outdoor air flow rate based on such data type.

(2) Type B occupancy

As the detected occupancy is often contaminated by random noise and the optimization for system operation is periodical, discrete occupant number with suitable time interval is preferable in many cyber energy models. In addition, fluctuations in occupancy could result in excessive adjustments. Therefore, Type B occupancy applies time window to average occupancy within its length.

$$p_r = p^r_{pred.B} = \frac{t_0}{T} * \sum_{i=0}^{T/t_0} x_i \tag{18}$$

Where $p^r_{pred.B}$ is the predicted occupancy. t_0 is the time resolution of the occupancy. T is the length of the averaging time window.

(3) Type C occupancy

Type C is a simplified categorical scale occupancy for the ease of system operation. In type C occupancy, the predicted results are divided into four levels, including zero, low, medium, and high. The mechanical system can switch between setting scenarios based on the building occupancy level.

In summary, the entire process of occupancy prediction with the ensemble algorithm is illustrated in Figure 2.

---

1. Feature abstraction from Wi-Fi dataset
2. Define occupancy patterns
3. Define Input $X = \{x_1, x_2, \ldots, x_n\}$, Output $y = \{y_1, y_2, \ldots, y_n\}$, a set of base estimators $F = \{f_1(x), f_2(x), \ldots, f_M(x)\}$. Loss function L
4. Select parameter in parameters tuning set

   For i = 1 to M number of iterations):

   (a). Compute residuals

   (b). Fit pseudo-residuals using base estimator

   i.e. set f_m to minimize $L(y, f_m(x))$

   (c). Find multiplier, $w_m = argmin_f\{L(y_i, F_{m-1}(x) + f_m(x)\}$

   (d). Update $F_{m-1}(x) + w_m f_m(x)$

   Output: occupancy model $F_m(x)$

   Calculate assessment metric (MAE, RMSE)

5. Output occupancy model to minimize assessment metric for occupancy patterns
6. Output occupancy pattern file

Fig. 5. The pseudocode of the ensemble algorithm for occupancy prediction

4. VALIDATION EXPERIMENT

4.1 Physical conditions of the experiment testbed

To examine the proposed occupancy linked e-CPSs, this study also conducted a validation experiment in a large office space. The testbed has an area of about 200 square meters and 20 long-term residents during the experiment period. Figure 6 shows the space layout and sensors setup. The room equipped with a dedicated outdoor air system to bring outdoor air into indoor areas without air handling process. The indoor air is conditioned by the fan coil unit with the variable refrigerant flow and the indoor air circulation is driven by positive pressure. The entire room has Wi-Fi coverage with three Wi-Fi probes. During the experiment, TA465-X sensor system (produced by TSI Co.) was utilized to monitor the indoor air temperature, relative humidity, and airflow rate. The CO2 concentration of return air of the fan coil unit was used to approximate the CO2 concentration of the indoor air after air mixing. To eliminate the uneven air mixing, three environmental sensors were evenly installed at the ceiling (3m). Air flow meters were installed near outdoor inlets to monitor the air flow rate of the ventilation system. Two overhead cameras were installed to record the entrance and exit events of occupants. During the experiment, the occupants aware of the Wi-Fi experiment and were instructed to switch on their Wi-Fi signal on their mobile devices. Table 1 shows the specifications of the installed sensors, including data storage types, sensing intervals, range, accuracy, and resolution. The experiment lasted for nine days.

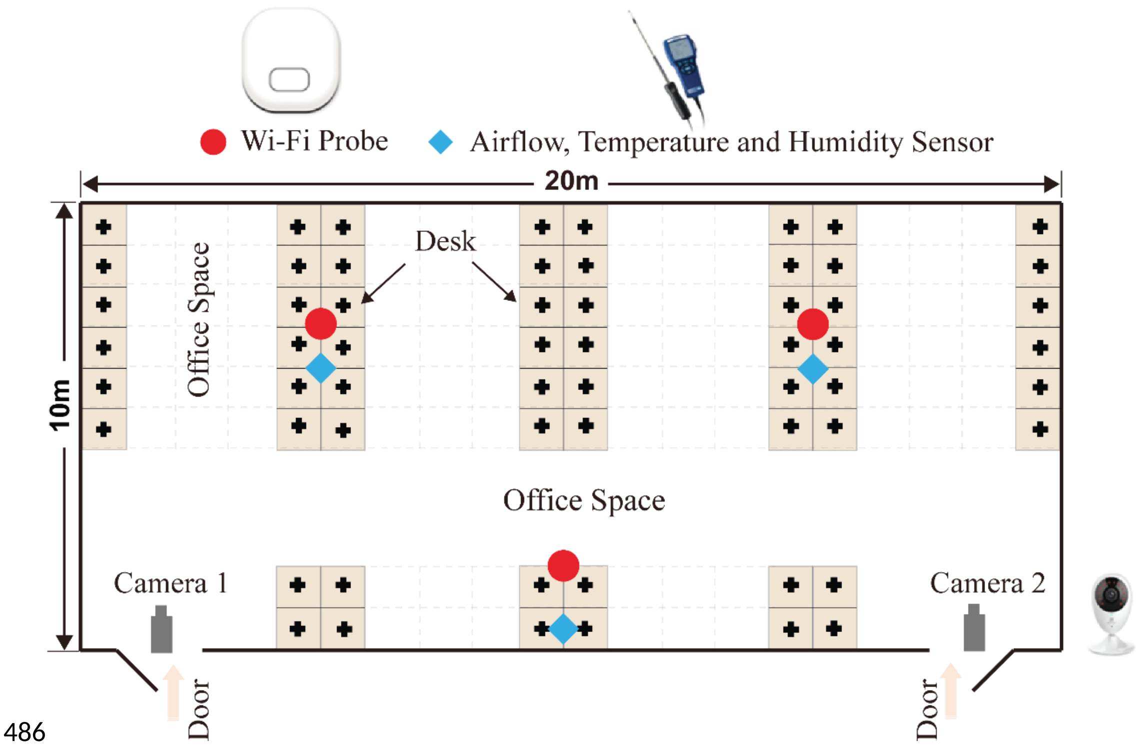

Fig. 6. Space layout and equipment setup.

Table 1. Sensors used in the experiment.

Sensors	Camera	Wi-Fi Probe	Environmental Sensors			
			Air flow rate	Temperature Sensors	Humidity Sensors	Other Sensors
Recorded Variables	Time, Actual occupancy	Time, MAC address, RSSIs	Time, Temperature, Relative humidity, Air flow rate, Air pressure			
Data Storage	Online	Online	Local			
Sensing interval		30s	1min	1min	1min	
Range			0 - 9999 ft/min	14 - 140 °F -10 – 60 °C	0 to 95%	
Accuracy			±3% or ±3 ft/min	±0.5°F (±0.3°C)	< 3%	
Resolution			1 ft/min	0.1°F (0.1°C)	0.10%	

4.2 Cyber model for energy management and simulation

Figure 7 shows the energy cyber model applied in this study. The model was developed with EnergyPlus and DOE2 to optimize facility operation. Based on BIM models, the energy cyber model is able to incorporate construction materials, building geometries, and schedule of operation to estimate the energy consumption of the building. With co-simulation with other programming languages, such as Matlab or Python, the model is capable of tuning system settings to minimize energy consumption. This study employed Eppy, a Python package that can manipulate EnergyPlus IDF files [61], to search for the optimal system settings. It takes full advantage of the rich data structure and idioms that are available in Python and provide availability of designing expected energy model and algorithm to integrate physical and cyber models. Eppy can help programmatically navigate, search, and modify EnergyPlus IDF files. Users can use Eppy to create one or multi new IDF files, make changes to original IDF files, change occupancy schedule in all the interior zones, and read data from the output files after EnergyPlus simulation run. Related to occupancy linked e-CPSs, Eppy provides an interface to link occupancy results from ensemble models as the input to cyber energy model with Python.

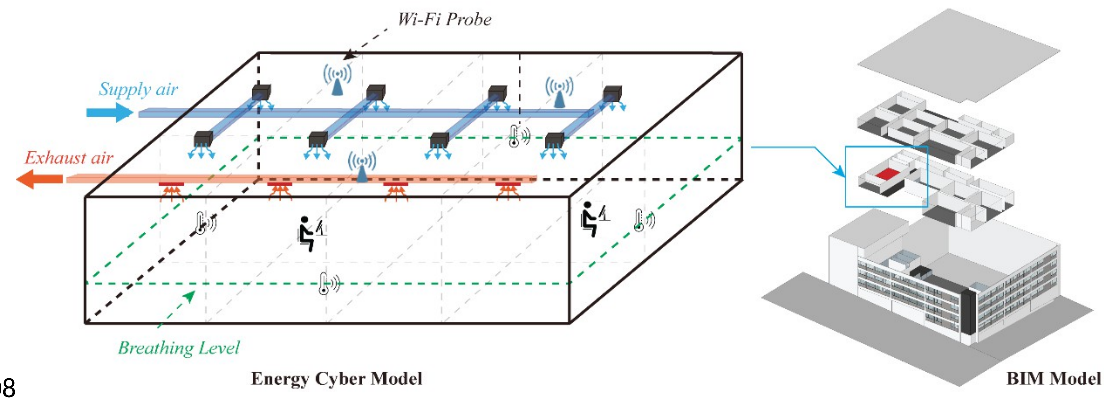

Fig. 7. The energy cyber model for the experiment testbed.

The cyber model matches the physical room with a size of 20 m (length) x 10 m (width) x 3 m (height) and 20 occupants. Internal heat sources were set as 75W for per person, 150W for per computer, and 35W for per lamp. The light schedule followed the on/off schedule and the schedule for computers was assumed to same as the occupancy schedule. Hong Kong has a subtropical climate and high-density highrise urban form. According to statistics [62], the typical mean, minimum, maximum values of monthly average temperature are around 23.4°C, 13.3°C, and 29.8°C, respectively. Also, relative humidity (RH) of Hong Kong is high and minimum, maximum values of monthly average RH are 78.2%, 60%, and 90%, respectively. The typical Hong Kong weather condition was used and the heat transfers from wall, floor, and ceiling were ignored since the experiment was conducted in one inner zone adjacent to conditioned zones. The cooling temperature setpoint is 24°C and there was no heating.

4.3 Data processing

4.3.1 Actual occupancy information

To collect the ground truth for training the ensemble learning algorithms and assessing the model errors, two cameras were installed above the two entrances of the experiment testbed. The number of occupants was counted through video analysis based on the camera records. The counted numbers were synchronized with the internet timestamp with a five-minute interval. To match the Type C occupancy data, the number also was also translated to categorical occupancy levels as specified in Table 2.

Table 2. The threshold setting for categorical occupancy levels

Occupancy level	Number of people
Zero (0)	0
Low (25%)	1-6
Medium (50%)	7-14
High (75%)	15-20

4.3.2 Model parameters tuning

To improve the facility operation with reliable occupancy information, it is necessary to identify, compare, and optimize the ensemble model through parameter tuning. The training model implemented n-fold cross-validation method. In this study, the raw dataset has total 882 samples and about 70% of dataset was used for model training and 30% for model validation and test. Table 3 shows the search space for the parameters tuning. The multi-variable comparison in the exhaustive grid search is applied to identify the best assembly of model parameters. For the RF classifier, the number of estimators determines the results precision and training time, while the number of features affects the accuracy and the diversity of results. For GTB and AdaBoost classifiers, learning rate affects the boosting step length of the gradient descent procedure.

Table 3. Parameters search space for the occupancy ensembled model

Algorithm	Parameter	Range
GTB	Number of estimators	[100; 150; 200; 250; 300; 400; 500; 600; 800; 1000; 1200]
	Learning rate	[0.01; 0.02; 0.05; 0.1; 0.2; 0.25; 0.3; 0.4; 0.5]
	Min_samples_split	[2; 3; 4; 5; 6; 8; 10; 15]
	Max_tree_depth	[3; 4; 5; 6; 7; 8; 9; 10; 12; 15]
AdaBoost	Number of estimators	[100; 150; 200; 250; 300; 400; 500; 600; 800; 1000; 1200]
	Learning rate	[0.01; 0.02; 0.05; 0.1; 0.2; 0.25; 0.3; 0.4; 0.5]
Random Forest	Number of estimators	[100; 150; 200; 250; 300; 400; 500; 600; 800; 1000; 1200]
	Max_ features	[‘all’; ‘sqrt’; ‘log2’]
	Min_samples_leaf	[1; 2; 3; 4; 5; 6; 7; 8; 9; 10]

4.3.3 Error assessment

To evaluate the effectiveness and accuracy of the model, both the mean average error (MAE) and root mean squared error (RMSE) metrics were used for Type A and Type B occupancy. For discrete Type C occupancy, the Accuracy (ACC) is defined with true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) of the confusion matrix.

$$TPR = \frac{TP}{TP + FP} \tag{19}$$

$$TNR = \frac{TN}{TN + FN} \tag{20}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{21}$$

Meanwhile, the value of the area under curve-receiver operating characteristic curve (AUC-ROC) is applied, which is created by the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. For the unbalanced dataset, Balanced Accuracy (bACC) can be used to average the TPR and TNR, which can be presented in the following formula:

$$bACC = \frac{TPR + TNR}{2} \tag{22}$$

According to ASHRAE standard 62.1-2013 [60], the fresh air volume of the ventilation system and the occupant-related thermal load of the air conditioning system are determined by the number of occupants. The errors in the occupancy assessment could directly affect the energy usage of the building. Therefore, the e-CPSs can be significantly improved with the occupancy information incorporated.

5. RESULTS

5.1 Environmental conditions

In the experiment field, dedicated outdoor air system and fan coil unit is under operation. The former system delivers the outdoor air to inner space directly without cooling and the latter cools indoor circulating air. Figures 8 show the environmental conditions during the experiment period. In Figure 8 (a), the outdoor air supply flow rate is 180 cfm (cubic feet per minute) for each outdoor air inlet consistently and the supply air flow rate for each supply air inlet is over 300 cfm but less than 400 cfm most of the time. The outdoor air was supplied uninterrupted during the night even if the cooling services from supply air terminals were closed. Figure 8 (b) shows that the measured supply air temperature varies periodically from 15°C to 25°C, which is caused by the periodical cycling operation of the fan coil system. During the experiment, the outdoor air temperature ranged from 30°C to 35°C, which is a typical summer day in Hong Kong. Figure 8 (c) reports the relative humidity.

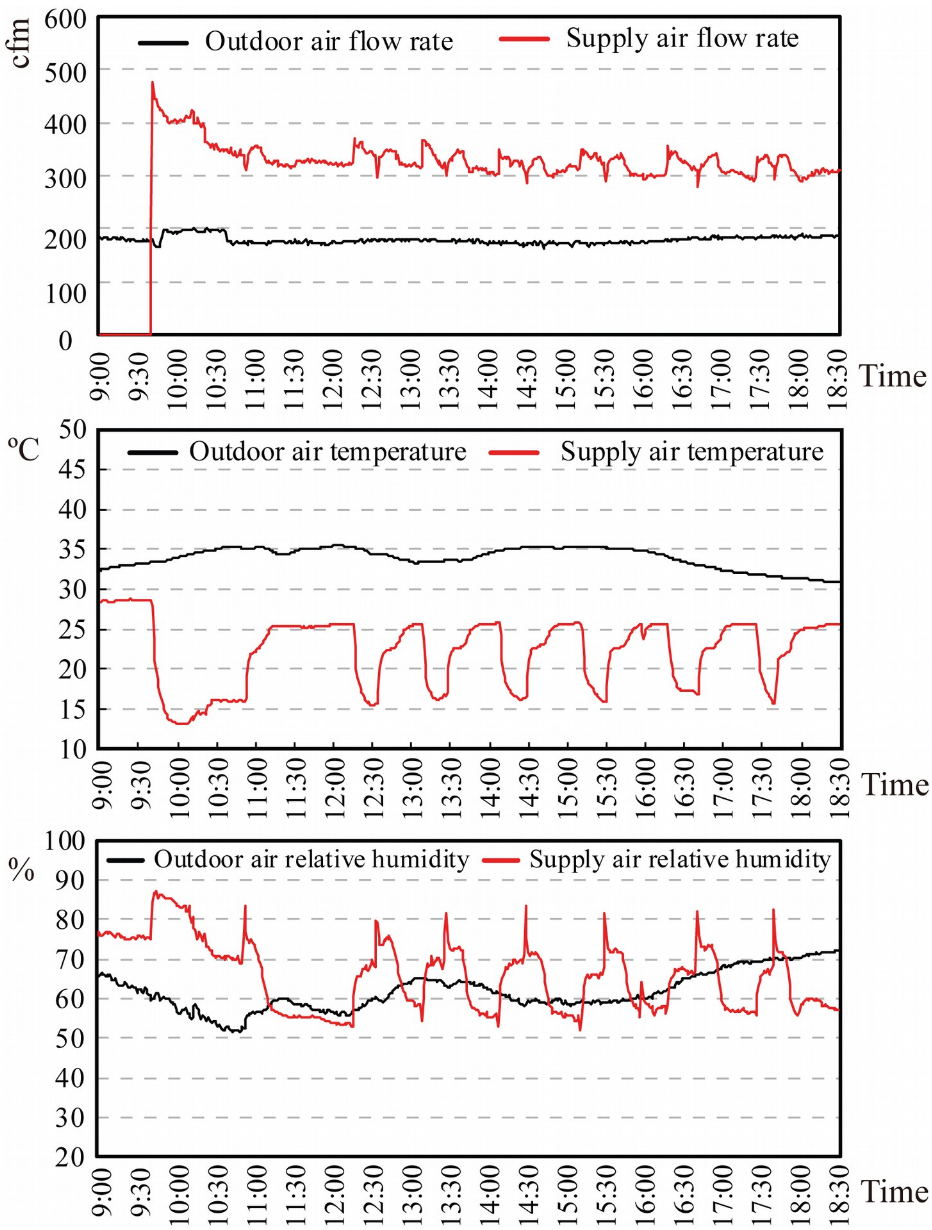

Fig. 8. Environmental conditions of a typical experiment day (a) Air flow rate (top) (b) temperature (middle) (c) relative humidity (bottom).

5.2 Predicted occupancy

This study performed a grid search to determine optimal values for the parameters of the tree-based ensembles. The features of Wi-Fi dataset described in Eq. 5 were considered as the input variables. The GTB classifier consists of 150 estimators with a learning rate of 0.01. To split an internal node, the model requires a minimum 8 samples and a maximum tree depth of 15. The AdaBoost classifier has 100 estimators with a learning rate of 0.2. The RF classifier has 250 estimators and 10 minimum sample leaf. Table 4 summaries the averaged errors of all three type of classifiers after tuning. Among all three types of classifiers, the AdaBoost classifier shows the highest accuracy.

Table 4. Averaged errors for the three ensemble learning algorithms.

	RFs			**GTB**			**AdaBoost**		
	MAE	RMSE	Accu.	MAE	RMSE	Accu.	MAE	RMSE	Accu.
Type A	2.66	3.31		2.89	3.58		2.54	3.30	
Type B	2.63	3.32		2.81	3.53		2.41	3.06	
Type C			71.0%			66.0%			72.7%

Figure 9 presents the predicted results for all three occupancy types with the AdaBoost classifier. Type B occupancy used a 30 minutes sliding time window to smooth the predicted occupancy. Type C occupancy levels are categorized as zero, low, medium, high. The detailed error comparison by days is listed in Table 5 and Table 6 shows the normalized confuse matrix of AdaBoost classifier for Type C occupancy. From detailed assessment results, it shows Day 5 and 7 have the almost best accuracies for type A occupancy with 1.88 and 1.91 of MAE and 2.40 and 2.30 of RMSE respectively. For type B occupancy, Day 3 shows the best accuracy with 1.48 of MAE and 2.48 of RMSE. For the detailed accuracy of Type C occupancy, it can be found that Day 2, 4, 6, and 7 have no “Zero” level occupancy, while Day 3, 5, and 7 have no “High” level occupancy. The best accuracy is shown on Day 7, where accuracies are 61.1% for “Low” and “Medium” levels occupancy, respectively. The total accuracy for Type C occupancy is 72.7% and AUC-ROC value is 0.82. According to Eq. 22, bACC in this study is 70%. The results suggest that although variance there is no significant differences or outlier are observed cross days for MAE and RMSE. Results of Type C occupancy indicate that the classifiers are more suitable for partial occupancy since the overall accuracy of medium occupancy level is much higher than the other levels.

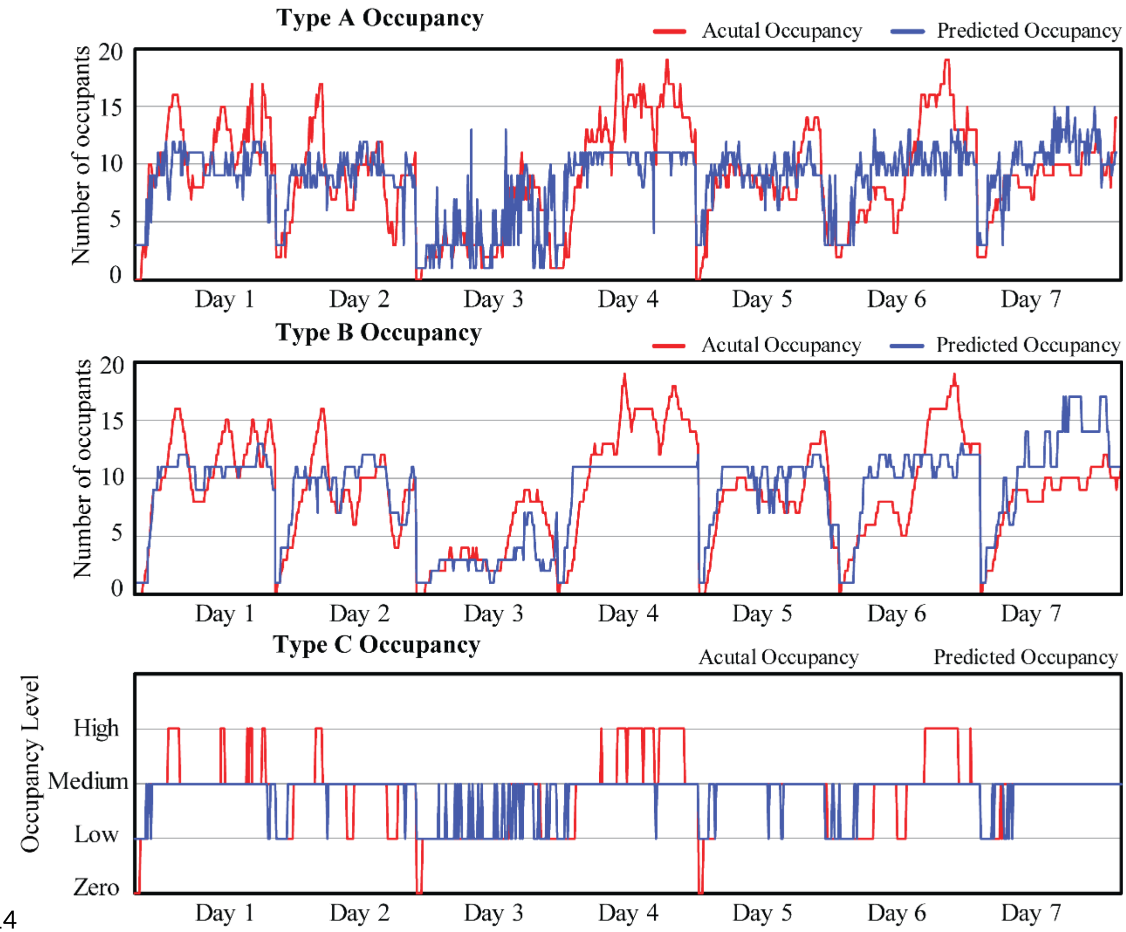

Fig. 9. The predicted occupancy (a) Type A Occupancy (top), (b) Type B Occupancy (middle), (c) Type C Occupancy (bottom).

Table 5. Averaged errors and accuracy of three occupancy types

	Type A Occupancy		Type B Occupancy		Type C Occupancy				
	MAE	*RMSE*	*MAE*	*RMSE*	*Zero*	*Low*	*Medium*	*High*	*Total*
Day 1	2.69	3.38	1.73	2.22	0	85.7%	96.8%	0	76.2%
Day 2	2.15	2.89	1.93	2.45	-	35.5%	93.3%	0	77.8%
Day 3	2.16	3.05	1.48	2.21	0	76.3%	64.2%	-	70.6%
Day 4	3.75	4.40	3.56	4.12	-	50.0%	98.2%	0	50.7%
Day 5	1.88	2.40	1.85	2.14	0	63.6%	95.0%	-	86.5%
Day 6	3.23	4.01	3.12	3.77	-	36.8%	100.0%	0	56.3%
Day 7	1.91	2.30	3.13	3.71	-	61.1%	95.4%	-	90.4%
Total	2.54	3.30	2.41	3.06	0	60.0%	95.0%	0	72.7%

Table 6. The normalized confusion matrix of Type C occupancy results

	Zero	Low	Medium	High
Zero	0.00	1.00	0.00	0.00
Low	0.00	0.60	0.40	0.00
Medium	0.00	0.05	0.95	0.00
High	0.00	0.00	1.00	0.00

5.3 Energy performance and analysis of the occupancy linked e-CPSs

To access the potential energy savings using occupancy-linked e-CPSs, this study simulated three scenarios of energy consumption for both the proposed model and traditional e-CPSs. The baseline model (BM1) is the traditional e-CPSs that use ASHRAE recommended occupancy (ASHRAE Standard 62.1-2013) schedule for energy management and facility operation. The occupancy-linked e-CPSs model (OLEM) implemented the three types of predicated occupancy as modeling input and updated the system operation with new optimized setting parameters. Another benchmarking model (BM2) implemented the actual occupancy information (captured by cameras) as the inputs for the occupancy linked e-CPSs model to estimate its energy saving potential and track the errors.

Figure 10 and 11 shows the simulated cooling load with different occupancy types. In the simulation, the thermostat HVAC terminals in BM1 were set to default temperature and the mechanical operation was mainly affected by the weather condition. From both figures, it can be seen that the energy consumption for the cooling load in BM1 is significantly higher than BM2 and OLEM, which included occupancy as inputs for load estimation. In addition, all three occupancy types are similar to each other and Type C seems closer to the actual demand.

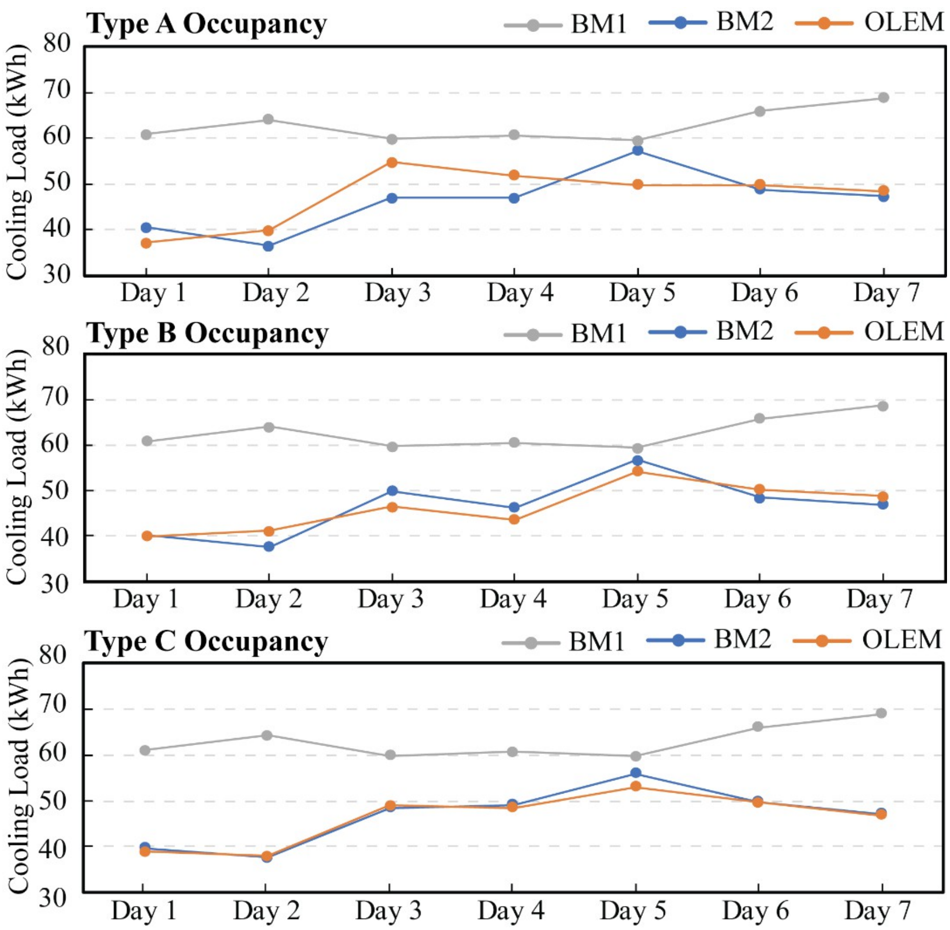

Fig. 10. Simulated daily cooling load based on three occupancy types.

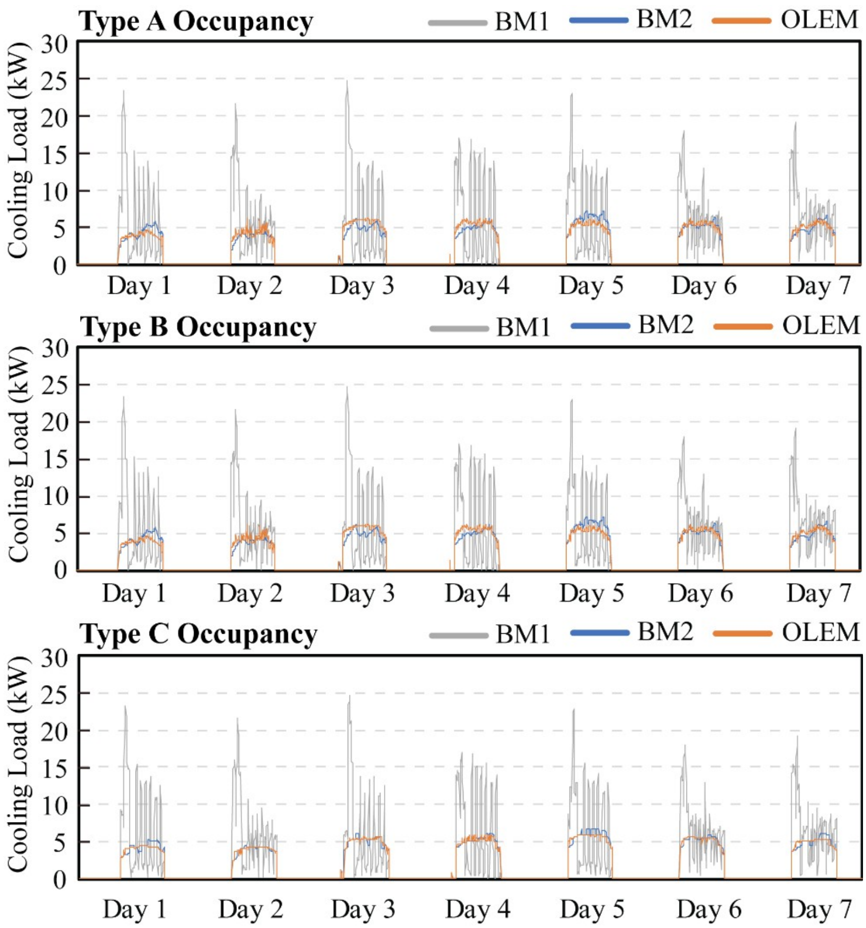

Fig. 11. Simulated hourly cooling load based on three occupancy types.

Another energy consumption component for the HVAC system is the fresh air amount. The mechanical drives and fans consume a large amount of energy when the air handling units deliver the outdoor air into indoor spaces. The physical building deploys on/off the system with a fixed flow rate about 1440 m3/h. However, according to ASHRAE standard, the flow amount is obviously insufficient given the number of occupants in the experiment office. Figure 12 and 13 show the simulated minimum outdoor air flow rate and amount. Both figures suggest that the outdoor air amount in BM1 is far less than the demand according to the number of occupants. Type A occupancy performs the worst among all three types, this could be caused by the tracking errors result from data fluctuation.

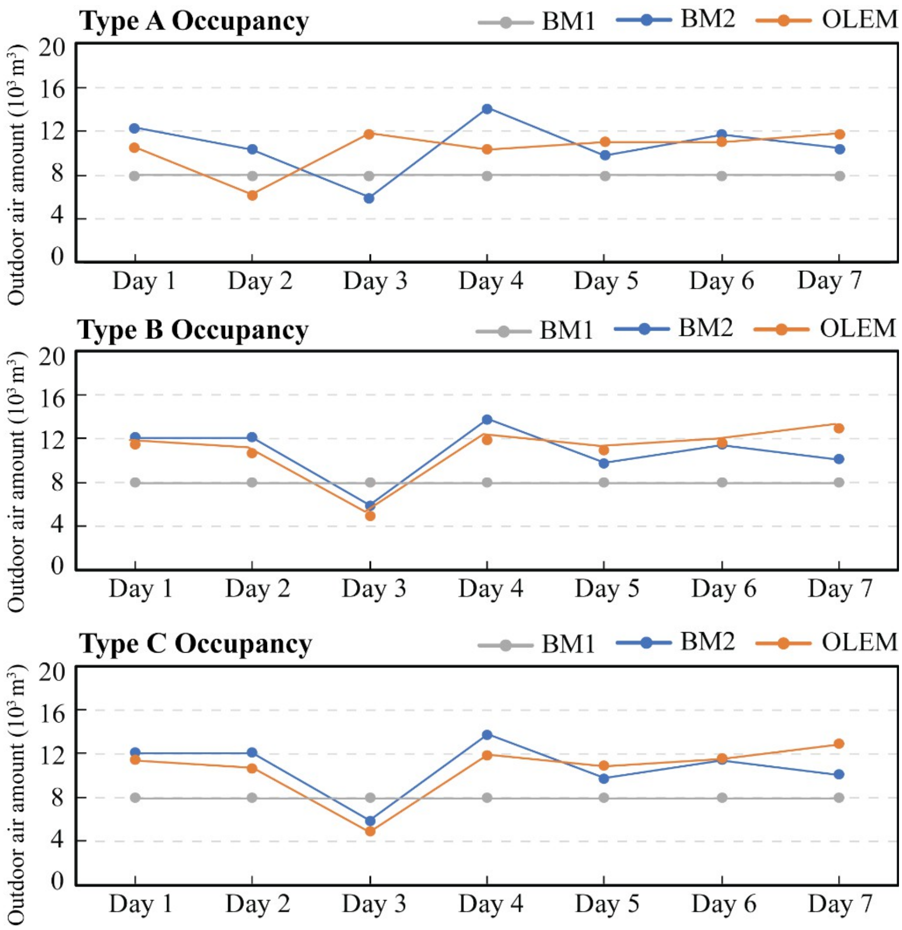

Fig. 12. Simulated daily outdoor air amount based on three occupancy types.

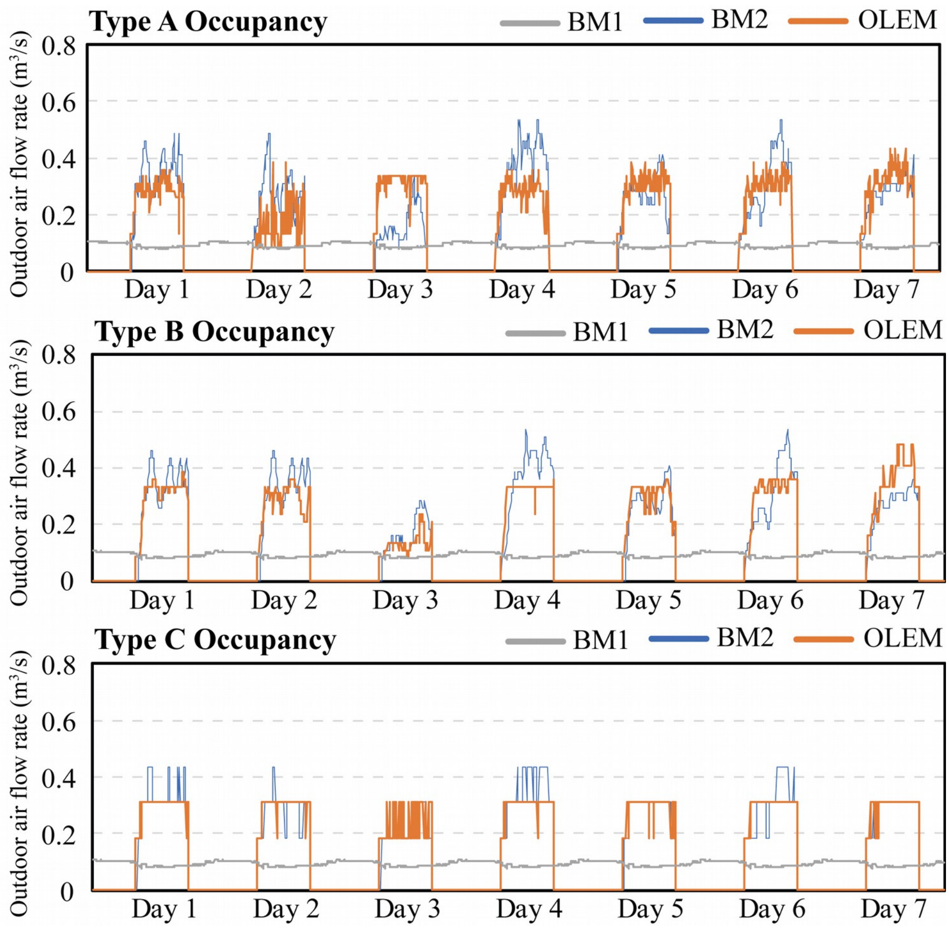

Fig. 13. Estimated hourly outdoor air flow rate based on three occupancy types.

Then the total energy consumption of air conditioning and ventilation was aggregated and compared for all three models. BM1 was used as the reference and potential savings are computed as a percentage less than the energy consumption of B1. Table 6 summaries the aggregated results. The averaged savings vary from 24.71% to 26.31% and all three occupancy types have a close performance. The results indicate that the fixed flow rate of conditioned air could easily result in over-cooling and energy wastes.

Table 6. Energy saving potentials for different occupancy types (compared with BM1)

	Type A		**Type B**		**Type C**	
	BM2 vs. BM1	OLEM vs. BM1	BM2 vs. BM1	OLEM vs. BM1	BM2 vs. BM1	OLEM vs. BM1
Day 1	33.46%	39.27%	34.27%	34.61%	34.65%	36.17%

Day 2	43.16%	38.07%	41.50%	36.08%	41.39%	40.77%
Day 3	21.54%	8.55%	16.38%	22.37%	18.99%	18.12%
Day 4	22.60%	14.61%	23.79%	28.01%	19.06%	20.08%
Day 5	3.86%	16.47%	4.69%	8.94%	6.12%	10.90%
Day 6	26.00%	24.62%	26.62%	23.87%	24.68%	24.83%
Day 7	31.26%	29.52%	31.81%	29.14%	31.31%	31.87%
Total	26.29%	24.71%	25.91%	26.31%	25.47%	26.37%

6. DISCUSSION

With the rapid technological development of ICT and IoT, an increasing number of buildings are encouraged to install various sensors and sensor networks to facility smarter management and control. Combining these technologies, e-CPSs allow new advances such as data analytics, artificial intelligence to be utilized in optimizing building control for higher energy efficiency and human-centric services. This study extended conventional e-CPSs by introducing occupancy detection and prediction components so that the occupancy information can be included for better service and less energy waste. The detected occupancy can be used as dynamic information exchange between the physical building and cyber models so that the optimization boundary conditions can be updated timely. For existing buildings, since all building features have been determined, the major uncertainties in e-CPSs arise from weather conditions and occupancy variations. The occupancy-linked e-CPSs mitigated the occupant-related uncertainty by incorporating a reliable occupancy prediction mechanism. Accurate occupancy information allows building management system to turn off certain functions when occupants are absent to avoid waste. The validation experiment results suggest that the accuracy can reach 72.7% and reveal that when incorporating occupancy information, the e-CPSs is capable of implementing the demand-based facility management to promote building energy efficiency. For example, the validation experiment suggests 24% of energy saving potential and 33.3% air amount compensation. With the proposed ensemble algorithm, e-CPSs can receive occupancy information with acceptable accuracy, especially when the occupancy was categorized. Also, it can be observed from the experiment that three types of occupancy information show no significant differences in the simulation and Type C occupancy is more suitable for practical implementation in e-CPSs control as it requires less computational power and is easier for practical deployment.

One challenge in conventional e-CPSs is that many predefined human-centric control approaches conflict with the occupants' actual preferences and activities since occupancy is stochastic and changeable in different buildings. This study contributes to the research gap by proposing a theoretical framework for occupancy-linked e-CPSs model and a feasible ensemble algorithm to predict occupancy with proper data sources. As WiFi networks become a premise of all cloud-based platforms and cyber models, it is naturally compatible with e-CPSs without additional cost. The highly accessible WiFi technologies in modern buildings can help boost applicability of proposed OLEM. For existing buildings with Wi-Fi installation, through deploying fast and reliable artificial intelligence technologies, such as the proposed ensemble algorithms, the occupancy becomes accessible to e-CPSs and creates a significant synergy among all cyber models. In addition, with the cumulation of the detected and predicated occupancy, designers also can rethink and refine the building space design and mechanical system selection for new buildings. For example, it is possible to integrate WiFi-based occupancy-driven lighting control for smart buildings [63] and include the lighting system into the e-CPSs platform. Additionally, the unprecedented increase of human activities in buildings, infrastructures, and vehicles generates a complex and interdependent system in modern cities. The advances in the world wide web technologies allow an efficient information sharing through cloud among e-CPSs. Under such a context, the occupancy studies for e-CPSs can also be extended to urban scale. For example, the occupancy information can be associated with the human mobility between buildings and can be used for inter-building energy demand assessment. The information gathered from occupancy linked e-CPSs can be used for regional electricity grid design and human-centric urban planning. Another inspiring research direction is to integrate OLEM with smart grids for dynamically computed demand at the building side to achieve smart girds or microgrids optimization. In addition, such implementation also requires new technologies to protect the occupants' security and privacy during occupancy detection [64].

This study also yields to limitations, which can be resolved in future studies. Firstly, the validation experiment constraint to small space (an office room). It is suggested to study a larger building space with multiple rooms so that the impact of indoor commutes can be included. Also, rooms with different functions also have their unique occupancy patterns and mechanical system selection. Secondly, the energy consumption in this study mainly results from cooling load and ventilation due to the tropical climate condition and short experiment period. However, there are various energy consuming services systems in buildings, such as lighting, security, heating, and etc., which are also closely associated with human behaviors and inter-dependent with each other.

7. CONCLUSION

This study proposed a theoretical framework for implementing occupancy information as dynamic links for e-CPSs. The framework adopted WiFi Probe technology and ensemble classifiers to interpret WiFi connections as reliable and usable occupancy information. Three occupancy types (Type A, B, and C) have been compared in a validation experiment to examine the accuracy and feasibility of the proposed occupancy-linked e-CPSs. After a validation experiment, the proposed model can accurately report occupant counts for system energy management. The AdaBoost method and type C occupancy report the highest detection accuracy of 72.7%. Type A occupancy has an absolute error and root mean squared error of 2.54 and 3.30, and both values for type B occupancy are 2.41 and 3.06, respectively. The energy simulation reports 24.7%, 26.4%, and 26.3% energy saving potentials by implementing these three types of occupancy information in e-CPSs, respectively.

This study contributes to the development of high-precision and large-scale human-centric services in e-CPSs. For future studies, it is suggested to investigate large-scale and more complicated system coordination and incorporate more information to bridge the energy system and CPSs, such as environmental conditions and occupants' feedback. In addition, the concept of occupancy-lined e-CPSs can be transplanted to smart grid management to optimize power supply across multiple buildings.

ACKNOWLEDGMENT

This work was financially supported by the Hong Kong General Research Fund

(GRF) – Early Career Scheme, #21204816, and the National Natural Science Foundation of China (NSFC), #51508487. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of GRF and NSFC.

REFERENCES

[1] U.S. Energy Information Administration 2017. https://www.eia.gov/energyexplained/index.cfm?page=us_energy_use.

[2] Rajkumar R, Lee ILI, Sha LSL, Stankovic J. Cyber-physical systems: The next computing revolution. Design Automation Conference (DAC), 2010 47th ACM/IEEE 2010:0–5. doi:10.1145/1837274.1837461.

[3] Gupta SKS, Mukherjee T, Varsamopoulos G, Banerjee A. Research directions in energy-sustainable cyber-physical systems. Sustainable Computing: Informatics and Systems 2011;1:57–74. doi:10.1016/j.suscom.2010.10.003.

[4] Peng Rong, Pedram M. Power-aware scheduling and dynamic voltage setting for tasks running on a hard real-time system. Asia and South Pacific Conference on Design Automation, 2006., IEEE; 2006, p. 473–8. doi:10.1109/ASPDAC.2006.1594730.

[5] Dobson I. Energy Cyber-Physical Systems : Research Challenges and Opportunities 2016.

[6] Szilagyi I, Wira P. An Intelligent System for Smart Buildings using Machine Learning and Semantic Technologies : A Hybrid Data-Knowledge Approach 2018:20–5.

[7] Kleissl J, Agarwal Y. Cyber-physical energy systems: focus on smart buildings. Proceedings of the 47th Design Automation Conference on - DAC ’10 2010:749. doi:10.1145/1837274.1837464.

[8] Balaji B, Abdullah M, Faruque A, Dutt N, Gupta R, Agarwal Y. Models , Abstractions , and Architectures : The Missing Links in Cyber-Physical Systems n.d.

[9] Zhao P, Simoes MG, Suryanarayanan S. A conceptual scheme for cyber-physical systems based energy management in building structures. 2010 9th IEEE/IAS International Conference on Industry Applications - INDUSCON 2010 2010:1–6. doi:10.1109/INDUSCON.2010.5739891.

[10] Paridari K, El-Din Mady A, La Porta S, Chabukswar R, Blanco J, Teixeira A, et al. Cyber-Physical-Security Framework for Building Energy Management System 2016. doi:10.1109/ICCPS.2016.7479072.

[11] Francisco A, Truong H, Khosrowpour A, Taylor JE, Mohammadi N. Occupant perceptions of building information model-based energy visualizations in eco-feedback systems. Applied Energy 2018;221:220–8. doi:10.1016/j.apenergy.2018.03.132.

[12] Hong T, Chou S., Bong T. Building simulation: an overview of developments and information sources. Building and Environment 2000;35:347–61.

doi:10.1016/S0360-1323(99)00023-2.

[13] Modelica and the Modelica Association — Modelica Association n.d. https://www.modelica.org/.

[14] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, et al. EnergyPlus: creating a new-generation building energy simulation program. Energy and Buildings 2001;33:319–31. doi:10.1016/S0378-7788(00)00114-6.

[15] Delwati M, Merema B, Breesch H, Helsen L, Sourbron M. Impact of demand controlled ventilation on system performance and energy use. Energy and Buildings 2018;174:111–23. doi:10.1016/j.enbuild.2018.06.015.

[16] Hong T, Sun K, Zhang R, Hinokuma R, Kasahara S, Yura Y. Development and validation of a new variable refrigerant flow system model in EnergyPlus. Energy and Buildings 2016;117:399–411. doi:10.1016/J.ENBUILD.2015.09.023.

[17] Stamatescu G, Stamatescu I, Arghira N, Calofir V, Fagarasan I. Building Cyber-Physical Energy Systems 2016.

[18] Behl M, Jain A, Mangharam R. Data-Driven Modeling, Control and Tools for Cyber-Physical Energy Systems. 2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems, ICCPS 2016 - Proceedings 2016. doi:10.1109/ICCPS.2016.7479093.

[19] Ferreira PM, Ruano AE, Silva S, Conceição EZE. Neural networks based predictive control for thermal comfort and energy savings in public buildings. Energy and Buildings 2012;55:238–51. doi:10.1016/J.ENBUILD.2012.08.002.

[20] Costanzo GT, Iacovella S, Ruelens F, Leurs T, Claessens BJ. Experimental analysis of data-driven control for a building heating system. Sustainable Energy, Grids and Networks 2016;6:81–90. doi:10.1016/J.SEGAN.2016.02.002.

[21] Belafi Z, Hong T, Reith A. Smart building management vs. intuitive human control—Lessons learnt from an office building in Hungary. Building Simulation 2017;10:811–28. doi:10.1007/s12273-017-0361-4.

[22] Wang Y, Shao L. Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning. Building and Environment 2017;114:106–17. doi:10.1016/j.buildenv.2016.12.015.

[23] Menezes AC, Cripps A, Bouchlaghem D, Buswell R. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. Applied Energy 2012;97:355–64. doi:10.1016/j.apenergy.2011.11.075.

[24] Liang X, Hong T, Shen GQ. Improving the accuracy of energy baseline models

for commercial buildings with occupancy data. Applied Energy 2016;179:247–60. doi:10.1016/j.apenergy.2016.06.141.
[25] Wang W, Chen J, Huang G, Lu Y. Energy efficient HVAC control for an IPS-enabled large space in commercial buildings through dynamic spatial occupancy distribution. Applied Energy 2017. doi:10.1016/J.APENERGY.2017.06.060.
[26] Barbeito I, Zaragoza S, Tarrío-Saavedra J, Naya S. Assessing thermal comfort and energy efficiency in buildings by statistical quality control for autocorrelated data. Applied Energy 2017;190:1–17. doi:10.1016/J.APENERGY.2016.12.100.
[27] Zhang S, Cheng Y, Fang Z, Huan C, Lin Z. Optimization of room air temperature in stratum-ventilated rooms for both thermal comfort and energy saving. Applied Energy 2017;204:420–31. doi:10.1016/J.APENERGY.2017.07.064.
[28] Korkas CD, Baldi S, Michailidis I, Kosmatopoulos EB. Occupancy-based demand response and thermal comfort optimization in microgrids with renewable energy sources and energy storage. Applied Energy 2016;163:93–104. doi:10.1016/j.apenergy.2015.10.140.
[29] Chen X, Wang Q, Srebric J. Occupant feedback based model predictive control for thermal comfort and energy optimization: A chamber experimental evaluation. Applied Energy 2016;164:341–51. doi:10.1016/j.apenergy.2015.11.065.
[30] Lim G-H, Keumala N, Ghafar NA. Energy saving potential and visual comfort of task light usage for offices in Malaysia. Energy and Buildings 2017;147:166–75. doi:10.1016/J.ENBUILD.2017.05.004.
[31] Shen E, Hu J, Patel M. Energy and visual comfort analysis of lighting and daylight control strategies. Building and Environment 2014;78:155–70. doi:10.1016/J.BUILDENV.2014.04.028.
[32] Siano P. Demand response and smart grids—A survey. Renewable and Sustainable Energy Reviews 2014;30:461–78. doi:10.1016/j.rser.2013.10.022.
[33] Strbac G. Demand side management: Benefits and challenges. Energy Policy 2008;36:4419–26. doi:10.1016/j.enpol.2008.09.030.
[34] Nguyen A-T, Reiter S, Rigo P. A review on simulation-based optimization methods applied to building performance analysis. Applied Energy 2014;113:1043–58. doi:10.1016/j.apenergy.2013.08.061.
[35] Díaz JA, Jiménez MJ. Experimental assessment of room occupancy patterns in an office building. Comparison of different approaches based on CO 2 concentrations and computer power consumption. Applied Energy

2017;199:121–41. doi:10.1016/j.apenergy.2017.04.082.

[36] Oldewurtel F, Sturzenegger D, Morari M. Importance of occupancy information for building climate control. Applied Energy 2013;101:521–32. doi:10.1016/j.apenergy.2012.06.014.

[37] Hong T, Yan D, D ’oca S, Chen C-F. Ten questions concerning occupant behavior in buildings: The big picture. Building and Environment 2017;114:518–30. doi:10.1016/j.buildenv.2016.12.006.

[38] Yan D, Hong T, Dong B, Mahdavi A, D’Oca S, Gaetani I, et al. IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. Energy and Buildings 2017;156:258–70. doi:10.1016/J.ENBUILD.2017.09.084.

[39] Hong T, Taylor-Lange SC, D ’oca S, Yan D, Corgnati SP, D’Oca S, et al. Advances in research and applications of energy-related occupant behavior in buildings. Energy and Buildings 2016;116:694–702. doi:10.1016/j.enbuild.2015.11.052.

[40] Yan D, O’Brien W, Hong T, Feng X, Burak Gunay H, Tahmasebi F, et al. Occupant behavior modeling for building performance simulation: Current state and future challenges. Energy and Buildings 2015;107:264–78. doi:10.1016/j.enbuild.2015.08.032.

[41] Hong T, Yan D, D ’oca S, Chen C-F. Ten questions concerning occupant behavior in buildings: The big picture. Building and Environment 2017;114:518–30. doi:10.1016/j.buildenv.2016.12.006.

[42] Kim Y-S, Heidarinejad M, Dahlhausen M, Srebric J. Building energy model calibration with schedules derived from electricity use data. Applied Energy 2017;190:997–1007. doi:10.1016/j.apenergy.2016.12.167.

[43] Yang J, Santamouris M, Lee SE, Deb C. Energy performance model development and occupancy number identification of institutional buildings. Energy and Buildings 2016;123:192–204. doi:10.1016/J.ENBUILD.2015.12.018.

[44] Yang Z, Becerik-Gerber B. The coupled effects of personalized occupancy profile based HVAC schedules and room reassignment on building energy use. Energy and Buildings 2014;78:113–22. doi:10.1016/j.enbuild.2014.04.002.

[45] Pisello AL, Asdrubali F. Human-based energy retrofits in residential buildings: A cost-effective alternative to traditional physical strategies. Applied Energy 2014;133:224–35. doi:10.1016/j.apenergy.2014.07.049.

[46] Chen Y, Liang X, Hong T, Luo X. Simulation and visualization of energy-related occupant behavior in office buildings. Building Simulation 2017;10:785–98. doi:10.1007/s12273-017-0355-2.

[47] Jin M, Bekiaris-Liberis N, Weekly K, Spanos CJ, Bayen AM. Occupancy

Detection via Environmental Sensing. IEEE Transactions on Automation Science and Engineering 2018;15:443–55. doi:10.1109/TASE.2016.2619720.

[48] Jin M, Jia R, Spanos CJ. Virtual Occupancy Sensing: Using Smart Meters to Indicate Your Presence. IEEE Transactions on Mobile Computing 2017;16:3264–77. doi:10.1109/TMC.2017.2684806.

[49] Weekly K, Zou H, Xie L, Jia QS, Bayen AM. Indoor occupant positioning system using active rfid deployment and particle filters. Proceedings - IEEE International Conference on Distributed Computing in Sensor Systems, DCOSS 2014 2014:35–42. doi:10.1109/DCOSS.2014.53.

[50] Wang W, Chen J, Lu Y, Wei H-H. Energy conservation through flexible HVAC management in large spaces: An IPS-based demand-driven control (IDC) system. Automation in Construction 2017;83:91–107. doi:10.1016/J.AUTCON.2017.08.021.

[51] Wang W, Lin Z, Chen J. Promoting Energy Efficiency of HVAC Operation in Large Office Spaces with a Wi-Fi Probe enabled Markov Time Window Occupancy Detection Approach. World Engineers Summit – Applied Energy Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference, WES-CUE 2017, Singapore: 2017.

[52] Chen J, Ahn C. Assessing occupants’ energy load variation through existing wireless network infrastructure in commercial and educational buildings. Energy and Buildings 2014;82:540–9. doi:10.1016/j.enbuild.2014.07.053.

[53] Balaji B, Xu J, Nwokafor A, Gupta R, Agarwal Y. Sentinel: Occupancy Based HVAC Actuation using Existing WiFi Infrastructure within Commercial Buildings. Conference: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, 2013. doi:10.1145/2517351.2517370.

[54] Jin M, Jia R, Kang Z, Konstantakopoulos IC, Spanos CJ. PresenceSense: zero-training algorithm for individual presence detection based on power monitoring. Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings - BuildSys ’14 2014:1–10. doi:10.1145/2674061.2674073.

[55] Zou H, Jiang H, Yang J, Xie L, Spanos CJ. Non-intrusive occupancy sensing in commercial buildings. Energy and Buildings 2017;154:633–43. doi:10.1016/j.enbuild.2017.08.045.

[56] Zou H, Zhou Y, Yang J, Spanos CJ. Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT. Energy and Buildings 2018;174:309–22. doi:10.1016/j.enbuild.2018.06.040.

[57] Zhao Y, Zeiler W, Boxem G, Labeodan T. Virtual occupancy sensors for real-time occupancy information in buildings. Building and Environment

2015;93:9–20. doi:10.1016/j.buildenv.2015.06.019.

[58] Wang W, Chen J, Hong T, Zhu N. Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology. Building and Environment 2018;138:160–70. doi:10.1016/J.BUILDENV.2018.04.034.

[59] Wang W, Chen J, Song X. Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference approach. Building and Environment 2017;124:130–42. doi:10.1016/J.BUILDENV.2017.08.003.

[60] ANSI/ASHRAE Standard 62.1. 2013:Ventilation for Acceptable Indoor Air Quality. 2013.

[61] Eppy Tutorial — eppy 0.5.44 documentation n.d. https://pythonhosted.org/eppy/Main_Tutorial.html.

[62] Ang BW, Wang H, Ma X. Climatic influence on electricity consumption: The case of Singapore and Hong Kong. Energy 2017;127:534–43. doi:10.1016/j.energy.2017.04.005.

[63] Zou H, Zhou Y, Jiang H, Chien S-C, Xie L, Spanos CJ. WinLight: A WiFi-based occupancy-driven lighting control system for smart building. Energy and Buildings 2018;158:924–38. doi:10.1016/J.ENBUILD.2017.09.001.

[64] Boroojeni KG, Amini MH, Iyengar SS. Smart Grids: Security and Privacy Issues. Cham: Springer International Publishing; 2017. doi:10.1007/978-3-319-45050-6.