

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

3 Wei Wang <sup>a</sup>, Tianzhen Hong <sup>b</sup>, Nan Li <sup>c</sup>, Ryan Qi Wang <sup>d</sup>, and Jiayu Chen <sup>e\*</sup>

4 <sup>a</sup> *School of Architecture, Southeast University, 210018, Nanjing, Jiangsu, China*

5 <sup>b</sup> *Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory,*  
6 *Cyclotron Road, Berkeley, CA 94720, USA*

7 <sup>c</sup> *Department of Construction Management, Tsinghua University, 100084, Beijing, China*

8 <sup>d</sup> *Department of Civil and Environment Engineering, Northeastern University, 433 SN, 360*  
9 *Huntington Avenue, Boston, MA 02115, USA*

10 <sup>e</sup> *Department of Architecture and Civil Engineering, City University of Hong Kong, Y6621,*  
11 *AC1, Tat Chee Ave, Kowloon, Hong Kong*

12  
13 **Abstract:**

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

---

\*Corresponding author. tel.: +852 3442 4696; fax: +852 3442 0427.

e-mail addresses: [jiayuchen@cityu.edu.hk](mailto:jiayuchen@cityu.edu.hk) (Jiayu Chen)

29 weak classifiers for WiFi signals. A validation experiment in a large office room was  
30 conducted to examine the performance of the proposed occupancy linked  
31 energy-cyber-physical systems. The experiment and simulation results suggest that,  
32 with a proper classifier and occupancy type, the proposed model can potentially save  
33 about 26.4% of energy consumption from the cooling and ventilation demands.

34

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

---

## Nomenclatures

---

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

---

37

38

## 39 **1. INTRODUCTION**

40 Buildings consume more than 40% of primary energy among all energy-consuming  
41 sectors [1] and energy bills become the largest overhead in building maintenance and  
42 operation budget. An increasing number of building owners and decision makers  
43 recognize promoting building energy efficiency as the most cost-effective approach  
44 for conservation. In modern buildings, the majority of energy is consumed by the  
45 mechanical/facility systems, which consists of heating, ventilation, air-conditioning  
46 (HVAC), lighting, water, safety, and similar allied subsystems. However, promoting  
47 energy efficiency of these facility systems is extremely challenging, as they usually  
48 have to comply with complicated working conditions, comfort requirements, and  
49 dynamic energy demand. In recent years, researchers propose to integrate both  
50 physical building systems with engineered cyber models so that building systems can  
51 be monitored, coordinated, controlled, optimized with a computing and  
52 communication core [2]. The integrated system is able to model, visualize, and  
53 operate complex building systems with various computing tools, and such systems are  
54 called cyber-physical systems (CPSs). With advances in the sensors, sensor networks,  
55 and embedded computing systems, CPSs unlocked the potential of optimizing  
56 building energy systems, such consolidated system is called energy-cyber-physical  
57 systems (e-CPSs) [3]. The ideal e-CPSs are designed to reduce the power demand  
58 though computational optimization so that the demand can be satisfied by the  
59 available power with minimum waste [3]. In this context, strategies were developed to  
60 optimize building facility operation through frequency control, voltage control, or  
61 sleep state scheduling [4]. However, dynamic demand caused by occupants and  
62 distributed operation cause poor system coordination in the centralized control system  
63 [5]. The physical facility systems require the computational outcomes from cyber  
64 model to optimize their operation, but the biggest challenge is the unreliable and  
65 incorrect demand estimation, which often results in energy wastes or unsatisfied  
66 thermal comfort.

67 Therefore, a well-integrated e-CPSs should ensure reliability of demand information,  
68 which is usually captured by the physical system. With accurate and meaningful data  
69 inputs, the cyber model can provide effective operation suggestions. However, a  
70 building's energy demand is mainly generated by occupants' thermal, lighting, and

71 functional requirements, which are extremely dynamic and difficult to be captured by  
72 the physical building system. Conventional e-CPSs can synchronize physical  
73 mechanical and energy management systems with digital models, but they lack the  
74 ability to response to uncertain demand of occupants. Due to this constraint, in  
75 practice, conventional e-CPSs usually are rigid and static systems that based on  
76 certain assumed operation schedules. To fill this research gap, this study proposes to  
77 implement structured occupancy information to bridge the cyber and physical systems  
78 and form a new occupancy linked e-CPSs. Such system incorporates WiFi probe  
79 technology and interpreters that are based on ensemble Wi-Fi signals classifiers. The  
80 WiFi probe infrastructure on the physical model side and the ensemble signal  
81 classifiers on the cyber model side can be integrated and bridged by the accurate and  
82 reliable occupancy estimation. With such occupancy information, accurate demand  
83 can be estimated and the facility operation can be optimized for the energy saving  
84 purpose.

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

91

## 92 **2. BACKGROUND**

### 93 **2.1 Energy management and cyber-physical system**

94 With the increased capability and decreased cost of wireless sensors, CPSs are  
95 capable of capture various building information through efficient networks and  
96 abundant computing powers. Thus, researches proposed to develop CPSs for building  
97 energy management systems in future smart buildings [6]. Kleissl and Agarwal looked  
98 at modern smart buildings entirely as a cyber-physical energy systems and examined  
99 the opportunities with joint optimization of energy use by occupants and information  
100 processing equipment [7]. Balaji et al. explored two case studies on smart buildings  
101 and electric vehicles to examine the feasibility of implementation of CPSs for energy

102 management [8]. Zhao et al. developed a conceptual scheme for CPSs based energy  
103 management in buildings that combines the building energy information system,  
104 net-zero energy system, and demand-driven system [9]. Paridari et al. proposed a  
105 cyber-physical-security framework that also includes building energy management  
106 system (BEMS) with resilient policy and security analytics [10]. Based on upon these  
107 efforts, researchers concluded that e-CPSs is one of most prominent platforms in  
108 promoting building efficiency by introducing energy management into the  
109 cyber-physical interaction loop.

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

128 As the premise of effective e-CPSs is to ensure human-centric services (e.g. thermal  
129 comfort, visual comfort) while saving as much as possible energy, researchers  
130 recognized that occupancy information played a central role to guarantee the e-CPSs'  
131 performance in smart buildings [21]. Latest studies suggest that accurate occupancy  
132 information not only links the physical building systems and cyber models but also  
133 mitigates the discrepancies between the designed/simulated and the actual building  
134 operation performance [22]. Menezes et al. conducted a comprehensive study on the

135 non-domestic buildings and concluded that occupancy information is significant to  
136 building energy and occupancy comfort benchmarking [23]. Liang et al. also stated  
137 occupancy data should be included to improve accuracy of building energy use  
138 predicting since occupancy is highly correlated with energy use and thermal comfort  
139 [24]. Wang et al. applied neural networks and WiFi technology to predict occupancy  
140 and integrate it to efficient building HVAC control and save 20% energy through  
141 avoiding overheating and overcooling [25]. Barbeito et al. assessed occupant thermal  
142 comfort and energy efficiency in buildings using statistical quality control (SQC) with  
143 integrated big data web energy platform [26]. Zhang et al optimized ventilation  
144 systems to satisfy occupant thermal comfort and saved 7.8% of total energy  
145 consumption [27]. Korkas et al. proposed a study of matching energy generation and  
146 consumption with occupant behavior to guarantee occupant thermal comfort and  
147 developing demand response in microgrids with renewable energy sources [28]. Chen  
148 et al. applied occupant feedback based model predictive control (MPC) for thermal  
149 comfort and energy optimization and proposed a novel dynamic thermal sensation  
150 model, saving 25% of energy use while maintaining thermal comfort level [29]. Lim  
151 et al. discussed occupant visual comfort in office spaces based on occupants'  
152 behaviors and reported 33.39% of lighting energy saving [30]. Shen et al. integrated  
153 lighting control strategies with occupancy state to guarantee visual comfort and  
154 resulted in a 48.8% saving [31].

155

## 156 **2.2 e-CPS and occupancy information**

157 Usable and efficient building cyber models require a good understanding of occupants'  
158 energy demand and meaningful inputs from physical building systems [32,33]. Many  
159 studies suggested that the actual energy consumption of physical buildings severely  
160 deviates from the estimations of cyber models due to incorrect estimation of  
161 occupancy behavior [34]. Significant discrepancies between actual and estimated  
162 energy performance have been observed due to the complicated interrelationship  
163 between occupancy and building facility operation and the uncertainty of human  
164 behavior [35]. Oldewirtel et al. investigates the potential of using occupancy  
165 information to realize a more energy efficient building climate control and in the  
166 simulations with alternating occupancy, the savings are in the range of 50% of the

167 savings with homogeneous occupancy [36]. Hong et al. discussed ten questions  
168 concerning occupant behavior and building energy performance [37]. The  
169 International Energy Agency (IEA) Energy in Building and Community (EBC)  
170 Programme Annex 66 also highlighted and concluded that occupancy and occupants'  
171 behaviors are the most significant role for various research of enhancing building  
172 performance and human-centric services [38]. However, both physical building and  
173 cyber model are seldom changed in CPSs after the building has been built and the  
174 system uncertainties mainly arise from dynamic occupants' behavior and weather  
175 conditions. Many studies concluded that the occupancy information is one of the most  
176 significant considerations in energy conservation or low energy building design  
177 [39,40]. Therefore, as occupancy is the most critical data sources in energy demand  
178 estimation, e-CPSs should allow accurate and reliable occupancy information  
179 exchange between the physical system and cyber model.

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

196 To acquire occupancy information, researchers have proposed various methods. Jin et  
197 al. detected occupancy information through environmental sensing based on proxy  
198 measurements, such as temperature and CO<sub>2</sub> concentrations, and achieved 0.6044  
199 mean squared error and 55% ventilation cost reduction [47]. Other researchers

200 focused on using smart meters to infer occupancy presence when no data or limited  
201 data is available and reported a detection accuracy of 93% for residences and 90% for  
202 offices, respectively [48]. On the other hand, Radio frequency identification (RFID)  
203 can be applied for indoor occupant positioning, e.g. Weekly applied RFID based  
204 sampling importance resampling particle filtering algorithm for occupant positioning  
205 in a real office and achieved an accuracy of 50% estimates within 3 m range and 90%  
206 estimates within 5 m range [49]. WiFi networks are the most preferable infrastructure  
207 in existing buildings, since they are efficient, affordable, and convenient [50]. In  
208 addition, WiFi access points are usually pre-installed in most modern buildings and  
209 multiple networks can cross-reference each other. The occupants' smartphones can  
210 serve as signal receivers or tags by measuring the signal strength indicators (RSSI)  
211 and hardware addresses. Thus, with these considerations, researchers developed  
212 various WiFi-based occupancy approaches to optimize HVAC operation [51]. For  
213 example, Chen et al. showed the number of Wi-Fi connections have a positive  
214 relationship with building energy consumption [52]. Balaji utilized WiFi networks  
215 and smartphones to adjust HVAC operation setting and achieved a 17.8% electricity  
216 saving [53]. Jin et al. proposed a PresenceSense research with data collection through  
217 multiple sensing sources, including ultrasonic sensors, acceleration sensors, and WiFi  
218 [54]. Zou et al. proposed a non-intrusive occupancy sensing system, called WinOSS,  
219 to count WiFi-enabled mobile devices, which can achieve 98.85% occupancy  
220 detection accuracy when occupants stay stationary [55]. Zou et al. claimed  
221 implementing Internet of Things (IoT) technologies the counting accuracy can be as  
222 high as 99.1% [56].

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

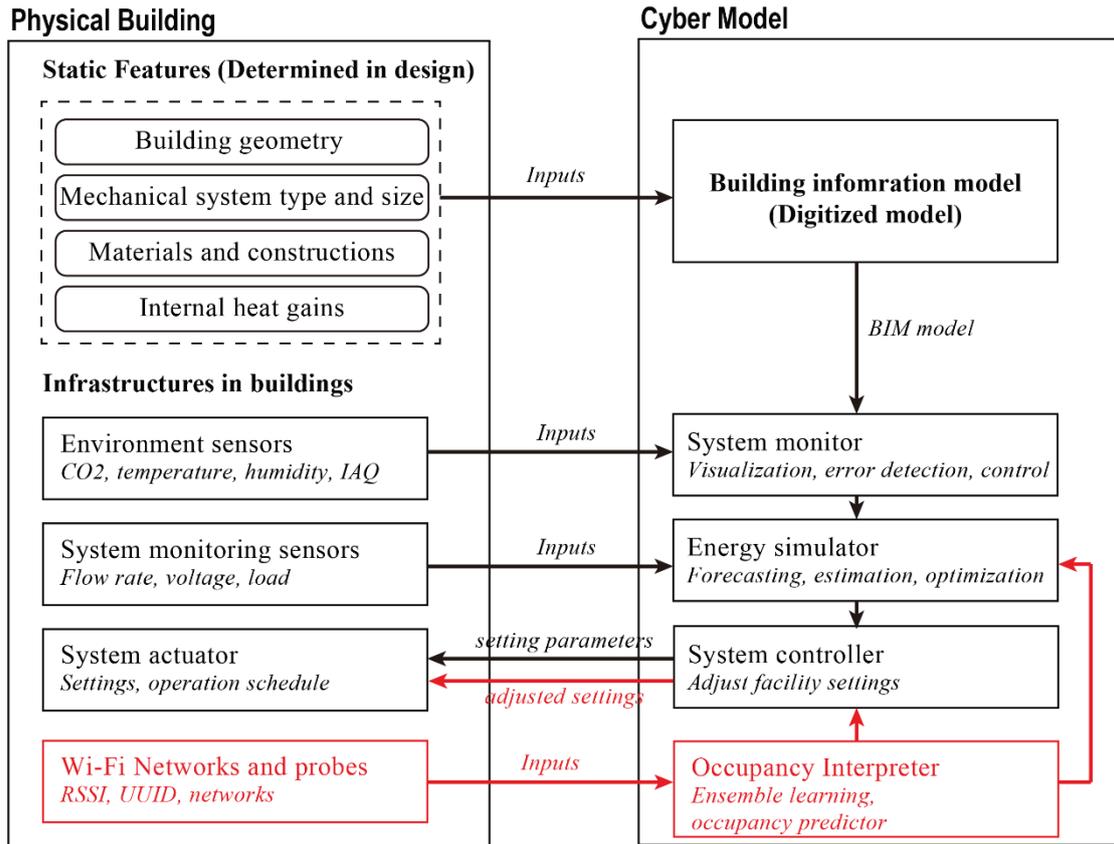
233

## 234 **3. METHODOLOGY**

### 235 **3.1 Occupancy linked e-CPSs**

236 A fundamental e-CPSs framework includes at least a physical building system and  
237 cyber model for energy management and optimization. The physical building model  
238 reflects the actual conditions and performance of a building while the cyber model is a  
239 digital twin that can be used for various computational processes. The physical  
240 buildings usually have sensors and sensor network installed which allows acquiring  
241 the various types of environmental information, such as temperatures, CO<sub>2</sub>  
242 concentration, and relative humidity (RH), and system operation information, such as  
243 supply/outdoor air flow rate and temperature, pump efficiency, and instantaneous  
244 energy load. The building information model is the key to associate both components  
245 and to create a dependable digital twin for the actual building. The building  
246 information model contains static features and dynamic operation settings. The static  
247 features include building materials, geometry, location, system type, and etc., while  
248 the dynamic operation settings include the operation schedule, efficiency, and settings  
249 of HVAC, lighting, and security systems.

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



\* Conventional e-CPSs

\* Occupancy linked e-CPSs

261

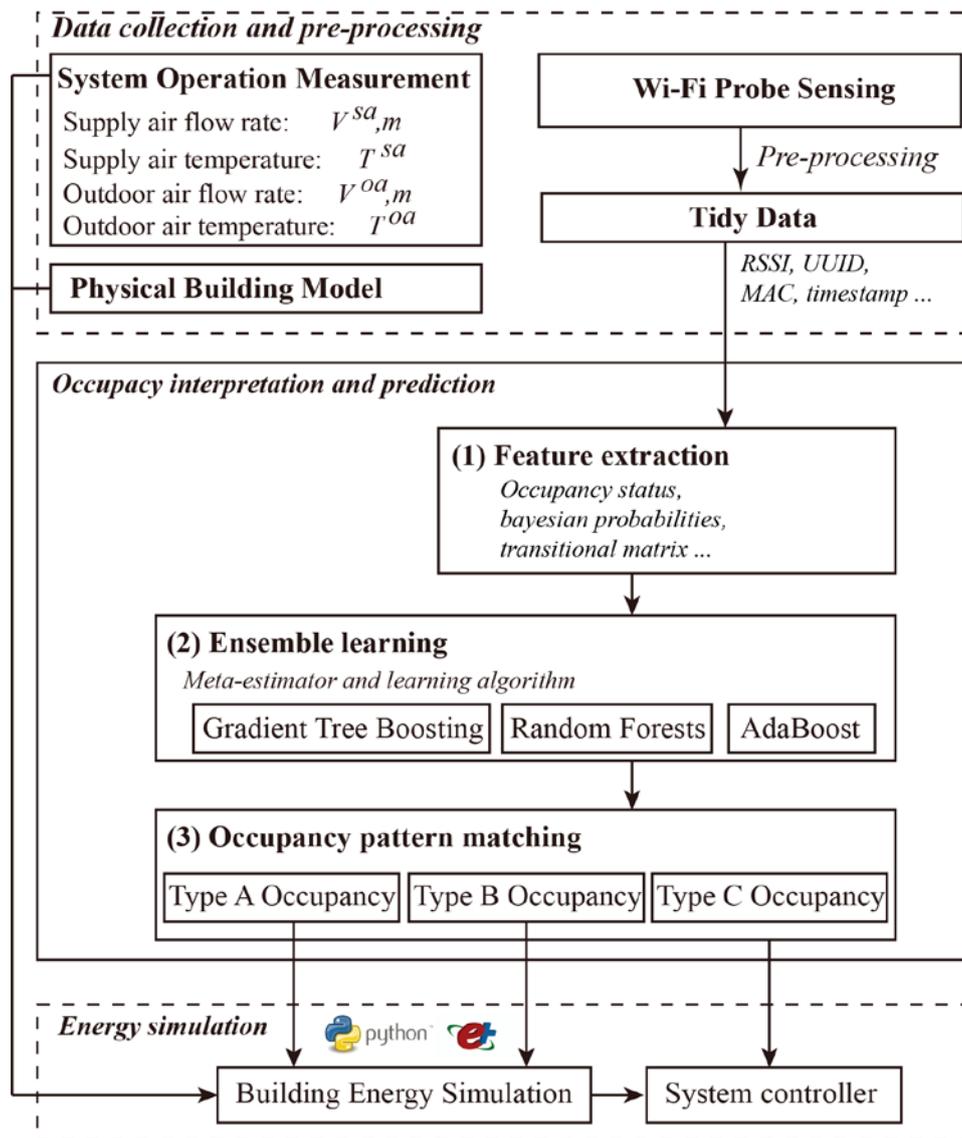
262

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

263

### 264 3.2 Wi-Fi Probe-based ensemble learning algorithm for occupancy prediction

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



273

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

275

### 276 3.2.1 Feature extraction

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

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

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

$$x_k^{i-i} = P(\text{observed state} = i | \text{observed state} = i) \quad (2)$$

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

$$x_k^{i-i} = \frac{\sum N_{i-i}}{\sum N_{i-i} + \sum N_{i-o}} \quad x_k^{o-o} = \frac{\sum N_{o-o}}{\sum N_{o-o} + \sum N_{o-i}} \quad (3)$$

292 Where  $N_{i-i}$  is the frequency in which the occupancy status transfers from “in” to  
 293 “in”.  $N_{i-o}$  is the frequency in which the occupancy status transfers from “in” to “out”.  
 294 Similarly,  $N_{o-o}$  and  $N_{o-i}$  represent the frequencies in which the occupancy status  
 295 transitioned from “out” to “out” and from “out” to “in”, respectively. With an  
 296 assigned probability for MAC addresses in the room. Each MAC address is formatted  
 297 as

$$x_k = \{x_k^{Mac}, x_k^{o-i}, x_k^{i-i}\} \quad (4)$$

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

$$X(t) = \{x_1^{Mac}, x_1^{o-i}, x_1^{i-i}, \dots, x_k^{Mac}, x_k^{o-i}, x_k^{i-i}, \dots, x_n^{Mac}, x_n^{o-i}, x_n^{i-i}\} \quad (5)$$

300

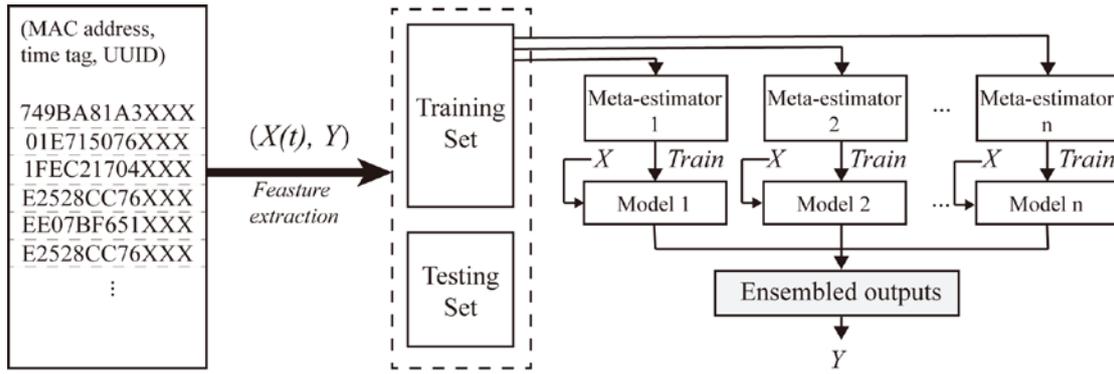
### 301 3.2.2 Ensemble learning algorithms

302 There are main families of ensemble methods. The first method is averaging, which  
 303 builds several estimators independently and then average predictions through  
 304 minimizing their prediction variance, such as Bagging methods and Forests of  
 305 Randomized Trees. The second method is boosting, which builds sequential  
 306 estimators to reduce the bias by combining several weak models, such as AdaBoost  
 307 and Gradient Tree Boosting. The ensemble learning algorithm in this study integrates

308 multiple meta-estimators through boosting method.

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

320



321

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

323

324 Suppose the ensemble outputs can be estimated from the aggregated results from  
 325 multiple meta-estimators as:

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

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

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

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

$$F_m(x) = F_{m-1}(x) + \underset{f}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + f(x)) \quad (8)$$

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

335

### 336 *(1) Gradient Tree Boosting (GTB)*

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

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

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

353

### 354 *(2) Random Forests (RF)*

355 Random Forests (RF) is another ensemble machine learning algorithm that follows

356 the bagging technique. The base estimators in random forest are decision trees. Unlike  
357 bagging meta estimator, RF classifier randomly selects a set of features which are  
358 used to decide the best split from the training set. By doing this, the sample bias can  
359 be eliminated and the best split among trees can be selected. With averaging, the  
360 variance of meta-estimators can be minimized, hence yielding a better model.

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

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

371

### 372 (3) *Adaptive Boosting (AdaBoost)*

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

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

- 383 • Assign equal weights to all observations in the dataset.
- 384 • Rule models are built for subsets and compute the predictions for the whole  
385 data set.

- 386 • Compute errors by comparing the predictions and actual data. Update the rule  
 387 models and assign higher weights for incorrectly predicted observations.
- 388 • Repeat above steps until errors are minimized.

389

### 390 3.2.3 Occupancy pattern matching

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

$$Q_{nor,r} = Q_{inf,r} + Q_{surf,r} \quad (9)$$

$$Q_{inf,r} = m_{inf,r} * C_p * (T_{in,r} - T_{air}) \quad (10)$$

$$Q_{surf,r} = A_{surf,r} * K_{surf} * (T_{in,r} - T_{air}) \quad (11)$$

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

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

$$Q_{or,r} = \sum_{P_r} G_p + \sum_{p_{eq}} G_{eq} + \sum G_{other} \quad (12)$$

$$Q_r = Q_{nor,r} + Q_{or,r} \quad (13)$$

403 Where  $P_r$  is the number of occupants and  $G_p$  is the heat gain from per occupant.  
 404  $G_{eq}$  contains the load from computers, water heaters, lights etc.;  $p_{eq}$  is the index of  
 405 equipment;  $Q_r$  is the total cooling load of a room. At room level, the ventilation and  
 406 air conditioning system should provide enough conditioned air to maintain proper  
 407 indoor temperature and the air handling system should supply sufficient fresh air.

$$E_r = Q_r = m_r * C_p * (T_{in,r} - T_{s,r}) \quad (14)$$

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

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

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

418 Then,

$$\begin{aligned} E_{ven,r} = Q_{vent,r} &= m_{OA,r} * (h_{OA} - h_{in}) \\ &= m_{OA,r} * (f(T_{air}, H_{air}) - f(T_{in,r}, H_{in,r})) \end{aligned} \quad (16)$$

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

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

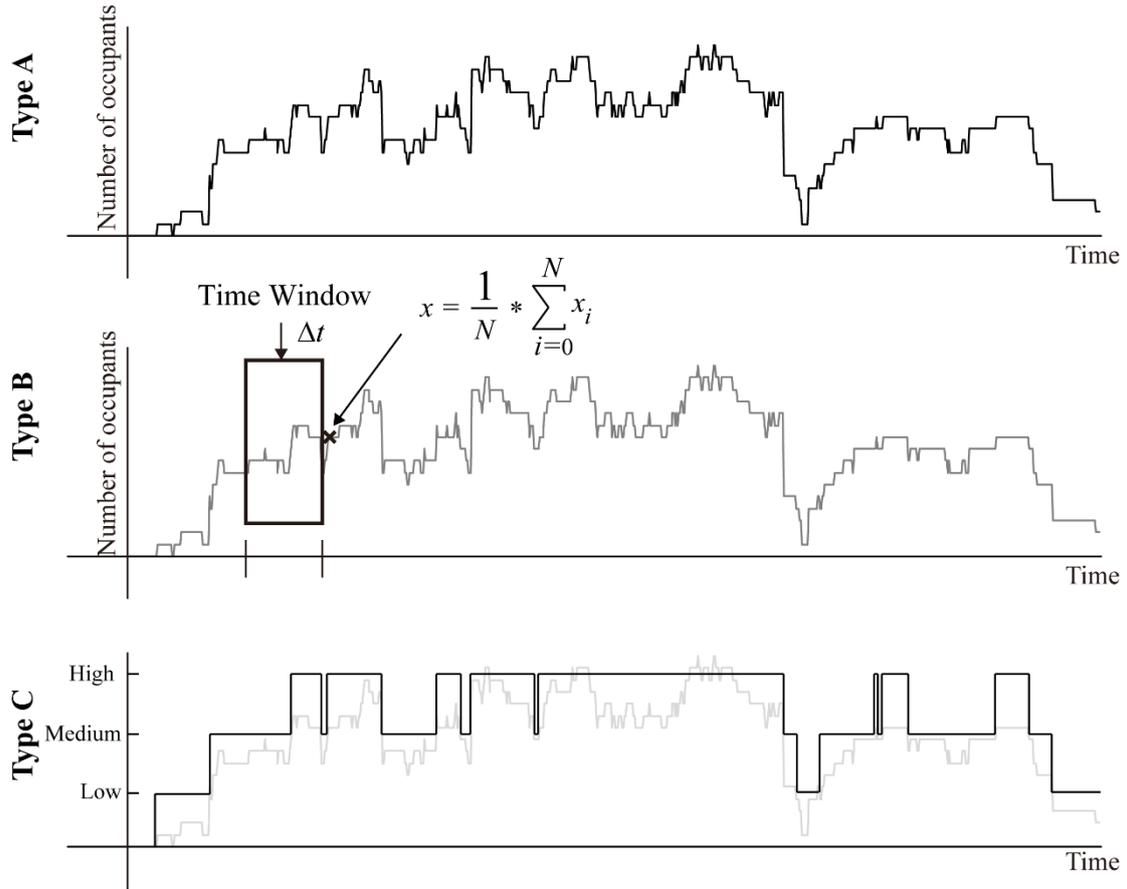


Fig. 4. Sample occupancy schedules for three occupancy types.

**(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.

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

Where  $T_{in,r} = T_{setting}$  and  $H_{in,r} = H_{setting}$  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**

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

$$p_r = p_{pred. B}^r = \frac{t_0}{T} * \sum_{i=0}^{T/t_0} x_i \quad (18)$$

450 Where  $p_{pred. B}^r$  is the predicted occupancy.  $t_0$  is the time resolution of the  
 451 occupancy.  $T$  is the length of the averaging time window.

452

### 453 (3) Type C occupancy

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

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

- 
1. Feature abstraction from Wi-Fi dataset
  2. Define occupancy patterns
  3. Define Input  $X = \{x_1, x_2, \dots, x_n\}$ , Output  $y = \{y_1, y_2, \dots, y_n\}$ , a set of base estimators  $F = \{f_1(x), f_2(x), \dots, f_M(x)\}$ . 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,  $w_m = \operatorname{argmin}_f \{L(y_i, F_{m-1}(x) + f_m(x))\}$
      - (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
- 

460

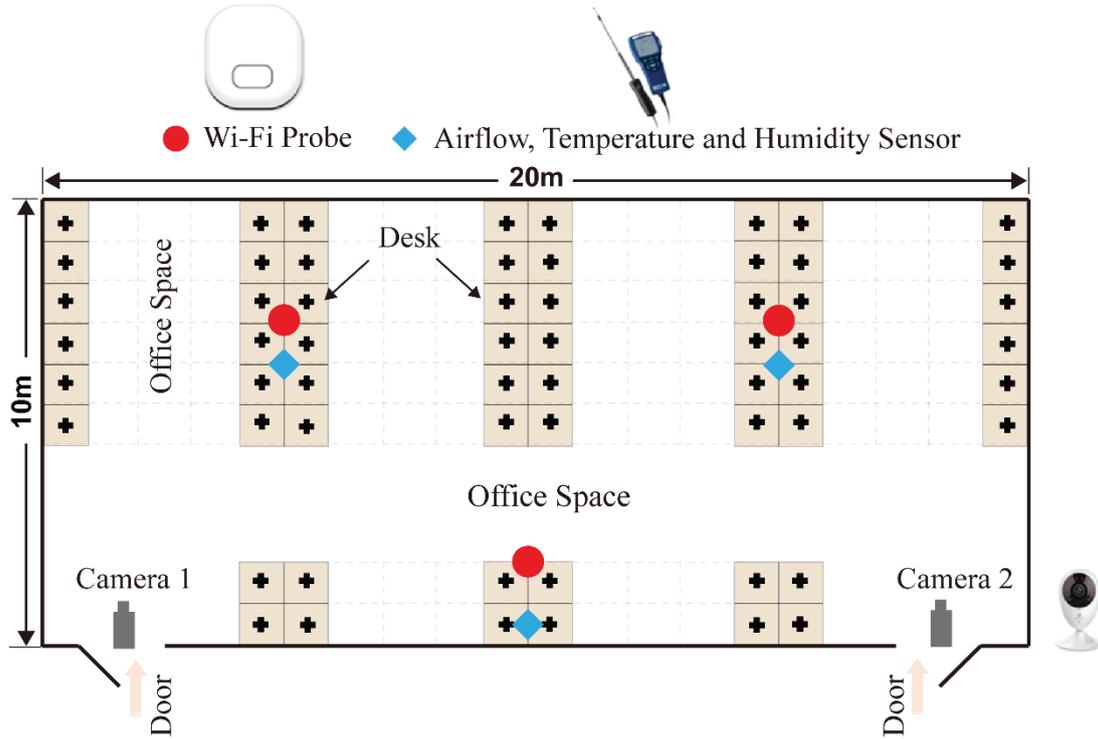
461 Fig. 5. The pseudocode of the ensemble algorithm for occupancy prediction

462

## 463 **4. VALIDATION EXPERIMENT**

### 464 **4.1 Physical conditions of the experiment testbed**

465 To examine the proposed occupancy linked e-CPSs, this study also conducted a  
466 validation experiment in a large office space. The testbed has an area of about 200  
467 square meters and 20 long-term residents during the experiment period. Figure 6  
468 shows the space layout and sensors setup. The room equipped with a dedicated  
469 outdoor air system to bring outdoor air into indoor areas without air handling process.  
470 The indoor air is conditioned by the fan coil unit with the variable refrigerant flow  
471 and the indoor air circulation is driven by positive pressure. The entire room has  
472 Wi-Fi coverage with three Wi-Fi probes. During the experiment, TA465-X sensor  
473 system (produced by TSI Co.) was utilized to monitor the indoor air temperature,  
474 relative humidity, and airflow rate. The CO<sub>2</sub> concentration of return air of the fan coil  
475 unit was used to approximate the CO<sub>2</sub> concentration of the indoor air after air mixing.  
476 To eliminate the uneven air mixing, three environmental sensors were evenly installed  
477 at the ceiling (3m). Air flow meters were installed near outdoor inlets to monitor the  
478 air flow rate of the ventilation system. Two overhead cameras were installed to record  
479 the entrance and exit events of occupants. During the experiment, the occupants aware  
480 of the Wi-Fi experiment and were instructed to switch on their Wi-Fi signal on their  
481 mobile devices. Table 1 shows the specifications of the installed sensors, including  
482 data storage types, sensing intervals, range, accuracy, and resolution. The experiment  
483 lasted for nine days.



484

485

Fig. 6. Space layout and equipment setup.

486

487

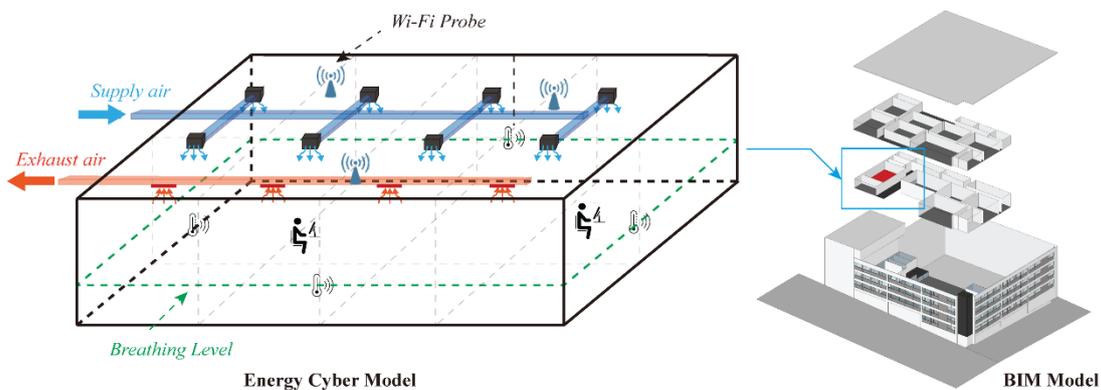
Table 1. Sensors used in the experiment.

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

488

489 **4.2 Cyber model for energy management and simulation**

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



506

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

508 The cyber model matches the physical room with a size of 20 m (length) x 10 m  
509 (width) x 3 m (height) and 20 occupants. Internal heat sources were set as 75W for  
510 per person, 150W for per computer, and 35W for per lamp. The light schedule  
511 followed the on/off schedule and the schedule for computers was assumed to same as  
512 the occupancy schedule. Hong Kong has a subtropical climate and high-density  
513 highrise urban form. According to statistics [62], the typical mean, minimum,

514 maximum values of monthly average temperature are around 23.4°C, 13.3°C, and  
515 29.8°C, respectively. Also, relative humidity (RH) of Hong Kong is high and  
516 minimum, maximum values of monthly average RH are 78.2%, 60%, and 90%,  
517 respectively. The typical Hong Kong weather condition was used and the heat  
518 transfers from wall, floor, and ceiling were ignored since the experiment was  
519 conducted in one inner zone adjacent to conditioned zones. The cooling temperature  
520 setpoint is 24°C and there was no heating.

521

## 522 **4.3 Data processing**

### 523 ***4.3.1 Actual occupancy information***

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

531 Table 2. The threshold setting for categorical occupancy levels

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

532

### 533 ***4.3.2 Model parameters tuning***

534 To improve the facility operation with reliable occupancy information, it is necessary  
535 to identify, compare, and optimize the ensemble model through parameter tuning. The  
536 training model implemented n-fold cross-validation method. In this study, the raw  
537 dataset has total 882 samples and about 70% of dataset was used for model training  
538 and 30% for model validation and test. Table 3 shows the search space for the  
539 parameters tuning. The multi-variable comparison in the exhaustive grid search is

540 applied to identify the best assembly of model parameters. For the RF classifier, the  
 541 number of estimators determines the results precision and training time, while the  
 542 number of features affects the accuracy and the diversity of results. For GTB and  
 543 AdaBoost classifiers, learning rate affects the boosting step length of the gradient  
 544 descent procedure.

545 Table 3. Parameters search space for the occupancy ensembled model

Algorithm	Parameter	Range
<b>GTB</b>	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]
<b>AdaBoost</b>	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]
<b>Random Forest</b>	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]

546

### 547 **4.3.3 Error assessment**

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

$$TPR = \frac{TP}{TP + FP} \quad (19)$$

$$TNR = \frac{TN}{TN + FN} \quad (20)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

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

$$bACC = \frac{TPR + TNR}{2} \quad (22)$$

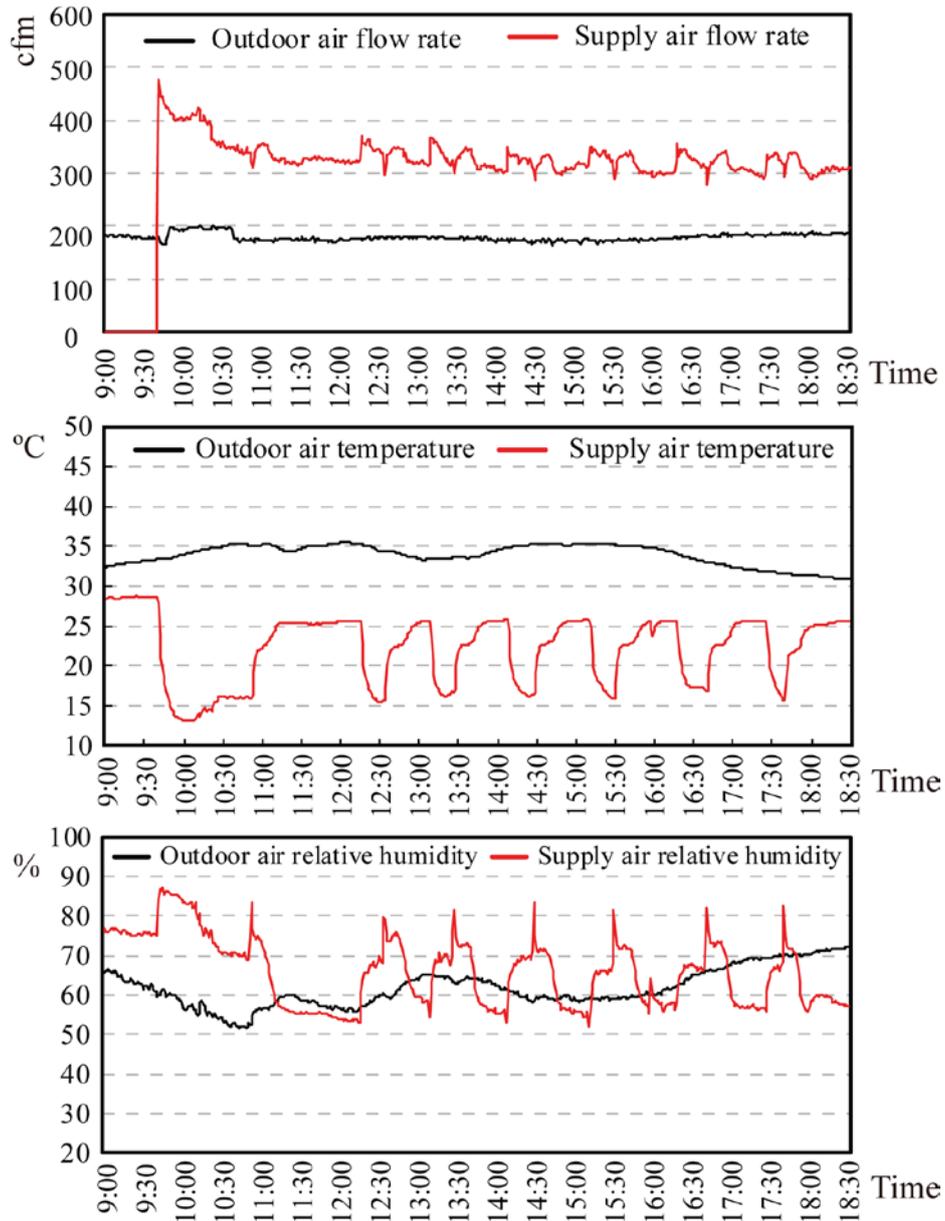
558 According to ASHRAE standard 62.1-2013 [60], the fresh air volume of the  
559 ventilation system and the occupant-related thermal load of the air conditioning  
560 system are determined by the number of occupants. The errors in the occupancy  
561 assessment could directly affect the energy usage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

563

## 564 **5. RESULTS**

### 565 **5.1 Environmental conditions**

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



578

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

581

## 582 5.2 Predicted occupancy

583 This study performed a grid search to determine optimal values for the parameters of  
 584 the tree-based ensembles. The features of Wi-Fi dataset described in Eq. 5 were  
 585 considered as the input variables. The GTB classifier consists of 150 estimators with a  
 586 learning rate of 0.01. To split an internal node, the model requires a minimum 8  
 587 samples and a maximum tree depth of 15. The AdaBoost classifier has 100 estimators

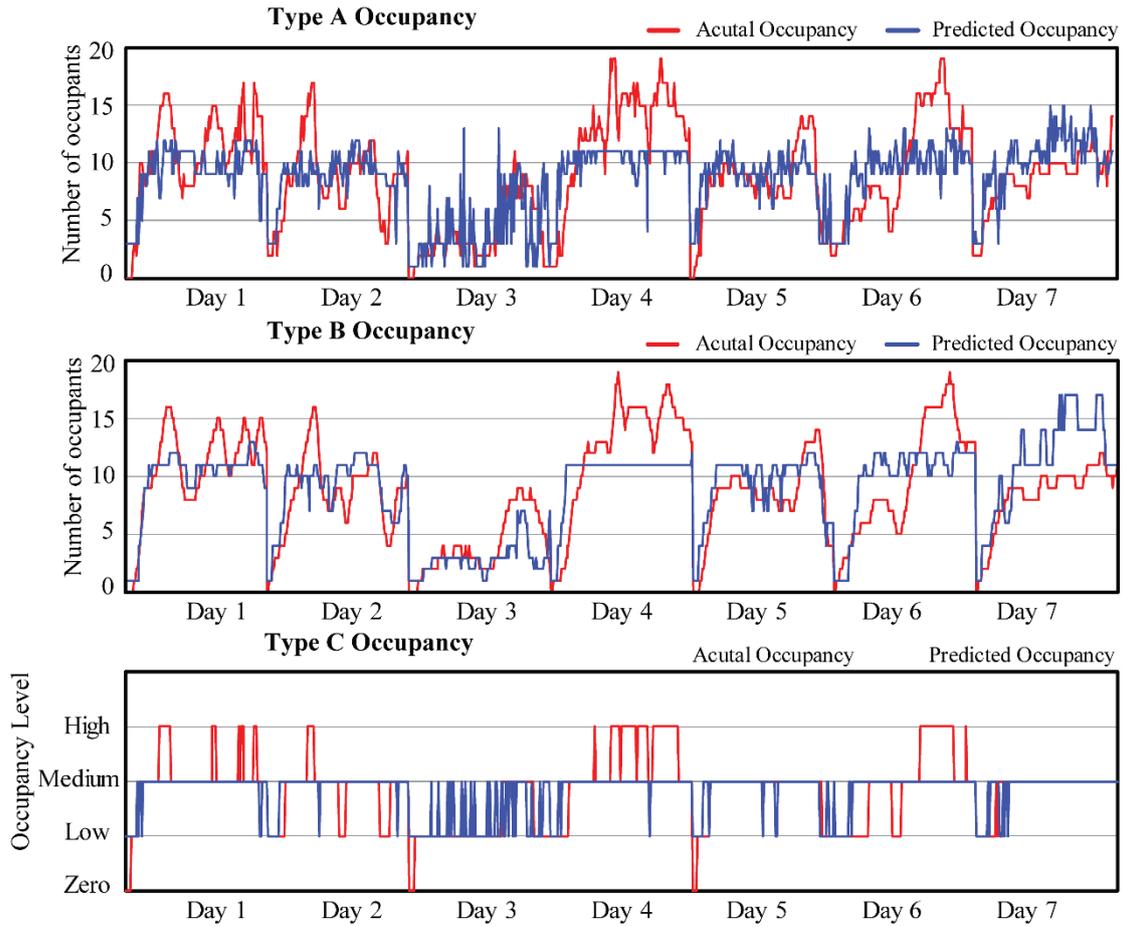
588 with a learning rate of 0.2. The RF classifier has 250 estimators and 10 minimum  
 589 sample leaf. Table 4 summaries the averaged errors of all three type of classifiers after  
 590 tuning. Among all three types of classifiers, the AdaBoost classifier shows the highest  
 591 accuracy.

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

	<b>RFs</b>			<b>GTB</b>			<b>AdaBoost</b>		
	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%

593

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



612

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

615

Table 5. Averaged errors and accuracy of three occupancy types

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

616

617

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

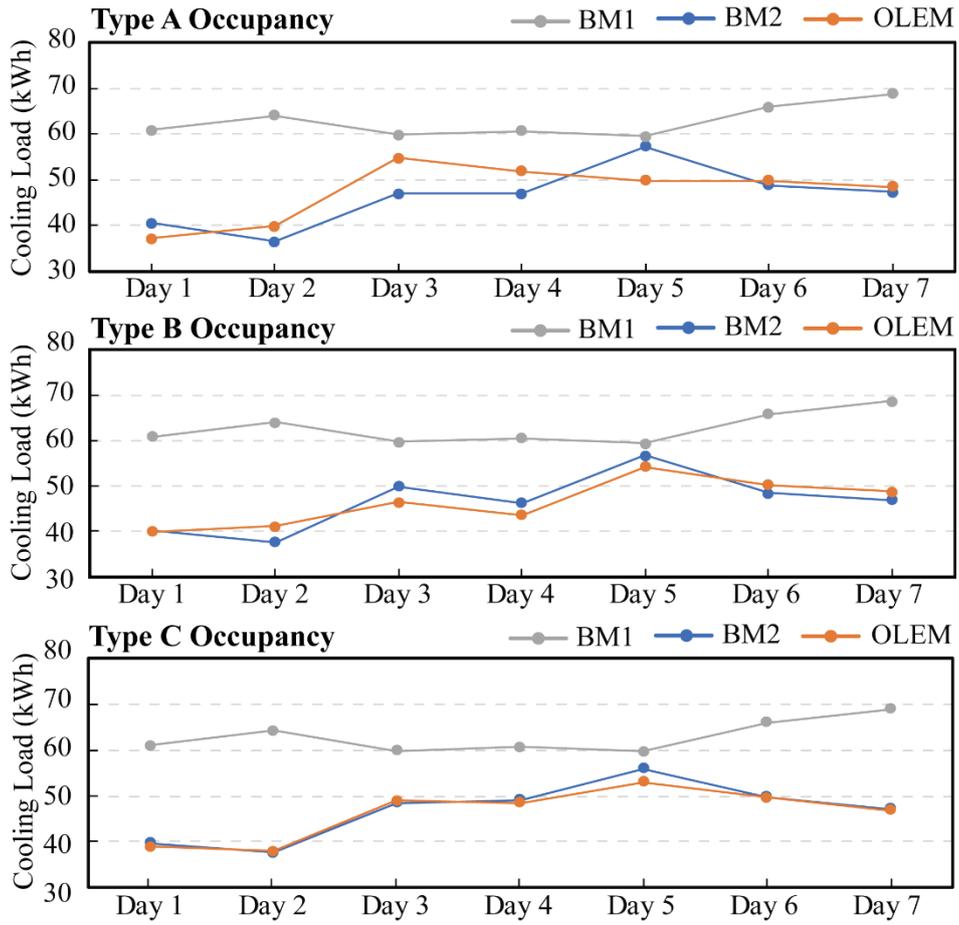
	<b>Zero</b>	<b>Low</b>	<b>Medium</b>	<b>High</b>
<b>Zero</b>	0.00	1.00	0.00	0.00
<b>Low</b>	0.00	0.60	0.40	0.00
<b>Medium</b>	0.00	0.05	0.95	0.00
<b>High</b>	0.00	0.00	1.00	0.00

618

### 619 **5.3 Energy performance and analysis of the occupancy linked e-CPSs**

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

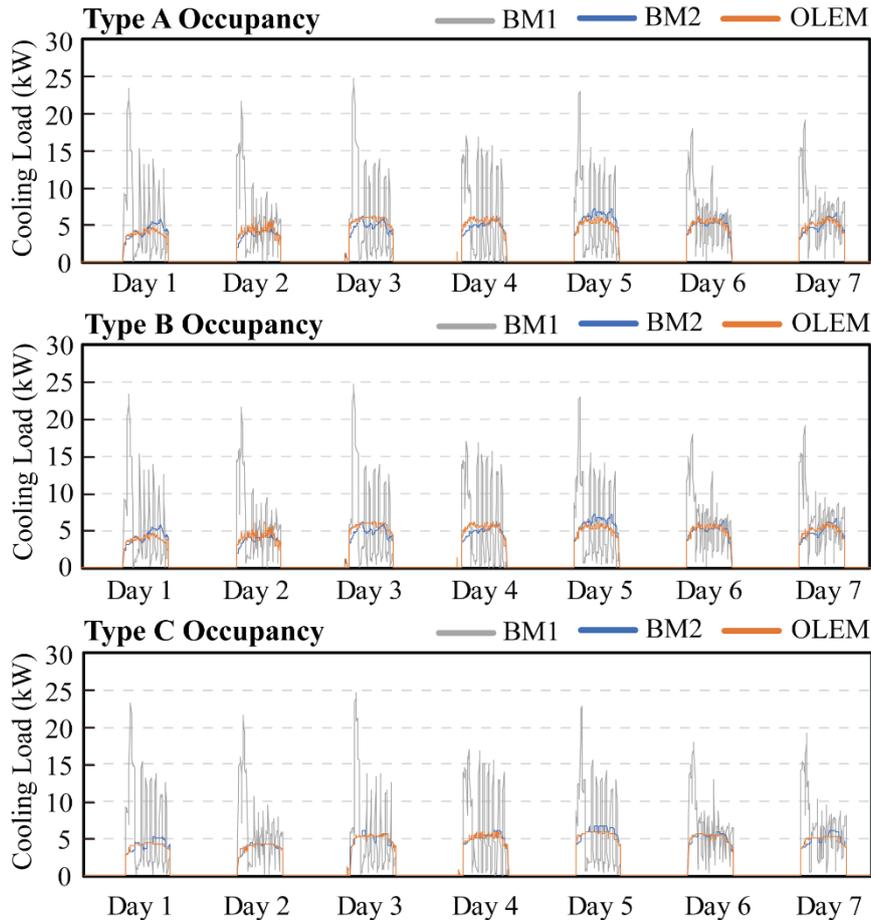
630 Figure 10 and 11 shows the simulated cooling load with different occupancy types. In  
631 the simulation, the thermostat HVAC terminals in BM1 were set to default  
632 temperature and the mechanical operation was mainly affected by the weather  
633 condition. From both figures, it can be seen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



637

638

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



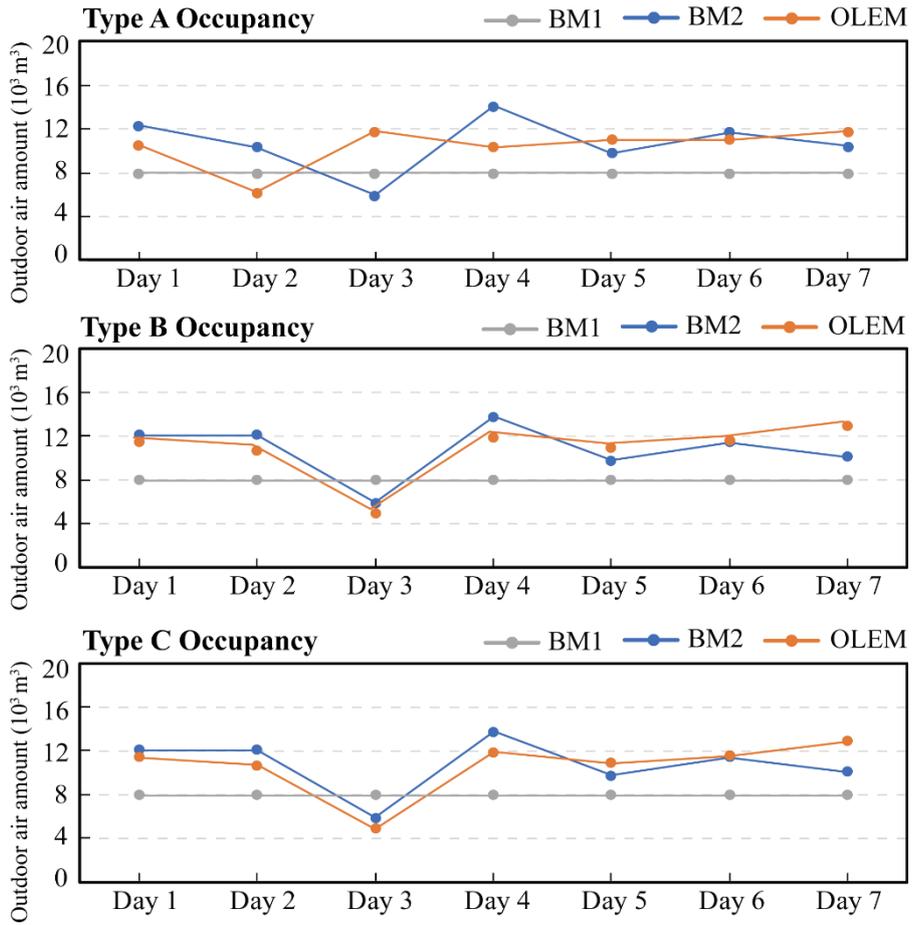
639

640

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

641

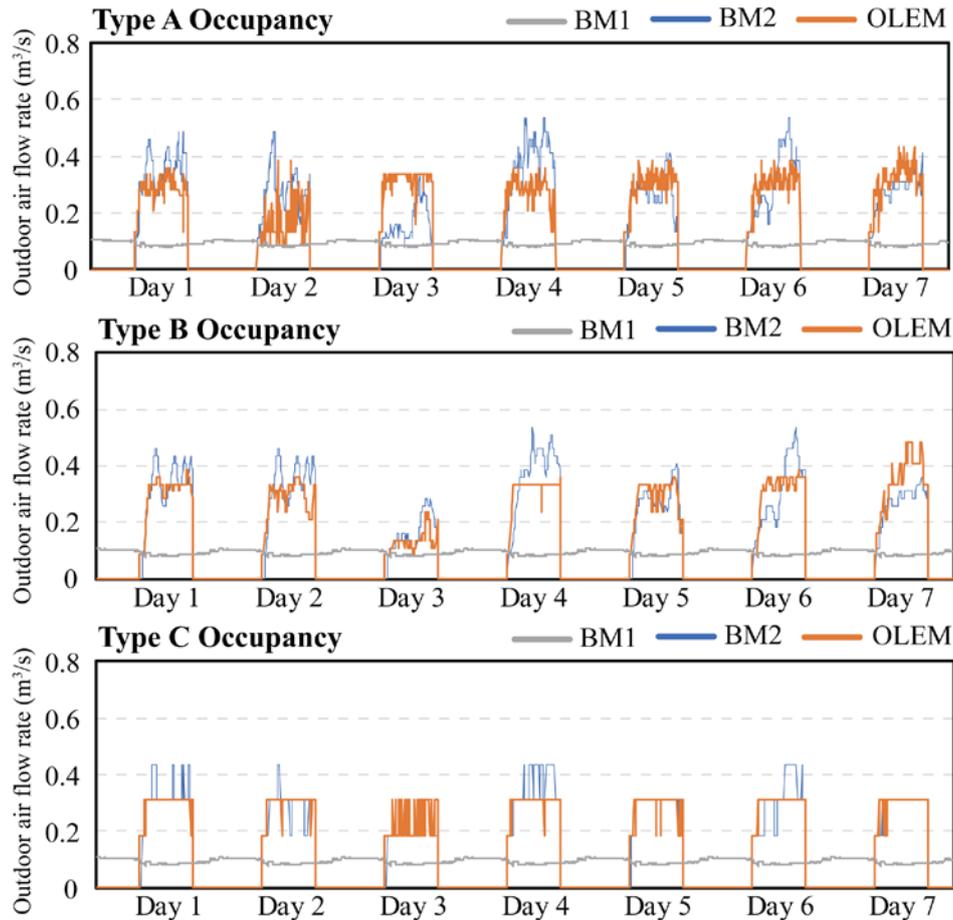
642 Another energy consumption component for the HVAC system is the fresh air amount.  
 643 The mechanical drives and fans consume a large amount of energy when the air  
 644 handling units deliver the outdoor air into indoor spaces. The physical building  
 645 deploys on/off the system with a fixed flow rate about 1440 m<sup>3</sup>/h. However,  
 646 according to ASHRAE standard, the flow amount is obviously insufficient given the  
 647 number of occupants in the experiment office. Figure 12 and 13 show the simulated  
 648 minimum outdoor air flow rate and amount. Both figures suggest that the outdoor air  
 649 amount in BM1 is far less than the demand according to the number of occupants.  
 650 Type A occupancy performs the worst among all three types, this could be caused by  
 651 the tracking errors result from data fluctuation.



652

653

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



654

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

656

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

664 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%

665

## 666 **6. DISCUSSION**

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

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

720 This study also yields to limitations, which can be resolved in future studies. Firstly,  
721 the validation experiment constraint to small space (an office room). It is suggested to  
722 study a larger building space with multiple rooms so that the impact of indoor  
723 commutes can be included. Also, rooms with different functions also have their  
724 unique occupancy patterns and mechanical system selection. Secondly, the energy

725 consumption in this study mainly results from cooling load and ventilation due to the  
726 tropical climate condition and short experiment period. However, there are various  
727 energy consuming services systems in buildings, such as lighting, security, heating,  
728 and etc., which are also closely associated with human behaviors and inter-dependent  
729 with each other.

730

## 731 **7. CONCLUSION**

732 This study proposed a theoretical framework for implementing occupancy  
733 information as dynamic links for e-CPSs. The framework adopted WiFi Probe  
734 technology and ensemble classifiers to interpret WiFi connections as reliable and  
735 usable occupancy information. Three occupancy types (Type A, B, and C) have been  
736 compared in a validation experiment to examine the accuracy and feasibility of the  
737 proposed occupancy-linked e-CPSs. After a validation experiment, the proposed  
738 model can accurately report occupant counts for system energy management. The  
739 AdaBoost method and type C occupancy report the highest detection accuracy of  
740 72.7%. Type A occupancy has an absolute error and root mean squared error of 2.54  
741 and 3.30, and both values for type B occupancy are 2.41 and 3.06, respectively. The  
742 energy simulation reports 24.7%, 26.4%, and 26.3% energy saving potentials by  
743 implementing these three types of occupancy information in e-CPSs, respectively.

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

751

## 752 **ACKNOWLEDGMENT**

753 This work was financially supported by the Hong Kong General Research Fund (GRF)  
754 – Early Career Scheme, #21204816, and the National Natural Science Foundation of  
755 China (NSFC), #51508487. Any opinions, findings, conclusions, or recommendations

756 expressed in this paper are those of the authors and do not necessarily reflect the  
757 views of GRF and NSFC.  
758

759 **REFERENCES**

- 760 [1] U.S. Energy Information Administration 2017.  
761 [https://www.eia.gov/energyexplained/index.cfm?page=us\\_energy\\_use](https://www.eia.gov/energyexplained/index.cfm?page=us_energy_use).
- 762 [2] Rajkumar R, Lee ILI, Sha LSL, Stankovic J. Cyber-physical systems: The next  
763 computing revolution. Design Automation Conference (DAC), 2010 47th  
764 ACM/IEEE 2010:0–5. doi:10.1145/1837274.1837461.
- 765 [3] Gupta SKS, Mukherjee T, Varsamopoulos G, Banerjee A. Research directions  
766 in energy-sustainable cyber-physical systems. Sustainable Computing:  
767 Informatics and Systems 2011;1:57–74. doi:10.1016/j.suscom.2010.10.003.
- 768 [4] Peng Rong, Pedram M. Power-aware scheduling and dynamic voltage setting  
769 for tasks running on a hard real-time system. Asia and South Pacific  
770 Conference on Design Automation, 2006., IEEE; 2006, p. 473–8.  
771 doi:10.1109/ASPDAC.2006.1594730.
- 772 [5] Dobson I. Energy Cyber-Physical Systems : Research Challenges and  
773 Opportunities 2016.
- 774 [6] SzilagyI I, Wira P. An Intelligent System for Smart Buildings using Machine  
775 Learning and Semantic Technologies : A Hybrid Data-Knowledge Approach  
776 2018:20–5.
- 777 [7] Kleissl J, Agarwal Y. Cyber-physical energy systems: focus on smart buildings.  
778 Proceedings of the 47th Design Automation Conference on - DAC '10  
779 2010:749. doi:10.1145/1837274.1837464.
- 780 [8] Balaji B, Abdullah M, Faruque A, Dutt N, Gupta R, Agarwal Y. Models ,  
781 Abstractions , and Architectures : The Missing Links in Cyber-Physical  
782 Systems n.d.
- 783 [9] Zhao P, Simoes MG, Suryanarayanan S. A conceptual scheme for  
784 cyber-physical systems based energy management in building structures. 2010  
785 9th IEEE/IAS International Conference on Industry Applications -  
786 INDUSCON 2010 2010:1–6. doi:10.1109/INDUSCON.2010.5739891.
- 787 [10] Paridari K, El-Din Mady A, La Porta S, Chabukswar R, Blanco J, Teixeira A,  
788 et al. Cyber-Physical-Security Framework for Building Energy Management  
789 System 2016. doi:10.1109/ICCPS.2016.7479072.
- 790 [11] Francisco A, Truong H, Khosrowpour A, Taylor JE, Mohammadi N. Occupant  
791 perceptions of building information model-based energy visualizations in  
792 eco-feedback systems. Applied Energy 2018;221:220–8.  
793 doi:10.1016/j.apenergy.2018.03.132.
- 794 [12] Hong T, Chou S., Bong T. Building simulation: an overview of developments  
795 and information sources. Building and Environment 2000;35:347–61.

- 796 doi:10.1016/S0360-1323(99)00023-2.
- 797 [13] Modelica and the Modelica Association — Modelica Association n.d.  
798 <https://www.modelica.org/>.
- 799 [14] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO,  
800 et al. EnergyPlus: creating a new-generation building energy simulation  
801 program. *Energy and Buildings* 2001;33:319–31.  
802 doi:10.1016/S0378-7788(00)00114-6.
- 803 [15] Delwati M, Merema B, Breesch H, Helsen L, Sourbron M. Impact of demand  
804 controlled ventilation on system performance and energy use. *Energy and*  
805 *Buildings* 2018;174:111–23. doi:10.1016/j.enbuild.2018.06.015.
- 806 [16] Hong T, Sun K, Zhang R, Hinokuma R, Kasahara S, Yura Y. Development and  
807 validation of a new variable refrigerant flow system model in EnergyPlus.  
808 *Energy and Buildings* 2016;117:399–411.  
809 doi:10.1016/J.ENBUILD.2015.09.023.
- 810 [17] Stamatescu G, Stamatescu I, Arghira N, Calofir V, Fagarasan I. *Building*  
811 *Cyber-Physical Energy Systems* 2016.
- 812 [18] Behl M, Jain A, Mangharam R. Data-Driven Modeling, Control and Tools for  
813 Cyber-Physical Energy Systems. 2016 ACM/IEEE 7th International  
814 Conference on Cyber-Physical Systems, ICCPS 2016 - Proceedings 2016.  
815 doi:10.1109/ICCPS.2016.7479093.
- 816 [19] Ferreira PM, Ruano AE, Silva S, Conceição EZE. Neural networks based  
817 predictive control for thermal comfort and energy savings in public buildings.  
818 *Energy and Buildings* 2012;55:238–51. doi:10.1016/J.ENBUILD.2012.08.002.
- 819 [20] Costanzo GT, Iacovella S, Ruelens F, Leurs T, Claessens BJ. Experimental  
820 analysis of data-driven control for a building heating system. *Sustainable*  
821 *Energy, Grids and Networks* 2016;6:81–90.  
822 doi:10.1016/J.SEGAN.2016.02.002.
- 823 [21] Belafi Z, Hong T, Reith A. Smart building management vs. intuitive human  
824 control—Lessons learnt from an office building in Hungary. *Building*  
825 *Simulation* 2017;10:811–28. doi:10.1007/s12273-017-0361-4.
- 826 [22] Wang Y, Shao L. Understanding occupancy pattern and improving building  
827 energy efficiency through Wi-Fi based indoor positioning. *Building and*  
828 *Environment* 2017;114:106–17. doi:10.1016/j.buildenv.2016.12.015.
- 829 [23] Menezes AC, Cripps A, Bouchlaghem D, Buswell R. Predicted vs. actual  
830 energy performance of non-domestic buildings: Using post-occupancy  
831 evaluation data to reduce the performance gap. *Applied Energy* 2012;97:355–  
832 64. doi:10.1016/j.apenergy.2011.11.075.
- 833 [24] Liang X, Hong T, Shen GQ. Improving the accuracy of energy baseline models

- 834 for commercial buildings with occupancy data. *Applied Energy* 2016;179:247–  
835 60. doi:10.1016/j.apenergy.2016.06.141.
- 836 [25] Wang W, Chen J, Huang G, Lu Y. Energy efficient HVAC control for an  
837 IPS-enabled large space in commercial buildings through dynamic spatial  
838 occupancy distribution. *Applied Energy* 2017.  
839 doi:10.1016/J.APENERGY.2017.06.060.
- 840 [26] Barbeito I, Zaragoza S, Tarrío-Saavedra J, Naya S. Assessing thermal comfort  
841 and energy efficiency in buildings by statistical quality control for  
842 autocorrelated data. *Applied Energy* 2017;190:1–17.  
843 doi:10.1016/J.APENERGY.2016.12.100.
- 844 [27] Zhang S, Cheng Y, Fang Z, Huan C, Lin Z. Optimization of room air  
845 temperature in stratum-ventilated rooms for both thermal comfort and energy  
846 saving. *Applied Energy* 2017;204:420–31.  
847 doi:10.1016/J.APENERGY.2017.07.064.
- 848 [28] Korkas CD, Baldi S, Michailidis I, Kosmatopoulos EB. Occupancy-based  
849 demand response and thermal comfort optimization in microgrids with  
850 renewable energy sources and energy storage. *Applied Energy* 2016;163:93–  
851 104. doi:10.1016/j.apenergy.2015.10.140.
- 852 [29] Chen X, Wang Q, Srebric J. Occupant feedback based model predictive control  
853 for thermal comfort and energy optimization: A chamber experimental  
854 evaluation. *Applied Energy* 2016;164:341–51.  
855 doi:10.1016/j.apenergy.2015.11.065.
- 856 [30] Lim G-H, Keumala N, Ghafar NA. Energy saving potential and visual comfort  
857 of task light usage for offices in Malaysia. *Energy and Buildings*  
858 2017;147:166–75. doi:10.1016/J.ENBUILD.2017.05.004.
- 859 [31] Shen E, Hu J, Patel M. Energy and visual comfort analysis of lighting and  
860 daylight control strategies. *Building and Environment* 2014;78:155–70.  
861 doi:10.1016/J.BUILDENV.2014.04.028.
- 862 [32] Siano P. Demand response and smart grids—A survey. *Renewable and*  
863 *Sustainable Energy Reviews* 2014;30:461–78. doi:10.1016/j.rser.2013.10.022.
- 864 [33] Strbac G. Demand side management: Benefits and challenges. *Energy Policy*  
865 2008;36:4419–26. doi:10.1016/j.enpol.2008.09.030.
- 866 [34] Nguyen A-T, Reiter S, Rigo P. A review on simulation-based optimization  
867 methods applied to building performance analysis. *Applied Energy*  
868 2014;113:1043–58. doi:10.1016/j.apenergy.2013.08.061.
- 869 [35] Díaz JA, Jiménez MJ. Experimental assessment of room occupancy patterns in  
870 an office building. Comparison of different approaches based on CO<sub>2</sub>  
871 concentrations and computer power consumption. *Applied Energy*

- 872 2017;199:121–41. doi:10.1016/j.apenergy.2017.04.082.
- 873 [36] Oldewurtel F, Sturzenegger D, Morari M. Importance of occupancy  
874 information for building climate control. *Applied Energy* 2013;101:521–32.  
875 doi:10.1016/j.apenergy.2012.06.014.
- 876 [37] Hong T, Yan D, D’oca S, Chen C-F. Ten questions concerning occupant  
877 behavior in buildings: The big picture. *Building and Environment*  
878 2017;114:518–30. doi:10.1016/j.buildenv.2016.12.006.
- 879 [38] Yan D, Hong T, Dong B, Mahdavi A, D’Oca S, Gaetani I, et al. IEA EBC  
880 Annex 66: Definition and simulation of occupant behavior in buildings. *Energy*  
881 and Buildings 2017;156:258–70. doi:10.1016/J.ENBUILD.2017.09.084.
- 882 [39] Hong T, Taylor-Lange SC, D’oca S, Yan D, Corgnati SP, D’Oca S, et al.  
883 Advances in research and applications of energy-related occupant behavior in  
884 buildings. *Energy and Buildings* 2016;116:694–702.  
885 doi:10.1016/j.enbuild.2015.11.052.
- 886 [40] Yan D, O’Brien W, Hong T, Feng X, Burak Gunay H, Tahmasebi F, et al.  
887 Occupant behavior modeling for building performance simulation: Current  
888 state and future challenges. *Energy and Buildings* 2015;107:264–78.  
889 doi:10.1016/j.enbuild.2015.08.032.
- 890 [41] Hong T, Yan D, D’oca S, Chen C-F. Ten questions concerning occupant  
891 behavior in buildings: The big picture. *Building and Environment*  
892 2017;114:518–30. doi:10.1016/j.buildenv.2016.12.006.
- 893 [42] Kim Y-S, Heidarinejad M, Dahlhausen M, Srebric J. Building energy model  
894 calibration with schedules derived from electricity use data. *Applied Energy*  
895 2017;190:997–1007. doi:10.1016/j.apenergy.2016.12.167.
- 896 [43] Yang J, Santamouris M, Lee SE, Deb C. Energy performance model  
897 development and occupancy number identification of institutional buildings.  
898 *Energy and Buildings* 2016;123:192–204.  
899 doi:10.1016/J.ENBUILD.2015.12.018.
- 900 [44] Yang Z, Becerik-Gerber B. The coupled effects of personalized occupancy  
901 profile based HVAC schedules and room reassignment on building energy use.  
902 *Energy and Buildings* 2014;78:113–22. doi:10.1016/j.enbuild.2014.04.002.
- 903 [45] Pisello AL, Asdrubali F. Human-based energy retrofits in residential buildings:  
904 A cost-effective alternative to traditional physical strategies. *Applied Energy*  
905 2014;133:224–35. doi:10.1016/j.apenergy.2014.07.049.
- 906 [46] Chen Y, Liang X, Hong T, Luo X. Simulation and visualization of  
907 energy-related occupant behavior in office buildings. *Building Simulation*  
908 2017;10:785–98. doi:10.1007/s12273-017-0355-2.
- 909 [47] Jin M, Bekiaris-Liberis N, Weekly K, Spanos CJ, Bayen AM. Occupancy

- 910 Detection via Environmental Sensing. *IEEE Transactions on Automation*  
911 *Science and Engineering* 2018;15:443–55. doi:10.1109/TASE.2016.2619720.
- 912 [48] Jin M, Jia R, Spanos CJ. Virtual Occupancy Sensing: Using Smart Meters to  
913 Indicate Your Presence. *IEEE Transactions on Mobile Computing*  
914 2017;16:3264–77. doi:10.1109/TMC.2017.2684806.
- 915 [49] Weekly K, Zou H, Xie L, Jia QS, Bayen AM. Indoor occupant positioning  
916 system using active rfid deployment and particle filters. *Proceedings - IEEE*  
917 *International Conference on Distributed Computing in Sensor Systems,*  
918 *DCOSS 2014* 2014:35–42. doi:10.1109/DCOSS.2014.53.
- 919 [50] Wang W, Chen J, Lu Y, Wei H-H. Energy conservation through flexible  
920 HVAC management in large spaces: An IPS-based demand-driven control  
921 (IDC) system. *Automation in Construction* 2017;83:91–107.  
922 doi:10.1016/J.AUTCON.2017.08.021.
- 923 [51] Wang W, Lin Z, Chen J. Promoting Energy Efficiency of HVAC Operation in  
924 Large Office Spaces with a Wi-Fi Probe enabled Markov Time Window  
925 Occupancy Detection Approach. *World Engineers Summit – Applied Energy*  
926 *Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference,*  
927 *WES-CUE 2017, Singapore: 2017.*
- 928 [52] Chen J, Ahn C. Assessing occupants’ energy load variation through existing  
929 wireless network infrastructure in commercial and educational buildings.  
930 *Energy and Buildings* 2014;82:540–9. doi:10.1016/j.enbuild.2014.07.053.
- 931 [53] Balaji B, Xu J, Nwokafor A, Gupta R, Agarwal Y. Sentinel: Occupancy Based  
932 HVAC Actuation using Existing WiFi Infrastructure within Commercial  
933 Buildings. *Conference: Proceedings of the 11th ACM Conference on*  
934 *Embedded Networked Sensor Systems, 2013.* doi:10.1145/2517351.2517370.
- 935 [54] Jin M, Jia R, Kang Z, Konstantakopoulos IC, Spanos CJ. PresenceSense:  
936 zero-training algorithm for individual presence detection based on power  
937 monitoring. *Proceedings of the 1st ACM Conference on Embedded Systems*  
938 *for Energy-Efficient Buildings - BuildSys ’14* 2014:1–10.  
939 doi:10.1145/2674061.2674073.
- 940 [55] Zou H, Jiang H, Yang J, Xie L, Spanos CJ. Non-intrusive occupancy sensing in  
941 commercial buildings. *Energy and Buildings* 2017;154:633–43.  
942 doi:10.1016/j.enbuild.2017.08.045.
- 943 [56] Zou H, Zhou Y, Yang J, Spanos CJ. Device-free occupancy detection and  
944 crowd counting in smart buildings with WiFi-enabled IoT. *Energy and*  
945 *Buildings* 2018;174:309–22. doi:10.1016/j.enbuild.2018.06.040.
- 946 [57] Zhao Y, Zeiler W, Boxem G, Labeodan T. Virtual occupancy sensors for  
947 real-time occupancy information in buildings. *Building and Environment*

- 948 2015;93:9–20. doi:10.1016/j.buildenv.2015.06.019.
- 949 [58] Wang W, Chen J, Hong T, Zhu N. Occupancy prediction through Markov  
950 based feedback recurrent neural network (M-FRNN) algorithm with WiFi  
951 probe technology. *Building and Environment* 2018;138:160–70.  
952 doi:10.1016/J.BUILDENV.2018.04.034.
- 953 [59] Wang W, Chen J, Song X. Modeling and predicting occupancy profile in office  
954 space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference  
955 approach. *Building and Environment* 2017;124:130–42.  
956 doi:10.1016/J.BUILDENV.2017.08.003.
- 957 [60] ANSI/ASHRAE Standard 62.1. 2013:Ventilation for Acceptable Indoor Air  
958 Quality. 2013.
- 959 [61] Eppy Tutorial — eppy 0.5.44 documentation n.d.  
960 [https://pythonhosted.org/eppy/Main\\_Tutorial.html](https://pythonhosted.org/eppy/Main_Tutorial.html).
- 961 [62] Ang BW, Wang H, Ma X. Climatic influence on electricity consumption: The  
962 case of Singapore and Hong Kong. *Energy* 2017;127:534–43.  
963 doi:10.1016/j.energy.2017.04.005.
- 964 [63] Zou H, Zhou Y, Jiang H, Chien S-C, Xie L, Spanos CJ. WinLight: A  
965 WiFi-based occupancy-driven lighting control system for smart building.  
966 *Energy and Buildings* 2018;158:924–38.  
967 doi:10.1016/J.ENBUILD.2017.09.001.
- 968 [64] Boroojeni KG, Amini MH, Iyengar SS. *Smart Grids: Security and Privacy*  
969 *Issues*. Cham: Springer International Publishing; 2017.  
970 doi:10.1007/978-3-319-45050-6.
- 971