Building and Environment 141 (2018) 1–15

Contents lists available at ScienceDirect

Building and Environment

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

Application of mobile positioning occupancy data for building energy simulation: An engineering case study

Zhihong Pang[a], Peng Xu[a,*], Zheng O'Neill[b], Jiefan Gu[a], Shunian Qiu[a], Xing Lu[c], Xin Li[d]

[a] College of Mechanical and Energy Engineering, Tongji University, Shanghai, 201804, China
[b] Department of Mechanical Engineering, The University of Alabama, Tuscaloosa, 35487, AL, USA
[c] Department of Civil, Environmental and Architectural Engineering, University of Colorado at Boulder, Boulder, CO, USA
[d] Tencent Holdings Limited, Beijing, 100080, China

ARTICLE INFO

Keywords:
Occupancy schedule
Mobile positioning data
Building simulation
Model calibration
Frequency-domain linear regression

ABSTRACT

Occupancy data is a critical input parameter for building energy simulation since it has a big impact on the precision and accuracy of building energy model performance. However, current approaches to get such data through the conventional occupancy detection technology require either implementation of a large-scale sensor network and/or sophisticated and time-consuming computational algorithms, which to some degree limits the application of the real-time occupancy data for building energy simulation. In the era of the mobile internet, the massive people position data, which is generated by smartphone users and stored on cloud servers, offers a potential to solve this important problem. Such mobile data source is precisely monitored, real-time updated, and accessible with affordable time and labor cost upon customer's agreements in some regions, and therefore could be one of the alternatives to traditional occupancy detection methods.

This paper presents an investigation of whether and how the mobile-internet positioning data can benefit building energy simulation. This paper first summarizes the pros and cons of several mainstream occupancy detection methods. Then, the principle of the proposed mobile-internet-based occupancy detection method is introduced. The methodology of using such occupancy data for building energy simulation is developed. An energy performance model of a complex building in Shanghai with a whole building simulation software EnergyPlus is used as a pilot case study to demonstrate the effectiveness of the proposed methodology. A calibration is performed using the building automation system data and the mobile-internet-based occupancy data. The simulation results show that mobile-internet-based occupancy data can help improve the building model prediction accuracy.

1. Introduction

The building sector is responsible for approximately 40% of total energy consumption in the world [1,2]. Among that part, nearly more than one half is used to support the operation of building heating, ventilation, and air-conditioning (HVAC) systems [3]. Such a significant level of consumption urges us to unravel the complexity of building's thermal behavior to optimize building operation and reduce building energy consumption [4]. Building energy performance simulation is one of the most powerful analytic tools to fulfill this purpose. A typical building energy model needs a number of inputs deriving from a wide range of fields including weather file, heating and cooling source, lighting, plug equipment, ventilation, etc. The accuracy of these inputs directly determines the credibility and effectiveness of the simulation results [5,6].

Recent studies show that building energy usage is highly correlated to the occupancy [7]. People influence building performance by both their presence and behaviors as illustrated in Fig. 1. Andersen et al. [8] conducted a simulating study to investigate the relationship between occupant behaviors and building energy consumptions. The results suggested that occupants' opening window behavior had a large effect on building energy usage. Yu et al. [9] examined the influences of the occupant behaviors on building energy usage with a basic data mining technique (i.e., cluster analysis). The authors organized similar buildings among all the investigated cases into various groups based on four user-behavior-unrelated factors. Grey relational grades were used as weighted coefficients of attributes in the cluster analysis. The results revealed that occupant behavior led to a huge difference in Energy Usage Intensity (EUI). A large variability of end-use loads that ranged from close to zero to about four times of the mean value was introduced

* Corresponding author. Postal address: Room A434, No. 4800 Cao'an Road, Department of Mechanical and Energy Engineering, Tongji University, Shanghai, 201804, China.
E-mail address: xupeng@tongji.edu.cn (P. Xu).

https://doi.org/10.1016/j.buildenv.2018.05.030
Received 29 January 2018; Received in revised form 12 May 2018; Accepted 14 May 2018
0360-1323/

Nomenclature

Variables, parameters, and indices

a_n	Fourier coefficients
AVG	Average value
b_n	Fourier coefficients
(CV)RMSE	Coefficient of variation of root-mean-square error
i	Parameter/input variable index
M	Monitored value
MBE	Mean bias error
MBE_{month}	Monthly data
MBE_{year}	Yearly data
n	Number of inputs
R	Coefficient of determination (R-squared value)
RMSE	Root-mean-square error
S	Simulating value
SC	Shading coefficient
ST_{summer}	Summer temperature setpoint
ST_{winter}	Winter temperature setpoint
U	Rate of heat transfer, W/(m2.°C)
VT	Visible transmission
x	Hour of the day
y	Target value of the regression
$\hat{y}$	Predicted value of the regression

Greek letters

ω	Angular frequency

Abbreviations

AEC	Architecture, engineering, and constriction
AMY	Actual meteorological year
ASHRAE	American Society of Heating, Ventilation, and Air-conditioning
BAS	Building automation system
COP	Coefficient of performance
EPD	Equipment power density
EUI	Energy usage intensity
FCU	Fan coil unit system
GPS	Global positioning system
HVAC	Heating, ventilation, and air-conditioning
LPD	Lighting power density
MPC	Model predictive control
PIR	Passive infrared sensor
TMY	Typical meteorological year
URE	Use range error
VAV	Variable air volume system

by occupant behaviors. Frauke et al. [10] applied a model predictive control (MPC) algorithm to analyze the energy saving potentials by using the dynamic occupant information for HVAC control. The results showed that the energy savings could be as high as 50% compared with the baseline cases which used occupancy for controls. In addition, an energy saving of 20% was also observed when the occupancy based demand response HVAC control strategy was performed [11]. Other occupancy-based HVAC control studies can also be found in Refs. [12–16].

Although the importance of occupancy information has become a common understanding of the HVAC community, there still lacks a time- and cost-efficient approach to obtain such data. Currently, the occupancy-related inputs of energy model are mainly acquired from the building codes or design manuals of different countries and organizations. This data source is derived based on the statistics and an assumption that buildings with the same type share similar occupancy

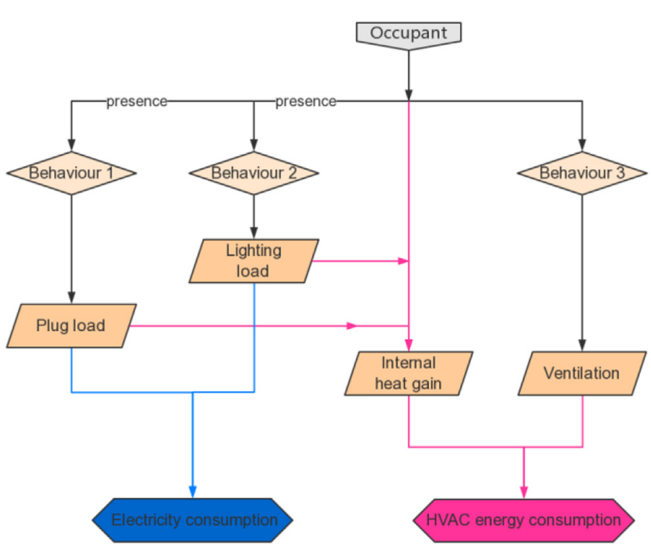

Fig. 1. The influences of occupants on building energy consumption [18].

schedules and densities. When such code-based method is applied, buildings are organized into various groups based on their types and other properties. Thus, through surveying the occupancy profiles in the sample buildings, we can get an averaged description of the occupancy in different building types. This method has been proved to be able to significantly reduce the workload to create an energy performance model since it offers a convenient and moderately accurate source of building occupancy information without requiring practitioners to conduct building surveys one by one.

However, using code-based occupancy inputs for building energy model has a deadly intrinsic problem: they are homogenous and static. So, although in general, it can be applied to buildings with the same type, when it comes to a specific case within the group, the accuracy may not be credible enough. And considering that the occupancy has such a profound influence on building energy consumption, inaccurate inputs associated with occupancy could have large contributions on the discrepancy between the simulated and measured energy consumptions. This mismatch weakens the credibility of the modeling results. Hence, more efforts are required to calibrate the model. Actually, Chang and Hong [17] pointed out that among the wide range of variables which affect building simulation results, the occupancy data is one of the most important ones to cause a model distortion.

Nowadays, to better refine and calibrate building energy performance model for a specified building, how to quickly obtain accurate input associated with occupancy information remains a challenging problem. As usual, two occupancy-related inputs, i.e., occupancy density and occupancy schedule, are required to identify the occupancy pattern of a building in traditional whole building energy simulation program such as EnergyPlus [19], eQUEST [20], and TRNSYS [21]. An occupancy schedule is a set of fractional multipliers which provides values for a 24-h period, starting at midnight. While an occupancy density is a constant parameter representing the maximum occupancy capacity of an occupied zone. It could be expressed as the number of people for a given zone or the number of people per the zone floor area. The product of two parameters gives the occupant number of a specific time for a given zone. For example, if an occupancy density is 200 people in a given zone, a schedule value of 0.5 means that 100 people are assumed to be in that zone at that time.

In practice, the occupancy density can be easily acquired by an onsite survey and a questionnaire since it tends to be a static constant with a limited variance while the occupancy schedule is hard to be scrutinized due to the variety and the indeterminacy of building occupant behaviors. Actually, the inherent unpredictability and substantial irregularity from the realistic presence of occupant schedule are the primary reasons hindering the generation of real occupancy profiles. There is an increased number of studies trying to solve the issue of occupancy detection currently. However, these occupancy detection methods are suffering from some intrinsic problems (e.g., insufficient accuracy, expensive initial investment, etc.) to some extent, which means they are usually limited to the laboratory study and may not be appropriated for practical projects on a large scale (This will be elaborated in Section 2.1). For the appreciation of the fine-grained occupancy data in building energy simulation, it's important and urgent to explore a new occupancy schedule detection method with high detection rates that is relatively unexpansive.

With the value of big data gradually recognized and appreciated by the society, the positioning information generated by smartphone users could offer an alternative solution to this thorny problem. It's well known that an extremely large volume of positioning data is generated during this internet era when mobile internet services are frequently used [22]. Due to the advances in high-resolution remote sensing technology, the mobile-internet-based occupancy positioning data source has huge potentials to benefit various areas such as urban planning [23] and the traffic flow analysis [24].

This paper presents an investigation of whether and how the mobile-internet positioning data can benefit building energy simulation through providing more accurate occupancy data. First, a brief summary of the pros and cons of several mainstream occupancy detection methods is provided. Then, the principles of the proposed mobile-internet-based occupant detection method are introduced, followed by a description of using such occupancy data for building energy model calibrations. Next, an energy performance model of a complex building in Shanghai with a whole building simulation software EnergyPlus is introduced. This model is used as a case study to demonstrate the effectiveness of the proposed method. A preliminary calibration is conducted using the history data from the building automation system as the first step. Then, the proposed mobile-internet-based occupancy data is used to replace the initial code-based occupancy for calibration. Last, simulation results are presented, which show that using BAS data and the mobile-internet-based occupancy data can help improve the building energy performance model prediction accuracy in this study.

2. Occupancy based on mobile positioning data

2.1. Current methods for occupancy detection

Knowing the presence, number, and the variation of people in a given zone of a building is a key component of occupant-oriented research such as occupancy based HVAC control. There is an increased number of literature and reviews that focus on a viable detection methodology to obtain occupancy data with high-quality to fuel building performance simulation and demand based HVAC control strategies in the last decade [25]. Some of the studies rely on the deployment of various sensors, i.e., CO_2 sensor, passive infrared sensor (PIR) sensor, etc., to detect occupants (e.g. [17,26–28]). In addition, some practitioners apply data-driven methods such as the clustering analysis and the decision tree to understand and predict occupant presence in buildings [29–31]. Also, there are studies trying to detect occupancy information with existing equipment of buildings such as sub-metering system [32] and the existing IT infrastructure [33,34]. However, most of these methods for occupancy detection are either lacking in accuracy (only the presence information of occupants is provided) [17,26,27,35] or too expensive (commercially and/or computationally) to implement [29–31,36,37], thus can hardly be put into practical use at this time.

One of the most-widely-used methods to obtain occupancy information is manual counting and questionnaires. The occupancy data obtained from this method is often used as the ground-truth values for the verification of other methods since such data is an accurate reflection of the actual occupied condition [25,38,39]. However, this method has two obvious shortcomings. First, it tends to require a lot of labor force to conduct the survey. Second, the result could be influenced by the questionnaire setting and the survey duration, which sometimes leads to issues related to the accuracy and the credibility of the occupancy data [39]. Especially when the occupancy data of a whole year is necessary to be recorded, the working load can be extremely heavy.

The continuing advances in sensor technology allow researchers to replace manual labor with various sensors to detect occupancy in buildings. Among them, one of the most widely and prominently applied sensors is PIR sensor with an output value of zero or one, which represents “occupied” and “unoccupied” status respectively, to describe the occupancy condition of the occupied area based on infrared (IR) light radiating from occupants [26]. It tends to require a direct line of sight between the sensor and occupants in the space and requires continuous motion of the occupants to function effectively for its passive characteristic. The ultrasonic sensor is also demonstrated being able to detect the occupant presence and location information through changes in the echo intensity and transmitted signal [35]. It can overcome the inherent passiveness problem of the PIR sensor, but it could be susceptible to some false ONs like the air turbulence caused by HVAC system. The measurement of sound waves provides an opportunity to implement sound sensors to detect the building occupant presence information [40]. The sound sensor functions by measuring the audible sound waves produced by building occupants to detect their presence and locations in buildings. But such sound sensor is easily triggered by sound waves from non-human sources, and it requires occupants to continually make the sound to avoid registering false OFF. These sensors (i.e., PIR sensor, ultrasonic sensor and sound sensor) are known as binary sensors because they can only provide a binary value to describe the occupied information of buildings. Besides, similar sensors such as light-switch sensor [17] and telephone off-hook sensor [27] can generate a binary output as well. The occupancy data detected by such binary sensors are discrete presence information (occupied/not occupied), which means the exact number of the building occupants is not available. This functional limitation makes such binary sensors mainly used in lighting control and some limited HVAC control (ON/OFF only) in practice. Regarding some more complicated cases like demand control ventilation, the occupancy information with presence data only is no longer applicable.

Some researchers proposed to deploy carbon dioxide sensors to acquire occupancy data of the number of people [41]. Carbon dioxide is a metabolic production of the respiration. The number of people in a given zone can be estimated by measuring the indoor carbon dioxide concentration and its variations as well as the outdoor air carbon dioxide concentration. Wang and Jin [41], and Wang et al. [42] applied carbon dioxide sensors for ventilation rate controls. Since it takes some time for the carbon dioxide exhaled by occupants to diffuse into the air, the consequent time lag is a common issue for this method. This delayed effect causes the problem that the estimated number of people from directly using the currently monitored value always has a time lag compared with the real occupancy data, which weakens the credibility.

Most buildings are implemented with an image recording system for security purposes. By taking the advantage of the security camera network, Erickson proposed a Smart Camera Occupancy Position Estimation System (SCOPES) [43] which is a sixteen-node sensor network of cameras to capture the occupancy. This system is based on the image sensor to catch the human movement. The system result can achieve approximately 80% accuracy in near real-time. In addition, this system is capable of predicting the probability of room usage. But such a system requires an advanced algorithm to process and interpret the raw data obtained from the sensors. Besides, a mass load of computation and programming is required for this method and restricts its practical application.

In Refs. [37] and [36], a comprehensive sensing testbed was deployed in Pittsburg to integrate the state-of-the-art IT science and sensing technology. This large-scale wired and wireless sensing network comprises a broad series of environmental sensors which can measure a variety of parameters including carbon monoxide, carbon dioxide, total volatile organic compounds, small particulates, acoustics, illuminance, etc. Such hybrid sensing network has promising potentials in human-centered HVAC control and building simulation. But the significant cost of implementations and the intensive labor required for continuous recalibrations of a variety of sensors makes it nearly impossible to be used for any current practical application in a cost-effective way.

Machine learning and data mining technologies have been attracting more and more attention in the last decade. Some researchers applied data-driven algorithms to acquire occupancy information. For example, Page et al. [18] generated a time series of the state of presence (absent or present) of each occupant of a zone by considering occupant presence as an inhomogeneous Markov chain interrupted by occasional periods of a long absence. The proposed model proved its capacity to reproduce the presence data through a validation using occupancy data from a private office. Hong et al. proposed a series of studies to apply data mining based approaches for occupancy schedule learning and predictions. They used the cluster analysis and the decision tree to predict the occupancy of office buildings [29,30]. Luo et al. applied an agent-based method to study the occupancy schedule simulation [31]. To effectively apply any machine learning algorithm, a complete training dataset with all scenarios should be available. In practice, it is often challenging to get such datasets.

Data generated by the utility sub-metering system is also an evident source of occupancy data. Kim [32] derived the occupants' profiles and schedules from the plug-load consumption and total electricity consumption in a building. This method is conveniently accessible without an additional cost and is exempted from a high level of computation. But the data accuracy is dependent on various factors such as the occupancy type and area usage type. Besides, the assumption that each occupant consumes the same amount of electricity could potentially cause some discrepancies.

Some scientists also proposed to use existing building IT infrastructures (Wi-Fi network, Ethernet resolution protocol) [33,34] to capture the occupancy information. This method is also known as electromagnetic signal (EM) detection. This method doesn't require the additional deployment of equipment but also has some disadvantages. For example, occupants often have more than one device which is enabled EM signal detection. Also, an advanced algorithm is required to process and interpret the raw data to generate occupancy profiles.

In summary, current detection approaches to get occupancy data are all suffering from some limitations. They either require intensive labor (manual counting and questionnaires [25,38,39]), cannot provide exact occupant numbers (binary sensor network [17,26,27,35]), requires the existing building infrastructure (security cameras [11,43], sub-metering system [32], and Wi-Fi network [33]), demands expensive initial investment (binary sensor network, carbon dioxide sensor network [41,42], comprehensive sensing network [36,37]), or requires a complete training dataset and/or massive computational resource (sensor networks, machine learning methods [29–31]). These problems restrict the practical application of occupancy data for building energy simulation.

2.2. Introduction to mobile positioning data

The limitations of conventional occupancy detection methods which focus on locating occupants in some specified zone spur scientists to solve this problem from another perspective. This new approach explores how to obtain occupancy information by tracking individuals. Yang et al. concluded [22] that with the popularity of mobile devices, massive location data can be generated by cell identifications with low accuracy or WLAN and Global Positioning System GPS usages with high accuracy. Deep processing of this data can generate the tracks of service users.

By cooperating with local telecom operator (Estonian Mobile Telephone), Rein et al. found that the accuracy of position data generated by passive mobile positioning data is of moderate quality. The location accuracy can reach as high as 100 m in urban area and 450 m in suburban areas. If combined with the population register data, the accuracy level is enough to specify the distribution of population on an urban scale [44]. With this method, an urban planning study was successfully conducted [45]. However, one limitation of this method is that the database only includes 60% of the mobile phone users in Estonia. Besides, the positioning data is not suitable for the studies of a smaller scale.

In the era of the mobile internet, the value of mobile positioning data seems to be steadily growing. The prosperity of the smartphone penetration makes it possible to apply big data technology to detect the occupancy of a specified area. A wide variety of internet services does not only facilitate the daily life of ordinary people but also generates a tremendous amount of positioning data through this process. Whenever smartphone users share their positions on social media, call Uber cars from a mobile client or use the navigation service on mobile devices, their real-time position will be generated and recorded in the cloud server of the service provider. Due to the maturity of the remote sensing science, the positioning data generated in this process is accurate and credible. Actually, the Global Positioning System (GPS) has achieved an average user range error (URE) of fewer than 0.8 m since 2016 [46].

A comprehensive and fine-grained building occupancy profile should provide six properties, i.e., presence, location, count, activity, identity, and track [33]. In building energy modeling, a good occupancy input should include the accurate number of occupants and its variation with time. This mobile-internet-based data source is different compared to traditional sensor-based sources in three ways. First, the positioning service is primarily delivered by trustworthy providers like GPS, Beidou, and Galileo. The level of positioning accuracy is highly ensured, and there is likely no need for the further calibration, as mentioned previously. Second, this whole occupancy capturing system does not require any additional implementation cost associated with sensors, yet does not need a frequent sensor system calibration. In addition, the workload of the deployment and the maintenance is relatively small.

2.3. Mobile positioning data acquisition and processing

Nowadays, petabytes of mobile positioning data are being generated, recorded, and stored every day. Gaining the authorized access to this database is the premise of the recognition of the values of position data. In this study, we established communication with an internet company in China and were successfully authorized limited access to their database under supervision. This company is a leading competitor in Chinese internet-service market, which holds a number of internet services including social network, online shopping, navigation, etc. As long as customers are consent to the user agreement and activate the location service on their devices, they will be able to use these online services. Customers can not only share their positions with friends and followers via a private message and/or a public check-in announcement but also enjoy other position-based services such as online shopping, online taxi-hailing, online tickets ordering, etc. The location services are provided by GPS, which has achieved a URE of fewer than 0.8 m since 2016, as has been mentioned. Such an accuracy level (less than 1 m) is credible enough for its application in building energy simulation.

The first step to acquire the occupancy data based on mobile positioning is to specify the geographical location of the intended research object or area. A graphical specification on the map should be provided to activate the monitoring process. All mobile positioning data in the specified area will be collected and trended by the system and saved on a cloud server. This includes who enters or leaves the area and the exact time of the movement. The records of single persons are then summed and processed to generate the number of occupants in the specified area in an interval. The number of the occupants is the mean value of the maximum and minimum occupant numbers recorded during the interval. Predefined by the user, the length of the interval can be flexibly varied from 5 min to one hour, which means that the frequency of the data updating can be as high as 12 times in one hour. There are two ways to use such data: Firstly, the data can be sent to the user in real time through the user interface. Secondly, this raw data can be downloaded in one batch after sufficient data has been collected.

Although the raw data obtained from the tracking system is directly the number of the occupants in the specified area, some post-processing is necessary to further improve the accuracy of the occupancy data. This post-processing procedure to get occupancy schedule is illustrated in Fig. 2, followed by the description of two additional observations for such procedure.

First, we should note that the mobile-internet-based data only takes into account some occupants who are the customers of the internet company we cooperated with. The location of an occupant will only be recorded when he or she shared his or her location via the mobile applications from this company. In other words, occupants who don't install and use the mobile applications from this company on his or her mobile device or don't often share locations via the internet will not be counted by the proposed system. This could make the collected occupancy data less persuasive and complete. Nonetheless, considering that our industry collaborator already had more than 900 million monthly active mobile-service users at the end of 2016 [47], and the locations service was used more than 50 billion times in a single day, we assume that this data source can well represent the relative variance of the occupancy. As mentioned above, in building performance simulation software, the occupancy condition of a building is mainly determined by two indicators, occupancy density and occupancy schedule. Therefore, we can directly use occupancy schedule collected from the mobile-internet-based data and just calibrate the occupancy density. We are assuming the occupancy density in a building is a relatively constant value, which could be obtained through an on-site survey. A correction coefficient could probably be introduced in the future to improve the data completeness. A largescale survey (such as whether use smart phone, whether use mobile-internet-service, etc.) will be necessary to help get such correction coefficient. It is expected that there will be variations of this coefficient in different regions and countries.

Secondly, to reduce the workload of the modeling, occupancy schedules with different day types (e.g., weekdays vs. weekend) instead of a 365-day time series schedule will be used. This simplification can make parameter inputs more efficient. For example, the selected building in this study is a shopping center and office complex, the daily variance of the occupancy doesn't experience a large variation in the same day type group. Therefore, this simplification is expected to yield a good accuracy. We use a frequency-domain linear regression method to generate occupancy schedules with different day types. This assumption and method will be elaborated in the case study in details.

Thirdly, it should be noted that the current method to specify the monitoring area is not accurate enough, especially for those buildings with irregular shapes. A smaller measuring scale should be provided in the future to reduce the related errors. However, it should also be noted that depending on the local laws and regulations, this data source may not be available in some countries and regions. This could be a potential limitation on the application of the mobile-internet positioning data. Authors fully understand and respect the privacy right of every internet user. The occupancy information obtained in this pilot case study was tracked under a supervision and only used for the presented building energy simulation research. There is a need to integrate cyber-security and privacy research with the mobile-internet-based occupancy data

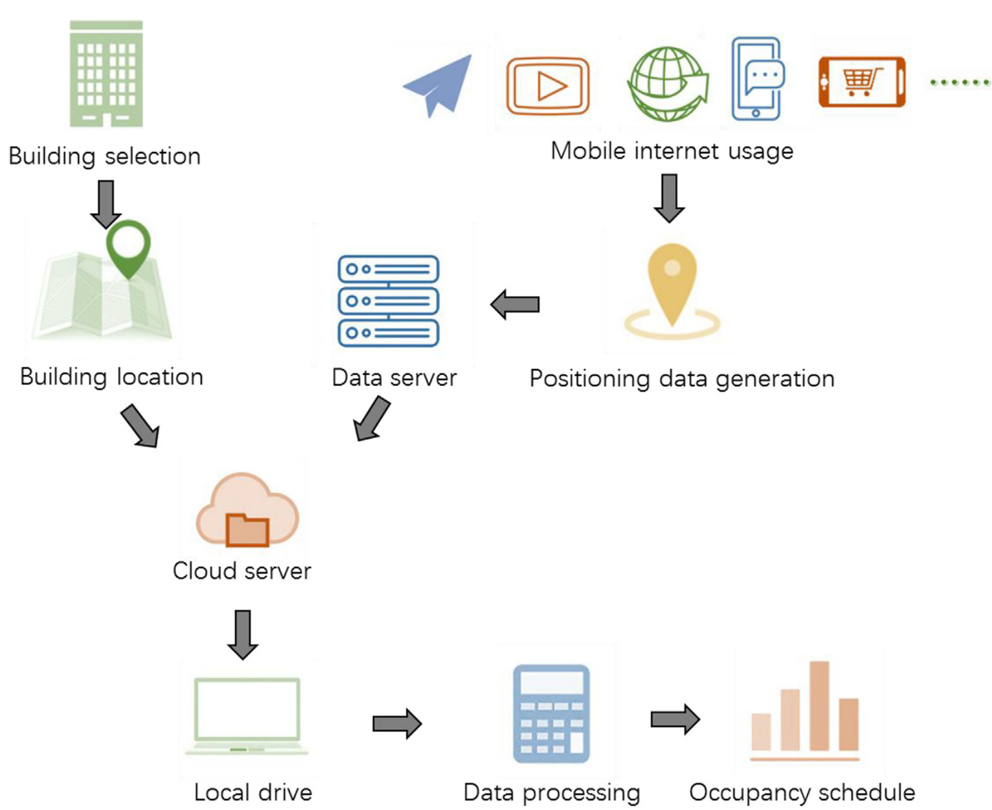

Fig. 2. Generation of occupancy based on mobile internet.

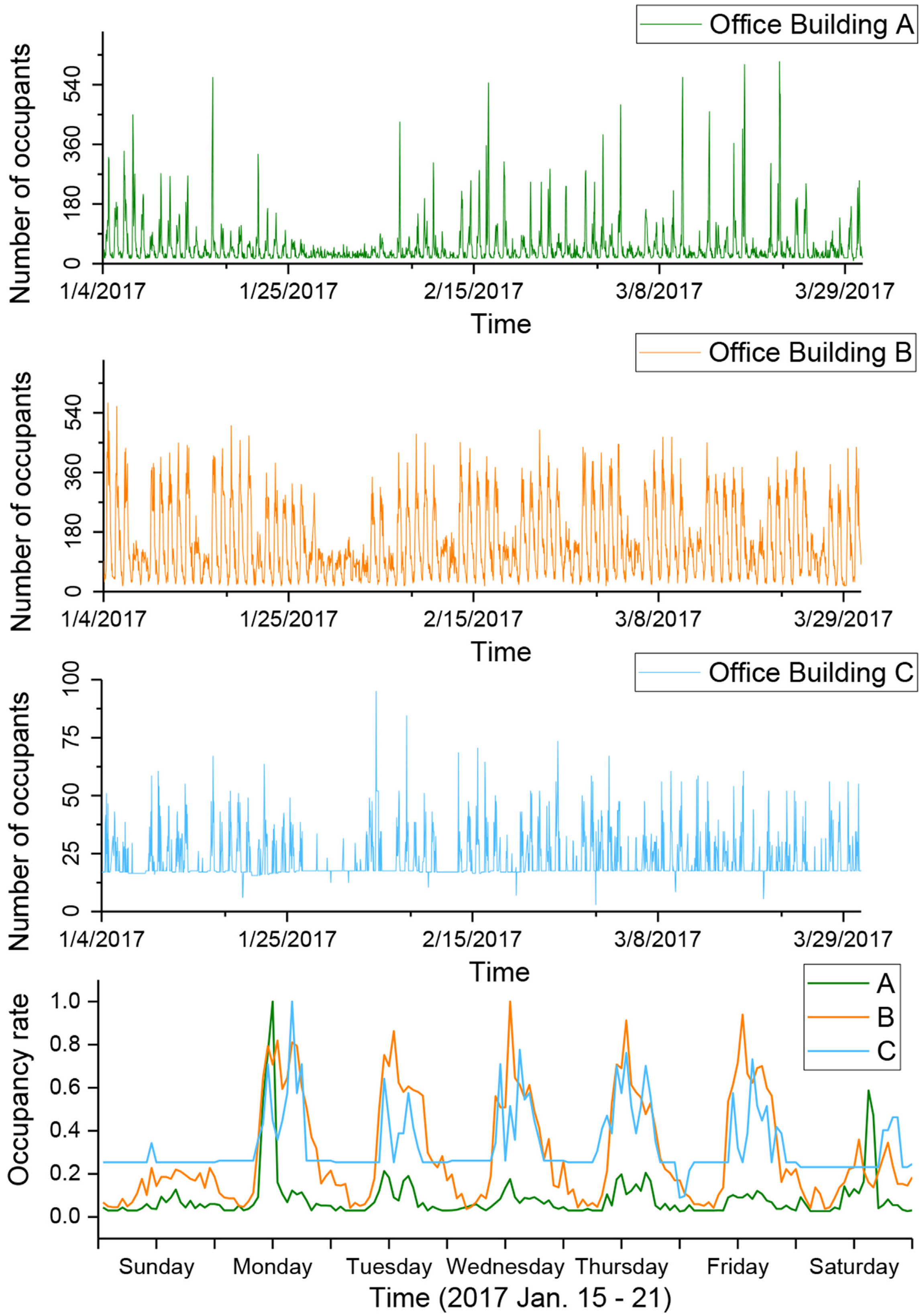

Fig. 3. Occupancy samples of three office buildings.

acquisition in the future work.

2.4. Evaluation of the mobile positioning data

The occupant profiles of three large offices in Shanghai are presented in Fig. 3. The data is acquired by the proposed mobile-internet-based occupancy detection method. This figure further verifies the inapplicability of code-based occupancy data for practical use. It's obvious from Fig. 3 that although these three sample buildings are of the same building type (i.e., large offices), their occupancy schedules are radically different from each other. The weekly schedules of Building A and C vary considerably with days while building B maintains on an almost stable level. There are noticeable fewer occupant activities for three building around Jan 28th, 2017, which is Chinese New Year day. Most of the companies in these buildings were closed for holidays. There is continuously a baseline amount of occupants in the office buildings because there is security personnel working at night. Even if there is a vast difference for these three large office buildings in terms of occupancy schedule, the same occupancy schedule input will be applied in the simulation if only occupancy data from the building codes and design manuals were used since these buildings are the same type.

Even for the same building, the daily and weekly schedules can still be much different with the time. For example, the occupancy of Building A shows a large unpredictability with the time. Its occupancy seems to vary randomly with the time, and there isn't an observable pattern to follow. In general, such characteristic of daily variation will most likely be ignored by building codes. Deru et al. had same comments on using occupancy information directly from building codes [48]. In their study, occupancy density of commercial building can vary in a huge range (i.e., from 334 ft^2/person to 300,000 ft^2/person) with a mean of over 25,000 ft^2/person.

In summary, the mobile positioning data can help us understand and identify key characteristics of building occupancy information, i.e., building occupancy schedule. Traditional methods which tend to organize building occupancy information into different groups always fail to capture the uniqueness of each building. Instead, this mobile-internet-based method takes advantages of the already-built mobile internet system and can potentially real-time monitor and update the occupancy information of every single building. Compared with other approaches, this method doesn't require the deployment of large-scale sensor network or massive computational power. The occupancy information such as occupancy schedule can be easily accessible with this method.

3. Methodology

3.1. Background

Building simulation is widely used at present, because it could help us achieve a sustainable built environment, and at the same time, improve indoor quality and occupant productivity. In addition, building simulation has been used to model future innovation and technological progress in the architecture, engineering, and construction (AEC) industry [49]. For the simulations used in the operation stage, the model is required to represent a building as accurately as it can; that is, the model should describe the building systems as-installed, as-operated, and as-used [50]. Standard building simulation programs, for example, EnergyPlus [19] and eQUEST [20], typically have energy usage (e. g., electricity, gas) and demand as the modeling outputs [51,52]. In general, the modeler calibrates the input parameters of a simulation program to minimize the discrepancy between the model outputs and the actual measurement counterparts.

However, in most modeling cases of existing building, the mismatch between the design phase and the operation phase could be significant due to the fact that most existing buildings do not operate as well or as efficiently as they could and should [49]. This results in a considerable discrepancy between the simulated and the measured energy consumptions [5,6]. Therefore, the model calibration serves as an essential and necessary step to ensure the accuracy and applicability of building models, in particularly, used in the operation stage [49,53](e.g., model-based controls, model-based diagnostics, etc.) (see Table 1).

Model calibration, which is also known as calibrated simulation (CS), refers to the process of tuning model's input parameters to narrow the disagreement between the simulated result and the real-monitored data. The calibration technique typically consists of four steps: (1) collect the data, (2) input the data and run the simulation, (3) compare simulation model outputs to measured data, and (4) decide on whether the desired accuracy has been achieved [54]. MBE_{month} (Mean Bias Error) and (CV) $RMSE_{month}$ (Coefficient of Variation of Root-mean-square Error), recommended by ASHRAE (American Society of Heating, Ventilation and Air-conditioning), are two commonly used indexes to evaluate the performance of CS. Only when the MBE_{month} and (CV) $RMSE_{month}$ fall within a specified range can the calibration be considered acceptable. Recommended values for these two indices from three frequently referred guidelines (ASHRAE [54], International Performance Measurement & Verification Protocol (IPMVP) [55], and Federal Energy Management Program (FEMP) [56]) are presented in Table 2.

Table 1
Summary of widely used occupancy detection methods.

Occupancy detection method	Presence/Number of occupants	Evaluation notes	Reference
Manual counting and questionnaire	Number	Requires a lot of labor force; not accurate enough	[25,38,39]
PIR	Presence	Can only output a binary value. Requires a direct line of sight to function effectively.	[26]
Ultrasonic sensor	Presence	Can only output a binary value. Susceptible to false signals such as air turbulence.	[35]
Sound sensor	Presence	Can only output a binary value. False OFFs when occupants make no sound.	[35]
Light-switch sensor	Presence	Can only output a binary value	[17]
Telephone off-hook	Presence	Can only output a binary value	[27]
Carbon dioxide sensor	Number	The estimated number of people always has a time lag compared with the real occupancy data if CO_2 data is directly used.	[28,41,42]
Image sensor	Number	Requires an advanced signal processing algorithm and mass computational power.	[11,43]
Large-scale sensor network	Number	The result is relatively accurate; but it requires implementations of an expensive sensor network, massive computational power, and complicated programming	[36,37]
Machine Learning	Number or presence	It can be used for prediction of occupancy. But it requires the availability and completeness of the training data.	[29–31]
Utility sub-metering	Number	Doesn't require any additional occupancy sensors. But the result is not accurate since each occupant is assumed to consume the same amount of electricity.	[32]
Wi-Fi network; Ethernet address resolution protocol	Number	Doesn't require any additional occupancy sensors. Occupants often have more than one device. Requires advanced algorithm.	[33,34]

Table 2
Regulations of MBE and CV(RMSE) in three guidelines.

Index	ASHRAE 14 (%)	IPMVP(%)	FEMP(%)
MBE_{month}	± 5	± 20	± 5
MBE_{year}	–	–	± 10
$CV(RMSE)_{month}$	± 15	–	± 15

$$MBE_{month} = \left[\frac{M_i - S_i}{M_i}\right] \times 100\% \quad \text{Eq. (1)}$$

$$MBE_{year} = \frac{\sum_{i=1}^{n}(M_i - S_i)}{\sum_{i=1}^{N} M_i} \times 100\% \quad \text{Eq. (2)}$$

$$AVG = \left[\frac{\sum_{i=1}^{n} M_i}{N}\right] \quad \text{Eq. (3)}$$

$$RMSE_{month} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(M_i - S_i)^2} \quad \text{Eq. (4)}$$

$$CV(RMSE)_{month} = \frac{RMSE_{month}}{AVG} \times 100\% \quad \text{Eq. (5)}$$

One of the major problems in current CS is that the modeler is required to have expertise in the HVAC domain since how well a simulation model is calibrated is highly relied on his or her subjective judgment and experience. In many cases, the process of calibrating a simulation can be extremely tedious and intensely laborious, especially with simulation programs that require a large number of input parameters. It is difficult for modelers to predict the consequent change in outputs after each parameter tuning, especially for inexperienced ones [49].

Additionally, the high non-linearity, multi-parameters characteristics, and the inherent complexity of the building system make the calibration incredibly time-consuming in practice. The modelers will typically "adjust" input parameters on a trial-and-error basis until some selected outputs match the actual measurements. Troncoso [57] considered this process as "fudging" because it often results in the manipulation of a large number of variables that may significantly decrease the credibility of the entire simulation.

There is an increased number of studies trying to propose a more efficient method to calibrate building models. Coakley et al. had a comprehensive review of the modern calibration methods [58]. In their review, calibration approaches were classified into two main categories, manual and automated. A manual approach consists of characterization techniques, advanced graphical approaches, and procedural extensions. An automated calibration needs some optimization techniques. Approaches like sensitivity analysis (SA), optimization techniques, and Meta-modeling need complicated mathematical algorithms either to find the key parameters or determine the cost or penalty function, etc. Clarke et al. [59] came up with a calibration method which emphasized the empirical data and applied this procedure using the ESP-r program. Heo et al. [50] quantified uncertainties associated with building models based on a Bayesian calibration approach. They demonstrated such method could correctly evaluate energy retrofit options. Liu et al. [60] presented a calibration signature method for the rapid calibration of heating and cooling energy consumption simulation of commercial buildings. O'Neill and Eisenhower [61] presented a systematic way for building energy model calibration using a parametric uncertainty analysis. Chaudhary et al. [62] conducted two case studies through using an "Autotune" approach to tune input parameters in EnergyPlus models. However, the most common approach to calibrate a model is still the trial-and-error method in practice, which is time-consuming and inefficient. The credibility of the parameters (e.g., occupancy schedule) autotuned by the algorithms is limited by the ranges pre-defined by the modelers.

This paper aims to investigate whether the application of the mobile-internet-based occupancy information can benefit building performance simulation. Although the occupancy information obtained with this method is a building-level data and lacks details on the location of each occupant, such precision level is probably sufficient for building calibrated simulation. This is evidenced by this pilot case study and will need to be further verified through other case studies in the future. On the one hand, a more accurate occupancy input can help better describe the building energy model, which helps to reduce the model distortion and make the simulation results more credible. On the other hand, the occupancy input, which is closer to the real occupied condition, can potentially help resolve the mismatch between the simulation result and the measured energy consumption and reduce the workload of the model calibration. In general, improved understanding occupancy in buildings can help improve building energy modeling,

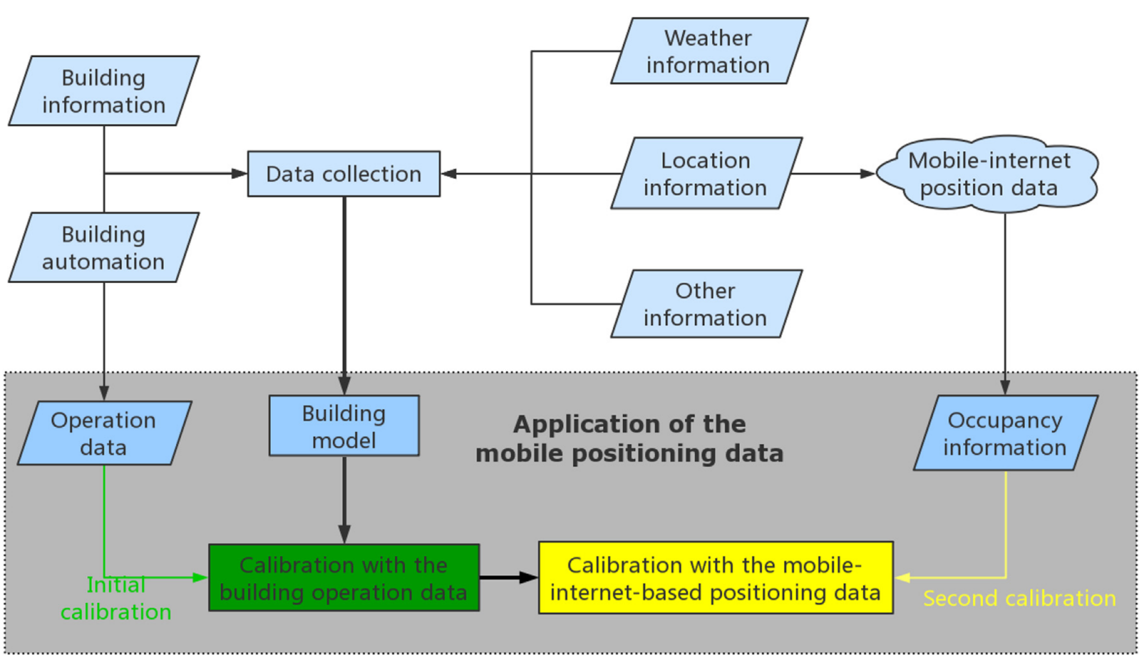

Fig. 4. Flow chart of the application of using mobile position data for building simulation.

which has been claimed in many recent studies by IEA Annex 66 [63].

3.2. Methodology

This study aims to prove that compared with the occupancy information derived from the building code and design manual, the mobile-internet-based occupancy data can benefit building performance simulation in terms of improvements of the model accuracy and reductions of the associated workload. Specifically, we conduct model calibrations for a chosen building with the building operation data from the BAS and the presented mobile-internet based occupancy data. Nowadays, it is not difficult even for practitioners to get building operation data from BAS particularly for modern new buildings [64]. The objective is to see whether the adding of mobile-internet-based occupancy can help improve simulation results. The detailed flow chart for the proposed calibration methodology using BAS data and mobile-internet-based occupancy data is shown in Fig. 4.

3.2.1. Data collection

A building appropriate for the proposed calibration method should be chosen. Ideally, the building candidate has the following features.

- This building should be equipped with a perfect building automation system (BAS) including a complete sub-metering system. All BAS and metering data are available and can be downloaded to a local drive. These data are usually available from building management companies.
- All building basic facts are available including building layouts, HVAC equipment schedules, practical usage, mechanical system, etc.
- An accurate geographical location of the building should be available. Building location can be graphically specified in the occupant tracking system. The mobile-internet service system will automatically track the occupied condition of the building and store the information on its cloud server using a predefined sampling frequency and tracking period.

Since the architectural drawings and the operation data of the office building A and C are not available to this research, this case study only modeled the office building B using EnergyPlus and the model calibration only covered this building as well.

3.2.2. Create building energy performance model

A building energy performance model can be created after all the building property information is collected. In terms of unavailable model input parameters, these values will be derived from building codes and design manuals, which is the current practice for the building simulation [65,66]. In other words, this initial model is built with a traditional method, i.e., no actually-measured data (indoor temperature setpoints (ST), lighting power density (LPD), etc.) is utilized or considered.

Then a simulation will be performed using this model to get some preliminary simulation results. By comparing such results with the BAS measured data and calculating MBE_{month} and $CV(RMSE)_{month}$, we can evaluate how good this model is and understand the accuracy of code-based inputs.

3.2.3. Initial calibration using BAS data

For most modelers, the first step to conduct a model calibration is to calibrate the ST, LPD, lighting schedule, equipment power density (EPD), equipment schedule and other parameters which can be easily calibrated using BAS measured data [65,66]. After this calibration, the MBE_{month} and $CV(RMSE)_{month}$ calculated using simulating and measured energy consumption can usually be reduced. The case study in this paper also includes this regular and initial calibration using BAS measured data, where the heating and cooling indoor air temperature setpoints, lighting density and schedule, equipment density and schedule are calibrated using BAS data.

3.2.4. Further calibration using the mobile-internet-based occupancy data

One of the biggest factors that contribute to this still mismatch is related to inaccurate inputs associated with building occupancy [18]. Thus, a further calibration which aims to minimize the mismatch by replacing the code-based occupancy inputs with mobile-internet-based occupancy information is conducted. This further calibration will follow the initial one which is mainly based on the BAS measured data. The mobile-internet-based occupancy data of the building is downloaded from the cloud server and then post-processed for generating occupancy schedules based on different day types, as described in section 2.3. These occupancy schedules are then input into the model to replace the original code-based values. The MBE_{month} and $CV(RMSE)_{month}$ will be calculated again to evaluate the effectiveness of this further calibration.

4. Case study

A calibration case study of a real building is conducted following the calibration methodology presented in section 3. In this section, a step-by-step calibration is performed. The initial calibration is conducted using the building operation data from the BAS data, and the further calibration is performed using the mobile-internet-based occupancy data.

4.1. Building description

As has been mentioned, only the architectural drawing and the

Fig. 5. Photo and model of the selected building in the case study.

operation data of the office building B were available for us, therefore, the office building A and C were not included in the case study. The selected building is a shopping center and office complex located in Changning District, Shanghai, China. The building is 132 m tall with a total building area of about 200,000 m^2. There are 30 storeys above the ground in total. A seven-storey podium is adjacent to the main tower. The podium together with the floors 1–7 of the main tower is used as the shopping and retailing areas. The top floor is designed to be a salon. Other floors of the main tower are designed to be offices. There are also four underground levels which contain civil air-defense facility, parking lots, miscellaneous equipment rooms, etc. The building model is shown in Fig. 5.

4.2. Model development

4.2.1. Modeling software

The building envelope model was developed by SketchUp Pro 2016 [67] and OpenStudio Legacy plug-in Ref. [68]. The simulating tool in this case study is EnergyPlus 8.6 [19].

4.2.2. Data input

The information of building envelope, building scheme, internal loads, mechanical systems, etc. are input into the building model using references from either architectural drawings and/or building codes. The selected input parameters associated with building envelope thermal characteristics are summarized in Table 3.

The typical meteorological year (TMY3) data of Shanghai developed by U.S. DOE [69] is used for the initial model running.

The initial values of input parameters associated with the internal loads are presented in Table 4. The initial schedules of lighting and equipment are presented in Fig. 6. These schedules are based on building codes [70].

4.2.3. Initial simulation

A preliminary check-up on the simulation results using a non-calibrated model shows that the simulated energy consumption of HVAC system highly deviates from the measurements. The simulated energy consumption in winter is much higher than the measurement while the trend is opposite (i.e., HVAC energy consumption in summer is much lower than measurement). The $CV(RMSE)_{month}$ of these two groups of data is 42.1%. And the MBE_{month} is even as high as over 40%. This makes it necessary for model calibration.

4.3. Initial calibration using BAS data

As described in the methodology section, an initial calibration using the BAS data is performed in this case study. The BAS monitors and records the changes of indoor parameters and sub-metering data of the building system. By referring to BAS data, we can adjust the indoor air temperatures and internal loads (i.e., LPD and EPD) to a reasonable level [65]. Ideally, an actual meteorological year (AMY) data should be used instead of the TMY data for the calibration [66]. Unfortunately, AMY data was not available for this case study at this moment. Therefore, a TMY3 weather data is still used during the calibration.

4.3.1. Parameters adjustment

Input parameter adjustments conducted during the initial calibration are listed in Table 4. The zone air temperature setpoints are adjusted based on the sensor reading values recorded in BAS. The internal heat gains, including LPD and EPD, are adjusted based on the sub-metering electricity consumption data of lighting and equipment. As shown in Fig. 7 and Fig. 8. After this initial calibration, the MBE_{month} and CV(RMSE) $_{month}$ of lighting electricity end-use decreases from 28.6% to 28.5%–2.5% and 5.8% respectively. The MBE_{month} and CV (RMSE) $_{month}$ of equipment electricity end-use decreases from 9.8% to 9.4%–0.8% and 1.5% respectively.

4.3.2. Initial calibration results

The result of the calibration with the BAS data is presented in Fig. 9. The MBE_{month} and CV(RMSE) $_{month}$ of these two groups of data are 13.6% and 18.9% respectively. The improvement in the simulation result is possibly because the LPD and EPD of the shopping and office areas were increased after the initial calibration. This change not only caused an increase in the energy consumptions of the lighting and equipment systems, but also resulted in a growth in the building cooling load and a reduction in the building heating load. This explained why the building energy consumption increased in the summer and decreased in the winter. Besides, it is interesting to note that the simulated energy consumption of March and November were close to the measured values after the calibration of the lighting and plug equipment system. This is probably because Shanghai is located in a climate zone with hot summer and cold winter. During the shoulder season such as March and November, the HVAC related energy usage is small. Therefore, only calibrations of the lighting and plug equipment system could lead to acceptable energy prediction errors at the building level.

It is clear that although the simulated energy consumption is closer to the real-monitored values from the BAS, there is a need for a further calibration to have the calibrated model meeting the criterion from the guidelines (e.g., ASHRAE guideline [54]).

4.4. Further calibration using the mobile-internet-based occupancy data

In this section, occupancy information acquired from mobile positioning data is tuned to calibrate the building.

4.4.1. Occupancy information acquisition and processing

The area monitored by the mobile-internet system is shown in Fig. 10. The occupancy information from September 2016 to September 2017 is monitored and saved. A sample of occupancy inside the monitored area is presented in Fig. 11. There are noticeable fewer occupant activities around September 15th, October 1st, January 1st, and Jan 28th which are Mid-Autumn Festival, National Day, New Year's Day, and Spring Festival in China respectively. The data source is then grouped based on the seasons and day types. Five groups of occupancy schedules which share the similar variance trends are generated. They are summer workday, winter workday, summer weekend, winter weekend, and holiday respectively. The summer covers the period from April to October and the winter covers the rest six months. Considering that the occupancy variance within each group is similar to each other, we assume that the hourly occupancy schedule of each day group is a periodic discrete time series every 24 h.

As mentioned in Section 2.3, these five groups of data are then processed with the frequency-domain linear regression method. The definition of the frequency-domain linear regression is presented in Eq. (6), where x denotes the hour of the day and ω denotes the angular frequency. This frequency-domain linear regression, which is also known as the Fourier Series Regression, is a way to represent a function as the sum of simple sine waves. Specifically, it decomposes any periodic function or periodic signal into the sum of a set of simple

Table 3
Building envelope characteristic.

Envelope	Values
External wall	U = 0.396 W/(m2.°C)
Internal wall	U = 1.926 W/(m2.°C)
Roof	U = 0.365 W/(m2.°C)
Door	U = 0.143 W/(m2.°C)
Floor	U = 1.043 W/(m2.°C)
Ceiling	U = 3.125 W/(m2.°C)
Window	U = 2.275 W/(m2.°C)
	SC = 0.22
	VT = 0.13

Table 4
Indoor air temperature setpoint and internal heat gains.

District	Parameter	Before Calibration	After Calibration
Office area	ST_{summer}	24 ± 2 °C	25 ± 2 °C
	ST_{winter}	24 ± 2 °C	25 ± 2 °C
	LPD	9 W/m2	30 W/m2
	EPD	15 W/m2	18 W/m2
Commercial area	ST_{summer}	20 ± 2 °C	22 ± 2 °C
	ST_{winter}	20 ± 2 °C	22 ± 2 °C
	LPD	10 W/m2	30 W/m2
	EPD	13 W/m2	18 W/m2
Garage	LPD	9 W/m2	30 W/m2
	EPD	5 W/m2	1 W/m2
Plant room	LPD	9 W/m2	30 W/m2
	EPD	5 W/m2	1 W/m2

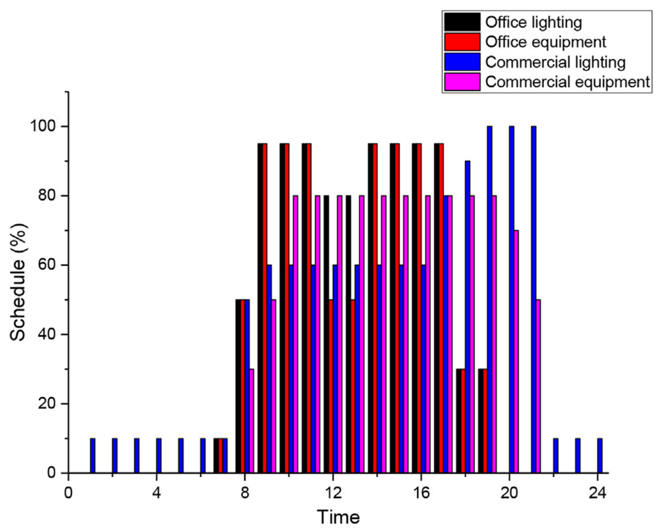

Fig. 6. Lighting and equipment schedules for office and commercial areas [70].

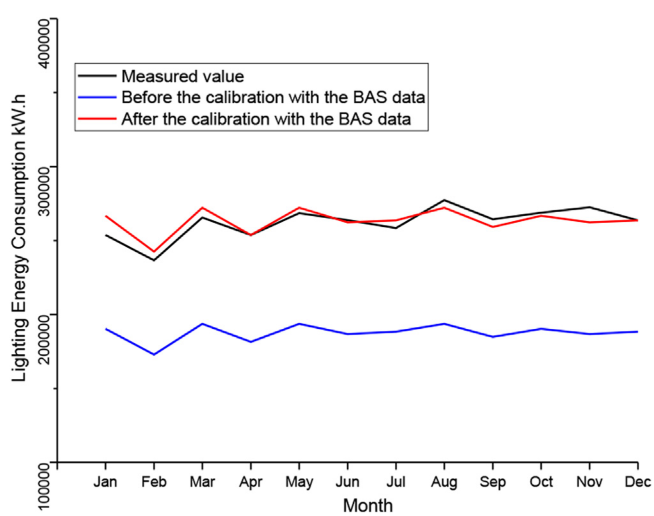

Fig. 7. The lighting electricity end-use of the measured values, initial simulation, and the calibration with the BAS data.

oscillating functions, namely sines and cosines. The discrete-time Fourier transform is a periodic function, often defined in terms of a Fourier series. It has been adopted in other studies to generate periodic time-series profiles. For example, Ji et al. [71]applied Fourier Series Model (FSM) to disaggregate HVAC terminal hourly end-use in commercial buildings. Niu and O'Neill [72,73] also applied an improved Fourier Series Decomposing to estimate HVAC electricity consumption

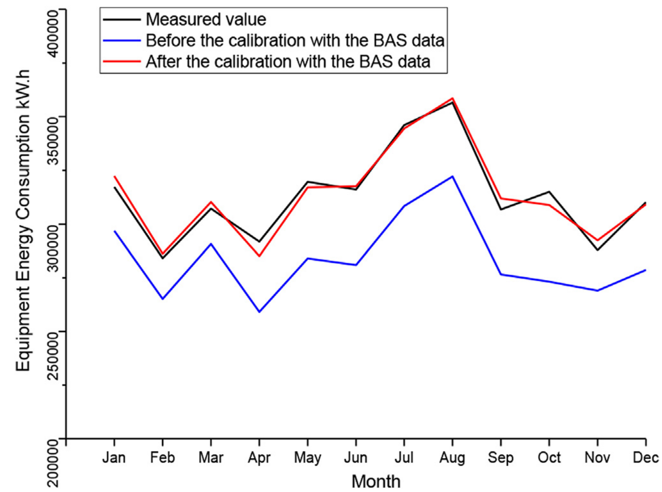

Fig. 8. The equipment electricity end-use of the measured values, initial simulation, and the calibration with the BAS data.

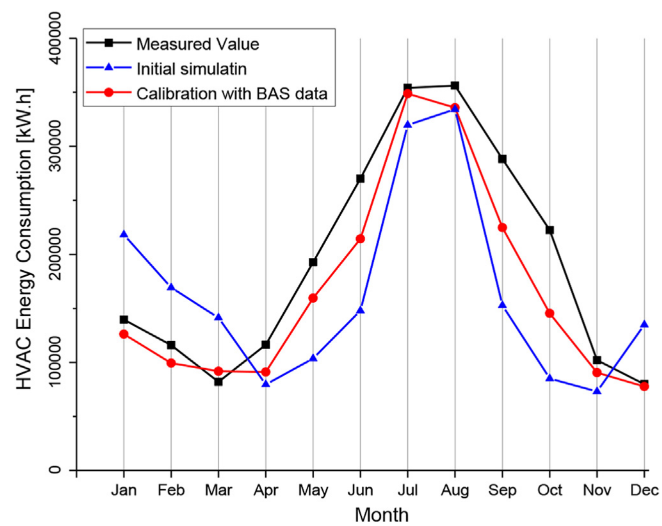

Fig. 9. Measured HVAC energy consumption, initial simulation, and the calibration with the BAS data.

in a dormitory building.

$$f(x) = a_0 + a_1{*}\cos(x{*}\ \omega) + b_1{*}\sin(x{*}\ \omega) + a_2{*}\cos(2{*}x{*}\ \omega) + b_2{*}\sin(2{*}x{*}\ \omega) + a_3{*}\cos(3{*}x{*}\ \omega) + b_3{*}\sin(3{*}x{*}\ \omega) + a_4{*}\cos(4{*}x{*}\ \omega) + b_4{*}\sin(4{*}x{*}\ \omega) + a_5{*}\cos(5{*}x{*}\ \omega) + b_5{*}\sin(5{*}x{*}\ \omega) + a_6{*}\cos(6{*}x{*}\ \omega) + b_6{*}\sin(6{*}x{*}\ \omega) \quad \text{Eq. (6)}$$

The regressions of winter workday with the regression residuals are illustrated in Fig. 12 as an example. It should be noted that there is an abnormal residual value around 1250. This is because this building holds several companies which have different holiday policies. Although the national standard for spring festival holiday is seven days, some of these companies may extend it to ten days or longer, so that the occupancy was noticeably less in this period.

The coefficient of determination (R-squared value) is often used as the indicator to evaluate the performance of the regression. It is defined in Eq. (7). R-squared is a statistical measure of how close the data are to the fitted regression line. The R-squared values of the five groups are all higher than 0.95 in this case study, which indicates that the regression process yields good accuracy, and five occupancy schedules are sufficient to represent the variation of the building occupancy in a whole year. In the future, we will investigate whether disaggregated schedules (e.g., daily by daily) will further improve the simulation accuracy. Five

Fig. 10. The geographical information of the simulated building.

occupancy schedules are presented in Fig. 13. It can be seen that compared with the summer condition, the occupant number in the winter condition is obviously smaller although the variation patterns are similar. Besides, compared with the code-based values, an apparent time lag is observed in the occupancy schedules obtained from the mobile-internet-based positioning data. The occupancy schedule of the winter workday obtained from the regression with raw data from 40 actual days was presented in Fig. 14. The CV(RMSE) is calculated to recognize the best and worst cases for the regression. The result suggests that the (CV)RMSEs of the best and worst cases are smaller than 4% and 10%, respectively, which demonstrates that the proposed regression yields a reasonable accuracy.

$$R=\frac{n\sum y\cdot\hat{y}-\sum y\cdot\sum\hat{y}}{\sqrt{(n\sum y^2-(\sum y)^2)(n\sum\hat{y}^2-(\sum\hat{y})^2)}} \qquad \text{Eq. (7)}$$

The occupancy density is generated from an on-site survey with the building manager. The occupancy densities of the shopping and office area were adjusted from 8 and 6 m^2 per person to 6 and 4.8 m^2 per person respectively. In this EnergyPlus based simulation case study, the occupants mainly influenced the building energy consumption by the related internal heat gains. The changes in the occupancy schedule (both amplitude and "timing") and density had the impacts on the building energy consumption since the building internal load was changed. However, in this study, we didn't study which occupancy related parameters (e.g., the people density or the timing of the time series occupancy) have the bigger impact on the simulation results. This will involve a systematic and complicated sensitivity study on the current EnergyPlus model such as the study from Refs. [74] [75], and [76].

It should be noted that the occupancy patterns of both areas (office and shopping) are assumed to be similar in this case study for the simplification. These two areas will be separately monitored in the future to improve the occupancy estimation accuracy.

4.4.2. Simulation result of the further calibration with an occupancy-schedule-based method

The simulation result of this calibration is presented in Fig. 15. It can be seen that the simulating energy consumption is in a good agreement with the measured value. The MBE_{year} and $CV(RMSE)_{month}$ are 2.5% and 4.9% respectively. MBE_{month} values are all less than 5%. All three indexes fall within the range of the recommended values of relevant guidelines [54]. This indicates that no further calibration is necessary. The possible reasons are as follows: Firstly, the occupant number in the summer increased after the new occupancy schedules and densities were used, thus, the internal loads related to the occupants were increased. This change caused a consequent growth of the energy consumption. Secondly, the occupant number in the winter decreased after new occupancy information was used. The higher building heating loads consumed more energy in the winter.

Besides, it's worth noting that the simulated energy consumption of February is slightly higher than the measured one. This is partly due to what has been mentioned above that the different holiday policies made the actual occupancy smaller than the regressed one. The possible reasons why the simulation result was good even if the AMY was not used are described as follows: Firstly, the building envelope performance for this new office-shopping center building was good that the

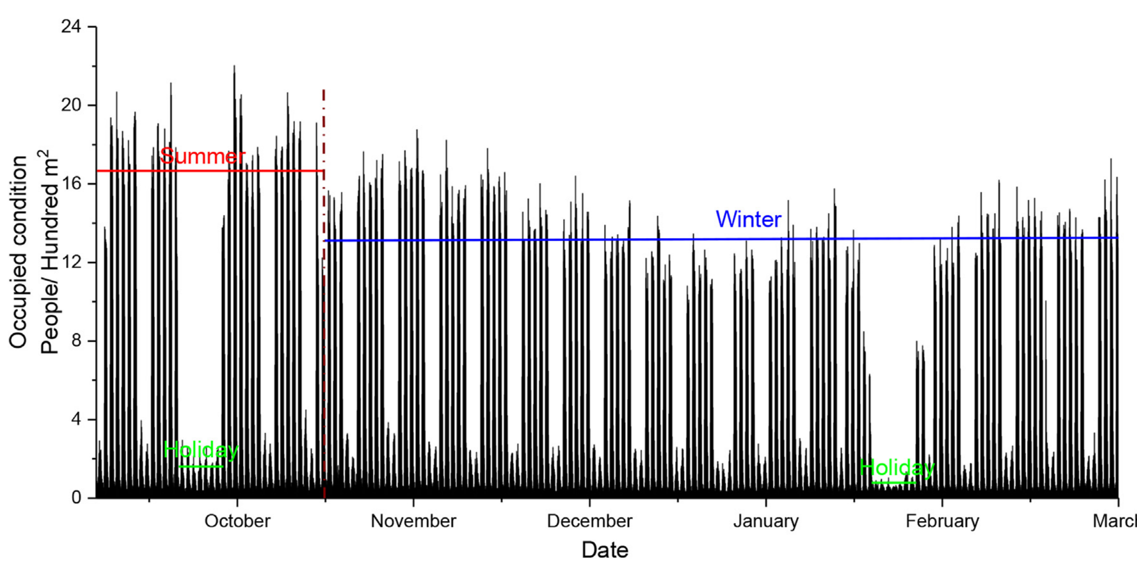

Fig. 11. Samples of data information (Sept. 2016 to Mar. 2017).

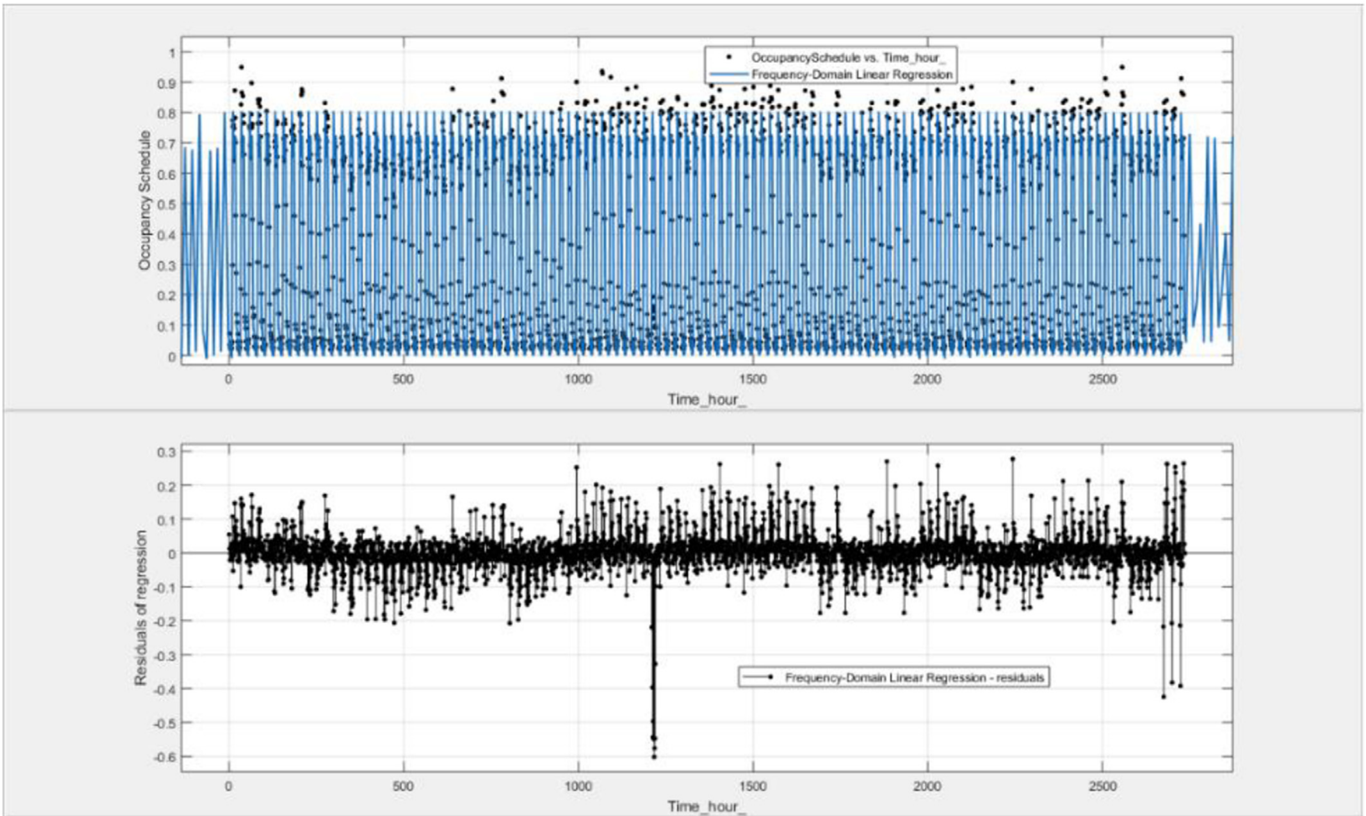

Fig. 12. Frequency domain linear regression of occupancy schedule (winter workday).

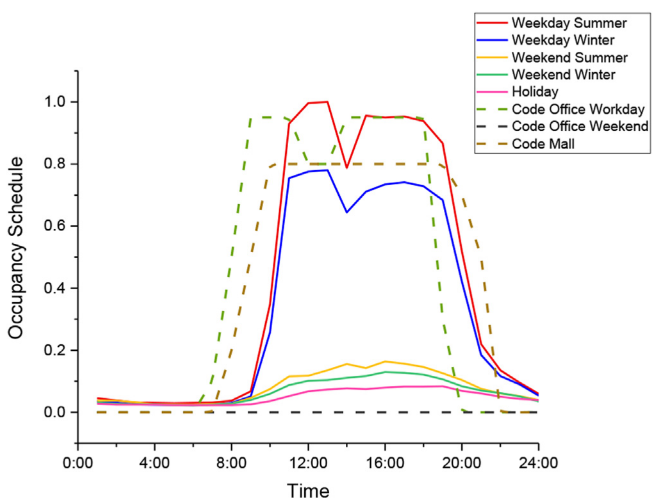

Fig. 13. Occupancy schedules derived from mobile-internet-based data.

impact of the weather was relatively small compared with internal heat gains. Secondly, this calibration was conducted using the lumped monthly energy usage instead of the daily/hourly energy consumption. If we conducted the calibration on a smaller time scale, the impact of the weather could be more predominant.

The good results of this case study prove that using the real occupancy schedule which is generated from the mobile-internet-based occupancy data can facilitate the building energy simulation. However, it should be noted that in this case study, the candidate building was newly built, and the HVAC system and other mechanical equipment are all in good conditions. Therefore, using the BAS data and a simple replacement of the code-based occupancy information is able to offer an acceptable energy prediction from the whole building energy performance simulation. Furthermore, this good result does not mean that the usage of the mobile-internet-based occupancy data can replace other calibration approaches including auto-tuning as described in Section 3.1.

5. Conclusions and future work

5.1. Conclusions

This study investigates the applicability of mobile-internet-based occupancy data in building energy simulation and whether it can simplify the calibration process and reduce the model distortion. An energy performance model of an office and commercial complex building in Shanghai is introduced as a case study. A preliminary calibration is conducted only using the history data (the electricity consumption of lighting and equipment) from the BAS in the first step. The lighting power density (LPD) and equipment power density (EPD) are adjusted in this calibration. The further calibration is then performed by replacing the code-based occupancy information with the mobile-internet-based occupancy information. The results suggest that the presented method can help overcome some intrinsic issues for building energy simulation. Both two indexes (i.e., MBE and CV(RMSE)) fall within the recommended range of the standard (e.g., ASHRAE Guideline [54]). Other key conclusions from this pilot case study are listed as follows:

- The proposed approach of using BAS data and the mobile-internet-based occupancy data for building energy simulation offers a potential alternative to building modelers and the practitioners who don't have a solid background in the HVAC domain.
- This proposed approach could lead to a less possible model distortion during the simulation and calibration process because it is relying on the actual BAS data and mobile-internet-based occupancy information. In other traditional calibration methods, although the

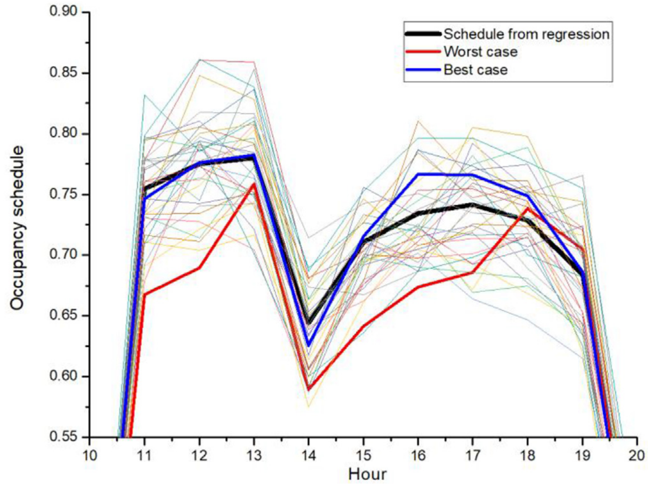

Fig. 14. The schedule from the regression vs. raw data from 40 days' actual cases for the winter workday group (11:00–19:00).

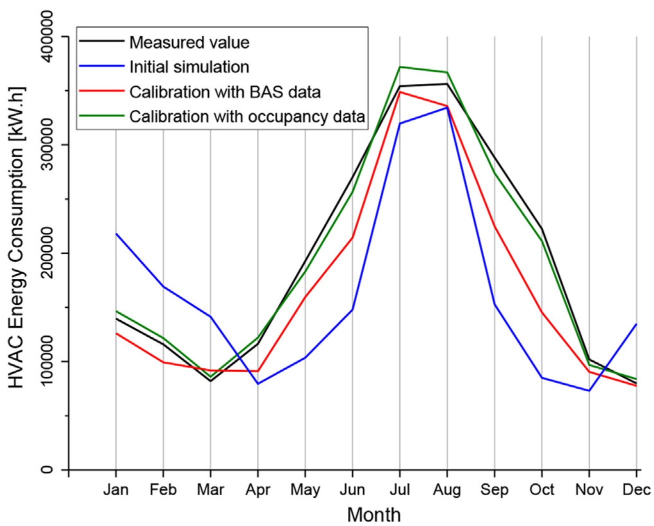

Fig. 15. The simulation results of the initial simulation, calibration with the BAS data, and the calibration with the occupancy data.

final result might meet the criterion of the standards, the model may be seriously distorted and cannot well represent the actual building.

- The mobile-positioning based occupancy detection method is advantageous to others for its low cost and convenience. There is no need to deploy any additional sensors, and thus a large amount of implementation and maintenance fees can be saved. Besides, it doesn't require an advanced algorithm to process the raw data. Such a supervised tracking system can be applied to any commercial building as long as the permission is granted.

5.2. Future work

Although the mobile-internet-based data has been proved to be effective in this pilot case study, there still remain some issues to be addressed in the further work.

- The weather condition could have some impacts on the building energy consumption. Actual Meteorological Year (AMY) data should be collected and used for the simulation and the calibration in the future.
- More case studies should be conducted to prove the scalability of this proposed approach.
- Some other machine learning methods, e.g., clustering analysis, will be introduced to process occupancy information from mobile positioning data.
- The current method to specify the monitoring area is not accurate enough, especially for those buildings with irregular shapes. A smaller measuring scale should be provided in the future to reduce the related errors.
- The different areas of a multi-purpose building should be separately monitored to improve the accuracy level of the mobile positioning data.
- Since mobile-internet-based occupancy data only takes into account those who are active smart phone users. A correction coefficient could probably be introduced to improve the data completeness. A large-scale survey (such as whether use smart phone, whether use mobile-internet-service, etc.) will be necessary to help get such correction coefficient. It is expected that there will be variations of this coefficient in different regions and countries.

Additionally, it should be noted that authors fully understand and respect the privacy right of every internet user. The occupancy information obtained in this pilot case study was tracked under a supervision and only used for the presented building energy simulation research. There is a need to integrate cyber-security and privacy research with the mobile-internet-based occupancy data acquisition. Depending on the local laws and regulations, this data source may not be available in some countries and regions.

Acknowledgement

This research is supported by a Grant from Ministry of Industry and Information Technology, China (Project No. 2017YFB0903400).

References

[1] T. Barker, I. Bashmakov, L. Bernstein, Technical summary, in: B. Metz, O.R. Davidson, P.R. Bosch, et al. (Eds.), Climate Change 2007: Mitigation, CUP, Cambridge, UK, 2007.
[2] J. Hou, et al., Implementation of expansion planning in existing district energy system: a case study in China, Appl. Energy 211 (2018) 269–281.
[3] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build. 40 (3) (2008) 394–398.
[4] W. Li, et al., Electricity demand response in China: status, feasible market schemes and pilots, Energy 114 (2016) 981–994.
[5] C. Turner, M. Frankel, Energy performance of LEED for new construction buildings, New Build. Inst. 4 (2008) 1–42.
[6] F. Karlsson, P. Rohdin, M.-L. Persson, Measured and predicted energy demand of a low energy building: important aspects when using building energy simulation, Build. Serv. Eng. Technol. 28 (3) (2007) 223–235.
[7] P. Hoes, et al., User behavior in whole building simulation, Energy Build. 41 (3) (2009) 295–302.
[8] V. Fabi, et al., Occupants' window opening behaviour: a literature review of factors influencing occupant behaviour and models, Build. Environ. 58 (2012) 188–198.
[9] Z. Yu, et al., A systematic procedure to study the influence of occupant behavior on building energy consumption, Energy Build. 43 (6) (2011) 1409–1417.
[10] F. Oldewurtel, D. Sturzenegger, M. Morari, Importance of occupancy information for building climate control, Appl. Energy 101 (2013) 521–532.
[11] V.L. Erickson, A.E. Cerpa, Occupancy based demand response HVAC control strategy, Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-efficiency in Building, ACM, 2010.
[12] J. Virote, R. Neves-Silva, Stochastic models for building energy prediction based on occupant behavior assessment, Energy Build. 53 (2012) 183–193.
[13] K. Mekhnacha, et al., Bayesian occupancy filter based" fast clustering-tracking" algorithm, IROS 2008, 2008.
[14] M. Yalcintas, Energy-savings predictions for building-equipment retrofits, Energy Build. 40 (12) (2008) 2111–2120.
[15] F.I. Vázquez, W. Kastner, Clustering methods for occupancy prediction in smart home control, Industrial Electronics (ISIE), 2011 IEEE International Symposium on, IEEE, 2011.
[16] S. Lee, et al., Occupancy prediction algorithms for thermostat control systems using mobile devices, IEEE Trans. Smart Grid 4 (3) (2013) 1332–1340.
[17] W.-k. Chang, T. Hong, Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data, Building Simulation, Springer, 2013.
[18] J. Page, et al., A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2) (2008) 83–98.

[19] https://energyplus.net/.
[20] http://www.doe2.com/equest/.
[21] http://www.trnsys.com/.
[22] T.T. Yang, K.Y. Yip, Use of MOBITEX wireless wide area networks as a solution to land-based positioning and navigation, IEEE Aero. Electron. Syst. Mag. 9 (7) (1994) 29–35.
[23] Z. Shen, M. Li, Big Data Support of Urban Planning and Management: the Experience in China, Springer, 2017.
[24] Y. Li, D. Chen, A learning-based comprehensive evaluation model for traffic data quality in intelligent transportation systems, Multimed. Tool. Appl. 75 (19) (2016) 11683–11698.
[25] T. Labeodan, et al., Occupancy measurement in commercial office buildings for demand-driven control applications—a survey and detection system evaluation, Energy Build. 93 (2015) 303–314.
[26] X. Guo, et al., The performance of occupancy-based lighting control systems: a review, Light. Res. Technol. 42 (4) (2010) 415–431.
[27] R.H. Dodier, et al., Building occupancy detection through sensor belief networks, Energy Build. 38 (9) (2006) 1033–1043.
[28] W.J. Fisk, D. Faulkner, D.P. Sullivan, Accuracy of CO2 Sensors in Commercial Buildings: a Pilot Study, (2006).
[29] X. Liang, T. Hong, G.Q. Shen, Occupancy data analytics and prediction: a case study, Build. Environ. 102 (2016) 179–192.
[30] S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408.
[31] X. Luo, et al., Performance evaluation of an agent-based occupancy simulation model, Build. Environ. 115 (2017) 42–53.
[32] Y.-S. Kim, Improvement of building energy simulation accuracy with occupancy schedules derived from hourly building electricity consumption, Build. Eng. 121 (2015) 353.
[33] K. Christensen, et al., Using existing network infrastructure to estimate building occupancy and control plugged-in devices in user workspaces, Int. J. Commun. Network. Distr. Syst. 12 (1) (2014) 4–29.
[34] C. Martani, et al., ENERNET: studying the dynamic relationship between building occupancy and energy consumption, Energy Build. 47 (2012) 584–591.
[35] S.P. Tarzia, et al., Sonar-based measurement of user presence and attention, Proceedings of the 11th International Conference on Ubiquitous Computing, ACM, 2009.
[36] B. Dong, et al., An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network, Energy Build. 42 (7) (2010) 1038–1046.
[37] K.P. Lam, et al., Occupancy detection through an extensive environmental sensor network in an open-plan office building, IBPSA Build. Simulat. 145 (2009) 1452–1459.
[38] J.A. Davis III, D.W. Nutter, Occupancy diversity factors for common university building types, Energy Build. 42 (9) (2010) 1543–1551.
[39] M.S. Gul, S. Patidar, Understanding the energy consumption and occupancy of a multi-purpose academic building, Energy Build. 87 (2015) 155–165.
[40] S. Uziel, et al., Networked embedded acoustic processing system for smart building applications, Design and Architectures for Signal and Image Processing (DASIP), 2013 Conference on, IEEE, 2013.
[41] S. Wang, X. Jin, CO2-based occupancy detection for on-line outdoor air flow control, Indoor Built Environ. 7 (3) (1998) 165–181.
[42] S. Wang, J. Burnett, H. Chong, Experimental validation of CO2-based occupancy detection for demand-controlled ventilation, Indoor Built Environ. 8 (6) (1999) 377–391.
[43] V.L. Erickson, M.Á. Carreira-Perpiñán, A.E. Cerpa, OBSERVE: occupancy-based system for efficient reduction of HVAC energy, Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on, IEEE, 2011.
[44] R. Ahas, et al., Using mobile positioning data to model locations meaningful to users of mobile phones, J. Urban Technol. 17 (1) (2010) 3–27.
[45] R. Ahas, et al., Evaluating passive mobile positioning data for tourism surveys: an Estonian case study, Tourism Manag. 29 (3) (2008) 469–486.
[46] https://www.gps.gov/systems/gps/performance/accuracy/.
[47] https://www.tencent.com/en-us/articles/17000341491836558.pdf.
[48] M. Deru, et al., US Department of Energy Commercial Reference Building Models of the National Building Stock, (2011).
[49] J.L. Hensen, R. Lamberts, Building Performance Simulation for Design and Operation, Routledge, 2012.
[50] Y. Heo, R. Choudhary, G. Augenbroe, Calibration of building energy models for retrofit analysis under uncertainty, Energy Build. 47 (2012) 550–560.
[51] T.A. Reddy, Literature review on calibration of building energy simulation programs: uses, problems, procedures, uncertainty, and tools, Build. Eng. 112 (1) (2006).
[52] Z. Pang, et al., Evaluation of the performance of a new solar ventilated window: modeling and experimental verification, J. Renew. Sustain. Energy 9 (6) (2017) 065101.
[53] Z. O'Neill, et al., Model-based real-time whole building energy performance monitoring and diagnostics, J. Build. Perform. Simulat. 7 (2) (2014) 83–99.
[54] A. Guideline, Guideline 14-2014, Measurement of Energy and Demand Savings, American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia, 2014.
[55] International Performance Measurement and Verification Protocol (IPMVP). 2007. vol. 1.
[56] (FEMP), F.E.M.P., Measurement and Verification Guideline (IPMVP), (2008) (Chapter 25).
[57] R. Troncoso, A hybrid monitoring-modeling procedure for analyzing the performance of large central chilling plants, Proceedings of Building Simulation, 1997.
[58] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renew. Sustain. Energy Rev. 37 (2014) 123–141.
[59] J. Clarke, P. Strachan, C. Pernot, An approach to the calibration of building energy simulation models, Trans. Am. Soc. Heat. Refrigerat. Air Condition. Eng. 99 (1993) 917–917.
[60] M. Liu, et al., Manual of Procedures for Calibrating Simulations of Building Systems, Lawrence Berkeley National Laboratory, Berkeley, CA, 2003.
[61] Z. O'Neill, B. Eisenhower, Leveraging the analysis of parametric uncertainty for building energy model calibration, Building Simulation, Springer, 2013.
[62] G. Chaudhary, et al., Evaluation of “Autotune” calibration against manual calibration of building energy models, Appl. Energy 182 (2016) 115–134.
[63] D. Yan, et al., IEA EBC Annex 66: definition and simulation of occupant behavior in buildings, Energy Build. 156 (2017) 258–270.
[64] F. Xiao, C. Fan, Data mining in building automation system for improving building operational performance, Energy Build. 75 (2014) 109–118.
[65] Y. Pan, et al., The application of building energy simulation and calibration in two high-rise commercial buildings in Shanghai, Proce. SimBuild 2 (1) (2006).
[66] Y. Pan, Z. Huang, G. Wu, Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai, Energy Build. 39 (6) (2007) 651–657.
[67] https://www.sketchup.com/products/sketchup-pro.
[68] https://www.openstudio.net/downloads.
[69] https://energyplus.net/weather.
[70] B.C. Bureau, Design Standard for Energy Efficiency of Public Buildings, China Architecture and Building Press, 2005. Also available at:http://bbs.topenergy.org/redirect.php.
[71] Y. Ji, P. Xu, Y. Ye, HVAC terminal hourly end-use disaggregation in commercial buildings with Fourier series model, Energy Build. 97 (2015) 33–46.
[72] Fuxin Niu, Z.D.O.N., bayesian network based HVAC energy consumption prediction using improved fourier series decomposition, 2016 ASHRAE Winter Meeting, 2016 (Orlando, FL).
[73] Niu, F., Z. O'Neill, and C. O'Neill. Data-driven based estimation of HVAC energy consumption using an improved Fourier series decomposition in buildings. in Building Simulation. Springer.
[74] Z. O'Neill, F. Niu, Uncertainty and sensitivity analysis of spatio-temporal occupant behaviors on residential building energy usage utilizing Karhunen-Loève expansion, Build. Environ. 115 (2017) 157–172.
[75] S. Qiu, Z. Li, Z. Pang, et al., A quick auto-calibration approach based on normative energy models, Energy Build. (2018).
[76] A.S. Silva, E. Ghisi, Uncertainty analysis of user behaviour and physical parameters in residential building performance simulation, Energy Build. 76 (2014) 381–391.