Building and Environment 169 (2020) 106549

Contents lists available at ScienceDirect

Building and Environment

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

# Dynamic occupant density models of commercial buildings for urban energy simulation

Chao Wang [a,b], Yue Wu [a,b], Xing Shi [a,b,*], Yanxia Li [a,b], Sijie Zhu [a,b], Xing Jin [a,b], Xin Zhou [a,b]

[a] School of Architecture, Southeast University, Nanjing, 210096, China
[b] Key Laboratory of Urban and Architectural Heritage Conservation, Ministry of Education, Nanjing, 210096, China

ARTICLE INFO

*Keywords:*
Big data
Commercial buildings
Urban-level
Dynamic occupant density models

ABSTRACT

The number of occupants and its changing pattern over time are key information for building and urban energy simulation. However, the commonly used assumption and simplification of a fixed occupancy schedule does not reflect the complicated reality, leading to significant errors in energy simulation. Therefore, dynamic occupant density models which describe the real-world situation more accurately should be developed. This paper presents a methodology to develop such a model for commercial buildings and expand it from the building level to urban level. First, a total of 2275 commercial buildings in Nanjing, a major city in China, are identified and classified into three sub-categories using Points of Interest and logistic regression. Then field measurement is conducted to obtain the hourly occupant density for 12 sample commercial buildings. The building-level dynamic occupant density model is developed by fitting normal distribution functions into the measured data. Finally, transportation accessibility and population level, two urban parameters, are defined and used to expand the building-level occupant density model to the urban-level one. The dynamic urban-level occupant density model is verified for all three sub-categories of commercial buildings and the overall results are acceptable.

## 1. Introduction

Facing the current global energy shortage and environmental deterioration, countries around the world have developed long-term targets for energy conservation and greenhouse gas (GHG) emission reduction. The United States of America (USA) has set targets of reducing GHG emission by 17% by 2020 and 26–28% by 2025, compared with the 2005 emission level [1]. The European Union (EU) has set the goal of reducing GHG emission by 40% by 2030 from the 1990 level [2]. China also sets the target of reducing carbon dioxide emission per unit of GDP by 40–45% by 2020, compared with 2005 emission level [3]. Since urban energy consumption accounts for a significant portion of the total energy consumption, planning, designing, building and managing energy-efficient and low-carbon cities becomes crucial to achieving these goals.

To make energy-efficient and low-carbon cities a reality, Urban Building Energy Models (UBEMs) have been developed to calculate building energy demands at urban scale [4]. UBEMs are bottom-up, physical simulation models of heat and mass flows in and around groups of buildings to predict energy consumption as well as indoor and outdoor environmental conditions [5]. Similar to Building Energy Models (BEMs), the generation of an UBEM requires numerous data inputs, including geometric data and non-geometric data [6]. The accuracy of the UBEM depends on the completeness and accuracy of the data inputs [7]. Thus, inaccurate data inputs could lead to unreliable UBEMs, which may have a negative impact on planning, designing, building and managing energy-efficient and low-carbon cities.

Occupant density, as one of the most fundamental data inputs for BEMs and UBEMs, has a great impact on building energy consumption and thus is one of the basic information required by energy simulation programs such as EnergyPlus and DeST [8]. D'Oca [9] and Monteiro [10] argued that a deep understanding of the relationship between occupant density and building energy consumption is one of the most effective ways to bridge the gap between predicted and actual energy consumption of buildings. Menezes [11] and Kwok [12] reported that the energy performance can be correctly predicted by using credible dynamic occupant information. Reliable occupant information is important for energy simulation because of many reasons [13], including the generation of heat gains within the building [14], the interaction with HVAC systems, lighting and hot water supplies [15–17], and the adjustment of windows [18–20].

Occupant density in almost all buildings is dynamic and continuous

* Corresponding author. School of Architecture, Southeast University, Nanjing, 210096, China.
*E-mail address:* shixing_seu@163.com (X. Shi).

https://doi.org/10.1016/j.buildenv.2019.106549
Received 1 July 2019; Received in revised form 12 October 2019; Accepted 12 November 2019
Available online 18 November 2019

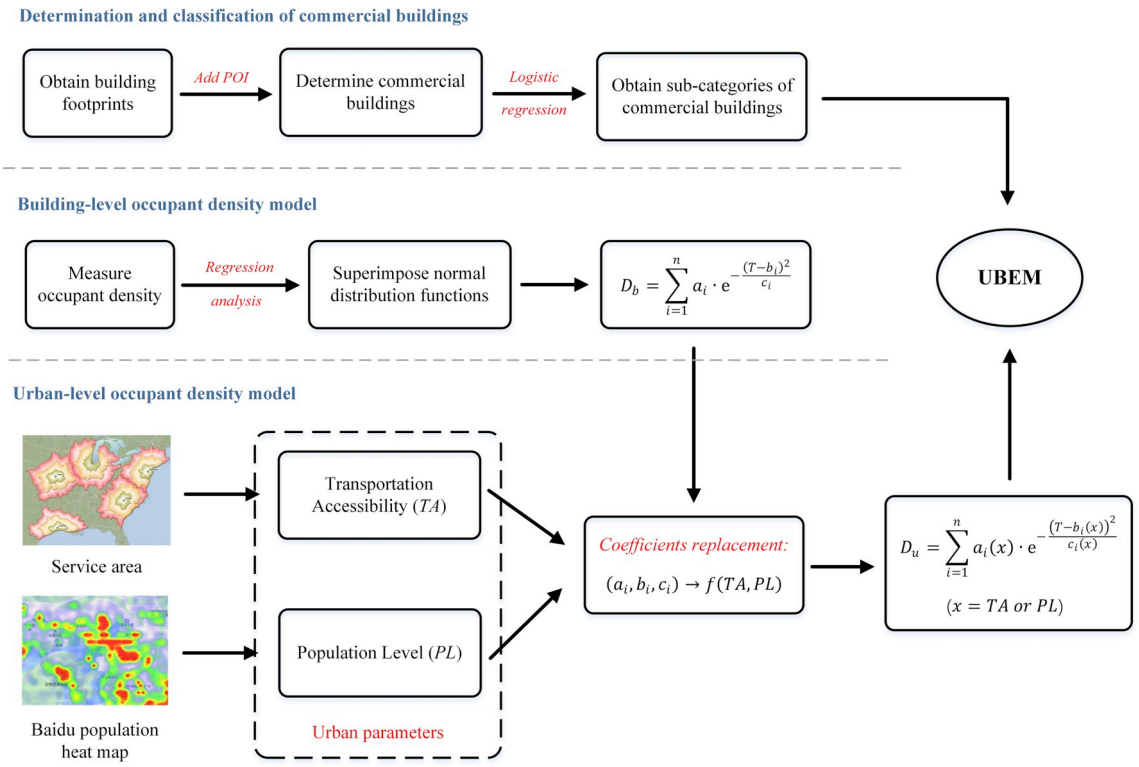



**Fig. 1.** Flowchart of the research.

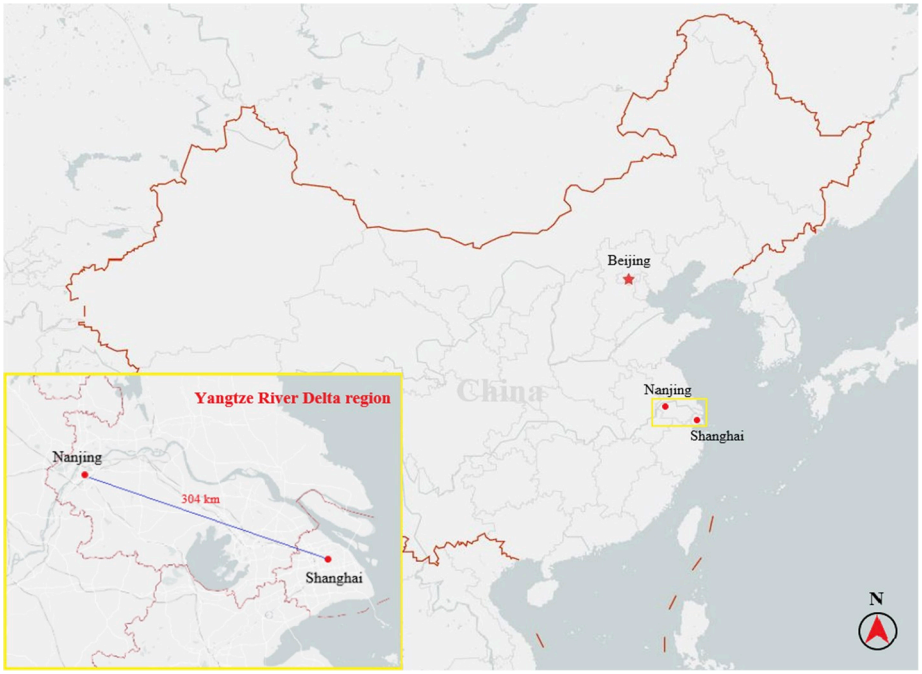



**Fig. 2.** The geographic location of Nanjing.

[21], i.e., it is not a constant number and changes with time. However, occupant density is usually assumed to follow a fixed schedule which describes when the building space is occupied and how many occupants are in the space [22]. A fixed occupant schedule is simple and easy to be incorporated into building energy simulation. However, it could differ up to 40% from the actual occupancy situation [23] and lead to significant simulation errors [24,25]. Thus, a dynamic occupant density model which can better reflect the changing state of building occupants is needed. Some research works have attempted to address this issue by investigating the number of occupants [26,27] and occupant presence [28,29] in buildings. Although they shed some light on the dynamic nature of occupant density in buildings, the problem of determining occupant information for UBEMs still remains because it requires the occupant information of hundreds of or even thousands of buildings with different occupancy characteristics. Thus, an innovative approach which can be applied to numerous urban buildings is needed.

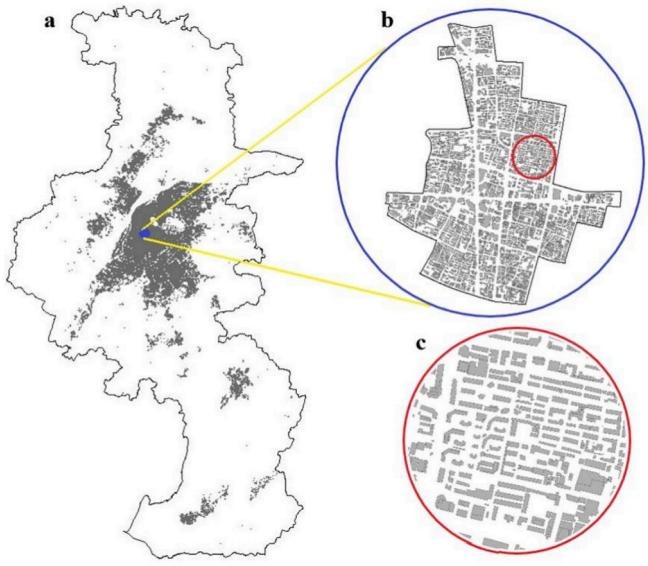



**Fig. 3.** Building footprints of Nanjing (a: administrative area of Nanjing; b: central business district of Nanjing; c: a representative neighborhood of Nanjing).

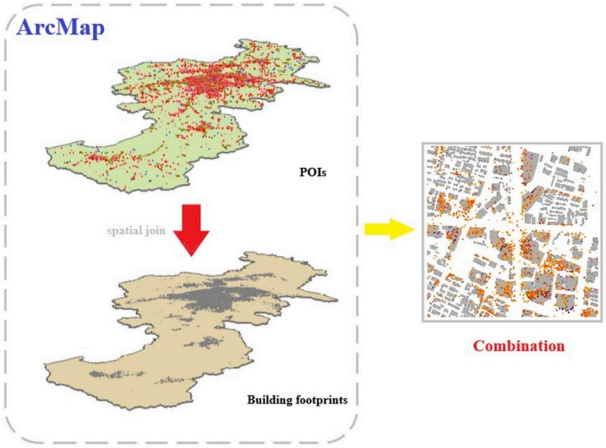



**Fig. 4.** The combination of POIs and building footprints.

**Table 1**
The number of eight categories of commercial POIs for 6 representative buildings.

| Building Name | LR | SR | SM | S | BM | EA | LF | SF |
|---|---|---|---|---|---|---|---|---|
| Deji Plaza | 33 | 47 | 2 | 1 | 3 | 10 | 1 | 2 |
| Wanda Plaza | 110 | 103 | 2 | 2 | 10 | 14 | 3 | 7 |
| Wal-Mart | 1 | 10 | 0 | 1 | 1 | 1 | 0 | 1 |
| Metro | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Home Mall | 0 | 1 | 0 | 0 | 177 | 9 | 78 | 99 |
| Gome | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |

**POI:** LR-Large Restaurant, SR-Small Restaurant, SM-Shopping Mall, S-Supermarket, BM-Building Material, EA-Electric Appliance, LF-Large Furniture, SF-Small Furniture.

Urban buildings can be categorized into multiple types, among which commercial buildings are an important one because of their high numbers, large area, and significant energy consumption. The definition of commercial buildings may not be exactly the same in different countries. In China, they are defined as a broad range of buildings such as shopping malls, supermarkets and specialty stores (the stores which

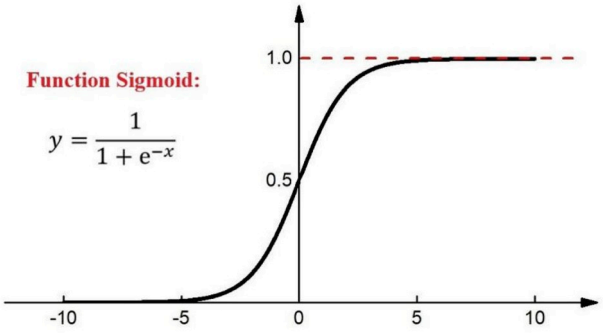



**Fig. 5.** The graph of Sigmoid.

are specialized in selling one category of commodities), according to *Code for design of store buildings* JGJ48-2014 [30]. While in the USA and EU, hotels and offices may be considered as commercial buildings [31,32]. The occupancy pattern of commercial buildings is generally more complex than that of other types of buildings such as offices and residential buildings. To be more specific, the number of occupants in a commercial building tends to fluctuate significantly over time. In addition, the dynamic nature of this fluctuation is more difficult to predict. Such a clear pattern cannot be easily identified for most, if not all, commercial buildings due to their business nature. Therefore, the dynamic occupant density model of commercial buildings is worth investigating. The understanding and knowledge gained can benefit UBEMs, BEMs, and other relevant research subjects. The methodology developed can be applied to other types of buildings as well.

Despite the importance and complexity of commercial buildings, the research work on their occupancy information is inadequate. It is revealed that the superposition of normal distribution functions, also called Gaussian Mixture Model (GMM), can well fit the occupancy profile in office buildings [21,29]. However, nearly no research has discussed whether the GMM can be applied into commercial buildings. Only a few of studies discussed the maximum occupant density of commercial buildings in the field of building fire protection. These researches offer methods in establishing the relationships between maximum occupant density and some parameters. Feng [33] treated business style, business periods and city economic level as the key parameters to describe maximum occupant density in commercial buildings, using general linear regression model. To be extended, the coefficient in occupant density models for urban commercial buildings could also be linked with some parameters, especially urban parameters. It is widely accepted that population density and transportation are two main urban parameters [34–36]. Besides, the type of commercial buildings may also have great effects on occupant density models. However, the classification of thousands of commercial buildings at urban level is quite hard and time-consuming. One method in dealing with this issue is to use Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Niu [37] used multi-source data, such as WeChat users' real-time location records, taxi GPS trajectories data and Points of Interest (POIs) data, and applies DBSCAN to infer the functions of urban buildings. It is undeniable that DBSCAN is an efficient attempt to determine the functions of buildings due to its simple structure and high computational efficiency. However, the data needed are complicated and hard to obtain, e.g. taxi GPS trajectories data is not free in China [38]. Thus, a method of classification based on open data source is more urgently needed, as it could be transplanted to other projects.

This paper presents an innovative method to develop dynamic occupant density models for commercial buildings in the context of UBEMs. It is organized as follows: Section 2 introduces the basic information needed for the research; Section 3 describes the methodology for the determination and classification of urban commercial buildings, which is based on the POI and logistic regression; Section 4 explains the

**Table 2**
The weights of POIs in each classifier.

| | $\theta_0$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\theta_5$ | $\theta_6$ | $\theta_7$ | $\theta_8$ |
|---|---|---|---|---|---|---|---|---|---|
| $y_1$ | −2.9482 | 0.5637 | −0.0509 | 2.3524 | −1.3836 | −0.3860 | −0.1901 | 0.2182 | 0.3958 |
| $y_2$ | 0.2684 | −0.3846 | 0.1161 | −1.9095 | 3.0519 | −0.9217 | −0.6649 | 0.2101 | −0.1317 |
| $y_3$ | 1.1464 | −0.0957 | −0.4202 | −0.4690 | −3.1597 | −0.1398 | 0.3083 | 0.6483 | 0.2119 |

$y_1 \sim y_3$: The classifiers used to judge shopping malls, supermarkets and specialty stores respectively.
$\theta_0$: Constant term; $\theta_1 \sim \theta_8$: The weights of POIs which represent large restaurant, small restaurant, shopping mall, supermarket, building material, electric appliance, small furniture, and large furniture respectively.

**Table 3**
Predicted and actual sub-categories of 20 commercial buildings.

| ID | $y_1$ | $y_2$ | $y_3$ | Predicted | Actual |
|---|---|---|---|---|---|
| 1 | 1.0000 | 0 | 0.5294 | SM | SM |
| 2 | 0.9998 | 0 | 0 | SM | SM |
| 3 | 1.0000 | 0 | 0 | SM | SM |
| 4 | 0.3758 | 0.0407 | 0.0001 | SM | SM |
| 5 | 0.9956 | 0.0032 | 0 | SM | SM |
| 6 | 0.9999 | 0 | 0 | SM | SM |
| 7 | 0.9981 | 0.0003 | 0.0001 | SM | SM |
| 8 | 1.0000 | 0 | 0 | SM | SM |
| 9 | 0.0350 | 0.6177 | 0.0158 | S | S |
| 10 | 0.0187 | 0.9064 | 0.1418 | S | S |
| 11 | 0.0204 | 0.9596 | 0.0498 | S | S |
| 12 | 0.0550 | 0.9328 | 0.0183 | S | S |
| 13 | 0.0321 | 0.9533 | 0.0201 | S | S |
| 14 | 0.0032 | 0 | 1.0000 | SS | SS |
| 15 | 0.0051 | 0 | 1.0000 | SS | SS |
| 16 | 0.0100 | 0 | 0.9984 | SS | SS |
| 17 | 0 | 0 | 1.0000 | SS | SS |
| 18 | 0.0361 | 0.2396 | 0.8969 | SS | SS |
| 19 | 0.1827 | 0.1924 | 0.6786 | SS | SS |
| 20 | 0.0675 | 0.3396 | 0.7188 | SS | SS |

SM: Shopping Mall, S: Supermarket, SS: Specialty Store.

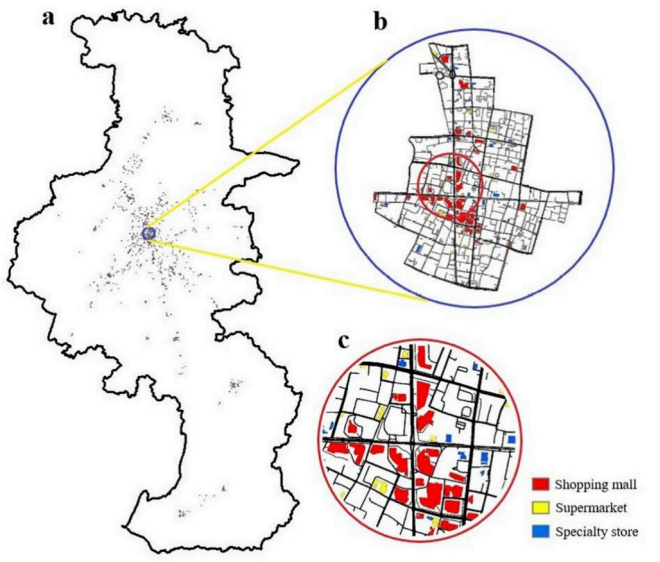



**Fig. 6.** Final classification of Nanjing commercial buildings (a: administrative area of Nanjing; b: central business district of Nanjing; c: a representative neighborhood of Nanjing).

development of the building-level dynamic occupant density model. The building-level model is obtained based on measured data and regression analysis, and can be described as the superposition of normal distribution functions; Section 5 presents the urban-level dynamic occupant density model which is expanded from the building-level model by establishing the relationship between the model coefficients and the key urban parameters; Section 6 offers further discussions on the subject; and finally, Section 7 summarizes the main conclusions. Fig. 1 illustrates the structure of the paper.

## 2. Basic information

### 2.1. Case city

Nanjing is selected as the case city to demonstrate the methodology developed in this study. It is a major city in China with more than 8 million people, located in the Yangtze River Delta region. The administrative area of Nanjing is 6587 $km^2$ and the urban built-up area is 1399 $km^2$ in 2017 compared with 513 $km^2$ in 2005 [39], suggesting its rapid urbanization. The climatic condition of Nanjing is categorized as "hot summer and cold winter" in accordance with *Code for thermal design of civil building* GB 50176–2016 [40]. Fig. 2 illustrates the geographic location of Nanjing.

### 2.2. Building footprints

Geometric information of urban buildings is essential for UBEMs. Three-dimensional (3D) geometric urban models include the information of building footprints and building heights. Compared with the information of building heights, the information of building footprints is more critical to the determination and classification of commercial buildings. In addition, the quality of 3D urban models relies heavily on the quality of building footprints [41]. To obtain building footprints, various data sources can be used, such as the existing 2D GIS (Geographic Information System) datasets [42], the remote sensing imagery [43], the air photogrammetric surveys [44], and OpenStreetMap [45].

In China, Baidu is one of the most successful Internet companies. It announced in 2010 that the APIs (Application Programming Interface) of Baidu Map would be open to the public for free. It soon became one of the most commonly used open source map for research in China [46,47]. The building footprints of Nanjing, as shown in Fig. 3, were obtained by extracting data through Baidu Map APIs. Note that these building footprints are in PNG (Portable Network Graphics) format, and should be transformed into vector data so that they can be further processed.

### 2.3. Point of interest

In accordance with the International Organization for Standardization (ISO) [48], POIs (Points of Interest) are defined as those physical features, especially the facilities closely related to people's lives, which can be abstracted as points and applied to electronic maps. A POI has its specific influence area and concerned groups. In addition, a POI usually contains some useful attributive information which can be used to classify urban buildings. The POIs of Nanjing were extracted from Baidu Map through the corresponding API. The keywords assigned are *shopping mall*, *supermarket*, *electrical appliance*, *building material*, *large furniture*, and *small furniture*, representing the most common business types in commercial buildings, namely shopping malls, supermarkets and specialty stores. In addition, since a large number of commercial buildings have rented spaces for dining, two more keywords, namely *large*


**Fig. 7.** The location and representative pictures of the 12 sample commercial buildings.

**Table 4**
The floor area and open hours of the 12 sample commercial buildings.

| Notation | Floor area ($m^2$) | Open hours (Workday) | Open hours (Weekend) | Type |
|---|---|---|---|---|
| SM1 | 48792 | 10:00–22:00 | 10:00–22:00 | Shopping mall |
| SM2 | 94144 | 10:00–22:00 | 10:00–22:00 | |
| SM3 | 39129 | 10:00–22:00 | 10:00–22:00 | |
| SM4 | 181739 | 10:00–22:00 | 10:00–22:00 | |
| S1 | 4428 | 08:00–22:00 | 08:00–22:00 | Supermarket |
| S2 | 15166 | 08:00–22:00 | 08:00–22:00 | |
| S3 | 11806 | 08:00–22:00 | 08:00–22:00 | |
| S4 | 27408 | 08:00–22:00 | 08:00–22:00 | |
| SS1 | 8212 | 10:00–21:00 | 10:00–21:00 | Specialty store |
| SS2 | 7855 | 09:00–21:00 | 09:00–21:00 | |
| SS3 | 19652 | 09:00–21:00 | 09:00–21:00 | |
| SS4 | 111184 | 10:00–18:00 | 10:00–18:00 | |

*restaurant* and *small restaurant*, were added. At last, a total of 161,172 POIs have been obtained.

## 3. Determination and classification of urban commercial buildings

### 3.1. Determination of urban commercial buildings

#### 3.1.1. Assignment of POIs to building footprints

The 161,172 POIs obtained need to be assigned to the building footprints to determine urban commercial buildings. The assignment is conducted in ArcMap, a GIS program, via the "Spatial Join" function. As demonstrated in Fig. 4, all POIs were combined with the whole building footprints of Nanjing in ArcMap, through the function of "spatial join".

The number of each kind of POIs on the building footprints is recorded. The results of six buildings are shown in Table 1 as an example. In total, 5303 buildings contain at least one POI that falls into the six categories (*shopping mall*, *supermarket*, *electrical appliance*, *building*

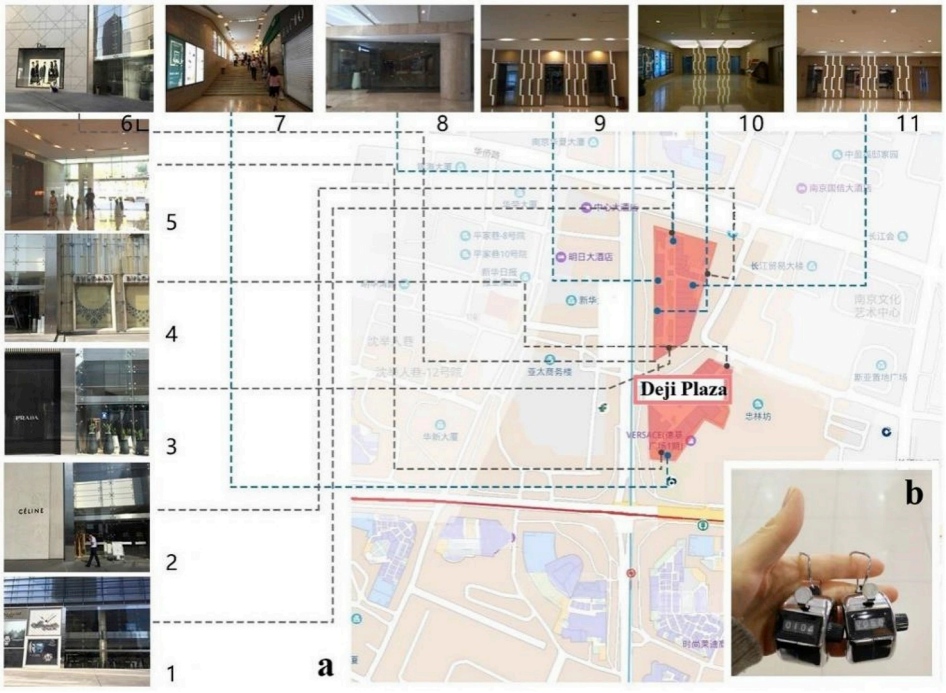



**Fig. 8.** The sample building: Deji Plaza (a: the 11 entrances of the building; b: the picture of hand-held counters).

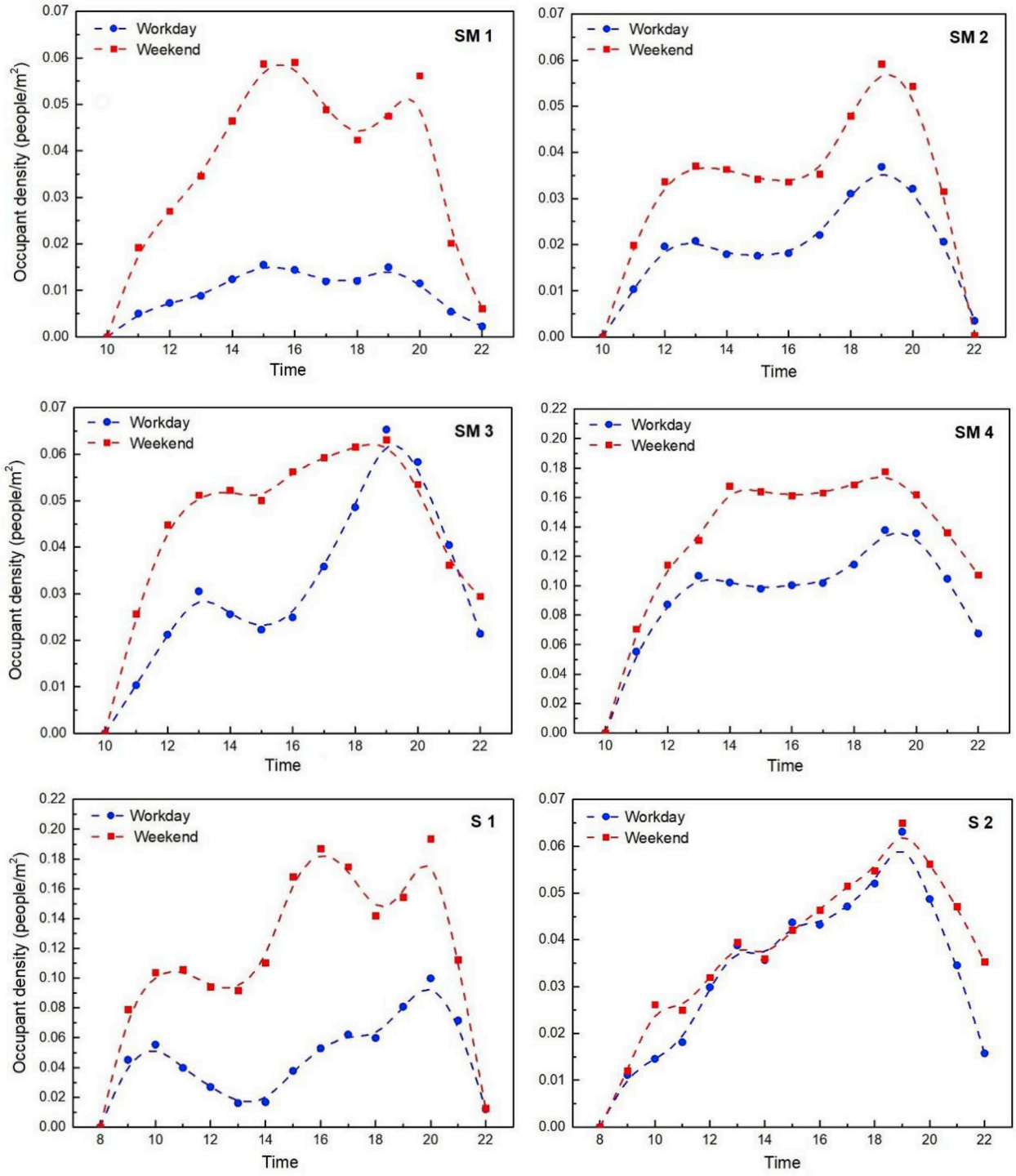



**Fig. 9.** Hourly occupant density of the 12 sample commercial buildings.

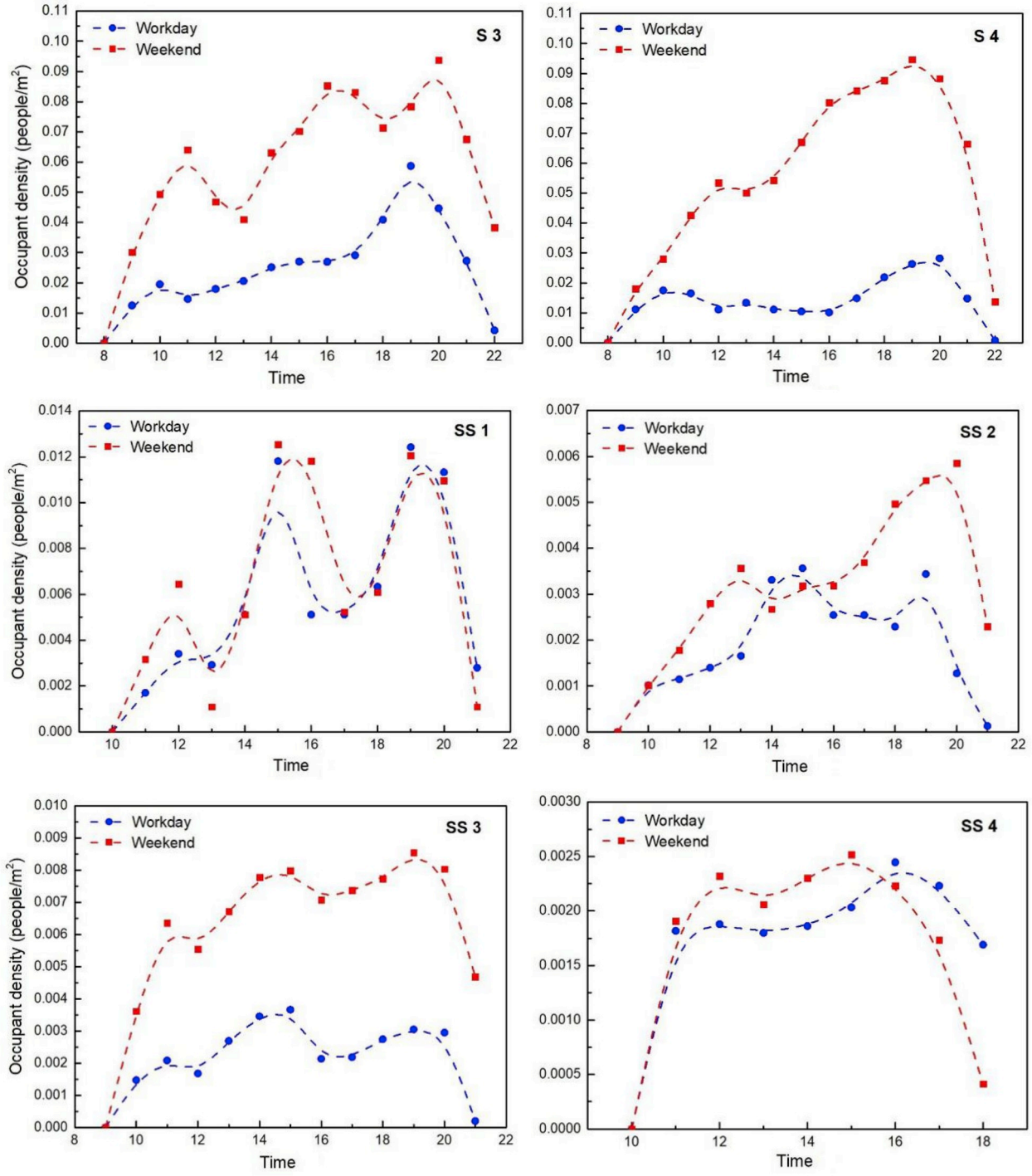



**Fig. 9.** (*continued*).

*material*, *large furniture*, and *small furniture*) and thus, are potential candidates for being determined to be commercial buildings.

### 3.1.2. Screening of urban commercial building candidates

Among the 5303 buildings considered to be potential candidates for commercial buildings, further screening was conducted according to two criteria. First, the footprint of the building needs to be larger than 500 $m^2$. Second, the building cannot be located within enclosed residential areas. The first criterion eliminates those small buildings such as convenient stores. The second criterion excludes those buildings with commercial POIs whose locations are within enclosed residential areas. The screening by the first criterion eliminated 1716 buildings and the second criterion 1312 buildings. In the end, 2275 buildings passed screenings and were determined to be urban commercial buildings.

## 3.2. Classification of urban commercial buildings

After the determination of urban commercial buildings, the next step is to classify these buildings into three sub-categories, namely shopping malls, supermarkets, and specialty stores.

### 3.2.1. Classifier and the weights of POIs

The logistic regression, which is based on function Sigmoid, was used to conduct classification. The value in function Sigmoid is always

**Table 5**
The coefficients of the occupant density model.

| Notation | $a_1$ | $b_1$ | $c_1$ | $a_2$ | $b_2$ | $c_2$ | $a_3$ | $b_3$ | $c_3$ | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| SM1 | 0.0147 | 15.24 | 12.80 | 0.0105 | 19.50 | 2.51 | / | / | / | 0.966 |
| SM1★ | 0.0581 | 15.63 | 15.19 | 0.0403 | 19.86 | 1.06 | / | / | / | 0.976 |
| SM2 | 0.0215 | 13.00 | 5.63 | 0.0366 | 18.85 | 6.89 | / | / | / | 0.941 |
| SM2★ | 0.0399 | 13.92 | 10.42 | 0.0591 | 19.24 | 4.05 | / | / | / | 0.933 |
| SM3 | 0.0291 | 13.30 | 4.95 | 0.0632 | 19.20 | 7.46 | / | / | / | 0.959 |
| SM3★ | 0.0429 | 13.00 | 5.05 | 0.0638 | 18.29 | 15.59 | / | / | / | 0.989 |
| SM4 | 0.1039 | 13.33 | 6.66 | 0.1363 | 19.24 | 11.69 | / | / | / | 0.963 |
| SM4★ | 0.1149 | 13.40 | 6.72 | 0.1733 | 18.84 | 20.46 | / | / | / | 0.995 |
| S1 | 0.0564 | 10.14 | 2.99 | 0.0609 | 17.09 | 8.72 | 0.0815 | 20.19 | 1.85 | 0.964 |
| S1★ | 0.1085 | 10.47 | 3.79 | 0.1880 | 16.17 | 9.66 | 0.1555 | 20.10 | 1.53 | 0.977 |
| S2 | 0.0033 | 10.42 | 4.92 | 0.0406 | 14.59 | 15.94 | 0.0484 | 19.36 | 5.75 | 0.977 |
| S2★ | 0.0173 | 10.89 | 5.16 | 0.0361 | 15.00 | 14.41 | 0.0522 | 19.81 | 10.18 | 0.972 |
| S3 | 0.0144 | 9.76 | 1.11 | 0.0268 | 14.82 | 15.84 | 0.0490 | 19.35 | 3.29 | 0.985 |
| S3★ | 0.0567 | 10.59 | 2.90 | 0.0839 | 16.09 | 11.12 | 0.0712 | 20.36 | 3.29 | 0.975 |
| S4 | 0.0142 | 10.10 | 1.91 | 0.0121 | 13.91 | 14.37 | 0.0279 | 19.34 | 3.87 | 0.945 |
| S4★ | 0.0457 | 11.72 | 6.17 | 0.0842 | 16.85 | 9.13 | 0.0646 | 20.07 | 2.67 | 0.986 |
| SS1 | 0.0036 | 12.15 | 1.61 | 0.0114 | 15.03 | 1.29 | 0.0128 | 19.22 | 2.79 | 0.943 |
| SS1★ | 0.0050 | 11.88 | 1.96 | 0.0137 | 15.47 | 1.97 | 0.0128 | 19.30 | 2.02 | 0.947 |
| SS2 | 0.0010 | 10.45 | 0.88 | 0.0034 | 15.00 | 8.14 | 0.0029 | 18.98 | 1.11 | 0.957 |
| SS2★ | 0.0026 | 12.22 | 3.71 | 0.0031 | 16.00 | 7.78 | 0.0054 | 19.45 | 3.24 | 0.962 |
| SS3 | 0.0021 | 10.53 | 0.54 | 0.0036 | 14.36 | 6.78 | 0.0033 | 19.06 | 2.65 | 0.927 |
| SS3★ | 0.0054 | 11.33 | 3.52 | 0.0077 | 14.83 | 5.21 | 0.0086 | 19.23 | 5.76 | 0.949 |
| SS4 | 0.0018 | 11.38 | 6.16 | 0.0024 | 16.33 | 8.46 | / | / | / | 0.965 |
| SS4★ | 0.0022 | 11.84 | 5.89 | 0.0023 | 15.63 | 4.17 | / | / | / | 0.957 |

Without ★: Occupant density models on workday; With ★: Occupant density models at weekend.

between 0 and 1, and the graph is shown in Fig. 5.

With the setting of threshold value $k$, a step function could be established:

$$Type = \begin{cases} 1, & y \geq k \\ 0, & y < k \end{cases} \quad (1)$$

Generally, k is set at 0.5 and the events can be categorized into two types: type 1 (y $\geq$ 0.5, Type $\rightarrow$ 1) and type 2 (y $<$ 0.5, Type $\rightarrow$ 0). However, if the determination of type 1 needs be stricter, k can be set at a larger value. A threshold value of 0.8 is chosen in this study.

To apply logistic regression into the classification of commercial buildings, those POIs, which are linked with building footprints, were used to establish the prediction function $h_\theta(x)$:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \quad (2)$$

$$\theta^T x = \sum_{i=0}^{8} \theta_i x_i = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_8 x_8 \quad (3)$$

where $x_i$ is the number of each kind of POI ($x_0 = 1$), $\theta_i$ is the weight of POI $x_i$ ($\theta_0$ = constant term).

If $h_\theta(x) \geq 0.8$, the target commercial building would be classified into type 1, otherwise, type 2. As three kinds of commercial buildings need to be classified, type 2 would be explained as the building that does not belong to type 1. For example, prediction function $h_\theta(x)$ is used to judge shopping malls. If $h_\theta(x) \geq 0.8$, the target building would be considered as the shopping mall. Otherwise, it does not belong to shopping malls. To be noted here, only when $h_\theta(x) \geq 0.8$, could the target building be judged as the named type. On this occasion, 3 types of prediction functions should be established in order to judge shopping malls, supermarkets and specialty stores respectively. We define each prediction function as a "classifier".

To obtain the weight of each kind of POIs, 80 training samples, including 26 shopping malls, 22 supermarkets and 32 specialty stores, were used to conduct machine learning. The results are summarized in Table 2.

#### 3.2.2. Validation of the classification approach

Among 3 classifiers, the one with the largest value determines the building type. The success rate of training samples is 79 out of 80 (98.75%). In order to further validate the proposed classification approach, 20 commercial buildings with their actual sub-categories known were selected as the predicting samples. Their calculated values and predicted sub-categories are summarized in Table 3. As shown in the table, the proposed classification approach successfully predicts all the samples. The success rate is 100%, which is considered adequate to be used to classify the 2275 commercial buildings identified previously.

#### 3.2.3. Classification results

By applying the approach presented above, the classifiers of the 2275 commercial buildings were calculated in ArcMap. The results are 219 shopping malls, 341 supermarkets, and 1715 specialty stores. Their distribution in the city of Nanjing is shown in Fig. 6.

## 4. Building-level occupant density model

### 4.1. Sample commercial buildings

A total of 12 sample commercial buildings were selected, 4 in each sub-category. They are located in 6 different districts of Nanjing and thus geographically diversified (Fig. 7). Shopping malls usually open from 10 a.m. to 10 p.m. and supermarkets from 8 a.m. to 10 p.m., on both workdays and weekends. The open hours for specialty stores are not consistent. Table 4 summarizes the floor areas, the open hours on workdays and weekends for the 12 sample commercial buildings.

### 4.2. Occupant density measurement

In recent years, numerous sensing techniques have been proposed for occupancy detection, i.e., using Passive Infrared Detection (PIR) sensor, camera, $CO_2$ sensor and Wi-Fi [49]. However, nearly no studies discussed their application in commercial buildings, perhaps because of the difficulties in obtaining building occupancy data. In fact, given some commercial buildings are large and complex with many entrances, as well as their privacy, it is hard to use automated methods. Instead, the manual measurement is chosen in this paper.

The number of occupants in each of the 12 sample commercial buildings was measured by recording the number of people in and out of

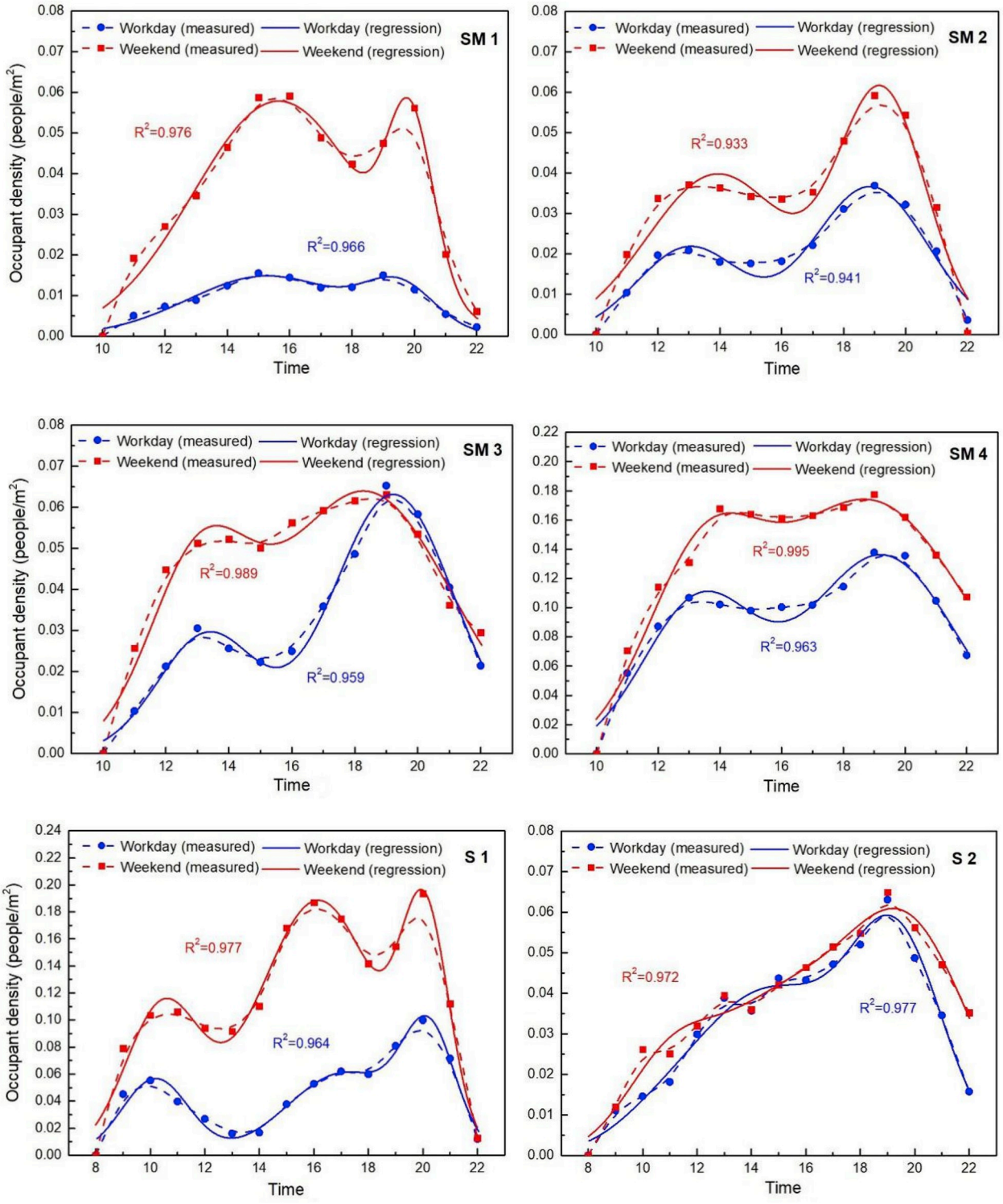



**Fig. 10.** The measured occupant density and their fitting models.

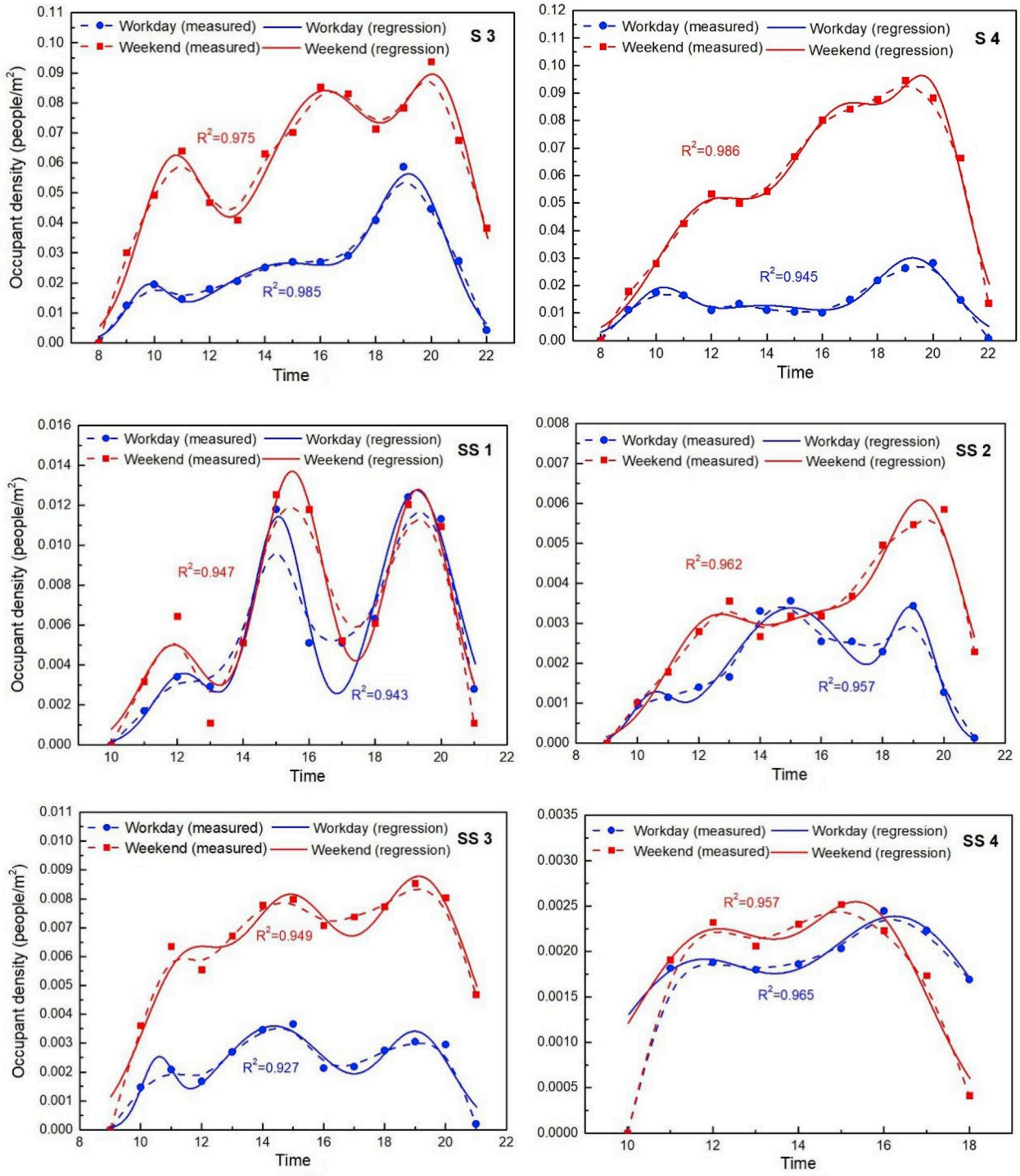



**Fig. 10.** (*continued*).

the building and calculated using Eq. (4).

$$c_i = c_{i-1} + \sum_{j=1}^{k} (m_{ij} - n_{ij}) \tag{4}$$

where $c_i$ is the number of occupants at $i$ o'clock; $m_{ij}$ is the number of people entering the building between $i$-1 o'clock and $i$ o'clock at the $j$-$th$ entrance; $n_{ij}$ is the number of people leaving the building between $i$-1 o'clock and $i$ o'clock at the $j$-$th$ entrance; $k$ is the number of entrances of the building.

All of the 12 sample buildings have more than one entrance. One person (if complicated, two people) was assigned to record the number of people entering and leaving the building in its entrances. Hand-held counters were used to facilitate the number recording. The measurement for each sample commercial building was conducted twice, one on a weekday and the other on a weekend. The occupant density was obtained by dividing the number of occupants by the total floor area. Fig. 8 illustrates one of the sample buildings, its 11 entrances, and the hand-held counters.

The measurement of the 12 sample commercial buildings was completed within 3 months. Their hourly occupant density patterns are illustrated in Fig. 9.

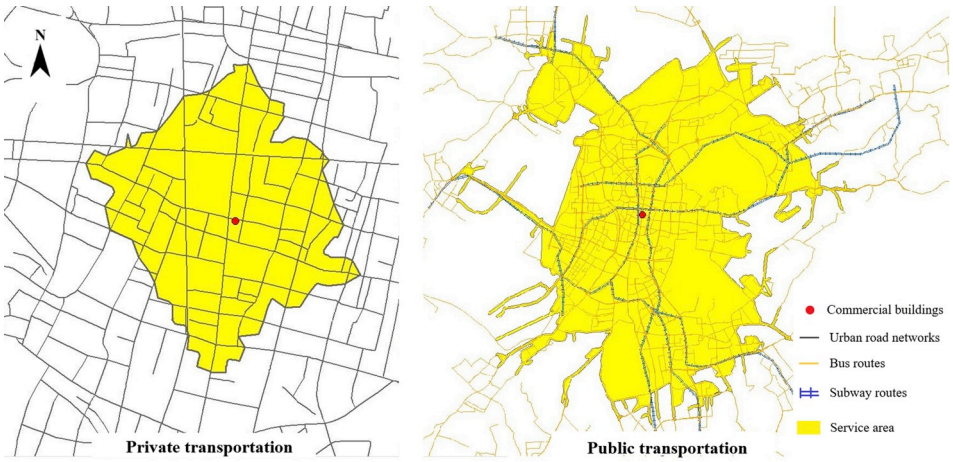



**Fig. 11.** The example of the 30-min service area.

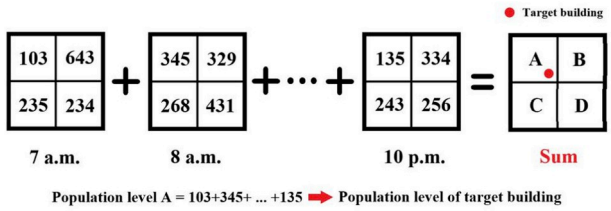



**Fig. 12.** The superposition of BHPMs.

**Table 6**
The urban parameters of the 12 sample commercial buildings.

| Notation | Transportation accessibility ($m^2$) | Population level (Workday) | Population level (Weekend) | Type |
|---|---|---|---|---|
| SM1 | 659151120.76 | 826 | 861 | Shopping mall |
| SM2 | 786484627.11 | 780 | 780 | |
| SM3 | 309047866.15 | 855 | 923 | |
| SM4 | 782976049.28 | 1699 | 1915 | |
| S1 | 776146560.61 | 1675 | 1712 | Supermarket |
| S2 | 317473092.70 | 780 | 780 | |
| S3 | 769477432.54 | 1150 | 1058 | |
| S4 | 746248429.67 | 809 | 853 | |
| SS1 | 1088875428.14 | 790 | 798 | Specialty store |
| SS2 | 624136111.51 | 870 | 891 | |
| SS3 | 778201163.34 | 1352 | 1335 | |
| SS4 | 649739151.71 | 789 | 780 | |

### 4.3. Occupant density model development

An examination of Fig. 9 indicates that the pattern of hourly occupant density for commercial buildings is multi-model and multi-peak. Therefore, it is plausible that the curves shown in Fig. 9 might be fitted by superimposing multiple normal distribution functions.

The occupant density model of shopping malls (SM 1–4 in Fig. 9) can be fitted by superimposing two normal distribution functions. And the occupant density model of supermarkets (S 1–4 in Fig. 9) is the superposition of 3 normal distribution functions. Their occupant density models can be established as Eq. (5) and Eq. (6) respectively.

$$D_b = a_1 \cdot e^{-\frac{(T-b_1)^2}{c_1}} + a_2 \cdot e^{-\frac{(T-b_2)^2}{c_2}} \tag{5}$$

$$D_b = a_1 \cdot e^{-\frac{(T-b_1)^2}{c_1}} + a_2 \cdot e^{-\frac{(T-b_2)^2}{c_2}} + a_3 \cdot e^{-\frac{(T-b_3)^2}{c_3}} \tag{6}$$

where $D_b$ is the building-level occupant density; $T$ is the time; $a_1$, $a_2$, $a_3$ is the first, second, third peak occupant density; $b_1$, $b_2$, $b_3$ is the time, in hours and counted from the opening time, required to reach the first, second, third peak occupant density; $c_1$, $c_2$, $c_3$ is a coefficient which represents the speed to reach peak occupant density; The smaller $c_1$, $c_2$, $c_3$ is, the larger the speed will be.

The occupant density model of specialty stores (SS 1–4 in Fig. 9) is more complicated. If the specialty store closes before 7 p.m., the model is the superposition of 2 normal distribution functions. Otherwise, the model is the superposition of 3 normal distribution functions. Therefore, the coefficient $p$ is used, and the relationship between occupant density $D_b$ and the time $T$ is:

$$D_b = a_1 \cdot e^{-\frac{(T-b_1)^2}{c_1}} + a_2 \cdot e^{-\frac{(T-b_2)^2}{c_2}} + p \cdot a_3 \cdot e^{-\frac{(T-b_3)^2}{c_3}} \tag{7}$$

$$p = \begin{cases} 0, & \textit{store closes before 7 pm} \\ 1, & \textit{store closes after 7 pm} \end{cases} \tag{8}$$

The curve fitting process is conducted in SPSS (Statistical Product and Service Solutions) using restricted non-linear regression. The model coefficients in Eqs. (5)–(7) are summarized in Table 5. The coefficient of determination, $R^2$, in each case, is larger than 0.9, suggesting generally good curve fit. The measured occupant density and the fitting model are illustrated in Fig. 10. It is evident that they match each other fairly well.

## 5. Urban-level occupant density model

In Section 4, the building-level occupant density models for the 12 sample commercial buildings were developed using measured occupancy data and regression analysis. This approach clearly cannot be expanded to all of the 2275 commercial buildings in the city of Nanjing because the time and effort required is simply too much to bear. This section presents an urban-level occupant density model, in which the coefficients defining the building-level occupant density models in Eqs. (5)–(7) are linked to the key urban parameters of the commercial buildings. These key urban parameters can be obtained using big data technology, enabling a fast calculation of all the coefficients for the 2275 commercial buildings.

### 5.1. Key urban parameters

The selection of key urban parameters needs to meet two criteria: (1) these parameters significantly affect the building-level occupant density model, (2) these parameters can be obtained for most, if not all, commercial buildings. Two urban parameters are selected, namely transportation accessibility *TA* and population level *PL*.

**Table 7**
The results of correlation analysis between coefficients and urban parameters.

| Type | Coefficient | *TA* | | *PL* | | *TA*★ | | *PL*★ | |
|---|---|---|---|---|---|---|---|---|---|
| | | C | S | C | S | C | S | C | S |
| SM | $a_1$ | 0.343 | 0.328 | 0.992** | 0.004 | 0.451 | 0.274 | 0.974** | 0.013 |
| | $b_1$ | 0.011 | 0.495 | −0.235 | 0.382 | 0.340 | 0.330 | −0.365 | 0.318 |
| | $c_1$ | 0.227 | 0.773 | −0.154 | 0.846 | 0.434 | 0.566 | −0.433 | 0.567 |
| | $a_2$ | 0.151 | 0.425 | 0.926** | 0.037 | 0.368 | 0.316 | 0.982** | 0.009 |
| | $b_2$ | −0.211 | 0.395 | 0.154 | 0.423 | 0.593 | 0.204 | −0.278 | 0.361 |
| | $c_2$ | 0.165 | 0.835 | 0.805 | 0.195 | −0.224 | 0.776 | 0.782 | 0.218 |
| S | $a_1$ | −0.083 | 0.458 | 0.920** | 0.040 | 0.726 | 0.137 | 0.970** | 0.015 |
| | $b_1$ | −0.783 | 0.108 | −0.287 | 0.357 | −0.020 | 0.490 | −0.626 | 0.187 |
| | $c_1$ | −0.874 | 0.126 | −0.190 | 0.810 | −0.351 | 0.649 | −0.537 | 0.463 |
| | $a_2$ | −0.128 | 0.436 | 0.946** | 0.027 | 0.673 | 0.164 | 0.967** | 0.016 |
| | $b_2$ | 0.293 | 0.354 | 0.750 | 0.215 | 0.870 | 0.065 | 0.218 | 0.391 |
| | $c_2$ | −0.461 | 0.539 | −0.862 | 0.138 | −0.922 | 0.078 | −0.482 | 0.518 |
| | $a_3$ | 0.150 | 0.425 | 0.902** | 0.049 | 0.512 | 0.244 | 0.989** | 0.005 |
| | $b_3$ | 0.350 | 0.325 | 0.813 | 0.073 | 0.826 | 0.087 | 0.317 | 0.342 |
| | $c_3$ | −0.876 | 0.124 | −0.886 | 0.114 | −0.885 | 0.075 | −0.630 | 0.370 |
| SS | $a_1$ | 0.972** | 0.014 | −0.105 | 0.448 | 0.940** | 0.013 | 0.596 | 0.202 |
| | $b_1$ | 0.761 | 0.120 | −0.599 | 0.201 | −0.199 | 0.400 | −0.708 | 0.108 |
| | $c_1$ | −0.294 | 0.706 | −0.521 | 0.479 | −0.798 | 0.202 | −0.141 | 0.859 |
| | $a_2$ | 0.961** | 0.020 | −0.305 | 0.347 | 0.983** | 0.009 | 0.067 | 0.467 |
| | $b_2$ | −0.313 | 0.343 | −0.712 | 0.144 | −0.323 | 0.339 | −0.800 | 0.100 |
| | $c_2$ | −0.891 | 0.059 | 0.185 | 0.815 | −0.814 | 0.186 | 0.275 | 0.725 |
| | $a_3$ | 0.955** | 0.046 | −0.585 | 0.301 | 0.993** | 0.036 | −0.238 | 0.423 |
| | $b_3$ | −0.714 | 0.133 | −0.313 | 0.399 | −0.520 | 0.326 | −0.620 | 0.287 |
| | $c_3$ | 0.800 | 0.410 | 0.315 | 0.796 | −0.495 | 0.670 | 0.987 | 0.104 |

SM: Shopping Mall, S: Supermarket, SS: Specialty Store; Without ★: Analysis on workday, With ★: Analysis at weekend; C; Correlation, S: Significance; With **: The correlation is significant, as the significance is no more than 0.05.

**Table 8**
The correlation of the relationships between coefficient c and a under 5 models.

| Type | Relationship | $R^2$ | | | | |
|---|---|---|---|---|---|---|
| | | L | QC | CC | LC | EC |
| SM | $c_1$, $a_1$ | 0.079 | 0.835 | 0.835 | 0.041 | 0.041 |
| | $c_2$, $a_2$ | 0.912 | 0.956 | 0.956 | 0.746 | 0.746 |
| SM★ | $c_1$, $a_1$ | 0.048 | 0.601 | 0.632 | 0.033 | 0.033 |
| | $c_2$, $a_2$ | 0.662 | 0.815 | 0.815 | 0.508 | 0.508 |
| S | $c_1$, $a_1$ | 0.278 | 0.953 | 0.953 | 0.388 | 0.388 |
| | $c_2$, $a_2$ | 0.494 | 0.983 | 0.983 | 0.537 | 0.537 |
| | $c_3$, $a_3$ | 0.362 | 0.615 | 0.613 | 0.532 | 0.532 |
| S★ | $c_1$, $a_1$ | 0.240 | 0.263 | 0.275 | 0.214 | 0.214 |
| | $c_2$, $a_2$ | 0.455 | 0.885 | 0.885 | 0.441 | 0.441 |
| | $c_3$, $a_3$ | 0.383 | 0.898 | 0.898 | 0.599 | 0.599 |
| SS | $c_1$, $a_1$ | 0.009 | 0.128 | 0.105 | 0.003 | 0.003 |
| | $c_2$, $a_2$ | 0.978 | 0.979 | 0.979 | 0.992 | 0.992 |
| | $c_3$, $a_3$ | 0.343 | 1.000 | 1.000 | 0.321 | 0.321 |
| SS★ | $c_1$, $a_1$ | 0.577 | 0.990 | 0.990 | 0.540 | 0.540 |
| | $c_2$, $a_2$ | 0.521 | 0.649 | 0.630 | 0.644 | 0.644 |
| | $c_3$, $a_3$ | 0.154 | 1.000 | 1.000 | 0.269 | 0.269 |

SM: Shopping Mall, S: Supermarket, SS: Specialty Store; Without ★: Analysis on workday, With ★: Analysis at weekend; L: Linear, QC: Quadratic Curve, CC: Cubic Curve, LC: Logarithmic Curve, EC: Exponential Curve.

**Table 9**
Urban-level occupant density models of shopping mall.

| | Workday | Weekend |
|---|---|---|
| Function | $D_u = a_1 \cdot e^{-\frac{(T-b_1)^2}{c_1}} + a_2 \cdot e^{-\frac{(T-b_2)^2}{c_2}}$ | |
| Diagram |  | |
| $a_1$ | $-0.05 + 9.34 \times 10^{-5} \cdot PL$ | $-7.36 \times 10^{-3} + 6.37 \times 10^{-5} \cdot PL$ |
| $b_1$ | 13.72 | 13.99 |
| $c_1$ | $22.69 - 838.30 \cdot a_1 + 6583.23 \cdot a_1^2$ | $-24.28 + 1070.57 \cdot a_1 - 6964.35 \cdot a_1^2$ |
| $a_2$ | $-0.06 + 1.14 \times 10^{-4} \cdot PL$ | $-0.04 + 1.11 \times 10^{-4} \cdot PL$ |
| $b_2$ | 19.20 | 19.06 |
| $c_2$ | $1.61 + 130.48 \cdot a_2 - 419.05 \cdot {a_2}^2$ | $-26.29 + 773.34 \cdot a_2 - 2905.22 \cdot a_2^2$ |

#### 5.1.1. Transportation accessibility

It is widely agreed that transportation accessibility has a significant impact on the occupant density of buildings. To access a commercial building, two transportation modes are available, namely private transportation (private cars) and public transportation (buses and subways).

To evaluate the transportation accessibility, the "service area" was used. The service area is defined as a region that can be reached from one point through all accessible routes within certain time. For instance, the 5-min service area for a point includes all the routes that can be reached within 5 min from that point. Thus, the service area of one commercial building can be determined by two parts: (1) the service area based on urban road networks, which is used for private transportation, and (2) the service area based on bus routes and subway routes. In addition, the driving time and the speed of the vehicles should be determined before calculating the service area. In this paper, the speed of private cars and buses is both defined as 500 m/min [50], and the speed of subways is based on Nanjing metro schedule. The driving time is defined as 30 min, which means that the 30-min service area is calculated here.

The information of urban road networks, bus routes and subway routes was obtained from A map, a product of Alibaba Group, by APIs. All the information was integrated in the software ArcMap, using the function of "service area analysis" to calculate the service area under private transportation and public transportation respectively. The example of the results is demonstrated in Fig. 11.

Finally, assuming the willingness for travel by private transportation is equal to public transportation, the transportation accessibility of each sample *TA* can be calculated as:

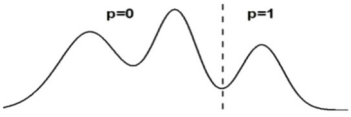
$$TA = 0.5 \cdot S_{pri} + 0.5 \cdot S_{pub} \tag{9}$$

where $S_{pri}$, $S_{pub}$ is the service area by private and public transportation respectively ($t = 30$ min).

**Table 10**
Urban-level occupant density models of supermarket.

| | Workday | Weekend |
|---|---|---|
| Function | $D_u = a_1 \cdot e^{-\frac{(T-b_1)^2}{c_1}} + a_2 \cdot e^{-\frac{(T-b_2)^2}{c_2}} + a_3 \cdot e^{-\frac{(T-b_3)^2}{c_3}}$ | |
| Diagram | | |
| $a_1$ | $-8.65 \times 10^{-3} + 3.46 \times 10^{-5} \cdot PL$ | $-0.04 + 8.71 \times 10^{-5} \cdot PL$ |
| $b_1$ | 10.11 | 10.92 |
| $c_1$ | $-4.00 + 474.34 \cdot a_1 - 6209.39 \cdot a_1^2$ | $6.19 - 44.58 \cdot a_1 + 198.53 \cdot a_1^2$ |
| $a_2$ | $-5.88 \times 10^{-3} + 3.71 \times 10^{-5} \cdot PL$ | $-0.06 + 1.46 \times 10^{-4} \cdot PL$ |
| $b_2$ | 15.10 | 16.03 |
| $c_2$ | $9.88 + 447.69 \cdot a_2 - 7637.82 \cdot a_2^2$ | $19.33 - 156.39 \cdot a_2 + 558.23 \cdot a_2^2$ |
| $a_3$ | $-1.30 \times 10^{-3} + 4.80 \times 10^{-5} \cdot PL$ | $-0.03 + 1.10 \times 10^{-4} \cdot PL$ |
| $b_3$ | 19.56 | 20.09 |
| $c_3$ | $0.21 + 189.96 \cdot a_3 - 2085.63 \cdot a_3^2$ | $44.60 - 868.12 \cdot a_3 + 3802.66 \cdot a_3^2$ |

**Table 11**
Urban-level occupant density models of specialty store.

| | Workday | Weekend |
|---|---|---|
| Function | $D_u = a_1 \cdot e^{-\frac{(T-b_1)^2}{c_1}} + a_2 \cdot e^{-\frac{(T-b_2)^2}{c_2}} + p \cdot a_3 \cdot e^{-\frac{(T-b_3)^2}{c_3}}$ | |
| Diagram |   | |
| $p$ | $p = 0$, store closes before 7 p.m.; $p = 1$, store close after 7 p.m. | |
| $a_1$ | $-1.73 \times 10^{-3} + 4.89 \times 10^{-12} \cdot TA$ | $-6.64 \times 10^{-4} + 5.68 \times 10^{-12} \cdot TA$ |
| $b_1$ | 11.13 | 11.64 |
| $c_1$ | $-44.67 + 66339.25 \cdot a_1 - 21280348.73 \cdot a_1^2$ | $26.45 - 12960.47 \cdot a_1 + 1614899.50 \cdot a_1^2$ |
| $a_2$ | $-9.60 \times 10^{-3} + 1.88 \times 10^{-11} \cdot TA$ | $-0.01 + 2.42 \times 10^{-11} \cdot TA$ |
| $b_2$ | 15.18 | 15.48 |
| $c_2$ | $11.03 - 1108.73 \cdot a_2 + 22330.31 \cdot a_2^2$ | $4.65 + 570.01 \cdot a_2 - 56227.61 \cdot a_2^2$ |
| $a_3$ | $-0.01 + 2.26 \times 10^{-11} \cdot TA$ | $-4.14 \times 10^{-3} + 1.57 \times 10^{-11} \cdot TA$ |
| $b_3$ | 19.09 | 19.33 |
| $c_3$ | $-16.27 + 7258.91 \cdot a_3 - 451422.27 \cdot a_3^2$ | $-11.16 + 3877.47 \cdot a_3 - 222118.65 \cdot a_3^2$ |

#### 5.1.2. Population level

The population in the vicinity of a commercial building is positively correlated with its occupant density and therefore, is selected as the second urban parameter. The population census data of Nanjing is available. However, it only covers the population number of the city and the city districts. Therefore, it is not precise enough to provide the population data in the vicinity of commercial buildings. Other data sources need to be considered, one of such being Baidu Population Heat Map (BPHM).

BPHM is developed using the mobile phone location technique. It provides grid data of the population level. For the city of Nanjing, the grid size is 1.34 × 1.34 km and there are a total of 1461 grids in the city. It should be noted that the BPHM grid data is not the actual number of population in the gird. It is rather an indicator of the population level in the grid, which is positively correlated with the actual number of population. In addition to being fine grid data, the BPHM has another important advantage, that is, it is updated hourly, which makes it a suitable data source for developing dynamic occupant density models for commercial buildings. Another advantage of the BPHM data worth mentioning is that it covers most of the major cities in China.

As the population level in each grid changes hourly and their difference may be huge during a day, it is unreasonable to regard any of them as the population level of the grid. The sum of the population level during open hours was used to represent the population level of the grid. Considering the open hour of commercial buildings is mainly between 7 a.m. and 10 p.m., 16 BHPMs have been superimposed, in order to calculate the sum of population level in each grid. As each of the 2275 commercial buildings is located in a BPHM grid, the population level of a commercial building will follow the sum of population level in its subordinated grid.

All the BHPMs were obtained through Baidu APIs, and the superposition of BHPMs was conducted in the ArcMap, via the function of "raster calculator". The superposition of BHPMs is demonstrated in Fig. 12, and this process has been conducted twice in order to generate two kinds of population level, one for workday and the other for weekend.

#### 5.1.3. Urban parameters of sample commercial buildings

After data acquisition, cleaning and processing, the urban parameters of the 12 sample commercial buildings are summarized in Table 6.

### 5.2. Urban-level occupant density model development

To establish urban-level occupant density $D_u$, urban parameters *TA, PL* were used to replace the coefficients $a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3, c_3$. This process can be described as:

$$D_b \sim f(T,\ a_1,\ b_1,\ c_1,\ a_2,\ b_2,\ c_2,\ a_3,\ b_3,\ c_3) \rightarrow D_u \sim f(T,\ TA,\ PL) \quad (10)$$

The correlation analysis was conducted firstly in the software SPSS, and the results are summarized in Table 7.

It is found that the coefficients $a_1$, $a_2$, $a_3$, which represent peak occupant density, have strong correlations with the urban parameter *TA* or *PL*, while the coefficients $b_1$, $b_2$, $b_3$ and $c_1$, $c_2$, $c_3$ have no significant correlation with the urban parameters *TA* and *PL*. Thus, the linear relationship between $a_1$, $a_2$, $a_3$ and *TA, PL* could be established. The mean value$b_1$,$b_2$,$b_3$ were calculated to represent the coefficients $b_1$, $b_2$, $b_3$ in occupant density models. As the coefficients $c_1$, $c_2$, $c_3$, which represent the speed to reach peak occupant density, only have mathematical meaning, the relationships between $c_1$ and $a_1$, $c_2$ and $a_2$, $c_3$ and $a_3$ were attempted to establish, based on linear model, quadratic curve model, cubic curve model, logarithmic curve model and exponential curve model. The results are shown in Table 8.

It is found that the quadratic curve model and cubic curve model best describe the relationships between $c_1$ and $a_1$, $c_2$ and $a_2$, $c_3$ and $a_3$. Considering the simple structure of models between coefficient $c$ and $a$, the quadratic curve model was selected to describe their relationship.

The results of the final urban-level occupant density models of shopping mall, supermarket and specialty store are summarized in Tables 9–11.

### 5.3. Validation of the urban-level occupant density model

#### 5.3.1. Evaluation of model regression

Fig. 13 and Table 12 present the urban-level occupant density models of the 12 sample commercial buildings shown in Table 6 and their statistical parameters for curve fitting. As shown in Table 12, the mean $R^2$ values for the three sub-categories are 0.66, 0.5335, and 0.5829, respectively. The overall mean $R^2$ value for all 12 sample buildings is 0.5921 and in 75% cases $R^2$ is larger than 0.5, which indicates relatively good curve fit. If the threshold of $R^2$ is increased to 0.75 indicating a very good curve fit, there are still 42.7% of sample buildings exceeding it.

#### 5.3.2. The test of prediction

Extra two shopping malls, one supermarket (tested for 2 days) and two specialty stores were selected to further verify the urban-level occupant density model. Their key information is summarized in

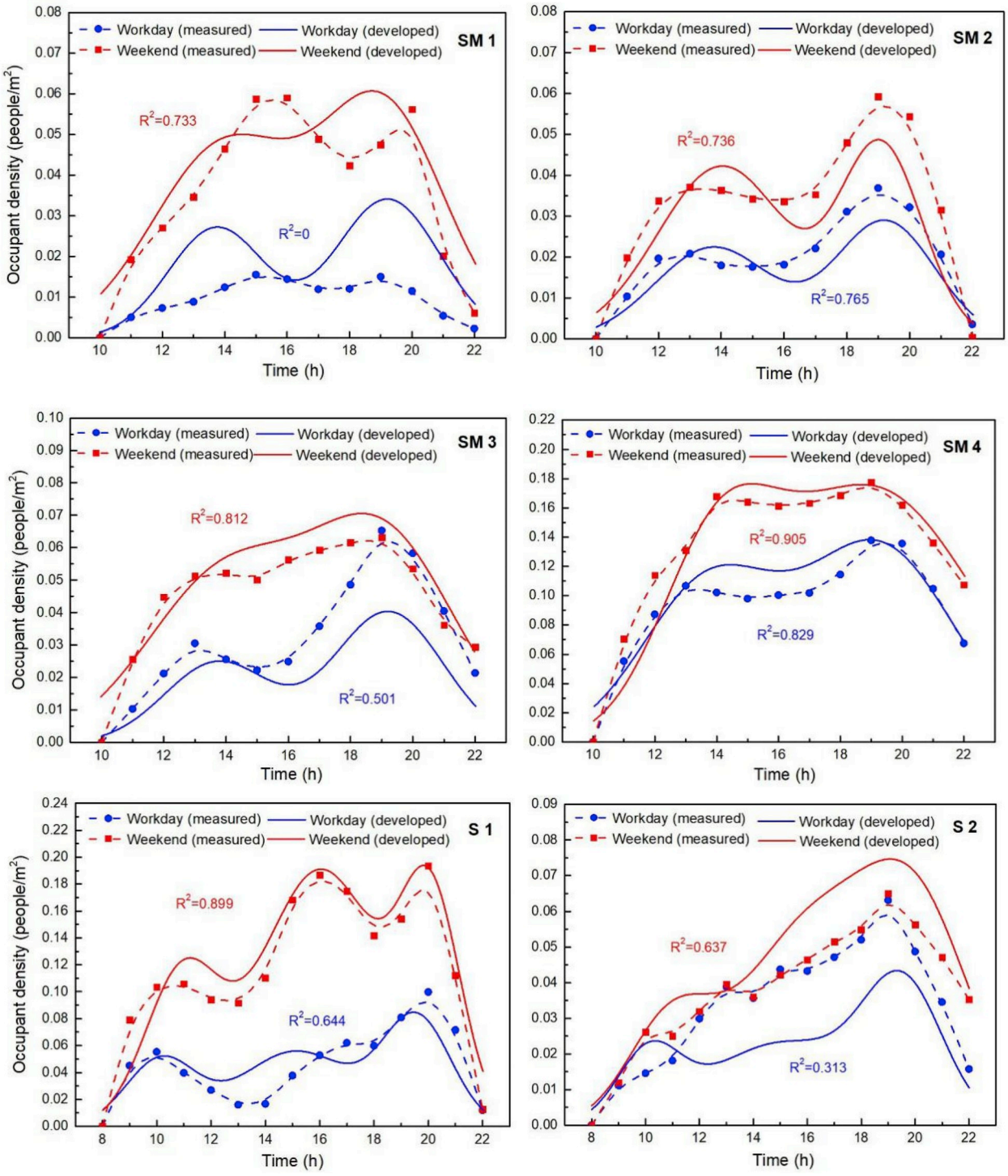



**Fig. 13.** The developed and measured occupant density models of 12 sample commercial buildings.

Table 13.

Using the method developed previously, their occupant density models are predicted as follows:

$$\textbf{ID}1:\ D_u = 0.0248\cdot e^{-\frac{(T-13.72)^2}{5.93}} + 0.0403\cdot e^{-\frac{(T-19.20)^2}{6.19}} \quad (11)$$

$$\textbf{ID}2:\ D_u = 0.0556\cdot e^{-\frac{(T-13.99)^2}{13.71}} + 0.0695\cdot e^{-\frac{(T-19.06)^2}{13.44}} \quad (12)$$

$$\textbf{ID}\ 3:\ D_u = 0.0196\cdot e^{-\frac{(T-10.11)^2}{2.91}} + 0.0244\cdot e^{-\frac{(T-15.10)^2}{16.26}} + 0.0378\cdot e^{-\frac{(T-19.56)^2}{4.41}} \quad (13)$$

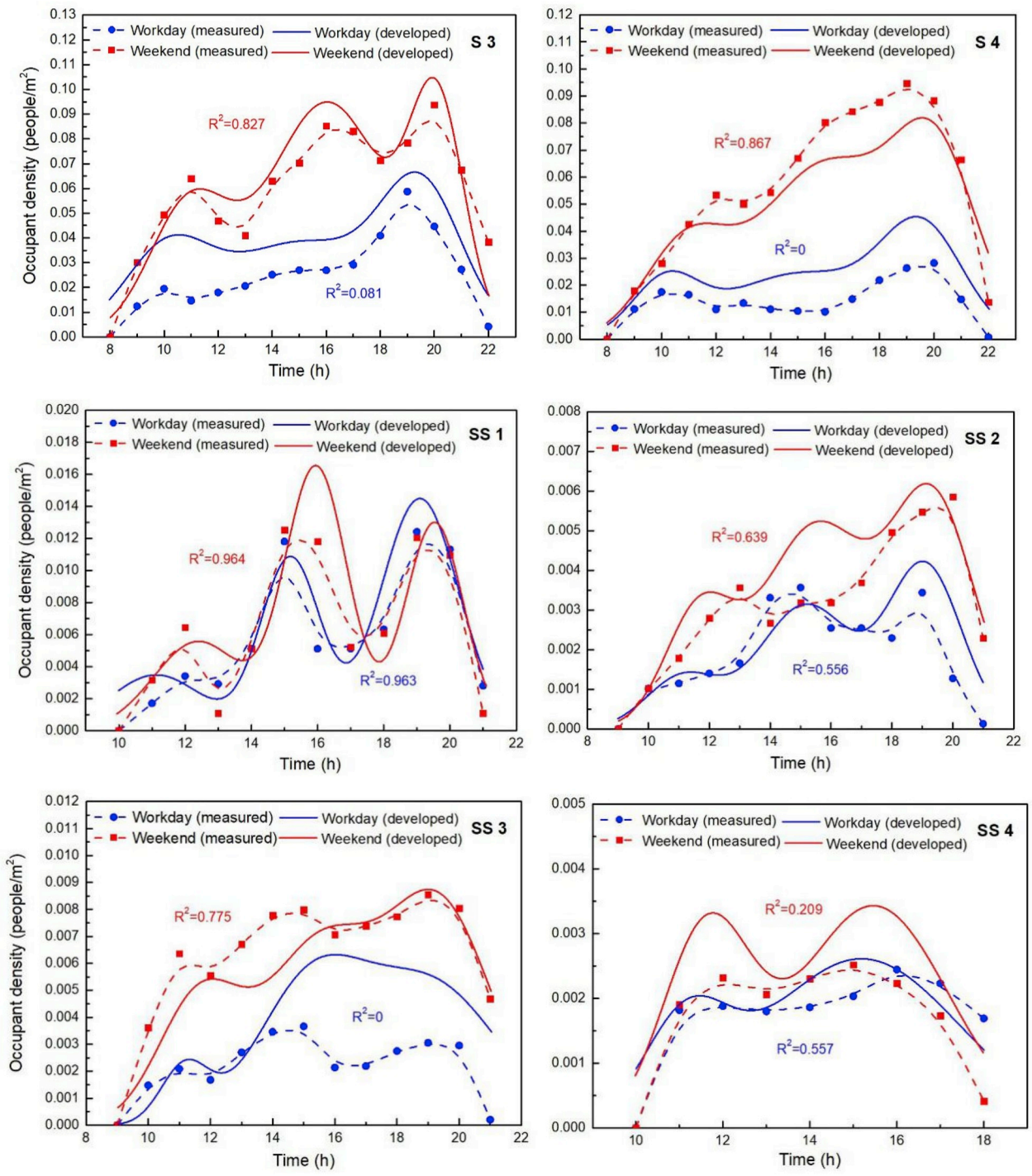



**Fig. 13.** (*continued*).

**Table 12**
Statistical parameters of the model regression for the sample commercial buildings.

| | MAE | RMSE | $R^2$ (mean) | $R^2 > 0.5$ (%) | $R^2 > 0.75$ (%) |
|---|---|---|---|---|---|
| Shopping mall | 0.0086 | 0.0110 | 0.6600 | 87.5% | 50% |
| Supermarket | 0.0111 | 0.0132 | 0.5335 | 62.5% | 37.5% |
| Specialty Store | 0.0010 | 0.0010 | 0.5829 | 75% | 37.5% |
| Overall | 0.0073 | 0.0103 | 0.5921 | 75% | 42.7% |

MAE: Mean Absolute Error, RMSE: Root Mean Squared Error.

$$\mathbf{ID}4: D_u = 0.0345 \cdot e^{-\frac{(T-10.92)^2}{4.88}} + 0.0601 \cdot e^{-\frac{(T-16.03)^2}{11.95}} + 0.0576 \cdot e^{-\frac{(T-20.09)^2}{7.19}} \quad (14)$$

$$\mathbf{ID}5: D_u = 0.0019 \cdot e^{-\frac{(T-11.13)^2}{3.80}} + 0.0045 \cdot e^{-\frac{(T-15.18)^2}{6.46}} + 0.0045 \cdot e^{-\frac{(T-19.09)^2}{7.35}} \quad (15)$$

$$\mathbf{ID}6: D_u = 0.0056 \cdot e^{-\frac{(T-11.64)^2}{4.51}} + 0.0144 \cdot e^{-\frac{(T-15.48)^2}{1.19}} \quad (16)$$

The comparison between the measured occupant density models and the predicted ones is illustrated in Fig. 14, and their coefficients $a_1$, $a_2$, $a_3$, $b_1$, $b_2$, $b_3$ are summarized in Table 14.

Shown in Fig. 14, the results are acceptable, as $R^2$ of 4 models (ID = 1, 2, 4, 5) are above 0.5. To be more specific, the absolute error and

**Table 13**
The information of the verification samples.

| ID | Floor area ($m^2$) | Transportation accessibility ($m^2$) | Population level | Type | Measurement day |
|---|---|---|---|---|---|
| 1 | 114327 | 693213373.32 | 853 | Shopping mall | Workday |
| 2 | 28320 | 673599415.03 | 988 | Shopping mall | Weekend |
| 3 | 6268 | 681566248.41 | 815 | Supermarket | Workday |
| 4 | 6268 | 681566248.41 | 842 | Supermarket | Weekend |
| 5 | 9424 | 750656742.91 | 1033 | Specialty store | Workday |
| 6 | 156235 | 1103222203.26 | 856 | Specialty store | Weekend |

relative error were used to describe the accuracy of prediction. These indicators were calculated in Eq. (17)–(19), and the results are shown in Table 15.

$$\Delta a_i = a_i^{'} - a_i \tag{17}$$

$$\delta a_i = |\Delta a_i| / a_i \tag{18}$$

$$\Delta b_i = b_i^{'} - b_i \tag{19}$$

where $\Delta a_i$ is the differences between the measured peak densities and the predicted ones; $\Delta a_i$ is the relative error between the measured peak densities and the predicted ones; $a_i$ is the measured peak densities; $a_i$' is the predicted peak densities; $\Delta b_i$ is the differences between their occurring time; $b_i$ is the measured occurring time; $b_i$' is the predicted occurring time.

It is clear that the relative error between the measured peak occupant density and the predicted one is small in shopping malls but larger in supermarkets and specialty stores. The curves of the measured and predicted models are basically consistent. The difference between the occurring time of peak occupant density is generally less than 1 h. Given the randomness of the verification samples, this result could be considered reasonable.

## 6. Discussion

### 6.1. Big data technique

In this paper, the acquisition of big data, such as building footprints and POIs for the classification, and urban parameters – transportation accessibility and population level for the establishment of occupant density models, is through APIs. The application for the keys of API is free, which gives the chance to apply the means in this paper into other projects. The websites of the application for keys are listed here: (1) Baidu map: http://lbsyun.baidu.com/; (2) A map: https://lbs.amap.com/.

However, the data acquisition still needs the programming techniques, and data processing challenges the skills of ArcMap use. Thus, to well reproduce the methods, it requires high professional skills of the researchers.

### 6.2. Improvements in the classification of building types

Logistic regression is used to classify urban commercial buildings. Although the method is able to quickly process a number of buildings, it has some room for further improvement. One difficulty is to handle the buildings with mixed functions such as commercial buildings with some office spaces. Since these buildings contain a large number of commercial POIs, they are categorized as commercial buildings. However, adding office POIs may alter the results and create new categories such as office-supermarket and office-shopping mall. In fact, these mixed-use buildings can be described by "shoeboxer" models [51]. The occupant density model developed in this paper would be linked with their commercial part while the non-commercial parts rely on other occupancy models.

### 6.3. Applicability to other building types

Although commercial buildings are the target of this study, the general methodology and framework, shown in Fig. 1, are applicable to other building types. For example, it is reported that the building-level occupant density model of office buildings can also be described by superimposing normal distribution functions [21,29]. It is important to realize that when the building-level model is extended to the urban-level, the proper urban parameters must be carefully selected. Transportation Accessibility and Population Level, two urban parameters sufficient for commercial buildings, may have to be modified or changed for other building types. The key is to find proper urban parameters to reflect the inner correlation between the building-level occupant model and the building's urban context.

## 7. Conclusions

The occupant information is critical for building and urban energy simulation. The dynamic nature of occupant density makes it challenging to be accurately predicted. In the context of UBEM, the occupant density of hundreds and even thousands of buildings need to be obtained. To address these problems, an innovative method is proposed to develop the urban-level dynamic occupant density models for commercial buildings.

The methodology starts from the determination and classification of commercial buildings. The POIs are used to determine which ones are commercial buildings. The sub-categories of commercial buildings are determined using logistic regression. The end results are 219 shopping malls, 341 supermarkets, and 1715 specialty stores in the city of Nanjing.

The hourly occupant densities of 12 sample buildings, including 4 shopping malls, 4 supermarkets, and 4 specialty stores, were measured. Through regression analysis, it is found that the superposition of normal distribution functions can accurately describe the measured occupancy profile, as demonstrated by an overall $R^2$ value above 0.9 This established building-level dynamic occupant density model serves as the foundation upon which the urban-level model is further developed.

The key step of expanding the building-level dynamic occupant density model to the urban-level is establishing the relationship between urban parameters of the commercial building and its occupancy pattern. Transportation accessibility and population level, two selected urban parameters, were obtained using big data. The coefficients defining the building-level dynamic occupant density models were correlated with these two urban parameters. So far, a complete urban-level dynamic occupant density model for commercial buildings has been established. The model was tested on 6 samples. According to the $R^2$ value, the overall goodness of model prediction is acceptable.

In addition to predicting the urban-level dynamic occupant density pattern of commercial buildings, the method and framework developed in this paper can be applied to other types of buildings as well. However, for non-commercial buildings transportation accessibility and population level may not be proper urban parameters. One has to modify them or select other urban parameters to expand the building-level model to the urban-level.

## Declaration of competing interest

None.

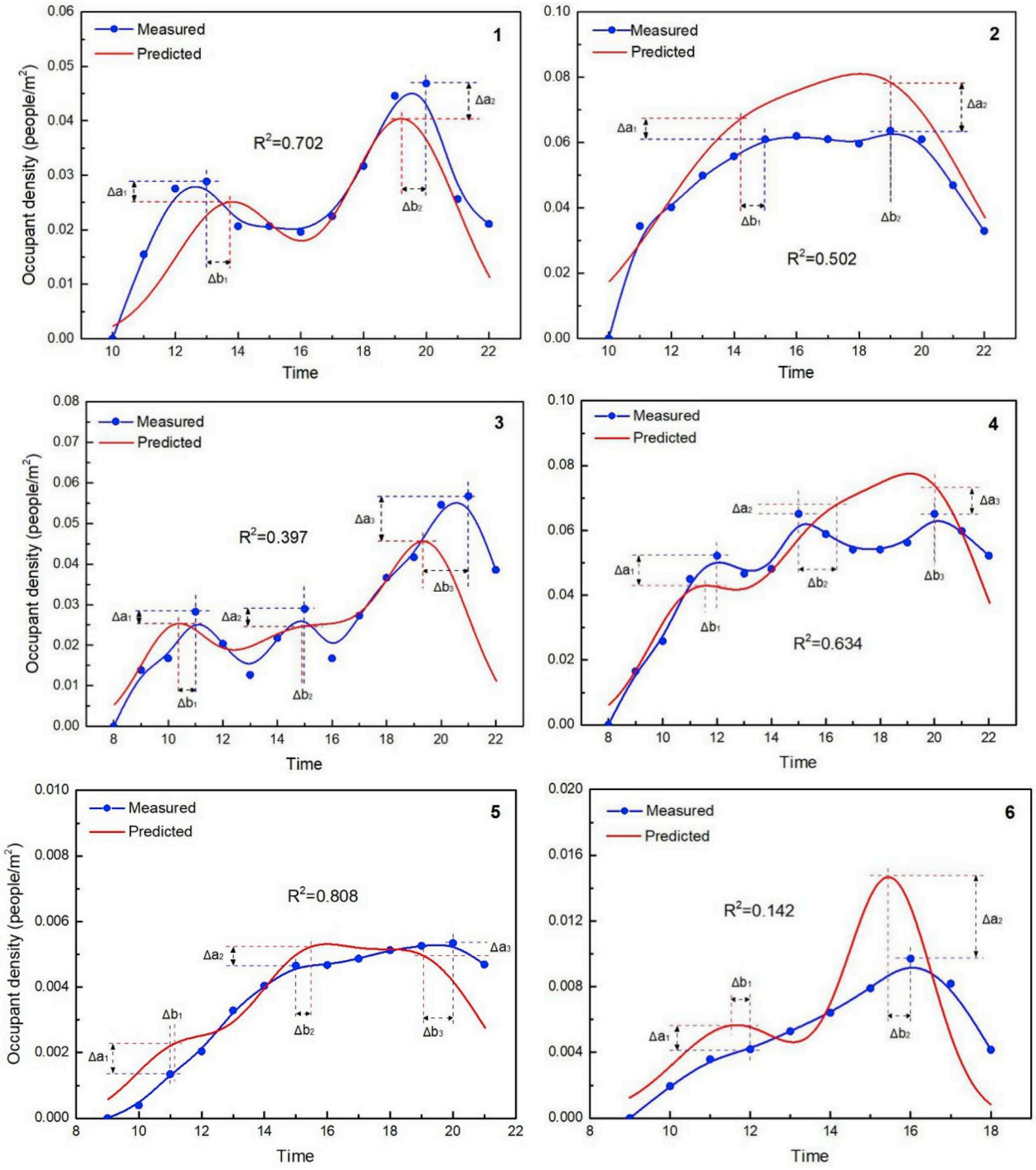



**Fig. 14.** The comparison between measured and predicted occupant density models.

**Table 14**
The key coefficients of measured and predicted models.

| ID | Measured models | | | | | | Predicted models | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $a_1$ | $a_2$ | $a_3$ | $b_1$ | $b_2$ | $b_3$ | $a_1$' | $a_2$' | $a_3$' | $b_1$' | $b_2$' | $b_3$' |
| 1 | 0.0289 | 0.0468 | / | 13 | 20 | / | 0.0248 | 0.0403 | / | 13.72 | 19.20 | / |
| 2 | 0.0610 | 0.0637 | / | 15 | 19 | / | 0.0556 | 0.0695 | / | 13.99 | 19.06 | / |
| 3 | 0.0282 | 0.0290 | 0.0567 | 11 | 15 | 21 | 0.0196 | 0.0244 | 0.0378 | 10.11 | 15.10 | 19.56 |
| 4 | 0.0522 | 0.0651 | 0.0651 | 12 | 15 | 20 | 0.0345 | 0.0601 | 0.0576 | 10.92 | 16.03 | 20.09 |
| 5 | 0.0013 | 0.0047 | 0.0053 | 11 | 15 | 20 | 0.0019 | 0.0045 | 0.0045 | 11.13 | 15.18 | 19.09 |
| 6 | 0.0042 | 0.0097 | / | 12 | 16 | / | 0.0056 | 0.0144 | / | 11.64 | 15.48 | / |

**Table 15**
The differences between the key coefficients.

| ID | $\Delta a_1$ | $\Delta a_2$ | $\Delta a_3$ | $\delta a_1$ | $\delta a_2$ | $\delta a_3$ | $\Delta b_1$ | $\Delta b_2$ | $\Delta b_3$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | −0.0041 | −0.0065 | / | 14.18% | 13.89% | / | 43min | −48min | / |
| 2 | −0.0054 | 0.0058 | / | 8.85% | 9.11% | / | −60min | 4min | / |
| 3 | −0.0086 | −0.0046 | −0.0189 | 30.50% | 15.86% | 33.33% | −53min | 6min | −86min |
| 4 | −0.0177 | −0.0050 | −0.0075 | 33.91% | 7.68% | 11.52% | −65min | 62min | 5min |
| 5 | 0.0006 | −0.0002 | −0.0008 | 46.15% | 42.55% | 15.09% | 8min | 11min | −55min |
| 6 | 0.0014 | 0.0047 | / | 33.33% | 48.45% | / | −22min | −31min | / |

## Acknowledgements

This paper is financially supported by the Ministry of Science and Technology of China (grant number: 2016YFC0700102) and the National Natural Science Foundation of China (grant number: 51678124).

## References

[1] T. Damassa, M. Ge, T. Fransen, The U.S. Greenhouse Gas Reduction Targets, 2014.
[2] European Commission, EU Climate Action, 2017.
[3] China's State Council, Approval of National Climate Change Response Planning (2014-2020), 2014.
[4] C. Cerezo Davila, C.F. Reinhart, J.L. Bemis, Modeling Boston: a workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, Energy 117 (2016) 237–250.
[5] C.F. Reinhart, C. Cerezo Davila, Urban building energy modeling–A review of a nascent field, Build. Environ. 97 (2016) 196–202.
[6] C. Cerezo, J. Sokol, S. AlKhaled, C. Reinhart, A. Al-Mumin, A. Hajiah, Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): a residential case study in Kuwait City, Energy Build. 154 (2017) 321–334.
[7] R. Nouvel, M. Zirak, V. Coors, U. Eicker, The influence of data quality on urban heating demand modeling using 3D city models, Comput. Environ. Urban Syst. 64 (2017) 68–80.
[8] X. Feng, D. Yan, T. Hong, Simulation of occupancy in buildings, Energy Build. 87 (2015) 348–359.
[9] S. D'Oca, T. Hong, A data-mining approach to discover patterns of window opening and closing behavior in offices, Build. Environ. 82 (2014) 726–739.
[10] C.S. Monteiro, C. Costa, A. Pina, M.Y. Santos, P. Ferrao, An urban building database (UBD) supporting a smart city information system, Energy Build. 158 (2018) 244–260.
[11] A.C. Menezes, A. Cripps, D. Bouchlaghem, R. Buswell, Predicted vs. actual energy performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap, Appl. Energy 97 (2012) 355–364.
[12] S.S.K. Kwok, E.W.M. Eric, A study of the importance of occupancy to building cooling load in prediction by intelligent approach, Energy Convers. Manag. 52 (7) (2011) 2555–2564.
[13] J. Page, D. Robinson, N. Morel, J.L. Scartezzini, A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2) (2008) 83–98.
[14] Q. Wang, G. Augenbroe, J. Kim, L. Gu, Meta-modeling of occupancy variables and analysis of their impact on energy outcomes of office buildings, Appl. Energy 174 (2016) 166–180.
[15] T. Hong, S.C. Taylor-Lange, S. D Oca, D. Yan, S.P. Corgnati, Advances in research and applications of energy-related occupant behavior in buildings, Energy Build. 116 (2016) 694–702.
[16] F. Haldi, D. Cali, R.K. Andersen, M. Wesseling, D. Muller, Modelling diversity in building occupant behaviour: a novel statistical approach, J. Build. Perform. Simul. 10 (5–6) (2017) 527–544.
[17] F. Haldi, D. Robinson, The impact of occupants' behaviour on building energy demand, J. Build. Perform. Simul. 4 (4) (2011) 323–338.
[18] F. Tahmasebi, A. Mandavi, On the utility of occupants' behavioural diversity information for building performance simulation: an exploratory case study, Energy Build. 176 (2018) 380–389.
[19] G.Y. Yun, K. Steemers, Night-time ventilated offices: statistical simulations of window-use patterns from field monitoring, Sol. Energy 84 (7) (2010) 1216–1231.
[20] Y. Zhang, P. Barrett, Factors influencing the occupants' window opening behaviour in a naturally ventilated office building, Build. Environ. 50 (2012) 125–134.
[21] W. O'Brien, H.B. Gunay, F. Tahmasebi, A. Mahdavi, A preliminary study of representing the inter-occupant diversity in occupant modelling, J. Build. Perform. Simul. 10 (5–6) (2017) 509–526.
[22] B. Dong, D. Yan, Z. Li, Y. Jin, X. Feng, H. Fontenot, Modelling occupancy and behavior for better building design and operation—A critical review, Build. Simul. 11 (3) (2018) 899–921.
[23] H.B. Gunay, W. O'Brien, I. Beausoleil-Morrison, A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices, Build. Environ. 70 (2013) 31–47.
[24] F. Haldi, D. Robinson, Adaptive actions on shading devices in response to local visual stimuli, J. Build. Perform. Simul. 3 (2) (2010) 135–153.
[25] Z. Yang, B. Becerik-Gerber, Coupling occupancy information with HVAC energy simulation: a systematic review of simulation programs, in: Proceedings of the 2014 Winter Simulation Conference: Savanah, USA, 2014.
[26] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, Build. Simul. 4 (2) (2011) 149–167.
[27] B. Ai, Z. Fan, R. Gao, Occupancy estimation for smart buildings by an auto-regressive hidden markov model, in: Proceedings of American Control Conference: Portland, USA, 2014.
[28] V.L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A.E. Cerpa, M.D. Sohn, S. Narayanan, Energy efficient building environment control strategies using real-time occupancy measurements, in: Proceedings of the 1st ACM on Embedded Sensing Systems for Energy-Efficiency in Buildings: Berkeley, USA, 2009.
[29] S. Gilani, W. O'Brien, A preliminary study of occupants' use of manual lighting controls in private offices: a case study, Energy Build. 159 (2018) 572–586.
[30] JGJ48-2014, Code for design of store buildings, 2014.
[31] S. Kimura, Y. Yamaguchi, B. Kim, Y. Miyachi, Y. Kou, Y. Shimoda, Urban scale energy demand modelling of commercial building stock considering the variety of HVAC system configuration, in: Proceedings of the 15th IBPSA Conference: San Francisco, USA, 2017.
[32] W. Li, Y. Zhou, K.S. Cetin, S. Yu, Y. Wang, B. Liang, Developing a landscape of urban building energy use with improved spatiotemporal representations in a cool-humid climate, Build. Environ. 136 (2018) 107–117.
[33] L. Feng, C. Tan, S. Lu, D. Kong, Statistical analysis and modeling for occupant density in commercial buildings, Fire Saf. Sci. 18 (3) (2014) 130–134.
[34] J. Li, Y. Long, A. Dang, Live-Work-Play centers of Chinese cities: identification and temporal evolution with emerging data, Comput. Environ. Urban Syst. 71 (2018) 58–66.
[35] C. Zhong, M. Schlapfer, S.M. Arisona, C. Ratti, M. Batty, G. Schmitt, Revealing the changing spatial structure of cities from human activity patterns, Urban Stud. 54 (2) (2017) 437–455.
[36] Y. Sun, H. Fan, M. Li, A. Zipf, Identifying the city center using human travel flows generated from location-based social networking data, Environ. Plan. Plan. Des. 43 (3) (2015) 480–498.
[37] N. Niu, X. Liu, H. Jin, X. Ye, Y. Liu, X. Li, Y. Chen, S. Li, Integting multi-source big data to infer building functions, Geogr. Inf. Sci. 31 (9) (2017) 1871–1890.
[38] G. Pan, G. Qi, Z. Wu, D. Zhang, S. Li, Land-use classification using taxi GPS traces, IEEE Trans. Intell. Transp. Syst. 14 (1) (2013) 113–123.
[39] Nanjing Statistic Bureau, Nanjing Municipal Bureau Statistics, 2017, 2017.
[40] GB 50176-2016, Code for Thermal Design of Civil Building, 2016.
[41] H. Fan, A. Zipf, Q. Fu, P. Neis, Quality assessment for building footprints data on OpenStreetMap, Geogr. Inf. Sci. 28 (4) (2014) 700–719.
[42] S.H. Holmes, N.B. Rajkovich, F. Baker, Optimizing a parametric energy model for use in citywide residential overheating analysis, in: Proceedings of the 15th IBPSA Conference: San Francisco, USA, 2017.
[43] J. Song, T. Lin, X. Li, A.V. Prishchepov, Mapping urban functional zones by integrating very high spatial resolution remote sensing imagery and Points of Interest: a case study of Xiamen, China, Remote Sens. 10 (11) (2018) 1737.
[44] P. Caputo, G. Costa, S. Ferrari, A supporting method for defining energy strategies in the building sector at urban scale, Energy Policy 55 (2013) 261–270.
[45] A. Alhamwi, W. Medjroubi, T. Vogt, C. Agert, OpenStreetMap data in modelling the urban energy infrastructure: a first assessment and analysis, in: Proceedings of the 9th International Conference on Applied Energy, Cardiff, UK, 2017.
[46] G. Gao, Z. Wang, X. Liu, Q. Li, W. Wang, J. Zhang, Travel behavior analysis using 2016 Qingdao's household traffic surveys and Baidu electric map API data, Adv. Transp. 2019 (1) (2019) 1–18.
[47] M. Liu, J. Wang, J. Wang, G. Tang, M. Sun, Design and implementation of emergency and evacuation system based on Baidu map API, in: Proceedings of the 26th International Conference on Geoinformatics: Yunnan, China, 2018.
[48] ISO 14825-2004, Intelligent Transport Systems-Geographic Data Files (GDF)-Overall Data Specification, 2004.
[49] H. Zou, Y. Zhou, J. Yang, C.J. Spanos, Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT, Energy Build. 174 (2018) 309–322.
[50] MOHURD, Ministry of Housing and Urban-Rural Development of People's Republic of China, Code for design of urban road engineering, 2012.
[51] T. Dogan, C. Reinhart, Shoeboxer, An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation, Energy Build. 140 (2017) 140–153.