Applied Energy 278 (2020) 115656

Contents lists available at ScienceDirect

Applied Energy

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

A novel mobility-based approach to derive urban-scale building occupant profiles and analyze impacts on building energy consumption

Wenbo Wu [a], Bing Dong [b,*], Qi (Ryan) Wang [c], Meng Kong [b], Da Yan [d], Jingjing An [e], Yapan Liu [b]

[a] Department of Management Science and Statistics, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, United States
[b] Department of Mechanical & Aerospace Engineering, Syracuse University, 223 Link Hall, Syracuse, NY 13244, United States
[c] Department of Civil and Environmental Engineering, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, United States
[d] Building Energy Research Centre, School of Architecture, Tsinghua University, Beijing 100084, China
[e] School of Environment and Energy Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

HIGHLIGHTS

- Develop a novel mobility-based approach to derive urban scale building occupant profiles.
- Compare empirically derived profiles with the existing DoE reference model.
- Quantify the differences using statistical methods.
- Urban scale energy simulations show up to 60% differences in heating and 40% in cooling energy consumption.

ARTICLE INFO

Keywords:
Occupancy profile
Urban mobility
Global positioning system
Urban-scale building energy modeling

ABSTRACT

In the US, people spend more than 90% of their time in buildings, which contributes to more than 70% of overall electricity usage in the country. Occupant behavior is becoming a leading factor impacting energy consumption in buildings. Existing occupant-behavior studies are often limited to a single building and individual behavior, such as presence or interactions in confined spaces. Moreover, studies modeling occupant behavior at the building or community level are limited. With the development of the Internet of Things, mobile positioning data are available through social media and location-based service applications. The goal of this study is to analyze the impacts of more representative occupancy profiles, derived from high resolution urban scale mobile position data, on building energy consumption. . A pilot study was conducted on more than 900 buildings in downtown San Antonio, Texas, with billions of mobile positioning data. We then compared these profiles with the existing Department of Energy prototype models and quantified the differences using a statistical method. On average, the differences in occupancy rates between the ones derived from the empirical profile and the ones from the Department of Energy reference ranged from −30% to 70%. The realistic derived profiles are then simulated in the CityBES. The results show that the predicted cooling energy demand is reduced by up to 40% while the heating energy demand is reduced by up to 60%. This study, therefore, advances knowledge of urban planning as well as urban-scale energy modeling and optimization.

1. Introduction

By 2050, 70% of the world's population is projected to live and work in cities [1], with two-thirds of global primary energy consumption attributed to cities, resulting in the production of 73% of global energy-related greenhouse gas emissions [2]. In the US, people currently spend more than 90% of their time in buildings [3], which contributes to more than 70% of overall electricity usage in the country [4]. Occupant behavior is a leading factor influencing energy consumption in buildings [5]. Existing occupant behavior studies are often isolated and focus on individual behavior, such as presence or interactions [6–11] in a single space or building. Recent studies have addressed various optimization, control, and occupancy-related challenges for the operation of individual buildings [12,13]. However, studies modeling occupant behavior at

* Corresponding author.
E-mail address: bidong@syr.edu (B. Dong).

https://doi.org/10.1016/j.apenergy.2020.115656
Received 9 April 2020; Received in revised form 9 July 2020; Accepted 2 August 2020
Available online 18 August 2020

community and urban levels are limited [14] but are necessary for urban-scale energy modeling and energy policy decision making.

Urban-scale energy modeling tools have been used for spatial analysis of energy consumption [15–17], retrofitting of energy-related policy development based on building occupancy profiles [18–21], and design and planning of new cities to optimize energy use [22]. However, most current urban-scale energy modeling tools use pre-defined or synthetic data to simulate occupancy profiles [5], which results in obvious energy differences compared with the use of practical occupancy profiles [23,24]. Existing studies did not use actual occupancy profiles owing to the dearth of such datasets. In addition, it is almost impossible to obtain information for a large number of buildings because of the high cost of occupant counting sensors. Nonetheless, with the development of the Internet of Things, researchers are considering data from mobile phones and Wi-Fi-based sensor data [25,26]. Recent research on urban mobility has uncovered the potential of deriving occupancy locations using mobile positioning data [12,27]. Furthermore, geosocial network data, such as Twitter and Foursquare, have been used to study the spatiotemporal patterns of occupants in urban environments [28–30].

In this study, we explored the possibility to use mobile position data to derive occupancy profiles at individual building level and evaluate their impacts on building energy consumption. The paper is organized as the following: we summarize the recent developments and research gaps in Section 2. In Section 3, we describe the overall methodology including data processing, deriving empirical occupancy profiles, statistical quantification method and urban scale energy modeling. We present the results in Section 4. We summarize our concluding remarks and limitation of this study in Section 7.

2. Literature review and research gaps

2.1. Literature review

Occupant presence modeling. Arrival, departure, and duration of occupancy absence are important factors influencing human–building interactions, such as lighting, thermostats, window blinds, and plug behaviors [31,32]. The above behaviors all depend on occupancy presence. Modern occupant presence modeling approaches fall into three general categories: scheduled occupancy patterns, stochastic models, and machine learning methods. Scheduled occupancy patterns are the most commonly used in industry but can lead to errors as great as 600% [33]. These patterns are popular because most building simulation software uses this type of model. A typical example is the diversity factor, with a previous study showing a 46% difference when compared to actual profiles [5,34,35]. Stochastic models such as the discrete-time Markov chain [36,37] used connected thermostat data to detect occupancy presence in residential buildings [38]. Most recently, the Gated Recurrent Unit Network has been applied to model the trace of occupant location [39] with a root mean square error of 4.79 cm for a single occupant in terms of spatial coordinates in a room.

Occupant sensing. Happle et al. [9] discussed the importance of considering occupant behavior models while planning building energy consumption and examined different occupancy behavior models, such as a deterministic space-based approach, which is a rule-based occupant behavior model, a stochastic space-based approach, which is based on occupant behavior in the built environment over different time spans on the building scale, and a stochastic person-based approach, which considers different time spans yet focuses on individuals. Because of the diversity of occupant behavior, the stochastic individual-based approach is superior to the stochastic space-based approach. To overcome the limitation of implementing this method on both the district and individual building levels, Happle et al. implemented a novel method, namely an "activity-based multi-agent approach." The occupant activity and appliance usage were monitored to model urban building occupancy.

General occupancy pattern detection methods for urban-scale modeling are commonly based on mobile internet-based data, cellular service data, Wi-Fi-based sensor data, and geosocial network data, such as Twitter and Foursquare. The prevalence of smartphone usage has enabled the tracking of human movement and occupancy patterns [25]. GPS sensor logs from smartphones have been used as a data source in fusion frameworks to study travel mode detection [40], since such location data have a high spatiotemporal resolution. Gu et al. [12] applied modeling occupancy behavior on an urban scale. K-means clustering has been applied to extract typical occupancy data. Occupancy data extracted from 60 buildings were classified into 7 categories and 19 sub-categories. Different building types showed different occupancy profiles and densities, and the same building types in different areas followed different occupancy patterns. Pang et al. [27] used mobile internet-based positioning data to model occupancy behavior in the built environment. Occupancy data were integrated with the building energy simulation [41] in EnergyPlus, and the results showed that mobile internet-based positioning data increased the precision accuracy of building energy models. Using the TimeGeo framework, Barbour et al. [42] estimated building occupancy in the city of Boston. Occupancy schedules were extracted for residential, commercial, and industrial buildings at urban scale. Compared with DOE standard occupancy schedules, their study showed potential energy savings up to 21% for commercial buildings and 15% for residential buildings. However, the building occupancy data generated by TimeGeo framework is not purely from raw GPS data. In order to simulate the whole population in each census tract, the TimeGeo framework expands active phone users to the population of each census tract in Boston's metro area. To generate individual building occupancy profiles, the TimeGeo framework uses a time-inhomogeneous Markov chain model to model temporal choices, and uses a rank-based exploration and preferential return (r-EPR) model to generate spatial choices. Hence, building occupancy profiles from TimeGEO framework are synthetical data. Happle et al. [43] created context-specific, data-driven occupancy schedules by utilizing location-based services data from Google Maps. Comparing with DOE references, it showed that the use of standard schedules could lead to overestimation of urban scale energy demand. However, Happle's study is limited in downtown areas of 13 selected big cities with occupancy data from commercial buildings only.

Mohammadi and Taylor [44] discussed the spatiotemporal relationship between human mobility and energy consumption in Chicago, IL. These authors collected Twitter and electricity consumption data for over one year. To determine the spatiotemporal pattern of occupants in an urban environment, they used a spatial autoregression model. The gyration radius was used to determine returners' intra-urban mobility. Building on previous work on the dichotomy of human mobility [45], Mohammadi and Taylor [28] explored two major network mobility patterns in urban environments, namely returners and explorers. Returners' mobility networks were dominated by a few frequently visited locations, but explorers' mobility network patterns were much larger. Energy consumption in urban environments did not solely depend on individual building occupancy level, as the authors provided insight into the spill-over effect of neighboring buildings. Thus, a better understanding of urban mobility patterns will likely improve the prediction, management, and efficient allocation of resources [5,8]. Wang et al. [30] explored urban human mobility patterns to investigate the social nature of people in 50 large American cities. Other studies also used geotagged tweets to estimate home and work locations and identify peoples' commutes on the city scale [14,46]. Riascos and Mateos [29] discussed human mobility in urban areas using data from Foursquare, a location-based networking platform. The co-presence of people was analyzed on a temporally and spatially. Social network data were used to study occupancy behavior patterns in New York and Tokyo metropolitan areas. Their travel distances and co-presence in public establishments, such as universities, bars, and restaurants, were examined to elucidate these patterns.

Urban scale energy modeling platform: Urban building energy

Fig. 1. Overview of the methodology.

modeling (UBEM) platform is essential for building energy studies at an urban scale, and it supports researchers to model a large group of buildings effectively. UBEMs within the literature can be classified into the top-down approach and bottom-up approach. Top-down (i.e., data-driven) models always include economic variables and statistical information, and those models are primarily based on macro-economic modeling principles and techniques. However, the bottom-up approach (i.e., Model-based building performance simulation) tends to group buildings with similar characteristics into one category which is the archetype. Those characteristics include building geometries, structural properties, building use, etc.

Bottom-up models include CityBES [47] UMI [48], HUES [49], TEASER [50], CitySim [51], and SUNtool [52]. CityBES was developed by Lawrence Berkeley National Laboratory. CityBES [47] is a web-based platform for urban-scale building energy modeling and analysis. It uses an international open data standard – CityGML for exchanging and representing 3D city building models. CityBES is built on OpenStudio and EnergyPlus to simulate building energy use and savings from energy-efficient retrofits. Reinhart et al. [48] developed a new Rhinoceros-based urban modeling design tool, UMI, to study operational energy, daylighting, outdoor comfort and walkability evaluations at the district level. UMI uses EnergyPlus, Radiance/Daysim as simulation engines. HUES platform is an extendable simulation environment for the study of urban multi-energy systems. TEASER, CitySim and SUNtool are only designed for studying operational building energy usages.

2.2. Summary and research gaps

Despite key findings from the abovementioned studies, critical knowledge gaps remain, as discussed below:

(1) Lack of an urban scale mobility data set to create urban scale occupancy profiles. Compared with previous studies [12,27,42,43], the uniqueness of this study is that it derived occupancy profiles for all DOE referenced building type at an urban scale. For example, Barbour's study only investigated occupancy data for 3 types of buildings, Happle's study only focused on commercial buildings, Gu et al. extracted occupancy data for 7 building types, Pang et al. investigated occupancy data for office buildings only, but this study covered 16 different building types.
(2) Lack of a statistical method to quantify the differences, in terms of occupancy rate and profiles, between derived and those from DOE prototype models. The existing studies focus on deriving and visualizing the occupancy profiles based on different data sources. However, how to quantify and statistically compare the differences between the derived empirical occupancy profiles to the DOE prototype models is still missing.
(3) Lack of a holistic assessment of the impacts of more representative occupancy profiles on urban-scale building energy usages. Previous studies either did not address the differences in at the individual building energy usage level or only focuses on a limited number and type of buildings. For example, Barbour's study only investigated building energy consumption at census tract level. In this paper, we integrated derived occupancy profiles from mobile position data with an urban scale energy simulation platform-CityBES to evaluate such impacts.

3. Methodology

3.1. Overview

Fig. 1 shows the overview of the methodology in this paper. First, ArcGIS and LiDAR data were processed to get basic building geometry data. Positioning data collected from mobile phones were processed used to extract the occupancy profile patterns of more than 900 buildings on an urban scale. We categorized these buildings according to the 16 reference building types of the US Department of Energy (DOE) [7], and we compared the derived schedules with the default occupancy profiles of these 16 building types. A statistical method was then developed to quantify the differences in these schedules. As a pilot study, this study focused on analysis methods and preliminary results with two exemplary building types as case studies. Finally, the assessment on the impacts of energy usages from the more representative occupancy profiles was conducted, comparing with those from DOE models.

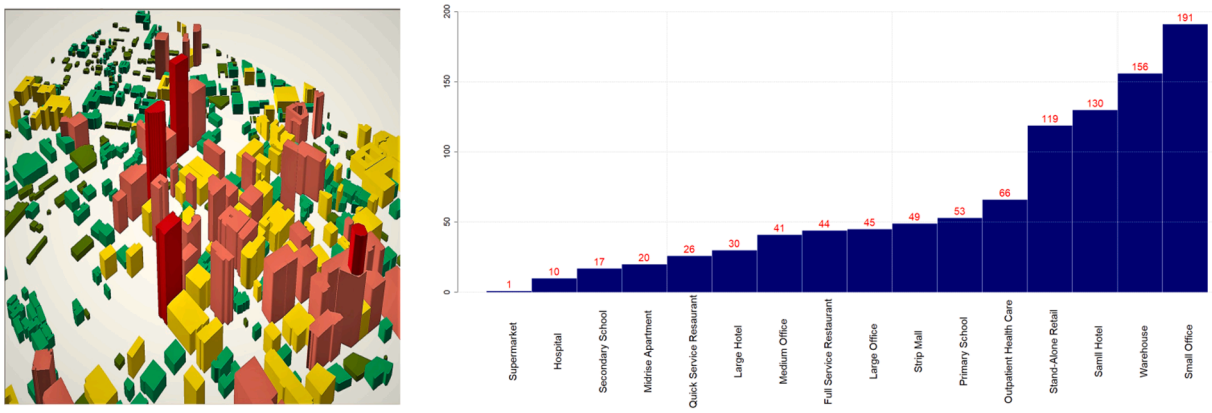

Fig. 2. Three-dimensional illustration and distribution of selected buildings from different Department of Energy reference building types.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Small Retail_Weekday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.50	0.50	0.70	0.70	0.70	0.70	0.80	0.70	0.50	0.50	0.30	0.30	0.00	0.00	0.00
Small Retail_Weekday_new	0.07	0.06	0.05	0.05	0.05	0.05	0.07	0.10	0.15	0.17	0.18	0.18	0.19	0.20	0.19	0.20	0.20	0.17	0.16	0.15	0.13	0.12	0.10	0.08
Small Retail_Sunday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.20	0.40	0.40	0.40	0.40	0.40	0.20	0.10	0.00	0.00	0.00	0.00	0.00
Small Retail_Sunday_new	0.12	0.10	0.08	0.07	0.07	0.07	0.08	0.09	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.17	0.20	0.18	0.18	0.19	0.20	0.20	0.18	0.15
Small Retail_Saturday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.50	0.60	0.80	0.80	0.80	0.80	0.80	0.80	0.60	0.20	0.20	0.20	0.10	0.00	0.00
Small Retail_Saturday_new	0.12	0.10	0.08	0.07	0.07	0.07	0.08	0.09	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.17	0.20	0.18	0.18	0.19	0.20	0.20	0.18	0.15
Small Office_Weekday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.95	0.95	0.95	0.95	0.50	0.95	0.95	0.95	0.95	0.30	0.10	0.10	0.10	0.10	0.05	0.05
Small Office_Weekday_new	0.04	0.04	0.04	0.03	0.03	0.04	0.06	0.12	0.21	0.24	0.26	0.26	0.26	0.27	0.25	0.24	0.21	0.16	0.11	0.09	0.08	0.07	0.06	0.04
Small Office_Sunday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Small Office_Sunday_new	0.08	0.06	0.06	0.05	0.05	0.04	0.05	0.06	0.10	0.13	0.12	0.14	0.12	0.12	0.14	0.13	0.12	0.13	0.15	0.14	0.13	0.13	0.11	0.08
Small Office_Saturday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.30	0.30	0.30	0.30	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.00	0.00	0.00	0.00	0.00
Small Office_Saturday_new	0.08	0.06	0.06	0.05	0.05	0.04	0.05	0.06	0.10	0.13	0.12	0.14	0.12	0.12	0.14	0.13	0.12	0.13	0.15	0.14	0.13	0.13	0.11	0.08
Medium Retail_Weekday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.50	0.50	0.70	0.70	0.70	0.70	0.80	0.70	0.50	0.50	0.30	0.30	0.00	0.00	0.00
Medium Retail_Weekday_new	0.07	0.06	0.05	0.05	0.05	0.05	0.07	0.10	0.15	0.17	0.18	0.18	0.19	0.20	0.19	0.20	0.20	0.17	0.16	0.15	0.13	0.12	0.10	0.08
Medium Retail_Sunday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.20	0.40	0.40	0.40	0.40	0.40	0.20	0.10	0.00	0.00	0.00	0.00	0.00
Medium Retail_Sunday_new	0.12	0.10	0.08	0.07	0.07	0.07	0.08	0.09	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.17	0.20	0.18	0.18	0.19	0.20	0.20	0.18	0.15
Medium Retail_Saturday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.50	0.60	0.80	0.80	0.80	0.80	0.80	0.80	0.60	0.20	0.20	0.20	0.10	0.00	0.00
Medium Retail_Saturday_new	0.12	0.10	0.08	0.07	0.07	0.07	0.08	0.09	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.17	0.20	0.18	0.18	0.19	0.20	0.20	0.18	0.15
Medium Office_Weekday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.95	0.95	0.95	0.95	0.50	0.95	0.95	0.95	0.95	0.30	0.10	0.10	0.10	0.10	0.05	0.05
Medium Office_Weekday_new	0.05	0.05	0.04	0.04	0.04	0.05	0.09	0.21	0.35	0.39	0.40	0.41	0.40	0.42	0.39	0.38	0.33	0.24	0.15	0.11	0.10	0.09	0.08	0.06
Medium Office_Sunday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Medium Office_Sunday_new	0.08	0.08	0.07	0.07	0.06	0.06	0.07	0.08	0.10	0.11	0.13	0.12	0.12	0.12	0.12	0.11	0.15	0.14	0.13	0.13	0.15	0.16	0.13	0.11
Medium Office_Saturday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.30	0.30	0.30	0.30	0.10	0.10	0.10	0.10	0.10	0.05	0.05	0.00	0.00	0.00	0.00	0.00
Medium Office_Saturday_new	0.08	0.08	0.07	0.07	0.06	0.06	0.07	0.08	0.10	0.11	0.13	0.12	0.12	0.12	0.12	0.11	0.15	0.14	0.13	0.13	0.15	0.16	0.13	0.11
Large Office_Weekday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.95	0.95	0.95	0.95	0.50	0.95	0.95	0.95	0.95	0.70	0.40	0.40	0.10	0.10	0.05	0.05
Large Office_Weekday_new	0.04	0.03	0.03	0.03	0.03	0.05	0.13	0.32	0.54	0.56	0.58	0.59	0.59	0.60	0.55	0.52	0.45	0.31	0.14	0.09	0.07	0.06	0.05	0.04
Large Office_Sunday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large Office_Sunday_new	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.06	0.08	0.08	0.09	0.09	0.09	0.09	0.08	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.05
Large Office_Saturday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.50	0.50	0.50	0.50	0.50	0.50	0.10	0.10	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large Office_Saturday_new	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.06	0.08	0.08	0.09	0.09	0.09	0.09	0.08	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.05
Large Office_basement_Weekday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.95	0.95	0.95	0.95	0.50	0.95	0.95	0.95	0.95	0.70	0.40	0.40	0.10	0.10	0.05	0.05
Large Office_basement_Weekday_new	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.95	0.95	0.95	0.95	0.50	0.95	0.95	0.95	0.95	0.70	0.40	0.40	0.10	0.10	0.05	0.05
Large Office_basement_Sunday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large Office_basement_Sunday_new	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large Office_basement_Saturday_ori	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.50	0.50	0.50	0.50	0.50	0.50	0.10	0.10	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large Office_basement_Saturday_new	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.50	0.50	0.50	0.50	0.50	0.50	0.10	0.10	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Full Service Restaurant_Weekday_ori	0.05	0.00	0.00	0.00	0.00	0.05	0.10	0.40	0.40	0.40	0.20	0.50	0.80	0.70	0.40	0.20	0.25	0.50	0.80	0.80	0.80	0.50	0.35	0.20
Full Service Restaurant_Weekday_new	0.07	0.06	0.05	0.05	0.04	0.05	0.06	0.09	0.11	0.13	0.14	0.17	0.20	0.19	0.17	0.16	0.17	0.17	0.19	0.21	0.21	0.19	0.15	0.12
Full Service Restaurant_Sunday_ori	0.05	0.00	0.00	0.00	0.00	0.00	0.05	0.50	0.50	0.20	0.20	0.30	0.50	0.50	0.30	0.20	0.25	0.35	0.55	0.65	0.70	0.35	0.20	0.20
Full Service Restaurant_Sunday_new	0.16	0.16	0.10	0.07	0.05	0.05	0.06	0.08	0.10	0.11	0.12	0.15	0.18	0.18	0.19	0.18	0.22	0.24	0.27	0.29	0.30	0.30	0.29	0.24
Full Service Restaurant_Saturday_ori	0.05	0.00	0.00	0.00	0.00	0.00	0.05	0.50	0.50	0.40	0.20	0.45	0.50	0.50	0.35	0.30	0.30	0.30	0.70	0.90	0.70	0.65	0.55	0.35
Full Service Restaurant_Saturday_new	0.16	0.16	0.10	0.07	0.05	0.05	0.06	0.08	0.10	0.11	0.12	0.15	0.18	0.18	0.19	0.18	0.22	0.24	0.27	0.29	0.30	0.30	0.29	0.24

Fig. 3. Comparison between the reference schedule and the actual schedule of all building types.

Table 1
Temperature set-point control for the actual schedule cases for each building type.

	Occupancy rate								Note
	<0.1	0.1–0.2	0.2–0.25	0.25–0.3	0.3–0.4	0.4–0.5	0.5–0.7	0.7–1.0	
Small Office	All Ts	~20% zones Tn, otherwise Ts	~25% zones Tn, otherwise Ts	~30% zones Tn, otherwise Ts	100% zones Tn				Setback set-point Ts: 29.44 °C for cooling and 15.56 °C for heating; Normal set-point Tn: 23.89 °C for cooling and 21.11 °C for heating
Medium Office	All Ts	~30% zones Tn, otherwise Ts			~40% zones Tn, otherwise Ts	~50% zones Tn, otherwise Ts	100% zones Tn		
Large Office	All Ts	~30% zones Tn, otherwise Ts			~50% zones Tn, otherwise Ts		~70% zones Tn, otherwise Ts	100% zones Tn	
Full-Service Restaurant	All Ts	All Tn							Setback set-point Ts: 30.00 °C for cooling and 15.60 °C for heating; Normal set-point Tn: 24.00 °C for cooling and 21.00 °C for heating
Small and Medium Retailers	All Ts	All Tn							Setback set-point Ts: 29.44 °C for cooling and 15.56 °C for heating; Normal set-point Tn: 23.89 °C for cooling and 21.11 °C for heating

Note: Tn: normal setpoint; Ts: setback setpoint.

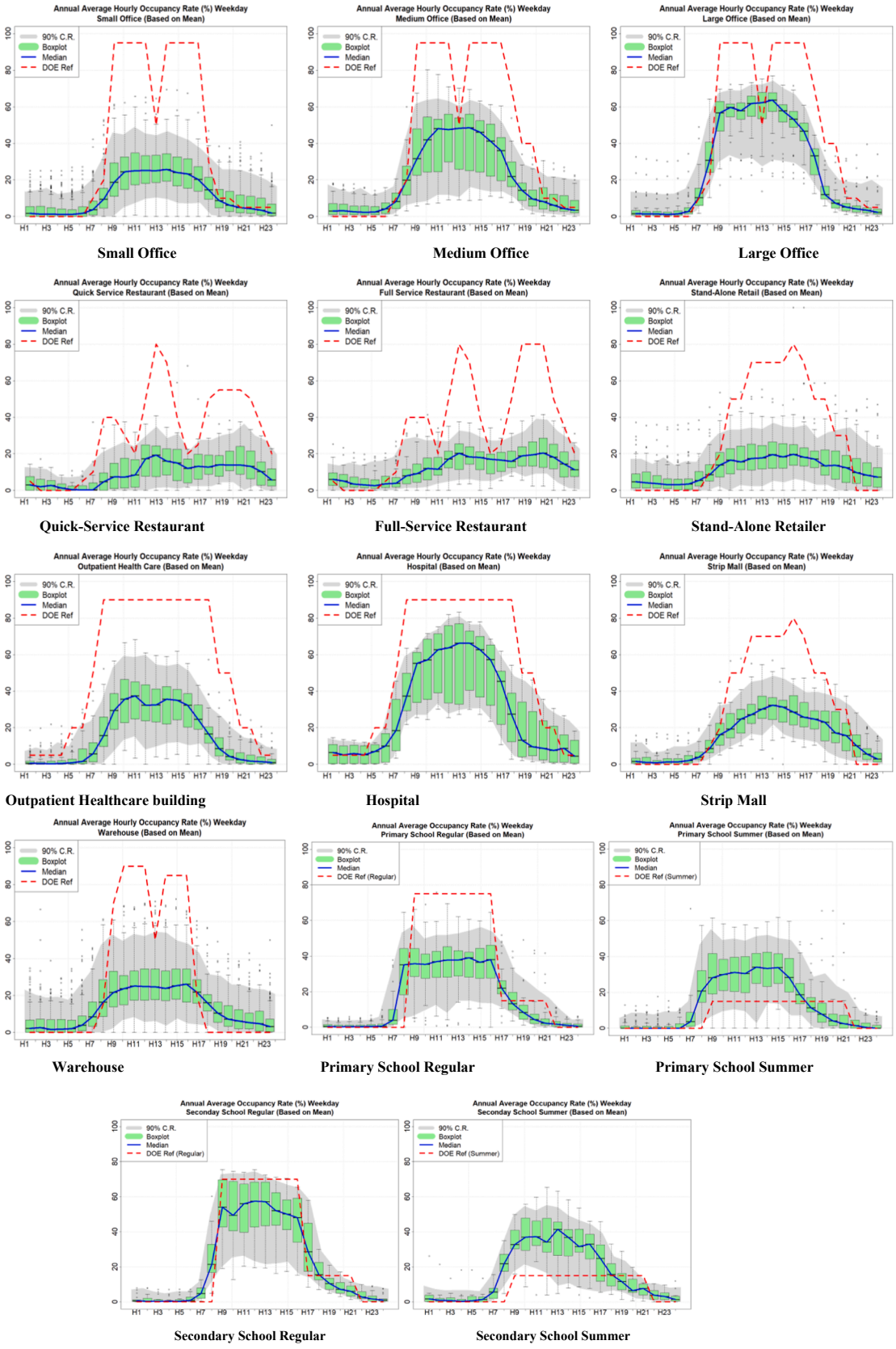

Fig. 4. Empirical confidence bands for median occupancy profiles in each Department of Energy building type on weekdays.

Table 2
Comparison of hourly occupancy rates (%) for selected hours and the overall daily average on weekdays for different building types.

Department of energy building type	7AM	8AM	9AM	6PM	7PM	8PM	All Day
Medium Office	21.33	34.69	39.44	14.89	11.10	9.97	**20.32**
Stand-Alone Retailer	10.33	14.94	16.81	15.56	14.87	13.40	**12.78**
Primary School Summer	23.64	32.08	30.98	9.41	6.52	5.25	**15.05**
Primary School Regular	33.83	37.40	36.17	9.75	5.90	3.88	**17.20**
Small Hotel	24.79	24.73	23.13	23.49	22.64	24.14	**21.77**
Strip Mall	9.92	17.17	21.23	22.15	19.86	16.02	**15.63**
Outpatient Healthcare Center	17.16	29.23	35.50	10.23	6.32	4.53	**15.68**
Warehouse	19.56	23.88	24.75	12.13	9.96	8.69	**15.12**
Large Office	32.03	53.77	56.38	14.24	8.59	6.78	**26.68**
Hospital	39.64	49.53	53.35	19.59	16.25	12.78	**28.53**
Small Office	12.14	20.82	24.23	10.71	8.96	8.11	**13.38**
Secondary School Summer	22.39	34.68	38.62	13.40	9.09	6.94	**16.95**
Secondary School Regular	26.78	55.00	53.34	11.89	8.37	6.02	**22.97**
Large Hotel	33.03	35.91	35.38	34.23	30.37	30.34	**29.51**
Supermarket	11.20	29.21	42.28	35.77	28.33	24.02	**25.49**
Full-Service Restaurant	9.00	10.69	12.66	19.42	20.90	21.07	**13.05**
Midrise Apartment	24.00	26.84	27.49	25.10	21.88	22.03	**21.05**
Quick-Service Restaurant	6.43	9.57	10.45	14.58	16.27	16.92	**10.26**

3.2. Data processing

Building data. We processed building data from San Antonio, Texas. First, the addresses, main use, and the construction year of 998 buildings were collected from Bexar County Appraisal District (Fig. 2a). The addresses were geocoded to their geographical coordinates using ArcGIS [53]. Geolocations were added to building footprints generated from ENVI [54]. Building heights were extracted from LiDAR open-source data [55] and added to building footprints. The ArcGIS Polygon-to-Point tool extracted the vertices of the building footprints. Finally, the studied buildings were categorized based on DOE references for commercial buildings. Fig. 2b shows the number distribution of building types.

Mobility data. This study processed human mobility data, collected from smartphone devices, from Cuebiq Inc. Cuebiq collaborates with over 100 smartphone apps that provide location-based services and thus is able to gather geolocation data when users utilize these apps. The data set is collected from approximately 20 percent of the population in San Antonio. It is from January 1 to September 31, 2017, and comprised anonymized device ID, latitude, longitude, and corresponding time (in seconds) data. The latitude and longitude from each entry of the data set allow us to spatially join the coordinate to the geometry of buildings and thus estimate occupancies. For better visualization and interpretation, we converted the time format from Unix timestamp to standard "HH:MM:SS" format. This dataset has also been used to study urban mobility and accessibility [56], social connectivity, and commuting and travel patterns [57,58]. The results from this dataset have shown high robustness [57–59] compared to those from other datasets collected from more traditional venues despite certain limitations [60].

3.3. Deriving building occupancy profiles

Obtaining individual building occupancy rate and profile. If the number of unique users at each hour $t \in \{1, 2, ..., 24\}$ in a building i on a given day j is C_{ijt}, we observed C_{ijt} by counting the distinct users appearing in the interior of each building using the latitude and longitude of users and building shape coordinates. To obtain the corresponding occupancy rate R_{ijt}, we divided C_{ijt} by the corresponding capacity $C_i^{\max}$, which was estimated using the maximum observed counts $C_i^{\max}$, which is the maximum occupancy count over a specific week. We estimated the building capacity using the weekly maximum count to incorporate changes in the total population tracked by mobility data. Hence, the hourly occupancy rates (percentages based on building capacity) could be obtained as follows:

$$R_{ijt} = \frac{C_{ijt}}{C_i^{\max}} \tag{1}$$

We constructed an occupancy profile for each building by obtaining a vector of average hourly occupancy rate: $\overline{\mathbf{R}}_i = (\overline{R}_{i,1}, \overline{R}_{i,2}, \cdots, \overline{R}_{i,24})'$, where $\overline{R}_{i,t} = \frac{1}{n_i}\sum_{j=1}^{n_i} R_{ijt}$, and n_i is the total number of days on which valid data were collected for building i. To comply with the DOE convention and make appropriate comparisons, we obtained $\overline{\mathbf{R}}_i$ for weekdays, Saturdays, and other days separately.

Building occupancy profile for each DOE type. Based on the individual building occupancy profile $\overline{\mathbf{R}}_i$, we could obtain the aggregated occupancy profile for each DOE type. As mentioned in the previous section, we collected individual building-level data of DOE types. If Ω_k is the building index set for all buildings belonging to DOE type k, we computed the average occupancy profile for DOE type k as follows:

$$\overline{\mathbf{R}}^k = \frac{1}{N_k}\sum_{i\in\Omega_k}\overline{\mathbf{R}}_i \tag{2}$$

where N_k is the total number of buildings belonging to DOE type k.

Statistical quantification of building occupancy profile. The mean vectors of hourly occupancy rates served as good representations of building occupancy profiles, and we further quantified the variations in occupancy rates by constructing a confidence region (band) based on daily observations at the individual building level. Specifically, for each $\overline{R}_{i,t}$ in $\overline{\mathbf{R}}_i$, we computed the lower and upper bounds of the confidence limits as $\overline{R}_{i,t} \pm 2 \times \text{stdev}_{it}(R_{ijt})$, where $\text{stdev}_{it}(R_{ijt})$ the is the standard deviation of R_{ijt}.

To quantify the deviation of the occupancy profile from the DOE reference occupancy profiles at the individual building level, we proposed a discrepancy score computed as follows:

$$\text{DS}_i = \frac{1}{24}\sum_{t=1}^{24} \max(0, D_{it} - U_{it}) + \min(0, L_{it} - D_{it}) \tag{3}$$

where D_{it} is the DOE reference occupancy rate at hour t, and $L_{it} = \overline{R}_{i,t} - 2 \times \text{stdev}_{it}(R_{ijt})$ and $U_{it} = \overline{R}_{i,t} + 2 \times \text{stdev}_{it}(R_{ijt})$ are the lower and upper empirical confidence limits at hour t, respectively. The interpretation of DS_i as the average hourly occupancy profile was significantly different ($p \le 0.05$) from the empirical occupancy profile of the DOE reference based on the empirical confidence band. Similarly, we computed the lower and upper limits for $\overline{\mathbf{R}}^k$ at the aggregated level, which we denoted as $\mathbf{L}^k = (L_1^k, L_2^k, .., L_{24}^k,)'$ and $\mathbf{U}^k = (U_1^k, U_2^k, .., U_{24}^k,)'$, respectively.

For each aggregated DOE type, we performed formal statistical hypothesis testing to determine whether the occupancy profile, i.e., the

Table 3
Summary statistics of the discrepancy scores for different building types on Weekday, Saturday, and other days.

	Weekday				Saturday				Other			
	Mean	St.dev.	Min.	Max.	Mean	St.dev.	Min.	Max.	Mean	St.dev.	Min.	Max.
Medium Office	5.81%	4.09%	1.15%	22.42%	5.31%	5.22%	0.00%	30.42%	0.18%	0.78%	0.00%	4.90%
Stand-Alone Retail	3.84%	5.06%	0.00%	25.83%	11.15%	8.76%	0.10%	39.58%	3.85%	5.75%	0.00%	35.00%
Primary School Summer	3.01%	4.87%	0.00%	27.29%	3.45%	5.67%	0.00%	27.08%	2.01%	2.97%	0.00%	12.50%
Primary School Regular	1.52%	3.62%	0.00%	23.61%	0.60%	1.80%	0.00%	8.33%	1.61%	3.77%	0.00%	16.67%
Small Hotel	10.54%	7.19%	0.00%	42.92%	7.22%	7.97%	0.28%	41.67%	6.25%	9.14%	0.00%	55.00%
Strip Mall	1.40%	1.95%	0.00%	7.75%	6.45%	5.53%	0.00%	22.08%	1.91%	3.31%	0.00%	15.00%
Outpatient Health Care	10.60%	5.36%	0.00%	25.42%	6.06%	4.96%	0.00%	18.96%	1.48%	3.13%	0.00%	15.49%
Warehouse	2.88%	4.82%	0.00%	33.13%	2.22%	3.89%	0.00%	27.50%	1.66%	5.63%	0.00%	60.42%
Large Office	3.79%	1.44%	0.83%	8.71%	7.23%	2.29%	0.52%	10.84%	0.24%	0.96%	0.00%	5.68%
Hospital	7.15%	4.76%	2.62%	14.20%	3.96%	4.00%	0.00%	9.86%	3.24%	3.79%	0.00%	10.81%
Small Office	4.59%	5.10%	0.00%	26.53%	3.60%	4.06%	0.00%	20.42%	1.17%	3.40%	0.00%	25.00%
Secondary School Summer	2.10%	2.91%	0.00%	12.08%	1.75%	4.02%	0.00%	15.97%	0.64%	1.71%	0.00%	6.46%
Secondary School Regular	0.38%	0.38%	0.00%	1.04%	0.18%	0.44%	0.00%	1.39%	0.78%	2.19%	0.00%	6.94%
Large Hotel	14.31%	5.46%	5.20%	27.64%	8.90%	5.63%	2.26%	27.85%	7.02%	6.89%	0.47%	26.62%
Full Service Restaurant	4.37%	3.44%	0.00%	15.66%	5.52%	4.41%	0.06%	20.07%	5.39%	4.30%	0.00%	14.90%
Midrise Apartment	23.82%	10.36%	2.08%	43.45%	14.24%	5.87%	4.86%	26.10%	1.65%	4.18%	0.00%	17.39%
Quick Service Restaurant	6.94%	5.70%	0.21%	18.96%	8.85%	6.38%	0.42%	23.96%	11.34%	8.57%	2.19%	31.46%

Table 4
Average discrepancy scores for different building types during work hours.

	Weekday		Saturday		Other	
DOEType	All Day	Work Hour	All Day	Work Hour	All Day	Work Hour
Medium Office	13.73%	23.47%	6.20%	10.49%	0.93%	1.37%
Stand-Alone Retail	12.21%	20.94%	11.20%	19.20%	0.38%	0.65%
Primary School Summer	0.09%	0.16%	0.00%	0.00%	0.00%	0.00%
Primary School Regular	7.39%	12.65%	0.00%	0.00%	0.00%	0.00%
Small Hotel	24.31%	6.20%	14.62%	1.94%	14.83%	5.25%
Strip Mall	8.86%	14.97%	4.77%	8.18%	0.00%	0.00%
Outpatient Health Care	22.56%	35.61%	0.00%	0.00%	0.00%	0.00%
Warehouse	9.94%	17.04%	0.00%	0.00%	0.00%	0.00%
Large Office	11.37%	19.37%	7.48%	12.76%	1.00%	1.64%
Hospital	11.13%	17.85%	0.00%	0.00%	0.30%	0.29%
Small Office	17.08%	29.29%	0.00%	0.00%	0.00%	0.00%
Secondary Summer	1.11%	1.87%	0.24%	0.41%	0.01%	0.01%
Secondary Regular	1.09%	1.74%	1.23%	2.11%	0.03%	0.05%
Large Hotel	24.10%	4.50%	12.15%	0.00%	14.83%	5.42%
Full-Service Restaurant	13.69%	21.71%	7.34%	11.87%	7.38%	12.03%
Midrise Apartment	35.15%	12.17%	20.95%	33.50%	3.99%	4.73%
Quick Service Restaurant	11.35%	17.45%	5.65%	8.88%	7.45%	9.54%

mean vector of hourly occupancy rates, was equivalent to the DOE reference given the observed data. We relied on Hotelling's *t*-squared statistic [10], which is an extension of Student's *t*-test in multivariate cases, to test the following hypothesis:

$$H_0 : \boldsymbol{\mu}_k = \boldsymbol{\mu}_0 vs H_1 : \boldsymbol{\mu}_k \neq \boldsymbol{\mu}_0 \tag{4}$$

where $\boldsymbol{\mu}_k$ is the true building occupancy profile for DOE type k, and $\boldsymbol{\mu}_0$ is the DOE reference corresponding to the specific building type. We calculated Hotelling's *t*-squared statistic as follows:

$$T^2 = \left(\overline{\mathbf{R}}^k - \boldsymbol{\mu}_0\right)' \hat{\Sigma}_{\overline{\mathbf{R}}^k}^{-1} \left(\overline{\mathbf{R}}^k - \boldsymbol{\mu}_0\right) \tag{5}$$

where

$$\hat{\Sigma}_{\overline{\mathbf{R}}^k} = \frac{1}{N_k(N_k-1)} \sum_{i \in \Omega_k} \left(\overline{\mathbf{R}}_i - \overline{\mathbf{R}}^k\right)\left(\overline{\mathbf{R}}_i - \overline{\mathbf{R}}^k\right)', \tag{6}$$

is the sample covariance matrix of $\overline{\mathbf{R}}^k$. Furthermore, it is known that

$$\frac{n_i - 24}{24(n_i - 1)} T^2 \; F_{p,N_k-p} \tag{7}$$

where F_{p,N_k-p} is the F distribution with numerator degrees of freedom p and denominator degrees of freedom $N_k - p$. A larger T^2 value provides more evidence against H_0, i.e., a large T^2 value indicates a higher probability that $\boldsymbol{\mu}_k$ is different from $\boldsymbol{\mu}_0$.

3.4. Analyze building energy consumption

Building occupancy behavior plays a key role in determining energy consumption in the built environment. On the one hand, it determines when building systems (e.g., lighting and heating, ventilation, and air-conditioning (HVAC) systems) need to be switched on. On the other hand, it partially defines the load of these systems, e.g., ventilation rate and temperature setpoints. Numerous studies have shown that occupants' behavior can help save up to 50% energy for single-occupancy offices [61], 23.5% energy consumption of central air-conditioning systems in campus buildings [62], and up to 41% HVAC energy for office buildings [63]. In addition, Model Predictive Control framework incorporating occupancy prediction model is often used to optimize HVAC controls [64,65,66]. In this study, we simply implemented temperature set-point changes based on the approximate percentage range of occupancy profiles.

We have chosen CityBES with a web-based platform, developed by Lawrence Berkeley National Laboratory, as our urban building energy modeling tool [65,66]. We have selected 359 out of 998 buildings due to the limitation of CityBES in simulating building types. CityBES uses the Commercial Building Energy Saver Toolkit [67,68], which builds on OpenStudio and EnergyPlus to provide energy retrofit analyses of

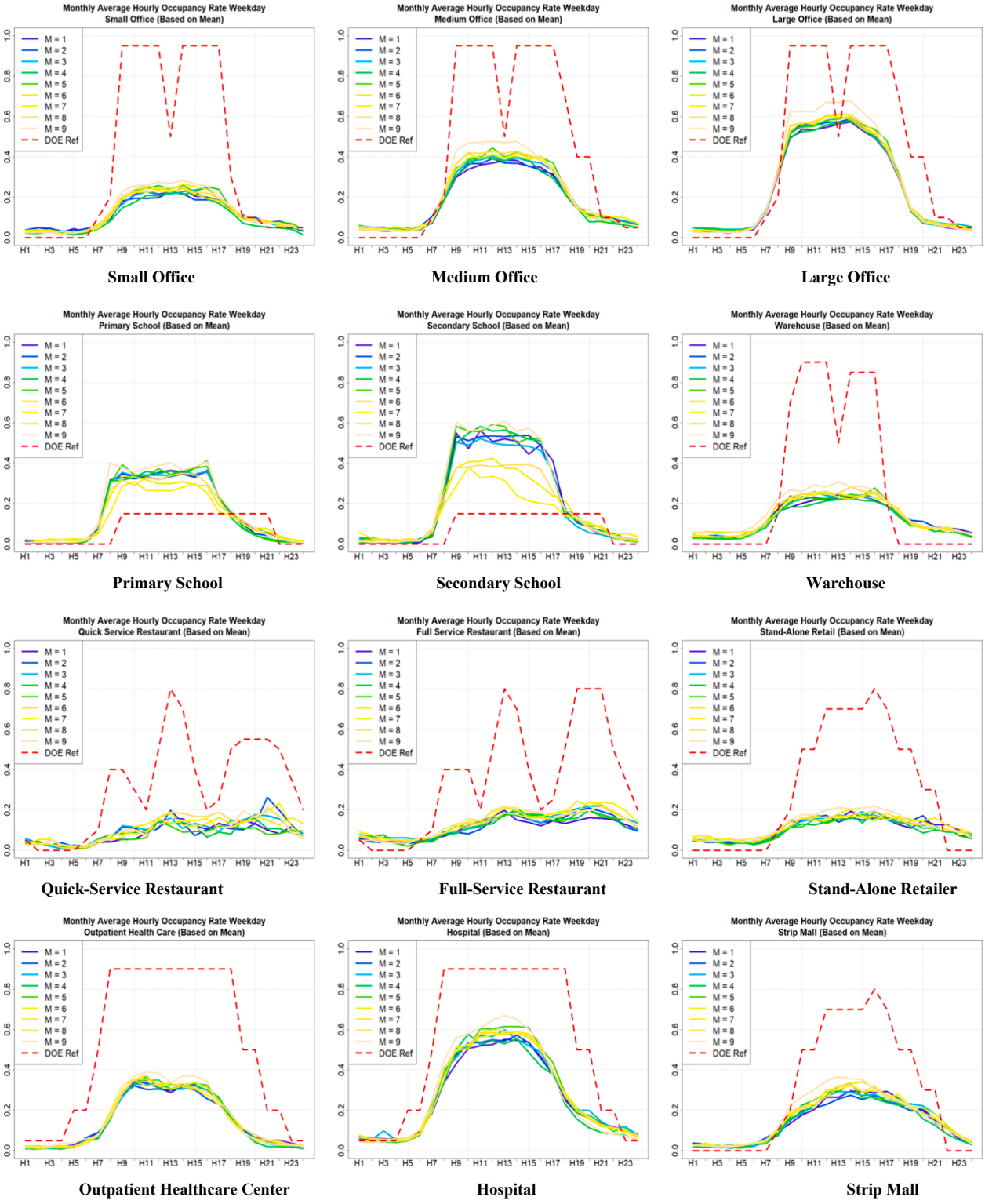

Fig. 5. Monthly average building occupancy profiles for each Department of Energy building type on weekdays.

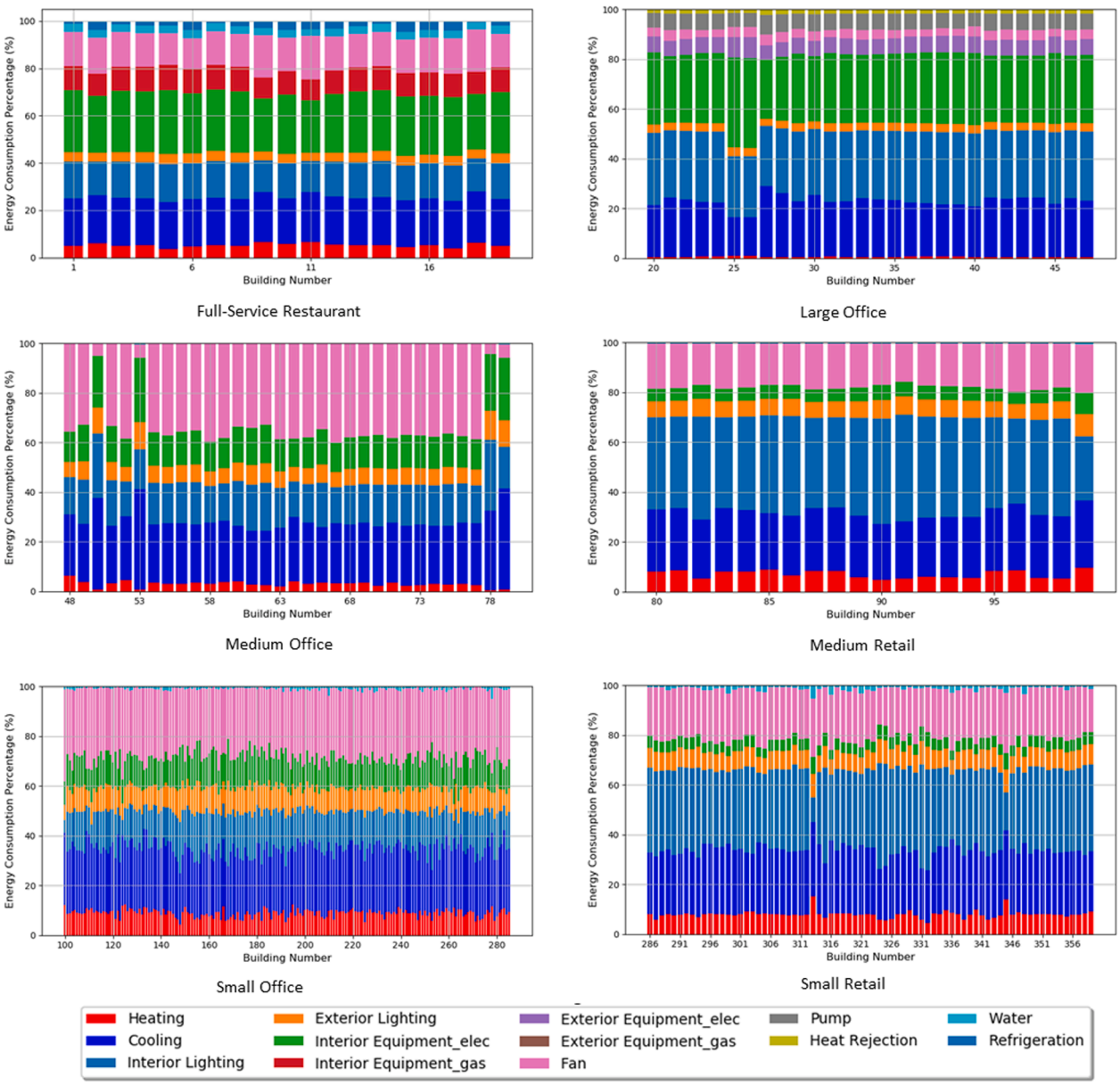

Fig. 6. Annual energy end-use percentage of all buildings with Department of Energy reference schedule.

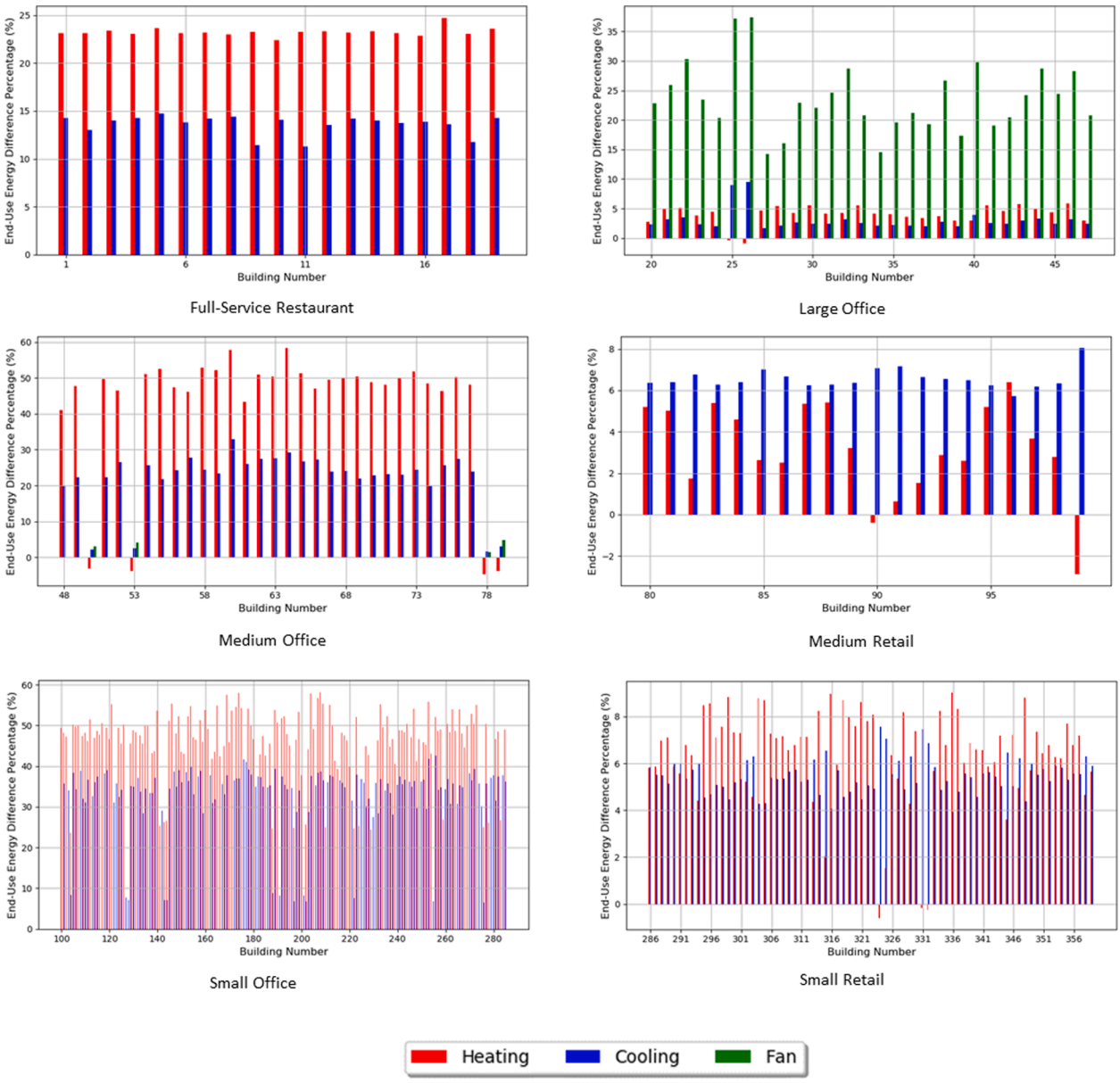

Fig. 7. Annual end-use energy difference of all the buildings on heating, cooling, and fans between the reference schedule and actual schedule.

individual commercial buildings in US cities. It used a prototype building database developed by the DOE [7] to generate EnergyPlus models for six building types in the study area, including small, medium, and large offices, full-service restaurants, and small and medium retailers. An occupancy profile based on the mean value for each building type was applied to these models and compared with the reference models (the basement schedule was not changed, Fig. 3). The HVAC temperature setpoint schedule was modified accordingly as shown in Table 1. One should note that the suffix '_ori' means the reference schedule and '_new' means the mean value from the measured data. A building was considered unoccupied if the occupancy rate was less than 0.1. Zonal control of the set-point was also applied to all office buildings (Table 1). For example, for small office buildings, the zones were sorted according to the ascending order of the floor area and grouped into four groups which representing around 20%, 5%, 5%, and 70% floor area; when the occupancy rate was below 0.1, the temperature set-point of all zones was set as set-back; when it was 0.1–0.2, the first group of zones were set to the normal set-point; when it was 0.2–0.25, the first two groups of zones were set to the normal set-point; when it was 0.25–0.3, the first three groups of zones were set to the normal set-point; and when it exceeded 0.3, all the groups of zones were set to the normal set-point. All these temperature set-points and set-back values were selected to be consistent with the DOE reference buildings.

Energy-differences. The energy differences on heating, cooling, fan and total energy are calculated by dividing energy difference through the actual schedule with end-use energy consumption in the reference model:

$$Energy\ \ Differences = \frac{\left(E^i_{ref} - E^i_{mod}\right)}{E^i_{ref}} \times 100\% \tag{8}$$

where E^i_{ref} and E^i_{mod} are the energy consumption by end-use i (heating, cooling, or fan) in the reference model or actual schedule, respectively.

4. Results and discussions

In this section, we present our empirical findings regarding the comparison of DOE-reference occupancy profiles to the empirically

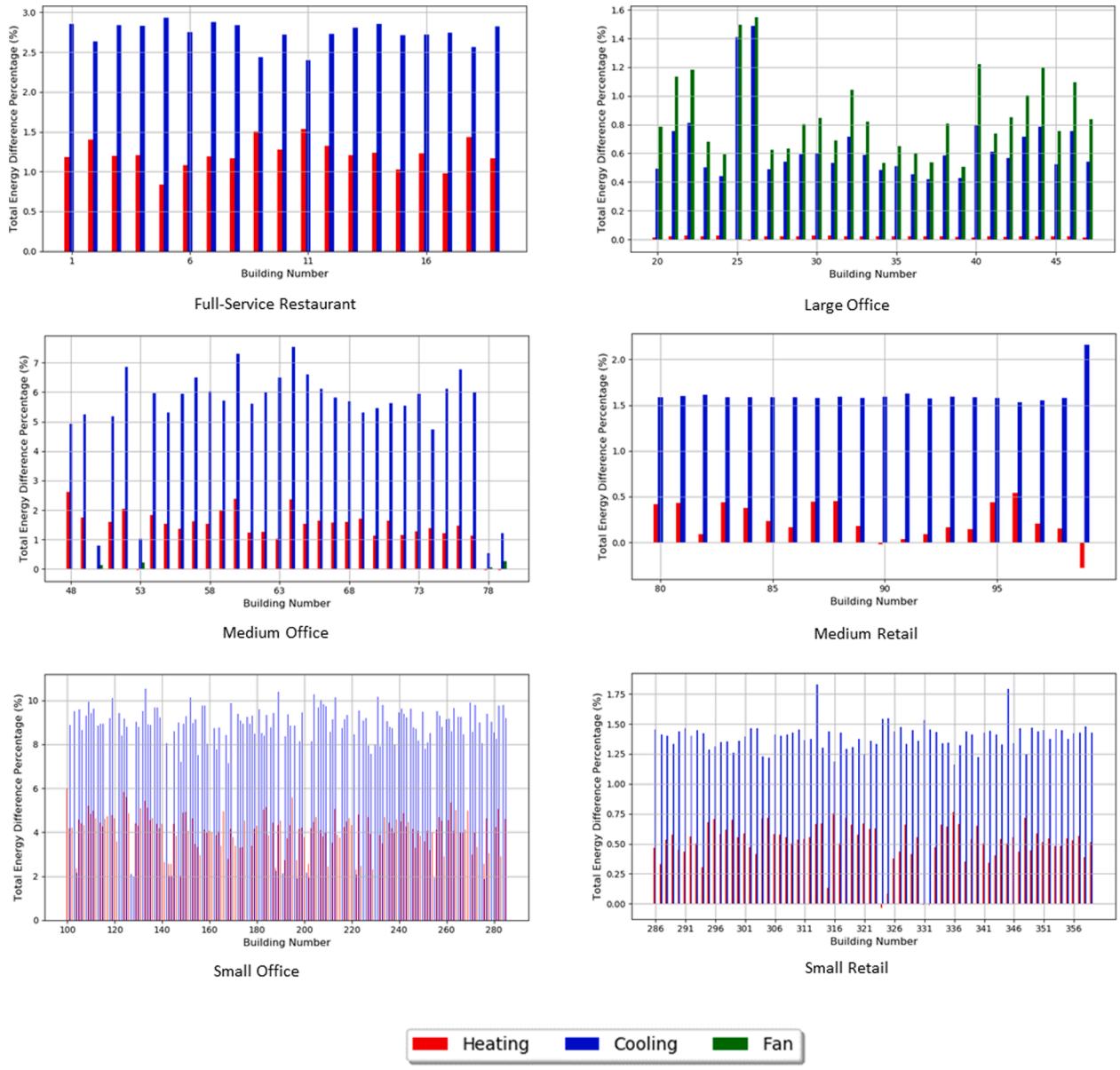

Fig. 8. Annual total energy difference of all the buildings by heating, cooling, and fans between the reference schedule and actual schedule.

observed occupancy profiles based on mobility data. We focused on the following DOE types: small office buildings, medium office buildings, large office buildings primary schools, secondary schools, warehouses, quick-service restaurants, full-service restaurants, stand-alone retailers, outpatient healthcare centers, hospitals, and strip malls. We present the results for weekdays here and show the results for Saturdays and other days in the appendix. In the discussion of the findings for Saturdays and other days, we refer to the relevant section in the appendix.

4.1. Empirical occupancy profiles

Fig. 4 presents the 90% confidence band for the hourly occupancy profiles of different DOE building types. The boxplots represent the hourly variations between different buildings of the same type. Except for some serviced-based building types (e.g., quick-service restaurants, full-service restaurants, and stand-alone retailers), the variations were generally larger during peak hours than during other times. Medium offices and hospitals had the largest hourly variations throughout the day, whereas large offices and strip malls had the smallest variations. The confidence bands enabled the comparison of empirical occupancy profiles to those provided by the DOE references. A DOE reference (red dashed line) outside the confidence band indicated that the empirical occupancy profile patterns were misaligned with those from the DOE references. We observed that for most building types, the DOE references overestimated the occupancy profile during the day, with the exception of schools, where the opposite was observed. These differences significantly impacted energy simulations, as demonstrated in Section 6.

To further assess the differences between the empirical occupancy profiles and DOE references, we used Hotelling's *t*-test (described in Section 4.3) to verify the statistical significance. All the differences were significant with low p-values. Hotelling's *t*-test tested the overall significance of the differences between the daily profiles and DOE references. Hence, for a particular building type, significant differences from the DOE reference may have occurred at some hours, whereas insignificant differences occurred at others. To demonstrate the hourly differences, we summarized the hourly occupancy profiles for the selected hours in Table 2 and compared them with the average daytime hourly occupancy rates. The red and blue values respectively indicate occupancy rates that were significantly higher and lower ($p \leq 0.05$) than the

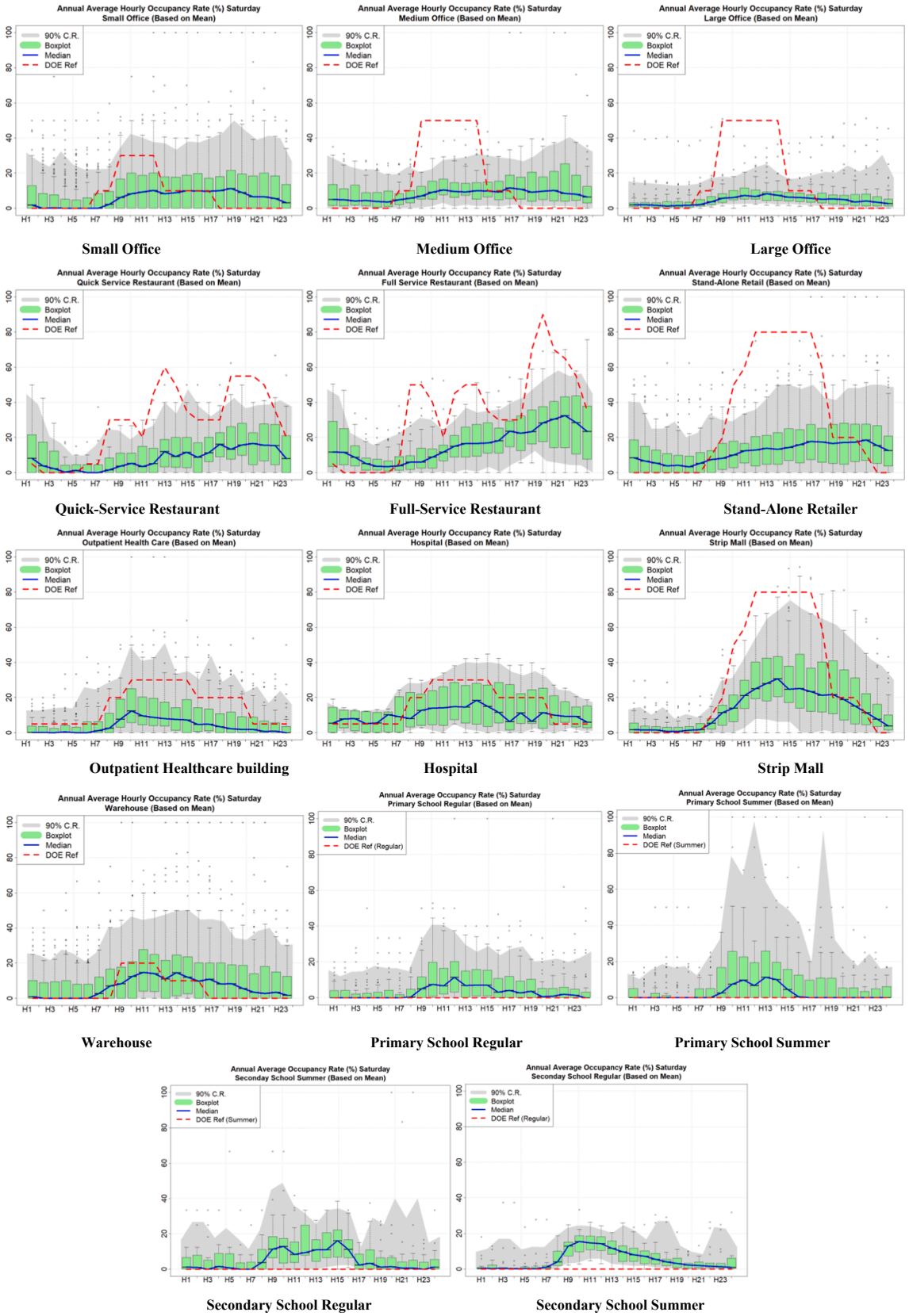

Fig. A1. Empirical confidence bands for median occupancy profiles in each Department of Energy (DoE) type on Saturdays.

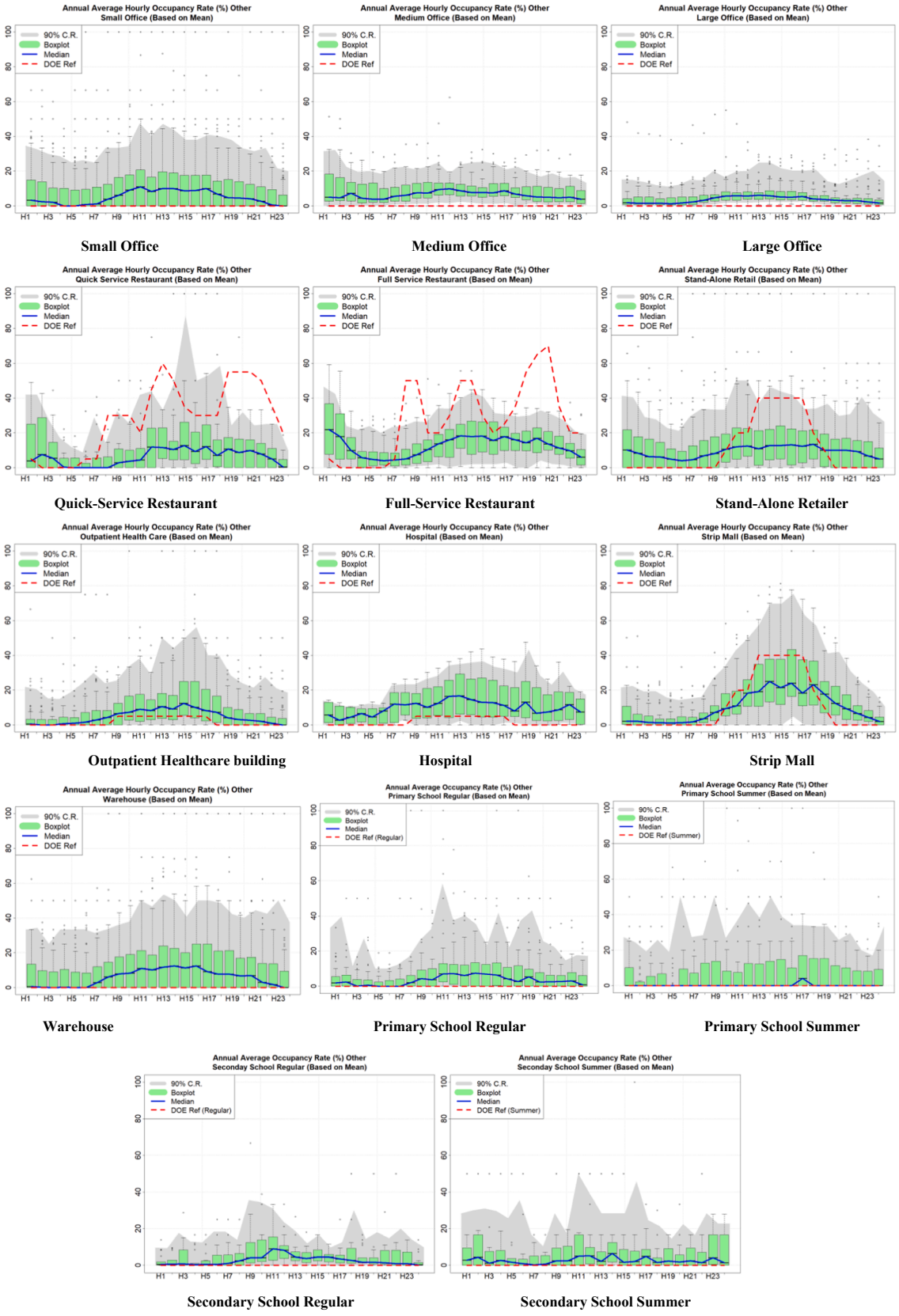

Fig. A2. Empirical confidence bands for median occupancy profiles in each Department of Energy (DoE) type on other days.

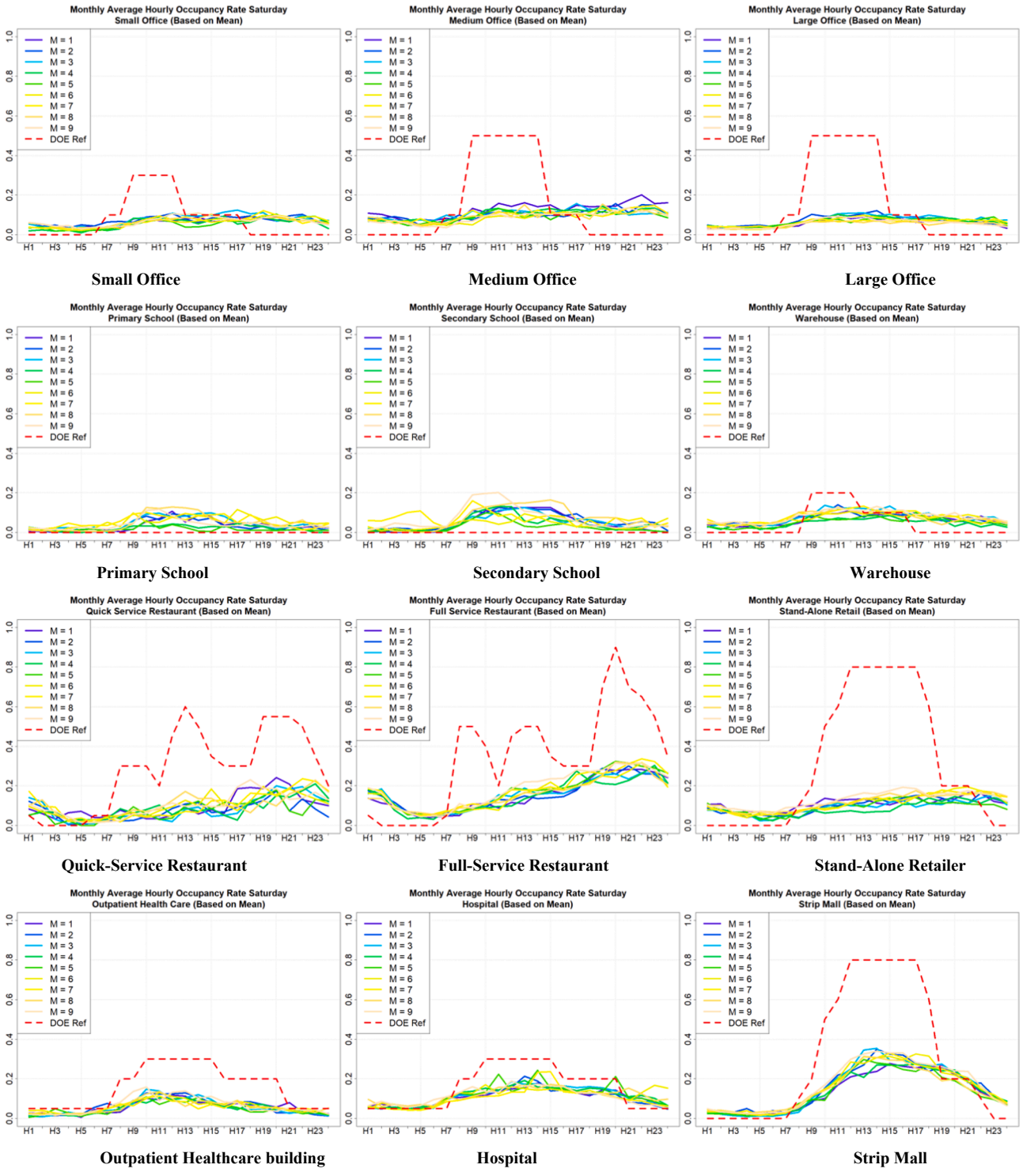

Fig. A3. Monthly average building occupancy profiles within each DOE type for Saturdays.

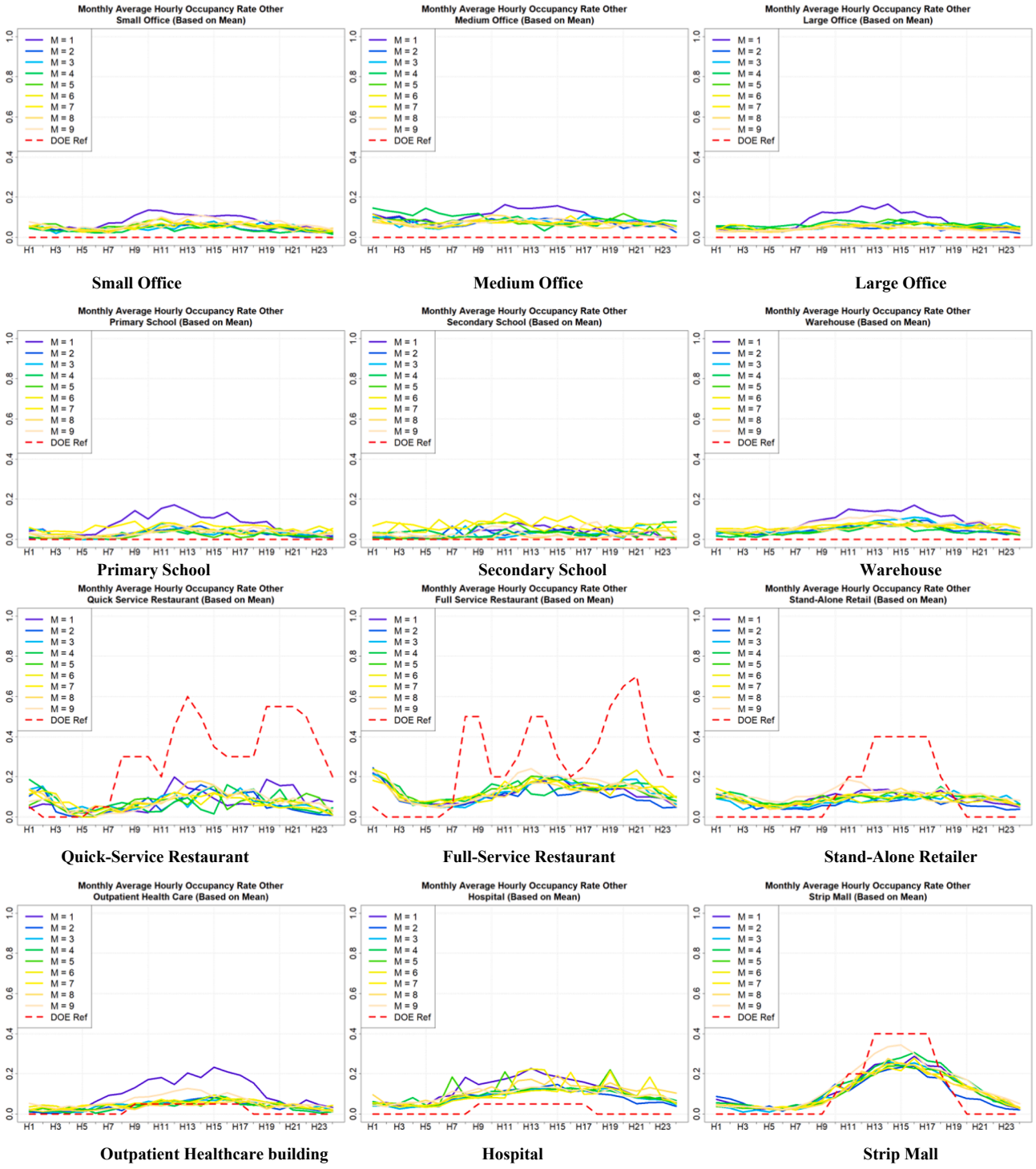

Fig. A4. Monthly average building occupancy profiles within each DOE type for other days.

daily average. The nonsignificant differences are denoted in gray. For offices and schools, the occupancy rates increased between 7 and 9 AM, when people started arriving for work and classes. In contrast, occupancy rates decreased between 6 and 8 PM when people started leaving for home. For service-related building types, such as restaurants, strip malls, and hotels, the opposite was observed, where occupancy rates increased during evening hours. For office buildings, the DOE reference models used the same occupancy profiles, whereas the actual schedules of the small, medium and large offices differed notably. The occupancy rates of the three office sizes during office hours were 25, 50, and 60%. A similar observation was made for healthcare buildings (outpatient healthcare centers and hospitals) and school buildings (primary and secondary schools). Hospitals and secondary schools had higher office-hour occupancy rates than did outpatient healthcare centers (60 vs. 40%) and primary schools (50 vs. 40%), respectively. However, similar occupancy rates or only slight differences were found for restaurants and retail buildings.

Table 3 summarizes the average discrepancy scores defined in Section 4.2 for each DOE reference building type. The discrepancy scores reflected the significant deviation from the empirical occupancy profiles to the DOE reference models. Notably, the large deviations (over 10% average hourly deviation) observed during weekdays were for midrise

apartments, large hotels, small hotels, and outpatient healthcare centers. On Saturdays, the large deviations were for midrise apartments and stand-alone retailers. On other days, only quick-service restaurants exhibited a large deviation. The deviations for large offices were much higher on Saturdays than on weekdays.

Since most of the divergence from the DOE references occur during work hours (7 AM to 9 PM), we recalculated the average discrepancy scores between 7 AM to 9 PM and report the results in Table 4. Comparing these results to Table 3, if the average discrepancy scores are higher in this table, it means that that the variations from the DOE references are more severe during the work hours. Among all the reported building types, most of them exhibit this phenomenal except for Primary School, where the differences are similar.

4.2. Monthly patterns of occupancy profiles

Both occupant behavior and energy usage vary from one month to another. We, therefore, assessed the monthly patterns of the occupancy profiles. Fig. 5 presents the monthly (average) occupancy profiles for each DOE reference building type. For primary and secondary schools, an expected drop in occupancy rates during summer months occurred. The variations between the occupancy profiles in different months were small for most building types. While the overall shape of the profiles appeared to be similar to those of the DOE references for most building types, no significant drop occurred at noon in the different office types, and the occupancy profiles of schools consistently shifted.

4.3. Impacts on building energy usage

Energy end-use with reference schedule. The energy end-use percentage of all buildings in the reference schedule is shown in Fig. 6. Since San Antonio is classified as Climate Zone 2A (hot and humid), cooling made up a large part of the total energy consumption (20–30%), whereas heating made up less than 5%. Lighting also consumed a large part of the total energy, with interior and exterior lighting constituting 20–40% of the total energy. In small offices, medium offices, full-service restaurants, small retailers, and medium retailers' fans consumed a large part (20–40%) of the total energy, but large offices spent less than 5% energy on fans. This was because these offices used chillers with multizone variable air volume systems, whereas the other buildings used a packaged air-conditioner unit with a constant air volume system. HVAC systems used more than 50% of the total energy in small and medium office buildings.

***Energy differences*:** HVAC end-use energy differences are described in Fig. 7. Among the office buildings, the most heating (~55%) and cooling (~40%) energy was reduced for small offices, but no energy was reduced on fans, because they use constant air volume systems, and fans constantly run at the same speed. There was slightly less energy reduced for medium offices than small offices (~50% for heating and ~30% for cooling), and even less (~5% for both heating and cooling) for large offices. This was because large offices had measured occupancy profiles, which most closely resembled the reference schedule: The peak occupancy rate on weekdays was 64%, whereas those for small and medium offices were 26 and 48%, respectively. In addition, around 25% fan energy was reduced for large offices compared to 0% for small and medium offices since large offices used variable air volume systems. For the other three building types, full-service restaurants and medium and small retailers, a similar amount of heating (−2 to 25%) and cooling (4–15%) energy, but 0% fan energy was reduced. This was also because these buildings used a constant air volume system.

In general, most total energy differences came from cooling (0.5–10%). Although 5–60% heating energy was reduced in most cases (Fig. 7), the total energy difference by heating was less than 5%, because San Antonio is categorized as Climate Zone 2A and thus needs little heating during the year (Fig. 8). In addition, heat generation from occupants also contributes to space heating, and the model with a modified schedule with a lower occupancy rate may need more heating in occupied zones. The most energy was reduced for small offices (~15%), followed by medium offices (~8%). The total energy difference for large offices dropped to around 1.5%, half of which was via cooling and half via fans. The reason for the energy difference by fans only existing in large buildings is explained above. Energy difference by heating was almost negligible because large office buildings need much less heating than other building types do because of their smaller surface/volume ratio and higher internal heat gain (by occupancy, internal lighting, and equipment). The total energy difference by the other three building types was relatively consistent (2–3.5%), with around one-third being contributed by heating and two-thirds by cooling.

5. Conclusions

In this paper, we utilize a rich urban scale mobile positioning data to introduce a new approach to capture empirical occupancy profiles for different Department of Energy building types in a city. The empirical occupancy profiles provide deeper insights regarding more realistic occupancy profiles compared to the Department of Energy prototype models. Visualization and statistical testing revealed significant differences between the empirical occupancy profile and that of almost all Department of Energy references except for the Primary School building. Department of Energy prototype models overestimated the occupancy rate during the daytime for most building types, except for schools, for which the occupancy rate was notably underestimated. For the same Department of Energy building type, the confidence band of the empirical occupancy profiles indicates that there are some variations that exist among the buildings belong to the same type. Our monthly occupancy profile plots demonstrate that there are some variations among different months to a certain degree, especially for secondary schools where the differences are most significant. That is the occupancy rate for the secondary schools that appear to be significantly lower in summer months than the winter months. The average discrepancy scores, which can be interpreted as the significant average hourly deviation from the Department of Energy prototype, are as much as 23.82% for weekdays, as much as 14.24% for Saturdays, and as much as 11.34% for other days.

The effect of the more realistic occupancy profiles on urban-scale building energy performance was investigated via an urban-scale building energy modeling platform. A comparison between models with derived and reference schedules demonstrated that (1) after using the derived schedule, there are up to 60% heating energy and 40% cooling energy differences among all the building types, (2) for restaurant and retail buildings, there are up to 25% heating energy and15% cooling energy differences, (3) small office buildings have more energy differences (15%) than medium (8%) or large ones (1.5%), because a larger schedule discrepancy was found in smaller offices, and (4) less than 5% in whole building energy differences were observed for buildings, except for small and medium offices, because these types of buildings consumed most energy (>50%) through HVAC systems. Those results can be applied in the real applications on: (1) Community and district level building design and operation; (2) Demand side management for a cluster of buildings; and (3) Grid interactive-efficient buildings.

The limitations for this study are: (1) Since this study only focused on the effect of occupancy profiles on HVAC systems, energy differences by lighting and plug loads due to reduced profile was not considered; (2) Only six building types were used in the selected district, and the effects of the measured occupancy profiles on other building types need to be studied in the future; (3) This study adopted deterministic occupancy profiles and ignored variations in occupancy profiles between different buildings and over time; (4) In this study, while analyzing the impacts of more representative occupancy profiles on buildings' energy usages, we assume that the building knows the rough occupancy rate in real-time. This is currently a challenge for an existing building. However, there

are on-going research projects that aim to address those challenges in both hardware and software levels [69]; (5) With the current approach, we observed that for Stand-alone retailers, the derived occupancy profile is much lower than Department of Energy references. This may be caused by the noises in the mobility data for this specific type of building. Further investigation is needed; (6) Urban scale energy analysis on peak heating or cooling demand, peak diversity and capacity factors, and load duration curves are out of the scope of this study, and will be analyzed in the future; (7) mobility data can suffer from representativeness issues caused by their socio-economic characteristics, meaning that more data can be generated from advanced (e.g., wealthy) neighborhoods and less from disadvantaged (e.g., poor) ones. Future research should, therefore, investigate and solve such potential biases. Moreover, we used all visited geolocations to indicate the occupancy presence, and future studies should apply more advanced algorithms to identify people's stays (i.e., visits) and time of occupancy.

CRediT authorship contribution statement

Wenbo Wu: Data curation, Formal analysis. **Bing Dong:** Conceptualization, Methodology, Supervision, Funding acquisition. **Qi (Ryan) Wang:** Data curation. **Meng Kong:** Software implementation, Data anaysis. **Da Yan:** Review, Resources. **Jingjing An:** Literature review. **Yapan Liu:** Literature review.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgments

To protect the confidentiality of any given individual's movement trajectory, all individuals' information from Cuebiq was encrypted, and all data are reported in nonidentifiable form. All data used in this paper were reviewed and exempted by the Northeastern University Institutional Review Board. The data from Cuebiq are proprietary and will not be shared. . This paper is part of the Annex 79 subtask 2.1 and 3.8 collaborative efforts, and is outcome of a research supported by the National Science Foundation (NSF) under Award Number: 1949372.

Appendix A

Figs. A1–A4.

References

[1] Department of Economic and Social Affairs of U.N., 2018 Revision of World Urbanization Prospects, (n.d.). https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html [accessed June 4, 2020].
[2] World Resource Institute, 4 Charts Explain Greenhouse Gas Emissions by Countries and Sectors, (n.d.). https://www.wri.org/blog/2020/02/greenhouse-gas-emissions-by-country-sector [accessed June 4, 2020].
[3] U.S. EPA, Indoor Air Quality | EPA's Report on the Environment (ROE), (n.d.). https://www.epa.gov/report-environment/indoor-air-quality [accessed June 4, 2020].
[4] U.S. Energy Information Administration (EIA), Total Energy Monthly Data, (n.d.). https://www.eia.gov/totalenergy/data/monthly/#consumption [accessed June 4, 2020].
[5] Yan D, O'Brien W, Hong T, Feng X, Burak Gunay H, Tahmasebi F, et al. Occupant behavior modeling for building performance simulation: current state and future challenges. Energy Build 2015;107:264–78. https://doi.org/10.1016/j.enbuild.2015.08.032.
[6] Cuebiq, Cuebiq, Cuebiq; 2018. https://www.cuebiq.com/. Technical report [accessed March 18, 2020].
[7] Deru M, Field K, Studer D, Benne K, Griffith B, Torcellini P et al. U.S. Department of Energy commercial reference building models of the national building stock, Publ. 2011:1–118. https://doi.org/NREL Report No. TP-5500-46861.
[8] Gaetani I, Hoes PJ, Hensen JLM. Occupant behavior in building energy simulation: towards a fit-for-purpose modeling strategy. Energy Build 2016;121:188–204. https://doi.org/10.1016/j.enbuild.2016.03.038.
[9] Happle G, Fonseca JA, Schlueter A. A review on occupant behavior in urban building energy models. Energy Build 2018;174:276–92. https://doi.org/10.1016/j.enbuild.2018.06.030.
[10] Hotelling H. In: The generalization of student's ratio. New York, NY: Springer; 1992. p. 54–65. https://doi.org/10.1007/978-1-4612-0919-5_4.
[11] Ren X, Yan D, Wang C. Air-conditioning usage conditional probability model for residential buildings. Build Environ 2014;81:172–82. https://doi.org/10.1016/j.buildenv.2014.06.022.
[12] Jiefan G, Peng X, Zhihong P, Yongbao C, Ying J, Zhe C. Extracting typical occupancy data of different buildings from mobile positioning data. Energy Build 2018;180:135–45. https://doi.org/10.1016/j.enbuild.2018.09.002.
[13] Yan D, Hong T, Dong B, Mahdavi A, D'Oca S, Gaetani I, et al. IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. Energy Build 2017;156:258–70. https://doi.org/10.1016/j.enbuild.2017.09.084.
[14] Kim J, Park J, Lee W. Why do people move? Enhancing human mobility prediction using local functions based on public records and SNS data. PLoS One 2018;13. https://doi.org/10.1371/journal.pone.0192698.
[15] Li Q, Jige Quan S, Pei-Ju Yang P. Building Energy Modelling at Urban Scale: Integration of Reduced Order Energy Model with Geographical Information; 2016. https://www.researchgate.net/publication/293958595 [accessed March 18, 2020].
[16] Kontokosta CE, Tull C. A data-driven predictive model of city-scale energy use in buildings. Appl Energy 2017;197:303–17. https://doi.org/10.1016/j.apenergy.2017.04.005.
[17] Abbasabadi N, Ashayeri M, Azari R, Stephens B, Heidarinejad M. An integrated data-driven framework for urban energy use modeling (UEUM). Appl Energy 2019. https://doi.org/10.1016/j.apenergy.2019.113550.
[18] Nutkiewicz A, Yang Z, Jain RK. Data-driven Urban Energy Simulation (DUE-S): Integrating machine learning into an urban building energy simulation workflow. Energy Proc, Elsevier Ltd 2017:2114–9. https://doi.org/10.1016/j.egypro.2017.12.614.
[19] Hu S, Yan D, Qian M. Using bottom-up model to analyze cooling energy consumption in China's urban residential building. Energy Build 2019;202. https://doi.org/10.1016/j.enbuild.2019.109352.
[20] Alhamwi A, Medjroubi W, Vogt T, Agert C. Modelling urban energy requirements using open source data and models. Appl Energy 2018;231:1100–8. https://doi.org/10.1016/j.apenergy.2018.09.164.
[21] Nutkiewicz A, Yang Z, Jain R, Jain R. Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. Appl Energy 2018;225(2018):1176–89.
[22] Sola A, Corchero C, Salom J, Sanmarti M. Simulation tools to build urban-scale energy models: a review. Energies 2018;11:3269. https://doi.org/10.3390/en11123269.
[23] Cuerda E, Guerra-Santin O, Sendra JJ, Javier F, González N. Comparing the impact of presence patterns on energy demand in residential buildings using measured data and simulation models Article History; n.d. https://doi.org/10.1007/s12273-019-0539-z.
[24] Muroni A, Gaetani I, Hoes P-J, Hensen JLM. Occupant behavior in identical residential buildings: A case study for occupancy profiles extraction and application to building performance simulation Keywords Article History; n.d. https://doi.org/10.1007/s12273-019-0573-x.
[25] Ahas R, Silm S, Järv O, Saluveer E, Tiru M. Using mobile positioning data to model locations meaningful to users of mobile phones. J Urban Technol 2010;17:3–27. https://doi.org/10.1080/10630731003597306.
[26] Zou H, Zhou Y, Jiang H, Chien SC, Xie L, Spanos CJ. WinLight: A WiFi-based occupancy-driven lighting control system for smart building. Energy Build 2018;158:924–38. https://doi.org/10.1016/j.enbuild.2017.09.001.
[27] Pang Z, Xu P, O'Neill Z, Gu J, Qiu S, Lu X, et al. Application of mobile positioning occupancy data for building energy simulation: An engineering case study. Build Environ 2018;141:1–15. https://doi.org/10.1016/j.buildenv.2018.05.030.
[28] Mohammadi N, Taylor JE. Urban infrastructure-mobility energy flux. Energy 2017;140:716–28. https://doi.org/10.1016/j.energy.2017.05.189.
[29] Riascos AP, Mateos JL. Emergence of encounter networks due to human mobility. PLoS ONE 2017;12. https://doi.org/10.1371/journal.pone.0184532.
[30] Wang Q, Phillips NE, Small ML, Sampson RJ. Urban mobility and neighborhood isolation in America's 50 largest cities. Proc Natl Acad Sci USA 2018;115:7735–40. https://doi.org/10.1073/pnas.1802537115.
[31] Hu S, Yan D, Dong B, Fu J. Exploring key factors impacting cooling usage patterns of Chinese urban household based on a large-scale questionnaire survey. Energy Build 2020;214:109885. https://doi.org/10.1016/j.enbuild.2020.109885.
[32] Dong B, Yan D, Li Z, Jin Y, Feng X, Fontenot H. Modeling occupancy and behavior for better building design and operation—a critical review. Build Simul 2018;11:899–921. https://doi.org/10.1007/s12273-018-0452-x.
[33] Haldi F, Robinson D. Adaptive actions on shading devices in response to local visual stimuli. J Build Perform Simul 2010;3:135–53. https://doi.org/10.1080/19401490903580759.
[34] Duarte C, Van Den Wymelenberg K, Rieger C. Revealing occupancy patterns in an office building through the use of occupancy sensor data. Energy Build 2013;67:587–95. https://doi.org/10.1016/j.enbuild.2013.08.062.
[35] Li J, (Jerry) Yu Z, Haghighat F, Zhang G. Development and improvement of occupant behavior models towards realistic building performance simulation: A review. Sustain Cities Soc 2019;50:101685. https://doi.org/10.1016/j.scs.2019.101685.
[36] Dong B, Lam KP. A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local

weather forecasting. Build Simul 2014;7:89–106. https://doi.org/10.1007/s12273-013-0142-7.
[37] Ahn KU, Kim DW, Park CS, de Wilde P. Predictability of occupant presence and performance gap in building energy simulation. Appl Energy 2017;208:1639–52. https://doi.org/10.1016/j.apenergy.2017.04.083.
[38] Huchuk B, Sanner S, O'Brien W. Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. Build Environ 2019;160:106177. https://doi.org/10.1016/j.buildenv.2019.106177.
[39] Das A, Kjærgaard MB. Precept: Occupancy presence prediction inside a commercial building. In: UbiComp/ISWC 2019- - Adjun. Proc. 2019 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Proc. 2019 ACM Int. Symp. Wearable Comput., Association for Computing Machinery, Inc, New York, New York, USA; 2019. p. 486–91. https://doi.org/10.1145/3341162.3345605.
[40] Miller H. Place-based versus people-based geographic information science. Geogr Compass 2007;1:503–35. https://doi.org/10.1111/j.1749-8198.2007.00025.x.
[41] Mosteiro-Romero M, Fonseca JA, Schlueter A. Seasonal effects of input parameters in urban-scale building energy simulation. Energy Proc, Elsevier Ltd 2017:433–8. https://doi.org/10.1016/j.egypro.2017.07.459.
[42] Barbour E, Davila CC, Gupta S, Reinhart C, Kaur J, González MC. Planning for sustainable cities by estimating building occupancy with mobile phones. Nat Commun 2019;10:3736. https://doi.org/10.1038/s41467-019-11685-w.
[43] Happle G, Fonseca JA, Schlueter A. Context-specific urban occupancy modeling using location-based services data. Build Environ 2020;175:106803. https://doi.org/10.1016/j.buildenv.2020.106803.
[44] Mohammadi N, Taylor JE. Urban energy flux: Spatiotemporal fluctuations of building energy consumption and human mobility-driven prediction. Appl Energy 2017;195:810–8. https://doi.org/10.1016/j.apenergy.2017.03.044.
[45] Pappalardo L, Simini F, Rinzivillo S, Pedreschi D, Giannotti F, Barabási AL. Returners and explorers dichotomy in human mobility. Nat Commun 2015;6:1–8. https://doi.org/10.1038/ncomms9166.
[46] Wang Q, Taylor JE. Process map for urban-human mobility and civil infrastructure data collection using geosocial networking platforms. J Comput Civ Eng 2016;30:04015004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000469.
[47] Hong T, Chen Y, Lee SH, Piette MA. CityBES: a web-based platform to support city-scale building energy efficiency. In: Conf 5th Int Urban Comput Work, San Francisco; 2016. https://www.researchgate.net/publication/304824985_CityBES_A_Web-based_Platform_to_Support_City-Scale_Building_Energy_Efficiency [accessed June 6, 2020].
[48] Reinhart CF, Dogan T, Jakubiec JA, Rakha T. UMI – an urban simulation environment for building energy use, daylighting and walkability. In: Proc Build Simul 2013, Chambery; 2013. https://www.researchgate.net/publication/264352980_UMI_-_An_urban_simulation_environment_for_building_energy_use_daylighting_and_walkability [accessed June 6, 2020].
[49] Bollinger LA, Evins R. HUES: A holistic urban energy simulation platform for effective model integration. In: Proc Int Conf CISBAT 2015 Futur Build Dist Sustain from Nano to Urban Scale; 2015. p. 841–6. https://hues.empa.ch. [accessed June 6, 2020].
[50] Remmen P, Lauster M, Mans M, Fuchs M, Osterhage T, Müller D. TEASER: an open tool for urban energy modelling of building stocks. J Build Perform Simul 2018;11:84–98. https://doi.org/10.1080/19401493.2017.1283539.
[51] Robinson D, Haldi F, Kämpf J, Leroux P, Perez D, Rasheed A et al. CitySim: comprehensive micro-simulation of resource flows for sustainable urban planning. In: Build Simul 2009, Glasgow, Scotland; 2009. https://www.aivc.org/resource/citysim-comprehensive-micro-simulation-resource-flows-sustainable-urban-planning [accessed June 6, 2020].
[52] Robinson D, Campbell N, Gaiser W, Kabel K, Le-Mouel A, Morel N, et al. SUNtool - A new modelling paradigm for simulating and optimising urban sustainability. Sol Energy 2007;81:1196–211. https://doi.org/10.1016/j.solener.2007.06.002.
[53] ArcGIS, ArcGIS Online; 2020. https://www.arcgis.com/index.html [accessed June 5, 2020].
[54] ENVI, ENVI - The Leading Geospatial Image Analysis Software; 2020. https://www.harrisgeospatial.com/Software-Technology/ENVI [accessed June 5, 2020].
[55] Lidar – Wikipedia; n.d. https://en.wikipedia.org/wiki/Lidar [accessed June 5, 2020].
[56] Akhavan A, Phillips NE, Du J, Chen J, Sadeghinasr B, Wang Q. Accessibility Inequality in Houston. IEEE Sens Lett 2018;3:1–4. https://doi.org/10.1109/lsens.2018.2882806.
[57] Sadeghinasr B, Akhavan A, Wang Q. Estimating commuting patterns from high resolution phone GPS data. Am Soc Civ Eng (ASCE) 2019:9–16. https://doi.org/10.1061/9780784482438.002.
[58] Wang F, Wang J, Cao J, Chen C, (Jeff) Ban X. Extracting trips from multi-sourced data for mobility pattern analysis: An app-based data example. Transp Res Part C Emerg Technol 2019;105:183–202. https://doi.org/10.1016/j.trc.2019.05.028.
[59] Phillips NE, Levy BL, Sampson RJ, Small ML, Wang RQ. The social integration of American cities: network measures of connectedness based on everyday mobility across neighborhoods. Sociol Methods Res 2019. https://doi.org/10.1177/0049124119852386. 004912411985238.
[60] Chen C, (Jeff) Ban X. Transportation big data: promises, issues, and implications. In: 97th Annu Meet Transp Res Board, Washington DC, United States; 2018.
[61] Hong T, Lin HW. Occupant behavior: impact on energy use of private offices | Energy Technologies Area, Berkeley, CA (United States); 2013. https://eta.lbl.gov/publications/occupant-behavior-impact-energy-use [accessed January 3, 2020].
[62] Tang R, Wang S, Sun S. Impacts of technology-guided occupant behavior on air-conditioning system control and building energy use Article History; n.d. https://doi.org/10.1007/s12273-020-0605-6.
[63] Sun K, Hong T. A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures. Energy Build 2017;146:383–96. https://doi.org/10.1016/j.enbuild.2017.04.065.
[64] Beltran A, Cerpa AE. Optimal HVAC building control with occupancy prediction. In: BuildSys 2014 - Proc. 1st ACM Conf. Embed. Syst. Energy-Efficient Build., Association for Computing Machinery, Inc, New York, New York, USA; 2014. p. 168–71. https://doi.org/10.1145/2674061.2674072.
[65] Chen Y, Hong T, Piette MA. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. Appl Energy 2017;205:323–35. https://doi.org/10.1016/j.apenergy.2017.07.128.
[66] Chen Y, Hong T, Luo X, Hooper B. Development of city buildings dataset for urban building energy modeling. Energy Build 2019;183:252–65. https://doi.org/10.1016/j.enbuild.2018.11.008.
[67] Hong T, Piette MA, Chen Y, Lee SH, Taylor-Lange SC, Zhang R, et al. Commercial Building Energy Saver: An energy retrofit analysis toolkit. Appl Energy 2015;159:298–309. https://doi.org/10.1016/j.apenergy.2015.09.002.
[68] Lee SH, Hong T, Piette MA, Sawaya G, Chen Y, Taylor-Lange SC. Accelerating the energy retrofit of commercial buildings using a database of energy efficiency performance. Energy 2015;90:738–47. https://doi.org/10.1016/j.energy.2015.07.107.
[69] ARPA-E. Saving Energy Nationwide in Structures with Occupancy Recognition (SENSOR); n.d. https://arpa-e.energy.gov/?q=arpa-e-programs/sensor [accessed June 4, 2020].