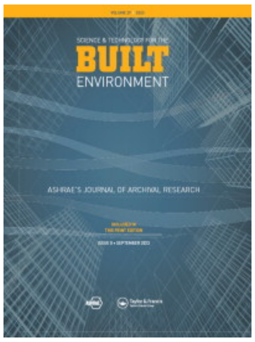

Science and Technology for the Built Environment

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uhvc21

ASHRAE URP-1883: Development and Analysis of the ASHRAE Global Occupant Behavior Database

Yapan Liu, Bing Dong, Tianzhen Hong, Bjarne Olesen, Thomas Lawrence & Zheng O'Neill

To cite this article: Yapan Liu, Bing Dong, Tianzhen Hong, Bjarne Olesen, Thomas Lawrence & Zheng O'Neill (2023) ASHRAE URP-1883: Development and Analysis of the ASHRAE Global Occupant Behavior Database, Science and Technology for the Built Environment, 29:8, 749-781, DOI: 10.1080/23744731.2023.2235971

To link to this article: https://doi.org/10.1080/23744731.2023.2235971

Published online: 26 Jul 2023.

Science and Technology for the Built Environment, (2023) **29**, 749–781

ISSN: 2374-4731 print / 2374-474X online
DOI: 10.1080/23744731.2023.2235971

ASHRAE URP-1883: Development and Analysis of the ASHRAE Global Occupant Behavior Database

YAPAN LIU[1], BING DONG[1]*, TIANZHEN HONG[2], BJARNE OLESEN[3], THOMAS LAWRENCE[4], and ZHENG O'NEILL[5]

[1]*Department of Mechanical and Aerospace Engineering, Syracuse University, 223 Link Hall, Syracuse, NY 13244, USA*
[2]*Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[3]*International Centre for Indoor Environment and Energy, Department of Environmental and Resource Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*
[4]*College of Engineering, Driftmier Engineering Center, University of Georgia, Athens, GA 30602, USA*
[5]*Department of Mechanical Engineering, Texas A&M University, 400 Bizzell St, College Station, TX 77843, USA*

In developed countries, people spend nearly 90% of their time in buildings or during transportation. Recent research studies demonstrated that occupant behaviors have a significant impact on building performance in relation to the indoor environment and energy use. This paper presents the ASHRAE Global Occupant Behavior Database which aims to advance the knowledge and understanding of realistic occupancy patterns and human-building interactions with building systems. This database includes 34 field-measured occupant behavior datasets for both commercial and residential buildings, contributed by researchers from 15 countries and 39 institutions covering 10 different climate zones. It includes occupancy patterns, occupant behaviors, indoor and outdoor environment measurements. The database is open source, a public website was developed for the users to interactively explore, query, and download datasets. This paper focuses on a detailed data analysis to investigate patterns of nine occupant behavior types, examining impacted factors such as building type, country, and climate zone. EnergyPlus simulations have been implemented based on the occupancy profiles derived from this database, and results showed overall building electricity consumption can be reduced up to around 27% in Summer and around 10% in Winter.

Introduction

In developed countries, people spend nearly 90% of their time in buildings or during transportation (Fontanini et al. 2016; US EPA. 2014; World Health Organization. Regional Office for Europe 2014). Recent research studies demonstrated that occupant behaviors have a significant impact on the building performance in relation to indoor environment and energy use. Building energy use is a systematic procedure comprehensively influenced by not only engineering technologies, but also cultural concept, occupant behavior and social equity. People spend nearly 90% of their lifetime in buildings (Klepeis et al. 2001), which makes occupant behavior one of the leading influences of energy consumption in buildings. Indeed, occupant actions such as adjusting a thermostat and opening/closing windows for thermal comfort, switching lights on/off and pulling window shadings up/down for visual comfort, using appliances, and moving between spaces can have a significant impact on both energy use and occupant comfort in buildings. Depending on the building type, climate, and degree of automation in operation and controls, such behaviors can increase or decrease energy use, for example, by a factor of up to three for residential buildings (Andersen 2012), and increase energy use by up to 80% or reduce energy use by up to 50% for single-occupancy offices (Hong and Lin 2013), while having a 41% Heating, Ventilation, and Air Conditioning (HVAC) energy savings potential for office buildings (Sun and Hong 2017).

Many research studies over the last decade have focused on the topic of occupant behavior. To better understand occupant behavior in buildings, prior studies conducted experiments to derive various mathematical models.

Received February 14, 2023; accepted June 26, 2023
Yapan Liu, Student Member ASHRAE, is a PhD Candidate. **Bing Dong, PhD, Full Member ASHRAE**, is an Associate Professor. **Tianzhen Hong, PhD, Fellow ASHRAE**, is a Senior Research Scientist. **Bjarne Olesen, PhD, Presidential Fellow/Life Member ASHRAE**, is a Professor. **Thomas Lawrence, PhD, Fellow ASHRAE**, is a Professor. **Zheng O'Neill, PhD, Fellow ASHRAE**, is an Associate Professor.
*Corresponding author e-mail: bidong@syr.edu

Occupancy models

Zhou et al. (2023) has collected time-use survey data to investigate the temporal changes in occupancy patterns and explore the impact factors for behavioral change. Another study (Fu et al. 2022) conducted national survey to study the occupancy patterns in residential buildings. Jin et al. (2021) developed machine learning model for building occupancy forecasting by integrating temporal-sequential analysis with artificial neural networks. Dobbs and Hencey (2014a, 2014b) developed a stochastic occupancy model based on Markov Chain analysis, using data collected over a three-month period from a conference room. The model was then combined with the building's thermal properties and local weather predictions to simulate the impact of occupancy on energy consumption and thermal comfort. Results showed that significant reductions in energy consumption were possible while maintaining occupant comfort. Li and Dong (2017) collected occupancy data from four residential houses and used it to create an inhomogeneous Markov model for occupancy prediction. This model outperformed other methods, achieving an average of 5% more accurate predictions. In another study, Mahdavi and Tahmasebi (2015) obtained high-resolution and long-term occupancy data from university workplaces. This data was used to evaluate the performance of existing probabilistic occupancy models against an original non-probabilistic model.

Occupant number count

A recent study (Alishahi, Ouf, and Nik-Bakht 2022) proposed a method to analyze long-term building occupant numbers based on Wi-Fi connections, and the method was applied in a university library building to collect occupant number data for three months. Another study (Choi et al. 2021) developed a vision-based occupancy counting method combined with deep learning models. Wang et al. (Wang et al. 2021) proposed a smart low-cost ventilation control strategy utilizing occupant-density-detection algorithm, it also balanced both infection prevention and energy efficiency. Accurately detecting the number of occupants in a room is essential for optimizing energy use and maintaining indoor comfort. Dong et al. (2010) and Dong and Lam (2011, 2014) deployed a complicated sensing network in a university office building and used the collected data to develop Gaussian Mixture Model-based Hidden Markov Models for room-level occupant number detection. Their models achieved an average accuracy of 83%, and EnergyPlus simulations showed that they could lead to energy savings of up to 18.5% while maintaining thermal comfort. Erickson et al. (2009), Erickson, Carreira-Perpiñán, and Cerpa (2011), Erickson, Carreira-Perpiñán, and Cerpa (2014), and Erickson and Cerpa (2010) estimated building occupancy using a wireless camera sensor network with an accuracy of 80%. They constructed multivariate Gaussian and agent-based models based on the collected data for room usage prediction. Simulation results showed that, on average, a 42% annual energy savings could be achieved while meeting ASHRAE thermal comfort standards. Manna et al. (2013) collected occupancy data from three different office buildings and proposed an algorithm for occupancy prediction by considering occupant behavior as an ensemble of multiple Markov models at different time lags. Recently, the increasing use of urban sensing, IoT, and big data in cities presents unique opportunities to gain a more profound comprehension of occupant behavior and energy consumption patterns on an urban level (Salim et al. 2020). Researchers (Dong et al. 2019; Kang et al. 2021; Wu et al. 2020) have developed mobility-based approaches to derive building occupancy profiles at urban scale.

Window operations

To detect building occupancy and human building interactions including manual window operations, a recent study (Tien et al. 2022) presented a vision-based deep learning framework and tested it in a university building. Another study (Niu et al. 2022) investigated occupant window opening behavior in a hospital during summer season. The study collected 10-minute level window operation, indoor and outdoor environmental data for three months. Verbruggen et al. conducted a study in residential buildings with a focus on habitual window opening behavior. Haldi and Robinson (2009) collected seven years of continuous measurement data and analyzed the correlation of window opening and closing behavior with occupancy patterns, indoor temperature, and outdoor climate parameters. They proposed a hybrid stochastic model to predict window operation behaviors. Schweiker, Kleber, and Wagner (2019) implemented field sensing measurements to monitor naturally ventilated office building window operations over a four-year period. Shi and Zhao (2016) conducted a field study in eight naturally ventilated residential apartments for 14 months and built stochastic models to represent occupants' window operation behaviors. Results indicated that outdoor air temperature is the most important explanatory variable affecting occupants' interactions with windows, among other measured parameters. Yun and Steemers (2008) investigated window-opening control by occupants in an office setting. They derived a statistical relationship between window operation behavior and indoor stimulus, such as indoor air temperature, in the summer. The authors proposed a stochastic model to predict window-opening behavior patterns, considering parameters such as indoor temperature, time of day, and the previous window state.

Shading and lighting operations

Ding et al. (2020) proposed a prediction model which coupled the lighting and shading control behavior accurately. This study collected hourly lighting and shading operation behaviors from 12 private offices in a 2-storey office building. Tabadkani et al. (2021) investigated the impact of automated shading controls on occupant's comfort and energy load. Some researchers investigated both shading and lighting operations together. Correia da Silva, Leal, and Andersen (2015) conducted a field campaign to monitor occupants' manual control of electric lighting and shading devices in single office spaces continuously. The study collected detailed measurements and observations. Reinhart et al. (Reinhart 2004; Reinhart and Voss 2003) proposed statistical models to simulate and predict the lighting energy performance of manually and automatically controlled electric lighting and shadings systems in an office setting. Similar studies (Haldi and Robinson 2009; Mahdavi et al. 2008) aimed to understand occupants' operation of shading and lighting systems in office buildings. Inkarojrit (2005) developed predictive manual control models for shading operations while considering occupants' satisfaction and preferences. Furthermore, researchers (Newsham and Arsenault 2009) developed a camera-based system to study lighting and shading control. Chang and Hong (2013) derived occupancy patterns from measured high granularity lighting-switch data in an open-plan offices. Such studies can provide valuable insights for developing efficient and effective building automation and control systems.

Thermostat adjustments

Recent study (Tamas, O'Brien, and Quintero 2021) investigated the correlation between residential thermostat usability and interface characteristics in Canada. The authors collected data from 51 participants through interviews. Vellei, Martinez, and Le Dréau (2021) proposed a framework which models occupants' thermostat usage behaviors in residential buildings. It utilized user interaction data from around 9,000 connected Canadian thermostats. Huchuk, O'brien, and Sanner (2021) examined smart thermostat users' schedule override behaviors and its energy consequences by analyzing a dataset which contains 20,000 smart thermostats. Previous study (Peffer et al. 2011) provided a comprehensive review of how occupants use thermostats and the evolution in technologies of residential thermostats. Fabi, Andersen, and Corgnati (2013) implemented filed experiments in 13 residential dwellings to study occupants' heating setpoint behavior. The developed models were later incorporated into simulations to study its influence on indoor climate quality and energy consumption. In a recent study (He et al. 2022), researchers collected 873 questionaries about the occupant behaviors in air-conditioned office buildings. The study developed a probability prediction model to represent the cooling temperature set-point adjustment behavior.

Air-conditioner operations

Rahman and Han (Rahman and Han 2021) evaluated the occupancy-based demand controlled ventilation strategies (time-based and CO_2) based in terms of applicability and performance. It collected occupancy and HVAC system data from a small office setting. Brackley, O'Brien, and Trudel (2020) implemented a study in 25 academic offices for over three months to study occupants' perceived control of HVAC system. The study utilized sensing data from building automation system during heating season. Fabi et al. (Fabi, Andersen, and Corgnati 2013) conducted a study to transit occupant behavior models from a deterministic method of building energy simulation to a probabilistic model to investigate the influence of occupants on building controls. To achieve this, a probabilistic approach was proposed and employed to accurately simulate occupant behavior. The methodology involved the probabilistic evaluation of input and output variables in building energy simulations, with the aim of comparing the results obtained to those from a conventional deterministic use of the simulation program. Models of occupant behavior patterns were utilized to explore how different behavior patterns impact indoor climate quality and energy consumption. The simulation results were presented as probability distributions of energy consumption and indoor environmental quality, based on the occupant's behavior. Ren et al. (Ren, Yan, and Wang 2014) introduced a model for air conditioning usage in residential buildings that takes into account the behavior of occupants. To develop the model, surveys and continuous monitoring of more than thirty households in 8 cities located in various climate zones were conducted, revealing a range of distinct patterns of air conditioning usage. The developed model is a quantitative stochastic model that considers the different factors affecting AC usage, including environmental and event triggers. These patterns are described mathematically through a series of conditional probabilities.

Models from these studies were built to describe occupant behavior in buildings in order to evaluate the performance of building design and operation. There are mainly four application areas in which occupant behavior modeling plays key roles, including 1) building energy performance analysis, 2) building architecture and engineering design, 3) intelligent building operation, and 4) building safety design.

Building energy performance analysis

Historically, occupant behavior is often modeled as a fixed input for building simulation tools to study total building energy consumption, and to size HVAC systems. Specifically, occupancy has been modeled as an hourly or sub-hourly schedule with values varying between 0 and 1 as the ratio of a predefined maximum number of occupants in a space. Recently, however, the stochastic nature of occupancy has captured a great deal of attention from researchers and engineers who are conducting building performance simulations studies (Dong and Lam 2011; Gram-Hanssen 2010; Hong et al. 2017; Jin et al. 2021; Kang et al. 2021; Saldanha and Beausoleil-Morrison 2012), and often stochastic results are given (Feng, Yan, and Wang 2017).

Building architecture and engineering design

Occupant behavior modeling is used for both architecture and engineering design, specifically building circulation design (Tomastik, Lin, and Banaszuk 2008; Yuhaski and Smith 1989) and HVAC sizing (Cook et al. 2003; Jacobs and Henderson 2002; Sun et al. 2014). As occupants move among spaces in a building for various activities, the success of a building circulation design lies in the level of convenience for people to transfer to another activity. To help the design, the preferences for different functions of rooms and the patterns of occupants' movements need to be captured. In the case of HVAC sizing, occupant schedules and behaviors such as adjusting thermostats and operating windows and shadings are two major impact factors.

Intelligent building operation

Advanced control design for building systems, such as lighting and HVAC, rely on the detection and modeling of occupant behavior. Many prior studies demonstrate, both in simulations and field experiments, that occupancy-based indoor climate control can save up to 30% energy consumption (Mirakhorli and Dong 2016). In addition, the scheduling of elevators strongly depends on the occupancy patterns in a building. The developed occupancy model often predicts the aggregation of occupants in order to reduce the time people spend waiting for the elevator to arrive.

Building safety design

Occurrences of intensive occupancy, or crowds (Helbing et al. 2005), have a significant impact on the building safety design. This is especially important to architects who design public buildings, such as theaters and shopping malls. It is important to have knowledge about how the crowd evolves when a specific condition has been reached to secure people's safety (Chow and Ng 2008). For example, in the subway system, crowd models can be used to evaluate the performance of the path design.

However, each research study has its own datasets and represents an individual case, although studies are across various countries globally. There are over 400+ papers published on the topic of occupant behavior over the last decade. Hence, it is time to consolidate those very valuable datasets into a large data repository. With such a large body of data to work on, occupant behavior researchers will be able to dive deeper to compare occupant behaviors across various building types and nations and derive valuable information for energy-efficient building design and operations.

Over the last decade, many research studies focused on the modeling and simulation of occupant behavior in buildings (e.g., IEA EBC Annex 53 (Yoshino, Hong, and Nord 2017), Annex 66 (Yan et al. 2017, 66) and Annex 79 (O'Brien et al. 2020, 79)), and their applications to building design and operation. Depending on the building types, climates, systems and controls, occupant behaviors could have favorable or adverse impacts on building performance. Thus, there is a need for a world-wide open-source database on occupant behavior in the built environment.

The ASHRAE Global Occupant Database we have developed includes 34 field-measured building occupant behavior datasets collected from 15 countries and 39 institutions across 10 climatic zones covering various building types in both commercial and residential sectors. This is a comprehensive global database of building occupant behavior. The database covers occupancy patterns (i.e., presence and people count), indoor and outdoor environment measurements, and occupant behaviors (i.e., interactions with devices, equipment, and technical systems in buildings). The database is open-access and provides data visualization, Application Programming Interface (API), and query tools. The database intends to support occupant behavior research that informs the design and operation of low or net-zero energy buildings with significant human-building interactions (HBI). And improve the understanding of human-building interactions, which is a key for design and operation of low-energy and high-performance buildings.

This paper is organized as follows: The method section provides an overview of our approach, covering data collection, processing, and database development. It includes the introduction of the database, query implementation, and organization of occupant behavior data. The data analysis section, which is the focus of this paper, presents an analysis of nine distinct occupant behaviors identified in this project. Afterwards, the case study section demonstrates the efficacy and utility of this database by integrating the collected data into EnergyPlus simulations. Then, the paper discussed its findings, limitations, and future work. Finally, we concluded the paper and outlined the potential applications of this database.

Methods

This section presents the overall approach we have adopted to develop the database, including data collection and processing, database development, along with the implementation of querying and organizing the occupant behavior data. As this paper concentrates on analyzing and showcasing the collected datasets, specifics regarding data collection and processing can be found in our earlier study (Dong et al. 2022) of this project. The development of building metadata modeling and the Brick schema extension are described in a separate study (Luo et al. 2022) and will not be discussed in this paper.

Data collection and processing

To obtain the most relevant data, a worldwide survey was conducted for researchers who have indicated their willingness to contribute to the database. The survey collects basic information of the building metadata and zones, building equipment, methods used to collect the data, dataset details, and additional information. With the information collected in this process, the project team later reached out to potential contributors with more detailed requirements Eventually, final collected datasets were contributed by 51 contributors from 39 institutes in 15 different countries. The database covers 10 different climate zones globally according to the Köppen-Geiger climate classification (https://en.climate-data.org).

After the data collection process, all datasets were evaluated based on the pre-defined requirements. Datasets contributors took steps to ensure the privacy of the occupant data, and further anonymization was applied during pre-processing. The datasets were then divided into three categories: survey type, in-situ type, and mixed type. The in-situ data includes continuously collected dynamic measurements within the building, such as the status of doors and windows (OPEN/CLOSED) and building equipment (ON/OFF), as well as indoor and outdoor environmental information (temperature, humidity, carbon dioxide concentration, illumination, etc.). Survey data consists of unique information specific to the study, including occupant questionnaires, static information about the building's envelope, and floor plan, project specific measurements, etc. Datasets without timestamps were also classified as survey type data. The mixed type of data includes both in-situ and survey type data; there was only one dataset identified as mixed type. Our recent study (Dong et al. 2022) provides a summary of all 34 datasets, including the country of origin, collection method, measurement categories, as well as publications related to each dataset. These datasets include 24 in-situ datasets, one mixed dataset, and nine Survey datasets.

Figure 1 shows a complete report of missing data rates, it can be observed that most datasets have less than 5% of missing data. To ensure that the datasets are both high quality and maintain their originality, we conducted a pre-processing procedure which includes removing empty columns from the raw data, filling missing values with − 999 in the raw data, anonymizing building and room information by assigning unique ID numbers, and applying a standardized data naming schema and format. A full description procedure of data pre-processing and quality control can be found in our data descriptor report (Dong et al. 2022). The project team has designed 11 templates to organize all the datasets as Table 1 shows, each template covers the specific type of behavior. During pre-processing, all the in-situ datasets were formatted to align with the templates and missing values in raw datasets were replaced with − 999, but survey datasets remain as their original form. Future contributors are suggested to follow the templates on the database's website when contributing to this database. The building metadata information has not been included in this database as not all researchers have provided this data. Nevertheless, there are alternative ways to access this information:1) From our raw dataset repository on figshare (Dong et al. 2021); 2) The database's website provides a list of publications related to the collected datasets.

Dataset Number	Missing Data Ratio (%)	Type
1	0	survey
2	0	in-situ
3	2.83	survey
4	0	in-situ
5	0.58	in-situ
6	2.7	in-situ
7	0	in-situ
8	17.47	in-situ
9	0	in-situ
10	1.4	in-situ
11	0	in-situ
12	0	survey
13	0	in-situ
14	7.36	in-situ
15	0	in-situ
16	0.15	in-situ
17	6.39	in-situ
18	3.39	in-situ
19	1.6	survey
20	0	in-situ
21	13.08	in-situ
22	1.1	in-situ
23	0.01	in-situ
24	15.81	in-situ
25	2.95	in-situ
26	11.53	mixed
27	0.45	survey
28	3.38	survey
29	1.44	survey
30	0	in-situ
31	0.03	survey
32	0	in-situ
33	0	in-situ
34	2.26	survey

Fig. 1. Report of missing data rate by dataset.

Database development

A website (https://ashraeobdatabase.com) was created as a data warehouse for public access. Query builder tools were developed based on different behavior types, cities, and countries, building types, study ID, and publication list. Users can select and download data from the database interactively through the query builder. Data analytic functions were developed to provide an interactive overview of the database and assist users to select the dataset. A Python package named "OBPlatform" was developed to access the database programmatically. The codes of this package are publicly accessible on the GitHub page (https://github.com/umonaca/obplatform) with beginner's tutorials. The website provides an API page that detailed out information to query and download datasets through Representational State Transfer (REST) APIs.

Figure 2 is the Entity Relationship Diagram (ERD) for the MySQL database. The database was developed using the hybrid of file-based database and traditional relational database (MySQL). The relational database is suitable for metadata queries, while the file-based database provides flexibility on the data schema and saves most of the time for the data wrangling process. Also, tests during the development have shown that the file-based approach is much faster than complex SQL joins with MySQL, even with all the optimization methods implemented, such as database indexing and manual semi-joins.

Database description

After data processing and quality control, the final ASHRAE Global Occupant Behavior Database consists of 34 field-measured building occupant behavior datasets collected from 2003 to 2020 covering various building types in both

Table 1. Dataset templates and their measurements.

Dataset Template	Variable		
Plug Load	Plug_Load_ID	Plug_ID	Room_ID
	Date_Time	Desk_ID	Building_ID
	Electric_Power[w]		
Door Status	Door_Status_ID	Door_ID	Building_ID
	Date_Time	Room_ID	
	Door_Status[0-Closed;1-Open]		
Fan Status	Fan_Status_ID	Date_Time	Room_ID
	Fan_Status[0-OFF;1-ON]	Fan_ID	Building_ID
HVAC Measurement	HVAC_Measurement_ID	VAV_Opening[%]	Return_Air_Temp[C]
	Date_Time	Supply_Air_Temp[C]	Duct_Air_Flowrate[cfm]
	Heating_Status[0-OFF;1-ON]		HVAC_Zone_ID
	Cooling_Status[0-OFF;1-ON]		Room_ID
	Temp_Setpoint[C]		Building_ID
Indoor Measurement	Indoor_Measurement_ID	Indoor_VOC[ppm]	Connected_Device_Number
	Date_Time	Indoor_Air_Speed[m/s]	Connected_Device_Type
	Indoor_Temp[C]	Indoor_Air_Pressure[Pa]	Desk_ID
	Indoor_RH[%]	Indoor_Illuminance[LUX]	Room_ID
	Indoor_CO2[ppm]		Building_ID
Lighting Status	Lighting_Status_ID		Lighting_Zone_ID
	Date_Time		Room_ID
	Ligthing_Status[0-OFF;1-ON]		Building_ID
Occupancy Measurement	Occupancy_Measurement_ID		Desk_ID
	Date_Time		Room_ID
	Occupancy_Measurement [0-Unoccupied;1-Occupied]		Building_ID
Occupant Number Measurement	Occupant_Number_Measurement_ID		Room_ID
	Date_Time		Building_ID
	Occupant_Number_Measurement		
Outdoor Measurement	Outdoor_Measurement_ID		Solar_Radiation[w/m2]
	Date_Time		IAQ
	Outdoor_Temp[C]		Particle_Level[ug/m3]
	Outdoor_RH[%]		Particle_Type
	Wind_Speed[m/s]		Precipitation
	Wind_Direction[deg]		Building_ID
	Outdoor_Air_Pressure[Pa]		
Shade Status	Shade_Status_ID		Shade_ID
	Date_Time		Room_ID
	Shade_Status[0-OFF;1-ON]		Building_ID
Window Status	Window_Status_ID		Window_ID
	Date_Time		Room_ID
	Window_Status[0-Closed;1-Open]		Building_ID

commercial and residential sectors. In total, 34 datasets (Dong et al. 2021) around 3.81 GB data records were included in this database, with 24 in-situ types of datasets, one mixed type dataset, and nine survey type datasets. Figure 1 shows the missing data report for each dataset. The Köppen-Geiger climate classification has been broadly used by researchers around the world in the smart building community (Amasyali and El-Gohary 2016; Carlucci et al. 2020; Kim et al. 2017). Since the datasets in this database were contributed by researchers around the globe, Köppen-Geiger climate classification was adopted to represent the different climate zones in the datasets. The database covers 10 different climate zones globally according to the Köppen-Geiger climate classification. Those climate zones include Af (Tropical rainforest climate), Aw (Tropical savanna, wet), Bwh (Hot deserts climate), Cfa (Humid subtropical climate), Cfb (Temperate oceanic climate), Csa (Hot-summer Mediterranean climate), Csc (Cool-summer Mediterranean climate), Dfa (Hot-summer humid continental climate), Dfb (Warm-summer humid continental climate), Dwa (Monsoon-influenced hot-summer humid continental climate). A website (https://ashraeobdatabase.com) was created to query and download the desired data from the database based on different selection criteria or through public accessible REST APIs.

The database covers field measurements from six different continents: Asia, Australia, Europe, Middle East, North America, and South America. Among those continents, about 36% of the data (by behavior type) comes from

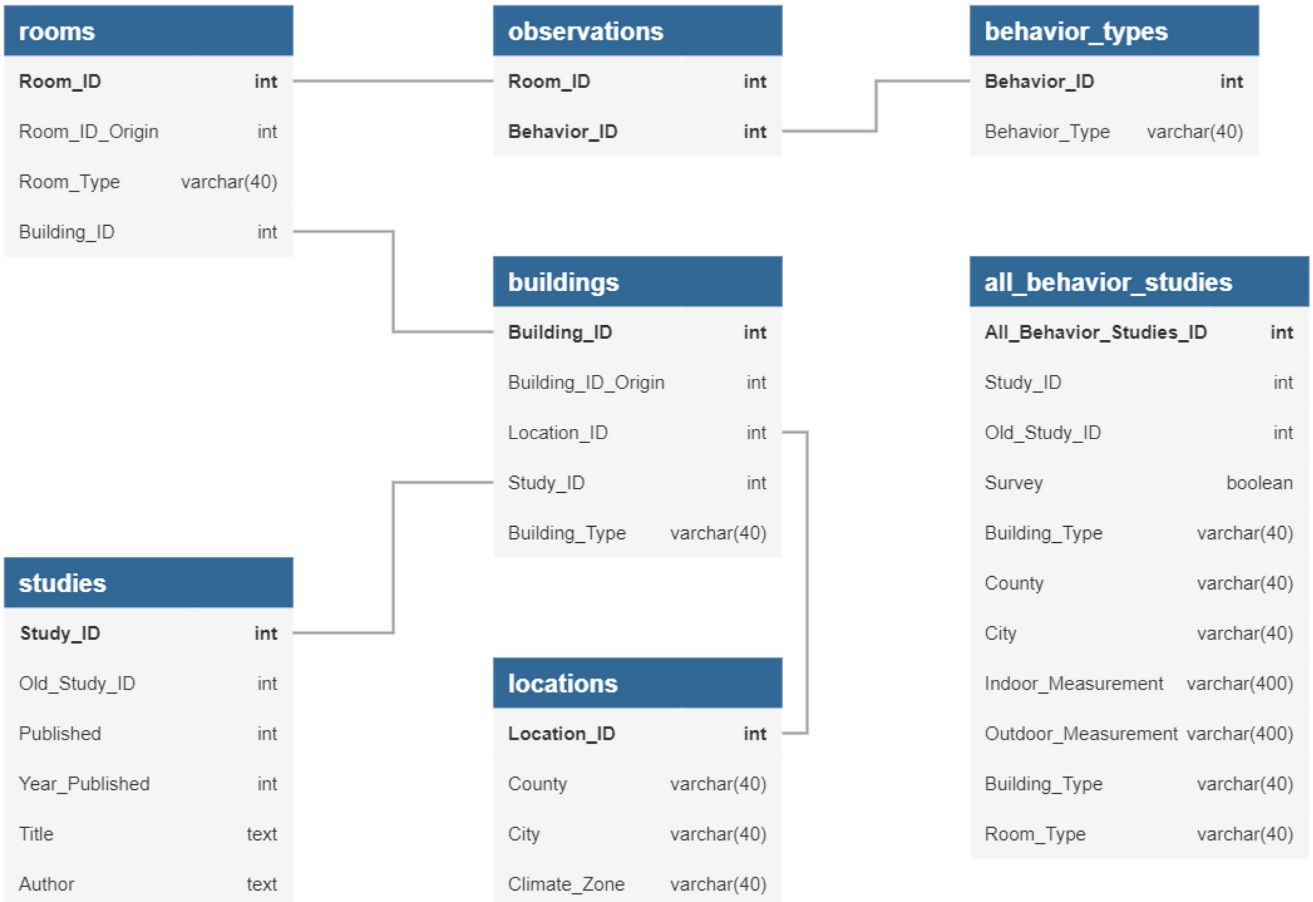

Fig. 2. Entity relationship diagram.

Fig. 3. Distribution of the locations and climate zones of the collected field studies.

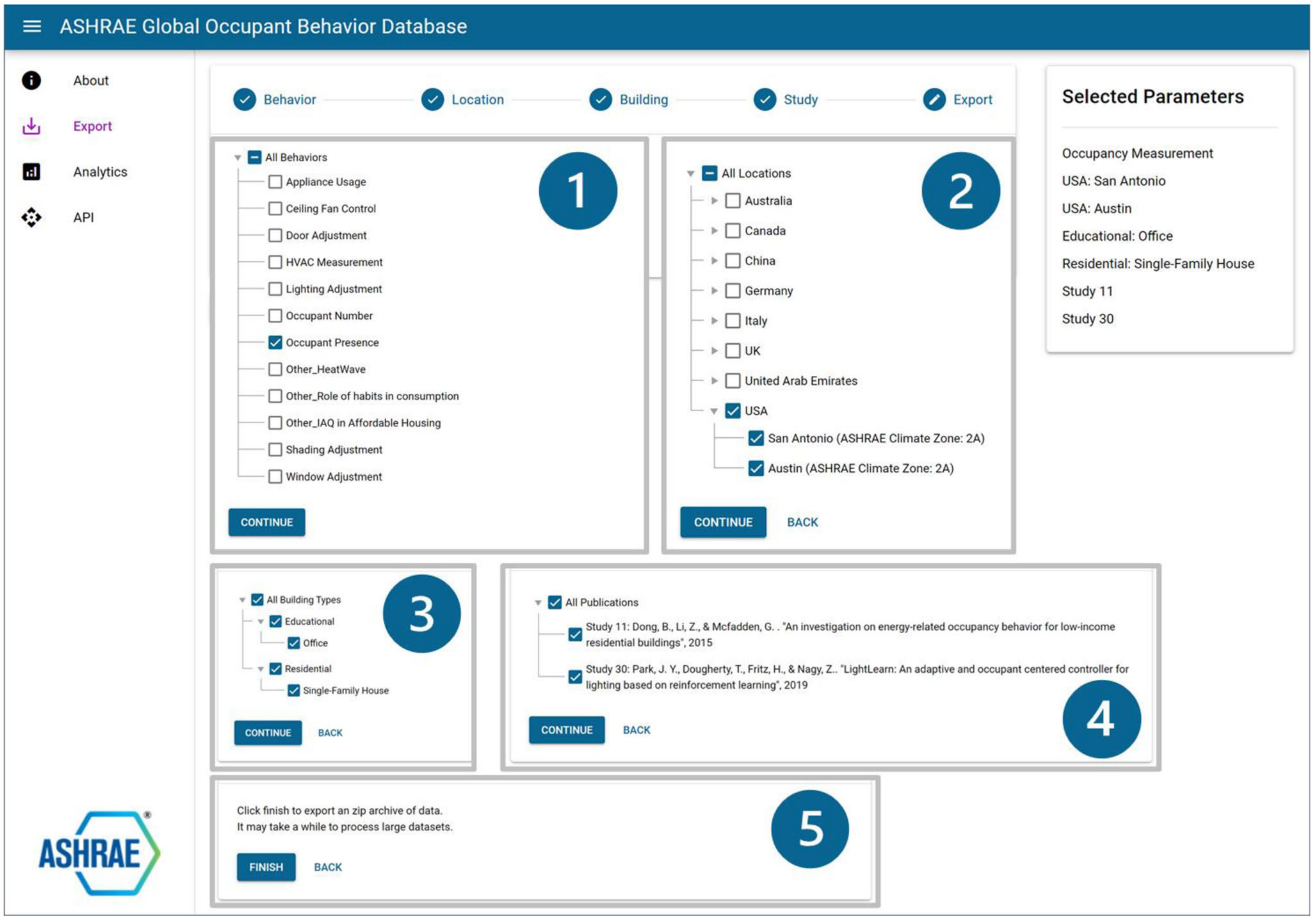

Fig. 4. The web interface of the ASHRAE Global Occupant Behavior Database.
1-Select behavior type; 2-Select location of interest; 3-Select building type; 4-Select study of interest; 5-Export data

Europe, and close to 35% data (by behavior type) contributed by researchers from Asia. Figure 3 shows the locations and climate zones of those field studies, each pin represents a city with its climate zone. In total, 35 cities across 10 climate zones were found from the datasets. A query builder which will be introduced in the following section, can assist users to select the desired dataset based on different parameters.

Three types of buildings were identified in the database: educational, commercial, and residential. The educational and commercial building types were separated because the data contributors are mostly university researchers who primarily collected data in university educational buildings. These two building types cover a wide range of behavior types in the database. Commercial buildings include office spaces, while educational buildings include classrooms, educational offices, and study zones. Residential buildings include single-family houses, apartments, and dormitories.

Occupant behavior data in this database include door status (on/off), fan status (on/off), window status (on/off), shade status (on/off), occupant number, lighting status (on/off), occupant presence (occupied or not), plug load (in watts), indoor measurements, outdoor measurements, and other types of study. Each type of measurement has a CSV template file associated with it as shown in Table 1. Based on the templates, all the raw data were pre-processed to be consistent in standard naming, data types and formats. The data types follow the entities and tags defined in the Brick schema, which is covered in the following section.

Database query builder

One of the main features of the ASHRAE Global Occupant Behavior Database is the query builder. It allows users to select and download data from different studies, filtered by behavior types and multiple other criteria. Figure 4 shows the web interface of this query builder with 5 different steps in the same figure. Step 1 shows a list of all behaviors in this database, one or more types of behaviors can be selected. Step 2 returns a list of countries and locations associated with available studies based on previous selections. Step 3 presents the available building types from selections made in step 2. Step 4 returns a list of publications of available studies. Once the user clicks “FINISH” button in step

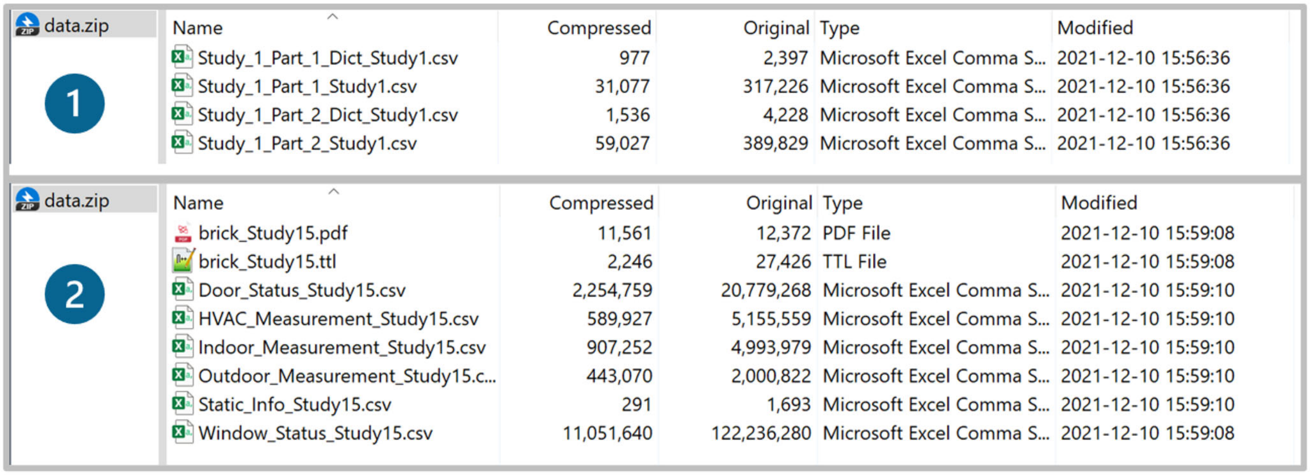

Fig. 5. Output data file from the query builder.
1-Ouput data for survey and mixed types of study; 2-Output data for in-situ type of study.

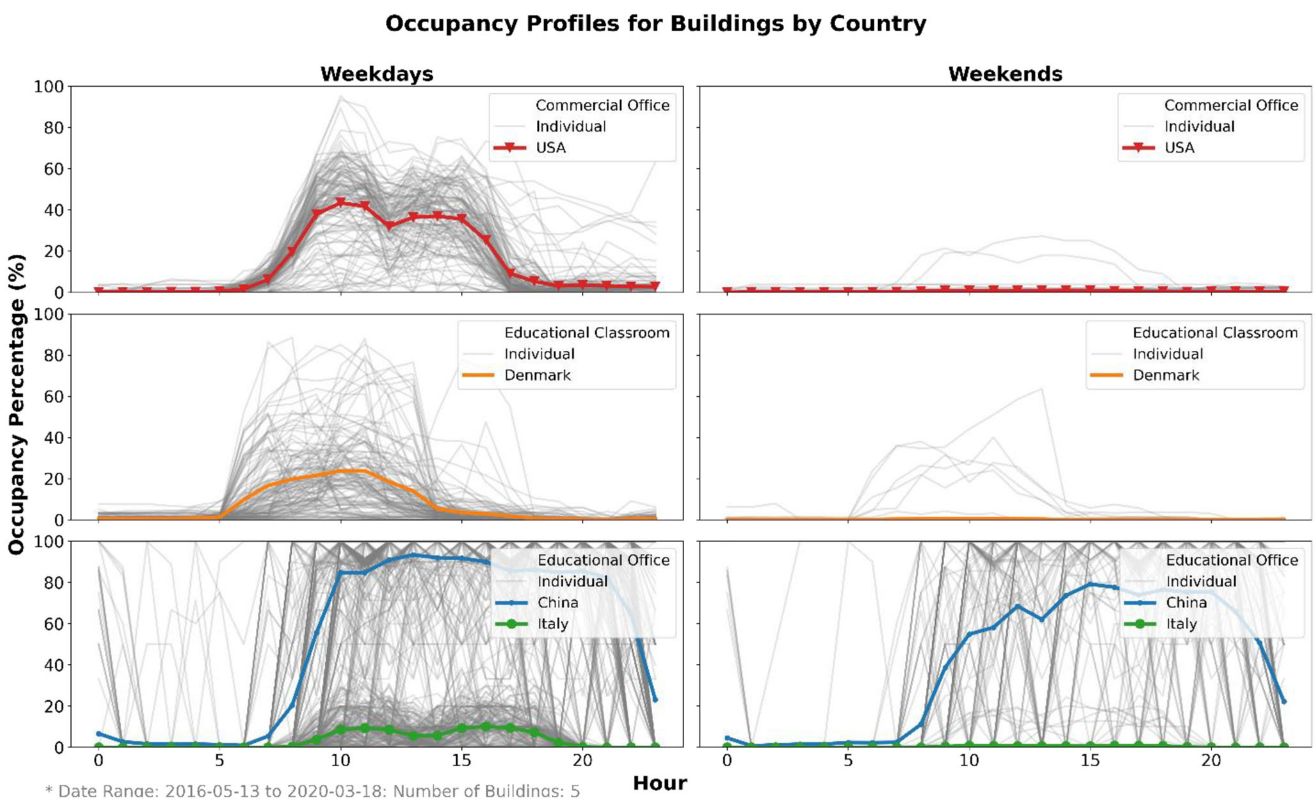

Fig. 6. Daily occupant number profiles by building type and country.

5, the query builder will pass all selected parameters to the server, and a compressed file will be returned from the server for download. For the survey and mixed types of dataset as Figure 5 step 1 shows, the output file includes processed data and a dictionary either in a separate file or within the data file, the dictionary provides detailed information of the different types of data collected in this study. For the in-situ type of dataset as Figure 5 step 2 shows, the output file includes static information of this study, measurement data of different behavior types, processed data, and the Brick model. Among those files, the static information provides detailed information of the location, building types, room numbers; the Brick model includes a Turtle file and a PDF document that visually summarized the metadata information of the study.

The available options in each step are decided by selected parameters in all previous steps. Parameters are organized into the following steps:

- Behavior: the behavior types that the user chooses to download.
- Location: countries and cities, combined with climate zone information.
- Building: building types and room types.
- Study: available studies based on selected parameters in previous steps.
- Export: button to download the data depending on selected study IDs and behavior types.

Data analysis

In this section, we explore the current datasets and perform various analysis on nine different occupant behavior patterns: 1) occupant number measurements; 2) plug load; 3) occupant presence; 4) window status; 5) door status; 5) lighting status; 6) shading status; 7) HVAC status including both cooling and heating scenarios; 9) fan status. Since the database covers occupant behavior patterns globally, the analysis was implemented at different aspects, including by country, by climate zone, and by building type. Since the occupant number measurements and plug load both include continuous measurements in the raw data with constantly changes over time, the individual daily profiles are included in the figures. However, for other behavior types that focus on the status (either 0 or 1) of the behavior, the results are presented in aggregate by different aspects.

Occupant number measurements

Occupant number measurements were collected by six different studies including both in-situ and survey datasets, from five individual buildings located in four countries (China, Denmark, Italy, and USA) between May 13, 2016, and March 18, 2020. The data includes commercial offices, education classrooms, and educational offices and was analyzed at the room level by building type. For each room, occupant numbers were divided by the historical maximum occupant number in the dataset and get the percentage of occupancy by hour of the day. Figure 6 illustrates the daily occupant number profiles for each building type by different countries at different hours of the day, with grey lines representing individual measurements and colored lines representing the averaged data by country. Arrival time and departure time of occupants to the space can be estimated from the data. Weekdays and weekends' difference can be clearly observed from commercial offices and educational classrooms, around zero percentage of average occupancy on weekends. But educational offices in China have shown a relatively higher occupancy percentage compared with the other two building types. Furthermore, the occupancy patterns for buildings in the USA, Denmark, and Italy showed a clear distinction between weekdays and weekends, while buildings in China have a high occupancy rate during both weekdays and weekends. And the dataset from Italy showed relatively low occupancy percentage both on weekdays and weekends. Figures A1 and A2 in Appendix A further break down the results by building type and by climate zone. Additionally, the data covers four different climate zones as shown in Figure A2, as well as one unknown climate zone due to a lack of information from the contributors. The results by climate zone showed similar trends as those by country.

Plug load

Plug load data was collected from six buildings located in Austria, Italy, the United Arab Emirates, and the USA between 2013 and 2020. The data analysis included commercial and educational offices at the building level. For each building, plug load measurements were divided by the historical maximum value in the dataset to get the percentage of load by hour of the day. Figure 7 illustrates the daily plug load profiles for each building type by country at different hours of the day, with grey lines representing individual measurements and colored lines representing the averaged data. Based on the load changes, one can estimate the arrival and departure time of occupant at the space. Weekdays and weekends' difference can be clearly observed for both building types, with a relatively low percentage of average plug load on weekends. The plug load patterns for buildings in Austria, Italy, and USA showed a clear distinction between weekdays and weekends, while buildings in UAE have similar patterns on both weekdays and weekends. Figures A3 and A4 in Appendix A further break down the results by building type and by climate zone. Six climate zones were identified from this data as shown in Figure A4. The load patterns by climate zone showed slightly different patterns as those by country because of multiple climate zones were covered by data from Italy.

Occupant presence

Occupant presence data was collected by nine in-situ datasets from 11 buildings located in seven different countries (Australia, Canada, China, Germany, Italy, UAE, and USA) between January 3, 2005, and July 23, 2020. The data covers commercial offices, educational classrooms, educational offices, and residential single-family houses. The behavior patterns were analyzed at the building level. For each building, the mean values were calculated for different hours of the day as the probability of a space was occupied. Figure 8 illustrates the daily occupant presence profiles for each building type by country at different hours of the day. From the changes of the occupancy pattern, one can estimate the arrival and departure time of occupant to the space. This can assist to model more realistic occupant behaviors in building energy simulation process. The data indicates a clear distinction between weekdays and weekends for commercial offices, educational classrooms, and educational offices, with a low probability of occupancy on weekends. However, residential single-family houses showed no significant difference between weekdays and weekends during the data collection period. The occupant presence patterns vary a lot based on the building type and country. Overall, commercial offices and educational classrooms showed distinct differences for weekdays and weekends. For educational offices, buildings

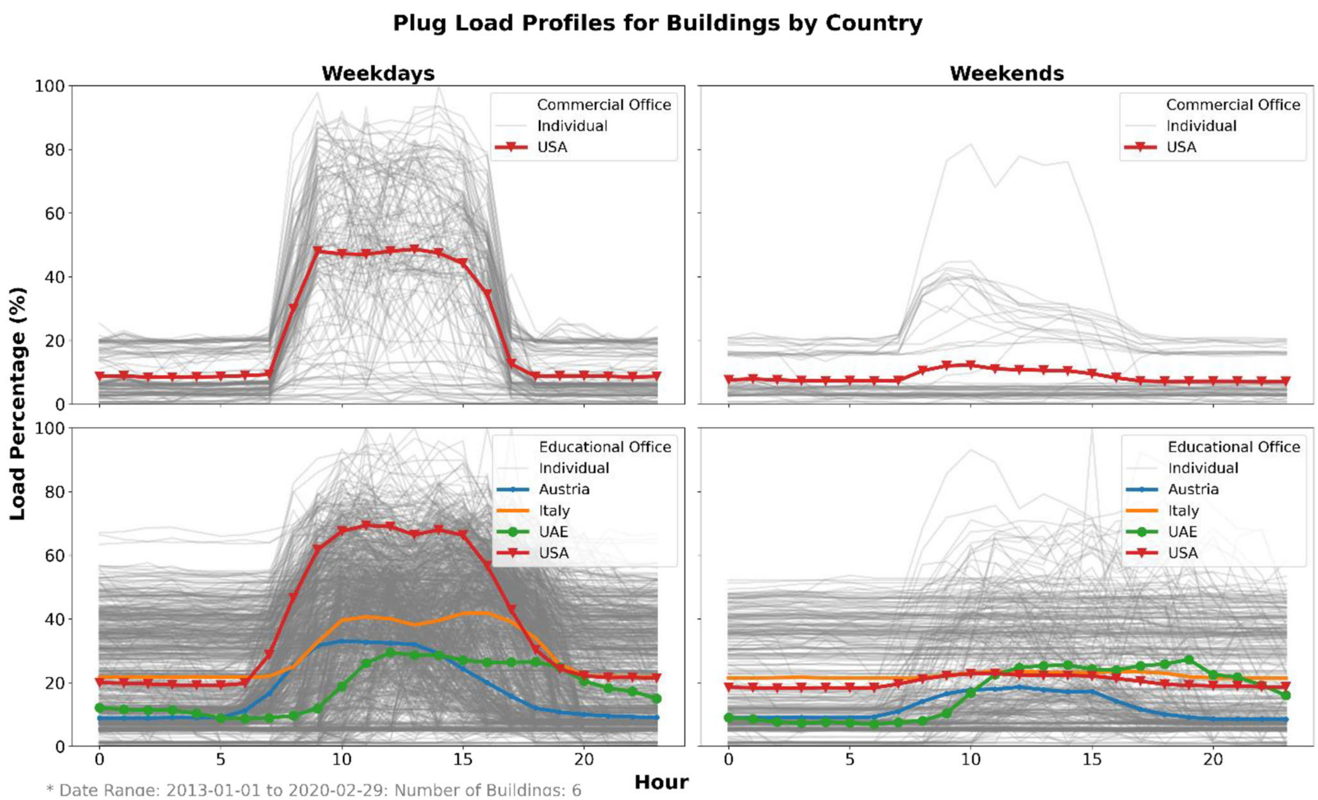

Fig. 7. Daily plug load profiles by building type and country.

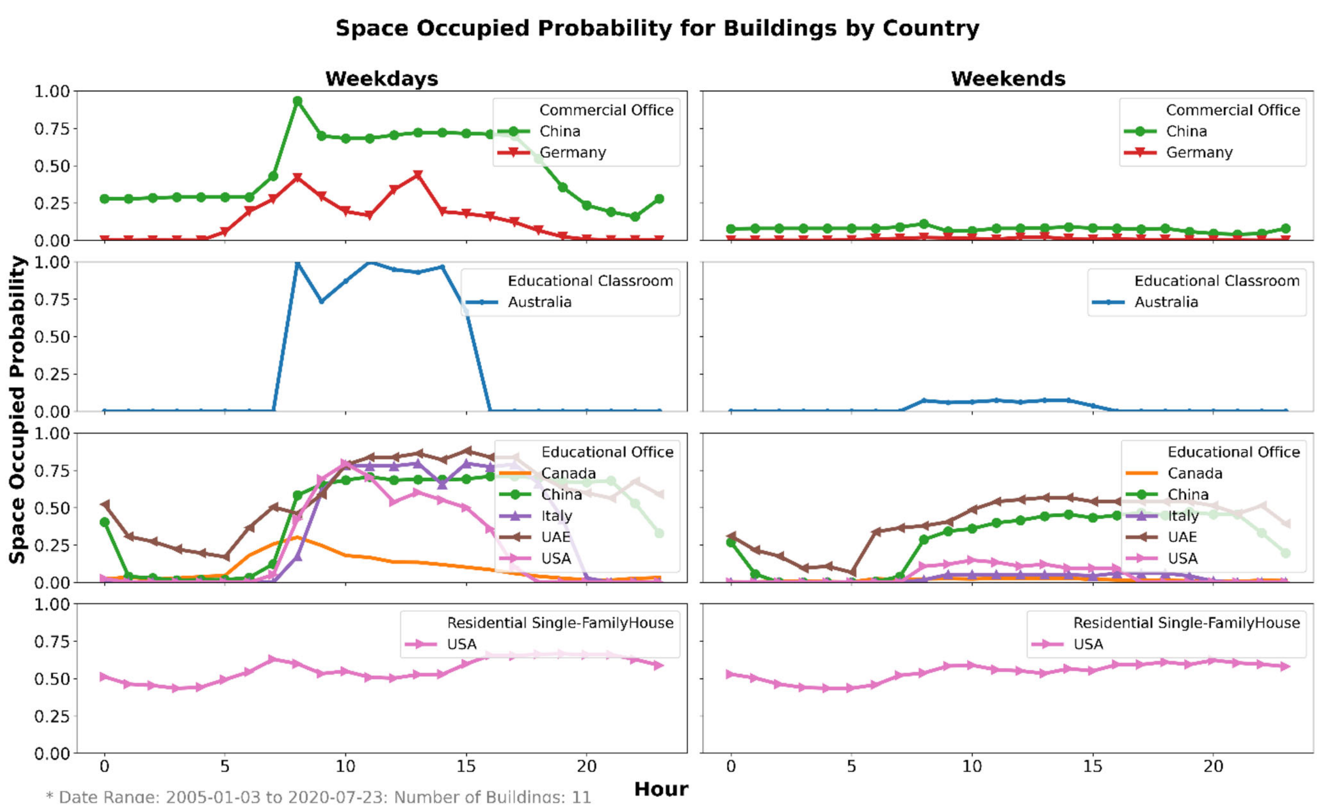

Fig. 8. Daily occupant presence profiles status by building type and country.

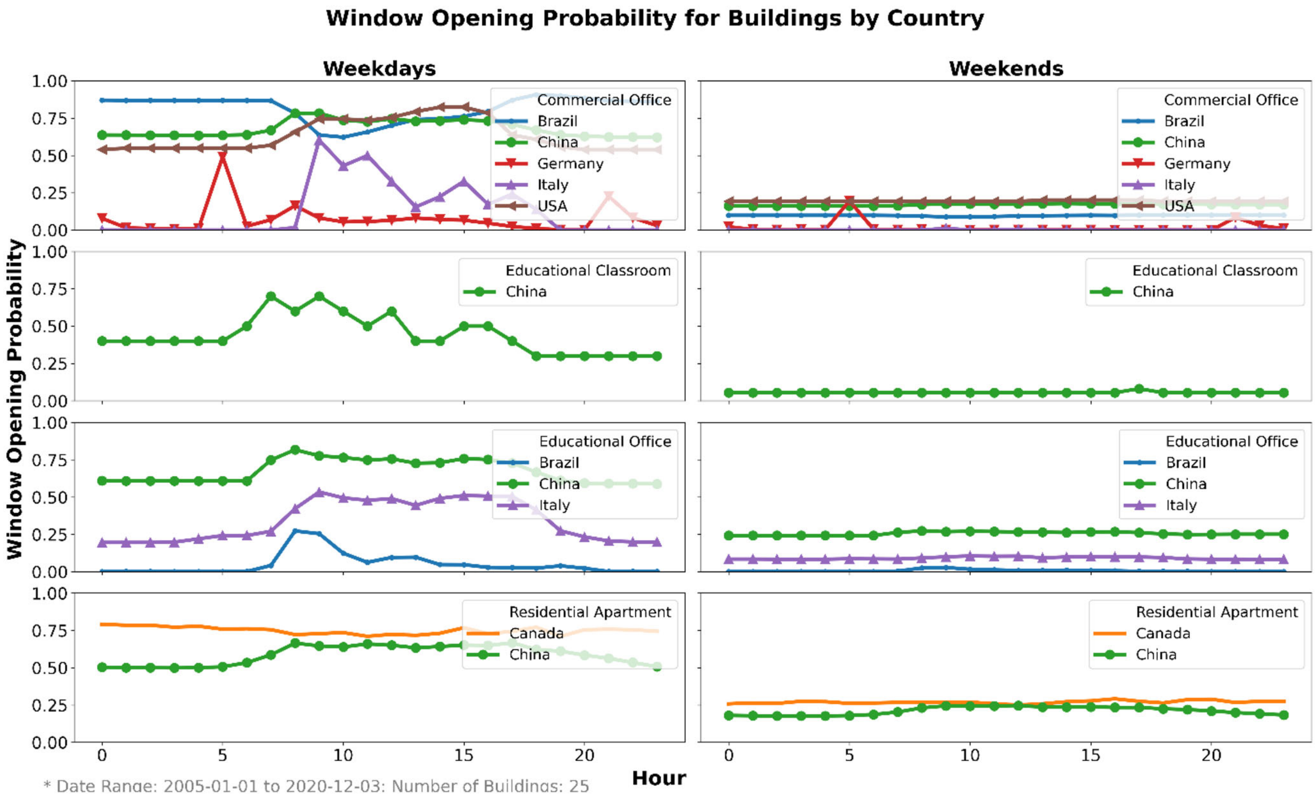

Fig. 9. Daily window opening profiles by building type and country.

in Canada, Italy, and USA showed clear difference between weekdays and weekends. Figures A5 and A6 in Appendix A further break down the results by building type and by climate zone. Six climate zones were identified from this data.

Window status

Data on window opening patterns was collected by 18 datasets which cover in-situ, survey, and mixed types. The data was from 25 buildings located in six different countries (Brazil, Canada, China, Germany, Italy, and USA) between January 1, 2005, and December 3, 2020. The data covers commercial offices, educational classrooms, educational offices, and residential apartments. The behavior patterns were analyzed at the building level. Using historical data, the probability of window opening behavior was calculated for each window in every building. The daily profiles of window opening behavior were obtained by averaging the results at the building level. Figure 9 illustrates the window opening profiles for each building type grouped by country at different hours of the day. Weekdays and weekends' difference can be clearly observed for all types of buildings, with quite a low percentage of window opening probability on weekends. On weekdays, most window opening behavior happened during the daytime, early morning and late night peaks were observed for commercial offices. Figures A7 and A8 in Appendix A further break down the results by building type and by climate zone. Six climate zones were identified from this data. As expected, results showed many variations of window operation patterns among different countries and climate zones. On weekdays, data in China and Brazil showed higher probability of window opening compared to data from other countries, while Canada had relatively low window operation activity at all times.

Door status

Data on door opening patterns was collected by seven datasets which cover in-situ, survey, and mixed types. The data was from 15 buildings located in four different countries (Canada, China, Italy, and USA) between June 6, 2008, and March 15, 2019. The data covers commercial offices, educational offices, and residential apartments. The behavior patterns were analyzed at the building level. For each building, the mean values were calculated for different hours of the day as the probability of a door that was opened. Figure 10 illustrates the door opening profiles for each building type grouped by country at different hours of the day. Weekdays and weekends' difference can be observed for commercial offices and educational offices, with quite a low percentage of door opening probability on weekends. And most door opening behavior occurred during daytime on weekdays. However, residential apartments showed flat patterns regardless the day of week with higher probability that doors are constantly open on weekdays. Figures A9 and A10 in Appendix A further

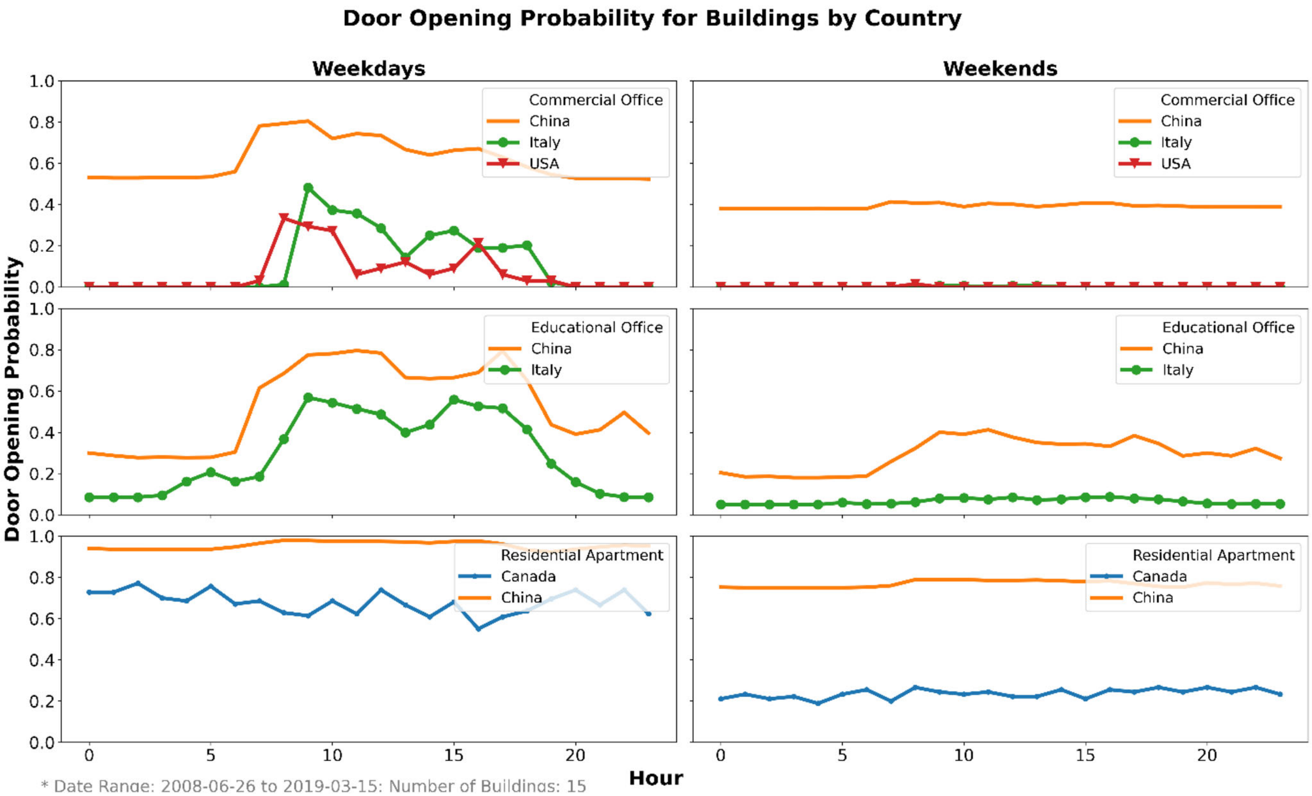

Fig. 10. Daily door opening profiles by building type and country.

break down the results by building type and by climate zone. Four climate zones were identified from this data including Cfa, Cfb, Csa, and Dfb. Patterns of residential apartment from Canada and China both showed relatively flat curve compared to other countries. This suggests a low probability of door operations being recorded in the data, and that doors were likely (80%-90%) to remain open on weekdays. Educational offices in climate zones Cfa, Cfb, and Csa showed similar patterns, indicating that an estimation of weekday office schedules could be extracted.

Lighting status

Data of lighting on patterns was collected by four datasets which cover in-situ and survey types. The data was from six buildings located in four different countries (Brazil, Canada, Italy, and USA) between June 28, 2008, and December 3, 2020. It covers commercial offices and educational offices. The behavior patterns were analyzed at the building level. For each building, the mean values were calculated for different hours of a day as the probability of a light that was turned on. Figure 11 illustrates the lighting on profiles for each building type grouped by country at different hours of the day. Weekdays and weekends' difference can be observed for both building types, with relatively low percentages of lighting on probability on weekends. And most lighting on operations occurred during daytime on weekdays. The results of aggregated data showed low probability of lighting on behaviors as the dataset from Canada has a significant influence on the overall results, while this dataset has lower profile pattern as shown in the results by country. Figures A11 and A12 in Appendix A further break down the results by building type and by climate zone. Three climate zones were identified from this data which are Cfa, Cfb, and Dfb. Compare with data from Canada, data from USA showed higher lighting on operations during the daytime on weekdays, followed by data from Italy. All results showed clear differences between weekdays and weekends regardless of country or climate zone.

Shading status

Data on shading operation patterns was collected by two datasets which cover in-situ and survey types. The data covers four buildings located in two different countries (Italy and USA) between June 28, 2008, and April 24, 2013. It covers commercial offices and educational offices. The behavior patterns were analyzed at the building level. For each building, the mean values were calculated for different hours of a day as the probability of a shading device that was on. Figure 12 illustrates the shading on status profiles for each building type grouped by country at different hours of the day. Weekdays and weekends' difference can be observed for both building types, with quite low percentages

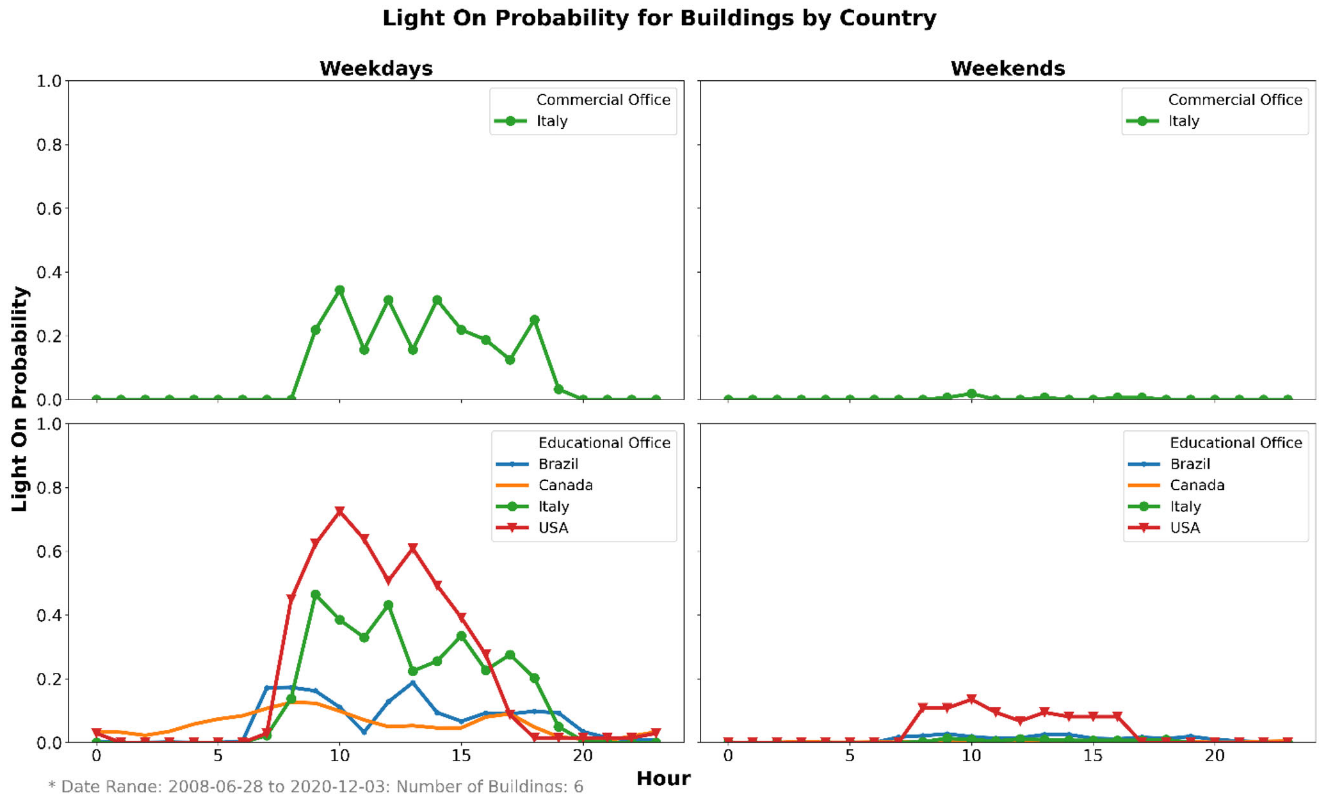

Fig. 11. Daily lighting on profiles by building type and country.

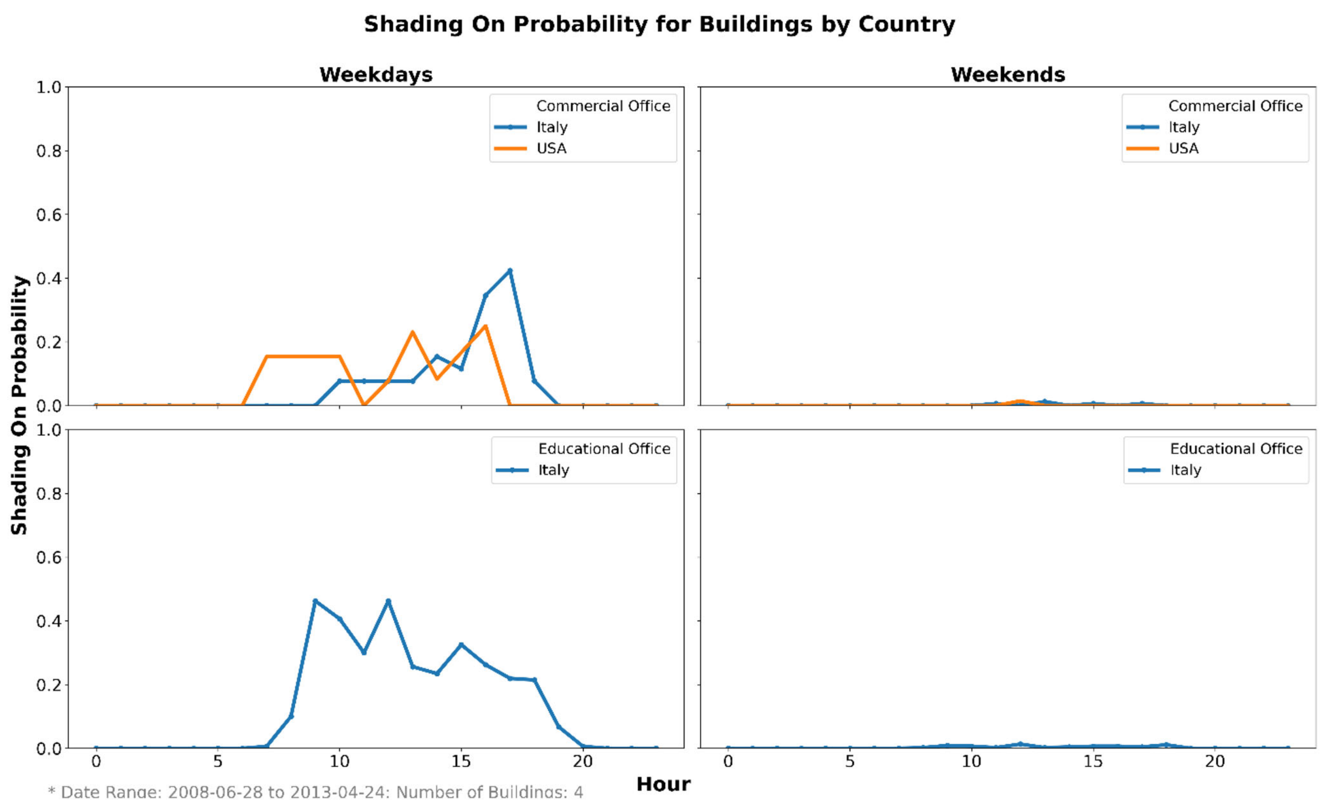

Fig. 12. Daily shading on profiles by building type and country.

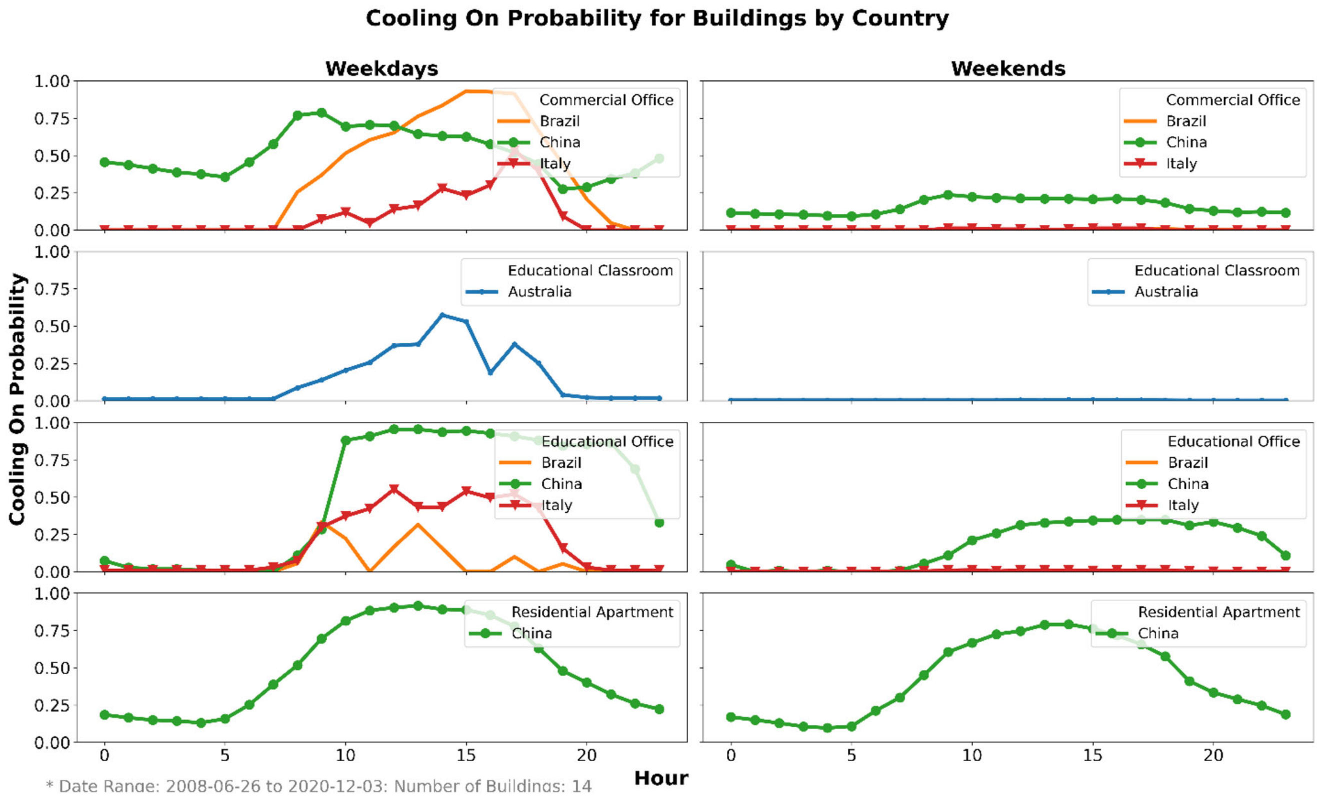

Fig. 13. Daily HVAC status (cooling) profiles by building type and country.

of shading on probability on weekends. And most shading on operations occurred during daytime on weekdays. Morning peaks can be observed in educational office buildings, while evening peaks were observed in commercial office buildings. Figures A13 and A14 in Appendix A further break down the results by building type and by climate zone. Two climate zones were identified from this data which are Cfa and Cfb. Compare with data from USA, data from Italy showed more variations in shading operation behavior and covered more climate zones. All results showed clear differences between weekdays and weekends regardless of country or climate zone. With only two datasets in the database measuring shading status, acquiring additional datasets is expected to enhance our understanding of shading operation behavior.

HVAC status

HVAC cooling

Data on HVAC cooling status was collected by 13 datasets which cover in-situ, mixed, and survey types. The data was from 14 buildings located in four different countries (Australia, Brazil, China, and Italy) between June 26, 2008, and December 3, 2020. It covers commercial offices, educational classrooms, educational offices, and residential apartments. The behavior patterns were analyzed at the building level. For each building, the mean values were calculated for different hours of a day as the probability of the HVAC cooling that was on. Figure 13 illustrates the profiles of HVAC cooling status for each building type grouped by country at different hours of the day. Weekdays and weekends' difference can be observed for nonresidential building types, with quite low percentages of cooling on probability on weekends. And most cooling operations occurred during daytime on weekdays. However, data from residential apartments showed similar patterns for both weekdays and weekends. Both commercial offices and residential apartments have a relatively high probability of cooling on during nighttime. Figures A15 and A16 in Appendix A further break down the results by building type and by climate zone. Three climate zones were identified from this data which are Cfa, Cfb, and Csa. Variations among countries and climate zones were observed, with commercial offices in Brazil and China more likely to have HVAC cooling on during daytime compared to Italy. Educational offices covered more climate zones than other building types.

HVAC heating

Data on HVAC heating status was collected by 13 datasets which cover in-situ, mixed, and survey types. The data was from 7 buildings located in five different countries (Australia, Brazil, China, Italy, and USA) between June 28, 2008, and March 23, 2020. It covers commercial offices, educational classrooms, and educational offices. The behavior patterns were analyzed at the building level. For each building, the mean values were calculated for different hours

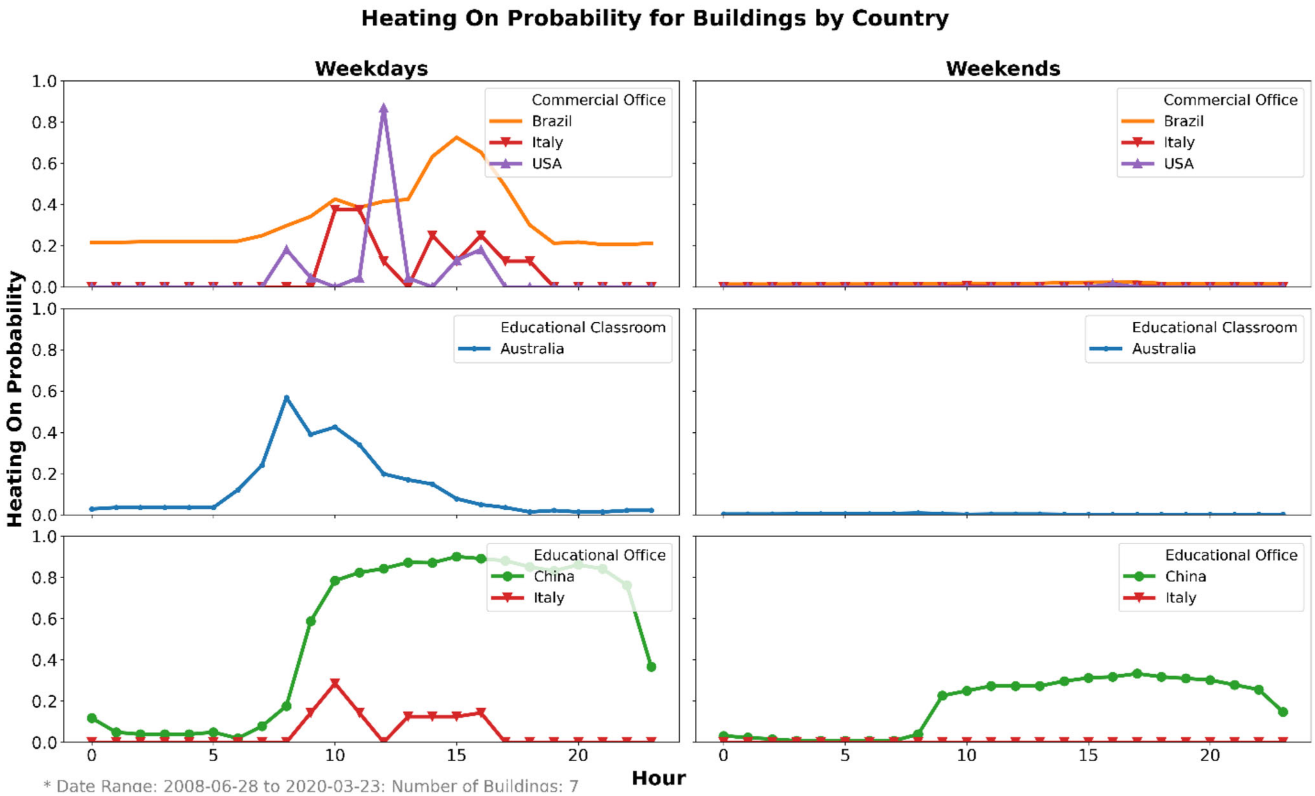

Fig. 14. Daily HVAC status (heating) profiles by building type and country.

of a day as the probability of the HVAC heating that was on. Figure 14 illustrates the profiles of HVAC heating status for each building type grouped by country at different hours of the day. Weekdays and weekends' difference can be observed for all building types, with quite low percentages of heating on probability on weekends. And most heating operations occurred during daytime on weekdays. But educational offices are more likely to have heating on for longer time compared with other building types. Figures A17 and A18 in Appendix A further break down the results by building type and by climate zone. Three climate zones were identified from this data which are Aw, Cfa, and Cfb. Variations among countries and climate zones were observed, with commercial offices in Brazil and educational offices in China are more likely to have HVAC heating on with longer times during daytime compared to Italy. Climate zones Aw and Cfa showed higher probability and longer duration of heating status on.

Fan status

Data on fan operation patterns were collected by three datasets which cover in-situ, survey, and mixed types. The data was from six buildings located in three different countries (Brazil, Italy, and USA) between June 28, 2008, and February 18, 2016. The data only covers commercial offices and educational offices. The behavior patterns were analyzed at the building level. For each building, the mean values were calculated for different hours of a day as the probability of a fan was on. Figure 15 illustrates the profiles of fan on status for each building type grouped by country at different hours of the day. Weekdays and weekends' difference can be observed for commercial offices and educational offices, with quite a low percentage of fan on status probability on weekends. And most fan operation behavior occurred during daytime on weekdays. Results showed that commercial offices have a high probability of fan on status compared with educational offices. Figures A19 and A20 in Appendix A further break down the results by building type and by climate zone. Two climate zones were identified from this data including Cfa and Cfb. Commercial offices have data from all three countries and showed three unique patterns of fan operation behavior patterns. Educational offices covered both climate zones with similar patterns and relatively low probability shown by each climate zone.

Case study

To demonstrate the efficacy and usefulness of the collected data, a case study was conducted using the datasets of occupant presence from commercial office, educational office, and educational classroom. We have derived occupancy profiles from one month in summer (July) and one month in winter (December). The profiles were incorporated into EnergyPlus simulation software (EnergyPlus^{TM} 2017) and

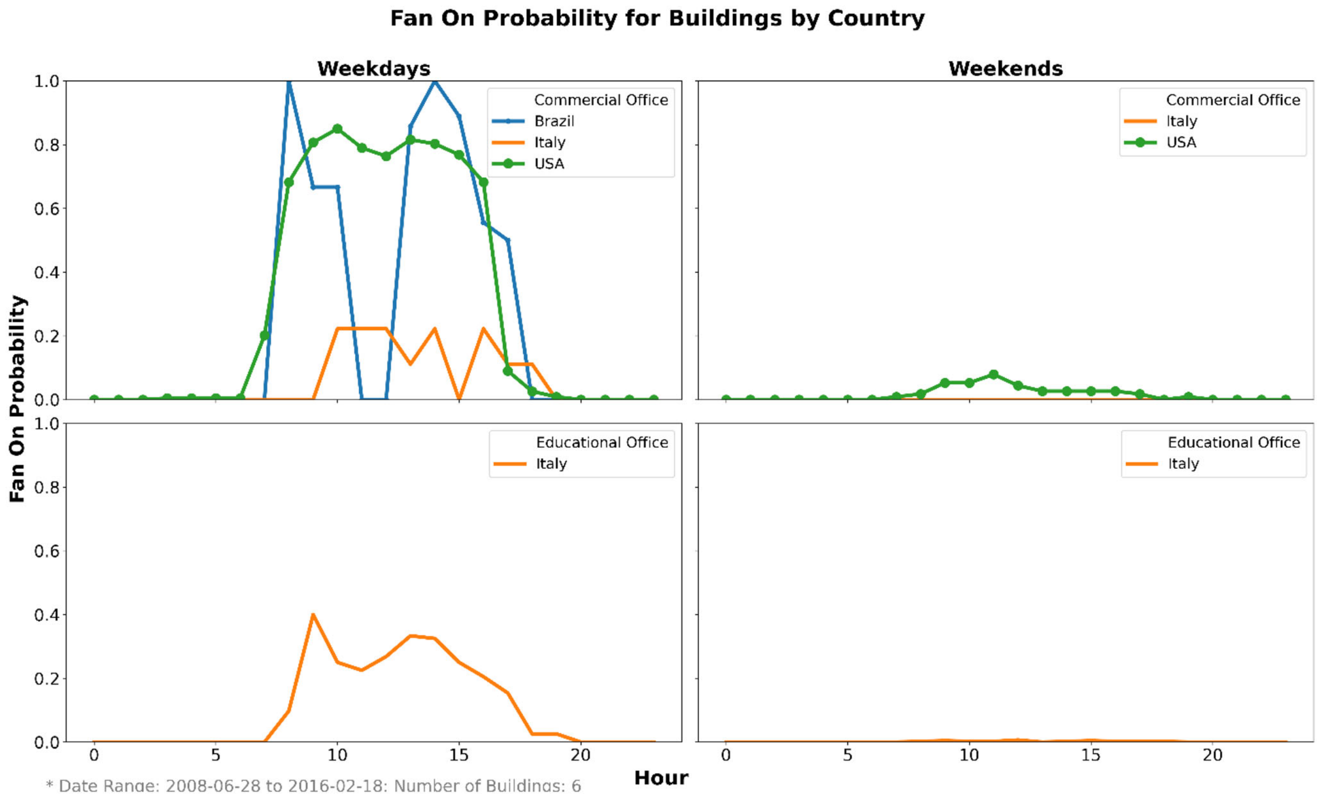

Fig. 15. Daily fan status profiles by building type and country.

compared with the default occupancy schedule, to investigate the difference on building overall electricity consumption. Since this case study focuses on the demonstration of collected dataset, we only replaced the default building-level building equipment schedule, lighting schedule, and occupancy schedule in the simulations. When the space is unoccupied, those schedules were set as its minimum in the simulation. DOE prototype building models (DOE 2023) under ASHRAE standard 90.1 (ASHRAE 90.1 2019) were used to run the simulations, including educational classroom represented by primary school model, educational office represented by small office model, and commercial office represented by medium office model. Both the building models and weather files cover the site of New York City.

Commercial office

Dataset 24 provides event-based occupant presence measurements which covers 17 different rooms in one commercial office building at Frankfurt, Germany. The data collection ranges from January 2005 to December 2008, resulted in 48,458 rows of measurements. We have picked the typical data from one of the rooms in July and December 2006 as input to the EnergyPlus simulations. The building prototype model is ASHRAE901 Medium Office at New York. It can be observed that the occupancy schedule from this data set (Occ Real) follows similar trends as the default schedule (Occ Default) in the beginning of July. However, there are significant variations since the space was empty most of the time after July 12th. As shown in Figure 16, building overall electricity consumption was reduced by 27.68% based on the Occ Real schedule. Winter simulation resulted in about 8.21% reduction in overall building electricity use as shown in Figure 17. Clear difference between default and real occupancy schedule can also be observed in the figure.

Educational office

Dataset 10 covers the minute level measurements of occupancy presence in one educational office building from Rende, Italy. The data collection period started from May 13, 2006 to May 12, 2017, with 525,600 rows of measurements. ASHRAE901 Small Office at New York prototype model is used to represent this educational building. Figure 18 shows that the real occupancy profile varies a lot from the default schedule in Summer and resulted in around 21.13% less overall building electricity consumption in the simulation. Winter simulation results in Figure 19 shows that the default schedule and real schedule have many overlaps, but the latter one has more variations. It resulted in around 13.23% less overall building electricity consumption.

Educational classroom

For educational classrooms, we picked dataset 7 from the database which is from Bangholme, Australia. It has 16 different classrooms measured every five minutes in an educational building. The date spans from October 2019 to March 2020. Since this dataset does not cover July, only the winter simulation was implemented. This building was represented

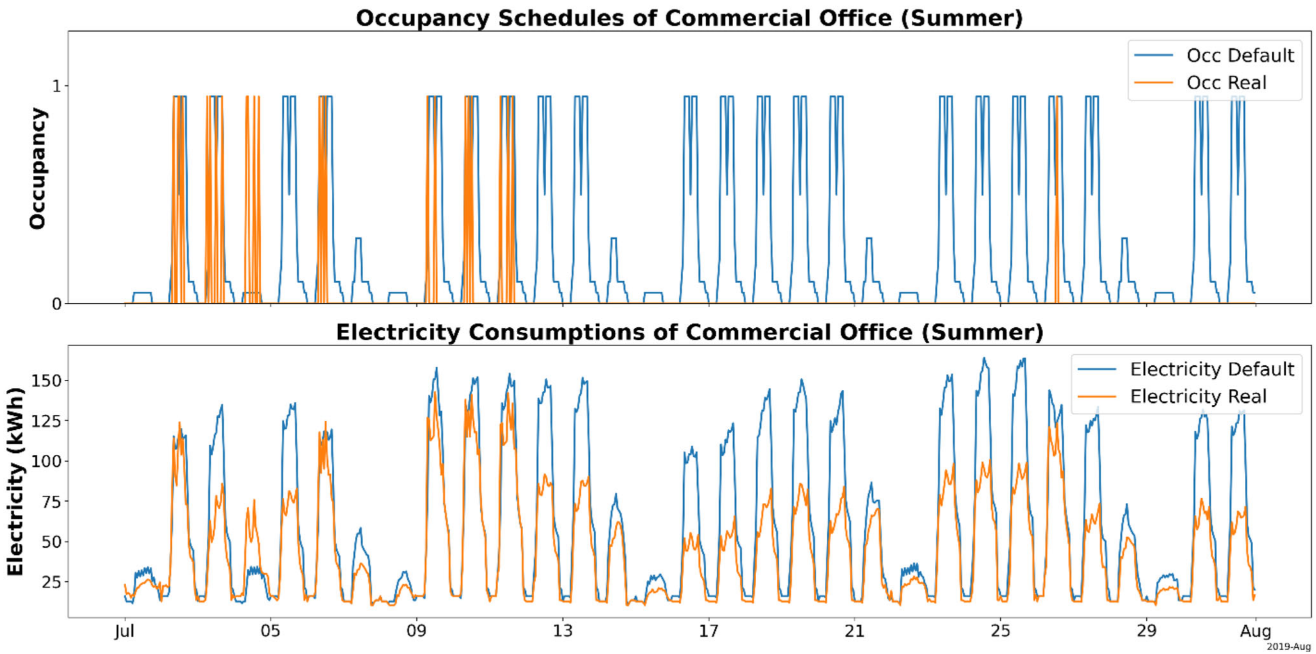

Fig. 16. Comparison of building occupancy schedule and overall electricity consumption (Commercial office in Summer).

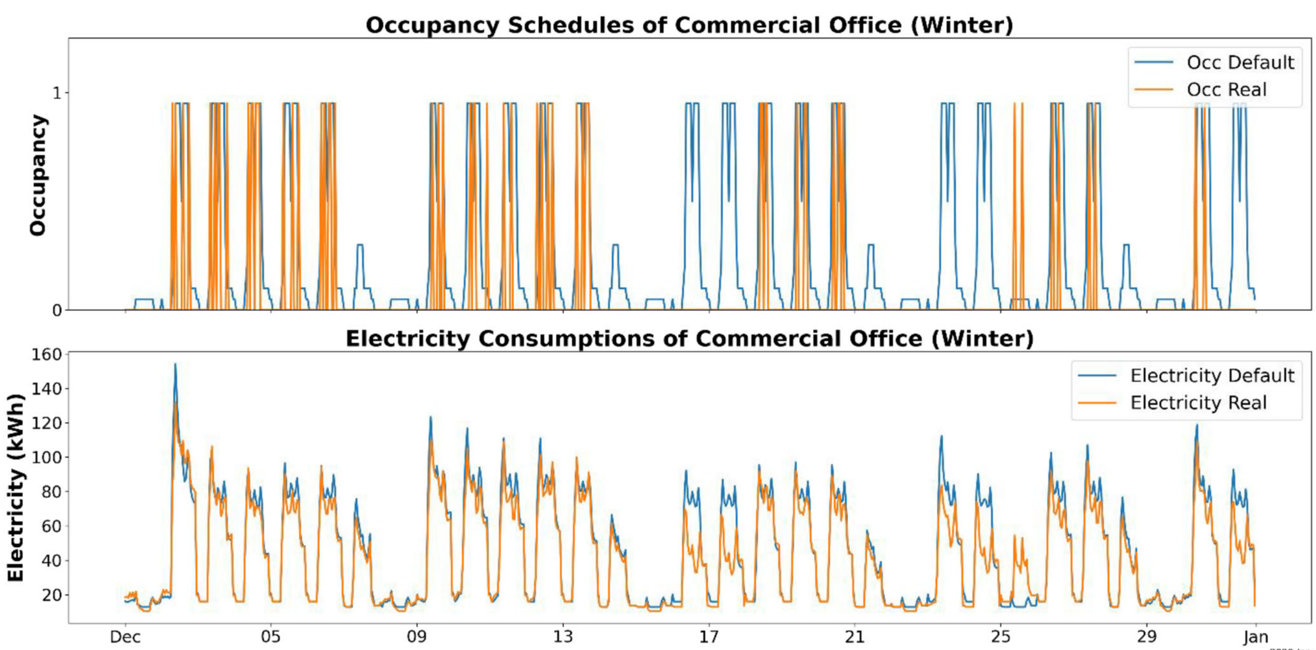

Fig. 17. Comparison of building occupancy schedule and overall electricity consumption (Commercial office in Winter).

by the ASHRAE901 Primary School at New York prototype model. Following the procedure defined in the beginning of this section, simulation was implemented in EnergyPlus software. As shown in the Figure 20, this building was unoccupied most of the time in December. Compared with the default occupancy schedule, the overall building electricity consumption was reduced by about 22.46% using the real occupancy data.

As an effort to investigate the efficacy and usefulness of the collected datasets in this global occupancy behavior database, we implemented the above simulations in EnergyPlus software. Results showed that the occupancy schedules derived from the database differ significantly from the default schedule, which leads to electricity consumption differences. The realistic occupant behavior data provides researchers with a better understanding of HBI, and also provides energy savings. We believe this

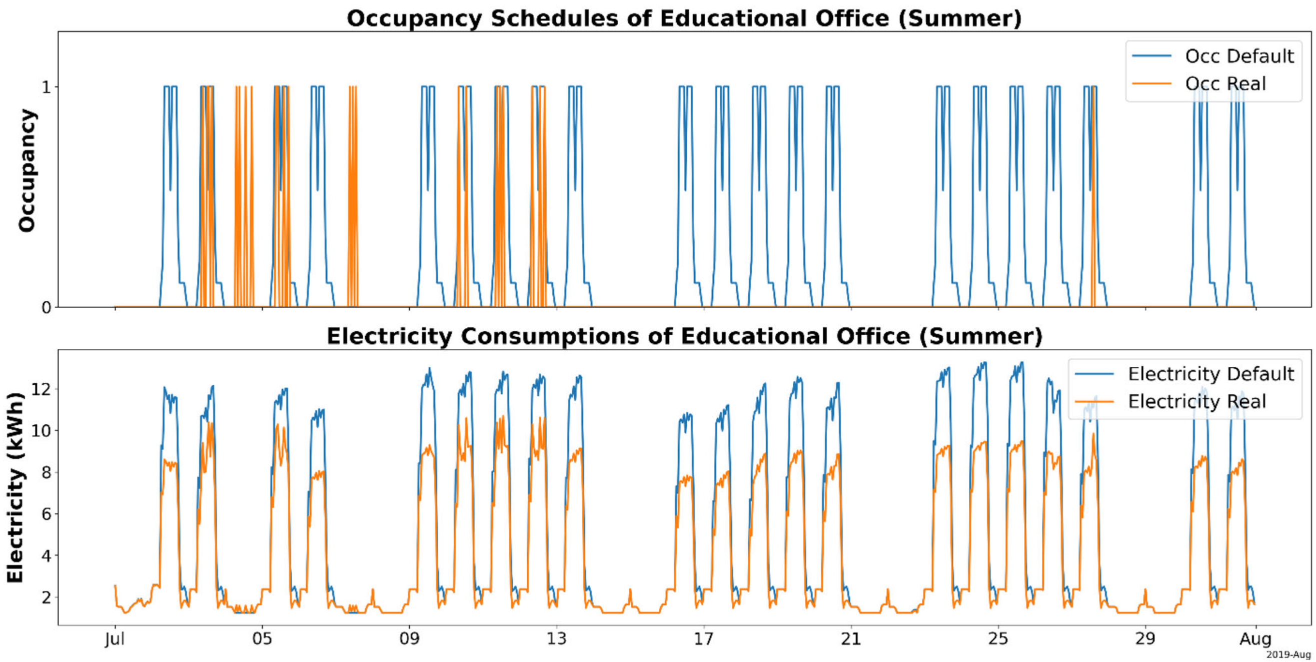

Fig. 18. Comparison of building occupancy schedule and overall electricity consumption (Educational office in Summer).

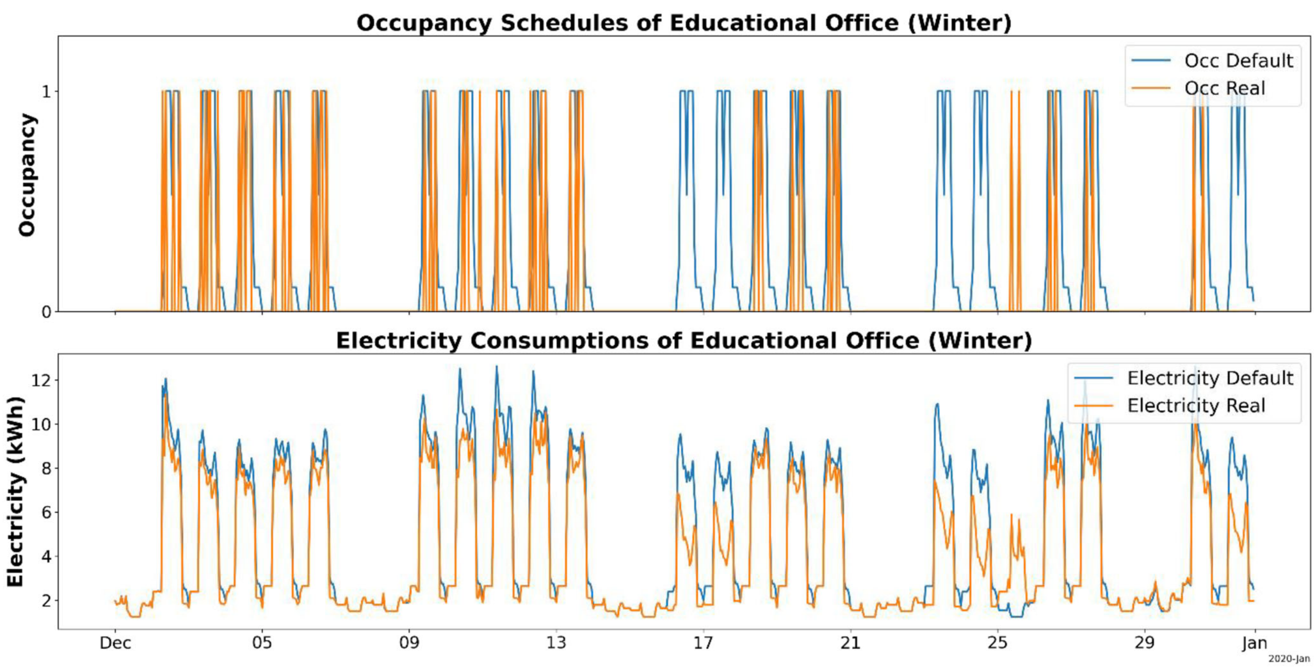

Fig. 19. Comparison of building occupancy schedule and overall electricity consumption (Educational office in Winter).

highlights the importance of collecting field measurement data for building energy simulations, since the default values are out of touch with practicality. With the database we have developed, researchers will have the opportunity to dive deeper into their work with the vast amount of data available, allowing them to compare occupant behaviors across different building types and countries. This will enable them to derive valuable insights that can inform the design and operation of energy-efficient buildings.

Discussions and limitations

This study has conducted high level analysis of nine occupant behavior types based on available data from the database. The occupant behavior patterns were examined in relation to different building types, countries, and climate zones. The results showed that different building types have distinct behavior patterns, which is expected. And these patterns can vary between different countries or climate zones. This indicates that the developed

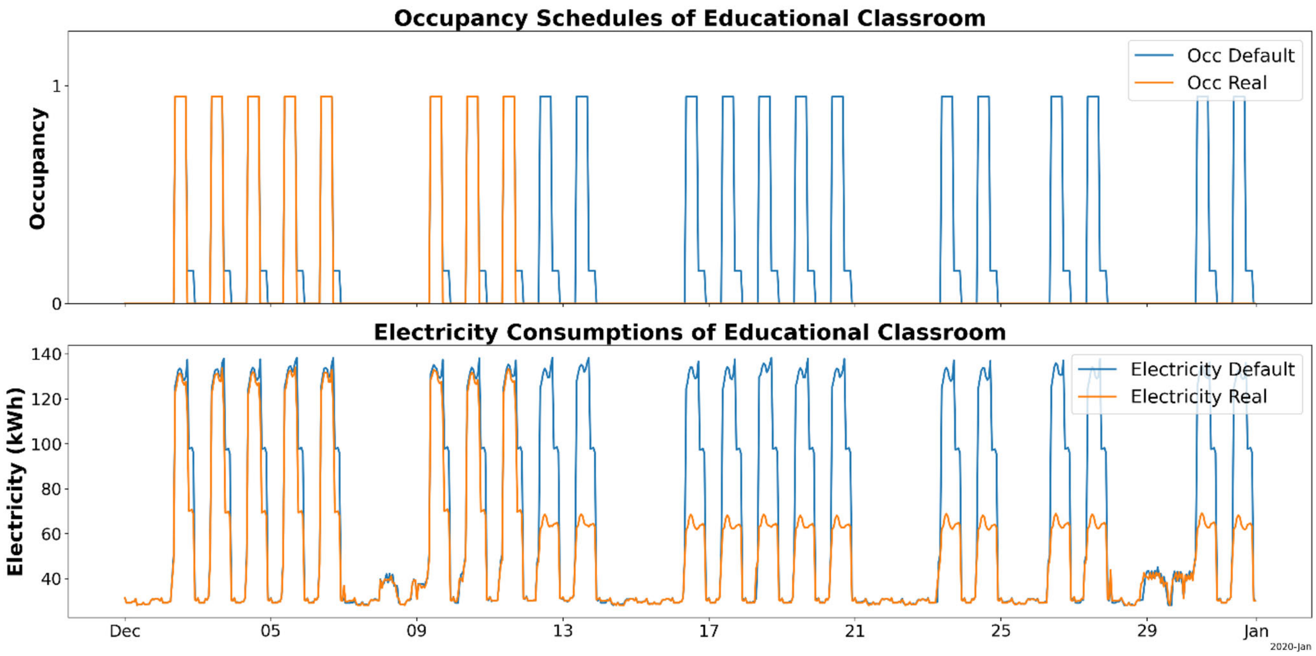

Fig. 20. Comparison of building occupancy schedule and overall electricity consumption (Educational classroom in Winter).

database contains abundant characteristics of the occupant behaviors. We found that in some case, the occupant behavior patterns by country were closely related to those by climate zones, as shown in Figures A17 and A18. This is due to the limited study of both country level and climate zone level data. Fluctuations have been also observed in the resulted profiles at country level and climate zone level for some types of behaviors, which can be attributed to the discrete time stamp in the survey and mixed types of datasets. Under this case, with more continuous measurement data in the future, these profiles can be smoothed and be more representative. Overall, we believe this study will contribute to understanding of occupant behavior simulation worldwide.

The limitations of this study include: 1) Seasonal effect is not analyzed since this study focused on investigating the occupant behavior patterns at a large spatial scale; 2) Even though this database has collected many datasets around the world, it is still limited in terms of the amount of data, so the results may not be very well representative for some behavior types or specific climate zone; 3) More detailed analysis like dataset by dataset comparison were not performed in this study, which can be the focus of future work; 4) The current case study only focuses on occupant number measurements; however, future work can include detailed case studies encompassing all behavior types.

Conclusions

This paper details out the development of the ASHRAE Global Occupant Behavior Database, including data collection and processing methods, data warehouse development for public access, query builders for users to select and download datasets The database also provides templates that were used to process the raw data, future data contributors are encouraged to follow the templates when contributing. Currently, the database has 34 field-measured building occupant behavior datasets contributed by researchers from 15 countries and 39 institutions, the datasets cover 10 different climatic zones and various building types in both commercial and residential sectors. This paper focused on conducting a comprehensive data analysis to examine patterns related to nine different occupant behavior types. We investigate various factors that influence these patterns, including building type, country, and climate zone. By implementing EnergyPlus simulations based on the occupancy profiles derived from this database, we demonstrate that significant reductions in overall building electricity consumption can be achieved. Specifically, our results indicate a potential reduction of approximately 27% during summer and around 10% during winter. Based on the analysis and simulations results, we can conclude that the availability of realistic occupant behavior data enhances researchers' comprehension of human-building interactions, leading to energy savings. This underscores the significance of gathering field measurement data for building energy simulations, as default values often lack practicality. Our developed database offers researchers an extensive collection of data, enabling them to delve deeper into their investigations. By comparing occupant behaviors across diverse building types and countries, valuable insights can be derived. These insights, in turn, can inform the design and operation of energy-efficient buildings.

Align with the scope of the ASHRAE Multidisciplinary Task Group of Occupant Behavior in Buildings (MTG.OBB), the intent of this database aims to incorporate the human aspect in the building life cycle and possibly help ASHRAE achieve its energy saving goals, by showing how behavior can influence passive and active building designs and revealing energy saving opportunities from behavior interventions. The database can support various use cases of occupant behavior research, including 1) Understand occupant behaviors in real buildings; 2) Compare and understanding the diversity and dynamics of occupant behaviors; 3) Develop mathematical models of occupant behaviors at various spatial and temporal resolutions by building types; 4) Benchmark various occupant behavior modeling approaches; 5) Generate typical occupant schedules and behavior models for use in building performance simulations, as well as building energy codes and standards.

Nomenclature

Door Status = A Boolean value of the state of a door
Window Status = A Boolean value of the state of a window
Fan Status = A Boolean value of the state of a fan
Shade Status = A Boolean value of the state of a shading device
Lighting Status = A Boolean value of the state of a lighting device
Occupant Number = An integer value of total number of occupants in a space or building
Occupant Presence = A Boolean value of the state of an occupant being in a space
Plug Load = Measurement of power usages from plug-in devices in a commercial building space
HVAC Measurement = Measurement of Heating, Ventilation, and Air Conditioning system typically includes heating or cooling status, temperature setpoint, percentage of variable air volume openings, etc.
Outdoor Measurement = A broad measure of an outdoor environment typically includes temperature, relative humidity, solar radiation, wind direction and speed, etc.
Indoor Measurement = A broad measure of an indoor environment typically including temperature, relative humidity
API = Application Programming Interface
HBI = Human-building interactions

Funding

This work was supported by American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), project number URP-1883. We are grateful for the support from ASHRAE Multidisciplinary Task Group of Occupant Behavior in Buildings (MTG.OBB). We acknowledge the support from Building Technologies Office of the United States Department of Energy, and Annex 79 research community promoted by the International Energy Agency, Energy in Buildings and Communities (IEA-EBC). We also give thanks to Professor Gabe Fierro from the Department of Computer Science at Colorado School of mines, for his kindly supports in developing Brick models.

ORCID

Yapan Liu http://orcid.org/0000-0002-7569-9544
Zheng O'Neill http://orcid.org/0000-0002-8839-7174

Data availability statement

The datasets collect by this project including raw data and metadata as well can be downloaded from a public repository hosted by figshare (https://doi.org/10.6084/m9.figshare.16920118.v6). A website (https://ashraeobdatabase.com) was created to query and download the desired data from the database based on different selection criteria. This database is open to researchers for future datasets, data templates can be found on the website (https://ashraeobdatabase.com/#/template), and contributions can be emailed directly to the corresponding author of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Alishahi, N., M. M. Ouf, and M. Nik-Bakht. 2022. Using WiFi connection counts and camera-based occupancy counts to estimate and predict building occupancy. *Energy and Buildings* 257: 111759. 10.1016/j.enbuild.2021.111759.

Amasyali, K., and N. M. El-Gohary. 2016. Energy-related values and satisfaction levels of residential and office building occupants. *Building and Environment* 95:251–63. 10.1016/j.buildenv.2015.08.005.

Andersen, R. 2012. The influence of occupants' behaviour on energy consumption investigated in 290 identical dwellings and in 35 apartments. In *10th International Conference on Healthy Buildings 2012*. Vol. 3.

ASHRAE 90.1. 2019. https://www.techstreet.com/ashrae/standards/ashrae-90-1-2022-i-p?product_id=2522082

Brackley, C. W. O'Brien, and C. Trudel. 2020. Evaluation of a modern commercial HVAC control system in the context of occupants' perceived control. *ASHRAE Transactions* 126 (2):174–83.

Carlucci, S., M. De Simone, S. K. Firth, M. B. Kjærgaard, R. Markovic, M. S. Rahaman, M. K. Annaqeeb, S. Biandrate, A. Das, J. W. Dziedzic, et al. 2020. Modeling occupant behavior in buildings. *Building and Environment* 174:106768. 10.1016/j.buildenv.2020.106768.

Chang, W.-K., and T. Hong. 2013. Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data. *Building Simulation* 6 (1):23–32. 10.1007/s12273-013-0106-y.

Choi, H., C. Y. Um, K. Kang, H. Kim, and T. Kim. 2021. Application of vision-based occupancy counting method using deep learning and performance analysis. *Energy and Buildings* 252:111389. 10.1016/j.enbuild.2021.111389.

Chow, W. K., and C. M. Y. Ng. 2008. Waiting time in emergency evacuation of crowded public transport terminals. *Safety Science* 46 (5):844–57. 10.1016/j.ssci.2007.01.015.

Cook, D. J., M. Youngblood, E. O. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja. 2003. MavHome: An agent-based smart home. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003)*, 521–4. 10.1109/PERCOM.2003.1192783

Correia da Silva, P., V. Leal, and M. Andersen. 2015. Occupants' behaviour in energy simulation tools: Lessons from a field monitoring campaign regarding lighting and shading control. *Journal of Building Performance Simulation* 8 (5):338–58. 10.1080/19401493.2014.953583.

Ding, Y., X. Ma, S. Wei, and W. Chen. 2020. A prediction model coupling occupant lighting and shading behaviors in private offices. *Energy and Buildings* 216:109939. 10.1016/j.enbuild.2020.109939.

Dobbs, J. R., and B. M. Hencey. 2014a. Predictive HVAC control using a Markov occupancy model. In *2014 American Control Conference*, 1057–62. 10.1109/ACC.2014.6859389

Dobbs, J. R., and B. M. Hencey. 2014b. Model predictive HVAC control with online occupancy model. *Energy and Buildings* 82: 675–84. 10.1016/j.enbuild.2014.07.051.

DOE. 2023. Prototype building models, May. https://www.energycodes.gov/prototype-building-models

Dong, B., B. Andrews, K. P. Lam, M. Höynck, R. Zhang, Y.-S. Chiou, and D. Benitez. 2010. An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network. *Energy and Buildings* 42 (7): 1038–46. 10.1016/j.enbuild.2010.01.016.

Dong, B., and K. P. Lam. 2011. Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network. *Journal of Building Performance Simulation* 4 (4):359–69. 10.1080/19401493.2011.577810.

Dong, B., and K. P. Lam. 2014. A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. *Building Simulation* 7 (1):89–106. 10.1007/s12273-013-0142-7.

Dong, B., Y. Liu, W. Mu, Z. Jiang, P. Pandey, T. Hong, B. Olesen, T. Lawrence, Z. O'Neil, C. Andrews, et al. 2022. A global building occupant behavior database. *Scientific Data* 9 (1), 369. 10.1038/s41597-022-01475-3.

Dong, B., Y. Liu, M. We, Z. Jiang, P. Pratik, and T. Hong. 2021. ASHRAE global occupant behavior database [Data set]. In *Figshare*. figshare. 10.6084/m9.figshare.16920118.v6

Dong, B., W. Wu, Q. Wang, S. Abdelmutaal, and V. Prakash. 2019. Derive urban scale occupant behavior profiles from mobile position data: A pilot study. In *16th International Conference of the International Building Performance Simulation Association, Building Simulation 2019*, 113–20. 10.26868/25222708.2019.210374

EnergyPlusTM (EnergyPlus^{TM} (e+); 005462MLTPL00). 2017. National Renewable Energy Lab. (NREL), Golden, CO (United States); Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). https://www.osti.gov/biblio/1395882

Erickson, V. L., M. Á. Carreira-Perpiñán, and A. E. Cerpa. 2011. OBSERVE: Occupancy-based system for efficient reduction of HVAC energy. In *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 258–69.

Erickson, V. L., M. Á. Carreira-Perpiñán, and A. E. Cerpa. 2014. Occupancy modeling and prediction for building energy management. *ACM Transactions on Sensor Networks* 10 (3):1–28. 10.1145/2594771.

Erickson, V. L., and A. E. Cerpa. 2010. Occupancy based demand response HVAC control strategy. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, 7–12. 10.1145/1878431.1878434

Erickson, V. L., Y. Lin, A. Kamthe, R. Brahme, A. Surana, A. E. Cerpa, M. D. Sohn, and S. Narayanan. 2009. Energy efficient building environment control strategies using real-time occupancy measurements. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 19–24. 10.1145/1810279.1810284

Fabi, V., R. V. Andersen, and S. P. Corgnati. 2013. Influence of occupant's heating set-point preferences on indoor environmental quality and heating demand in residential buildings. *HVAC&R Research* 19 (5):635–45. 10.1080/10789669.2013.789372.

Feng, X., D. Yan, and C. Wang. 2017. On the simulation repetition and temporal discretization of stochastic occupant behaviour models in building performance simulation. *Journal of Building Performance Simulation* 10 (5–6):612–24. 10.1080/19401493.2016.1236838.

Fontanini, A. D., U. Vaidya, A. Passalacqua, and B. Ganapathysubramanian. 2016. Quantifying mechanical ventilation performance: The connection between transport equations and Markov matrices. *Building and Environment* 104:253–62. 10.1016/j.buildenv.2016.05.019.

Fu, J., S. Hu, X. He, S. Managi, and D. Yan. 2022. Identifying residential building occupancy profiles with demographic characteristics: Using a national time use survey data. *Energy and Buildings* 277:112560. 10.1016/j.enbuild.2022.112560.

Gram-Hanssen, K. 2010. Residential heat comfort practices: Understanding users. *Building Research & Information* 38 (2): 175–86. 10.1080/09613210903541527.

Haldi, F., and D. Robinson. 2009. Interactions with window openings by office occupants. *Building and Environment* 44 (12):2378–95. 10.1016/j.buildenv.2009.03.025.

He, M., H. Pen, M. Li, Y. Huang, D. Yan, S. Lou, and L. Wen. 2022. Investigation on typical occupant behavior in air-conditioned office buildings for South China's Pearl River Delta. *Architectural Intelligence* 1 (1):8. 10.1007/s44223-022-00005-w.

Helbing, D., L. Buzna, A. Johansson, and T. Werner. 2005. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science* 39 (1):1–24. 10.1287/trsc.1040.0108.

Hong, T., D. Yan, S. D'Oca, and C-f Chen. 2017. Ten questions concerning occupant behavior in buildings: The big picture. *Building and Environment* 114:518–30. 10.1016/j.buildenv.2016.12.006.

Hong, T., and H.-W. Lin. 2013. Occupant behavior: Impact on energy use of private offices. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA, 12.

Huchuk, B., W. O'brien, and S. Sanner. 2021. Exploring smart thermostat users' schedule override behaviors and the energy consequences. *Science and Technology for the Built Environment* 27 (2):195–210. 10.1080/23744731.2020.1814668.

Inkarojrit, V.. 2005. *Balancing comfort: Occupants' control of window blinds in private offices*. UC Berkeley: Center for the Built Environment. https://escholarship.org/uc/item/3rd2f2bg

Jacobs, P., and H. Henderson. 2002. State-of-the-art review whole building, building envelope, and HVAC component and system simulation and design tools (DOEOR22674/ARTI-21CR/605-30010-30020-01). Architectural Energy Corporation, Boulder, CO (US); CDH Energy Corp., Cazenovia, NY (US). https://www.osti.gov/biblio/795581-state-art-review-whole-building-building-envelope-hvac-component-system-simulation-design-tools.

Jin, Y., D. Yan, X. Kang, A. Chong, H. Sun, and S. Zhan. 2021. Forecasting building occupancy: A temporal-sequential analysis and machine learning integrated approach. *Energy and Buildings* 252:111362. 10.1016/j.enbuild.2021.111362.

Kang, X., D. Yan, J. An, Y. Jin, and H. Sun. 2021. Typical weekly occupancy profiles in non-residential buildings based on mobile positioning data. *Energy and Buildings* 250:111264. 10.1016/j.enbuild.2021.111264.

Kim, J., T. Hong, J. Jeong, M. Lee, M. Lee, K. Jeong, C. Koo, and J. Jeong. 2017. Establishment of an optimal occupant behavior considering the energy consumption and indoor environmental quality by region. *Applied Energy* 204:1431–43. 10.1016/j.apenergy.2017.05.017.

Klepeis, N. E., W. C. Nelson, W. R. Ott, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, and W. H. Engelmann. 2001. The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants.

Journal of Exposure Analysis and Environmental Epidemiology 11 (3):231–52. 10.1038/sj.jea.7500165.

Li, Z., and B. Dong. 2017. A new modeling approach for short-term prediction of occupancy in residential buildings. *Building and Environment* 121:277–90. 10.1016/j.buildenv.2017.05.005.

Luo, N., G. Fierro, Y. Liu, B. Dong, and T. Hong. 2022. Extending the Brick schema to represent metadata of occupants. *Automation in Construction* 139:104307. 10.1016/j.autcon.2022.104307

Mahdavi, A., A. Mohammadi, E. Kabir, and L. Lambeva. 2008. Occupants' operation of lighting and shading systems in office buildings. *Journal of Building Performance Simulation* 1 (1):57–65. 10.1080/19401490801906502.

Mahdavi, A., and F. Tahmasebi. 2015. Predicting people's presence in buildings: An empirically based model performance analysis. *Energy and Buildings* 86:349–55. 10.1016/j.enbuild.2014.10.027.

Manna, C., D. Fay, K. N. Brown, and N. Wilson. 2013. Learning Occupancy in Single Person Offices with Mixtures of Multi-lag Markov Chains. In *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, 151–8. 10.1109/ICTAI.2013.32

Mirakhorli, A., and B. Dong. 2016. Occupancy behavior based model predictive control for building indoor climate—A critical review. *Energy and Buildings* 129:499–513. 10.1016/j.enbuild.2016.07.036.

Newsham, G., and C. Arsenault. 2009. A camera as a sensor for lighting and shading control. *Lighting Research & Technology* 41 (2):143–63. 10.1177/1477153508099889.

Niu, B., D. Li, H. Yu, N. Mahyuddin, Y. Liu, J. Wu, X. Wang, and S. Pan. 2022. Investigation of occupant window opening behaviour during the summer period in a Beijing maternity hospital. *Journal of Building Engineering* 45:103441. 10.1016/j.jobe.2021.103441.

O'Brien, W., A. Wagner, M. Schweiker, A. Mahdavi, J. Day, M. B. Kjærgaard, S. Carlucci, B. Dong, F. Tahmasebi, D. Yan, et al. 2020. Introducing IEA EBC annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation. *Building and Environment* 178:106738. 10.1016/j.buildenv.2020.106738.

Peffer, T., M. Pritoni, A. Meier, C. Aragon, and D. Perry. 2011. How people use thermostats in homes: A review. *Building and Environment* 46 (12):2529–41. 10.1016/j.buildenv.2011.06.002.

Rahman, H., and H. Han. 2021. Real-time ventilation control based on a Bayesian estimation of occupancy. *Building Simulation* 14 (5): 1487–97. 10.1007/s12273-020-0746-7.

Reinhart, C. 2004. Lightswitch-2002: A model for manual and automated control of electric lighting and blinds. *Solar Energy* 77 (1):15–28. 10.1016/j.solener.2004.04.003.

Reinhart, C., and K. Voss. 2003. Monitoring manual control of electric lighting and blinds. *Lighting Research & Technology* 35 (3):243–58. 10.1191/1365782803li064oa.

Ren, X., D. Yan, and C. Wang. 2014. Air-conditioning usage conditional probability model for residential buildings. *Building and Environment* 81:172–82. 10.1016/j.buildenv.2014.06.022.

Saldanha, N., and I. Beausoleil-Morrison. 2012. Measured end-use electric load profiles for 12 Canadian houses at high temporal resolution. *Energy and Buildings* 49:519–30. 10.1016/j.enbuild.2012.02.050.

Salim, F. D., B. Dong, M. Ouf, Q. Wang, I. Pigliautile, X. Kang, T. Hong, W. Wu, Y. Liu, S. K. Rumi, et al. 2020. Modelling urban-scale occupant behaviour, mobility, and energy in buildings: A survey. *Building and Environment* 183:106964. 10.1016/j.buildenv.2020.106964.

Schweiker, M., M. Kleber, and A. Wagner. 2019. Long-term monitoring data from a naturally ventilated office building. *Scientific Data* 6 (1):293. 10.1038/s41597-019-0283-3.

Shi, S., and B. Zhao. 2016. Occupants' interactions with windows in 8 residential apartments in Beijing and Nanjing, China. *Building Simulation* 9 (2):221–31. 10.1007/s12273-015-0266-z.

Sun, Y., L. Gu, C. F. Wu, and G. Augenbroe. 2014. Exploring HVAC system sizing under uncertainty. *Energy and Buildings* 81:243–52. 10.1016/J.ENBUILD.2014.06.026

Sun, K., and T. Hong. 2017. A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures. *Energy and Buildings* 146:383–96. 10.1016/j.enbuild.2017.04.065.

Tabadkani, A., A. Roetzel, H. Xian Li, A. Tsangrassoulis, and S. Attia. 2021. Analysis of the impact of automatic shading control scenarios on occupant's comfort and energy load. *Applied Energy* 294:116904. 10.1016/j.apenergy.2021.116904.

Tamas, R., W. O'Brien, and M. S. Quintero. 2021. Residential thermostat usability: Comparing manual, programmable, and smart devices. *Building and Environment* 203:108104. 10.1016/j.buildenv.2021.108104.

Tien, P. W., S. Wei, J. K. Calautit, J. Darkwa, and C. Wood. 2022. Real-time monitoring of occupancy activities and window opening within buildings using an integrated deep learning-based approach for reducing energy demand. *Applied Energy* 308:118336. 10.1016/j.apenergy.2021.118336.

Tomastik, R., Y. Lin, and A. Banaszuk. 2008. Video-based estimation of building occupancy during emergency egress. In *2008 American Control Conference*, 894–901. 10.1109/ACC.2008.4586606

US EPA. 2014. Indoor air quality (IAQ) [collections and lists], July 3. https://www.epa.gov/indoor-air-quality-iaq

Vellei, M., S. Martinez, and J. Le Dréau. 2021. Agent-based stochastic model of thermostat adjustments: A demand response application. *Energy and Buildings* 238:110846. 10.1016/j.enbuild.2021.110846.

Wang, J., J. Huang, Z. Feng, S.-J. Cao, and F. Haghighat. 2021. Occupant-density-detection based energy efficient ventilation system: Prevention of infection transmission. *Energy and Buildings* 240:110883. 10.1016/j.enbuild.2021.110883.

World Health Organization. Regional Office for Europe. 2014. Combined or multiple exposure to health stressors in indoor built environments: An evidence-based review prepared for the WHO training workshop "Multiple environmental exposures and risks": 16–18 October 2013, Bonn, Germany. World Health Organization Press. Regional Office for Europe. https://apps.who.int/iris/handle/10665/350495.

Wu, W., B. Dong, Q. Wang, M. (Ryan), Kong, D. Yan, J. An, and Y. Liu. 2020. A novel mobility-based approach to derive urban-scale building occupant profiles and analyze impacts on building energy consumption. *Applied Energy* 278:115656. 10.1016/j.apenergy.2020.115656.

Yan, D., T. Hong, B. Dong, A. Mahdavi, S. D'Oca, I. Gaetani, and X. Feng. 2017. IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. *Energy and Buildings* 156:258–70. 10.1016/j.enbuild.2017.09.084.

Yoshino, H., T. Hong, and N. Nord. 2017. IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods. *Energy and Buildings* 152:124–36. 10.1016/j.enbuild.2017.07.038.

Yuhaski, S. J., and J. M. Smith. 1989. Modeling circulation systems in buildings using state dependent queueing models. *Queueing Systems* 4 (4):319–38. 10.1007/BF01159471.

Yun, G. Y., and K. Steemers. 2008. Time-dependent occupant behaviour models of window control in summer. *Building and Environment* 43 (9):1471–82. 10.1016/j.buildenv.2007.08.001.

Zhou, X., Y. Lu, S. Hu, Z. Yang, and D. Yan. 2023. New perspectives on temporal changes in occupancy characteristics of residential buildings. *Journal of Building Engineering* 64:105590. 10.1016/j.jobe.2022.105590.

Appendix A: Results of data analysis by building type and climate zone

This appendix comprises figures generated through data analysis, presenting the outcomes related to nine occupant behavior patterns across various building types and climate zones.

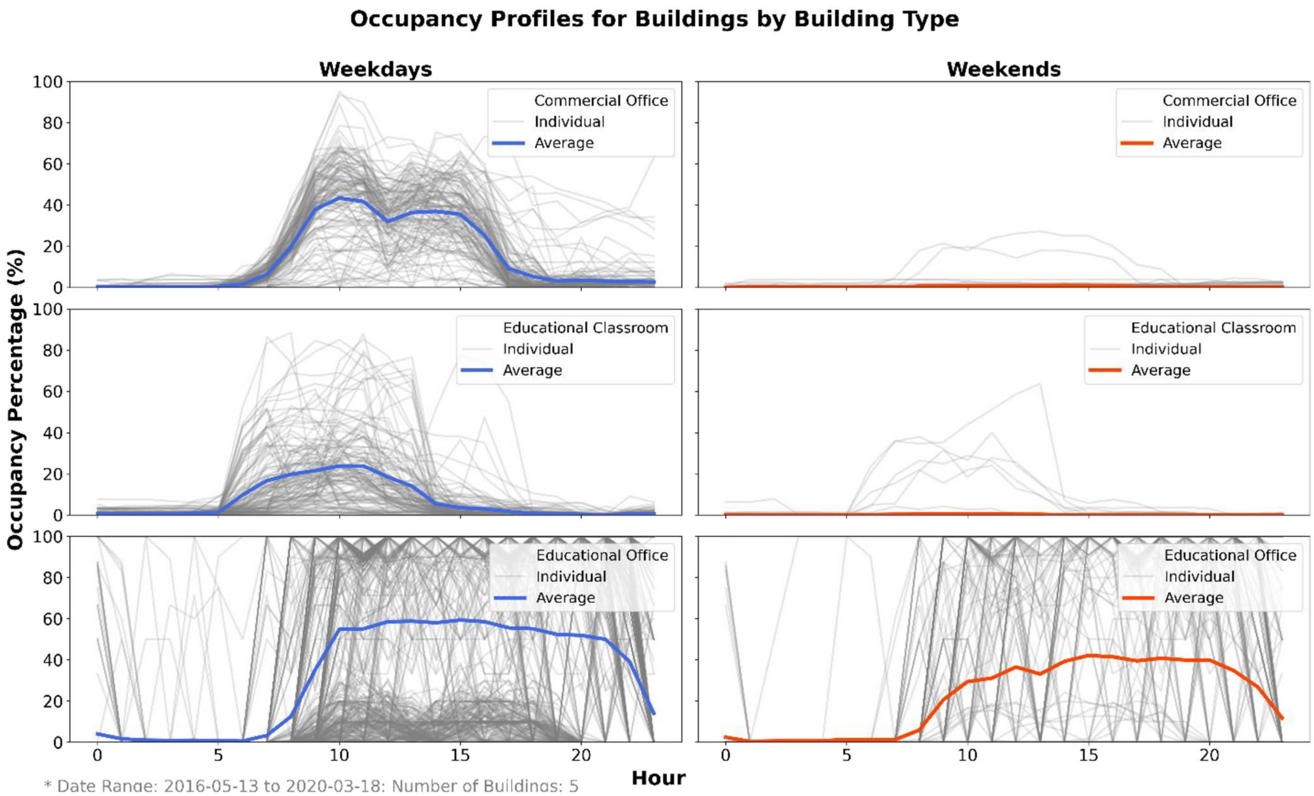

Fig. A1. Daily occupant number profiles by building type.

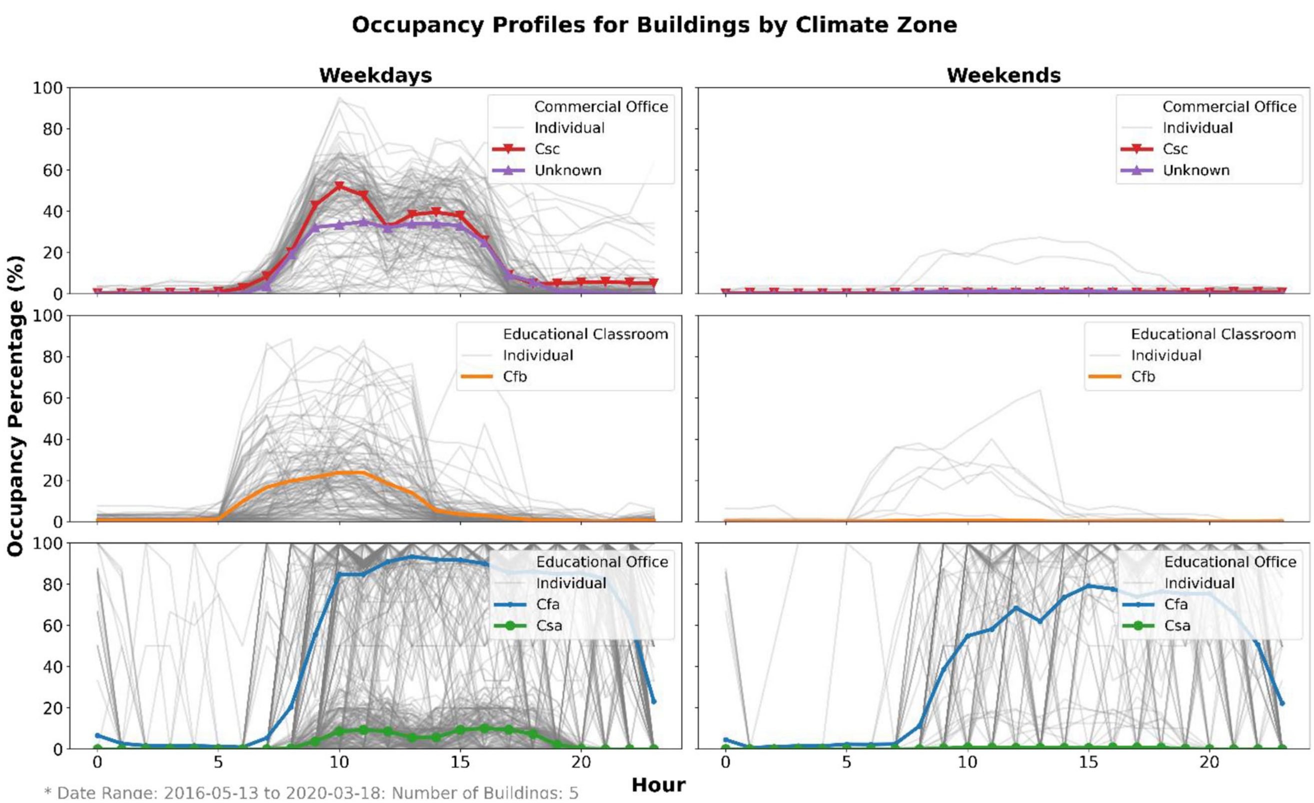

Fig. A2. Daily occupant number profiles by building type and climate zone.

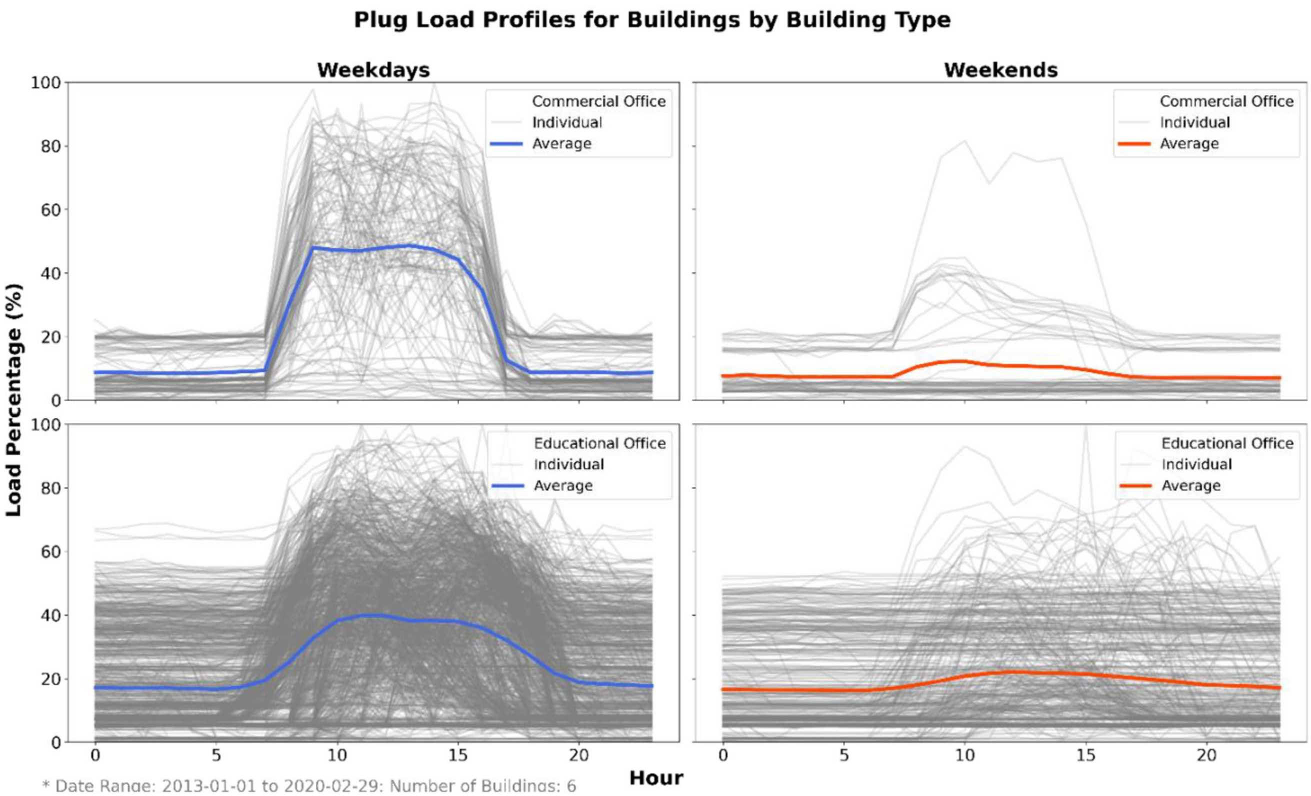

Fig. A3. Daily plug load profiles by building type.

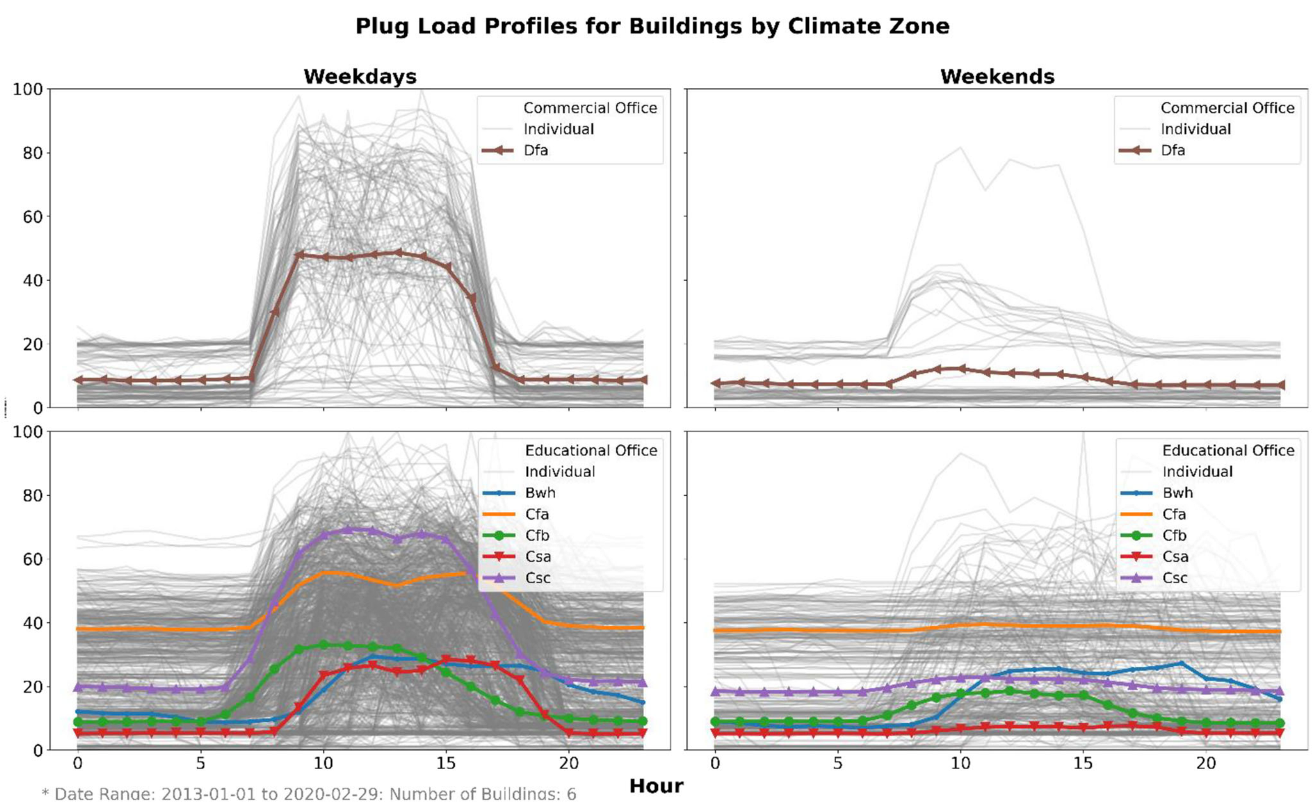

Fig. A4. Daily plug load profiles by building type and climate zone.

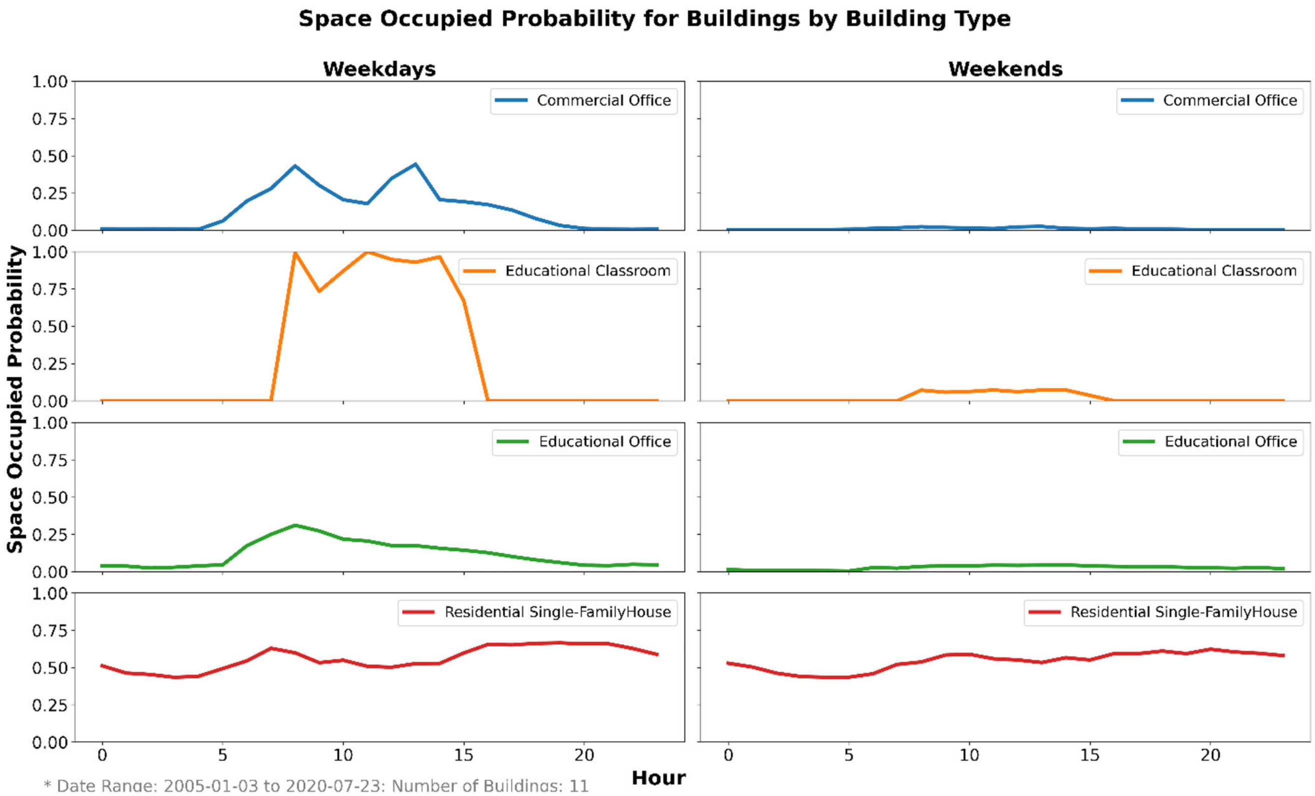

Fig. A5. Daily occupant presence profiles by building type.

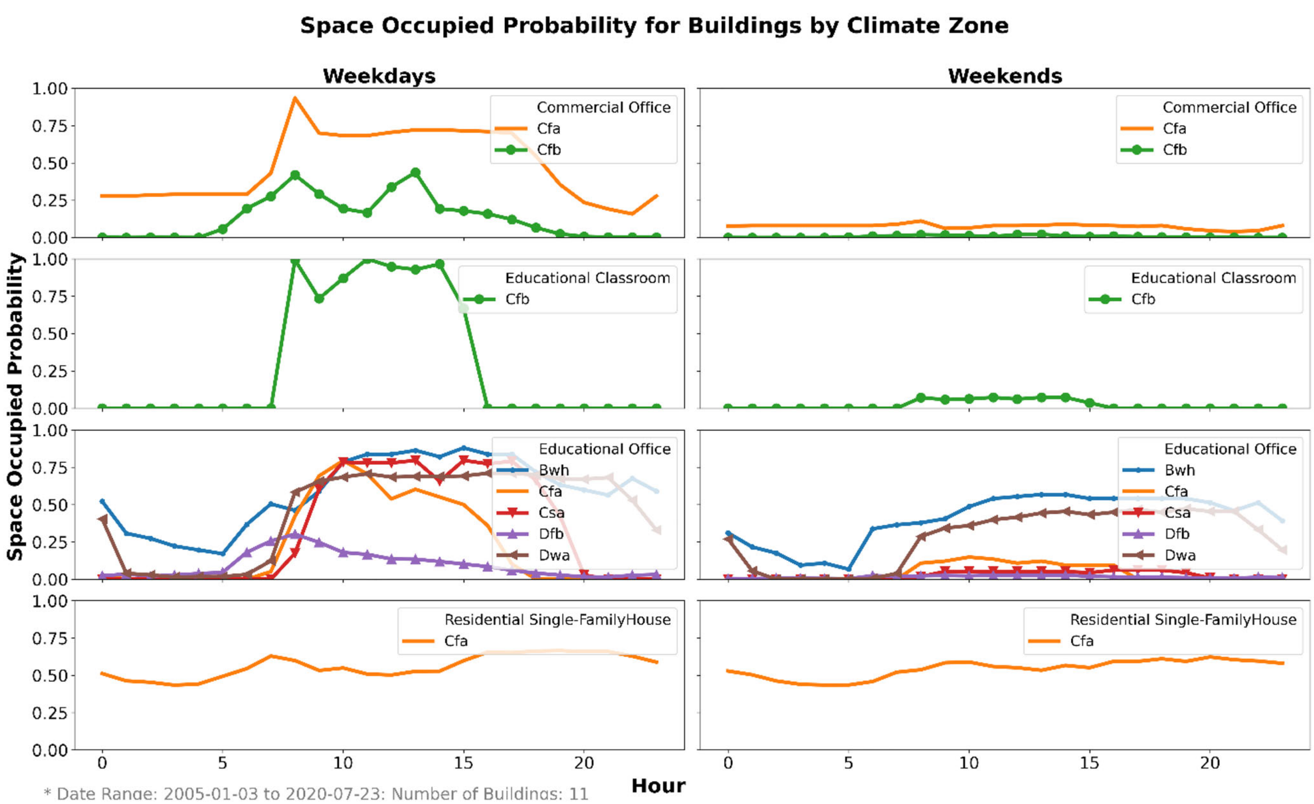

Fig. A6. Daily occupant presence profiles by building type and climate zone.

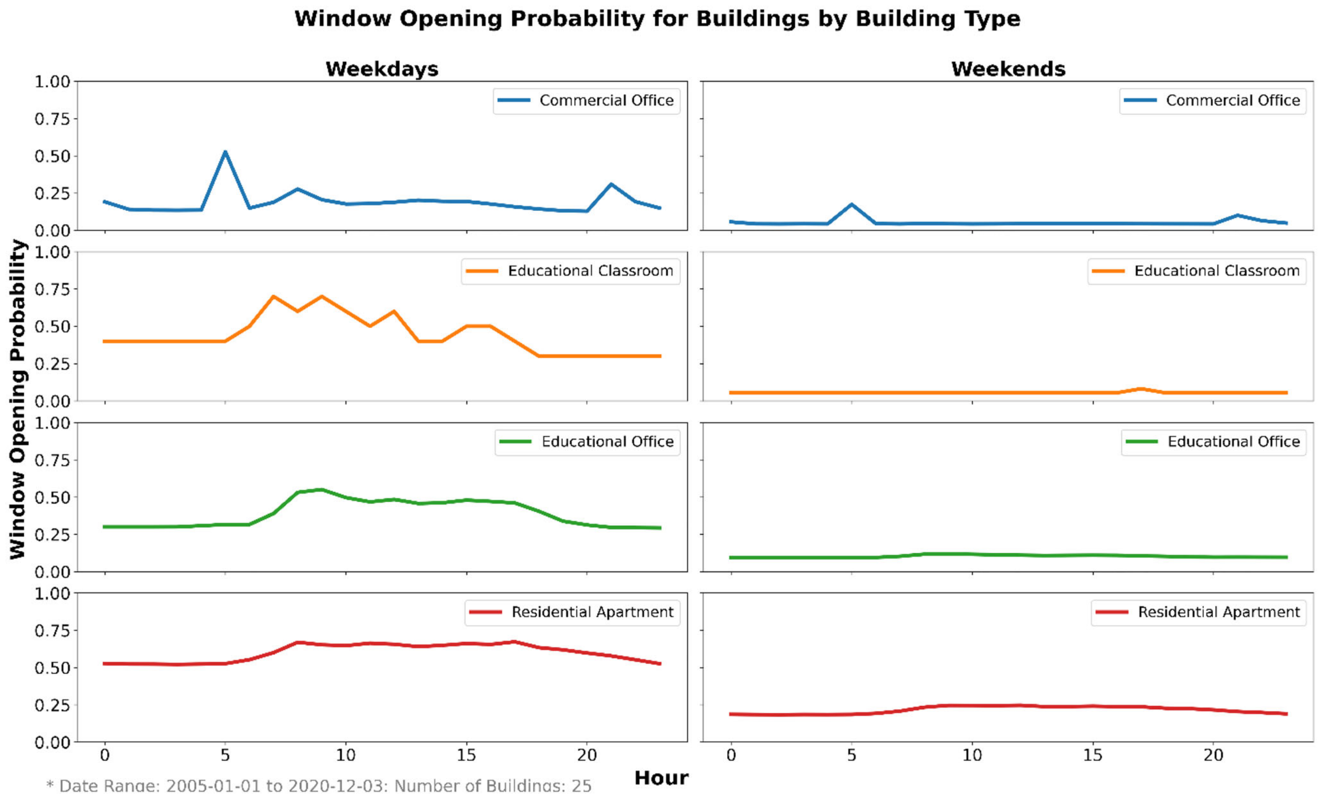

Fig. A7. Daily window opening profiles by building type.

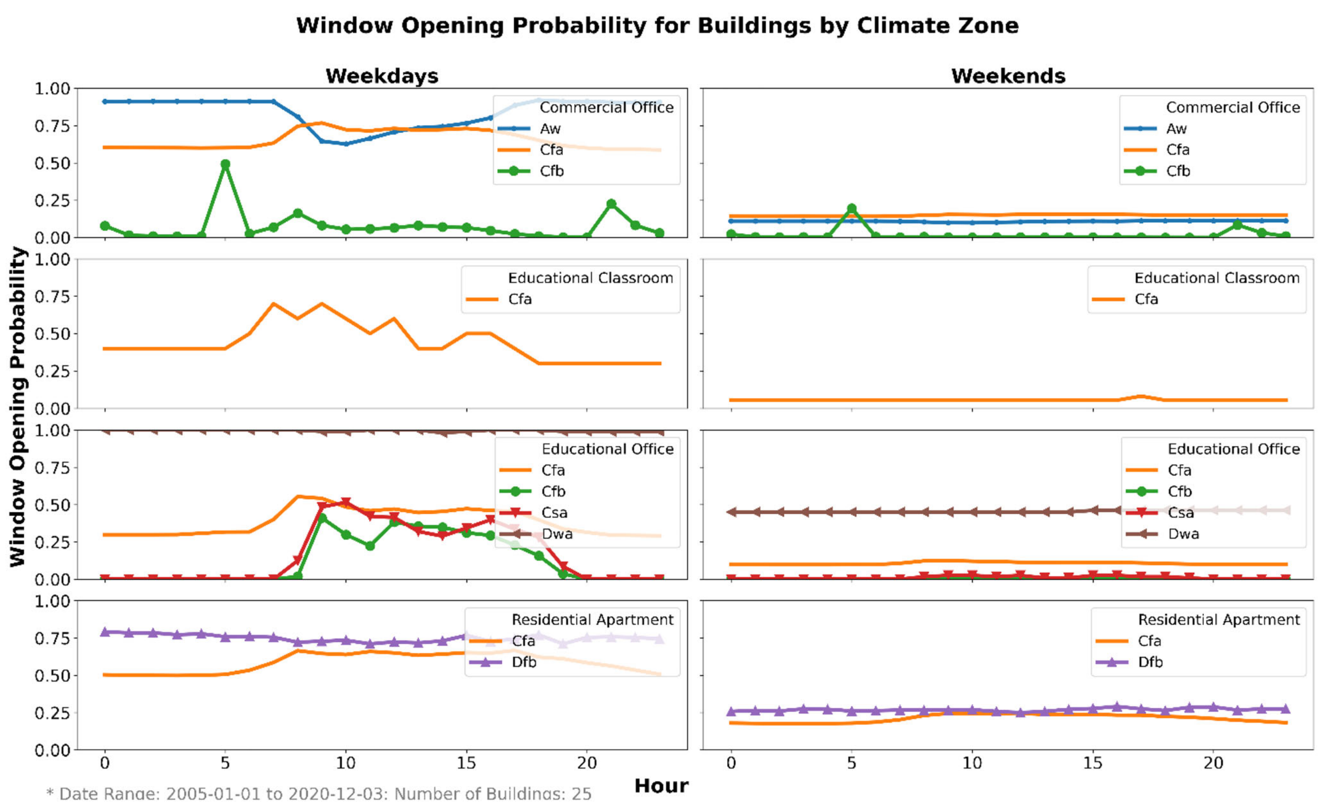

Fig. A8. Daily window opening profiles by building type and climate zone.

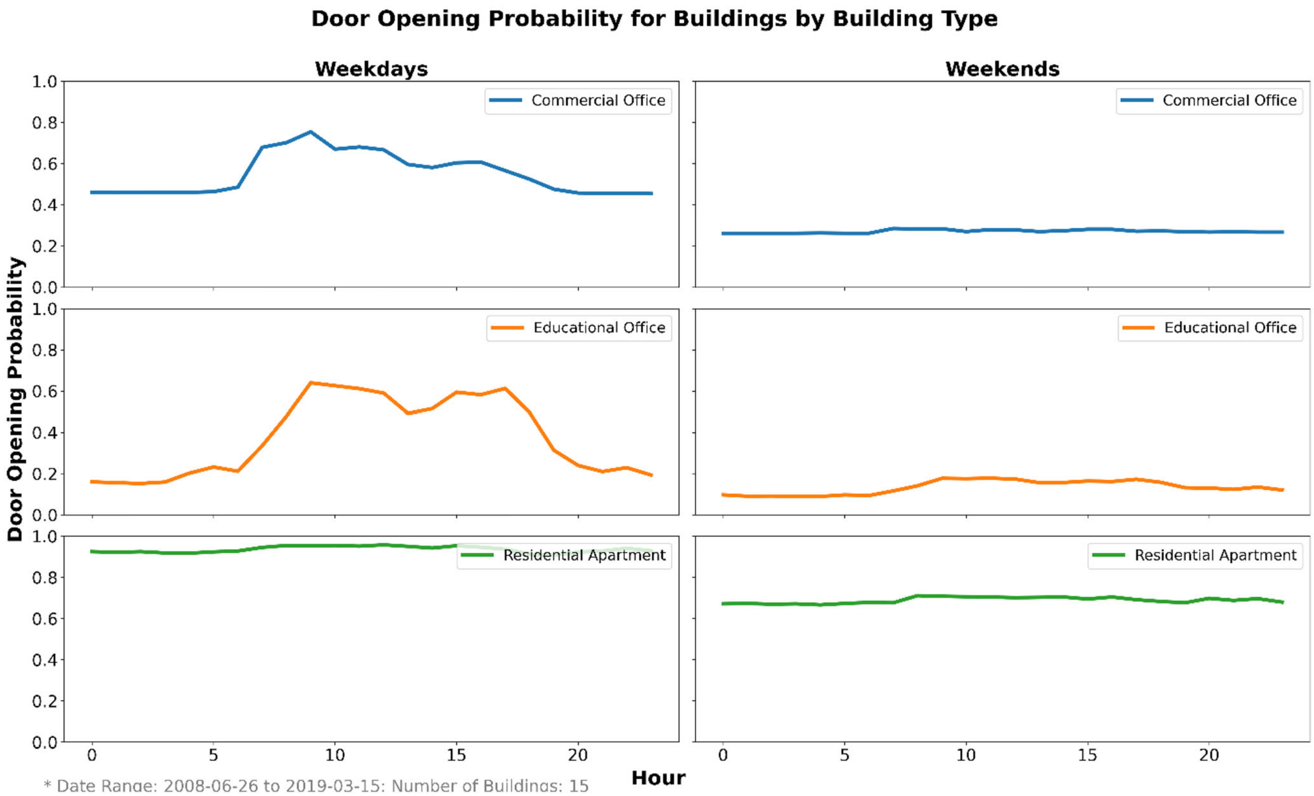

Fig. A9. Daily door opening profiles by building type.

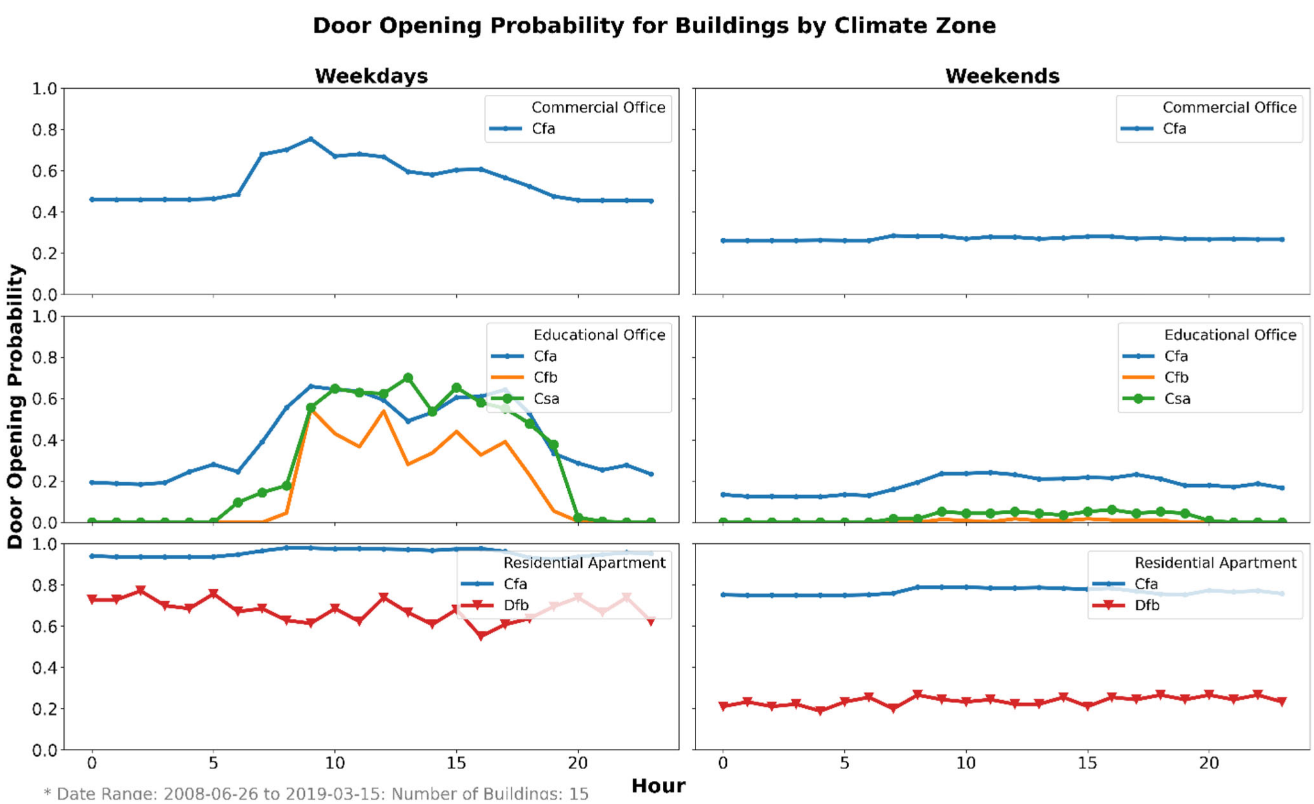

Fig. A10. Daily door opening profiles by building type and climate zone.

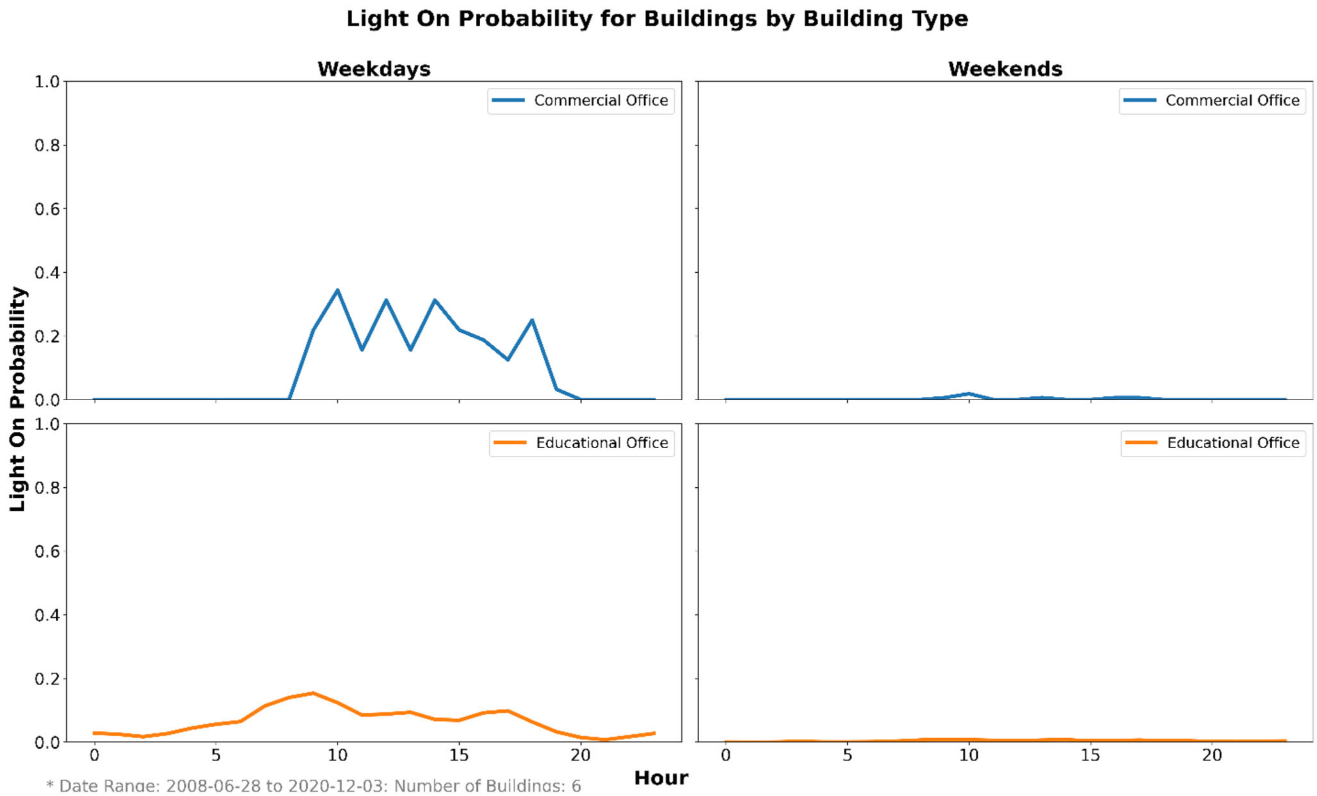

Fig. A11. Daily lighting on profiles by building type.

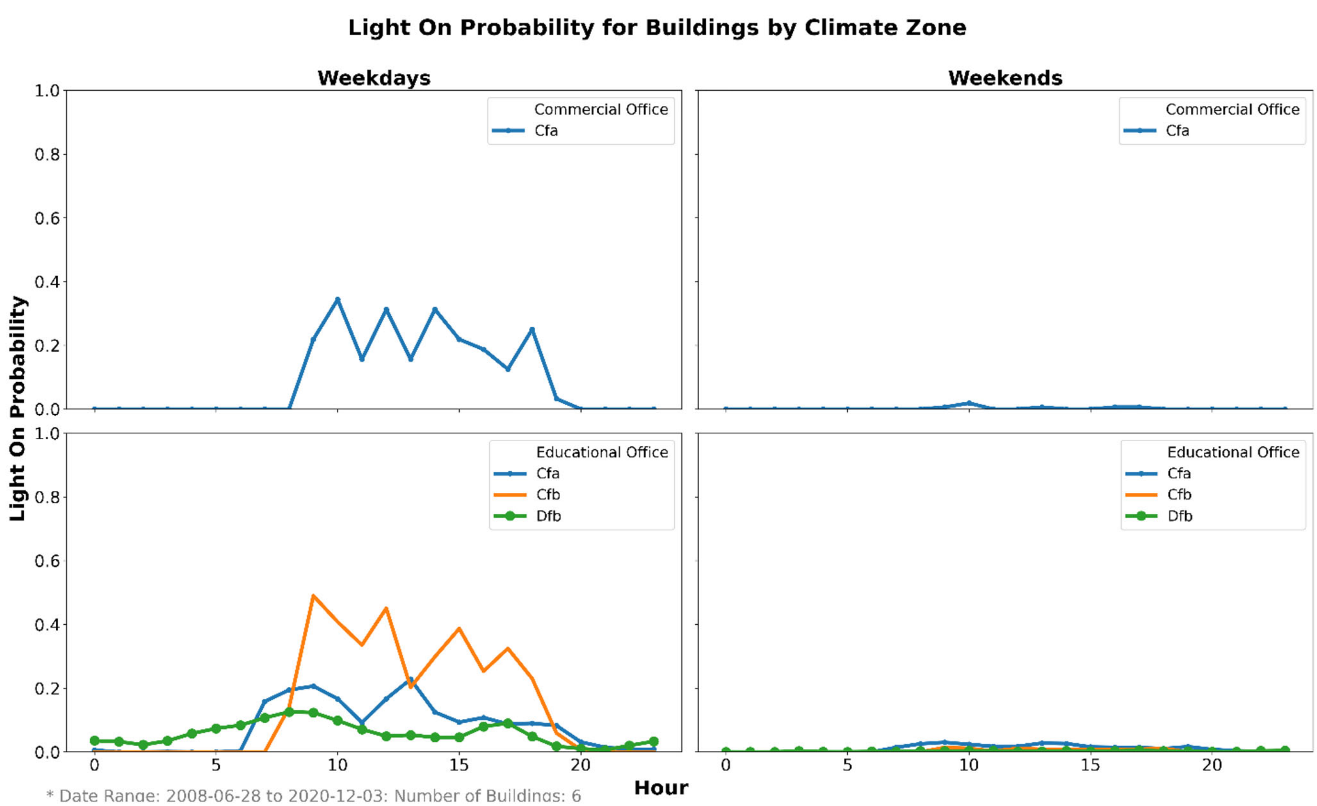

Fig. A12. Daily lighting on profiles by building type and climate zone.

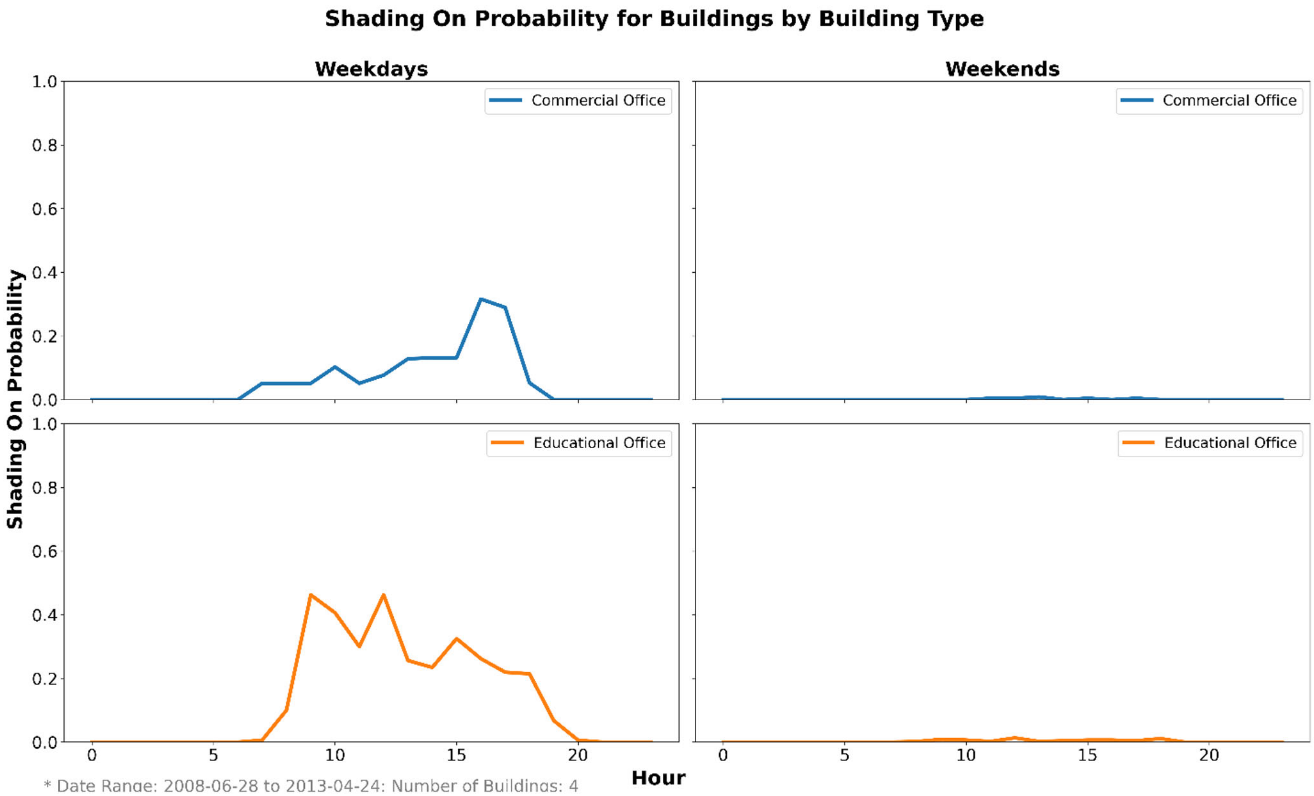

Fig. A13. Daily shading on profiles by building type.

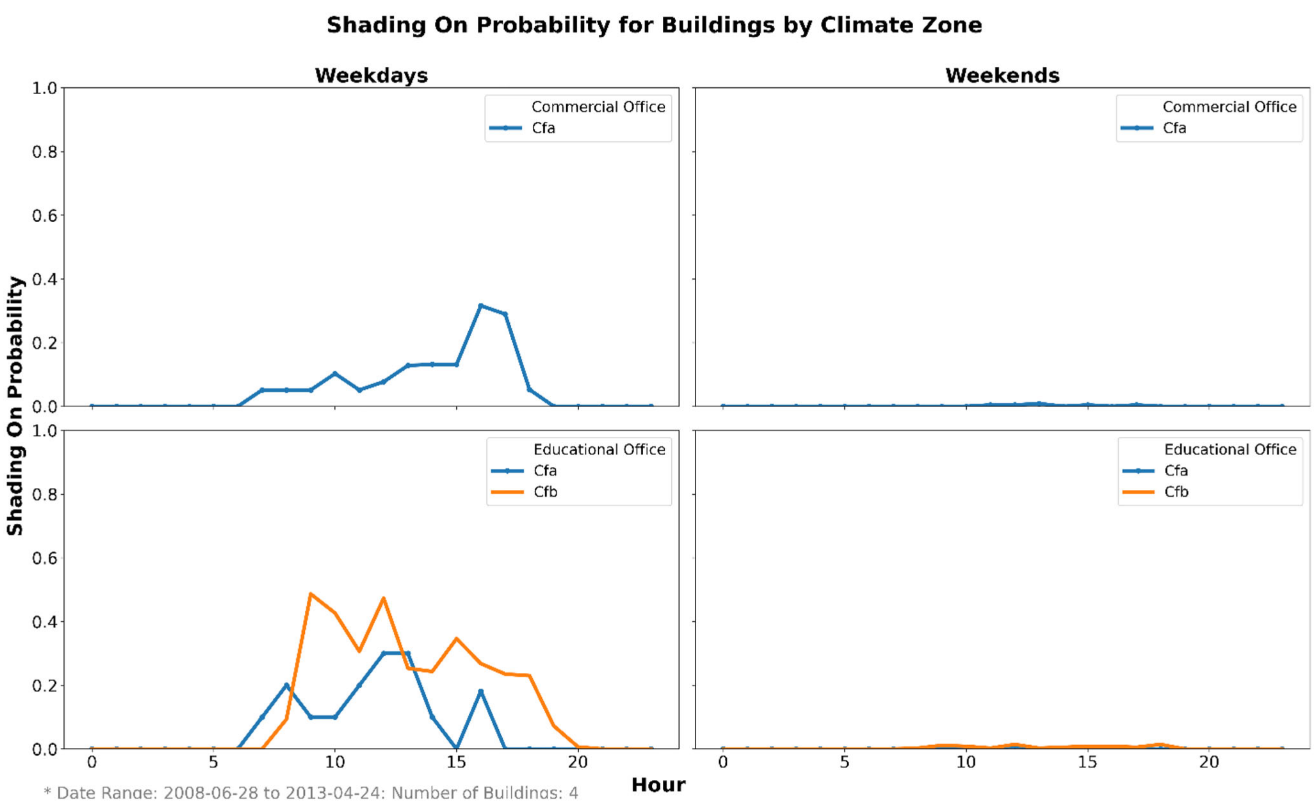

Fig. A14. Daily shading on profiles by building type and climate zone.

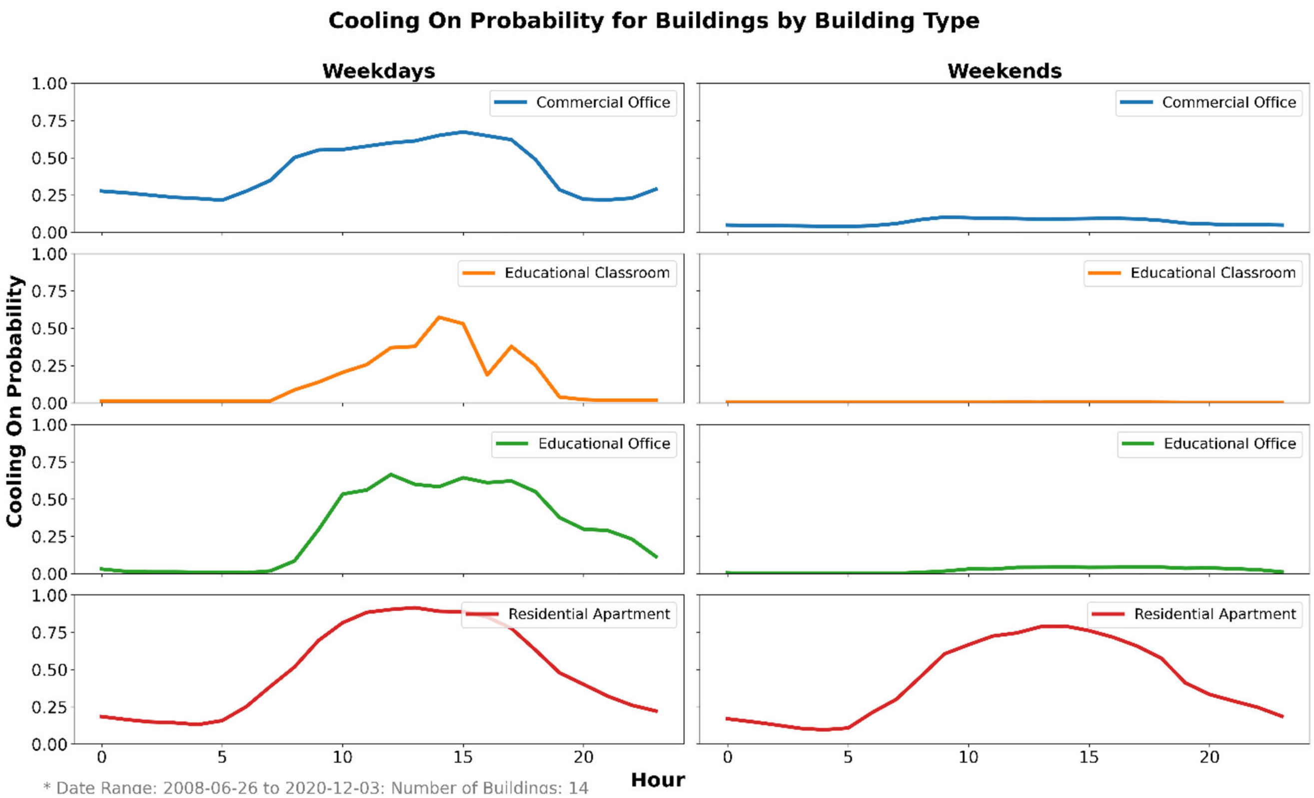

Fig. A15. Daily HVAC status (cooling) profiles by building type.

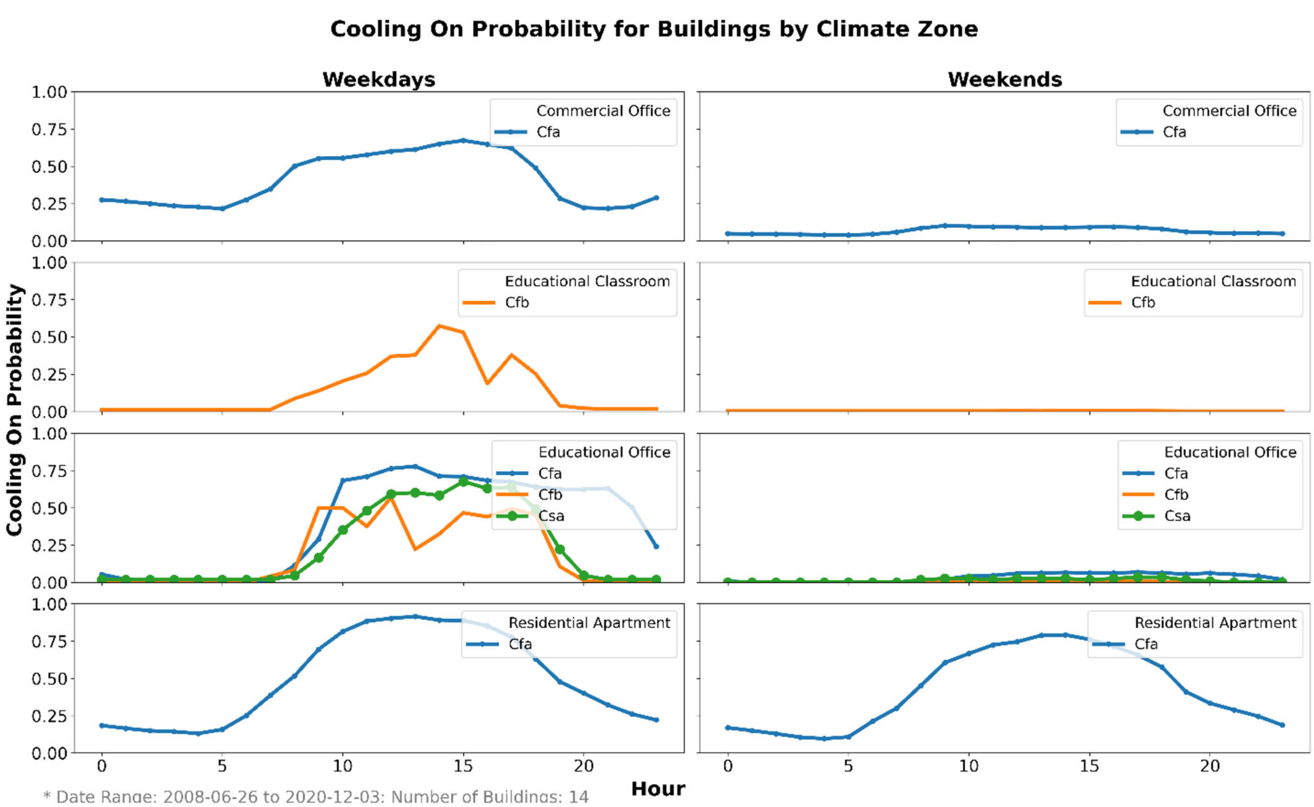

Fig. A16. Daily HVAC status (cooling) profiles by building type and climate zone.

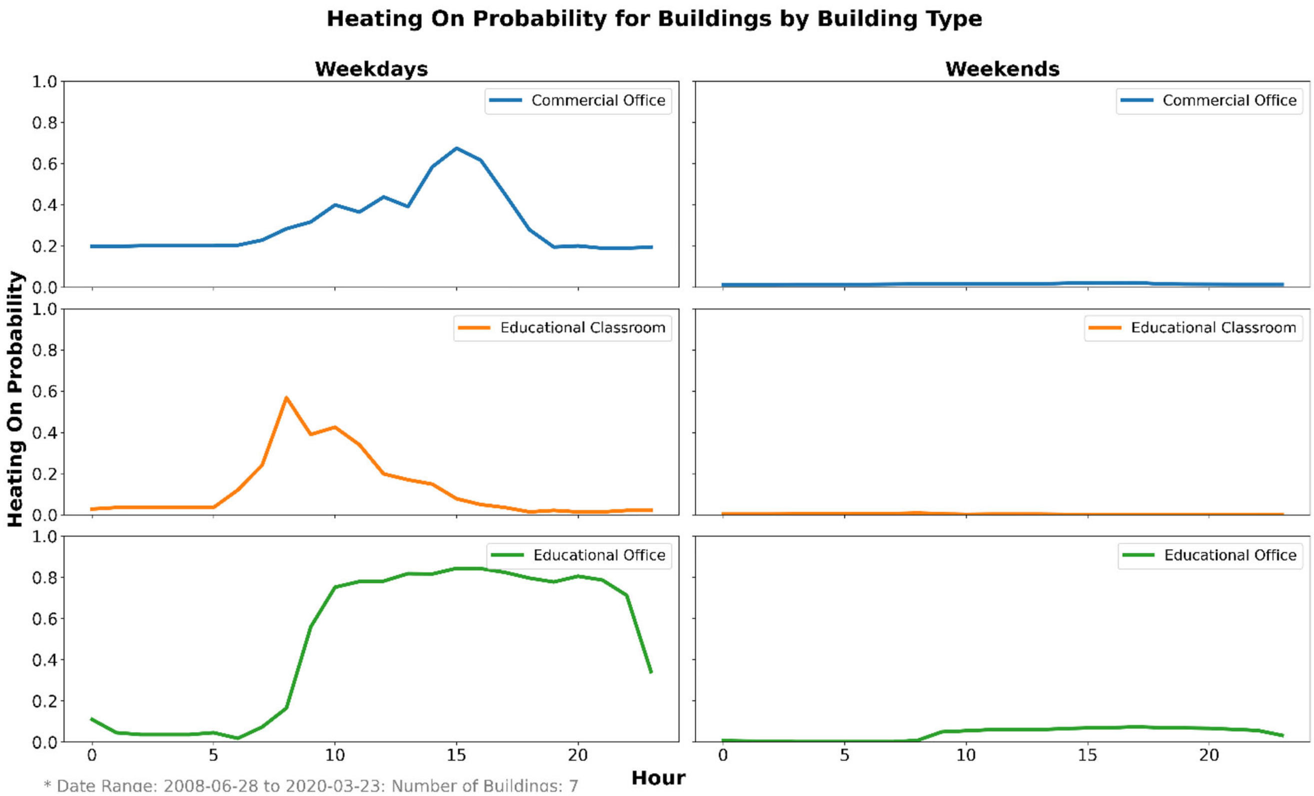

Fig. A17. Daily HVAC status (heating) profiles by building type.

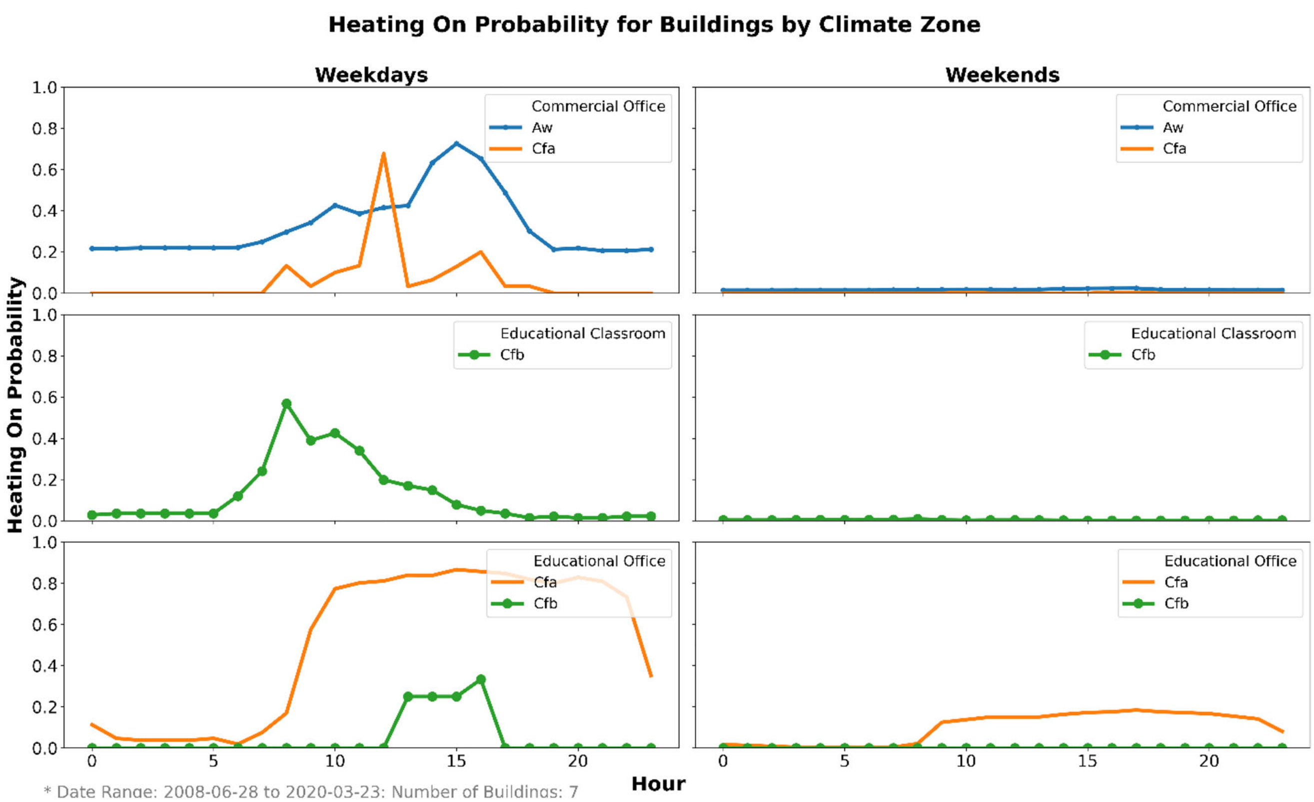

Fig. A18. Daily HVAC status (heating) profiles by building type and climate zone.

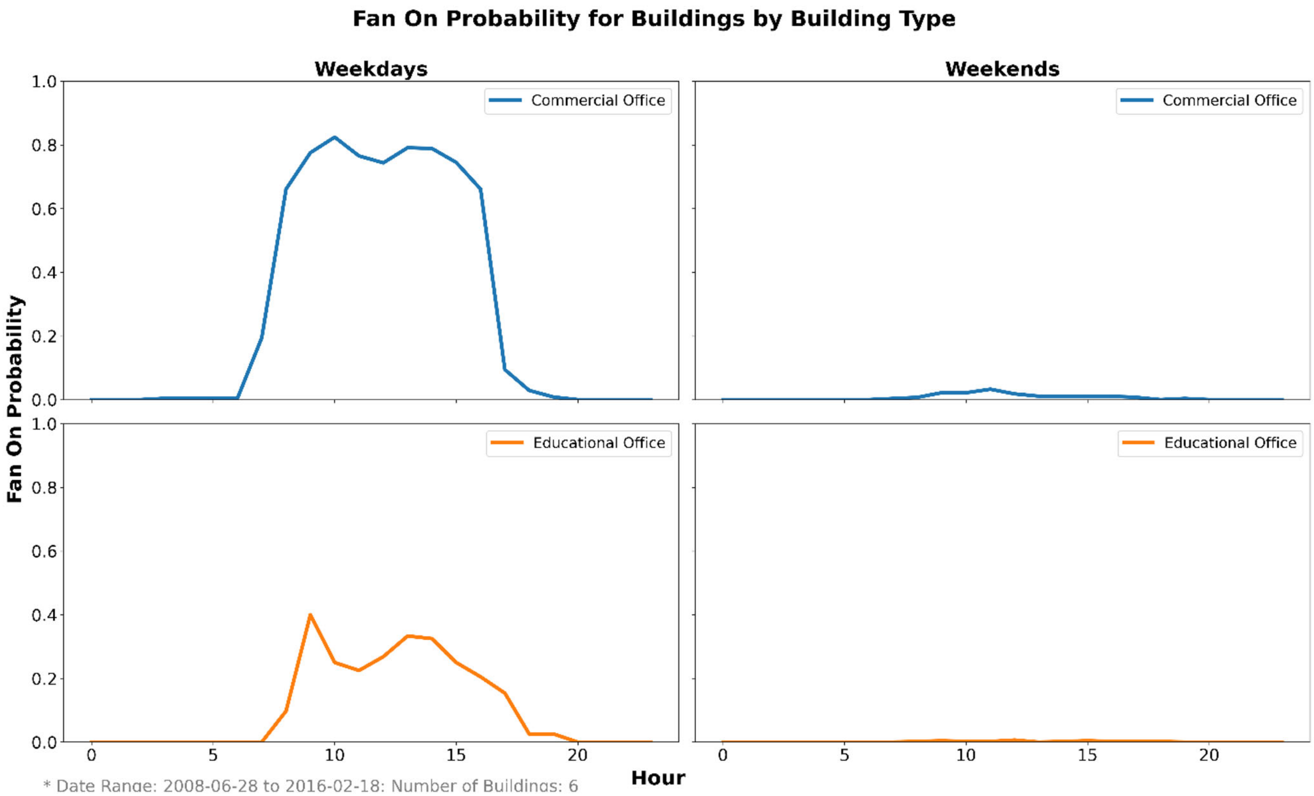

Fig. A19. Daily fan status profiles by building type.

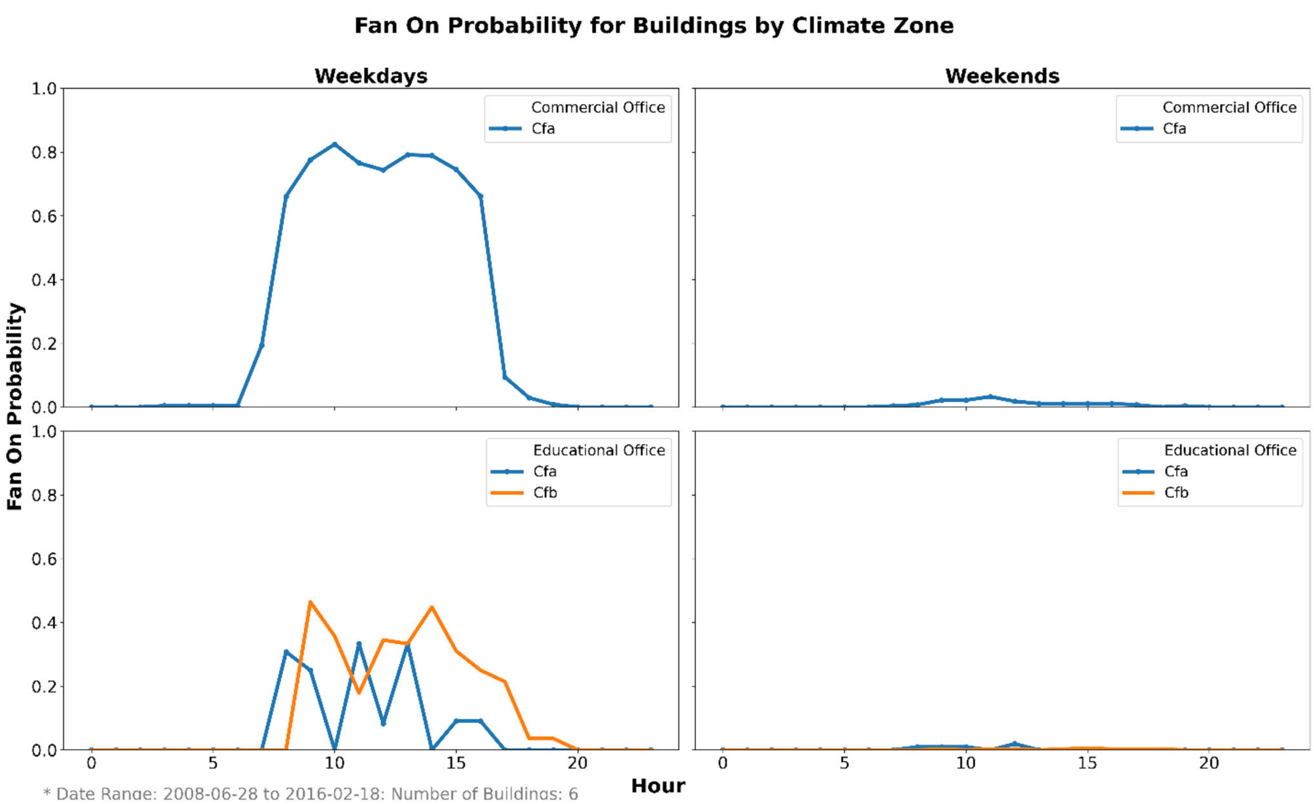

Fig. A20. Daily fan status profiles by building type and climate zone.