Building and Environment 275 (2025) 112792

Contents lists available at ScienceDirect

Building and Environment

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

Analysis of occupant thermal comfort and energy-saving potential based on cooling behaviors in residential buildings: A case study of Shanghai

Junmeng Lyu [a], Yuxin Yang [a], Dayi Lai [a], Li Lan [a], Zisheng Zhao [b,c], Heng Du [d], Zhiwei Lian [a,e,*]

[a] Department of Architecture, School of Design, Shanghai Jiao Tong University, Shanghai, China
[b] School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China
[c] Guangdong Midea Air-Conditioning Equipment Co., Ltd. Guangdong, China
[d] Department of Built Environment, National University of Singapore, Singapore
[e] China Institute of Urban Governance, Shanghai Jiao Tong University, Shanghai, 200030, China

ARTICLE INFO

Keywords:
Usage behavior
Sustainability
Occupant-centric control
Thermal comfort
Urban governance
Building energy

ABSTRACT

Buildings account for a significant share of global energy consumption and carbon emissions, with the cooling sector representing a critical challenge in achieving net-zero emissions. The urgent need lies in developing strategies that effectively balance cooling energy efficiency with thermal comfort of building occupants. The development of the Internet of Things has made it possible to collect users' cooling usage data. Mining historical data for thermal comfort analysis, thermal environment control and fine-scale energy consumption prediction is a promising approach, but research gaps remain. To implement this approach, this study develops a Residential Cooling Behavior Analytics Framework (RE-CBA), which leverages IoT-enabled air conditioner data to optimize cooling performance. By analyzing usage patterns, identifying cooling periods, and modeling user-specific comfort temperatures, the RE-CBA framework generates personalized setpoint schedules for automatic air conditioner control. Integrated with building energy simulations, it quantifies energy performance under dynamic controls. Applying the RE-CBA framework to 920 air conditioners in Shanghai, the results reveal diverse individual comfort preferences. Personalized control strategies derived from the framework resulted in an 11.04% energy saving ($\Delta E = 116{,}657$ kWh) compared to the static 26 °C baseline. By uniting real-time data insights with occupant-centric control strategies, this framework advances both thermal comfort and energy efficiency. With its scalability and ease of adoption, the RE-CBA framework offers an effective pathway to sustainable cooling in urban residential buildings.

1. Introduction

Climate change is a global issue of significant concern, threatening the development of human societies and the very existence of life on Earth [1]. The building sector, according to the International Energy Agency [2–4], accounts for over 40% of global energy consumption, and this figure continues to rise. While buildings utilize various forms of energy, electricity is the primary source [5], primarily powering heating, ventilation, and air conditioning (HVAC) systems, which are essential for maintaining thermal comfort for occupants [6–8]. Data from the National Bureau of Statistics of China (NBSC) reveals that there are currently 540 million residential air conditioners in the country [9]. Projections indicate that by 2030, up to 85% of households will own at least one air conditioner, with the total number of residential cooling devices, including fans and dehumidifiers, expected to exceed 1.1 billion units [3]. For China, a key challenge over the next decade will be to meet the increasing demand for residential comfort without significantly raising energy consumption [4].

Unlike in commercial and public buildings, where building managers centrally regulate the thermal environment, residents in residential buildings have a high degree of manual control [10,11]. In public buildings, where individuals cannot control the environment and are not responsible for energy costs, people often exhibit more critical attitudes [12]. Conversely, in residential buildings, individuals tend to balance economic considerations and environmental awareness before making thermal adjustments, rather than relying solely on their thermal perception. This results in a higher tolerance for thermal discomfort and a greater likelihood of adapting to the thermal environment. Accurately

* Corresponding author.
E-mail address: zwlian@sjtu.edu.cn (Z. Lian).

https://doi.org/10.1016/j.buildenv.2025.112792
Received 26 December 2024; Received in revised form 18 February 2025; Accepted 25 February 2025
Available online 25 February 2025

analyzing thermal demands in residential buildings, therefore, poses a significant challenge. Currently, the exploration of occupants' thermal demands within buildings predominantly relies on climate chamber experiments [13,14]. These experiments can reflect a specific group's direct perception of the thermal environment, but do not account for the influence of non-environmental factors on their judgment in real-world usage scenarios, such as the increased costs associated with maintaining lower or higher temperatures [15]. However, long-term field surveys on specific buildings to capture such context effects require large expense and labor.

Analyzing air conditioner usage behavior provides an effective method to accurately assess residential users' thermal demands, as these behaviors directly reflect their actual needs [16]. Early studies on occupant behavior primarily approached the topic from a sociological perspective, lacking in-depth quantitative analysis [17–20]. Quantitative studies on air conditioning behavior typically involve placing sensors in households to monitor indoor environmental parameters, combined with survey questionnaires to analyze usage patterns and influencing factors [21–23]. The EBC Annex 66 project was the first to establish a systematic quantitative simulation method to understand the impact of occupant behavior on building energy use and indoor environments [24]. This project enhanced the accuracy of energy consumption simulation software such as EnergyPlus and DeST. However, current studies on occupant energy behavior within buildings in China are limited by costs and participant cooperation, resulting in small sample sizes, usually fewer than 50 households [25–27]. This limitation poses challenges in accurately identifying regional residential energy use characteristics. Additionally, these studies often use comfort parameters defined by international thermal environment design standards, such as ASHRAE-55 [28] and ISO-7730 [29], as benchmarks for energy consumption analysis. These standards may not effectively account for the high level of occupant autonomy in residential buildings and its impact on thermal adaptability. For energy management and conservation in buildings, it is crucial to explore energy-saving potential based on a thorough understanding of occupant thermal demands; otherwise, energy consumption estimates may deviate from reality. With the development of IoT technology and big data, it is now possible to analyze historical usage data from connected air conditioners to infer residents' thermal demands.

Current research on thermal comfort primarily relies on subjective survey feedback, which serves as the foundational method for developing widely applied models and thermal environment design guidelines [28,30,31]. Given the complexity of multiple factors (e.g., high costs, difficulties in occupant cooperation, and variability in survey quality), subjective survey-based approaches do present certain challenges and limitations, particularly when applied to real-world buildings [25–27]. Mining historical data for thermal comfort analysis and fine-scale energy consumption prediction is a promising approach. This may represent a highly effective approach for future data-driven automatic thermal comfort environment control. While more research focusing on HVAC historical data has been dedicated to large-scale energy consumption prediction (e.g., at the whole-building level) [32], studies specifically aimed at mining thermal comfort information from historical data are underdeveloped and have been somewhat underestimated, lacking systematic and comprehensive frameworks [33–35]. To implement this approach, this study aims to develop a Residential Cooling Behavior Analytics Framework (RE-CBA), which systematically extracts information by identifying annual cooling periods, analyzing usage scenarios, and modeling residents' comfort thermostat setpoints. The individualized indoor thermostat setpoints schedules generated by the framework enable automatic, comfort-driven adjustments to air conditioning systems. To evaluate the energy implications of this model-driven approach, the framework also integrates building energy simulations, providing a method to assess air conditioner energy consumption under this occupant-centric control strategy. By implementing this approach through the RE-CBA framework, this research seeks to demonstrate its effectiveness and highlight its potential applications.

The design of this framework is intended to ensure broad applicability, enabling the analysis of air conditioner usage data across different regions. Shanghai, a densely populated metropolis, is characterized by significant cooling demands driven by rapid urbanization and extensive cooling infrastructure [36]. With its representative climate and electricity consumption patterns, Shanghai serves as a typical example of high-density urban environments with substantial summer cooling demands. Therefore, this study selects Shanghai as a case study to clarify the application process of the established framework. Using historical IoT air conditioner usage data from residential buildings in Shanghai, collected between May and October 2020, an in-depth analysis of residential cooling behavior and thermal demand was conducted. The analysis revealed residents' preferred indoor air temperature ranges and their variations, identifying opportunities for energy savings while maintaining occupant comfort. This study aims to propose a viable approach for occupant-centric control and energy prediction in residential settings.

2. Development of residential cooling behavior analytics framework

The proposed RE-CBA framework integrates data preprocessing, comfort-based indoor air temperature analysis, and energy consumption simulation to evaluate the energy performance of control strategies driven by the framework (Fig. 1). The first module, **data preprocessing**, improves the quality and consistency of raw IoT-enabled air conditioner usage data (Section 2.1). The second, **comfort-based indoor air temperature analysis**, applies statistics and an unsupervised learning algorithm to explore usage scenarios and derive personalized comfortable indoor air temperatures (Section 2.2 and 2.3). The third, **energy consumption simulation**, integrates the personalized comfortable thermostat setpoints derived in the second module with EnergyPlus simulations to evaluate air conditioner energy performance (Section 2.4).

2.1. Raw data preprocessing

2.1.1. Data cleaning

The historical air conditioner usage data was sourced from sensors integrated within the air conditioner units and collected via IoT reporting. The dataset comprises the following parameters: IoT timestamp, indoor air temperature (T_{ai}), outdoor air temperature (T_{ao}), setpoint temperature (T_{set}), operation mode, and operational status. Indoor and outdoor air temperatures were measured using dedicated sensors on the air conditioner units. User interactions are reflected in the setpoint temperature, operation mode, and operational status, where T_{set} indicates the user's desired indoor air temperature. The operation mode includes heating, cooling, ventilation, and dehumidification, while the operational status specifies whether the air conditioner is active or inactive. Existing studies have highlighted that this data acquisition method may introduce invalid or missing values, necessitating preprocessing to ensure analytical accuracy. Following existing studies [26,35], the data cleaning process employed the following rules:

·Samples with T_{set} above 30 °C and "NaN" were deleted.
·Samples with T_{ai} above 35 °C and T_{ao} above 45 °C were deleted.
·Samples in cooling and dehumidification modes were retained.

2.1.2. Time interpolation

Historical air conditioner usage data is typically uploaded at fixed intervals, with additional uploads triggered by user adjustments to the air conditioner settings. While this approach captures the dynamic characteristics of user interactions, it introduces irregularities into the time series, complicating temporal analysis. To address these inconsistencies, a time-resampling approach was applied to standardize

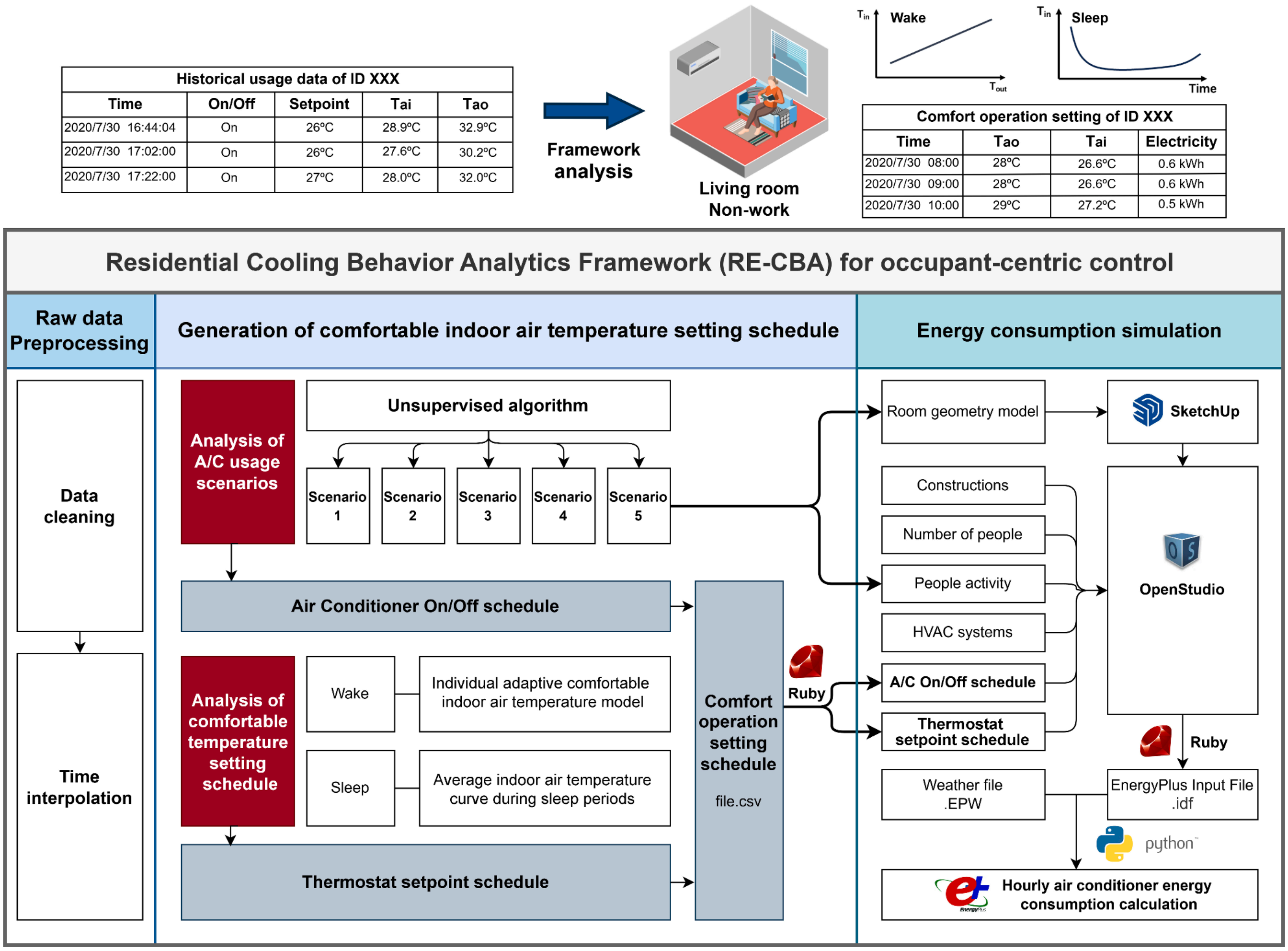

Fig. 1. Residential cooling behavior analytics framework (RE-CBA) for occupant-centric control and its energy-saving potential evaluation.

the dataset to fixed time intervals. The resampling process aligned all records to a consistent n-minute interval, ensuring data continuity and comparability. A complete n-minute time series spanning the dataset's temporal range was first generated, after which the original records were merged with this standardized time series. Missing time points were filled as follows:

·For T_{ai} and T_{ao}, linear interpolation was employed to estimate missing values, providing smooth transitions over time and reducing potential errors caused by data gaps.
·For T_{set} and operational status, forward filling was applied, propagating the most recent recorded values to maintain stability and consistency.

2.2. Usage scenario analysis

Key information, such as the type of room where the air conditioner is installed and the user's lifestyle habits, is often missing from historical air conditioner usage data. However, this information is critical for developing intelligent air conditioner control strategies and conducting accurate energy consumption assessments, as it reflects core factors influencing operational optimization. In the absence of predefined labels required for supervised classification of usage scenarios, unsupervised learning methods, particularly clustering analysis, provide an effective solution. By identifying intrinsic similarities within the data, clustering can automatically group air conditioner usage patterns into distinct scenarios, even without prior labels.

Clustering analysis organizes air conditioner usage data by grouping samples with similar usage probabilities across specific time periods, thereby reflecting potential lifestyle patterns or usage scenarios. The average characteristics of each cluster can be analyzed to infer possible room types or user routines. For example, a cluster exhibiting higher usage probabilities at night may correspond to air conditioners installed in bedrooms. Typical air conditioner usage scenarios in residential buildings include units serving bedrooms, living rooms, or multifunctional spaces (e.g., studios that combine sleeping and living areas). Furthermore, user groups can be categorized based on lifestyle routines, such as individuals with regular work or study schedules (e.g., employees and students) and those who spend extended periods at home, such as retirees. By capturing these variations, clustering analysis provides critical insights into user behavior, enabling the development of tailored operational strategies that balance energy efficiency and thermal comfort.

Given its simplicity and effectiveness for low-dimensional data, the K-means clustering algorithm is widely used for time series and behavioral data analysis. In this study, K-means clustering was applied to group air conditioners based on their 24 h usage probability profiles. Specifically, the usage probability for each hour of the day was calculated for each unit, resulting in a 24-dimensional feature vector. Each feature represents the ratio of the number of minutes the air conditioner was active in that hour to the total active minutes recorded over the observation period. This calculation ensures that the feature vectors capture the temporal distribution of usage patterns while normalizing differences in overall operational duration across units. The formulation for deriving the usage probability vector is presented in Eqs. (1) and (2). The Davies-Bouldin (DB) Score was utilized to evaluate the clustering performance, as it effectively quantifies the trade-off between intra-cluster compactness and inter-cluster separation [37]. This metric is particularly suitable for assessing clustering quality in multi-dimensional feature spaces, such as the 24-dimensional usage probability profiles analyzed in this study.

$$P_{i,h} = \frac{M_{i,h}}{T_i} \tag{1}$$

$$T_i = \sum_{h=1}^{24} M_{i,h} \tag{2}$$

Here, $P_{i,h}$ represents the usage probability of the i-th air conditioner during the h-th hour over the observation period. This probability is calculated as the ratio of $M_{i,h}$, the cumulative active minutes of the i-th air conditioner during the h-th hour, to T_i, the total active minutes of the i-th air conditioner across the entire recording period.

2.3. Comfortable indoor air temperature analysis

In air conditioner usage scenarios, users balance factors such as economic cost and thermal comfort, shaping their perception of environmental comfort. This subjective sense of comfort is a critical component of building thermal environment research, as it directly impacts the accuracy of energy consumption assessments. Traditionally, thermal comfort feedback has been collected through subjective votes in climate chamber experiments or in-home surveys [16,38]. However, such real-time feedback is often unavailable in air conditioner operation datasets. Analyzing user adjustment behaviors during air conditioner operation offers an alternative approach for assessing thermal comfort.

For awake states, comfortable indoor air temperature can be analyzed through two primary methods based on Operational Response Threshold (ORT) and Steady-state Average Temperature (SAT) [33,39]. The ORT method extracts data samples during temperature increase and decrease adjustments made by users, statistically deriving the upper and lower limits of unacceptable indoor temperatures. The SAT method, inspired by climate chamber studies, assumes that the indoor temperature stabilizes after a certain period (typically 10–20 mins) following air conditioner adjustments. The average indoor air temperature from the onset of this steady state until the user changes the thermostat setpoint or turns off the unit is defined as the comfortable indoor air temperature for that operational period, as described in Eq. (3) [14]:

$$\overline{T_{ai,comfort}} = 1 \Big/ (\Delta t - t') \int_{t_0+t'}^{t_0+\Delta t} T_{ai}(t)dt \tag{3}$$

Where Δt represents the total time in minutes between two consecutive user actions, such as changing the thermostat setpoint or switching the air conditioner on/off. t_0 is the timestamp of the last user action (either switch on or change the thermostat setpoint). $T_{ai}(t)$ is the indoor air temperature at time t. The parameter t' denotes the time required to reach the assumed steady-state condition.

Notably, existing research has yet to conduct a comparative analysis of the two methods for deriving comfortable indoor air temperature or to examine how the assumed "steady-state" time (t') in the SAT method influences the calculated comfortable temperature. To address this gap, this study performs a comparative analysis of these two methods and investigates the sensitivity of the SAT method to variations in the steady-state assumption.

For sleep states, determining comfortable indoor air temperatures based on thermostat setpoint adjustments is impractical, as users typically do not adjust air conditioner settings during sleep. For these periods, the mean indoor air temperature curve derived from historical usage data is used as a representative profile of relatively comfortable setpoints. Specifically, for rooms classified as bedrooms or other sleep-associated spaces, the comfort setting curve for sleep time operation is defined by the hourly average indoor air temperature curve from 22:00 to 8:00 in the historical dataset.

2.4. Energy calculation

After completing the usage scenario and comfortable indoor air temperature analyses, this study developed an EnergyPlus-based simulation workflow to evaluate the electricity consumption performance of air conditioner operation strategies derived from historical operation data when applied to automatic control. The process begins with identifying the annual cooling season based on the historical usage data . Subsequently, the clustering method described in *Section 2.2* is applied to categorize usage scenarios, with each unit sample assigned to its respective scenario. Using these scenarios, an hourly operational status schedule is generated for the cooling season. In parallel, a comfortable indoor air temperature model, developed as outlined in *Section 2.3*, incorporates outdoor meteorological data to produce an hourly thermostat setpoint schedule. Together, these schedules form a comfort-based operational settings table, which serves as the basis for simulation inputs.

Room geometry is modeled in SketchUp, with typical Chinese room layouts used as examples: bedroom areas range from 12 to 15 m^2, living rooms from 15 to 40 m^2, and mixed-use studio spaces from 20 to 40 m^2. These room models are then imported into OpenStudio, where detailed configurations are applied, including envelope properties, occupancy density, occupant activity schedules, and air conditioning systems. The comfort-based operational settings table, derived for individual air conditioner units, is integrated into OpenStudio using Ruby scripts to generate the necessary IDF file. The eppy library in Python is employed to combine the IDF file with the corresponding EPW weather file, enabling hourly simulations of air conditioner electricity consumption using the EnergyPlus solver. After completing the energy consumption simulation, the energy-saving performance of RE-CBA is quantified by comparing the total cooling consumption under dynamic control ($E_{\text{RE-CBA}}$) against the baseline (E_{baseline}), as expressed by Eq. (4).

$$\textit{Saving rate} = (E_{\text{baseline}} - E_{\text{RE-CBA}})/E_{\text{baseline}} \times 100\% \tag{4}$$

3. Dataset of the case study

This study utilized a historical air conditioner usage dataset provided by a Chinese air conditioner manufacturer, comprising operational data from 920 IoT-enabled air conditioners installed in residential buildings in Shanghai. The typical meteorological summer in Shanghai generally spans from late May to late September. To capture a comprehensive and representative set of user cooling behaviors, the data collection period from May to October 2020 was selected. This period includes both the peak cooling demand during the summer high-temperature months and the transitional period, enabling the analysis of variations in air conditioner usage as the seasons change. The dataset includes six core measurement parameters provided by air conditioner sensors and IoT-

Table 1
Basic measurements provided by air conditioner sensors and IoT report parameters.

Symbol	Measurement	Category	Accuracy
t	IoT time	IoT	1 s
T_{ai}	Indoor air temperature	Sensor	± 0.7 °C
T_{ao}	Outdoor air temperature	Sensor	± 0.7 °C
T_{set}	Thermostat setpoint	User settings	- (resolution: ± 0.5 °C)
M_{ac}	Operating mode	User settings	–
O	Operation state	User settings	–

reported data, detailed in Table 1. The IoT timestamp, sourced from internet time servers, has an accuracy of up to one second. Measurements were collected at a fixed 15 min interval, with additional data uploaded whenever users adjusted air conditioner settings. Approximately 2,500,000 raw operational data entries were initially recorded between May 1 and October 31, 2020. Following data cleaning, this number was refined to 2,463,679 entries.

Selecting an appropriate time granularity is crucial to minimize information loss while avoiding the introduction of unnecessary influences. Fig. 2 illustrates the adjustment patterns of thermostat setpoint by users in the dataset. Fig. 2a shows that most adjustments occur within the first minute after switching on the air conditioner. As the indoor air temperature approaches the setpoint value, the frequency of adjustments significantly decreases. Twenty minutes after switching on the air conditioner, the number of adjustments accounts for only 5.4 % of the total adjustments made after switching it on (Fig. 2b). Although it generally takes more than 10 mins for a household air conditioner to achieve a steady thermal environment [40], the data reveals that most thermostat setpoint adjustments occur within the first 1 min of operation. This suggests that the rapid and frequent adjustments to the setpoint immediately after switching on the air conditioner are likely not primarily driven by the users' thermal sensations. Instead, users tend to make these adjustments based on the air conditioner's temperature display, aiming to meet their psychological expectations. To avoid interference from high-frequency actions shortly after startup and to prevent the loss of valuable information due to excessive time intervals, a 1 min interval was selected for time granularity in this dataset. After time resampling, the dataset contains a total of 150,535,014 air conditioner historical usage data entries.

4. Results

4.1. Descriptive analysis of the case dataset

Fig. 3 illustrates the monthly distribution of indoor air temperatures triggering air conditioner on/off operations. The trigger temperatures for switching on air conditioners, representing the indoor air temperatures at which users initiate cooling. The trigger temperatures for switching off air conditioners, corresponding to the indoor air temperatures at which users deactivate cooling. The results show a gradual increase in trigger temperatures from May to August, followed by a decline. Specifically, in May, the trigger temperature for switching the air conditioner on is 25.7 °C (SD = 1.8, IQR = 2.0), while the off-trigger temperature is 24.6 °C (SD = 2.0, IQR = 2.3). By August, these thresholds rise to 29.1 °C (SD = 1.7, IQR = 2.1) and 26.9 °C (SD = 1.5, IQR = 1.8), respectively.

Fig. 4 presents the statistics of air conditioner usage duration throughout the study period, highlighting both daily and hourly variations. Fig. 4a presents the daily outdoor air temperature, while Fig. 4b shows the daily air conditioner usage duration. A strong correlation exists between average daily air conditioner usage duration and outdoor air temperature (Pearson's $r = 0.840$, two-sided $P < 0.0001$). The usage duration increases stepwise from May 15, peaking in mid-August, before sharply declining after this period, dropping below 10% by mid-September. Between mid-May (May 15) and early June (June 10), the average daily usage duration is 2.4 h (SD = 0.67, IQR = 0.81). Usage reaches its peak on August 14, with an average daily duration of approximately 7.8 h (SD = 0.94, IQR = 0.96). This peak period lasts about ten days before decreasing. By early September, the daily usage duration significantly drops, from 6.3 h on September 1 to just 0.63 h by September 15. Fig. 4c illustrates the average hourly usage rate throughout the day with nighttime usage consistently higher than daytime usage. From June to early July and again in early September, air conditioner is predominantly used between 20:00 and 07:00, with an average nighttime usage rate of 23.0 % (SD = 8.09%, IQR = 12.32%) and daytime rates below 20%. In August, the usage rate rises substantially, with nighttime (20:00–07:00) usage rate averaging 43.6% (SD = 6.87%, IQR = 9.34%) and daytime (08:00 –19:00) usage rate reaching an average of 33.3% (SD = 11.25%, IQR = 19.70%).

4.2. Usage scenario identification

Following the methodology outlined in Section 2.2, the hourly usage probability of each air conditioner over a 24 h period was analyzed, resulting in a 24-dimensional feature vector for each sample. An initial K-Means clustering analysis identified three distinct groups, each characterized by unique daily operational patterns. The average hourly usage probability curves for these clusters are presented in Fig. 5a

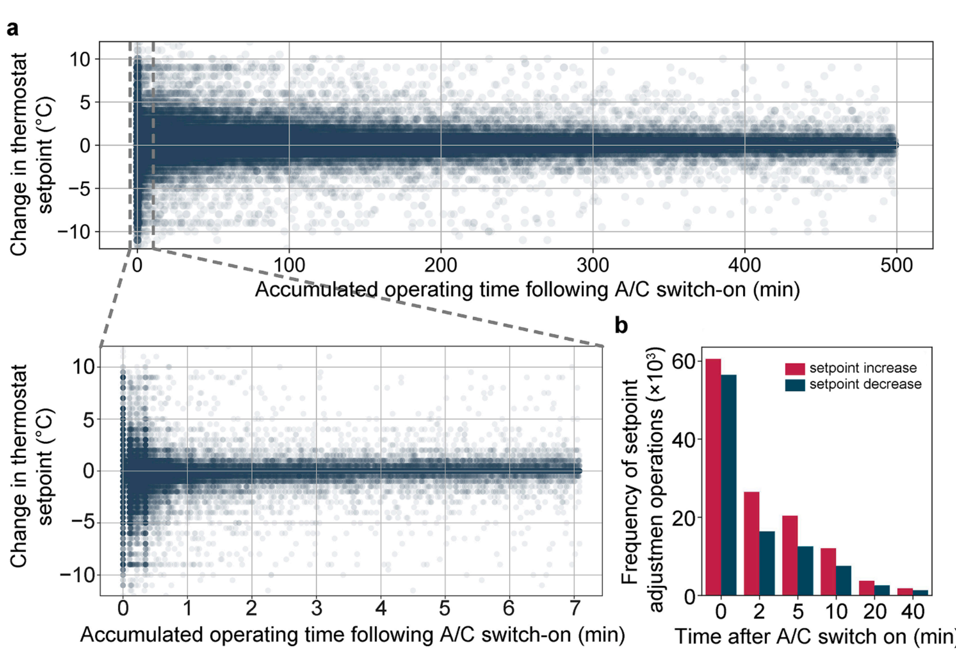

Fig. 2. Characteristics of thermostat setpoint adjustments (samples selected with operating durations ≤ Q3, 75th percentile). **a.** Magnitude of setpoint increases and decreases following A/C switch-on. **b.** Frequency of setpoint adjustment operations at 0, 2, 5, 10, 20, and 40 mins after A/C switch-on.

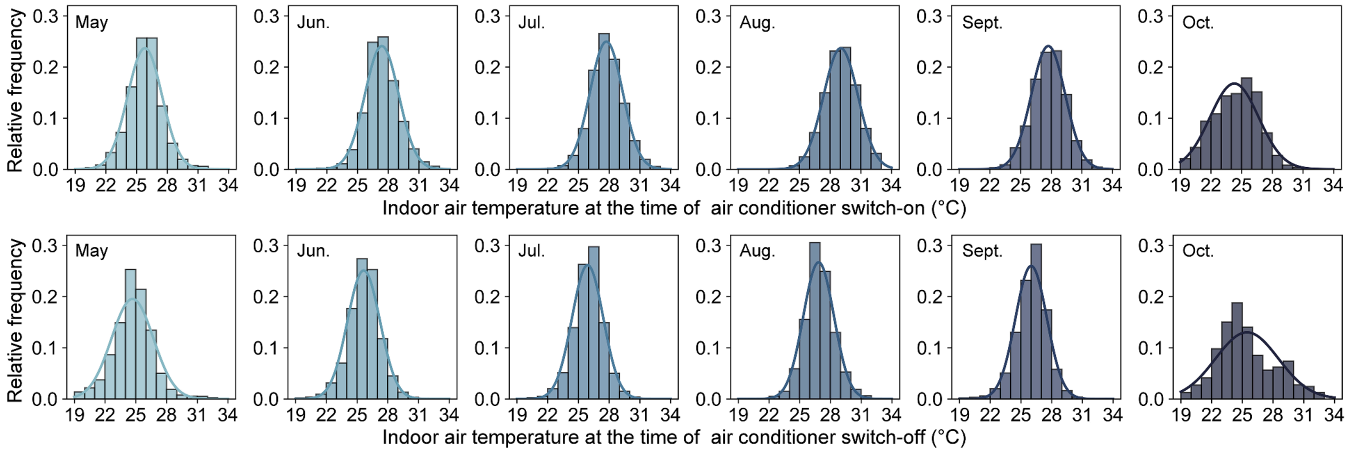

Fig. 3. Monthly distribution of indoor air temperatures triggering air conditioner on/off events. The upper illustrates the trigger temperatures for switching on air conditioners. The lower depicts the trigger temperatures for switching off air conditioners.

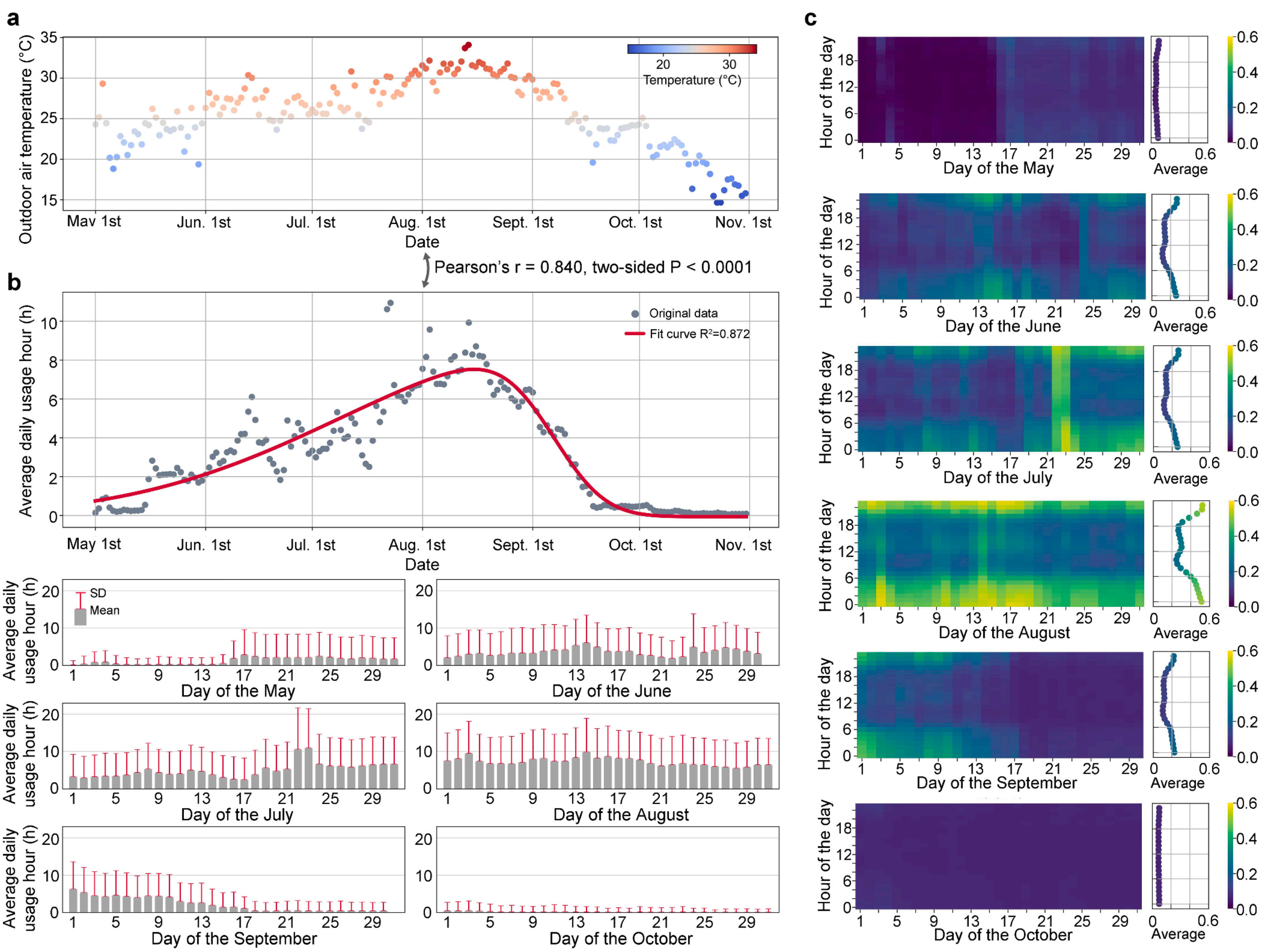

Fig. 4. Air conditioner usage duration statistics and daily outdoor air temperature. **a.** Daily outdoor air temperature throughout the study period. **b.** Daily average air conditioner usage duration. **c.** Hourly average air conditioner usage rate (usage duration/24 h).

(Within-Cluster Variance = 4.415, DB = 1.402). The usage characteristics of air conditioner samples across different clusters are summarized in Table 2, highlighting distinct operational behaviors. A secondary clustering analysis was performed on air conditioners within the mixed-use and living room groups to further refine the categorization, as shown in Fig. 5b (Within-Cluster Variance = 0.992, DB = 1.799) and Fig. 5c (Within-Cluster Variance = 1.163, DB Score = 1.739). Considering the inherent variability in human behavior, the clustering performance is within acceptable ranges, making the results a reasonable representation of actual usage patterns. Based on these clustering results, the entire

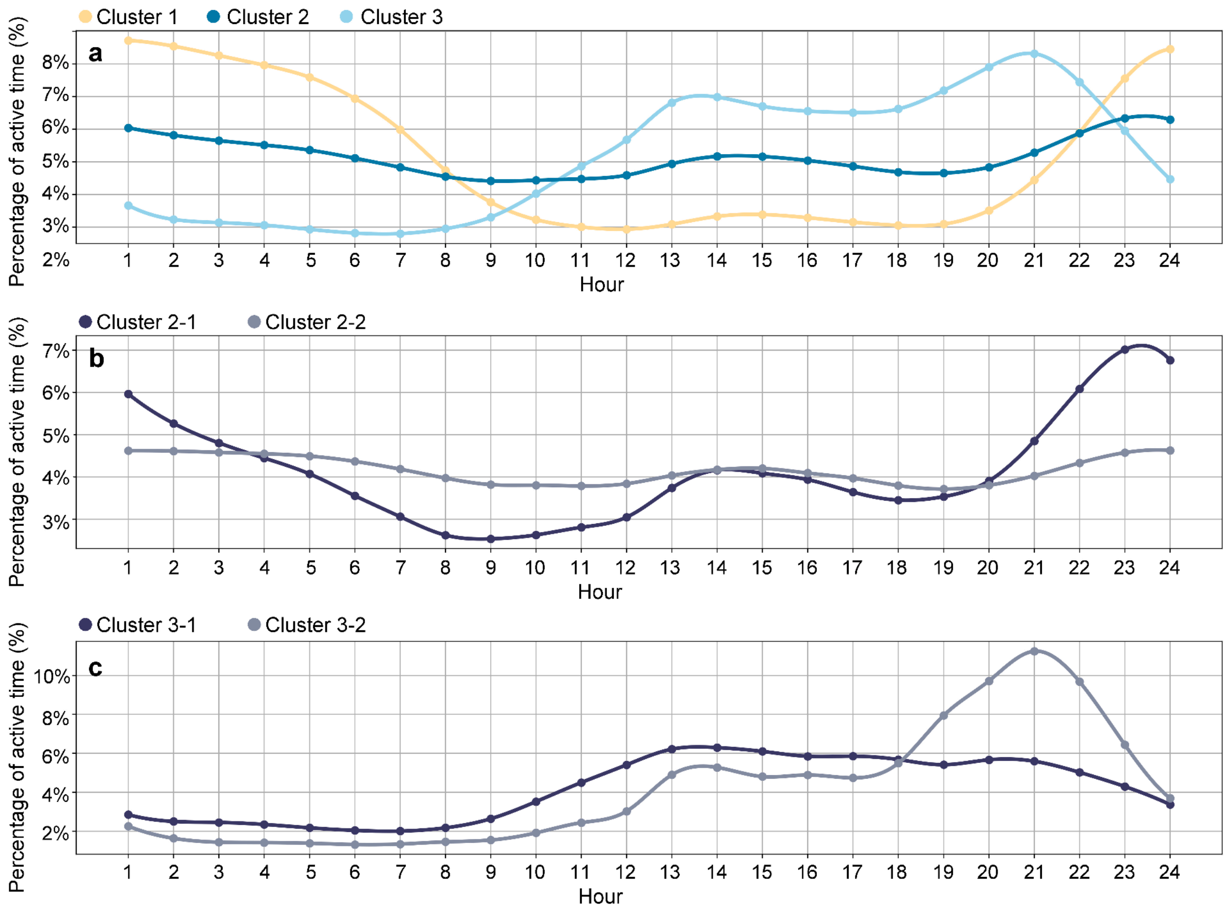

Fig. 5. Hourly average usage probability of each cluster group following K-means clustering analysis.

Table 2
Usage characteristics of air conditioner samples in different clusters.

Cluster	Usage characteristics	Similar room type	Similar user group	Number of Samples
Cluster1	Usage rate gradually increases after 22:00, peaks at night, and decreases to a low value after 09:00 in the morning.	Bedroom	No typical	330
Cluster2–1	Usage rate gradually increases after 22:00, decreases after 07:00, shows a peak from 12:00 to 13:00, then declines again.	Mixed room	Work	156
Cluster2–2	Usage rate remains consistent throughout the day.	Mixed room	Non-work	339
Cluster3–1	Usage rate increases after 09:00, peaks during the day, and drops to a low value after 22:00 at night.	Living room	Non-work	66
Cluster3–2	Usage rate Shows a minor peak from 12:00 to 13:00, high usage from 19:00 to 22:00, and consistently low usage from 23:00 to 09:00 the following morning.	Living room	Work	29

dataset was classified into five usage scenarios. It should be noted that the "Work" category refers to users who are away from their residential buildings during typical working hours. Due to the unique characteristics of the dataset, individuals working from home during the pandemic exhibit air conditioner usage patterns more similar to the "Non-work" category.

4.3. *Comfortable indoor air temperature model for awake and sleep states*

This study compares ORT method and SAT method and examines the impact of the "steady-state" determination time on the results, with a detailed comparative analysis is provided in the ***Discussion***. Based on this comparison, the optimal method was selected to analyze the comfortable indoor air temperature for each user in the dataset. Specifically, when a user does not adjust the thermostat, the mean indoor air temperature during the period starting 15 mins after stability is reached and ending when the user changes the setpoint or turns off the unit is defined as the comfort indoor air temperature for that operational period.

Adaptive thermal comfort theory highlights the influence of environmental context [31,41], emphasizing how individuals adapt to thermal discomfort through environmental interactions (e.g., adjusting clothing or activity levels [42]) and psychological expectations shaped by climatic and building conditions [43,44]. Based on this theory, this study investigates the relationship between comfortable indoor air temperature and outdoor air temperature. For 590 non-bedroom air conditioner samples, Fig. 6a shows the daily mean comfortable indoor air temperature alongside corresponding outdoor air temperatures. From May 15 to September 15, a significant correlation was observed between daily comfortable indoor air temperature and outdoor air temperature ($\beta = 0.164$, 95% CI = [0.144, 0.183], $R^2 = 0.691$, $P < 0.0001$), with significant synchrony in their trends ($\beta = 0.097$, 95% CI = [0.076, 0.110], $R^2 = 0.404$, $P < 0.0001$) (Fig. 6b).

Further data exploration was conducted for individual air

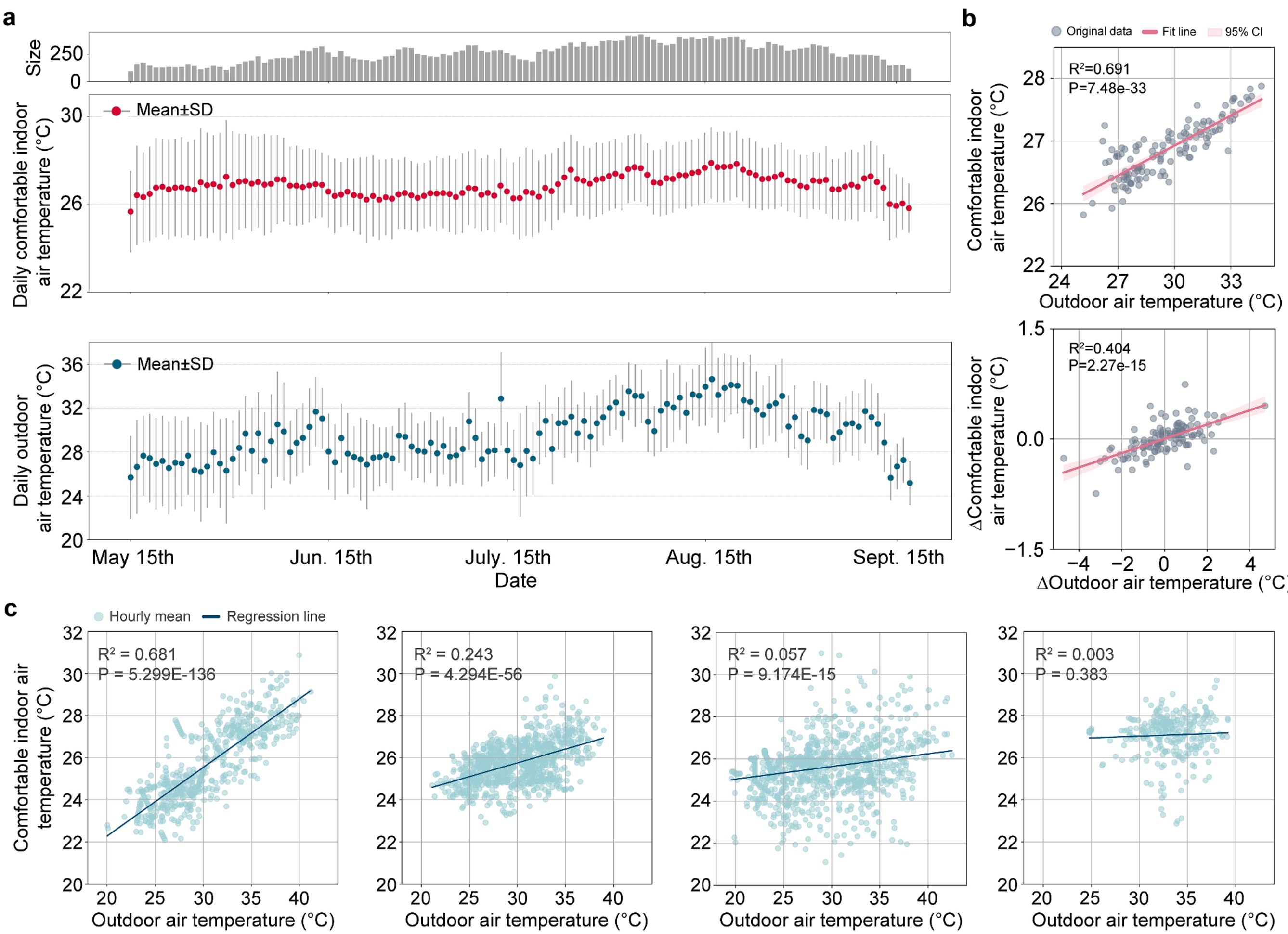

Fig. 6. User comfortable indoor air temperature. **a.** Daily comfortable indoor air temperature and corresponding outdoor air temperature. **b.** Relationship between daily comfortable indoor air temperature and outdoor air temperature (From May 15th to Sept 15th). **c.** Linear regression results for four representative users.

conditioners. For the 590 non-bedroom air conditioner samples, individual "comfortable indoor air temperature model" were developed. The analysis revealed significant variability in comfortable indoor air temperatures among users, both in absolute values and in patterns of change. The regression model slopes represent each user's adaptability to outdoor temperatures, with Fig. 6c displaying regression results for four representative users. Among the 590 users, 129 users exhibited a strong correlation between comfortable indoor air temperature and outdoor air temperature ($P < 0.05$, $R^2 > 0.5$), indicating higher adaptability. In contrast, 161 users showed no significant relationship between comfortable indoor air temperature and outdoor air temperature ($P > 0.05$), indicating consistent comfort preferences regardless of external conditions. For the remaining users, while a significant correlation was observed, their lower R^2 values and gentler slopes indicate reduced sensitivity of comfortable indoor air temperatures to outdoor air temperature fluctuations.

For the 825 air conditioner samples categorized as mixed-use rooms and bedrooms, Fig. 7a illustrates the distribution of average indoor air temperatures during sleep periods. The results show that 84.2 % of the samples had nighttime air temperatures between 25 °C and 29 °C, with the majority falling within the 26 °C to 27 °C range. The proportion of users experiencing indoor air temperatures between 24 °C and 25 °C was similar to those experiencing temperatures between 29 °C and 30 °C. Fig. 7b displays the hourly average indoor air temperature curve, showing that nighttime air temperatures remain relatively stable. This stability is likely due to the lack of manual adjustments to air conditioner settings during sleep. Compared to the comfortable indoor air temperature range of 26 °C to 28 °C observed during wake states, indoor air temperatures during sleep time exhibit greater individual variability. Fig. 7c presents the hourly average indoor air temperatures during sleep for four representative air conditioner users, highlighting significant individual differences. Notably, the preferred nighttime indoor air temperatures for two of these users differ by more than 5 °C.

4.4. Energy consumption for comfort-based air conditioner control

The cooling season energy consumption for different air conditioner samples in the dataset, driven by the individualized comfort-based indoor air temperature model, was evaluated using the simulation module within the proposed framework. The comprehensive description of the simulation inputs is provided in the ***Supplementary Materials Text S1***. To evaluate the performance of model-driven control, energy consumption for each usage scenario was also simulated with a constant thermostat setpoint of 26 °C as a baseline for comparison. This setpoint is recommended by the Chinese national standard GB 50736–2012 [45]. The results are presented in Fig. 8. The simulation results reveal that for a small subset of users with a preference for lower indoor air temperatures, the model-driven control resulted in higher energy consumption compared to the baseline. In non-bedroom scenarios, approximately 20% of users fell into this category, while in bedroom scenarios, the proportion was higher, at around 40%. Fig. 9 illustrates the daily energy consumption of the highest-consuming air conditioner sample within each usage scenario, along with its hourly consumption and indoor air temperature on the day with the highest daily consumption.

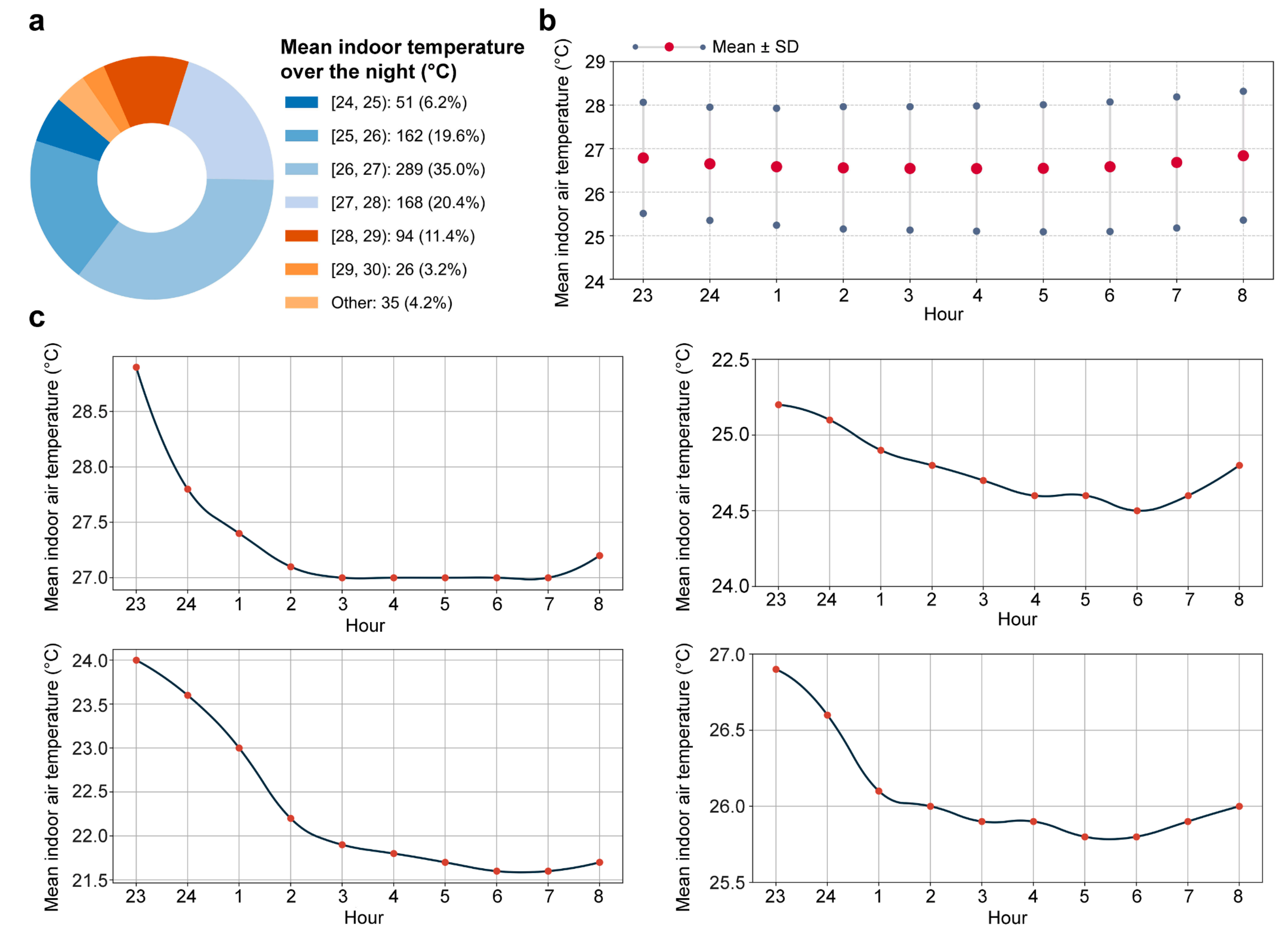

Fig. 7. Indoor air temperature preferences and patterns in mixed rooms and bedrooms during sleep periods. **a.** Proportion of air conditioner samples across different ranges of mean indoor air temperatures during sleep periods. **b.** Hourly mean indoor air temperatures during sleep periods averaged across 825 air conditioner samples. **c.** Hourly mean indoor air temperatures for four representative air conditioner users during sleep periods.

Users' preference for cooler air temperatures during sleep is a significant factor contributing to higher energy consumption. Furthermore, for users with a historical preference for lower comfortable indoor air temperatures, such as 24 °C or below, the air conditioner is required to maintain a lower setpoint for longer periods under the automatic control based on this framework, leading to relatively higher energy consumption. However, for the majority of users, model-driven air conditioner control resulted in lower energy consumption compared to a constant thermostat setpoint of 26 °C. Across the dataset's 920 air conditioner samples, model-driven control achieves 11.04% energy savings ($\Delta E =$ 116,657 kWh) compared to the static 26 °C baseline ($E_{\text{baseline}} =$ 1,057,131 kWh). This energy-saving benefit primarily arises from the model's ability to dynamically adjust comfortable setpoints based on outdoor air temperature changes. For most users, the comfortable setpoint calculated by the model is higher than the static 26 °C during certain periods, which reduces energy consumption compared to maintaining a constant 26 °C setpoint. This demonstrates the model's effectiveness in capturing energy-saving potential by accounting for users' adaptive comfort preferences. Assuming a 70% adoption rate among Shanghai's 8 million households, the deployment of this approach could yield citywide seasonal savings of 710,089,034 kWh (0.7×8 M $\times$ 116,657 kWh / 920). This is equivalent to saving 87,340 tons of standard coal (conversion factor: 0.123 kgce/kWh [46]) and reducing carbon dioxide emissions by 707,959 tons (conversion factor: 0.997 kg/kWh [47]). This conservative estimate excludes potential grid-interactive optimizations, which suggests even greater systemic benefits.

5. Discussion

5.1. Calculation method for comfortable indoor air temperature

Existing studies differ in their methods for analyzing comfortable indoor air temperature, primarily including the Operational Response Threshold (ORT) and the Steady-state Average Temperature (SAT) methods. A subset of non-bedroom air conditioner samples was used to analyze the impact of the assumed steady-state period length on results when applying the SAT method and conducted a comparative analysis of these two methods.

One of the critical issues when using the SAT method to determine comfortable indoor air temperature is selecting an appropriate duration as the starting point for the steady-state period. To address this issue, the time after each setpoint adjustment or start-up was divided into intervals, each representing the duration during which the user did not make any adjustments to the air conditioner, and calculated the average indoor air temperature for each interval. Fig. 10 shows the results for the 590 samples. Overall, as the time without user adjustments increased, indoor air temperature gradually stabilized. Table 3 presents the results of comparative analysis for different time intervals. The results show that data from the first 15 mins captures transient changes in the

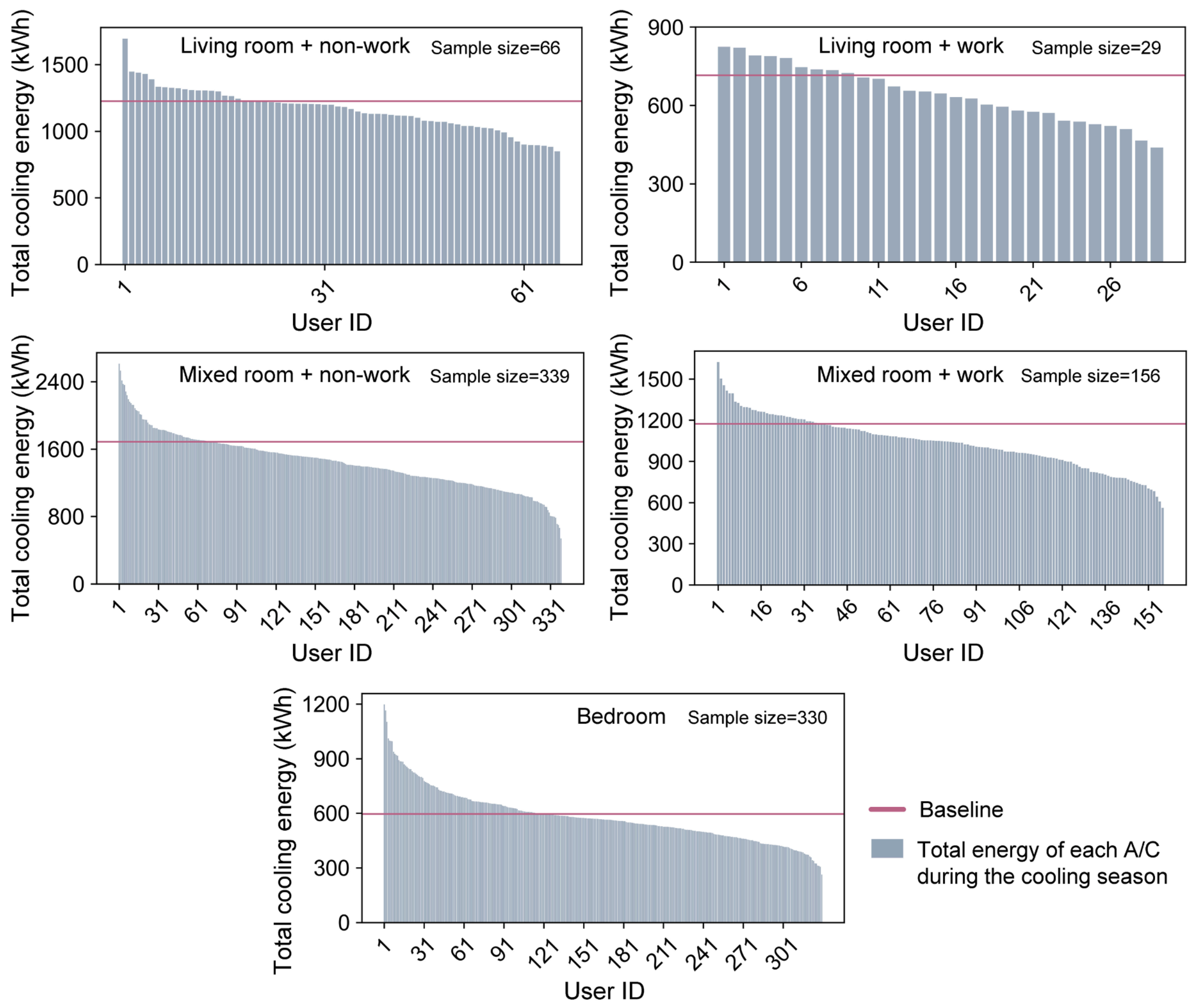

Fig. 8. Simulation results of total energy consumption during the cooling season for different air conditioner samples in the dataset.

environment following a thermostat setpoint adjustment, leading to significant differences from indoor air temperatures observed after longer periods. However, beyond 15–20 mins, variations in indoor air temperature showed no statistically significant differences. Based on this analysis, 15 mins is considered as the minimum duration required to reach a steady-state period, effectively filtering out short-term fluctuations while maintaining sufficient data volume and enhancing the validity of the analysis.

The basis of the ORT method is that users must perform enough setpoint adjustment operations. Fig. 11a summarizes the number of non-bedroom air conditioner samples corresponding to different frequencies of setpoint adjustment actions. During the data collection period, some samples had very few setpoint adjustment operations throughout the cooling season was observed. This may be due to users either being satisfied with the "default setpoint" upon switching-on or rarely using the air conditioner during the data collection period. To better represent the sample distribution, a logarithmic transformation to the number of actions was applied, resulting in a normal distribution (Fig. 11b). Subsequently, using anomaly detection algorithm for normal distributions, samples with fewer operations than $\mu - \sigma$ (109 samples) were identified as unsuitable for the ORT method due to insufficient recorded temperature adjustment operations. For samples with operation counts greater than or equal to $\mu - \sigma$ (481 samples), regression models for indoor air temperature against outdoor air temperature under both setpoint increase and decrease adjustments were developed. This resulted in the "lower bound equation" and "upper bound equation" for the acceptable indoor air temperature range.

To clearly compare the output differences between these two methods, 19 outdoor air temperature values were generated at 0.5 °C intervals between 26 °C and 35 °C. Each air conditioner sample was then analyzed using comfortable indoor air temperature model based on the ORT and SAT methods to predict either the comfortable indoor air temperature range or the specific comfortable indoor air temperature value corresponding to each outdoor air temperature.

To evaluate the differences between these two methods, an "Overall Score" was introduced as a comprehensive metric (See Eqs. (5) – (7)). The Overall Score consists of two components: cumulative distance and logical penalty. For each value generated by the SAT method, an assessment was made as to whether it fell within the range defined by the "lower bound equation" and "upper bound equation". If it fell outside this range, the cumulative distance between these points and the boundary was calculated (Fig. 11c). Additionally, due to the inherent randomness in some users' setpoint adjustment behaviors, the fitted "upper bound equation" values may occasionally exceed the "lower bound equation" values, violating the logical hierarchy of thermal boundaries (Fig. 11d). This discrepancy was therefore incorporated as a logical penalty in the overall score. The cumulative distance component measures the deviation of predicted temperatures from the boundary, while the logical penalty addresses instances that violate the thermal boundary hierarchy.

$$\text{Overall Score} = \sum_{i=1}^{N=19} (d_i + p_i) \tag{5}$$

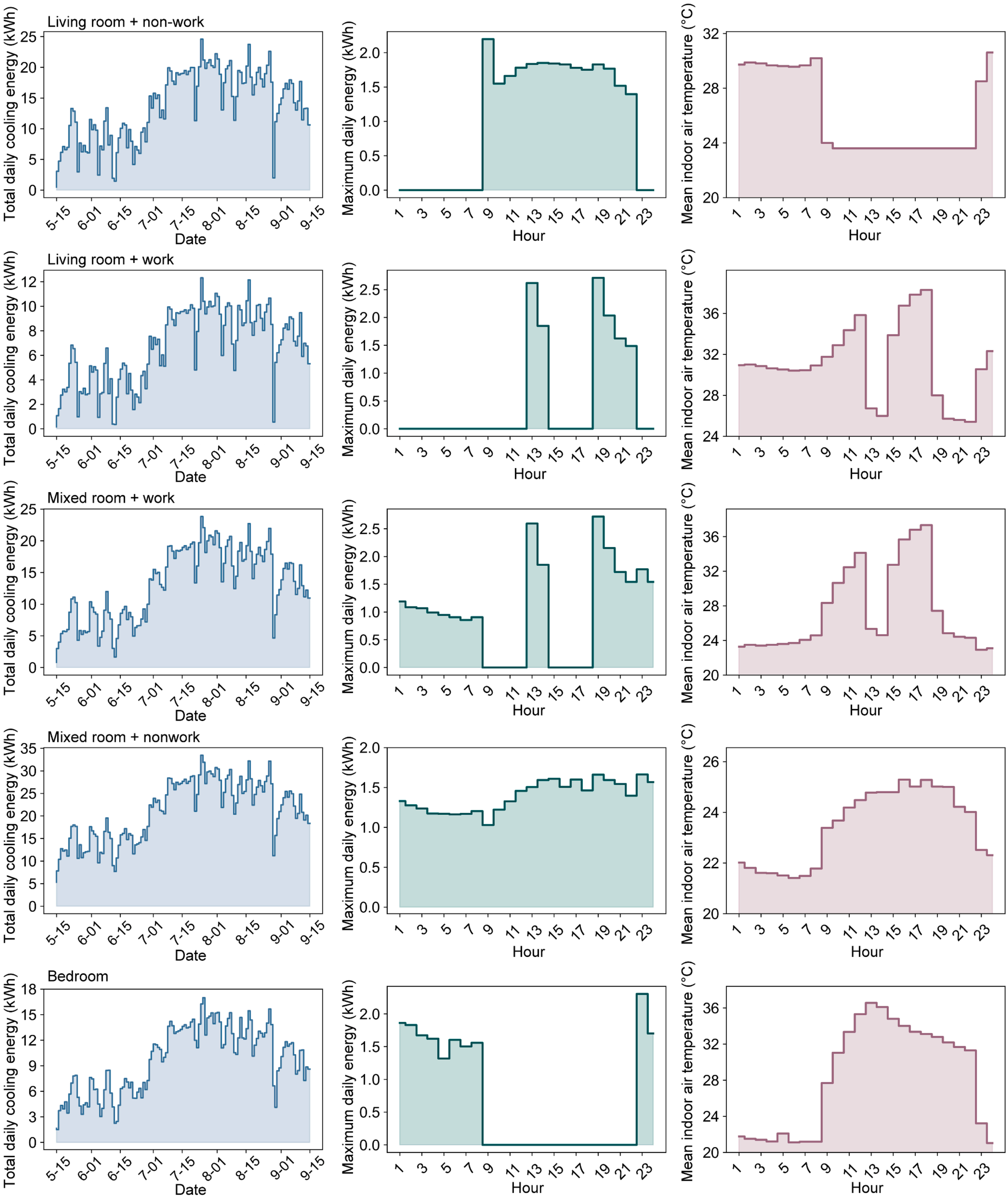

Fig. 9. Daily energy consumption of the highest-consuming air conditioner sample for each usage scenario, along with its hourly consumption and indoor air temperature on the peak consumption day.

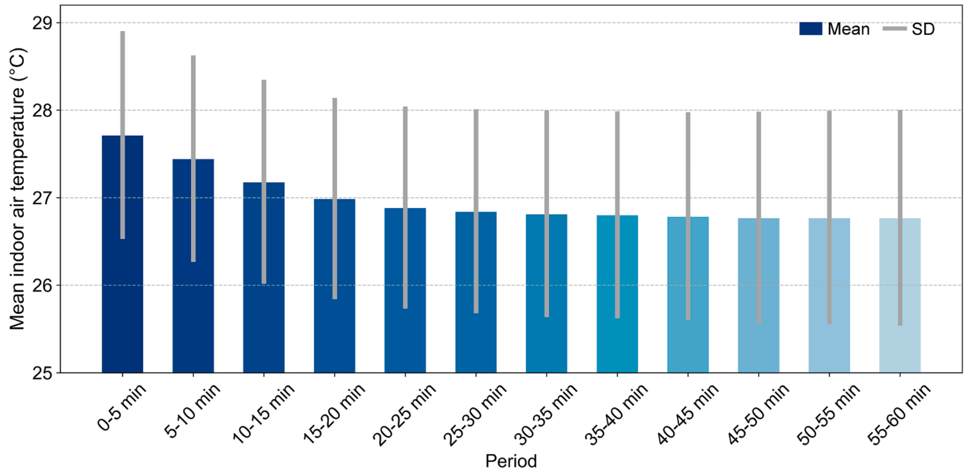

Fig. 10. Mean indoor air temperature at various time intervals following thermostat setpoint adjustments.

Table 3
Results of comparative analysis for different time intervals.

Comparison	t-stat	Original p-value	Adjusted p-value (Bonferroni)	Cohen's d	Sample Size (N)
0–5 min vs 5–10 min	26.27	4.74E-101	5.21E-100	0.25	590
5–10 min vs 10–15 min	32.86	5.43E-135	5.97E-134	0.23	590
10–15 min vs 15–20 min	22.54	1.92E-81	2.11E-80	0.20	590
15–20 min vs 20–25 min	17.71	1.63E-56	1.79E-55	0.09	590
20–25 min vs 25–30 min	10.42	1.92E-23	2.11E-22	0.04	590
25–30 min vs 30–35 min	5.01	7.25E-07	7.98E-06	0.02	590
30–35 min vs 35–40 min	2.33	0.02	0.222	0.01	590
35–40 min vs 40–45 min	4.74	2.70E-06	2.97E-05	0.01	590
40–45 min vs 45–50 min	2.27	0.02	0.261	0.01	590
45–50 min vs 50–55 min	−0.02	0.98	1	0.00	590
50–55 min vs 55–60 min	0.35	0.72	1	0.00	590

$$d_i = \begin{cases} \left| y_{comfort} - y_{lower} \right|, & y_{comfort} < y_{lower} \\ 0, & y_{lower} \leq y_{comfort} \leq y_{upper} \\ \left| y_{comfort} - y_{upper} \right|, & y_{comfort} > y_{upper} \end{cases} \tag{6}$$

$$p_i = \begin{cases} \left| y_{upper} - y_{lower} \right|, & y_{upper} < y_{lower} \\ 0, & y_{lower} \leq y_{upper} \end{cases} \tag{7}$$

Fig. 11e presents the Overall Score for different samples. The findings indicate that when users performed fewer than 100 setpoint adjustments, the output results of the models derived from the ORT method and the SAT method differed significantly, with the ORT model occasionally violating thermal boundary logic. In contrast, for users who frequently adjusted the setpoint, the results from both methods were more consistent. Although actual thermal sensation votes from users are lacking to further validate the accuracy of these methods, the comparative results suggest that the comfortable indoor air temperature model based on the SAT method appears to be a more robust choice.

5.2. Dynamic response of indoor thermal environment to outdoor climate

According to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), global warming has significantly increased the frequency of extreme weather events, posing substantial threats to natural ecosystems. This study, using real historical usage data, demonstrates the correlation between extreme climate conditions and cooling demand. Historical meteorological data for Shanghai indicates that July 22 and July 23, 2020, were extreme heat days. On these days, the daily air conditioner usage rate increased by more than 20%, and the hourly usage rate nearly doubled compared to adjacent days (Fig. 4). During sudden high-temperature events, residents may find it difficult to adapt quickly, resulting in increased reliance on air-conditioned spaces for comfort. This reliance not only reduces outdoor activities but also leads to a significant increase in indoor energy consumption. According to the IEA's statistical report, in East and Central China, peak cooling demand during summer heatwaves can persist for up to 10 days or longer, aligning with the patterns observed in this study [4]. Extreme weather conditions pose significant challenges for electricity producers, making it increasingly difficult to meet cooling demand during summer heatwaves, thereby driving up capacity costs and energy prices [48]. As expectations for thermal comfort rise and the number of hot weather days increases, air conditioner usage continues to grow. However, advancements in equipment performance require time, and market adoption of high-efficiency devices remains challenging [49]. Recent trends indicate a declining focus among builders and architects on designing buildings with high-efficiency, low-energy cooling systems [4]. Therefore, exploring strategies to curb the rapid growth of cooling demand in China's building sector is essential [50]. Enhancing data collection and collaborating with manufacturers can help identify emerging trends, technological requirements, and energy efficiency opportunities to achieve sustainable cooling. With advancements in sensor technology, the proliferation of IoT, and algorithm optimization, building cooling energy can now be effectively optimized using big data.

In defining comfortable indoor air temperature ranges, China primarily refers to ASHRAE-55 [28] and the national standard GB 50736–2012 [45]. ASHRAE-55 recommends Fanger's predicted mean vote (PMV) model [51] for defining neutral ranges in mechanically cooled environments, while de Dear's adaptive model is applied to naturally ventilated environments. For a typical summer room, China's national standard GB 50736–2012 and the PMV model define fixed comfortable or neutral indoor air temperature ranges for mechanically cooled environments (see Fig. 12a). However, analysis of historical IoT air conditioner usage data from residential buildings in Shanghai reveals that some users exhibit significant adaptability to indoor thermal

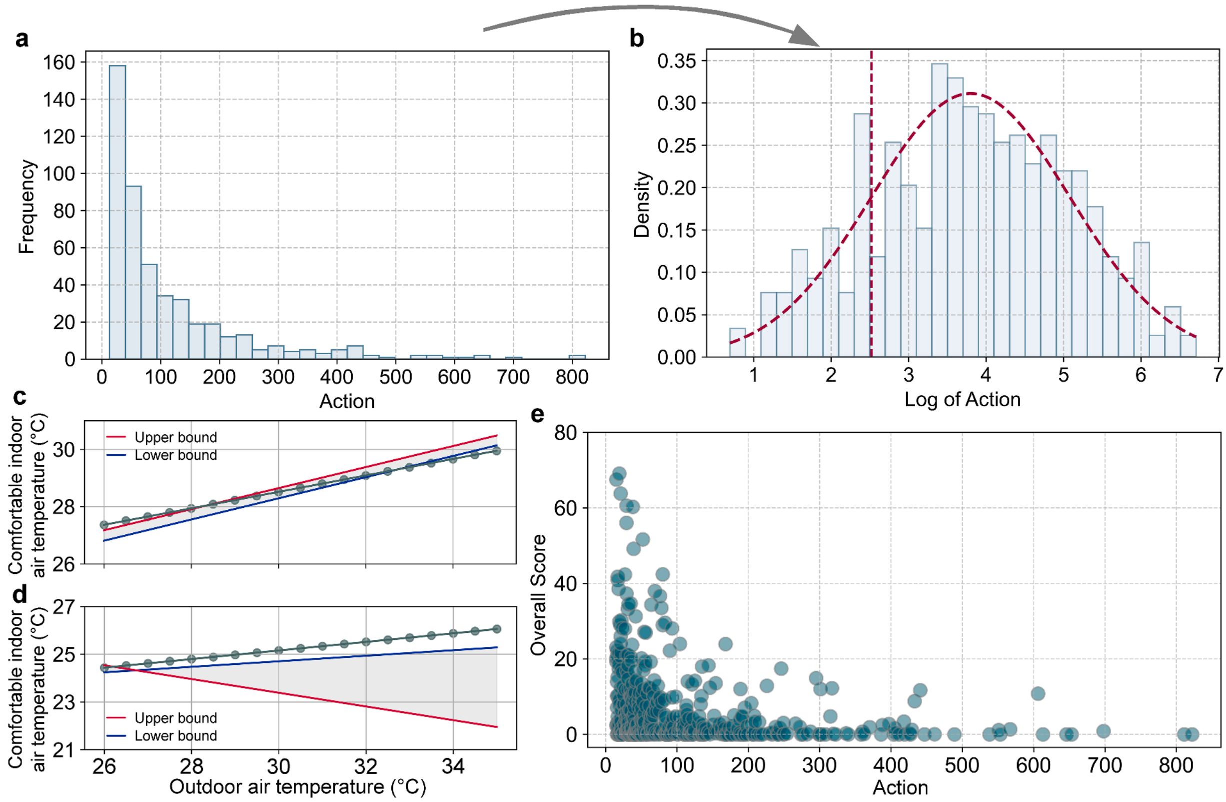

Fig. 11. Comparison of calculation methods for comfortable indoor air temperature. **a.** Number of non-bedroom air conditioner samples corresponding to different frequencies of setpoint adjustment actions. **b.** Frequency distribution of the number of actions after logarithmic transformation. **c.** Typical case: minimal output difference between methods **d.** Typical case: large output differences with thermal boundary logical hierarchy in OTR results. **e.** Overall score for different air conditioner samples with varying action frequencies.

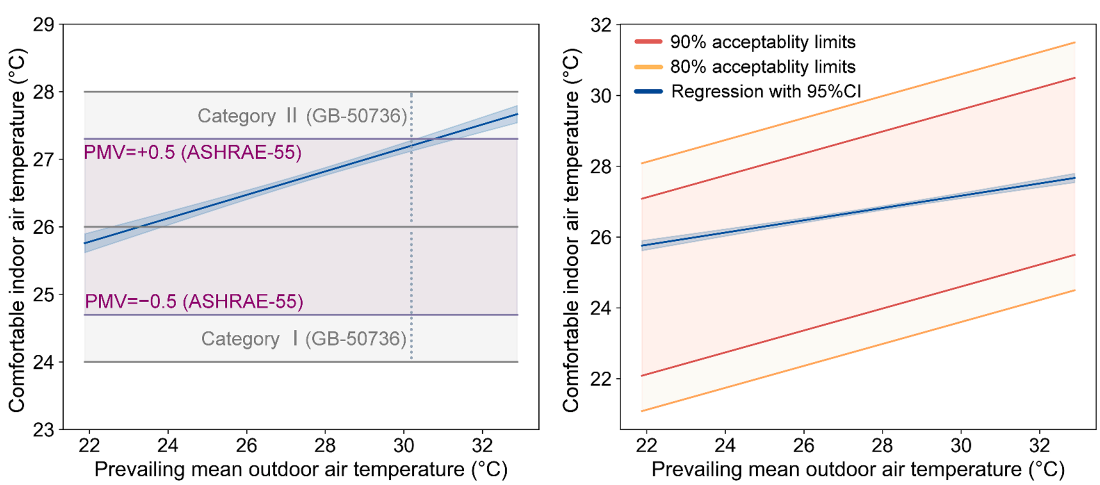

Fig. 12. Comparison of comfortable indoor air temperatures in Shanghai residences with existing standards. **a.** Comparison with the comfortable indoor air temperature range for air-conditioned rooms defined by ASHRAE-55 and the Chinese National Standard GB 50736–2012 (grey lines indicate comfort zone limits from GB 50736; purple lines indicate neutral indoor air temperature limits from ASHRAE-55). **b.** Comparison with the comfortable indoor air temperature range for naturally ventilated rooms defined by ASHRAE-55.

environments, even in mechanically cooled conditions. This adaptability is reflected in two key observations: first, both the air conditioner switch-on and switch-off trigger temperatures show a notable upward trend with increasing outdoor air temperatures; second, a significant positive correlation is observed between most users' comfortable indoor air temperatures and outdoor air temperatures. These findings align closely with de Dear's adaptive thermal comfort theory. Using data from 590 non-bedroom air conditioners during awake periods, de Dear's adaptive thermal comfort model was fitted to establish a linear relationship between the prevailing mean outdoor air temperature and the daily average comfortable indoor air temperature (regression statistics are detailed in Table 4) [31]. The regression results reveal that the ranges specified by the national standard GB 50736–2012 and the PMV model fail to fully capture the dynamic nature of residents' comfortale air temperatures in residential buildings as outdoor conditions change. While the fitted regression line generally falls within the ranges defined

Table 4
Linear regression results between prevailing mean outdoor air temperature and daily mean comfortable indoor air temperature.

	Coef	Std	t	$P > t$	95 % CI	
					0.025	0.975
Intercept	21.960	0.305	71.998	0.000	21.356	22.564
Prevailing mean outdoor air temperature	0.174	0.011	15.823	0.000	0.152	0.195

by these standards, the slope of the linear regression between the prevailing mean outdoor air temperature and residents' daily average comfortable air temperature is statistically significant, although slightly lower than the slope of de Dear's adaptive model (see Fig. 12b). This finding reinforces the insights that, even in mechanically cooled residential buildings, residents' thermal comfort exhibits dynamic adaptability to outdoor air temperatures rather than adhering to fixed comfort ranges, aligning with previous studies [12,31,42,52].

A key unresolved issue is whether this adaptability to indoor air temperatures is active or passive. On the one hand, residents may actively adjust to higher indoor air temperatures due to physiological thermoregulation mechanisms, which allow their bodies to perceive similar thermal sensations at higher outdoor air temperatures. On the other hand, the adaptability might be passive, driven by the inability of air conditioning systems to achieve lower air temperatures under extreme heat conditions. In such cases, residents might resort to alternative cooling strategies, such as adjusting clothing, using fans, or consuming cold beverages, to compensate. Regardless of whether this adaptability is active or passive, the observed upward shift in the acceptable indoor air temperature range under higher outdoor air temperatures represents a significant opportunity for energy savings. In high-temperature climates, it may not be necessary to rely solely on air conditioning to create excessively low-temperature environments that promote habitual reliance on cooling for comfort. This study supports the energy-saving potential of adaptive thermal comfort theory when applied to automatic adjustments in residential air conditioning systems [16].

Another critical aspect deserving attention is the determination of the cooling season. Unlike the well-defined heating periods for urban areas during winter in China, there is no consensus on recommended cooling periods for summer. This study reveals that residents' cooling demand deviates from the meteorological definition of summer (period during daily maximum temperature of 25 °C [29]), with the cooling period starting earlier and ending sooner than expected. Dynamically determining cooling periods based on the operational patterns of IoT-enabled air conditioners may provide a more accurate definition of cooling periods, supporting improved electricity load forecasting and energy-efficient management.

5.3. Limitations

The findings underscore the potential of such frameworks in advancing the understanding of occupant behavior and energy efficiency in residential cooling systems. However, the study has inherent limitations. The most critical limitation lies in the current control strategy, which is entirely based on analyzing historical user behavior to drive dynamic air conditioner control and evaluate energy consumption. While this approach has demonstrated energy savings compared to a constant thermostat setpoint of 26 °C, it does not explore further energy-saving opportunities through behavioral optimization. For instance, if the thermostat were to automatically increase by 1 °C from a 27 °C setpoint without notifying the user, would occupants still accept the adjusted 28 °C as comfortable? Shifting from merely learning user behavior to intelligently optimizing it through enhanced human-machine interaction presents a promising pathway for achieving greater energy savings. Additionally, privacy and security considerations limited access to user profile information, such as age distribution and gender composition, limiting the analysis of how demographic factors influence air conditioner usage. The lack of precise geographic coordinates also precluded detailed investigations into the impact of localized microclimatic variations on air conditioner operation. Furthermore, the dataset spans only a single year. Although based on one year of air conditioner usage data, we have identified adaptive patterns in comfortable indoor air temperatures in relation to outdoor air temperatures, but this dataset constrains the ability to account for interannual variability in thermal demand patterns. Notably, the dataset was collected during the COVID-19 pandemic. While Shanghai did not implement a full lockdown in 2020, the outbreak led to the implementation of work-from-home and online education policies in certain workplaces and schools. By April, work and study began to return to normal. The data for this study was collected from May to October 2020, which somewhat minimizes the impact of pandemic-related restrictions. However, residents' behaviors during this period may not have fully returned to baseline [53]. Some residents may have increased their demand for fresh air due to health concerns, potentially opening windows more frequently while operating air conditioning, which could have raised indoor air temperatures or increased the air conditioning load. Additionally, the economic uncertainty and health anxiety in the first half of 2020 may have led some households to adopt more conservative cooling usage behaviors. Despite these constraints, the extensive data collected by air conditioner manufacturers offers significant opportunities. Encouraging the sharing of anonymized, multi-year datasets could enable more comprehensive analyses using the proposed framework. Such advancements would refine the understanding of thermal behavior across diverse climatic and demographic contexts, providing actionable insights for optimizing energy efficiency and shaping adaptive cooling strategies to meet future challenges.

6. Conclusions

As global cooling demands rise, balancing occupant thermal comfort with energy efficiency has become a critical challenge in residential buildings. In smart residential buildings, automation of cooling systems is key to providing occupant comfort without requiring manual adjustments, while capturing energy-saving opportunities. This study develops an analytics framework for historical air conditioning usage behavior, aimed at automatically generating personalized and comfortable air conditioning settings. Based on the framework, a case study in Shanghai was conducted. The conclusions are as follows:

1. Cooling periods, usage scenarios, and personalized indoor air temperature preferences can be derived from IoT-enabled air conditioner usage data. These could be further used to drive air conditioner automatic control and to optimize energy simulation in residential buildings. Based on this, this study developed an occupant-centric residential cooling behavior analytics framework (RE-CBA) to realize this information extraction procedure.
2. Application of the RE-CBA framework to a dataset of 920 residential air conditioners in Shanghai revealed that users' comfortable indoor air temperatures adapt dynamically to outdoor conditions, increasing with rising outdoor air temperatures. This supports adaptive thermal comfort theory and making adaptive theory could be practically used for automatic control without relying on subjective feedback.
3. Energy simulations of the case study dataset demonstrated that the personalized control strategy derived from the framework, when compared to a fixed setpoint of 26 °C, more effectively aligns with users' adaptive thermal behaviors. Across the 920 air conditioner samples in the dataset, the personalized control strategy resulted in an 11.04 % energy savings ($\Delta E = 116{,}657$ kWh) compared to the static 26 °C baseline.

CRediT authorship contribution statement

Junmeng Lyu: Writing – original draft, Visualization, Resources, Methodology, Formal analysis, Conceptualization. **Yuxin Yang:** Writing – original draft, Methodology, Formal analysis. **Dayi Lai:** Writing – review & editing, Supervision. **Li Lan:** Writing – review & editing. **Zisheng Zhao:** Resources, Data curation. **Heng Du:** Writing – review & editing, Methodology. **Zhiwei Lian:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the National Key R&D Program of China (2022YFC3803201).

Code availability

The RE-CBA framework was developed using custom-made scripts in Python 12.0. The framework has been open-sourced and is available on GitHub for use and further development (https://github.com/Lyu2Patrick/Residential-Cooling-Behavior-Analytics-Framework). The air conditioner comfort setting table (.cvs) file has also been integrated into the Ruby script for OpenStudio, which has been made available on GitHub for public access and use. All statistical analyses and visualizations were performed using Python libraries, including pandas [54], NumPy [55], SciPy [56], Matplotlib [57], and Seaborn [58].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.buildenv.2025.112792.

Data availability

Data will be made available on request.

References

[1] Intergovernmental Panel on Climate Change, Climate Change 2022: Impacts, Adaptation and Vulnerability, IPCC, Geneva, 2022.

[2] International Energy Agency, Building Energy Performance Metrics, IEA, Paris, 2015. https://www.iea.org/reports/building-energy-performance-metrics.

[3] International Energy Agency, The Future of Cooling (2018). https://webstore.iea.org/the-future-of-cooling.

[4] International Energy Agency, The Future of Cooling in China, IEA, Paris, 2019. https://www.iea.org/reports/the-future-of-cooling-in-china.

[5] W. Wu, H.M. Skye, Residential net-zero energy buildings: review and perspective, Renew. Sustain. Energy Rev. 142 (2021) 110859, https://doi.org/10.1016/j.rser.2021.110859.

[6] Z. (Jerry) Yu, F. Haghighat, B.C.M. Fung, E. Morofsky, H. Yoshino, A methodology for identifying and improving occupant behavior in residential buildings, Energy 36 (2011) 6596–6608, https://doi.org/10.1016/j.energy.2011.09.002.

[7] X. Xu, H. Yu, Q. Sun, V.W.Y. Tam, A critical review of occupant energy consumption behavior in buildings: how we got here, where we are, and where we are headed, Renew. Sustain. Energy Rev. 182 (2023) 113396, https://doi.org/10.1016/j.rser.2023.113396.

[8] T.N.T. Lam, K.K.W. Wan, S.L. Wong, J.C. Lam, Impact of climate change on commercial sector air conditioning energy consumption in subtropical Hong Kong, Appl. Energy 87 (2010) 2321–2327, https://doi.org/10.1016/j.apenergy.2009.11.003.

[9] National Bureau of Statistics, Number of Durable Consumer Goods Owned Per 100 Urban Households at Year-End, Yearbook Press of China, 2020, https://doi.org/10.39163/y.cnki.yfggh.2021.001161.

[10] P.H. Shaikh, N.B.M. Nor, P. Nallagownden, I. Elamvazuthi, T. Ibrahim, A review on optimized control systems for building energy and comfort management of smart sustainable buildings, Renew. Sustain. Energy Rev. 34 (2014) 409–429, https://doi.org/10.1016/j.rser.2014.03.027.

[11] H. Yu, F. Zhong, Y. Du, X. Xie, Y. Wang, X. Zhang, S. Huang, Short-term cooling and heating loads forecasting of building district energy system based on data-driven models, Energy Build. 298 (2023) 113513, https://doi.org/10.1016/j.enbuild.2023.113513.

[12] R. de Dear, J. Kim, T. Parkinson, Residential adaptive comfort in a humid subtropical climate—Sydney Australia, Energy Build. 158 (2018) 1296–1305, https://doi.org/10.1016/j.enbuild.2017.11.028.

[13] S. Karjalainen, Thermal comfort and gender: a literature review, Indoor Air 22 (2012) 96–109, https://doi.org/10.1111/j.1600-0668.2011.00747.x.

[14] R.F. Rupp, N.G. Vásquez, R. Lamberts, A review of human thermal comfort in the built environment, Energy Build. 105 (2015) 178–205, https://doi.org/10.1016/j.enbuild.2015.07.047.

[15] Y. Song, Y. Sun, S. Luo, Z. Tian, J. Hou, J. Kim, T. Parkinson, R. de Dear, Residential adaptive comfort in a humid continental climate – Tianjin China, Energy Build. 170 (2018) 115–121, https://doi.org/10.1016/j.enbuild.2018.03.083.

[16] Y. Yang, J. Lyu, Z. Lian, Y. Xie, Y. Jiang, J. Lin, J. Niu, Subjective information in thermal comfort evaluation methods: a critical review, Energy Build. 325 (2024) 115019, https://doi.org/10.1016/j.enbuild.2024.115019.

[17] M. Rahimi, Classification and analysis of occupant behavior during earthquake shaking, in: Earthquake Engineering Tenth World Conference, 1992, pp. 6085–6090, 1992.

[18] J.S. Weihl, P.M. Gladhart, Occupant behavior and successful energy conservation: findings and implications of behavioral monitoring, in: Proceedings of the 1990 ACEEE Summer Study on Energy Efficiency in Buildings, 1990, pp. 171–180.

[19] A.F. Emery, C.J. Kippenhan, A long term study of residential home heating consumption and the effect of occupant behavior on homes in the Pacific Northwest constructed according to improved thermal standards, Energy 31 (2006) 677–693, https://doi.org/10.1016/j.energy.2005.04.006.

[20] W. Kempton, D. Feuermann, A.E. McGarity, I always turn it on super": user decisions about when and how to operate room air conditioners, Energy Build. 18 (1992) 177–191, https://doi.org/10.1016/0378-7788(92)90012-6.

[21] Z. Lin, S. Deng, A questionnaire survey on sleeping thermal environment and bedroom air conditioning in high-rise residences in Hong Kong, Energy Build. 38 (2006) 1302–1307, https://doi.org/10.1016/j.enbuild.2006.04.004.

[22] G.Y. Yun, K. Steemers, Behavioural, physical and socio-economic factors in household cooling energy consumption, Appl. Energy 88 (2011) 2191–2200, https://doi.org/10.1016/j.apenergy.2011.01.010.

[23] M. Schweiker, M. Shukuya, Comparison of theoretical and statistical models of air-conditioning-unit usage behaviour in a residential setting under Japanese climatic conditions, Build. Environ. 44 (2009) 2137–2149, https://doi.org/10.1016/j.buildenv.2009.03.004.

[24] D. Yan, T. Hong, B. Dong, A. Mahdavi, S. D'Oca, I. Gaetani, X. Feng, IEA EBC Annex 66: definition and simulation of occupant behavior in buildings, Energy Build. 156 (2017) 258–270, https://doi.org/10.1016/j.enbuild.2017.09.084.

[25] H. Liu, H. Sun, H. Mo, J. Liu, Analysis and modeling of air conditioner usage behavior in residential buildings using monitoring data during hot and humid season, Energy Build. 250 (2021) 111297, https://doi.org/10.1016/j.enbuild.2021.111297.

[26] M. Luo, K. Jiang, J. Wang, W. Feng, L. Ma, X. Shi, X. Zhou, Data-driven thermal preference prediction model with embodied air-conditioning sensors and historical usage behaviors, Build. Environ. 220 (2022) 109269, https://doi.org/10.1016/j.buildenv.2022.109269.

[27] Q. Fan, X. Xu, P. Liu, H. Zhang, S. Tang, A data-driven framework for thermal comfort assessment method based on user interaction, J. Build. Eng. 82 (2024), https://doi.org/10.1016/j.jobe.2023.108294.

[28] S. ASHRAE, Thermal environmental conditions for human occupancy. ANSI/ASHRAE Standard 55-2023, American Society of Heating, Refrigerat. Air Condition. Eng. 145 (2023).

[29] International Organization for Standardization, ISO 7730:2005 ergonomics of the thermal environment: analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria, (2005).

[30] P.O. Fanger, Thermal comfort, in: Analysis and Applications in Environmental Engineering, 45, Danish Technical Press, Copenhagen, 1970, p. 244. http://www.cabdirect.org/abstracts/19722700268.html%5Cnpapers2://publication/uuid/5CE163C3-F9AC-4937-A143-1238F1D806C5.

[31] R. de Dear, G. Brager, Developing an Adaptive Model of Thermal Comfort and Preference, United States: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States), United States, 1998, pp. 145–167.

[32] W. Zhou, Y. Chen, Y. Liu, Z. Yu, D. Wang, A new method for dynamic load simulation of urban building complexes' heating considering building level occupancy based on shoebox algorithm, Build. Environ. 261 (2024) 111710, https://doi.org/10.1016/j.buildenv.2024.111710.

[33] K. Kaspar, K. Nweye, G. Buscemi, A. Capozzoli, Z. Nagy, G. Pinto, U. Eicker, M. M. Ouf, Effects of occupant thermostat preferences and override behavior on residential demand response in CityLearn, Energy Build. 324 (2024) 114830, https://doi.org/10.1016/j.enbuild.2024.114830.

[34] B. Huchuk, W. O'Brien, S. Sanner, A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data, Build. Environ. 139 (2018) 199–210, https://doi.org/10.1016/j.buildenv.2018.05.003.

[35] J. Lyu, J. Li, Z. Zhao, X. Miao, H. Du, D. Lai, Y. Yang, Z. Lian, How do people set air conditioning temperature setpoint in urban domestic–Behavior model in Chinese three climate zones based on historical usage data, Energy Build. 284 (2023) 112856, https://doi.org/10.1016/j.enbuild.2023.112856.
[36] L. Cui, J. Shi, Urbanization and its environmental effects in Shanghai, China, Urban Clim 2 (2012) 1–15.
[37] M. Mughnyanti, S. Efendi, M. Zarlis, Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation, IOP Conf. Ser. Mater. Sci. Eng. 725 (2020) 012128, https://doi.org/10.1088/1757-899X/725/1/012128.
[38] S. Das, S. Subudhi, A review on different methodologies to study thermal comfort, Int. J. Environ. Sci. Technol. 19 (2022) 2155–2171, https://doi.org/10.1007/s13762-021-03210-8.
[39] Q. Fan, X. Xu, P. Liu, H. Zhang, S. Tang, A data-driven framework for thermal comfort assessment method based on user interaction, J. Build. Eng. 82 (2024) 108294, https://doi.org/10.1016/j.jobe.2023.108294.
[40] J. Yu, Q. Ouyang, Y. Zhu, H. Shen, G. Cao, W. Cui, A comparison of the thermal adaptability of people accustomed to air-conditioned environments and naturally ventilated environments, Indoor Air 22 (2012) 110–118, https://doi.org/10.1111/j.1600-0668.2011.00746.x.
[41] J.F. Nicol, M.A. Humphreys, Thermal comfort as part of a self-regulating system, Build. Res. Practice 1 (1973) 174–179, https://doi.org/10.1080/09613217308550237.
[42] R. de Dear, J. Xiong, J. Kim, B. Cao, A review of adaptive thermal comfort research since 1998, Energy Build. 214 (2020) 109893, https://doi.org/10.1016/j.enbuild.2020.109893.
[43] F. Zhang, R. de Dear, Impacts of demographic, contextual and interaction effects on thermal sensation—Evidence from a global database, Build. Environ. 162 (2019) 106286, https://doi.org/10.1016/j.buildenv.2019.106286.
[44] R.F. Rupp, J. Kim, E. Ghisi, R. de Dear, Thermal sensitivity of occupants in different building typologies: the Griffiths constant is a variable, Energy Build. 200 (2019) 11–20, https://doi.org/10.1016/j.enbuild.2019.07.048.
[45] Ministry of Housing and Urban-Rural Development of the People's Republic of China, Design Code for Heating Ventilation and Air Conditioning of Civil Buildings, China Planning Press, Beijing, China, 2012.
[46] X. Zhang, S. Nie, M. He, J. Wang, Charging system analysis, energy consumption, and carbon dioxide emissions of battery electric buses in Beijing, Case. Studies Thermal. Eng. 26 (2021) 101197, https://doi.org/10.1016/j.csite.2021.101197.
[47] X.Z. Jiang, D. Zheng, Y. Mi, Carbon footprint analysis of a combined cooling heating and power system, Energy Convers. Manag. 103 (2015) 36–42, https://doi.org/10.1016/j.enconman.2015.06.036.
[48] F. Meng, L. Zhang, G. Ren, R. Zhang, Impacts of UHI on variations in cooling loads in buildings during heatwaves: a case study of Beijing and Tianjin, China, Energy 273 (2023) 127189, https://doi.org/10.1016/j.energy.2023.127189.
[49] J. Park, J. Woo, Analyzing consumers' willingness to purchase energy-efficient appliances in response to energy price changes: case study of South Korea, Energy Econ. 127 (2023) 107088, https://doi.org/10.1016/j.eneco.2023.107088.
[50] X. Lin, C. Guo, P. Wargocki, S. Tanabe, K.W. Tham, L. Lan, The effects of temperature on work performance in the typical office environment: A meta-analysis of the current evidence, Build Environ 269 (2025) 112488, https://doi.org/10.1016/j.buildenv.2024.112488.
[51] P.O. Fanger, Thermal comfort. Analysis and applications in environmental engineering. Thermal Comfort. Analysis and Applications in Environmental Engineering, 1970.
[52] R. de Dear, G. Brager, Thermal comfort in naturally ventilated buildings: revisions to ASHRAE Standard 55, Energy Build. 34 (2002) 549–561, https://doi.org/10.1016/S0378-7788(02)00005-1.
[53] L. Yan, J. Li, M. Liu, M. Hu, Z. Xu, K. Xue, Heating behavior using household air-conditioners during the COVID-19 lockdown in Wuhan: an exploratory and comparative study, Build. Environ. 195 (2021) 107731, https://doi.org/10.1016/j.buildenv.2021.107731.
[54] Wes McKinney, pandas: a foundational Python library for data analysis and statistics, (2020). https://pandas.pydata.org (accessed April 18, 2023).
[55] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, Array programming with NumPy, Nature 585 (2020) 357–362.
[56] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Ilhan Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 contributors, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.
[57] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (2007) 90–95, https://doi.org/10.1109/MCSE.2007.55.
[58] M.L. Waskom, Seaborn: statistical data visualization, J. Open Source Softw. 6 (2021) 3021, https://doi.org/10.21105/joss.03021.