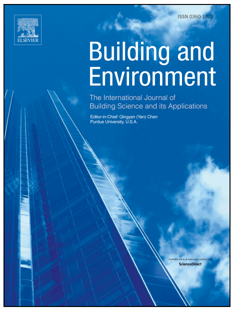

An investigation of the influencing factors for occupants' operation of windows in apartments equipped with portable air purifiers

Yan Wang, Farhang Tahmasebi, Elizabeth Cooper, Samuel Stamp, Zaid Chalabi,
Esfandiar Burman, Dejan Mumovic

PII: S0360-1323(21)00660-0

DOI: https://doi.org/10.1016/j.buildenv.2021.108260

Reference: BAE 108260

To appear in: *Building and Environment*

Received Date: 17 June 2021

Revised Date: 28 July 2021

Accepted Date: 12 August 2021

Please cite this article as: Wang Y, Tahmasebi F, Cooper E, Stamp S, Chalabi Z, Burman E, Mumovic D, An investigation of the influencing factors for occupants' operation of windows in apartments equipped with portable air purifiers, *Building and Environment* (2021), doi: https://doi.org/10.1016/j.buildenv.2021.108260.

An investigation of the influencing factors for occupants' operation of windows in apartments equipped with portable air purifiers

Yan Wang[1*], Farhang Tahmasebi[1], Elizabeth Cooper[1], Samuel Stamp[1],
Zaid Chalabi [1], Esfandiar Burman[1], Dejan Mumovic[1]

[1]UCL Institute for Environmental Design and Engineering, London, United Kingdom

*Corresponding author:
y.wang.18@ucl.ac.uk

UCL Institute for Environmental Design and Engineering

The Bartlett School for Environment, Energy and Resources

14 Upper Woburn Place

London WC1H 0NN

United Kingdom

Keywords:
Occupant behaviour, Window operation, Residential buildings, Logistic regression, Air purifier, Air quality

Abstract

As operations of windows by occupants can greatly affect building energy consumption and indoor air quality, understanding the driving factors of this adaptive control behaviour is of great importance. The present paper reports on an investigation into the influencing factors for window operation behaviour in eighteen newly-built, low-energy apartments in London, UK. A range of indoor and outdoor environmental variables (including temperature, relative humidity, CO_2 and particulate matter) as well as the window status (open/closed) were monitored for 6 - 11 months. About half of the apartments included monitoring for nearly three months during a national pandemic lockdown. Additionally, each apartment was provided with a portable home air purifier (HAP) to use during most of the study period. The effects of environmental variables and the use of HAPs on occupants' operations of windows in the main bedroom were analysed according to different periods (free-running, heating and lockdown period) and occupancy stages (arrival, departure and intermediate occupied times). Results indicated that analysing the heating period alone could lead to explanations of window operation behaviour that were contradictory to those from analysing other periods, and separating the dataset based on different occupancy stages to develop behaviour models was of little value. The results of statistical significance tests showed that indoor temperature was the leading driving factor for occupants' window opening and closing behaviour, whereas neither air quality-related variables nor the use of air purifiers had a statistically significant impact on window operation behaviour.

1. Introduction

1.1 Impacts of occupant behaviour

Occupant behaviour, including people's presence and interactions with building components such as windows and lights, is one of the most important parameters that impact building energy consumption [1]. It is also acknowledged as a contributor to the building performance gap, the discrepancy between the predicted and actual building energy use [2, 3]. In particular, as one of the most recurring adaptive actions, building users' interactions with windows has been gaining increasing research attention due to the close relationship between window operations, energy consumption, and indoor air quality (IAQ). The opening and closing of windows can exert a substantial impact on the air change rate, a crucial parameter to both thermal load and indoor pollutant level [4]. In this regard, understanding the influencing factors for occupants' window operation behaviour is of profound significance.

1.2 Modelling approaches

Various kinds of statistical modelling approaches were used in previous studies to achieve different aims, such as survival analysis (e.g., [5, 6]), Artificial Neural Network (e.g., [7]), Bayesian Network (e.g., [8]) and logistic regression models (e.g., [9] [10]). Among these analysis methods, the logistic regression model has been widely adopted to estimate the probability of building environmental control systems being in a specific state or for specific occupant actions to occur [11]. In the former case, for example, the logistic regression model can be used to predict the window state, namely a Bernoulli model to predict whether the window is open or not under different environmental conditions (with exemplar work seen in Nicol [9] and Rijal et al. [10]). However, using indoor environmental variables to predict window state may be problematic due to environmental feedback [12, 13]. That is, explanatory variables such as indoor temperature (as the model input) can change as a direct consequence of different window states (as the model output). Consequently, researchers developed a series of logistic models to predict occupants' actions of opening and closing windows (e.g., [6, 14, 15]).

1.3 Influencing factors for window operations

Occupants' window operation behaviour is a complex product of multiple factors, including environmental, contextual, psychological, physiological and social factors [16]. Historically, the correlation between window operations and a range of environmental variables has been widely studied. Temperature is one of the most examined environmental parameters, given that indoor and/or outdoor temperature was considered as a driver in almost every model (e.g., [9] [6] [17] [13]). Other environmental variables that were modelled as explanatory variables in prior researches include relative humidity (e.g., [12] [6]), CO_2 (e.g., [12] [15]), illuminance (e.g., [12]), and wind speed and direction (e.g., [6]). Occupants' interactions with windows was also found to be related to occupancy patterns. For example, occupants were observed to be more likely to open windows on arrival [14] in office buildings.

Growing concerns about the health impacts of air pollutants have led to an increased interest in studying the role ambient $PM_{2.5}$ may play in occupants' operations of windows. Outdoor $PM_{2.5}$ concentration was used as one of the variables to develop models to predict window state in a study in 2015 [18] and window opening probability in research from 2017 [19], and to compare the predictive performance of different modelling approaches in work published in 2019 [7]. Although these investigations extended the range of studied environmental variables, there is still scope for improvement not least in the data collection. In most previous studies (e.g., [18] [19]), the ambient $PM_{2.5}$ concentration data was obtained from a monitoring station several kilometres away from the monitored buildings. This, often third-party data, may not accurately reflect the actual ambient pollution level at the site, as local conditions can greatly influence pollutant concentrations [20]. Additionally, most of the studies investigating the influence of $PM_{2.5}$ on window operations were carried out in east Asia, especially China, but rarely in European countries where both ambient air quality and social norms may differ. Moreover, compared to temperature and relative humidity, $PM_{2.5}$ has remained largely unexplored in research on window operation behaviour.

1.4 Portable air purifier

Due to indoor air quality concerns, standalone home air purifiers (HAPs) utilising high-efficiency particulate air (HEPA) filters for particle filtration have been gaining increasing popularity. This technology, designed to reduce indoor $PM_{2.5}$ (as well as allergens and larger particles), is widely available for domestic use. The utilisation of HAPs was proven to contribute to significant reductions in indoor $PM_{2.5}$ concentration in dwellings in several studies [21, 22]. A primary benefit of using the air purifier in residential buildings is the potential health-related effects [23]. Additionally, air purifiers equipped with HEPA filters have been recommended as a supplementary measure to remove SARS-CoV-2 aerosols to reduce people's exposure to virus-laden aerosols [24]. However, consideration for occupants' window operations in the HAP-centric studies was not often reported, even though pollutants from outdoor air introduced through window openings may weaken the practical effectiveness of HAPs. Besides, whether there exists a conflict between window operations and the use of HAPs has not yet been explored in occupant behaviour research.

1.5 Research aim

This study set out to investigate the influencing factors for window operation behaviour in the context of UK apartment buildings, with particular interest paid to the use of portable air purifiers. Overall, this paper aims to answer the following research questions:

A. To what extent does occupants’ operation of windows vary across different periods (free-running, heating and lockdown period) and do the occupant behaviour models need to treat these periods separately?

B. Is there a significant difference between occupants’ operation of windows at different occupancy stages that justifies such a distinction in the behaviour models?

C. To what degree can thermal comfort-related variables explain the operation of windows by occupants?

D. To what degree can air quality-related variables explain the operation of windows by occupants?

E. Does the use of HAPs affect the way occupants operate windows?

2. Material and methods

2.1 Data collection

Eighteen apartments were selected from two residential building sites (Site A and B) in East London for the monitoring study. These two sites are approximately 2 km apart, both located in busy urban areas and adjacent to high-traffic roads. All apartment buildings were constructed within the last 15 years. They are representative of many modern low-energy apartments in the UK. The U-values of the building envelopes are lower than the regulatory limits in England. The Energy Performance Certificates (EPCs) [25] for these apartments from both sites are band B, with band D being the average rating for a dwelling in England and Wales (as indicated on the EPCs issued for these apartments). Eleven flats from Site A are equipped with decentralised mechanical ventilation with heat recovery (MVHR) units, while seven flats from Site B fully rely on natural ventilation through window and door openings. The size of the master bedroom in all apartments is similar, with a floor area of approximately 10 to 12 m^2 and a ceiling height of 2.7 m. More information about surveyed residences is detailed in Table 1.

Monitoring for this project was carried out from early July of 2019 to mid-June of 2020. However, due to issues such as participant availability, study fatigue, the duration of monitoring differed amongst the flats (refer to Table 1 for start and end dates for each flat). Briefly, about half of flats participated for the entire duration of the monitoring period, approximately 11 months. The other half of the participants discontinued the study in early 2020 providing data for about 6 to 8 months. Three periods are defined in this study: free-running period (1st July to 31st October 2019), heating period (1st November 2019 to 22nd March 2020) and Covid pandemic lockdown period (23rd March to 15th June 2020). The heating period was defined as such based on the fact that participants generally used the heating system between the end of October and the end of March, while the lockdown period referred to the first national pandemic lockdown in England which started on 23rd March 2020.

The on-site measurement of a range of environmental variables was conducted both indoors and outdoors, as detailed in Table 2. The indoor sensors (Eltek GD47B for the bedroom and Eltek IAQ 110 for the living room) were installed onto the internal wall about 1.60 m above the floor (to minimise disruption to occupants). The outdoor sensors (Eltek IAQ 110) were placed outside of each building site on the ground floor level in a protected location. The sampling frequency of these sensors was every 5 minutes and the equipment specifications are detailed in Table 3.

Table 1. General description of surveyed apartments

Flat index	Site	Year of construction	Floor level	Number of bedrooms	Floor area (m^2)	Number of occupants	Monitoring period
Flat 1	A	2015	ground - 3rd	4	127	5	08/07/2019 – 27/12/2019
Flat 2			9th - 13th	3	100	4	08/07/2019 – 14/06/2020
Flat 3			4th - 8th	3		4	09/07/2019 – 14/06/2020
Flat 4			9th - 13th	3		5	11/07/2019 – 14/06/2020
Flat 5				1	50	2	13/07/2019 – 14/06/2020
Flat 6				2	70	2	19/07/2019 – 08/01/2020
Flat 7			4th - 8th	3	100	4	11/07/2019 – 05/03/2020
Flat 8				3		4	12/07/2019 – 10/01/2020
Flat 9			ground - 3rd	3		4	18/07/2019 – 07/01/2020
Flat 10			4th - 8th	3	106	4	08/07/2019 – 23/12/2019
Flat 11			9th - 13th	1	50	1	12/07/2019 – 27/12/2019
Flat 12	B	2006	4th - 8th	1	49	2	09/07/2019 – 15/06/2020
Flat 13			ground - 3rd	2	65	1	15/07/2019 – 23/03/2020
Flat 14			4th - 8th	2		2	17/07/2019 – 15/06/2020
Flat 15			ground - 3rd	1	46	2	15/07/2019 – 27/12/2019
Flat 16			9th - 13th	2	59	1	15/07/2019 – 15/06/2020
Flat 17				1	46	1	16/07/2019 – 01/03/2020
Flat 18				2	59	2	22/07/2019 – 15/06/2020

Table 2. A summary of monitored environmental parameters

	Sensor	Temperature	Relative Humidity	CO_2	$PM_{2.5}$ (≤2.5 µm)	PM_{10} (≤10.0 µm)
Bedroom	Eltek GD47B	√	√	√		
Living room	Eltek IAQ 110	√	√	√	√	√
Outdoor		√	√	√	√	√

Table 3. Specifications of sensors

Sensor	Parameter	Range	Resolution	Accuracy
Eltek GD47B	Temperature	-30.0 to 65.0°C	0.1°C	± 0.2°C (at 20°C) ± 0.4°C (-5 to 40°C) ± 1.0°C (-20 to 65°C)
	Relative Humidity (RH)	0.0 to 100.0%	0.1%	± 2% RH (0 to 90% RH) ± 4% RH (0 to 100% RH)
	CO_2	0 to 5000ppm	1 ppm	± 50 ppm
Eltek IAQ 110	Temperature	-30.0 to 65.0°C	0.1°C	± 0.2°C (at 20°C) ± 0.4°C (-5 to 40°C) ± 1.0°C (-20 to 65°C)
	RH	0.0 to 100.0%	0.1%	± 2% RH (0 to 90% RH) ± 4% RH (0 to 100% RH)
	CO_2	0 to 5000ppm	1 ppm	± 50 ppm
	$PM_{2.5}$ (≤2.5 µm) PM_{10} (≤10.0 µm)	0.00 to 500.00 µg/m^3	0.01 µg/m^3	
HOBO UX90	PIR	82° detection angle, 0-10m detection range		
Eltek GS34/ Easylog EL-USB-5+	Window status	0 (closed) or 1 (open)		

The status of all operable windows and balcony doors in the bedrooms, namely open or closed, was recorded using the magnetic reed switches-based sensor Eltek GS34, except for 2 windows which used Easylog EL-USB-5+. Windows were permanently sealed in the bedrooms of Flat 8 and 9 and therefore 24 windows in total were monitored from the 16 remaining flats. The windows across the two sites were of different types; Site A had top-hung casement windows whereas Site B had tilt and turn windows. Window opening restrictors were installed on all the windows at both sites such that these windows could only be opened to a limited distance to reduce fall risks. All residents confirmed that the windows were typically left open in the maximum-opening position and the balcony doors, when open, were usually open fully.

The HAPs were placed in the main bedrooms of all surveyed apartments. All HAP operation information was automatically stored in the password-secured cloud server of the device manufacturer with participants' consent. The HAPs had several different fan speeds, but for this research, the HAP status is simplified to either 'On', irrespective of the fan speed, or 'Off'. More details of the HAP study can be found in a previous publication [26]. Note that the HAP was provided to participants from July 2019 to January 2020. For the nine flats that continued with the monitoring study beyond January 2020, there were a few months when residents did not have HAPs. The HAP status was considered as 'Off' in this case, as the focus of modelling was on the effect of using HAPs on occupants' window operation behaviour.

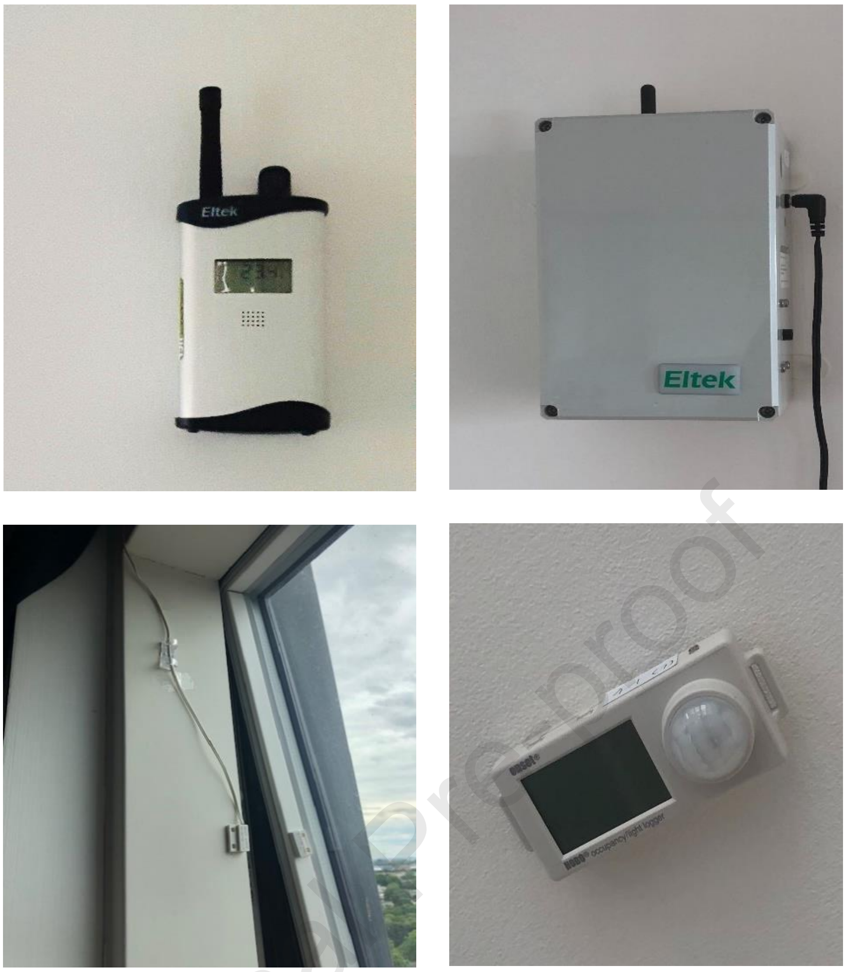

Figure 1. Photos of sensors (Eltek GD47B (top left), Eltek IAQ 110 (top right), magnetic switches of Eltek GS34 (bottom left) and HOBO UX90 (bottom right))

2.2 Occupancy detection

Passive infrared (PIR) sensors (HOBO UX90) were installed on the centre of the ceiling in both bedrooms and living rooms. There were no pets in any flats, which largely reduced the uncertainty of positive detection results from non-human occupants. In this study, PIR and CO_2 data were jointly used to determine whether any person was present at a given time, due to inherent limitations of each sensor type, such as the inability of PIR to sense immobile people, or the time delay in CO_2 concentration change to capture changes in occupancy.

The strategy was to primarily rely upon the positive values of PIR data with correction of potential false-negative PIR data by weighing the CO_2 concentration. Several past studies compared the CO_2 concentration to a predefined threshold to infer the occupancy status. Andersen et al. [12] considered the room as unoccupied when all windows were closed and the CO_2 concentration was no greater than 420 ppm; similarly, the CO_2 concentration of 460 ppm was adopted as the cut-off value to discriminate vacant times in Lai et al. [27]; Few and Elwell [28] used the CO_2 gradient over time to estimate occupancy status. Inspired by prior researches, the CO_2 concentration and its gradient with respect to time, $\frac{dCO_2}{dt}$, were both used to re-evaluate the unoccupied times detected by PIR sensors. More specifically, when any window or door was open, only the gradient of CO_2 concentration was analysed to determine if the room was occupied (postive gradient). When all windows and doors were closed, the room was determined to be occupied if the CO_2 concentration was over 480 ppm and the gradient was positive. The rationale for choosing the value of 480 ppm as the threshold was that the median outdoor CO_2 concentration for Site A and B was approximately 460 ppm, and a safety margin of 20 ppm was given. The positive CO_2 concentration gradient was used to eliminate the CO_2 decay period when the absolute value could still be in a high range.

The flat was determined to be vacant if no one was in the bedroom or the living room at a given time; otherwise, the flat was considered occupied. Then, based upon the initially estimated flat-level occupancy, if the vacancy status lasted for more than one hour, the flat was considered unoccupied for this period; otherwise, it was considered occupied. A threshold of one hour was used to reflect household activities in other rooms that were not monitored (e.g., the kitchen and bathroom). There were limitations of the method used to determine the flat-level occupancy based on monitoring in bedrooms and living rooms. Nevertheless, the proposed occupancy detection method was deemed practically acceptable for the following reasons: for one four-bedroom and seven three-bedroom flats, the children's activity was dependent on parents'; for five two-bedroom flats, no one occupied the second bedroom during the study period.

The validation of this rule-based occupancy detection method was conducted by comparing the estimated flat-level occupancy profiles against 1-week occupancy diaries completed by participants from nine of the flats. Occupants recorded the time when they left home and when they came back for seven randomly picked days including two holidays or weekend days. As shown in Table 4, the accuracies of the occupancy detection method were satisfying.

Table 4. Accuracy of the occupancy detection method compared to occupants' diaries

	Day 1	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**
Flat 2	99.3%	99.7%	100.0%	98.3%	97.9%	96.9%	98.6%
Flat 3	100.0%	100.0%	100.0%	100.0%	98.6%	99.7%	99.3%
Flat 4	97.9%	100.0%	100.0%	99.3%	99.7%	99.7%	98.3%
Flat 5	93.8%	95.1%	93.8%	95.8%	97.6%	94.1%	95.5%
Flat 12	100.0%	100.0%	71.5%	92.0%	88.9%	89.9%	95.5%
Flat 13	85.1%	100.0%	100.0%	100.0%	83.3%	80.6%	
Flat 16	93.8%	89.9%	92.0%	97.2%	82.6%	94.8%	
Flat 17	85.4%	95.8%	94.1%	92.4%	93.8%	97.6%	92.7%
Flat 18	71.5%	89.9%	87.9%	94.4%	93.4%		

2.3 Data processing

The data processing was conducted in the MATLAB (Mathworks®) environment. First, a time-series dataset in a 5-minute time step for the main bedroom of measured flats with operable windows was generated, consisting of window status, environmental variables, HAP operation status and PIR data. All windows in this study were analysed independently, without aggregating the data from several windows in the same room, to avoid potential information loss.

A rather strict data interpolation strategy was applied to deal with missing data: only one missing value of each variable between two valid readings could be linearly interpolated for continuous variables and replaced by the next value for binary variables. For two, or more, missing values between valid sensor readings, data were neither interpolated nor replaced.

Data for unoccupied times and sleep times were excluded from the statistical modelling, as window operations could only happen during people's presence and only awake people's behaviour was of interest. The unoccupied times were determined based on the occupancy schedules estimated using the method described in Section 2.2. Occupants' sleeping patterns were estimated from the semi-structured interviews conducted during the first site visit as part of the sleep quality survey using the Pittsburgh Sleep Quality Index (PSQI) [29] (further details available in [26]). Reported sleeping hours were generally consistent between midnight and 6 am. However, if any window opening or closing action occurred between midnight and 6 am, no sleep time was defined for that particular day.

Three occupancy stages for occupied time intervals were also defined: arrivals, intermediate occupied times and departures. The arrival time refers to the time step when the flat transitions from unoccupied to occupied state, while the departure time is when the flat transitions from occupied to vacant. Note that if any opening or closing actions happened within 10 mins after the arrival or before the departure, the actions were included in the arrival or departure group. The time intervals other than arrivals and departures are intermediate occupied times. Accordingly, there are two types of absence duration: prior absence duration (Abs1) for the arrivals, namely how long the flat had been left empty before anyone came back, and subsequent absence duration (Abs2) for the departures, namely how long the flat remained vacant.

2.4 Variable selection

A common method for variable selection is through a stepwise selection procedure based on various criteria, e.g., the Akaike information criterion (AIC) in Andersen et al. [12]. In the work presented here, for better control of multi-collinearity issues, a different approach driven by correlation analysis and variance inflation factor (VIF) /generalized variance inflation factor (GVIF) check was adopted.

Multicollinearity can arise when two or more explanatory variables in the regression models are highly correlated. Overlooking the multicollinearity issue could cause adverse impacts to the interpretation of the regression analysis results, e.g., sign reversals of the regression coefficients, enlarged confidence intervals and unstable p-values for the estimated coefficients of predictor variables [30, 31]. In our dataset, some pairs of variables can be naturally correlated, such as temperature and relative humidity, $PM_{2.5}$ and PM_{10}. Multicollinearity is not quite a problem for model prediction performance in which all explanatory variables count, but is problematic for identifying the impact of explanatory variables on the outcome variable [31].

Common diagnostic methods include the inspection of pairwise correlations and VIFs. There are no universally agreed cut-off values for the correlation coefficients and VIFs. Some researchers from other disciplines adopted a correlation coefficient of 0.5 as the threshold [32] and others used 0.8 [33]. In our work, the correlation coefficient above 0.5 was deemed as a high correlation which was worth careful consideration. A rule of thumb for using VIF is whether they are greater than 5 or 10 [34]. In this analysis, the GVIF was calculated using ‘the car Package’ in R language [35]. For continuous variables, VIF is the same as GVIF using the calculation method as described by Fox and Monette [36]. The pairwise correlation check was used as a screening tool for the variable selection, and subsequently, the GVIF was used to further diagnose the multicollinearity condition of multivariate models. Adjustments of predictors were then made when the GVIF was over 5 for continuous variables and over 25 for categorical variables. Given the degree of freedom (Df) for all categorical variables is 2 in our case, and requiring $(\mathrm{GVIF}^{(1/2 \cdot \mathrm{Df})})^2 < 5$ for categorical variables is equivalent to requiring VIF < 5 for continuous variables, that is why the cut-off value of GVIF was 25 for categorical variables here.

2.5 Statistical modelling

The logistic regression model, one of the most widely used techniques in occupant behaviour research (e.g., [6, 14, 15]), was chosen to model occupants’ window operation behaviour. Due to the environmental feedback issue of the window state model aforementioned, this work set to model window opening and closing actions. The probability of window operation actions was estimated for each window, with a generic mathematical expression shown in Equation (1), where P is the probability of opening or closing windows, k is the total number of explanatory variables, b_0 refers to the intercept, b_i represents the slope (regression coefficient) of each type of independent variable x_i included in the model. The p-value obtained from the t-test is used to judge the statistical significance of each independent variable at the 0.05 confidence level. That is, one predictor is determined as statistically valid if the estimated p-value for the regression coefficient is less than or equal to 0.05, equivalently translated into a statistically significant relationship between window opening or closing behaviour and this parameter. More than that, the sign of the coefficient was also used to assist in the interpretation of the physical meaning of the fitted models.

$$\log\left(\frac{P}{1-P}\right) = b_0 + \sum_{i=1}^{k} b_i \cdot x_i \qquad (1)$$

3. Results

This section starts with a preliminary analysis including an overview of environmental conditions and window operations and correlation analysis. Next are analyses of occupants' window operation behaviour in different periods and occupancy stages to answer questions A and B (in section 1.5). Lastly, the effect of each variable on occupants' window operations is evaluated to address questions C, D and E.

3.1 Preliminary analysis

3.1.1 Environmental conditions and window operations

Overall, missing data accounted for less than 10% of the monitored period for each flat. Figure 2 presents descriptive statistics of various types of environmental variables monitored in this study.

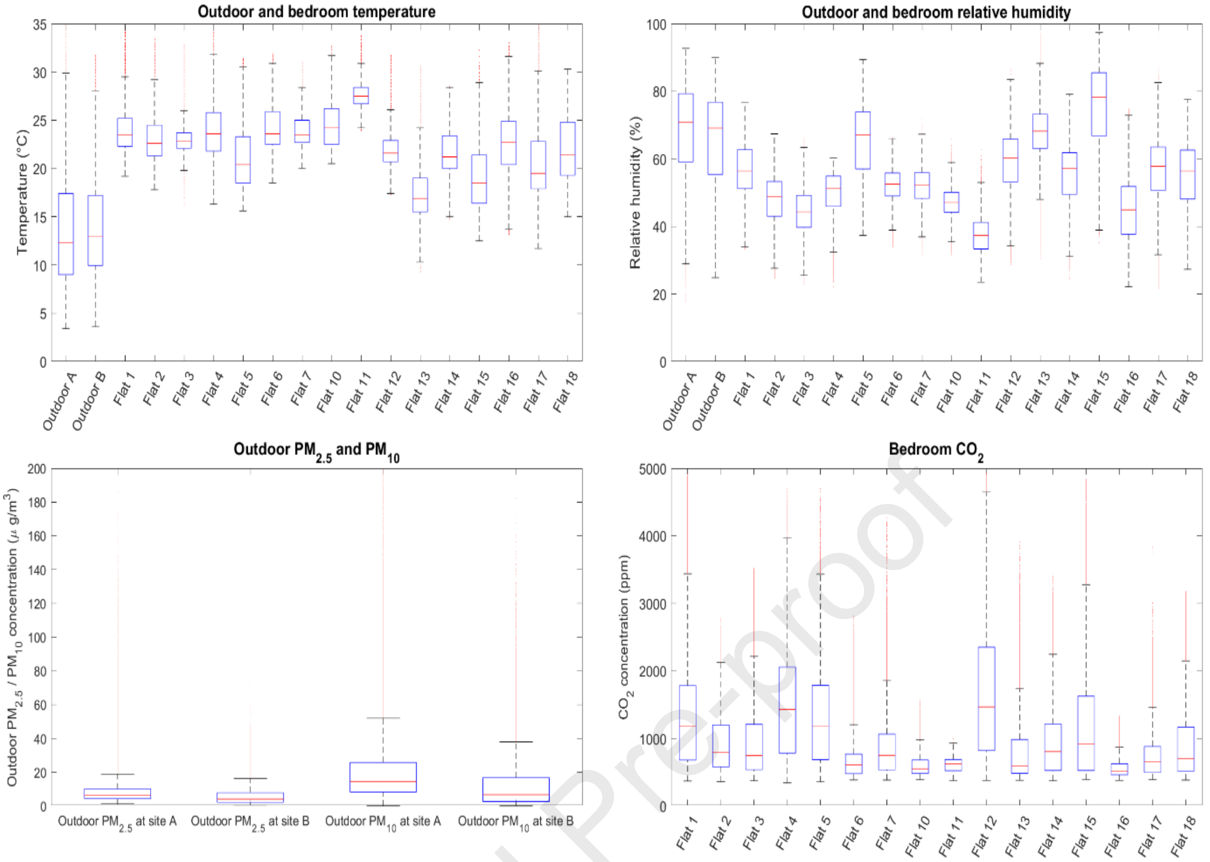

Figure 2. Descriptive statistics of outdoor and bedroom environmental variables. Outdoor A and Outdoor B on the x-axes refer to the variable at each site. The blue boxes represent the 75th and 25th percentiles, and the medians are shown as red lines.

The ventilation systems of the two building sites are different as described in section 2.1. The necessity of separately evaluating the flats based on ventilation types (e.g., natural ventilation alone) was examined early in the analysis. The apartments at Site A (Flat 1–11) with the MVHR systems were, in general, slightly warmer and less humid than naturally ventilated flats at Site B (Flat 12–18). The airflow rates were not measured in this study, but the indoor CO_2 concentrations at Site A and B were comparable, indicating that the ventilation conditions were unlikely to be fundamentally different.

To further examine the validity of separating the building sites for analysis, three characteristics of occupants' window behaviour were evaluated. The three metrics used were ‘overall percentage of time in the open state’, ‘median open state duration’ and ‘window opening rate in occupied intervals’, with results shown in Table 5. Occupants living in the flats equipped with MVHR systems left the bedroom window open, on average, about twice as long during the free-running period, a similar proportion of time during the heating period, and less than half as much time during the lockdown period, compared to occupants from naturally ventilated flats. The frequency of the window opening action was fairly close in the apartments at the two building sites. Thus, the presence of the MVHR system did not discourage occupants from opening the windows. Since there was no strong evidence to support separate modelling analyses of apartments by ventilation system, and doing so would come at the cost of reducing the sample sizes, the windows from the two building sites have been collectively analysed.

Table 5. Basic window operation metrics in different periods. ‘D’ in the annotation stands for door and ‘W’ for window with a number attached to differentiate multiple windows in the same bedroom. The means and standard deviations were calculated after removing the outliers, which were defined as more than three standard deviations from the mean.

Building site	**Window**	**Overall percentage of time in the open state**			**Median open state duration [hours]**			**Window Opening rate in occupied intervals [1/h]**		
		Free-running period	**Heating period**	**Lockdown period**	**Free-running period**	**Heating period**	**Lockdown period**	**Free-running period**	**Heating period**	**Lockdown period**
A	Flat 1 D	6.5%	0.9%	-	1.75	0.08	-	0.020	0.013	-
	W1	21.4%	0.4%	-	5.58	5.42	-	0.039	0.001	-
	W2	19.0%	0.9%	-	5.83	2.58	-	0.037	0.005	-
	Flat 2 W1	19.7%	0.0%	1.1%	1.00	0.08	5.13	0.079	0.000	0.004
	W2	9.8%	0.0%	3.5%	0.17	0.00	6.96	0.048	0.000	0.006
	Flat 3 W	63.9%	6.8%	35.2%	1.42	0.38	0.75	0.082	0.060	0.110
	Flat 4 W1	11.1%	0.7%	8.5%	7.6	0.5	1.3	0.017	0.004	0.034
	W2	42.0%	2.4%	24.6%	3.0	0.5	1.3	0.071	0.021	0.099
	Flat 5 W	14.6%	1.7%	20.3%	2.4	0.3	1.2	0.080	0.052	0.143
	Flat 6 W1	38.6%	2.1%	-	8.0	1.0	-	0.095	0.014	-
	W2	0.0%	0.0%	-	0.0	0.0	-	0.000	0.000	-
	Flat 7 W	64.0%	24.5%		1.1	11.6	-	0.043	0.029	-
	Flat 10 D	42.1%	0.2%	-	1.6	0.1	-	0.039	0.022	-
	W	93.7%	34.9%	-	67.0	6.6	-	0.013	0.068	-
	Flat 11 W	25.6%	0.1%	-	0.3	0.1	-	0.181	0.007	-
	Mean	31.5%	5.0%	15.5%	2.8	1.9	2.8	0.056	0.020	0.066
	Std.dev.	26.0%	10.3%	13.4%	2.8	3.4	2.6	0.045	0.023	0.059
B	Flat 12 W	6.1%	9.3%	23.7%	10.4	2.8	6.4	0.013	0.025	0.050
	Flat 13 D	5.2%	0.5%	-	0.7	0.2	-	0.124	0.017	-
	W	20.8%	4.5%	-	1.1	0.5	-	0.203	0.123	-
	Flat 14 D	7.7%	1.2%	16.9%	1.1	0.3	1.0	0.074	0.027	0.082
	W	3.5%	0.5%	30.5%	12.1	1.0	5.3	0.006	0.006	0.039
	Flat 15 W	3.4%	0.0%	-	13.3	0.0	-	0.005	0.000	-
	Flat 16 W	62.7%	24.6%	72.2%	11.6	9.2	12.0	0.049	0.063	0.035
	Flat 17 W	18.8%	0.0%	-	1.0	0.4	-	0.021	0.001	-
	Flat 18 W	2.9%	0.1%	57.5%	6.9	0.1	9.3	0.008	0.003	0.012
	Mean	14.6%	4.5%	40.2%	6.4	1.6	6.8	0.056	0.029	0.044
	Std.dev.	19.3%	8.1%	23.6%	5.5	3.0	4.2	0.068	0.040	0.026
Overall	Mean	25.1%	3.5%	26.7%	4.2	1.4	4.6	0.056	0.019	0.056
	Std.dev.	24.7%	7.0%	21.9%	4.3	2.4	3.8	0.053	0.022	0.046

3.1.2 Correlation analysis and model specification

The correlation analysis conducted in this study served as a screening tool to select variables for inclusion in the first round of modelling. Starting from period-based models, the mean and standard error of the correlation coefficients for the four pairs of variables across the cohort of monitored windows are shown in Figure 3. The overall period refers to the entire monitoring time for each flat.

- Indoor temperature (T_{in}) and outdoor temperature (T_{out}): They were highly correlated, with the correlation coefficient up to over 0.8 in both window opening and closing models, except in the heating period. Therefore, T_{in} was kept as an independent variable in all sets of models, while T_{out} was only included in models for the heating period. The rationale behind this choice is that it is the indoor environment that people directly sense and have control over and thus, the indoor variable should be prioritised over the outdoor one when they are highly correlated.
- T_{in} and indoor relative humidity (RH_{in}): The correlation between them was relatively low, below 0.5, so both were included in the regression model in the preliminary screening phase.
- Indoor relative humidity (RH_{in}) and outdoor relative humidity (RH_{out}): For both window opening and closing models, the correlation between RH_{in} and RH_{out} was moderate in the heating period but high in other periods. so RH_{out} was only kept in the heating-period models.
- Outdoor $PM_{2.5}$ ($PM_{2.5out}$) and outdoor PM_{10} (PM_{10out}): High correlations between them were consistently observed in all periods for both window opening and closing models. $PM_{2.5out}$ was chosen due to the known significant

health impacts of $PM_{2.5}$ [37]. As can be seen in Figure 2, the outdoor $PM_{2.5}$ concentration can be as high as 200 μg/m^3 across short periods, a much greater range than temperature and relative humidity. To make the estimated regression coefficients of the independent variables in the model more comparable, the transformation of $\log(PM_{2.5out}+1)$ was applied for the statistical modelling. A similar transformation of $\log(CO_2)$ was made for the same reason.

There are three sets of models for different occupancy stages. The correlations between variables in three occupancy phases were similar to those in the overall period model. Therefore, the explanatory variables included in the model for intermediate occupied times were the same as in the overall-period model. In the window opening models for the arrival stage, it was slightly different in that the prior absence duration (Abs1) was also modelled as a predictor in addition to variables in the model for intermediate occupied times, as the contrast between the indoor and outdoor air perceived at the point of people returning home may elicit windows openings [38]. In the window closing models for the departure stage, the subsequent absence duration (Abs2) is the only predictor, as occupants' decisions to close windows before departure may be strongly associated with issues of security.

Given the correlation analysis above, the explanatory variables for the first-round models for different periods and occupancy stages are as in Table 6. Note that for seven multi-window bedrooms, the state of the additional window(s) was used as a categorical predictor, referred to as other window status (OWS) in the models. For example, in the case of a bedroom with two windows, if the other window is open, the OWS is 1 in the models for the analysed window.

Final predictor settings were subject to refinement in the second round of modelling wherever the GVIF was over the defined threshold. Detailed model information including estimated coefficients, p-values and GVIFs is provided in Table S1-S4 in the supplementary material. The GVIFs from the first round of models were generally below the threshold. Scrutiny of the first-round models showed that the problematic GVIFs were mainly caused by the inclusion of RH_{in} or RH_{out}. All minor changes of predictors (including the removal of RH_{in} or RH_{out}) can be easily traced by cross-comparison of predictors retained in the final models in Table S1-S4 with the initial predictor settings in Table 6.

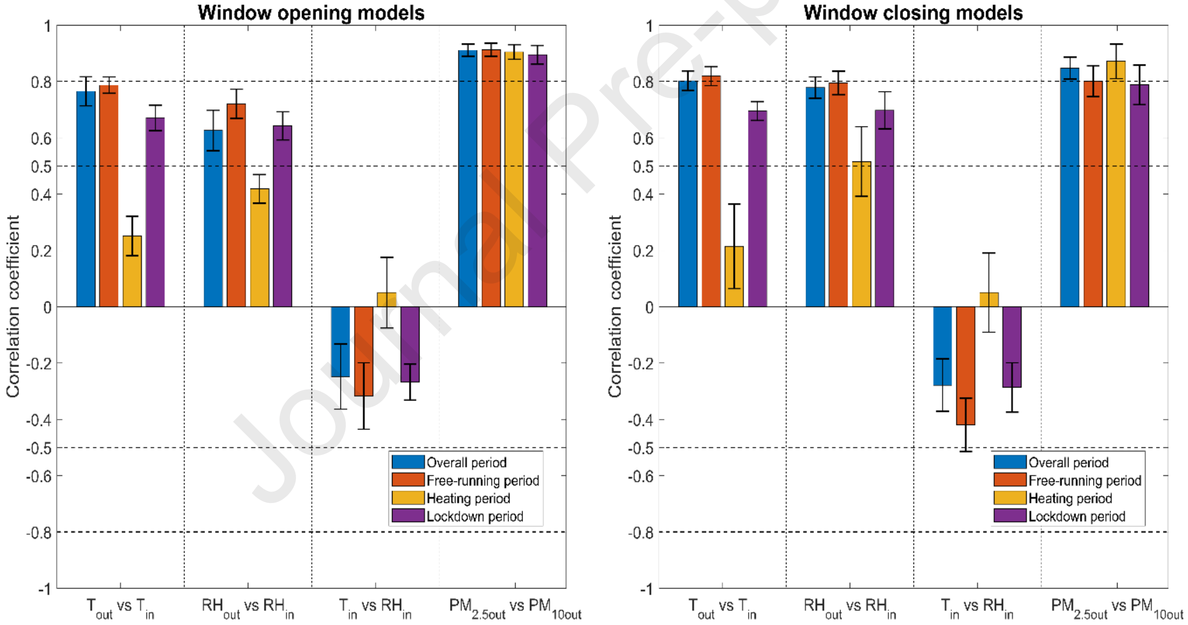

Figure 3. Pair-wise correlation analysis for period-based models

Table 6. Predictor settings for different models

Category		Model type	Explanatory variables
Period	Overall period, Free-running, Lockdown	Opening	$PM_{2.5out}$, T_{in}, RH_{in}, CO_2, HAP, OWS (OWS only applicable for multi-window bedrooms)
		Closing	
	Heating period	Opening	$PM_{2.5out}$, T_{in}, T_{out}, RH_{in}, RH_{out}, CO_2, HAP, OWS
		Closing	
Occupancy	Intermediate occupied times	Opening	$PM_{2.5out}$, T_{in}, RH_{in}, CO_2, HAP, OWS
		Closing	
	Arrival	Opening	$PM_{2.5out}$, T_{in}, RH_{in}, CO_2, HAP, OWS, Abs1
	Departure	Closing	Abs2

3.2 Question A: Variations of occupants' operation of windows in free-running, heating and lockdown period

Distinct features regarding occupants' window operations were observed in different periods. In the heating period, windows were open for a smaller amount of time, and there were fewer window opening actions, compared to in the free-running and lockdown period. On average, the windows were open for about 25% of the time in both free-running and lockdown periods, but only 3.5% in the heating period, as can be seen in Table 5. The median of window open state duration in the free-running period was very close to that in the lockdown period, and both were approximately two times higher than that in the heating period. Similarly, the frequency of opening windows at occupied times also dropped by varying degrees for most of the windows in the heating period.

Another interesting finding was that the preferential use of windows in the multi-window bedrooms generally did not change across different periods. For most occupants having control of more than one window, such as Flat 4, one window (or balcony door) was preferentially opened over the other in the same room. The exception was Flat 2, where two windows were operated equally in the heating period but one window was more frequently used in other periods. This observation backs up our strategy of modelling the operation of each window instead of aggregating the operations of multiple windows in the same room.

The results of the statistical significance tests for each variable categorised by different periods are summarised in Table *7*. The 'Valid' column shows the percentage of models where each type of parameter is determined to be statistically significant (p-value ≤ 0.05). The 'sign (+)' and the 'sign (-)' columns specify the signs of the regression coefficients for those statistically valid predictors. For example, the outdoor $PM_{2.5}$ was found statistically valid in about 40% of window opening models for the overall period, 100% of which have a coefficient for $PM_{2.5out}$ with a negative sign; the OWS achieved the defined significance level in only around 35% of overall-period models for window openings, 60% of which have a coefficient for OWS with a negative sign. A period-specific analysis is provided below.

Table 7. Summary of statistical significance tests for period-based models. The ‘Valid’ column shows the percentage of models where each type of parameter reaches the defined statistical significance level. The ‘sign (+)’ and the ‘sign (-)’ columns specify the signs of the regression coefficients for those statistically valid predictors.

		Overall period			Free-running period			Heating period			Lockdown period		
	Variable	Valid	sign (+)	sign (-)	Valid	sign (+)	sign (-)	Valid	sign (+)	sign (-)	Valid	sign (+)	sign (-)
Window opening model	T_{out}	-	-	-	-	-	-	19.0%	100.0%	0.0%	-	-	-
	RH_{out}	-	-	-	-	-	-	11.1%	0.0%	100.0%	-	-	-
	$PM_{2.5out}$	39.1%	0.0%	100.0%	26.1%	16.7%	83.3%	4.3%	0.0%	100.0%	27.3%	0.0%	100.0%
	T_{in}	87.0%	100.0%	0.0%	78.3%	94.4%	5.6%	34.8%	62.5%	37.5%	72.7%	87.5%	12.5%
	RH_{in}	30.4%	57.1%	42.9%	27.3%	100.0%	0.0%	22.7%	60.0%	40.0%	36.4%	75.0%	25.0%
	CO_2	60.9%	14.3%	85.7%	43.5%	0.0%	100.0%	36.4%	62.5%	37.5%	45.5%	20.0%	80.0%
	HAP	34.8%	75.0%	25.0%	17.4%	100.0%	0.0%	15.0%	100.0%	0.0%	-	-	-
	OWS	35.7%	40.0%	60.0%	35.7%	60.0%	40.0%	16.7%	100.0%	0.0%	33.3%	50.0%	50.0%
Window closing model	T_{out}	-	-	-	-	-	-	12.5%	100.0%	0.0%	-	-	-
	RH_{out}	-	-	-	-	-	-	12.5%	100.0%	0.0%	-	-	-
	$PM_{2.5out}$	26.1%	83.3%	16.7%	21.7%	60.0%	40.0%	17.6%	33.3%	66.7%	45.5%	100.0%	0.0%
	T_{in}	91.3%	0.0%	100.0%	78.3%	0.0%	100.0%	41.2%	14.3%	85.7%	63.6%	0.0%	100.0%
	RH_{in}	54.5%	50.0%	50.0%	9.5%	100.0%	0.0%	35.7%	40.0%	60.0%	44.4%	50.0%	50.0%
	CO_2	52.2%	83.3%	16.7%	34.8%	75.0%	25.0%	31.3%	60.0%	40.0%	27.3%	66.7%	33.3%
	HAP	26.1%	33.3%	66.7%	27.3%	50.0%	50.0%	30.0%	33.3%	66.7%	-	-	-
	OWS	35.7%	40.0%	60.0%	23.1%	66.7%	33.3%	0.0%	-	-	50.0%	33.3%	66.7%

Overall period: It can be clearly seen in Table *7* that indoor temperature is the most influential and consistent explanatory variable. Firstly, it is statistically significant in about 90% of both window opening and closing models, substantially higher than other covariates. Secondly, the sign of the regression coefficient estimates for indoor temperature remains consistent, positive and negative in the window opening and closing models, respectively. The sign of the regression coefficient matters in the interpretations of the physical meanings of the models. Taking the indoor temperature as an example, a positive sign in the window opening models means people are more likely to open the closed window at higher indoor temperatures, while the probability of closing action occurring is higher at lower indoor temperatures given the negative sign. The calculated statistical significance and the consistent sign of the coefficient doubly confirm the key role of the indoor temperature in affecting occupants' window operation behaviour.

Free-running and lockdown period: Similar to the overall-period models, the indoor temperature is the leading influencing factor, reflected in both the high percentage of models where it passes the significance test and the consistency in the sign of the regression coefficient. Although the absolute percentage drops compared to models for the overall period and the sign differs in a few window opening models, the impact of indoor temperature on occupants' operation of windows is still dominant, in contrast with other variables. Interestingly, people's behaviour seemed not to fundamentally change during the pandemic lockdown in that the key determinant for both window openings and closings was still indoor temperature despite significantly altered occupancy and working schedules.

Heating period: Unlike the above-mentioned periods, none of the modelled variables, indoor temperature included, can justifiably be considered a determinant to window operations given the results of the statistical significance tests.

It is not surprising to find that variables like $PM_{2.5out}$ or RH_{in} pass the statistical significance test in a low percentage of models for the heating period, as this is also the case for free-running, lockdown and the overall period. But for indoor temperature, there was a dramatic performance change between models for the heating period and other periods. A possible explanation is the lack of window opening and closing actions during the heating period, as reflected in the overall low window opening action rates in Table 5, making it hard to develop statistically valid logistic regression models. In one extreme case, the bedroom window in Flat 15 was never opened during the heating period, so it was impossible to develop any meaningful logistic regression models in this case. Therefore, it seems unreasonable to judge the statistical significance of each predictor based on logistic models for the heating period.

To answer research question A, occupants' operation of windows differed significantly between the heating period and the non-heating period, and analysing the heating period alone could lead to explanations of window operation behaviour that were contradictory to those from analysing other periods.

3.3 Question B: Differences in occupants' operation of windows in three occupancy stages

The three occupancy stages defined in this study were arrivals, intermediate occupied times and departures, as described in section 2.3. Table 8 shows the frequency of window opening and closing actions at each occupancy stage. The blank cell (-) indicates that the action did not happen at one particular occupancy stage for one window. Overall, most window openings and closings occurred at intermediate occupied times. However, for some windows, the action was more frequent on arrival or at departure. More specifically, the window opening action rarely occurred at departure (in only 3 windows) and happened on arrival in 10 windows. Notably, for these 10 windows, the window opening frequency was generally much higher on arrival than during intermediate occupied times. As for the window closing action, it seldomly happened on arrivals (in only 2 windows) and occurred at departure times in 7 windows. Particularly, for these 7 windows, the closing action rate at departure was generally higher than during intermediate occupied times.

Table 8. Basic window operation metrics at different occupancy stages

Window	**Window opening rate in different occupancy stages [1/h]**			**Window closing rate in different occupancy stages [1/h]**		
	Departure	**Arrival**	**Intermediate occupied times**	**Departure**	**Arrival**	**Intermediate occupied times**
Flat1 D	–	–	0.015	–	–	0.015
Flat1 W1	–	–	0.023	–	–	0.023
Flat1 W2	–	–	0.023	–	–	0.023
Flat2 W1	–	–	0.021	–	–	0.021
Flat2 W2	–	–	0.013	–	–	0.013
Flat3 W	–	–	0.073	–	–	0.073
Flat4 W1	–	–	0.015	–	–	0.015
Flat4 W2	–	0.141	0.053	0.141	–	0.053
Flat5 W	–	0.367	0.077	–	–	0.078
Flat6 W1	0.351	0.640	0.034	0.281	–	0.037
Flat6 W2	–	–	–	–	–	–
Flat7 W	1.714	–	0.018	–	–	0.019
Flat10 D	–	–	0.021	0.571	–	0.020
Flat10 W	–	–	0.020	–	–	0.020
Flat11 W	–	0.923	0.109	0.235	0.231	0.110
Flat12 W	–	0.085	0.025	–	–	0.025
Flat13 D	–	0.072	0.033	–	–	0.033
Flat13 W	–	1.157	0.087	–	–	0.091
Flat14 D	–	0.077	0.046	0.038	0.038	0.046
Flat14 W	–	–	0.015	0.076	–	0.015
Flat15 W	–	–	0.003	–	–	0.003
Flat16 W	0.261	2.039	0.031	0.470	–	0.037
Flat17 W	–	0.064	0.008	–	–	0.008
Flat18 W	–	–	0.005	–	–	0.005

The results of the significance tests for each variable in models based on three different occupancy stages are shown in Table 9. The same message from the models for intermediate occupied times and those for the overall period is that the indoor temperature is the leading driving factor, dominant in the percentage and consistent in the sign of the coefficients. However, in the regression models for arrivals and departures, the included explanatory variables barely reached the significance level (p-value ≤ 0.05), suggesting that partitioning data into different occupancy phases for modelling in our case is unnecessary. In particular, the absence durations (Abs1 and Abs2) were found not to be statistically significant, indicating that time away from the flat was not strongly associated with window operation patterns in the bedroom.

To answer research question B, little information is gained by developing sub-models for arrival and departure stages, and the model for intermediate occupied times is very similar to that for the overall period.

Given the outcome of the analysis in Section 3.2 and 3.3, the following analysis of each variable will be based on the overall-period models.

Table 9. Summary of statistical significance tests for occupancy-based models. The ‘Valid’ column shows the percentage of models where each type of parameter reaches the defined statistical significance level. The ‘sign (+)’ and the ‘sign (-)’ columns specify the signs of the regression coefficients for those statistically valid predictors.

	Departure				**Arrival**				**Intermediate occupied times**			
	Variable	Valid	sign (+)	sign (-)	Variable	Valid	sign (+)	sign (-)	Variable	Valid	sign (+)	sign (-)
Window opening models	–	–	–	–	$PM_{2.5out}$	0.0%	0.0%	0.0%	$PM_{2.5out}$	47.8%	9.1%	90.9%
	–	–	–	–	T_{in}	20.0%	100.0%	0.0%	T_{in}	87.0%	100.0%	0.0%
	–	–	–	–	RH_{in}	10.0%	0.0%	100.0%	RH_{in}	34.8%	62.5%	37.5%
	–	–	–	–	CO_2	10.0%	0.0%	100.0%	CO_2	56.5%	15.4%	84.6%
	–	–	–	–	HAP	0.0%	–	–	HAP	39.1%	77.8%	22.2%
	–	–	–	–	OWS	0.0%	–	–	OWS	50.0%	57.1%	42.9%
	–	–	–	–	Abs1	0.0%	–	–				
Window closing models	$PM_{2.5out}$	–	–	–	–	–	–	–	$PM_{2.5out}$	26.1%	83.3%	16.7%
	T_{in}	–	–	–	–	–	–	–	T_{in}	91.3%	0.0%	100.0%
	RH_{in}	–	–	–	–	–	–	–	RH_{in}	50.0%	45.5%	54.5%
	CO_2	–	–	–	–	–	–	–	CO_2	52.2%	75.0%	25.0%
	HAP	–	–	–	–	–	–	–	HAP	26.1%	33.3%	66.7%
	OWS	–	–	–	–	–	–	–	OWS	35.7%	40.0%	60.0%
	Abs2	14.3%	100.0%	0.0%	–	–	–	–				

3.4 Question C: Effects of thermal comfort-related factors on window operations

Both indoor temperature and relative humidity fall into the category of thermal comfort. Figure 4 gives an illustration of the effect of indoor temperature or relative humidity on the probability of opening and closing actions to occur, respectively. Note that only models in which one type of variable is proven as a statistically significant predictor are plotted, so the number of curves in Figure 4, Figure 5 and Figure 7 is equivalent to the percentage in the 'valid' column for the overall period in Table 7.

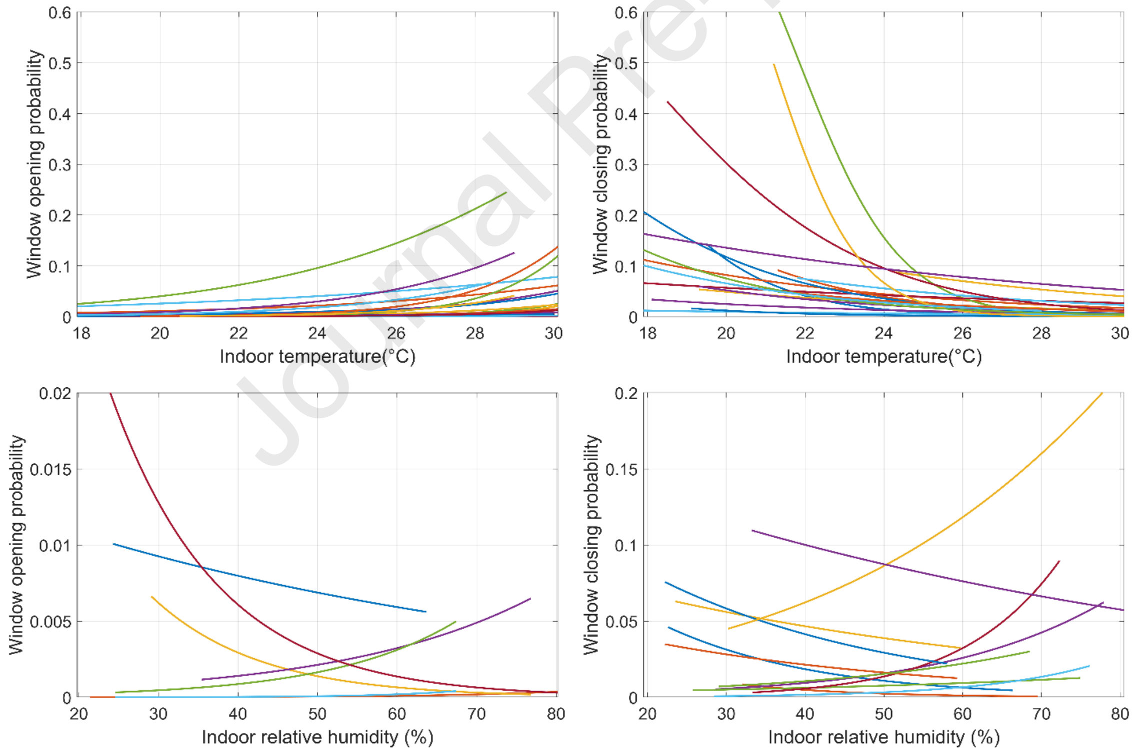

Figure 4. Plots of window opening and closing models based on T_{in} or RH_{in} for different windows. The probability is calculated using Equation (1) based on T_{in} or RH_{in} with other continuous variables fixed at their mean levels and the categorical variables fixed at 0.

The same trend shared by all models is that the probability of opening a closed window increases with increasing temperatures, while the probability of closing an open window gets greater with decreasing temperature. In comparison, there are very few curves for models in which RH_{in} was proven to be statistically significant. There also exist two opposite trends, representing people's opposing responses to the indoor relative humidity change. Some residents tend to open the window when the indoor air gets drier, while for others more humid air is more likely to trigger the window opening action. Differing patterns are also observed in window closing behaviours associated with RH_{in}.

In short, to answer research question C, in relation to the thermal comfort-related factors, indoor temperature in our study can statistically explain occupants' window opening and closing behaviour to a great extent.

3.5 Question D: Effects of air quality-related factors on window operations

Outdoor $PM_{2.5}$ and indoor CO_2 belong to the air quality group of variables. As reflected in the number of plot curves in Figure 5, neither variable reaches the statistical significance level in most models. For CO_2, only two curves show that the probability of opening windows increases in higher CO_2 concentrations, while the rest of the models exhibit the opposite trend that the window opening action is more likely to happen in lower CO_2 levels. A similar situation is observed for window closing models.

As for outdoor $PM_{2.5}$, it was demonstrated as a valid explanatory variable in only about 40% and 25% for window opening and closing models, respectively, suggesting a weak connection between people's window operation behaviour and outdoor $PM_{2.5}$. The trend, as shown in Figure 5, is that the probability of opening windows decreases with increasing outdoor $PM_{2.5}$ concentrations and that of closing windows increases with increasing outdoor $PM_{2.5}$ concentrations.

In brief, to answer the research question D, air quality-related factors in our study do not statistically explain occupants' window opening and closing behaviour.

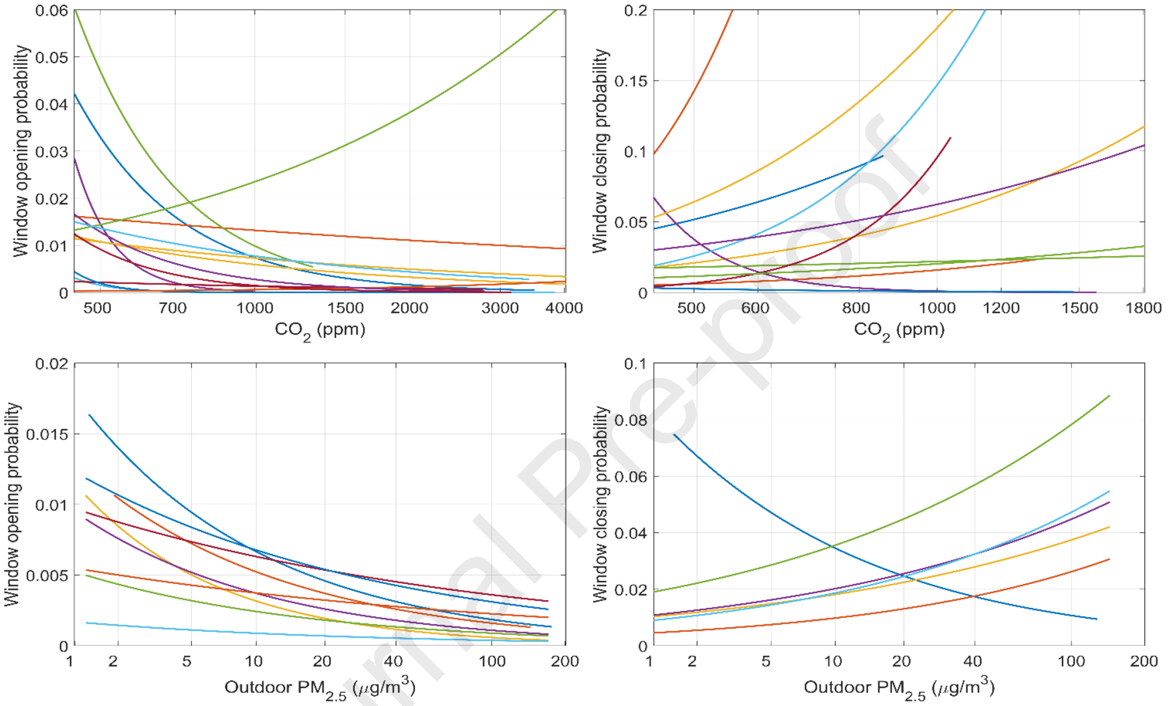

Figure 5. Plots of window opening and closing models based on $PM_{2.5out}$ or CO_2 for different windows in the studied flats. The probability is calculated using Equation (1) based on $PM_{2.5out}$ or CO_2 with other continuous variables fixed at their mean levels and the categorical variables fixed at 0.

3.6 Question E: Effects of home air purifier on window operation

It is worthwhile to gain an overview of the air purifier usage before delving into the statistical analysis. Figure 6 shows the HAP use duration on a daily basis throughout the six months that residents had the HAPs. Overall, HAPs were for a very limited amount of time by all participants. No user consistently utilised the HAP, but rather it was used intensively for short periods and then barely used at all, which can be seen in the clear cluster of peaks in Figure 6. However, all participants used the HAP for at least a short time, even if, such as the occupants in Flat 16, only for a few days, providing the opportunity to study if the use of HAPs could influence people's window operation behaviour.

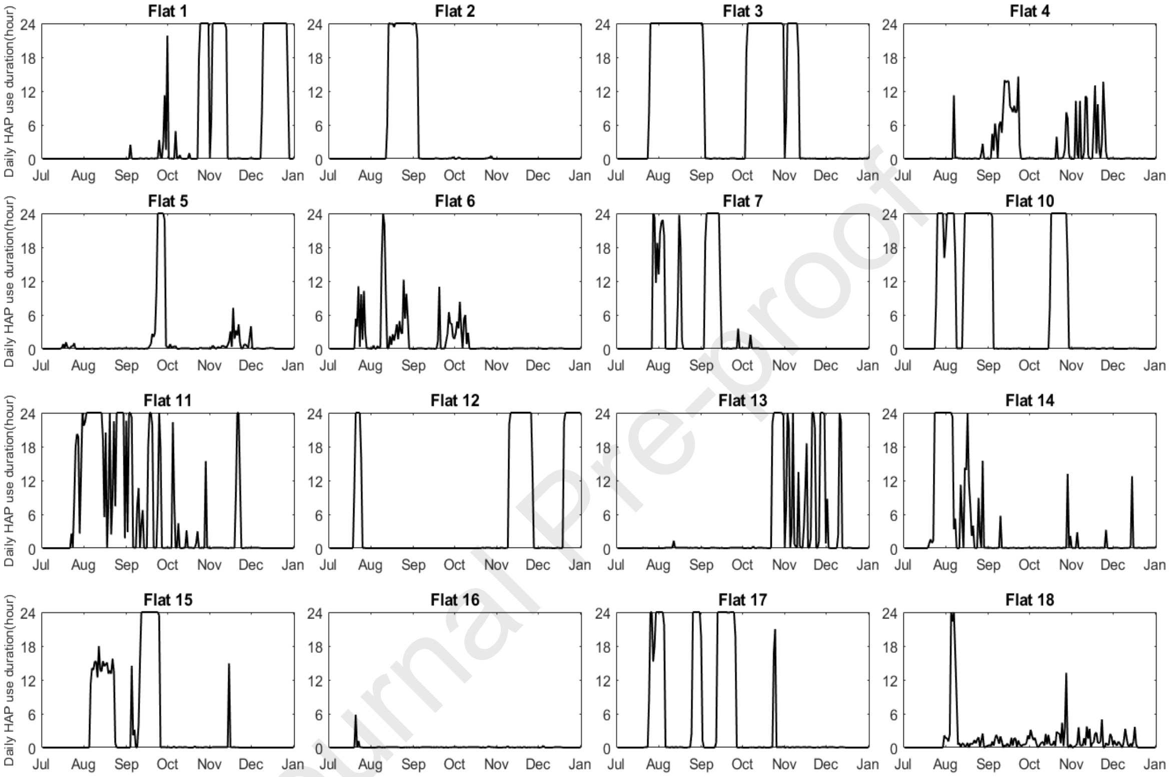

Figure 6. Daily duration of HAP use for different flats

The statistical analysis aims to explore if there exists any potential conflict between HAP use and people's window operations. For example, people may attempt to minimise the amount of outdoor air coming into their home by closing or not opening windows while the air purifier is working. However, as it stands, the HAP was found to be a statistically significant determinant in only around 30% of both window opening and closing models as shown in Table 7, indicative of a weak statistical relationship between HAP use and window operations. It is in line with surveys from a previous paper [26], where none of the participants related HAP use to window operations.

Figure 7 provides an illustration of two window opening models with a positive and negative sign of HAP in the regression model against the indoor temperature separately. A negative sign of HAP indicates that the window opening action is estimated to be less likely to happen, given the same indoor temperature when the HAP is working compared to the HAP not working, as is illustrated by Flat 1 W1. A positive sign of HAP shows the opposite, increasing the probability, as is seen in the case of Flat 4 W2.

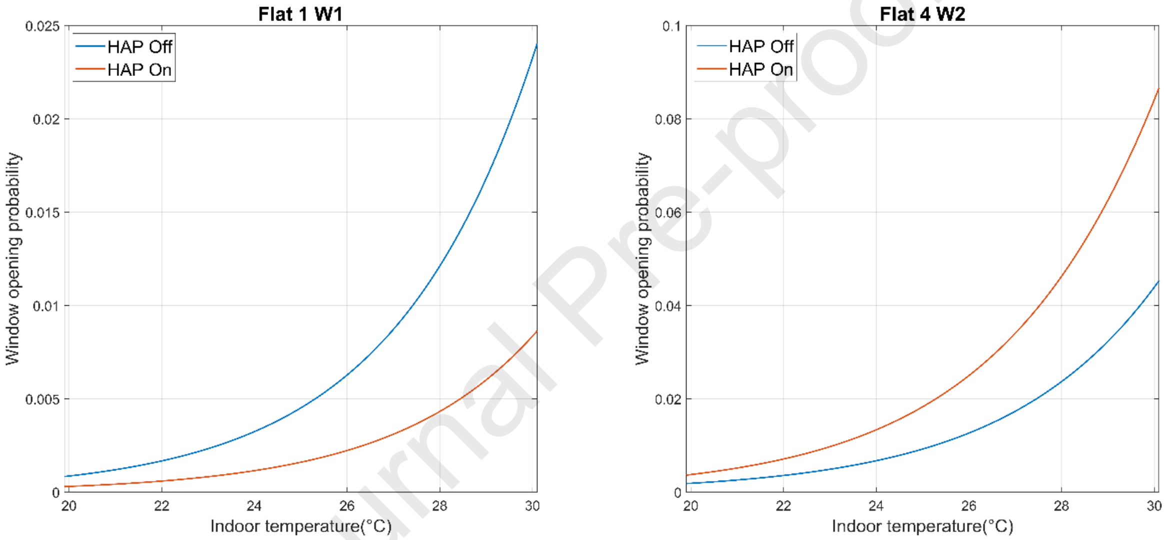

Figure 7. Plots of window opening models based on T_{in} and HAP for different windows in the studied flats. The probability is calculated using Equation (1) based on T_{in} and HAP with other continuous variables fixed at their mean levels and OWS fixed at 0.

In summary, to answer the research question E, the use of home air purifiers in our study is not significantly correlated with occupants’ window opening and closing behaviour.

4. Discussion

4.1 Different periods and occupancy stages

The operation of windows by occupants differed between the heating and non-heating period. The window openings during the heating period were found to be very scarce in this study, which is in line with many previous studies (e.g., [10, 12, 39]). One potential consequence, as found in this study, was that it was technically difficult to develop statistically valid logistic regression models for window openings and closings in the heating period. Authors further argued that using a dataset with too few actions (such as data from the heating period) for developing logistic models and conducting statistical significance tests could lead to unreliable explanations of window operation behaviour.

Window operations were found to be closely linked with occupancy stages in office buildings and many researchers stressed the importance of segmenting occupancy stages for modelling occupant behaviour [6, 14]. However, our analysis, in the context of domestic buildings, suggests that modelling bedroom window operations for flat-level arrivals and departures was of little value. A plausible explanation is that people did not often occupy the bedroom shortly after coming back home or before leaving home, causing the window openings and closings in the bedroom to be less likely to happen during the occupancy transition phases.

4.2 Thermal comfort

Indoor temperature is deemed the most important influencing factor for both window opening and closing behaviour based on our statistical analysis. This is consistent with the main findings from the literature. Indoor temperature was determined as a key driver for occupants' control of windows in previous researches (e.g., [6, 10, 39, 40]). On the other hand, relative humidity was shown to have a limited influence on occupants' window operations. This finding is consistent with those from similar prior studies, as very few reported relative humidity as a main influencing factor (e.g., [41]).

4.3 Air quality

There are potential reasons for considering CO_2 to be a potential influencing factor for window operations. Although humans are less sensitive to CO_2 compared to temperature [42], high concentrations of CO_2 can be associated with discomfort and health symptoms [43]. Besides, CO_2 can serve as an air quality indicator for its representativeness of human bio-effluent [44]. However, CO_2, as a predictor of window operating behaviour, does not reach the statistical significance level in many regression models. More importantly, our analysis does not show that CO_2 at higher concentrations motivates window opening actions or depresses window closing actions. This finding is notable and is supported by some prior studies. Jeong et al. [45] and Yao and Zhao [19] found a negative impact of indoor CO_2 on the window state and window opening action in residential buildings, respectively. Stazi et al. [46] concluded that CO_2 concentration was not a driver for window operation at school, as occupants were unaware of indoor CO_2 change. In light of our finding and similar ones from other studies, CO_2 indicators that can alert occupants of high CO_2 concentrations and prompt them to open windows may be recommended in some situations.

There is no evidence that people can directly sense $PM_{2.5}$, but outdoor $PM_{2.5}$ can be considered as a potential proxy parameter of other impacting factors, such as traffic noise and people’s perceived external air pollution level, which can effectively affect people’s window operations. Nevertheless, the relationship between people’s window operation behaviour and outdoor $PM_{2.5}$ was identified to be weak in our study. This finding was similar to that from a previous study [18] that determined outdoor $PM_{2.5}$ concentration was a less important explanatory variable than other environmental variables. This could indicate that there may be a need to implement a sensor-based alert system to notify occupants of the ambient pollution condition and recommend the best action to take.

4.4 Home air purifiers

The consideration of the use of localised air purifiers is non-trivial, given their effectiveness in reducing indoor pollutants with minimal energy costs, especially when mitigation of the pollutant source or upgrading the existing ventilation system is not feasible. Understanding the relationship between residents’ HAP use and window operations is vital to making informed assumptions in scenarios of using HAPs as an intervention measure to improve indoor air quality. Given the evidence presented in our study, windows are expected to be freely operated when the HAP is working. To the best of the authors’ knowledge, this is the first time that evidence has been presented on the relationship between HAP use and window operation. However, due to the rather limited period (six months) that occupants had access to the HAPs, this study cannot determine if the long-term use of HAP would affect occupants’ window operations.

Traditional HAP-centric studies typically do not account for actual window operations. However, pollutants of outdoor origin introduced through window opening-led natural ventilation could reduce purification effectiveness. For example, in our case, the window was open in the same room for an average of 38% of the time when HAPs were running, varying between 4% and 97% for individual apartments. Therefore, the authors argue that local outdoor pollutants, window operation behaviour, and purification effectiveness of the HAP, should be considered collectively in cases of using HAP as a supplementary or precautionary measure to reduce the indoor air contaminants.

4.5 Limitations

To the best of our knowledge, this work is one of the few research studies that used on-site PIR sensors in domestic buildings. A sensor fusion method (PIR with CO_2 sensors) was meant to compensate for the disadvantages of each sensor type. Yet, the authors acknowledge that there were uncertainties arising from apartment-level occupancy estimation based on data collected from a living room and a bedroom, especially for multi-bedroom apartments. Moreover, the proposed occupancy detection rule method was validated for only a short period by half of the participants. It is highly recommended that future in-situ monitoring cover as many rooms of the dwelling as possible to gain a holistic picture of the complex occupancy patterns of residences. Additionally, a longer period of ground-truth occupancy information allowing for a long-term validation would be strongly desired.

As aforementioned, the measured variables of the air quality variables were intended to be a proxy of the actual drivers for occupants’ window operations, such as outdoor traffic noise, indoor odour, air stuffiness and discomfort experienced by people, most of which are subjective measures and technically difficult to quantify. It would be worthwhile to investigate if other types of pollutant variables such as indoor $PM_{2.5}$ or total volatile organic compounds (TVOCs) can offer a better approximation of occupants' perception of the air quality and its effects on window operations. Future work is suggested to extend the scope of environmental variables beyond currently well-studied temperature and relative humidity.

5. Conclusion

Statistical analysis was performed to identify the key determinants of window operations observed in eighteen low-energy modern apartments in the UK. Results suggested that analysing the heating period alone could lead to explanations of window operation behaviour that were contradictory to those from analysing other periods, and separating the dataset based on different occupancy stages to develop window opening and closing models was of little value. Moreover, statistical significance tests for different types of variables and the signs of their coefficients in the logistic regression models for the overall period were reported. The indoor temperature was identified as the most consistent and influential variable in explaining occupants' window opening and closing behaviour. Air quality-related variables (outdoor $PM_{2.5}$ and indoor CO_2) and the use of air purifiers, in contrast, had limited statistically significant impacts on occupants' window operation behaviour.

This research also has implications for future research. That thermal comfort, rather than air quality, was the leading influencing factor for occupants' window operation behaviour suggests a need for an alert system to inform occupants of the indoor and outdoor air quality conditions and trigger the optimal action to reduce indoor air pollutants. Additionally, the authors believe that future research into the deployment of local air purifiers should consider the concurrent operation of windows by occupants, as the present study found a weak statistical relationship between HAP use and window operations by occupants.

Acknowledgements

The research presented was funded by the project “Quality of Indoor Air on Sites Matched with Outdoor Air Quality Datasets to Improve Wellbeing Outcomes” (activity number 19144), EPSRC IAA (Project number 559487) and CIBSE Support for Research. In addition, the study largely benefited from the authors’ participation in IEA-EBC Annex 79.

Appendix A. Supplementary material

Supplementary material to this article can be found online.

Reference

1. Yoshino, H., T. Hong, and N. Nord, *IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods.* Energy and Buildings, 2017. **152**: p. 124-136.
2. Branco, G., et al., *Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data.* Energy and Buildings, 2004. **36**(6): p. 543-555.
3. Majcen, D., L.C.M. Itard, and H. Visscher, *Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications.* Energy Policy, 2013. **54**: p. 125-136.
4. Wallace, L.A., S.J. Emmerich, and C. Howard-Reed, *Continuous measurements of air change rates in an occupied house for 1 year: the effect of temperature, wind, fans, and windows.* J Expo Anal Environ Epidemiol, 2002. **12**(4): p. 296-306.
5. Chen, W., et al., *Research on occupants’ window opening behavior in residential buildings based on the survival model.* Sustainable Cities and Society, 2020. **60**.
6. Haldi, F. and D. Robinson, *Interactions with window openings by office occupants.* Building and Environment, 2009. **44**(12): p. 2378-2395.
7. Wei, Y., et al., *Comparison of different window behavior modeling approaches during transition season in Beijing, China.* Building and Environment, 2019. **157**: p. 1-15.
8. Barthelmes, V.M., et al., *Exploration of the Bayesian Network framework for modelling window control behaviour.* Building and Environment, 2017. **126**: p. 318-330.
9. Nicol, J.F. *Characterising occupant behaviour in buildings: towards a stochastic model of occupant use of windows, lights, blinds, heaters and fans*. in *Proceedings of the seventh international IBPSA conference, Rio*. 2001.

10. Rijal, H.B., et al., *Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings.* Energy and Buildings, 2007. **39**(7): p. 823-836.
11. Gaetani, I., P.-J. Hoes, and J.L.M. Hensen, *Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy.* Energy and Buildings, 2016. **121**: p. 188-204.
12. Andersen, R., et al., *Window opening behaviour modelled from measurements in Danish dwellings.* Building and Environment, 2013. **69**: p. 101-113.
13. Tahmasebi, F. and A. Mahdavi, *On the utility of occupants' behavioural diversity information for building performance simulation: An exploratory case study.* Energy and Buildings, 2018. **176**: p. 380-389.
14. Yun, G.Y. and K. Steemers, *Time-dependent occupant behaviour models of window control in summer.* Building and Environment, 2008. **43**(9): p. 1471-1482.
15. Calì, D., et al., *Analysis of occupants' behavior related to the use of windows in German households.* Building and Environment, 2016. **103**: p. 54-69.
16. Fabi, V., et al., *Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models.* Building and Environment, 2012. **58**: p. 188-198.
17. Rijal, H.B., et al., *Development of adaptive algorithms for the operation of windows, fans, and doors to predict thermal comfort and energy use in Pakistani buildings.* 2008. **114**(2): p. 555-573.
18. Shi, S. and B. Zhao, *Occupants' interactions with windows in 8 residential apartments in Beijing and Nanjing, China.* Building Simulation, 2015. **9**(2): p. 221-231.
19. Yao, M. and B. Zhao, *Window opening behavior of occupants in residential buildings in Beijing.* Building and Environment, 2017. **124**: p. 441-449.
20. Pinto, J.P., A.S. Lefohn, and D.S. Shadwick, *Spatial variability of PM2.5 in urban areas in the United States.* J Air Waste Manag Assoc, 2004. **54**(4): p. 440-9.
21. Cheng, K.-C., et al., *Mixing and sink effects of air purifiers on indoor PM2.5 concentrations: A pilot study of eight residential homes in Fresno, California.* Aerosol Science and Technology, 2016. **50**(8): p. 835-845.
22. Zhan, Y., et al., *The influence of air cleaners on indoor particulate matter components and oxidative potential in residential households in Beijing.* Sci Total Environ, 2018. **626**: p. 507-518.
23. Fisk, W.J., *Health benefits of particle filtration.* Indoor Air, 2013. **23**(5): p. 357-68.
24. Zhao, B.B., Y. Liu, and C. Chen, *Air purifiers: A supplementary measure to remove airborne SARS-CoV-2.* Build Environ, 2020: p. 106918.
25. Arcipowska, A., et al., *Energy performance certificates across the EU–a mapping of national approaches.* BPIE, 2014.
26. Cooper, E., et al., *Use of portable air purifiers in homes: Operating behaviour, effect on indoor PM2.5 and perceived indoor air quality.* Building and Environment, 2021. **191**.
27. Lai, D., et al., *Window-opening behavior in Chinese residential buildings across different climate zones.* Building and Environment, 2018. **142**: p. 234-243.

28. Few, J. and C. Elwell, *Measuring the ventilation rate in occupied buildings and adapting the CO2 tracer gas technique.*
29. Buysse, D.J., et al., *The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.* 1989. **28**(2): p. 193-213.
30. Mela, C.F. and P.K.J.A.E. Kopalle, *The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations.* 2002. **34**(6): p. 667-677.
31. Vatcheva, K.P., et al., *Multicollinearity in regression analyses conducted in epidemiologic studies.* 2016. **6**(2).
32. Donath, C., et al., *Predictors of binge drinking in adolescents: ultimate and distal factors-a representative study.* 2012. **12**(1): p. 1-15.
33. Berry, W.D., S. Feldman, and D. Stanley Feldman, *Multiple regression in practice*. 1985: Sage.
34. Neter, J., et al., *Applied linear statistical models.* 1996.
35. Fox, J. and S. Weisberg, *An R Companion to Applied Regression, Third edition.* Sage, Thousand Oaks CA. , 2019.
36. Fox, J. and G. Monette, *Generalized collinearity diagnostics.* Journal of the American Statistical Association, 1992. **87**(417): p. 178-183.
37. WHO, *Ambient air pollution: Pollutants.* 2019.
38. Roetzel, A., et al., *A review of occupant control on natural ventilation.* Renewable and Sustainable Energy Reviews, 2010. **14**(3): p. 1001-1013.
39. Herkel, S., U. Knapp, and J. Pfafferott, *Towards a model of user behaviour regarding the manual control of windows in office buildings.* Building and Environment, 2008. **43**(4): p. 588-600.
40. D'Oca, S. and T. Hong, *A data-mining approach to discover patterns of window opening and closing behavior in offices.* Building and Environment, 2014. **82**: p. 726-739.
41. Zhang, Y. and P. Barrett, *Factors influencing the occupants' window opening behaviour in a naturally ventilated office building.* Building and Environment, 2012. **50**: p. 125-134.
42. Zhang, X., et al., *Human responses to carbon dioxide, a follow-up study at recommended exposure limits in non-industrial environments.* 2016. **100**: p. 162-171.
43. Allen, J.G., et al., *Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office environments.* 2016. **124**(6): p. 805-812.
44. CEN, C., European Committee for Standardization, *Ventilation for buildings: Design criteria for the indoor environment.* 1998.
45. Jeong, B., J.-W. Jeong, and J.S. Park, *Occupant behavior regarding the manual control of windows in residential buildings.* Energy and Buildings, 2016. **127**: p. 206-216.
46. Stazi, F., F. Naspi, and M. D'Orazio, *Modelling window status in school classrooms. Results from a case study in Italy.* Building and Environment, 2017. **111**: p. 24-32.

Highlights:

- There was a weak relationship between air purifier use and occupants' window operations.
- Thermal comfort rather than air quality was the driving factor of occupants' window operations.
- Using the heating-season data alone to develop window operation models may be unreliable.
- Little information was gained by developing sub-models for arrival and departure stages.

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: