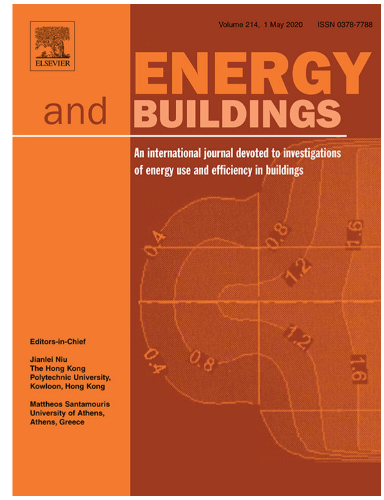
Journal Pre-proofs



A multilevel window state model based on outdoor environmental conditions that captures behavioural variation at room and apartment levels

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

PII: S0378-7788(22)00733-2
DOI: https://doi.org/10.1016/j.enbuild.2022.112562
Reference: ENB 112562

To appear in: *Energy & Buildings*

Received Date: 11 August 2022
Revised Date: 5 October 2022
Accepted Date: 6 October 2022

Please cite this article as: Y. Wang, E. Cooper, F. Tahmasebi, Z. Chalabi, S. Stamp, E. Burman, D. Mumovic, A multilevel window state model based on outdoor environmental conditions that captures behavioural variation at room and apartment levels, *Energy & Buildings* (2022), doi: https://doi.org/10.1016/j.enbuild.2022.112562

A multilevel window state model based on outdoor environmental conditions that captures behavioural variation at room and apartment levels

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

[a] 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

Abstract

Occupants' use of windows can influence the building energy demand, thermal conditions and indoor air quality. Researchers have made substantial efforts to develop probabilistic models to predict the window open/closed state. However, the hierarchical data structure and the heterogeneity in occupant behaviour have been generally neglected in previous modelling efforts. Multilevel modelling can provide an appropriate framework to handle this type of data structure and variability, but this method has rarely been used in the field. This study investigated *room*- and *apartment*-level variations in the effects of outdoor environmental variables on the window open state in low-energy apartment buildings in the UK using a multilevel modelling approach. The results showed that the *room*-level, rather than *apartment*-level, variation was statistically significant. Meanwhile, the room type (i.e., living room or bedroom) did not significantly affect the relationship between outdoor environmental variables and the window open state. The strength of this study is that the modelling accounted for the hierarchical structure of the data by simultaneously considering room-and apartment- level behavioural variations. By quantifying the significant diversity of occupant behaviour in the natural ventilation of residences, future research can more accurately estimate the variation in building energy and indoor air quality impacts.

Keywords: window open state, behavioural diversity, multilevel modelling, residential buildings, environmental factors

1. Introduction

The pursuit of reducing carbon emissions and energy consumption drives the need for improving building energy efficiency. Various kinds of building efficiency measures, such as highly insulated windows and energy-efficient building energy systems, have been implemented to improve building energy performance [1, 2]. In addition to technological solutions, human factors should not be ignored in the global effort toward a decarbonised society [3]. It has been well acknowledged that occupants can exert a substantial impact on building energy performance [4-7] [8], and their behaviour can have an even larger impact in low-energy buildings [9].

Within the domain of occupant behaviour research, building occupants' use of windows has been a popular research topic in recent decades. Opening windows is a simple but important way to improve ventilation for occupants in residential buildings. The window state (i.e., open or closed) can strongly affect the air change rate in buildings [10], which can, in turn, influence the building energy demand [11], occupants' thermal comfort and indoor air quality (IAQ) [12]. Therefore, proper control of window openings could achieve a good balance between energy savings and comfortable and healthy building environments [13].

Researchers have made substantial research efforts to develop probabilistic models based on field monitoring data to predict the probability of either the window open state ([14]) or the probability of window opening and closing action to occur ([15]). The core concept of the probabilistic model in this context is that people's adaptive behaviour should be considered stochastic rather than deterministic [16]. One of the most common methods used to develop probabilistic models for predicting the window state is logistic regression, with examples seen in previous studies by Haldi and Robinson [17] and Andersen et al. [18].

Previous studies found that the window open/closed state could be affected by environmental factors [17], time-related factors such as time of day and occupancy stages [19], as well as psychological and social factors [12]. Historically, the effects of environmental variables on the window state have been widely studied. Indoor and outdoor temperatures were frequently reported as the key factors impacting the window state in the literature [20, 21]. Additionally, there were other types of environmental variables that were identified as influencing factors, such as indoor CO_2 concentration [22], outdoor relative humidity [23] [24], outdoor wind speed [23] [24], and outdoor $PM_{2.5}$ concentration [23].

Although a large number of probabilistic models have been developed for predicting either the window open state or window opening and closing actions, most models ignored the hierarchical structure inherent in the data and occupants' behavioural diversity. An example of such a hierarchical structure can be that individual rooms where the monitoring is conducted are nested within apartments, and apartments are nested within buildings; as such, occupant behaviour in the rooms in the same apartments/ buildings may share more common traits than rooms from different apartments/ buildings. It is not a trivial issue, as reliable information about inter-occupant variation regarding occupant behaviour that affects building performance would be helpful for predictions of the extremes of building energy demand and evaluation of the robustness of the building design [25, 26]. Many studies aggregated all data collected from different rooms to create a meta-model to predict the average behaviour of the sample ([17] [14] [18]). However, it could be problematic to assume that the statistically typical occupant behaviour is representative of all individuals, as using this average behaviour could cause a substantial difference between predicted and actual building energy use [27]. In contrast, other studies chose to model every single occupant ([28] [29]), but a full picture of the heterogeneity in occupant behaviour could not be captured. Additionally, a few studies classified occupants into several categories to characterise different behaviour types. For instance, window operation patterns were clustered as 'active operation', 'neutral operation' and 'passive operation' by D'Oca and Hong [30]. However, such discrete classification may be very specific to the analysed dataset and may not be generalisable to other settings. For example, the active window user in one place might be miscategorised as the average user in another place with more active window operations.

Multilevel modelling can be a powerful method to handle hierarchical data structure and occupants' behavioural diversity. This method divides the variance of the outcome variables into *between-group/ level* variance (namely variance between different groups/ levels) and *within-group/ level* variance (namely variance between individual units within the same group/ level) [31]. There are two components in the multilevel model, *fixed* effect and *random* effect. The *fixed* effect part represents the average effect of the independent variables on the dependent variables at a population level, while the *random* effect part allows such effects to vary within a group/ level. Compared to conventional modelling methods which aggregate findings based on single-level group-specific means, a primary advantage of the multilevel modelling method lies in its ability to accurately represent the variability in the data across hierarchical structures, which could result in more reliable statistical inference including *p*-values and confidence intervals [32].

The multilevel modelling approach has been applied in some branches of building research. For example, Li et al. [33] used the multilevel model to study the effect of urban form on electricity consumption in residential buildings in China; Prignon et al. [34] developed multilevel models to quantify the uncertainties in airtightness measurements in apartment buildings in Belgium; Belaïd et al. [35] used this modelling method to analyse the geographic and household effects on residential energy demand in France; Kent et al. [36] conducted multilevel modelling analysis to evaluate the effect of the time of day on people's glare sensation in the UK. However, to the best of the authors' knowledge, only two previous studies developed multilevel models to predict window operation or window open state based on environmental variables.

In 2016, Haldi et al. [37] proposed to use the multilevel model to quantify the effects of behavioural diversity and applied it to datasets from long-term monitoring campaigns in an office building in Switzerland and residences in Denmark and Germany. Multilevel logistic regression models were developed for occupants' window opening actions based on separate datasets. In the case of the Danish dwellings, multilevel models were developed based on a number of environmental variables with random effects at the household level. As for the German dwellings, sets of multilevel models were developed for different room typologies (e.g., bathroom, kitchen, living room and bedroom). The authors recommended adopting this method to express behavioural diversity as a systematic description of occupant behaviour patterns. Nevertheless, the room and household levels were not simultaneously taken into account following the hierarchical order (e.g., rooms nested within households) in their modelling, and the effects of room type differences on occupant behaviour remained unknown. In 2020, Shi et al. [23] used multilevel logistic regression models to analyse the effects of household features on window open state in Chinese apartments. It was found that the household features significantly affected the relationship between the window open state and outdoor environmental variables. However, their models only included random intercepts, which means that the slope associated with each environmental variable was assumed to be uniform across different occupants. This assumption is very likely to be oversimplified, given what has been known about how diverse individual behaviour could be [25].

To contribute to the research on multilevel models for predicting the window open state, the research described here set to answer the following questions in the context of recently-built low-energy apartment buildings in the UK:

A. Is there *room*-level variation in the effects of outdoor environmental variables on the window open state that should be accounted for?

B. Is there *apartment*-level variation in the effects of outdoor environmental variables on the window open state that should be accounted for?

C. Does the *room type* make a difference in the relationship between outdoor environmental variables and the window open state?

Regarding the multilevel model structure in this study, the fixed effect component refers to the average effects of outdoor environmental variables on the window open state; the random effect component allows such a relationship to vary, at the *room* level for question A, and both *room* and *apartment* levels for question B. Simply put, question A investigates whether the *room*-level random effect is statistically significant, while question B examines whether the *apartment*-level random effect is statistically significant. Different from the other two questions, Question C explores the effect of the *room type*, namely taking all living rooms as one group versus all bedrooms as the other group, on the relationship between outdoor environmental variables and the window open state.

The novelty of this paper comes from both the statistical modelling method applied and the research questions addressed. There were a limited number of multilevel models for occupant behaviour identified in the literature, including any on window open state. This study fills this research gap by developing two-level and three-level window state models following the hierarchical structure of rooms and apartments, which are novel developments compared to the previously published multilevel occupant behaviour models. Furthermore, this research analysed *room*- and *apartment*-level variations, along with the potential behavioural differences associated with room types, from the perspective of the relationship between studied outdoor environmental variables and the window open state in the residences. To the best of the authors' knowledge, these research questions have not previously been addressed.

2. Methodology

2.1 Data collection

This study used data collected from a recent field measurement project [28] [38] which was carried out between July 2019 and June 2020. The indoor monitoring was conducted in both living rooms and bedrooms of 18 apartments from 2 apartment buildings (referred to as apartment buildings A and B) in London, UK. These two apartment buildings were about 2 km apart. 11 apartments from apartment building A had decentralised mechanical ventilation with heat recovery (MVHR) systems, while 7 apartments from apartment building B were not equipped with mechanical ventilation systems. All apartments had exhaust fans in bathrooms and kitchens. More detailed information about the surveyed apartments and residents can be found in a previous paper [39]. In brief, our dataset covered a wide range of apartment samples, which is favourable for the analysis of behavioural diversity in different apartment settings. The number of regular occupants varied between 1 and 5 (mostly 2 or 4); the minimum apartment floor area was 46 m^2 (1-bedroom) and the maximum was 127 m^2 (4-bedroom), with most around 60 m^2 (2-bedroom) or 100 m^2 (3-bedroom). Although both windows (1.1 m (height) by 0.9 m (width)) and balcony doors (1.9 m (height) by 0.9 m (width)) existed in some rooms, they were all considered, and referred to, as windows in this study. About half of the monitored bedrooms had 1 window and the other half had more than 1 window (mostly 2), likewise for the living room windows; a similar number of windows faced each direction (southeast, southwest, northeast and northwest).

The indoor monitoring was conducted in the living room and master bedroom of each apartment. The individual rooms within the apartments, such as Bedroom 1 and Living room 1 in Figure 1, are the room level, the finest analysis unit in this study. Above the room level, is the apartment level, for example, Bedroom 1 and Living room 1 were from, and nested within, Apartment 1. In each monitored room, the window state was recorded for all operable windows and doors by magnetic contact sensors (Eltek GS34), and passive infrared (PIR) sensors (HOBO UX90) were placed in the centre of the room ceiling to collect occupancy-related information. This work focused on outdoor environmental variables, but more information about indoor environmental monitoring (such as indoor temperature, relative humidity, and CO_2 concentration) can be found in previous publications [28] [38]. A range of outdoor environmental variables was measured. The thermal (e.g., outdoor temperature, outdoor relative humidity) and air quality variables (e.g., outdoor $PM_{2.5}$) were measured by the integrated environmental sensor unit (Eltek AQ110, https://www.eltekdataloggers.co.uk/pdf/user_instructions/TU1082_AQ110_from_serial_no_31544.pdf) at the ground level of both apartment buildings. Ground level measurement of $PM_{2.5}$ concentration was used, irrespective of the actual apartment level modelled, based on the relatively small differences between ground floor $PM_{2.5}$ concentrations and those at the 16th floor (about 65 m) reported in the literature [40]. The Alphasense OPC-N2 $PM_{2.5}$ sensor used in this study has been in other past field studies of indoor and outdoor particulate matter [41-43], and evaluations showed that it had good agreement with reference instruments, within the limitations of small and low-cost sensors [44, 45]. The wind speed was recorded by an anemometer (Davis 6410, https://shop.weatherstations.co.uk/davis-anemometer-6410-157-p.asp) at an open roof of a primary school about 100m away from apartment building A. The specification of the outdoor environmental sensors is detailed in Table 1, with photos of the equipment shown in Figure 2. All measurement data were logged every 5 minutes and stored in a cloud server.

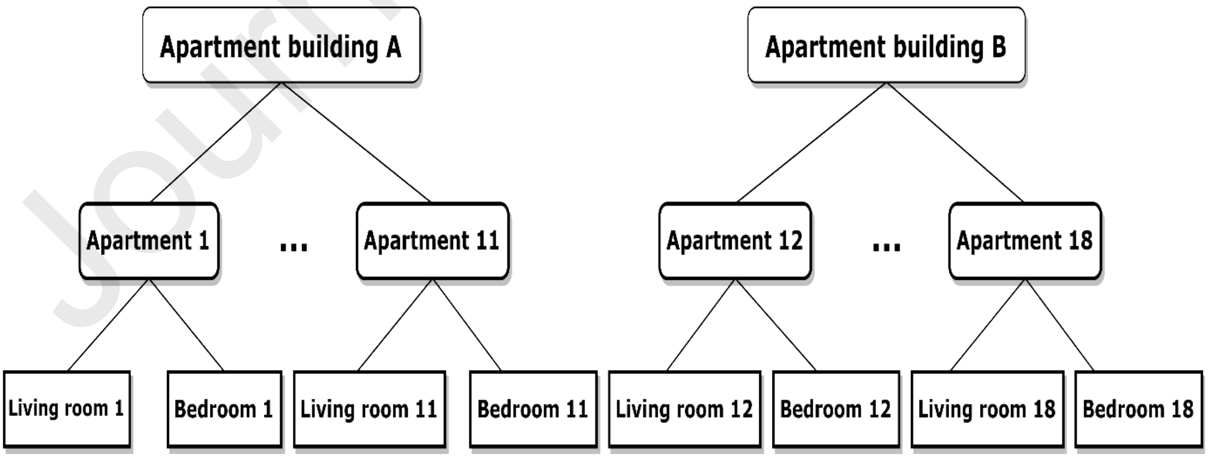


Figure 1. Schematic diagram of the hierarchical data structure.

Table 1. Specification for the outdoor environmental sensors

Equipment	Parameter	Range	Resolution	Accuracy
Eltek IAQ 110	Temperature	-30.0 - 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	0.0 - 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)	0.00 - 500.00 μg/m^3	0.01 μg/m^3	Detection range for particulates: 0.38–17 μm
Davis 6410	Wind speed	0 to 89 m/s	0.1 m/s	± 5%

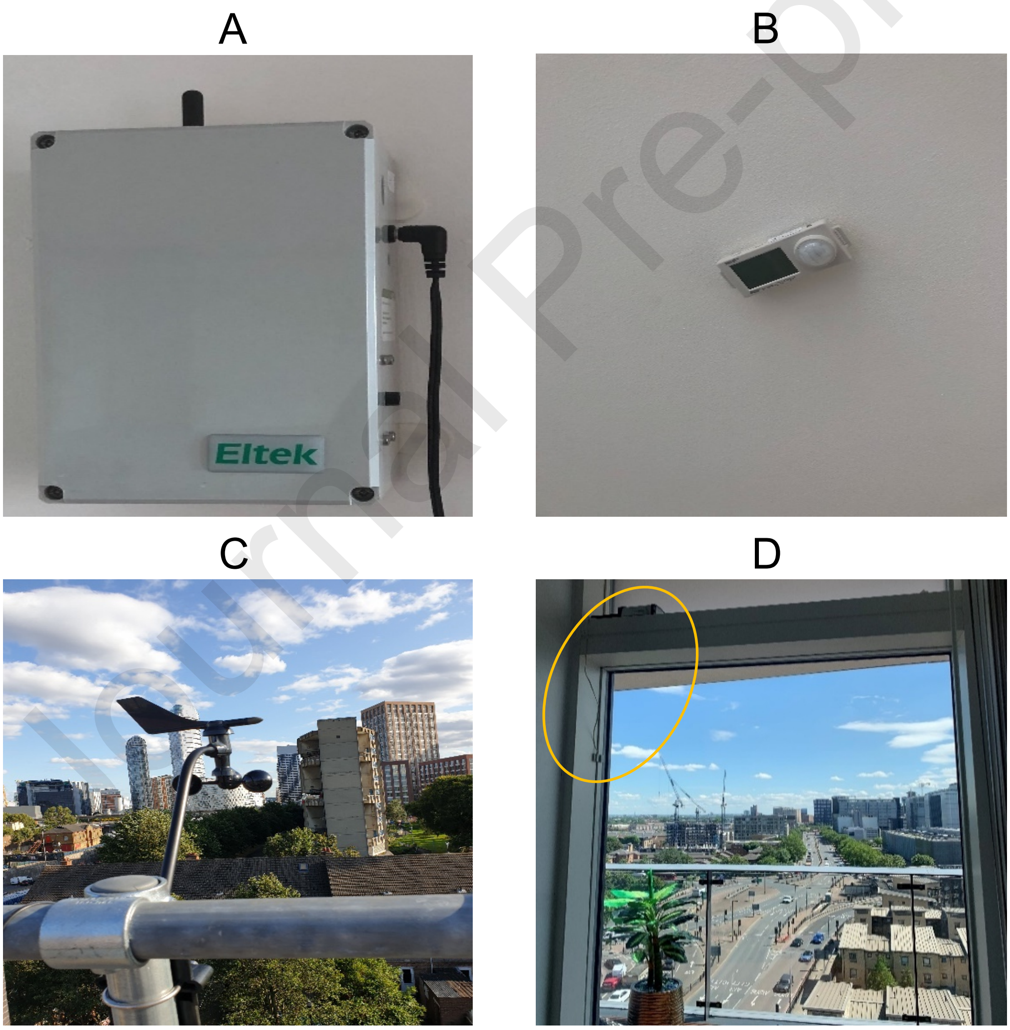


Figure 2. Photos of equipment. A: Eltek IAQ 110; B: HOBO UX90; C: Davis 6410; D: Eltek GS34.

2.2 Data preparation

Data on four outdoor environmental variables, namely outdoor temperature, outdoor relative humidity, outdoor $PM_{2.5}$ concentration and outdoor wind speed, were used to construct multilevel models in this study. The reasons behind this choice were two-fold. First, these four environmental variables were previously found to be significant factors influencing the window open state in residential buildings [23]. Second, using the indoor environmental variables to predict the window state can cause 'environmental feedback' [46]. That is, the indoor environmental conditions (independent variables) can be directly affected by the window state (dependent variable). However, using the outdoor environmental variables as predictors can eliminate this bias. Note that the whole monitoring project lasted continuously for about one year, but the wind data from the on-site weather station was only available for the period between September 2019 and early January 2020. Therefore, the modelled period in this work was set to be from 1st September 2019 to 31st December 2019, including both free-running and heating periods.

Bedroom windows were sealed in two apartments, and the monitoring did not cover the living rooms in another two apartments. Therefore, in total, 32 rooms out of 18 apartments were used for this multilevel modelling work. If any window was open at a given time in a room, the window state is 1; otherwise, 0. Different types of measured outdoor environmental variables were at varying scales, as seen in Table 2. To make the regression coefficients for different explanatory variables comparable, the outdoor environmental variables were standardised (i.e., centred around their means and scaled by their standard deviations), ahead of the statistical modelling, with descriptive statistics for all variables provided in Table 2. Data for the unoccupied time intervals were removed before the statistical modelling since the window state could change only during occupants' presence. The occupancy schedules were estimated using a customised occupancy detection method which used the indoor CO_2 concentration data to partially correct possible false negative values of the PIR data. More details of the occupancy determination method and the validation results are available in a previous publication [39].

Table 2. The statistical description of outdoor environmental variables

Environmental variables	Mean	Standard deviation	Minimum	Maximum
Outdoor temperature (℃)	12.3	4.2	3.4	24.4
Outdoor relative humidity (%)	74.9	10.5	31.4	92.3
Outdoor $PM_{2.5}$ concentration (μg/m^3)	12.32	16.54	1.22	186.22
Outdoor wind speed (m/s)	1.1	1.0	0.0	22.4

2.3 Multilevel logistic regression models

The multilevel model development considered 3 levels that fit the data structure:

- **Level 1** (low level): Fixed effects of outdoor environmental variables on the probability of the window being in the open state.
- **Level 2** (intermediate level): Random effects due to the room-level variation in the effects of outdoor environmental variables on the probability of the window being in the open state. The status of windows in the bedroom and living room of the same apartment were accounted for separately because the occupants in bedrooms and living rooms at any one point in time could be different. Additionally, room features (such as floor area) have been reported to have modifying effects on the relationship between the probability of the window being in the open state and environmental variables [23].

- **Level 3** (high level): Random effects due to apartment-level variation in the effects of outdoor environmental variables on the probability of the window being in the open state.

The apartment building level was not considered in this work, because there were only two apartment buildings in our dataset. The detailed steps for developing multilevel models are described below and were performed in MATLAB 2021b (Mathworks®).

Step 1: Develop a logistic regression model (M1) that predicts the probability of the window being in the open state based on outdoor environmental variables. The regression expression for M1 is shown in Equation (1). All available outdoor environmental variables in our dataset (i.e., outdoor temperature, outdoor relative humidity, outdoor $PM_{2.5}$ concentration and outdoor wind speed) were considered potential predictor candidates and fit into the model, as all of them were previously reported to be correlated with the window open state [23]. The *p*-value was used to judge the statistical significance of each variable at the confidence level of 0.05. The variance inflation factors (VIFs) were used to evaluate the multicollinearity in the model.

$$logit(P_j) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + e_j \qquad (1)$$

where $i = 1 \ldots k$ $(k = 4)$ denotes the index for the environmental variable, 1 for outdoor temperature, 2 for outdoor relative humidity, 3 for outdoor $PM_{2.5}$ concentration and 4 for outdoor wind speed; x_i $(i = 1 \ldots k)$ denotes the value of the environmental variable; β_0 is the estimated intercept; β_i is the estimated slope for the environmental variable x_i; P_j is the probability of the window being in the open state where $j (j = 1 \ldots 32)$ refers to the room index; e_j is the residual.

Step 2: Develop a 2-level logistic regression model (M2) which considers random effects from *room*-level behavioural variation, with the regression expression given in Equation (2).

$$logit(P_j) = (\beta_0 + u_{0j}) + \sum_{i=1}^{k}(\beta_i + u_{ij})\,x_i + e_j \quad (2)$$

where u_{0j} and u_{ij} are the random effects associated with the behavioural differences at the room level; u_{0j} is the room-level deviation from the population mean of the intercept; u_{ij} is the room-level deviation from the population mean of the slope for the environmental variable; $u_{0j}, u_{1j}, \ldots, u_{kj}$ are assumed to be distributed as a multivariate normal distribution

$N_{k+1}(0, \Sigma)$ with mean zero and the corresponding variance-covariance matrix Σ given in Equation (3).

$$\Sigma = \begin{bmatrix} \delta_{u0j}^2 & & & & \\ \rho_{0,1}\delta_{uoj}\delta_{u1j} & \delta_{u1j}^2 & & & \\ \rho_{0,2}\delta_{uoj}\delta_{u2j} & \rho_{1,2}\delta_{u1j}\delta_{u2j} & \delta_{u2j}^2 & & \\ \rho_{0,3}\delta_{uoj}\delta_{u3j} & \rho_{1,3}\delta_{u1j}\delta_{u3j} & \rho_{2,3}\delta_{u2j}\delta_{u3j} & \delta_{u3j}^2 & \\ \rho_{0,4}\delta_{uoj}\delta_{u4j} & \rho_{1,4}\delta_{u1j}\delta_{u4j} & \rho_{2,4}\delta_{u2j}\delta_{u4j} & \rho_{3,4}\delta_{u3j}\delta_{u4j} & \delta_{u4j}^2 \end{bmatrix} \quad (3)$$

The variance-covariance matrix Σ is symmetric, and only the main diagonal and the lower triangle are presented in Equation (3). δ_{u0j}^2 ($j = 1 \ldots 32$) is the variance of the random intercept and δ_{uij}^2 ($i = 1 \ldots 4$, $j = 1 \ldots 32$) is the variance of the random slope for the relevant environmental variable. Each term in the covariance matrix is given by the product of the correlation coefficient and two standard deviations. ρ (with appropriate subscripts) is the correlation coefficient between different model components, with subscript 0 corresponding to the intercept and 1- 4 corresponding to the index for the environmental variable. For example, $\rho_{0,1}\delta_{uoj}\delta_{u1j}$ is the covariance between the random intercept and the random slope for outdoor temperature; $\rho_{0,1}$ is the correlation coefficient between the variation of the random intercept and the variation of the random slope for outdoor temperature; δ_{uoj} is the standard deviation of the random intercept; δ_{u1j} is the standard deviation of the random slope for outdoor temperature.

Step 3: Compare the goodness-of-fit between models M2 and M1 using the likelihood ratio (LR) test [47] to examine if adding the room-level random effect is meaningful. It is noteworthy that the *p*-value or standard error for each variance-covariance component was not used to judge the statistical significance of the random effect in this work. That was because the interpretation of the standard errors of the coefficients for the random effects can be problematic since variance cannot be negative [48].

Step 4: Develop a 3-level logistic regression model (M3) which considers random effects from apartment-level behavioural variations, with the regression expression in Equation (4).

$$logit(P_{jg}) = (\beta_0 + u_{0jg} + u_{0g}) + \sum_{i=1}^{k}(\beta_i + u_{ijg} + u_{ig})\, x_i + e_{jg} \quad (4)$$

where g $(g = 1...18)$ is the apartment index; u_{0g} and u_{ig} are random effects related to the occupants' behavioural differences at the apartment level; u_{0g} is the apartment-level deviation from the population mean of the intercept; u_{ig} is the apartment-level deviation from the population mean of the slope for the environmental variable; u_{0g}, u_{1g}, …, u_{kg} are assumed to be distributed as a multivariate normal distribution N_{k+1} $(0, \Sigma)$ with mean zero and the corresponding variance-covariance matrix with the same structure as Equation (3). u_{0jg} and u_{ijg} are random effects related to the occupants' behavioural differences at the room level but nested with the apartment;

u_{0jg} is the room-level deviation from the apartment-level mean of the intercept; u_{ijg} is the room-level deviation from the apartment-level mean of the slope of the environmental variable; u_{0jg}, u_{1jg}, …, u_{kjg} are assumed to be distributed as a multivariate normal distribution $N_{k+1}\ (0, \Sigma)$ with mean zero and the corresponding variance-covariance matrix similar to Equation (3).

Step 5: Compare the goodness-of-fit between models M3 and M2 using the LR test to examine if adding the apartment-level random effect is meaningful.

Step 6: Develop a multilevel logistic regression model (M4) which considers the room type difference in occupants' behaviour, namely bedroom versus living room. The regression expression is given in Equation (5), where the room type difference was added to the regression equation for M2.

$$logit(P_j) = (\beta_0 + u_{0j} + \lambda_0 Z_j) + \sum_{i=1}^{k} (\beta_i + u_{ij} + \lambda_i Z_j)\, x_i + e_j \quad (5)$$

where Z_j is a binary variable (i.e., bedroom or living room), λ_0 and λ_i represent the change on the intercept and the slope, respectively, when the room type changes from the reference type (i.e., bedroom) to the living room. It should be noted that the room type difference could be added to either the regression equation for M2 or that for M3 depending upon whether modelling the apartment-level random effects was meaningful (i.e. the result of Step 5). For ease of presentation, only models to be discussed below in the result sections are described here.

Step 7: Comparing the goodness-of-fit between models M4 and M2 to examine if adding the room-type difference is meaningful based on the LR test.

2.4 Model projection

The best model was identified after conducting the modelling process described above in sections 2.1-2.3. Then, for illustration, the projections of the best model were simulated by randomly drawing coefficients for the intercept and the slopes for environmental variables 20 times from the obtained multilevel model to generate 20 different occupant behaviour models to predict the probability of window being in the open state.

3. Results

3.1 M1: Environmental variables

This study focused on answering the research questions by reporting and analysing the results of the multilevel models, but more information on window open status can be found in the supplementary file. The detailed model information for M1 which used four outdoor environmental variables (i.e., outdoor temperature, outdoor relative humidity, outdoor $PM_{2.5}$ concentration and wind speed) to predict the probability of the window being in the open state is shown in Table 3. More detailed model results can be found in the supplementary material. Note that all the outdoor environmental variables were centred around their means and scaled by their standard deviations prior to fitting the model as described above. The VIFs for all explanatory variables were calculated to be below 5, indicating that the multicollinearity in this model was not an issue of concern. All outdoor environmental variables were found to be statistically significant in M1 (p-value < 0.05). In general, the window was more likely to be open at higher outdoor temperatures given the positive sign of the regression coefficient for outdoor temperature. In contrast, the other three variables (outdoor relative humidity, outdoor $PM_{2.5}$ concentration and wind speed) were negatively correlated with the probability of the window being in the open state. Based on the magnitudes of the estimated regression coefficients, the outdoor temperature had the largest impact on the probability of the window being in the open state, while outdoor $PM_{2.5}$ concentration and wind speed had minimal effects.

Table 3. Results of the logistic regression model M1.

	β_0	β_1	β_2	β_3	β_4
Estimate	-1.5507*** [a]	0.8813***	-0.2222***	-0.0528***	-0.0320***
VIF		1.2	1.2	1.1	1.0
Goodness-of-fit					
AIC: 408375		Deviance: 408367			

a: Significance levels: *** for p < 0.001, ** for 0.001 < p < 0.01, * 0.01 < p < 0.05, NS: not significant.

3.2 M2: Room-level variations as random effects

The details for the model M2, a 2-level model that considers random effects from the room-level behavioural variation, are provided in Table 4. For the fixed effect part, all outdoor environmental variables remained statistically significant (p-value < 0.05). Noticeably, compared to M1, the absolute value of the regression coefficient for each explanatory variable in M2 increased to some extent, but the sign stayed the same. For the random effect part, the standard deviations of the intercept and slopes were considerable relative to the absolute values of the intercept and slopes in the fixed effect, for example, 0.3351 (random effect) versus -0.3388 (fixed effect), regarding the slope for outdoor relative humidity. This suggests that the inter-occupant differences in the effects of environmental variables on the window open state were considerable. On the other hand, the correlations between random effects associated with slopes for different environmental variables were mostly below 0.5. This implies that the room-level behavioural variations related to the environmental variables were not strongly correlated with each other. As can be seen in Table 5, the *p*-value for the LR test of M2 versus M1 was less than 0.05, indicating that the multilevel model M2 fit the data significantly better than the single-level model M1. In other words, accounting for the room-level random effect was necessary.

As a consequence, the answer to the first research question is that there is room-level variation in the effects of outdoor environmental variables on the window open state that should be accounted for.

Table 4. Results of the multilevel model M2.

Fixed effects					
	β_0	β_1	β_2	β_3	β_4
Estimate	-2.5410***[a]	1.1103***	-0.3388***	-0.3276**	-0.1019*
VIF		1.2	1.1	1.3	1.2
Random effects					
	δ_{uoj}[b]	δ_{u1j}	δ_{u2j}	δ_{u3j}	δ_{u4j}
Estimate	2.1681	0.5061	0.3351	0.5608	0.2507
	$\rho_{0,1}$[c]	$\rho_{0,2}$	$\rho_{0,3}$	$\rho_{0,4}$	$\rho_{1,2}$
Estimate	0.0322	0.3499	0.5940	0.3324	0.2921
	$\rho_{1,3}$	$\rho_{1,4}$	$\rho_{2,3}$	$\rho_{2,4}$	$\rho_{3,4}$
Estimate	-0.2323	0.0144	0.1215	0.1768	0.4786
Goodness-of-fit					
AIC: 294197		Deviance: 294157			

a: Significance levels: *** for p < 0.001, ** for 0.001 < p < 0.01, * 0.01 < p < 0.05, NS: not significant.
b: δ_{uoj} $(j = 1\ldots32)$ is the standard deviation of the random intercept, and δ_{uij} $(i = 1\ldots4,\ j = 1\ldots32)$ is the standard deviation of the random slope for the environmental variable, as denoted in Equation (3).
c: ρ is the correlation coefficient between the deviations of different parts of random effects, as given in Equation (3), with the subscript 0 corresponding to the intercept and 1- 4 corresponding to the index for the environmental variable.

Table 5. Results of LR tests.

Model	LR test		
Model 1		ΔDf[a]	Chi-squared
Model 2	Model 1 vs Model 2	15	114208***[b]
Model 3	Model 2 vs Model 3	15	16~~15.6~~^{NS}
Model 4	Model 2 vs Model 4	5	9~~8.5~~^{NS}

a: ΔDf: Difference in degrees of freedom

b: Significance levels: *** for $p < 0.001$, ** for $0.001 < p < 0.01$, * $0.01 < p < 0.05$, NS: not significant.

3.3 M3: Apartment-level variations as random effects

The results for model M3, a 3-level model that considered random effects at both apartment and room levels, are shown in Table 6. For the fixed effect, again, all outdoor environmental variables were statistically significant (p-value < 0.05), and compared to M2, the coefficient for each explanatory variable in M3 was rather similar. There are two parts of random effects in M3: between-apartment variation and within-apartment-between-room variation (referred to as ‘Apartment’ and ‘Room: Apartment’, respectively, in Table 6). At both levels, the standard deviations for both the intercept and the slopes for each variable were considerable in relation to the absolute values of the fixed effect. However, as shown in Table 5, the p-value for the LR test of comparing M3 with M2 was greater than 0.05, meaning that M3 was not a better fit to the data than M2. That is, adding apartment-level random effects was not meaningful.

Therefore, to answer research question B, there is no significant apartment-level variation in the effects of outdoor environmental variables on the window open state that should be accounted for.

Table 6. Results of the multilevel model M3.

Fixed effects						
		β_0	β_1	β_2	β_3	β_4
	Estimate	-2.7513***[a]	1.1052***	-0.3379***	-0.3553**	-0.1085*
	VIF		1.2	1.2	1.3	1.3
Random effects						
Room: Apartment		δ_{uojg}[b]	δ_{u1jg}	δ_{u2jg}	δ_{u3j}	δ_{u4j}
	Estimate	1.2408	0.4504	0.2541	0.4966	0.2401
		$\rho^R_{0,1}$[c]	$\rho^R_{0,2}$	$\rho^R_{0,3}$	$\rho^R_{0,4}$	$\rho^R_{1,2}$
	Estimate	-0.2661	0.4606	0.6509	0.2346	0.0972
		$\rho^R_{1,3}$	$\rho^R_{1,4}$	$\rho^R_{2,3}$	$\rho^R_{2,4}$	$\rho^R_{3,4}$
	Estimate	-0.2419	-0.0834	0.4094	0.0566	0.5247
Apartment		δ_{uog}[d]	δ_{u1g}	δ_{u2g}	δ_{u3g}	δ_{u4g}
	Estimate	2.0628	0.2313	0.2159	0.2862	0.0811
		$\rho^A_{0,1}$	$\rho^A_{0,2}$	$\rho^A_{0,3}$	$\rho^A_{0,4}$	$\rho^A_{1,2}$
	Estimate	0.4802	0.2741	0.7621	0.8881	0.7503
		$\rho^A_{1,3}$	$\rho^A_{1,4}$	$\rho^A_{2,3}$	$\rho^A_{2,4}$	$\rho^A_{3,4}$
	Estimate	-0.0727	0.6460	-0.4127	0.6719	0.3932
Goodness-of-fit						
AIC: 294212			Deviance: 294142			

a: Significance levels: *** for p < 0.001, ** for 0.001 < p < 0.01, * 0.01 < p < 0.05, NS: not significant.

b: $\delta_{uojg}(j = 1\ldots32,\ g = 1\ldots18)$ is the standard deviation of the random intercept at the room level but nested with the apartment, and δ_{uijg} $(i = 1\ldots4,\ j = 1\ldots32,\ g = 1\ldots18)$ is the standard deviation of the random slope for the environmental variable at the room level but nested with the apartment.

c: ρ is the correlation coefficient between the deviations of different parts of the random effects; the subscript 0 corresponds to the intercept and 1- 4 to the index for the environmental variable; the superscript 'R' refers to the apartment-room level and 'A' to the apartment level.

d. : δ_{uog} $(g = 1\ldots18)$ is the standard deviation of the random intercept at the apartment level, and δ_{uig} $(i = 1\ldots4, g = 1\ldots18)$ is the standard deviation of the random slope for the environmental variable at the apartment level.

3.4 M4: Differences between room types

Model M4 was developed by adding the room type difference to model M2, since the previous results showed that M3 was not significantly better than M2. The details for fitting model M4 are given in Table 7. In the fixed effect part, all outdoor environmental variables were statistically significant (p-value < 0.05), but the room type (λ_0) and the interaction terms between the binary categorical variable and the continuous environmental variables (e.g., λ_1) were all statistically insignificant. These results suggested that the room type had no statistically significant effect on the relationship between outdoor environmental variables and the window open state. The results of the LR test point to the same finding. As shown in Table 5, the p-value in the LR test for comparing M4 with M2 was greater than 0.05, meaning that model M4 was not a better fit for the data than model M2.

As a result, the answer to research question C is that the relationship between studied environmental variables and the window open state in the living room is not statistically significantly different from that in the bedroom.

Table 7. Results of the multilevel model M4.

			Fixed effects		
	β_0	β_1	β_2	β_3	β_4
Estimates	-2.5381***[a]	1.1111***	-0.3378***	-0.3206**	-0.1034*
VIF		2.3	2.1	2.2	2.3
	λ_0	λ_1	λ_2	λ_3	λ_4
Estimates	0.2925[NS]	-0.0041[NS]	-0.0552[NS]	-0.0766[NS]	0.0479[NS]
VIF	1.4	2.3	2.3	2.5	2.3
			Random effects		
	δ_{uoj}[b]	δ_{u1j}	δ_{u2j}	δ_{u3j}	δ_{u4j}
Estimates	2.1424	0.5054	0.3295	0.5464	0.2487
	$\rho_{0,1}$[c]	$\rho_{0,2}$	$\rho_{0,3}$	$\rho_{0,4}$	$\rho_{1,2}$
Estimates	0.0278	0.3725	0.6082	0.3231	0.2905
	$\rho_{1,3}$	$\rho_{1,4}$	$\rho_{2,3}$	$\rho_{2,4}$	$\rho_{3,4}$
Estimates	-0.2538	0.0199	0.0847	0.2186	0.4991
			Goodness-of-fit		
	AIC: 294199		Deviance: 294149		

a: Significance levels: *** for p < 0.001, ** for 0.001 < p < 0.01, * 0.01 < p < 0.05, NS: not significant.

b: δ_{uoj} $(j = 1 \ldots 32)$ is the standard deviation of the random intercept, and δ_{uij} $(i = 1 \ldots 4,\ j = 1 \ldots 32)$ is the standard deviation of the random slope for the environmental variable.

c: ρ is the correlation coefficient between the deviations of different parts of random effects, with the subscript 0 corresponding to the intercept and 1- 4 corresponding to the index for the environmental variable.

3.5 Model projections

Given the results presented in previous sections, the multilevel model M2 which included both fixed effects of environmental variables and room-level random effects was identified to be the best model. To simulate the projections of model M2, as described in section 2.4, coefficients for the intercept and the slopes for environmental variables were randomly drawn 20 times from the obtained multivariate normal distribution for M2 (as per Table 4) to generate different occupant behaviour models. It is worth noting that the fixed effect part of the multilevel model represents the estimated population mean of the slope for the respective environmental variable and the intercept, whereas the random number drawn from the random effect part of the multilevel model is the deviation from the population mean. The latter is analogous to the variation in occupant behaviour relative to the statistically typical behaviour. Then, for each model, the probability that the window is in the open state was calculated and plotted against each environmental variable separately (outdoor temperature, outdoor relative humidity, outdoor $PM_{2.5}$ concentration and wind speed), as shown in Figures 3-6. Note that when plotting the multivariate model against one variable, other variables were fixed at their standardised means; the standardised outdoor environmental variables in these figures correspond to the measured outdoor conditions during the modelling period, as described in section 2.2.

In relation to the outdoor temperature, the inter-occupant behaviour diversity is reflected in how the probability that the window is in the open state is displayed as a function of the regression slope (i.e., the coefficient for standardised T_{out}) as shown in Figure 3. Noticeably, the general trends of all curves in this figure are consistent, i.e., the higher temperature, the greater the probability of the window being in the open state. This is not purely coincident, because the coefficient for T_{out} (1.1103) in the fixed effect is greater than twice the standard deviation for T_{out} (0.5061) in the random effect, and thus, the chance of drawing a negative coefficient for T_{out} is slight.

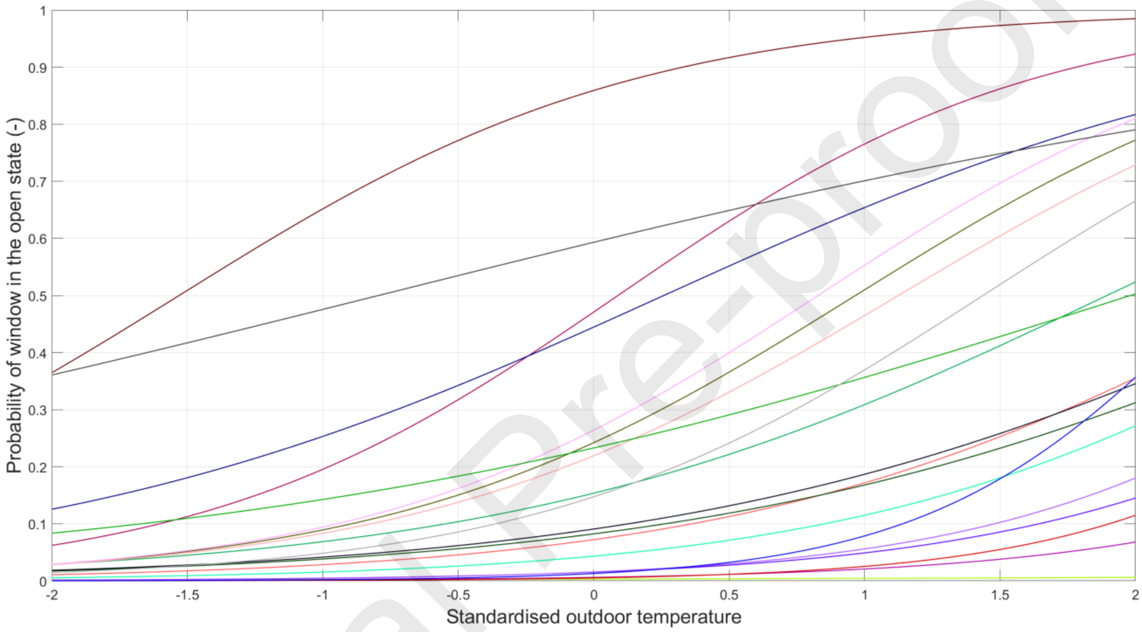


Figure 3. Probability of window in the open state based on outdoor temperature, with other environmental variables fixed at their standardised means. Note that different colours represent the model projections associated with different values of coefficients randomly generated for the developed multilevel model.

In terms of outdoor relative humidity, the absolute value of the regression coefficient for RH_{out} (-0.3388) in the fixed effect part was very close to the standard deviation for RH_{out} (0.3351) in the random effect part. Therefore, in this case, if the randomly drawn coefficient for outdoor relative humidity is higher than the estimated population mean by around 1 standard deviation, the curve is very flat, such as those shown at the bottom of Figure 4; for example, a random slope for RH_{out} can be -0.0037 which is 1 standard deviation higher than the population mean of the slope for RH_{out}. If the value of the randomly drawn slope is higher than the population mean by more than 1 standard deviation, the curve displays an increasing trend; for example, a random slope for RH_{out} can be 0.3314 which is 2 standard deviations higher than the population mean of the slope for RH_{out}. If the random slope is lower than the population mean, the curve would show a declining trend; for example, a random slope for RH_{out} can be -0.6739 which is 1 standard deviation lower than the population mean of the slope for RH_{out}.

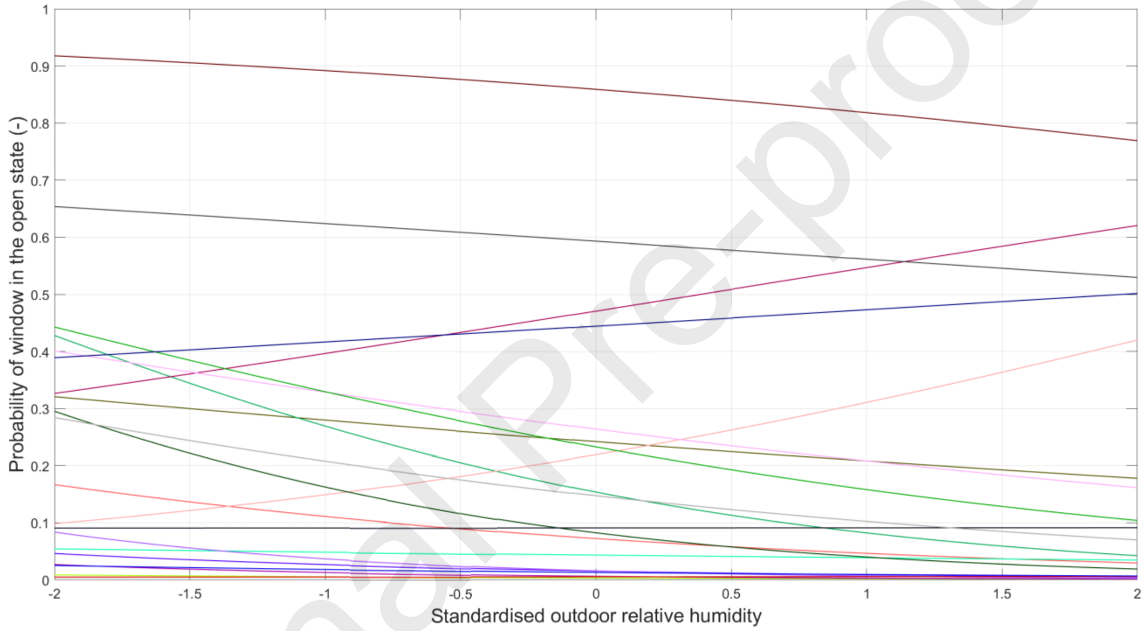


Figure 4. Probability of window in the open state based on outdoor relative humidity, with other environmental variables fixed at their standardised means. Note that different colours represent the model projections associated with different values of coefficients randomly generated for the developed multilevel model.

The absolute values of the regression coefficients for outdoor $PM_{2.5}$ concentration (-0.3276) and wind speed (-0.1019) in the fixed effect part were significantly less than the standard deviations for outdoor $PM_{2.5}$ concentration (0.5608) and wind speed (0.2507) in the random effect part, respectively. Thus, different trends and varying slopes of curves are expected as shown in Figure 5 and Figure 6.

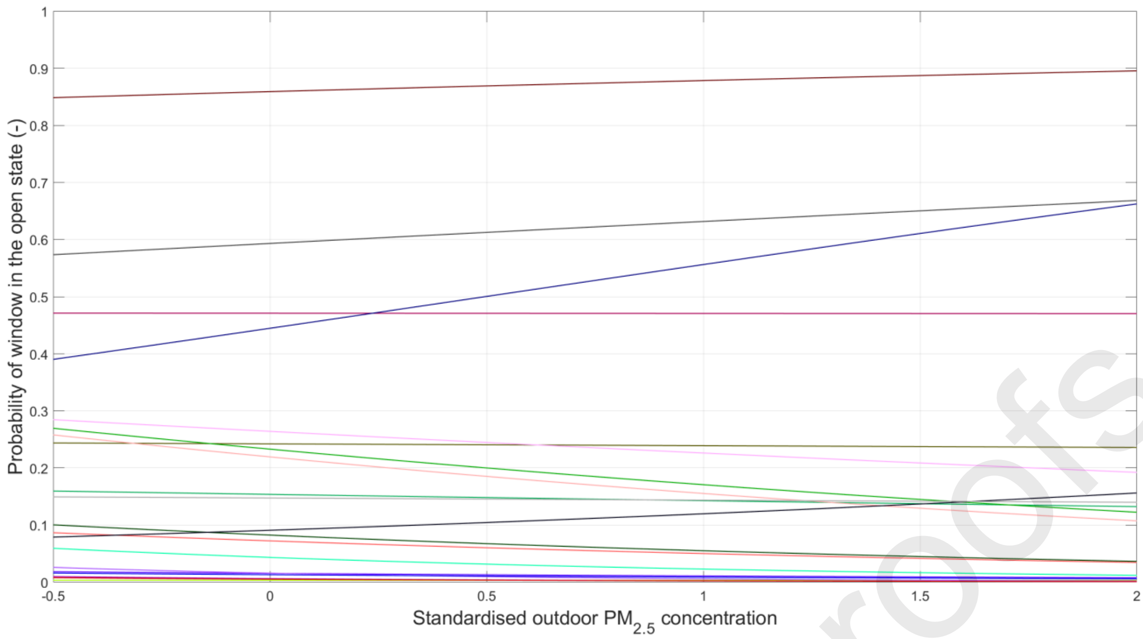


Figure 5. Probability of window in the open state based on outdoor $PM_{2.5}$ concentration, with other environmental variables fixed at their means. Note that different colours represent the model projections associated with different values of coefficients randomly generated for the developed multilevel model.

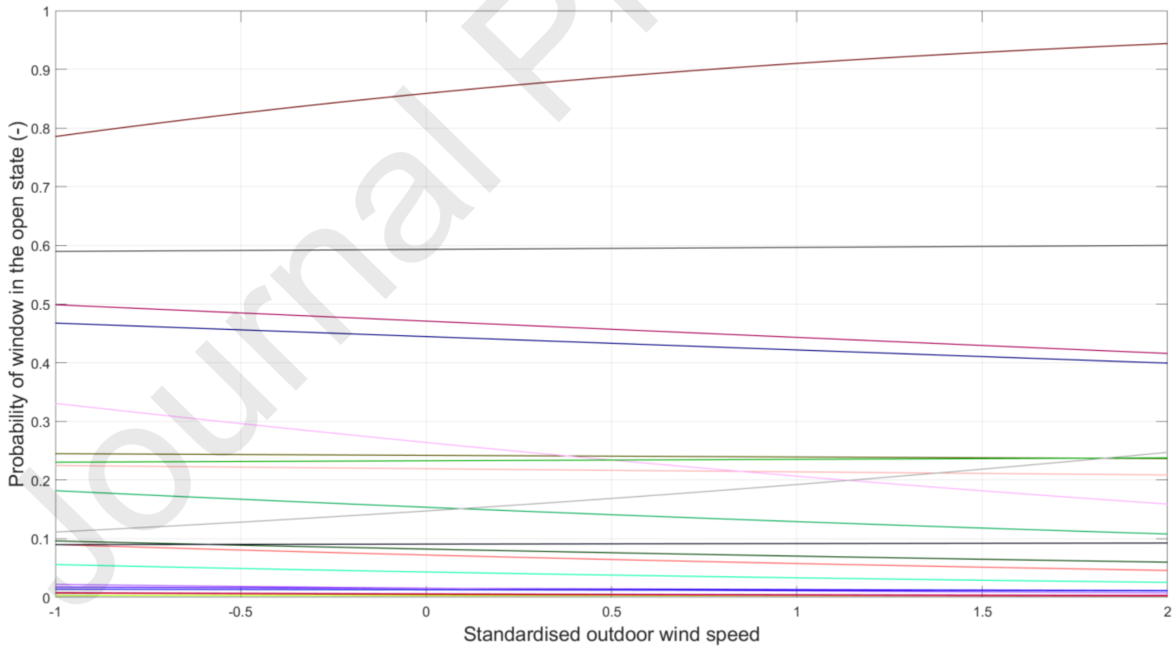


Figure 6. Probability of window in the open state based on outdoor wind speed, with other environmental variables fixed at their means. Note that different colours represent the model projections associated with different values of coefficients randomly generated for the developed multilevel model.

Beyond the examples above, one could infer that the relationship between the absolute values of the population means in the fixed effect part and the corresponding standard deviations in the random effect part is a key determinant of the degree of variation in occupant behaviour. If the standard deviations are much larger than the absolute values of the respective means, more diverse behaviours are expected across different occupants. On the other hand, if the standard deviations are relatively small compared to the absolute values of the means, inter-occupant behaviours are more alike.

4. Discussion

4.1 Main findings

The room-level behavioural variation was found to be statistically significant. This finding is understandable and expected, as both spatial and human factors could play an important role in affecting occupant behaviour in buildings [49]. However, adding the random effect from the apartment-level variation in the effects of outdoor environmental variables on the window open state was not statistically significant. This could be because much of the behavioural variation has already been captured at the room level. This finding can facilitate the multilevel modelling process by only modelling the room-level occupant behaviour for studying the diversity of occupants' use of windows in residential buildings. In contrast, Shi et al. [23] reported significant apartment-level variation in the probability of windows in the open state in Chinese apartment buildings. However, it should be noted that the results cannot be interpreted separately from the multilevel model structure. Shi et al.'s model only considered apartment-level variation and environmental variables, namely a 2-level model. Therefore, their conclusion about apartment-level behavioural variation was not directly comparable to the one reported in this study, where the apartment-level variation was modelled in addition to the room-level variation. Nevertheless, behaviour diversity was significant at the finest analysis unit level in both our study (i.e., room-level) and theirs (i.e., apartment-level).

This study also examined the potential behavioural differences between different types of rooms in the apartment building, namely living room versus bedroom. The statistical evidence suggested there were no significant differences between the two types of rooms in the relationship between the window open state and outdoor environmental variables. This finding suggests that when occupancy schedules are available for building performance simulations, outdoor environmental variables can be used in a similar way to predict the window open state for either the living room or bedroom.

4.2 Strengths and contributions

Compared to previous studies that developed probabilistic occupant behaviour models, this study adopted the multilevel modelling approach that has rarely been applied in the domain of occupant behaviour in buildings. In comparison with very few studies that developed multilevel models for window open state or window operation, the research presented here accounted for the hierarchical structure of the data at a fine scale by distinguishing room and apartment levels. In addition, given the paucity of similar models, this study makes several contributions to the literature:

- A multilevel logistic regression model for predicting the window open state based on outdoor environmental variables in recently-built low-energy apartment buildings in the UK has been established.

- A step-by-step methodological framework for modelling occupant behaviour following a hierarchical structure of room and apartment levels is presented in detail. This statistical modelling framework can be applied to other building settings.
- The multilevel model developed for predicting the window open state can be useful for supporting building design and operation for similar low-energy apartment buildings in moderate climatic conditions, with potential applications discussed later in section 4.3.
- This work does not directly contribute to our understanding of building energy use and performance, but instead provides valuable information for others in, for example, building energy modelling, as reliable diversity information on occupant behaviour is a necessity for occupant behaviour models to provide effective support for simulation-aided building design [15]. By helping to identify the room- and apartment-level characteristics of occupant behaviour that are meaningful and significant in the natural ventilation of residences, future research can more accurately and robustly estimate the variation in building energy and IAQ impacts.

4.3 Applications of multilevel window state models

The random effects in the multilevel models are representative of ~~inter-occupant behavioural variation~~ variations of the effects of environmental variables on the window open state, and the associated statistical expression can provide a sound basis to implement occupant behaviour models in the building simulation framework. ~~For example,~~ A general application of multilevel window state models in building simulations is described as follows. T~~t~~he process described in section 2.4, which randomly draws model coefficients based on the multilevel model, can be used to generate different sets of window state models. Then, through Monte-Carlo simulation, each window state model can produce time-series window state profiles when the appropriate environmental data are given as model inputs, as illustrated in previous work [15]. Even given the same outdoor environmental conditions, window state models with different combinations of slopes and intercepts could lead to different window state profiles. Finally, these window state schedules can be fed into building performance simulation tools for probabilistic predictions of, for example, the concentration of indoor air pollutants and space heating and cooling demands. The statistical distribution of such simulation results could be useful for the evaluation of the robustness of the building design against the variability of occupant behaviour.

4.4 Limitations and future work

The current work only covered living rooms and bedrooms in recently-built low-energy apartment buildings. Future work can extend to different room types (e.g., bathroom, kitchen), building typologies (e.g., houses), and countries (e.g., with different climates and cultures), including more building context details (e.g., room and window orientations), using the multilevel modelling framework described here. Moreover, due to data availability, this modelling work was based on only a few months covering the autumn and winter seasons, and the study sample was also restricted to UK apartment buildings. Therefore, it is hard to extrapolate the findings presented in this study to different building settings and periods. Additionally, in the current multilevel model, only four outdoor environmental variables were used as explanatory variables. Beyond environmental variables, occupancy stages (e.g., arrivals and departures) were not considered in the multilevel modelling of this study. Occupants' use of windows was found to show different patterns in different occupancy phases in offices ([17, 19]), but very little is known about residences.

Therefore, i~~I~~t would be desirable to include a wider range of environmental variables (such as ambient noise level) and other contextual information (such as the occupancy phases) in future work ~~to give a holistic representation of the environment people experience~~.

~~Modelling behavioural diversity is not the ultimate goal of our research.~~ In this study, the room-level variation in the effects of outdoor environmental variables on the window open state was found to be significant, while other studies reported the spatial variation in indoor air quality [50], occupant comfort and building energy consumption [51]. Given that it is well acknowledged that occupant behaviour could significantly affect building performance [4] [11] [12], a question emerged naturally - how does the diversity of occupant behaviour relate to the diversity of the indoor environment quality, building energy demand patterns and building users' satisfaction? It is hoped that our research team would be able to carry out future monitoring and modelling work to examine the complex relationship between the diversity of occupant behaviour and the diversity of building performance, with this work as the first step.

5. Conclusion

To answer three research questions, multilevel logistic regression models were developed to predict the probability of the window being in the open state in eighteen low-energy apartments in the UK based on measured outdoor environmental variables. The results showed that there was significant variation in the effects of outdoor environmental variables on the window open state at the *room* level which should be accounted for, but not at the *apartment* level. Additionally, the statistical relationship between studied outdoor environmental variables and the window open state in the living room was not significantly different from that in the bedroom. As an original contribution, this study considered the *room*- and *apartment*-level behavioural heterogeneity, following the hierarchical data structure, in developing multilevel logistic regression models for predicting the window open state in residential buildings. The developed multilevel ~~window state~~ model can be ~~further~~ used to simulate different occupant profiles by generating different sets of window state models based on environmental variables; then, different window state models can yield corresponding window state schedules as essential inputs to ~~inte~~ ~~grated~~ ~~into~~ building performance simulations for various applications such as probabilistic predictions of indoor air pollutants and heating/ cooling demands. Building simulation to predict future building performance during the design phase, and to assess buildings post-occupancy, is important in the optimisation of building energy use and indoor air quality. Therefore, tools and techniques to improve the accuracy in predicting complex systems, such as occupants' use of windows, are critical to achieving building energy, carbon reduction and indoor air quality goals.

Acknowledgements

The research presented was supported by the EIT-Digital project, "Quality of Indoor Air on Sites Matched with Outdoor Air Quality Datasets to Improve Wellbeing Outcomes" (activity number 19144). In addition, this study largely benefited from the authors' participation in IEA-EBC Annex 79. All aspects of the work involving human participants were in accordance with, and approved by, the ethical standards of the institution (University College London) and all collected data was handled in accordance with the provisions of the Regulation (EU) 2016/679: EU General Data Protection Regulation (GDPR).

Appendix A. Supplementary file

The supplementary file to this article can be found online.

References

1. Symonds, P., et al., *Home Energy Efficiency and Subjective Health in Greater London.* J Urban Health, 2021. **98**(3): p. 362-374.
2. Zhang, Y., et al., *Rethinking the role of occupant behavior in building energy performance: A review.* Energy and Buildings, 2018. **172**: p. 279-294.
3. Tong, H., et al., *Increases in noise complaints during the COVID-19 lockdown in Spring 2020: A case study in Greater London, UK.* Science of the Total Environment, 2021. **785**: p. 147213.
4. 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.
5. Azar, E. and C.C. Menassa, *A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings.* Energy and Buildings, 2012. **55**: p. 841-853.
6. Yan, D., et al., *Occupant behavior modeling for building performance simulation: Current state and future challenges.* Energy and Buildings, 2015. **107**: p. 264-278.
7. Hong, T., et al., *Advances in research and applications of energy-related occupant behavior in buildings.* Energy and Buildings, 2016. **116**: p. 694-702.
8. Dong, B., et al., *Modeling occupancy and behavior for better building design and operation—A critical review.* Building Simulation, 2018. **11**(5): p. 899-921.
9. Nord, N., et al., *Influence of occupant behavior and operation on performance of a residential Zero Emission Building in Norway.* Energy and Buildings, 2018. **159**: p. 75-88.
10. 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.
11. Haldi, F. and D. Robinson, *The impact of occupants' behaviour on building energy demand.* Journal of Building Performance Simulation, 2011. **4**(4): p. 323-338.
12. 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.
13. Wang, Y., Elizabeth Cooper, Farhang Tahmasebi, Jonathon Taylor, Samuel Stamp, Phil Symonds, Esfandiar Burman, Dejan Mumovic, *Improving indoor air quality and occupant health through smart control of windows and portable air purifiers in residential buildings.* Building Services Engineering Research and Technology, 2022.
14. Nicol, J.F. and M.A. Humphreys, *A Stochastic Approach to Thermal Comfort-Occupant Behavior and Energy Use in Buildings/Discussion.* ASHRAE transactions, 2004. **110**: p. 554.

15. 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.
16. Raja, I.A., et al., *Thermal comfort: use of controls in naturally ventilated buildings.* Energy and buildings, 2001. **33**(3): p. 235-244.
17. Haldi, F. and D. Robinson, *Interactions with window openings by office occupants.* Building and Environment, 2009. **44**(12): p. 2378-2395.
18. Andersen, R.K., V. Fabi, and S.P. Corgnati, *Predicted and actual indoor environmental quality: Verification of occupants' behaviour models in residential buildings.* Energy and Buildings, 2016. **127**: p. 105-115.
19. 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.
20. Stazi, F., F. Naspi, and M. D'Orazio, *A literature review on driving factors and contextual events influencing occupants' behaviours in buildings.* Building and Environment, 2017. **118**: p. 40-66.
21. Dai, X., J. Liu, and X. Zhang, *A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings.* Energy and Buildings, 2020. **223**.
22. Fabi, V., R.K. Andersen, and S. Corgnati, *Verification of stochastic behavioural models of occupants' interactions with windows in residential buildings.* Building and Environment, 2015. **94**: p. 371-383.
23. Shi, S., et al., *Effects of household features on residential window opening behaviors: A multilevel logistic regression study.* Building and Environment, 2020. **170**.
24. 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.
25. O'Brien, W., et al., *A preliminary study of representing the inter-occupant diversity in occupant modelling.* Journal of Building Performance Simulation, 2016. **10**(5-6): p. 509-526.
26. Tahmasebi, F. and A. Mahdavi, *The sensitivity of building performance simulation results to the choice of occupants' presence models: a case study.* Journal of Building Performance Simulation, 2015. **10**(5-6): p. 625-635.
27. de Wilde, P., *The gap between predicted and measured energy performance of buildings: A framework for investigation.* Automation in Construction, 2014. **41**: p. 40-49.
28. Wang, Y., et al., *Exploring the relationship between window operation behavior and thermal and air quality factors: A case study of UK residential buildings.* Journal of Building Engineering, 2022. **48**.
29. 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.
30. 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.
31. Qian, S.S., et al., *On the application of multilevel modeling in environmental and ecological studies.* Ecology, 2010. **91**(2): p. 355-361.
32. Bingenheimer, J.B. and S.W. Raudenbush, *Statistical and substantive inferences in public health: issues in the application of multilevel models.* Annu Rev Public Health, 2004. **25**: p. 53-77.

33. Li, C., Y. Song, and N. Kaza, *Urban form and household electricity consumption: A multilevel study.* Energy and Buildings, 2018. **158**: p. 181-193.
34. Prignon, M., et al., *A method to quantify uncertainties in airtightness measurements: Zero-flow and envelope pressure.* Energy and Buildings, 2019. **188-189**: p. 12-24.
35. Belaïd, F., D. Roubaud, and E. Galariotis, *Features of residential energy consumption: Evidence from France using an innovative multilevel modelling approach.* Energy Policy, 2019. **125**: p. 277-285.
36. Kent, M.G., et al., *Temporal effects on glare response from daylight.* Building and Environment, 2017. **113**: p. 49-64.
37. Haldi, F., et al., *Modelling diversity in building occupant behaviour: a novel statistical approach.* Journal of Building Performance Simulation, 2016. **10**(5-6): p. 527-544.
38. 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**.
39. Wang, Y., et al., *An investigation of the influencing factors for occupants' operation of windows in apartments equipped with portable air purifiers.* Building and Environment, 2021. **205**: p. 108260.
40. Zauli Sajani, S., et al., *Vertical variation of PM2.5 mass and chemical composition, particle size distribution, NO2, and BTEX at a high rise building.* Environmental Pollution, 2018. **235**: p. 339-349.
41. Stamp, S., et al., *Long-term, continuous air quality monitoring in a cross-sectional study of three UK non-domestic buildings.* Building and Environment, 2020. **180**.
42. Chatzidiakou, L., D. Mumovic, and A. Summerfield, *Is CO2 a good proxy for indoor air quality in classrooms? Part 1: The interrelationships between thermal conditions, CO2 levels, ventilation rates and selected indoor pollutants.* Building Services Engineering Research and Technology, 2015. **36**(2): p. 129-161.
43. Cheriyan, D. and J.H. Choi, *Data on different sized particulate matter concentration produced from a construction activity.* Data Brief, 2020. **33**: p. 106467.
44. Chatzidiakou, L., et al., *Characterising low-cost sensors in highly portable platforms to quantify personal exposure in diverse environments.* Atmospheric Measurement Techniques, 2019. **12**(8): p. 4643-4657.
45. Crilley, L.R., et al., *Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring.* Atmospheric Measurement Techniques, 2018. **11**(2): p. 709-720.
46. Mahdavi, A. and F. Tahmasebi, *On the quality evaluation of behavioural models for building performance applications.* Journal of Building Performance Simulation, 2016. **10**(5-6): p. 554-564.
47. Buse, A., *The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note.* The American Statistician, 1982. **36**(3a): p. 153-157.
48. Harrison, X.A., et al., *A brief introduction to mixed effects modelling and multi-model inference in ecology.* PeerJ, 2018. **6**: p. e4794.
49. Marín-Restrepo, L., M. Trebilcock, and M. Gillott, *Occupant action patterns regarding spatial and human factors in office environments.* Energy and Buildings, 2020. **214**.

50. Stamp, S., et al., *Seasonal variations and the influence of ventilation rates on IAQ: A case study of five low-energy London apartments.* Indoor and Built Environment, 2021. **31**(3): p. 607-623.
51. Shahzad, S., et al., *Energy and comfort in contemporary open plan and traditional personal offices.* Applied Energy, 2017. **185**: p. 1542-1555.

Highlights:

- Between-room variation in the effect of outdoor conditions on window openings was significant.
- Between-flat variation in the effect of outdoor conditions on window openings was not significant.
- The room type did not significantly affect occupants' window opening behaviour.
- The developed model accounted for hierarchical data structure in occupant behaviour.