

A Model Based on Gauss Distribution for Predicting Window Behavior in Building

Song Pan^{1,2,3}, Yiye Han¹, Shen Wei⁴, Yixuan Wei^{5,*}, Liang Xia^{5,*}, Lang Xie⁶,
Xiangrui Kong⁷, Wei Yu⁶

¹ Beijing Key Laboratory of Green Built Environment and Energy Efficient Technology, Beijing University of Technology, Beijing 100124, China

² Engineering Research Center of Digital Community, Ministry of Education, Beijing, 100124, China

³ Beijing Laboratory for Urban Mass Transit, Beijing, 100044, China

⁴ The Bartlett School of Construction and Project Management, University College London (UCL), London, WC1E 7HB, UK

⁵ Research Centre for Fluids and Thermal Engineering, University of Nottingham Ningbo China, Ningbo 315100, China

⁶ Department of Architectural Engineering, North China Institute of Aerospace Engineering, Hebei 065000, China

⁷ Department of Architectural Engineering, North China institute of science and technology, Hebei 065201, China

* Corresponding authors.

E-mail addresses: Yixuan.WEI@nottingham.edu.cn (Y. Wei), Liang.Xia@nottingham.edu.cn (L. Xia)

Abstract

Modeling of window behavior is a key component for building performance simulation, due to the significant impact of opening/closing windows on indoor environment and energy consumption. The predictions of existing models cannot well reflect actual window behavior, the prediction accuracy still needs to be improved. The Gauss distribution model is a new machine-learning technique which has achieved successful applications in many fields because of its special advantages (i.e. simple structure, strong operability and flexible nonparametric inference ability) compared to existing models. This paper presents results from a study using the Gauss distribution model to predict window behavior in office building. The data used in this study were from a real building located in Beijing, China, and covered two transitional seasons (from October 1 to November 15, 2014 and from March 15 to May 16, 2015), when natural ventilation was fully applied. When modeling, three types of input variables, i.e., indoor temperature, outdoor temperature and their combination were used. This work validates the importance of selecting suitable input variables when developing Gauss distribution model. This study also compared the prediction performance between the Gauss distribution modeling approach and the Logistic regression modeling approach, which is the most popular method used to model occupant window behavior in buildings. The results showed that Gauss distribution models could provide higher prediction accuracy, with 9.5% higher than Logistic regression model when using suitable inputs. This paper provided a novel modeling method that can be used to predict window states more accurately in office buildings.

Keywords

Window behavior, Gauss distribution, Logistic regression, Modeling, Office building.

1 Introduction

1.1 State of the art

The rapid development of computer technologies has enabled the usage of building performance simulation for building design and operation. The energy consumption of buildings is mainly affected by six factors: meteorological conditions, building envelope, building equipment, indoor environmental parameters, operation management and occupant behavior. In the past decade, significant progresses and advanced technologies have been achieved in terms of above aspects, expect occupant behavior [1-2]. Some methods, such as the survey-based approach [3-9], Data-driven approaches [10-11], and building performance simulation (BPS) [12] have been used to evaluate the impact of occupant behavior on buildings performance.

Numerous studies [4-9,12-28] have confirmed that occupant's interaction with building systems would attributes to sizeable variation in building energy consumption. For example, Takasu et al. [9] conducted questionnaire based field surveys to record thermal comfort responses of occupants and found that find behavioral adaptation related to window-opening leading to variation in the comfort temperature across different seasons. However, in conventional simulation packages, occupant behavior is described in the form of either fixed schedules or rule-based methods, which fail to capture the stochastic nature of occupant behavior. This simplification of occupant behavior will significantly reduce the reliability and accuracy of results from building performance simulation [24-28]. On the other hand, many simulated measurements or retrofits [29-35] with significantly energy-saving potential often fail to reach expected performance in real situations. Sometimes the situation would be even worse that energy consumption of real buildings is increased after simulated energy-efficient measurements are adopted [36-40], one of the

important reasons for this phenomenon is the huge deviation between the behavioral modeling in simulation and actual occupants behavior. Therefore, designers should consider the complexity and variation of occupant behavior when designing buildings. Undoubtedly, inaccurate descriptions concerning occupant behavior would inevitably result in great deviation between building design and operation, and this deviation is often referred as “performance gap” [41-42]. Therefore, better understanding and more accurate modeling of occupant behavior in buildings is vital to bridge the gap between simulation result and actual building performance, especially for those buildings that largely depend on passive design features and occupancy controlled technologies [43-44].

As one important type of occupant behavior, window behavior has been proven to have significant impact on building performance [45-46]. Operable windows provide opportunities for occupants to improve the indoor air quality and thermal comfort by nature ventilation. To better describe and predict window behavior, stochastic modeling approaches have been adopted because they can capture the random characteristics of window behavior and provide probabilistic distributions based on performance indicators [47]. Many efforts have been carried out to develop stochastic models based on large amount of measured data to better predict window behavior in buildings [48-51].

In early 1990, Fritsch et al. [52] first attempted to mathematically model window states in buildings, using a stochastic model based on the Markov chains process. The time series model of window angle was generated based on half-hourly measured data from four windows in a real building. Rijal et al. [53-54] have proposed a behavioral model based on multiple logit distributions to predict the probability that a window is open. In this model, both indoor and outdoor temperatures were considered as

independent variables. Based on almost seven years of continuous measurement from 14 south-facing cellular offices, Haldi and Robinson [55] used three different modeling methods to model occupant window behavior, including a Bernoulli process based on a logit probability distribution, a discrete-time Markov process with sub-models for different occupancy states and a continuous-time random process. Then, they ranked these methods according to their prediction performance.

Since 1990, many studies on window behavior in office buildings have been carried out and Logistic regression has been popularly adopted as the modelling approach. Zhang and Barret [56] carried out a field study of occupant window behavior in a naturally ventilated office building in the UK, lasting for 16 months. A probabilistic model for predicting window opening and closing actions was developed considering outdoor temperature as the only independent variable. Moreover, in the UK, Wei et al. [57] monitored occupant decisions on the end-of-day window positions in an office building and analyzed potential factors that may drive their decision-making, using Logistic regression analysis. Based on a more comprehensive set of indoor and outdoor environmental variables, Andersen et al. [58] used multivariate Logistic regression to describe the probability of window state change (closed to open and open to closed). Their study was carried out in 15 dwellings located in Denmark. Additionally, four models were developed in their study according to the ownership and ventilation type of the building. Li et al. [59] have used Logistic regression analysis to propose a probability model of occupant window operation in China. The measured data were selected from a two-month field observation in a natural-ventilation office building during transitional seasons in Chongqing, China. Simona et al. [49] developed a framework combining statistical analysis with two data-mining techniques, cluster analysis and association rules mining, to identify valid window

operational patterns in the measured data. They analyzed a dataset with measured indoor and outdoor physical parameters and human interaction with operable windows in 16 offices. Logistic regression has been used to identify factors influencing window-opening and window-closing behavior in their study. Moghadam et al. [60] attempted to simulate the impact of individual behavior on the window state, by implementing Logistic regression models to describe three distinct types of window behavior from active, passive and generalized window users. In 2018, Pan et al. [61] analyzed the effects of both environmental and non-environmental factors on window opening/closing behaviors and developed window models using Logistic regression. The study was based on a 9.5-month field measurement for a total of 5 windows in 5 offices on the second floor in Beijing, China. Naspi et al. [62] conducted a survey to determine the presence of people and their interactions with windows in three offices. In this study, trigger parameters of window opening and closing actions were investigated, and Logistic regression models were proposed to predict openings at arrival and openings/closings during intermediate periods.

Besides, there was a lot of studies on window behavior in residential buildings and Logistic regression was used as the main modelling approach. Cali et al. [63] analyzed window behavior in residential buildings, and investigated drivers who lead occupants to interact with windows, while focusing on how these actions can be modeled. A method to analyze the probability of a state change of the windows, based on Logistic regression, was used on the data from a four-year measurement in two refurbished demonstration buildings. Jones et al. [64] conducted a field study in ten UK dwellings over the period of a year, and used multivariate Logistic regression to investigate the probability of window opening and closing of a main bedroom, based on indoor and outdoor environment factors, considering the time of the day and

season as well. Meanwhile, Yao and Zhao [65] conducted an investigation on the factors influencing occupant window behavior in 19 residences in Beijing, based on the monitored state of windows and eight environmental parameters. In their study, multivariate linear Logistic regression was also used to establish predictive models for occupant window behavior, and the results indicated that outdoor temperature was the most influential factor. Stazi et al. [66] investigated the relationship between window use and environmental stimuli in an Italian classroom, and developed a window behavior model using Logistic regression. In their study, it was found that indoor temperature was the best predictor for both opening and closing windows. In this study, outdoor temperature also had a significant impact on the window states but not as strong as indoor temperature. In the same year, Kim et al. [8] conducted longitudinal field observations and used the logistic function based on a second-order polynomial regression to create fit curves of window behavior.

The summaries of potential factors that influence window-opening behavior and existing modeling approaches of occupants' window behavior are shown in [Table 1](#) and [Table 2](#), respectively.

Table 1. The influencing factors of occupants' window behavior.

Environmental factors	Outdoor temperature; outdoor relative humidity; indoor temperature; indoor relative humidity; indoor CO ₂ concentration.
Non-environmental factors	Occupants' age, gender; time of day; season; previous status of the window; heating modes; house property ownership; window type and orientation; building and room types; floor level; smoking; occupancy pattern; personal preference; rule, special events.

Table 2. Existing modeling approaches of occupants' window behavior.

Modeling approaches	Used as/by (examples)
• Deterministic model	Simplified model
• Threshold model	Simplified model
• Markov chain with Logistic regression and survival analysis model	Haldi and Robinson (2009)
• Markov with Logistic regression model	Fritsch et al. (1990)/ Andersen et al. (2013) Cali et al. (2016)/ Federica et al.(2018)
• Logistic regression model	Rijal et al.(2007)/ Zhang et al. (2012) Wei et al. (2012)/ Nan Li et al. (2014) Simona et al.(2014)/ Maghadam et al.(2015) Jones et al. (2017)/ Stazi et al. (2017) Naspi et al. (2017)/ Yao et al.(2017) Kim et al. (2017)/ Pan et al. (2018)

1.2 Statements of objectives

Most recent studies conducted to model occupants' window behavior were developed based on mathematical approaches, i.e. Markov chains process, Bernoulli process, and Logistic regression. However, it is difficult to develop an elaborated mathematical model to describe the stochastic occupancy behavior, which is influencing by multiple factors simultaneously. Hence, the prediction accuracy of these proposed models is not always satisfying. In order to further improve the model performance, the Gauss distribution approach, a new machine-learning technique, has been tested and used to model window opening behavior.

2 Research Methodology

2.1 Basic principle of the Gauss Distribution model

Window behavior is a very complex phenomenon with a high degree of nonlinearity and randomness and may be driven by many factors. In existing studies, environmental factors, such as indoor temperature and outdoor temperature, have been identified as main driving factors and were usually used as inputs of predicting models for window behavior. Hence, multivariate Gauss distribution model has been adopted in this study because more than one environmental factor were considered, expressed as a vector superposition of Gaussian distribution. The specific calculation of window opening probability is shown in Formula (1):

$$P(X) = \sum_{i=1}^t m_i F(x_i | \mu_i, \sigma_i^2) \quad (1)$$

where $P(X)$ represents probability of window opening; X is a vector representing t kinds of environmental factors ($x_i, i = 1, \dots, t$); m_i represents the weight coefficient of the i^{th} influential factor; $F(x_i | \mu_i, \sigma_i^2)$ represents the cumulative distribution function of the Gaussian distribution of the i^{th} influential factor; x_i represents the specific value of the i^{th} influential factor, and μ_i and σ_i^2 represent the corresponding mean value and variance, respectively.

As seen from Formula (1), to calculate the probability of window opening, $P(X)$, the cumulative distribution function of Gaussian distribution $F(x | \mu, \sigma^2)$ and the weight coefficients for all input variables need to be determined, following the specific steps described below:

1. Calculating the cumulative distribution function of Gaussian distribution $F(x | \mu, \sigma^2)$

The cumulative distribution function of the Gaussian distribution $F(x | \mu, \sigma^2)$ refers to the probability that random variable X is less than or equal to X , expressed as a density function and shown in Formula (2):

$$F(x | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (2)$$

Before calculating $F(x | \mu, \sigma^2)$ for each input variable, corresponding values of μ and σ^2 need to be determined. In this model, the least-square method has been adopted to find the corresponding μ and σ^2 for the training dataset, following the steps below:

- a. For each environmental parameter that has a significant impact on window behavior, the corresponding probability of window opening needs to be determined. Since all environmental parameters are time-continuous, a suitable interval length should be selected;
- b. There is a set of datasets of environmental parameters for each interval selected in step a, the probability of window opening based on each dataset is calculated and the median is adopted to present the probability of window opening at that interval;
- c. The least-squares method is used to fit these datasets and determine the mean and variance of each environment variable.

2. Calculating weight coefficient m

In reality, there is more than one environmental factor that has a significant influence on occupant window behavior. Therefore, to predict window behavior more

accurately, it is necessary to describe the influence degree, named weight, for each factor, as shown in Formula (1). Then, in view of the normalization requirement, the influential factors selected as model inputs are considered to impact window behavior, leading to Formula (3):

$$\sum_{i=1}^t m_i = 1 \quad (3)$$

In this study, the following assumptions have been used: the window was “closed” when the window opening probability $P(X) < 0.5$, and it was “open” when the window opening probability $P(X) \geq 0.5$, as shown in Formula (4):

$$\begin{cases} \sum_{i=1}^t m_i \sum_{l=1}^k F(x_{il} | \mu_{il}, \sigma_{il}^2) < 0.5, \text{ window is closed} \\ \sum_{i=1}^t m_i \sum_{l=k+1}^p F(x_{il} | \mu_{il}, \sigma_{il}^2) \geq 0.5, \text{ window is open} \end{cases} \quad (4)$$

where k is the number of closed window states, and p is the total number of window states for the training dataset.

As mentioned above, the cumulative distribution function of the Gaussian distribution $F(x | \mu, \sigma^2)$ could be determined in accordance with Steps a–c. Afterwards, the weight coefficient m could be determined by Formulas (3) and (4). In addition, the Monte Carlo method is used to determine the final window state (opened or closed) in this model. A random number in the range of 0 to 1 is generated and compared with calculated probability of window behavior. If the random number is less than the calculated probability of window behavior, the output is “opened”, otherwise it is “closed”.

The frame structure of the Gauss distribution model is shown in [Fig. 1](#).

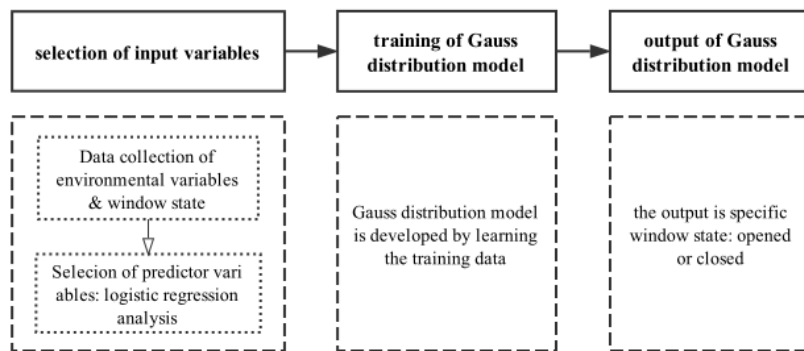


Fig. 1 The frame structure of the Gauss distribution model

2.2 Data collection

To verify the applicability of using the Gaussian distribution to model window states in buildings, both environmental data and window state data were measured and collected from a typical free-running office building located in a university campus of Beijing, China. The impact of outside noise on occupants' window behavior could be negligible in this study because the onsite questionnaires show that the surrounding is usually quiet. In addition, there were no other tall buildings or trees blocking solar gains. The building had two floors, with some laboratories in the ground floor and nine 10m² offices in the second floor, as shown in Fig. 2. There was one south-facing push-pull window which is controlled by the occupant, as shown in Fig. 3. Measurements were carried out in the five offices in the second floor. Each office was single occupied by the same occupant throughout the entire measurement period. In addition, the occupants in the five offices were all non-smoking teachers, two of them are males and remains are females. They had lived in Beijing for many years and got used to Beijing climate already. In winter, radiator heating was used to heat the building, and in summer, split-type air conditioners were adopted. During transitional seasons, typically between October 1 and November 15, and between March 15 and May 16, natural ventilation was used as the main cooling strategy. The

data used in this study were collected from the two transitional seasons.



Fig. 2 The case study building



Fig.3 A typical office

Occupant window operation is highly correlated to the change of indoor and outdoor environments, which has been widely justified in existing studies [56-60,63-66]. For this reason, important environmental data were collected during the study. Environmental sensors were installed inside each room where the sensor is rarely impacted by outdoor disturbances and indoor heating resources. Outdoor parameters included air temperature, air humidity, PM_{2.5} concentration, solar radiation, sunshine hours, wind speed and direction were measured by the sensors installed on the roof of the case study building. In addition to these, pyroelectric infrared (PIR) sensors were used to record the occupancy of the monitored offices, and window displacement testers were installed to detect and record the state of windows. The main features of the monitoring equipment are shown in [Table 3](#).

Table 3 Measurement range and accuracy of monitoring equipment

Monitoring instruments/parameters	Recording interval	Sensitivity	Accuracy
Infrared instrument	1 min	5 m	
Window displacement tester	10 min	3 cm	
Indoor temperature sensor	10 min		$\pm 0.5^{\circ}\text{C}$
outdoor temperature	1min		$\pm 1.0^{\circ}\text{C}$
outdoor humidity	1min		$\pm 5\%$
solar radiation	1min		$\pm 10\text{W}/\text{m}^3$
sunshine hours	1min		$\pm 0.5\text{h}$
wind speed	1min		$\pm 1\text{m}/\text{s}$
wind direction	1min		$\pm 10^{\circ}$
PM _{2.5} concentration	1h		$\pm 10\mu\text{g}/\text{m}^3$

2.3 Selection of input parameters

Basic information about Gauss distribution models has been presented in Section 2.1. When modeling occupant window behavior, another issue to consider is the selection of input parameters, as inter-correlated input parameters both decrease the modeling accuracy and increase the computational time. To handle this issue and remove insignificant variables, the stepwise regression method was adopted to determine input parameters for the Gauss distribution model.

The stepwise regression method has been used in many fields [67-70], and it introduce variables one by one and test the imported variables simultaneously. If the introduction of a new variable makes any existing variables less significant, that existing variable(s) is eliminated to ensure that only significant variables are included in the regression model. This process is repeated until no more input variables are to

be added or eliminated from the regression model. In this study, SPSS22, a professional statistical package, was used to for the model development.

2.4 Evaluate criteria

Validation is crucial for model development to test the accuracy and usability of the developed models. It is commonly performed by comparing the predicted values against field measurements, which should be different from the training data. In this study, the validation dataset mentioned in Section 2.2 was prepared and used for this purpose. In this study, 2000 datasets were collected from the five offices. The datasets were equally and randomly split into two groups, one for training the model and the other for validating the model.

The distribution of the datasets is shown in Fig. 4 and Fig. 5. It can be seen that the datasets from two transitional seasons cover a similar indoor/outdoor temperature range. In addition, p-p diagram generated by SPSS22 was used to validate that the datasets obey normal distribution if majority of them are approximately located near the diagonal line. The results from Fig. 6 show that the both indoor and outdoor temperature datasets approximately follow a normal distribution. On the other hand, it also reflects rationality of data which was selected in random.

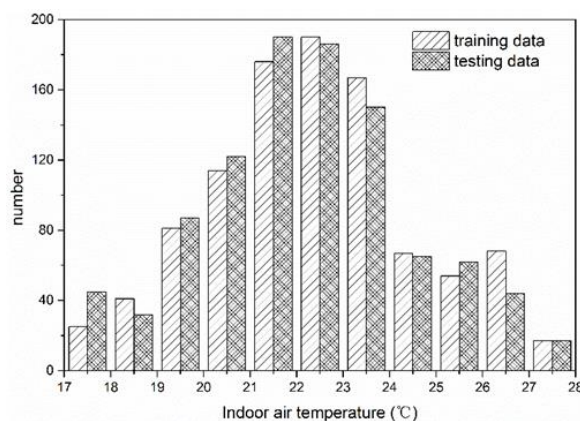


Fig. 4 Distribution of indoor temperature for two datasets

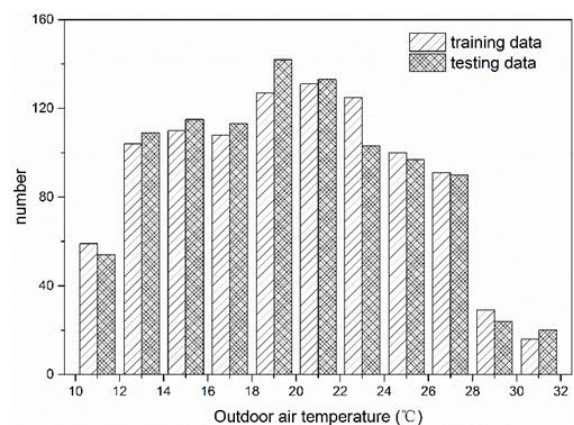
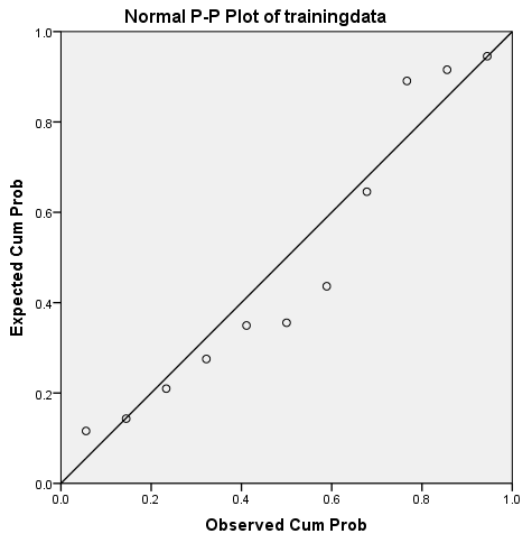
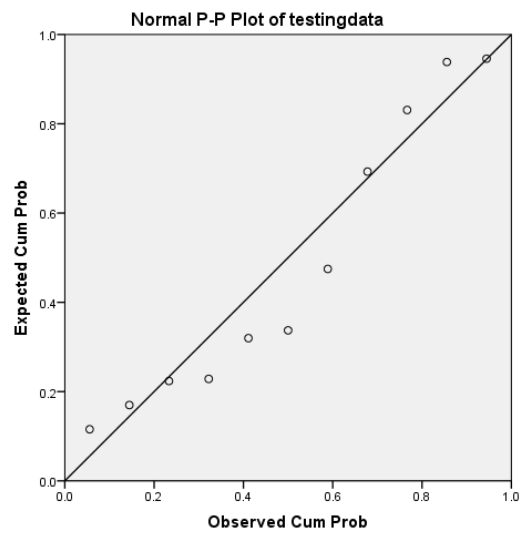


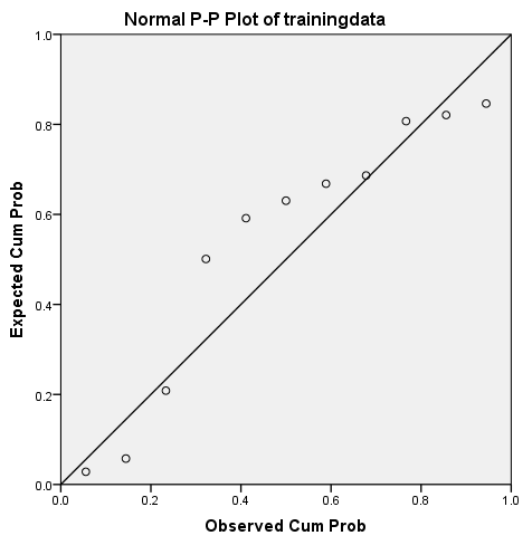
Fig. 5 Distribution of outdoor temperature for two datasets



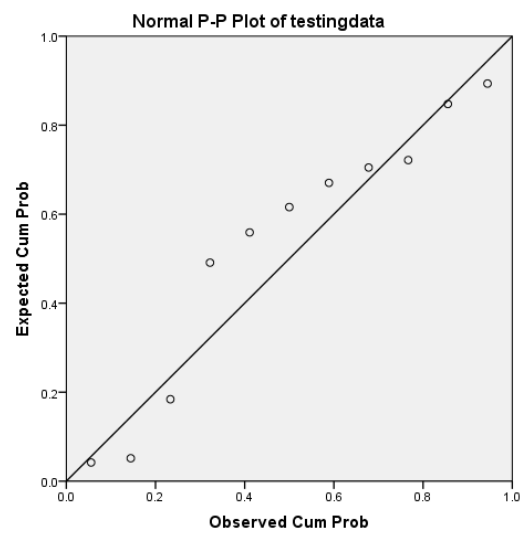
(a)



(b)



(c)



(d)

Fig. 6 normal distribution test of datasets for indoor temperature (a/b) and outdoor temperature (c/d)

Haldi and Robison [71] and Schweiker et al. [72] assessed performance of prediction models by comparing the observed window states and the predicted window states. Two states of windows were considered in their study, as positive was defined for

open windows and negative was defined for closed windows, resulting in definitions of predicted outcomes as true (truly positive, i.e. both predicted and measured window states were open, or truly negative, i.e. both were closed) or false (falsely positive, i.e. the predicted window state was open but the measured state was closed, or falsely negative, i.e. the predicted window state was closed but the measured state was open). Based on these classifications, Schweiker et al. [72] used the true positive rate (TPR) as the proportion of actual open windows that was correctly predicted as open, and true negative rate (TNR) as the proportion of actual closed windows that were correctly predicted as closed. Finally, the accuracy of the model (ACC) was defined as the proportion of correct predictions weighting the proportion of true outcomes (positive or negative) on the total amount of window states measured. In this paper, TPR, TNR and ACC were also adopted as testing criteria to assess the performance of the model, shown as formulas (5)-(7):

$$TPR = c/m \quad (5)$$

$$TNR = d/n \quad (6)$$

$$ACC = (c + d)/(m + n) \quad (7)$$

where c is the number of correctly predicted open states; d is the number of correctly predicted closed states; m is the total number of open states; and n is the total number of closed states.

2.5 Model Comparison

In existing studies, in terms of modeling occupant window behavior, Logistic regression has been widely chosen as the modeling method. In this study, to justify the performance of Gauss distribution models, predictions from Logistic regression

models were obtained and compared.

Logistic regression analysis is a statistical method with outputs as binary variables. Due to its flexibilities regarding to input parameters, which can be both categorical and continues, it has been extensively used for modeling occupant window behavior in buildings. In Logistic regression analysis, the relationship between the probability of a binary response and individual explanatory variables could be expressed by Formula (8),

$$\text{logit}P = \ln\left(\frac{P}{1-P}\right) = a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n + b \quad (8)$$

where P is the probability of opening or closing a window; a_1, a_2, \dots, a_n are coefficients for explanatory variables, such as indoor temperature and outdoor temperature, and b is a constant.

3 Results

3.1 A Gauss distribution Model for Window State Prediction

The training dataset introduced in Section 2.2 was used to develop the Gauss distribution model. The approach introduced in Section 2.3 was used to select input parameters for the model using SPSS22, with results shown in [Table 4](#).

Table 4 Results of stepwise regression

Variables	Beta	T	Sig.	Partial correlation	Tolerance
indoor temperature(°C)	0.006	0.105	0.916	0.005	0.881
outdoor air humidity(%)	-0.046	-0.937	0.349	-0.048	0.996
outdoor PM _{2.5} concentration(µg/m ³)	-0.155	-2.833	0.105	-0.143	0.800
solar radiation(W/m ³)	0.034	0.680	0.497	0.035	0.984
sunshine hours(h)	0.302	6.167	0.015	0.299	0.927
wind speed(m/s)	-0.027	-0.553	0.580	-0.028	1.000
wind direction(°)	-0.058	-1.158	0.248	-0.059	0.968
	B	T	Sig.	S.E	
Outdoor temperature	0.023	4.738	<0.001	0.005	

where *B* was regression coefficient; *Beta* referred to standardized regression coefficient and the magnitude of its absolute value directly reflected the influence of independent variable on dependent variable; *t* was the result of hypothesis testing on *B/Beta*; the *Sig.* value was the probability value corresponding to *t* and when the *Sig.* value was less than 0.05, independent variable was considered as significant; The partial correlation referred to the correlation between one variable that excludes the influence of other independent variables and *Y* that excludes the part that other independent variables can explain; Tolerance was used to test multicollinearity between variables, and when it was less than 0.1, there was multicollinearity between variables; *S.E* represented standard error and reflected the degree of dispersion between sample means.

It can be seen from Table 4 that except for outdoor temperature, the Sig. values of all other variables were greater than 0.05, meaning that the paramount environmental factor influencing occupant window behavior was outdoor temperature for the selected dataset. This could be explained by the use of natural ventilation, which is highly dependent on the outdoor conditions. However, in many existing studies, indoor temperature has been identified as another factor influencing occupant window behavior in buildings [52-55,57-58,60-65]. Therefore, indoor temperature has been used to train the model as well. When using different independent variables, the main parameters of the trained Gauss distribution models are listed in Table 5. And the test results from using different independent variables as input parameters for Gauss distribution models to predict window state are as shown in Table 6.

Table 5. Corresponding mean value and variance of Gauss distribution models.

Model	μ	σ^2
Gauss dist. (with θ_{in})	28	8
Gauss dist. (with θ_{out})	32	10
Gauss dist. (with θ_{in} and θ_{out})	28/32	8/10

Table 6. Predicting results of Gauss distribution model based on validation datasets.

Model	TPR(%)	TNR(%)	ACC(%)
Exact	100.0	100.0	100.0
Gauss dist.(with θ_{in})	32.1	66.6	58.3
Gauss dist.(with θ_{out})	14.2	93.0	74.1
Gauss dist.(with θ_{in} and θ_{out})	15.8	87.8	70.5

where θ_{in} refers to indoor temperature and θ_{out} refers to outdoor temperature.

3.2 Model Validation

The validation of the models is based on the validation datasets. Table 6 shows the prediction performance of the Gauss distribution models discussed above when using different input parameters. Outdoor temperature is still shown to be the most significant input parameter, giving the highest prediction accuracy, i.e. ACC equals to 74.1%, comparing to 58.3% when using indoor temperature only and 70.5% when using both indoor and outdoor temperatures.

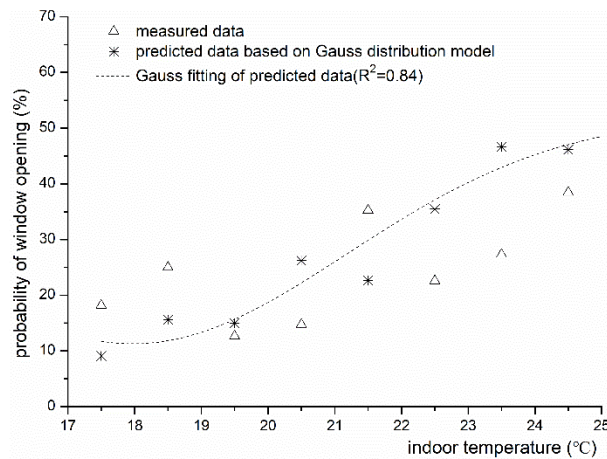


Fig. 7 Results of Gauss distribution model using θ_{in} as input

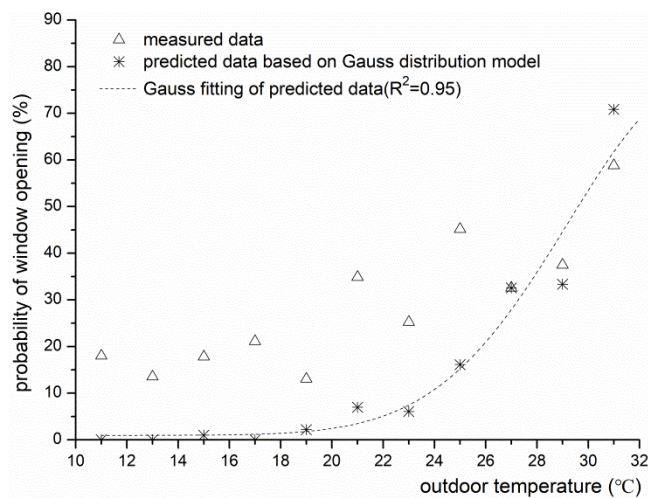
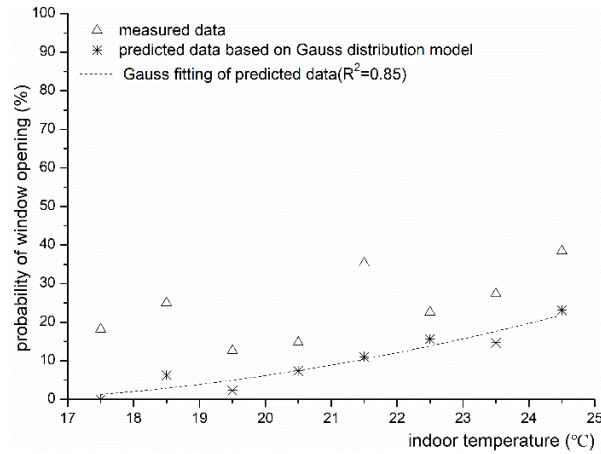
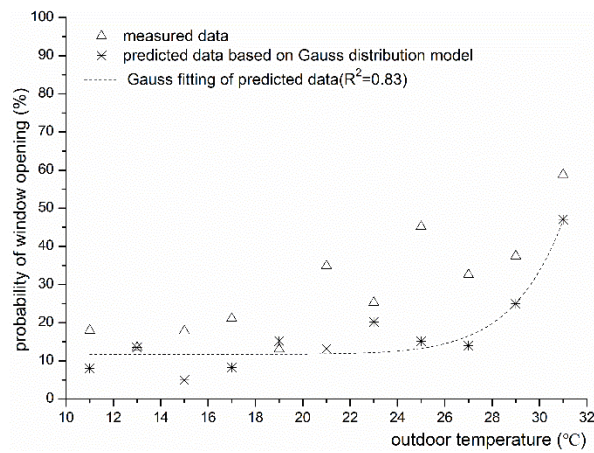


Fig. 8 Results of Gauss distribution model using θ_{out} as input



(a)



(b)

Fig. 9 Results of Gauss distribution model using both θ_{in} (a) and θ_{out} (b) as inputs

Fig. 7 and Fig. 8 show the results of Gauss distribution models whose input variables are indoor temperature and outdoor temperature, respectively. When the inputs are two variables, the probability of window opening for each variable is calculated and exhibited in Fig. 9 (a) and (b). It can be clearly seen that aiming at this measurement of window behavior during transitional seasons, the predicted probability of window opening shows a similarly increasing tendency as the outdoor or indoor temperature, which is similar with measured probability. Compared with the results of Fig. 8 and Fig. 9, the results of Fig. 7 seem to show better fitting. One possible reason for this is that Gauss distribution model whose input variable is indoor temperature has a

highest true positive rate than the other two models. In addition, Fig. 9(a) and Fig. 9(b) show that though there is the same tendency of probability of window opening, increasing with outdoor/indoor temperature, the predicted probability are generally lower than measured values. One possible reason is that the model predicting window state based on two variables (indoor and outdoor temperatures), hence the statistical relationship between predicting result and each single variable is weak. However, we can conclude that the predicted probability of window opening can track the measured data fairly well. Gauss distribution model has been validated as a qualified model for window state prediction.

3.3 Comparison with Logistic regression

For comparison, the same input parameters have been applied to the training dataset and corresponding Logistic regression models were developed, with key values listed in Table 7.

Table 7. Coefficients of Logistic regression models.

Model	a_1 for θ_{in}	a_2 for θ_{out}	b
Logit dist. (with θ_{in})	0.037	0	-1.913
Logit dist. (with θ_{out})	0	0.0317	-1.688
Logit dist. (with θ_{in} and θ_{out})	-0.009	0.33	-1.535

where a_1 and a_2 refer to the coefficients for indoor temperature and outdoor temperature, respectively, and b is a constant.

Table 8. Predicting results of logistic regression model based on validation datasets.

Model	TPR(%)	TNR(%)	ACC(%)
Exact	100.0	100.0	100.0
Logit dist. (with θ_{in})	21.7	73.8	61.3
Logit dist. (with θ_{out})	28.3	76.4	64.6
Logit dist. (with θ_{in} and θ_{out})	27.1	75.4	64.1

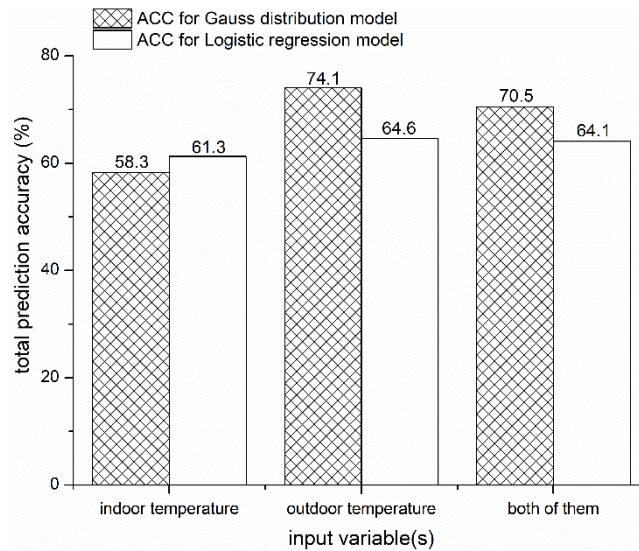


Fig. 10 Comparison of results between Gauss Distribution models and Logistic regression models when using different input parameters

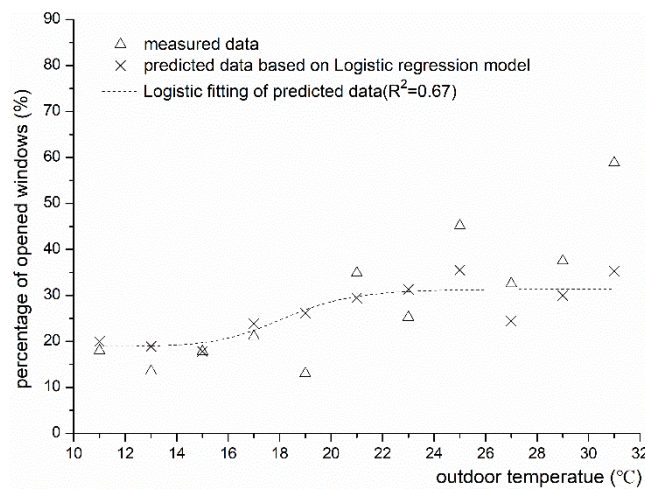


Fig. 11 Results of Logistic regression model using θ_{out} as input

Table 8 shows the prediction results when using the Logistic regression models developed for comparison. Same as Gauss distribution model, the ACC is highest for Logistic regression models when using only one input parameter (outdoor temperature), which can be easily seen in Fig. 10. Fig. 11 shows the results of Logistic regression model using optimum input (outdoor temperature). It can be seen from Fig. 11 that the performance of Logistic regression models is satisfying when outdoor temperature is low. However, the predicted accuracy is decreased with the increase of outdoor temperature. In addition, when comparing the ACC listed in Table 6 and Table 8, it could be seen that when the influential factor was appropriately selected, the Gauss distribution model gave a more accurate prediction (74.1% of ACC) for window states, 9.5% higher than Logistic regression model (64.6% of ACC) when using outdoor temperature as input. However, Gauss distribution models seem to be more sensitive to input variables than Logistic regression models because the predicted accuracy of Gauss distribution models is varied from 58.3% to 74.1% while Logistic regression models give smaller variance of predicted accuracy of 61.3% to 64.6%. Therefore, the selection of input variables is significant to Gauss distribution models. It is observed that when suitable input parameters are selected, Gauss distribution models provide more accurate prediction of occupants' window behavior than Logistic regression models.

4 Conclusions and discussion

Window behavior has significant influences on indoor air quality, energy consumption, and thermal comfort, especially for naturally ventilation buildings. This study provided an exploration of using Gauss distribution for modeling occupant window behavior and compared it with the more conventional method, i.e. Logistic

regression analysis. The actual behavioral data for two transitional seasons collected from a field measurement were used to model validation and comparison. The main conclusions of this paper are summarized as follows:

1. Three sets of input parameters, indoor temperature only, outdoor temperature only and a combination of indoor and outdoor temperatures, have been used as input parameters to model window behavior. According to the results, the Gauss distribution models with outdoor temperature only give best prediction accuracy.
2. Gauss distribution model shows a better prediction performance than Logistic regression model when input parameters are selected carefully. For this study, when outdoor temperature is selected as input, the Gauss distribution model has 9.5% higher prediction. However, Gauss distribution models seem to be more sensitive to input variables than Logistic regression models, the selection of input variables is significant to Gauss distribution models. It is observed that when suitable input parameters are selected, Gauss distribution models provide more accurate prediction of occupants' window behavior than Logistic regression models.

Although the Gauss distribution model is validated as a qualified method to predict the occupants' window behavior, it should be pointed out that the field measurement is not comprehensive. For example, the indoor CO₂ concentration which is identified as the key parameter in existing studies has not been included in this study. A similar study conducted with comprehensive sensor network is necessary in the future work. In addition, we only analyze the window opening state in this study, the window control actions (opening and closing) merits further investigation in following study. What's more, occupants potentially exhibit a tendency to adjust the window depending on the thermal condition. Hence, the deviation from thermal comfortable temperature remains to be analyzed for window behavior modeling based on Gauss

distribution. Also, comparison of the model performance between Gauss distribution model and other existing models would be conducted in the future work.

Acknowledgements

This work was supported by “the 13th Five-Year” National Science and Technology Major Project of China (Grant No. 2017YFC0702202), National Natural Science Foundation of China (Grant No. 51578011), Beijing Natural Science Foundation (Grant No. 3172041) and Beijing municipal education commission.

References

- [1] International Energy Agency, Total energy use in buildings: analysis and evaluation methods, 2013. [http://www.iea-ebc.org/fileadmin/user_upload/images/Pictures/EBC Annex 53 Main Report.pdf](http://www.iea-ebc.org/fileadmin/user_upload/images/Pictures/EBC%20Annex%2053%20Main%20Report.pdf).
- [2] D. Yan, T. Hong, IEA EBC Annex 66: Definition and Simulation of Occupant Behavior in Buildings, 2014. <http://www.annex66.org/>.
- [3] Belafi ZD, Naspi F, Arnesano M, Reith A, Revel GM, Investigation on window opening and closing behavior in schools through measurements and surveys: A case study in Budapest, Building and Environment (2018), doi: 10.1016/j.buildenv.2018.07.022.
- [4] Emery A F, Kippenhan C J. A long term study of residential home heating consumption and the effect of occupant behavior on homes in the Pacific Northwest constructed according to improved thermal standards[J]. Energy, 2006, 31(5):677-693.
- [5] H.B. Rijal, P. Tuohy, M.A. Humphreys, et al. Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings[J].

- Energy & Buildings, 2007, 39(7):823-836.
- [6] Andersen R V, Toftum J, Andersen K K, et al. Survey of occupant behaviour and control of indoor environment in Danish dwellings[J]. Energy & Buildings, 2009, 41(1):11-16.
- [7] Andersen R V, Toftum J, Andersen K K, et al. Survey of occupant behaviour and control of indoor environment in Danish dwellings[J]. Energy & Buildings, 2009, 41(1):11-16.
- [8] Hu S , Yan D , Guo S , et al. A survey on energy consumption and energy usage behavior of households and residential building in urban China[J]. Energy & Buildings, 2017, 148.
- [9] Takasu M., Ooka R., Rijal H.B., Indraganti M., Singh M.K. (2017), Study on adaptive thermal comfort in Japanese offices under various operation modes[J]. Building and Environment, 2017, 118(6), 273-288.
- [10] Wei Y, Zhang X, Shi Y, et al. A review of data-driven approaches for prediction and classification of building energy consumption[J]. Renewable & Sustainable Energy Reviews, 2018, 82:1027-1047.
- [11] Xin Z . Data Analysis and Modeling of Lighting Energy Use in Large Office Buildings[J]. Energy & Buildings, 2015, 86:275-287.
- [12] State and Local Energy Efficiency Action Network, Evaluation, Measurement, and Verification of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations, 2012. <http://behavioranalytics.lbl.gov>.
- [13] Sun K, Hong T. A simulation approach to estimate energy savings potential of occupant behavior measures[J]. Energy & Buildings, 2017, 136:43-62.
- [14] Diao L, Sun Y, Chen Z, et al. Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and

- stochastic simulation[J]. *Energy & Buildings*, 2017, 147:47-66.
- [15] D'Oca S, Fabi V, Corgnati S P, et al. Effect of thermostat and window opening occupant behavior models on energy use in homes[J]. *Building Simulation*, 2014, 7(6):683-694.
- [16] D'Oca S. Influence of occupants' behaviour on heating energy consumption and thermal comfort in residential buildings.[J]. 2012.
- [17] Zhou X , Yan D , Feng X , et al. Influence of household air-conditioning use modes on the energy performance of residential district cooling systems[J]. *Building Simulation*, 2016, 9(4):429-441.
- [18] Hong T, Taylor-Lange S C, D'Oca S, et al. Advances in research and applications of energy-related occupant behavior in buildings ☆[J]. *Energy & Buildings*, 2016, 116:694-702.
- [19] Schakib-Ekbatan K, Çakıcı F Z, Schweiker M, et al. Does the occupant behavior match the energy concept of the building? – Analysis of a German naturally ventilated office building[J]. *Building & Environment*, 2015, 84(10):142-150.
- [20] Daniel L, Soebarto V, Williamson T. House energy rating schemes and low energy dwellings: The impact of occupant behaviours in Australia[J]. *Energy & Buildings*, 2015, 88(88):34-44.
- [21] Wang L, Greenberg S. Window operation and impacts on building energy consumption[J]. *Energy & Buildings*, 2015, 92:313-321.
- [22] Santin O G, Itard L, Visscher H. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock[J]. *Energy & Buildings*, 2009, 41(11):1223-1232.
- [23] Annex 66 Final Report: Definition and Simulation of Occupant Behavior in Buildings. <https://annex66.org/?q=Publication/2018FinalReport/>, 2018 (accessed

May 2018)

- [24] V. Fabi, R. Andersen, S. Corgnati, Description of occupant behavior in building energy simulation: state-of-art and concepts for improvements, *Build. Simul.* 2011 12th Conf. Int. Build. Perform. Simul. Assoc. Syd. 14-16 Novemb. (2011) 2882-2889.
- [25] D. Yan, et al., "Occupant behavior modeling for building performance simulation: current state and future challenges, *Energy Build.* 107 (2015) 264-278.
- [26] M. Schweiker, Understanding occupants' behaviour for energy efficiency in buildings, *Curr. Sustain. Energy Rep.* (2017) 1-7.
- [27] T. Hong, S.C. Taylor-Lange, S. D'Oca, D. Yan, S.P. Corgnati, Advances in research and applications of energy-related occupant behavior in buildings, *Energy Build.* 116 (2016) 694-702.
- [28] Nord N, Tereshchenko T, Qvistgaard L H, et al. Influence of Occupant Behavior and Operation on Performance of a Residential Zero Emission Building in Norway[J]. *Energy & Buildings*, 2018, 159.
- [29] Vasilyev A, Yarmoshenko I. Effect of energy-efficient measures in building construction on indoor radon in Russia[J]. *Radiation Protection Dosimetry*, 2016, 174(3).
- [30] Shi G Z. Energy-Efficient Measures in Architecture and its Application[J]. *Journal of Anyang Institute of Technology*, 2006.
- [31] Yong-Zheng Y I. Energy-efficiency design and measures of architecture[J]. *Shanxi Architecture*, 2007.
- [32] Saari A, Kalamees T, Jokisalo J, et al. Financial viability of energy-efficiency measures in a new detached house design in Finland[J]. *Applied Energy*, 2012, 92(2):76-83.
- [33] Heravi G, Qaemi M. Energy performance of buildings: The evaluation of design

and construction measures concerning building energy efficiency in Iran[J]. *Energy & Buildings*, 2014, 75(75):456-464.

[34] Vasilyev A, Yarmoshenko I. Effect of energy-efficient measures in building construction on indoor radon in Russia[J]. *Radiation Protection Dosimetry*, 2016, 174(3).

[35] Huang Y, Niu J L, Chung T M. Study on performance of energy-efficient retrofitting measures on commercial building external walls in cooling-dominant cities[J]. *Applied Energy*, 2013, 103(1):97-108.

[36] Vine E L, Craig P P, Cramer J C, et al. The applicability of energy models to occupied houses: Summer electric use in Davis. *Energy*, 1982, 7(11):909-925.

[37] Yi Jiang, Qingpeng Wei, Xiu Yang. Results based on data-scientific development of building energy conservation. *Construction technology*, 2009, 7:20-24.

[38] Yi Jiang, Xiu Yang. The state of building energy consumption and problems in building energy conservation in China. *China construction*, 2006, 2:12-18.

[39] Yu Liu. Classification and system development of green building tools [J]. *Journal of architecture*, 2006, 07: 36-40.

[40] DB. Crawley, JW. Hand. Contrasting the Capabilities of Building Energy Performance Simulation Programs [J], *Building and Environment*, 2008, 43(4): 661-673.

[41] P. de Wilde, The gap between predicted and measured energy performance of buildings: a framework for investigation, *Autom. Constr.* 41 (2014) 40-49.

[42] C. Menezes, A. Cripps, D. Bouchlaghem, R. Buswell, Predicted vs. actual energy performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap, *Appl. Energy* 97 (2012) 355-364.

[43] C.J. Andrews, D. Yi, U. Krogmann, J.A. Senick, R.E. Wener, *Designing*

- buildings for real occupants: an agent-based approach, *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* 41 (2011) 1077–1091, <http://dx.doi.org/10.1109/TSMCA.2011.2116116>.
- [44] R.J. Meyers, E.D. Williams, H.S. Matthews, Scoping the potential of monitoring and control technologies to reduce energy use in homes, *Energy Build.* 42 (2010) 563–569, <http://dx.doi.org/10.1016/j.enbuild.2009.10.026>.
- [45] Wang L, Greenberg S. Window operation and impacts on building energy consumption[J]. *Energy & Buildings*, 2015, 92:313-321.
- [46] Wang L. Window Operation and Its Impacts on Thermal Comfort and Energy Use[M]// *Proceedings of the 8th International Symposium on Heating, Ventilation and Air Conditioning*. Springer Berlin Heidelberg, 2014:259-266.
- [47] I. Gaetani, P.J. Hoes, J.L.M. Hensen, Occupant behavior in building energy simulation: towards a fit-for-purpose modeling strategy, *Energy Build.* 121 (2016)188-204.
- [48] H. Burak Gunay, William O'Brien, Ian Beausoleil-Morrison, A critical review of observation studies, modeling, and simulation of adaptive occupant behaviours in offices, *Build. Environ.* 70 (December 2013) 31-47, <http://dx.doi.org/10.1016/j.buildenv.2013.07.020>.
- [49] Simona D'Oca, Tianzhen Hong, A data-mining approach to discover patterns of window opening and closing behavior in offices, *Build. Environ.* 82 (2014) 726e739, <http://dx.doi.org/10.1016/j.buildenv.2014.10.021>.
- [50] Seyed Amir Sadeghi, Panagiota Karava, Iason Konstantzos, Athanasios Tzempelikos, Occupant interactions with shading and lighting systems using different control interfaces: a pilot field study, *Build. Environ.* 97 (15 February 2016) 177-195, <http://dx.doi.org/10.1016/j.buildenv.2015.12.008>.

- [51] Siddharth Goyal, Prabir Barooah, Timothy Middelkoop, Experimental study of occupancy-based control of HVAC zones, *Appl. Energy* 140 (15 February 2015) 75e84, <http://dx.doi.org/10.1016/j.apenergy.2014.11.064>.
- [52] R. Fritsch, A. Kohler, M. Nygård-Ferguson, J.-L. Scartezzini, A stochastic model of user behaviour regarding ventilation, *Build. Environ.* 25 (1990) 173-181.
- [53] Rijal HB, Tuohy P, Humphreys MA, Nicol JF, Samuel A, Clarke J. 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):823-36.
- [54] Rijal HB, Tuohy P, Nicol JF, Humphreys MA, Samuel A, Clarke J. Development of an adaptive window-opening algorithm to predict the thermal comfort, energy use and overheating in buildings. *Journal of Building Performance Simulation* 2008; 1(1):17-30.
- [55] F. Haldi, D. Robinson, Interactions with window openings by office occupants, *Build. Environ.* 44 (2009) 2378-2395.
- [56] Y. Zhang, P. Barrett, Factors influencing the occupants' window opening behaviour in a naturally ventilated office building, *Build. Environ.* 50 (2012) 125-134.
- [57] S. Wei. Preference-based modelling and prediction of occupants window behaviour in non-air-conditioned office buildings[J]. Loughborough University, 2013.
- [58] R. Andersen, V. Fabi, J. Toftum, S.P. Corgnati, B.W. Olesen, Window opening behaviour modeled from measurements in Danish dwellings, *Build. Environ.* 69 (2013) 101-113.
- [59] Li N, Li J, Fan R, et al. Probability of occupant operation of windows during transition seasons in office buildings[J]. *Renewable Energy*, 2015, 73:84-91.
- [60] Moghadam S T, Soncini F, Fabi V, et al. Simulating Window Behaviour of Passive and Active Users ☆[J]. *Energy Procedia*, 2015, 78:621-626.

- [61] Pan S, Xiong Y, Han Y, et al. A study on influential factors of occupant window-opening behavior in an office building in China[J]. *Building & Environment*, 2018.
- [62] Naspi F, Arnesano M, Zampetti L, et al. Experimental study on occupants' interaction with windows and lights in Mediterranean offices during the non-heating season[J]. *Building & Environment*, 2017, 127.
- [63] Calì D, Andersen R K, Müller D, et al. Analysis of occupants' behavior related to the use of windows in German households[J]. *Building & Environment*, 2016, 103:54-69.
- [64] Jones R V, Fuertes A, Gregori E, et al. Stochastic behavioural models of occupants' main bedroom window operation for UK residential buildings[J]. *Building & Environment*, 2017, 118:144-158.
- [65] Yao M, Zhao B. Window behavior of occupants in residential buildings in Beijing[J]. *Building & Environment*, 2017, 124.
- [66] Stazi F, Naspi F, D'Orazio M. Modeling window state in school classrooms. Results from a case study in Italy[J]. *Building & Environment*, 2017, 111:24-32.
- [67] Wu X J, Dong Y. Identify of Runoff Influence Factors in Coal Mining Area Based on Stepwise Regression Analysis[C]// *International Forum on Energy, Environment Science and Materials*. 2018.
- [68] Luo J, Luo J H, Zhou Y M, et al. Stepwise Regression Analysis on the Factors Influencing the Life Quality of the Elderly of the Va Nationality in Rural Areas[J]. *Journal of Kunming Medical University*, 2017.
- [69] Ma W, Tan S, Hei X, et al. A Prediction Method Based on Stepwise Regression Analysis for Train Axle Temperature[C]// *International Conference on Computational Intelligence and Security*. IEEE, 2017:386-390.
- [70] You S, Yan Y. Stepwise Regression Analysis and Its Application[J]. *Statistics &*

Decision, 2017.

[71] F. Haldi, D. Robinson, Interactions with window openings by office occupants, *Build. Environ.* 44 (12) (2009) 2378e2395.

[72] M. Schweiker, F. Haldi, M. Shukuya, D. Robinson, Verification of stochastic models of window behavior for residential buildings, *J. Build. Perform. Simul.* 5 (1) (2012) 55-74.