This document is the accepted manuscript version of the following article:
Fu, Y., Zhou, T., Lun, I., Khayatian, F., Deng, W., & Su, W. (2021). A data-driven approach for window opening predictions in non-air-conditioned buildings. Intelligent Buildings International, 1-17.
https://doi.org/10.1080/17508975.2021.1963651

A data-driven approach for window opening predictions in non-air-conditioned buildings

Abstract: In non-air-conditioned buildings, opening or closing of windows is one of the most common behaviours that occupants tend to carry out to restore their thermal comfort. However, occupant behaviour is a sophisticated concept and is an integration of physical environment, thermal sensation etc. that varies from one person to another. This work, therefore, attempts to predict the occupant behaviours in terms of window openings based on thermal comfort through a data-driven machine learning approach. The training set is composed of the key weather information, the main characteristics of buildings, the Adaptive Predicted Mean Vote (APMV) values on thermal comfort and its corresponding window status. Building simulation results of 95 cities in China covering all the climate zones, in total 828,360 groups are adopted. The predictor achieves a high accuracy of approximately 95%, and therefore enables the users to estimate the window openings directly from weather conditions and building characteristics. As an original contribution, the study shows that conditioned to availability of adequate simulation data, a machine learning predictor trained solely on simulation data can accurately predict realistic window opening behaviours, without relying on any indoor measurement.

Keywords: occupant behaviour; thermal comfort; machine learning; building simulation; natural ventilation

1. **Introduction**

Worldwide, people spend around 80% of their time indoors (Zhao, Sun and Ding, 2004). Therefore, with the need to improve the thermal environment and to reduce building energy consumption, numerous indoor thermal comfort studies have been carried out(von

Grabe, 2016; Deng and Chen, 2018; Wu *et al.*, 2018). It has been found that people are not inert recipients of the environment, but do interact with building facilities to optimise their comfort (Nicol and Michael, 2004). People usually respond to their thermal state by either changing the environmental conditions when they are uncomfortable or keeping the environment unaltered when they are comfortable. For instance, Brager et al. (2004) stated that in non-air-conditioned buildings, people primarily use the action of opening/closing windows to retain thermal comfort because it has an immediate effect on changing the indoor thermal environment. It is worth to highlight that window opening behaviour is a result of the continuous combination of many factors, such as weather condition (e.g. air temperature, humidity etc.), building features (e.g. orientation, window size etc.) and even occupant characteristics (e.g. personal background, psychology etc.) (Fabi *et al.*, 2012) .

Plenty of works have been devoted to investigating the key influential factors, especially the weather conditions. Herkel et al. (2008) studied 21 south-facing offices in Germany over one year and found that window opening behaviours had a strong correlation with the indoor and outdoor temperature. Meanwhile, solar radiation has little correlation with window opening as compared with indoor and outdoor temperature. Haldi and Robinson (2009) produced a field study of a south-facing cellular office building for seven years to investigate the relationship between window opening behaviours and outdoor weather. Based on the result, they concluded that wind speed was reversely proportional to the opening of windows although the wind direction had no direct correlation to window opening behaviour. It differs from D'Oca and Hong (2014) who studied an office building in Germany and found that wind direction had a certain influence on the window opening. Dutton and Shao (2010) reported a study of a naturally ventilated elementary school in the UK, and found that besides the aforementioned factors, atmospheric pressure also had an impact on the indoor thermal comfort, hence, the action of opening the window. A summary of field studies about the window opening behaviours is shown in Table 1. It reveals that outdoor temperature and wind are the most investigated variables. It is also observed that most of the field studies in this area were constrained by limited on-site measurements.

Recently, data-driven approaches and computational statistics like machine learning have been widely applied to building research where abundant data is available. Khayatian et al. (2016) used Artificial Neural Network (ANN) to evaluate energy performance certificates of residential buildings. They trained a set of 100 models with 187587 entries and 12 variables to obtain confidence interval, expressing that 95% of entries fall within ±3 standard deviation in a confidence interval. Wu et al. (2018) produced a field study on 24 dormitory buildings which include both non-air-conditioned and split air-conditioning buildings, and resorted to Bagging, an ensemble learning algorithm, to predict thermal comfort in buildings. They found that Bagging returned the best Predicted Mean Vote (PMV) prediction with a coefficient of determination (R^2) of 0.99, surpassing ANN and support vector machine. Markovic et al. (2018) used a deep learning architecture with 5 hidden layers to predict window opening behaviour based on a more than 3100 hours dataset that collected from 3 independent buildings in Aachen, Frankfurt and Philadelphia. The model was tested on a building simulation program and achieved a final accuracy ranging from 86% to 89%.

Deng and Chen (2018) also used ANN to predict thermal comfort based on data collected from 10 offices, 6 apartments and 4 houses in the USA. They found that the comfortable air temperature in apartments/houses was 1.7 °C lower than that of offices, and the comfort zone obtained from ANN was narrower than that of ASHRAE Standard 55.

In many of these investigations, for non-air-conditioned buildings, there is a common agreement that indoor thermal comfort is the key driver for the window opening behaviour especially in transition seasons (Brager, Paliaga and de Dear, 2004); and indoor thermal comfort is closely related to the outdoor conditions. Among many thermal comfort prediction models, the PMV (Predicted Mean Vote) is a widely used rational approach. However, PMV model is unable to account the effect of occupants' adaptions to the thermal environment(Van Hoof, 2008), and it has been sometimes reported to overestimate thermal discomfort (Manu *et al.*, 2016) . By taking the advantages of PMV and adaptive approaches, the adaptive PMV (APMV) has been developed by researchers (Yao, Li and Liu, 2009). The APMV takes into account the four environmental factors, the two occupant-related factors and their adaptations. Due to its convenience of calculations, APMV has been an effective index in practice to evaluate the thermal comfort, especially in non-air conditioned buildings, in different countries and regions (Conceição *et al.*, 2012)(Kim *et al.*, 2015)(Costanzo *et al.*, 2019). Despite the existing distinction between the prediction and reality, it is worth mentioning that the APMV model yields reliable comparative results, i.e., it is reliable to estimate the relative effect of changing a condition to the thermal comfort level. Therefore, APMV is often used as an essential factor in studying the thermal environment changes and the induced occupant behaviours, or vice versa (Vallianos, Athienitis and Rao, 2019). Since the year 2012, the APMV also has been adopted in the Chinese standard of thermal comfort evaluation of civil buildings (GB/T50785-2012) (China Academy of Building Research and Chongqing University, 2012), in where the adaptive coefficients for different climate zones in China are provided. Therefore, it would be of interest to us to find out the trends of window openings in various climatic conditions in China.

Nowadays, with building simulation programs, the prediction of indoor thermal comfort levels in terms of APMV etc., can be obtained with satisfying accuracy. However, relying on building simulation programs alone makes it difficult to predict particular human behaviours, i.e. window opening, or seeking correlations between window opening behaviour patterns and its influencing factors. On the other hand, machine learning is a promising approach that can make predictions based on a training dataset by possessing abundant representative data.

Therefore, the aim of this work is to study the correlation of weather condition and building characteristics with the window opening behaviour under a variety of climatic regions. Although it is practically impossible to conduct large scale on-site surveys and collect real data covering hundreds of cities in China, we proposed a new method which is able to generate reliable comparative results from simulation reflecting the changes in weather conditions and building characteristics. To this end, a data-driven predictor is developed by using the simulation results from 95 cities covering all five climatic regions in China. The introduced framework enables discovering the patterns of window opening behaviour in non- air-conditioned buildings in transient seasons. To assess the viability of the proposed method, the framework is applied to a case study of 200 residential units with field observations. As an original contribution, the study shows that conditioned to availability of adequate simulation data, a machine learning predictor trained solely on simulation data can accurately predict realistic window opening behaviours, without relying on any indoor measurement. This is of great importance, specifically for residential buildings, where privacy concerns are the main challenge for collecting data, and resorting to simulations is the only viable option for large-scale prediction of window operation.

2. Methodology

The current study includes building simulation, identification of window opening states, data training and validation. A well-recognised building simulation tool, namely Integrated Environmental Solutions - Virtual Environment (IES-VE), is used to simulate the indoor environment. Weather data and building characteristics are taken as input feature of the training set; whilst the corresponding window states is used as the target feature.

2.1. Building model and variable selection

According to ANSI/ASHRAE Standard 169-2013 and Building Design Standard in China GB50189-2015, as shown in Figure 1, in total 95 cities are selected from five climate zones in China, i.e. Severe Cold, Cold, Hot summer and cold winter, Hot summer and warm winter and Temperate zone. Transition season periods of each climate zone is adopted from

Regional Buildings Design Standards in China (JGJ134-2010, JGJ26-2018, JGJ75-2012 and JGJ475-2019).

In non-air-conditioned buildings, local climate and building parameters play main roles in indoor thermal conditions (Raja, Nicol and McCartney, 1998). Table 2 summarises the key weather features and building parameters extracted from previous studies (Herkel, Knapp and Pfafferott, 2008; Haldi and Robinson, 2009; Dutton and Shao, 2010; D'Oca and Hong, 2014; Shi *et al.*, 2018; Zhou *et al.*, 2018).

To minimise the possible bias due to the complexity of building types, we decided to use the same simple building model for all the cities. The building model has a dimension of 8m * 8m * 3m, as shown in Figure 2. It has four 4m * 4m rooms, and each room has one window facing east, south, west or north, respectively. The model is assumed to be located at a middle floor of a building, and there is no heat transfer with adjacent levels through floor or ceiling. In this work, the dimension of the model remains constant for all 5 climatic zones. However, different building characteristics are chosen to meet the requirements of the local building design standards, corresponding to the climatic differences. The key building characteristics including window orientation, glazing ratio, U-values of wall and window, shading coefficient, internal gain, and air exchange rate vary with different climate zones and are tabulated in Table 3.

2.2. Identification of window opening states

In non-air-conditioned buildings, people tend to restore the thermal comfort by changing the window opening states (i.e. from close to open or from open to close) if they feel uncomfortable, and retain existing state if feeling comfortable (Andersen *et al.*, 2009). In this study, the thermal comfort is described by APMV in where the perception of comfort level is defined from cold (-3) to neutral (0) to hot (+3). The APMV is proportional to the PMV value as given:

$$\mathrm{APMV} = \mathrm{PMV}/(1 + \lambda \times \mathrm{PMV}) \quad (1)$$

where the adaptive coefficient λ is designated based on China's context (Wang *et al.*, 2018), as shown in Table 4. If the APMV value is within the range of ±0.5, it can be assumed that people feel comfortable in this condition (ASHRAE, 2013). In addition, the clothing insulation and metabolic rate are assumed as 0.7 clo and 1.2 met, respectively, in accordance with the typical garment in transition seasons and sedentary activity.

As illustrated in Figure 3, each of the simulated cases is run twice with the windows opened and closed at each hour, respectively. After that, the APMV values of two conditions are compared with each other. If both APMV values fall in the range of ±0.5, the window retains the same state of the previous hour (i.e. people tend to retain the state if feel comfortable). Otherwise, the window opening state is changed to the condition that has an APMV value closer to 0 (i.e. people tend to restore the thermal comfort by changing the window opening state). In this work, a total training dataset containing 828,360 groups of window opening samples from 95 cities are examined and formatted by Python.

2.3. Machine learning algorithm

Window opening behaviour could be driven by many factors of dissimilar importance. In this study Gradient Boosting Decision Trees (GBDT) (Friedman, 2001) is selected as the preferred machine learning algorithm, as it is not prone to collinear features. The GBDT is a black-box boosting machine learning algorithm of ensemble learning and is based on the decision tree method. The methodology of the decision tree is one of the most widely used data mining techniques (Quinlan, 1986; Han, Kamber and Pei, 2006). The decision tree algorithm categorises all training set into various classes, thereby the data description and final classification can be provided in a flowchart which looks resembles tree structure. In this work, the training set consists of almost one million samples, and each sample has 15 input features and 1 target feature as shown in Table 2. These samples are used to train the decision tree model to find the relationship between the features and the window opening conditions.

A decision tree has a generation algorithm, and it has three modes: ID3, C4.5 as well as classification and regression trees (CART). CART adopts recursive binary partitions to classify or regress all the training data and is the main regressor of GBDT. Each CART can be seen as a weak learner, and GBDT sequentially combines all weak learners to reduce prediction errors and force them to become a strong learner.

The importance of each feature in GBDT model can be then calculated. The global importance of feature *j* is given by:

$$J_j^2 = \frac{1}{M}\sum_{m=1}^{M} J_j^2(T_m) \qquad (1)$$

where, M is the number of trees model, and the importance of feature j at one tree is given by:

$$J_j^2(T) = \sum_t^{L-1} i_t^2(v_t = j) \qquad (2)$$

where, L is the number of leaf nodes, thus $\mathrm{L} - 1$ is the number of none-leaf nodes, v_t is the node of feature j and i_t^2 is the decreasing value of square loss.

3. Results and discussion

3.1. Predictor performance

As a well-accepted technique for evaluating predictive models, the 10-fold cross-validation method (Kohavi, 1995) is used to assess the accuracy of the trained model. In figure 4, the whole training set is shuffled at the beginning, of which 10% is randomly taken as testing set while the remaining data constitutes the training set. This process is repeated ten times, and then the mean value is taken as the final accuracy. All of the models have been properly tuned based on the parameter learning curve figure to avoid overfitting and unnecessary calculation.

Figure 4 describes the accuracy changes with proportions of input samples, where the final accuracy is around 95.4%. It can be seen that the learning curve is convergent; hence this GBDT model has a good bias-variance trade-off without overfitting. As shown in Figure 5, among the 82836 testing samples (10% of the total training set), only 1748 samples are incorrectly predicted as “closing of the window” and 2046 samples are wrongly labelled as "opening of the window". The rest of the 79042 samples (i.e. 95.4%) are correctly classified, which implies that the GBDT model performs well in the testing set.

3.2. Importance of features

Figure 6 shows the importance of the features. In general, the weather plays a much more crucial role in window opening behaviour than the building parameters in non-air-conditioned buildings. The outdoor temperature reaches over significant high importance of 84.23% among all features, followed by solar radiation 4.65%, humidity 3.44% and atmospheric pressure 3.03%, etc. The window orientation 0.52%, window area 0.45%, U-value of window 0.45% and wall area 0.22% are the four most important features among building parameters.

The importance of various features is further investigated under different climate zones in Table 5. It shows that apart from the outdoor temperature, the importance of other features varies under different climate zones. The calculated ranking agrees well with previous findings in the literature. For instance, Zhang and Barrett (2012) conducted a field measurement in Sheffield city, northern England, and their results indicated that outdoor temperature has crucial role among all the weather features, along with solar radiation, humidity and wind velocity. These studies have feature importances similar to the ones reported in this work, implying the feasibility and reliability of the developed predictor.

It is worth mentioning that wind velocity in hot summer and cold winter climate zone is ranked the second most influential factor, showing a more critical role than in other climate zones. This is in line with the findings of Haldi and Robinson (2009) and D'Oca and Hong

(2014), in which the significant impact of wind velocity are highlighted on decreasing window opening proportion in Lausanne, Switzerland and Frankfurt, Germany, respectively. These two European cities have a similar climate as the hot summer and cold winter zone in China. Meanwhile, in those places similar to other climate zones in China, the wind speed seems less important to the window opening states. For instance, Andersen et al. (2009) demonstrate that wind speed has no effect on the windows opening behaviour in Denmark, where climatic conditions are similar to the cold climate zone in China.

3.3. Visualisation of the prediction process

Figure 7 shows the distribution of all predicted window states (open/close) with the correlation of outdoor temperature and other important features including solar radiation, humidity, atmospheric pressure, wind velocity and room orientation. A few interesting findings are summarised:

- A clear borderline can be visualised between outdoor temperature and other weather features (i.e. (a) solar radiation, (b) relative humidity, (c) atmospheric pressure and (d) window velocity). When it is on the right-hand side of this borderline, especially when the outdoor temperature is higher than about 21℃, it can be confidently stated that the window will be opened, regardless of other conditions. When the outdoor temperature is lower than approximate 14 ℃, most of the windows will be closed, except for south-facing window, which may remain open till the outdoor temperature falls below 11 ℃, as shown in Figure 7 (e).

- It is found that the humidity and atmospheric pressure are in positive correlation with the window opening, while high solar radiation and wind velocity may decrease the probabilities of opening windows.

The borderlines are in the intermediate temperature range, where the window opening behaviours are difficult to predict precisely with only one feature. Figure 8 depicts the prediction error ratio under different outdoor temperatures in the testing set. It can be seen that the temperatures between 16 °C and 21°C return high error ratios, indicating that opening and closing window conditions have minimum thermal comfort discrimination in this temperature range.

A more detailed correlation of wind direction and velocity to the window opening behaviours of different orientated rooms are illustrated as wind-rose-like diagrams in Figure 9. The length of each sector represents the percentage of the opened windows under different wind directions, when compared to the total number of all opened windows in the entire transition season. These diagrams are useful to provide a general trend, which highlights the likelihood of opening windows under different wind directions in China. It is found that there is a higher tendency to open windows during south and east winds, rather than north and west winds, which is in line with the typical situation in China. Interestingly, it is also found that Figures 9. (a) ~ (d) display a very high similarity, i.e. the room orientation has little impact on the window opening behaviours under the same wind direction.

3.4 Validation of the prediction results

A typical student dormitory (Figure 10) located in Ningbo China was used for the validation of the developed predictor. Ningbo is located on the east coast of China near Shanghai and is categorised in a "hot summer and cold winter" climate zone. The dormitory is assumed as a non-air-conditioned building in transition seasons. The general information and the configuration of the building are shown in Table 6 and Figure 10, respectively. In total 9 groups of data were collected over the period from 15th September 2018 to 15th November 2018, i.e. a typical transition season in Ningbo. The proportion of opened windows, among approximately 200 occupied rooms, was counted manually, and the weather data was obtained from the local weather station in Ningbo, as summarised in Table 7. To ensure only the occupied rooms were counted, data was collected at night when occupants were home and turned on the lighting. Consequently, the solar gain was always equal to zero in the records.

The weather data and building information were fed into the predictor to anticipate the window opening states. The result is shown in Figure 11. In all 9 groups, the anticipated window opening states are in good agreement with the on-site observations, returning a root mean squared error (RMSE) of 12%.

4. Conclusions

Building simulation programs can establish the correlation between physical factors and the indoor environment in a fast manner. With an abundant amount of simulation data, machine learning is a promising approach to investigate more complicated interactions between the occupants and physical environments. This paper used IESVE to simulate window opening behaviour in transition seasons under different climate zones to obtain abundant groups of training data sets. Features of weather data and building characteristics of 95 cities from five different climate zones in China, (in total 828, 360 groups of data) were employed to train the predictor. Due to the diversity and complicated relationship of different features, a boosting machine learning algorithm using GBDT was adopted to investigate the correlation between features and window opening behaviours. Some key findings are summarised as followings:

- The predictor successfully predicted window opening in the testing set with around 95.4% accuracy;
- The importance of each feature to the window opening behaviour patterns was identified. The outdoor temperature has the dominant influence, and other weather features have different feature importance rankings under various climate zones;
- Humidity and atmospheric pressure have a positive correlation with window opening while wind velocity and solar radiation showed a negative correlation with the window opening;
- South and East winds may lead to more window opening in China, when compared to North and West winds, regardless of the room orientation;
- A good agreement was found between the real data and predicted data, returning approximately 12% of RMSE.

The window opening behaviour is a complex combination of many factors which are in interaction. This work provides a promising data-driven approach that enables people to accurately predict realistic window opening behaviours, solely based on adequate simulation data. Meanwhile, the predictor could be integrated into building simulation to reflect the dynamic window opening states. However, some influencing factors like air quality and acoustics are not yet considered in this predictor, which deserve to be investigated in future work. Moreover, more complex personal models, e.g., multi-nodes thermal comfort models, would be implemented in where more on-site measurements are available, to improve the predictor's performance.

Acknowledgements: This work is funded by the Ningbo Natural Science Programme (project codes: 2018A610358 and 2019A610393).

References

Andersen, R. V. *et al.* (2009) 'Survey of occupant behaviour and control of indoor environment in Danish dwellings', *Energy and Buildings*, 41(1), pp. 11–16. doi: https://doi.org/10.1016/j.enbuild.2008.07.004.

ASHRAE (2013) 'Thermal Environmental Conditions for Human Occupancy'.

Brager, G., Paliaga, G. and de Dear, R. (2004) 'Operable Windows, Personal Control,and Occupant Comfort', *ASHRAE Transactions*, 110, p. Part 2. Available at: https://escholarship.org/uc/item/4x57v1pf.

China Academy of Building Research and Chongqing University, B. (2012) (2012) *Chinese Standard GB/T50785-2012 - Evaluation Standard for Indoor Thermal Environments in Civil Buildings*.

Conceição, E. Z. E. *et al.* (2012) 'Application of a developed adaptive model in the evaluation of thermal comfort in ventilated kindergarten occupied spaces', *Building and Environment*. Elsevier Ltd, 50, pp. 190–201. doi: 10.1016/j.buildenv.2011.10.013.

Costanzo, V. *et al.* (2019) 'Natural ventilation potential for residential buildings in a densely built-up and highly polluted environment. A case study', *Renewable Energy*. Elsevier Ltd, 138, pp. 340–353. doi: 10.1016/j.renene.2019.01.111.

D'Oca, S. and Hong, T. (2014) 'A data-mining approach to discover patterns of window opening and closing behavior in offices', *Building and Environment*, 82, pp. 726–739. doi: https://doi.org/10.1016/j.buildenv.2014.10.021.

Deng, Z. and Chen, Q. (2018) 'Artificial neural network models using thermal sensations and occupants' behavior for predicting thermal comfort', *Energy and Buildings*, 174, pp. 587–602. doi: https://doi.org/10.1016/j.enbuild.2018.06.060.

Dutton, S. and Shao, L. (2010) 'Window opening behaviour in a naturally ventilated school', *Proceedings of SimBuild*, 4(1), pp. 260–268.

Fabi, V. *et al.* (2012) 'Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models', *Building and Environment*, 58, pp. 188–198. doi: https://doi.org/10.1016/j.buildenv.2012.07.009.

Friedman, J. H. (2001) 'Greedy Function Approximation: A Gradient Boosting Machine', *The Annals of Statistics*. Institute of Mathematical Statistics, 29(5), pp. 1189–

1232. Available at: http://www.jstor.org/stable/2699986.

von Grabe, J. (2016) ‘Potential of artificial neural networks to predict thermal sensation votes’, *Applied Energy*, 161, pp. 412–424. doi: https://doi.org/10.1016/j.apenergy.2015.10.061.

Haldi, F. and Robinson, D. (2009) ‘Interactions with window openings by office occupants’, *Building and Environment*, 44(12), pp. 2378–2395. doi: https://doi.org/10.1016/j.buildenv.2009.03.025.

Han, J., Kamber, M. and Pei, J. (2006) ‘Data mining: concepts and techniques’, *Elsevier Inc*.

Herkel, S., Knapp, U. and Pfafferott, J. (2008) ‘Towards a model of user behaviour regarding the manual control of windows in office buildings’, *Building and Environment*, 43(4), pp. 588–600. doi: https://doi.org/10.1016/j.buildenv.2006.06.031.

Van Hoof, J. (2008) ‘Forty years of Fanger’s model of thermal comfort: Comfort for all?’, *Indoor Air*, 18(3), pp. 182–201. doi: 10.1111/j.1600-0668.2007.00516.x.

Khayatian, F., Sarto, L. and Dall’O’, G. (2016) ‘Application of neural networks for evaluating energy performance certificates of residential buildings’, *Energy and Buildings*, 125, pp. 45–54. doi: https://doi.org/10.1016/j.enbuild.2016.04.067.

Kim, J. T. *et al.* (2015) ‘Development of the adaptive PMV model for improving prediction performances’, *Energy and Buildings*. Elsevier B.V., 98, pp. 100–105. doi: 10.1016/j.enbuild.2014.08.051.

Kohavi, R. (1995) ‘A study of cross-validation and bootstrap for accuracy estimation and model selection’, in *Ijcai*. Montreal, Canada, pp. 1137–1145.

Li, N. *et al.* (2015) ‘Probability of occupant operation of windows during transition seasons in office buildings’, *Renewable Energy*, 73, pp. 84–91. doi: https://doi.org/10.1016/j.renene.2014.05.065.

Manu, S. *et al.* (2016) ‘Field studies of thermal comfort across multiple climate zones for the subcontinent: India Model for Adaptive Comfort (IMAC)’, *Building and Environment*. Elsevier Ltd, 98, pp. 55–70. doi: 10.1016/j.buildenv.2015.12.019.

Markovic, R. *et al.* (2018) ‘Window opening model using deep learning methods’, *Building and Environment*, 145, pp. 319–329. doi:

https://doi.org/10.1016/j.buildenv.2018.09.024.

Nicol, J. F. and Michael, H. (2004) ‘A Stochastic Approach to Thermal Comfort-Occupant Behavior and Energy Use in Buildings’, *ASHRAE Transactions*, 110, pp. 554–568.

Pan, S. *et al.* (2018) ‘A study on influential factors of occupant window-opening behavior in an office building in China’, *Building and Environment*, 133, pp. 41–50. doi: https://doi.org/10.1016/j.buildenv.2018.02.008.

Quinlan, J. R. (1986) ‘Induction of decision trees’, *Machine Learning*, 1(1), pp. 81–106. doi: 10.1007/bf00116251.

Raja, I. A., Nicol, J. F. and McCartney, K. J. (1998) ‘Natural ventilated buildings: Use of controls for changing indoor climate’, *Renewable Energy*, 15(1), pp. 391–394. doi: https://doi.org/10.1016/S0960-1481(98)00193-1.

Schakib-Ekbatan, K. *et al.* (2015) ‘Does the occupant behavior match the energy concept of the building? – Analysis of a German naturally ventilated office building’, *Building and Environment*, 84, pp. 142–150. doi: https://doi.org/10.1016/j.buildenv.2014.10.018.

Shi, Z. *et al.* (2018) ‘Seasonal variation of window opening behaviors in two naturally ventilated hospital wards’, *Building and Environment*, 130, pp. 85–93. doi: https://doi.org/10.1016/j.buildenv.2017.12.019.

Takasu, M. *et al.* (2017) ‘Study on adaptive thermal comfort in Japanese offices under various operation modes’, *Building and Environment*, 118, pp. 273–288. doi: https://doi.org/10.1016/j.buildenv.2017.02.023.

Vallianos, C., Athienitis, A. and Rao, J. (2019) ‘Hybrid ventilation in an institutional building: Modeling and predictive control’, *Building and Environment*. Elsevier, 166(September), p. 106405. doi: 10.1016/j.buildenv.2019.106405.

Wang, Z. *et al.* (2018) ‘Individual difference in thermal comfort: A literature review’, *Building and Environment*, 138, pp. 181–193. doi: https://doi.org/10.1016/j.buildenv.2018.04.040.

Wu, Z. *et al.* (2018) ‘Using an ensemble machine learning methodology-Bagging to predict occupants’ thermal comfort in buildings’, *Energy and Buildings*, 173, pp. 117–127. doi: https://doi.org/10.1016/j.enbuild.2018.05.031.

Yao, R., Li, B. and Liu, J. (2009) ‘A theoretical adaptive model of thermal comfort – Adaptive Predicted Mean Vote (aPMV)’, *Building and Environment*, 44(10), pp. 2089–2096. doi: https://doi.org/10.1016/j.buildenv.2009.02.014.

Yun, G. Y. and Steemers, K. (2008) ‘Time-dependent occupant behaviour models of window control in summer’, *Building and Environment*, 43(9), pp. 1471–1482. doi: https://doi.org/10.1016/j.buildenv.2007.08.001.

Zhang, Y. and Barrett, P. (2012) ‘Factors influencing the occupants’ window opening behaviour in a naturally ventilated office building’, *Building and Environment*, 50, pp. 125–134. doi: https://doi.org/10.1016/j.buildenv.2011.10.018.

Zhao, R., Sun, S. and Ding, R. (2004) ‘Conditioning strategies of indoor thermal environment in warm climates’, *Energy and Buildings*, 36(12), pp. 1281–1286. doi: https://doi.org/10.1016/j.enbuild.2003.07.002.

Zhou, X. *et al.* (2018) ‘Case study of window operating behavior patterns in an open-plan office in the summer’, *Energy and Buildings*, 165, pp. 15–24. doi: https://doi.org/10.1016/j.enbuild.2018.01.037.

Table 1. Variables for window opening in other studies

	Outdoor temperature	Wind velocity	Wind direction	Global radiation	Atmospheric pressure	Relative humidity	Room Orientation	Data size	Location
Herkel et al. (2008)	●	●						21 offices (13 months)	Germany
Haldi and Robinson (2009)	●	●	●			●		14 offices	Lausanne, Swaziland
D'Oca and Hong (2014)	●	●	●	●		●		16 offices (1 year)	Frankfurt, Germany
Dutton and Shao (2010)	●			●	●			1 school (1 year)	UK
Takasu et al.(2017)	●							5 office buildings	Tokyo and Kanagawa, Japan
Zhang and Barrett (2012)	●	●	●	●		●	●	1 building (5 months)	Sheffield, UK
Pan et al. (2018)	●	●	●	●		●		5 offices (3 and half months)	Beijing, China
Shi et al. (2018)	●	●				●		1 hospital (1 year)	Nanjing, China
Schakib-Ekbatan et al. (2015)	●						●	1 office building (6 years)	Frankfurt, Germany
Zhou et al.(2018)	●	●				●		1 office (1 month)	Nanjing, China
Yun and Steemers (2008)	●						●	6 offices (3 months)	Cambridge, UK
Li et al. (2015)	●							1 building (2 months)	Chongqing, China

Table 1. Utilised features

	Input features	Target feature
Weather	temperature, wind velocity, wind direction, global radiation, atmospheric pressure, relative humidity	Window states (open/close)
Building parameters	window orientation, U value of walls, U value of windows, shading coefficient, floor height, opaque surface area, glazed surface area, average internal gain, infiltration	

Table 2. Building characteristics under different climate zones

Climate zone	Window Orientation	Glazing ratio	U-value of wall (W/m^2K)	U-value of window (W/m^2K)	Total shading coefficient	Average internal gain (W/m^2)	Infiltration (ACH)	Transition season
Hot summer and cold winter	East	0.35	1	2.8	0.31	4.3	1	1st Mar - 15th June & 1st Sept - 30th Nov
	South	0.45			0.34			
	West	0.35			0.31			
	North	0.4			0.457			
Cold	East	0.35	0.7	2.3	0.38	3.8	0.5	16th Mar - 30th June & 1st Sept - 15th Nov
	South	0.5			0.455			
	West	0.35			0.38			
	North	0.3			0.455			
Severe Cold	East	0.3	0.55	1.8	0.35	3.8	0.5	21st Apr - 30th June & 1st Sept - 20th Oct
	South	0.45			0.411			
	West	0.3			0.35			
	North	0.25			0.411			
Hot summer and warm winter	East	0.3	0.7	3.5	0.41	4.2	1	1st Jan - 30th Apr & 1st Nov - 31st Dec
	South	0.4			0.43			
	West	0.3			0.41			
	North	0.4			0.43			
Temperate zone	East	0.35	0.8	2.8	0.38	3.8	1	1st Feb - 30th Nov
	South	0.45			0.456			
	West	0.35			0.38			
	North	0.4			0.456			

Table 3. Adaptive coefficients under different climate zones

Climate zones		Adaptive coefficient (λ)
Cold and Severe Cold	PMV ≥ 0	0.24
	PMV ≤ 0	-0.50
Hot summer and cold winter, Hot summer and warm winter, Temperate	PMV ≥ 0	0.21
	PMV ≤ 0	-0.49

Table 4. Ranking and feature importance under different climate zones

Features	Ranking / Feature importance*				
	Cold	Temperate	Hot summer and warm winter	Hot summer and cold winter	Severe cold
Outdoor Temperature	**1**/0.869	**1**/0.833	**1**/0.797	**1**/0.694	**1**/0.836
Solar radiation	**2**/0.043	**2**/0.051	**2**/0.058	**3**/0.048	**2**/0.042
Humidity	**3**/0.028	**3**/0.044	**4**/0.040	**5**/0.037	**3**/0.041
Atmospheric pressure	**4**/0.023	**4**/0.038	**3**/0.052	**4**/0.040	**4**/0.036
Wind direction	**5**/0.009	**5**/0.013	**5**/0.023	**6**/0.016	**5**/0.013
Wall area	**6**/0.007	**7**/0.004	**8**/0.001	**8**/0.002	**8**/0.007
Wind Velocity	**7**/0.007	**6**/0.008	**6**/0.017	**2**/0.156	**6**/0.010
Window area	**8**/0.006	**8**/0.004	**9**/0.001	**9**/0.001	**9**/0.006
Orientation	**9**/0.006	**9**/0.004	**7**/0.010	**7**/0.004	**7**/0.008

*The features importance keeps 3 decimal places

Table 5. General information of the dormitory building

Type of building	Student dormitory
Location	Ningbo, China
Number of floors	Total floor number: 9 Experimental targets: 2nd - 8th floors
Window orientations	South and North
Window opening	Double horizontal sliding window
Thermal characteristics	U-values walls: 1 W/m^2k, windows: 2.8 W/m^2k

Table 6. Weather data and window opening states of each group

	Outdoor temperature (℃)	Humidity (%)	Atmospheric pressure (kPa)	Wind velocity (m/s)	Wind direction (degree)	Proportion of opening window
Group 1	19.0	53	102.3	0.6	45	0.65
Group 2	20.3	62	102.1	0.8	90	0.67
Group 3	18.9	68	102.2	0.8	90	0.69
Group 4	17.1	91	102.0	1.2	45	0.60
Group 5	16.4	90	101.6	1.2	0	0.47
Group 6	16.6	90	101.7	1.5	0	0.31
Group 7	19.0	53	101.7	1.3	45	0.51
Group 8	18.0	91	101.5	3.0	315	0.65
Group 9	14.0	82	102.1	0.6	0	0.21

Note: The data was collected during the night, and the solar radiation of each group is equal to zero.

Climate zones	Cities
Hot summer and cold winter	Shanghai, Dongtai, Wenzhou, Hefei, Nanping, Nanchang, Wuhan, Zhumadian, Ganzhou, Changsha, Laohekou, Wugang, Xinyang, Youyang, Wanyuan, Hanzhong, Zunyi, Nanchong, Luzhou, Chengdu.
Cold	Shangqiu, Weifang, Xinyang, Jinan, Dalian, Tianjin, Yueshe, Beijing, Yuanping, Yanan, Guyuan, Lanzhou, Jiulong, Garze, Qamdo, Nyingchi, Turpan, Ruoqiang, Minfeng, Shache.
Severe Cold	Dunhuang, Minqin, Harbin, Mudanjiang, Baicheng, Linjiang, Siping, Yingkou, Chifeng, Ejin Qi, Hailar, Hohhot, Tongliao, Yinchuan, Golmud, Madoi, Nangqen, Lhasa, Altay, Karamay.
Hot summer and warm winter	Congwu, Xiamen, Shanghang, Shantou, Shaoguan, Heyuan, Shanwei, Guangzhou, Guilin, Wuzhou, Yangjiang, Guiping, Dianbai, Nanning, Qinzhou, Haikou, Baise, Qionghai, Longzhou, Dongfang.
Temperate zone	Guiyang, Bijie, Xingyi, Weining, Xichang, Huili, Kunming, Mengzi, Chuxiong, Yuanjiang, Lijiang, Simao, Mengla, Lincang, Tengchong.

Figure 1. China climate zones and selected cities

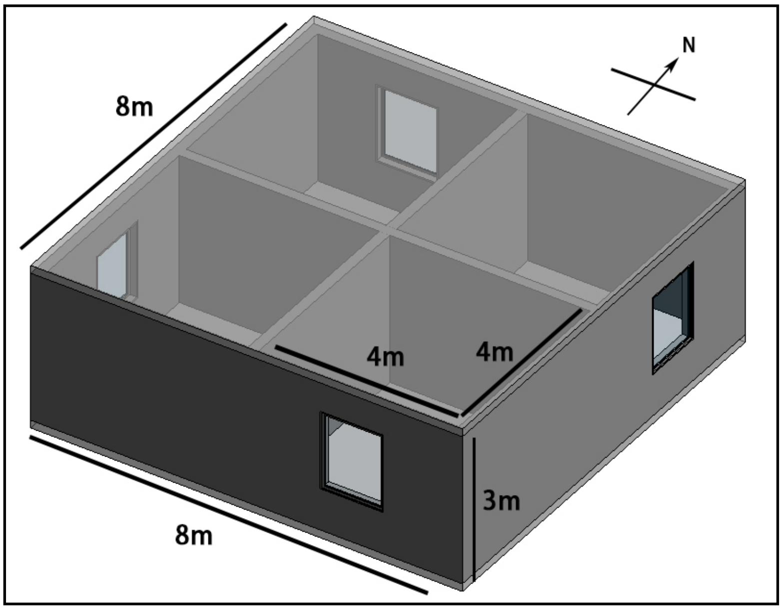


Figure 1. The configuration of the building model

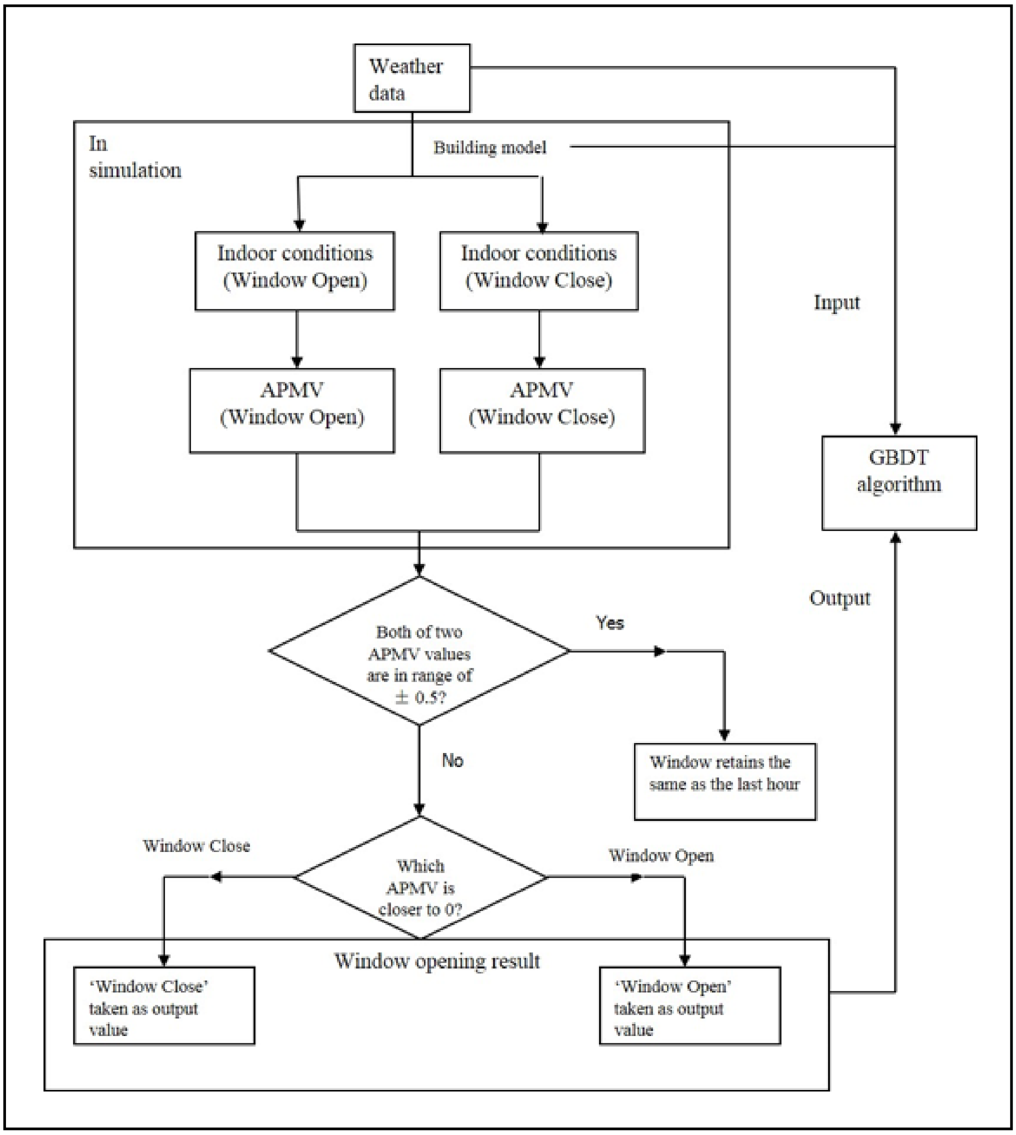


Figure 2. Identification of window opening behaviour

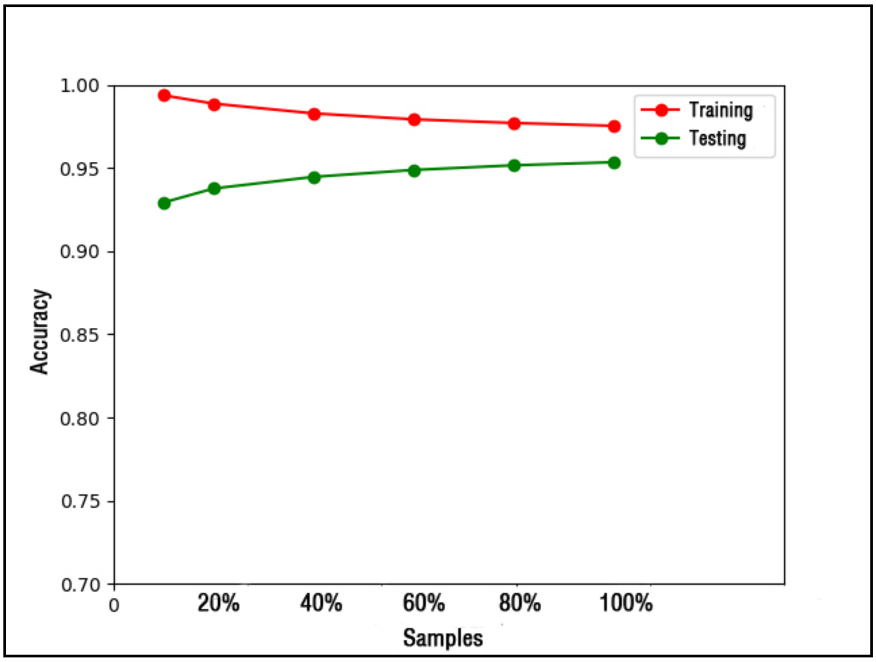


Figure 3. The learning curve of GBDT model

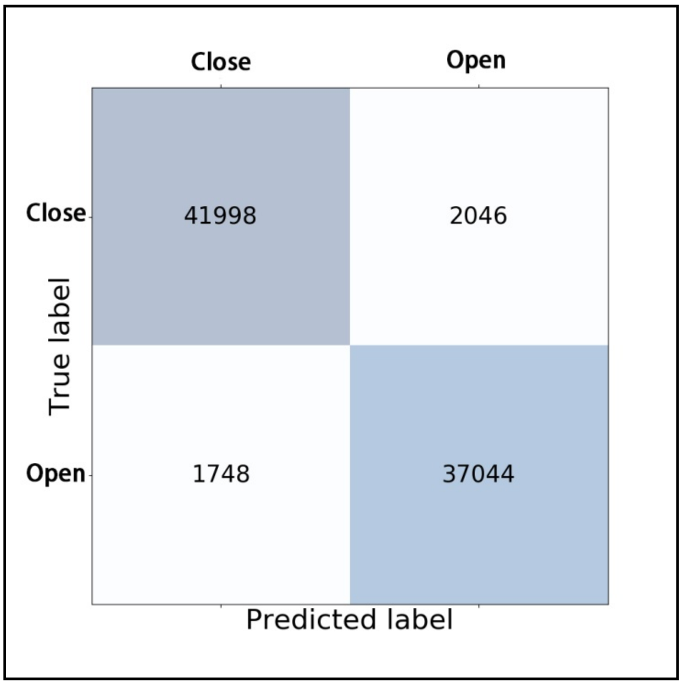


Figure 4. Confusion matrix of the testing set

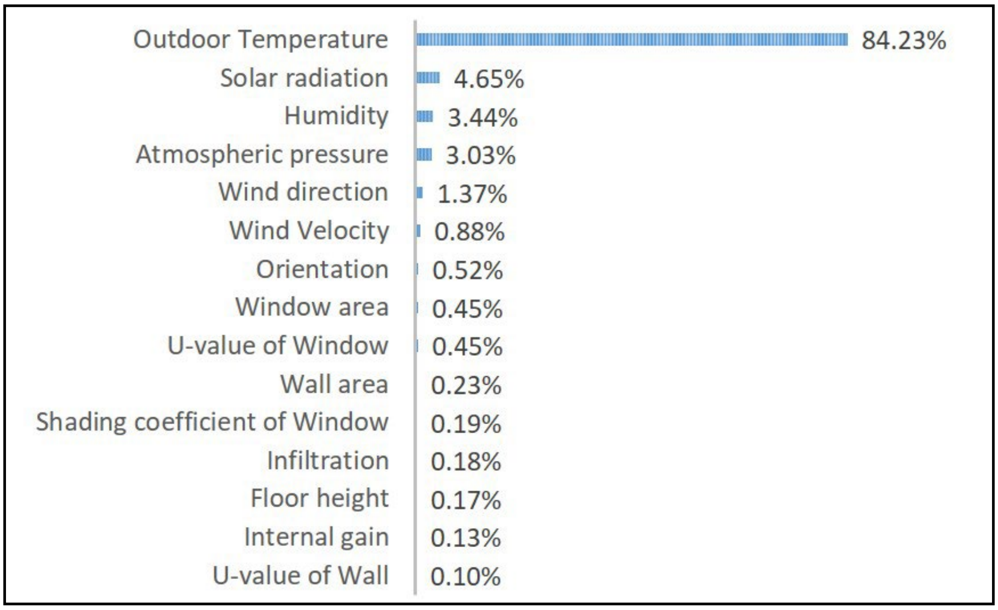


Figure 5. Feature importance in all samples

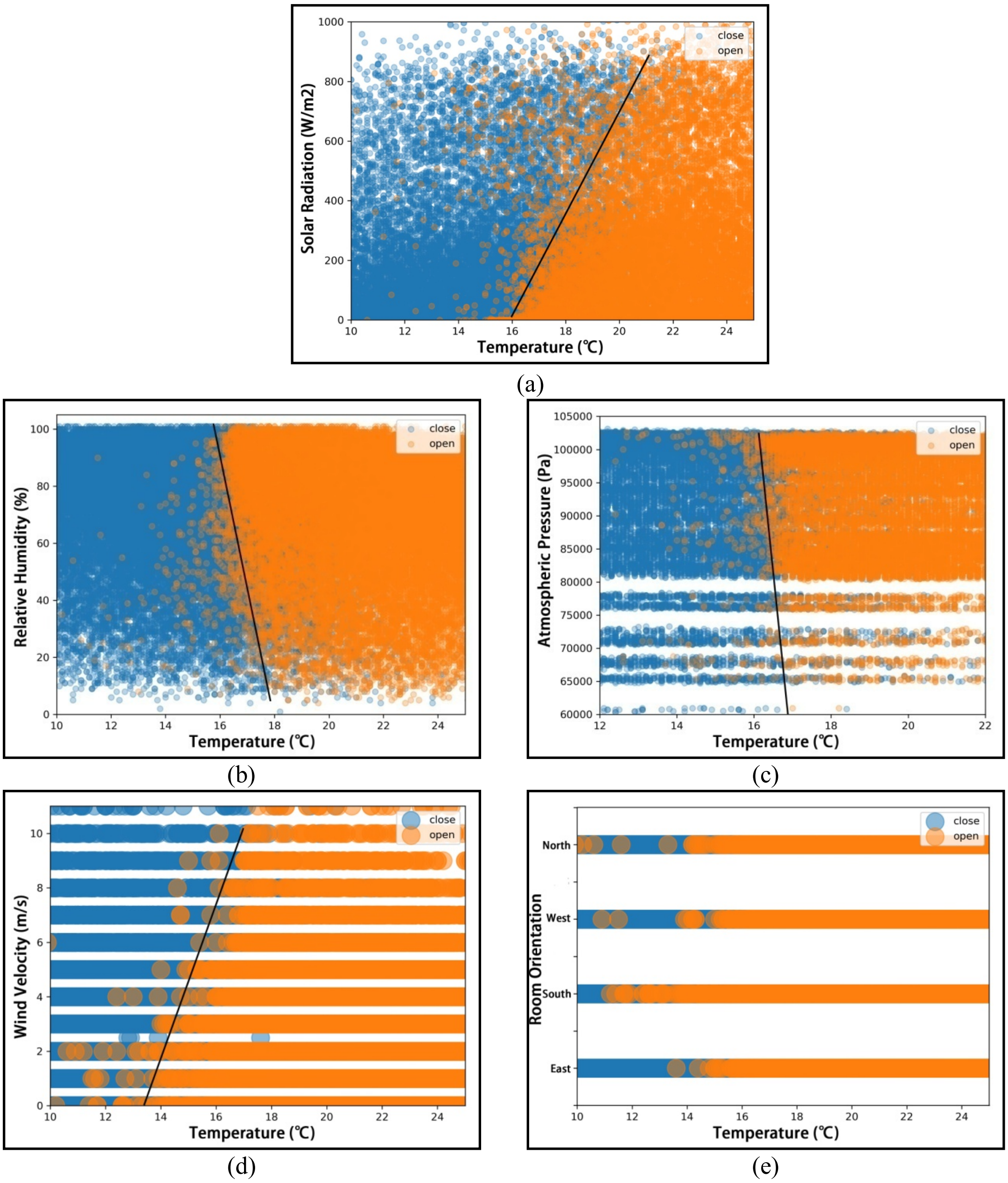


Figure 6. The tendency of various features to window opening against outdoor temperature: (a) Solar radiation; (b) Relative humidity; (c) Atmospheric pressure; (d) Wind velocity; (e) Room orientation

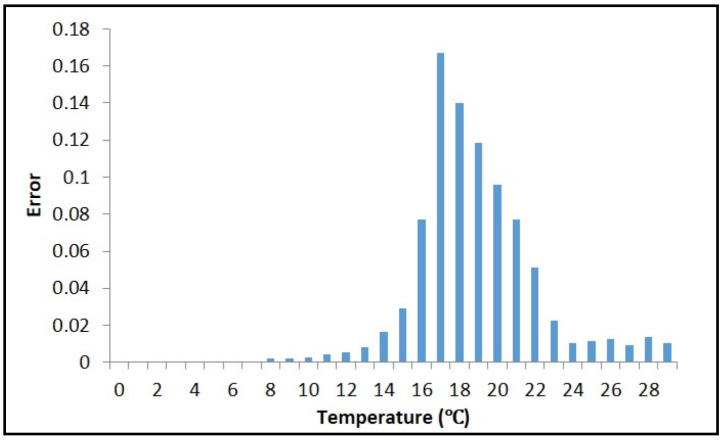
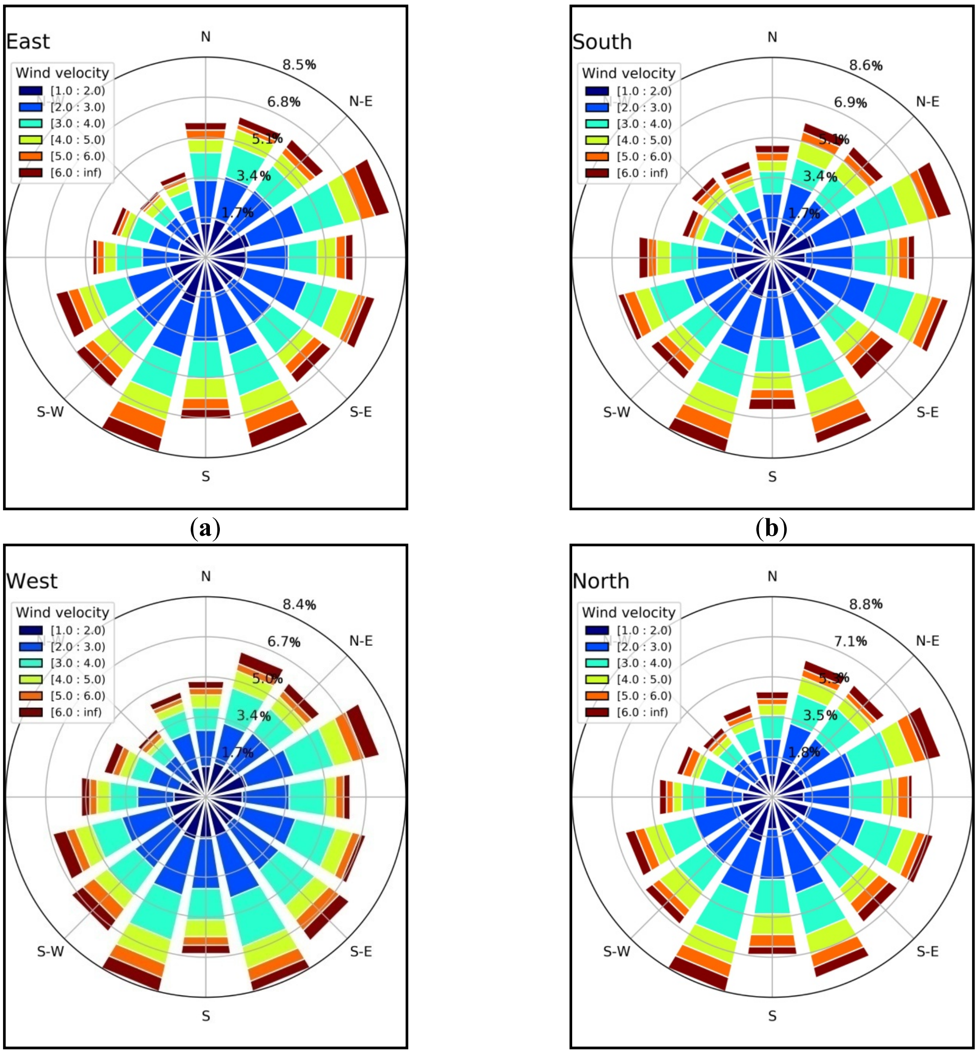


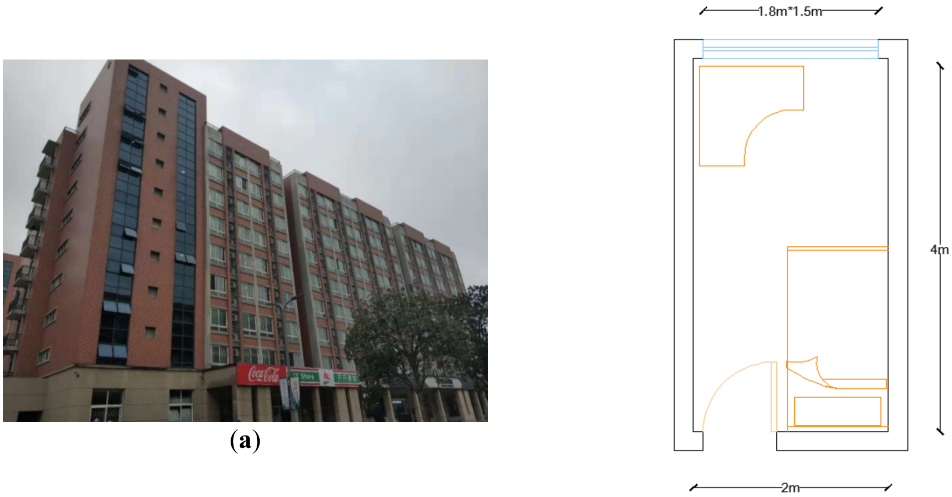
Figure 7. Error ratio under different outdoor temperature



(a) (b)

(c) (d)

Figure 8. Opened window distribution under different wind directions and room orientations (in percentage): (a) East; (b) South; (c) West; (d) North



(a)

(b)

Figure 9. The configuration of the dormitory building: (a) Outlook of the building; (b) Dimension of each room unit

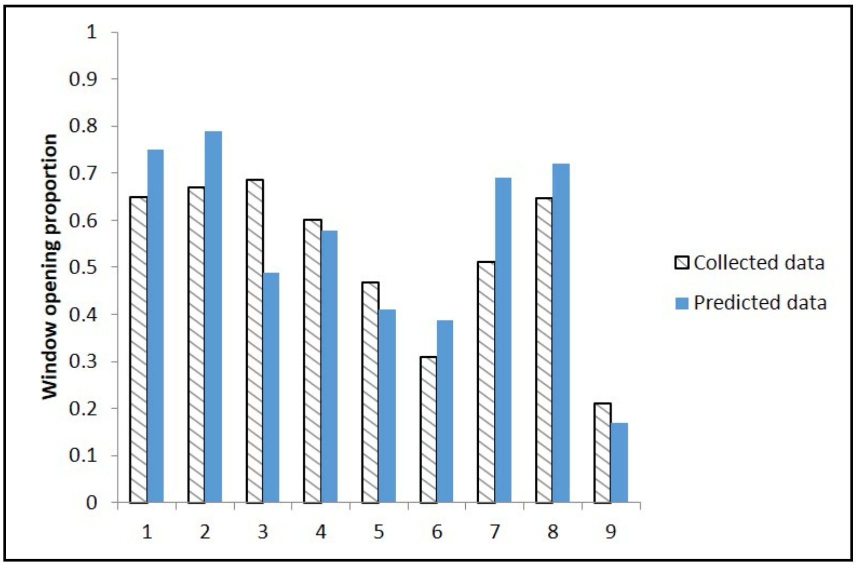


Figure 10. The comparison of predicted window opening proportion and reality