# Mining and modeling occupant-behavior patterns in residential buildings based on field survey results

H. Zhou[1], B.R. Lin[1, *], Y. Jiang[2], H. J. Sun[3]

[1]School of Architecture, Tsinghua University, Beijing 100084, China
[2]Intelligent Green Solutions Ltd., Beijing 100084, China
[2]School of Environmental Science and Engineering, Tianjin University, Tianjin 300072, China

## ABSTRACT
The variety of occupant behaviors in buildings have led to a significant gap between simulated building energy performance and measured one. It is crucial to explore and model occupant-behavior patterns based on field survey data to achieve more accurate simulation purpose. The team continuously collected environment and behavioral data in 3 residential rooms for over half a year, and then proposed a data mining approach, frequent-pattern mining, to mine the correlation rules between behaviors and related driving forces from collected data. Occupant-behavior patterns were finally extracted and modeled based on the generated correlation rules, and expected to be applied in housing energy simulation.

## KEYWORDS
Occupant-behavior, Frequent-pattern mining, Driving force, Correlation rule, Probit analysis

## INTRODUCTION
Most building-energy simulation programs use fixed, statistically averaged data on occupants to predict energy demands (Hirsch 2010). However, energy-related behaviors may differ greatly among occupants, especially in residencies where the occupants have greater freedom than in offices to adjust the indoor environment through behaviors. Li et al. (2007) conducted a field survey of summer air-conditioner usage in 25 apartments within a single building in China, and found the largest electricity consumption was about 15 kW/m$^2$ while the smallest was almost 0 kW/m$^2$. As a result, averaged behavior patterns in simulation have led to a performance gap between simulated and measured building energy (De Wilde 2014). Thus, it is essential to learn occupant behaviors in residential buildings to predict the building energy performance more accurately.

Modeling of the occupant-behavior patterns is a process of quantitatively describing the relationship between behaviors and driving forces, namely, why and how behaviors

* Corresponding author email: linbr@tsinghua.edu.cn

happen. The majority of previous studies used probit analysis with single parameter to describe the patterns (Nicol 2001). Such models are readily understood but often one-sided due to limit of single driving force. Fabi et al. (2012) investigated previous field surveys and summarized that occupants' behaviors in buildings could be driven by a diversity of drivers including physical environment, context, and psychological. Some researchers (Herkel et al. 2008) tried to identify the main forces by statistical method such as odds ratio, but still concluded single-factor models. Wang (2014) proposed a theoretical model which covered almost all the scenarios when behaviors happened. Although comprehensive, this model required a large amount of continuous field data and appropriate data analysis method to identify the constants in its functions. Therefore, this study continuously collected environment and behavioral data in 3 residential rooms for over half a year, and proposed a data mining approach, frequent-pattern mining, to mine the correlation rules between behaviors and related driving forces from collected data, and modeled the behavior patterns based on Wang's theoretical model.

## MATERIALS/METHODS

### Experiment/survey setup

The survey was conducted from July to December 2014 in 3 residential rooms, as listed in Table 1. All the adults in the experiment were office workers, aging from 27 to 43.

***Table 1.*** *Room details*

| *Label* | *Function* | *Family Composition* |
|---|---|---|
| A | Living room | An adult with retired parents |
| B | Bedroom | A couple |
| C | Bedroom | A single adult |

Figure 1 shows a schematic of the indoor environment monitoring system (Zhou et al. 2016). Indoor environment data was collected 5 $min^{-1}$ on six parameters: air temperature, relative humidity, noise, illuminance, $CO_2$ and formaldehyde concentration. To reduce intrusiveness in the families' daily life, the placement of sensing terminals was at the participants' discretion. Outdoor weather and air quality data was captured from official websites by a Python program, and stored in the central server. Besides, an application installed on occupants' phones was used to collect the ground truth data (Figs. 1(b) and (c)). These data included information on: 1) family composition, 2) when an occupant entered or left the room, and the number of occupants present in the room, and 3) use of air-conditioning, light switches, windows, and other activities such as watching TV.

### Mining of correlation rules

Correlation rules in data mining could often be expressed by an item combination as "$\alpha$, $\beta$, … => ω [*support*, *confidence*]", where $\alpha$, $\beta$, … and $\omega$ are items correlated with each other. Items relevant with occupant behaviors could be divided into 5 group: behaviors (actions), events, status of devices and room occupancy, environment, and time, and items of each group are listed in Table 2.

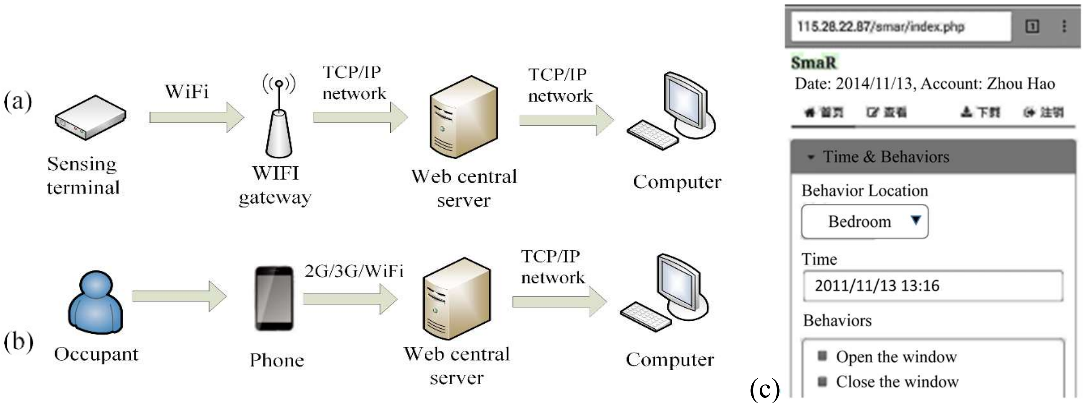


***Figure 1.*** *Data collection system: (a) environment data, (b) and (c) ground truth data*

*Table 2 Items relevant with occupant behaviors*

| *Group* | *Item* |
|---|---|
| Behaviors (actions) | Lk/Lg (turn on/off the light),<br>ACk/ACg (turn on/off the air-conditioner), St (change the temperature set point of air-conditioner), Wk/Wg (open/close the window) |
| Events | Rj/Rc (enter/leave the room), Rs/Rq (go to bed/get up) |
| Status of devices and room occupancy | Lon/Loff (light is on/off), ACon/ACoff (air-conditioner is on/off),<br>Wo/Wc (window is open/closed),<br>Rin/Rout (room is occupied/unoccupied) |
| Environment | it/ot (indoor/outdoor temperature), discretized by 1$^o$C<br>irh/orh (indoor/outdoor relative humidity), discretized by 5%<br>i (room illuminance), discretized by 5lux<br>n (noise), n0=[30, 45) dB, n1=[45, 60) dB, and n2=[60, ) dB<br>d (formaldehyde concentration), discretized by 0.01 mg/m$^3$<br>c ($CO_2$ concentration), discretized by 200ppm<br>p (PM2.5), p0=[0, 75) µg/m$^3$, p1=[75,150) µg/m$^3$, and p2=[150, ) µg/m$^3$ |
| Time | Seasons: summer (6.1-8.31), autumn (9.1-11.14), and winter (11.15-3.15)<br>A day: discretized by 0.5h, eg. T0=[0:00, 0:25], T1=[0:30, 0:55] |

Assuming that occurrence of every behavior is memoryless, and could be considered as one independent random test under the current scenario (comprised of events, status of devices and room occupancy, environment, and time). Therefore, if we classify the results (behaviors or actions) that happen under the same scenario as one group, the *N* samples in the group could be considered as *N* tests. And if Result *w* happens *N*r times among the *N* tests, then the *support* of Combination "$\alpha$, $\beta$, … => $\omega$" is *N*r, and the *confidence* can be calculated as *N*r/*N*. Only both the *support* and *confidence* are higher than their lower limits can the combination be a frequent pattern, and then selected as a correlation rule. In this study, the lower limit of *support* is 3, and that of *confidence* is 5%. FP-growth algorithm (Han et al. 2000) was used to explore the frequent patterns from field surveyed data because of its high effectiveness and scalable performance.

When all the frequent patterns are found out, a process of pruning should be conducted to remove "overlapped" correlation rules. For example, if $\omega$ corresponds to two item groups, "$\alpha_1$, $\beta_1$" and "$\alpha_1$, $\beta_2$, $\gamma$", and *confidence* ($\alpha_1$, $\beta_1$ => $\omega$) = *confidence* ($\alpha_1$, $\beta_1$, $\gamma$ => $\omega$), which means the additive of Item $\gamma$ brings no improvement to the *confidence*, then the Pattern "$\alpha_1$, $\beta_1$, $\gamma$ => $\omega$" could be removed from the set of correlation rules.

**Modeling of occupant-behavior patterns**

This study referred to Wang's work (2014) in modeling the occupant-behavior patterns, and defined the models as a set of probability functions, as is shown in Table 3. The model of AC temperature set point adjustment is also defined, shown in Table 4. Then, the process of modeling occupant-behavior patterns is to identify the driving forces and calculate the features in the tables from the generated correlation rules. The values of features related to driving forces such as time and events could often be equal to the confidences of correlation rules, while the values of features related to environmental driving forces should be calculated by regression analysis among the correlation rules.

*Table 3 Models of occupant-behaviors*

| *Action* | *Driving force* | *Probability function* | *Feature* |
|---|---|---|---|
| Turn on (Open) | Never | $P = p$ | $P$ |
| | All the time | $P = p$ | $P$ |
| | Enter or leave the room | $P = \begin{cases} p & \text{Enter or left the room} \\ 0 & \text{Others} \end{cases}$ | $P$ |
| | Go to bed or get up | $P = \begin{cases} p & \text{Go to bed or get up} \\ 0 & \text{Others} \end{cases}$ | $P$ |
| | Window, when the AC is turned off | $P = \begin{cases} p & \text{AC is turned off} \\ 0 & \text{Others} \end{cases}$ | $P$ |
| | Window or AC, when it is hot/cold;<br>Window, when it is stuffy or there is strong smell;<br>Light, when it is dark | $P = \begin{cases} 1 - e^{-\left(\frac{x-u}{l}\right)^k \Delta\tau} & x \le u \\ 0 & x > u \end{cases}$ | $u$, $l$, $k$, $x$* |
| Turn off (Close) | Never | $P = p$ | $P$ |
| | All the time | $P = p$ | $P$ |
| | Enter or leave the room | $P = \begin{cases} p & \text{Enter or left the room} \\ 0 & \text{Others} \end{cases}$ | $P$ |
| | Go to bed or get up | $P = \begin{cases} p & \text{Go to bed or get up} \\ 0 & \text{Others} \end{cases}$ | $P$ |
| | Window, when the AC is turned on | $P = \begin{cases} p & \text{AC is turned on} \\ 0 & \text{Others} \end{cases}$ | $P$ |
| | Window or AC, when it is hot/cold;<br>Window, when it is noisy or polluted outside;<br>Light, when it is bright | $P = \begin{cases} 1 - e^{-\left(\frac{x-u}{l}\right)^k \Delta\tau} & x > u \\ 0 & x \le u \end{cases}$ | $u$, $l$, $k$, $x$* |

*$u$, $l$, $k$ are constants obtained from regression analysis; $x$ is the environmental parameters related to the driving forces.

*Table 4 Model of AC temperature set point adjustment*

| *Action* | *Driving force* | *Probability function* | *Feature* |
|---|---|---|---|
| Adjustment of AC temperature set point | Fixed set point $T_0$ | $T_{set} = T_0$ | $T_0$ |
| | Adjust to $T_1$ when feel cold | $T_{set} = \begin{cases} T_0 & \text{AC is turned on} \\ T_1 & T \le T_c \end{cases}$ | $T_0$, $T_1$, $T_c^*$ |
| | Adjust to $T_2$ when go to bed | $T_{set} = \begin{cases} T_0 & \text{AC is turned on} \\ T_2 & \text{Go to bed} \end{cases}$ | $T_0$, $T_2$ |
| | Adjust to $T_3$ when leave the room | $T_{set} = \begin{cases} T_0 & \text{AC is turned on} \\ T_3 & \text{Left the room} \end{cases}$ | $T_0$, $T_3$ |

*$T_c$ is the lower limit temperature of the occupant's thermal comfort zone.

## RESULTS

Before the data mining, samples with missing data (environment or ground truth) due to power or network failures were excluded. All the mining processes were conducted using a Python program specifically developed for the task.

### Mining of correlation rules

Take the summer light actions in Room A as an example. The frequent patterns mined from Room A are listed in Table 5. As the basic premise of the action of turning on light is either that the occupant is already present in the room or right enter the room, the rules should contain items Rin or (Rout, Rj). Besides, as the key feature in the driving force of light operations, illuminance should also be included in the rules. Therefore, Rule 3, 4, 7, 9 and 11-16 in Table 4 were selected as the correlation rule candidates.

*Table 5 Pattern-mining result of actions related to light in Room A in summer*

| *Rule* | "$\alpha$, $\beta$, …" | $\omega$ | *Support* | *Confidence* |
|---|---|---|---|---|
| 1 | Doff, T38 | Dk | 25 | 0.097 |
| 2 | Doff, Rin, T38 | Dk | 24 | 0.136 |
| 3 | Doff, i0, Rout, Rj | Dk | 19 | 0.864 |
| 4 | Doff, i0, Rin | Dk | 18 | 0.144 |
| 5 | Doff, i0, T38 | Dk | 14 | 0.141 |
| 6 | Doff, T39 | Dk | 13 | 0.102 |
| 7 | Doff, i0, Rin, T38 | Dk | 13 | 0.250 |
| 8 | Doff, i0, T39 | Dk | 10 | 0.085 |
| 9 | Doff, i1, Rin | Dk | 9 | 0.058 |
| 10 | Doff, Rin, T39 | Dk | 8 | 0.200 |
| 11 | Doff, i1, Rin, T38 | Dk | 6 | 0.097 |
| 12 | Doff, i0, Rout, Rj, T40 | Dk | 6 | 1.000 |
| 13 | Doff, i0, Rout, Rj, T39 | Dk | 5 | 1.000 |
| 14 | Doff, i0, Rout, Rj, T41 | Dk | 5 | 1.000 |
| 15 | Doff, i0, Rin, T39 | Dk | 5 | 0.161 |
| 16 | Doff, i2, Rin, T38 | Dk | 4 | 0.129 |

Except for Rule 3, 4, and 9, all the other rules have a time item in them. However, time item is not included in the driving forces of light actions shown in Table 3. Theoretically, the lower the room illuminance is, the higher the probability of turning on the light will be. However, every room has windows for nature lighting, and some of the occupants were used to not turning on the light until it was totally dark outside, before which the indoor measured illuminance had already dropped down to 0lux. Thus, the probability increases when it gets later (shown in Table 5), and the rooms with better daylighting was also found more likely to have later night-lighting. Although, scenarios of Rule 7 and 11-16 were included in Rule 3, 4, and 9. In order to simplify the behavioral model, only Rule 3, 4, and 9 were finally used to develop the patterns.

**Modeling of occupant-behavior patterns**

According to Rule 3, the probabilities of turning on the light under different illuminance within i0 ([0, 5) lux) could be obtained by interpolation, namely, 0.864 ($i$=0lux), 0.691 ($i$=1lux), 0.518 ($i$=2lux), 0.346 ($i$=3lux), 0.173 ($i$=4lux), and 0 ($i$=5lux). Then, the pattern generated from Rule 3 could be regressed as Eq. (1) ($R^2$=0.992). Similarly, the pattern generated from Rule 4 and 9 together could be expressed as Eq. (2) ($R^2$=0.998).

$$p=\begin{cases}1-e^{-\left(\frac{x-5}{-10.01}\right)^{1.53}\times 5} & x<5\\ 0 & x\geq 5\end{cases} \quad \text{(Turn on the light when enter the room)} \tag{1}$$

$$p=\begin{cases}1-e^{-\left(\frac{x-10}{-153.99}\right)^{1.27}\times 5} & x<10\\ 0 & x\geq 10\end{cases} \quad \text{(Turn on the light when it is dark in the room)} \tag{2}$$

Then, all the three seasons of behavior patterns related to light in Room A could be modeled in the same way, as is shown in Table 6. The driving forces kept unchanged over the seasons, suggesting a relatively fixed lifestyle in Room A. The lower limits of illuminance ($u$) when the occupants turned on the light had small fluctuations through the seasons, and so did the probability of turning on the light under similar illuminance.

*Table 6 Behavior patterns related to light in Room A*

| *Season* | *Action* | *Driving force* | *Probability function* | *Feature* |
|---|---|---|---|---|
| Summer | Turn on | Enter the room | $p=\begin{cases}1-e^{-\left(\frac{x-5}{-10.01}\right)^{1.53}\times 5} & x<5\\ 0 & x\geq 5\end{cases}$ | Illuminance, lux |
| | | It is dark inside | $p=\begin{cases}1-e^{-\left(\frac{x-10}{-153.99}\right)^{1.27}\times 5} & x<10\\ 0 & x\geq 10\end{cases}$ | Illuminance, lux |
| | Turn off | Leave the room | $p$=1 | |

| *Season* | *Action* | *Driving force* | *Probability function* | *Feature* |
|---|---|---|---|---|
| Autumn | Turn on | Enter the room | $p=\begin{cases}1-e^{-\left(\frac{x-5}{-17.07}\right)^{1.31}\times 5} & x<5\\ 0 & x\geq 5\end{cases}$ | Illuminance, lux |
| | | It is dark inside | $p=\begin{cases}1-e^{-\left(\frac{x-5}{-234.18}\right)^{1.03}\times 5} & x<5\\ 0 & x\geq 5\end{cases}$ | Illuminance, lux |
| | Turn off | Leave the room | $p$=1 | |
| Winter | Turn on | Enter the room | $p=\begin{cases}1-e^{-\left(\frac{x-5}{-17.16}\right)^{1.31}\times 5} & x<5\\ 0 & x\geq 5\end{cases}$ | Illuminance, lux |
| | | It is dark inside | $p=\begin{cases}1-e^{-\left(\frac{x-5}{-278.55}\right)^{1.03}\times 5} & x<5\\ 0 & x\geq 5\end{cases}$ | Illuminance, lux |
| | Turn off | Leave the room | $p$=1 | |

Table 7 shows the behavior patterns related to AC in summer in Room B. Occupants in Room B didn't turn on the AC until the indoor air temperature raised up to 29 $^{o}$C, and turned on it either when they entered the room or it was hot inside. They turned it off when it was too cold inside or they left the room. Mostly, they remained the temperature set point of AC unchanged.

*Table 7 Behavior patterns related to AC in summer in Room B*

| *Season* | *Action* | *Driving force* | *Probability function* | *Feature* |
|---|---|---|---|---|
| Summer | Turn on | Enter the room | $p=\begin{cases}1-e^{-\left(\frac{x-29}{3.86}\right)^{1.28}\times 5} & x\geq 29\\ 0 & x<29\end{cases}$ | Indoor air temperature, $^{o}$C |
| | | It is hot inside | $p=\begin{cases}1-e^{-\left(\frac{x-29}{908.05}\right)^{1.06}\times 5} & x\geq 29\\ 0 & x<29\end{cases}$ | Indoor air temperature, $^{o}$C |
| | Turn off | It is cold inside | $p=\begin{cases}1-e^{-\left(\frac{x-30}{-8703.71}\right)^{0.73}\times 5} & x<30\\ 0 & x\geq 30\end{cases}$ | Indoor air temperature, $^{o}$C |
| | | Leave the room | $p$=0.9 | |
| | Temperature set point | Fixed | *T*set=28 $^{o}$C (*confidence*=0.88) | |

Table 8 shows the behavior patterns related to window in Room C. The main driving forces for the occupants' window actions were the events, other than any environmental parameter. However, the probabilities of window opening gradually dropped from summer to winter, indicating the impacts from the outdoor temperature.

*Table 8 Behavior patterns related to window in the three seasons in Room C*

| *Season* | *Action* | *Driving force* | *Probability fuction* | *Feature* |
|---|---|---|---|---|

| | | | |
|---|---|---|---|
| Summer | Open | Get up in the morning | $p$=0.96 |
| | | Enter the room | $p$=0.93 |
| | Close | Leave the room | $p$=0.53 |
| | | Go to bed in the evening | $p$=0.95 |
| | | Turn on the AC | $p$=1 |
| Autumn | Open | Get up in the morning | $p$=0.85 |
| | | Enter the room | $p$=0.56 |
| | Close | Leave the room | $p$=0.67 |
| | | Enter the room | $p$=0.64 |
| | | Go to bed in the evening | $p$=1 |
| Winter | Open | Get up in the morning | $p$=0.3 |
| | Close | Leave the room | $p$=0.5 |
| | | Go to bed in the evening | $p$=1 |

**Preliminary results of occupant-behavior simulations**

Using the patterns generated above, occupant-behavior simulations were carried out in DeST, and the preliminary result (take Room A as an example) is shown in Figure 2. The simulation result is well consistent with the actual occupants' lifestyle in Room A.

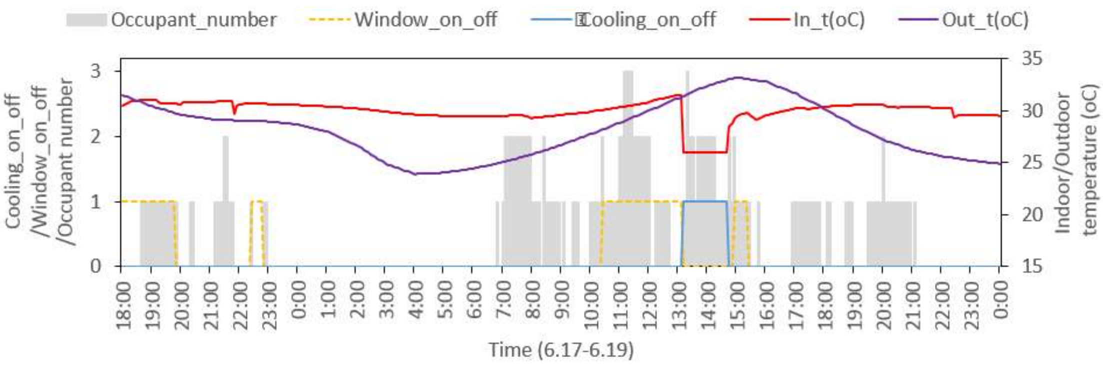


*Figure 2. Preliminary result of occupant-behavior simulation in Room A*

## DISCUSSION

The behavior patterns modeled in this study were generated from filed measurement data, so the results could match the actual one as closely as possible. However, different from that obtained from laboratory, the experiment scenarios could not be controlled in real buildings, so that we often couldn't get enough samples with continuously changed environmental parameters. For example, even though the occupants in Room B turned on the AC the most frequently among all (64 times), the indoor air temperature when the AC was turned on were focused on a small range from 29 to 31 ºC. Moreover, the samples still should be divided into groups with different driving forces, which leads to much less and discrete samples in each group. Thus, we need to add samples artificially by interpolation, and then ideally predict the probabilities of turning on AC when the temperature is lower than 29 ºC or higher than 31 ºC by the regressed functions. Error could be expected. Continuous data collection in long-time scale can be a solution.

The driving forces in our models are found less than those of Wang's theoretical models, especially when related to window. This may be partly due to the limit of sample space discussed above, and more partly because human have strong adaptability to dynamic building environment, and are less sensitive to the environment change than expected.

## CONCLUSION AND IMPLICATIONS

According to the results, the following conclusions can be drawn: 1) The data mining algorithm, frequent-pattern mining, could be used to identify the correlations between occupant-behaviors and driving forces in residences; 2) The occupant-behavior patterns could be generated from correlation rules, and modeled as a set of probability functions. The data-mining methods could be extended to similar researches in offices, and models could be further applied in building energy simulation programs such as DeST and E+.

## ACKNOWLEDGMENTS

This study was supported by the project "Total Performance of Low Carbon Buildings in China and the UK" (funded by the Engineering and Physical Sciences Research Council grant (EP/N009703/1) and the National Natural Science Foundation of China grant (51561135001)). The authors gratefully acknowledge the enthusiastic cooperation of the families during the experiment. Several workers from Intelligent Green Solutions Ltd. helped with the data collection system. Their contributions are greatly appreciated.

## REFERENCES

De Wilde P. 2014. The gap between predicted and measured energy performance of buildings: A framework for investigation, *Automation in Construction.* 41: 40-49.

Fabi V. et al.. 2012. Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models, *Building and Environment*. 58: 188-198.

Han J, Pei J, and Yin Y. 2000. Mining frequent patterns without candidate generation. *ACM Sigmod Record*. Vol. 29, No. 2, pp 1-12.

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: 588-600.

Hirsch JJ. 2010. *eQuest Introductory Tutorial*, Version 3.64, Camarillo, CA, 59.

Li ZJ, Yi J, and Wei Q. 2007. Survey and analysis on influence of environment parameters and residents' behaviors on air conditioning energy consumption in a residential building, *HV&AC*, 37: 67–71 (in Chinese).

Nicol JF. 2001. Characterising occupant behaviour in buildings: towards a stochastic model of occupant use of windows, lights, blinds, heaters and fans, *Proceedings of the Seventh International IBPSA Conference*. Rio., Vol 2, pp 1073-1078.

Wang C. 2014. Research on simulation of energy-related occupant behaviors, *Ph.D. Thesis*, Tsinghua University (in Chinese).

Zhou H et al.. 2016. Recognition of air-conditioner operation from indoor air temperature and relative humidity by a data mining approach, *Energy and Buildings*, 111, 233-241.