Building and Environment 82 (2014) 726–739

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

A data-mining approach to discover patterns of window opening and closing behavior in offices

Simona D'Oca [a], Tianzhen Hong [b,*]

[a] Polytechnic of Turin, Energy Department, TEBE Group, Technology Energy Building Environment, Italy
[b] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

A R T I C L E I N F O

Article history:
Received 1 July 2014
Received in revised form
12 September 2014
Accepted 23 October 2014
Available online 30 October 2014

Keywords:
Data mining
Behavioral pattern
Occupant behavior
Office buildings
Window closing
Window opening

A B S T R A C T

Understanding the relationship between occupant behaviors and building energy consumption is one of the most effective ways to bridge the gap between predicted and actual energy consumption in buildings. However effective methodologies to remove the impact of other variables on building energy consumption and isolate the leverage of the human factor precisely are still poorly investigated. Moreover, the effectiveness of statistical and data mining approaches in finding meaningful correlations in data is largely undiscussed in literature. This study develops a framework combining statistical analysis with two data-mining techniques, cluster analysis and association rules mining, to identify valid window operational patterns in measured data. Analyses are performed on a data set with measured indoor and outdoor physical parameters and human interaction with operable windows in 16 offices. Logistic regression was first used to identify factors influencing window opening and closing behavior. Clustering procedures were employed to obtain distinct behavioral patterns, including motivational, opening duration, interactivity and window position patterns. Finally the clustered patterns constituted a base for association rules segmenting the window opening behaviors into two archetypal office user profiles for which different natural ventilation strategies as well as robust building design recommendations that may be appropriate. Moreover, discerned working user profiles represent more accurate input to building energy modeling programs, to investigate the impacts of typical window opening behavior scenarios on energy use, thermal comfort and productivity in office buildings.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

To secure sustainable energy development in the building sector, occupant behavior needs to be modified towards a more efficient and conscious energy usage. The development of energy-conserving technologies is a necessary but incomplete step toward reduced energy consumption in buildings. Achieving energy conservation becomes a double challenge, partly technical and partly human, since energy consumption may vary largely due to how occupants interact with system controls and the building envelope. Currently, building simulation tools can only imitate some typical occupant activities in a rigid and pre-defined way (occupancy, use of windows, thermostat, shadings, and lighting). Nevertheless, occupant behavior and comfort is stochastic, complex, and multi-disciplinary therefore more realistic behavioral patterns need to be developed.

As a matter of fact, a deeper understanding of the relationship between occupant behavior and building energy consumption can be seen as one of the most effective ways to bridge the gap between predicted and actual energy consumption in buildings Several studies underlined that huge variability exists in terms of default settings and day-to-day use of control systems and appliances in buildings, where 'behavior' is central to consumption levels [1–3]. In this context, the 'dark side of occupant behavior on building energy use' was demonstrated by Masoso et al., in 2000 [4]. The work showed that more energy was used during non-working hours (56%) than during working hours (44%) in one office building. This arises largely from occupants' behavior of leaving lights and equipment on at the end of the day, and partly due to poor zoning and controls. In 2004 Bordass et al. [5] referred to this occurrence as the 'credibility gap', alluding to the loss of credibility when design expectations of energy efficiency and actual building consumption outcomes differ substantially. They suggest that credibility gaps arise not so much because occupants preform 'wrong', but because the assumptions often used are not well enough informed by what really happens in practice. In the last

* Corresponding author. Tel.: +1 510 486 7082; fax: +1 510 486 4089.
E-mail addresses: simona.doca@polito.it (S. D'Oca), thong@lbl.gov (T. Hong).

http://dx.doi.org/10.1016/j.buildenv.2014.10.021
0360-1323/© 2014 Elsevier Ltd. All rights reserved.

decades, a number of studies focused on overcoming this barrier, testing valid, applicable and robust methodologies and analysis techniques to predict building occupant behavior seriously [6–16]. Describing, predicting or influencing energy related individual behavior are challenging tasks that must start with the non-trivial understanding of the stochastic nature of human beings. In this view, the scientific community is addressing rising interest around the issue of energy efficient buildings and specifically toward the need of a more robust description of the motivations driving humans to interact with building envelope and control systems (fans, windows, thermostats, lights, etc.) in order to bring about desired comfort conditions [21].

The most important issue in between perceived indoor environmental quality and outdoors, in the built environment, is the building envelope [17]. As a consequence, window operation is one of the most relevant tools that allow occupants to bring about desired indoor thermal and air quality conditions, by moving air through the building. Further, since the building envelope is getting always more thermally efficient, ventilation and air infiltrations due to window opening are increasing their influence with respect to energy use, becoming the most dominant source of thermal loss of the heat balance mechanism. Fitting the Humphrey's adaptive principle that if a change occurs such as to produce discomfort, people react in a way which tend to restore their comfort [18] to the findings in literature [19,20], it is demonstrated that occupants in naturally ventilated buildings accepted and actually preferred a significant wider range of temperatures compared to users of mechanically ventilated buildings. As a matter of fact, naturally ventilated buildings allow occupants' a greater degree of control over indoor hydro-thermal conditions than air conditioned buildings that strongly influence their satisfaction with working spaces [19]. In 2004 de Dear and Brager [20] highlighted that the variation of indoor environmental conditions caused from a human operable control source such as windows lead occupants to a relaxation in expectations and higher tolerance of temperature excursions.

1.1. Statistical analysis of factors influencing occupant behavior in buildings

Statistical analysis techniques are extensively applied to discover associations and relationships among the various factors influencing building energy performance and occupant behavior in buildings. Different suitable user behavioral models were defined by means of statistical analysis (Markov Chain, Generalized Linear Models, etc …) [7–16,23–32]. An extensive review of these studies has been conducted in the context of Annex 53 – Total Energy Use in Buildings, under the International Energy Agency Energy in Buildings and Communities Program [21], in order to understand the correlation between window opening and the parameters, also called drivers, influencing users' interaction in buildings with natural ventilation. The parameters are divided into five categories of influencing factors:

- Physical (indoor and outdoor environment);
- Psychological (preferences, attitudes);
- Physiological (age, sex);
- Contextual (type of environment where the occupants are located);
- Social (income, lifestyle).

Specifically for window opening in office buildings, a literature review was carried out in 2012 by Fabi et al. [22] of more than 70 scientific papers, indicating that window operation was not only influenced by perceived thermal condition, but it was also seen as a response of sensed indoor air quality, external (outdoor temperature, solar radiation, wind speed, rain) and internal (indoor temperature) environmental conditions as well as contextual factors (window type, time of the day, season of the year) and personal and cultural preferences. In these studies, statistical analysis techniques were applied to identify the influential variables on user behavior in buildings. The strength of this methodology was the simplicity and widespread familiarity.

- Indoor and outdoor temperatures were found as paramount factors influencing window opening and closing by several studies [23–27]. For instance, Fabi et al. suggested that rising indoor temperatures might drive the opening of windows, but how long the window stayed open might depend more on outdoor temperature. More specifically, Andersen et al. [28] found that the CO_2 concentration was the most important driver for opening the windows, while the outdoor temperature was the most dominant driver for closing the windows.
- Solar radiation was found by Herkel et al. [29] to have little correlation with window openings. Solar radiation was a relatively small factor when compared with the correlation of indoor and outdoor temperatures
- Wind speed was reported by Roetzel et al. [30] as a driver for closing the windows when the sensation of draft was producing a predominant discomfort.
- Time of arrival and departure as well as the time of the day had been found having a strong correlation between window adjustments by several researches. [28,29,31]
- The season of the year was found by Herkel et al. [29] to have a strong correlation with window opening [6]. Usually, the interactivity with openings was higher in summer and during the midseason (autumn and spring) and lower in winter.
- The current state of the window was also underlined by several studies [30,32] as a key aspect to take into account when concerning user's willingness to open and close windows.

1.2. A data mining framework for behavioral pattern discovery

Currently, there is no comprehensive consensus about the way people interact with building controls or the motivating factors that influence their decisions. However, there is a substantial body of research that offers guidance on patterns of behaviors. Patterns are expressions describing typical behaviors or models applicable to a subset of the data to anticipate and replicate common actions. Moreover, patterns correlate repetitive behaviors and actions to user profiles. Guerra Santin [33] statistically determined behavioral patterns of HVAC system interactions and associated energy spent on heating. From this, household and building characteristics that could contribute to the development of energy-user profiles, were identified [33]. A study conducted by Van Den Wymelenberg [34] reviewed data from more than 50 buildings and identified patterns of occupant interaction with window blind controls. Moreover, Yun and Steemers [35,36] provided evidence of a statistically significant relationship between window-opening behavior patterns and clusters of indoor stimuli. In 1983 Van Raaij and Verhallen [37] carried out a study in 145 Dutch dwellings and defined five patterns of energy behavior (conservers, spenders, cool, warm and average) in relation to the use of heating systems and ventilation habits. Findings of this research showed that the energy uses of these five pattern groups differed considerably, up to 31% [37].

Data mining techniques to discover patterns of data are largely applied to research fields such as marketing, medicine, biology, engineering, medicine, and social sciences [38]. Even so, the application of data mining framework to building energy consumption and operational data is still under investigation and nevertheless could be potentially highly effective.

Data mining was defined in 2001 by Hand et al. [39] as: *"The analysis of large observation data sets to find unsuspected relationships and to summarize the data in novel ways so that owners can fully understand and make use of the data"*. Another definition was given in 1998 by Cabena et al. [40] as: "*An interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases and visualization to address the issue of information extraction from large databases*".

Applying data mining techniques, with the scope to discern behavioral patterns, were tested by several studies both in residential and office buildings. Between 2011 and 2012 Yu et al. [41–43] tested several systematic data mining methodologies for identifying and improving occupant behavior in buildings. The results showed that this analysis methodology proved powerful in providing insights into energy pattern related to the occupant behavior, facilitating evaluations of building saving potential by improving users' energy profiles as well as driving building energy policy formulation.

2. Methodology

In this study, a methodology was proposed in order to identify valid, novel, potential useful and understandable patterns of window opening and closing behavior in offices.

Statistical analysis and data-mining techniques were applied to measured building energy and environmental data. Statistical analysis provided leverage in identifying the influencing factors on occupant energy-related behavior and removed the effects of other insignificant variables on building energy performance. In the literature, examples could be found of logistic regression analyses to discover the variables influencing energy-related behavior [21]. This technique was borrowed from the natural sciences literature, where several investigations focused on the relations between energy-related behavior and (mainly physical) drivers of this behavior [44,45]. According to Nicol [46], energy-related behavior was clearly affected by physical parameters, but the relationship tended to be stochastic. For example, there was no exact temperature at which every occupant would open a window, but for increasing temperatures, the probability of the occupant opening the window, increased.

However, associations among variables found to have little statistical correlation on isolated occupant behaviors or small data set may lead to the understanding of more general patterns of behavior in a large data set, helping direct future research. In this context, data mining techniques such as cluster analysis and association rules algorithms were applied in the proposed framework with the scope to discern typical office user profiles which may allow for more accurate assumptions on group behaviors, overcoming the lack of personalization of statistical patterns.

The proposed framework suggests an improvement of the notion of behavioral patterns not only as merely statistical relevant clusters, but also incorporating the driver-response conditioning dimension with typical window opening habits.

Fig. 1 shows the proposed framework in this study:

- In *step 1*, a statistical analysis technique (logistic regression) was applied to the given data set. The goal was to discover the factors (variables and coefficients) influencing window opening and closing behavior.

A two steps cluster-then-association rules mining approach was applied to the given data set.

- In *step 2*, clustering procedures were employed in order to obtain distinct behavioral patterns. The goal was to estimate the motivational, opening duration, interactivity, degree of opening and behavioral patterns. In this aim, the research was estimating why (*motivational* pattern), for how long (*opening duration* pattern), how often (*interactivity* pattern) and how much (*position* pattern) working users open and close windows in offices of the same building.
- In *step 3*, the clustered patterns constitute a base for association rules segmenting the building occupants into typical office *user profiles*.

2.1. Statistical analysis technique

Generalized linear models (GLMs) [47] are a class of statistical models for describing linear combination of predictor and dependent variables. The GLM allows the statistical model to be related to a dependent variable via a link function of its predicted values. In the specific case, logistic regression is a sigmoidal classification GLM able to predict the probability of an event having binary outcome (0–1) occurrences based upon predictor variables and coefficients. Logistic regression also allows to express the magnitude of the coefficients of each dependent variables as a function of the binary outcome.

Formula (1) describes the relationship:

$$\mathrm{Log}\left(\frac{P}{1-P}\right) = a + b_1 \cdot X_1 + \ldots + b_n \cdot X_n + \ldots \tag{1}$$

where:

- P is the probability
- a is the intercept
- b_{1-n} are coefficients
- x_{1-n} are variables

2.2. Data mining techniques

Two descriptive data mining approaches: 1) cluster analysis (k-means algorithm) and 2) association rules mining (Frequent Pattern FP-Growth Algorithm) were employed to discover patterns of windows opening and closing [38].

Cluster analysis is the process of merging data into different clusters, so that instances in the same cluster have high similarity and instances in different clusters have low similarity. The similarity between clusters was computed based on the distance between the clusters. The distance measure was described using the Euclidian distance Formula (2) where:

$$d(a,b) = d(b,a) = \sqrt{(b_1-a_1)^2 + (b_2-a_2)^2 + \ldots + (b_n-a_n)^2} \tag{2}$$

where:

- $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$ are two points in an Euclidean n-space

The *k-means algorithm* is a method of vector quantization for cluster analysis in data mining. Given the simple nature of the algorithm, it is one of the widely used classification technique. Assumed a data set D, containing a number n of records (instances), the number of clusters K must be specified.

The performance of the cluster models was evaluated by means a Cluster Distance Performance operator. In this study, the

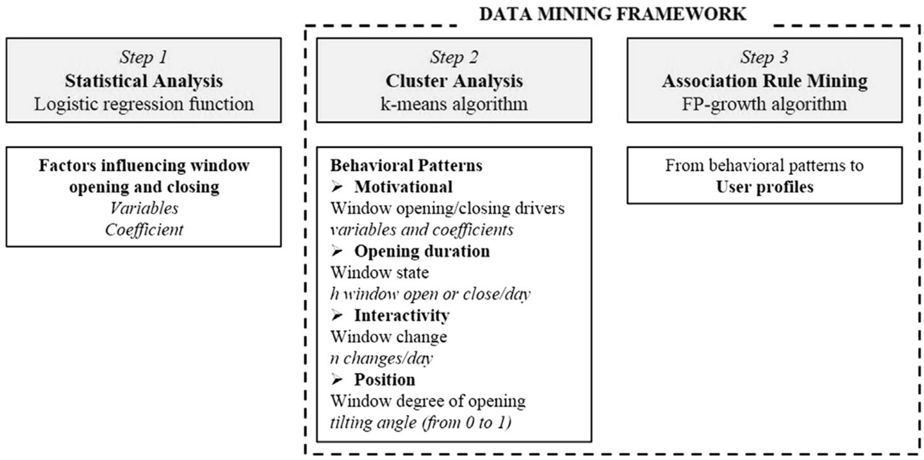

Fig. 1. Proposed framework of the research.

Davies–Bouldin index was used for performance evaluation. The $k = n$ algorithm that produces clusters with low intra-cluster distances (high intra-cluster similarity) and high inter-cluster distances (low inter-cluster similarity) will have a low Davies–Bouldin index, and will be considered the $k = n_{opt}$ cluster algorithm for the specific data set. Each cluster was associated with a centroid (center point), the mean of the points in the cluster and each point was assigned to the cluster with the closest centroid.

Association rule mining (ARM) is a classification technique used to identify associations and correlations between parameters (attributes). Main objective was to extract frequent correlations or patterns (association rules) from a database. Given a data set D, containing a number n of non-ordered records, the association rule was used and described by the Formula (3):

$$A,\ B \Rightarrow C \tag{3}$$

where:

- A, B = items in the rule body
- C = item in the rule head

The validity of the association rules was indicated by support, confidence and lift.

Support, represents the fraction of transactions (T) containing both A and B, shown in Formula (4).

$$\text{Support} = \frac{\#(A,\ B)}{|T|} \tag{4}$$

Confidence (5) represents the conditional probability of finding B having found A, and gives strength to the "if, then" statement of the association rules. Mathematically, confidence can be calculated as the frequency of B in transactions containing A.

$$\text{Confidence} = \frac{\text{sup}(A,\ B)}{\text{sup}(A)} \tag{5}$$

To discover reliable and valid rules in data set, minimum value for confidence and support must be pre-defined. Accordingly, in this study association rules that satisfy the minimum support of 0.3 and minimum confidence of 0.8 in the given data set, were used.

Lift (6) represents the ratio of the observed support to that expected, if A and B were independent. The *Lift* value must be different to 1, to avoid the occurrence of A being independent of the occurrence of B. The higher the lift value, the more likely that a correlation between A and B exists.

$$\text{Lift} = \frac{\text{sup}(A,\ B)}{\text{sup}(A)^*\text{sup}(B)} \tag{6}$$

The *frequent pattern growth algorithm* (FP growth), the most commonly used algorithm to discover patterns into a given data set, generated a classification tree (FP-tree) that exploited a memory compressed representation of the database. This dense data distribution was used to mine frequent patterns of the smaller subsets.

2.3. The data set

An office building based in Frankfurt am Main, Germany, was used as case study (Table 1). Frankfurt am Main has generally a temperate-oceanic climate, with relatively cold winters and warm summers. The building was naturally ventilated and cooled in summer and every office was equipped with an operable window that may be opened and shut to accommodate the occupant's ventilation needs. Moreover, the building showed very strict design criteria in terms of energy efficiency and energy optimization for heating, cooling, ventilation and lighting, having an average transparent and opaque envelope U-value of 0.54 W/m^2 K.

In this study, the following data set [21] (Figs. 2 and 3) was used and includes:

Table 1
Building characteristics.

Type of building	Multi-story office building
Dimension	17,402 m^2 (8585 m^2 heated)
No. of employee	~350 employees
Location	Frankfurt, Germany
Thermal characteristics	Low energy standard of building envelope (U-values walls 0.24–0.5 W/m^2 K, windows 1.5 W/m^2 K)
Annual primary energy consumption	Less than 100 Wh/m^2
Type of observed spaces	Office rooms
Year of construction	2002
No. of floors	2-level underground car park + 4 office floors + 1 floor apartments on top
Windows, orientation	Mostly E and W
Window opening	Tilt-and turn (automatic BMS + occupant driven mode)
Shading devices	External sun protection (automatic BMS + occupant driven mode)

Fig. 2. Two-part sun protection enables glare-free use of daylight.

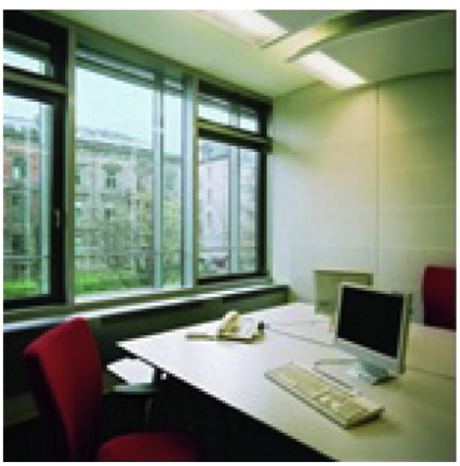

Fig. 3. Offices with operable windows and sun protection, allowing natural ventilation and natural lighting.

a) 16 private offices with single or dual occupancy (Table 2). E01–E11 are eleven offices facing the east while W01–W05 are five offices facing the west.
b) 10-min interval data over two complete years (Table 3)
c) measured indoor and outdoor physical parameters (Table 4)
d) measured behavior and energy use (Table 4)

2.4. Statistical analysis and data normalization

In this study, *logistic regression analysis* was performed to compare the leverage (b_{1-n} coefficient's impact factors) of the x_{1-n} variables influencing the window opening and closing probability.

Table 2
Database characteristics.

Number of offices	16
Period of measurement	2006 and 2007
Type of observed spaces with sensors	Standard offices
Dimension of observed spaces	20 m^2
Occupancy level of observed spaces	1 or 2 persons
Number of observed spaces with indoor CO_2-concentration	3
Orientation	East and West

Table 3
Data characteristics.

	Items	Interval
Climate	Outdoor air temperature, outdoor humidity, wind speed, solar radiance	10 min
Building envelope	Not in database	
Building service & Systems		10 min
Operation & Maintenance	Monitoring of heating, cooling, lighting and ventilation system, and related energy flows	10 min
Indoor environmental quality	Indoor (operative) temperature, humidity, (CO_2)	10 min
Occupants' activities and behavior	Window state (open/closed) Presence State of sun protection (open/closed) Usage of lighting equipment	Event
Social and economical aspects	None	

Table 4
Monitored parameters characteristics.

Outdoor	Indoor	Behavior
Solar radiation [W/m^2] Rain – amount [l/m^2] Rain – event [yes/no] Light intensity – horizontal [l×] Light intensity – South [l×] Light intensity – East [l×] Light intensity – North [l×] Light intensity – West [l×] Outdoor temperature [°C] Wind – velocity [m/s] Wind – direction [°] CO_2 content in air [ppm] Outdoor humidity [%rh]	Room air temperature [°C] Surface temperature [°C] Ceiling slab temperature [°C] CO_2 concentration [ppm]	Occupancy [0/1]* Window contact [0/1; reed contacts]* Top light control [0/1; reed contacts]* Sun protection [% of closure: 0% = open to 100% = closed] Electricity consumption [kWh]

Accordingly, the literature findings [22] suggested that the probability of opening and closing a window was calculated as function of 15 non-numerical and numerical variables for the 16 offices.

2.4.1. Non-numerical variables

1. Season (Summer, Spring, Autumn, Spring)
2. Day of the week (Monday to Sunday)
3. Time of the day (Early Morning 6–9 am, Morning 9 am–12 pm, Noon 12–3 pm, Afternoon 3–6 pm, Evening 6–9 pm, Night 9 pm–6 am)
4. Window State (0 = close, 1 = open)
5. Occupancy State (0 = vacant, 1 = present)
6. Window Change (if occupancy state $t_{n-1} = t_n$ then = no change, otherwise = change)
7. Occupancy Change (if occupancy state $t_{n-1} < t_n$ then = arriving time, if occupancy state $t_{n-1} > t_n$ then = leaving time, otherwise = no change)
8. Precipitation (event 0–1)

2.4.2. Numerical variables

9. Indoor air temperature
10. Outdoor air temperature
11. Outdoor relative humidity
12. Solar radiation horizontal
13. Illuminance level
14. Wind velocity
15. Wind direction.

In logistic regression modeling, it is normal practice in to undertake a process of parameter selection to identify the minimum number of variables required to predict observed behavior, based on their significance and usefulness. Nonetheless, cluster analysis was used in this study to group of variables influencing the window opening and closing behavior in the 16 offices. In this view, all the selected variables are potentially assumed equally significant and useful motivational stimuli driving occupants to satisfy their needs with respect to the natural ventilation of their offices.

Data normalization was applied to numerical variables in order to scale each coefficient within a comparable range and to normalize east-west office orientation. In order to determine the coefficient's impact factors, x_{max} and x_{min} were assumed as the original maximum and minimum coefficient values of the numerical variables selected for the statistical analysis. By rank normalization, a value x of the coefficient was transformed into x' in the new specific coefficient range for each of the numerical variables in the east and west orientations (Table 5).

$$x' = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}$$

Logistic regression analysis was performed along the open source statistical analysis program R [48]. The east-west normalized coefficients' impact factor of every variable on the window opening (Table 6) and closing (Table 7) probability, was then calculated for the 16 offices.

The following key points could be observed from the statistical analysis results:

Avg. value?

- Indoor air temperature, arrival time, occupant presence, time of the day (early morning) and outdoor temperature were the main factors influencing window opening behavior.
- Indoor air temperature, leaving time, occupant presence and time of the day (evening) were the main factors influencing window closing behavior.
- Window opening and closing occupant behavior was equally affected by common physical and non-physical drivers.
- Occupants in the building interact with windows principally driven by thermal discomfort (indoor air temperature) but also behave according to a daily routine (time of the day) and/or habits (arriving and leaving time).

2.5. Cluster analysis of behavioral patterns

The clusters in the present study disaggregate occupant behavior into patterns. Specifically, four patterns of behavior were mined in the given data set: *motivational*, *energy intensity*, *activity* and *position*.

- *Motivational patterns* clustered the factors which drive the users to open or close windows. Clusters were labeled according to the impact (b_{1-n} coefficient's impact factors) the x_{1-n} influencing variables had on the window opening and closing actions.
- *Opening duration patterns* cluster occupant behavior based on the number of hours the window state was recorded open every day.
- *Interactivity patterns* cluster occupant behavior based on the number of window position changes recorded every day.
- *Position patterns* cluster occupant behavior according to the most frequent window degree of opening every day.

Table 5
Coefficient range of the numerical variables for the East and West facing offices.

Numerical Variables	East offices	West offices
Indoor air temperature (C°)	23	18
Solar radiation horizontal	1092	1092
Illuminance level	98,824	97,646
Outdoor temperature	44	44
Wind velocity	13	13
Wind direction	360	360
Outdoor relative humidity	73	73

Four distinct data sets, based on different parameters, were used to mine window opening drivers, state, change and position (Table 8).

The k-means algorithm was employed along with the open source data mining program Rapid Miner 6.0 [49] to perform cluster analysis.

The value $2 > k < 10$ was adjusted in this study in order to find the k_{opt} by using Cluster Distance Performance operator. In this study, the Davies–Bouldin index was used for performance evaluation. The $k = n$ algorithm that produced clusters with low intra-cluster distances (high intra-cluster similarity) and high inter-cluster distances (low inter-cluster similarity) had a low Davies–Bouldin index, and were considered at the $k = n_{opt}$ cluster algorithm, for the specific data set.

2.5.1. Motivational behavioral patterns

Patterns of window opening and closing drivers in offices were clustered based on the impact that the influencing variables played on these actions. The optimal k-means algorithm, validated by means the Davies–Bouldin index, grouped $k_{opt} = 3$ clusters for factors influencing window opening and $k_{opt} = 2$ clusters for factors influencing window closing.

Each office was assigned to a cluster both considering window opening and closing actions.

- **Opening Cluster 1**: 31% offices assigned (E01, E03, E04, E11, W02)
- **Opening Cluster 2**: 31% offices assigned (E02, E05, E09, W04, W05)
- **Opening Cluster 3**: 38% offices assigned (E06, E07, E08, E10, W01, W03)
- **Closing Cluster 1**: 44% offices assigned (E01, E02, E03, E04, E09, E11, W02)
- **Closing Cluster 2** 56% offices assigned (E05, E06, E07, E08, E10, W01, W03, W04, W05)

将每个房间看成一个多维空间下的点，维数等于影响因子的数目

The cluster centroids of the $k = $ opt means algorithms were plotted to provide a visualization of the emerged occupancy patterns. Among the 15 numerical and non-numerical variables, Tables 9 and 10 highlight the top five influencing variables and coefficients for window opening and window closing, respectively. The results from Table 9 suggest the top five drivers for window opening were indoor air temperature, outdoor air temperature, time of the day (office arriving time and early morning) and occupancy presence. From Table 10, the top five drivers for window closing were indoor air temperature, time of the day (office leaving time and evening), occupancy presence and outdoor air temperature.

Fig. 4 shows the impacts (absolute value) that the driving forces have on the window opening and closing, towards the pursuit of occupant comfort. The key findings are as follows:

- Opening Cluster 1 appeared to be significantly more influenced by physical parameters such as indoor (6.49) and outdoor (2.25) air temperature than the other two clusters. Hence, offices

Table 6
Calculated variables and coefficients' impact factors for window opening probability.

	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10	E11	W01	W02	W03	W04	W05
Intercept	−2.86	−0.44	0.24	1.59	3.33	−10.77	−2.38	1.04	1.81	−22.97	−19.97	2.97	1.99	7.49	−0.62	2.51
Occupancy presence	−1.13	−0.65	−1.7	−1.79	−1.66	−0.12	−1.24	−1.45	−2.78	16.67	14.85	−1.84	−1.61	−1.29	−1.15	−1.80
Air temperature	−2.92	0.40	−9.08	−7.49	−4.47	4.66	1.37	−5.31	−4.32	−0.58	−2.35	−6.81	−21.35	−8.16	−0.20	−5.04
Arriving time	−1.10	−1.69	−14.23	−2.25	−2.13	−2.61	−3.15	−2.53	−3.35	−0.83	−0.22	−2.47	0.34	−14.92	−19.18	−2.67
Season spring	0.74	−1.38	−1.80	0.17	0.15	−0.35	−0.68	−0.49	−0.59	1.32	−0.09	−0.69	−0.50	2.54	0.37	−0.03
Season summer	0.55	−1.54	0.05	−0.81	0.52	0.15	−0.73	−0.72	−0.32	−14.86	−1.91	−0.70	−14.20	1.08	−15.42	−0.20
Early morning	−2.76	−2.46	0.47	−0.88	−16.34	−0.22	−0.03	0.01	−1.91	−17.10	−1.46	−1.87	0.98	−2.73	−1.07	−0.98
Outdoor temperature	1.77	−0.17	3.99	2.99	−0.95	2.94	1.66	1.11	0.98	5.00	1.33	2.46	13.61	2.07	1.02	2.46
Outdoor RH	−0.80	−2.17	−4.54	−0.59	−1.65	1.06	−1.60	−0.34	−1.02	−0.31	−2.41	−0.60	−0.31	−2.15	−3.12	−0.32
Noon	−0.69	−0.11	1.06	0.00	−0.47	0.30	−0.31	0.30	−0.55	−17.48	0.57	−0.20	17.06	−0.28	−0.28	0.44
Solar radiation horizon	−2.03	−0.79	−2.64	−0.81	−0.29	0.06	−1.55	0.27	−1.44	−2.40	−1.56	−0.56	−3.31	−0.57	0.44	−2.03
Season winter	0.41	−0.28	0.05	0.09	−0.10	−0.38	−1.24	0.12	0.33	0.42	−0.52	−0.21	1.48	2.88	0.47	−0.05
Morning	−0.27	0.13	2.78	0.07	−1.08	−0.26	0.00	0.00	0.74	−16.95	−0.50	0.04	19.03	0.45	0.54	0.35
Illuminance level	1.01	0.24	−0.29	0.24	−1.56	1.50	0.74	0.21	0.43	−2.55	1.29	0.36	−4.77	−0.38	0.17	−1.04
Wind velocity	2.14	0.97	1.54	0.58	−0.08	0.42	0.34	1.52	1.43	2.97	−0.40	0.16	−1.20	−2.44	−0.30	−0.28
Wind direction	−0.38	−0.60	0.82	0.02	0.07	−0.19	−0.29	−0.50	−0.05	1.86	0.37	0.05	0.40	−0.11	0.61	0.36

assigned to this cluster were associated to a *thermal-driven* window opening behavior.

- Opening Cluster 3 appeared to be more influenced by time-dependent parameters such as office arrival time (2.65) and time of the day (2.1) than physical parameters. This cluster of behavior tend to open windows as a response to preference and attitudes which were psychological (preference and attitudes) and contextual more than physical drivers. Offices assigned to this cluster were therefore associated to a *time-driven* window opening behavior.
- Opening Cluster 2 was mainly driven by a combination of a physical parameter such as indoor air temperature (3.51) and psychological and contextual factors such as office arriving time (2.53). Offices assigned to this cluster were therefore associated to a *thermal-time driven* window opening behavior.
- Closing Cluster 1 was mainly influenced by indoor air temperature (4.93) and outdoor air temperature (3.87) when closing windows and time-dependent parameters were significant but secondary driving forces. Hence, offices assigned to this cluster were associated to a *thermal-driven* window closing behavior.
- Closing Cluster 2 was mainly influenced by time-dependent parameters such as time of the day (3.34) and office leaving time (3.23) than physical parameters. Offices assigned to this cluster were therefore associated to a *time-driven* window closing behavior.

Occupancy presence clearly emerged as one of the top five influencing factors for both window opening and closing actions.

2.5.2. *Window opening duration behavioral patterns*

The two-year data set was organized based on the number of hours the window state was recorded to be open in one day in each of the 16 monitored offices. The optimal k-means algorithm, validated by means the Davies–Bouldin index, grouped $k_{opt} = 4$ clusters of window opening duration during the four seasons of the year. Hence, three window opening *duration patterns* were clustered in the data set (Fig. 5):

- Long Openings: 19% offices assigned (E10, E04, W05)
- Medium Openings: 31% offices assigned (E07, W01, E08, E05, E09)
- Short Openings: 50% offices assigned (W04, E02, W03, E06, E01, E11, W02, E03)

四维（四季）空间点的聚类

Generally, window is kept open for longer periods during summer months and for shorter periods during winter months. Even following this tendency, office E10 was labeled as isolated cluster with respect to the average window opening duration in the data set records. Office E10 presented extreme window opening *duration patterns* where the window position was recorded (i) open almost all day long during summer months, (ii) around 16 h per day during autumn and spring and, (iii) around 12 h per day during the winter season. For simplicity to further consideration, this cluster was incorporated to the closest cluster and associated to the long openings behavioral pattern.

The variation of the average duration for which the window was kept open in every office ranged from 0.04 h/day (office E04) to 6 h/day (office E03) and not considering the extreme case (office E10, in which window state is recorded open on average for more than 17.2 h/day).

- Long Openings: windows stay open for an average of 6–17.2 h per day
- Medium Openings: windows stay open for an average of 1–2.2 h per day
- Short Openings: windows stay open for an average of less than 0.7 h per day

2.5.3. *Window interactivity behavioral patterns*

The same two-year data was reorganized based upon the average number of window state changes in one day, for each of the 16 offices. The optimal k-means algorithm, validated by means the Davies–Bouldin index, grouped $k_{opt} = 3$ clusters of window interactivity behavioral patterns during the four seasons of the year. Great variation among the number of daily window interaction was found among seasons of the year even in a same office. For these reasons, the number of daily window position changes during winter, summer, spring and autumn was used as indicators of the office user interactivity with the natural ventilation system.

Three *interactivity* behavioral patterns were clustered in the data set (Fig. 6):

- Active Operation: 31% offices assigned (E02, E04, E07, E08, W01)
- Neutral Operation: 25% offices assigned (E05, E06, E09, W05)
- Passive Operation: 44% offices assigned (E01, E03, E10, E11, W02, W03, W04)

The average number of changes varies from 0.04 to 3.8 changes per day.

Table 7
Calculated variables and coefficients' impact factors for window closing probability.

	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10	E11	W01	W02	W03	W04	W05
Intercept	−15.9	−4.52	−12.55	−4.74	−3.93	−11.52	−5.69	−4.60	−4.54	−23.80	−7.71	−5.47	−11.48	1.34	−1.70	−5.53
Occupancy presence	−1.74	−1.52	−2.28	−2.31	−2.19	−1.98	−2.99	−2.05	−1.83	−3.23	−1.78	−2.40	−1.33	−2.91	−3.18	−2.41
Leaving time	−2.05	−2.68	−19.01	−5.13	−4.12	−3.17	−3.04	−3.35	−2.24	−20.52	−1.76	−3.46	−18.35	−16.08	−4.53	−4.68
Evening	−3.62	−3.57	−15.25	−2.72	−3.16	−4.96	−2.47	−5.03	−2.79	−1.94	−1.39	−17.20	1.24	−2.54	−3.00	−17.19
Season spring	0.72	−1.53	−2.25	−0.34	−0.33	−0.45	−0.79	−0.65	−1.25	16.90	−1.21	−0.71	−1.03	2.46	0.24	−0.38
Early Morning	0.04	−1.61	0.06	0.84	−1.95	0.30	0.88	0.69	0.77	1.10	0.42	0.45	4.09	1.00	1.16	1.34
Illuminance. level	0.36	0.70	0.09	1.12	1.57	1.07	1.54	0.20	1.53	−0.02	2.77	−1.29	5.14	−0.01	0.30	0.69
Season winter	0.31	−1.69	−0.02	−1.18	0.26	0.12	−1.00	−0.97	−0.59	−0.63	−2.31	−0.96	−14.60	0.92	−15.39	−0.84
Season summer	0.17	−0.32	−0.68	−0.56	−0.49	−0.52	−1.30	−0.21	−0.37	17.78	−1.43	−0.55	1.01	2.73	0.22	−0.50
Noon	0.21	0.15	0.77	0.41	0.28	0.42	0.18	0.52	−0.72	−0.50	1.41	0.64	3.70	0.28	0.62	0.94
Air temperature	8.65	2.78	1.54	0.77	1.41	5.78	3.20	0.76	1.24	0.99	2.72	2.75	−8.56	−1.31	2.33	2.42
Morning	0.12	0.03	1.80	−0.09	0.27	−0.43	0.18	0.32	−0.26	−0.03	1.26	0.83	4.38	1.48	1.43	0.93
Outdoor temperature	−0.37	−1.01	1.74	−0.38	−1.85	3.12	0.53	−0.24	−1.16	−2.84	−0.78	−1.12	11.02	−0.86	0.39	−1.11
Solar radiation horizon	0.00	1.05	3.43	−0.04	−0.41	0.73	0.12	0.69	0.17	1.48	−1.00	0.82	−1.51	−0.21	0.45	−0.20
Wind velocity	0.63	0.21	3.07	1.55	0.63	−0.09	−0.06	0.80	1.29	0.89	−1.90	−0.98	0.70	−2.04	−1.18	0.17
Outdoor RH	0.19	−0.47	1.42	0.23	0.19	1.37	0.67	0.01	0.12	−2.20	−2.23	0.36	−1.21	−2.21	−2.66	0.33
Wind direction	−0.15	−0.34	0.35	−0.15	0.17	−0.46	0.11	−0.23	0.07	0.01	1.47	−0.10	1.55	−0.66	0.25	−0.17

Table 8
Discerned behavioral patterns.

Patterns of behavior	Data mining	Parameters
Motivational	Window opening/ closing drivers	Coefficients and variables (statistical analysis)
Opening duration	Window state	*h* window open or close/day
Interactivity	Window changes	*n* changes/day
Position	Window degree of opening	Tilting angle (from 0 to 1)

Table 9
Clustered top five influencing variables and coefficients for window opening probability.

Opening Cluster_1		Opening Cluster_2		Opening Cluster_3	
Variables	Coeff.	Variables	Coeff.	Variables	Coeff.
Indoor air temperature	−6.49	Indoor air temperature	−3.51	Arriving time	−2.65
Outdoor air temperature	−2.25	Arriving time	−2.53	Early morning	−2.1
Arriving time	−2.01	Early morning	−2.21	Air temperature	−1.64
Occupancy presence	−1.6	Occupancy presence	−1.82	Outdoor air temperature	1.57
Early morning	−1.09	Evening	−1.41	Occupancy presence	−1.48

- Active Operation: window position changes on the average from 2.1 to 3.8 times per day
- Neutral Operation: window position changes on the average from 1 to 1.7 times per day
- Passive Operation: window position changes on the average from 0 to 0.7 times per day

Table 10
Clustered top five influencing variables and coefficients for window closing probability.

Closing Cluster_1		Closing Cluster_2	
Variables	Coeff.	Variables	Coeff.
Air temperature	−4.93	Evening	−3.34
Outdoor air temperature	−3.87	Occupancy leaving	−3.23
Evening	−2.81	Occupancy presence	−2.31
Occupancy presence	−1.93	Indoor air temperature	1.19
Occupancy leaving	1.45	Outdoor air temperature	0.75

2.5.4. *Window position behavioral patterns*

The sole parameter of the number of window changes or the duration of the window state, was not indicative of the user preference regarding natural ventilation in indoor environment. Accordingly, the same two-year data was organized based on the most frequent window tilting angle position (where 0 indicates window totally closed and 1 window totally opened) recorded for each of the 16 offices. Hence, the most frequent window tilting angles were clustered into three *window position* behavioral patterns, named as small, intermediate and big opening (Fig. 7). Outlier behavioral patterns (uncommon window opening position) were isolated and associated to the most extreme behavioral pattern for further considerations (big openings).

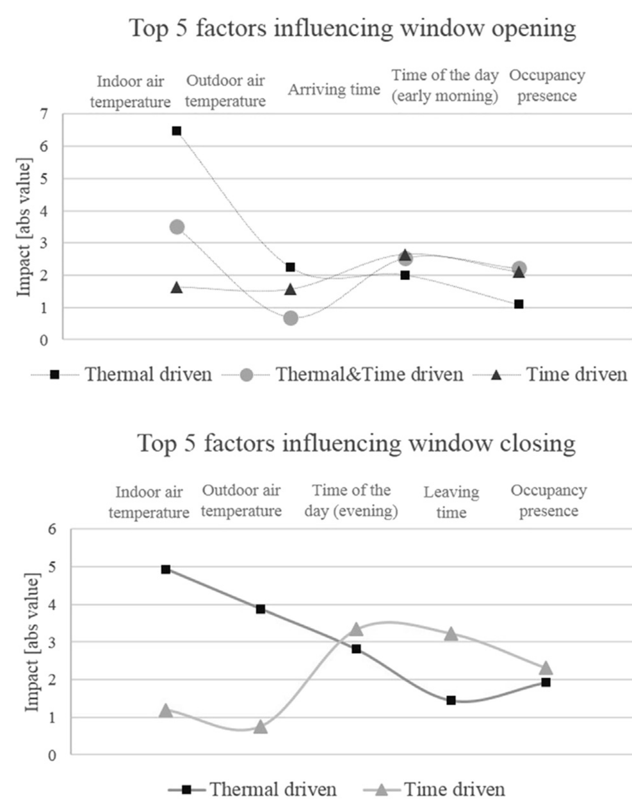

Fig. 4. Top 5 influencing factors for window opening and closing.

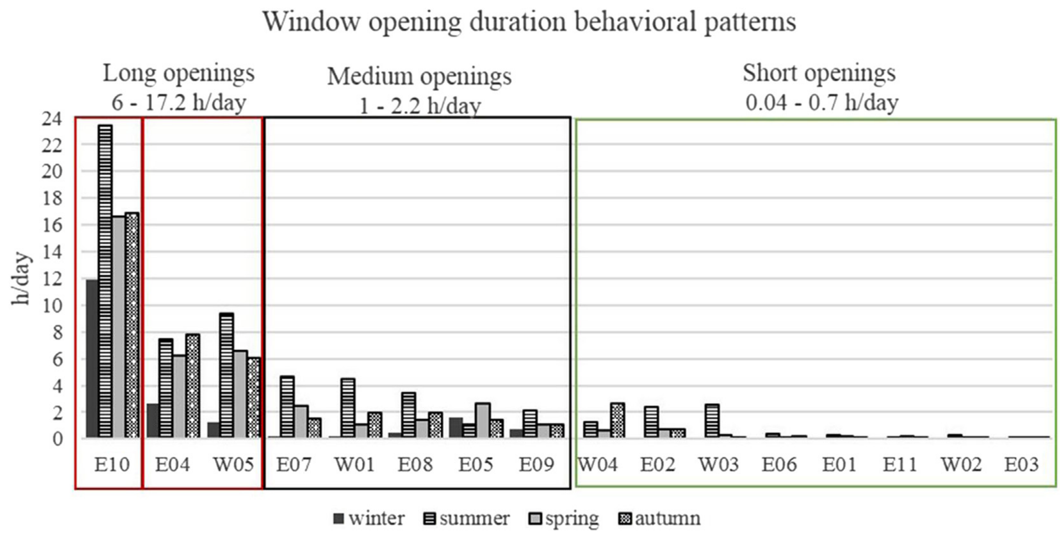

Fig. 5. Window opening duration behavioral patterns in 16 offices.

- Small Openings: 50% offices assigned (E01, E02, E03, E06, E09, E11, W02, W03)
- Intermediate Openings: 25% offices assigned (E05, E07, E08, W01, W04)
- Big Openings: 25% offices assigned (E04, E10, W05)

The average recorded window tilting angle varied based upon the hour of the day.

- Big Openings: window tilting angle position varied on the average from 0.8° around noon to 0.1 during night time.
- Intermediate Openings: window tilting angle position varied on the average from 0.6° around noon to a total close position during night time.
- Small Openings: window tilting angle position varied on the average from 0.3° around noon to a total close position during night time.

Interestingly, the typical window tilting angle of single offices varied broadly, when the data set was broken down into seasons. For these reasons, the window tilting angle recorded during winter, summer, spring and autumn was used to draw schedules of the window opening positions (values from 0 = totally closed to 1 = totally open) over the 24 h of the day, for the four season of the year (Fig. 8).

The findings presented in Fig. 7 allow for the patterns of window tilting angle preferences on energy use and design of natural ventilation in office buildings, to be considered in future building energy modeling programs. The discerned schedules, sorted by season, day of the week and time of the day, represent more robust inputs for building energy modeling programs, like EnergyPlus [50] or IDA-ICE [51].

2.6. *Association rules mining among behavioral patterns*

Based on the information gained from the cluster analysis conducted in this study, each office was associated to a *motivational, duration, interactivity* and *position* behavioral pattern concerning window use (Table 11).

Association rules were mined with the objective to extract frequent and meaningful correlations among the four window behavioral patterns. The *frequent pattern growth algorithm* (FP growth) was the most commonly used algorithm to discover patterns into a given data set. The FP-growth algorithm was employed along with the open source data mining program Rapid Miner to mine the association rule mining (ARM) analysis.

In order to obtain significant results from the ARM analysis, support of 30%, confidence of 80% and a lift of 1, were set as the minimum thresholds. Such criteria indicated that for each association rule mined, at least 30% of all the data records in the given data set contained both premise and conclusion, with the probability that a specific premise lead to a specific conclusion was more than 80%. Moreover, all of the rules mined had positive correlations (lift > 1). Such mining generated 12 rules which provided useful information for the demonstration purposes in this study (Table 12).

From the information gained by the 12 rules mined, two typical working user profiles can be drawn:

- *User* α was a working user type (rules 1, 2, 3, 4, 5, 6, 10, 11) which tended to open the window for short periods of time (0.04–0.7 h/day), interacting on the average in between 0.7 and 0.04 times per day (passive operation) and usually preferred small openings (<0.3° of tilting angle). Moreover, users α was mainly influenced by thermal parameters both when opening and closing windows (rule 9).
- User β was a working user type (rule 8, 12) whom tended to open the window on the average from 1 to 2.2 h per day (medium openings), interacting on the average in between 1.0 and 1.7 times per day (neutral operation) and usually preferred intermediate openings (<0.6° of tilting angle). Moreover, user β was mainly influenced by time-dependent parameters both when opening and closing window (rule 7).

3. Discussion

In a view of the complexity of human behavior, distinguishing singular diversity in big office building becomes a challenging task. Parameter selection methods such as regression and correlation analysis are commonly utilized to identify the factors influencing occupant behavior in buildings and to cluster driver-response conditioning behavioral patterns. The strength of these statistical analysis techniques is their widespread familiarity among researcher and data analysts. Nonetheless, their outcomes are

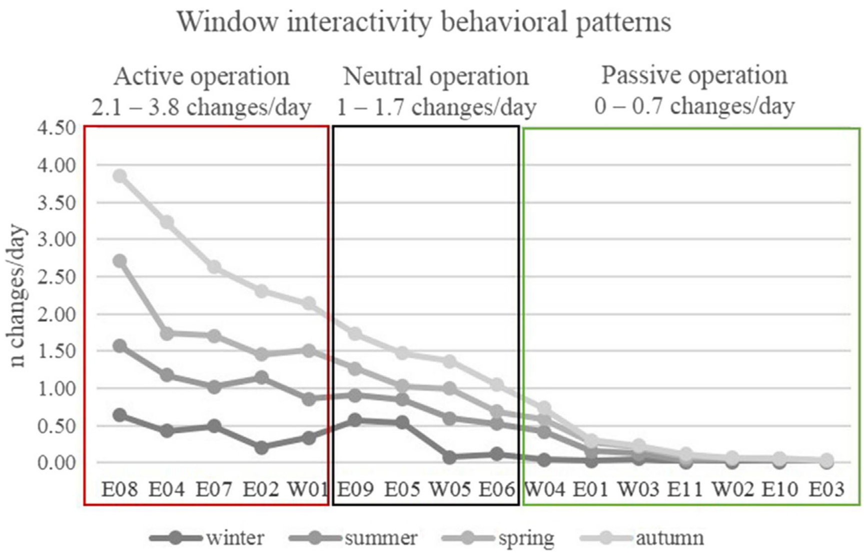

Fig. 6. Window interactivity behavioral patterns in 16 offices.

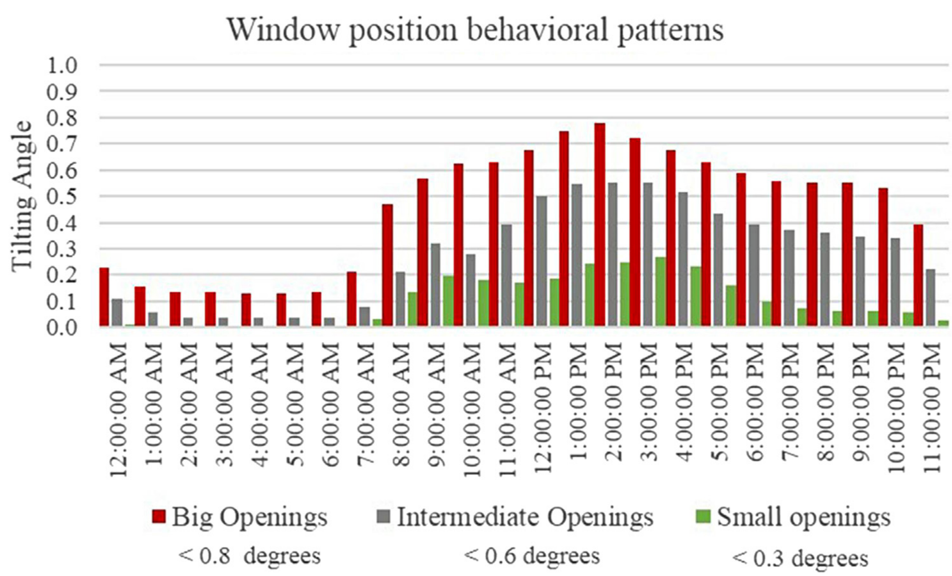

Fig. 7. Window position behavioral patterns in 16 monitored offices.

usually complex equations which may not be easily understandable and interpreted especially for non-expert users without advanced statistical knowledge (i.e. building operators and managers, building designers, energy modelers). Statistical analysis helps to identity repetitive behaviors, which may or may not be significant in terms of schedules of operation incorporated into energy models. Moreover, "standard" behavior does not exist in the real world, and the concept of pattern encompass much more than what is normally defined as expressions describing the most frequent behaviors in a building.

In a view of these facts, our data mining framework suggests an improvement of the notion of behavioral patterns not only as statistical relevant driver-response conditioning clusters, but also incorporating the motivational dimension with typical window opening habits. In this context, cluster analysis gain information from key determinants for individual behavior by revealing a set of rules which may allow more accurate assumption on group behaviors overcoming the lack of personalization of statistical methods. In a view of these facts, nevertheless the mined patterns of ventilation behavior are circumstantial to the given data set, the proposed framework was conceived generic enough to provide solutions to represent the diversity of typical office user profiles in real buildings. The further implementation of the discerned user profiles into building energy simulation tools provides an opportunity to establish an experience base for the assessment of real obtainable energy savings in buildings, equally in the design, retrofit and operation and maintenance contexts as well as for driving future energy policies (Fig. 9).

3.1. From driving factors to motivational patterns of behavior

Factors influencing window opening and closing, which could be named under the general term "drivers", are the stimuli leading to a reaction in the building occupants in ways to restore their comfort with respect to natural ventilation.

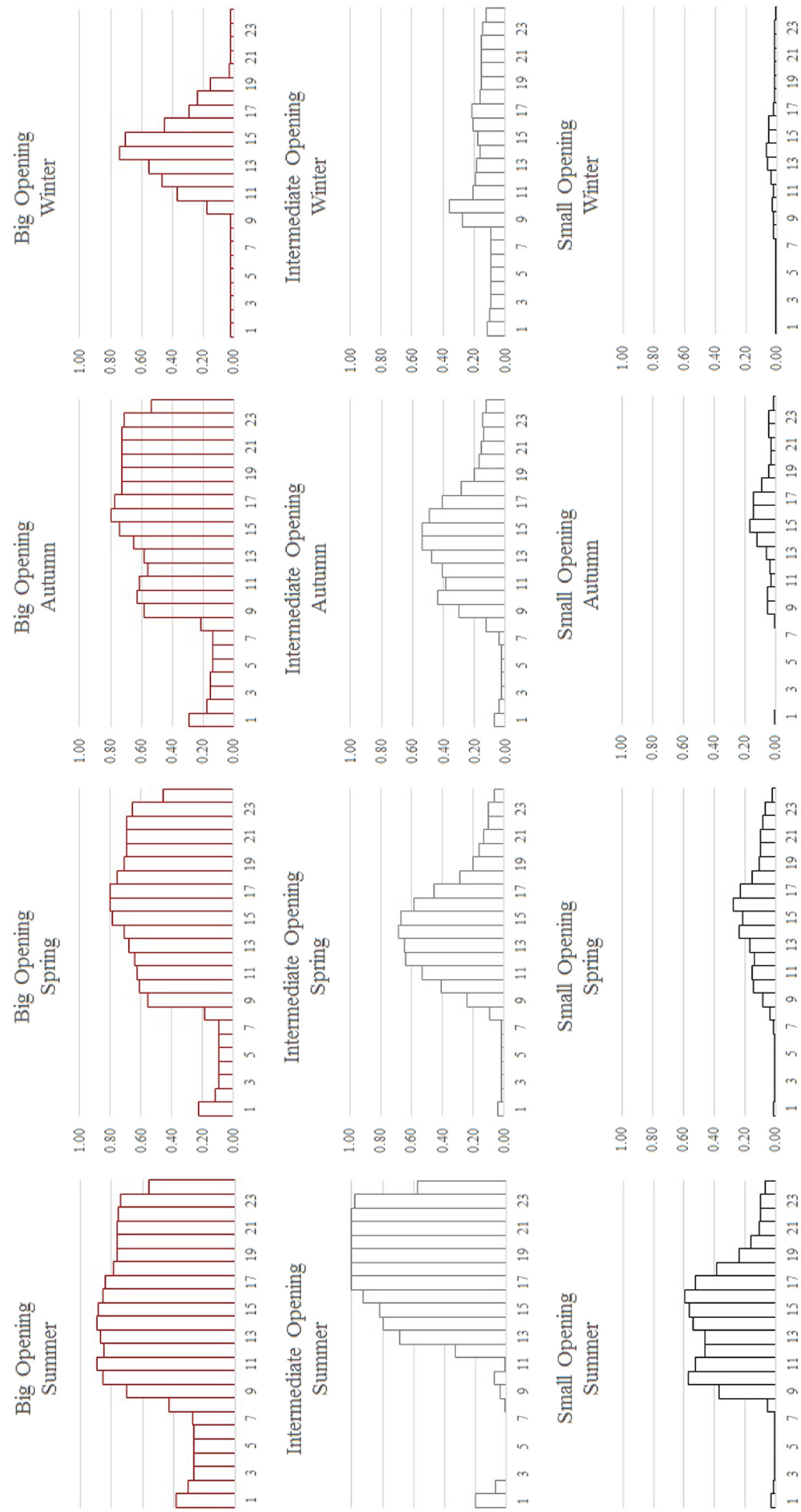

Fig. 8. Behavioral pattern schedules for window tilting angle.

Table 11
Clustered behavioral patterns in 16 offices.

Office	Motivational		Duration	Interactivity	Position
	Window opening	Window closing	Window state	Window change	Window tilting angle
E01	Thermal driven	Thermal driven	Short openings	Passive operation	Small openings
E02	Thermal/time driven	Thermal driven	Short openings	Active operation	Small openings
E03	Thermal driven	Thermal driven	Short openings	Passive operation	Small openings
E04	Thermal driven	Thermal driven	Long openings	Active operation	Big openings
E05	Thermal/time driven	Time driven	Medium openings	Neutral operation	Intermediate openings
E06	Time driven	Time driven	Short openings	Neutral operation	Small openings
E07	Time driven	Time driven	Medium openings	Active operation	Intermediate openings
E08	Time driven	Time driven	Medium openings	Active operation	Intermediate openings
E09	Thermal/time driven	Thermal driven	Medium openings	Neutral operation	Small openings
E10	Time driven	Time driven	Long openings[a]	Passive operation	Big openings[a]
E11	Thermal driven	Thermal driven	Short openings	Passive operation	Small openings
W01	Time driven	Time driven	Medium openings	Active operation	Intermediate openings
W02	Thermal driven	Thermal driven	Short openings	Passive operation	Small openings
W03	Time driven	Time driven	Short openings	Passive operation	Small openings
W04	Thermal/time driven	Time driven	Short openings	Passive operation	Intermediate openings
W05	Thermal/time driven	Time driven	Long openings	Neutral operation	Big openings

[a] Outlier.

Window operation is not only influenced by perceived thermal condition, but it is also seen as a response of sensed indoor air quality, external (outdoor temperature, solar radiation, wind speed, rain) and internal (indoor temperature) environmental conditions as well as contextual factors (window type, time of the day, season of the year) and personal and cultural preferences.

Different time scales of time dependent parameters such as 1) season of the year time 2) day of the week and 3) time of the day, were included in the statistical analysis as predictor of the window opening and closing probability. Moreover, window and occupancy were expressed in terms of 4) window state (open/closed) 5) occupancy state (present/vacant) and 6) window and 7) occupancy change of state. These predictors, even if closely related to the same parameters, were not surrogates of the others and were not duplicative of the same action. Instead, they were indicators of time dependence, occupant presence and movement respectively. Altogether they describe the intricate dynamics of different occupant behaviors in buildings. In our view, this overlap provides clarity in describing the complexity of occupant behavior and addresses the inadequacy of current practices based upon simplistic standardized schedules and input.

From the analysis it emerged that top drivers for window opening and closing were physical thermal (indoor air temperature, outdoor air temperature) and time-dependent contextual (time of arriving and leaving the office) parameters, apart from occupancy presence. These results strengthen the belief that not only physical factors, such as indoor and outdoor environmental parameters, influence human energy behavior, but also non-physical drivers, such as personal preference, habit, context and attitude, play an important role in understanding occupant behavior.

The results demonstrated that, in the specific office buildings, three motivational patterns of window opening (*thermal-driven*, *time-driven*, *thermal-time driven*) and two motivational patterns of window closing (*thermal-driven*, *time-driven*) stimulated an occupant to open a window.

3.2. From occupant behavior to user profiles

Clustering procedures were employed in order to analyze different aspects of the window opening and closing behavior. The goal was to estimate why, for how long, how often and how much similar patterns of occupant open and close windows in offices of the same building. In this aim, the research was clustering 1) *motivational*, 2) *opening duration*, 3) *interactivity* and 4) degree of opening *position* behavioral patterns which would further constitute a base for association rules segmenting the building occupants into attitudinal typical working user profiles. From the information gained by the 12 rules mined, two typical working user profiles were drawn. *User* α was a mainly *physical environmental driven* working user type which tends to open the window for short periods of time (0.04–0.7 h/day), interacting infrequently (on the average in between 0.7 and 0.04 times per day and usually preferred small openings (<0.3° of tilting angle). On the other side, *user* β was mainly *contextual driven* working user type which

Table 12
Association rules mining of behavioral patterns.

Rules	Premise	Conclusion	Support	Confidence	Lift
1	Window tilting angle_small openings, window closing_thermal driven	Window state_short openings	0.31	0.83	1.67
2	Window state_short openings, window change_passive operation	Window tilting angle_small openings	0.31	0.83	1.67
3	Window closing_thermal driven	Window tilting angle_small openings	0.38	0.86	1.71
4	Window change_passive operation	Window state_short openings	0.38	0.86	1.71
5	Window tilting angle_small openings	Window state_short openings	0.44	0.88	1.75
6	Window state_short openings	Window tilting angle_small openings	0.44	0.88	1.75
7	Window opening_time driven	Window closing_time driven	0.38	1	1.78
8	Window tilting angle_ intermediate openings	Window closing_time driven	0.31	1	1.78
9	Window opening_thermal driven	Window closing_thermal driven	0.31	1	2.29
10	Window state_short openings, window closing_thermal driven	Window tilting angle_small openings	0.31	1	2
11	Window tilting angle_small openings, window change_passive operation	Window state_short openings	0.31	1	2

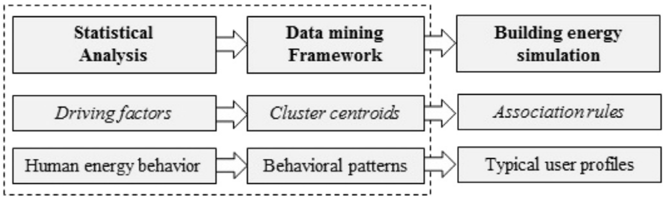

Fig. 9. Schema explaining actual and further steps of the proposed methodology.

tended to open the window for longer periods (on the average from 1.0 to 2.2 h per day), interacting more frequently (on the average in between 1 and 1.7 times per day) and usually preferred intermediate openings (<0.6° of tilting angle).

4. Conclusions

A framework combining statistical analysis with two data-mining techniques, clustering and association rules, was employed to identify occupant behavior patterns of window opening and closing in a natural ventilated office building in Germany, using detailed time-interval measured building data.

Goal of the research was to identify 1) *motivational*, 2) *opening duration*, 3) *interactivity* and 4) degree of opening *position* behavioral patterns. In this aim, four aspects of window operations were clustered:

1. three (*thermal-driven*, *thermal/time-driven*, *time-driven*) motivational patterns clustering the factors driving window opening and closing behavior according to the impact that the factors play on the two actions
2. three (*long*, *medium*, *short*) opening duration patterns clustering occupant behavior based on the number of hours the window state was recorded open every day
3. three (*active*, *neutral*, *passive*) interactivity patterns clustering occupant behavior based on the number of window position changes recorded every day
4. three (*small*, *intermediate*, *big*) opening position patterns clustering occupant behavior according to the most frequent window degree of opening every day.

Analysis of the results indicated indoor air temperature, outdoor air temperature, time of the day (office arriving time and early morning) and occupancy presence are the top drivers for window opening. On the other hand, indoor air temperature, time of the day (office leaving time and evening), occupancy presence and outdoor air temperature emerged as top drivers for window closing.

The four behavioral patterns were further mined using association rules to produce two typical window opening office user profiles, one mainly *physical environmental driven* and one mainly *contextual driven*. The results indicated that office users interact with windows principally driven by thermal discomfort (indoor air temperature) but also behave accordingly to daily routine (time of the day) and habits (arriving and leaving time). The implications of these findings suggest that occupant behavior was somewhat predictive and subject to the constraints or motivating factors of thermal comfort and time management.

From the association rule, it emerged that when interacting with windows to restore the indoor environmental quality, users mainly driven by *physical environmental parameters* had less impact on natural ventilation than users driven by *contextual factors and habits*, opening windows for shorter periods of time, interacting less frequently and usually preferring smaller openings.

In the bigger picture this implies that behavioral patterns are not only statistical relevant driver-response conditioning clusters, but also incorporate the motivational dimension with typical window opening habits. In a view of these facts, any improvement of the notion of behavioral patterns associating the driver-response conditioning motivational dimension with typical window opening habits data mining, overcoming the lack of personalization of statistical methods, is strongly required in order to bridge the gap between predicted and actual building energy performance.

Occupants in naturally ventilated buildings are demonstrated to accept and actually prefer a significant wider range of temperatures compared to users of mechanically ventilated buildings, positively influencing their satisfaction with working spaces and leading them to higher productivity. However, while providing manual ventilation opportunities seems to be beneficial, in doing so the behavior of the occupants gained a larger degree of influence on the indoor environment and energy performance, especially when a robust variation of motivations leading to window opening and closing, duration and number of opening and typical degree of opening was demonstrated. In this view, the persistent patterns of operation and non-homogeneous working user profiles drawn by this study could be broadly applied in further studies to:

1) provide more accurate assumption of actual natural ventilation scenarios in big office buildings that may allow building designers and operating manager to tailor more efficient and robust control strategies and system and envelope design;
2) quantify the energy and economic impacts of diverse ventilation office user profiles in a block of buildings, as well as the sensitivity of physical-environment and contextual time-dependent influencing factors on occupancy, space optimization, thermal comfort and productivity in offices;
3) deliver a set of behavioral rules at the office level to direct specific operation and maintenance and ventilation energy saving strategies with a high replication potential and low capital investment, as well as future energy saving policy in the commercial building sector.

Acknowledgment

This work was sponsored by the U.S. Department of Energy (Contract No. DE-AC02-05CH11231) under the U.S.-China Clean Energy Research Center for Building Energy Efficiency. The authors very appreciated Marcel Schweiker and Andreas Wagner of Karlsruhe Institute of Technology, Germany for sharing the dataset and answering our questions. This work is also part of the research of Annex 66, Definition and Simulation of Occupant Behavior in Buildings, under the International Energy Agency Energy in Buildings and Communities Program.

References

[1] Hong T, Lin HW. Occupant behavior: impact on energy use of private offices. In: ASim 2012–1st Asia conference of international building performance simulation association; 2012.
[2] Haldi F, Robinson D. On the behaviour and adaptation of office occupants. Build Environ 2008;43:2163–77.
[3] Karjalainen S. Thermal comfort and use of thermostats in Finnish homes and offices. Build Environ 2009;44:1237–45.
[4] Masoso OT, Grobler LJ. The dark side of occupants' behaviour on building energy use. Energy Build 2010;42:173–7.
[5] Bordass W, Cohen R, Field J. Energy performance of non-domestic buildings: closing the credibility gap. In: Building performance congress. Frankfurt, Germany: IEECB'04; 19–22 April 2004.
[6] Nordford LK, Socolow RH, Hsieh ES, Spadaro GV. Two-to-one discrepancy between measured and predicted performance of a "low energy" office building: insights from a reconciliation based on the DOE-2 model. Energy Build 1994;21:121–31.

[7] Nicol JF, Humphreys M. A stochastic approach to thermal comfort occupant behavior and energy use in buildings. ASHRAE Trans 2004;110:554–68.
[8] Andersen RV, Toftum J, Andersen KK, Olesen BW. Simulation of the effects of occupant behavior on indoor climate and energy consumption. In: Proceedings of Clima2007: 9th REHVA world congress: wellbeing indoors, Helsinki, Finland; 2007.
[9] Peng C, Yan D, Wu R, Wang C, Zhou X, Jiang Y. Quantitative description and simulation of human behavior in residential buildings. Build Simul 2012;50:85–94.
[10] Wang C, Yan D, Jiang Y. A novel approach for building occupancy simulation. Build Simul 2001;4:149–67.
[11] Sun K, Yan D, Hong T, Guo S. Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration. Build Environ 2014;79:1–12.
[12] Chang WK, Hong T. Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data. Build Simul 2013;6(1):23–32.
[13] Page J, Robinson D, Morel N, Scartezzini JL. A generalized stochastic model for the simulation of occupant presence. Energy Build 2008;40:83–98.
[14] Pfafferott J, Herkel S. Statistical simulation of user behaviour in low-energy office buildings. Sol Energy 2007;81:676–82.
[15] Karjalainen S. Gender differences in thermal comfort and use of thermostats in everyday thermal environments. Build Environ 2007;42:1594–603.
[16] Nicol JF. Characterizing occupant behavior in buildings: towards a stochastic model of occupant use of windows, lights, blinds, heaters and fans. In: Proceedings of the 7th international IBPSA conference, Rio 2; 2001. p. 1073–8.
[17] Raja IA, Nicol JF, McCartney MA, Humphreys M. Thermal comfort: use of controls in naturally ventilated buildings. Energy Build 2001;33:235–44.
[18] Humphreys MA. Field studies of thermal comfort compared and applied. J Inst Heat Vent Eng 1976;44:5–27.
[19] Wagner A, Gossauer E, Moosmann C, Gropp T, Leonhart R. Thermal comfort and workplace occupant satisfaction – results of field studies in German low energy office buildings. Energy Build 2007;39:758–69.
[20] De Dear R, Brager G, Cooper D. Developing an adaptive model of thermal comfort and preference. Final report ASHRAE RP-884. 1997.
[21] IEA Annex 53 Task Force. Total energy use in residential buildings – the modeling and simulation of occupant behavior. Final report. 2012.
[22] Fabi V, Andersen RV, Corgnati SP, Olesen BW. Occupants' window opening behaviour: a literature review of factors influencing occupant behaviour and model. Build Environ 2012;58:188–98.
[23] Rijal HB, Tuohy P, Humphreys MA, Nicol JF. Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings. Energy Build 2007;39:823–36.
[24] Brundrett GW. Ventilation: a behavioural approach. Int J Energy Res 1997;1:289–98.
[25] Warren PR, Parkins LM. Window-opening behaviour in office buildings. ASHRAE Trans 1984;90:1056–76.
[26] Inkarojrit V, Paliaga G. Indoor climatic influences on the operation of windows in a naturally ventilated building. In: Proceedings of Plea2004. The 21th conference on passive and low energy architecture; 2004.
[27] Haldi F, Robinson D. Interactions with window openings by office occupants. Build Environ 2009;44:2378–95.
[28] Andersen RV, Olesen BW, Toftum J. Modelling window opening behaviour in Danish dwellings. Proceedings of indoor air 2011: the 12th international conference on indoor air quality and climate, Austin, Texas.
[29] Herkel S, Knapp U, Pfafferott J. Towards a model of user behaviour regarding the manual control of windows in office buildings. Build Environ 2008;43:588–600.
[30] Roetzel A, Tsangrassoulis A, Dietrich U, Bushing S. A review of occupant control on natural ventilation. Renew Sustain Energy Rev 2010;14:1001–13.
[31] Borgeson S, Brager G. Occupant control of windows: accounting for human behavior in building simulation. Internal Report. Center for Environmental Design Research; 2008.
[32] Fritsch R, Kohler A, Nygard-Ferguson M, Scartezzini JL. A stochastic model of user behaviour regarding ventilation. Build Environ 1990;25:73–81.
[33] Santin OG. Behavioural patterns and user profiles related to energy consumption for heating. Energy Build 2011;43:2662–72.
[34] Wymelenberg VDK. Patterns of occupant interaction with window blinds: a literature review. Energy Build 2012;51:165–76.
[35] Yun GY, Steemers K, Baker N. Natural ventilation in practice: linking façade design, thermal performance, occupant perception and control. Build Res Inf 2008;36:608–24.
[36] Yun GY, Steemers K. Time-dependent occupant behaviour models of window control in summer. Build Environ 2008;43:1471–82.
[37] Van Raaij WF, Verhallen TM. Patterns of residential energy behavior. J Econ Psychol 1983;4:85–106.
[38] Han J, Kamber M. Data mining: concepts and techniques. Morgan Kaufmann Publishers; 2001.
[39] Hand DJ, Mannila H, Smyth P. Principles of data mining. MIT Press; 2001.
[40] Cabena P, Hadjinian P, Stadler R, Verhees J, Zanasi A. Discovering data mining: from concept to implementation. Prentice-Hall Inc; 1998.
[41] Yu Z, Haghighat F, Fung BCM, Yoshino H. A decision tree method for building energy demand modeling. Energy Build 2010;42:1637–46.
[42] Yu Z, Fung BCM, Haghighat F, Yoshino H, Morofsky E. A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy Build 2011;43:1409–17.
[43] Yu Z, Haghighat F, Fung BCM, Zhou L. A novel methodology for knowledge discovery through mining associations between building operational data. Energy Build 2012;47:430–40.
[44] Reinhart CF. Lightswitch-2002: a model for manual and automated control of electric lighting and blinds. Sol Energy 2004;77:15–28.
[45] Rijal HB, Tuohy P, Humpreys MA, Nicol JF, Samuel A, Raja IA, et al. Development of adaptive algorithms for the operation of windows, fans, and doors to predict thermal comfort and energy use in Pakistani buildings. ASHRAE Trans 2008;114:555–73.
[46] Nicol JF. Characterising occupant behavior in buildings: towards a stochastic model of occupant use of windows, lights, blinds, heaters and fans. In: Proceedings of the seventh international IBPSA conference. Rio de Janeiro, Brazil; 2001.
[47] McCullagh P. Generalized linear models. Eur J Oper Res 1984;16(3):285–92.
[48] R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.
[49] Rapid–I–Rapid miner, V 6.0 http://rapid-i.com/content/view/181/190/.
[50] EnergyPlus V 8.1.0 http://apps1.eere.energy.gov/buildings/energyplus/energyplus_about.cfm.
[51] IDA indoor climate and energy V. 4.0: http://www.equa.se/ice/intro.html.