Experimental and numerical study to evaluate the effect of thermostat settings on building energetic demands during the heating and transition seasons

Walid Belazi, Salah-Eddine Ouldboukhitine, Alaa Chateauneuf, Abdelhamid Bouchaïr

Experimental and numerical study to evaluate the effect of thermostat settings on building energetic demands during the heating and transition seasons

Walid Belazi[1, 2], Salah-Eddine Ouldboukhitine[1, 2], Alaa Chateauneuf[1], Abdelhamid Bouchair[1]

[1.] Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal, F-63000 Clermont-Ferrand, France
[2.] IUT d'Allier, CS82235, 03100 Montluçon, France

Abstract

Occupant behavior towards heating and cooling system setting is a very complex process that has been under investigation in the past years. As most of dynamic energy simulation tools consider energy consumption as fully deterministic with fixed and unrealistic schedules, the ability to predict properly the energy consumption is poor because of occupant interaction with indoor environment. In this study, the occupant in residential buildings is modeled as a probabilistic process. The occupant behavior related to thermostat settings is studied through experimental measurements collected in eleven buildings in France over a period of one year, by monitoring various parameters, including indoor air temperature, ambient temperature, indoor and outdoor relative humidity and indoor CO_2. The occupant attitude was classified into three groups, active, normal and passive, according to the number of setting changes per year. The Logistic regression is adopted to calculate the probability of changing the thermostat setting by an occupant, in terms of different environment parameters. The results yield to a proposed model that can be implemented in simulation software, in order to take into account the occupant behavior in the assessment of realistic energy consumption.

Keywords: Thermostat setting points; Occupant behavior; Probabilistic model; Building energy performance simulation; Logistic regression.

I. Introduction

During building operation, occupants interact with different equipment to satisfy the comfort requirements. Occupant behavior include lighting switch on/off, window opening/closing, thermostat settings, etc. However, the behavior strongly differs from one occupant to another, leading to very large variations in indoor environment and energy consumption [1, 2]. Many investigations show a significant gap between calculated and actual energy consumption, exceeding up to 300% in extreme cases [3]. For this, it is very important to take into account the occupant interactions with indoor environment since the design stage.

Various activities have been modeled such as window opening, lighting and blind use [4-7]. These activities have shown to produce significant changes in energy use. In this study, a probabilistic model is developed to consider the effect of thermostat setting on energy consumption. Traditionally, the temperature regulation is typically modeled through deterministic scenarios in the dynamic thermal simulation tools, though this does not correspond to the real use of heating. For these reasons, several researchers proposed stochastic models to define the behavior patterns to be implemented in simulation software.

In the past decade, the probabilistic approach has been used to model the human behavior, as occupants have important influence on indoor environment. The institute of health watching (InVS) has found that occupants in France spend between 13.5 and 20 hours/day of their time at home [8]. Therefore, opening windows, switching lights and controlling HAVAC systems have important effects on energy consumption.

Recently, Shi and Zhao [9] presented a stochastic model of occupant interactions with windows in 8 residential apartments in Beijing and Nanjing, relating the use of windows to many indoor and outdoor environmental factors, such as outdoor air temperature, outdoor relative humidity, outdoor wind speed, etc. They analyzed the influence of each parameter on the probability of window opening, using univariate and multivariate logistic regression. On the other hand, Haldi and Robinson [5] proposed a model of window opening/closing by the occupants of the LESO building on the basis of 7 years of continuous measurements. Indoor air temperature and many environmental variables have been incorporated into the model. Several modeling approaches have been tested including logistic distributions, Markov chains and continuous time processes. The model has been validated and integrated in dynamic thermal simulation tools. Also, Fritsch et al. [10] proposed a model to predict the window opening angle in winter, on the basis of 8-months measurements taken every half hour in four office rooms facing the south in the LESO building. During the working hours, the transition probabilities between six possible opening angles over four orientations were defined.

In addition, several studies considered other occupant interactions such as blind and thermostat setting in building. Sutter et al. [7] studied the use of Venetian blinds in eight individual offices, every 15 min, over a period of 30 weeks. Their study showed that the number of use of manual systems is three times lower than that of automated systems. In addition, they concluded that when the screen quality is high in terms of emitted brightness level, the probability that the occupant accepts the high level of diffuse reflection increases, benefiting more from the available daylights. Furthermore, Andersen et al. [11] studied the preferences in terms of temperature set, through a campaign of measurements on fifteen dwellings in Denmark in 2008. A linear regression was used to model the thermostatic valve positions of the radiators and the environmental variables (air temperature, relative humidity, wind speed and solar radiations). The results of this study indicate that the most influencing variables in the determination of the thermostatic valve positions are the outdoor temperature, relative humidity and wind speed.

Regarding uncertainties, Simona D'Oca et al. [2] have modeled the occupant behavior in residential buildings by using probabilistic approaches developed to simulate both the effect of thermostat settings and window openings. The aim of this study was to compare mean values of the probabilistic distribution with a singular heating energy consumption value obtained by means of standard deterministic simulations. In order to highlight the effect of occupant behavior on energy consumption, simulations were performed for three different climate locations (Mediterranean, Continental and Nordic) and three comfort category conditions (Categories I, II, III). It can be noticed, that the deterministic approach (deterministic input) underestimated the heating consumption, when compared to probabilistic predictions taking into account the occupant behavior. As example, in Mediterranean climate and for Category of Comfort II, that the probabilistic distribution of heating loads ranges from 46 kWh/m²/year to 68 kWh/m²/year, with a maximum variation of 45%. Unlike, the heating load was 47 kWh/m²/year for the deterministic scenario. Whereas, the variation of the probabilistic distribution of heating loads in the Continental and the Nordic climates are 36% and 26% respectively for the same comfort category (Category II). Otherwise, the largest impact in delivered energy variation from deterministic to probabilistic simulations was observed for category III, with a maximum variation of 61%, 47% and 35% in Mediterranean, Continental and Nordic climates respectively

In general, determinist scenarios are mostly applied to describe occupant activities in buildings. The occupant behavior is a complex and stochastic process [12]. Indeed, several parameters influence the occupant behavior and the energy consumption, such as gender, income, energy cost, health situation, clothing, activity level etc. [13-16]. These studies showed that the energy need is greatly affected by occupant behavior. Moreover, the probabilistic modeling presents a way to quantify the uncertainties and to provide an interval on possible outputs. For example, Haldi and Robinson [5] presented a mathematical method to model the occupant adaptive actions. This mathematical model, describing occupant behavior, needs to be implemented in simulation software. There is a need to consider the occupant interaction with indoor environment in order to decrease the gap between predicted and real energy use.

In the literature, heating profile is important parameter that influences the indoor temperature of residential buildings, with a large variation between different apartments [17-20]. The occupant has strong influence due to his presence and activities in the building on one hand, and due to his control actions that aim to improve indoor environmental conditions, on the other hand [21]. Xu et al. [22] carried out investigations and field observations in China to study how occupants interact with thermostatic radiator valves (TRVs). They found that 26 % of the occupants never changed the set-point and 46 % of them rarely adjusted the TRV setting, which has remained unchanged for several months. The rest of the occupants modified the TRVs frequently. This study showed large variations in the use of thermostats and these differences are the results of occupant habits. Karjalainen [15] conducted a questionnaire survey in Finland in order to investigate gender effects on the adjustment of thermostat

setting. The results have shown significant differences between the genders in terms of thermal comfort and use of thermostats. As a result, females are less satisfied with room temperatures than males and preferred a higher heating set-point than males, but males adjusted the thermostat setting more often than females did.

The studies mentioned above show that the set-point temperature is adjusted by occupants. Therefore, using a constant set-point temperature in the simulations will result in wrong estimations. Indeed, the designers calculate the energy consumption of a building by assuming that the heating set-point is always 19 °C and the cooling set-point is always 26 °C [23]. It is therefore mandatory to consider the variations in occupants' attitudes, preferences in indoor temperature and reactions with the indoor environment.

The development of behavioral models to predict occupant behavior inside buildings is a significant area of focus in the building sector. However, office buildings have been the objective of several surveys and behavior in residential building is not specifically treated. Furthermore, the most of the buildings monitored have been located in the UK, Switzerland and Asia, while France has been scarcely examined. In addition, in the literature, there is a lack of studies that evaluate the impact of thermostat changes in building energy consumption. As well, studies analyzing occupant behavior observed specificities in summer, but seasonal variations in behavior have yet to be taken into account. To answer the identified lacks, the present work provides new experimental results to show how the occupant behavior can impact the building energy use. In that case, models of the thermostat use, in residential buildings, were proposed to evaluate the impact of occupant behavior in building energy consumption when changing the thermostat. The description of the instrumented building is presented with statistical analysis of the obtained data. Afterward, the probabilistic analysis of the building energy performance is carried out to define the interaction patterns. The approach is constituted of two models: the univariate and multivariate logistic model of thermostat changes, as a function of indoor and outdoor environments and for different types of occupant behaviors (active, medium and passive) according to the seasons of the year (heating season and transition seasons). Finally, the study can be considered as a starting point defining the methodology of analysis and its advantages. It should be continued in the future with the purpose to not only show the importance of the probabilistic models of occupants' behavior but also to show how they can be coupled with building energy simulation software.

II. Data and methods

1. Experimental set-up

In this section, the building for which data have been collected is presented in order to provide the basis for the development model. The survey was performed on a collective social building (Figure 1) located at the city of "Cébazat" in France. It is a low energy building with 45.44 kWh/m²/year regarding the climate and geographical area.

The building has 18 apartments of type T2, T3 and T4[1]. It is divided into two sections separated by an expansion joint. Only the eleven apartments (T2 and T3) of the southern part were instrumented due to the limited budget. Table 1 lists the main features of the eleven monitored apartments.

The building is considered as independent building with an important mask at the north side (Figure 1). Built in 2010, the building has a concrete load-bearing structure. The eleven apartments are occupied by one to four people, as indicated in table 1.

1.1 Why considering this building

This building was chosen for several reasons: it is a low-energy building in the sense of the LEB-effinergie label. In addition being a social building, it was built to provide the maximum reduction of construction costs. The orientation North-South, due to the land geometry, is the opposite of bioclimatic logic and monitoring of comfort will make it possible to survey problems of summer overheating. Finally, the owner wishes to make this building a showcase of its commitment to sustainable development and a prototype of his future constructions.

1.2 The envelope

The external envelope consists of the same type of wall found on most of the opaque vertical walls (Table 2). The Reinforced concrete walls insulated with Xtherm Itex 32SE expanded polystyrene from Knauf, with thermal resistance R = 3.40 m².K/W. The walls of unheated area are insulated only from the inside. In several apartments, there are load-bearing walls which are used as partitions. The load-bearing walls are un-insulated and allow thermal transmission through the building. The roof is partially in wooden structure covered with Roman tiles insulated with 35 cm of cellulose wadding (R = 8 m².K/W). The remained roof is covered by green vegetation for which the insulation is provided by rigid polyurethane foam panels with composite facings, or expanded polystyrene (R = 5.15 m².K/W). The first floors are insulated to compensate any possible vacancy of shops in the ground Floor, the insulation is of type Unimat Sol Supra thickness 70 mm (R = 2.60 m².K/W).

Each intermediate floor is composed and insulated as follows:

- Concrete slab;
- 30 mm thick rock wool panels Isover Domisol LR30 on all surfaces;
- Extruded polystyrene foam panels of Isover Styrofoam 2000-A type on all surfaces, U = 0.29 W/ m².K, R = 2 .60m².K/ W;

---

[1] **T** is used for **type**, which corresponds to a type of apartment with a number of main rooms (living room and bedrooms are the main rooms. While, the kitchen and bathrooms are considered as included in all the apartments). In the real estate sector in France, a T2, T3 or T4 is an expression used to designate a category of housing, which can be furnished or not.
T2: Apartment 2 pieces including 1 bedroom and 1living room.
T3: Apartment 3 pieces including 2 bedrooms and 1living room.
T4: Apartment 4 pieces including 3 bedrooms and 1living room.

- Polyane film;
- 5 cm screed;
- Peripheral strip composed of 8 mm thick polyethylene foam.

1.3 Installed equipment

The ventilation is provided by a hygro-B single flow CMV[2] (controlled mechanical ventilation). The building is divided into two parts by an expansion joint. The apartments at the north of this joint are ventilated by a common extraction unit located on the roof. The southern part (the instrumented building) of the joint, the third floor is ventilated by two extraction groups. The heating is provided for each apartment by individual gas condensing boiler. The common areas are not heated. The T2 and T3 have combined boiler / DHW (domestic hot water) with micro accumulation of 23 kW. The building is uncooled. The domestic hot water (DHW) production is provided by individual gas boilers.

2. Data collection

A monitoring system was set up to record environmental parameters. The monitoring started on June 2011 and continues till now.

Table 3 presents the sensors and weather station characteristics. The time step used for data collection is one hour. The data is stored for a period of 21 days on the acquisition system located in the building's technical room. The controllers are connected via Ethernet line to a server located in the local in owner's office, on which supervision software is installed. This software transmits by e-mails the raw data to be processed.

In each of the eleven instrumented apartments, the assessment of thermal comfort of occupants is carried out by measuring the indoor temperature T_{int} (°C), the indoor relative humidity RH_{int} (%) and the CO_2 concentration (ppm). The sensors are placed on the internal walls, about 1.5 m from the ground and at least 50 cm distant from the ceiling, the partition angles and any other dead zone. They are placed far from direct radiations, air streams, heat source and occupant proximity.

The layouts of the four apartments in the first-floor as well as the location of sensors in every measured room are specified in Figure 2. The plans of the two other floors and the sensors location are similar to those of the first floor. Each apartment has a central heating controller which transmits the set-point temperature. The heating season (winter) and the transition seasons (autumn and spring) corresponds to the period during which the central

---

[2] A detailed explication on CMV is available in [24].

heating plant is operational, from September to May. One bedroom among the twelve apartments is equipped with carbon dioxide sensor. These sensors are used both to assess occupancy scenarios of the rooms where the measurements took place, in addition to indoor air quality.

To quantify the actual outdoor climate conditions applied to the building, a meteorological station is installed. This station allows the measurement of the following quantities:

- Dry bulb temperature (°C) and external relative humidity (%)
- Diffuse and direct solar radiation on horizontal surface (W/m²).
- Wind speed (m/s).
- CO_2[3] concentration of the outside air (ppm).

This station is located on the roof of the building, away from solar masks as shown in Figure 3.

All the measurements were carried out in the living room and the bedroom of the apartments. The detail information relating to the location of the sensors of these apartments is shown in table 4. Moreover, occupant interactions with the set-point temperature were gathered by measurements of the most representative zones of the building. In this study, one thermostatic radiator valve (°C) in the living room is installed in each apartment.

III. Method
1. Statistical analysis

The objective of the study is to evaluate the influence on thermostat changing. The controls of the set-point temperature inside the apartments were monitored by the thermostat. Through these thermostat units, the occupants were permitted to change their set-points temperature for the rest of the day.

Figure 4 shows the distribution of heating consumption for each apartment during the heating season and the transition seasons. The heating consumption of the monitored apartments ranges from 441 kWh to 3157 kWh. The ratio between the highest and the lowest consumptions is equal to 6 for type T2 and 7 for type T3. Moreover, regarding the season, important variation can be observed in the heating consumption of the 11 apartments, resulting in a ratio between highest and lowest heating consumption of 25, 25 and 17 in winter, spring and autumn respectively. Consequently, regarding the same apartments, a high variation is observed in heating consumption. These differences must be a result of differences in occupancy and differences in the occupants' behavior.

Large variations in occupant behaviors have been recorded. As shown in table 5 and Figure 5, the total number of thermostat changes strongly varies between apartments. The apartments were grouped according to the number of interactions with thermostat during the monitoring period. Three occupant profiles have been observed:

---

[3] A detailed explanation on outdoor CO2 concentration is available in [25, 26].

Active, Passive and Medium. The probability of thermostat changes can be deduced for three different statistical models. It can be seen that the number of thermostat change is the highest in winter and spring and the lowest in autumn. It can also be noted that the apartment number twelve had the highest number of thermostat changes in winter, while the apartment number two and four had the lowest changes in autumn and spring.

A summary of indoor and outdoor parameters monitored during the survey is reported in Table 6. This gives the variation ranges for each parameter, as well as statistical analyses including the mean, max, min, and standard deviation values. Based on the results of the survey, the living room is considered to be occupied permanently in all the apartments. Apartment 04 had the highest indoor temperatures in winter and spring because of its orientation (East-facing), which provides direct solar radiation only in the morning. In contrarily, apartment 03 has the highest mean indoor temperature in spring (23.5 °C) owing to South-East its orientation which provides direct solar radiation all the day. Indoor humidity is very similar between the apartments, with a mean value of approximately 36% in winter, 42% in spring and 48% in autumn.

Figure 6 presents the distribution of the indoor temperatures during the heating season (winter) and transition seasons (autumn and spring). The median of indoor temperature fluctuate between 18 and 22°C, which indicates that the level of satisfaction in terms of comfort is very different between the apartments. This difference can be explained by household-specific attributes, such as age, income or social status. Contrary to expectations, the apartment orientation has shown low impact on the set-point temperature (table 1 and Figure 6). Moreover, the set temperature increases with the number of people in the apartment since the indoor temperature increase by 1°C to 3°C for identical apartments (same surface and orientation) with a larger number of occupants. Indeed, Kelly et al. [27, 28] showed that the number of occupants in a building have the largest effect on energy consumption. These results prove that for each extra person living in the building, the mean daily internal temperature increases by nearly 0.25 °C.

In addition, the apartment surface has no impact on the set temperature. The apartments of the intermediate floor have lower indoor temperature that in the other floors. Based on these measurements, a thermostat changes model is proposed. This model provides the probability to act on the thermostat of the living room.

2. Stochastic model of occupants interaction with thermostat in residential building

In this section, the model of occupant action is developed using logistic regression. In fact, the logistic regression models are very appropriate to predict probability p of a binary response variable when the explanatory variables $x = (x_1, x_2,…, x_n)$ are at particular values [5].

In our study, the thermostat change is a binary response variable: the value of 1 indicates that the occupant changed the set-point temperature and 0 indicates that no change happened. The outdoor variables including T_{ext}, RH_{ext}, V_{wind}, $CO2_{ext}$, F and indoor variable such as $T_{int-upper/lower}$ and RH_{int} are explanatory variables, as given in

table 7. The indoor temperature $T_{int\text{-}upper/lower}$ represents the temperature of the upper and/lower apartment compared to the investigate apartment (Active, medium or passive).

The relationship between the set-point changes probability p of the response and each explanatory variable can be expressed by a univariate logistic regression as following:

$$logit\ p = log\left(\frac{p}{1-p}\right) = \beta 0 + \beta x \qquad (1)$$

where β_0 is the intercept and β is the coefficient of the explanatory variable x. In the other hand, the relationship between the probability (p) of the binary response and multiple explanatory variables can be expressed by a multivariate logistic regression. Thus, in classical linear models, the probability can be assumed to take the form $p(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$ where β_n are the coefficients to be determined. However, this model would allow p to take values outside the interval [0 1]. It is then necessary to use a suitable transformation g of $p(x)$ [5].

The classical choice is the logistic transformation $g(x) = \log(\frac{p(x)}{1-p(x)})$. The probability distribution is called logit distribution [5]. The probability function is described by:

$$p(x) = p(x1, \ldots, xn) = \frac{\exp(\beta 0 + \beta 1 x 1 + \cdots + \beta n x n)}{1 + \exp(\beta 0 + \beta 1 x 1 + \cdots + \beta n x n)} \qquad (2)$$

where p is the probability of an event, β_0 is the intercept, $\beta_1 + \ldots + \beta_n$ are coefficients associated to each explanatory variable and $x_1 + \ldots + x_n$ are the explanatory variables.

In our study, first the relationship between the probability of thermostat changing and single explanatory variable was analyzed using univariate logistic regression. Then, the multivariate logistic regression models were determined using the 'backward selection' procedure with the studied explanatory variables presented in table 7.

IV. Results

From the data collected over the seasons, the obtained univariate logistic regression parameters are given in table 8 and tables 9 to 10 (in appendix A). The regressions are presented in Figure 7 and the multivariate logistic regression parameters are listed in table 11. The statistical software R was used for all data analyses.

1. Univariate logistic model

Univariate logistic model allows as analyzing the relationship between the probability of thermostat changes and each single explanatory variable (table 7). The analysis of thermostat changes for the active occupant considers only the apartments where there is maximum action on the thermostat. The variables of the indoor and outdoor environment are considered in this evaluation.

Appendix C gives a summary of the possible criteria of goodness-of-fit for each of the models. Therefore, table 8 summarizes the regression parameters and the goodness-of-fit (GOF) estimators. The p-values are less than 0.001 for the outdoor temperature in autumn, winter and all the year, while they are higher for outdoor relative humidity, the solar flux and wind speed in the three seasons. Additionally, the indoor relative humidity and the indoor temperature of the upper/lower apartments had p-value < 0.001 in the different seasons and all the year. Although, the p-value of the CO_2 concentration is less than 0.001 in autumn, spring and all the year only.

The above information indicates that the probability of thermostat changes has a significant correlation. Indeed, a significant correlation is noticed in autumn with outdoor temperature, CO_2 concentration, indoor relative humidity and the indoor temperature of the upper/lower apartment. Regarding winter, a correlation is noticed with outdoor temperature, indoor relative humidity and the indoor temperature of the upper/lower apartment. In spring, a correlation is noticed with outdoor CO_2 concentration, indoor relative humidity and the indoor temperature of the upper/lower apartment. A correlation is noticed for all the year with the all explanatory variables except the wind speed. The probability of thermostat changes positively related to the outdoor relative humidity and negatively correlated with the rest of the explanatory variables.

In autumn, the univariate logistic regression model with CO_2 concentration, indoor relative humidity and indoor temperature of the upper/lower apartment had the highest values of Nagelkerke's R^2, McFadden's R^2 and AUC. This indicates that the univariate logistic regression model with these variables had the best predictive capability. Hence, the lowest values of the goodness-of-fit indicate the poor predictive capabilities of the model. In winter, the univariate logistic regression model with the outdoor temperature, indoor relative humidity and the indoor temperature of the upper/lower apartments had the highest values of Nagelkerke's R^2, McFadden's R^2 and AUC. In the other hand, the CO_2 concentration and indoor relative humidity had the higher values of Nagelkerke's R^2, McFadden's R^2 and AUC for the univariate logistic regression model in spring. Finally, the goodness-of-fit estimators indicate that the outdoor temperature, the indoor relative humidity and indoor temperature of the upper/lower apartment are the principal driver for the univariate logistic regression model for all the year.

Figure 7 shows the observed probabilities and the corresponding logistic model during the all year (heating season and transition seasons), for the active occupant, for each explanatory variable. Thus, Figures 8 and 9 in appendix B show the observed probabilities and the corresponding logistic model for the medium and passive occupants.

In general, the probability of thermostat changes decreases when the outdoor temperature increases. When the outdoor relative humidity is higher than 30%, the probability of thermostat changes tends to increase when the outdoor relative humidity increases. The probability of thermostat changes first increases as the wind speed

increases. Then, when the wind speed becomes more than 9 m/s the probability decreases to 0 when the wind speed is kept increasing. The relationship between the solar flux, CO2 concentration and the probability of thermostat changes is less pronounced. As for the indoor relative humidity and indoor temperature for the upper/lower apartments, the probability related to the thermostat changes is monotonically decreasing function with respect to both variables.

2. Multivariate logistic model

The multivariate logistic regression is based on the assumption that the studied explanatory variables are independent. The explanatory variables in each multivariate logistic regression model were determined based on the forward and backward selection procedure using the Akaike information criterion [5, 6, and 29]. The purpose of this procedure is to create a model containing only explanatory variables that had a consistent effect on the probability. Rory et al. [29] presented the steps to be considered in practice.

The correlation between the explanatory variables for thermostat change models is assessed using the generalized variance inflation factor (GVIF). Table 12 shows the VIF and the GVIF analyses for thermostat change models. The $GVIF^{1/(2*Df)}$ was used to assess all the thermostat changes models for the whole year and season in order to evaluate when the multicollinearity exists or not between the explanatory variables. A GVIF of 1 indicates that the explanatory variables are not correlated. As long as, a GVIF between 1and 5 indicates a moderate correlation between the explanatory variables, and more than 5 indicates a highly correlation [19]. It can be seen that, the GVIF value of all the explanatory variables, are between 1 and 2, turned to be small (less than 5), therefore we can consider that there is no collinearity between the explanatory variables [29].

In order to compare the influence of the different explanatory variables, it is required to normalize these variables as following:

$$x_{Normal} = \frac{x - x_{min}}{x_{max} - x_{min}}$$ (3)

where x_{normal} is the normalized explanatory variable, x is the explanatory variable, x_{max} and x_{min} are respectively the maximum and the minimum observed values.

The multivariate logistic regression was carried out based on normalized explanatory variables and via backward and forward selection procedure. The parameters of the logistic model have been obtained using the

'step' in function R <mark>[30]. Table 11</mark> presents the coefficients of the explanatory variables included in the thermostat changes models.

For the active occupant, four multivariate thermostat models were obtained, according to seasons. Equations 4-7 show the most influential variables on thermostat changes behaviors in monitored apartments for the active occupant.

The multivariate logistic regression model of thermostat changes during the heating season is expressed in Equation 4. The Outdoor Temperature, Wind Speed, and Outdoor CO2 concentration are removed from the model because they do not have a consistent effect on the probability of the probability of thermostat changes. The most influential factors are the outdoor relative humidity, the solar flux, the indoor temperature of the lower apartment and the indoor relative humidity. The regression coefficients of the indoor temperature of the lower apartment and indoor relative humidity are negative, indicating the negative correlation between thermostat changes and these two variables.

- <u>Winter</u> $\quad : \ln(\dfrac{p}{1-p}) = -1.029 + 0.6195 RH_{ext} + 0.5549 F - 1.297 RH_{int}$
  $\qquad\qquad -1.287 T_{int\_lower}$

$$(4)$$

where p is the probability of thermostat changes in the living room, RH_{ext} is the outdoor relative humidity, F is the solar flux, RH_{int} is the indoor relative humidity and $T_{int\text{-}lower}$ is the indoor temperature of the lower apartment.

In transition seasons, the multivariate logistic regression model of thermostat changes behaviors is expressed by equation 5 which introduces the spring season and equation 6 which introduces the autumn season. During autumn, solar radiation and outdoor CO2 concentration variables are removed from the model because they do not have a consistent effect on the probability of a thermostat changes. The indoor temperature of the upper apartment is the most important factor. Wind speed is the least significant factor in the model, which varies from 0 to 2 m/s. The regression coefficient of the indoor relative humidity is positive, which indicates the overall increasing trend of indoor relative humidity with the increase of the thermostat change probability. On the other hand, during the spring season the variables of the outdoor temperature, the external and internal relative humidity, the solar radiation and the wind speed are eliminated from the model because they do not have a coherent effect on the probability of thermostat changes. The indoor temperature of the upper apartment is the least significant factor in the model. The regression coefficient of the outdoor CO2 concentration is negative, indicating the negative correlation between thermostat changes behaviors and this variable.

- Spring : $\ln(\dfrac{p}{1-p}) = -1.920 - 2.2295 CO2_{ext} + 1.795 T_{int\_upper}$ (5)

- Autumn : $\ln(\dfrac{p}{1-p}) = 2.247 - 3.39 T_{ext} - 3.36 RH_{ext} - 1.032 V_{wind}$
  $+ 2.17 RH_{int} - 4.29 T_{int\_upper}$ (6)

where p is the probability of thermostat changes in the living room, RH_{ext} is the outdoor relative humidity, RH_{int} is the indoor relative humidity, T_{ext} is the outdoor temperature, V_{wind} is the wind speed, $CO2_{ext}$ is the outdoor CO2 concentration and $T_{int-upper}$ is the indoor temperature of the upper apartment.

In the all year model (heating season + transition seasons) expressed in Equation 7, solar radiation and relative indoor relative humidity are eliminated because they are insignificant. Instead of that, the thermostat changes behavior is more affected by the indoor temperature of the lower apartment and the outdoor temperature, followed by the wind speed, the outdoor relative humidity, and the outdoor CO2 concentration. The regression coefficient of the outdoor temperature is negative, which is similar to that of the autumn season. The external concentration of CO2 and the external relative humidity are the least significant factors in this model.

- All the year : $\ln(\dfrac{p}{1-p}) = -0.1002 - 4.27 T_{ext} + 0.949 RH_{ext} + 0.947 CO2_{ext}$
  $+ 1.090 V_{wind} - 4.90 T_{int-lower}$ (7)

where p is the probability of thermostat changes in the living room, RH_{ext} is the outdoor relative humidity, T_{ext} is the outdoor temperature, V_{wind} is the wind speed, $CO2_{ext}$ is the outdoor CO2 concentration and $T_{int-lower}$ is the indoor temperature of the lower apartment.

The indoor relative humidity, the indoor temperature of the upper/lower apartment as well as the outdoor relative humidity and the outside temperature contribute as a dominant factor in the majority of the proposed models. Solar radiation is significant only during the heating season (winter), which can be explained by the fact that the windows are toward North-East without direct sunlight coming in.

V.   Implementation in Energy Models

The results obtained in the study provide the possibility of defining behavioral models of thermostat use to be implemented as a tool for energy simulation. Usually, the energy simulation software uses a deterministic scenario in nature, consequently it needs to interpret the probability of an action such as thermostat changes.

The implementation steps of the probabilistic models of thermostat changes to predict the effect on energy consumption are presented. Figure 10 shows a general scheme describing the implementation procedure for a dynamic simulation of thermostat changes for each time step. The first step consists of verifying the occupancy status. If the occupant is absent, the thermostat state is set as identical to its previous state. If the occupant is present, the logistic models are used to calculate the probability of thermostat changes according to the occupant profile (active occupant, medium occupant or passive occupant).

These probabilities are compared to random number in order to determine if the event takes place or not. The occupant changes the thermostat when the probability of the thermostat changes is more than the random number. If the probability of thermostat changes is large than the random number, Predicted Mean Vote (PMV) is calculated. Indeed, the PMV is used to determine an index that allows knowing the percentage of dissatisfied. Depending on the ranges of admissible PMV, three kinds of comfort zones can be considered [31]. The PMV can be calculated using conventional parameters of the thermal environment (indoor temperature, indoor humidity), activity and clothing of occupant. In this work, the thermal comfort zone is usually spread between the sensation of light freshness (-1) and the sensation of slight heat (+1). If the PMV less than (-1), the occupant turns up the thermostat and if the PMV is higher than (+1) the occupant turn down the thermostat. Finally, the set-point was changed and the next time step starts.

VI.   Discussion and limitation

In this study, it was noticed that the indoor conditions (indoor relative humidity and indoor temperature of the upper/lower apartments) impacts and describes thermostat changes action better than the outdoor conditions in the case of active occupant and univariate models. Moreover, the probability of thermostat changes given for multivariate models for active occupants tends to be always presented by the indoor environmental variables in all the cases (winter, spring, autumn and all the year). The main finding of this study underlines that the indoor environment of the upper/lower apartments is strictly interrelated to the thermostat changes.

The different probabilistic patterns of thermostat changes (active, medium and passive occupant for different seasons) are investigated. The behavior models could be considered as models of specific occupants. Nevertheless, as the models were deducted from data of 11 different apartments including many residents, they

could be considered as a representation of standard behavior patterns, enveloping the variation observed in the 11 apartments. For that, Haldi et al [32] discussed how different occupants could be assigned different model coefficients, to model the fact that different occupants will have different behavior patterns.

The present study is focused on developing thermostat changing models based on indoor and outdoor environment variables. However, other variables such as the psychological, social and economic behaviors should be considered. Gill et al. [33] show that the behavioral and social factors account for 51%, 37%, and 11% of the variance in heat, electricity and water consumption respectively. So, future research should try to collect larger dataset.

In this study, the weather data for the all monitored apartments are specific only for one city. Hence, this study ignored the spatial difference of weather data. So, for the future analysis of occupant's interaction with thermostat, it is better to take on a large interval of meteorological data in order to have an international model and improve the accuracy of the research. Additionally, the behavior due to psychological and many social factors which were not evaluated in this study need to be studied as they may have stronger relationship with the thermostat changing.

VII. Conclusion

This study investigated occupant behavior in residential building during the heating period. Occupant interaction with thermostat was evaluated. The adaptive actions executed on thermostat changes have been correlated with the different season as well as the indoor and outdoor environmental variables. Hence, behavioral models have been suggested. The aim of this study is to assess whether thermostat changes are driven by the outdoor variables and/or the indoor variables.

Based on one year of observations of thermostat changes and indoor/outdoor variables in eleven apartments, and regarding the results, three occupant behaviors; active, medium and passive were noticed.

A model has been established for occupant interaction with thermostat and indoor/outdoor environmental variables. The indoor environment variables and the outdoor temperature are the most important variables in determining the univariate logistic regression models of thermostat changing. In this context, four multivariate regression logistic models of thermostat changes were proposed in this study. The results of multivariate logistic regression based on normalized explanatory variables illustrate that the thermostat changes is mainly affected by: indoor relative humidity and temperature of the lower apartment in winter, and by the outdoor CO_2 concentration in the other seasons.

These models can be used in building performance simulation software to predict the energy consumption of residential buildings. The results of this study, recommended that when analyzing and modeling occupant behaviors in building, other variables, than just indoor/outdoor, should be taken into account. Ongoing studies have been initiated to take into account the psychological, social and economic factors on energy building simulation for a better energy use prediction.

Appendix C. Statistical tests for logistic models

1. p-value for models

The p-value is the probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value. The p-value for a model determines the significance of the model compared with the null model. So the p-value for the model indicates if there is a significant relationship described by the model.

A p-value provides one approach to summarizing the incompatibility between a particular set of data and a proposed model for the data. The most common context is a model, constructed under a set of assumptions, together with a so called "null hypothesis." Often the null hypothesis postulates the absence of an effect, such as no difference between two groups, or the absence of a relationship between a factor and an outcome. Smaller p-values do not necessarily imply the presence of larger or more important effects, and larger p-values do not imply a lack of importance or even lack of effect. Any effect, no matter how tiny, can produce a small p-value if the sample size or measurement precision is high enough, and large effects may produce unimpressive p-values if the sample size is small or measurements are imprecise. Similarly, identical estimated effects will have different p-values if the precision of the estimates differs [34, 35].

2. Area under ROC curve

In this study, the area under ROC curve is called the AUC index. It is a direct measure of the discriminating power of a given model. The AUC of the logistic model have been obtained using the software R. The area under the ROC (Receiver Operating Characteristic) curve range from zero to one and provides a measure of the model's ability to discriminate between those subjects who experience the outcome of interest versus those who do not. ROC curves plot sensitivity as a function of specificity for different values of the cutpoint between 0 and 1. A higher value of sensitivity for a given value of specificity indicates better performance [5, 36].

As a general rule:

- If $AUC = 0.5$: this suggests no discrimination.
- If $0.6 \leq AUC < 0.8$: this is considered acceptable discrimination.
- If $0.8 \leq AUC < 0.9$: this is considered acceptable discrimination.
- If $AUC \geq 0.9$: this is considered outstanding discrimination.

3. The Nagelkerke R² and the McFadden R²

The Nagelkerke R^2 or the McFadden R^2, called the coefficient of determination and generalized R^2, measures the proportion of explained deviance in a model and are defined as [37]:

$$R^2_{Nagelkerke} = \frac{1 - (\frac{-2LL_{Null}}{-2LL_M})^{\frac{2}{n}}}{1 - (-2LL_M)^{\frac{2}{n}}}$$

(8)

$$R^2_{McFadden} = 1 - \frac{LL_M}{LL_{null}}$$

(9)

where the null model is the logistic model with just the constant and the M model contains all the predictors in the model, based on n observations. The McFadden and Nagelkerke R² are the easiest to understand: when the regression is useless, the explanatory variables explain nothing, the indicator is 0; when the regression is perfect, the indicator is 1

References

[1] Andersen, Rune Korsholm, "The influence of occupants' behaviour on energy consumption investigated in 290 identical dwellings and in 35 apartments", 10th International Conference on Healthy Buildings - Brisbane, Australia 2012.

[2] Simona D'Oca, Valentina Fabi, Stefano P.Corgnati, Rune Korsholm Andersen, "Effect of thermostat and window opening occupant behavior models on energy use in homes", Build Simul, 2014.

[3] Rune Andersen, "The influence of occupants' behaviour on energy consumption investigated in 290 identical dwellings and in 35 apartments", 10th International Conference on Healthy Buildings - Brisbane, Australia 2012.

[4] P.C. da Silva, V. Leal, M. Andersen, "Occupant's interaction with electric lighting and shading systems in real single-occupied offices: Results from a monitoring campaign", Building and Environment, 2013.

[5] F. Haldi, D. Robinson, "Interactions with window openings by office occupants", Building and Environment, 2009.

[6] Rune Andersen, Valentina Fabi, Jorn Toftum, Stefano P. Corgnati, Bjarne W. Olesen, "Window opening behaviour modelled from measurements in Danish dwellings", Building and Environment, 2013.

[7] Y. Sutter, D. Dumortier et M. Fontoynont, "The use of shading systems in VDU task offices: A pilot study ", Energy and Buildings, 2006.

[8] Institut de veille sanitaire, "Description du budget espace-temps et estimation de l'exposition de la population française dans son logement", 2003-2005.

[9] Shanshan Shi, Bin Zhao, "occupants' interactions with windows in 8 residential apartments in Beijing and Nanjing, China", Building Simulation, 2016, Vol. 9, Issue 2, pp 221–231.

[10] R. Fritsch,A. Kohler,M. Nygård-Ferguson etJ.-L. Scartezzini, "A stochastic model of user behaviour regarding ventilation", Building and Environment, Vol. 25, No. 2, pp. 173 181, 1990.

[11] Rune Vinther Andersen, Bjarne W. Olesen, and Jørn Toftum, "Modelling occupants' heating set-point prefferences", Proceedings of Building Simulation 2011. 12th Conference of International Building Performance Simulation Association, Sydney, 14-16 November.

[12] Ayesha Kashif, Stéphane Ploixa, Julie Dugdaleb, Xuan Hoa Binh Le, "Simulating the dynamics of occupant behaviour for power management in residential buildings", Energy and Buildings, 56 (2013) 85-93.

[13] Cohen C, Lenzen M, Schaeffer R, "Energy requirements of households in Brazil", Energy Policy, 2005.

[14] Gesche M. Huebner, Megan McMichael, David Shipworth, Michelle Shipworth, Mathieu Durand-Daubin, Alex Summerfield. "The reality of English living rooms, A comparison of internal temperatures against common model assumptions", Energy and Buildings 2013, Vol. 66, pp. 688-696.

[15] Karjalainen S, "Gender differences in thermal comfort and use of thermostats in everyday thermal environments", Building and Environment, 2007.

[16] Luis Lopes, Shuichi Hokoi , Hisashi Miura, Kondo Shuhei. "Energy efficiency and energy savings in Japanese residential building-research methodology and surveyed results", Energy and building, 37 (2005) 698-706.

[17] Valentina Fabi, Rune Korsholm Andersen, Stefano Paolo Corgnati, " Influence of occupant's heating set-point preferences on indoor environmental quality and heating demand in n residential buildings ", HVAC&R Research, 2013.

[18] Huebner G. M., McMichael M. ea, "The reality of English living rooms, A comparison of internal temperatures against common model assumptions", Energy and Buildings 2013.

[19] FIRTH, S.K., LOMAS, K.J. and WRIGHT, A.J., "Targeting household energy efficiency measures using sensitivity analysis", Building Research & Information, 2010.

[20] SHIPWORTH, M. et al, "Central heating thermostat settings and timing: building demographics", Building Research and Information, 2010.

[21] Hoes, P, Hensen, J.L.M, Loomans, M.G.L.C, Vries, B, Bourgeois, D, "User behavior in whole building simulation", Energy and Buildings, 2009.

[22] Xu Baoping, Lin Fu, Hongfa Di, "Field investigation on consumer behavior and hydraulic performance of a district heating system in Tianj in China", Building and Environment, 2009.

[23] EN15251, " Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics", 2007-2008.

[24] G. Evola, A. Gagliano, L. Marletta, F. Nocera. Controlled mechanical ventilation systems in residential buildings: Primary energy balances and financial issues. Journal of Building Engineering, 2017.

[25] Jeyhun I. Mikayilov, Marzio Galeotti, Fakhri J. Hasanov. The Impact of Economic Growth on CO2 Emissions in Azerbaijan. Journal of Cleaner Production, 2018.

[26] Siyuan Yang, Bin Chen, Sergio Ulgiati. Co-benefits of CO2 and PM2.5 emission reduction. Energy Procedia, 2016.

[27] Kelly S, "Do homes that are more efficient consume less energy?: a structural equation model of the English residential sector", Energy, 2011, Vol. 36, Issue 9, September 2011, pp. 5610-5620.

[28] Kelly S, Shipworth M, Shipworth D, Gentry M, Wright A, Pollitt M, Crawford-Brown D, Lomas K, "Predicting the diversity of internal temperatures from the English residential sector using panel methods", Applied energy, 2013, Vol. 102, pp. 601-621.

[29] Roy V. Jones, Alba Fuertes, Elisa Gregori, Alberto Giretti, "Stochastic behavioural models of occupants' main bedroom window operation for UK residential buildings", Building and environment, 2017, Vol. 118, pp. 144-158.

[30] Vincent ISOZ, Daname KOLANI. R statistical software, one software to rule them all. V0.27 revision 202, 2017.

[31] N. Djongyang, R. Tchinda, D. Njomo, Thermal comfort: a review paper, Renewable and Sustainable Energy Reviews. 2010.

[32] Haldi F. A probabilistic model to predict building occupants' diversity towards their interactions with the building envelope. Proceeding of IBPSA International Conference, Chambéry, France. 2013.

[33] Zachary M. Gill, Michael, J. Tierney, Ian M. Pegg, Neil Allan. Low-energy dwellings: the contribution of behaviours to actual performance", Building Research and information, 2010.

[34] R.L. Wasserstein, N.A. Lazar, The ASA's statement on p-values: context, process, and purpose, Am. Stat, 2016.

[35] Perturbations on the uniform distribution of p-values can lead to misleading inferences from null-hypothesis testing. Trends in Neuroscience and Education, 2017.

[36] Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York, USA, 2000.

[37] Ricco Rakotomalala. Pratique de la Régression Logistique. Université Lumière Lyon 2, 2017.

List of Figures

Fig.1. External view of monitored building
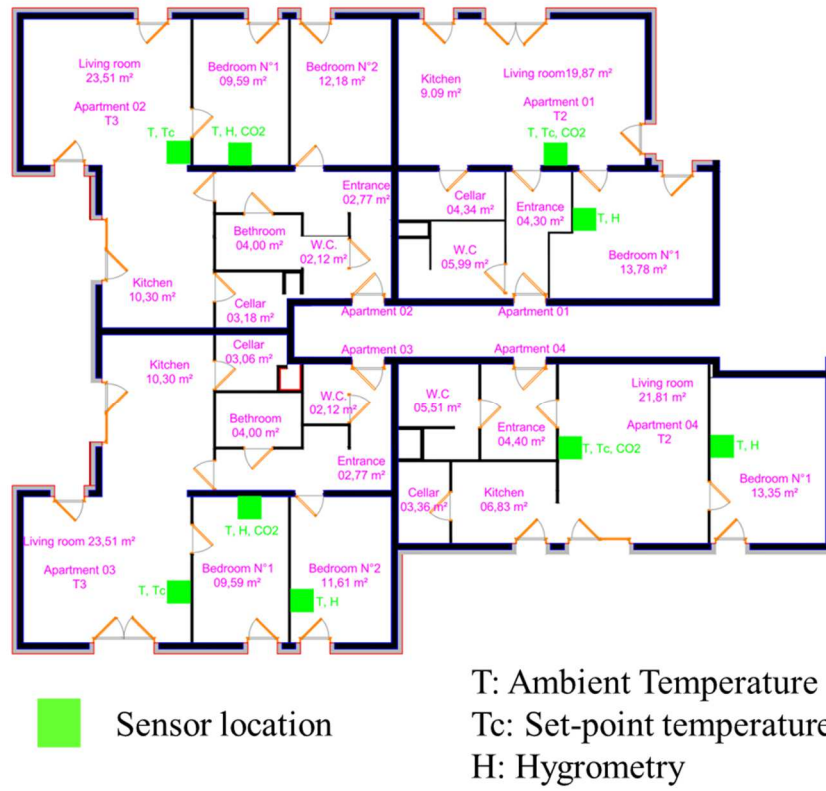
Fig.2. View plan of the first-floor apartments with sensor locations

Fig.3. Weather station

Fig.4. Heating consumption for monitored apartments in different seasons
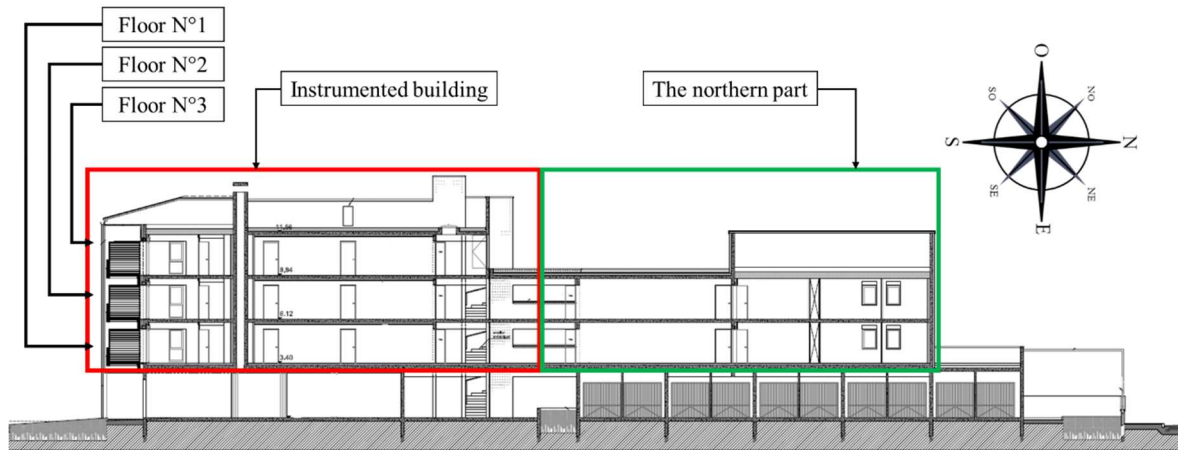
Fig.5. Season average thermostat changes

Fig.6. Distribution of indoor temperatures during the heating season and the transition seasons

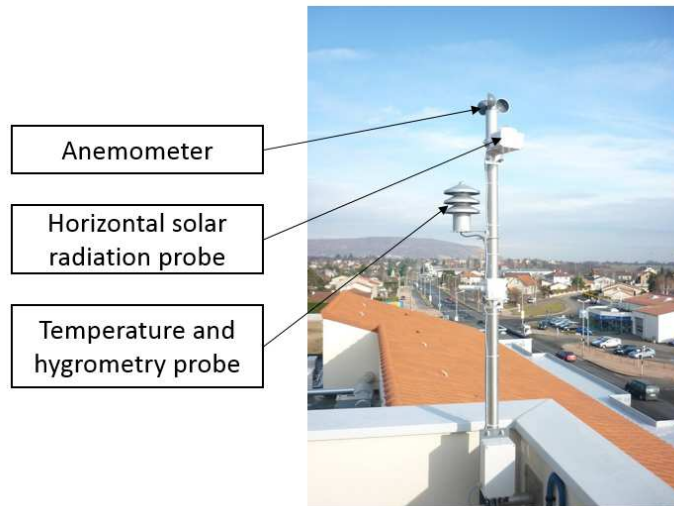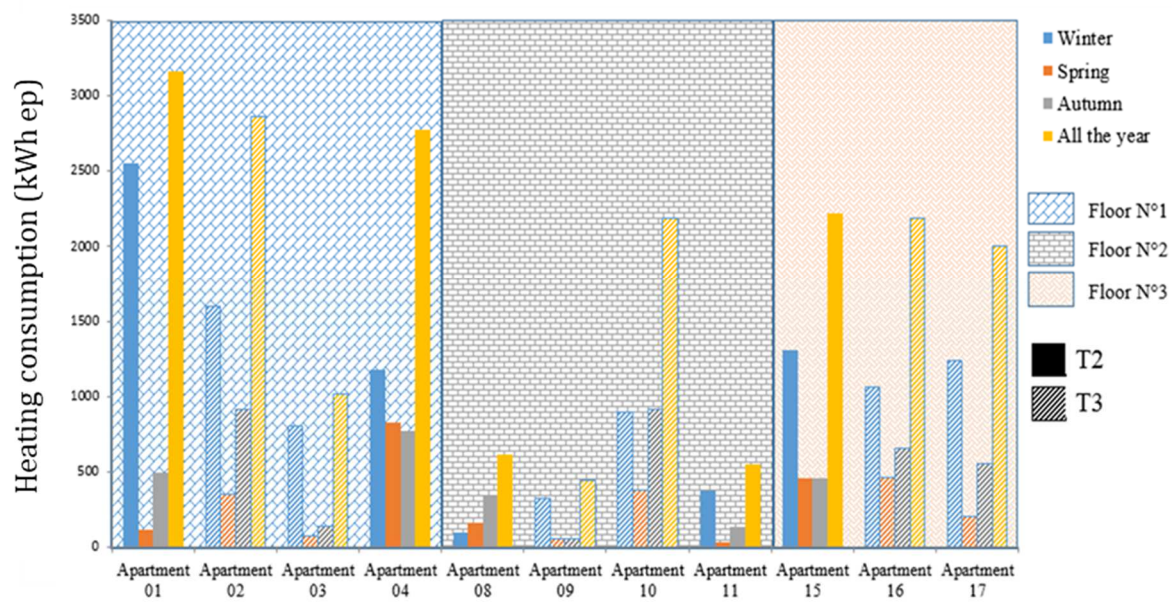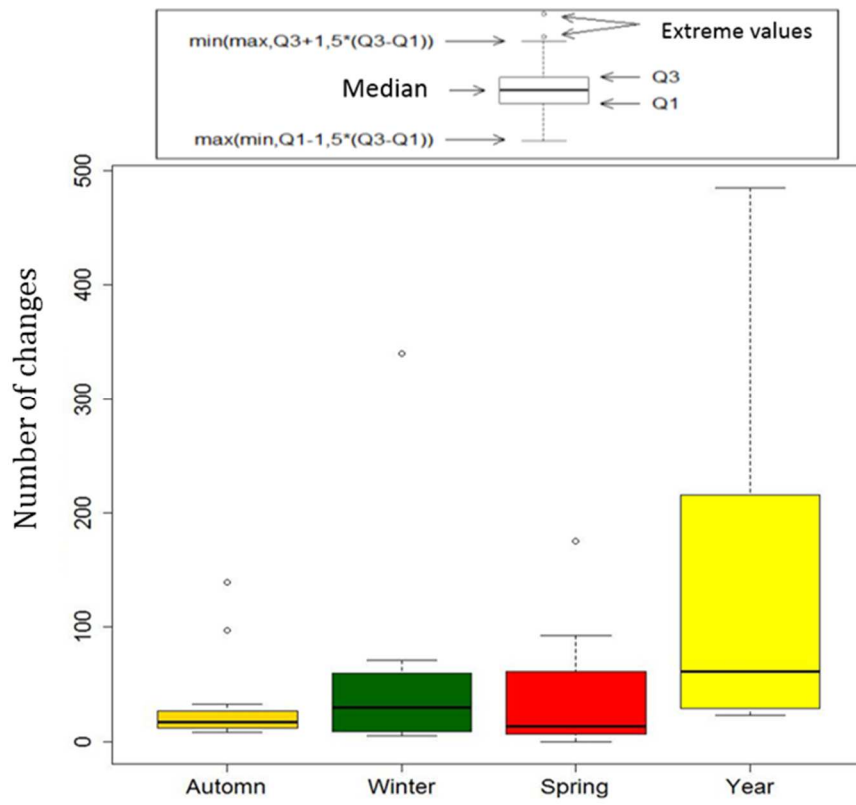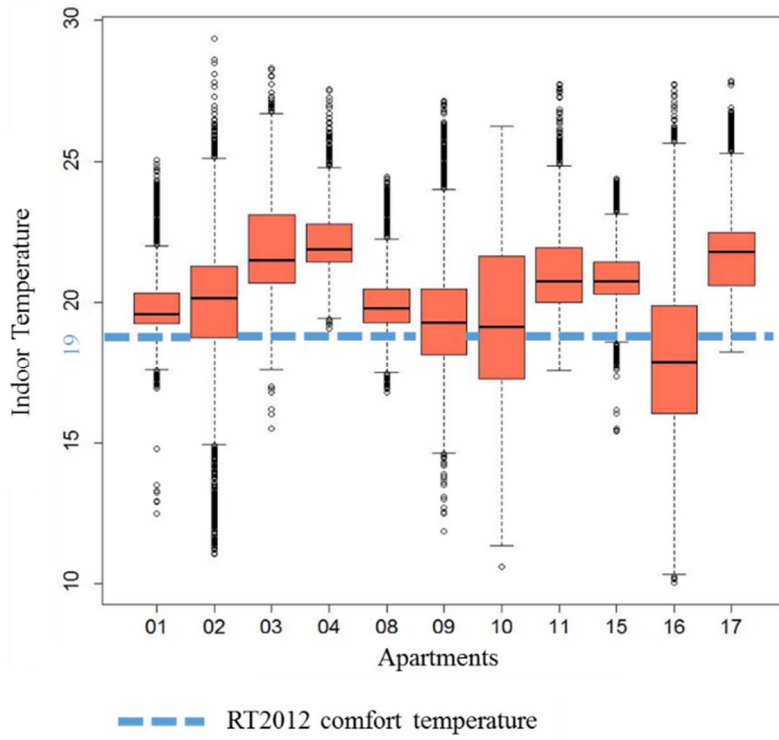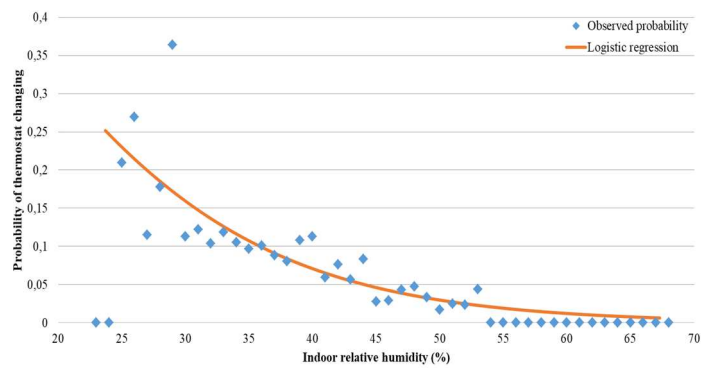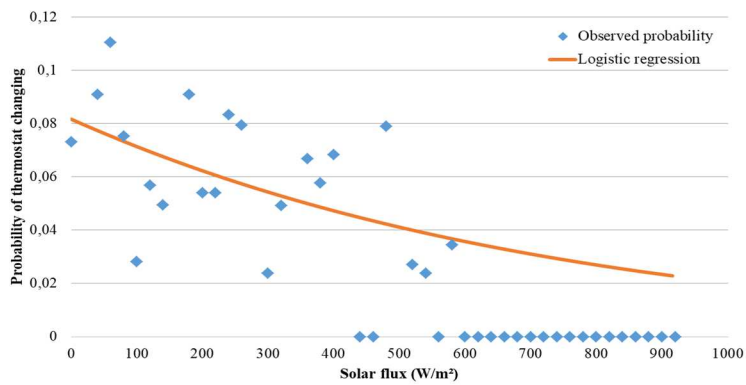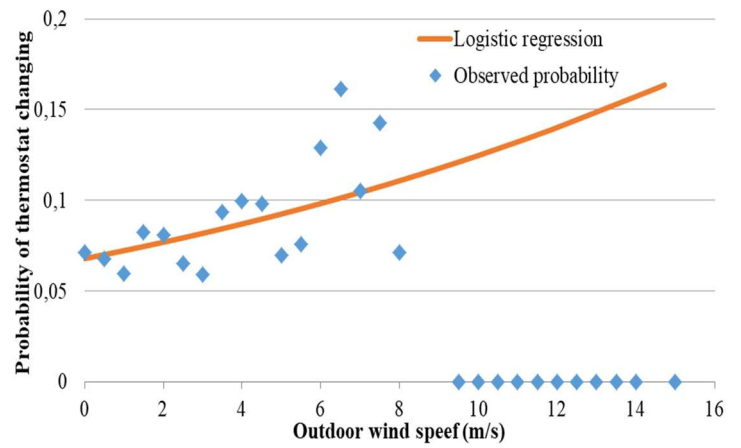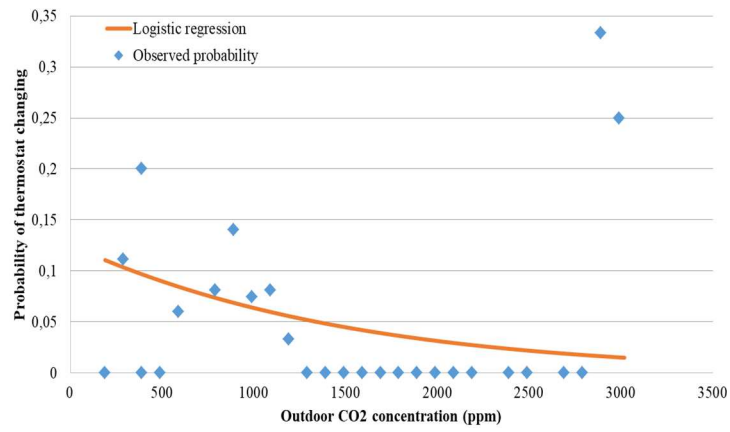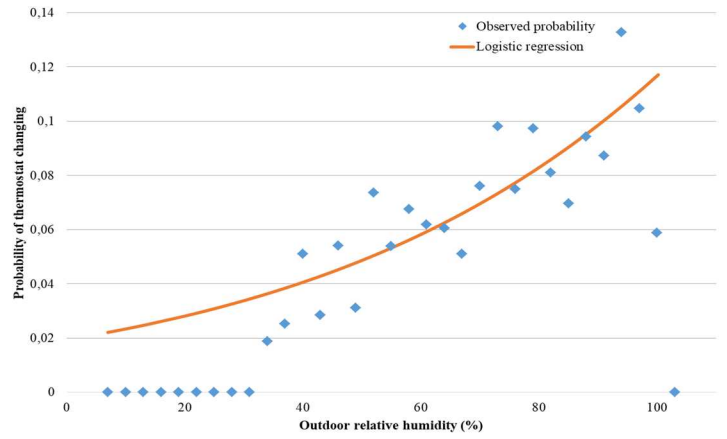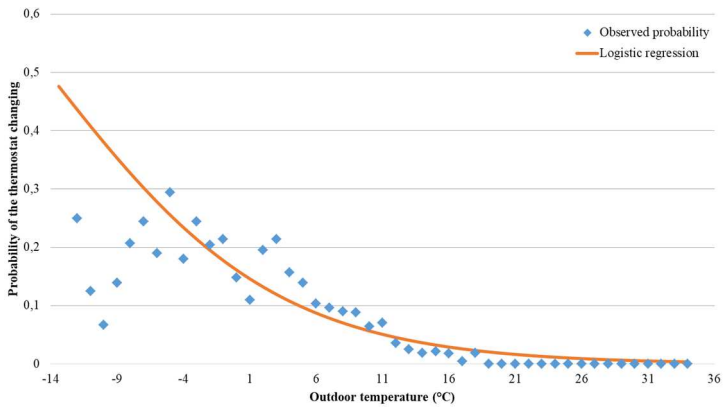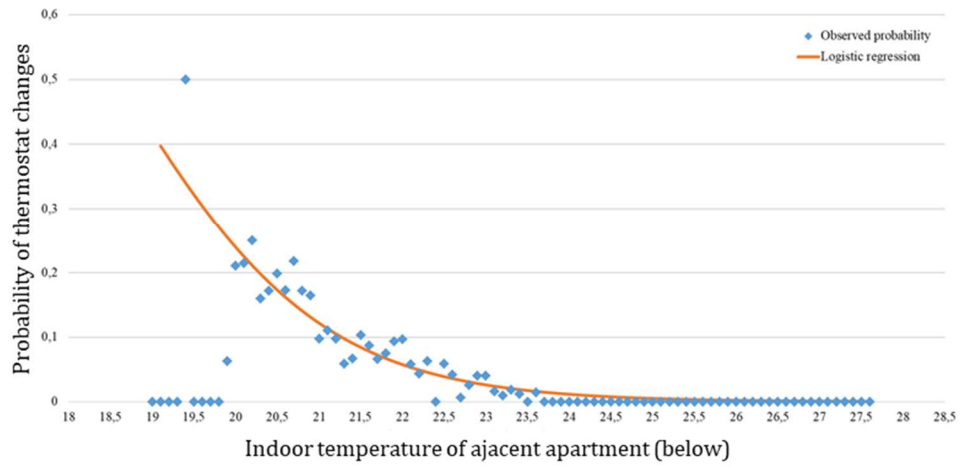**Fig.7.** Thermostat change probabilities in terms of explanatory variables for the active occupant: univariate model

Appendix B. Thermostat change probabilities in terms of explanatory variables for Medium and Passive occupants: univariate model
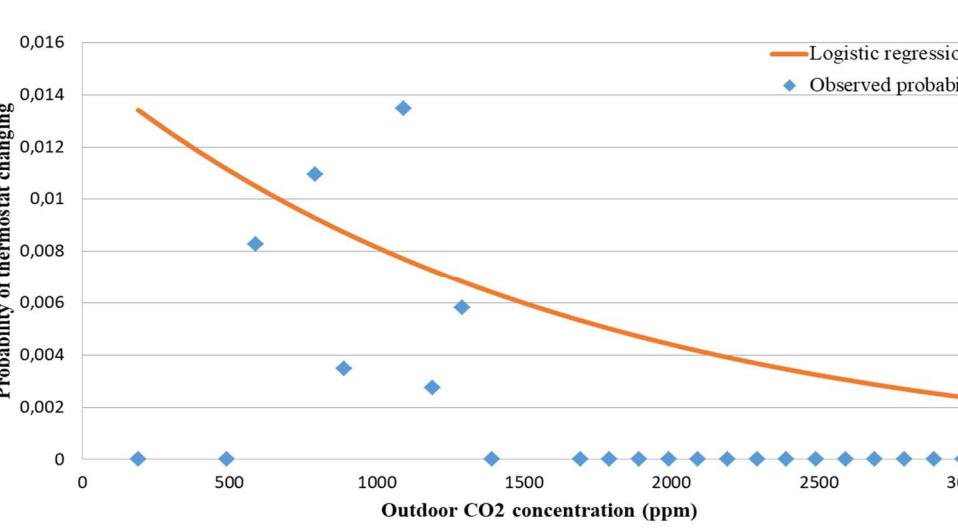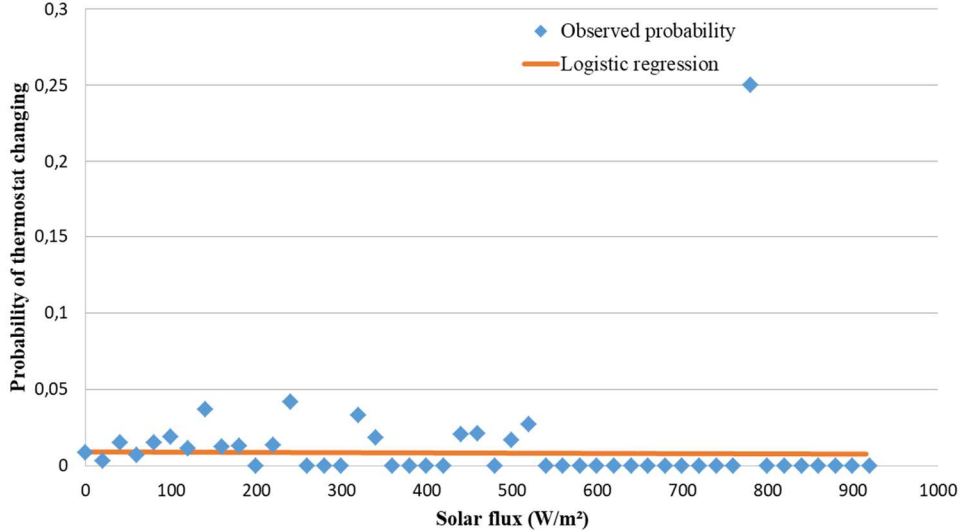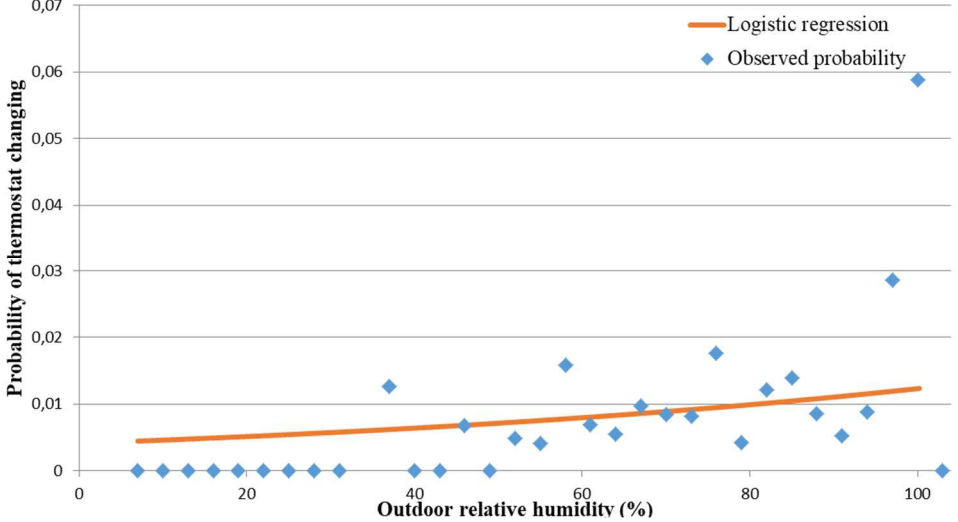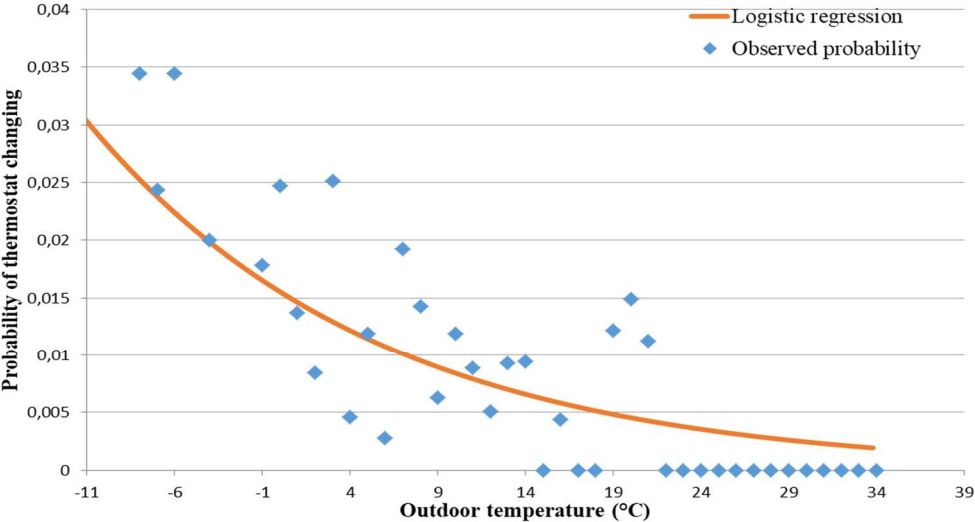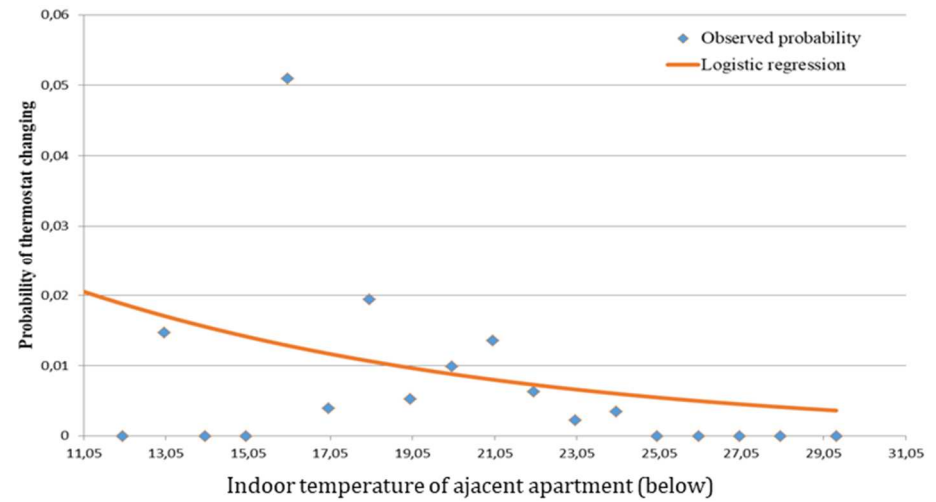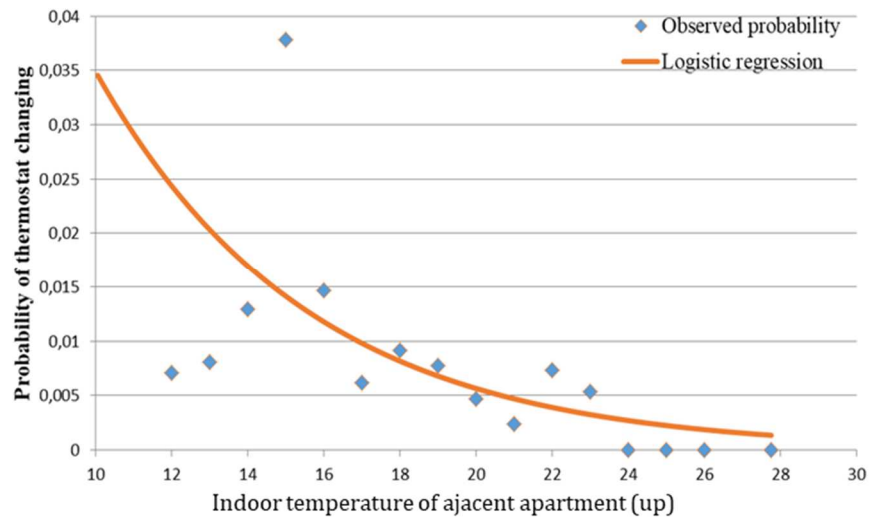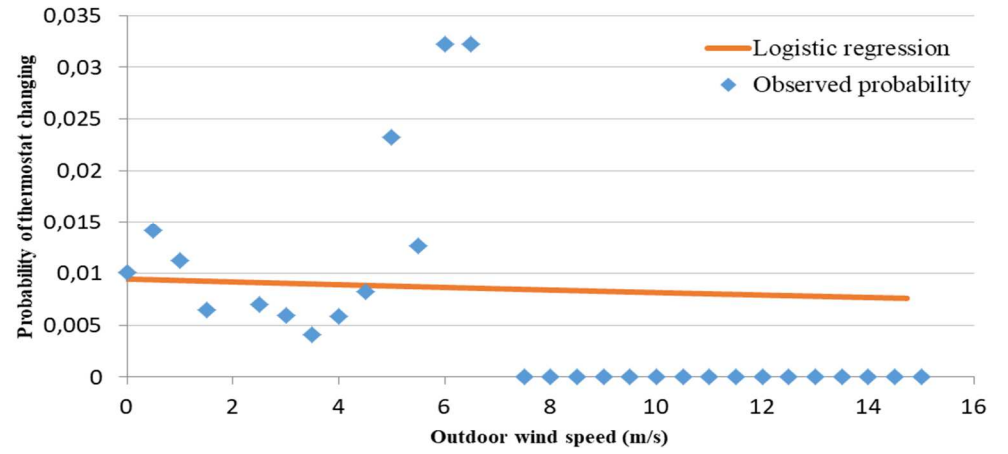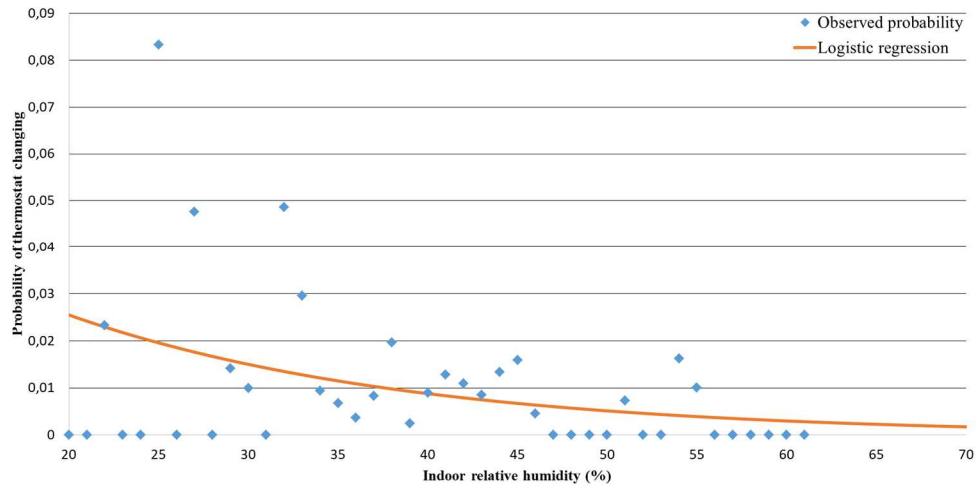
Fig.8. Thermostat change probabilities in terms of explanatory variables for Medium occupant: univariate model
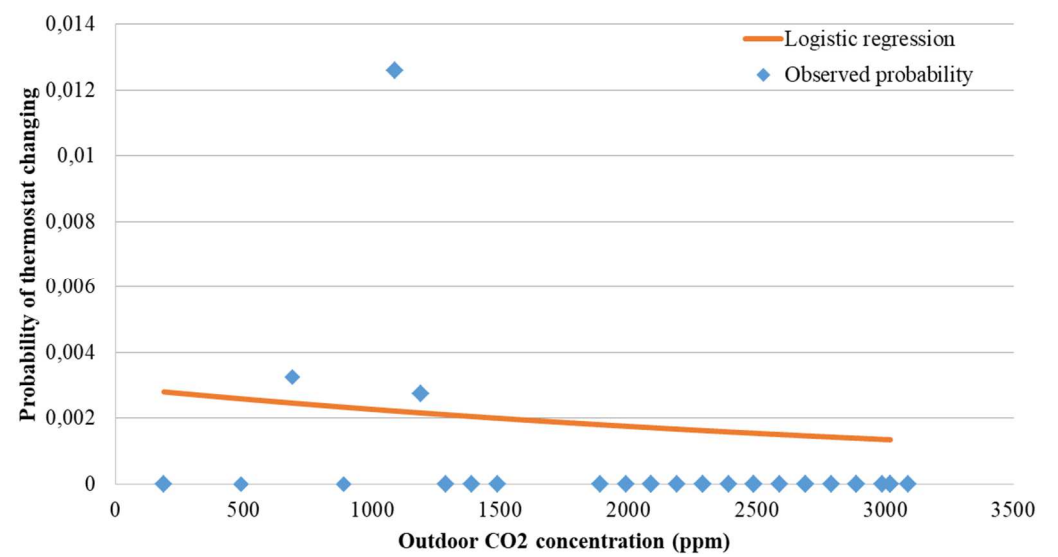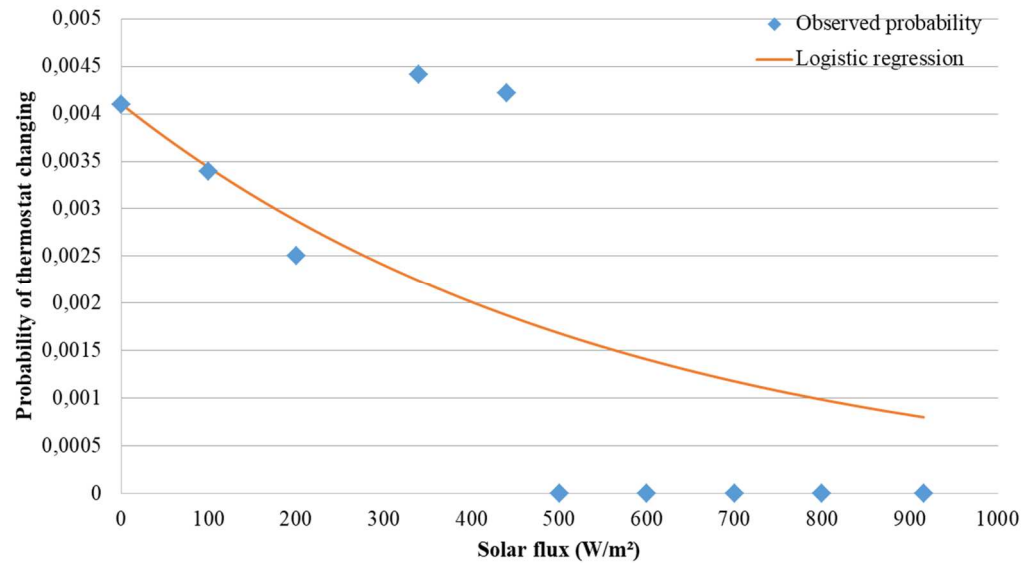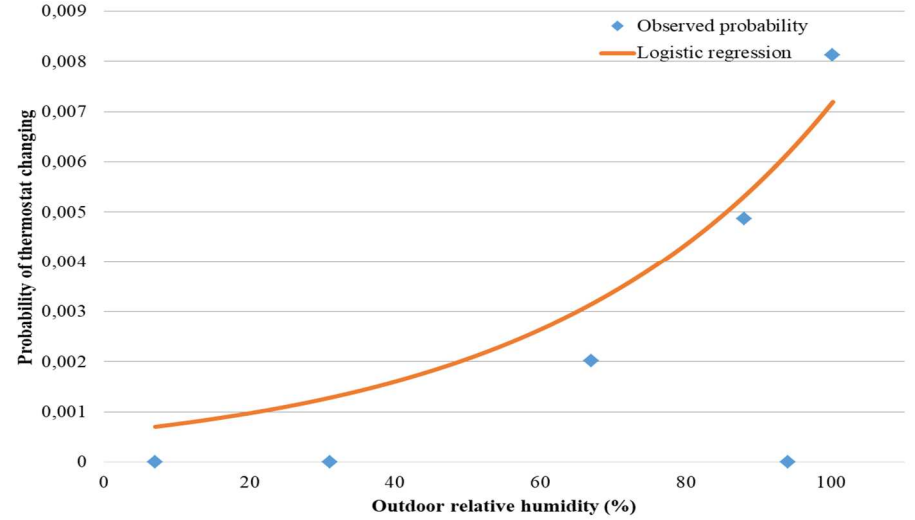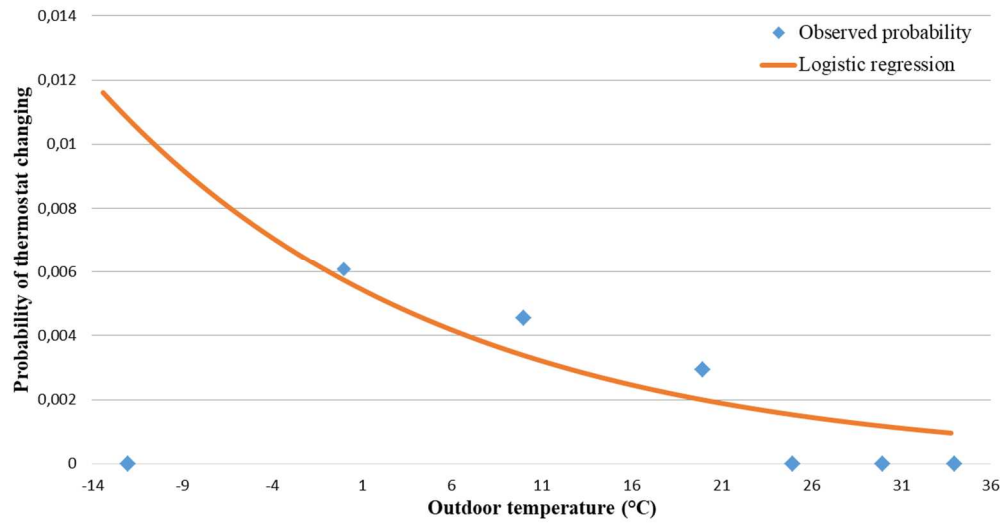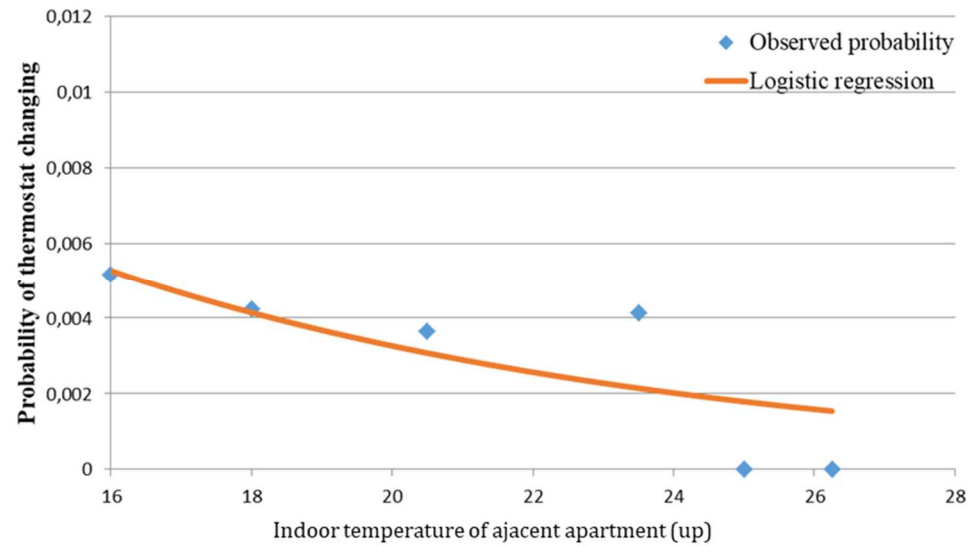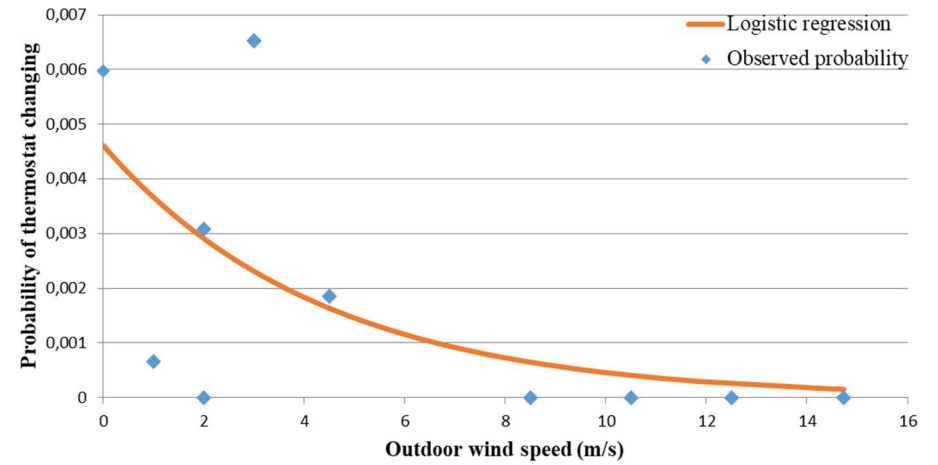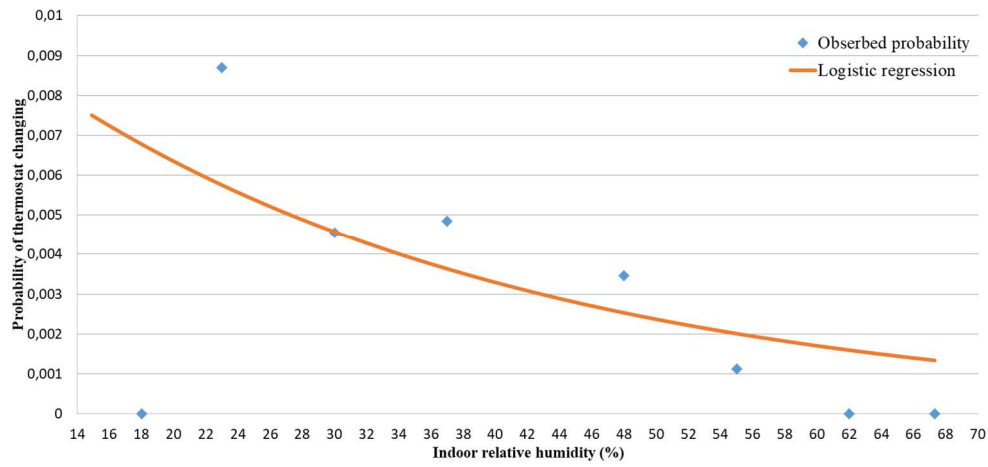
Fig.9. Thermostat change probabilities in terms of explanatory variables for Passive occupant: univariate model
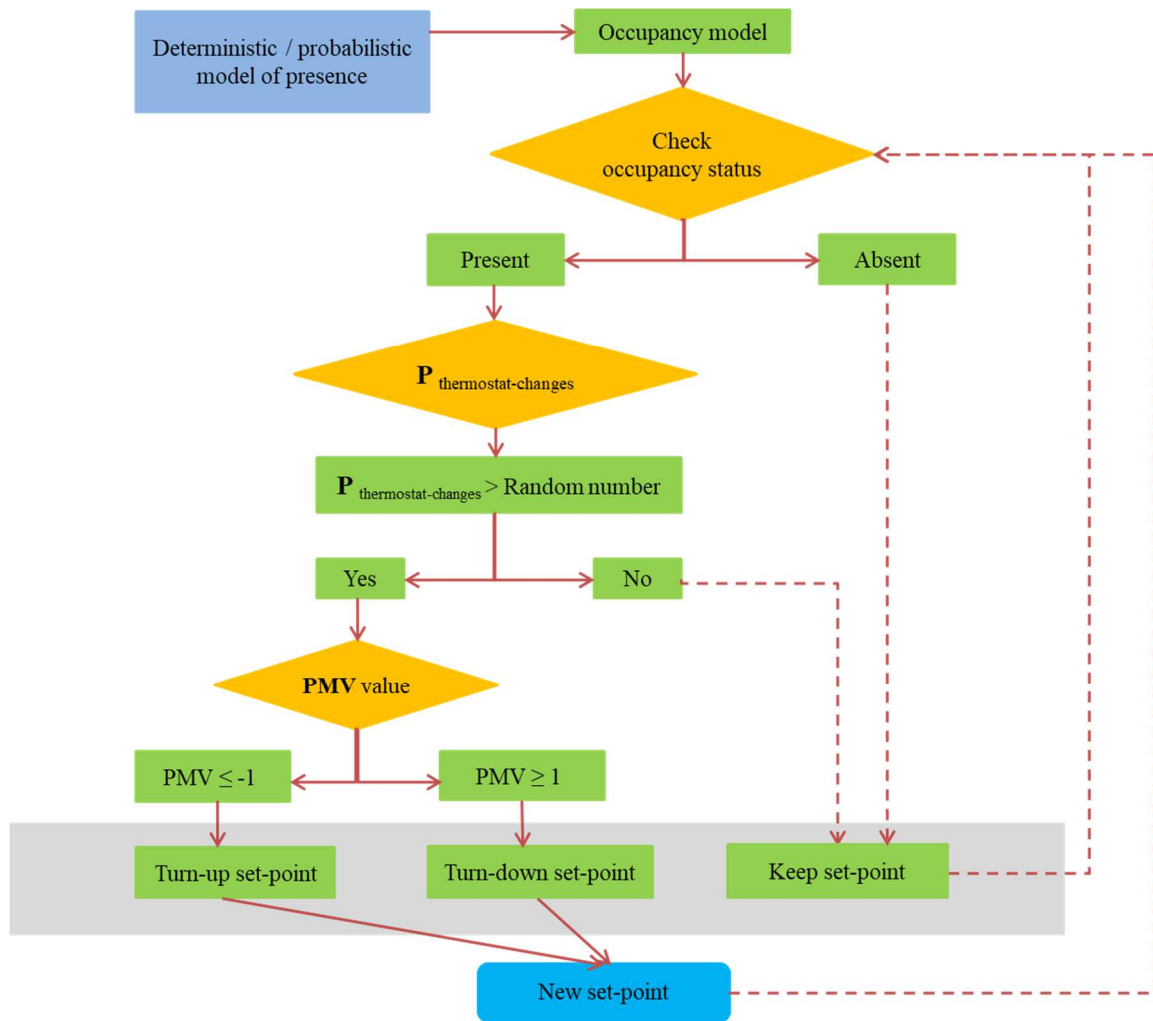
Fig.10. Implementation scheme of the thermostat changes

List of Tables

Table 1 Main features of the monitored spaces

Floor N°	Apartment	Type	Area (m²)	Number of residents	Orientation
1	01	T2	59.15	1 adult	West
1	02	T3	76.01	2 adults and 1 child	South - West
1	03	T3	75.03	2 adults	South -East
1	04	T2	57.55	1 adult	East
2	08	T2	59.15	1 adult	West
2	09	T3	76.01	2 adults and 2 children	South - West
2	10	T3	75.03	1 adults and 1 child	South -East
2	11	T2	57.55	2 adults	East
3	15	T2	59.15	2 adults	West
3	16	T3	76.01	2 adults	South - West
3	17	T3	74.88	2 adults and 1 child	South -East

Table 2 Thermo-physical properties of the opaque components

	Components	Layers (Out to in)	Thickness (cm)	R value (m2. °C/W)
Opaque walls	External walls isolated from inside and outside	Siding	1	8.33
		Polystyrene	15	
		Concrete	18	
		Polystyrene	10	
		Plasterboard	1,3	
	External walls isolated from inside	Concrete	18	4
		Polystyrene	15	
	Internal walls isolated from inside	Concrete	18	4
		Expanded polystyrene	15	
Glass walls	PVC double glazing Uf= 1.5 or 1.8	Glass	0.4	0.71
		Air	1.6	
		Glass	0.4	

Table 3 Main features of the installed sensors

(a) Sensor characteristics			
Parameter	**Sensor**	**Accuracy**	**Range**
Indoor temperature (°C)	RFF	± 0.35 °C	5-40°C
Relative Humidity %	RFF	± 3 %	25 % ➔95 % at 20°C
CO_2 (ppm)	NDIR	± 50 ppm	0➔5000 ppm
Set point temperature (°C)	Thermostat	± 0.6°C	10°C ➔30°C
(b) Weather station			
Air temperature (°C)	----------------	± 0.35°C at 25°C	- 40°C ➔ 75°C
Relative Humidity %	----------------	± 3 % [0°C to 50°C]	0.00 % ➔100 %
Air velocity (m/s)	RB-WT1000	± 0.5 m/s	0.5 m/s➔54 m/s
Global solar radiation (w/m²)	Spektron	± 10 W/m² or ± 5%	0 W/m²➔1280 W/m²
CO_2 (ppm)	----------------	± 100 ppm	0➔2000 ppm

Table 4 Detailed information about monitored apartments

Apartment N°	Location	Sensors	Quantity
01 to 04	Bedroom	Temperature, Hygrometry	1 per apartment
08 to 12	Living room	Set-point temperature, temperature probe	1 per apartment
15 to 17		CO_2	1 per apartment

Table 5 Total thermostat changing for monitored apartments considering living room variation

Apartment N°	Number of residents	Heating season	Transition seasons		Heating and transition seasons
		Winter*	Spring**	Autumn***	All the year****
01	1	5	8	17	30
02	3	61	74	139	274
03	2	7	5	12	24
04	1	59	175	12	246
08	1	5	11	12	28
09	4	30	13	18	61
10	2	10	5	8	23
11	2	340	48	97	485
15	2	71	93	21	185
16	2	12	0	33	45
17	3	37	27	16	80

*: From 1st December to 30th February **:From 1st March to 30th May

: From 1st September to 30th November   *: From 1st September to 30th May

<div style="text-align:center">**Table 6** Summary of measured settings</div>

	Variable	Max			Min			Mean			St. Dev		
		W*	S*	Au*	W*	S*	Au*	W*	S*	Au*	W*	S*	Au*
Outdoor	Temperature (°C)	17.2	32.8	28	-13.4	0.1	0	4.12	11.89	13.67	5.84	5.42	5.73
	Solar flux (w/m²)	589.4	915.6	672	0	0	0	46.39	121.8	25.10	105.12	204.17	95.50
	Outdoor RH (%)	98.9	98.4	98	28.4	7	24	73.77	63.72	71.35	12.43	16.88	16.26
	CO_2 concentration (ppm)	3020	770	2180	190	330	840	709.12	579.56	1169.93	241.45	64.31	202.77
	Wind speed (m/s)	17.73	11.94	2	0	0	0	1.42	1.52	0.014	1.67	1.62	0.17
Apartment 01	Indoor temperature (°C)	21.6	23.4	24	18.75	12.5	20	19.6	19.17	21.32	0.34	1.26	1.3
	Indoor RH (%)	82.6	81.7	71	32.4	32.6	44	48.62	55.79	57.79	5.93	7.06	5.36
Apartment 02	Indoor temperature (°C)	22.15	25.35	25	11.05	12.95	18	17.07	20.6	21.20	2.65	1.71	1.57
	Indoor RH (%)	79.1	88.3	72	16.4	25.7	31	35.87	42.06	49.31	8.41	8.55	7.67
Apartment 03	Indoor temperature (°C)	23.85	26.8	28	15.5	16	20	20.82	21.88	23.25	0.75	1.54	1.61
	Indoor RH (%)	54.9	67.3	60	14.9	16.1	30	30.85	38.19	43.78	7.04	7.45	7.38
Apartment 04	Indoor temperature (°C)	23.65	25.4	27	19.1	19.55	20	21.4	22.22	22.7	0.88	0.89	1.21
	Indoor RH (%)	51.5	66.1	62	19.8	26.2	38	36.03	40.74	47.21	5.3	6.72	5.48
Apartment 08	Indoor temperature (°C)	21	22.6	24	18.3	16.8	19	19.65	19.6	20.95	0.48	1.22	1.37
	Indoor RH (%)	53.9	64.2	71	18.4	27.9	36	39.22	43.61	50.65	5.96	5.63	7.24
Apartment 09	Indoor temperature (°C)	22.3	25.85	27	12.55	11.85	19	17.74	19.43	21.91	1.27	1.43	1.9
	Indoor RH (%)	56	77.6	69	17.4	24.4	31	35.07	41.70	47	7.08	9.07	7.33
Apartment 10	Indoor temperature (°C)	24.75	26.25	26	11.35	10.6	18	17.19	19.14	22.33	1.71	2.27	1.8
	Indoor RH (%)	56	83.4	63	16.8	24.3	29	35.11	41.83	45.29	7.59	9.82	6.88
Apartment 11	Indoor temperature (°C)	22.35	25.65	26	17.55	18.85	20	20	21.33	22.18	0.65	1.36	1.55
	Indoor RH (%)	58.4	64.9	67	23.7	28.3	33	37.15	40.40	47.17	6	6.68	6.65
Apartment 15	Indoor temperature (°C)	22.2	23.9	24	15.45	15.4	18	21.14	20.27	21.77	0.53	0.98	1.62

	Indoor RH (%)	48	75.7	65	25.6	32.2	39	37.03	43.52	50.37	4.66	8.60	6.30
Apartment 16	Indoor temperature (°C)	23.85	25.75	27	10.05	10.7	17	15.68	18.03	20.21	2.21	2.25	2.28
	Indoor RH (%)	80.4	80.2	68	19.3	27.4	31	40.72	45.33	48.49	9.33	9.74	7.88
Apartment 17	Indoor temperature (°C)	24.55	27.85	26	18.6	18.25	19	21.38	21.04	22.66	1.03	1.6	1.4
	Indoor RH (%)	54.2	71	61	12.8	23.9	30	29.32	38.49	44.16	6.98	8.69	7.83

W*: winter, **S***: spring, **Au***: autumn

Table 7 List of explanatory variables used to infer the models of thermostat changing

Variable	Unit
Outdoor temperature T_{ext}	°C
Outdoor relative humidity RH_{ext}	%
Solar flux F	w/m²
Wind speed V_{wind}	m/s
Outdoor CO2 concentration $CO2_{ext}$	ppm
Indoor temperature $T_{int\text{-}upper\ (or\ lower)}$	°C
Indoor relative humidity RH_{int}	%

Table 8 Univariate logistic parameters for changing thermostat states for active occupants

| Variable | Season | Thermostat Changing | | p-value of | Goodness of fit | | |
| | | Regression coefficient | | | | | |
		$\beta 0$	$\beta 1$	$\beta 1$	AUC	Nagelkerke's R²	McFadden's R²
Outdoor temperature T_{ext}	Autumn	-1.968±0.218	-0.054±0.016	0.001	0.60	0.015	0.012
	Winter	-1.544±0.067	-0.040±0.010	< 0,0001	0.60	0.013	0.009
	Spring	-2.708±0.192	0.021±0.014	0.137	0.55	0.24	0.18
	All the year	-1.650±0.061	-0.116±0.007	< 0,0001	0.75	0.11	0.09
Outdoor relative humidity RH_{ext}	Autumn	-2.575±0.399	-1.06×10^{-3}±0.005	0.771	0.50	1.02×10^{-4}	8.15×10^{-5}
	Winter	-1.738±0.356	6.42×10^{-4}±0.005	0.892	0.51	1.44×10^{-5}	9.7×10^{-6}
	Spring	-2.589±0.312	2.13×10^{-3}±0.005	0.649	0.51	0.50	0.43
	All the year	-3.924±0.245	0.019±0.003	< 0,0001	0.578	0.014	0.011
Solar Flux F	Autumn	-2.711±0.101	2.36×10^{-4}±0.001	0.646	0.58	2.51×10^{-4}	2.00×10^{-4}
	Winter	-1.738±0.065	9.49×10^{-4}±0.001	0.066	0.55	0.0025	0.0017
	Spring	-2.458±0.092	4.24×10^{-5}±0	0.912	0.57	0.32	0.27
	All the year	-2.432±0.052	-1.37×10^{-3}±0	< 0,0001	0.514	0.007	0.005
Outdoor CO2 concentration $CO2_{ext}$	Autumn	0.440±0.583	-2.77×10^{-3}±0.001	< 0,0001	0.64	0.038	0.030
	Winter	-1.659±0.186	-4.42×10^{-5}±0.001	0.859	0.54	2.51×10^{-5}	1.68×10^{-5}
	Spring	-0.365±0.721	-3.64×10^{-3}±0.001	0.004	0.56	0.009	0.007
	All the year	-1.823±0.142	-9.01×10^{-4}±0.001	< 0,0001	0.54	0.011	0.008
Wind speed V_{wind}	Autumn	-2.656±0.106	-0.042±0.079	0.598	0.52	0.00034	0.00027
	Winter	-1.721±0.078	0.021±0.035	0.540	0.50	0.0003	0.00019
	Spring	-2.429±0.107	-0.016±0.049	0.750	0.49	0.18	0.14
	All the year	-2.618±0.061	0.067±0.028	0.018	0.523	0.002	0.001
	Autumn	-0.591±0.584	-0.044±0.012	0	0.60	0.016	0.013

Indoor relative humidity RH_{int}	Winter	-0.215±0.355	-0.04±0.010	< 0,0001	0.60	0.014	0.009
	Spring	-3.416±0.460	0.023±0.011	0.003	0.54	0.30	0.25
	All the year	1.067±0.276	-0.091±0.007	< 0,0001	0.667	0.064	0.051
Indoor temperature $T_{int\text{-}lower}$		—	—	—	—	—	—
Indoor temperature $T_{int\text{-}upper}$	Autumn	3.822±1.155	-0.310±0.056	< 0,0001	0.63	0.044	0.035
Indoor temperature $T_{int\text{-}lower}$		5.029±1.489	-0.315±0.070	< 0,0001	0.6	0.017	0.011
Indoor temperature $T_{int\text{-}upper}$	Winter	—	—	—	—	—	—
Indoor temperature $T_{int\text{-}lower}$		—	—	—	—	—	—
Indoor temperature $T_{int\text{-}upper}$	Spring	-3.862±0.608	-0.032±0.014	0.018	0.57	0.19	0.15
Indoor temperature $T_{int\text{-}lower}$		15.70±1.173	-0.840±0.055	< 0,0001	0.714	0.10	0.08
Indoor temperature $T_{int\text{-}upper}$	All the year	—	—	—	—	—	—

Appendix A. Coefficients and GOF of logistic regression models for Medium and passive occupant (Table 9 + Table 10)

Table 9 Regression results and statistical tests of the univariate logistic regression model for changing thermostat states: Medium occupants

| Variable | Season | Thermostat Changing | | p-value of | ± | Goodness of fit | |
| | | Regression coefficient | | | | | |
		$\beta 0$	$\beta 1$	$\beta 1$	AUC	Nagelkerke's R^2	McFadden's R^2
Outdoor temperature T_{ext}	Autumn	-4.897±0.492	0.049±0.029	0.098	0.62	0.0084	0.0078
	Winter	-3.725±0.180	0.031±0.024	0.204	0.55	0.0035	0.0031
	Spring	-1.237±0.336	-0.288±0.044	< 0,0001	0.80	0.4	0.37
	All the year	-4.148±0.175	-0.062±0.017	0	0.63	0.019	0.018
Outdoor relative humidity RH_{ext}	Autumn	-1.767±0.647	-3.62×10^{-2}±0.010	0	0.67	0.039	0.036
	Winter	-4.226±0.828	8.62×10^{-3}±0.011	0.428	0.53	0.0013	0.0012
	Spring	-5.270±0.668	0.022±0.009	0.019	0.60	0.57	0.55
	All the year	-5.483±0.636	0.011±0.009	0.184	0.55	0.0028	0.0026
Solar Flux F	Autumn	-4.594±0.235	2.95×10^{-3}±0.001	0	0.71	0.040	0.037
	Winter	-3.599±0.145	3.19×10^{-4}±0.001	0.791	0.51	1.42×10^{-4}	1.26×10^{-4}
	Spring	-3.651±0.161	-0.002±0.001	0.083	0.46	0.22	0.20
	All the year	-4.688±0.146	-1.91×10^{-4}±0.001	0.798	0.52	9.8×10^{-5}	9.3×10^{-5}
Outdoor CO2 concentration $CO2_{ext}$	Autumn	-1.901±1.125	-2.00×10^{-3}±0.001	0.047	0.59	0.014	0.013
	Winter	-4.272±0.270	9.20×10^{-4}±0	0.002	0.66	0.014	0.013
	Spring	-0.993±1.336	-0.005±0.002	0.037	0.58	0.011	0.0097
	All the year	-4.180±0.379	-6.21×10^{-4}±0	0.182	0.50	0.003	0.003
Wind speed V_{wind}	Autumn	-4.488±0.226	0.296±0.105	0.005	0.67	0.020	0.019
	Winter	-3.908±0.179	0.189±0.058	0.001	0.60	0.018	0.016
	Spring	-4.151±0.211	0.192±0.072	0.008	0.60	0.33	0.31
	All the year	-4.652±0.165	-0.015±0.086	0.858	0.53	4.94×10^{-5}	9.7×10^{-5}
	Autumn	-2.722±1.12	-0.030±0.023	0.196	0.56	0.0054	0.0050

Indoor relative humidity RH_{int}	Winter	-6.119±1.068	0.069±0.028	0.015	0.60	0.013	0.012
	Spring	1.734±1.018	-0.144±0.028	< 0,0001	0.80	0.39	0.37
	All the year	-2.564±0.578	-0.054±0.015	0	0.63	0.02	0.019
Indoor temperature $T_{int-lower}$	Autumn	-4.863±1.175	0.014±0.024	0.552	0.54	0.034	0.032
Indoor temperature $T_{int-upper}$		—	—	—	—	—	—
Indoor temperature $T_{int-lower}$	Winter	—	—	—	—	—	—
Indoor temperature $T_{int-upper}$		-5.709±0.924	0.056±0.023	0.017	0.60	0.0034	0.0030
Indoor temperature $T_{int-lower}$	Spring	-3.205±1.463	-0.184±0.040	< 0,0001	0.75	0.20	0.18
Indoor temperature $T_{int-upper}$		—	—	—	—	—	—
Indoor temperature $T_{int-lower}$	All the year	-2.803±0.856	-0.096±0.044	0.03	0.60	0.0066	0.0063
Indoor temperature $T_{int-upper}$		-1.471±0.765	-0.185±0.046	< 0,0001	0.66	0.026	0.025

Table 10 Regression results and statistical tests of the univariate logistic regression model for changing thermostat states: Passive occupants

Variable	Season	Thermostat Changing		p-value of β1	Goodness of fit		
		Regression coefficient			AUC	Nagelkerke's R²	McFadden's R²
		β0	β1				
Outdoor temperature T_{ext}	Autumn	-3.453±0.749	-0.192±0.075	0.011	0.77	0.073	0.072
	Winter	-6.905±0.918	0.137±0.106	0.195	0.658	0.030	0.029
	Spring	-5.521±1.051	-0.051±0.09	0.574	0.60	0.51	0.50
	All the year	-5.154±0.288	-0.053±0.028	0.058	0.62	0.011	0.010
Outdoor relative humidity RH_{ext}	Autumn	-6.513±1.802	0.012±0.024	0.601	0.52	0.0028	0.0027
	Winter	-3.956±2.334	-0.030±0.033	0.372	0.65	0.011	0.010
	Spring	-5.951±1.724	-0.002±0.026	0.935	0.49	0.64	0.63
	All the year	-7.432±1.129	0.025±0.015	0.089	0.61	0.010	0.010
Solar Flux F	Autumn	-5.462±0.383	-0.002±0.003	0.477	0.67	0.0063	0.0062
	Winter	-6.680±0.619	0.006±0.002	0.010	0.71	0.071	0.070
	Spring	-5.532±0.448	-8.423	0.9	0.66	0.079	0.078
	All the year	-5.491±0.221	-1.79×10^{-3}±0.002	0.291	0.548	0.0045	0.0044
Outdoor CO2 concentration $CO2_{ext}$	Autumn	4.504±3.545	-0.010±0.004	0.008	0.83	0.118	0.116
	Winter	-6.638±0.926	0.001±0.001	0.461	0.75	0.0054	0.0053
	Spring	-5.299	0.001±0.007	0.846	0.55	5.44×10^{-4}	5.35×10^{-4}
	All the year	-5.825±0.556	2.59×10^{-4}±0.001	0.671	0.54	5.56×10^{-4}	5.43×10^{-4}
Wind speed V_{wind}	Autumn	-5.749±0.441	0.156±0.249	0.530	0.565	0.0033	0.0032
	Winter	-5.469±0.528	-0.738±0.249	0.213	0.63	0.04	0.036
	Spring	-6.037±0.607	-0.035±0.287	0.904	0.50	0.40	0.39
	All the year	-5.374±0.253	-0.231±0.177	0.193	0.58	0.0066	0.0064
Indoor relative humidity RH_{int}	Autumn	-0.530±2.209	-0.119±0.054	0.029	0.74	0.050	0.049
	Winter	-8.701±3.548	0.065±0.085	0.445	0.59	0.0091	0.0089
	Spring	-3.131±2.899	-0.082±0.082	0.322	0.62	0.52	0.51
	All the year	-8.701±3.548	0.065±0.085			0.0070	0.0069
Indoor temperature $T_{int-lower}$		1.002±2.393	-0.161±0.063	0.010	0.80	0.065	0.064

Indoor temperature $T_{int\text{-}upper}$	Autumn	1.391±2.602	-0.175±0.070	0.012	0.79	0.10	0.088
Indoor temperature $T_{int\text{-}lower}$		-24.39	0.933±0.993	0.348	0.60	0.0104	0.0103
Indoor temperature $T_{int\text{-}upper}$	Winter	8.388	-0.688±0.502	0.171	0.66	0.017	0.016
Indoor temperature $T_{int\text{-}lower}$		—	—	—	—	—	—
Indoor temperature $T_{int\text{-}upper}$	Spring	-4.448	-0.087±0.211	0.681	0.58	0.49	0.49
Indoor temperature $T_{int\text{-}lower}$		—	—	—	—	—	—
Indoor temperature $T_{int\text{-}upper}$	All the year	-3.321±1.437	-0.120±0.076	0.116	0.60	0.0084	0.0082

Table 11 Regression results and statistical tests of the multi-variate logistic regression model for changing thermostat states: Active occupants

Variable	Season	Regression coefficient	p-value of β	Goodness of fit	
				Nagelkerke's R^2	McFadden's R^2
Intercept β_0		2.247±0.850	—		
Outdoor temperature T_{ext}		-3.39±1.426	0.0174		
Outdoor relative humidity RH_{ext}		-3.36±0.949	0.0003		
Solar Flux F		—	—		
Outdoor CO2 concentration $CO2_{ext}$	Autumn	—	—	0.066	0.053
Wind speed V_{wind}		-1.032±0.704	0.1426		
Indoor relative humidity RH_{int}		2.17±1.029	0.0350		
Indoor temperature $T_{int-lower}$		—	—		
Indoor temperature $T_{int-upper}$		-4.29±0.939	< 0,0001		
Intercept β_0		-1.029±0.299	0.00060		
Outdoor temperature T_{ext}		—	—		
Outdoor relative humidity RH_{ext}		0.6195±0.3721	0.0959		
Solar Flux F		0.5549±0.3348	0.0974		
Outdoor CO2 concentration $CO2_{ext}$		—	—	0.0324	0.0219
Wind speed V_{wind}	Winter	—	—		
Indoor relative humidity RH_{int}		-1.297±0.3658	0.00039		
Indoor temperature $T_{int-lower}$		-1.287±0.3325	0.00010		
Indoor temperature $T_{int-upper}$		—	—		
Intercept β_0		-1.920±0.3090	—		
Outdoor temperature T_{ext}		—	—		
Outdoor relative humidity RH_{ext}		—	—		
Solar Flux F		—	—		

Parameter	Season	Coefficient	p-value		
Outdoor CO2 concentration $CO2_{ext}$		-2.2295±0.5612	< 0,0001		
Wind speed V_{wind}	**Spring**	—	—	0.035	0.027
Indoor relative humidity RH_{int}		—	—		
Indoor temperature $T_{int\text{-}lower}$		—	—		
Indoor temperature $T_{int\text{-}upper}$		1.795±0.3954	< 0,0001		
Intercept β_0		-0.1002 ±0.329	—		
Outdoor temperature T_{ext}		-4.27 ±0.387	< 0,0001		
Outdoor relative humidity RH_{ext}		0.945±0.356	0.0079		
Solar Flux F		—	—		
Outdoor CO2 concentration $CO2_{ext}$		0.947 ±0.483	0.050		
Wind speed V_{wind}	**All the year**	1.090 ±0.450	0.015	0.1529	0.1223
Indoor relative humidity RH_{int}		—	—		
Indoor temperature $T_{int\text{-}lower}$		-4.90 ±0.519	< 0,0001		
Indoor temperature $T_{int\text{-}upper}$		—	—		

Table 12 Results of the VIF and GVIF analyses for the explanatory variables for all the year and different season thermostat changing models

Variable	Autumn			Winter			Spring			All the year		
	VIF*	Df*	GVIF$^{1/(2*Df)}$	VIF*	Df*	GVIF$^{1/(2*Df)}$	VIF*	Df*	GVIF$^{1/(2*Df)}$	VIF*	Df*	GVIF$^{1/(2*Df)}$
Outdoor temperature T_{ext}	6.54	1	2.55	—	—	—	—	—	—	1.22	1	1.10
Outdoor relative humidity RH_{ext}	4.58	1	2.14	1.32	1	1.15	—	—	—	1.17	1	1.08
Solar Flux F	—	—	—	1.21	1	1.1	1.09	1	1.04	—	—	—
Outdoor CO2 concentration $CO2_{ext}$	—	—	—	—	—	—	—	—	—	1.14	1	1.07
Wind speed V_{wind}	1.18	1	1.08	—	—	—	—	—	—	1.14	1	1.07
Indoor relative humidity RH_{int}	3.56	1	1.88	1.16	1	1.08	—	—	—	—	—	—
Indoor temperature $T_{int-lower}$	—	—	—	1.01	1	1.00	—	—	—	1.13	1	1.06
Indoor temperature $T_{int-upper}$	1.66	1	1.28	—	—	—	1.09	1	1.04	—	—	—

VIF*: Variance Inflation Factor.

Df*: Degree of freedom.