Modelling of occupant behaviour in non-residential mixed-mode buildings: The distinctive features of tropical climates

Maareva Payet, Mathieu David, Philippe Lauret, Manar Amayri, Stéphane Ploix, François Garde

Modelling of occupant behaviour in non-residential mixed-mode buildings : the distinctive features of tropical climates

Maareva Payet[a,b], Mathieu David[b], Philippe Lauret[b], Manar Amayri[c], Stéphane Ploix[c], François Garde[b]

[a] *Corresponding author : maareva.payet@univ-reunion.fr*
[b] *University of La Réunion - PIMENT laboratory, 15 avenue René Cassin, 97715 Saint-Denis, Reunion*
[c] *GSCOP laboratory, 46 Avenue Félix Viallet 38000 Grenoble*

Abstract

This paper focuses on the modelling of occupant behaviour in the case of a non-residential mixed-mode building on the tropical island of La Réunion. For such areas and types of buildings, occupants can operate passive solutions to achieve comfort while energy-consuming ones can offer alternatives during the hottest months. Yet, compared to other climatic zones, specific knowledge on occupant comfort and behaviour is limited, making the work of engineers difficult during the design phase. In this work, occupants' operations on hygrothermal comfort controls, such as windows and fans, were first measured and analysed. Secondly, these behaviours were modelled using two deterministic methods based on machine learning techniques and a probabilistic graphical model. A model was also implemented to estimate the number of people, using the power demand of the electrical outlets. The estimation ability of the behavioural models was evaluated and led to F1 scores greater than 0.7. A two-classifier model was proposed to estimate the level of ceiling fan use. This combined model slightly improves the F1 scores by more that 2 %, which demonstrates the necessity of taking into account the links between the different controls.

Keywords: Tropical climate, bioclimatic mixed-mode buildings, occupant behaviour modelling, machine learning techniques, operable openings, ceiling fans, thermal comfort

1. Introduction

1.1. Context

Buildings in tropical climates are facing a significant increase in energy needs. In particular, high energy-consuming systems such as cooling systems account for a significant portion of the building energy consumption [1]. Reducing this usage is therefore an important challenge. This study takes place in La Réunion, a French tropical island situated in the Indian Ocean. Classified as category A by Köppen - Geiger[2], the climatic conditions of La Réunion in summer (the wettest period) have a strong impact on the occupants' comfort due to high temperature and humidity. However, on this island, the trade winds, which are common regular winds blowing from east to west in the intertropical areas, temper the hottest days. These specific climatic conditions foster the development of mixed-mode buildings.

1.2. Definition of mixed-mode buildings

According to Brager et al. [3], a well-designed mixed-mode building first optimises the facades design using as much as possible passive solutions to minimize cooling loads. Then the conception promotes the use of natural ventilation, whenever it is feasible, to maximize comfort while reducing needs of energy-consuming systems. And finally, this type of buildings integrates the use of energy-consuming systems like air-conditioning when and where it is ultimately necessary. Mixed-mode buildings are classified into subcategories according to the operation strategy. The study of Kim et al.[4] revealed that occupants of a mixed-mode building are more tolerant or adapted to warm indoor thermal conditions when it is operated with natural ventilation, rather than air conditioning, in particular by adjusting their clothing. Mixed-mode buildings are becoming increasingly popular as they offer a more energy efficient alternative than conventional mechanical solutions. However, when indoor temperatures and humidity levels are rising, it can be more difficult to maintain comfortable conditions using only passive solutions, and occupants are more likely to add energy-consuming cooling systems. This is the principle of adaptive thermal comfort addressed by Nicol et al. [5] : "If a change occurs such as to produce discomfort, people react in ways, which tend to restore their comfort".

Controls, like windows or fans [6], are equipment available to occupants to achieve comfort, in particular hygrothermal comfort. Occupants both adapt to the building and behave to adapt the building to their individual needs[7][8] .

Within mixed-mode buildings, the usual hygrothermal controls available are :

- Passive systems such as numerous operable openings on opposite facades. In addition to providing a solution for air quality and thermal load removal, when airflow rates are sufficient, they offer the occupants the possibility to have natural cross-ventilation that decreases their perceived temperature and therefore the level of comfort, without using any energy. Openings are the key elements of a naturally ventilated design. According to a review done by Roetzel, A. et al. [9], the type of window depends on the climate. For instance, in warm climates, the effective size of the openings and their operability may be more important factors than the ones used to protect the building from outdoor conditions ;
- Low energy-consuming systems such as variable speed ceiling fans. Ceiling fans also have the task of reducing the occupants' perceived temperature, and thus increasing the comfort temperature, providing forced convective cooling by increasing the air movement. Indraganti, M et al. [10] found that air speeds of about 1.0 m/s provided by ceiling fans push the comfort temperature up by about 2.7 °C in naturally ventilated offices in the warm and humid climate of India. The term low energy-consuming is justified by their ability to increase the level of thermal comfort with very low energy requirements, unlike high energy-consuming systems such as air conditioning. Energy is only needed for the motor to turn the blades and thus to increase the air speed. A ceiling fan at full power needs only 50-75 W on average. They play a key role and are systematically made available to occupants in new efficient buildings in La Réunion ;

— High energy-consuming systems such as air conditioning. These systems offer an alternative during the hottest months but at the cost of a much higher energy consumption.

In non-residential mixed-mode buildings in tropical climates, the switch between these different operation strategies can be determined either manually by the occupants, or automatically by a building management system. It is a common practice to give occupants the opportunity to manage their own comfort by being active on the systems.

1.3. Occupant behaviour modelling : state of the art

The scientific community is already convinced that occupant behaviour has a significant impact on the energy performance of buildings [11]. Among other elements (such as the climatic conditions, the building envelope and the efficiency of the systems) Liu, Y et al [12] and Yu, Z. et al [13] listed occupant behaviour as one of the main element that affects energy consumption. Occupants have an impact by their presence and also by their actions on passive (i.e. windows), low energy-consuming (i.e. ceiling fans) and high energy-consuming systems (i.e. air conditioning) of the building. Presence is obviously a precondition for action to take place. In addition to the level of energy consumption, the designers must also take into account occupant behaviour and their needs, in order to make the right choices by designing the building itself [14]. For example, in the case of a naturally ventilated building, they have to determine the position of openings, their size and their type. If these systems are not ultimately accepted and used, the building does not operate efficiently. Another example, reported in [15], highlights that spatial layout and distance to the system are among the main factors influencing the opportunities for action. This study shows that occupants of open spaces are more "passive" regarding windows and thermostats, due to a greater distance with the control system, and also because they feel less free to perform actions when several people are present. Ignoring occupant behaviour can therefore be a potential cause of error in the design phase, when engineers have to estimate the future energy consumption of the building. The operational phase is often neglected and basic assumptions are made about how the building will operate, thus leading to significant gaps between prediction and actual performance. The work of the International Energy Agency (IEA), Annex 66 and more specifically Annex 79 : Occupant-Centric Building Design and Operation[16], has reinforced the knowledge in this area by emphasising the need to focus on practical implementation of behavioural models and their design consequences.

Many models have been developed to address this issue, but few are available for tropical climates. Indeed, the recent state of the art done by Carlucci, et al [17] stresses that few models have been implemented in non-air-conditioned buildings and especially in tropical climates. Among these few studies, energy demand prediction models have been developed. For instance, Yu et al. [18] trained a decision tree on data measured in residential buildings situated in the humid subtropical climate of Japan, but there are very limited feedbacks on behaviours and on how occupants actually operate on buildings controls in these conditions. In [19], the responsibility of the internal heat gains and occupant behaviour on the air-conditioning was revealed as the main source of the discrepancy between the models employed (i.e. artificial neural network and white-box via EnergyPlus) and the actual values

of energy demand in a university building located in Sao Paulo. Bavaresco, et al. [20] investigated the influence of occupant behaviour towards internal blinds on the energy efficiency of an office located in the south of Brazil, using surveys. The information collected was directly used to identify the internal loads of a building modeled with EnergyPlus. From surveys and environmental measurements, Indraganti, et al. [21] also emphasised the role of fans as controls to offset discomfort in both naturally ventilated and air conditioned office buildings in the hot and humid climate of India. A Logistic regression model was used to predict the probability of fan use according to different indoor environmental conditions.

The use of ceiling fans and their combination with the use of windows is poorly studied. Indeed, in the review done by Stazi et al. [22] about the driving factors and the contextual events influencing occupant behaviour in buildings, only 4 papers dealt with fan use. In addition, only one took place in the tropical climate. In the work of Rijal et al. [23], a logistic regression model estimates the level of use of windows and fans and also evaluates the change in the indoor environment caused by these actions. The buildings studied in this work are naturally ventilated offices located in different sites, including a hot and humid climate in Pakistan. A method for implementing the prediction algorithms in the ESP-r dynamic thermal simulation software was also depicted.

1.4. Problem statement and objectives of the study

Even if methods for predicting occupant behaviour already exist in the literature, we believe that the implementation of specific equipment like windows and fans in tropical mixed-mode buildings leads to specific occupant behaviors that is significantly different from temperate climates. To the best of our knowledge, these specific behaviors are not fully described and modelled in the literature. As a consequence, it is necessary to develop new specific models. A report of the Australian James Cook University [24] predicts that almost 50% of the world's population will be living in the tropics by 2050. It is therefore important to boost the level of knowledge in these areas, where models are lacking and where a significant increase in the population and construction rate is expected in the coming years.

Considering the different issues highlighted above, the following questions raise up :

— How the occupants of mixed-mode buildings in a tropical climate use the different available controls ?

— How do they combine these different controls ?

This works aims at providing new tools to understand and to accurately model the occupant behaviour of mixed-mode buildings in tropical climates, where controls on ceiling fans and windows are available. More specifically, the developed models will estimate the use of the passive (kind of windows called louvers) and low energy-consuming (fans) controls by taking into-account the expected level of hygrothermal comfort of the occupants. Furthermore, as the operation of the considered controls are dependent, we also propose to link some of the developed models to represent the corresponding relationships.

To do this, we first monitored occupants' actions in a mixed mode building located in a tropical climate. Behavioural data on window opening, ceiling fan use and environmental variables such as indoor air temperature, indoor relative humidity and meteorological data

were collected over a one-year period. Occupancy was not directly recorded. But as mentioned above, building's occupancy is a prerequisite key factor. Thus, a method for estimating the number of occupants, based on a regression decision tree algorithm and on the consumption data of the electrical outlets, is proposed. To model occupant behaviour, three types of machine learning techniques have been implemented : decision trees (DT), random forests (RF), and bayesian networks (BN). The development of these models is part of an overall desire and need to improve the assumptions made in the building design phase.

This paper is organised as follows. Section 2 describes the case study and the monitoring experimental set-up. Section 3 outlines the methodology and the different modelling techniques. The results of the study are discussed in section 4. Finally, section 5 gives some concluding remarks .

2. Case study

2.1. Description of the building

The case study is a design office of 310 m^2, occupying the ground floor of a residential building, located in La Réunion. Its 30 employees are architects, building engineers, landscape architects or urban planners, all using bioclimatic design principles in their projects, i.e. with an environmental awareness. This bioclimatic non-residential building is architecturally representative of the current construction trends. It is also easy to exploit since metering systems are already partially installed. Theses features justified the selection of this building as a case study.

The office is composed of two floors divided into several areas : open-spaces, a meeting room, a computer room and two individual offices. A floorplan of the case study is provided in Appendix A. Even if the building is located in an urban site, it is surrounded by vegetation, which helps to improve the hygrothermal, acoustic and aeraulic comfort conditions. The building was delivered in 2008, when there was no building regulations adapted to the tropical climate of Réunion Island. During the design phase the engineers followed the principles of natural ventilation and solar protection (see Figure 1). Figure 2 shows an interior view of the building under study. In the office areas, the temperature is regulated by natural crossing airflows. The latter are made possible trough the use of many manually adjustable and full-height openings, called louvers. The top and bottom parts are adjustable independently. Ceiling fans can also help to reduce the temperature felt by occupants, on the hottest days, when air temperatures are high and when there isn't enough wind. The computer room is air-conditioned thanks to a split system. Finally, the meeting room is equipped with louvers allowing natural ventilation, a ceiling fan and it is also air-conditioned by a split system. According to Brager et al. [3], we will refer to a zoned mixed-mode building because natural ventilation and mechanical cooling operate in different areas of the building. Occupants are active and have manual control over the systems. Note that in this study, we focused on occupant behaviour in the open space area.

2.2. Data collection

Environmental and occupant action data were monitored for approximately 1 year, from November 2019 to March 2021. The environmental data are the indoor air temperature

FIGURE 1. Overview of the North facade of the case study : green patio with large solar protections

FIGURE 2. View of the louvers and ceiling fans in open-spaces

FIGURE 3. Opened window (called louvers) with magnetic contact

and humidity measured by 9 sensors TESTO 174H. The outdoor temperature, humidity, wind and solar radiation were recorded using a weather station. Point measurements of the operative temperature (including radiation effects from walls) were made at 3 different positions inside the building, and compared with the indoor air temperature. The data related to the occupants' operation are the average power demand for the ceiling fans and for the electrical outlets, recorded every 10 minutes by energy meters. The status of the louvers was given by 37 magnetic contacts (see Figure 3) providing binary signals (0 = the louver is open, 1 = the louver is closed). For these kind of sensors, the data flow is asynchronous since new information only appears when there is a new action. An overview of the monitoring equipment is provided in Table 1.

The data recorded by the different sensors located inside the building show that the whole office area behaves as a single thermal zone. In fact, in this zone, the temperature and humidity differences measured were within the accuracy range of the sensors and were therefore ignored. The individual offices, with probably less thermal loads, were excluded from the data because their thermal conditions were significantly different. As a consequence, in this work, we used the average indoor air temperature and humidity of the remaining 8 sensors located in the open space area.

The number of occupants was not directly recorded. However, occupancy is a prerequisite

Table 1. Monitoring equipment inside the building

Equipment	Denomination	Number
Air temperature and hygrometry sensor	Testo 174H	9
Magnetic contact	NODON SDO-2-1-05	37
Multichannel concentrator + meters modules	OMEGAWATT	1
Data acquisition system	JEEDOM Pro v2	1

for action to take place and it is important to estimate occupancy prior building occupant behaviour models. The classical methods employed to infer the occupants' presence, such as CO2 sensors, occupancy detectors like passive infrared sensors (PIR)[25], acoustic sensors, RFID tags on ID cards, sonar sensors or video cameras [26], were not suitable for an open and naturally ventilated building where important air change rates are observed. In addition, the information obtained by some of these sensors, such as video cameras, sometimes requires heavy post-processing. The employees' time schedules can also provide knowledge, but we chose not to record them, because they can vary considerably from one individual to another and from one week to another. For example, in the present context, the specific work of building engineers (who are mobile workers) did not allow for the setting of regular time schedules. To estimate this data, we therefore developed a new model based on the total power demand of the electrical outlets, detailed in section 3.

3. Methodology

3.1. Data analysis

Prior to the modelling of occupant behavior, a statistical analysis of the data was carried out. The first objective was to understand the evolution of occupant behaviour regarding the opening of windows and the use of ceiling fans over a full year. The second one was to determine the proportion of control use as a function of the environmental variables in order to highlight thresholds for triggering controls.

3.2. Overall methodology

Let us recall that our main objective is to model occupant behaviour in mixed mode buildings in tropical climates. Specifically, we are interested in modelling behaviours related to the level of ceiling fans and window opening. To this end, three machine learning classification techniques were implemented. In order to complete the measured data, the "Electrical outlets power" was used as an input feature of a regression decision tree model to estimate the number of people. All the measured data and the estimated number of people were scaled to an hourly time step in order to build the final behavioural models. Figure 4 provides a general overview of the methodology employed. Note that the modelling components (i.e. occupancy and occupant behaviour models) of the proposed methodology are based on machine learning techniques namely Decision Tree (DT), Random Forests (RF) and Bayesian Networks (BN). These techniques are depicted at length in Appendix B.

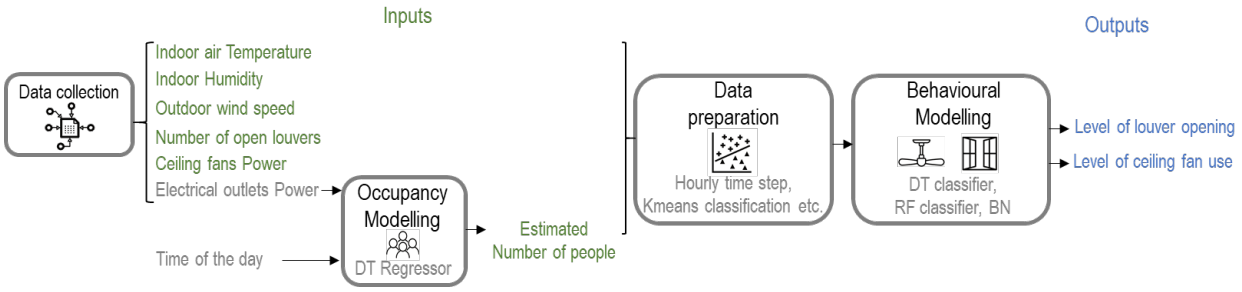


FIGURE 4. Overall methodology. In this scheme, DT stands for Decision Tree, RF for Random Forest and BN for Bayesian Networks

3.3. Occupancy modelling

Zhao et al. [27] proposed to estimate the occupancy of an open-space, based on the individual behaviours of 15 employees, using 3 techniques namely C4.5 decision tree, locally weighted naïve Bayes, and support vector machine. The data collected over 3 months were the electric consumption of the office appliances for each occupant, like the computers, monitors, desk lamps and other office equipment. In addition to these individual consumption data, pedometers were used as field data to train and validate the individual presence models. By combining this individual presence and absence information, the group occupancy schedule was generated for the open space. This study showed the correlation between presence and consumption of office equipment but at the expense of many expensive and intrusive sensors. Furthermore, it was noted that the results may be biased and dependent on occupant stringency, since occupants may forget to remove their pedometers outside the office.

In this work, we developed a simple occupancy model that does not require specific sensors (like pedometers) to estimate the number of people. Instead, we assumed that the overall occupancy can be inferred from the total power demand of the electrical outlets, which mainly include the computers of each person. The latter is directly measured by the energy meters. To build the training database, the first step was to count manually, every hour, the number of people present at their workstation, during 1 week (5 working days and 2 weekend days). The selected week was considered to be representative of a typical working week. Secondly, we extracted the total power demand of the electrical outlets recorded during this week. Then, we approximated the number of people present every hour, with a regression decision tree (DT) model, which is depicted in Appendix B. The time of the day and the total power demand of the electrical outlets were used as input features. Figure 5 describes the implementation of the model. 55% of the data were taken for training and 5-fold cross-validation, 45% for the test phase. Finally, the results of the model were extended to the whole year, in order to obtain a realistic estimation of the number of people over the year.

The output of this first regression decision tree was used as input for all subsequent behaviour models.

3.4. Behavioural modelling : Data preparation

The measured data and the estimated occupancy were pre-processed to get an homogeneous hourly time step for every variable. The choice of an hourly time step was motivated

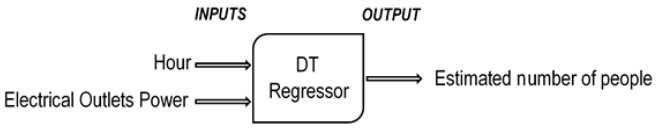


FIGURE 5. Implementation of Decision Tree (DT) regressor for the prediction of the number of people

by the fact that no significant variation in the number of occupant actions was detected below the 1h granularity.

Different input features were tested to build the behavioural models. Only the most relevant ones are listed in Table 2.

TABLE 2. Summary of the data used to create the models

Input variable	**Type**	**Class / Range**
Indoor air temperature [°C]	Numerical	[19.95 : 31.51]
Indoor humidity [%]	Numerical	[46.50 : 93.14]
Outdoor wind speed [m/s]	Numerical	[0 : 15.90]
Level of wind	Categorical	Low, Medium, High (Low for wind speed $\leq$ 4.7 m/s and High wind speed $\geq$ 9 m/s)
Level of comfort	Categorical	Discomfort, Comfort at 0 m/s, Comfort at 0.5 m/s, Comfort at 1 m/s
Number of people [u]	Numerical	[0 : 23]
Occupancy	Categorical	Absent, Medium, High (Absent for Occupancy $\leq$ 2 and High for Occupancy $\geq$ 19)
Number of open louvers [u]	Numerical	[0 : 43]
Ceiling fans Power [W]	Numerical	[0 : 750]
Output variable	**Type**	**Class / Range**
Level of louver opening	Categorical	Low, Medium, High (Low for a louver opening rate $\leq$ 16% and High for a louver opening rate $\geq$ 47 %)
Level of ceiling fans	Categorical	Low, Medium, High (Low for ceiling fans power $\leq$ 17% of the installed power and High for ceiling fans power $\geq$ 45% of the installed power)

Occupant behaviours were modelled using the classification techniques presented in Appendix B, namely classification decision trees, random forest and Bayesian networks. The models were constructed using either numerical or categorical variables, or both. Only categorical data were used in the case of BN. Regarding the outputs of the behavioral models i.e. level of fans and louvers opening, classes were created using the unsupervised clustering K-means algorithm, to reduce the number of possible values for each variable. A detailed description of this algorithm is provided in [28]. The goal is to group similar data points together from data without output labels. Furthermore, we believe that estimating behaviours in terms of classes and thus intervals of values is more realistic than discrete numerical outputs. 3 categories for each variable were defined, for a trade-off between modelling performance and physical consistency in the interpretation of the classes. For example, with only 2 categories for the variable "Level of ceiling fans", the difference between the minimum and maximum bounds of the "High" level was several hundred watts. This does not permit to

accurately ascertain the impact of this control. Table 2 summarises the variables and their possible values. For example, the occupancy variable in categorical form is called "Occupancy" and can take one of the following 3 levels : Absent, Medium or High while occupancy in discrete numerical form is called "Number of people" and ranges between 0 and 23. In the same way, the "Level of louver opening" and the "level of ceiling fans" were classified as : "Low", "Medium" or "High", based respectively on their numerical forms called "Number of open louvers" and "Ceiling fans Power".

For classification decision tree and random forest, thermal comfort data were directly expressed using the continuous variables of indoor air temperature and humidity. For bayesian network, the class form of thermal comfort level was defined based on the comfort levels inspired by the Givoni diagram, a tool widely used in design offices in tropical climates. This graphical display identifies comfort zones based on air temperature and relative humidity [29]. A comfort level of 0.5 m/s means that an air speed of 0.5 m/s is required to be in a comfortable situation. It should be noted that an air speed of 1 m/s can be achieved with the help of ceiling fans. For the comfort zone at 0 m/s, no wind is necessary to be comfortable The reader is refered to Appendix C for more details regarding the Givoni diagram.

Finally, it must be stressed that, in our case study, we are dealing with a dataset composed of multiple imbalanced classes, i. e. the number of instances available for the different classes is unequal. For example, the "Low" class for the level of ceiling fans contains 7620 data points while the "Medium" class contains 528 data points.

Original data of this study are available at Mendeley Data (doi : 10.17632/cyh3z6dnr3.1).

3.5. Behavioural modelling : Implementation

We implemented the following classification models :

1) Estimation of the level of louver opening deduced from indoor air temperature, indoor humidity and the number of people. The impact of the external wind speed on these results was also analysed (See Figure 6) ;
2) Estimation of the level of ceiling fan deduced from the same inputs, i.e. indoor air temperature, indoor humidity and the number of people (See Figure 7) ;

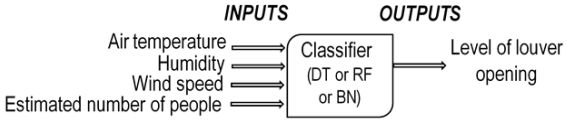


FIGURE 6. Model 1 : Classifier description for the estimation of the level of louver opening. Note that the classifier can be either a decision tree (DT), a random forest (RF) or a bayesian network (BN).

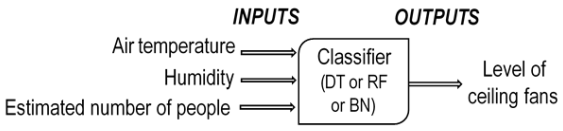


FIGURE 7. Model 2 : Classifier description for the estimation of the level of ceiling fan. Note that the classifier can be either a decision tree (DT), a random forest (RF) or a bayesian network (BN).

It is common to assume that occupants will use the openings first before turning on the fans. This is evidenced by the data collected in the present study (see Section 4) which demonstrate that the controls are linked. As a consequence, in this work, we designed models

that link the louver openings and the fan operations. Therefore, two additional variants to estimate the level of ceiling fan, incorporating information on the level of louver opening in addition to the previous inputs, were assessed :

3) Estimation of the level of ceiling fan deduced from measured level of louver opening (see Figure 8) . This is an "ideal" case where field data on the opening of louvers is available.

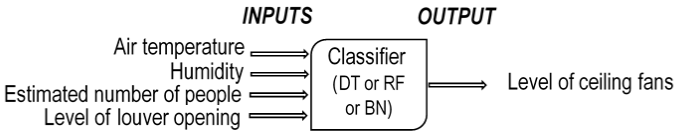


FIGURE 8. Model 3 : Classifier description for the estimation of the level of ceiling fan with information about louver opening from measurements. Note that the classifier can be either a decision tree (DT), a random forest (RF) or a bayesian network (BN).

4) Estimation of the level of ceiling fan adding estimated level of louver opening as input, i.e. linking with model 1. This model could be used when field data on louver opening are not available. Figure 9 explains the principle of linking classifiers. The structure of this model, i.e. the link between the opening of louvers and then the use of ceiling fans, is consistent with the results obtained in section 4 showing that an occupant first acts on the louvers and then on the ceiling fans in order to satisfy his comfort needs.

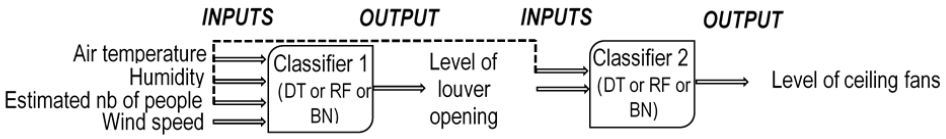


FIGURE 9. Model 4 : Combination of 2 classifiers to estimate the level of ceiling fan with estimated level of louver opening as input. Note that the classifiers can be either decision trees (DT), random forests (RF) or bayesian networks (BN).

Regarding the building and the assessment of the models, 70% of the data were used respectively for the training phase and 30% for the testing phase. It must be noted that we also employed a 10-fold cross-validation in order to optimize the models' hyper-parameters.

Finally, following the framework proposed by [30] in their review of machine learning in building load prediction, we propose a synthetic description of our models in Appendix D.

3.6. Performance evaluation

The performance of the occupancy model by regression decision tree was evaluated using the classical determination coefficient (R^2) and the Root Mean Square Error (RMSE) [31]. R^2 must be maximised while RMSE must be minimised. RMSE is described as follow :

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\text{ Estimated }_i - \text{ Actual }_i)^2}{N}} \quad (1)$$

where Actual is the value taken by variable to predict, Estimated the value estimated by the model and N the number of observations.

To assess the performance of classifiers, the accuracy metric is the most employed indicator [32]. However, as mentioned in section 3.4, we are dealing with imbalanced datasets and some metrics are more appropriate than the accuracy metric. For instance, the F1 score is used in [33] to compare the performance of several classification models on an imbalanced multiclass dataset. Hence, in this work, the F1 score, also known as balanced F-score or F-measure, is used to assess the performance of the different classifiers used to model occupant behaviours .

The F1 score is a harmonic mean combining model precision and recall into a single number (see Equation 4). Precision and Recall are calculated from the confusion matrix (see Figure 10) for each class, according to equations 2 and 3. Finally, the F1 scores of every classes are averaged and combined into a single value, the overall F1 score. More precisely, it should be noted that, to deal with the imbalance in the data, the average F1 scores were weighted by the number of true instances of each class.

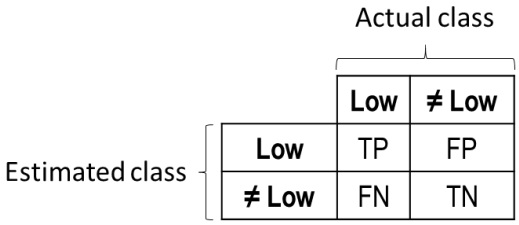


FIGURE 10. Confusion matrix related to a specific class

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$F1 = \frac{2 \cdot \text{ precision } \cdot \text{ recall}}{\text{precision } + \text{ recall}} \tag{4}$$

The F1 score metric should be maximised, i.e. an F1-score reaches its best value at 1 and its worst score at 0.

4. Results

4.1. Data analysis

Temporal analysis. The operation of controls fluctuates depending on the period of the year. In summer [1], their use is more frequent than in winter. Figure 11 shows the evolution of the

1. Note that summer in La Reunion runs from November to April.

level of ceiling fans along the year. The hours of use of the fans are regular, between 9 :00 am and 8 :00 pm, i.e. during the occupancy hours. The highest levels of fans are found between December and April, the warmest months. In 2020, the fans are less employed in March-April than in 2021 at the same period. This situation may relate to the COVID-19 pandemic that happened during these two particular years. Indeed, a total lockdown (red zone on Figure 11) took place for a few days in March 2020, with no employee able to access the offices. During the two periods of partial lockdown (orange zone on Figure 11) came the implementation of teleworking and shift work. During this phase, employees were present but in a reduced number. Even if the number of occupants was lower during these periods, there were still actions on the controls of hygrothermal comfort. Indeed, occupation is the starting point of an action on louvers or fans that is why our models are built to take into account a decrease in occupancy. However, it can be assumed that for a "typical" year the period of high fan use would extend over the whole period from December to April with higher levels in February-March. Similarly, louvers are employed in high proportions during the month of March. Between April and May, it is a medium use and between June and August it is a low use.

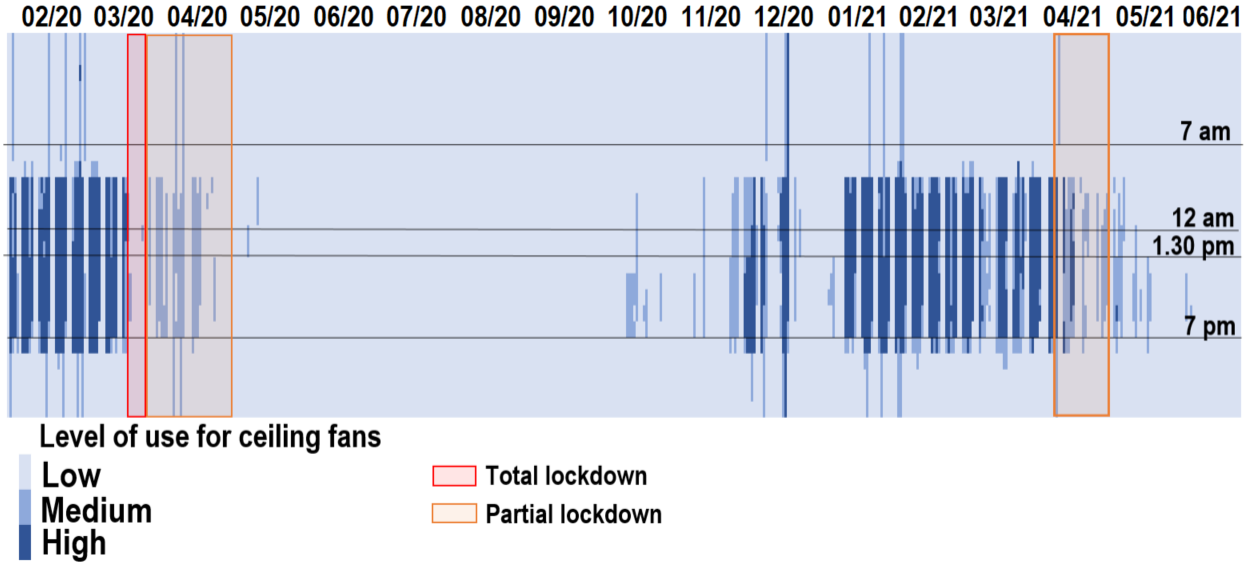


FIGURE 11. Evolution of power levels for ceiling fans. The y-axis gives the hourly evolution of the fan use while the x-axis provides around 1.5 year of daily operation.

Time periods of high levels of use for both controls are those of hygrothermal discomfort as well.

Statistics and thresholds. Figure 12 shows the ratio of opened windows and fan use during occupancy hours, for each level of comfort given by the Givoni chart. The distribution of the hygrothermal data is plotted in the Givoni diagram (see Appendix C). For both controls, the proportion of use increases as the comfort decreases, which is in line with the principle of adaptive thermal comfort mentioned in the introduction [5]. Windows are employed alone

in the comfort zone with no wind (0 m/s). The fans start to be employed when an air speed of 0.5 m/s is required to be comfortable.

Figure 13 specifically focuses on the proportion of control use as a function of the indoor air temperature only. As shown by this figure, when the interior temperature is 25°C, in average 25% of the louvers are opened and all the fans are off. The occupants switch on the fans once this threshold of 25°C is overtaken, while the louvers are already opened. Therefore, users operate first on louvers then on fans. A proportion of open louvers of 0.5 is obtained for an average indoor air temperature equals to 27.6°C, while for the ceiling fan control, this proportion is reached at 31.5°C. It should be noted that our data showed that the effect of wall radiation was not significant for our case study due to the large solar shading on the facades. The indoor air temperature is therefore considered equivalent to the average radiant temperature. In the study by Kumar, S. et al [34], the window opening proportion of 0.5 was obtained for 28.5°C. In addition, the maximum proportions they found were 75% for window opening and 81% for fan use. In our case study, the occupants operate more on the openings since up to 88% are opened at 32°C. However, the maximum rate of use of the fans did not exceed 62% over the year of measurement. The operation of controls increases as comfort decreases, but they are never used at their maximum.

This suggests either the controls are not easily accessible or available, or there are too many systems available compared to the occupants' needs to achieve thermal comfort.

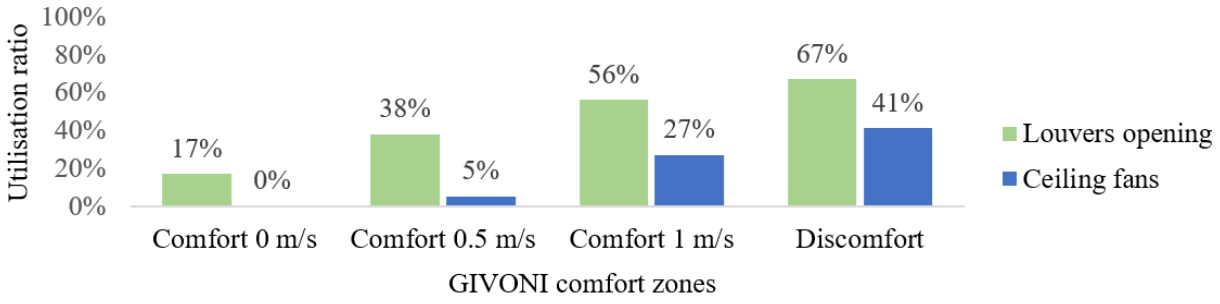


FIGURE 12. Controls use ratio according to the level of comfort

Figure 14 reveals a strong correlation between the operation of the controls during hours of occupancy and the indoor air temperature. Indeed linear regressions give coefficients of determination $r^2 = 0.7$ for ceiling fans power and $r^2 = 0.82$ for the number of open louvers. According to our measurements, the relationship between external wind speed and the use of the controls studied does not appear to be linear.

The first control threshold is 22°C for louver opening. The following one is above 25 °C for turning on the ceiling fans. This is consistent with our model construction 3 and 4 where the estimation of the level of ceiling fans is linked to the estimation of the louver opening level. When the temperature exceeds 26°C, the minimum power demand of the ceiling fans is 200W. This power demand increases in accordance with the indoor air temperature augmentation. This means that either the speed of the selected fans increases, or the number of fans switched on increases, or both at the same time. The maximum power demand is observed between

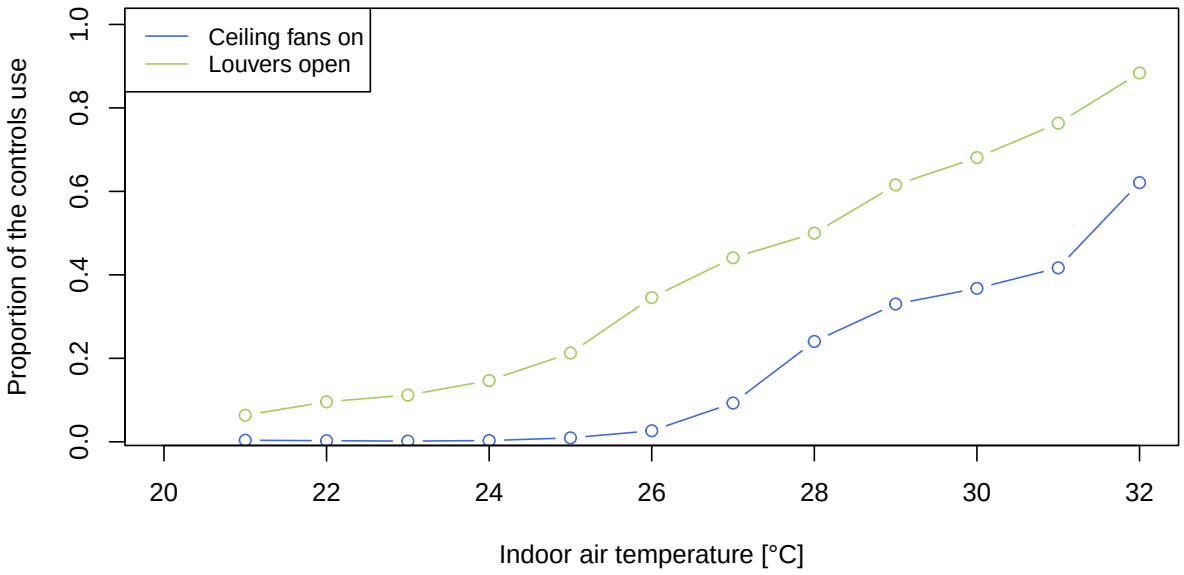


FIGURE 13. Proportion of the controls use according to indoor air temperature

28°C and 30°C. On the other hand, an average of less than 10 louvers are opened below 26°C. Above 26°C, the number of open louvers is more than 20. Whatever the comfort level, some louvers always remain closed since we never reach the maximum number of louvers present in the building. An analysis of the spatial distribution of the louvers showed that the ones situated furthest from the workstations were the least opened. In the study of Rijal et al. [23], the threshold at which controls were activated is an indoor temperature of 20°C. In our study, the higher thresholds suggest that the building is better designed towards ventilation, the controls more efficient and the occupants may be more adapted to higher temperatures. Note also that the "discomfort" and "comfort at 1m/s" zones overlap for the most important uses of controls, which questions the Temperature/Relative Humidity bounds defined for these two zones.

4.2. Occupancy modelling

The regression DT implemented to estimate the occupancy has a fairly good determination coefficient $R^2 = 0.937$ and an Root Mean Square Error (RMSE) = 2.176 people, highlighting the link between the occupancy, the time of the day and the electrical outlets power demand. Figure 15 shows the resulting tree, for the first 4 levels of depth, that highlights the structure of the model. The final leaves are the possible values that occupancy can take, i.e. the number of people. The general rule is that a higher power demand means a higher number of people. Above a power demand of 1279.5 W, i.e. on the right side of the tree, the occupancy can range from 10 to 26 people, except during the lunch break when no

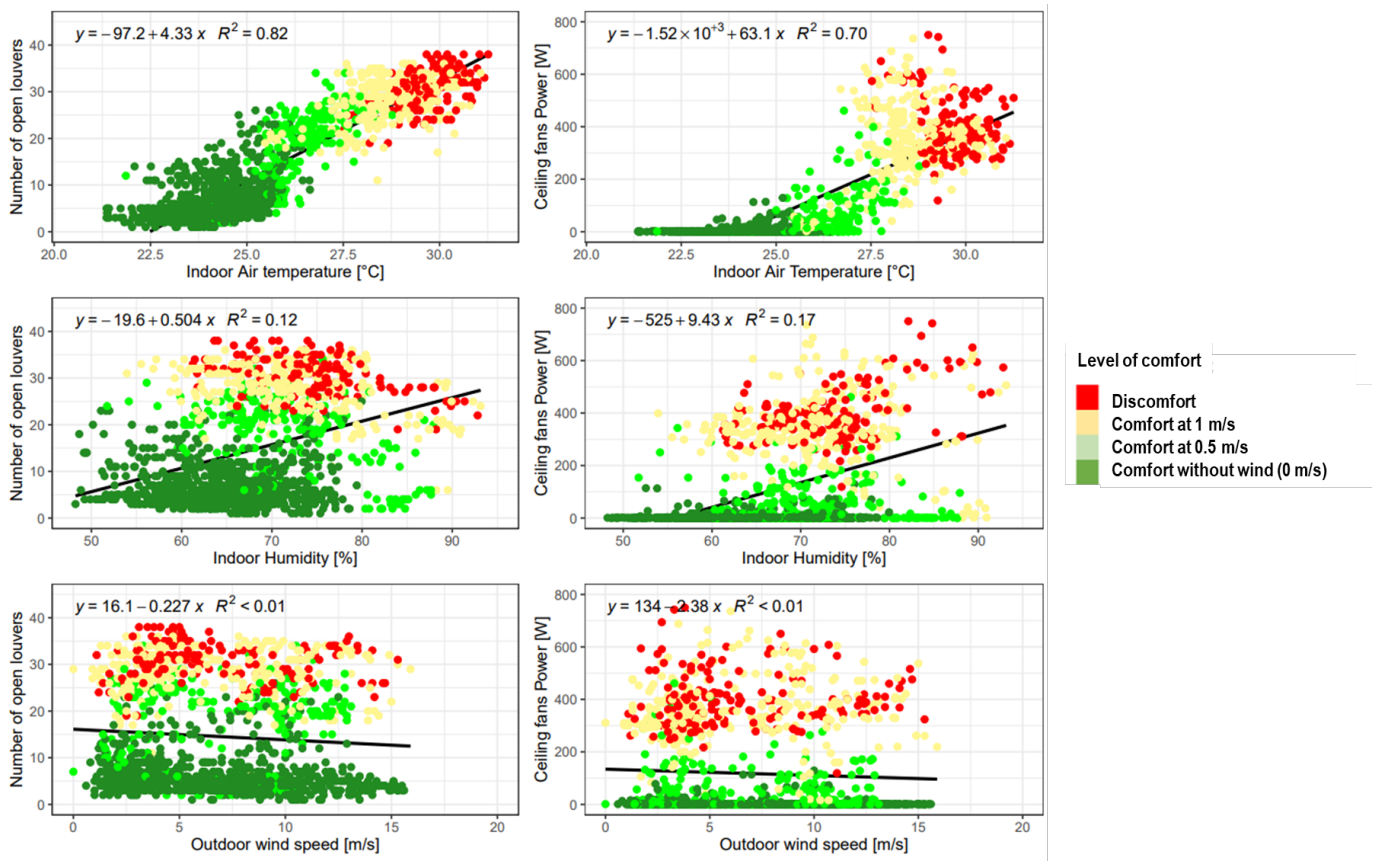


FIGURE 14. Relationships between occupant behaviour, i.e ceiling fan use and windows opening, and the indoor air temperature, indoor humidity and outdoor wind speed

one is present. Indeed, on this branch, we can see a specific rule which takes into account the decrease in the occupation between 12 :00 pm and 1 :30 pm. Below a power demand of 1279.5 W, i.e. on the left side of the tree, the occupancy does not exceed 5 persons. A minimum power demand of 990 W is also observed, even during the night. This lower bound of the power demand refers to a fixed consumption of equipment always plugged in, such as the water cooler or refrigerator, as well as the standby power of appliances such as the printer. Some computers are also not switched off during night. In addition, arrival and departure times seem to be well taken into account since specific rules exist at 7 :00 am and 4 :30 pm, which are key times for some employees.

We built a first DT that only took the average energy demand of electrical outlets and the hour as unique input features. Unfortunately, as some people leave their computers on during lunch time, this early model was unable to take into account the drop in occupancy at noon. Incorporating the time of day into the features enables more realistic estimations during the lunch break, when the number of people decreases. A weakness of this model is the small amount of data taken for the validation phase, and it would be desirable to collect a new set of real occupancy data over a new week to confirm the results.

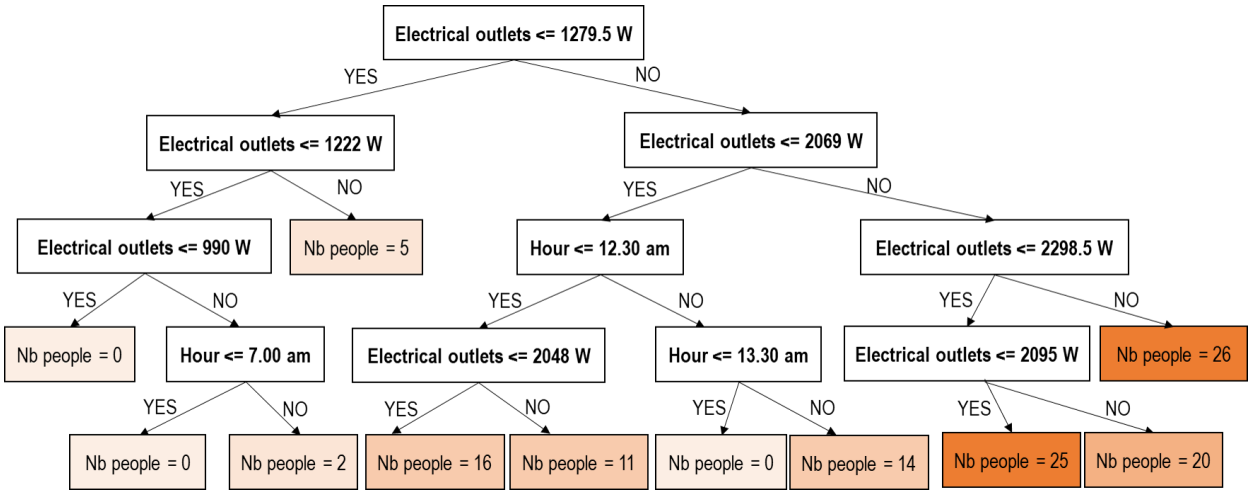


FIGURE 15. Regression Decision tree obtained for occupancy estimation for the first 4 levels of depth

4.3. Behaviour modelling

Figure 16 represents the overall F1 scores obtained for the estimation of the level of louver opening (model 1). The hatched bars are the results obtained when the outdoor wind speed is included as an input feature. As shown by Figure 16, overall the RF classifier performs slightly better than the two other models. In addition, the inclusion of the external wind speed therefore provides additional information and slightly improves the performance of the DT and RF model for the level of window opening.

Figure 17 compares the ceiling fan level estimation with models 2 (light-coloured bars), 3 (bright-coloured bars) and 4 (gridded bars). Again, when no information on louver is used, the RF classifier slightly outperforms the two other classifiers, although the others perform also well. First, it must be noted that the addition of louver data as input (i.e. models 3 and 4) improves the modelling results for the estimation of the level of ceiling fans. This confirms the link between these two controls. Second, and more interestingly, the integration of louver opening estimated by model 1 (i.e. model 4) does not significantly degrades the performances of the models. Model 4 can therefore be used when no window opening data is available.

This section only presents the overall results of the weighted average F1 score. The detailed scores obtained by each class can be found in Appendix E.

4.4. Discussion

The analysis of a full year of monitoring data suggests that dynamic thermal simulations in the design phase should take into account the dependency between the indoor air temperature, the surface of opened windows and the level of fans. Indeed, above 22°C, an increasing air change rate due to window opening should be defined and, in the same way, above 25°C the electrical consumption of ceiling fans should rise. The thermal simulation models should also assume seasonal calendars for the operation of the systems and not just a single one for the whole year.

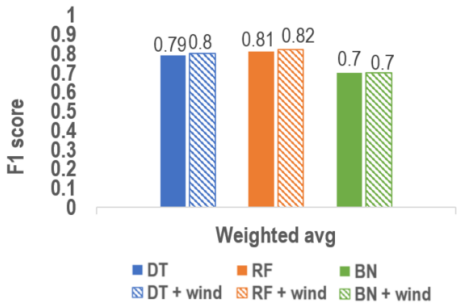


FIGURE 16. F1 score for the estimation of the level of louver opening (model 1)

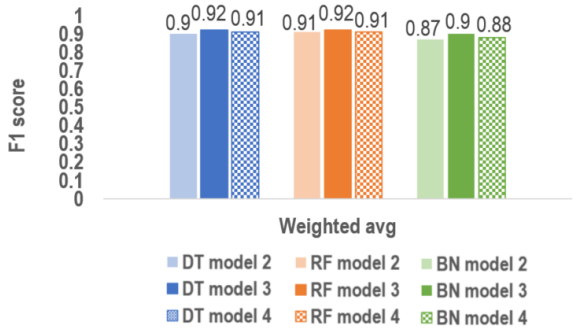


FIGURE 17. F1 scores for estimating the level of ceiling fans without louver data (model 2) VS integrating louver data : optimal case from measured data (model 3) VS combined case from estimated louver data (model 4)

Occupancy data is difficult to capture in a mixed-mode building composed mostly of open spaces and using natural ventilation most of the time. To overcome this difficulty, a new occupancy model was implemented using the power demand of the electrical outlets and the hour of the day. This non-intrusive method provides a reliable estimate of the number of people in the building. Collecting occupancy data for an additional week would complete and confirm the current validation phase. In addition, the occupancy model proposed in this work estimates the number of people based on the energy consumption of the building, and is therefore valid for this specific case study. In order to generalize the approach, a possible solution would be to use dimensionless ratios such as "electrical outlets/floor area" and "percentage of people".

In this work, three modelling techniques were investigated to estimate the level of window (called louvers) opening and ceiling fan use. In both cases, the Random Forest classifier obtained the highest F1 scores. However, each of the other models also provides useful information for designers. Decision trees have the advantage of providing readable classification rules by highlighting the most important key variables. Bayesian networks can generate additional information by calculating the probabilities of class membership.

Different input features data sets were tested in the modelling process. Better results were obtained by using directly the numerical values of indoor air temperature and of humidity instead of using the classes derived from the Givoni comfort zones. This result highlights the discrepancies between this comfort model and the comfort actually perceived by occupants. An indoor condition corresponding to a comfort zone different from "comfort without wind" does not necessarily mean that the controls are employed. On the other hand, they can be operated outside the comfort zone of 1 m/s. Similarly, the addition of outdoor wind speed as an input feature also slightly improved the F1 score for the estimation of the level of window opening. However, for an application on other case studies, it may be more suitable to choose a model without this feature. Indeed, this data is not always reliable and available on-site.

As demonstrated in this work, the estimation of ceiling fan use was improved by adding

information on window opening as an input feature, which demonstrates that these two controls are linked. A model combining 2 classifiers was therefore proposed, which provides an answer to Carlucci's statement [17] about the need to model "combined actions providing a wider view of human actions and their impact in terms of energy consumption and occupants' comfort". Hence, since the design phase of the building, these two systems should not be considered separately.

The lockdown and teleworking periods occurred during the summer, when the controls are the most solicited. The temperature and humidity data for these periods indicate that the "Medium" and "High" levels classes for the use of controls should be better populated under normal conditions of operation. This particular distribution of the data has two effect on the results. First, the models were trained with a higher proportion of low occupancy data than would be expected for a normal year. Consequently, they probably perform better under these low occupancy conditions. However, even if the lockdown periods had been removed from the data set, the training samples corresponding to medium and high level of behaviours, would have been the same and the models should perform identically for these conditions. Second, for a distribution of the data corresponding to a normal year, the differences in terms of F1 scores between the different classes should be significantly reduced. Oversampling or undersampling techniques could also be applied to reduce these differences. However, we use the F1 score to compare the performance of the different models developed in this work and the particular distribution of the classes resulting on the lockdown period has no effect on the results of the comparison.

5. Conclusion and outlooks

The main objective of this work was to model occupant behaviour in mixed mode buildings in tropical climates. The behaviors were related to the use of ceiling fans and window opening. A mixed mode office building, representative of the current construction trend, located in the tropical climate of La Réunion was therefore monitored during one year. Data related to hygrothermal comfort and occupant behaviour towards ceiling fans and window opening were collected. A new occupancy model based on a regression decision tree was first implemented to estimate the number of people. In a second step, the occupant behaviour was modelled using classification decision trees, random forests and bayesian networks.

A statistical analysis allowed to determine thresholds for the operation of the controls. Above 22°C, the rate of indoor air exchange increases due to the increase in the level of window opening, and then, above 25°C, the power consumption of the ceiling fans increases. The estimation ability of the models was evaluated and random forests led to the best F1 scores. However, the other methods also demonstrated their capacity to represent occupant behaviour reliably and provided useful information. A combined model of two classifiers was proposed to estimate the level of ceiling fans. The relationship between occupancy and behaviours, comfort level and behaviours, and between the controls themselves was highlighted. More importantly, it was shown that the use of windows and ceiling fans are linked and should not be considered separately.

This study is a new contribution to the understanding and consideration of occupant behaviour in tropical climates. The solution of mixed mode buildings, which take advantage of natural ventilation and reduce the use of air conditioning, is an interesting option in tropical climates, or even in other climates which are likely to become warmer in the coming years. The methodology proposed in this work for estimating, firstly, occupancy in a non-intrusive way, and secondly, the level of use of the controls, provides a framework that can be replicated to collect more data in other buildings. We believe that the amount of field data measured for different building configurations and occupant types needs to be increased in order to help understand behaviours in tropical climates.

Further work will consist in implementing these models with a dynamic thermal simulation tool like EnergyPlus, in order to improve hypotheses in the design phase.

We hope these models will permit reducing the gap between predicted and actual energy consumption in buildings, while enlightening designers on the potential of passive or low energy-consuming solutions, towards an efficient architecture where occupants are placed at the centre of the thinking.

6. Acknowledgement

This work is part of an on-going PhD thesis funded by the National Association for Research and Technology (ANRT) and the design Office LEU Reunion about the impact of occupant behaviour in non-residential bioclimatic buildings in tropical climates.

Références

[1] S. H. Réunion, Bilan énergétique de la réunion 2018, édition 2019, Tech. rep., SPL Horizon Réunion (2019).
URL https://energies-reunion.com/nos-actions/observation/bilan-energetique-de-la-reunion-2/

[2] M. C. Peel, B. L. Finlayson, T. A. McMahon, Updated world map of the köppen-geiger climate classification, Hydrology and earth system sciences 11 (5) (2007) 1633–1644.

[3] G. Brager, S. Borgeson, Y. Lee, Summary report : control strategies for mixed-mode buildings, Tech. rep., Center for the Built Environment (2007).

[4] J. Kim, F. Tartarini, T. Parkinson, P. Cooper, R. De Dear, Thermal comfort in a mixed-mode building : Are occupants more adaptive ?, Energy and Buildings 203 (2019) 109436.

[5] J. F. Nicol, M. A. Humphreys, A stochastic approach to thermal comfort-occupant behavior and energy use in buildings/discussion, ASHRAE transactions 110 (2004) 554.

[6] I. Raja, J. Nicol, K. McCartney, M. Humphreys, Thermal comfort : use of controls in naturally ventilated buildings, Energy and Buildings 33.

[7] A. Mahdavi, The human dimension of building performance simulation, in : 12th International IBPSA Conference : Building Simulation, 2011, pp. 14–16.

[8] W. Turner, T. Hong, A technical framework to describe occupant behavior for building energy simulations, Tech. rep., Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States) (2013).

[9] A. Roetzel, A. Tsangrassoulis, U. Dietrich, S. Busching, A review of occupant control on natural ventilation, Renewable and Sustainable Energy Reviews 14 (3) (2010) 1001–1013.

[10] M. Indraganti, R. Ooka, H. B. Rijal, G. S. Brager, Adaptive model of thermal comfort for offices in hot and humid climates of india, Building and Environment 74 (2014) 39–53.

[11] É. Vorger, Étude de l'influence du comportement des habitants sur la performance énergétique du bâtiment, Ph.D. thesis, Paris, ENMP (2014).

[12] Y. Liu, L. Yang, W. Zheng, T. Liu, X. Zhang, J. Liu, A novel building energy efficiency evaluation index : Establishment of calculation model and application, Energy Conversion and Management 166 (2018) 522–533.
[13] Z. Yu, B. C. Fung, F. Haghighat, H. Yoshino, E. Morofsky, A systematic procedure to study the influence of occupant behavior on building energy consumption, Energy and buildings 43 (6) (2011) 1409–1417.
[14] M. Marschall, J. Burry, Can the use of stochastic models of occupants' environmental control behavior influence architectural design outcomes ?-how field data can influence design outcomes, Proceedings of the 24th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 1 (2019) 715–724.
[15] L. Marín-Restrepo, M. Trebilcock, M. Gillott, Occupant action patterns regarding spatial and human factors in office environments, Energy and Buildings 214 (2020) 109889.
[16] E. ANNEX, Definition and simulation of occupant behavior in buildings, Tech. rep., IAE (2019).
[17] S. Carlucci, M. De Simone, S. K. Firth, M. B. Kjærgaard, R. Markovic, M. S. Rahaman, M. K. Annaqeeb, S. Biandrate, A. Das, J. W. Dziedzic, et al., Modeling occupant behavior in buildings, Building and Environment 174 (2020) 106768.
[18] Z. Yu, F. Haghighat, B. C. Fung, H. Yoshino, A decision tree method for building energy demand modeling, Energy and Buildings 42 (10) (2010) 1637–1646.
[19] A. H. Neto, F. A. S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy and buildings 40 (12) (2008) 2169–2176.
[20] M. V. Bavaresco, E. Ghisi, Influence of user interaction with internal blinds on the energy efficiency of office buildings, Energy and Buildings 166 (2018) 538–549.
[21] M. Indraganti, R. Ooka, H. B. Rijal, G. S. Brager, Adaptive model of thermal comfort for offices in hot and humid climates of india, Building and Environment 74 (2014) 39–53.
[22] F. Stazi, F. Naspi, M. D'Orazio, A literature review on driving factors and contextual events influencing occupants' behaviours in buildings, Building and Environment 118 (2017) 40–66.
[23] H. B. Rijal, P. Tuohy, M. A. Humphreys, J. F. Nicol, A. Samuel, I. A. Raja, J. Clarke, Development of adaptive algorithms for the operation of windows, fans, and doors to predict thermal comfort and energy use in pakistani buildings, American Society of Heating Refrigerating and Air Conditioning Engineers (ASHRAE) Transactions 114 (2) (2008) 555–573.
[24] University James Cook Australia, State of the tropics - report, Tech. rep., University James Cook Australia (2014).
URL `https://www.jcu.edu.au/state-of-the-tropics`
[25] R. H. Dodier, G. P. Henze, D. K. Tiller, X. Guo, Building occupancy detection through sensor belief networks, Energy and buildings 38 (9) (2006) 1033–1043.
[26] M. Amayri, A. Arora, S. Ploix, S. Bandhyopadyay, Q.-D. Ngo, V. R. Badarla, Estimating occupancy in heterogeneous sensor environment, Energy and Buildings 129 (2016) 46–58.
[27] J. Zhao, B. Lasternas, K. P. Lam, R. Yun, V. Loftness, Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining, Energy and Buildings 82 (2014) 341–355.
[28] X. Gao, A. Malkawi, A new methodology for building energy performance benchmarking : An approach based on intelligent clustering algorithm, Energy and Buildings 84 (2014) 607–616.
[29] A. Lenoir, On comfort in tropical climates. the design and operation of net zero energy buildings, Ph.D. thesis, Université de la Réunion (2013).
[30] L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, W. Livingood, A review of machine learning in building load prediction, Applied Energy 285 (2021) 116452.
[31] T. Liu, Z. Tan, C. Xu, H. Chen, Z. Li, Study on deep reinforcement learning techniques for building energy consumption forecasting, Energy and Buildings 208 (2020) 109675.
[32] P. Branco, L. Torgo, R. Ribeiro, A survey of predictive modelling under imbalanced distributions, arXiv preprint : 1505.01658.
[33] B. Jeong, H. Cho, J. Kim, S. K. Kwon, S. Hong, C. Lee, T. Kim, M. S. Park, S. Hong, T.-Y. Heo, Comparison between statistical models and machine learning methods on classification for highly imbalanced

multiclass kidney data, Diagnostics 10 (6) (2020) 415.

[34] S. Kumar, M. K. Singh, V. Loftness, J. Mathur, S. Mathur, Thermal comfort assessment and characteristics of occupant's behaviour in naturally ventilated buildings in composite climate of india, Energy for Sustainable Development 33 (2016) 108–121.

[35] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees, Routledge, 2017.

[36] R. Genuer, J.-M. Poggi, Arbres CART et Forêts aléatoires, Importance et sélection de variables, hal-01387654v2, 2017.

[37] R. Garreta, G. Moncecchi, Learning scikit-learn : machine learning in python, Packt Publishing Ltd, 2013.

[38] P. Naïm, P.-H. Wuillemin, P. Leray, O. Pourret, A. Becker, Réseaux bayésiens, Eyrolles, Paris 3 (2008) 120.

[39] A. Ankan, A. Panda, pgmpy : Probabilistic graphical models using python, Proceedings of the 14th Python in Science Conference. Citeseer 10.

Appendices

A. Floorplan of the case study

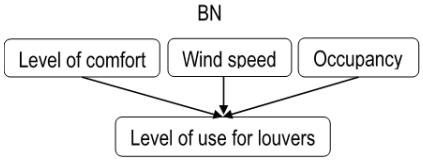
B. Machine learning techniques selected in this study

Decision Trees. The first technique we employed is the supervised algorithm of decision trees (DT). Decision trees are part of the data-driven deterministic methods and are able to predict either a numerical value (regression tree, as applied for the estimation of the number of people) or a class (classification tree, as applied for the behavioural models). A tree is a decision aid for a plausible value of the target variable Y (the label) we want to explain, when the values of the input variables (the features) are known. A tree is read by going down from the root to one of the final leaves where the final decisions (i.e. the possible values of Y) are located, passing through intermediate nodes. It is built iteratively. At each step, or "node", the data is separated into two subsets, following an "If, Then" rule applied to an input variable. For each step, these input features are chosen to provide the best possible separation of Y values. The objective is to obtain the optimal sequences of rules to explain the different possible values of Y[35]. For a new dataset, the resulting tree will lead to a decision, following the path obtained when applying the rules. In this way, we can understand the relationships between the variables in an explicit way [18]. Moreover, decision trees require relatively little effort for data preparation, which represents a great advantage during the design phase. Several algorithms have been developed over the years for decision trees. We used the decision tree algorithm of the scikit-learn library available for Python, which is based on an optimised version of the CART algorithm [36].

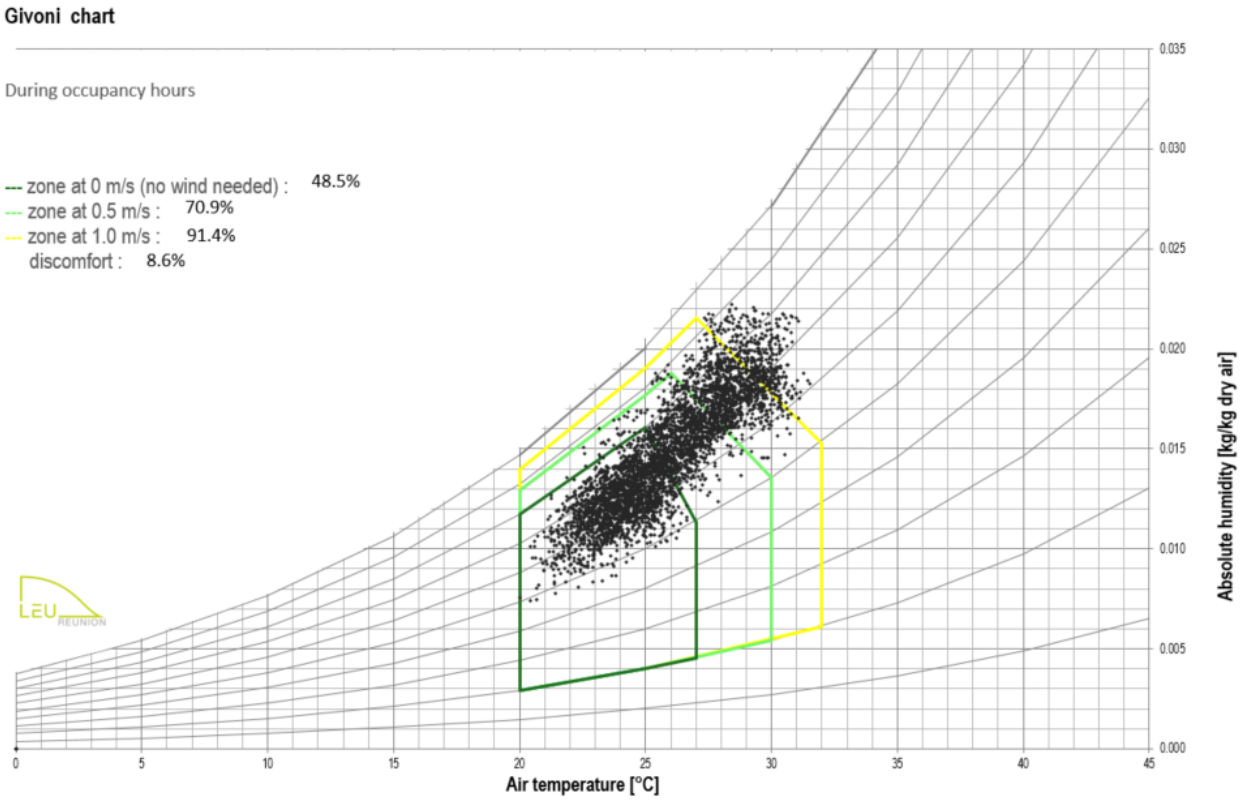
Random Forests. Trees can are easily prone to overfitting i.e. over complex trees will not perform well on unseen data. Trees can also become unstable to small variations in the data. To avoid this situation, we can control the model complexity by using mechanisms such as pruning the tree, setting the minimum number of samples required at a leaf node or setting the maximum depth of the tree. Another solution is the use of derived models such as random forests (RF), also belonging to the data-driven deterministic models. RF will fit a number of decision tree classifiers to various boostrapped subsamples of the data set and take the average (for a numerical variable) or the majority answer (for categorical variables) to improve the performance of the model during the testing phase. Several decision trees will therefore be trained. The strengths (and weaknesses) of each tree are aggregated. However, contrary to DT, the results obtained by a RF are not easily readable. In this work, we used the random forest classifier algorithm from the scikit-learn library of Python [37].

Bayesian Networks. The third type of technique is the Bayesian network (BN) which belongs to the probabilistic graphical methods. A BN contains conditional probabilities which is the main difference with deterministic trees. The principle of this method is to calculate a probability for the target variable, given specific observations. Like a decision tree, a BN provides an explicit and intuitive graphical representation. In addition, a BN provides additional information included in the probability distributions. The figure below represents the structure of the BN implemented in this work to estimate the level of louver opening. The nodes represent the random variables associated with a conditional probability table. The arcs that connect the nodes are oriented and indicate not only a simple correlation, but

also the cause/effect relationship between 2 variables [38]. To build the BN models, we used an algorithm from the pgmpy python library [39].



C. Hygrothermal data on the Givoni chart



8.6% of the hygrothermal data are in the discomfort zone during occupancy hours, 48.5% are in the zone where no wind is needed to be comfortable (0 m/s), while 42.9% are in the zone where an air speed of 0.5 or 1 m/s is necessary to achieve comfort. These air speeds can be reached by natural ventilation (use of louvers) or by the use of ceiling fans.

D. Identity card of the models implemented in this work

This identity card framework was proposed by Zhang et al. [30].

General information	Building type	Office building
	Building numbers	Single building
	Type to predict	level of ceiling fans / level of louver opening
	Application scenario	Building parameter design
	Forecasting horizon	Hours
Data description	Source of data	Real Building Automation System / independent sensors
	Sampling interval of data	Hours
	Data cleaning method	Timestamp formatting, Data cleaning, Data resolution processing, Scaling
	Training/validation/testing data	70 % training + 10-fold cross validation dataset 30 % testing dataset
	Nb of data	8640 points from March 2020 to February 2021
Feature engineering	Feature extraction	From measurements, Occupancy estimated by regression decision tree
	Final features used	Nb of people/level of louver opening/Indoor temperature & humidity
Algorithm	Main algorithm structure	Decision trees / Random Forest / Bayesian network
	Other support techniques	Clustering technique : kmeans
Performance evaluation	Error metrics	Weighted average F1 score

E. Detailed results

	Model 1			Model 2			Model 3			Model 4		
Class	BN + wind	RF + wind	DT + wind	DT	RF	BN	DT	RF	BN	DT	RF	BN
Low	0.76	0.83	0.81	0.96	0.96	0.95	0.97	0.97	0.97	0.96	0.96	0.95
Medium	0.68	0.82	0.80	0.19	0.16	0.00	0.26	0.19	0.04	0.20	0.16	0.00
High	0.56	0.74	0.75	0.80	0.83	0.62	0.86	0.87	0.88	0.82	0.84	0.73
Weighted avg	0.70	0.82	0.80	0.90	0.91	0.87	0.92	0.92	0.90	0.91	0.91	0.88

The imbalanced nature of the dataset can be seen in the estimation performance of the fan use. The class Medium scored lower than the classes “High” and “low”. Indeed, this class has the lowest amount of data, thus penalizing the training phase. As a result, the overall weighted average F1 score depends significantly on the score obtained for the class “low” since it is the most populated in terms of number of data. However, the most energy intensive class is the class “High”, which had relatively good F1 scores, up to 0.9 even for the BN, making all the tested models suitable.