Algorithm to simulate occupant behavior in mixed-mode office buildings

Zur Erlangung des akademischen Grades einer

Doktor-Ingenieurin (Dr. –Ing.)

von der KIT-Fakultät für Architektur
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Camila Chagas Anchieta Grassi

Tag der mündlichen Prüfung:	28.04.2021
Erster Gutachter:	Prof. Dr. Marcel Schweiker
Zweiter Gutachter:	Prof. Dr. Karin Chvatal

Dedicated to my very present and loving family; my husband, my parents, my brother and my grandmas.

Acknowledgments

First and foremost I thank God for the opportunities given to me, for allowing me to come this far and for giving me the ability to do all the work.

To my husband for always being by my side and always so understanding. For supporting me in every decision and helping me in every possible way. For always reminding me that everything will be ok. Stay calm.

To my family, my parents for always cheering me on, being excited for my accomplishments (big or small), so present and loving. Always believing in me and telling me I can do anything and everything, even when I doubt myself. And to my brother, who makes things lighter, makes fun of me to make it all easier, and helps me whenever I need.

A very special thank you to dear Hartmut and Regina, for being the most wonderful and loving people, taking me in, giving me a lovely home and treating me like family. You certainly made my days in Germany brighter, more cheerful and full of energy! I was never alone and always happy.

Thank you to my supervisors, Karin Chvatal and Marcel Schweiker, for all the support and input throughout this work, and for making the cotutelle possible. Thank you to Professor Andreas Wagner, who also made this possible accepting me as a member of the Building Science groups (Fbta) and assisting during the the entire process. And thank you to everyone in the group who welcomed me and where always there to help.

I would also like to thank everyone who was somehow involved in the monitoring campaign; the offices who opened their doors and let us take measurements, those who helped with the measurements, loaning equipment, data treatment and so much more. It was a team effort, every step of the way I needed partners, and I was never alone. I am deeply grateful for that.

And lastly, I would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and the Brazilian National Council For Scientific and Technological Development (CNPq) for their financial support, these are highly appreciated.

Declaration

I certify that the work in this thesis entitled “Algorithm to simulate occupant behavior in mixed-mode office buildings" has not previously been submitted for a degree in this or any other university. I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis and the regulations of the Karlsruhe Institute of Technology (KIT) about good scientific practices have been obeyed.

January 2021.

Camila Chagas Anchieta Grassi

Kurzfassung

Die Art und Weise, wie Gebäude betrieben werden, hat großen Einfluss auf den Energieverbrauch, und die Bewohner spielen dabei eine wichtige Rolle. Die Vorhersage der Anwesenheit und des Verhaltens der Bewohner ist jedoch eine komplexe Aufgabe, die es auch schwierig macht, den Energieverbrauch eines Gebäudes genau vorherzusagen. Infolgedessen haben Studien eine Lücke zwischen gemessenen und simulierten Daten festgestellt, die zum großen Teil darauf zurückzuführen ist, wie das Verhalten der Insassen in die Energiesimulationssoftware eingegeben wird. Um einen Beitrag zur Überbrückung dieser Lücke zu leisten, wird diese Arbeit vorgestellt, mit dem Ziel, einen Algorithmus zur Vorhersage des Insassenverhaltens in Bezug auf das Öffnen / Schließen von Fenstern und die Aktivierung / Deaktivierung von Wechselstrom in Bürogebäuden im gemischten Modus in großer Höhe zu erstellen tropisches Klima, insbesondere die Stadt São Carlos, SP, Brasilien.

Die generierten prädiktiven Verhaltensmodelle basierten auf Daten, die in einer 18-monatigen Überwachungskampagne gesammelt wurden. Umweltfaktoren wie Innen- und Außentemperaturen sowie die relative Luftfeuchtigkeit im Innen- und Außenbereich sowie die Maßnahmen im Zusammenhang mit den untersuchten Fenster- und Klimaanlagensteuerungen wurden gemessen. Zwei Arten statistischer Methoden wurden auf den Datensatz angewendet, verallgemeinerte lineare gemischte Modelle und Markov-Kette. Die resultierenden Modelle wurden in EnergyPlus implementiert, um ihre Genauigkeit bei der Darstellung des Insassenverhaltens in Simulationen zu bewerten. Zu Vergleichszwecken wurde auch ein Szenario mit festen Zeitplänen ausgeführt, das üblicherweise in gemischten Gebäudesimulationen angewendet wird.

Die Modellergebnisse zeigten hohe Wahrscheinlichkeiten für die Verwendung beider Kontrollen bei Ankunft und Abreise, insbesondere für Fenster. Die Innentemperatur um 20 °C steht im Zusammenhang mit geringen Verwendungswahrscheinlichkeiten für beide Kontrollen, während bei Innentemperaturen über 24 °C die Wahrscheinlichkeit der Verwendung von Wechselstrom zunimmt und das Öffnen des Fensters abnimmt. In Bezug auf die Außentemperatur sagen die Modelle eine höhere Verwendung von Fenstern mit den niedrigsten Außentemperaturwerten und eine höhere Wahrscheinlichkeit einer AC-Aktivierung mit höheren Werten voraus.

Bei der Implementierung in EnergyPlus zeigten die Simulationsergebnisse, dass das Modell, das das überwachte Insassenverhalten am besten repräsentierte, das Markov-Kettenmodell war, das mit synthetischen Daten durch zufällige Überstichprobenbeispiele (ROSE) generiert wurde, was auch zeigt, dass die häufig verwendeten festen Zeitpläne das Insassenverhalten nicht korrekt darstellen in gemischten Gebäuden. Die gemessenen Daten sowie das Modell mit der besten Leistung zeigten, dass die Insassen in allen Temperaturbereichen zwischen beiden Kontrollen wechseln und sich nur in der Häufigkeit der Verwendung jeder Kontrolle in jedem Temperaturbereich unterscheiden. Daher wird ein solches Verhalten mit festen Zeitplänen, die die Verwendung von Kontrollen basierend auf der Temperatur einschränken, nicht korrekt dargestellt.

Die verschiedenen angewandten statistischen Methoden erzeugten Modelle mit unterschiedlichen Anwendungen. Die verallgemeinerten linearen gemischten Modelle können zur Überprüfung der Verwendung von Steuerelementen in einem Büro verwendet werden, da deren Einschränkungen bekannt sind und bei der Analyse der bereitgestellten Ergebnisse berücksichtigt werden. Die Markov-Kettenmodelle eignen sich besser für die Implementierung in Computersimulationsprogrammen, da sie den Wechsel des gemessenen Insassenverhaltens in einem Szenario im gemischten Modus erfassen können. Daher können die in dieser Studie generierten Markov-Kettenmodelle in ähnlichen Büros im gemischten Modus innerhalb desselben Klimas angewendet werden. Es gibt jedoch immer noch Einschränkungen bei den Modellen, und es ist wichtig, umfangreichere Messkampagnen durchzuführen, damit Daten zu mehr Arten von Insassen

und Klima gesammelt werden können, sodass umfassendere Modelle erstellt werden können.

Abstract

The way buildings are operated greatly influence energy consumption, and occupants play a significant role in it. However, the prediction of occupants' presence and behavior is a complex task, making it also challenging to accurately predict a building's energy consumption. As a consequence, studies have identified a gap between measured and simulated data, in great part attributed to how occupant behavior is inputted into energy simulation software. In an effort to contribute to bridge this gap, this work is presented, with the objective of creating an algorithm to predict occupant behavior, in relation to window opening/closing and AC activation/deactivation, in mixed-mode office buildings in a high altitude tropical climate, specifically, the city of São Carlos, SP, Brazil.

The predictive behavioral models generated were based on data collected in an 18-month monitoring campaign. Environmental factors, such as indoor and outdoor temperatures, and indoor and outdoor relative humidity were measured, as well as the actions related to window and air conditioning controls being studied. Two types of statistical methods were applied to the data set, generalized linear mixed models and Markov chain. The resulting models were implemented in EnergyPlus to assess their accuracy in depicting occupant behavior in simulations. A fixed schedules scenario, commonly applied in mixed-mode buildings simulations, was also run for comparison purposes.

Models results showed high probabilities of use of both controls upon arrival and departure, especially for windows. Indoor temperature around 20 ^{o}C is related to low probabilities of use for both controls, whereas with indoor temperatures above 24 ^{o}C the probabilities of AC use increase, and of window opening it begins to decrease. As for outdoor temperature, the models predict higher use of windows with the lowest outdoor temperature values, and higher probability of AC activation with higher values.

When implemented in EnergyPlus, simulation results showed that the model that best represented the monitored occupant behavior was the Markov chain model generated with Synthetic Data By Randomly Over Sampling Examples (ROSE), also evidencing that the commonly used fixed schedules do not correctly represent occupant behavior in mixed-mode buildings. Measured data, as well as the best performing model, showed that occupants alternate between both controls in all temperatures' ranges, varying only in the frequency of use of each control within each temperature range. Therefore, such behavior is not correctly depicted with fixed schedules that limit the use of controls based on temperature.

The different statistical methods applied generated models that have different applications. The generalized linear mixed models can be used for verification of the use of controls in an office, given its limitations are known and accounted for when analyzing the results provided. The Markov chain models are more adequate to be implemented in computer simulation programs, as it is able to capture the alternating nature of the measured occupant behavior within a mixed-mode scenario. Therefore, the Markov chain models generated in this study can be applied in similar mixed-mode offices within the same climate. However, there are still limitations to the models, and it is important to conduct more extensive measuring campaigns, so data on more types of occupants and climates can be collected, allowing more comprehensive models to be generated.

Resumo

A maneira como os edifícios são operados tem grande influência no consumo energético, e os ocupantes têm um papel importante. No entanto, o comportamento do usuário é de difícil predição, o também torna difícil a predição do consumo energético de edifícios. Consequentemente, estudos têm identificado uma discrepância entre os dados medidos e simulados, em grande parte atribuída a maneira como o comportamento do usuário é inserido em programas de simulação energética. Para contribuir para a diminuição desta discrepância, este trabalho é apresentado, com o objetivo de criar um algoritmo para predição do comportamento do usuário, em relação a operação de janelas e do ar-condicionado em escritórios de modo-misto em um clima tropical de altitude, especificamente na cidade de São Carlos, SP, Brasil.

Os modelos preditivos gerados foram baseados em coleta de dados de uma campanha de monitoramento com duração de 18 meses. Fatores ambientais, como temperaturas internas e externas, e umidade relativa interna e externa, foram medidas, assim como as ações relacionadas às janelas e ao ar-condicionado, chamados também de controles.

Dois métodos estatísticos foram aplicados aos dados coletados, modelos mistos lineares generalizados e cadeia markoviana. Os modelos resultantes foram implementados no EnergyPlus para avaliar a precisão dos mesmos em reproduzir o comportamento do usuário em simulações. Um cenário com os dados fixos sobre o comportamento do usuário, ou schedules fixos, comumente utilizados em simulações de edifícios de modo-misto, também foi simulado para propósito de comparação.

Os resultados dos modelos mostraram altas probabilidades de uso de ambos os controles no momento da chegada e saída, especialmente as janelas. Com a temperatura interna aproximadamente a 20 °C, observou-se baixas probabilidades de uso de ambos os controles, ao passo que com temperaturas internas acima de 24 °C, as observou-se um aumento na probabilidades de uso do ar-condicionado e uma diminuição do uso das janelas. Quanto à temperatura externa, os modelos predizem um maior uso de janelas com temperaturas externas mais baixas, e maior probabilidade de uso do ar-condicionado com valores mais altos.

Quando os modelos foram implementados no EnergyPlus, os resultados mostraram que o modelo que melhor representou o comportamento do usuário monitorado foi o modelo gerado com cadeia markoviana utilizando dados sintéticos com amostragem aleatória (Synthetic Data By Randomly Over Sampling Examples (ROSE)), também evidenciando que schedule fixos comumente utilizados não representam corretamente o comportamento do usuário em edifícios de modo-misto. Os dados medidos, assim como o modelo com melhores resultados, mostraram que os usuários alternam entre o uso dos controles em todas as faixas de temperatura, variando apenas a frequência com que operam cada controle em cada faixa de temperatura. Assim, tal comportamento não é retratado com schedules fixos que limitam o uso dos controles com base nas temperaturas.

Os diferentes modelos estatísticos aplicados geraram modelos que apresentam aplicações diferentes. Os modelos mistos lineares generalizados têm uma melhor aplicação nos estágios iniciais de projeto, para uma verificação rápida do uso dos controles em um escritório, dado que suas limitações são sabidas e levadas em consideração quando analisando os resultados. Os modelos de cadeia markoviana são mais adequados para serem implementados em programas de simulação computacional, pois são capazes de capturar a alternação entre os controles conforme observado no comportamento medido em ambientes de modo-misto.

Assim, os modelos de cadeia markoviana gerados neste estudo podem ser aplicados em escritórios de modo-misto similares inseridos em um mesmo clima. No entanto, ainda há limitações aos modelos, e é importante conduzir campanhas de medição mais extensas, para que dados sobre mais tipos de usuários possam ser coletados, permitindo que modelos mais completos sejam gerados.

Contents

Acronyms and symbols

Acronyms

ABNT Associação Brasileira de Normas Técnicas (Brazilian Standard)

ASHRAE American Society of Heating, Refrigerating and Air-Conditioning Engineers

MMV Mixed-Mode Ventilation

AIC Akaike Information Criterion

HVAC Heating Ventilation and Air Conditioning

NV Natural Ventilation

RH Relative Humidity

GLMM Generalized Linear Mixed Model

MC Markov Chain

WO Window Opening

ACA Air Conditioning Activation

TPR True Positive Rate

FPR False Positive Rate

TNR True Negative Rate

FNR False Negative Rate

ROC Receiver Operating Characteristic

AUROC Area Under the Receiver Operating Characteristic

ROSE Randomly Over Sampling Examples

WWR Window to Wall Ratio

EMS Energy Management System

SD Standard Deviation

INMET Instituto Nacional de Meteorologia (Meteorology Institute)

PMV Predicted Mean Vote

IAQ Indoor Air Quality

DOE Department of Energy

Greek symbols and variables

λ lambda

β coefficient

α absorptance

1 Introduction

Over the past 15 years there has been an increase in energy use and also in the studies analyzing energy use in buildings. The need for more sustainable solutions has become a global concern and the analysis of energy use has been one of the focuses of such discussions [42, 80, 144]. An increase in energy consumption has occurred due to the constant technological development and shown itself in alterations in the use pattern in a global context. Such consumption has presented an impact on the urban and global contexts, resulting in a raise in green house gas emissions (GHG), global warming and a decrease in natural resources [60]. As foreseen, one of the consequences are climate changes, which have become a threat. Several international agreements, such as the Kyoto Protocol have taken place in an effort to remedy the situation [37]. However, climate changes have already occurred, resulting in higher temperatures, leading to an intensified use of air conditioning in buildings [15]. The building sector is continuously growing, consequently increasing the levels of energy globally consumed [66]. Within this sector, office and retail buildings, composing the commercial sector, display the highest electricity consumption and CO_2 emissions [100]. In Brazil, office buildings contribute to 17.4% of the energy consumed [14], with similar consumption rates seen in the UK and the US in previous years, with 17% and 18%, respectively [13]. Nevertheless, this same sector presents great potential to reduce its contribution climate changes, since it presents great opportunities of energy savings [11].

It is agreed around the world that air-conditioning in buildings represent a significant contribution to energy consumption. In Brazil, the use of air-conditioning constitutes approximately 50% of office buildings' electric energy consumption [36], and the way air conditioning is operated in buildings can have high impacts on their energy consumption, given that the same building can present very divergent energy uses depending on how it is operated by occupants [31,146]. However, there are strategies that can be implemented to decrease energy consumption, one of which is mixed-mode ventilation (MMV). MMV is a combination of natural ventilation and a mechanical ventilation system, which can be activated when solely using natural ventilation is not enough to maintain the environment comfortable. In this system, natural ventilation is used when the external conditions are favorable, thus activating the mechanical system when the opposite situation occurs [38].

The mixed-mode, or hybrid ventilation strategy, shows great potential to motivate a greater use of natural ventilation over air-conditioning with the goal of reducing energy consumption. However, even though passive strategies, such as natural ventilation (NV) reduce the needs for energy use, they increase the levels of uncertainty by reinforcing the central role of occupants, as they become the essential elements in control of the building and its environment [52]. With the growing public concern regarding climate change, more attention has been drawn towards energy consumption in buildings [33, 69], and predicting energy demand has gained significant relevance in designing and constructing buildings, from the early design stages to post occupancy [37]. Such is the need for the decrease of energy consumption that regulatory conditions have been established for all European countries, in an effort to decrease the energy required for heating, cooling, ventilation and lighting [37].

One of the ways to verify the impact of different design solutions and buildings' properties in energy use is to conduct simulations. There are several different simulation techniques, and their application in building performance simulation and building environmental design have been increasing [135]. They can also be used as a tool to comply to the established norms and standards, as a way to verify if adequate consumption levels are being reached. Computer simulation is an important and useful tool to predict energy consumption in buildings based on design information. However, several studies [24, 46, 118, 130, 135, 146] have shown that there is a discrepancy between the real and predicted data regarding energy consumption in buildings. Studies have demonstrated that the real energy consumed in buildings can be up to three times greater than the predicted values [24, 46]. This suggests a performance gap, which can be attributed to, among other factors, such as weather files [43] or building construction materials [25], a disregard of occupant behavior in the simulation process. [85] conducted five different studies to show that buildings do not perform as predicted even when using very accurate simulations, and concluded that occupant behavior and preferences are significant contributors to the identified gap. [48] estimated the influence of occupant behavior in heating and cooling energy in a simulation run, showing a high variability in energy demand resulting from the effect of occupant behavior within the studied building. [118] identified occupant behavior as possibly the most overlooked parameter, and that it might not be properly considered as an integral part of the energy design, thus resulting in discrepancies in the data.

Occupants adapt within their environment in order to achieve comfort in ways that are convenient to them, and not necessarily energy conserving [93, 105]. They may act in unexpected ways to respond to a crisis of discomfort, and such actions are contrary to the static assumptions designers tend to make when portraying such parameter in simulations. Therefore, occupant behavior can be considered one of the variables that generates great amounts of uncertainties in simulation results [26]. Occupant behavior is a variable known to have a significant impact on how buildings are operated, thus influencing their energy consumption [56, 95, 101, 132, 136, 139, 141, 147]. Occupants can affect the indoor environment, and as a result, energy consumption as well, depending on how they behave in the environment and on how they interact with the building and the controls available to them, such as windows, ventilation systems and shading devices [4]. Energy savings from occupant behavior presents great potential, and can constitute from 5% to 30% of savings in commercial buildings [61]. Nonetheless, even though it is a variable with great savings potential, it is also one with many uncertainties, due to occupants' personal experiences and preferences. Occupants' behaviors and feelings are complex and very difficult to predict [134], therefore making it a challenging task to accurately predict a building's performance [59].

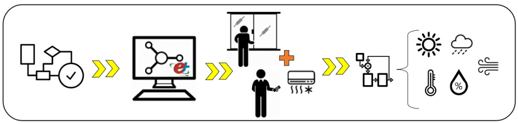
Several studies have addressed the issue of occupant behavior within different types of buildings and with emphasis on different influential parameters [23, 50, 102, 108, 118, 131].

However, even with efforts being made to diminish the gap between predicted and actual data, and to better understand occupant's actions, there is still a need to develop research studies to contribute to a wider range of knowledge in the field [27]. Not having enough models to depict a variety of behaviors in different buildings types and climates, such as mixed-mode buildings in a high-altitude tropical climate, presents a difficulty to computer simulation users, who have little or no information on how to model such behavior to acquire precise results.

In addition, there is no guide on how to simulate or even design mixed-mode buildings specifically [116], making it difficult to simulate this scenario in computational programs, which can often lead to non reliable results. Even though the number of researches on the topic has been growing over the past years, one of the gaps that remain is related to the use of the controls when operated by occupants and not a system [36]. This is specially relevant to the Brazilian reality, where, as stated by [36], "a significant portion of the older office building stock has had air conditioning retrofitted into a basic, naturally ventilated design, rendering it mixed-mode ventilation". The authors also state that few field research studies have been conducted in these environments, and that in Brazil, buildings rely on artificial mechanisms for cooling all year long, dispensing the need for heating.

Considering this context and the identified gaps, this research proposes the study of occupant behavior in mixed-mode office buildings related to the opening/closing of windows and the activation/deactivation of the air conditioning in the city of São Carlos, SP, Brazil. The study focuses on buildings displaying operable windows and individual air conditioning (AC) units in the offices, all of which the occupants are free to operate at any given time. These buildings do not have automation systems, nor any temperature indication of any kind to inform occupants of the indoor environment conditions, except when selecting the AC set point. This work intended to create an algorithm to be implemented in computer simulation programs, to provide more accurate input data for simulations of mixed-mode office buildings related to energy use, once there is a discrepancy between the real and simulated data referring to occupant behavior [94]. To this date, there are no other works presenting behavioral models to be implemented in building performance simulation programs in Brazil, especially for the climate and type of building in question. [27] identified that there are several contributions in the field from North America, Europe, and China, and well established collaborations between groups from these locations. However, the authors also identify that there are very few contributions from Latin America, specifically from Brazil, highlighting the relevance of such studies.

1.1 Objectives



The main objective of this work was to create an algorithm to predict window opening/closing and air conditioning activation/deactivation in mixed-mode office buildings specific to a high-altitude tropical climate, and to implement it in simulation programs to provide more precise input data on occupant behavior.

The specific objectives to be achieved as the research progressed are presented with respective research questions specific to each objective:

- To identify the main driving factors that lead to window opening and AC activation based on the analysis of the measured data.
    - Is there a seasonal effect to the way occupants operate the studied controls?

  - What indoor and outdoor temperature values show higher frequency of window opening?
  - What indoor and outdoor temperature values show higher frequency of AC activation?
  - Is there a difference in the use of the studied controls as the day progresses?
  - Does the building envelope and office layout influence the way occupants operate the studied controls?

- To apply statistical methods to develop behavioral models that represent the measured data well.
  - What statistical methods can be applied to the type of data collected?
  - What are the limitations and advantages of the applied statistical methods?
  - Are the models able to capture the measured behavior, thus accurately predicting the most likely action to be taken given the environmental variables being considered?

- To implement and test the developed models in a computer simulation program and assess their representation of the measured data.
  - Are behavioral models more representative of occupant behavior than the deterministic or fixed schedules commonly used in building performance simulations?
  - Do the developed behavioral models represent the measured data well in computer simulations?
  - Is it possible to use the developed behavioral models to predict energy use more accurately?

– Is it possible to use the developed behavioral models to simulate different types of occupants?

1.2 Thesis Structure

This thesis was structured in order to achieve the proposed objectives as described in the previous section. Figure 1.1 illustrates the general organization of the work.

Chapter 2 Literature

Presents the current knowledge of the main themes addressed in this work, thus divided into four main parts, namely; occupant behavior, mixed-mode ventilation, computer simulation and statistical methods. This chapter also counts with a final section, where the above-mentioned themes are combined and presented as relevant studies within the same field.

Chapter 3 Method

This chapter describes the method developed and applied to achieve the proposed objectives. The *Initial Stages* section describes the development of the theoretical model, followed by the definition of the unit of analysis, population and scope. The *Model Creation* section is subdivided into four main parts, describing the monitoring campaign and model creation, with the application of statistical methods. Furthermore, it describes the validation procedure and the simulation tests performed.

Chapter 4 Results and Chapter 5 Discussion

Chapter 4 presents the results and validation for all the models created, as well as the results of all models' combinations simulated in EnergyPlus. Chapter 5 presents a discussion of the models and simulation tests results based on the literature.

Chapter 6 Conclusions

Lastly, chapter 6 presents the findings in a summarized manner, discussing the limitations of the work and proposing further studies that can improve the work presented here.

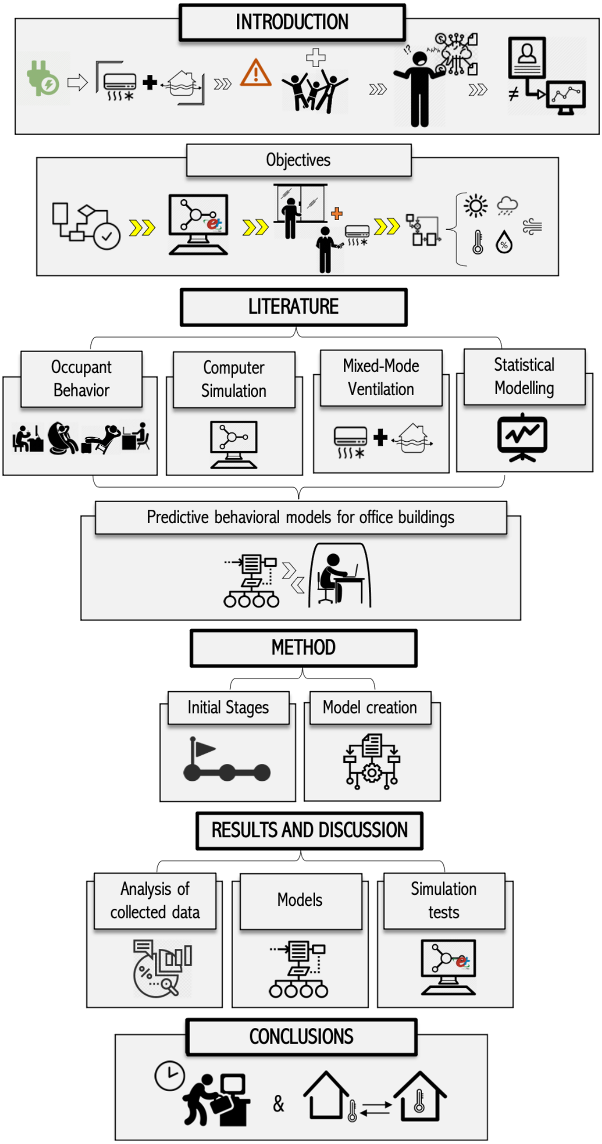


Figure 1.1: Thesis Structure

2 Literature

This chapter presents the main themes studied in this research, such as occupant behavior, energy use, mixed-mode ventilation, computer simulation and statistical methods applied to data sets to create stochastic models portraying occupant behavior. This chapter also counts with a section that describes studies developed applying similar methodologies.

2.1 Occupant Behavior

In a general sense, occupant behavior can be defined as the interactions with a building's systems in an effort to control the indoor environment for indoor air quality, health and comfort, be it thermal, visual and/or acoustic [37]. Occupant behavior is among the six most influencing factors of a building's performance, along with climate, building envelope, equipment operation and maintenance, and indoor environment conditions [149]. There are prerequisites, established by buildings' occupants, so they are able to adjust and adapt systems and components according to their preferences. Such requisites involve improving air quality, by ventilating and eliminating odor and pollution, visual or lighting quality, by controlling glare, reflections and the amount of illuminance, acoustical conditions, by avoiding noise, and aesthetics, as well as improving thermal comfort in the indoor environment [16].

As stated by [93], "if there is a change that causes discomfort, people will act in ways to regain their comfort". The term ‘alliesthesia’ is the combination of the words ‘changed’ (allios) and ‘sensation’ (aisthesis) [22], and can be used to describe that an external stimulus can be understood as pleasant or unpleasant, depending on the signals sent from the body. It is human nature to search for pleasant conditions and avoid unpleasant ones. However, because of people’s different backgrounds and preferences, also their physical, physiological and psychological differences, and several other influential external drivers such as economic and regulatory, for example, they do not perceive and respond in the same manner [16, 119].

Occupants play a very significant role in a building’s energy performance, as they are present, move around and interact with the building and its systems to better fit their purposes and comfort needs, resulting on an impact in energy consumption [63]. As there is a growing concern for sustainability and low energy buildings using passive strategies, such as natural ventilation, the role of occupants in a building is reinforced, as they become the most important elements controlling their environment [52]. There is great diversity in occupants’ actions, and the main objective in studying occupant behavior is to determine the triggers of their actions [62].

Occupants present the individual aspect that is related to personal experiences, preferences and expectations, altering the way each person uses the available controls, which affects the total energy consumption in a building. Their active use of energy, as in the way they interact with control systems and their available building elements to reach their desired levels of thermal comfort, have a significant impact on the total amount of energy consumed [37]. Such interactions happen in different ways; window opening and closing, lighting, shading devices, HVAC (Heating, Ventilation and Air-Conditioning) systems, hot water and appliances (Figure 2.1).

[62] identified actions and inactions that occupants may take to regain thermal comfort. Actions can be adjusting the level of clothing, drinking a hot or cold beverage, opening a window or adjusting the thermostat. As for inactions, it could

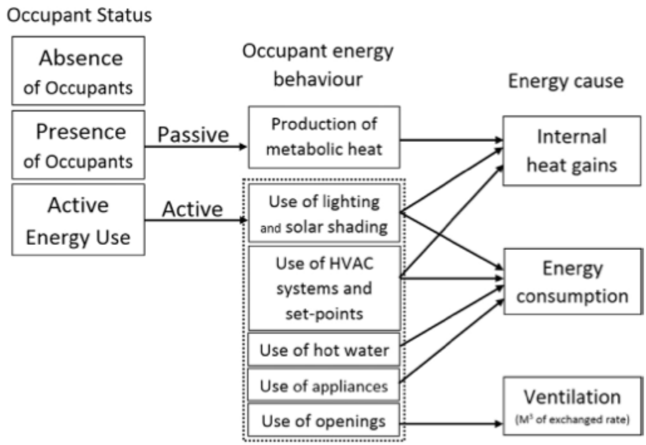


Figure 2.1: Occupants' types of activities affecting building energy consumption [37]

be, for example, moving to a different location and tolerating some degree of discomfort. Different actions and inactions can be taken by different occupants in response to the same kind of situations, thus impacting in very different ways how energy is consumed in the same environment. Therefore, it is critical to understand the relationship between the building and its occupants' activities, lifestyle and behavior.

In an effort to better portray occupants and their impact on the built environment, researchers have categorized occupants into different groups according to their energy use. [39] created the groups active, medium and passive, regarding occupants and their energy use. The active occupants are the ones that change the heating/cooling set point, whereas the passive occupant does nothing, related to operating available controls, and continues to tolerate some degree of discomfort. Other categorizations, such as the one made by [62], described with more accuracy people's actions, classifying them as; "energy frugal", "energy indifferent" and "energy profligate". Using another method, which classified behavioral factors in residential buildings, [32] categorized occupants into three levels referring to their complexity, that is; simple, intermediate and complex. Each level was developed for a different application; the simple level was for statistical analysis, the intermediate, with more parameters, for case studies, and the complex one was meant for detailed simulations. This division, using occupants' behavior to create profiles, could contribute to more accurate assumptions when performing energy analysis. Nonetheless, there is still a need to perform large scale field studies to gather comprehensive data to create such profiles.

Occupant behavior is a contributing variable to the uncertainty of building performance, and it can significantly affect building energy consumption. According to [59], it is the leading source of uncertainty in predicting energy use in buildings.

The International Energy Agency (IEA), lists occupant behavior, among other factors such as climate and building envelope, as one of the driving forces of energy use in buildings [65]. [120] defines occupant behavior as "a human being's unconscious and conscious actions to control the physical parameters of the surrounding built environment based on the comparison of the perceived environment to the sum of past experiences." Some of these actions can be interactions with windows, lights, blinds, thermostats, air conditioning and plug-in appliances.

Based on the definition above, it is possible to state that occupant behavior is difficult to predict because it can be influenced by a number of factors, be them external to the occupant, such as air temperature and wind speed, or internal or individual, such as personal background and preferences, as well as building properties, which can be perceived as ownership or the availability of heating devices, for example [45].

Traditionally, occupant behavior has been connected to indoor and outdoor thermal conditions, leading to interactions with building control systems, which are only one aspect of occupant behavior. Human behavior, when in the field of social sciences, can be set in relation to causes that are called "internal or individual factors", which in addition to external factors, influence occupant behavior with a variety of perceptions and actions in complex ways [120].

2.1.1 Parameters Influencing Occupant Behavior

Energy consumption is largely affected by occupant behavior, as they respond differently to regain their levels of thermal comfort, which can vary according to their personal (psychological, physiological) and social characteristics [37]. In addition, parameters such as climate, economy, regulations and policies, architecture and interior design of the spaces can also influence energy use (Figure 2.2).

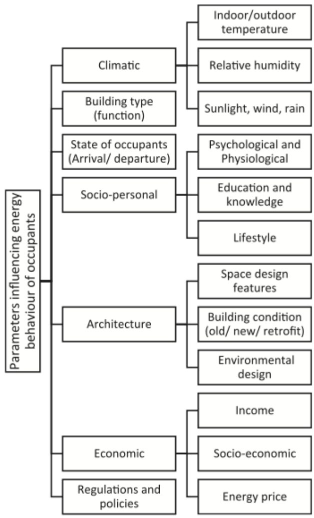


Figure 2.2: Factors influencing occupants' behavior regarding energy use [37]

Climatic parameters, including outdoor temperature, relative humidity, solar radiation, wind and rain are very significant influences on occupant behavior and their interaction with a building's systems to achieve thermal comfort [37]. Several researchers have studied the influence of climatic parameters on occupant behavior [79, 107, 118], and because such parameters are time/date dependent, stochastic models are the most common method in these studies, estimating the probability of given outcomes.

In order to attain more accurate results from building performance simulation programs, it is important to consider and understand the reasons that lead occupants to take actions and interact with the building and its systems. Factors that influence occupant behavior, either external or individual, have been categorized and denominated as “drivers”, which are described as the reasons that lead the occupant to react in a building and suggests that such occupant acts, thus driving the occupant to take an action [45]. Occupants needs and preferences drive them to react and interact with their environment, thus the available controls, to satisfy their needs [77].

Several researchers have divided driving factors into categories, [145] and [71] considered the existence of two groups, internal and external factors, as potential triggers to actions. [99], when conducting a residential study, classified behaviors in three main categories:

- Environmentally related: actions triggered by environmental factors
- Time related: actions repeated within certain time frames
- Random: actions taken depending on uncertain/non quantifiable factors.

The main three categories defined by [99] can be applied to other types of building, such as an office building, where environmental factors will be an important influence, as will time related actions, which can be related to routine. Random related actions can also be observed in both types of building. Depending on the action under consideration, some drivers can have a greater impact on triggering an action than others.

[45] classified drivers into five groups, namely: physical environmental factors, psychological factors, physiological factors, social factors and contextual factors. The latter was included when the authors made a review on window opening, showing how one action can indicate to different drivers. Their descriptions are as follows:

a) Physical environmental: environmental aspects that lead occupants to act and that have an impact in energy consumption: temperature, humidity, air velocity, noise, illumination and odor.

b) Contextual: factors that do not influence in a direct way, but are determined by the context, such as building insulation, facade orientation and heating system type, for example.

c) Psychological: occupants seek to satisfy their needs regarding thermal comfort, acoustic comfort, health, among other factors. They also have expectations for their environment, such as temperature and indoor environmental quality.

d) Physiological: some factors, such as age, gender, health condition, clothing and activity level are physiological driving forces that can determine the physiological condition of occupants.

e) Social: social driving forces are related to the interaction between the occupants.

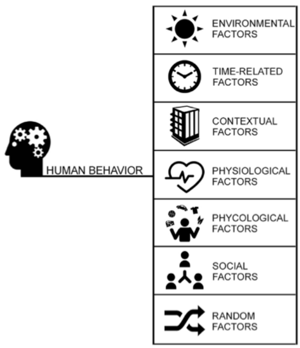


Figure 2.3: Categories of the factors that influence occupant behavior in building [130]

All operations taken to improve or maintain adequate thermal comfort levels have an impact and a consequence on the indoor environment, and consequently in energy consumption. By taking actions, the occupant becomes the central operator who controls the energy consumption and indoor environmental quality in a building [45]. Figure 2.3 presents a combination of the above-mentioned drivers' categories that can influence occupant behavior. Figure 2.4 shows sub-groups, as defined by [121], displaying an overview of the triggers and contextual factors that have been studied and discussed in the literature as having an influence on occupant behavior. The sub-groups were based on the five above-mentioned categories as defined by [45].

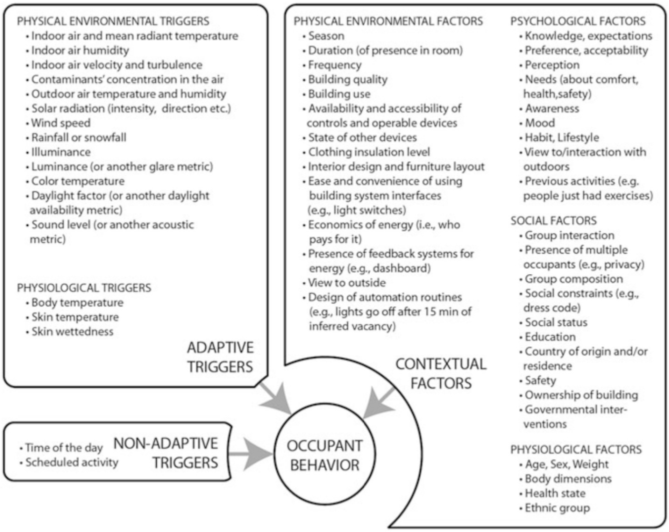


Figure 2.4: Potential influencing factors of occupant behavior [122]

The authors [121] describe the physical environmental triggers as the properties that describe the indoor and outdoor environments. Such properties cause a response from the body, thus triggering occupants to take an action. As for the contextual factors, they are divided into physical environmental factors, psychological factors, related to social factors; and physiological factors, which do trigger a response, just not an immediate one [94].

Another influential parameter is the building type and the types of activities performed by the occupants within the built space. The building type usually determines the activities performed in it, which in turn, sets a clothing type, metabolic rates and the occupants' specific needs and expectations related to such activities, as well as the way they interact with the building.

Social and personal parameters are also critical when investigating occupants' comfort and energy-related attitude. Social and personal factors have been identified as also being influential on energy behavior in residencies, such as occupants' awareness of energy issues, gender, age, employment, family size and social-cultural belonging [85]. In an office environment, social parameters can also be seen as the number of occupants sharing the space. [125] state that perceived control decreases with higher numbers of occupants in an office, which in turn affects how occupants operate the available controls. [69] highlights the effect of education and growth in awareness in people's attitude towards energy use. Other relevant and influencing factors, such as energy regulations, policies and economical parameters, as energy price and employment, have been discussed in the literature. As part of such discussion, several studies have investigated the influence of the above-mentioned parameters on occupants' energy consumption behavior [24, 85, 110]. Specifically related to energy costs, studies conducted in residential buildings identified that this variable affects the way energy is consumed. [97] identified, by means of questionnaires, that occupants tolerated a certain amount of discomfort on account of energy prices. [128] found that, due to the implications on their energy bills, occupants of a low to middle income housing development in South Australia ranked air conditioning as their least preferred strategy.

The state of occupants in office buildings, that is, their arrival, presence and departure, has been studied and revealed that occupants tend to adjust given building systems upon arrival more than at departure [58, 152]. This parameter has been considered and modeled in several studies to investigate the connection between occupants' movements and their behavior [95, 152]. Architecture and interior design can also influence occupants in the way that the space may change their perception [35]. However, the impact of interior design has not been broadly studied [37, 82].

2.2 Building performance simulation and occupant behavior

Computer simulation increasingly became more affordable and possible, both to researches and designers, due to the fast progress seen in the computer industry and in computational techniques [142]. During the 1990's, there was a considerable growth in the use of computer simulation tools, due to a greater availability both of hardware and software. Personal computers became more affordable in industry and research, leading to a progress in computer simulation in the field of techniques and tools. Today, as stated by [142], building simulation is involved in several stages, such as design, engineering, operation and management of buildings, becoming an integral part of building design and industry. Such tools and techniques are believed to accelerate the design process and optimize building performance at what can be considered a low cost [8].

There are several advantages to the use of simulation tools, with the growing need for energy savings and reduction of environmental impacts, simulation based design is an important tool to achieve such targets. According to [142], the main goal for developing building simulation techniques and tools is to assist in the creation of a built environment that meets all the existing needs and criteria and presents the least cost with construction and operation, as well as a low amount of resource consumption. One of the greatest advantages to simulation tools is that they are able to provide quantitative data and thus aid in the decision making process. One of the main reasons why such tools are being promoted and enforced is the increasing number of building performance codes and standards. Therefore, the adoption of simulation tools has been enforced to evaluate a building's performance compliance. As exemplified by [142], ASHRAE 90.1 [7] requires that the whole year building energy simulation results are presented in order to rate the building's energy performance. The Green Building Rating System adopts the same policy. Various governments also demand, and with that accelerate, the use of simulation tools for building design to achieve low energy buildings and to meet their energy and greenhouse gas (GHG) reduction targets [142].

Building energy simulation programs are based on a building's basic physical parameters, such as heat transfer, occupant density and operation schedules. They are very precise in simulating deterministic factors influencing buildings' energy balance. However, when representing non-deterministic variables, their accuracy to faithfully represent reality diminishes. Occupant behavior is by nature stochastic, and many times remains neglected when represented in simulation programs because it is accounted for as a fixed presence, such as schedules or deterministic interaction strategies [52], as is the traditional method [7]. [146] state that the adoption of standard schedules to portray occupant behavior is an incorrect way to represent the dynamic human-building interaction. [34] state that using "average behavior can be a major cause of the gap between the actual and predicted energy use of the building". Specifically to the operation of windows, studies have found that it largely differs from the fixed-schedule scenario, resulting in a discrepancy between building design and actual performance of up to 10 times when looking at energy consumption [63].

This is mainly the reason why energy simulation of buildings offering adaptive opportunities to their occupants present such discrepancies between the simulated and real data. Buildings with the same physical features can show great differences in energy consumption, which can be related to occupancy patterns, occupants' lifestyle, comfort preferences and interactions with the buildings systems. However, most programs have little consideration of the impact of occupant behavior on energy use [79].

This situation is true for residential and office buildings. Several studies conducting field survey monitoring indicate a large difference between identical buildings, attributed to differences in occupant behavior. As stated by [52], it is necessary to reliably represent occupants within a building, so low-energy free running buildings can be correctly designed and thus avoid contradictions between occupants' freedom and sustainability.

2.2.1 Studies implementing behavioral models in building energy simulations

Stochastic models provide the opportunity to represent occupant behavior by considering the several factors that can influence occupants' actions [77]. Elaborating such models is key, and integrating building occupant behavior models and energy simulation programs is extremely helpful to quantify the influence of occupant behavior on a building's performance, as a means to increase building energy prediction accuracy [71]. This implementation is part of the solution to overcome the gap between measured and simulated data so often found [12, 44].

There are several studies that present the implementation of behavioral models in building simulation programs, identifying the different results achieved by using fixed, or deterministic schedules, and the models created based on field studies [39, 40, 59, 106]. [5] investigated the influence of occupant behavior in energy consumption by simulating a room occupied by one person who could operate six different controls. Two behavioral modes were simulated, an energy consuming and an energy efficient behavioral mode. In addition, a reference scenario where the occupant had no control was also simulated. Predicted Mean Vote (PMV) limits were established and simulated within each behavior mode. The authors reported that there was not a lot of difference on energy consumption between the different PMV limits within each behavior mode. However, they reported that the energy consuming behavior mode showed energy consumption of up to 330% higher than the energy efficient mode, even though PMV limits were close to neutral during most of the year in both modes. This study shows that, even though PMV limits did not change that much, there can be significant different energy consumption levels depending on how the available controls are operated, which depends on how the occupant behaves.

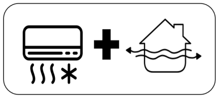
Specifically to window use, [105] used results from field surveys to create an adaptive algorithm to simulate window opening in office buildings. By means of logistic regression analysis, the authors identified that the proportion of open windows is highly correlated to indoor and outdoor conditions, also showing that the applied statistical method is appropriate to establish a relationship between the observed action and environmental factors. Lastly, the authors implemented the algorithm in a simulation software, concluding that using the adaptive algorithm is more suitable than the more common simulation methods, as it is more sensitive to changes in the building's design parameters. The authors also state that using an adaptive algorithm better represents how occupants use windows, thus allowing more accurate assessments of thermal comfort conditions and energy use.

Studies have explored several aspects of occupant behavior and actions within the built environment by means of simulation. [101] generated occupancy models based on two years of monitored data. The authors compared simulation results using two static standard models, one stochastic model and case specific data-driven models developed from the monitored data. Results demonstrated that standard occupancy models are not good representations of occupants attitudes of the energy peak demand, or of their seasonal behavior. The authors state that the standardized procedures do not take into account occupants' adaptability to the environment associated with human comfort perception. Even though the data-driven occupancy scenarios still show discrepancies, they are a better representation of occupants real daily attitudes.

Because occupant behavior models still show discrepancies and limitations, the work of [51] demonstrates how the different ways that models are built can result in different simulation results. Specifically for window opening, the models by [54, 55, 106, 152] were implemented in EnergyPlus. In general, results results indicated a window opening behavior following a seasonal trend, with a peak in the summer (July). The model by [152] predicted windows remaining open for 90% of the time in July, while this value was of 50% for the other models. The authors [51] state that these results are a reflection of the data with which the model was created. [152]'s model was based on data collected only during the summer, while the other models were based on summer and winter data collections. Nonetheless, simulations results of all four models presented an increase in heating load and a decrease in cooling load. The authors conclude that, even though there are differences between the results respective to each window operation model, the models consistently show a seasonal trend for window opening, as well as a decrease in cooling loads.

As exemplified by the above-mentioned studies, stochastic models are more representative of occupant behavior in buildings simulation programs than fixed schedules. Even though they may still present limitations, they provide more realistic information about occupants and the consequent energy consumption depending on how occupants operate the available controls. In this sense, it is important for researchers to continue developing representative models of occupant behavior through a human centered design.

2.3 Mixed-Mode Ventilation



Natural ventilation is a passive strategy that can provide better air quality while improving thermal comfort by increasing air speed during daytime, ventilation rates during the night and by removing heat from the environment, all without consuming a significant amount of energy. In hot climates, the most common method to maintain adequate levels of thermal comfort is air conditioning, which consumes large amounts of energy, resulting in increased costs and green house emissions [76, 140]. In light of the amount of energy consumed with cooling and mechanical ventilation in commercial buildings, it is critical that such strategies are considered when designing.

Due to an increased concern to improve energy efficiency and the need to pursue a greater use of passive strategies for thermal comfort, new alternatives for design strategies discourage the use of mechanical cooling systems where and when natural ventilation can be used [116]. However, studies have shown that when in extreme weather conditions, using only natural ventilation can lead to higher levels of discomfort [9]. As a response, a type of hybrid ventilation system, denominated “mixed-mode” has been studied and evaluated during the past years [116]. According to [60], the Mixed-Mode Ventilation (MMV) system is a hybrid ventilation approach to condition the indoor environment using a combination of natural ventilation from windows, that can be either manually or automatically controlled, and mechanical air conditioning that can provide air distribution and cooling when (or where) needed.

The mixed-mode ventilation system is especially significant in hot climates, for example, where it is not possible to solely rely on natural ventilation, and thermal comfort levels are harder to achieve using only passive strategies. Comfort levels are expected to be met to ensure good indoor environmental quality for occupants. Using only natural ventilation in such climates has been proven insufficient to meet requirements and occupant satisfaction. Therefore, a possible strategy is the combination of natural ventilation and mechanical cooling systems, in a mixed-mode ventilation system [21]. It is a viable solution to provide cooling, air ventilation, indoor air quality (IAQ) and thermal comfort for the occupants of a given building [116].

There are other kinds of ventilation modes that can be combined with natural ventilation, such as low power fans and passive inlet vents. Whichever choice the designer makes, the main objective is to maximize thermal comfort in the building, while avoiding unnecessary use of energy during the year with mechanical air conditioning [29].

According to the Center of the Built Environment [29] at the University of California, there are three types of mixed-mode ventilation buildings, classified as shown in Figure 2.5. The *concurrent mixed-mode operation* uses natural ventilation and air conditioning in the same space and at the same time; the air conditioning serves as a complement to natural ventilation, and occupants are free to operate windows based on their preferences. In the *change-over design*, both situations occur in the same space, though at different times; the change-over can happen on a seasonal or daily basis. This design uses one system or the other, and in several cases it involves an automated system that shuts down the air conditioning when windows are open and vice-versa. As for the *zoned system*, it refers to different zones within the same building, each making use of a different cooling system.

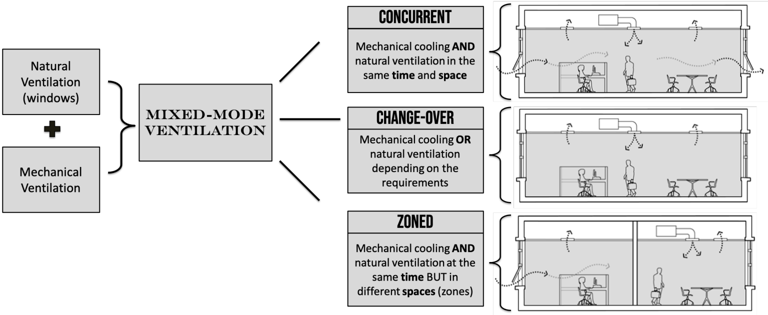


Figure 2.5: Types of MMV buildings (Adapted from Salcido; Raheem; Issa, 2016 and CBE, 2013)

As stated by [116], mixed-mode ventilation strategies have been used and shown effective results in energy saving, while still maintaining indoor air quality for occupants. The authors also state that mixed-mode buildings can save up to 40% of HVAC energy when optimizing window operation schedules, thus considering such strategy a sustainable way to condition buildings. The studies by [115, 117], also evidence the potential in energy savings of such strategy, specifically for Brazilian climates. However, there are no complete guides on how to simulate and/or design mixed-mode office buildings.

2.4 Statistical Methods

Over the years, researches have made use of a variety of statistical methods to model occupant behavior, in an effort to predict a given action. A mathematical model of a natural phenomenon is a quantitative description of that phenomenon [138]. There are several examples in different subject areas, such as biology and physics, of mathematical models portraying natural phenomenons. In the context of architecture, for instance, a model can provide qualitative information about the relationships between several factors that might influence a given event. Ultimately, a model is judged by its usefulness, and such criterion allows the existence of more than just one model for the same event; that is, there isn't one best model for a given phenomenon [138]. Within the realm of models, there are deterministic and the stochastic models. The word deterministic means "certain", and these models are able to predict a single outcome from a given set of circumstances, whereas stochastic models, the word "stochastic" meaning "random", "predict a set of possible outcomes weighted by their likelihoods, or probabilities" [138].

According to [138], there are three components to scientific modeling: (a) a natural phenomenon being studied; (b) a logical system to deduce the implications about the phenomenon; and (c) a connection that links the elements of the natural systems being studied to the logical system applied to model it. Several methods, by using the logical system, can be applied to generate stochastic models to be used in simulation programs. According to [125], simulating occupant behavior in a given context of building modeling from an energetic point of view is mostly done by aggregated stochastic models based on logistic regression, Markov chain or similar methods [46,55,105,124]. Several authors have used the above-mentioned methods in their studies [3,23], and some have combined such methods to achieve their desired results [72, 126]. The following sections are a brief description of commonly used methods to create such models.

2.4.1 Logistic Regression

"Regression is the process of learning relationships between inputs and continuous outputs from example data, which enables predictions for novel inputs" [133]. Regression methods are an integral part of any data analysis that intends to describe the relation between a response variable and one or more explanatory variables [64].

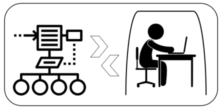
Logistic regression is a statistical method commonly applied to predict the probability of a binary response variable when explanatory variables are at given values [68]. Such method has been used in several studies [55, 72], analyzing occupant behavior in relation to the opening/closing of windows to describe the probability of a window being open or closed, thus configuring the window's state [55, 104, 126], or alterations to the window's state [3, 23] with a set of explanatory variables. The relation between the 'correct' probability for the binary response and the several explanatory variables can be described by multivariate logistic regression. The univariate logistic regression expresses the probability function of a certain event occurring, and can be used in the multivariate logistic regression, which describes the probability of an event occurring depending on an explanatory variable [23]. To select the most significant variables and to build a better model, the Akaike Information Criterion (AIC) is commonly applied, where the forward and backwards procedure is used.

2.4.2 Markov Chain

A stochastic process is a random process, which is a collection of random values in a common probability space [17]. It is an abstract notion that describes quantities that happen at random and can be altered with the passing of time [67]. It is possible to notice in random processes that the outcome of a given trial usually depends on the previous trial. When that outcome is known, there is practically no dependence on the preceding trial, which is known as the Markov property. A random sequence that takes values in a measurable space is a Markov Chain [17]. More specifically, a Markov chain is a collection of random variables that, given the present, past and future states are independent [49]. Therefore, a Markovian process is a stochastic process that only considers the previous state to predict the next one, being independent of the process [55].

According to [47], a Markovian process does not have a memory, and the prediction of the next state will depend only on the present state and no other. The authors made a first attempt to develop a mathematical model to predict the state of windows by using a discrete-time Markov process model to predict transitions between sets of data of opening angles for four office rooms. Studies that use Markov chain technique, intend to generate synthetic data that can portray the overall statistics of the measured data [103].

2.5 Predictive behavioral models for office buildings



Given the growing need to reduce energy consumption, there have been numerous studies on occupant behavior focused on minimizing the gap between measured and simulated data. Most of the studies focus on residential and office buildings, since these show greater impact on global energy use. Great part of the work in this area focuses on one particular type of occupant interaction; the use of electricity and plug loads being the most researched subject, followed by window opening behavior and the use of fans and/or air conditioning [37].

Designers, architects, engineers and researchers need to improve the way energy consumption is calculated in buildings, and considering occupant behavior is one way to achieve that. However, this poses a challenge, due to the complexity and dynamic nature of behavior, that can be influenced by internal and/or external, individual and contextual factors. Also, because the data is monitored on site, the models created can be “locally" applied, that is, the scope of applicability of such models can be limited by the location where the measurements took place. This characteristic reinforces the need to create models for different locations and climates, thus expanding their applicability. Several researchers have developed models from measured data as a means to provide more accurate input data on occupant behavior to achieve results showing less discrepancy from the real data when simulating.

In that sense, operation of different controls have been studied, in an effort to better understand and relate the different drivers to the selected actions being considered. [27] conducted a study reviewing different approaches, methods and findings related to occupants' presence and actions in different building types. The set of occupants' actions in the study includes window, solar shading, electric lighting, thermostat adjustment, clothing adjustment and appliance use. In the database provided by the authors [27], the most recurring action studied in office buildings is related to lighting, followed by window operation.

As shown in the above-mentioned study, researchers have relied on stochastic and data-driven models to predict the operation of different controls. When investigating window use, the desired output of the models have varied, such as window opening/closing, window opening, window state transition and window opening probability. In some occasions, the window model also considered lighting operation in the output [53, 89]. The most recurring methods applied to the prediction of such actions have been Markov chain and logit analysis, although other methods such as generalized linear models, probit analysis and generalized linear mixed models, for example, have also been used. Recently, deep learning techniques have been used to model occupant behavior, both to explain and predict it [27, 83].

The work by [47] was the first attempt to create a stochastic model to predict occupant behavior in regards to natural ventilation in office buildings. The authors performed measurements in four offices in the LESO test facility in Switzerland during a heating season (October to May). Indoor variables such as, room temperature, wind speed, radiation and ambient temperature were registered. Using Markov chain to develop the models, the authors found that outdoor temperature acted as a driver for window opening and closing.

Based on measurements and from surveys conducted, [92] proposed the first coherent probability distribution to predict the state of windows as logit functions for indoor and outdoor temperatures. As stated by the author, this method assumes that the probability of an event happening increases as its stimulus increases, or the intensity of it. The measuring campaigns took place in Pakistan, the United

Kingdom and five European countries, from 1993 to 1996. It was observed that in most of the cases, the correlation with indoor temperature is similar to the one with outdoor temperature, and in this study, the author recommends the use of outdoor temperature, based on the fact that it is an input in any simulation tool, whereas indoor temperature is an output. However, the author later reported that a more consistent predictor for the use of windows was the indoor and not the outdoor temperature [91]. In the latter study, the authors also analyzed the relationship between temperature (indoor and outdoor), and the use of controls in mixed-mode buildings, with mechanical systems available for both cooling and heating. Their results showed that the AC was used for cooling with outdoor temperatures starting at 25 ^{o}C.

[55], based on nearly seven years of continuous measurements, investigated the influence of occupancy patterns, indoor temperature and outdoor variables on window opening and closing behavior. The measurements were recorded in office buildings in Switzerland. The authors tried different modeling approaches, such as logistic probability distribution, Markov chains and continuous-time random processes. Combinations of the approaches were tested, and a hybrid model was selected to be implemented in building simulation tools. The authors report, along with other observations, that indoor and outdoor temperatures, among other possible parameters, are the main drivers for actions on windows.

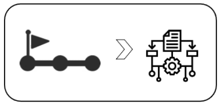
In addition to establishing a relationship between temperature and window operation, studies have also investigated the influence of non-thermal stimuli on occupant behavior. [58], based on measurements conducted in 21 offices in Freiburg, Germany, identified a relationship between window opening/closing and occupancy, showing that most window openings occur at time of arrival. [151] conducted measurements in four offices in South Korea, to demonstrate the link between temperature, CO_2 concentrations, occupancy schedules and window use patterns. The objective was to predict window operation as a function of both thermal and non-thermal stimuli. The authors state that seasonal effects are highly influential in window use, and that the drivers for window operation differ with the seasons. More recently, a study developed by [113] conducted field measurements in mixed-mode office buildings in the south of Brazil, in a city characterized by a subtropical climate. The study generated models applying different statistical methods to analyze relevant variables based on their field measurements' results. The authors focused on analyzing behaviors such as changing clothing or the intake of hot or cold beverages, revealing that such behaviors were affected by different conditions. Data about window and AC operation showed that indoor and outdoor temperatures were strong predictors for these actions. There was a higher probability of AC use with outdoor temperature above 25 ^{o}C, and lower probability below 19 ^{o}C.

2.6 Considerations

Studies on different actions taken by occupants have been developed, mainly focusing on residential and office buildings, although there are researches on commercial and educational buildings as well. The study conducted by [27] revealed that, although there are many studies on window operation in office buildings, there are no works investigating occupant behavior in mixed-mode buildings to create predictive behavioral models to be implemented in simulation programs. The mixed-mode buildings presented in the papers reviewed by [27] differ from those in this work. In other locations, such as Japan [137], Pakistan and Europe [91], the monitored mixed-mode buildings also allowed the mechanical system to be used for heating when needed, and the climates are significantly different from the one in this study (Cwa, for the Koppen classification; [30]), for which there are no reported models. In addition, several actions are contextually sensitive and differ according to personal characteristics, demonstrating the need for studies in different contexts (building type, climate, culture) to better understand a given action [119, 124, 125].

This scenario emphasizes the relevance of this study, since there are low rates of scientific production and collaboration from Latin American countries, specifically from Brazil. As identified by [27], there is a need to develop research studies outside the consolidated and well established domains, so there is a wider range and coverage of knowledge in the field, especially in those climates where there are no models yet. This work comes as a contribution to the field, as it attempts to generate predictive models to a climate, type of building, and consequently control operation that, to this date, are nonexistent.

3 Method



This chapter describes the methodology used in this research, describing the methods and tools to conduct the monitoring campaign, data treatment and statistical approaches to create the models. It also presents the steps taken to validate and test the proposed models.

3.1 Method Overview

The method for the research was developed following the guidelines provided by [127], and its general overview is illustrated in Figure 3.1. It is composed of two initial stages entitled *Development of Theoretical Model* and *Definition of Units of analysis, population and scope*. The *Development of Theoretical Model* was a step to help define what would be measured given the actions being studied. The *Definition of Units of analysis, population and scope* delimited the applicability of the models, since it defines where the findings can be applied. Once these initial stages were complete, the model and its applicability were defined.

The following step was the models' creation, comprising the main body of the method and divided into four main stages (Figure 3.2):

- *Buildings' data collection and pre-test*, where the buildings to be monitored were selected, and a pre-test was conducted to verify the use of equipment and measuring details to minimize errors during the monitoring period;
- *In situ measurements and data analysis*, composed of the monitoring campaign and subsequent data treatment and analysis;
- *Statistical methods' application and creation of algorithm*, application of statistical methods to generate the models predicting window opening and AC activation; and,
- *Algorithm's validation and tests in EnergyPlus*, where the generated models were validated using a test set taken from the data set, followed by tests where the models were implemented in a building performance simulation program, namely, EnergyPlus.

All the above mentioned stages of the method are described in detail in the following sections in this chapter.

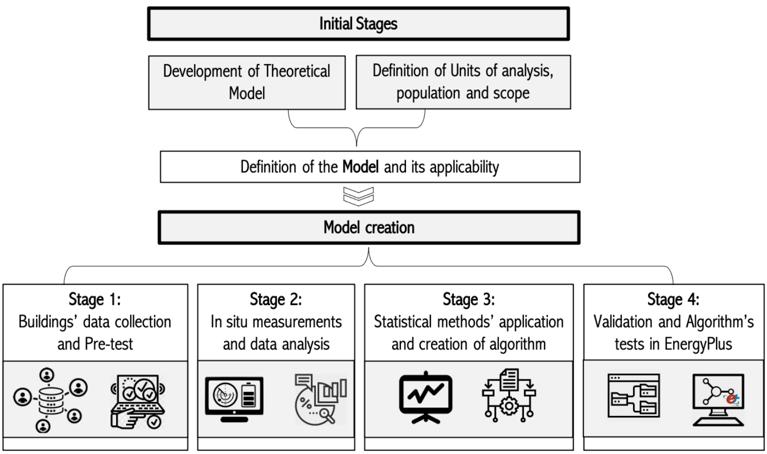


Figure 3.1: Method Overview

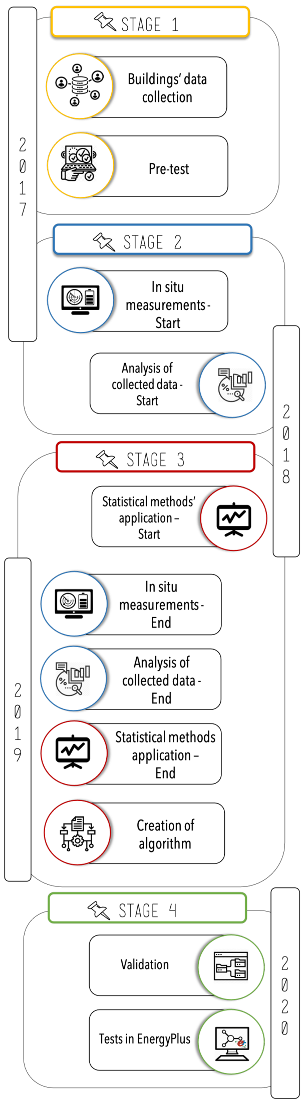


Figure 3.2: Timeline of *Main Body* of method

3.2 Initial Stage: Development of Theoretical Model

When studying occupants in buildings, relevant concepts may include temperature, comfort, and glare, for example, which can be measured and then used to predict occupant behavior. In the process of defining what to be measure and why, it is important to identify concepts to measure and see how they link to each other, by drawing out a theoretical model [127]. Figure 3.3 shows the theoretical model developed for this study.

The theoretical model in this study (Figure 3.3) defines that occupants, in order to achieve or regain thermal comfort, will take actions, specifically open/close windows and/or activate the AC. By taking these actions, they will affect the levels of energy consumption in the building. Links and relationships were established between possible measured and observed variables that will result in a combination of the available actions. For example, when looking at measured variables, high outdoor temperature can result in closing windows and activating the AC. When analyzing observed variables, activation of either control can be made upon arrival and their deactivation upon departure.

At this stage, it is useful to make a distinction between concepts and constructs. Concepts in the field of research on occupants, can be temperature, comfort, glare, and productivity, for example. While constructs are the occurrences of such in a given population [84]. Both can be considered the same thing when inserted in the same population, however, the distinction becomes relevant in international comparative work, where concepts can transfer between populations, and constructs cannot [127]. In the area of occupant behavior, the distinction becomes especially relevant due to the fact that this is a highly international area, and researchers can measure the same concepts, but knowing how they are constructed and conducted demands taking into consideration the differences in climate and culture, for example. With the constructs being well distinguished from the concepts, the process of determining how to measure such constructs begins, which is further detailed in Section 3.4.2.

To aid in further specifying the relationships between the concepts being studied, it is important to establish questions, such as; (a) what triggers occupants to open/close windows and/or activate/deactivate AC?; (b) Do occupants open/close and/or activate AC more frequently as indoor (or outdoor) temperature rises?; and (c) How differently do occupants respond to environmental changes during each season and in between seasons (transition periods)? From such questions, hypothesis can be drawn with the aid of the graphical representation (Figure 3.3), as for example, as outdoor temperature rises, so does the indoor temperature and the probability of occupants closing windows and activating the AC is higher.

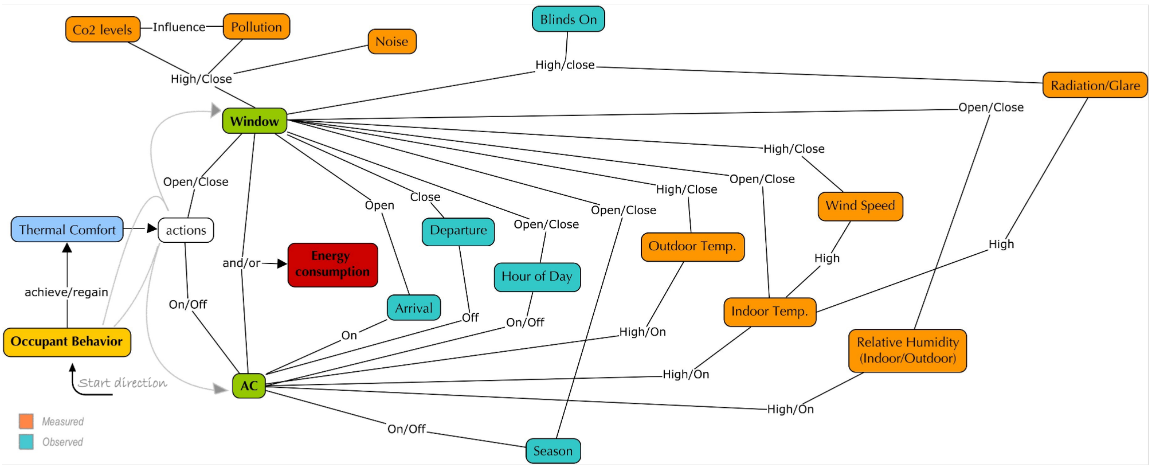


Figure 3.3: Graphical representation of the relationships between concepts.

To aid in further specifying the relationships between the concepts being studied, it is important to establish questions, such as; (a) what triggers occupants to open/close windows and/or activate/deactivate AC?; (b) Do occupants open/close and/or activate AC more frequently as indoor (or outdoor) temperature rises?; and (c) How differently do occupants respond to environmental changes during each season and in between seasons (transition periods)? From such questions, hypothesis can be drawn with the aid of the graphical representation (Figure 3.3), as for example, as outdoor temperature rises, so does the indoor temperature and the probability of occupants closing windows and activating the AC is higher.

3.3 Initial Stage: Definition of units of analysis, population and scope

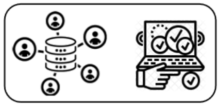
Once the theoretical model is established, it is then necessary to define the scope of its applicability, which requires a statement of the population of interest to which the findings will apply [127]. The first step here is defining the unit of analysis, which is what the data is collected about. In the context of this study, the unit of analysis is the mixed-mode office, which can be studied within a multistory or one story building[1]. Because the mixed-mode office is the unit of analysis in this study, the population of interest and scope's definition are also related to it.

In regards to the population of interest, the samples drawn for the study were taken from a population of buildings found in the selected city for the study, São Carlos, in the state of São Paulo, Brazil. The city is located at latitude 22º 02' South at 863 meters and classified as a high-altitude tropical climate (Cwa, for the Koppen classification; [30]), with a generally non-rigorous dry winter and moderately rainy and hot summer. As for the temporal scope, the monitoring period lasted 18 months, encompassing all four seasons. Therefore, the findings of this study are

[1] Further detailing of such buildings is provided in Section 3.4

restricted to mixed-mode offices in a high-altitude tropical climate for summer, spring, winter and fall.

3.4 Stage 1: Buildings' data collection and Pre-test



3.4.1 Buildings' data collection

The first step in this stage was to create a sample to represent the unit of analysis in the study. As mentioned above, the sample is drawn from a population of buildings within the city of the study. As stated by [127], there is a wide range of sampling strategies that can be used. The "gold standard", as the authors name it, is pure random sampling. However, it is not always possible to attain the ideal case, and other sampling strategies can be applied. In this study, a *purposive sampling* was used, in which "population members are recruited based on certain characteristics considered useful to the study" [127].

In this work, the population members can be understood as the mixed-mode offices in the city of São Carlos, SP. At first, a data collection was performed to create a database with potential multistory and one story buildings. Such one story buildings, mostly former residencies, were entirely adapted to become a commercial building with offices, changing the rooms within the house into offices. The initial data collection was performed using the work of [28], which listed office buildings in the city of São Carlos, SP. In addition to the buildings provided in the cited work, more recently built office buildings in the city were included in this initial data set. Three main, and broad, criteria were established to begin selecting the offices to be included in the sample (Table 3.1), (1) a range for the floor area for the offices within the buildings in the data set (10 m^2 to 50 m^2); (2) the existence of operable windows and an AC unit in each office, where occupants were free to operate either. The AC units here refer to split and window units; excluding central AC systems, and (3) no automation systems nor any other kind of equipment or screen that could somehow influence or inform occupants' or their decisions as to which action to take. After applying the main criteria, part of the buildings within the initial data set was excluded. Next, the specific criteria were defined and the offices were visited to verify if they met the established specific criteria, as listed in Table 3.1.

Table 3.1: Main and specific criteria for building selection

Questions	**Expected Answers**
Main Criteria	
Office's floor area	10 - 50m^2
Are there operable windows and an AC unit?	Yes
Are there automation systems?	No
Specific Criteria	
Are the windows unobstructed?	Yes
Does the user operate the window(s)?	Yes
Does the user stay all day?	Yes
Type of office	Not in health care
Activity	Sedentary
Main equipment	Computers and printers
Number of occupants	1 - 7

Upon visits, as mentioned above, occupants were asked if they alternated between using windows and the AC, to which some responded that they did not operate the windows, using the AC at all times, causing these offices to be excluded from the data set. As for presence, occupants informed their working hours, and the criteria questioned if they remained all day in the office, that is, if they occupied the office during most of the reported working hours, from 9 a.m. to 6 p.m., with a lunch break from 12 p.m. to 2 p.m. Because there was no equipment to monitor occupancy, units where occupants did not remain for the most part of the commonly reported working hours, were also excluded from the database.

The common activity performed in the selected offices was sedentary and mainly using a computer. Offices related to health care, such as medical or dental, were excluded, due to the difference in the activities performed, fluctuation of the number of occupants during the day and difference in equipment. The selected offices only used computers and printers, which account for a specific load that can be estimated when simulating. The number of occupants varied from 1 to 7, given the difficulty to select all offices with the same amount of occupants. The number of occupants in each office can influence the way occupants behave, and this was taken into consideration when building the model. After applying the criteria and visiting each location, a total of ten offices was selected. Table 3.2 presents a general characterization of the selected offices. 11 offices are presented in this table, because office *K* was measured for a short period and later replaced. Nonetheless, the data collected in office *K* was still used in this work, therefore it is included in the offices' characterization. During the entire monitoring period a total of 10 offices (A-J) was monitored in a rotation scheme described in the following sections. Figure 3.4 shows the façades of the selected buildings.

3.4.2 Pre-test

The pre-test was performed during November/2017, and was designed as a prior step to the actual monitoring campaign, with the objectives of (a) determining how best to measure the selected constructs (operationalizing constructs [127]); (b) testing the selected equipment (Figure 3.6); and (c) testing the location of the measuring equipment. Figure 3.5a shows the building façade where the pre-test was conducted, and Figure 3.5b displays the floor plan of the two offices where the test took place. The gray area is the office located on the 6th floor, while the hatched area is a smaller office on the 7th floor.

(a) Building 1 (b) Building 2 (c) Building 3

(d) Building 4 (e) Building 5 (f) Building 6

Figure 3.4: Buildings in final data set.

The indoor variables measured were air temperature and air relative humidity, as well as the actions being studied; window opening and AC activation. The equipment to measure such variables are listed in Table 3.3. The variables were selected based on other field studies in the literature [55, 108, 154]. The variables that show greater influence on the studied occupants' actions, are indoor and outdoor temperatures and relative humidity, and therefore were monitored [54, 55, 152]. Wind speed was not included, since [90] performed a field monitoring in São Paulo, SP, Brazil, and reported very low wind speed rates even when windows were open (maximum registered value of 0.17 m/s). From the observations during this study, the authors concluded that such low values were a consequence of the offices being unilaterally ventilated, as is the case of all the offices in the data set of the present research. [36] reported very similar results when performing

(a) Building where pre-test took place

(b) Offices where pre-test took place

Figure 3.5: Location of pre-test. No scale.

Table 3.2: Offices characterization

Office	Building	Floor	Window orientation	Numb. of occupants	Numb. of windows	Type of window	Blinds	View	Type of ventilation	Numb. of ext. walls	Area
A	01	10^{th}	west	2 (2018) 1 (2019)	5	casement	yes	street	unilateral	1	$35m^2$
B	02	3^{rd}	west	1	1	sliding	yes	street	unilateral	2	
C	01	1^{st}	south	5 (2018) 7 (2019)	5	casement	yes	street	unilateral	2	$50m^2$
D	01	1^{st}	west	3	5	casement	yes	parking lot	unilateral	1	$35m^2$
E	03	0	west	1	1	sliding	yes	wall	unilateral	1	$10m^2$
F	03	0	east	1 (2018) 2 (2019)	1	sliding	yes	wall	unilateral	1	$13m^2$
G	04	9^{th}	south	5	6	casement	yes	street	unilateral	1	
H	04	-	-	1 (2018) 2 (2019)	2	casement	yes	street	unilateral	2	
I	05	7^{th}	west	1	3	casement	yes	street	unilateral	1	$15m^2$
J	05	6^{th}	west	3	8	casement	yes	street	unilateral	1	$32m^2$
K	06	3^{rd}	east	1	3	casement	yes	street	unilateral	1	

a study for the same type of building in Florianópolis, Brazil. Also, [55] stated that window operation was correlated with outdoor air temperature, and not wind speed.

The temporal sampling established in the pre-test to measure indoor temperature and relative humidity was set at a 10-minute interval, based on previous studies [3, 72, 108]. The AC temperature was initially also set at a 10-minute interval, later tested at 5-minute intervals to verify if there were any significant changes at a shorter interval. The 10-minute interval was then selected, since the temperature difference recorded between the 5-minute intervals was not significantly different from the ones recorded on the 10-minute interval trial.

Table 3.3: Equipment for measuring indoor variables

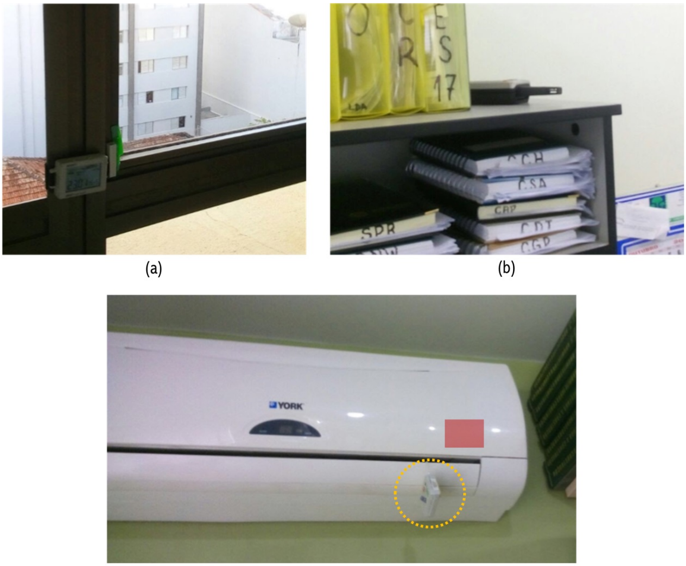
Equipment	Variable	Model	Brand
Air Temp. + RH	AC monitoring	HOBO H08-003-02	Onset
Air Temp. + RH	Indoor environment	175H1	Testo
State data logger	Window monitoring	HOBO UX90-001	Onset

Figure 3.6: Indoor measuring equipment

The windows were not monitored based on time intervals; a state logger (Figure 3.6c) was used to record the changes in state. The equipment continually logged every event (date and time) when there was a change in state. The pre-test allowed a better comprehension of how sensitive this equipment is to displacement. The state logger is composed of a sensor and a magnet; when the sensor is close to the magnet, it records as closed, and when it is away from the magnet, as open. The office room where the pre-test was conducted had top-hung windows, and the sensor was attached to the fixed window frame with a velcro band, while the magnet was attached to the movable part with a tape (Figure 3.7a). The data from the pre-test showed that the windows remained open the entire time, while occupants reported that the common routine was that they opened the windows upon arrival, closed when using the AC and always closed upon departure. It was possible to conclude that the sensor displaced by very little, and thus the minimal distance between the sensor and the magnet affected the recording of the closed periods, showing as always open. These details were all taken into consideration and attended to when the measuring campaign began. This specific issue was addressed by using a double-sided heavy duty wall tape.

The location for each equipment was also tested, and details, such as positioning on the window, how to fix the equipment on the AC unit were all verified and tested, so there were as few errors as possible during the measuring periods. Indoor temperature and relative humidity loggers were positioned away from windows and heat generating equipment. In addition, these equipment were positioned at heights between 80 cm and 160 cm, depending on each office (Figure 3.7b). The intent was to record the measurements at the same height of the mid-section of a sitting person, though that was not always possible.

The positioning of the equipment on the AC was initially done by attaching the equipment on the AC's flap using a plastic clamp (Figure 3.7c - yellow circle), which was later changed to being attached with tape with the sensor facing the flap (red rectangle), to identify the change in temperature more quickly. The outdoor variables were taken from a meteorological station set by the National Meteorology Institute (INMET- Station A711 OMM Code: 86845), which takes hourly measurements. This station is located approximately 5 Km away from the monitored building.



(a) (b) (c)

Figure 3.7: Equipment positioning in monitored office

Prior to the pre-test, an additional procedure was conducted to verify the equipment's calibration. Measurements performed during the pre-test also confirmed that the equipment were functioning well. The software for each equipment was also verified, updated, and tested, as well as all the cables for downloading and uploading data. The pre-test allowed the verification and adaptation of several details for the measuring campaign, which minimized errors during the monitoring period.

3.5 Stage 2: In situ measurements and data analysis

A descriptive, or correlational design, is the basis of studies on the impact of occupant behavior in building energy consumption. It is when data is gathered using sensors, be they installed or virtual, or if the data are gathered for other purposes and possibly enriched with paper or electronically surveys, by use of smart-phones or computers [127]. Once the data are collected, they are analyzed in search of correlations between the variables. This type of study allows researchers to understand the relationships in the data, but not state that a change in one variable is the cause for a change in another. It is common to relate, for example, window opening behavior to high indoor temperature, or to control the indoor air quality. However, as described by [45], there are different types of driving factors, which can be alternative reasons for occupants to take actions, such as hour of day, routine, lifestyle or safety, to name a few.

There are many descriptive research designs; the one adopted for this specific work was the cross-sectional design. As described by [127], this design is one that collects data at a specific point in time covering a range of units of analysis, such as occupants and buildings, for example. This type of design can be conducted once or several times at different moments, thus creating a repeat cross-sectional design.

3.5.1 In Situ Measurements

The measuring campaign for this study lasted 18 months; from December/2017 to May/2019. The objective was to have enough data to train (build) and test (validate) the model, as well as to collect data during all the seasons, in an effort to broaden the applicability scope of the model. The train and test sets were created using statistical methods to randomly select the data entries for each set (Section 3.6). Measurements were conducted in the city of São Carlos, in the state of São Paulo, Brazil (Figure 3.8). According to the Brazilian Standard 15 220-3 [1], Brazil is divided into eight bioclimatic zones, and São Carlos is located in Zone 04, representing an intermediate climate within the range of climates in the country. Figure 3.9 presents monthly characteristics for the city of São Carlos regarding dry bulb temperature and relative humidity taken from the weather data provided by [111]. Based on the same weather data, a wind rose was created (Figure 3.10), showing that the predominant wind direction in the city is southeast.

Figure 3.8: (a) São Paulo state within Brazilian national territory; (b) São Carlos' location within the state of São Paulo.

The measuring campaign for this study lasted 18 months; from December/2017 to May/2019. The objective was to have enough data to train (build) and test (validate) the model, as well as to collect data during all the seasons, in an effort to broaden the applicability scope of the model. The training and test sets were created using statistical methods to randomly select the data entries for each set (Section 3.6).

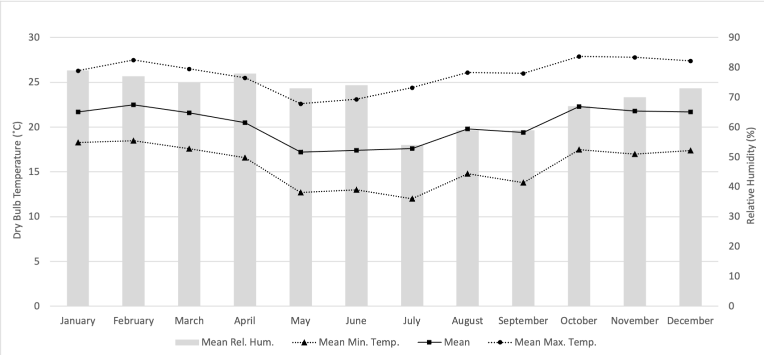


Figure 3.9: Climatic characteristics for São Carlos, SP. Adapted from *EPView* [111].

Each set of five offices, due to the availability of five equipment kits, were measured for a period of two weeks. After this period, the equipment were collected, the data downloaded and the equipment reprogrammed for the next two-week measuring period in five different offices. This rotation was necessary due to the limitation in the amount of available equipment, combined with the offices availability. The objective of such rotation - five kits in five offices every two weeks - was to measure all ten offices every month. However, it was not possible to strictly follow it due to the availability of each office. Nonetheless, measurements were taken in all offices during summer, winter and at least one transition season, with the exception of offices H and K. Figure 3.11 presents a timeline of the measurements. The 2-week periods identified in the timeline as first, middle and last weeks of the month are approximations of such periods. The exact dates can be verified in the complementary material provided and available in Mendeley Data[2].

[2] https://data.mendeley.com/datasets/9v5vgkcykh/1

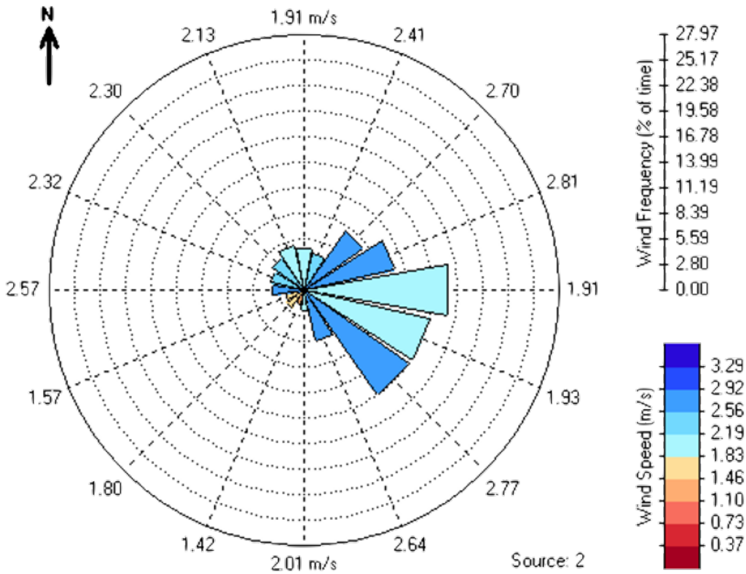
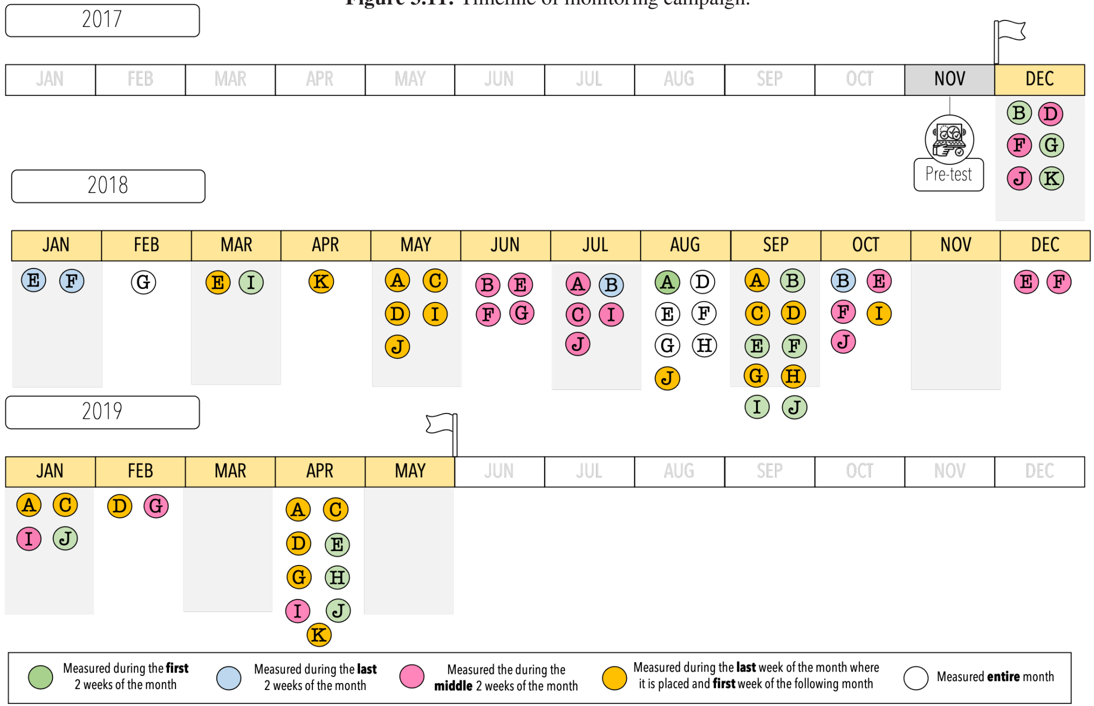


Figure 3.10: Wind rose for São Carlos, SP. Adapted from *EPView* [111]

Figure 3.11: Timeline of monitoring campaign.



The measurements conducted in the campaign followed the guidelines tested and defined in the pre-test (Section 3.4.2) and are presented in Table 3.4. Window opening and AC activation were the two actions monitored and observed in this study. The equipment recording window opening and closing was a state logger that registered when one of these events occurred. The availability of this equipment was limited, so there was one state logger in each office, and not one in each window of each office. Because most offices displayed more than one window, occupants were asked which window was the most used one, and that is where the equipment was placed.

Table 3.4: Measuring guidelines

Variable	Measuring Interval	Details
Occupancy	Occupied period: Mon-Fri, 9 a.m. to 6 p.m.	Diminished or no occupancy during lunch break: 12 p.m. to 2 p.m.
Indoor air temperature (°C)	10 minutes	Equip. placed near occupants at height between 80-160 cm
Indoor air relative humidity (%)	10 minutes	
Outdoor air temperature (°C)	Hourly	Meteorological station A711 OMM Code: 86845*
Outdoor relative humidity (%)	Hourly	
Window State	Every state change	Only one window in each office was monitored
AC Temperature (°C)	10 minutes	AC state calculated from this variable

*www.inmet.gov.br/sonabra

Since there was no equipment to monitor occupancy, if and when there were any unusual activities or if an occupant worked during the weekend or after hours, for example, one of the occupants was designated as a point of contact in each office to report such occasions to be included in the data collection. If there were no uncommon activities, it was assumed that the occupied period was from Monday to Friday, from 9 a.m. to 6 p.m., with a lunch break ranging from 12 p.m. to 2 p.m, as it is the common practice. The measurements' schedule was designed intentionally avoiding national holidays, as to minimize the gaps in the data collection.

The outdoor variables were taken from a meteorological station in the city of São Carlos, as specified in Table 3.4, with a maximum distance of 7 Km from the measured buildings, except for one unit that was more distant from the perimeter containing most of the buildings. This building (Building 04) was approximately 10 Km away from the meteorological station.

Some of the offices were within the same building, but on different floors and sometimes with different facade orientations, as presented in Section 3.4.2, Table 3.2. The measurements were repeated in the same units during each designated period, though occupants in some of the offices varied throughout the measuring period.

3.5.2 Data Treatment

As described in the Pre-test section, the data logger (HOBO UX90-001) was not programmed to measure at given time intervals, but rather when there was a change in state. The equipment is binary and programmed to register an opening event as zero (0) and a closing event as one (1). When reading the collected data, the interpretation for such was that when there was a date and time with a given event, zero for example, it was understood that the window was opened, and that state only changed when the next entry was recorded. Consequently, the next state read one (1) and that was read as closed, and so on for the entire monitored period. Because the other indoor variables were monitored at 10-minute intervals, the window state data had to be treated to fill the 10-minute interval rows respective to the readings on the other equipment. During posterior data treatment to create the models, the numbers respective to each window state, open and closed, were inverted to be in accordance with what is commonly used in literature; 1 (one) as open, and 0 (zero) as closed.

The outdoor variables were also not recorded in 10-minute intervals, the meteorological station is set by the National Meteorology Institute (INMET), and provides the hourly values for each variable (instantaneous), as well as the minimum and maximum values within the hour. These data were also treated to be used in this study, given the established 10-minute interval. The criteria applied to this was to use the instantaneous value and linearly decrease or increase it, at 10-minute intervals, to match the next hour's instantaneous value, thus providing the same measuring interval as the other variables being recorded in the study. This procedure was applied to outdoor temperature and relative humidity.

The equipment registering the AC temperature followed the 10-minute interval specified in the Pre-test. Although there was no need to treat these data to fit the required interval, it was necessary to establish a guideline for its interpretation. Because the equipment registered temperature and not the date and time when the AC unit was activated or deactivated, it was necessary to analyze the data set for the measured period and identify when the AC activation and deactivation occurred, thus establishing the activation and deactivation periods.

The first method used to identify AC operation consisted in applying Equation 3.1, here specified as 'thermal velocity'. Where T_n is the nth temperature within the measured AC temperatures, divided by 10, which is the amount of minutes in the measuring interval. When the module's value was greater than 0.3, it was se to indicate AC operation. However, this method only identified AC *operation*, meaning that it also identified AC operation while the equipment was already activated. The following step was to identify when there was only AC activation and deactivation considering the recorded temperatures during the monitoring campaign. When using Equation 3.1, the result value of 0.3 was set to indicate operation in all offices, since this first treatment was only for measurements during the summer and this value was adequate to indicate operation of AC. However, as the monitoring campaign progressed, and because each office displayed different use patterns according to each season, this value was changed and became specific to each monitored period and office as needed.

$$ThermalVelocity = (T_n - T(n-1))/10 \qquad (3.1)$$

In each of the spreadsheets for the data treatment, (available in Mendeley Data[3]), the following four columns were created; **Activation** and **Deactivation**, and **Activation value** and **Deactivation value**. Each line of the **Activation** and **Deactivation** columns was linked to each of its respective line of the 'thermal velocity' value established on the previous step by the given equation, and to the **Activation value** and **Deactivation value** specified. The **Activation** and **Deactivation** columns displayed a 1 (one) if activation or deactivation of the AC was identified. This was achieved by using a formula stating that if the thermal velocity value was smaller than the activation value established, it should be indicated with a 1 (one). The same logic was applied to establish deactivation, although using the inverse. Therefore, if the thermal velocity value was greater than the specified deactivation value, it was indicated with a 1 (one).

Lastly, a formula was applied in another column, with the intention of combining the information in the **Activation** and **Deactivation** columns. This merging allowed to fill every 10 minute interval with a state for the AC equipment, thus creating the AC state column. The following step was to verify if every calculated AC state period was in agreement with the measured data. This was achieved by creating a line graph including the calculated AC state, measured window state (to check if they were used alternately or not), AC temperature, as well as indoor and outdoor temperatures. The graphs allowed to verify if the calculated AC state was coherent with the monitored AC temperatures. Some manual corrections had to be made in some cases after verification with the graphs.

After all variables were treated to fit the 10-minute interval, filters were applied on all spreadsheets (one for each measured period of each office), so that only

[3] https://data.mendeley.com/datasets/9v5vgkcykh/1

days of the week were selected and all holidays excluded. Days when occupants reported as being on vacation or not at the office for varied reasons were also excluded. In addition, a filter for time (in minutes, with 0 minutes at 12 a.m.) was applied, so that the data displayed was only that referring to the occupied period. The filter applied here ranged from around 8:00 a.m. to 6 p.m. (510 to 1080 minutes), so that the time of arrival would be captured more easily. The actual reported time of arrival was 9 a.m., but the filter was applied for an earlier time to ensure the capture of arrival, since it varied around 9 a.m. Because the models were generated using the data resulting from the filters' application, the following sections refer to the reported working hours as from 8 a.m. to 6 pm. The reported 2-hour lunch break was not excluded from the data, since not all occupants took their lunch break at the same time and with the same duration.

3.5.2.1 Data treatment for Markov Chain models

The data set resulting from the data treatment described in the previous sections was used to create the models using two different methods; GLMM and Markov chain (Section 3.6). However, for the Markov chain models to be generated, there was an additional step to the data treatment procedure, as this type of model predicts the transition probabilities, which differs it from the GLMM models.

Because one file was created combining all the spreadsheets of all the measured offices and periods, the first step was to sort the data set, first ordering it by *Office ID*, then by *Day of Year*, and lastly by *Time*. This was necessary because this method needs to identify when a transition takes place, and thus indicate the instance right before the transition occurs. Therefore, having the days and times in order is important when identifying such instances.

As mentioned above, this method is based on the transitions that occur. In the case of this study, that means when a window goes from *Open* to *Closed*, and vice versa, or when the AC goes from *On* to *Off*, and vice versa. The following step was to identify such transitions, and every step described from here on out was performed for the Window State and the AC State separately. This was accomplished by applying functions to read the entire data set and identify each line as 00, 01, 10 and 11. Table 3.5 details what each of these instances mean.

Table 3.5: Markov chain data treatment: states and transitions

State	Identification	Detail	Classification
00	Closed/Off	Current and Next states were *closed/off*	Closed/Off
01	Transition	Current state was *closed/off*, Next state was *open/on*	Open/On
10	Transition	Current state was *open/on*, Next state was *closed/off*	Closed/Off
11	Open/On	Current and Next states were *open/on*	Open/On

The Markov chain models, like the GLMM ones, will predict the probability of an outcome, the probability of state 1 occurring, therefore each state identified as described in Table 3.5 had to be converted into *ones* (1s) and *zeros* (0s), though in this data set the 1s represent a transition. Every 00 and 11 instance was converted into 0, as there was no transition, and every 10 and 01 was converted into 1s, as they implied a transition took place. With this, two separate data sets were created for each control; Window Open, Window Closed, AC On and AC Off. Every 00 and 01 states were grouped into the Open or On data set, and every 11 and 10 were grouped into the Closed or Off set, resulting in four data sets.

Once the four data sets were complete, each one was split into the *Training* and *Test* sets. Each model was generated using their respective *Train* set, leaving the *Test* set for validation.

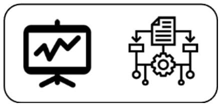
3.5.3 Analysis of collected data

Once all the data was treated and the spreadsheets were completed only with data relevant to the study, an analysis of the studied actions in relation to the environmental data collected was conducted. This was a step taken to identify the relationships between such variables, as thought out in the theoretical model (Section 3.2). Relationships mostly between temperature and actions were identified, providing a basis of what to expect from the statistical models' predictions.

The first step to this analysis was to look at the characterization of each monitored office, with its key features as listed on Table 3.2, together with a graph combining indoor, outdoor and AC temperatures, as well as window opening/closing and AC activation/deactivation periods, which was created for each monitoring period (2 weeks) for each office (Available in Mendeley Data[4]). In addition, histograms portraying the frequency of window opening and AC activation by indoor and outdoor temperatures, and hour of day were also created for the measurements, as well as histograms for each control use during each season, box plots and density graphs to aid in the interpretation of the measured data.

By combining the information from the room's characterization and the graphs, it was possible to create a general overview of how the monitored offices performed differently and how occupants responded to the building's characteristics, in combination with possible personal preferences and routine. The graphs aided in the analysis of the measured variables and enabled further analysis, allowing to infer how routine and hour of day, for example, influence occupants' actions.

3.6 Stage 3: Statistical Methods' Application and Creation of algorithm



Based on previous studies presented in the literature [3,23,46,55,72,105,124,126], there is a variety of statistical methods that can be used to create occupant behavior

[4] https://data.mendeley.com/datasets/9v5vgkcykh/1

models. The most common methods are Markov chain and logistic regression. Given the type of data collected, and the desired outcome of the predictions, the models were first generated using logistic regression. Next, Markov chain models were created. One model for each control (ACA and WO) was created for the logistic regression models, while 2 models for each control were created for the Markov chain models, namely, ACAOn, ACAOff, WOpen and WClosed.

After the data treatment described in the above sections, statistical data treatment was performed using the statistical software RStudio 1.3.959 [112], which was also the software used to create the models. The *"rosner test"* [75] was applied to identify and remove any outliers from the data set, which resulted in removing 20 data entries from the set. Next, the complete data set (displaying data from the 18-month monitoring campaign) was randomly divided to create the *"training"* and *"test"* sets. The function *'sample_fraction'* was used, determining a set with 70% of the complete data set (approximate amount of entries respective to 12 months of measurements), and the function *'anti_join'* was applied to separate the remaining 30% of the set. These functions are part of the 'dplyr' package [143].

3.6.1 Logistic Regression Models

With the data treated and the sets established, the following step was to create the statistical models using the *"training"* set to predict the studied actions. In order to address the issue of behavioral diversity, a *Generalized linear mixed model* (GLMM) was used. The function *"glmer"* from the package "lme4" [10] was applied. According to [52], behavioral diversity is not correctly handled when simply adding a further predictor to a *generalized linear model* (GLM). According to the authors, this issue is correctly addressed by adding a variable of random nature, which results in a GLMM. Design or environmental factors are taken to be *fixed effects*, which are definite values, measured, as opposed to the individual's characteristics taken randomly from a population, thus treated as a *random effect*. Therefore, the combination of these effects is what composes a *mixed-effects logistic model* to predict the probabilities of a given action as shown in Equation 3.2

$$\text{logit}(p) = \beta_0 + b_0 + \textstyle\sum_{k=1,\ldots,n}(\beta_k x_k + b_k x_k) \tag{3.2}$$

where x_k is a fixed effect predictor, β_k is the regression parameter for a fixed predictor x_k, and b_k is the estimated random effect associated to a predictor x_k [52].

The environmental variables included in the ACA model were indoor and outdoor temperatures, indoor and outdoor relative humidity and time, which were considered as *fixed effects*. All the same environmental variables, and therefore *fixed effects*, were included in the WO model, except for indoor relative humidity, which displayed a very high correlation to outdoor relative humidity, and was therefore not used in this model to avoid non convergence of the model. Although indoor and outdoor relative humidity were included in the models, an analysis specific to this variable was not included, as its behavior, and therefore the predictions given by the models based on this variable, were very similar to indoor and outdoor temperatures. *Office ID*, which portrayed an individual or a group of occupants, was included as a *random effect* in both models. For the ACA model, window state was added to the model as a factor and vice-versa.

Polynomial functions were used for the variables time and indoor and outdoor temperatures. Time and its polynomials to the 4$^{\text{th}}$ degree were included, and temperatures and its polynomials to the 3$^{\text{rd}}$ degree were included. No interactions between variables were included in the models. All variables were scaled, given the different ranges that they presented. Table 3.6 presents the mean (center) and standard deviation (scaling factor) for each variable.

3.6.2 Markov Chain Models

Markov models are modeled to predict the transition probabilities of each control. They were formulated as logistic models (Equation 3.3), with P_{01} being the

Table 3.6: Mean and Standard Deviation of scaled variables

Variable	**Center (mean)**	**Scaling Factor (SD)**
Time (minutes)	795.3	166.5
Indoor Temperature (°C)	24.4	2.4
Outdoor Temperature (°C)	23.1	4.4
Indoor Relative Humidity (%)	50.6	11
Outdoor Relative Humidity (%)	61.5	19.9

probability of a transition from closed to open, or off to on, and vice-versa (P_{10}) [123]. These models were also generated with the RStudio software [112], and the *lrm* function [57] was used to create the models. All measured variables were included with no interactions (indoor and outdoor temperatures, and indoor and outdoor relative humidity). Time and temperature (indoor and outdoor) were included with their respective polynomials with the same degrees as in the GLMM models, 4th and 3rd, respectively. This model did not include *Office ID* as one of its variables, as this was the *random effect* variable particular to the GLMM models.

$$P(X_1, \ldots, X_p) = \frac{exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p}{1 + exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)} \tag{3.3}$$

3.6.2.1 Penalty Factor

When a model is generated using an imbalanced set, that is, when there is a predominance of entries of one of the classes over the other, applying a penalty factor is a method that helps improve the predictions made by the model. In this work, because of the imbalanced sets resulting from the data treatment to create the Markov Models (Sections 3.5.2.1 and 4.3), the *pentrace* function [57] was applied to determine the penalty factor for each model.

The *pentrace* function is applied, where a variety of λ's, with a range from 0 to 60, is used to estimate the corresponding AIC. The function selects the λ that maximizes the modified AIC as the optimal penalty factor.

Next, the original model is updated to include the penalty factor. An approximation of the full penalized model is made, which is simplifying or reducing it. This step verifies if the original variables in the model will remain or if there are variables that can be excluded, given their small significance to predict the desired outcomes.

More detailed information of the applied method, including the specific functions applied for each step can be found in the additional data available in [86].

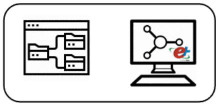
3.6.2.2 ROSE Method

The ROSE package [81] allows to generate artificial data based on sampling methods. The package has functions that allows the creation of sample synthetic data, achieved by enlarging the features space of minority and majority class examples.

The *ovun.sample* function gives four possible options to create the synthetic sample data, namely; over sampling, under sampling, both and ROSE. Within the function one can define which of the four above-mentioned methods to be used. The method then instructs the algorithm as to which type of sampling to perform. The number of variables within the new sample data is defined by *N*. Therefore, when the method *over* is selected, for example, it defines over sampling, which over samples the minority class to match the amount of samples in the majority class. *N* is then defined as the number of samples in the majority class multiplied by two. The opposite is true for the method *under*, which performs under sampling and is done without replacement. For the method *both*, as the name itself suggests, does both over and under sampling, therefore the minority class is over sampled and the majority class under sampled without replacement. With this method, variable *P* is also defined, and refers to the probability of positive class in the newly generated sample. In this work, P was set at the default value of 0.5 for all four models.

The last method used with the ROSE package created synthetic data, which can help avoid limitations found in the above described methods. The newly generated data has the same size as the original set. Once all the newly generated data sets were ready, the models were run using each of the sets, yielding a total of 16 models. Next, predictions were made using the test sets, and the accuracy of each new model was verified using the in-built *roc.curve* function to determine the best performing model for each of the studied actions. The models generated using this method are referred to in this study as ROSE models.

3.7 Stage 4: Validation and algorithm's tests in EnergyPlus



3.7.1 Validation

Validation and verification are procedures employed to determine realistic and confidence expectations [147]. For this research area, specifically, [146] describe that validation can include a careful collection and preparation of sufficient and representative data, and a systematic separation of subsets to (a) generate a model and (b) to evaluate the model and generate a clear discussion of its limitations based on statistical significance and application.

In this study, RStudio [112] was used to randomly generate the training and test sets, splitting the data into two sets, of 70 and 30% of the data for each of the tasks, respectively (Figure 3.12). Next, the test set was used as input for the models to predict the desired outcomes. Lastly, the outcomes given by the model while using the test set, were compared to the measured values in that same set and

Confusion Matrices were generated to evaluate the models' performance given the different values (thresholds) established to create the classes.

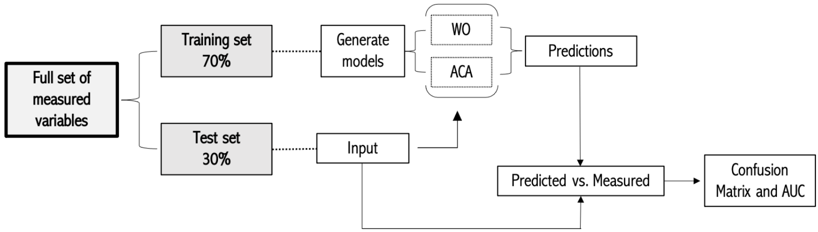


Figure 3.12: Validation process overview

3.7.1.1 Confusion Matrix and AUC-ROC Curve

After creating the models and getting the output in probabilities, it is important to measure the effectiveness of the models. For this measurement, Confusion Matrices can be used as a way to measure a models performance. A Confusion Matrix is a performance measurement to classify machine learning problems [88]. Figure 3.13 illustrates the concept, showing the four different classes of predicted and measured values. This method is useful to calculate Recall, Precision, Specificity and Accuracy, as well as the AUC-ROC Curve, which are described below.

Once the model is fitted and predictions are made using the test set, the following step is to turn the probabilities into classes and verify their frequencies. In this study, probabilities of 40, 50 and 70% were turned into classes and analyzed for the GLMM models, (Section 4), while a value of 90% was used for the ROSE models (Section 4.3.2).

As shown in Figure 3.13, the predicted data can fall into four different categories, namely; True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) [78]. The predicted values are described as Positive and Negative, whereas the Measured (or actual) values are described as True and False. In terms of the studied actions, the interpretations are as follows when using window opening as an example, also shown in Figure 3.14:

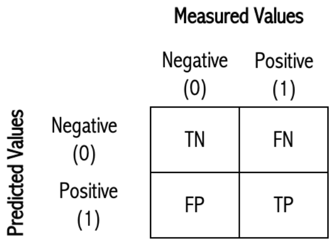


Figure 3.13: Confusion Matrix: Combinations of predicted and actual values

True Positive: The model predicted the window as open (positive), and it was true, the window was open in the measured data.

True Negative: The model predicted the window as closed (negative), and it was true, the window was closed in the measured data.

False Positive: The model predicted the window as open (positive), and it was false (Type 1 error), the window was closed in the measured data.

False Negative: The model predicted the window as closed (negative), and it was false (Type 2 error), the window was open in the measured data.

Once the amount of predictions of each category is given, it is then possible to calculate the Recall, Precision, Accuracy and F-score of the models. The Recall is calculated using Equation 3.4, which gives how much was predicted correctly out of all the positive classes. Ideally, it should be as high as possible.

$$Recall = \frac{TP}{TP + FN} \tag{3.4}$$

Precision (Equation 3.5) is given by calculating how many are actually positive given all the positive classes that were predicted correctly. The Accuracy will be given by how much was predicted correctly out of all classes.

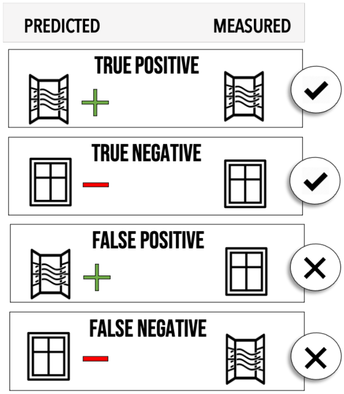


Figure 3.14: Confusion Matrix: Possible outcomes of window example

$$Precision = \frac{TP}{TP + FP} \tag{3.5}$$

And lastly, because of the difficulty of comparing two models with low precision and high recall, or vice versa, the F-score (Equation 3.6) is used to make the models comparable. This method combines Precision and Recall using Harmonic Mean instead of Arithmetic Mean by punishing extreme values more.

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3.6}$$

Once the Confusion Matrices are complete, the following step is calculating the AUROC Curve. This can be considered an additional step, since it is another performance measurement. This method, which stands for Area Under the Receiver Operating Characteristic curve, measures the performance of classification of the given problem at different threshold settings [87]. The AUC plots the values of the False Positive Rate (FPR) (x-axis) against those of the True Positive Rate (TPR, also Sensitivity or Recall) (y-axis) for specified cutoff values. This is useful since it shows how well the model can distinguish between the classes. The higher the AUC, the better the model can predict zeros and ones correctly.

The following terms are then defined to be used in the AUROC Curve:

$$TPR = \frac{TP}{TP + FN} \tag{3.7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.8}$$

$$FPR = 1 - Specificity \tag{3.9}$$

When analyzing the AUROC Curve, the model can be classified into one of four possible general categories. A model that is predicting very well will have its AUC near to 1, meaning it has a very good measure of separability. On the other hand, a poor model will display an AUC near to zero, meaning the exact opposite. A third scenario is when the AUC is 0.5, meaning the model does not have any separation capacity. And lastly, when the distributions overlap, the model's capability of distinguishing between the classes will be classified given the AUC value, as described in more detail ahead.

Figure 3.15 displays the four possible outcomes as mentioned above. Figure 3.15a illustrates the ideal situation, when the curves do not overlap, meaning the model has a good measure of separability and is able to distinguish well between the positive and negative classes (AUC=1). Figure 3.15b shows some degree of overlapping, which is where Type 1 and Type 2 errors are introduced. Depending on the threshold, these errors can be minimized or maximized. In this example, with an AUC=0.7, it means that there is a probability of 70% that the model will be able to correctly distinguish between the classes. Figure 3.15c shows the worst situation, where the model has no capability of distinguishing between the classes (AUC=0.5). And lastly, Figure 3.15d, with AUC=0, illustrates a scenario where the model is reciprocating the classes, that is, it is predicting a negative class as positive and vice-versa.

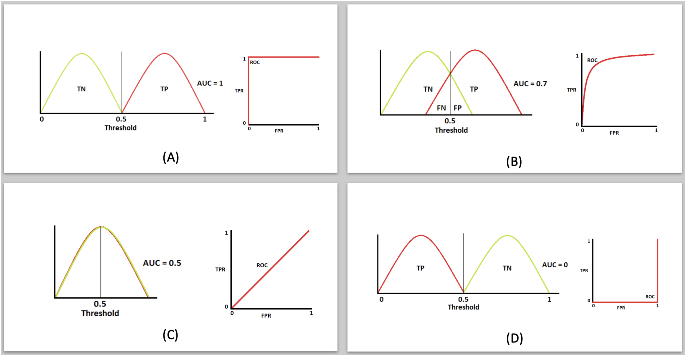


Figure 3.15: Possible outcomes of the AUROC Curve. Adapted from [87].

3.7.2 Algorithm's tests in EnergyPlus

It is well known and established that there is a performance gap between measured and simulated data concerning occupant behavior [78, 89, 101]. As this work's objective was to contribute to bridge such gap, the objective of the tests was to compare the energy consumption outcome of a mixed-mode office building when using the generated models versus the commonly applied fixed schedule for occupant behavior.

3.7.2.1 EnergyPlus

EnergyPlus (EP) is an open source program that models heating, ventilation, cooling, lighting, water use and renewable energy generation among other building energy flows. It is supported by the Department of Energy of the United States and validated by ASHRAE-140 [6]. The program includes some capabilities that can be considered innovative, such as sub-hourly time-steps, natural ventilation and thermal comfort. It is a reliable program, since its heat balance approach has the potential to be the most accurate method to solve heating and cooling loads, since it accounts for all energy flows in their most fundamental form [142]. Interfaces and modules were created to make the program easier to use in engineering.

3.7.2.2 Building configuration

All simulations were based on the EP model by [117], therefore containing the same general information about the building. The selected EP model has very similar characteristics to the measured offices in this work. The above-mentioned authors defined their models after analysing a database provided by a real estate company [20]. After applying a filter to display only offices with floor area of less than 100 m^2, the authors verified that the offices' floor areas varied between 13 and 67 m^2, with an average of 36 m^2, and most of them with a floor-to-ceiling height of 2.7 m, as well as a split air conditioning system. These characteristics are very similar to those of the monitored offices in this work, in addition to the simulated model also displaying unilateral ventilation, as was the case of all the monitored offices in this work. Further detailing of the model, such as materials, internal gains and occupation, are also in agreement with the collected data, therefore the simulations outputs for the selected EP model are comparable to the measured data.

The definition of the constructive components was based on the work of [19], where the author defined common office buildings' characteristics based on the data collection performed by [28]. Each constructive components' properties, such as thermal transmittance (U), thermal capacity (TC), thermal conductivity (λ), specific heat (h) and visible solar absorptance (α), were calculated based on the Brazilian Standard NBR 15220 [2] and are presented in Tables 3.7 and 3.8.

The internal gains provided by [117] and shown in Table 3.9 were calculated taking into account occupants, equipment and artificial lighting. A total of 5 occupants was considered, with occupation hours from 8 a.m. to 6 p.m. During the occupation period, the equipment, such as computers and printers, was considered to be constantly in use, while artificial lighting was activated when needed.

Figure 3.16 shows the model's geometry (Figure 3.16a) and simulation module (Figure 3.16b). All comparisons to the model in the work of [117] were in relation to the north facing room. A window to wall ratio (WWR) of 40% was used. The same weather file for the city of São Carlos, SP, Brazil [111] was also used for

Table 3.7: Input data for EnergyPlus simulation: Constructive Components

Component	Material
Roof	Concrete slab (8 cm)
External Wall	External Plaster (2.5 cm) Perforated concrete block (19 cm) Internal Plaster (2.5 cm)
Internal Wall	Plaster (2.5 cm) Perforated concrete block (9 cm) Plaster (2.5 cm)
Floor	Gravel (3 cm) Concrete (5 cm) Plaster (2.5 cm) Ceramic floor (0.5 cm)
Door	Wood (3.5 cm)
Simple Glazing	Clear (6 mm)*

*EnergyPlus database

all simulations. The type of windows and air conditioning system characteristics remained the same in all simulations. The AC was modeled in EP as objects in the class *HVAC Template*, with a cooling set point of 25°C and running as an ideal system. Natural ventilation was modeled using the *Airflow Network* object, where the unilateral ventilation was also modeled. The type of windows used in this model represent an effective opening area of 100%. Lastly, to alternate the use of AC and windows with the fixed schedules, a hybrid ventilation manager was used (*Availability Manager Hybrid Ventilation*), where the specific parameters for the use of each control are defined. Figure 3.17 illustrates the temperature limits that allow AC and window use within the fixed schedules scenario. The variation from one simulation scenario to another was the way the controls were allowed to be operated, and not their type.

Table 3.8: Specification of constructive components

Material	**U (W/m².K)**	**TC (KJ/m².K)**	λ **(W/m.K)**	**h (J/kg.K)**	α
Concrete slab	3.31	1.76	1.75	1,000	
External Plaster	2,62	323	1.75	1,000	
Perforated concrete block (19cm)	2,62	323	1.15	1,000	0.11*
Internal Plaster	2,62	323	1.15	1,000	
Plaster	2,27	206	1.15	1,000	
Perforated concrete block (9cm)	2,27	206	1.75	1,000	
Gravel	3.08	281	0.7	800	
Concrete	3.08	281	1.75	1,000	-
Ceramic floor	3.08	281	1.15	1,000	

*Matt acrylic white paint (Branco Neve) [41]

3.7.2.3 Implementation of behavioral models

The models created in this work were implemented in EnergyPlus using the Energy Management System, following the guidelines as described by [51]. The first step is to define the sensors (*EnergyManagementSystem: Sensor*), which are variables that will be used in the model, such as indoor and outdoor temperatures, for example. Next, actuators are defined (*EnergyManagementSystem: Actuator*), and these are linked to each actions' respective schedules already defined in the *.idf* file. When the model calculates an output, it will set a value for the actuator, meaning action or no action. The actuator will then allow the action in the schedule that it is linked to. In this case, actions will be taken to control the window and the AC, therefore the actuators' schedules are related to these actions. The following step is to identify the programs that will be used in the simulation (*EnergyManagementSystem: ProgramCallingManager*). The programs are the models themselves, in this section their respective names are listed, and the order

Table 3.9: Input data for EnergyPlus simulation: Internal Gains

Internal Gains	
Users	Occupation: 7 m^2/person Metabolic rate: 65 W/m^2
Equipment (computers)	Density: 10.7 W/m^2 Radiant factor: 0.3
Lighting	Density: 9.7W/m^2 Use pattern (hours): Dimmer

in which they are listed is the order in which EP will read and run them. For this study, the window models are listed first, since the measured data showed a higher frequency of window opening at time of arrival.

And lastly, the program is written (*EnergyManagementSystem: Program*) using the variables and their respective coefficients as previously defined.[5]. The parameters defined to be used in each program must be entered as Global Variables (*EnergyManagementSystem: GlobalVariable*). All models were created using scaled variables, therefore the scaling factors (mean and SD) had to be included when writing the models in EP. These variables were configured in the same section as the other ones in the model. As a way to verify if EP calculated all scaled variables correctly, all variables defined in EMS were selected in the section *EnergyManagementSystem: OutputVariable*, which displays the scaled variables as calculated by EP and used in the models for each simulation.

3.7.2.4 Simulation scenarios: models' combinations

Different models combinations were tested in an attempt to find the best representation of occupant behavior in mixed-mode office buildings, taking into

5 The coefficients for each model generated in this work can be found in the Appendix (Section A)

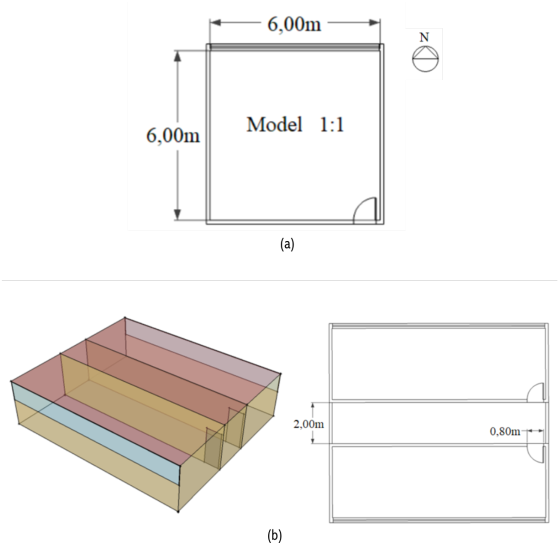


Figure 3.16: (a) Dimensions of selected model; (b) Simulation module. No scale. Adapted from [117]
.

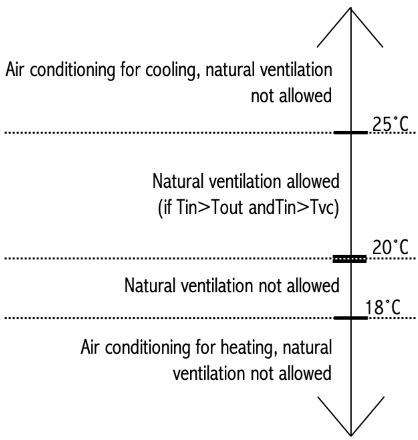


Figure 3.17: Temperature control for air conditioning and natural ventilation use during the occupied period. Adapted from [117]
.

consideration comparisons to the measured data. *Simulation A* and the measured indoor temperature were used as the basis of comparison. *Simulation A* is representative of the way mixed-mode offices are modeled in the Brazilian scenario of energy simulation. Energy consumption was not one of the measured variables in the monitoring campaign, thus the choice of comparing the indoor temperature between measured and simulated data. However, energy consumption was compared between the outputs of all simulated scenarios.

The following scenarios were tested in this specific order, as the outputs of each scenario were analysed and compared to the measured data. This comparison indicated if the behavioral models used in the simulation were accurately depicting the measured behavior. As limitations were identified, different models were used, as well as different combinations of the available models.

- *Simulation A* | Model using fixed schedules by temperature to allow window, that is, natural ventilation (NV), or air conditioning (AC) use, and thus alternate between these controls (Figure 3.17).
- *Simulation B* | Window opening and AC activation models to control the studied actions. Both models in this simulation were GLMM models.
- *Simulation C* | GLMM Window opening model combined with the AC by temperature as used in *Simulation A*. The use of AC was only allowed when the window was closed and the remaining parameters were met as well.
- *Simulation D* | Window opening and AC activation models generated using Markov chain.
- *Simulation E* | GLMM Window opening model and AC activation model using Markov chain.
- *Simulation F* | Window opening and AC activation models generated using Markov Chain with Penalty Factor applied.
- *Simulation G* | GLMM Window opening model and AC activation model using Markov chain model with Penalty Factor.

- *Simulation H* | Window opening and AC activation ROSE models (Markov chain).

4 Results

This chapter presents the results of an analysis of the collected data, followed by the results of the models generated in this work with their respective validation procedures, and lastly, the results of the simulations where the models were implemented and tested in EnergyPlus. Models coefficients for the GLMM and MC Rose models are presented along with the models' results. Coefficients of all models are available in Section A (Appendices A-D).

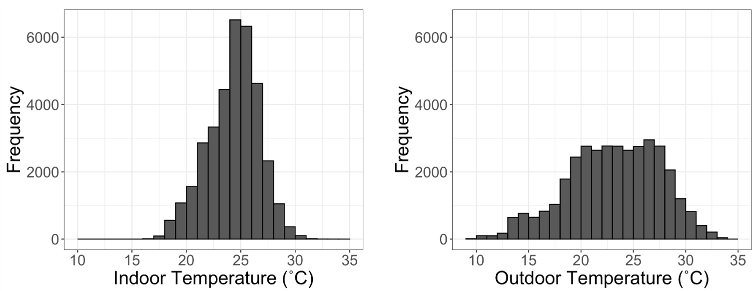
4.1 Analysis of collected data

This data analysis was based on the full treated data set, which comprises 18 months of monitoring. This analysis was conducted to provide a better understanding of the complete data set, such as the temperature ranges and the temperatures at which the controls were mostly in use or not. It was also important to identify key information, for example, most frequent outdoor temperature at which the AC was used, to verify if the models predictions reflected the measured data well.

In addition, this analysis allowed to identify differences in how the controls were used in each office. The information of the graphs provided here, combined with the information about the offices provided in Table 3.2 (Section 3.4), as well as the graphs available in Mendeley Data[1], allowed to draw conclusions about some of the offices.

4.1.1 Temperature

The indoor temperature ranged from 16 to 34.5 °C, while the outdoor temperature ranged from 9.5 to 33.4 °C (Figure 4.1). Figure 4.2 shows a density graph for the entire measured period, displaying the indoor, outdoor and AC temperatures. As it can be seen, indoor temperature is mostly maintained at 25 °C. Figure 4.3 presents the temperature density specific for each control use.



(a) (b)

Figure 4.1: Indoor and outdoor temperature frequencies during the measured period

Figure 4.3a shows two peaks of AC temperatures. A smaller one, showing that the AC was set at low temperatures, around 10 °C, and a bigger one, with a peak at around 25 °C. This figure also allows to see that the AC was activated with higher outdoor temperatures

As for window use, Figure 4.3c shows a higher density of outdoor temperature around 20 °C when windows were open, while Figure 4.3d shows that they were closed with higher outdoor temperatures, likely coinciding with the use of AC.

[1] https://data.mendeley.com/datasets/9v5vgkcykh/1

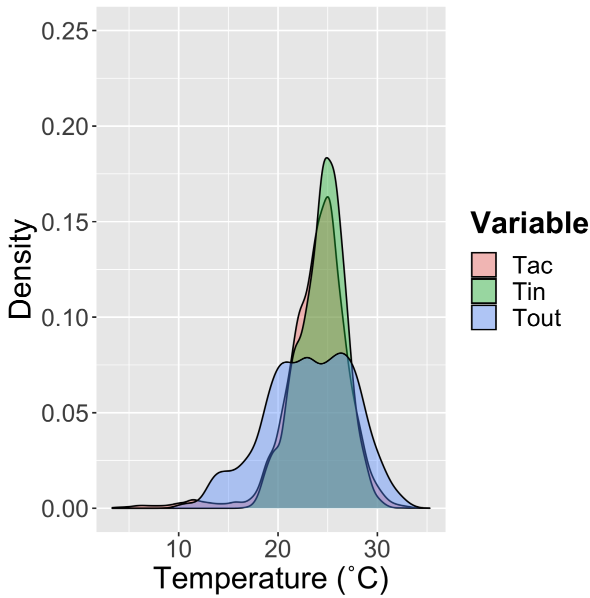


Figure 4.2: Density graphs of measured data

As for indoor temperature, it is clear that occupants alternate between controls to maintain it around 25 ^{o}C, as it constantly shows a peak at this value.

Figure 4.4 shows the frequency of use of controls based on time, and indoor and outdoor temperatures. The column to the left displays AC use, while window use is shown to the right. As it can be seen, AC is mostly activated in the afternoon, while windows are mostly used in the morning, although both controls are used throughout the entire day (Figure 4.4a and Figure 4.4b). In relation to indoor temperature (Figure 4.4c and Figure 4.4d), both controls show higher frequency of use with values around 25 ^{o}C. However, windows are also used at temperatures below 20 ^{o}C, and the AC use below that temperature is nearly zero. As for outdoor temperature (Figure 4.4e and Figure 4.4f), it is possible to see that the AC has a higher frequency of activation with outdoor temperature above 25 ^{o}C, while

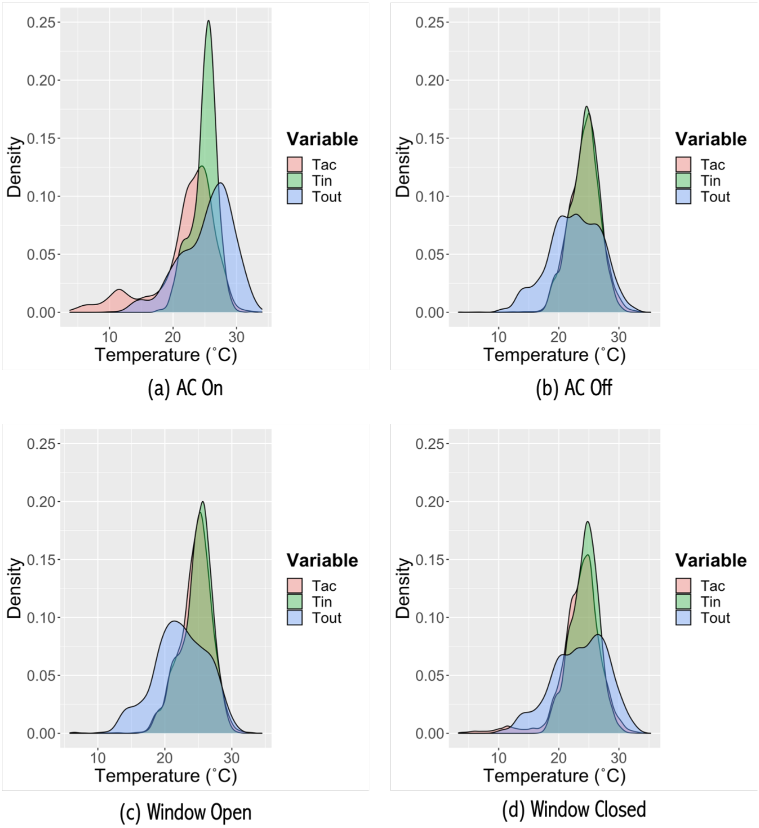


Figure 4.3: Density graphs of measured data for each control state

windows show their highest frequency of use with such temperatures below 25 ^{o}C, and a second peak of use above 25 ^{o}C, around 27 ^{o}C.

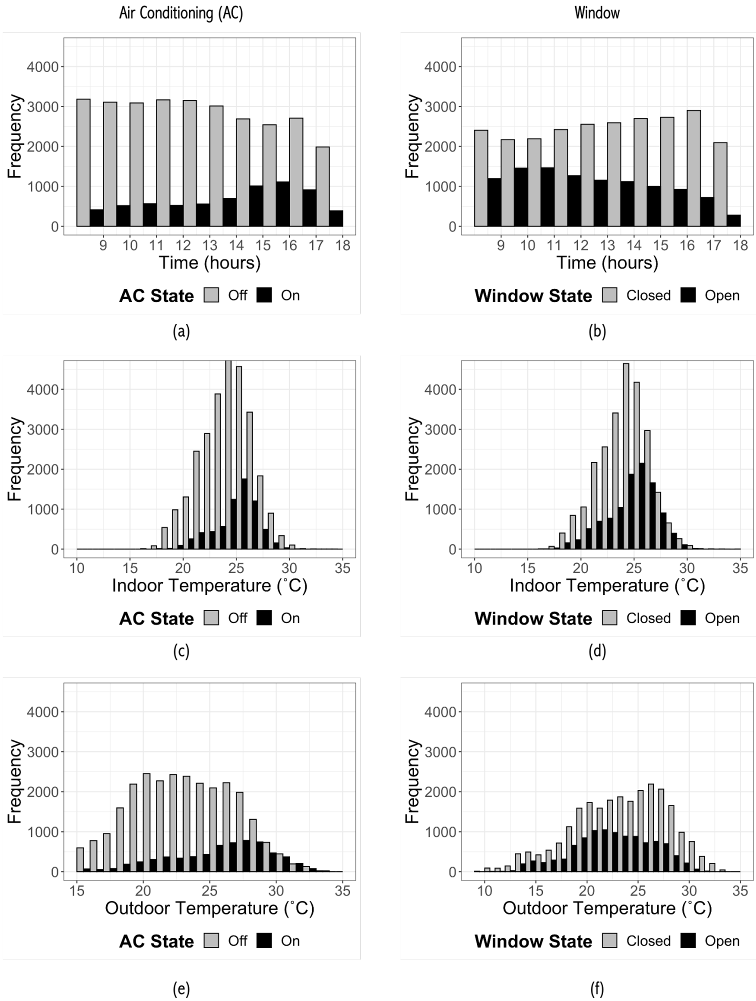


Figure 4.4: Frequency of use of the measured controls

4.1.2 Seasons

Figures 4.5 and 4.6 present the distribution of the indoor and outdoor temperatures by season for each of the studied actions. As it can be seen, there is a seasonal effect to the way the controls are operated. The AC is mostly operated with indoor temperature close to or above 25 ^{o}C, except during the winter, when it is operated at temperatures lower than 25 ^{o}C. Window operation is distributed along a wider range of indoor temperature, showing higher frequencies of use around 25 ^{o}C, except during the summer, when its highest frequency of use is around 27 ^{o}C.

When analyzing the use of controls and outdoor temperature in each season, the peaks of use of the AC are constantly seen around 27 ^{o}C, with the peak closer to 30 ^{o}C during the summer. As for window use, the highest frequencies of use shift with each season. During the summer, the highest frequencies of use are seen around 20 ^{o}C, which is also true for fall and spring. However, during fall there is more use of windows with outdoor temperatures below 20 ^{o}C, which is seen with less frequency during the summer and spring. During the winter, the highest frequencies of use are seen around 27 ^{o}C, showing a similar frequency as seen in the fall for values below 20 ^{o}C.

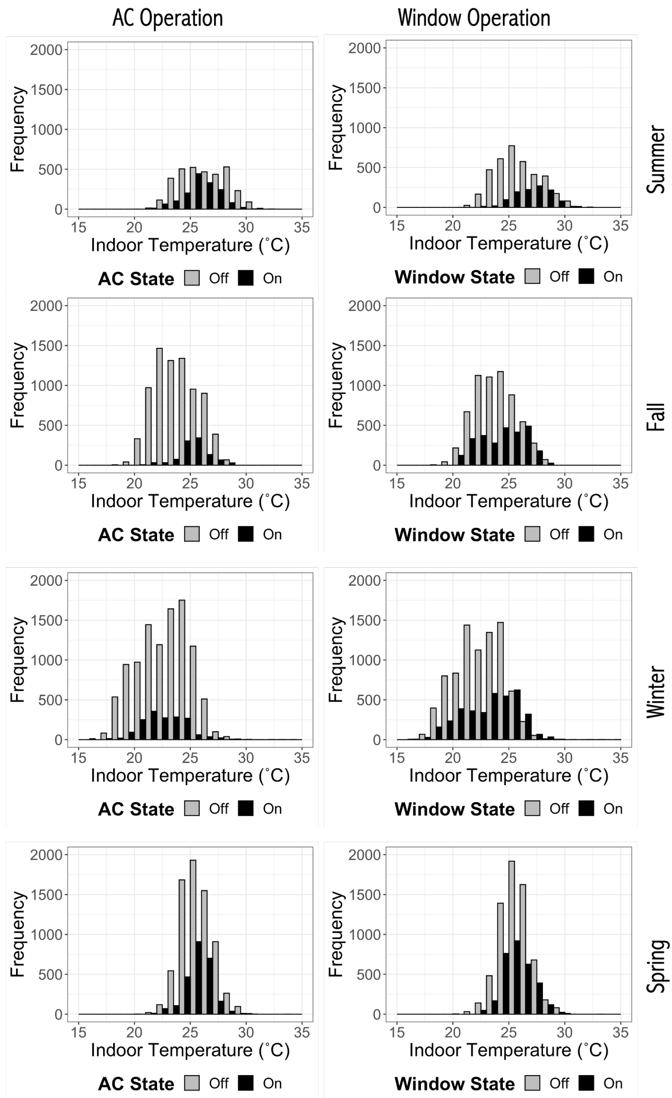


Figure 4.5: Distribution of the indoor temperature by season and operation mode

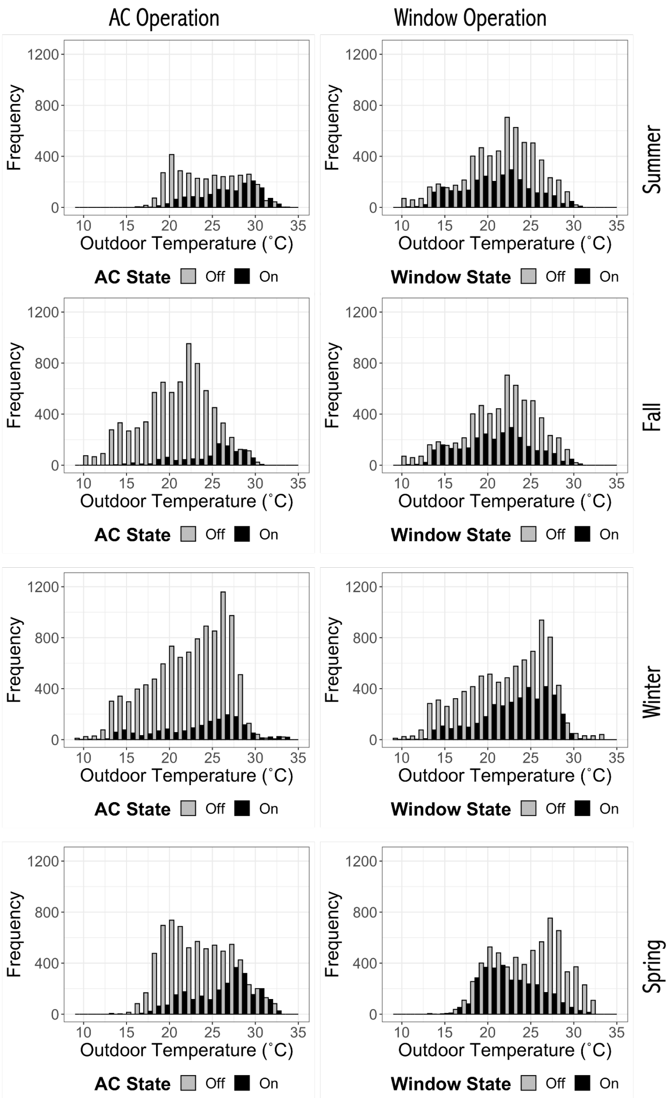


Figure 4.6: Distribution of the outdoor temperature by season and operation mode

4.1.3 Offices: general observations

For the offices analysis, Figure 4.7 shows the AC, indoor and outdoor temperatures ranges for each office, while Figures 4.8 and 4.9 show the temperatures ranges for each control state.

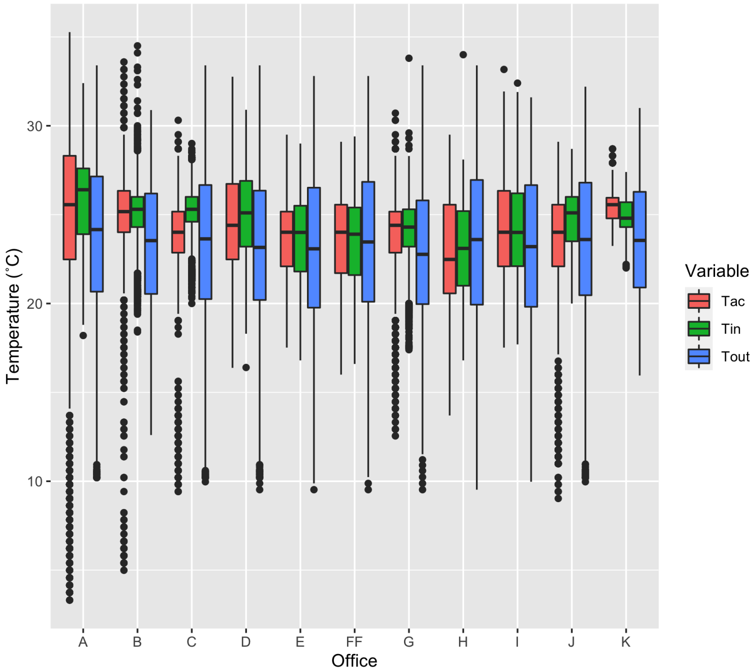
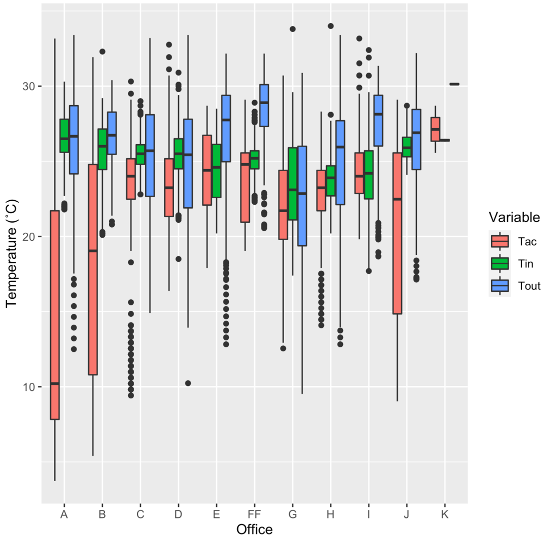


Figure 4.7: Temperature box plots of full data set

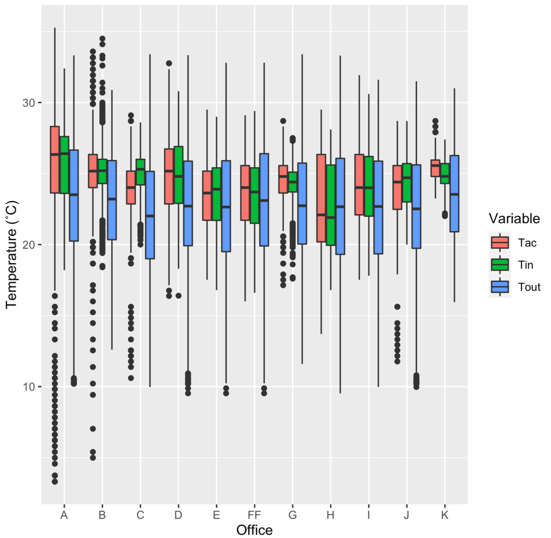
Figure 4.7 shows that even with different outdoor and AC temperature means, the indoor temperature was, for the most part, maintained around 25 °C. It can also be observed that outdoor temperature mean is around the same value for all offices, and yet their indoor, and especially AC temperature means vary, indicating the influence of different building properties, office characteristics, and occupants preferences and thus behaviors.

Figure 4.8 allows a more detailed insight on how the offices used the AC and maintained indoor temperatures around 25 °C. Figure 4.8a, showing the temperatures during the periods of AC activation, illustrates how differently each office used the AC in terms of the temperature they set. Offices A, B, G and J show the lowest AC temperatures, and yet, with the exception of office G, they all show indoor temperature means very similar to the other offices. In addition, all offices show AC activation with outdoor temperature means above 25 °C, once again with the exception of office G. As for the periods with the AC not in use, Figure 4.8b shows that the mean indoor temperature was below 25 °C, and the indoor temperature mean was maintained at, but mostly below, 25 °C, except for office A.

As for window use, Figure 4.9a shows that occupants opened windows with outdoor temperature means below 25 °C, mostly around 22 °C, which caused indoor temperature means to remain close to or below 25 °C. Figure 4.9b shows that windows were closed with outdoor temperature means approximately between 22 to 25 °C. In these cases, some offices show more clearly the use of AC, displaying AC temperature means lower than indoor temperature means. However, based on this plot, other offices may have (a) closed the windows and not used the AC, as AC temperature means are the same as indoor, or (b) used the AC at a temperature very close to the indoor temperature, thus displaying similar values.

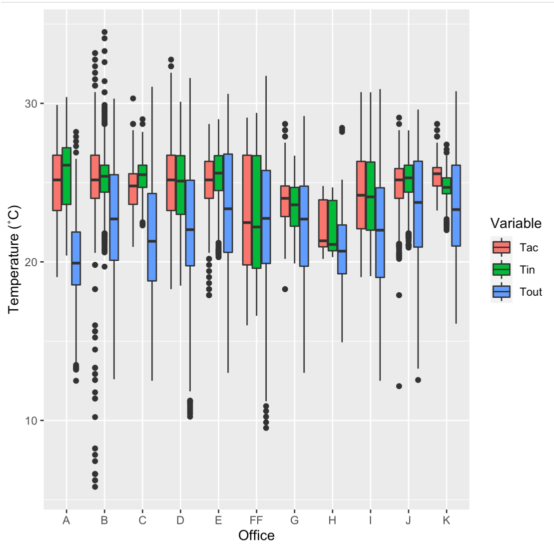


(a) AC On

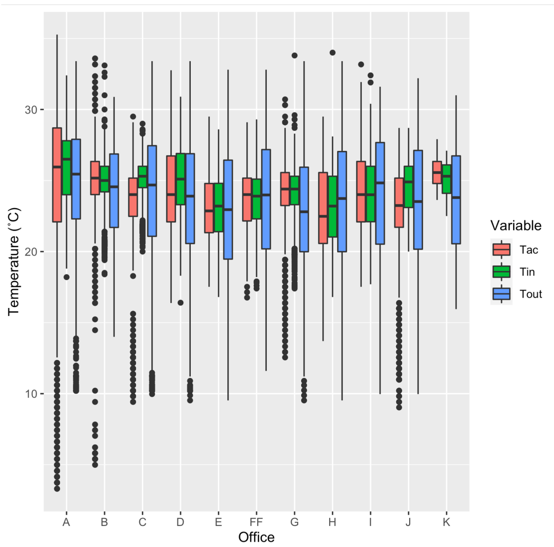


(b) AC Off

Figure 4.8: Temperature box plots of measured data for AC state in each office



(a) Window Open



(b) Window Closed

Figure 4.9: Temperature box plots of measured data for window state in each office

4.1.4 Offices: specific observations

Offices A, C and D: Offices within the same building. AC equipment in office A is an older window model and distant from occupants. Offices C and D have split models and the equipment is closer to the occupants. However, all three offices show indoor temperature means above 25 ^{o}C when the AC is activated (Figure 4.8a), which can indicate that the building orientation and materials may be playing a significant role in this building's thermal performance, thus affecting the way the studied controls were operated.

Office B: West facing balcony window. This feature may have influenced the occupant to use the AC because of high temperatures, but also because the occupant may have had to close the window to be able to use the blinds because of wind and glare.

Offices E, FF, and I : These offices, all occupied by one female (two female occupants in office FF in 2019), show AC use with the highest outdoor temperature means, and yet also the highest AC temperature means.

Office J: Indoor temperature mean above 25 ^{o}C even when the AC was in use. AC was activated every day, and some days there was only AC use and windows were not opened. AC always activated at a very low temperature, but indoor temperature did not decrease much. Possibly location of AC unit is not favorable to cooling down the entire office. Current location probably chosen due to window proximity and installation viability.

Office K: AC was not activated during the entire monitored period. Possible unusual occupant preference, AC unit location, health conditions, etc.

4.2 Generalized Linear Mixed Models (GLMM)

4.2.1 Validation: Confusion Matrix

The validation method applied to each study can vary according to the model obtained after applying the selected statistical method. As described in Section 3.6, 70% of the collected data was used to generate, or train, the models, leaving the remaining 30% to be used to test them, and thus establish their performance.

The models' results are given in percentages, that is, the probability of the desired outcome happening. Two GLMM models were generated, one that predicts the probability of AC activation and the other of window opening. The higher the percentage given by the model, the more likely it is that the action will take place. As described in Section 3.7, the probabilities were turned into classes to verify their frequencies. Probabilities of 40, 50 and 70% were calculated. This means that for the 40% probability, for example, if the model predicted a probability of 0.4 or above, this prediction was classified as a 1, otherwise as a 0. Confusion Matrices and their respective metrics were calculated for each probability for each model (Table 4.1). This test set contained a total of 10,598 observations.

As shown in Table 4.1 and Figure 4.10, both models predict more negatives than positives, that is, more zeros than ones. From those, the negatives are true much more than the positives are, meaning that the models are better at predicting when there will not be an action, instead of the opposite. Ideally, when looking at Figure 4.10, the highest numbers would be in a diagonal from the upper left corner to the lower right corner of each matrix, which would mean high numbers of True Negatives and True Positives. As shown, the higher the threshold (higher probability), or when the specificity is increased, the lower the sensitivity, and the result is more negative values predicted. Because sensitivity and specificity

Table 4.1: WO and ACA models performance calculated using Confusion Matrix

Probabilities	**40%**		**50%**		**70%**	
Models	**WO**	**ACA**	**WO**	**ACA**	**WO**	**ACA**
Predictions (0)	7 380	8 980	8 741	9 511	10 298	10 200
Predictions (1)	3 218	1 618	1 857	1 087	300	398
Precision	.58	.58	.66	.62	.75	.67
Recall	.59	.49	.39	.35	.07	.14
Accuracy	.75	.84	.76	.84	.72	.83
F-score	.58	.53	.49	.45	.13	.23

are inversely proportional, when the threshold is lower, in this case at 40%, more positive values are predicted.

Overall, the accuracy for the WO model remains around 70%, and for the ACA around 80%. However, the higher the threshold for the classes, the lower their F-scores are. From 40 to 50% there is a decrease of about 0.1 in the F-scores. But when the threshold is increased from 50 to 70%, the values decrease by 0.36 and 0.22, for WO and ACA, respectively. The accuracy rates remain around 0.7 and 0.8 for WO and ACA, respectively, in all 3 instances (40, 50 and 70%), mostly due to the fact that they predict a good amount of True Negatives, when, preferably, there would be high rates of True Positives as well.

4.2.2 Validation: AUROC Curve

The AUROC Curve plots the model's *Specificity* (x-axis) against its *Sensitivity* (y-axis). The AUC shows how well the model is at predicting zeros as zeros and ones as ones. Figure 4.11 shows the curves for both models. The AUC for the ACA model is of 0.84, while for the WO model it is slightly lower, with a value of 0.78. The AUC values show that both models can correctly distinguish between the classes around 80% of the time.

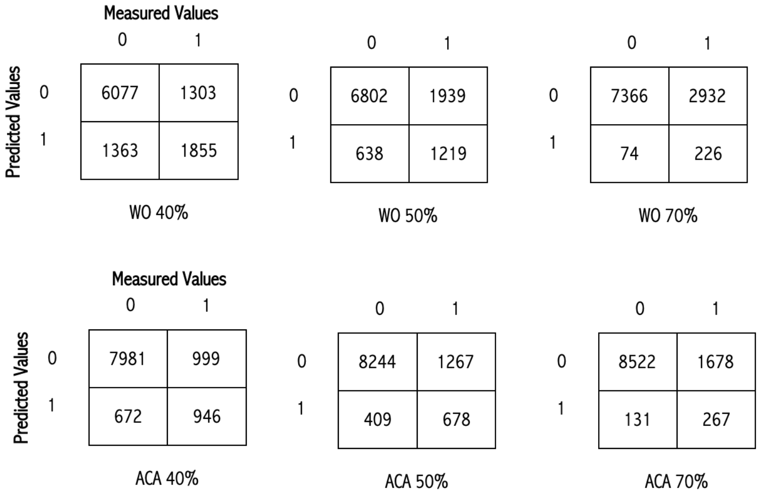


Figure 4.10: Confusion matrices of models predictions

Although the AUC values for both models are around 0.8, which is close to the ideal value of 1, it is important to highlight that a large portion of the classes correctly classified is of zeros, as shown in Section 4.2.1. Therefore, there are improvements that can be made to the models to achieve higher AUC values, or even maintain them around 0.8, but correctly classifying a larger portion of ones.

4.2.3 Results of GLMM

Occupants manually controlled windows and the AC based on their preferences and needs. Tables 4.2 and 4.3 show the GLMM models for the operation of such controls.

Hour of day and routine are variables known to have an impact on occupant behavior. Studies have shown that occupants tend to operate controls upon arrival and departure [55,58,152], thus the respective hours of day corresponding to such

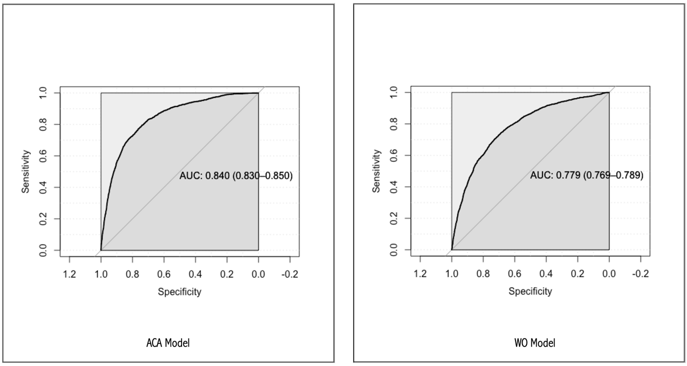


Figure 4.11: AUROC Curves and confidence intervals for ACA and WO models (GLMM)

actions show a higher frequency of operation. The models developed in this work show behavior consistent with what is reported in the literature.

Figure 4.12 shows that high probability of control use, for both ACA and WO, is in the morning around 8 a.m., which is associated with arrival. The lowest probability for ACA is at 1 p.m., around lunch time, as reported by occupants (Figure 4.12 (a) and (b)). WO has the highest probability of use during this hour, which can be related to occupants turning the AC off and leaving windows open during the lunch break, while the office is either not occupied or with lower occupancy. This behavior can also indicate an energy saving attitude (Figure 4.12 (c) and (d)). As the day progresses, WO probabilities are lower and ACA probabilities increase. This is due to the increase in outdoor temperature during the afternoon, and consequently of indoor temperature as well. At the end of the day WO probabilities reach zero, meaning that they are closed, increasing the probabilities of ACA. Because the models were generated using the data related to the working hours, the models do not take into account the later hours when both controls were deactivated, hence the probabilities of WO being low and ACA high. The models are shown as inverses of each other, demonstrating that occupants alternate between controls. The pattern displayed by the probabilities of control use based on time also depict an occupancy pattern, seen more clearly with the ACA model. The pattern seen in Figure 4.12 shows a higher probability of use

Table 4.2: Generalized Linear Mixed Model for Air-Conditioning Activation (ACA)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-2.20	$p < 0.001$
Time 1	0.21	$p < 0.001$
Time 2	0.93	$p < 0.001$
Time 3	-0.21	$p < 0.001$
Time 4	-0.36	$p < 0.001$
Indoor Temperature 1	-0.50	$p < 0.001$
Indoor Temperature 2	-0.01	.55
Indoor Temperature 3	-0.003	.71
Outdoor Temperature 1	1.85	$p < 0.001$
Outdoor Temperature 2	0.01	.45
Outdoor Temperature 3	-0.10	$p < 0.001$
Indoor Rel. Humidity	-0.62	$p < 0.001$
Outdoor Rel. Humidity	1.27	$p < 0.001$
Window State	-1.00	$p < 0.001$

in the beginning of the day, when the office is occupied, the lowest between the reported lunch break (12-2 p.m.), when it is expected to be empty or with lower occupancy, and then an increase during the afternoon, when occupants are once again in the office. In relation to time and temperature, during the entire day, indoor temperatures at 24 ^{o}C shows the lowest activation probabilities for both controls. As for the outdoor temperature, the lowest probabilities for WO are seen at 23 ^{o}C, but for ACA they are at 32 ^{o}C.

Indoor and outdoor temperatures are considered the main drivers for occupants' actions, such as window opening [55]. Lower indoor temperatures present higher probabilities of use for both controls (Figure 4.13). For the ACA, Figure 4.13a shows the highest probabilities of use when indoor temperature is at its highest and lowest values, with the lowest value's probability (19 ^{o}C) decreasing as outdoor temperature increases. The highest probability at the lowest indoor temperature

Table 4.3: Generalized Linear Mixed Model for Window Opening (WO)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.63	$p < 0.001$
Time 1	0.03	.45
Time 2	0.51	$p < 0.001$
Time 3	-0.11	$p < 0.001$
Time 4	-0.30	$p < 0.001$
Indoor Temperature 1	-0.54	$p < 0.001$
Indoor Temperature 2	-0.10	$p < 0.001$
Indoor Temperature 3	-0.03	$p < 0.001$
Outdoor Temperature 1	-0.56	$p < 0.001$
Outdoor Temperature 2	-0.24	$p < 0.001$
Outdoor Temperature 3	-0.02	.17
Outdoor Rel. Humidity	-0.15	$p < 0.001$
AC State	-1.35	$p < 0.001$

is a result of how this type of model is built. The models use indoor temperature to predict the desired outcomes. However, this same variable is affected by the use of the AC, and the models capture both effects. The same can be considered for indoor temperatures at 22 °C and 26 °C, with almost identical probabilities of ACA. Indoor temperature at 22 °C can be a result of the AC already being in use, and at 26 °C occupants activate the AC. The lowest probabilities are seen for indoor temperature at 24 °C, which can be considered a temperature at which occupants are most comfortable, not needing to change it.

Indoor temperatures ranging from 22 to 29 °C show a peak in ACA probability with outdoor temperature at 20 °C, followed by a decrease. This peak in the probability of ACA with outdoor temperatures between 18 and 23 °C, can be related to arrival as well. As it can be seen in Figure 4.14, most of the instances with outdoor temperature between the aforementioned range can be seen in the morning (between 500 and 650 minutes, or 8 a.m. and 11 a.m.), where indoor

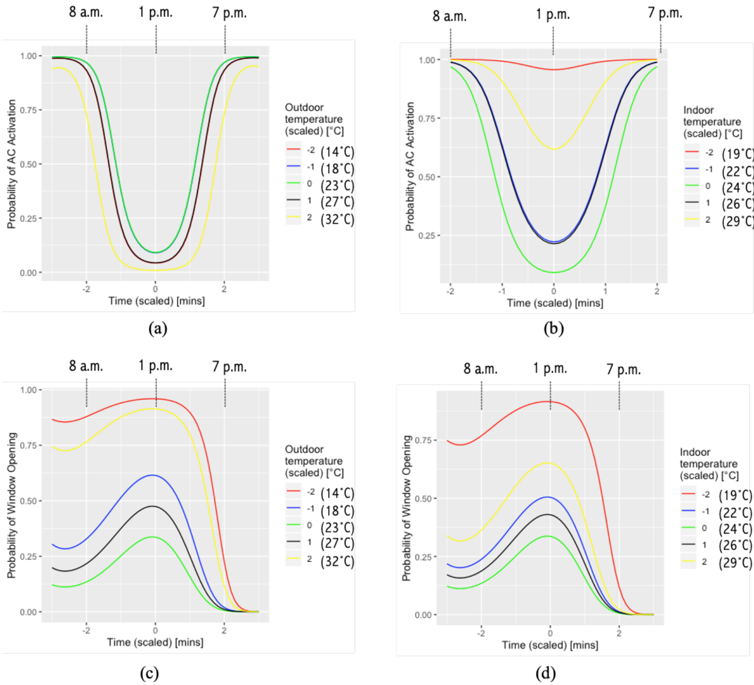


Figure 4.12: Models' predictions based on Time and Temperature

temperatures vary around 22 and 30 °C. Therefore, the probabilities observed in Figure 4.13a can be a result of the combination between indoor and outdoor temperatures.

When looking at the outdoor temperature, probabilities decrease as outdoor temperature increases after showing a peak. Figure 4.13a depicts this behavior, which is likely a result of how occupants operate the AC. Figure 4.14 illustrates how values of outdoor temperature above 30 °C are seen after 1 p.m. (800 minutes), with the highest values more concentrated towards the end of the work day. In agreement with the measured data, the AC is deactivated some time before the hour of departure, and not exactly at the moment of departure, as can happen with windows. This is so because once the AC is deactivated, indoor temperatures are

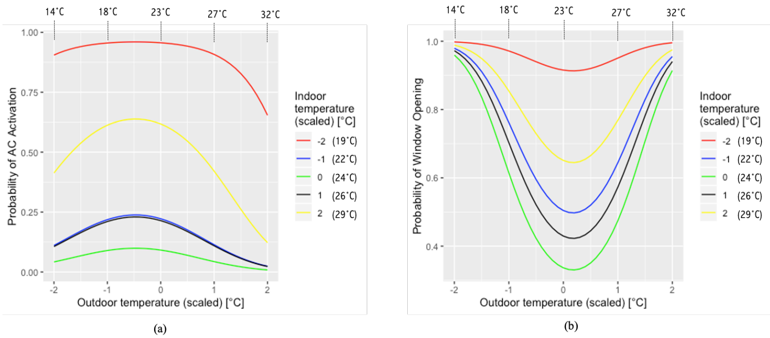


Figure 4.13: Models' predictions based on indoor and outdoor temperatures

still maintained at acceptable levels, making it possible to deactivate it some time before departure. In addition, it can be a preventive measure taken by occupants, to not forget the AC activated when they leave.

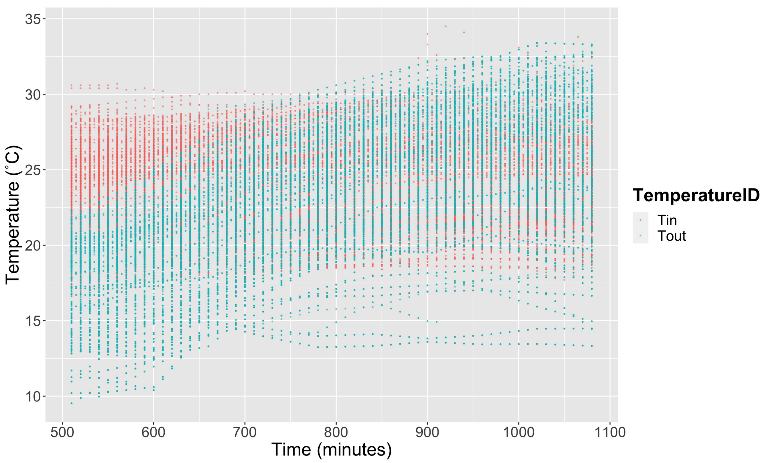


Figure 4.14: Measured Indoor and Outdoor temperatures' progressions during the work hours

When analyzing WO, once again the behavior pattern shown by the models are inverses of each other. Again, the highest probabilities are seen with the highest and lowest values of indoor temperature. High probabilities of WO with indoor temperature around 20 ^{o}C indicates that occupants open windows to allow the warmer outdoor air to enter the environment. As for when indoor temperature is around 28 ^{o}C, the same logic is followed, though inverted; thus allowing the cooler air inside. This behavior also suggests that at higher indoor temperatures, occupants may desire some air movement, thus allowing natural ventilation with open windows. Lastly, indoor temperature at 24 ^{o}C shows the lowest probabilities of WO, as it does for ACA, once again showing that this value may be associated with a comfort temperature for occupants.

In response to the limitations found in the GLMMs, the following sections report the results of the additional methods implemented in this work in an effort to create more precise models.

4.3 Markov Chain Models

As mentioned in Section 4.2.3, the GLMMs show high probabilities of control activation with low indoor temperatures, indicating when the AC is already on. Because this is a result of how the models were built, a second approach to building the models was taken; Markov chain (MC). The first models built using MC showed that the models almost always predicted an outcome of zeros, due to the greater amount of zeros in the data set, configuring the data as imbalanced. Table 4.4 shows the resulting amount of zeros and ones for each control state after the data treatment for the Markov chain models was performed, as described in Section 3.5.2.1. The following sections, 4.3.1 and 4.3.2, present the results of the additional methods applied to increase the MC models' accuracy, Penalty Factor and Synthetic Data, respectively.

For the validation of such models, *train* and *test* data sets were created by applying the method as described in Section 3.6. The AUC, calculated for each model using the *test* set, was used as the metric to validate the models and indicate their accuracy. These values are reported along with the results of the MC models in their respective sections.

Table 4.4: Data sets' classes distribution

	0	**1**
Window Open	0.98	0.02
Window Closed	0.96	0.04
AC On	0.99	0.01
AC Off	0.94	0.06

4.3.1 Markov Chain models with Penalty Factor

As a way to address the imbalanced data, a Penalty Factor was applied to each model. Table 4.5 presents the values of each Penalty Factor calculated and applied, as well as the AUC values of each model resulting from the validation procedure.

There was a slight difference in probabilities of the predicted outcomes after applying the penalty factors for each model, but because the difference was very small, results only for the MC models with Penalty Factor are presented. Such models still display very low probabilities of prediction and still predict more zeros than ones. There was no improvement on the AUC values, therefore only the AUC for the MC models with Penalty Factor are reported. Nonetheless, the behavior pattern shown in each instance is in agreement with the measured data, and with the previously built models, with some improvements.

All the MC models were built using time and its polynomials to the 4 th degree, indoor and outdoor temperatures and their polynomials to the 3 rd degree, and indoor and outdoor relative humidity. Figure 4.15 shows each models' predictions

Table 4.5: MC models and Penalty factors

Model	Penalty Factor	AUC
Window Open	1	0.74
Window Closed	8	0.70
AC On	2	0.72
AC Off	4	0.64

based on time, indoor and outdoor temperatures. The predicted outcome of each model is the probability of a transition occurring, hence the four different models, as opposed to the GLMMs, where each model predicted the probability of control activation.

When analyzing the predictions made based on time, also during the reported work hours (8 a.m. to 6 p.m.), the behaviors seen are closely related to routine, as the GLMMs also showed. Even though the predicted probabilities for AC On are low, the curve shows a pattern of higher activation before 9 a.m., which is in agreement with the reported time of arrival, around 8 a.m. This is followed by a drop during the morning period, later showing an increase, from 1 p.m. until around 5:30 p.m., 30 minutes before departure time. The period displaying higher probabilities are coherent with the measured data, which showed the highest activation frequency of AC during the afternoon. Around 5:30 p.m. the probability decreases, meaning the AC is deactivated when occupants are close to leaving the office. This is confirmed when observing the AC Off model, as it displays the inverse of the above described behavior. The AC Off model shows higher probabilities during the morning period, and its highest probabilities around 5:30 p.m.

As for the window models, they both show their respective higher probabilities around 5:30 p.m. For the Window Closed (WC) model, this is related to the hour of departure, as occupants close windows to leave the office. When looking at the Window Open (WO) model, this behavior can reflect occupants deactivating the AC around the same time, and thus opening windows, later closing them upon departure. The second highest probability of Window Opening is in the morning, before 9 a.m., related to arrival, consequently causing the WC model to display its lowest probabilities at the same time. The Window Open model shows a slight increase in probabilities between 12 and 1 p.m., which is similar to the results seen for the GLMM WO model.

When analysing the models predictions based on temperature, the AC On model shows an improvement in comparison to the GLMM ACA model in relation to correctly capturing the desired effects. The AC On model displays the lowest probabilities with the lowest values of indoor temperature, showing that it correctly captures the temperature of AC activation, that is with higher indoor temperatures, beginning to rise around 26 ^{o}C, and not including the instances when the AC was already on. For the outdoor temperature, the model predicts higher probabilities for the lower temperature value, and then again for the highest values. The high probability of transition with the lowest outdoor temperature value can also be related to arrival, as the lowest outdoor temperatures were mostly measured during the morning, as seen in Figure 4.14 with the measured temperatures fluctuations throughout the day. As for the AC Off model, it correctly predicts higher probabilities of AC Off with lower indoor temperatures, displaying the opposite for outdoor temperatures.

For the windows, the predictions for Window Open are higher with higher indoor temperatures, and the opposite for outdoor temperatures, considering the range of probabilities seen for this model. Higher probabilities of window opening with higher indoor temperatures and lower outdoor temperatures is an effect that was also captured by the GLMM WO model (Figure 4.13b). However, the GLMM WO model shows high probabilities of control use with the lowest indoor temperature, which is not shown by the MC Window Open model. The MC Window Closed model shows higher probabilities for higher indoor temperatures, in agreement with the AC On, also evidencing that occupants tend to alternate between controls, once again, just as the GLMMs also showed. As for the outdoor temperature, predictions for all values are very low, only showing a slightly higher probability for the lowest values. The fact that window closing probabilities are low for all outdoor values can indicate that this variable does not have a strong effect on this action.

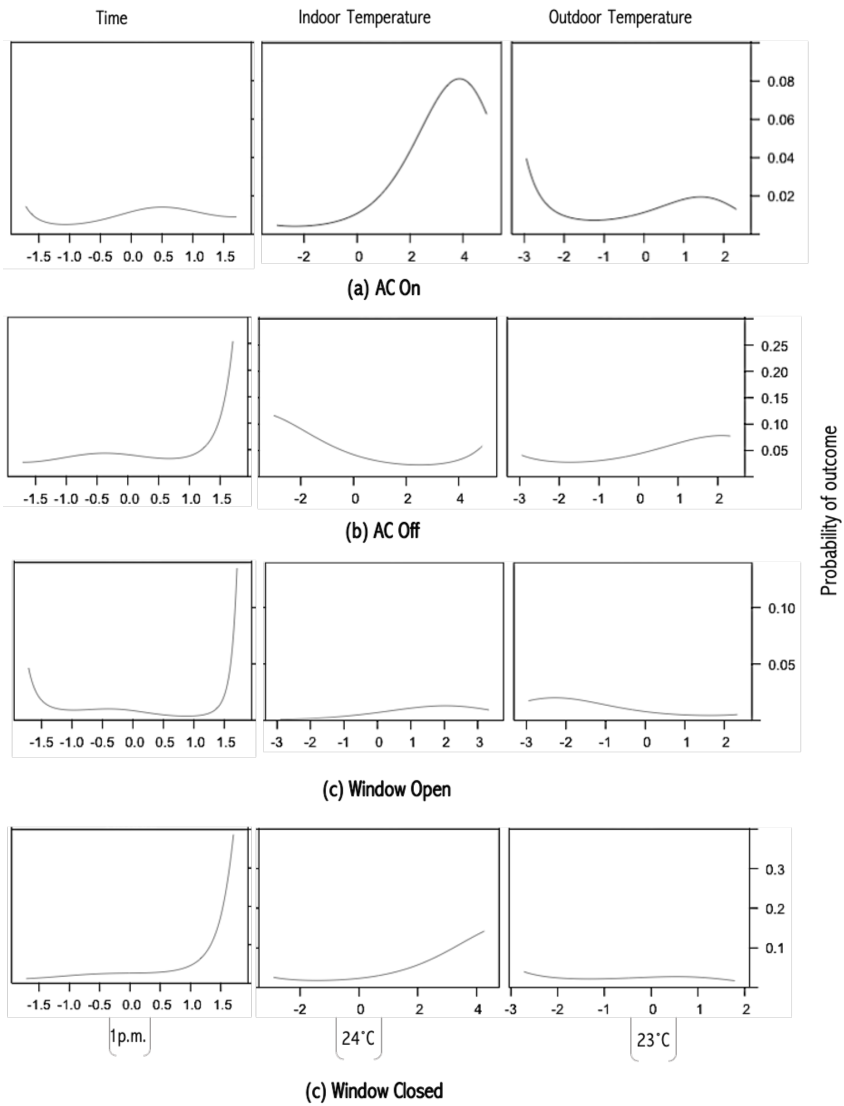


Figure 4.15: Predictions of Markov Chain Models with Penalty Factor

4.3.2 Markov Chain ROSE Models

The results for the MC models with penalty factor presented in the previous section showed that MC is a method that can address the limitation found in the GLMMs. However, even after applying the penalty factor, AUC values were still low, as were the prediction probabilities given by the models.

This section presents MC models created using newly generated data sets for each MC model by using four different approaches available with the ROSE RStudio package [81] (Tables 4.6-4.9). The new data sets were used to generate the MC ROSE models. Over sampling, under sampling, both of these methods combined and the creation of synthetic data were implemented and their respective results analyzed in comparison with each of the four models previously generated.

For Window Opening (WO), Window Closing (WC), and ACon, the method *both* showed the best results, while the ACoff model had best results with the oversampling method. To determine the best outcome of the four tested methods for each model, both their AUC and accuracy values were considered. Table 4.10 shows such values for each model and the selected ones are shown in boldface type, while Figure 4.16 shows the Confusion Matrices for the original models and for the MC ROSE selected ones. These confusion matrices were calculated with the threshold set at 0.9. The selection of the best performing ROSE model was based on comparisons made among the four newly generated models (only ROSE models), and not the ROSE models against the original ones. When selecting the WO model, the *over* and *both* methods yielded the same results. In this case, the threshold for the predictions was increased from 0.9 to 0.99, and the accuracy of the model using *both* had a slight increase in relation to the other model, and was therefore selected.

Even though the AUC values show small improvements when compared to the values of the original models, and in some occasions with lower accuracy, the advantage of the ROSE generated models is that they predict ones, which the original models do not. Despite the new models not having a high True Positive

Table 4.6: Markov Chain Rose Model for Window Opening (WO)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.47	$p < 0.001$
Time 1	-0.22	$p < 0.001$
Time 2	-0.71	$p < 0.001$
Time 3	0.12	$p < 0.001$
Time 4	0.48	$p < 0.001$
Indoor Temperature 1	0.35	$p < 0.001$
Indoor Temperature 2	-0.12	$p < 0.001$
Indoor Temperature 3	0.01	.05
Outdoor Temperature 1	-0.66	$p < 0.001$
Outdoor Temperature 2	-0.002	.92
Outdoor Temperature 3	-0.05	$p < 0.001$
Indoor Rel. Humidity	0.33	$p < 0.001$
Outdoor Rel. Humidity	-0.36	$p < 0.001$
R^2=0.28		

rate (TPR), they predict differently from the original models, and therefore were also implemented and used in the EnergyPlus simulations.

Even with a low TPR, this procedure showed that given a more balanced set, the models are able to correctly predict both classes. Ideally, this correction in the data sets' imbalance would be addressed by a more extensive data collection. Since this option was not available, the ROSE methods were applied.

In addition, as Figure 4.17 shows, the ROSE generated models present the same patterns of prediction as the previously presented GLMM and MC models (Sections 4.2 and 4.3, respectively). The difference with the ROSE models is that their predicted probabilities for each outcome are higher than those of the Markov Chain models with the Penalty Factor.

Table 4.7: Markov Chain Rose Model for Window Closing (WC)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.44	$p < 0.001$
Time 1	-0.08	.24
Time 2	0.02	.86
Time 3	0.44	$p < 0.001$
Time 4	0.15	.0003
Indoor Temperature 1	0.05	.43
Indoor Temperature 2	0.10	$p < 0.001$
Indoor Temperature 3	0.03	.02
Outdoor Temperature 1	0.60	$p < 0.001$
Outdoor Temperature 2	-0.07	.04
Outdoor Temperature 3	-0.1	$p < 0.001$
Indoor Rel. Humidity	-0.50	$p < 0.001$
Outdoor Rel. Humidity	0.81	$p < 0.001$
R^2=0.21		

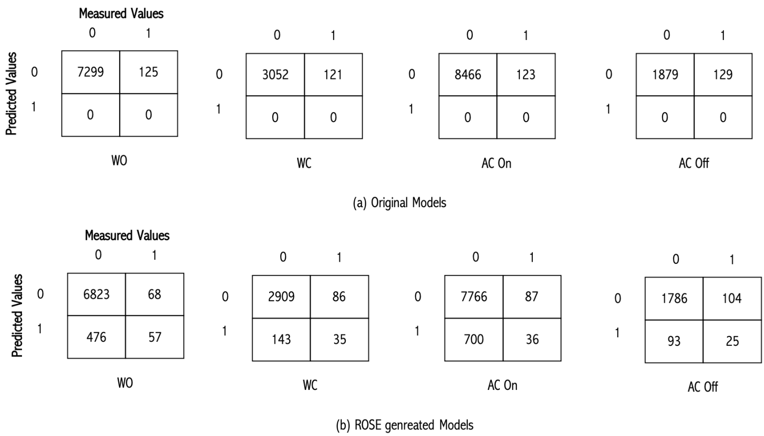


Figure 4.16: Confusion Matrices of ROSE methods

Table 4.8: Markov Chain Rose Model for Ac On (ACon)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.19	$p < 0.001$
Time 1	0.69	$p < 0.001$
Time 2	-0.73	$p < 0.001$
Time 3	-0.27	$p < 0.001$
Time 4	0.23	$p < 0.001$
Indoor Temperature 1	0.72	$p < 0.001$
Indoor Temperature 2	0.09	$p < 0.001$
Indoor Temperature 3	-0.03	$p < 0.001$
Outdoor Temperature 1	0.53	$p < 0.001$
Outdoor Temperature 2	0.06	$p < 0.001$
Outdoor Temperature 3	-0.09	$p < 0.001$
Indoor Rel. Humidity	-0.18	$p < 0.001$
Outdoor Rel. Humidity	0.36	$p < 0.001$
R^2=0.23		

Table 4.9: Markov Chain Rose Model for AC Off (ACoff)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.33	$p < 0.001$
Time 1	-0.35	$p < 0.001$
Time 2	-0.16	.08
Time 3	0.40	$p < 0.001$
Time 4	0.14	$p < 0.001$
Indoor Temperature 1	-0.43	$p < 0.001$
Indoor Temperature 2	0.002	.94
Indoor Temperature 3	0.04	.0003
Outdoor Temperature 1	0.30	$p < 0.001$
Outdoor Temperature 2	0.04	.06
Outdoor Temperature 3	-0.05	.001
Indoor Rel. Humidity	-0.01	.70
Outdoor Rel. Humidity	0.17	.0002
R^2=0.14		

Table 4.10: ROSE Models AUC and Accuracy values

	AUC	**Acc.**	**AUC**	**Acc.**
	WO		**WC**	
Original	0.74	0.98	0.70	0.96
Over	0.74	0.92	0.70	0.93
Under	0.74	0.88	0.70	0.85
Both	**0.74**	**0.92**	**0.71**	**0.93**
Rose	0.73	0.98	0.69	0.95
	AC On		**Ac Off**	
Original	0.72	0.99	0.64	0.94
Over	0.73	0.90	**0.65**	**0.90**
Under	0.73	0.88	0.63	0.91
Both	**0.73**	**0.91**	0.64	0.90
Rose	0.72	0.93	0.63	0.93

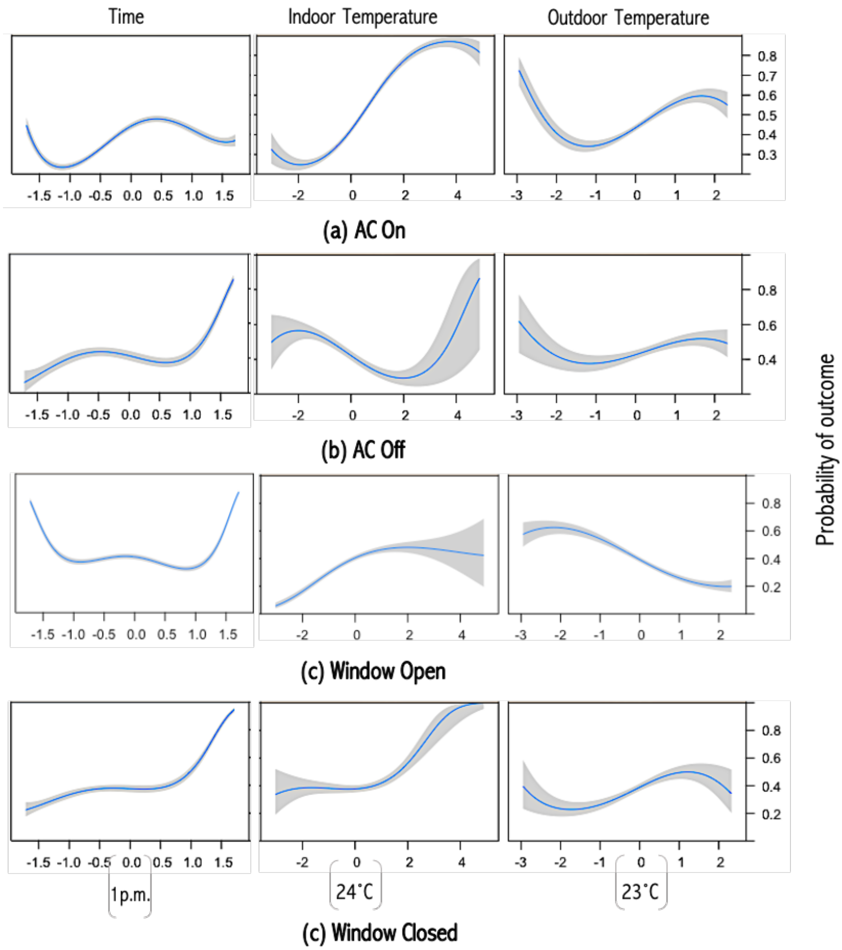


Figure 4.17: Predictions of Markov Chain Models with ROSE methods

4.4 Simulation tests in EnergyPlus

4.4.1 Overview of the different simulations scenarios outputs

A total of seven scenarios with varied models of occupant behavior for mixed-mode offices were simulated in EnergyPlus (Figure 4.18). In addition, the fixed schedules scenario, for means of comparison, was also simulated. The initial analyzed units of comparison were indoor temperature and cooling energy loads. Indoor temperature was also used to compare the simulated outcomes to the measured data. In a second step of the analysis, the use of the controls, here specified as behavior graphs, was analyzed for the scenarios that showed results closer to the measured data.

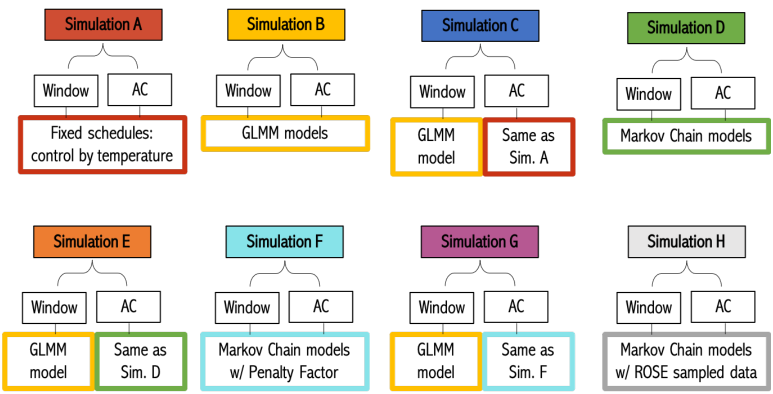


Figure 4.18: Simulation scenarios: models combinations

Figure 4.19 shows the average values for indoor temperature and cooling energy consumption of each simulated scenario, while Table 4.11 shows the values for the mean, standard error and standard deviation of each scenario. Figure 4.20 shows the indoor temperature amplitude for each simulation scenario, while Figure 4.21 shows the different indoor/outdoor temperature distributions for all scenarios. It was not possible to compare cooling energy consumption levels to the measured data, as this variable was not monitored in the campaign.

Each scenario presented different energy consumption levels, and average indoor temperature, even though in some cases the difference was of 0.1 °C (Table 4.11). The measured average outdoor temperature was the same as the simulated one. All simulations used the same air conditioning and natural ventilation parameters as those established by [117]. The difference from one scenario to another was the way these controls were operated, but their characteristics remained the same. All reported values were considering the occupied period, from 8 a.m. to 6 p.m.

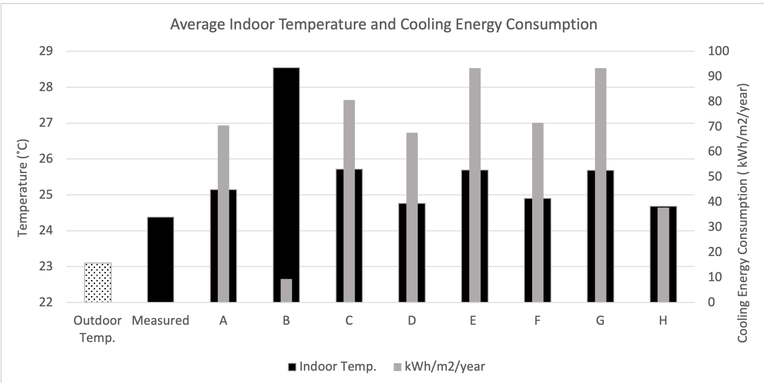


Figure 4.19: Average indoor temperature and cooling energy consumption of measured and simulated scenarios

Simulation A, with the fixed schedules, presents a higher indoor temperature average than the measured value, as well as higher levels of energy consumption when compared to *Simulations D* and *H*, which present lower indoor temperature averages than *Simulation A*. *Simulation A* also presents very similar levels of energy consumption as *Simulation F*, although the average indoor temperature for the fixed schedules scenario is slightly higher.

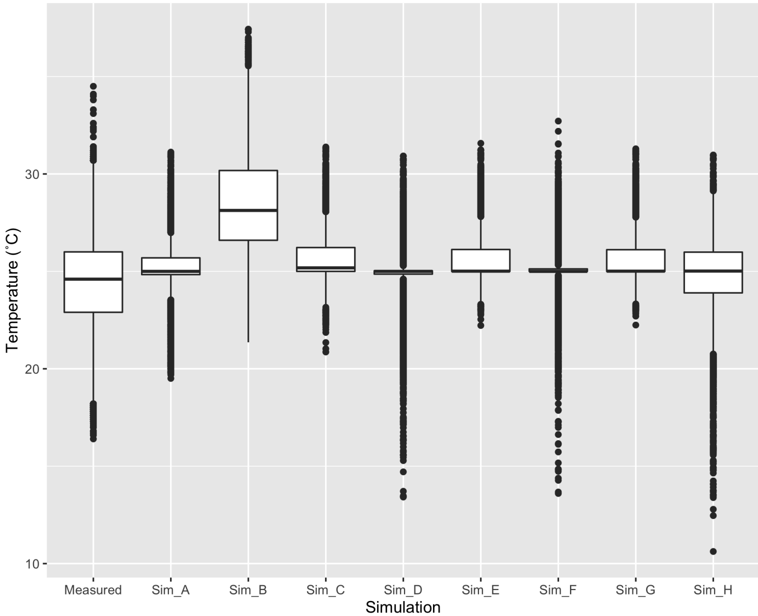


Figure 4.20: Box plots of indoor temperature for each simulation scenario

Simulation B shows the highest indoor temperature values, and consequently the lowest values for energy consumption. *Simulations C, E* and *G*, where the GLMM window model was combined with different options of AC models or control, all displayed indoor temperature averages above 25 ^{o}C, as well as the highest levels of energy consumption.

As it can be seen, *Simulations D* and *H* are the scenarios with the closest indoor temperature average values to the measured data, followed by *Simulation F*. However, their energy consumption levels differ. In terms of maintaining indoor temperature averages closer to the measured value, all three types of Markov models performed satisfactorily. As for energy consumption, values varied, with lower values seen with *Simulation H*. As a way to verify which scenario best represented the measured data, excerpts from summer and winter were taken from each scenario and compared to each other and the measured data (Section 4.4.3.)

Table 4.11: Mean, Standard Error and Standard Deviation of indoor temperature and cooling energy consumption of measured data and simulated scenarios

	Tin			**kWh**		
Simulation	Mean	SE	SD	Mean	SE	SD
Measured	24.4	0.012	2.35	-	-	-
A	25.1	0.019	1.78	0.63	0.0076	0.71
B	28.5	0.030	2.81	0.08	0.0027	0.25
C	25.7	0.015	1.38	0.72	0.0069	0.65
D	24.8	0.020	1.88	0.61	0.0065	0.61
E	25.7	0.014	1.33	0.84	0.0073	0.69
F	24.9	0.025	2.34	0.64	0.0067	0.63
G	25.7	0.014	1.32	0.84	0.0073	0.69
H	24.7	0.026	2.43	0.34	0.0050	0.47

The following section brings further analysis and discussion of the differences between the scenarios.

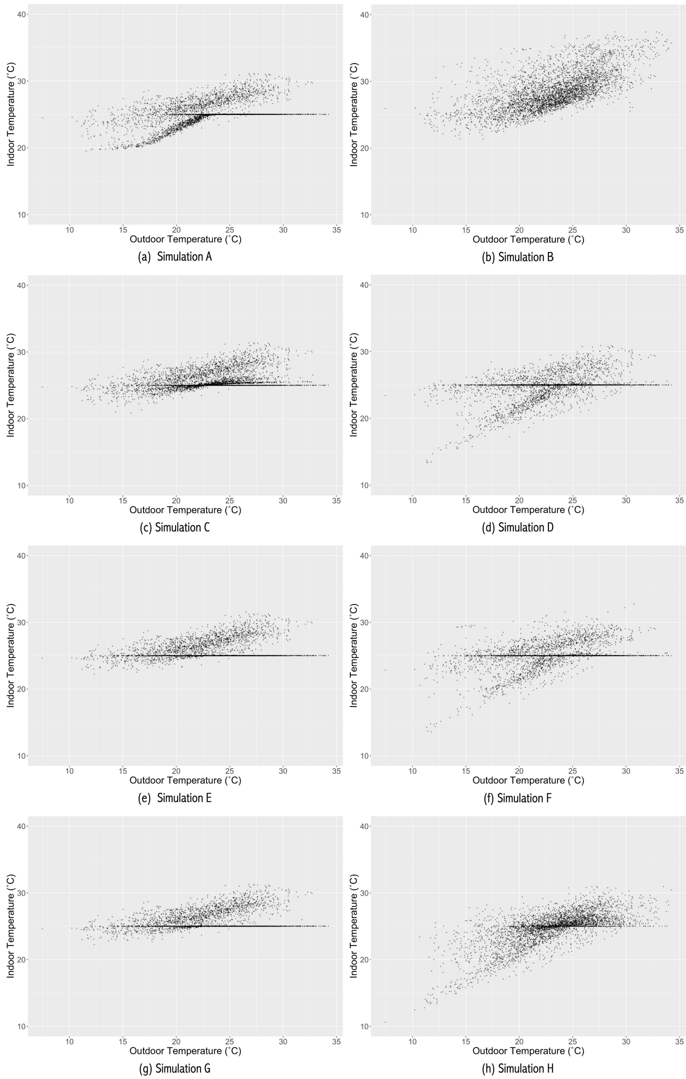


Figure 4.21: Scatterplots of indoor and outdoor temperatures for all simulation scenarios

4.4.2 Simulation Scenarios

Simulation A presented an average indoor temperature of 25.1 ^{o}C, which is 0.7 ^{o}C above the measured average, and 0.63 kWh average related to the cooling energy consumed. However, because this model alternates between the studied controls based on temperature while it attempts to represent occupant behavior, it does not precisely capture how occupants behave, since it is known that occupants may tolerate higher or lower temperatures than the fixed ones. This is illustrated with Figure 4.22, which shows the frequency of use of both controls in the measured data, while Figure 4.23 shows the frequency of use of the controls as allowed by the schedules in this simulation scenario. In addition, mixed-mode offices of the same type as the one simulated, and as the ones measured, do not have any temperature displays, nor any type of system to inform occupants of the indoor and/or outdoor temperatures, meaning that occupants do not operate the controls at specific temperatures, as established in the schedules of this scenario, but rather on how hot or cold it feels to them.

Simulation B, using both models generated using the GLMM method, presented the highest indoor temperatures among the simulated scenarios, ranging from 21 to 37 ^{o}C, with an average of 28.5 ^{o}C, and the lowest cooling energy consumption. By observing the measured data (Figure 4.22), it is possible to see a higher frequency of ACA with indoor temperature ranging from approximately 25 to 27 ^{o}C, and outdoor temperatures from 25 to 30 ^{o}C. However, *Simulation B* presented very little use of the AC, which is not coherent with the measured data, especially during the warmer seasons. As stated in Section 4.2, and shown by the outputs from *Simulation B*, this ACA model gives the highest probabilities of AC use when the indoor temperature is at 19 ^{o}C, which is a result of the AC already being on, and the second highest probability given with indoor temperature at 29 ^{o}C. Given the indoor temperature range in this scenario, there were no instances with indoor temperature at 19 ^{o}C in order for the model to predict AC use with these values. Also, the highest frequency of indoor temperature was 27.5 ^{o}C, and the GLMM ACA model has its second highest probabilities for indoor temperature at

29 ^{o}C, also contributing to mostly window and very little AC use in this scenario (Figure 4.24).

As a consequence, the cooling energy consumption was much lower than the other scenarios. To address this issue, two actions were taken, (a) to combine the GLMM WO model with the AC activation control by temperature, the same one used in *Simulation A*; and (b) to use the models generated using Markov Chain.

The first above-mentioned action resulted in *Simulation C* and the second in *Simulation D*. *Simulation C* allowed the AC to be activated if/when the windows were closed, and all the temperature parameters established in *Simulation A* for AC were also met. This instance presented higher indoor temperature averages than the measured data, as well as *Simulation A*, possibly due to the fact that there was very little use of windows (Figure 4.25). This scenario shows almost the opposite of *Simulation B*, where the AC had very little use. These extremes show that, in this kind of office in a high-altitude tropical climate, using only one of the controls can result in high indoor temperature values. In addition, *Simulation C* presented higher cooling energy consumption than *Simulation A*, because the indoor temperature ranged from 20 to 31 ^{o}C, therefore meeting the criteria for AC use during practically the entire period. As shown in Figure 4.25, there was almost no use of windows, resulting in higher energy consumption.

The second action taken, *Simulation D*, used the models generated using Markov Chain with no additional treatment for both controls. This scenario presented average indoor temperatures very close to the measured ones, and a cooling energy consumption average lower than *Simulation A*. Given the average indoor temperatures observed, it can be said that the models used in *Simulation D* were a better representation of occupant behavior than *Simulation A*. In comparison to *Simulation C*, *Simulation D* is more representative of occupant behavior because it allows both controls to be activated at different temperatures (Figure 4.26), unlike the fixed schedules situation. Also, if this model is a closer representation of occupant behavior than *Simulation A*, and the cooling energy consumption is lower than that of *Simulation A*, it can be argued that occupants make more use of natural ventilation than what is set in the fixed schedules commonly used. This shows, once again, that occupants tolerate higher temperatures before alternating to the AC. In addition, it shows that it is possible to maintain lower indoor temperature values and consume less energy by using windows more frequently.

Next, *Simulation E* was run to combine the window model of *Simulation B* and the AC model from *Simulation D*. Such combination was made given that the AC model from *Simulation B* did not perform well, therefore the AC model in *Simulation D* was created. This instance, *Simulation E*, was to verify how the GLMM WO model performed when combined with a different AC model, as *Simulation C* combined it with a fixed schedule option. However, this combination presented higher indoor temperature averages than all other instances, except for *Simulation G*. Results from simulations *E* and *G* will be addressed in the sequence (Figure 4.27).

Simulation F was run using Markov Models with the Penalty Factors. Results from this scenario were very similar to those in *Simulation D*, as these models presented very similar results (Section 4.3), however with slightly higher values for both the variables being considered (Figure 4.28). *Simulation G* (Figure 4.29) was once again an attempt to combine the newly generated AC Markov Models with Penalty factor and the GLMM WO model. As with the previous similar instance (*Simulation E*), this also resulted in higher indoor temperatures and energy consumption.

All the combinations that used the GLMM WO model and another model/control for the AC resulted in higher indoor temperature values, as well as higher levels of energy consumption. These combinations resulted in the AC being used most of the time, allowing very little use of windows. As a consequence, indoor temperatures were higher, consequently causing the cooling energy loads to also be higher. Although these combinations (*Simulations C, E, G*) didn't accurately represent the measured data from this study in the simulations, they can be representative of a type of occupant, one that gives preference to the AC over the window. As shown by these results, occupants with similar behavior are likely to cause higher levels of energy consumption.

The final simulation, *Simulation H*, was run using the MC ROSE models. This scenario presented the closest indoor temperature average to the measured data, and the lowest energy consumption, with the exception of *Simulation B*. Simulations *D, F* and *H* had very similar outcomes in terms of indoor temperature values. Interestingly, *Simulation H* had the lowest average values for temperature and for energy consumption. Figure 4.30 shows the frequency of use of each control for this scenario, illustrating a higher frequency of window use when compared to most of the other scenarios. The average indoor temperature values of these three scenarios (D, F and H) were very similar, but their energy consumption varied, so samples from summer (January) and winter (July) were taken from their simulation outputs to analyze how the models predicted occupant behavior. In addition, comparisons between the measured data and these outputs were made.

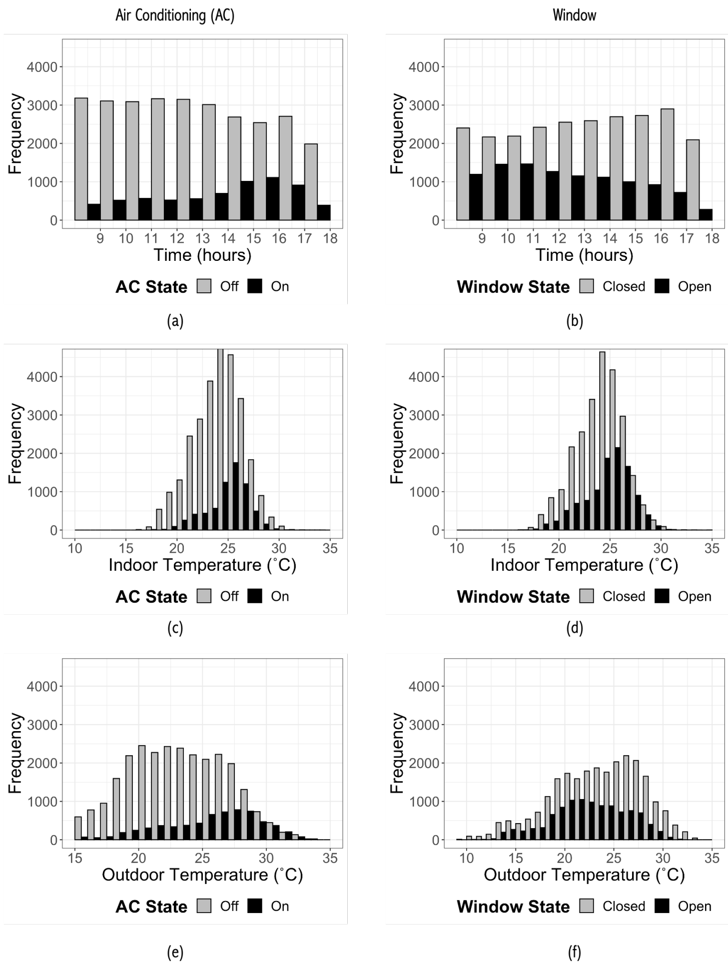


Figure 4.22: Measured Data: histograms for AC and window use

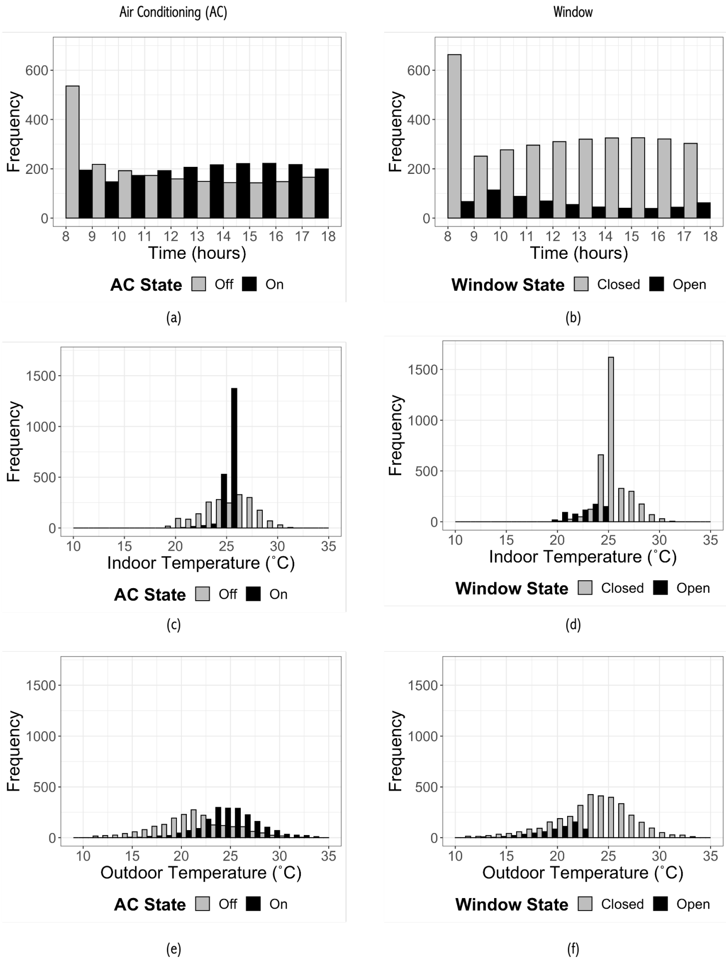


Figure 4.23: Simulation A: histograms for AC and window use

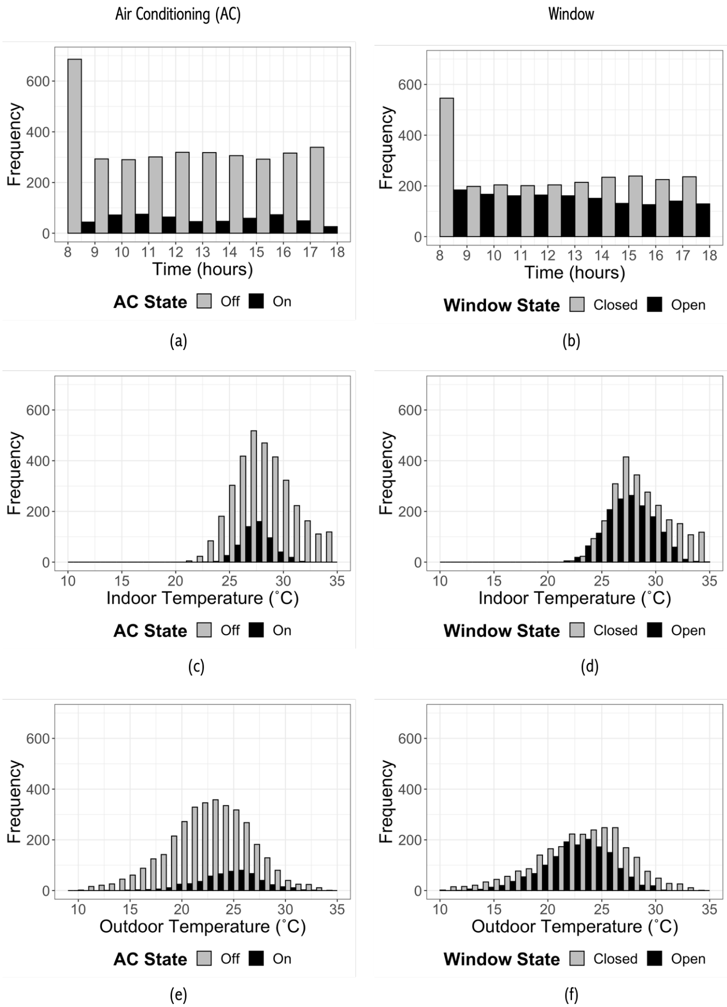


Figure 4.24: Simulation B: histograms of AC and window use

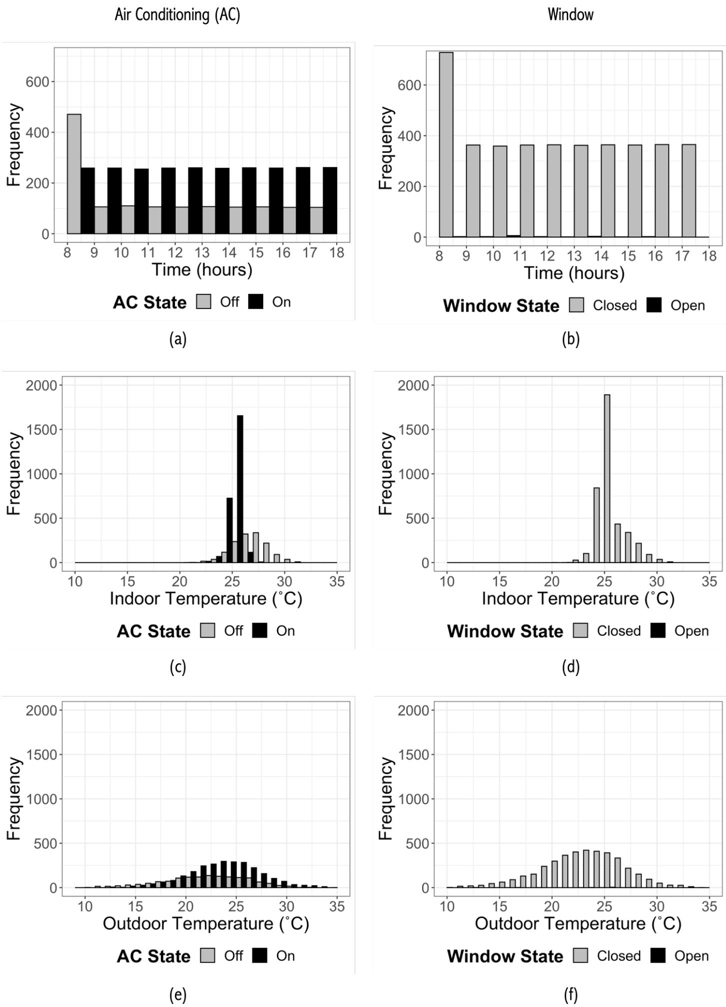


Figure 4.25: Simulation C: histograms for AC and window use use

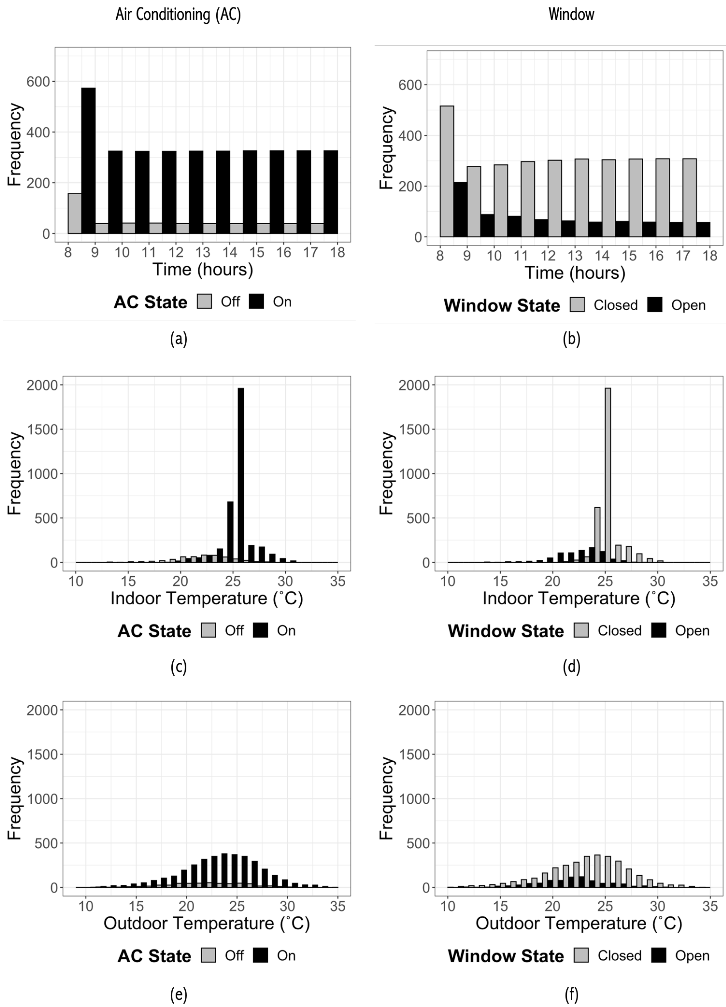


Figure 4.26: Simulation D: histograms for AC and window use

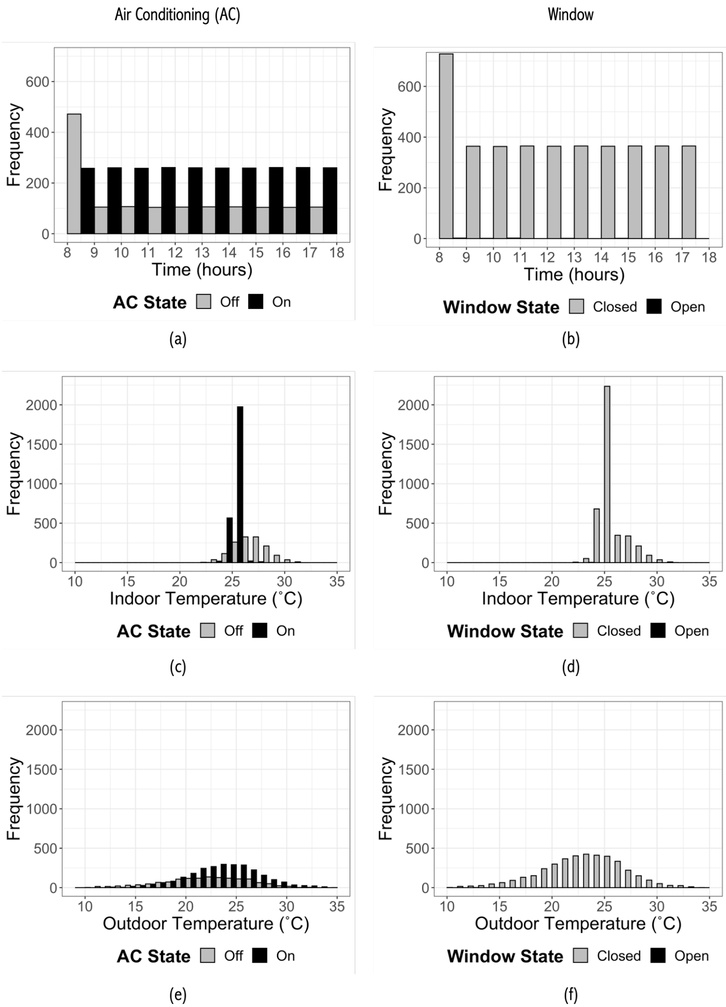


Figure 4.27: Simulation E: histograms for AC and window use

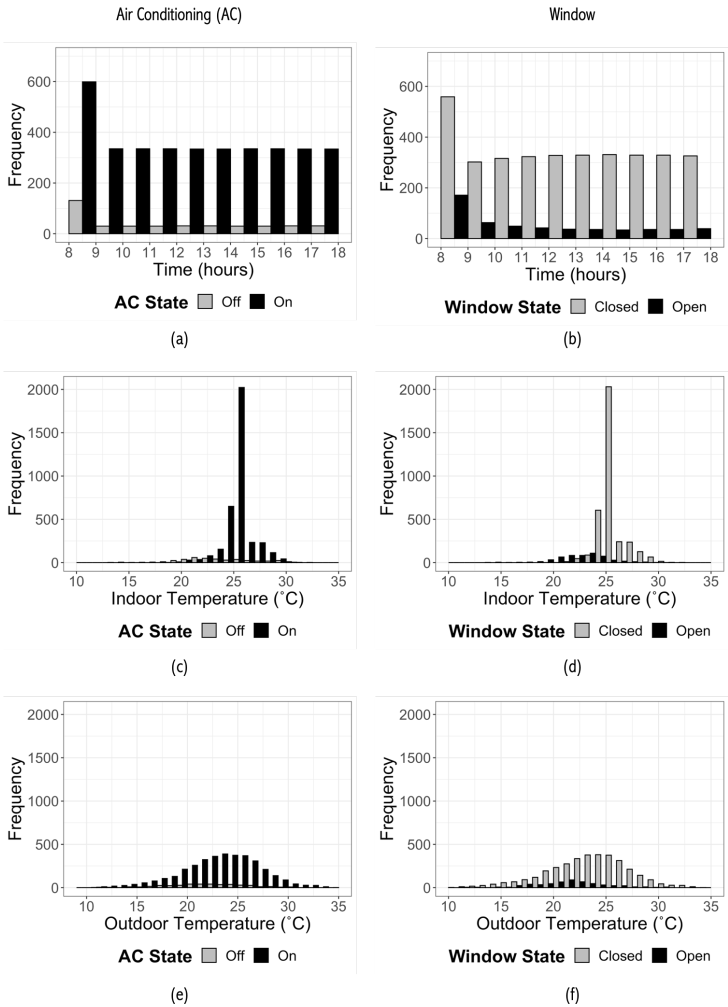


Figure 4.28: Simulation F: histograms for AC and window use

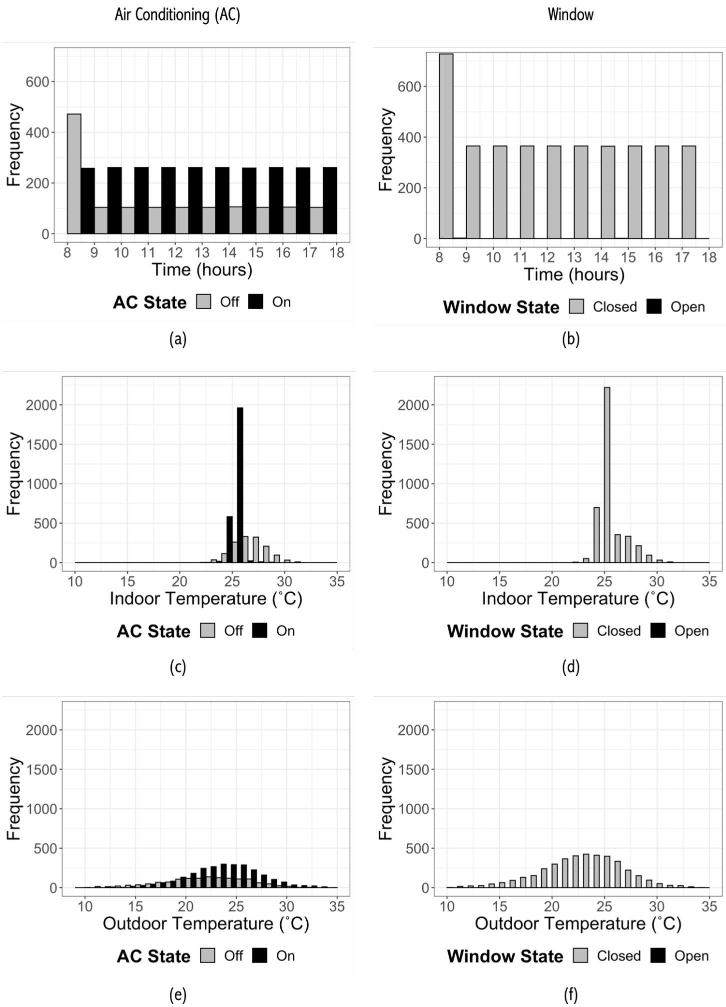


Figure 4.29: Simulation G: histograms for AC and window use

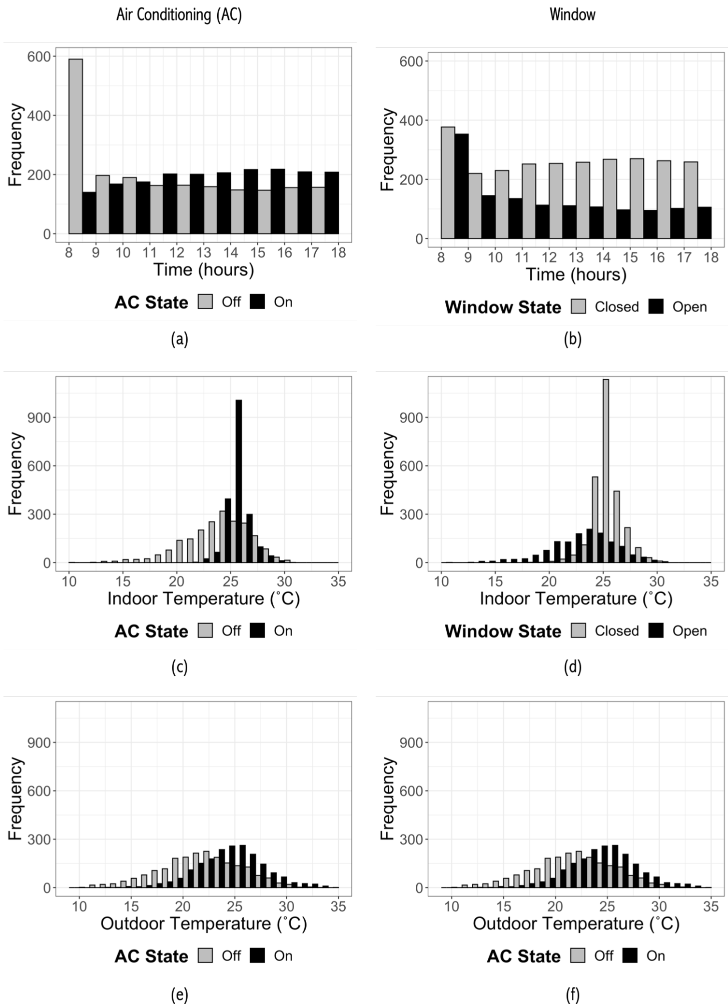


Figure 4.30: Simulation H: histograms for AC and window use

4.4.3 Behavior Graphs

Exerts of one week from summer and one from winter were taken from the outputs of *Simulations D, F* and *H* to compare how the different models allowed different control activation. The selected days refer to the second week (Mon-Fri) of each month, January 8 to 12, and July 9 to 13.

The considerations made here are based on the portrayed days, from 8 a.m. to 6 p.m., as the objective of such comparison is to demonstrate how differently each model portrayed occupant behavior in the simulations, even though the resulting average indoor temperatures were very similar. The results from these three scenarios show that, given the different use of the available controls, the same indoor temperature can be obtained. However, because of such difference in control use, it is possible to also achieve lower levels of energy use.

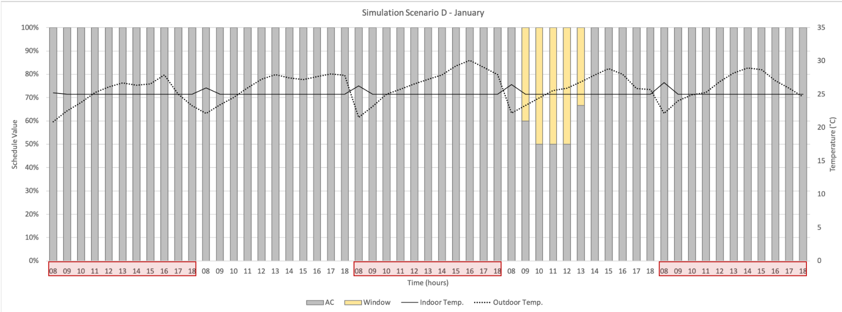
During these days the outdoor temperature ranged from 20.9 to 30 °C during the summer, and from 10.7 to 23.6 °C during the winter. The schedule values shown in Figures 4.31, 4.32 and 4.33 refer to the average value of each control, meaning how much that control was allowed to be activated within the respective hour. This does not mean that the control was in fact used because it was allowed. There were occasions when the AC was allowed to be activated, but so was the window, and the latter took precedence. These figures are to illustrate how much each model allowed each of the studied controls to be activated each hour.

As shown in Figures 4.31a and 4.32a, these models (*Simulations D* and *F*) allowed very little use of window during the summer. Because the AC was in use during most of the time, the indoor temperature was maintained at 25 °C. It is interesting to see that these two models allowed very little use of window, however in different occasions within the same week. In *Simulation D*, window use was allowed during the morning of day 4, while in *Simulation F*, it was allowed mostly during the afternoon of day 5. As for the winter, *Simulation D* allowed more use of windows in comparison to *F*, and these behaviors are in agreement with the energy consumption relative to each scenario, which, as shown in Figure 4.19, *Simulation F* presents higher levels of energy consumption.

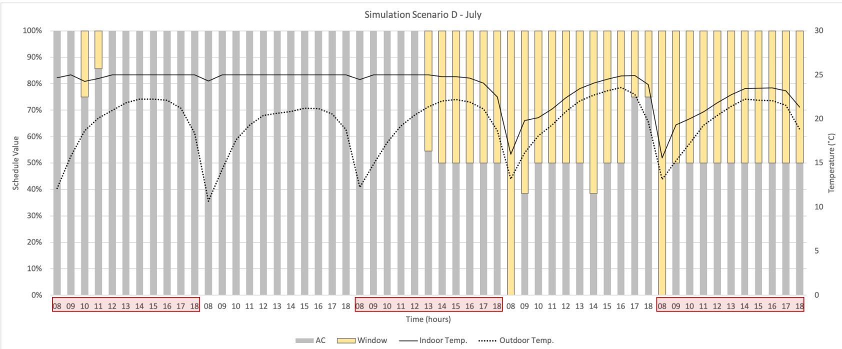
When analyzing Figures 4.31b and 4.32b, it is possible to see that *Simulation D* allowed the most window use during the last three days of the represented winter period, while *Simulation F* allowed it during the last day. However, in both situations, when natural ventilation was used, because the outdoor temperature was close to, or below 20 ^{o}C, indoor temperatures decreased, and were maintained below, or slightly above 25 ^{o}C. Even though these figures show values of how much each control was allowed to be activated during each hour, one can identify when natural ventilation was in use when allowed, as the indoor temperatures oscillate more, accompanying the pattern seen for the outdoor temperatures, as opposed to being maintained at 25 ^{o}C, as this was the cooling set point for the AC.

In addition, Figure 4.22 shows when occupants most used each control based on time, indoor and outdoor temperatures according to the measured data. When looking at the results from *Simulations D* and *F* and at the histograms of the measured data (Figures 4.26, 4.28 and 4.22), it can be observed that these simulated models did not allow much window use with outdoor temperature between 20 and 25 ^{o}C, even though the measured data shows higher frequency of AC off (Figure 4.22e) and window open (Figure 4.22f) for the same temperature range. Figures 4.26e and 4.28e, referent to *Simulations D* and *F*, respectively, also bring evidence to the above statement, showing high frequencies of AC use with outdoor temperature between 20 and 27 ^{o}C, consequently yielding high frequencies of window closed, which in turn resulted in higher energy consumption than *Simulation H*.

As for *Simulation H*, it is the scenario that better represents the alternating behavior seen in the measured data, as shown in Figure 4.33. One of the main differences that can be seen between these three scenarios when looking at Figures 4.31, 4.32 and 4.33, is that in *Simulation H*, the AC and Window models took both indoor and outdoor temperatures into account, balancing the use of both controls, consequently resulting in lower levels of energy consumption (Figure 4.19), while maintaining the indoor temperature around 25 ^{o}C. Figure 4.33 clearly shows the relation between indoor and outdoor temperatures and the resulting control that was allowed for a longer period of time within each hour. Figure 4.33a shows more



(a)

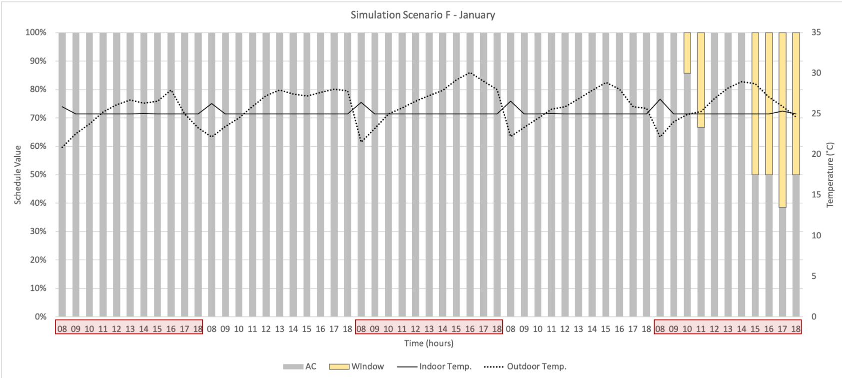


(b)

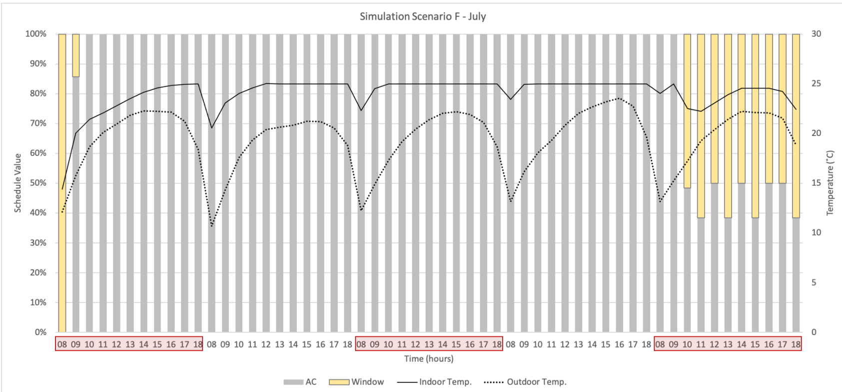
Figure 4.31: Simulation D: summer and winter

clearly that when outdoor temperature is equal to or above indoor temperature, the AC is the control most allowed, and likely the one activated. This response to the interaction between the temperatures is also in agreement with the measured data. In addition, when comparing Figure 4.30 to the measured data (Figure 4.22) it can also be seen how this scenario, among the ones presented, more closely represented the measured data and thus better simulated occupant behavior.

Figure 4.34 shows the indoor and outdoor temperatures distribution for the measured data, as well as for the fixed schedules scenario (Figure 4.34b), and the three analyzed scenarios in this section (Figures 4.34c-e). There is a clear difference



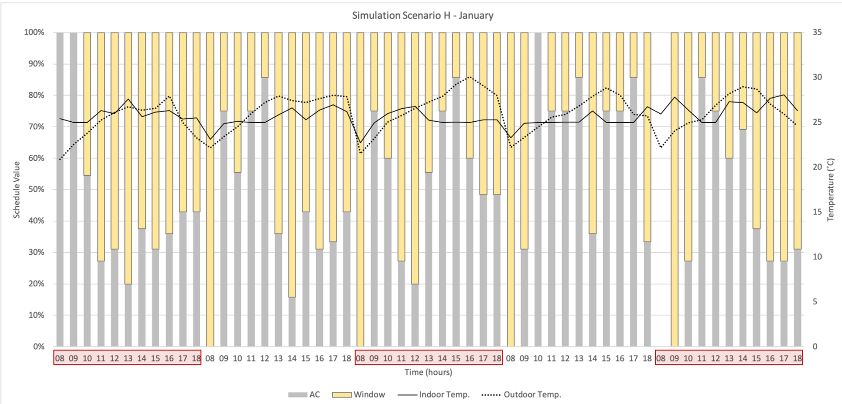
(a)



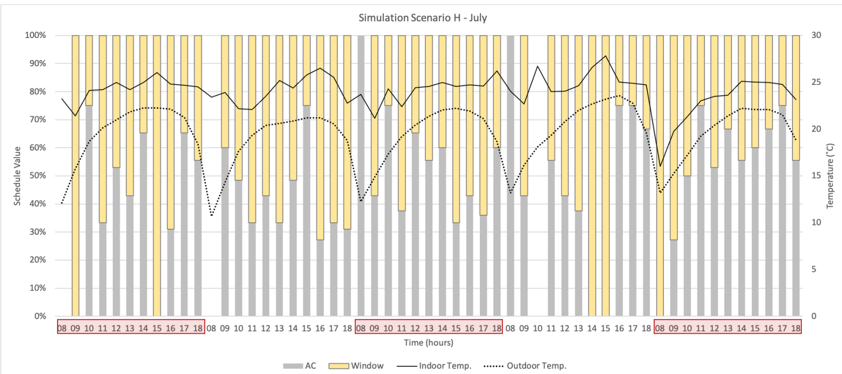
(b)

Figure 4.32: Simulation F: summer and winter

from *Simulation A* to the three scenarios using the behavior models, as there is also a clear difference from *Simulations D and F* to *H*. Figure 4.34e depicts the least use of AC and a more distributed relation between indoor and outdoor temperatures. In this scenario, the cloud is dense with outdoor temperature between 20 and 25 °C, corresponding to indoor temperature ranging close to and above 25 °C, resembling the measured data (Figure 4.34a).

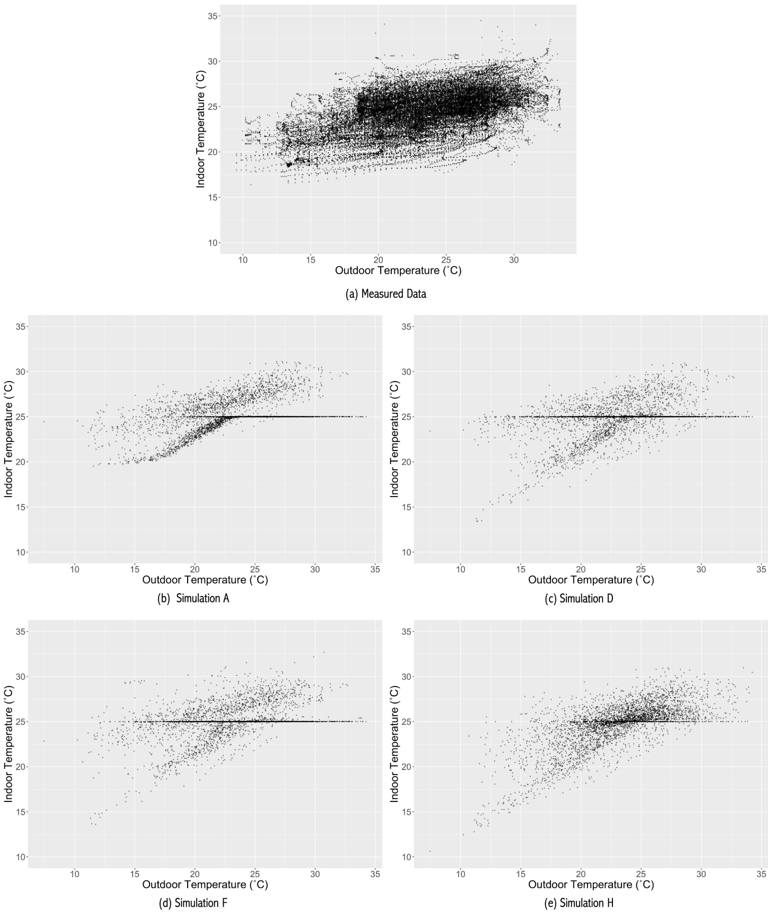


(a)



(b)

Figure 4.33: Simulation H: summer and winter



(a) Measured Data

(b) Simulation A

(c) Simulation D

(d) Simulation F

(e) Simulation H

Figure 4.34: Indoor and Outdoor temperatures' distribution

4.4.4 Final Considerations

The previous sections have shown that, among the simulated scenarios, *Simulation H* is the one that best represented occupant behavior in EP. With that in mind, it is important to highlight some main differences between the fixed schedules scenario based on temperature, *Simulation A* and *Simulation H*, using the MC ROSE models. Once again, taking into account the metrics here analyzed, indoor temperature and cooling energy load, it is possible to see that *Simulation A* resulted in a higher indoor temperature average, 0.9 ^{o}C higher than the measured value, and 0.4 ^{o}C higher than the average for *Simulation H*. However, as shown in Figure 4.19, a higher indoor temperature average did not result in lower energy consumption. This result evidences that the way the controls are operated greatly influence on how energy is consumed, in addition to impacting the indoor environment.

This test analyzed one small mixed-mode office facing north, and different results were obtained using fixed schedules vs. occupant behavior models. If this data were to be extrapolated, and one considered, for instance, a 20 story building with 10 north facing offices, the differences seen in the simulation results would be much higher, having a more significant impact. Not to mention if the offices were larger, had more windows, etc. Even the small differences among the results between *Simulations D, F* and *H* could have a high impact when considering a large building. The tests presented in this work show the need to have models that accurately represent the real data, as well as their relevance in the results. Specifically, two aspects of behavioral models can be highlighted, (a) the different way that the models are built can result in different outcomes in simulations, and (b) they have a clear impact in energy simulations, and therefore having accurate models is key to having precise simulation results.

The fixed schedules commonly used do not realistically capture and portray how occupants behave. Figure 4.23 shows how this method does not allow window use with indoor temperature above 25 ^{o}C and outdoor temperature above 22.5 ^{o}C. However, as the measured data shows (Figure 4.22), occupants operate both controls within a vast range of both indoor and outdoor temperatures, granted at different frequencies. Therefore, in this work, *Simulation H* (Figure 4.30) best represents the measured occupant behavior. Given the differences identified in control use, it is advised to use occupant behavior models when these are available and reliable, to represent occupant behavior instead of fixed schedules.

4.4.5 Application of Models

In this study, occupant behavior models were generated using different statistical methods, GLMM and Markov Chain. The models were generated to predict the use of window and AC in mixed-mode offices. As described in the previous sections, these models performed in different ways, each presenting different applications, advantages and limitations.

The GLMMs are simpler models to create, also considering the fewer steps to treat the data to generate such models, and there is the need for only one for each action. Considering these aspects of their simplicity, and taking their limitations into account, the application of these models is easier, thus being a viable option as an early approach to a study. In addition, this model reflects a pattern of occupancy based on the use of the controls, even though occupancy was not measured, which is an interesting feature not present in the MC models.

As for the Markov Chain models, they require a more extensive data collection so more precise models can be created based on real data, in addition to a more complex data treatment to generate the models. However, this type of model addresses part of the limitations found in the GLMMs, while correctly predicting each outcome. Therefore, this type of model is better suited to be used in simulation software when predicting occupant behavior in mixed-mode offices.

5 Discussion

This chapter brings a further discussion to the collected data analysis, to the models results, and the results of the simulation tests run in EnergyPlus, all presented in Chapter 4.

5.1 Collected data

The collected data showed that season has an effect on how occupants operate the controls available to them, which is in agreement with the findings of other studies in the literature. The study by [113], also conducted in a mixed-mode office setting, showed a predominant use of AC, and nearly no use of windows, during the summer, while the present study identified use of both controls during all seasons. In studies conducted in naturally ventilated offices, [79] reported that window opening probability linearly increases with outdoor temperature increase, and [73] showed that windows are more likely to be open during warmer seasons. Seasonal window opening behavior reported for naturally ventilated buildings differ from the one observed in this study, as occupants in mixed-mode offices have different choices of actions to take. Nonetheless, different patterns of behavior are observed in each season in naturally ventilated offices as well as in mixed-mode ones.

The results from the analysis on the collected data highlights the importance of generating models specific for each climate and culture. The measured data in this study showed higher frequency of use of windows in the winter and fall (Figures 4.5 and 4.6), with the lowest use of windows during the summer, while studies in different locations show different results. For example, [151] conducted a study in Korea, and while the authors also state that seasonal effects have an impact on window use patterns, their results present a higher frequency of window use during the summer and the lowest during the fall.

Even within the same country, however different climates, different window use patterns are seen in each season. The work of [113], who also conducted measurements in mixed-mode offices, although in the south region of Brazil, in a sub-tropical climate, shows that there was almost no use of windows during the summer, and very little use of AC during the winter. In the work presented here, located in a high-altitude tropical climate within the same country, the analysis showed higher use of windows during the summer in comparison to [113], as well as a higher use of AC during the winter, evidencing a very different behavior pattern.

These findings show that many actions taken by occupants are contextually sensitive and will differ according to their personal characteristics. Therefore, studies in several different contexts, for example, cultures and climates, need to take place in order to collect data on a broad range of these influencing variables [124, 125].

5.2 Models

This study aimed to create models to predict the use of windows and the AC based on data collected during a monitoring campaign. The objective of the models was to be representative of occupant behavior in mixed-mode office buildings in a high-altitude tropical climate, and therefore be implemented in building simulation software in order to provide more accurate results on energy consumption. Two statistical methods were applied to generate the models, namely, Generalized Linear Mixed Models (GLMM) and Markov Chain (MC). GLMM was the first method applied, and given the limitations found in the models, MC models were generated as a way to address such limitations.

Both statistical approaches resulted in models showing the same general behavioral trends, satisfactorily portraying the behavior seen in the measured data. The models demonstrated that indoor and outdoor conditions are highly influential on how occupants operate windows [77, 105], and consequently the AC. The overall identified behavior, given the type of office, actions studied, and freedom of occupants to operate both controls as they pleased, was that occupants alternate between the use of windows and the AC “adequately". That is, occupants will mostly use windows, until indoor and/or outdoor temperatures reach 25 ^{o}C or above, when they will close windows and switch to AC. Similar results have been reported in the literature. [74], although in a residential setting in Sydney (Cfa, for the Köppen classification [98]), reported an increase in AC use with outdoor temperature above 25 ^{o}C, with a decrease in window and doors use. [109] reported a steeper increase in AC use with indoor temperature above 24 ^{o}C in mixed-mode offices in five European countries and five cities in Pakistan, and [113] showed a higher probability of AC use with outdoor temperature above 25 ^{o}C in Florianópolis, Brazil.

The alternating behavior shows great potential for energy savings, as this is already an advantage of the mixed-mode setting discussed in other studies, such as a 52% reduction in AC use [117], and a 64.9% reduction in energy use when combining the use of natural lighting and hybrid ventilation [114]. In addition to the energy savings potential presented by the mixed-mode scenario as studied in this work, it is also important to highlight that this setting presents great freedom for occupants to operate the controls, contributing to high levels of perceived control, which, in turn result in higher levels of control operation [82, 153], occupant satisfaction [18], and energy savings [150].

The models indicate a higher use of controls upon arrival and departure, before 9 a.m. and around 6 p.m., respectively, especially of windows upon arrival, which is consistent with the findings reported by [55, 58, 152]. Where [55] reported their results based data collected from an experiment building with operable windows located in Lausanne, Switzerland. [152] conducted measurements in naturally ventilated offices located in Cambridge, UK. And [58] reported findings based on the monitoring of naturally ventilated individual offices in Freiburg, Germany. This behavior can also be associated with occupants' routine, which also plays a role in their decisions to use a given control, as window operation could be a function of routine, as pointed out by [45], and is identified as a predominant feature in AC use [130]. In addition, upon entering a room occupants are more sensitive to bad air quality, and WO could also be in response to low IAQ perception [3].

Indoor and outdoor temperatures were key predictors to window operation, in agreement with several studies in the literature [96,131,148]. Considering window and AC operation in the same environment, these variables were also strong predictors, especially outdoor temperature for the AC, in accordance with the results found by [113].

Low outdoor temperature showed high probabilities of window opening with both types of model, a behavior that allows to cool the indoor environment [106], and also let fresh air inside [70, 156].

GLMMs show higher probabilities of window use with higher outdoor temperature, while the MC Window Open model shows decreasing probabilities of window use as outdoor temperature increases. The behavior shown by the GLMM model, although seemingly counter intuitive, may indicate that occupants will open windows after having used the AC for longer periods. When the AC is activated, as the measured data showed, windows were closed, resulting in no air renovations, since these offices are equipped with split or window AC equipment. If one considers that occupants activate the AC with higher outdoor temperatures, and remain with the control active for as long as the temperature remains high, then it is possible to consider that the AC is on and windows closed for several hours, resulting in occupants opening the windows while outdoor temperatures are still high. There were no IAQ measures in this study to correlate to the observed behavior, and such measures should be conducted in future studies.

Lower indoor temperature values (19 ^{o}C) showed high probabilities of window opening, which depicts the same logic of window opening to avoid overheating, though reversed. That is, occupants open windows when indoor temperatures are low to let the outside air in, likely expecting the outdoor air temperature to be higher than the indoor. Once again, this effect was only captured in the GLMMs, and not in the MC models.

In general, the two different types of models predicted similar behavior trends. However, when analyzing how each model predicted given each variable identified as a significant predictor, it is possible to see that there are details that differ them, which are consequences of the different ways they were built. These differences, though seemingly minor, show an impact on building energy simulation results, as evidenced with the results of the simulation tests performed.

5.3 Simulation tests

The models created in this work were implemented in EnergyPlus. The tests were designed to allow comparisons between the commonly applied fixed schedules determining window and AC use, and the created models generated.

Eight different controls combinations were simulated (Figure 4.18), including the fixed schedules scenario. As shown in Figure 4.19, each different model combination presented different energy consumption and indoor temperature averages, as these were the main indexes analyzed. As shown in the work of [51], different ways of creating the models can cause different simulations results. In this work, although all the models were created using the same data, the different modeling techniques and different models combinations generated different results, evidencing that the way these models are built can greatly influence the results they provide in building simulations.

The fixed schedules scenario, *Simulation A*, resulted in indoor temperature average slightly above 25 ^{o}C, which is above the measured data average of 24.4 ^{o}C, and an intermediate energy consumption among the simulated scenarios. In this scenario, no windows were operated with outdoor temperature above approximately 23 ^{o}C, causing higher levels of energy consumption. A greater, although not sole, use of windows can result in less energy use [155]. This fixed schedules scenario did not accurately portray the observed occupant behavior, which is in agreement with what is commonly reported in the literature [59]. The simplifications seen in fixed schedules do not accurately simulate the complexity of occupant behavior and their influence on energy use and the indoor environment [39, 40].

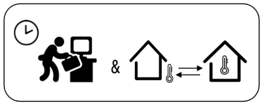
Simulation B presented the highest indoor temperature averages, 28.5 ^{o}C, and lowest energy consumption. This scenario presented a predominance of window use, showing that in the studied climate, the sole use of windows can lead to overheating. On the other hand, among the three scenarios presenting the closest indoor temperature averages to the measured data, *Simulations D and F*, presented the highest energy consumption averages. These two scenarios show a counterpoint to *Simulation B*, since they presented mostly the use of AC. *Simulations B, D and F* exemplify how different types of occupants, and therefore the different ways the available controls can be operated, result in different energy consumption levels [5].

The scenarios where the different generated window models were combined with a fixed AC schedule showed the highest energy consumption levels. Studies have shown that the actual operation of windows largely differ from fixed schedules [63], as seen in *Simulation A*. However, *Simulations C, E and G* show that the actual use of AC, in this kind of mixed-mode configuration, also largely differs from the fixed schedules. These three simulation scenarios present a predominance of the use of AC, when in fact, the measured data shows that occupants alternated between the use of both controls within a considerable range of indoor and outdoor temperatures.

Simulation H was the scenario that best represented the measured data in this study, consequently presenting indoor temperature average values closest to the measured data. This scenario showed evidence that indoor and outdoor temperatures were taken into account to allow the use of one control or the other, portraying the “alternating" behavior identified in the measured data. By allowing more use of windows, more amenable indoor temperatures were achieved, and less energy consumed [155,156]. The reduction seen in energy consumption, here occasioned by the actual use of both controls, is also in agreement with the results from other studies on the advantages of the mixed-mode strategy in Brazil [114, 117].

The simulation tests evidenced the discrepancy between the fixed schedules commonly implemented to simulate the mixed-mode strategy, and the actual data on occupant behavior. This discrepancy can be attributed to the inability of the fixed schedules to take into account occupant adaptability [101], which is a key element of occupant behavior. The simulation tests results showed that occupant behavior models are indeed necessary to achieve more accurate simulation results. Furthermore, accurate behavior models, although difficult to create, are essential in order to achieve more precise simulation outputs for energy consumption.

6 Conclusions



This work created predictive behavioral models for mixed-mode office buildings, focusing on window and AC operation, with the purpose of implementing such models in computer simulation programs. The models were created using data collected during an 18-month monitoring campaign in ten offices in the city of São Carlos, SP, Brazil, classified as a city of high-altitude tropical climate. The monitored office buildings displayed operable windows and AC units, all of which occupants were free to operate at any given time, with no aid from systems or displays to indicate the indoor or outdoor environmental conditions, which is the most recurring type of mixed-mode building in the city of this study. During the monitoring campaign, indoor and outdoor environmental measurements were taken, as well as measurements related to occupants' actions related to the study. Indoor variables, such as air temperature and relative humidity were measured, while outdoor environmental variables were taken from a meteorological station. Occupants' actions, that is, window opening and AC activation, were also monitored.

The complete data set provided an overview of occupants behaviors, considering time of day and temperature, within the studied climate and type of building. Occupants tend to operate controls upon arrival and departure. Specifically, they tend to open windows upon arrival. As for the studied actions in relation to indoor and outdoor temperatures, windows are opened when outdoor temperature is around 20 ^{o}C and indoor temperature is close to, but mostly below, 25 ^{o}C. As for AC, it is mostly activated when indoor and outdoor temperatures are close to or above 25 ^{o}C.

Using the measured data, two statistical modeling techniques were applied to create the models: (a) *Generalized Linear Mixed Models* (GLMM), which generated two models, one predicting the probability of window opening, and another the probability of AC activation, and (b) *Markov Chain* (MC), which resulted in four models, two for each of the studied controls, one model for opening/activation and another for closing/deactivation. Due to the prediction patterns given by the GLMMs, simulation results did not accurately depict the measured occupant behavior, and the MC models were implemented. Several combinations of the different models created were tested, and results showed that the model that most accurately portrayed the monitored occupant behavior in simulations, given the type of controls being studied, were MC models that are able to accurately predict both classes (zeros and ones). These models, in this work the MC ROSE models, best depicted the way occupants alternate between one control and the other given the fluctuations in, and relation between, indoor and outdoor temperatures throughout the day.

Nonetheless, both modeling techniques present their advantages to the further investigation of occupant behavior in mixed-mode buildings. GLMMs are able to address the issue of behavioral diversity by adding a variable of random nature - in this study the different monitored offices. In addition, this type of model can be built with smaller samples than MC models and still provide good results, allowing researchers to use this method even when shorter monitoring campaigns take place. However, it is important that one knows and understands the measured data well, so that the limitations in the resulting models can be identified and accounted for when using the model. On the other hand, although the MC models presented a better performance in the simulation tests, they require a more extensive data collection and/or additional steps to generate a model that can correctly identify both classes, and thus perform well in simulation programs.

Because the models are built based on data collected on site, they are locally applicable. The models created in this work can be applied to building types of the same kind, i.e. mixed-mode, within the same climate, a high altitude tropical climate. To this date, there are no predictive behavioral models for mixed-mode offices in Brazil to be implemented in simulation programs, especially for a high-altitude tropical climate, highlighting the importance of future studies on occupant behavior for this context. Brazil is a large country with several different climates and cultures. Because occupant behavior is influenced by a number of factors, physical environmental ones, as well as contextual factors, such as culture, it is crucial to investigate the different drivers that triggers actions. Furthermore, the mixed-mode configuration poses an additional challenge in predicting actions, since occupants have another factor that influences their decisions; the option to use a different control whenever they feel necessary. Within the mixed-mode context it is important to study and understand what causes occupants to alternate from one control to the other, thus giving insights on how this combination can be better used to minimize energy consumption.

6.1 Limitations

The monitoring campaign in this study lasted 18 months, and 10 offices were monitored in alternated periods within the 18-month period. There were limitations in equipment availability, which were reflected as limitations in the study. The amount of state loggers was very low, resulting in an impossibility to monitor all the windows in offices with multiple windows. Another equipment-related limitation was the AC state measurement. Since there was no equipment to measure the actual AC state by monitoring energy consumption, for example, AC temperature was measured and the state calculated based on such measurements. This poses two implications, (1) the resulting calculated AC state can present flaws and indicate some incorrect records of this control's state, and (2) it was not possible to quantify the amount of energy consumed by the AC equipment in each office to make more accurate comparisons to the simulated data. There were also no equipment to monitor occupancy, and the study relied on the general schedule of occupancy provided by occupants, eliminating unoccupied days reported by the occupants.

In relation to the occupants themselves, because the equipment remained in each office for a period of two weeks at a time, the Hawthorne effect [129] can be mentioned as a limitation, since occupants were aware they were being monitored. However, there is no way of quantifying any Hawthorne effects in this study if they do exist.

The sample size in the study can be considered a limitation. First, sample size related to the amount of monitored offices, since being able to monitor more offices, and therefore more occupants, would be ideal to capture a broader range of behavioral diversity. And second, to generate Markov chain models a larger data set is ideal, which enables the models to correctly identify both classes, and thus be more precise in portraying the desired behavior.

And, lastly, although the monitoring campaign in this study occurred prior to the COVID pandemic, this new reality and subsequent requirements show that the studied mixed-mode office configuration does not present enough air renovations to maintain adequate air quality in these offices, mainly when the windows are closed and the AC is in use.

6.2 Future Work

In order to improve the models generated in this study, and to further investigate the driving factors of occupant behavior in mixed-mode office buildings, it is suggested that future works apply questionnaires related to occupants' personal preferences, in relation to window and AC use, to learn how those preferences interfere with the actions taken in the office. Also, because this study is focused on mixed-mode offices, it would be interesting to analyze occupant behavior in solely naturally ventilated offices, with the aid of questionnaires, in order to compare which actions occupants in NV buildings take, given the same season or environmental conditions, when occupants of MM buildings activate the AC.

This would further the understanding of what leads occupants to use the AC, and also provide ways to suggest alternatives to MM building occupants to space out the use of AC as much as possible, implementing the techniques observed in the behavior of NV buildings' occupants.

And, to address the issue of accuracy in the data collection, the ideal scenario is to conduct measurements where all the windows are monitored, and AC activation is precisely registered. Monitoring doors, or the main/entrance door to the office, is also suggested, since it is a way to verify; (a) time of arrival and departure, thus aiding in occupancy as well; (b) if occupants combine the use of windows and doors to improve air movement, quality, etc.; and (c) if occupants open doors before activating the AC, as an intermediate action to improve the indoor environment, and if occupants close the door when using the AC.

Lastly, it is important that future works based on monitoring campaigns consider the implications that the COVID pandemic brought to the way occupants operate windows and the AC. Post-pandemic occupant behavior may differ from previous years, when the concern for the risk of contamination in an office environment, for example, was not something so evident and frequent in occupants' minds.

A Appendix

This section provides the models coefficients for all the models generated as part of this study. These coefficients were used to implement the models in EnergyPlus.

A.1 Generalized Linear Mixed Models Coefficients

Table A.1: Generalized Linear Mixed Model for Air-Conditioning Activation (ACA)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-2.20	$p < 0.001$
Time 1	0.21	$p < 0.001$
Time 2	0.93	$p < 0.001$
Time 3	-0.21	$p < 0.001$
Time 4	-0.36	$p < 0.001$
Indoor Temperature 1	-0.50	$p < 0.001$
Indoor Temperature 2	-0.01	.55
Indoor Temperature 3	-0.003	.71
Outdoor Temperature 1	1.85	$p < 0.001$
Outdoor Temperature 2	0.01	.45
Outdoor Temperature 3	-0.10	$p < 0.001$
Indoor Rel. Humidity	-0.62	$p < 0.001$
Outdoor Rel. Humidity	1.27	$p < 0.001$
Window State	-1.00	$p < 0.001$

Table A.2: Generalized Linear Mixed Model for Window Opening (WO)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.63	$p < 0.001$
Time 1	0.03	.45
Time 2	0.51	$p < 0.001$
Time 3	-0.11	$p < 0.001$
Time 4	-0.30	$p < 0.001$
Indoor Temperature 1	-0.54	$p < 0.001$
Indoor Temperature 2	-0.10	$p < 0.001$
Indoor Temperature 3	-0.03	$p < 0.001$
Outdoor Temperature 1	-0.56	$p < 0.001$
Outdoor Temperature 2	-0.24	$p < 0.001$
Outdoor Temperature 3	-0.02	.17
Outdoor Rel. Humidity	-0.15	$p < 0.001$
AC State	-1.35	$p < 0.001$

A.2 Markov Chain Models Coefficients

Table A.3: Markov Chain Model for Window Opening (WO)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-4.66	$p < 0.001$
Time 1	-0.92	$p < 0.001$
Time 2	-1.16	$p < 0.001$
Time 3	0.44	$p < 0.001$
Time 4	0.67	$p < 0.001$
Indoor Temperature 1	0.58	$p < 0.001$
Indoor Temperature 2	-0.09	.08
Indoor Temperature 3	-0.02	.57
Outdoor Temperature 1	-0.60	.0008
Outdoor Temperature 2	0.06	.28
Outdoor Temperature 3	0.06	.11
Indoor Rel. Humidity	0.33	.0005
Outdoor Rel. Humidity	-0.34	.007
R^2=0.14		

Table A.4: Markov Chain Model for Window Closing (WC)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-3.65	$p < 0.001$
Time 1	-0.03	.88
Time 2	-0.04	.87
Time 3	0.43	$p < 0.001$
Time 4	0.15	.12
Indoor Temperature 1	0.34	.002
Indoor Temperature 2	0.10	.04
Indoor Temperature 3	-0.02	.05
Outdoor Temperature 1	0.30	.16
Outdoor Temperature 2	-0.12	.14
Outdoor Temperature 3	-0.10	.06
Indoor Rel. Humidity	-0.38	.001
Outdoor Rel. Humidity	0.51	.002
R^2=0.11		

Table A.5: Markov Chain Model for Ac On (ACon)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-4.38	$p < 0.001$
Time 1	0.82	$p < 0.001$
Time 2	-0.73	.003
Time 3	-0.33	$p < 0.001$
Time 4	0.25	.005
Indoor Temperature 1	0.70	$p < 0.001$
Indoor Temperature 2	0.06	.27
Indoor Temperature 3	-0.03	.06
Outdoor Temperature 1	0.57	.004
Outdoor Temperature 2	0.02	.71
Outdoor Temperature 3	-0.11	.004
Indoor Rel. Humidity	-0.12	.20
Outdoor Rel. Humidity	0.31	.02
R^2=0.08		

Table A.6: Markov Chain Model for AC Off (ACoff)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-3.11	$p < 0.001$
Time 1	-0.43	.03
Time 2	-0.38	.16
Time 3	0.41	$p < 0.001$
Time 4	0.22	.03
Indoor Temperature 1	-0.45	$p < 0.001$
Indoor Temperature 2	0.03	.65
Indoor Temperature 3	0.02	.24
Outdoor Temperature 1	0.51	.007
Outdoor Temperature 2	0.03	.63
Outdoor Temperature 3	-0.05	.22
Indoor Rel. Humidity	-0.12	.16
Outdoor Rel. Humidity	0.30	.02
R^2=0.08		

A.3 Markov Chain Models with Penalty Factor Coefficients

As a result of the additional statistical methods applied to create the models with the Penalty Factors included, all models have R^2 values of 1.0, and all variables show significance values $p < 0.001$, therefore only the models coefficients are displayed in the following tables in this section.

Table A.7: Markov Chain Model with Penalty Factor for Window Opening (WO)

Predictor	*Coefficient (b)*
Intercept	-0.47
Time 1	-0.85
Time 2	-1.0
Time 3	0.41
Time 4	0.61
Indoor Temperature 1	0.55
Indoor Temperature 2	-0.09
Indoor Temperature 3	-0.02
Outdoor Temperature 1	-0.57
Outdoor Temperature 2	0.05
Outdoor Temperature 3	0.05
Indoor Rel. Humidity	0.32
Outdoor Rel. Humidity	-0.32

Table A.8: Markov Chain Model with Penalty Factor for Window Closing (WC)

Predictor	*Coefficient (b)*
Intercept	-3.67
Time 1	0.07
Time 2	0.01
Time 3	0.37
Time 4	0.13
Indoor Temperature 1	0.35
Indoor Temperature 2	0.09
Indoor Temperature 3	-0.02
Outdoor Temperature 1	0.26
Outdoor Temperature 2	-0.09
Outdoor Temperature 3	-0.08
Indoor Rel. Humidity	-0.29
Outdoor Rel. Humidity	0.38

Table A.9: Markov Chain Model with Penalty Factor for Ac On (ACon)

Predictor	*Coefficient (b)*
Intercept	-4.40
Time 1	0.73
Time 2	-0.61
Time 3	-0.30
Time 4	0.21
Indoor Temperature 1	0.70
Indoor Temperature 2	0.06
Indoor Temperature 3	-0.03
Outdoor Temperature 1	0.56
Outdoor Temperature 2	0.03
Outdoor Temperature 3	-0.10
Indoor Rel. Humidity	-0.11
Outdoor Rel. Humidity	0.29

Table A.10: Markov Chain Model with Penalty Factor for AC Off (ACoff)

Predictor	*Coefficient (b)*
Intercept	-3.11
Time 1	-0.33
Time 2	-0.30
Time 3	0.37
Time 4	0.20
Indoor Temperature 1	-0.42
Indoor Temperature 2	0.03
Indoor Temperature 3	0.01
Outdoor Temperature 1	0.43
Outdoor Temperature 2	0.02
Outdoor Temperature 3	-0.04
Indoor Rel. Humidity	-0.12
Outdoor Rel. Humidity	0.26

A.4 Markov Chain ROSE Models Coefficients

Table A.11: Markov Chain Rose Model for Window Opening (WO)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.47	$p < 0.001$
Time 1	-0.22	$p < 0.001$
Time 2	-0.71	$p < 0.001$
Time 3	0.12	$p < 0.001$
Time 4	0.48	$p < 0.001$
Indoor Temperature 1	0.35	$p < 0.001$
Indoor Temperature 2	-0.12	$p < 0.001$
Indoor Temperature 3	0.01	.05
Outdoor Temperature 1	-0.66	$p < 0.001$
Outdoor Temperature 2	-0.002	.92
Outdoor Temperature 3	-0.05	$p < 0.001$
Indoor Rel. Humidity	0.33	$p < 0.001$
Outdoor Rel. Humidity	-0.36	$p < 0.001$
R^2=0.28		

Table A.12: Markov Chain Rose Model for Window Closing (WC)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.44	$p < 0.001$
Time 1	-0.08	.24
Time 2	0.02	.86
Time 3	0.44	$p < 0.001$
Time 4	0.15	.0003
Indoor Temperature 1	0.05	.43
Indoor Temperature 2	0.10	$p < 0.001$
Indoor Temperature 3	0.03	.02
Outdoor Temperature 1	0.60	$p < 0.001$
Outdoor Temperature 2	-0.07	.04
Outdoor Temperature 3	-0.1	$p < 0.001$
Indoor Rel. Humidity	-0.50	$p < 0.001$
Outdoor Rel. Humidity	0.81	$p < 0.001$
R^2=0.21		

Table A.13: Markov Chain Rose Model for Ac On (ACon)

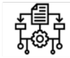
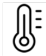
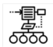
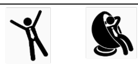
Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.19	$p < 0.001$
Time 1	0.69	$p < 0.001$
Time 2	-0.73	$p < 0.001$
Time 3	-0.27	$p < 0.001$
Time 4	0.23	$p < 0.001$
Indoor Temperature 1	0.72	$p < 0.001$
Indoor Temperature 2	0.09	$p < 0.001$
Indoor Temperature 3	-0.03	$p < 0.001$
Outdoor Temperature 1	0.53	$p < 0.001$
Outdoor Temperature 2	0.06	$p < 0.001$
Outdoor Temperature 3	-0.09	$p < 0.001$
Indoor Rel. Humidity	-0.18	$p < 0.001$
Outdoor Rel. Humidity	0.36	$p < 0.001$
R^2=0.23		

Table A.14: Markov Chain Rose Model for AC Off (ACoff)

Predictor	*Coefficient (b)*	*Sig.*
Intercept	-0.33	$p < 0.001$
Time 1	-0.35	$p < 0.001$
Time 2	-0.16	.08
Time 3	0.40	$p < 0.001$
Time 4	0.14	$p < 0.001$
Indoor Temperature 1	-0.43	$p < 0.001$
Indoor Temperature 2	0.002	.94
Indoor Temperature 3	0.04	.0003
Outdoor Temperature 1	0.30	$p < 0.001$
Outdoor Temperature 2	0.04	.06
Outdoor Temperature 3	-0.05	.001
Indoor Rel. Humidity	-0.01	.70
Outdoor Rel. Humidity	0.17	.0002
R^2=0.14		

A.5 Icons

Credit attributions of the icons used in this work. Icons not listed here either did not require attribution or were created by the author of this work.

	designed by Eucalyp from Flaticon
	designed by Pixel perfect from Flaticon
	designed by Becris from Flaticon
	Designed by H, Alberto Gongorra
	Designed by Iconsmind www.iconsmind.com
	Designed by Saeful Muslim, ID
	Designed by Studio 365
	Provided by Icons8
	Designed by Gan Khoon Lay
	Dessigned by Freepik from Flaticon
	Freedom icon Freedom Icon #72187
	Freedom icon Freedom Icon #72179
	Designed by S Leonov
	Designed by ahasoft

List of Figures

List of Tables

Bibliography

[1] ABNT. NBR 15220 Desempenho térmico de edificações - Parte 3: Zoneamento bioclimático brasileiro e diretrizes construtivas para habitações unifamiliares de interesse social. *Associação Brasileira de Normas Técnicas*, pages 1–23, 2003.

[2] ABNT. *ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 15220: Desempenho Térmico de edificações*. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, Rio de Janeiro, 2005.

[3] Rune Andersen, Valentina Fabi, Jorn Toftum, Stefano P. Corgnati, and Bjarne W. Olesen. Window opening behaviour modelled from measurements in Danish dwellings. *Building and Environment*, 69:101–113, 2013.

[4] Rune K. Andersen, Valentina Fabi, and Stefano P. Corgnati. Predicted and actual indoor environmental quality: Verification of occupants' behaviour models in residential buildings. *Energy and Buildings*, 127:105–115, 2016.

[5] Rune Vinther Andersen, Bjarne Olesen, and Jørn Toftum. Simulation of the Effects of Occupant Behaviour on Indoor Climate and Energy Consumption. In *Clima 2007 WellBeing Indoors*, page 7, 2007.

[6] ASHRAE-140. AMERICAN SOCIETY OF HEATING, REFRIGERATING AND AIR-CONDITIONING ENGINEERS. *ANSI/ASHRAE Standard 140: Standard method of test for the evaluation of building energy analysis computer programs*, 2004.

[7] ASHRAE-90.1. Energy standard for building except low-rise residential buildings. *American Society of Heating, Refrigerating and Air-Conditioning Engineers*, 2019.

[8] Godfried Augenbroe. Trends in building simulation. *Building and Environment*, 37(8-9):891–902, 8 2002.

[9] Nick Baker and Mark Standeven. Thermal comfort for free-running buildings. *Energy and Buildings*, 23(3):175–182, 3 1996.

[10] Douglas Bates, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 10 2015.

[11] Marco Beccali, Giuseppina Ciulla, Valerio Lo Brano, Alessandra Galatioto, and Marina Bonomolo. Artificial neural network decision support tool for assessment of the energy performance and the refurbishment actions for the non-residential building stock in Southern Italy. *Energy*, 137:1201–1218, 10 2017.

[12] C. Becchio, C. Bello, S.P. Corgnati, and L. Ingaramo. Influence of Occupant Behaviour Lifestyle on an Italian Social Housing. *Energy Procedia*, 101:1034–1041, 11 2016.

[13] BEN. Balanço Energético Nacional 2016: Ano base 2015. Technical report, Empresa de Pesquisa Energética, Rio de Janeiro, 2016.

[14] BEN. Balanço Energético Nacional 2020: Ano base 2019. Technical report, Empresa de Pesquisa Energética, Rio de Janeiro, 2020.

[15] Tania Berger, Christoph Amann, Herbert Formayer, Azra Korjenic, Bernhard Pospischal, Christoph Neururer, and Roman Smutny. Impacts of climate change upon cooling and heating energy demand of office buildings in Vienna, Austria. *Energy and Buildings*, 80:517–530, 9 2014.

[16] Philomena M Bluyssen. The Indoor Environment Handbook. In *Proceedings of the ICE - Engineering Sustainability*, pages 234–234, London, 2012. Earthscan.

[17] Konstantin Borovkov. *Elements of Stochastic Modelling*. World Scientific, Singapore, 2nd edition, 2014.

[18] Gail S. Brager, Gwelen Paliaga, and Richard de Dear. Operable Windows, Personal Control, and Occupant Comfort (RP-1161). *ASHRAE Transactions*, 110:17–35, 2004.

[19] R. R. Brugnera. *Potencial de economia de energia em edifícios de escritórios com estratégias de ventilação híbrida*. Dissertação (Mestrado)-Instituto de Arquitetura e Urbanismo, Universidade de São Paulo, São Carlos, 2014.

[20] Buildings. Pesquisa Imobiliária Corporativa, 2021.

[21] Annamaria Buonomano and Max H Sherman. Analysis of residential hybrid ventilation performance in US climates. In *30th AIVC Conference*, pages 1–6, 2009.

[22] M. Cabanac. Physiological Role of Pleasure. *Science*, 173(4002):1103–1107, 1971.

[23] Davide Calì, Rune Korsholm Andersen, Dirk Müller, and Bjarne W. Olesen. Analysis of occupants' behavior related to the use of windows in German households. *Building and Environment*, 103:54–69, 2016.

[24] Davide Calì, Tanja Osterhage, Rita Streblow, and Dirk Müller. Energy performance gap in refurbished German dwellings: Lesson learned from a field test. *Energy and Buildings*, 127:1146–1158, 9 2016.

[25] Gloria Calleja Rodríguez, Antonio Carrillo Andrés, Fernando Domínguez Muñoz, José Manuel Cejudo López, and Yi Zhang. Uncertainties and sensitivity analysis in building energy simulation using macroparameters. *Energy and Buildings*, 67:79–87, 12 2013.

[26] Salvatore Carlucci, Francesco Causone, Silvia Biandrate, Martina Ferrando, Amin Moazami, and Silvia Erba. On the impact of stochastic modeling of occupant behavior on the energy use of office buildings. *Energy and Buildings*, 246:111049, 9 2021.

[27] Salvatore Carlucci, Marilena De Simone, Steven K. Firth, Mikkel B. Kjærgaard, Romana Markovic, Mohammad Saiedur Rahaman, Masab Khalid Annaqeeb, Silvia Biandrate, Anooshmita Das, Jakub Wladyslaw Dziedzic, Gianmarco Fajilla, Matteo Favero, Martina Ferrando, Jakob Hahn, Mengjie Han, Yuzhen Peng, Flora Salim, Arno Schlüter, and Christoph van Treeck. Modeling occupant behavior in buildings. *Building and Environment*, 174(February):106768, 2020.

[28] K.R.M. Carrières. *Desempenho Térmico e Consumo Energético de Edifícios de Escritórios em São Carlos , SP*. PhD thesis, Universidade Estadual de Campinas, UNICAMP, 2007.

[29] CBE. Center for the Built Environment (CBE), About Mixed-Mode[Online]. Available: http://www.cbe.berkeley.edu/mixedmode/aboutmm.html, 2017 (accessed: 31.05.17)., 2013.

[30] CEPAGRI. Centro de Pesquisas Meteorológicas e Climáticas Aplicadas a Agricultura, 2018.

[31] Shuo Chen, Guomin Zhang, Xiaobo Xia, Yixing Chen, Sujeeva Setunge, and Long Shi. The impacts of occupant behavior on building energy consumption: A review. *Sustainable Energy Technologies and Assessments*, 45:101212, 6 2021.

[32] Shuqin Chen, Weiwei Yang, Hiroshi Yoshino, Mark D. Levine, Katy Newhouse, and Adam Hinge. Definition of occupant behavior in residential buildings and its application to behavior analysis in case studies. *Energy and Buildings*, 104:1–13, 10 2015.

[33] S. P. Corgnati, F. Cotana, S. D'Oca, A. L. Pisello, and F. Rosso. A Cost-Effective Human-Based Energy-Retrofitting Approach. In *Cost-Effective Energy Efficient Building Retrofitting: Materials, Technologies, Optimization and Case Studies*, pages 219–255. Elsevier Inc., 2 2017.

[34] Usman Ijaz Dar, Laurent Georges, Igor Sartori, and Vojislav Novakovic. Influence of occupant's behavior on heating needs and energy system performance: A case of well-insulated detached houses in cold climates. *Building Simulation*, 8(5):499–513, 2015.

[35] Julia Day, Judy Theodorson, and Kevin Van Den Wymelenberg. Understanding controls, behaviors and satisfaction in the daylit perimeter office: A daylight design case study. *Journal of Interior Design*, 37(1):17–34, 2012.

[36] R. De Vecchi, C. Candido, R. de Dear, and R. Lamberts. Thermal comfort in office buildings: Findings from a field study in mixed-mode and fully-air conditioning environments under humid subtropical conditions. *Building and Environment*, 123:672–683, 10 2017.

[37] Elham Delzendeh, Song Wu, Angela Lee, and Ying Zhou. The impact of occupants' behaviours on building energy analysis: A research review. *Renewable and Sustainable Energy Reviews*, 80:1061–1071, 12 2017.

[38] Max Paul Deuble and Richard John de Dear. Mixed-mode buildings: A double standard in occupants' comfort expectations. *Building and Environment*, 54:53–60, 8 2012.

[39] Simona D'Oca, Valentina Fabi, Stefano P. Corgnati, and Rune Korsholm Andersen. Effect of thermostat and window opening occupant behavior models on energy use in homes. *Building Simulation*, 7(6):683–694, 12 2014.

[40] Bing Dong and Khee Poh Lam. A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. *Building Simulation*, 7(1):89–106, 2 2014.

[41] Kelen Almeida Dornelles. Absortância solar de superfícies opacas: base de dados de tintas látex acrílica e PVA e a influência da rugosidade superficial. In *X Encontro nacional e VI Encontro latino americado de conforto no ambiente construído - ENCAC*, pages 612–621, Natal, 2009. ANTAC.

[42] Simona D'Oca, Tianzhen Hong, and Jared Langevin. The human dimensions of energy use in buildings: A review. *Renewable and Sustainable Energy Reviews*, 81:731–742, 1 2018.

[43] Silvia Erba, Francesco Causone, and Roberto Armani. The effect of weather datasets on building energy simulation outputs. In *Energy Procedia*, volume 134, pages 545–554. Elsevier Ltd, 10 2017.

[44] Valentina Fabi, Rune Korsholm Andersen, and Stefano Corgnati. Verification of stochastic behavioural models of occupants' interactions with windows in residential buildings. *Building and Environment*, 94(P1):371–383, 2015.

[45] Valentina Fabi, Rune Vinther Andersen, Stefano Corgnati, and Bjarne W. Olesen. Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models. *Building and Environment*, 58:188–198, 2012.

[46] Valentina Fabi, Rune Vinther Andersen, Stefano P. Corgnati, and Bjarne W. Olesen. A methodology for modelling energy-related human behaviour: Application to window opening behaviour in residential buildings. *Building Simulation*, 6(4):415–427, 2013.

[47] R Fritsch, A Kohler, M Nyg, / ˜rd-Ferguson, and J.-L Scartezzini. A Stochastic Model of User Regarding Ventilation Behaviour. *Building and Environment*, 25(2):173–181, 1990.

[48] Isabella Gaetani, Pieter Jan Hoes, and Jan L.M. Hensen. Estimating the influence of occupant behavior on building heating and cooling energy in one simulation run. *Applied Energy*, 223:159–171, 8 2018.

[49] Dani Gamerman and Hedibert Lopes. *Markov Chain Monte Carlo: Stochastic simulation for Bayesian Inference*. Chapman & Hall/CRC Taylor & Francis Group, Boca Raton, 2nd edition, 2006.

[50] Yaxiu Gu, Tong Cui, Kun Liu, Fei Yang, Shengpeng Wang, Hui Song, Qian Qi, Qinglong Meng, and Yanpeng Li. Study on influencing factors for occupant window-opening behavior: Case study of an office building in Xi'an during the transition season. *Building and Environment*, 200:107977, 8 2021.

[51] H. Burak Gunay, William O'Brien, and Ian Beausoleil-Morrison. Implementation and comparison of existing occupant behaviour models in EnergyPlus. *Journal of Building Performance Simulation*, 9(6):567–588, 2016.

[52] Frédéric Haldi, Davide Calì, Rune Korsholm Andersen, Mark Wesseling, and Dirk Müller. Modelling diversity in building occupant behaviour: a novel statistical approach. *Journal of Building Performance Simulation*, 1493(January):1–18, 2016.

[53] Frédéric Haldi, Davide Calì, Rune Korsholm Andersen, Mark Wesseling, and Dirk Müller. Modelling diversity in building occupant behaviour: a novel statistical approach. *Journal of Building Performance Simulation*, 10(5-6):527–544, 2017.

[54] Frédéric Haldi and Darren Robinson. On the behaviour and adaptation of office occupants. *Building and Environment*, 43(12):2163–2177, 2008.

[55] Frédéric Haldi and Darren Robinson. Interactions with window openings by office occupants. *Building and Environment*, 44(12):2378–2395, 2009.

[56] Frédéric Haldi and Darren Robinson. The impact of occupants' behaviour on building energy demand. *Journal of Building Performance Simulation*, 4(4):323–338, 12 2011.

[57] E Harrell and Maintainer Frank. Package 'rms' Title Regression Modeling Strategies, 2018.

[58] Sebastian Herkel, Ulla Knapp, and Jens Pfafferott. Towards a model of user behaviour regarding the manual control of windows in office buildings. *Building and Environment*, 43(4):588–600, 2008.

[59] P. Hoes, J.L.M. Hensen, M.G.L.C. Loomans, B. de Vries, and D. Bourgeois. User behavior in whole building simulation. *Energy and Buildings*, 41(3):295–302, 3 2009.

[60] Michael J. Holmes and Jacob N. Hacker. Climate change, thermal comfort and energy: Meeting the design challenges of the 21st century. *Energy and Buildings*, 39(7):802–814, 7 2007.

[61] Tianzhen Hong. Occupant Behavior: impact on energy use of private offices. In *ASim 2012 - 1st Asia conference of International Building Performance Simulation Association*, page 10, 2013.

[62] Tianzhen Hong, Simona D'Oca, Sarah C. Taylor-Lange, William J N Turner, Yixing Chen, and Stefano P. Corgnati. An ontology to represent energy-related occupant behavior in buildings. Part II: Implementation of the DNAS framework using an XML schema. *Building and Environment*, 94(P1):196–205, 2015.

[63] Tianzhen Hong, Da Yan, Simona D'Oca, and Chien-fei Chen. Ten questions concerning occupant behavior in buildings: The big picture. *Building and Environment*, 114:518–530, 2017.

[64] David W Hosmer and Stanley Lemeshow. *Applied Logistic Regression*. John Wiley & Sons, 1989.

[65] IEA. International Energy Agency, 2012.

[66] IEA. International Energy Agency, 2018.

[67] Kiyosi Itô. *Translations of Mathematical Monographs: Essentials of Stochastic Processes*. American Mathematical Society, Providence, Rhode Island, 2006.

[68] Gareth James, Daniela Witen, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R.* Springer, 2007.

[69] Kathryn B. Janda. Buildings don't use energy: People do. *Architectural Science Review*, 54(1):15–22, 2011.

[70] Bongchan Jeong, Jae-Weon Jeong, and J.S. Park. Occupant behavior regarding the manual control of windows in residential buildings. *Energy and Buildings*, 127:206–216, 9 2016.

[71] Mengda Jia, Ravi S. Srinivasan, and Adeeba A. Raheem. From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. *Renewable and Sustainable Energy Reviews*, 68:525–540, 2 2017.

[72] Rory V. Jones, Alba Fuertes, Elisa Gregori, and Alberto Giretti. Stochastic behavioural models of occupants' main bedroom window operation for UK residential buildings. *Building and Environment*, 118:144–158, 2017.

[73] Amy Kim, Shuoqi Wang, Ji-Eun Kim, and Dorothy Reed. Indoor/Outdoor Environmental Parameters and Window-Opening Behavior: A Structural Equation Modeling Analysis. *Buildings*, 9(4):94, 4 2019.

[74] Jungsoo Kim, Richard de Dear, Thomas Parkinson, and Christhina Candido. Understanding patterns of adaptive comfort behaviour in the Sydney mixed-mode residential context. *Energy and Buildings*, 141:274–283, 4 2017.

[75] Alexander Kowarik. EnvStats, 2018.

[76] Birgit Krausse, Malcolm Cook, and Kevin Lomas. Environmental performance of a naturally ventilated city centre library. *Energy and Buildings*, 39(7):792–801, 2007.

[77] Y. Laaroussi, M. Bahrar, M. El Mankibi, A. Draoui, and A. Si-Larbi. Occupant presence and behavior: A major issue for building energy performance

simulation and assessment. *Sustainable Cities and Society*, 63:102420, 12 2020.

[78] Jun Li, Zhun (Jerry) Yu, Fariborz Haghighat, and Guoqiang Zhang. Development and improvement of occupant behavior models towards realistic building performance simulation: A review. *Sustainable Cities and Society*, 50:101685, 10 2019.

[79] Nan Li, Juncheng Li, Ruijuan Fan, and Hongyuan Jia. Probability of occupant operation of windows during transition seasons in office buildings. *Renewable Energy*, 73:84–91, 2015.

[80] Marta A. R. Lopes, Carlos Henggeler Antunes, Ana Reis, and Nelson Martins. Estimating energy savings from behaviours using building performance simulations. *Building Research & Information*, 3218(April):1–17, 2016.

[81] Nicola Lunardon, Giovanna Menardi, and Nicola Torelli. ROSE: a Package for Binary Imbalanced Learning. *R Journal*, 6:82–92, 2014.

[82] Laura Marín-Restrepo, Maureen Trebilcock, and Mark Gillott. Occupant action patterns regarding spatial and human factors in office environments. *Energy and Buildings*, 214:109889, 5 2020.

[83] Romana Markovic, Eva Grintal, Daniel Wölki, Jérôme Frisch, and Christoph van Treeck. Window opening model using deep learning methods. *Building and Environment*, 145:319–329, 11 2018.

[84] Keith A. Markus. Constructs, Concepts and the Worlds of Possibility: Connecting the Measurement, Manipulation, and Meaning of Variables. *Measurement: Interdisciplinary Research & Perspective*, 6(1-2):54–77, 2008.

[85] Vytautas Martinaitis, Edmundas Kazimieras Zavadskas, Violeta Motuzienė, and Tatjana Vilutienė. Importance of occupancy information when simulating energy demand of energy efficient house: A case study. *Energy and Buildings*, 101:64–75, 8 2015.

[86] K. G.M. Moons, A. Rogier T. Donders, E. W. Steyerberg, and F. E. Harrell. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: A clinical example. *Journal of Clinical Epidemiology*, 57(12):1262–1270, 2004.

[87] Sarang Narkhede. Understanding AUC - ROC Curve, 2018.

[88] Sarang Narkhede. Understanding Cofusion Matrix, 2018.

[89] Federica Naspi, Marco Arnesano, Francesca Stazi, Marco D'Orazio, and Gian Marco Revel. Measuring occupants' behaviour for buildings' dynamic cosimulation. *Journal of Sensors*, 2018, 2018.

[90] Leticia De Oliveira Neves, Eduardo Rodrigues Quesada, Camila Anchieta, and Karin Maria Soares Chvatal. Developing user profiles for mixed-mode office buildings operation based on occupant behaviour evaluation. In *Windsor conference: Rethinking comfort*, page 10, Windsor, UK, 2018.

[91] J. Fergus Nicol, Michael A. Humphreys, and Bjarne Olesen. A stochastic approach to thermal comfort - Occupant behavior and energy use in buildings. In *ASHRAE Transactions*, volume 110 PART I, pages 554–568, 2004.

[92] J.F. Nicol. Characterising occupant behavior in buildings: Towards a stochastic model of occupant use of windows, lights, blinds heaters and fans. *Seventh International IBPSA Conference*, pages 1073–1078, 2001.

[93] J.F. Nicol and M.A. Humphreys. Adaptive thermal comfort and sustainable thermal standards for buildings. *Energy and Buildings*, 34(6):563–572, 7 2002.

[94] William O'Brien and H. Burak Gunay. The contextual factors contributing to occupants' adaptive comfort behaviors in offices - A review and proposed modeling framework. *Building and Environment*, 77:77–87, 2014.

[95] J. Page, D. Robinson, N. Morel, and J. L. Scartezzini. A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings*, 40(2):83–98, 2008.

[96] Song Pan, Yingzi Xiong, Yiye Han, Xingxing Zhang, Liang Xia, Shen Wei, Jinshun Wu, and Mengjie Han. A study on influential factors of occupant window-opening behavior in an office building in China. *Building and Environment*, 133:41–50, 4 2018.

[97] J.S. Park and H.J. Kim. A field study of occupant behavior and energy consumption in apartments with mechanical ventilation. *Energy and Buildings*, 50:19–25, 7 2012.

[98] M. C. Peel, B. L. Finlayson, and T. A. McMahon. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11(5):1633–1644, 2007.

[99] Chen Peng, Da Yan, Ruhong Wu, Chuang Wang, Xin Zhou, and Yi Jiang. Quantitative description and simulation of human behavior in residential buildings. *Building Simulation*, 5(2):85–94, 6 2012.

[100] Luis Pérez-Lombard, José Ortiz, and Christine Pout. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 1 2008.

[101] Cristina Piselli and Anna Laura Pisello. Occupant behavior long-term continuous monitoring integrated to prediction models: Impact on office building energy performance. *Energy*, 176:667–681, 6 2019.

[102] Anna Laura Pisello, Veronica Lucia Castaldo, Cristina Piselli, Claudia Fabiani, and Franco Cotana. How peers' personal attitudes affect indoor microclimate and energy need in an institutional building: Results from a continuous monitoring campaign in summer and winter conditions. *Energy and Buildings*, 126:485–497, 2016.

[103] I Richardson, M Thomson, and D Infield. A high-resolution domestic building occupancy model for energy demand simulations A high-resolution domestic building occupancy model for energy demand simula- tions. Energy and. *Buildings*, 40(8):1560–1566, 2008.

[104] H B. Rijal and F Stevenson. Thermal comfort in UK housing to avoid overheating: lessons from a 'Zero Carbon' case study. In *Proceedings of Conference: Adapting to Change: New Thinkingon Comfort Cumberland Lodge, Windsor, UK, 9-11 April 2010. London: Network for Comfort and Energy Use in Buildings, http://nceub.org.uk*, pages 9–11, 2010.

[105] H B Rijal, P Tuohy, M A Humphreys, J F Nicol, A Samuel, and J Clarke. Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings. *Energy and Buildings*, 39(7):823–836, 7 2007.

[106] H B Rijal, P Tuohy, F Nicol, M A Humphreys, A Samuel, and J Clarke. Development of an adaptive window-opening algorithm to predict the thermal comfort, energy use and overheating in buildings. *Journal of Building Performance Simulation*, 1(1):17–30, 2008.

[107] Hom Rijal, Michael Humphreys, and Fergus Nicol. Adaptive Thermal Comfort in Japanese Houses during the Summer Season: Behavioral Adaptation and the Effect of Humidity. *Buildings*, 5(3):1037–1054, 2015.

[108] Hom B Rijal, Michael A. Humphreys, and J. Fergus Nicol. Development of a Window Opening Algorithm to Predict Occupant Behavior in Japanese Houses. *PLEA 2014 - 30th International PLEA Conference*, 2013(December):90–97, 2014.

[109] Hom B. Rijal, Michael Alexander Humphreys, and J. F. Nicol. How do the occupants control the temperature in mixed-mode buildings? Predicting the use of passive and active controls. In *Proceedings of Conference: Air Conditioning and the Low Carbon Cooling Challenge - Windsor 2008 Conference*, 2008.

[110] Hom B. Rijal, Paul Tuohy, Michael A. Humphreys, J. Fergus Nicol, and Aizaz Samuel. An algorithm to represent occupant use of windows and fans including situation-specific motivations and constraints. *Building Simulation*, 4(2):117–134, 2011.

[111] M Roriz. Arquivos Climáticos de Municípios Brasileiros. *ANTAC – Associação Nacional de Tecnologia do Ambiente Construído. Grupo de Trabalho sobre Conforto e Eficiência Energética de Edificações*, 2012.

[112] RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., 2020.

[113] Ricardo Forgiarini Rupp, Rune Korsholm Andersen, Jørn Toftum, and Enedir Ghisi. Occupant behaviour in mixed-mode office buildings in a subtropical climate: Beyond typical models of adaptive actions. *Building and Environment*, 190:107541, 3 2021.

[114] Ricardo Forgiarini Rupp and Enedir Ghisi. Potencial de economia de energia elétrica através do uso da luz natural e da ventilação híbrida em edifícios comerciais em Florianópolis. *Ambiente Construído*, 13(4):75–86, 12 2013.

[115] Ricardo Forgiarini Rupp and Enedir Ghisi. What is the most adequate method to assess thermal comfort in hybrid commercial buildings located in hot-humid summer climate? *Renewable and Sustainable Energy Reviews*, 29:449–462, 1 2014.

[116] Juan C. Salcido, Adeeba Abdul Raheem, and Raja. R.A. Issa. From simulation to monitoring: Evaluating the potential of mixed-mode ventilation (MMV) systems for integrating natural ventilation in office buildings through a comprehensive literature review. *Energy and Buildings*, 127:1008–1018, 2016.

[117] Caroline Antonelli Santesso and Karin Maria Soares Chvatal. Consumo energético de salas de escritório climatizadas de modo misto em São Paulo: considerações sobre o projeto. *Ambiente Construído*, 18(4):119–137, 2018.

[118] Karin Schakib-Ekbatan, Fatma Zehra Çakıcı, Marcel Schweiker, and Andreas Wagner. Does the occupant behavior match the energy concept of the building? – Analysis of a German naturally ventilated office building. *Building and Environment*, 84:142–150, 1 2015.

[119] M Schweiker, M Hawighorst, and A Wagner. The influence of personality traits on occupant behavioural patterns. *Energy and Buildings*, 131:–, 2016.

[120] Marcel Schweiker. *Occupant Behaviour and the Related Reference Levels for Heating and Cooling*. PhD thesis, Tokyo City University, 2010.

[121] Marcel Schweiker, Salvatore Carlucci, Rune Korsholm Andersen, Bing Dong, and William O'Brien. Occupancy and Occupants' Actions. In A. Wagner, William O'Brien, and Bing Dong, editors, *Exploring Occupant Behavior in Buildings. Methods and Challenges*, chapter 2, page 333. Springer, Cham, Switzerland, 2018.

[122] Marcel Schweiker, Salvatore Carlucci, Rune Korsholm Andersen, Bing Dong, and William O'Brien. Occupancy and Occupants' Actions. In Andreas Wagner, William O'Brien, and Bing Dong, editors, *Exploring Occupant Behavior in Buildings. Methids and Challenges*, chapter 2, page 333. Springer, Cham, Switzerland, 2018.

[123] Marcel Schweiker, Frédéric Haldi, Masanori Shukuya, and Darren Robinson. Verification of stochastic models of window opening behaviour for residential buildings. *Journal of Building Performance Simulation*, 5(1):55–74, 2012.

[124] Marcel Schweiker and Masanori Shukuya. Comparison of theoretical and statistical models of air-conditioning-unit usage behaviour in a residential setting under Japanese climatic conditions. *Building and Environment*, 44(10):2137–2149, 2009.

[125] Marcel Schweiker and Andreas Wagner. The effect of occupancy on perceived control, neutral temperature, and behavioral patterns. *Energy and Buildings*, 117:246–259, 2016.

[126] Shanshan Shi and Bin Zhao. Occupants interactions with windows in 8 residential apartments in Beijing and Nanjing, China. *Building Simulation*, 9(2):221–231, 2016.

[127] D Shipworth and G. M. Huebner. Designing Research. In A Wagner, W O'Brien, and B Dong, editors, *Exploring Occupant Behavior in Buildings. Methods and Challenges*, chapter 3, page 333. Springer, Cham, Switzerland, 2018.

[128] Veronica Soebarto and Helen Bennetts. Thermal comfort and occupant responses during summer in a low to middle income housing development in South Australia. *Building and Environment*, 75:19–29, 5 2014.

[129] Francesca Stazi and Federica Naspi. *Impact of Occupants' Behaviour on Zero-Energy Buildings*. Springer, Cham, Switzerland, 2018.

[130] Francesca Stazi, Federica Naspi, and Marco D'Orazio. A literature review on driving factors and contextual events influencing occupants' behaviours in buildings. *Building and Environment*, 118:40–66, 6 2017.

[131] Francesca Stazi, Federica Naspi, and Marco D'Orazio. Modelling window status in school classrooms. Results from a case study in Italy. *Building and Environment*, 111:24–32, 2017.

[132] Koen Steemers and Geun Young Yun. Household energy consumption: A study of the role of occupants. *Building Research and Information*, 37(5-6):625–637, 2009.

[133] Freek Stulp and Olivier Sigaud. Many regression algorithms, one unified model: A review. *Neural Networks*, 69:60–79, 9 2015.

[134] Cheng Sun, Ran Zhang, Steve Sharples, Yunsong Han, and Hongrui Zhang. A longitudinal study of summertime occupant behaviour and thermal comfort in office buildings in northern China. *Building and Environment*, 143:404–420, 10 2018.

[135] Cheng Sun, Ran Zhang, Steve Sharples, Yunsong Han, and Hongrui Zhang. Thermal comfort, occupant control behaviour and performance gap – A study of office buildings in north-east China using data mining. *Building and Environment*, 149:305–321, 2 2019.

[136] Kaiyu Sun and Tianzhen Hong. A simulation approach to estimate energy savings potential of occupant behavior measures. *Energy and Buildings*, 136:43–62, 2 2017.

[137] Marina Takasu, Ryozo Ooka, Hom B. Rijal, Madhavi Indraganti, and Manoj Kumar Singh. Study on adaptive thermal comfort in Japanese offices under various operation modes. *Building and Environment*, 118:273–288, 6 2017.

[138] Howard M Taylor and Samuel Karlin. *An introduction to stochastic modeling*. Academic Press, 1994.

[139] Paula van den Brom, Arjen Meijer, and Henk Visscher. Performance gaps in energy consumption: household groups and building characteristics. *Building Research and Information*, 46(1):54–70, 2018.

[140] Prapapong Vangtook and Surapong Chirarattananon. Application of radiant cooling as a passive cooling option in hot humid climate. *Building and Environment*, 42(2):543–556, 2007.

[141] Jörn Von Grabe. The systematic identification and organization of the context of energy-relevant human interaction with buildings - A pilot study in Germany. *Energy Research and Social Science*, 12:75–95, 2 2016.

[142] Haidong Wang and Zhiqiang John. Advances in building simulation and computational techniques : A review between 1987 and 2014. *Energy & Buildings*, 128:319–335, 2016.

[143] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. dplyr: A Grammar of Data Manipulation, 2021.

[144] Xiaoqi Xu, John E. Taylor, Anna Laura Pisello, and Patricia J. Culligan. The impact of place-based affiliation networks on energy conservation: An holistic model that integrates the influence of buildings, residents and the neighborhood context. *Energy and Buildings*, 55:637–646, 12 2012.

[145] Da Yan, Tianzhen Hong, Bing Dong, Ardeshir Mahdavi, Simona D'Oca, Isabella Gaetani, and Xiaohang Feng. IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. *Energy and Buildings*, 156:258–270, 12 2017.

[146] Da Yan, William O'brien, Tianzhen Hong, Xiaohang Feng, H. Burak Gunay, Farhang Tahmasebi, and Ardeshir Mahdavi. Occupant behavior modeling for building performance simulation: current state and future challenges. *Energy and Buildings*, 107:264–278, 8 2015.

[147] Da Yan, Chuang Wang, Xiaohang Feng, and Bing Dong. Validation and Ground Truths. In Andreas Wagner, William O'Brien, and Bing Dong, editors, *Exploring Occupant Behavior in Buildings. Methids and Challenges*, chapter 9, page 333. Springer, Cham, Switzerland, 2018.

[148] Mingyao Yao and Bin Zhao. Window opening behavior of occupants in residential buildings in Beijing. *Building and Environment*, 124:441–449, 11 2017.

[149] Hiroshi Yoshino. Total energy use in buildings Analysis and evaluation methods. Technical report, Institute for Building Environment and Energy Conservation, 2013.

[150] Geun Young Yun. Influences of perceived control on thermal comfort and energy use in buildings. *Energy and Buildings*, 158:822–830, 1 2018.

[151] Geun Young Yun, Hyoin Kim, and Jeong Tai Kim. Thermal and non-thermal stimuli for the use of windows in offices. *Indoor and Built Environment*, 21(1):109–121, 2 2012.

[152] Geun Young Yun and Koen Steemers. Time-dependent occupant behaviour models of window control in summer. *Building and Environment*, 43(9):1471–1482, 2008.

[153] Geun Young Yun, Koen Steemers, and Nick Baker. Natural ventilation in practice: linking facade design, thermal performance, occupant perception and control. *Building Research & Information*, 36(6):608–624, 12 2008.

[154] Yufan Zhang and Peter Barrett. Factors influencing the occupants' window opening behaviour in a naturally ventilated office building. *Building and Environment*, 50:125–134, 2012.

[155] Hengjie Zheng, Fei Li, Hao Cai, and Kai Zhang. Non-intrusive measurement method for the window opening behavior. *Energy and Buildings*, 197:171–176, 8 2019.

[156] Xin Zhou, Tiance Liu, Xing Shi, and Xing Jin. Case study of window operating behavior patterns in an open-plan office in the summer. *Energy and Buildings*, 165:15–24, 4 2018.