AUGUST 2017

INDOOR-AIR QUALITY IMPROVEMENT AND REAL-TIME OCCUPANCY DETECTION BY MEANS OF SENSOR DATA

PATRICIA-ALEXANDRA DIACONU

CONSTRUCTION MANAGEMENT
AND ENGINEERING 2015-2017

Technische Universiteit
Eindhoven
University of Technology

Indoor-air quality improvement and real-time occupancy detection by means of sensor data

In partial fulfillment of the requirements for the degree of *Master of Science in Construction Management and Engineering* at *Eindhoven University of Technology*
To be defended publicly on *Thursday, 24th August 2017*

Personal information:
Author: Patricia-Alexandra Diaconu
Student ID: 0978799
E-mail address: patriciadiaconu90@gmail.com

Graduation Committee:
Chairman: prof. dr. ir. B. (Bauke) de Vries
Graduation supervisor: dr. D. (Dujuan) Yang
Graduation supervisor: dr. ir. A.D.A.M. (Astrid) Kemperman

Institute:
University: Eindhoven University of Technology
Faculty: Faculty of the Built Environment
Department: Construction Management and Engineering

Graduation Company:
Name: Inteliments Netherlands B.V.
Address: Apparaten Fabriek STRIJP-S
Torenallee 26-50, 5617 BD Eindhoven

Technische Universiteit Eindhoven University of Technology

Table of Contents

Acknowledgements

The present report represents my graduation thesis concluding the Masters of Science Construction Management and Engineering at Eindhoven University of Technology. Performing this research would not have been possible without the support of a number of people, who helped me throughout the process, and provided important feedback for the good development of my project. Furthermore, studying at the TU/e was an insightful experience, which gave me a good overview on the existing innovative concepts in the built environment. Performing this research contributed to my personal development and helped me understand which aspects are important when designing the occupancy of a building, as well as how much technology influences our daily lives.

I would like to thank my supervisors, for their guidance, constructive feedback and the time they invested in my project. I would like to thank my first supervisor, Dr. Dujuan Yang, for helping me in the process of deciding the topic of my thesis, for the inspiring discussions on the chosen topic and for putting me in contact with the Strijp-S smart city project, for potential companies where I could perform my research. Furthermore, I am grateful to my chairman, prof. dr. ir. Bauke de Vries, for providing me guidance in reshaping my topic, for his support in acquiring the data, and for the good cooperation in the first months of developing my thesis. I would also like to thank dr. ir. Astrid Kemperman for the explanations given on the chosen method (Bayesian Belief Network and descriptive analyses), her thorough feedback and for stepping in as my second supervisor.

I would also like to express my regards to the company (Inteliments) where I performed my research, for their warm welcome in the team and for their support. More specifically, I want to thank my contact person Rene Schouwman, for his constant support and answers to all my e-mails and texts, for his explanations and feedback, and to Michal Puza and Tomas Puza for their help with the data collection. Furthermore, I would like to thank Roger Boersma, for putting me in contact with Inteliments, and for providing his support in gathering ground truth data for the thesis.

Conducting this research was a real academic experience, but above all it was an experience contributing to my personal growth. I could not have overcome the barriers along the way without the faith and encouraging words of my colleagues, friends and family. I would like to thank my parents for supporting me along these two years of my Masters, financially and morally, and I hope I made them proud of my achievements. I would also like to thank my dear friend Ruxi, for her creative and innovative design skills, for helping me with the cover design and layout of this master project. Lastly but certainly not least, I am most grateful to Matija, for believing in me and for his encouragement to do better, for his patience, and for the tough job of carefully proof reading my report.

Executive Summary

Real-time occupancy detection in offices positively impacts energy savings in lighting and the Heating, Ventilation and Air Conditioning (HVAC) system. Moreover, research also connected occupancy to the health of the building and its employees, specifically to the indoor air quality and the thermal comfort. The detection of occupancy and the improvements in indoor air quality have become easier, due to the current developments in the Internet of Things (IoT) technology. Sensors can detect a variety of parameters, such as relative humidity, temperature, carbon monoxide (CO) and carbon dioxide (CO_2) emission, amount of plugged in devices or devices connected to the wireless network. These parameters can be connected to the occupancy in the room, and emission patterns can be easily registered for giving suggestions for air quality improvement. Furthermore, compared to private homes, the office space provides an essential challenge in terms of indoor air quality, due to the increased time people are currently spending at their work.

This leads to the following research objective: to perform an analysis of existing indoor conditions for formulating recommendations in terms of air quality and thermal comfort in the office and to examine the usefulness of virtual sensors in real-time occupancy detection. The thesis also studies the impact of occupancy on the indoor air quality controllers (CO_2, humidity and temperature). This research aims to answer three main research questions structured in two parts. The first part refers to the Indoor Air Quality improvement and aims to answer the question of how future states of CO_2, humidity and temperature can be predicted based on historical data. The second part focuses on the real-time occupancy detection, and on checking the accuracy of sensors based on literature review. It also includes an experiment on CO_2 and passive infrared (PIR) sensors for checking the validity of this type of information fusion. Moreover, the third and last research question bridges the two parts, focusing on the influence of real-time occupancy on indoor air quality and the prediction of future states of CO_2, humidity and temperature.

The second chapter of the thesis presents the literature review created based on a concept centric approach. Important concepts were addressed and the conclusions of different authors are discussed based on these concepts. Firstly, the importance of indoor air quality and thermal comfort are discussed, along with their impact on work performance and health related issues. Secondly, the relevance of real-time occupancy detection is illustrated, as well as different occupancy patterns adopted in the thesis are explained. Such patterns refer to the peaks in occupancy during different months, weekdays and times of the day. Different authors actually identify diverse peaks for the months and weekdays, whereas for the time of the day, peaks usually occur before lunch. Thirdly, the usefulness of virtual sensors is evaluated, for answering the second research question. In this respect, the CO_2 sensors offer a non-intrusive approach to occupancy detection providing up to 93% accuracy, while the PIR sensors can provide false registrations. Furthermore, several methods, such as pair placement, or an increased amount of sensors are used for improving the accuracy of the PIR sensors. According to the literature, electromagnetic signals offer the best occupancy detection information, however privacy

concerns need to be taken into account. Lastly, the Bayesian Belief Network method, which is used for detecting occupancy patterns as well as for the future prediction of CO_2, humidity and temperature levels, is discussed and evaluated. Based on previous research the Bayesian Belief Network is selected as a research method in this study.

The third chapter provides the research approaches. Two types of data are collected: historical data and data coming from an experiment. The historical data consists of data on CO_2, humidity and temperature over five months, in two office spaces. The data for the experiment consists of ground truth data (the actual amount of people counted in the room), and data on CO_2, humidity, temperature and PIR collected for one week. First, descriptive analyses are used to analyze the data according to the research objective. Subsequently, Bayesian Belief Networks are estimated, to identify connections and the possibility to generate the mentioned patterns.

In the fourth chapter the results of the analysis are explained. The processed data is used to estimate the Bayesian Belief Networks and to apply different scenarios for answering the research questions. The first conclusions refer to the patterns identified in the two rooms. Some examples for the first room are: primarily healthy levels of CO_2 (up to 450 parts per million - ppm), high temperatures (26 to 30°) and optimal humidity levels (40 to 60%) in September, adequate comfortable temperatures for working in December and uncomfortable humidity (below 40%) for all other months. In what concerns the experiment the results show correlation between the indoor temperature and the occupancy, but no correlation between the humidity and the occupancy. In terms of sensors, the fusion between the PIR and CO_2 strengthens the results, compared to the situation when only one sensor is used. Furthermore, the PIR and CO_2 sensors present a strong correlation in the Bayesian Belief Network.

The study ends with the conclusion, recommendations and discussion chapter, including the answers to the research questions as well as the comparison among literature and results. From a societal point of view, this study brings added value to the field of indoor air quality control, by suggesting improvements on certain months, weeks and weekdays in two office spaces. Also, the study presents potential adjustments of the HVAC system for decreasing energy consumption. Moreover, from a scientific point of view, this research strengthens existing literature by connecting the indoor air quality system with the occupancy. The estimated model shows a high correlation between the PIR and the CO_2 sensor and between the CO_2 emission and the number of people present in the room. Recommendations for the company where the research was performed are illustrated. Some recommendation examples are: introducing a humidifier in the winter months, increasing the ventilation and decreasing the heating before employees arrive in the office, or including the occupancy in the air quality platform based on the CO_2 emission. Lastly, the limitations and future research are presented, and the study ends with a personal reflection on the topic and on the achievements realized in the research.

Abstract

Indoor air quality and thermal comfort play important roles in the productivity and health of the employees in an office space. Moreover, detecting the actual occupancy of the building contributes to the air quality of the indoor space and the increase of energy savings, through the adjustments made in the HVAC system based on this actual occupancy. Furthermore, detecting real-time occupancy can be realized with various virtual sensors such as: Electromagnetic Signals, CO_2 or passive-infrared (PIR); each sensor presents several benefits and downsides. Therefore, the current research investigated mainly two aspects: indoor air quality and real-time occupancy sensors; and the correlation between these two. More specifically, indoor air quality is quantified in this research by CO_2 emission, temperature and humidity levels. Patterns in two office spaces are predicted based on historical data, based on the month, weekday and time of the day. The accuracy of different sensors for occupancy detection is analyzed based on literature and the connection between occupancy and the indoor air quality parameters is assessed. For this purpose, descriptive analyses and the Bayesian Belief Network estimation are used, by generating scenarios that could help analyze the occupancy and indoor air quality patterns. The estimated network models show that the CO_2 is dependent on the actual occupancy in the office space. The temperature depends on the CO_2 emission and therefore on the occupancy, however the humidity is not influenced by any of the two.

List of Abbreviations

(according to the appearance in the report)

HVAC = Heating, Water and Air-Conditioning
IAQ = Indoor air quality
CO_2 = carbon dioxide
RH = relative humidity
PIR = Passive-Infrared,
GPS = Global Positioning System
IoT = Internet of Things
BBN = Bayesian Belief Network
RQ = Research Question
SQ = Sub-Question
RFID = Radio frequency identification
ASHRAE = American Society of Heating, Refrigerating and Air-Conditioning Engineers
RTLS = Real-Time Location System
UWB = Ultra-wideband technology
RF = Radio Frequency
WLAN = wireless local area network
GSM = Global System for Mobile Communications
WSAN = Wireless Sensors and Actuators Network
CRF = Conditional Random Field
CPT = Conditional probability table
EM = Expectation maximization algorithm
GTT = Greedy Thick Thinning
S2M = Seats2Meet
Ppm = parts per million

List of figures

List of tables

1

INTRODUCTION

1.1. Introduction

Due to the increase of time spent indoors at work, people are sitting eight hours or more in office rooms. With this increase, the satisfaction with the thermal environment is and becomes more important as it increases productivity and leads to health benefits (Huizenga, Abbaszadeh, Zagreus, & Arens, 2006). Depending on the type of work and schedule, the time spent inside the office can vary and therefore room occupancy differs. The demand on energy systems can be inaccurate due to the systems being operational but the room being actually unoccupied for a long period of time. This happens because the real-time occupancy is usually less than the designed population for the office building. According to Zhao, Zeiler, Boxem, & Labeodan (2015), energy waste occurs when the supplies from energy systems are higher than the demands. More specifically, about 10-20% of Heating, Water and Air-Conditioning (HVAC) consumption and 30% of the lighting consumption could be reduced if real-time occupancy is taken into account.

This chapter presents an introduction to the topic of indoor air quality (IAQ) and real-time occupancy in the office buildings. The problem statement is described from two perspectives: the practical aspect and the scientific one. Furthermore, the research objective, research framework, questions, and sub-questions are formulated for solving the defined problems. Finally, the importance from a social and scientific perspective is explained, together with the steps taken in each chapter of the study.

1.2. Problem Statement

In this section, the problem addressed in the research is described. Section 1.2.1 addresses the problem from a practical point of view, making reference to the impact of real-time occupancy on the air quality in the room and the indoor comfort, as well as the energy waste generated by the currently designed occupancy in comparison to the real-time occupancy. The second part of the problem statement, section 1.2.2, describes the problem from a scientific and theoretical perspective. This section addresses the relevance of using different sensors for detecting occupancy and using different sensors for assessing indoor air quality.

1.2.1. Practical Problem

Practical knowledge can be compared to applied knowledge, or the knowledge gained by doing. The practical problem refers, therefore, to the social implications of the studied topic. Specifically, the problem addressed is, on one hand, the increase in energy usage caused by the excessive and inappropriate use of office appliances and the HVAC system. On the other hand, the decrease in indoor comfort and air quality, caused by increased room occupancy and use of the HVAC system is also a problem addressed by this research.

Energy waste

The major energy waste is coming from two important energy consumers: lighting and the HVAC systems. On this matter, several authors proved a significant reduction in energy usage or electricity consumption based on real-time occupancy measurements.

Specifically, Erickson, et al. (2009) demonstrated a reduction in energy usage after creating a control strategy based on real-time occupancy measurements of a multi-function building. The created models predict room usage and help in controlling the HVAC systems. Following the same line of ideas, Rosiek & Batlles (2013) also demonstrated a reduction in electrical energy usage, by connecting the real-time occupancy detection with the chilled-water storage tanks. In their experiment they managed to improve the efficiency of solar-assisted air-conditioning systems and hence, to reduce their electricity consumption.

The above-mentioned achievements suggest that there is a continuous need for reducing energy consumption and that detecting real-time occupancy can play an important role in achieving this. To complement their generated occupancy profile, Masoudifar, Hammad, & Rezaee (2014) suggest a future study where "this profile can be used to proactively optimize the building energy consumption while responding to comfort preferences of the occupants and to support the implementation of demand-driven HVAC, lighting, and IT equipment operations."

Indoor air quality (IAQ)

Indoor comfort is influenced by a combination of complex factors, such as meteorological conditions, building characteristics, the HVAC system, or the occupancy, as mentioned by Szczurek, et al. (2016). Actually, Szczurek, et al. (2016) mention that "occupancy is one of the major factors, which affect indoor air quality." The impact of occupancy refers to the impact of the human activity on the building (e.g.: presence or movement within the space, opening/closing windows, operating the HVAC systems etc.), the lifestyle (e.g.: tobacco smoking, plant and pet drugs, cooking etc.) or the operation schedule of the building. Furthermore, the number of people inside the building is related to the emission of various chemical substances: carbon dioxide (CO_2), water vapor or organic compounds, which affect IAQ (Szczurek, et al., 2016).

On the topic of the correlation between occupancy and the indoor air quality, Szczurek, et al. (2016) characterized IAQ by the variation of CO_2 concentration in time and reflected the occupancy profile through the internal structure of CO_2 concentration time series. It was observed that when repeating a certain occupancy profile the IAQ is not influenced in the same way. However, the occurrence of a particular occupancy profile can be detected through considering the CO_2 concentration in time.

The above-mentioned research indicates the need of IAQ improvement, and while indoor air quality is impacted by various factors, real-time occupancy can be an important determinant. Occupancy can determine the heating and cooling loads or ventilation rates necessary to maintain a certain thermal comfort, and it affects CO_2 concentration (which impacts the IAQ), as well as air quality indexes such as temperature and relative humidity (RH) (Szczurek, Maciejewska, & Pietrucha, 2017).

1.2.2. Scientific Problem

From a scientific perspective, the studied problem should bring added value in the theoretical field and cover an existing gap in the literature. Moreover, the scientific knowledge creates the context behind the practical problem and it is based on the experience of different authors. Hence, by constructing on the identified practical problems, this study addresses two scientific problems: the detection of real-time occupancy in the office building by means of sensor data and the improvement of IAQ with the help of sensors.

Detecting real-time occupancy in the office

Real-time occupancy in the office building has become a problem that many authors are addressing, mainly due to its impact on the energy consumption. More specifically, research is currently ongoing in the field of occupancy detection by means of sensor data.

In this regard, various types of sensors are being used for detecting occupancy. On this subject, Zhao, Zeiler, Boxem, & Labeodan (2015) conducted an experiment where different sensor types were used for detecting real-time occupancy. The study aims at developing a generic and low-cost solution, therefore the used sensors are Passive-Infrared (PIR), keyboard and mouse sensors, chair and door sensors, light switch, Wi-Fi connection and Global Positioning System (GPS) location. Their results show that the virtual occupancy sensors (Wi-Fi and GPS), are very effective in occupancy detection. Furthermore, in the experiment conducted by Labeodan, Aduda, Zeiler, & Hoving (2015), a mechanical switch is introduced in cushions placed on chairs, in a conference room. The study shows the benefit of chair sensors for detecting occupancy over systems using sensing techniques such as strain and vibration. Literature provides research and experiments on real-time occupancy coming from sensor data. Several sensors can be used to assess whether the office is occupied or not. In this respect, there are two types of sensors that can be used, as described by Zhao, Zeiler, Boxem, & Labeodan (2015):

- virtual occupancy sensors developed at the room level (chair and table sensors, keyboard and mouse sensors, or PIR sensors detecting motion in different room areas)
- occupancy access information from building management systems or real-time Wi-Fi connection from smart devices

However, in regards to environmental sensors detecting additional aspects that can be connected to occupancy (levels of humidity, the amount of emitted CO_2, noise levels etc.), there is limited literature suggesting relevant results. Using one single type of sensors can most often fail, due to, for example, the uncertainties in the physical activities of the occupants, or the lack of movement for a certain period of time. More concretely, each sensor has a series of drawbacks that can be solved by using a fusion of sensors. In this regard, many authors suggest a combination of sensors when it comes to occupancy detection. The first scientific problem addressed in this study is, therefore, the suitability of the sensors used for detecting occupancy in the office building.

Monitoring IAQ with sensors

Building the context in respect to the practical problem of IAQ improvement, many authors address the monitoring of IAQ through the Internet of Things (IoT) technology. Moreover, Marques & Pitarma (2016) emphasize that the ongoing technological developments make it possible to build smart objects in the indoor environment that can communicate with each other. The IAQ, as a wireless, low cost, indoor air quality system, accesses real-time data through web access, which can be used for creating different medical diagnostics. Furthermore, the IAQ allows for various types of sensors to be included, which check different types of contamination (air temperature, humidity, carbon monoxide, carbon dioxide and luminosity), and the results of these sensors can contribute to the improvement of the indoor air quality.

Marques & Pitarma (2016) also mention that this system can be used for the building manager for providing a healthy and productive work environment. The system is useful for improving the air quality and for better understanding the behavior of the environmental parameters. Moreover, Kumar, et al., (2016), highlight that IAQ monitoring by means of sensor technologies would result in temporarily-resolved indoor pollutant data which could give benefits in health impact assessment or emission patterns.

The second scientific problem addressed in the study is, therefore, connected to the indoor air quality monitoring and the possibility to identify emission patterns in CO_2 and RH in the working environment. These could affect the health of the people in the office if not maintained in certain ranges. In this study another problem analyzed is the patterns in temperature, which impact mainly the concentration and productivity in the office environment.

1.3. Research Objective and Research Framework

According to Verschuren & Doorewaard (2010), the research objective concerns the use of knowledge that the research produces, and it is the actual goal of the project. Subsequently, the researcher must determine the information that contributes to achieving the research objective and should create the research framework, which is a schematic representation of the most important research phases (Verschuren & Doorewaard, 2010). The current section is meant to define the aim of the present research and the framework created around the research objective.

Research Objective

The objective of the thesis is to investigate the usefulness of certain virtual occupancy sensors at the room level for determining the real-time occupancy in the different types of office spaces. Furthermore, the proposed research also develops a comprehensive analysis on indoor air quality controllers (CO_2, humidity, and temperature), based on occupancy, time span and meteorological factors.

Research Framework

The research framework captures the research objective in a brief definition and can involve establishing a hypothesis, testing a theory, making a diagnostic, evaluating an intervention or formulating recommendations (Verschuren & Doorewaard, 2010).

Furthermore, achieving the result can be reached by studying an object. "The research object is the phenomenon in empirical reality that you are going to study and that will lead to statements based on the empirical research to be carried out. Depending on the research objective, research objects can take on widely divergent forms. They may involve people, organizational structures, workflows, laws, policy problems, information campaigns, policy memoranda, and so forth." (Verschuren & Doorewaard, 2010). Furthermore, an important part of the research framework is played by the research perspective or the conceptual model, which is used for studying the research object. Together, the elements constitute the research framework, as described by Verschuren & Doorewaard (2010).

The current research project uses the *research objective* as the basis for formulating conclusions and recommendations. By using a reversed method for defining the elements of the research framework, this project uses *data analysis results*, coming from the *research perspective/conceptual models*, which are the Bayesian Belief Network (BBN) models on IAQ and occupancy. The Bayesian Belief Networks are estimated based on the *research objects* which are the data sources: historical data on CO_2, temperature, humidity, time, meteorological information and data coming from an experiment: CO_2, temperature, humidity, time, PIR and ground-truth data. A basic research framework, as described above, can be seen in Figure 1. A complete research framework, which includes also the methodology for obtaining the data files and a classification of the concepts addressed in the literature can be found in Appendix 1.

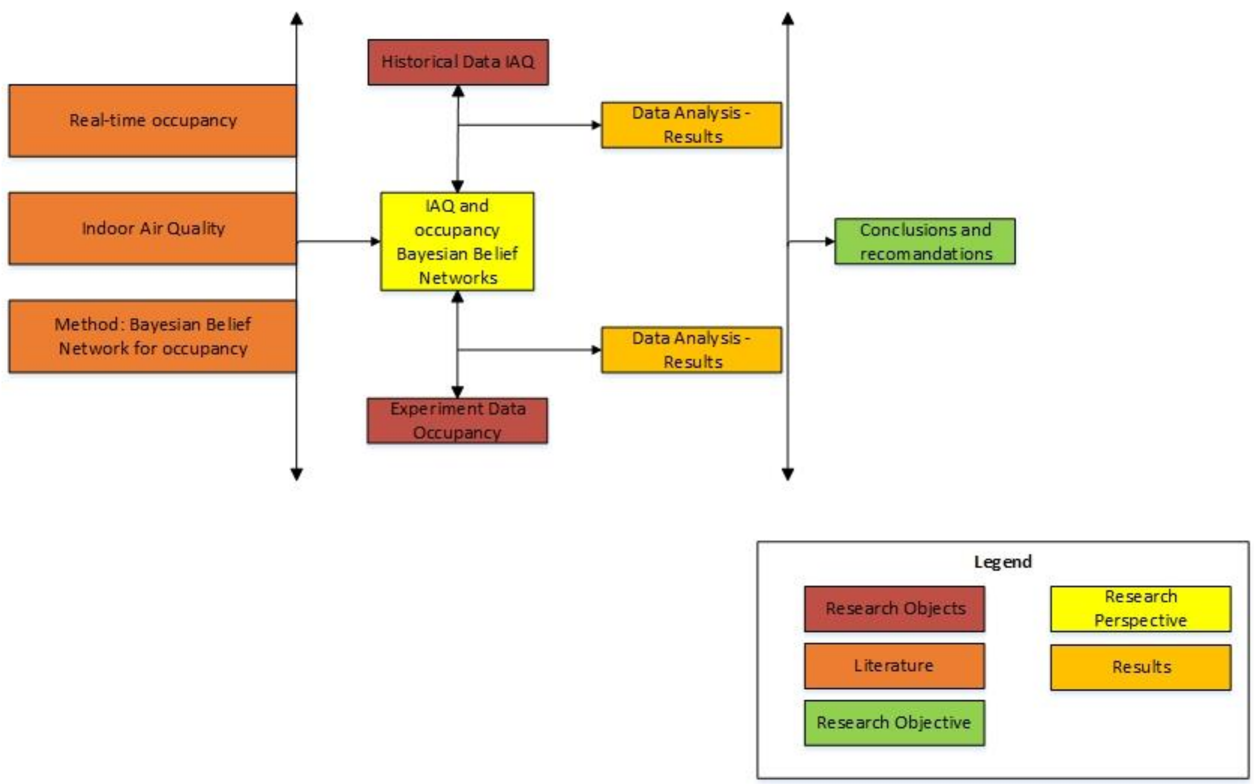


Figure 1: Theoretical Framework

1.4. Research Questions and research sub-questions

Based on the research objective presented in the previous section, the research questions and research sub-questions are formulated. With the aim of investigating both the scientific and the practical problems described, the research questions are split into two parts:

Part 1: Indoor Air Quality improvement (IAQ)

Research Question (RQ): How can future states of CO_2, humidity, and temperature in an office room be predicted based on historical sensor data for improving the indoor air quality?

In order to help to answer this question, one sub-question was formulated. This sub-question aims at providing more insight into the historical sensor data that should be approached:

Sub-Question (SQ): How do the indoor CO_2, humidity, and temperature levels change according to the season, month, day of the week and time of the day?

For answering this sub-question both a literature study and a practical part will be conducted. Actual historical sensor data will be collected for analyzing its impact on the future predictions of the indoor parameters.

Part 2: Real-time occupancy detection

RQ: What sensor data can provide the most accurate information in detecting real-time occupancy information?

This research question will be answered through the literature review, by emphasizing the benefits and drawbacks in using different types of sensor data for occupancy detection and the information fusion that gave positive results in previous studies. Further, into the analysis, a case is developed specifically on PIR and CO_2 sensors for occupancy detection. A research sub-question focusing on this case is therefore formulated:

SQ: How accurate is the information fusion given by the CO_2 and PIR sensors in detecting real-time occupancy?

For answering this sub-question an experiment will be conducted, where information from these two sensor types is analyzed, and furthermore ground truth data is collected.

Finally, occupancy will be linked to the indoor air quality, and for this purpose, a final research question is formulated:

RQ: How do the indoor CO_2, humidity, and temperature levels change according to the real-time occupancy?

The corresponding chapters, where all the research questions and sub-questions will be answered are presented in Table 1.

Table 1: Research questions, sub-questions and chapters

Research Question	Chapter(s)
RQ1: How can future states of CO_2, humidity, and temperature in an office room be predicted based on historical sensor data for improving the indoor air quality?	4.2. Historical Data 5.2. Research question answers
SQ: How do the indoor CO_2, humidity, and temperature levels change according to the season, month, day of the week, and time of the day?	4.2. Historical Data 5.2. Research question answers
RQ2: What sensor data can provide the most accurate information in detecting real-time occupancy information?	2.3. Real-time occupancy detection 2.3.3. Virtual sensors for occupancy detection 5.2. Research question answers
SQ: How accurate is the information fusion given by the CO_2 and PIR sensors in detecting real-time occupancy?	4.3. Experimental Data 5.2. Research question answers
RQ: How do the indoor CO_2, humidity, and temperature levels change according to the real-time occupancy?	4.3. Experimental Data 5.2. Research question answers

1.5. Importance of the thesis

1.5.1. Scientific relevance

From an academic perspective, this study contributes to the existing literature through the link that will be created between the real-time occupancy detection and the improvement in the IAQ parameters. Furthermore, the study analyses the existing literature for concluding which is the best information fusion in terms of occupancy detection, and an experiment is conducted for analyzing the information fusion between the CO_2 and PIR sensors as occupancy detection sensors.

These aspects (analysis of sensors and link between occupancy and IAQ) make this study relevant from a scientific perspective, strengthening the existing literature and bringing added value to this field in the built environment.

1.5.2. Social relevance

From a social perspective, this graduation project contributes to improving the quality of life and energy costs of office occupants. More specifically, the results of the conducted analysis can be used as an input for creating automated adjustments to the HVAC system or giving office occupants indications of the air quality and the recommended actions for improving the air. The temperature, CO_2, and humidity will be monitored, together with the occupancy information coming from PIR sensors. A comparison between the two types of data gathered will be made (the environmental data and the occupancy data), for predicting future CO_2, humidity, and temperature according to presence or absence in the office space. These aspects contribute to the social relevance of this thesis.

1.6. Reading Guide

This thesis is structured into five chapters, where the first chapter captures the introduction of the developed research. In the first chapter the problem is stated, based on which the research objective, theoretical framework and research questions are formulated. The introductive chapter also presents the importance of the thesis from scientific and practical points of view.

The second chapter provides an insight into the existing literature in the topic of indoor air quality and real-time occupancy, as well as into the types of virtual sensors that can be used for approaching these two topics. Furthermore, the literature review provides an analysis on the usability of existing sensor types for detecting real-time occupancy based on previous authors' results.

The third chapter introduces the research approach of the thesis, where the method used for analysis is described, together with the programs and the variables used. Furthermore the method for data collection, processing and analysis is described. This chapter provides an understanding on the methodology used, for a better grasp of the results, which are presented in the fourth chapter.

The fourth chapter is divided in two parts: the historical data analysis and the experimental data analysis, based on the research questions formulated. Both parts illustrate the outcomes, based on the descriptive analyses and the Bayesian Belief Network scenarios.

The fifth chapter presents the conclusions of the research, based on the results and data analysis, the recommendations for the company where the research is performed, limitations of the research as well as future research opportunities.

2

LITERATURE REVIEW

2.1. Introduction

The second chapter presents a review of existing academic literature on the topic of indoor air quality control and real-time occupancy detection in the office building, by means of sensor data. This literature review is created based on the Research Framework exemplified in section 1.3, by showing the theories/concepts connected to the approached topic and the relationship between them. Furthermore, a concept-centric approach to the literature review, as pointed out by Webster & Watson (2002) is used and illustrated in Appendix 2: Research concepts. Hence, the selection of articles covers the most interesting concepts present, while the concepts are also indicated in the concept matrix and linked to each article.

The remaining of this chapter is structured based on the identified concepts. Firstly, for the first part of the research and RQ1, the need for indoor air quality and thermal comfort is presented. This is followed by the experiments realized in detecting real-time occupancy for several indicated reasons, which cover the introduction into the second part of the research. Patterns in occupancy are also illustrated in regards to this concept. Subsequently, the benefits and drawbacks of virtual sensors are presented, for determining the best accuracy for detecting occupancy and for answering RQ2. Finally the method used for analysis is discussed from the point of view of multiple authors.

2.2. Indoor air quality (IAQ) and thermal comfort

With poor air quality in the large, urbanized cities, people become aware of the involved health risks and therefore, an increased amount of government initiatives for resolving air quality problems can be seen (Chau, Hui, & Tse, 2006). According to Kumar, et al. (2016), people typically spend over 90% of their daily time indoors and levels of indoor pollution often exceed those of outdoor pollution.

In terms of performance in the office work, Wargocki & Wyon (2016) specify that air quality conditions, currently accepted by the building occupants, reduce performance with 5-10% for adults. Also, it is stated that the thermal state of the body is what affects performance and that mental performance such as creative thinking, is improved by exposure to a few degrees above thermal neutrality or thermal comfort. Furthermore, based on a literature review, Wargocki & Wyon (2016), determined that 70% of the employees were dissatisfied with the indoor air quality when the quality was altered by the presence of a major pollution source. Ultimately, Wargocki & Wyon (2016) specify that it is well documented that both thermal conditions and IAQ affect the performance of the office work and that the motivation to perform well may be influenced by the indoor environment. On the topic of environmental comfort, Cociorva & Iftene (2017) also point out that comfort is strongly linked to the physical and mental state (e.g.: sadness, happiness, sickness etc.), with the performed activity in the space (e.g.: work, rest, sleep etc.), or the social environment (e.g.: alone, at school, at work, with family etc.).

According to Kumar, et al. (2016), thermal comfort, indoor air pollution, and temperature are interrelated. They are generally governed by the ventilation, and therefore by the HVAC system.

Hence, tackling this problem will be strongly connected to making certain adjustments in the HVAC system. Furthermore, Kik, Tang, Boxem, & Zeiler (2016), conclude that upgrading or installing a balanced ventilation system may achieve expected IAQ and the thermal comfort. Furthermore, according to Stazi, Naspi, & D`Orazio (2017), the indoor temperature is the second driver to intervene on the opening or closing of windows, after the outdoor temperature. In particular, when the indoor temperature reaches 20° C, a substantial increase in windows opening is noticed. At the same time, when the temperature reaches 27° C, window openings are decreased for forbidding the heat to enter from the outside. Also, the interaction between the users and the windows opening are time-related events, different window openings occurring according to the temperature, at different times.

To combat the indoor pollution, Kumar, et al. (2016) mention that air sensors provide a novel way of assessing the environment and the human exposure in real-time by transmitting data to digital platforms for real-time analysis and visualization. Furthermore, Cociorva & Iftene (2017), conclude that the realization of a smart building is conditioned on strict air quality control, by using an intelligent system for air quality assessment, such as an "electronic nose". This concept comprises a network of micro-sensors (gas sensors in this case), a data acquisition system and a system of pattern recognition. The pattern recognition system takes input from the data acquisition system and compares it with previously received data through the learning process. A classification of the received data in defined classes is realized. The functionality of the "electronic nose" concept is particularly interesting for the present graduation project, as similarly, data is collected from air sensors, with the potential integration of patterns from the collected historical data.

Most developed countries regulate different indoor parameters such as temperature range, the RH, and the CO_2. Temperature and RH relate to the thermal comfort, while the CO_2 relates to the human metabolism and can increase in inadequately ventilated indoor spaces (Kumar, et al., 2016). The three parameters are the parameters considered by the present graduation thesis and analyzed in terms of IAQ. This graduation thesis focuses on identifying, based on historical data, the different indoor temperature, RH and CO_2 levels in accordance with the presence in the room and previously registered data. As illustrated in this literature review, the human performance in the office is strongly correlated to the indoor air and thermal comfort. Therefore, the estimated levels for the three parameters and occupancy are meant to contribute to the increase of IAQ and to enhance performance at work. This increase will be realized by giving indications on window openings or control of HVAC based on the future occupancy and desired IAQ levels.

2.3. Real-time occupancy detection

This section gives an insight into the real-time occupancy detection concept by focusing on the second part of the research, and aims to answer RQ2 by providing a discussion on virtual occupancy sensors.

2.3.1. Introduction

Building control is mainly done manually, including the control of heating or switching lights and appliances, and it is typically limited to simple motion detection and a simple indoor climate control based on temperature and CO_2 level (Nguyen & Aiello, 2012). According to the Oxford Dictionaries (2017), the word "real-time" can be defined as: "The actual time during which a process or event occurs". The real-time occupancy can be regarded as the actual time when a person is present in an indoor environment, while the occupancy-based control can be defined as the control of indoor appliances based on human activity. According to Nguyen & Aiello (2012) occupancy-based control can result in up to 40% energy saving for HVAC system, however, an increase in the comfort level continues to be an essential success criterion for ICT-based solutions in order to improve user acceptance of the system.

In their research, Labeodan, Zeiler, Boxem, & Zhao (2015) identified six spatial-temporal properties related to real-time occupancy detection, namely:

- Presence: in a particular room or in a thermal zone
- Location: occupants coordinates in relation to the thermal zone
- Count: number of people present in the thermal zone
- Activity: different body metabolic rate and CO_2 production
- Identity: who is the person
- Track: where was the person before or movement in particular thermal zones

In the same research, Labeodan, Zeiler, Boxem, & Zhao (2015), also mention the methods, functions, and infrastructure needed for occupancy detection.

Method:

- Terminal-based: mobile phone, radio frequency identification (RFID); carried by the user
- Non-terminal: PIR, CO_2; not carried by the user

Function:

- Individualized: provide information based on the individual
- Non-individualized: provide information without knowledge of users identities

Infrastructure:

- Explicit: primary function to detect occupancy (such as PIR)
- Implicit: use of building appliances (such as printers or computers)

The present study is detecting the presence, with a non-terminal method, non-individualized function and with the use of explicit infrastructure. These properties were given based on already existing data and sensors in the research areas, and to avoid privacy issues related to detecting identity, activity or track.

2.3.2. Occupancy patterns

According to Duarte, Wymelenberg, & Rieger (2013), diversity factors in occupancy are hourly fractions for a day, and a profile can be created to make a representative week of general occupancy or equipment operations in a building. For their study, Duarte, Wymelenberg, & Rieger (2013) collected long-term data, allowing to show different results for the time of the day, the day of the week and even month of the year or holidays. This method does not assume any change in occupancy throughout the year, and is referred to it as a deterministic model.

Results show that the months cluster at three different profiles. The months of January, March to June, September and October, cluster together forming a high-level profile, December, February, and July are the medium level profile while November and August create the low-level profile. In what concerns the weekdays, Mondays have the highest level of occupancy, while Friday the lowest, with Tuesday, Wednesday and Thursday not showing any significant difference. In terms of occupancy hours, the peaks occur before lunch and tend to return approximately the same level after lunch, except for Fridays. The study makes a comparison to the ASHRAE 90.1 2004 diversity factors, which is illustrated in Figure 2. The ASHRAE stands for the American Society of Heating, Refrigerating and Air-Conditioning Engineers and is an energy standard for buildings except for low-rise Residential buildings (American Society of Heating, 2007). The comparison illustrates that the factors actually differ with 46%, which may result in misleading results and incorrect system designs.

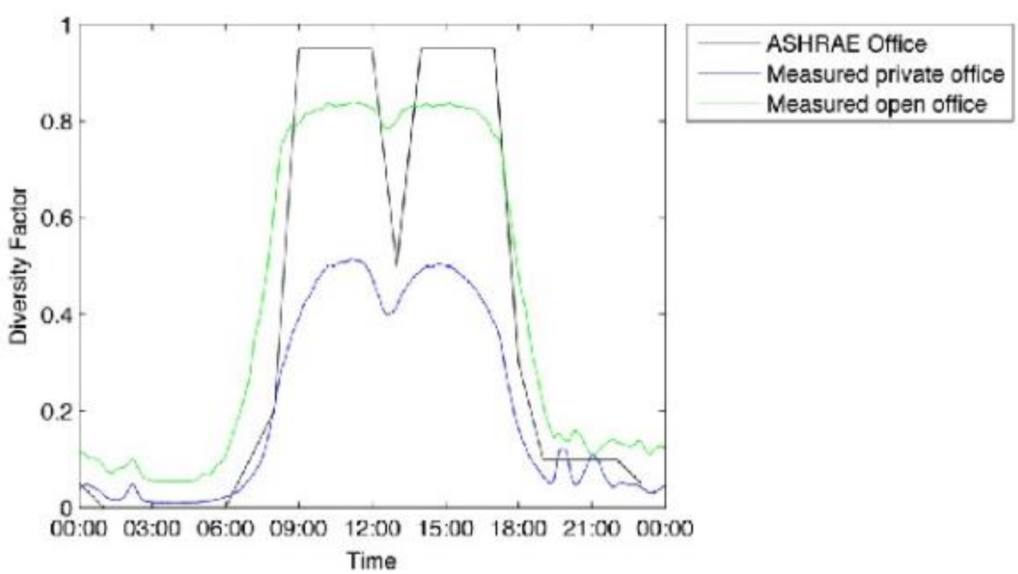


Figure 2: Comparison between ASHRAE 90.1 2004 and the experiment of Duarte, Wymelenberg, & Rieger (2013)

Occupancy patterns were also investigated by Labeodan, Maaijen, & Zeiler (2013). Their results show that the floor occupancy was below 50% for the duration of the experiment and opposite to the results from Duarte, Wymelenberg, & Rieger (2013), the patterns observed vary every day. Highest floor occupancy was found on Tuesdays, while the lowest was on Thursdays. By the means of a linear regression analysis, a comparison between the occupancy profiles and the use of electrical appliances was developed. A high overall coefficient denoted a strong correlation between the occupancy and the electrical appliances. However, the work appliances (personal computers) denote a weak correlation. Surprisingly, no significant change in the energy usage was observed during lunch break, even with the absence of the occupants. Similarly, most rooms were maintained at the same thermal comfort level despite human absence.

In terms of occupancy schedules, Liang, Hong, & Shen (2016) conduct a case study based on occupancy schedule learning and prediction in office buildings. When looking at the occupants' presence based on time periods the occupancy is divided into six periods:

- The night period (7 pm – 6 am) – with 10% occupancy rate
- The going-to-work period (7 am – 9 am) – with growing occupancy rate from 10% to 70%
- The morning period (10 am – 12 pm) – where occupancy rate stays at 80%
- The lunch period (12 pm – 1 pm) – with the occupancy rate dropping below 80%
- The afternoon period (2 pm – 3 pm) – where the occupancy is slightly higher than 80%
- The going-home period (4 pm – 6 pm) – with the decreasing occupancy rate from 70% to 10%

Similar research, created by Masoudifar, Hammad, & Rezaee (2014) shows three possible states for each occupant of an office: in the office and within his/her zone, in the office but not in his/her zone, and absent. Similarly to the study performed by Duarte, Wymelenberg, & Rieger (2013), occupancy charts illustrate the daily presence, while occupancy profiles illustrate daily and weekly occupancy. However, Masoudifar, Hammad, & Rezaee (2014) complement the existing research by using the profiles for calculating energy waste related to the occupant's behavior. Future work is recommended for a proactive optimization of the energy consumption with respect to the comfort preferences of the occupants. As previously described, the comfort inside the office is one of the focuses of this graduation project. In terms of user comfort, Labeodan, de Bakker, Rosemann, & Zeiler (2016), also identified that it is strongly connected to user adaptation. Questionnaires and diaries were sent to occupants, for writing down distractions sensed in the Indoor Environmental Quality parameters and the level of satisfaction with the lighting adjustments. Artificial lighting was the key distraction during the first week the occupancy-based control was implemented, occupants being unsatisfied with the level of light uniformity. This improved in the following week, which was attributed to the user adaptation.

According to Wahl, Despenic, & Amft (2012), by estimating people count per office space, the building control system could adjust the climate and building appliances. However, this is more challenging as people move between office spaces, and due to the large office sizes, installation and maintenance of sensors can become expensive processes. Therefore, for exploring different algorithms for estimating the occupancy in the office, a simulation environment was implemented with the use of inexpensive ambient sensors that can be retrofitted into older buildings. A floor map reflecting an actual office environment (with sensors and occupants) was configured. The concurrent behavior pattern of occupants was reflected through batch scripts, PIR sensors being triggered when occupants moved in their field of view. Different occupant activities, such as getting coffee, visiting colleagues, or attending meetings, were simulated. This study suggests, therefore, that accurate people count needs to be complemented with an additional simulation detecting the movement between the spaces.

In what concerns the people count, with the IoT expansion, and the possibility to collect a massive amount of data for localization and tracking of people in commercial buildings, Akkaya, Guvenc, Aygun, Pala, & Kadri (2015), illustrate two major challenges:

a) to achieve occupancy monitoring in the least intrusive manner (by using existing infrastructure, or by not requiring users to install any applications on their mobile devices)
b) to develop data fusion techniques by using a multitude of sources

The present study intends to approach both challenges identified by Akkaya, Guvenc, Aygun, Pala, & Kadri (2015), by making use of non-intrusive sensing devices and by using a multitude of sources for information and data fusion. Furthermore, this study makes use of the patterns identified in previous studies in terms of weekly, monthly and hourly occupation, for performing a comparison with existing literature. This research also uses the same time shifts as identified by Liang, Hong, & Shen (2016), for looking at the presence in the office, but also at the indoor parameters (temperature, humidity, and CO_2).

2.3.3. Virtual sensors for occupancy detection

In terms of sensors used for detecting occupancy, this section aims to provide an overview of the existing possibilities used in previous studies, but also to give an insight into the two sensor types used for the experiment performed in this thesis (PIR and CO_2). Additionally, this section also aims at answering the second research question RQ2: What sensor data can provide the most accurate information in detecting real-time occupancy information, by providing a study on the benefits and drawbacks of three selected occupancy sensor types (PIR, CO_2 and Wi-Fi).

Labeodan, Zeiler, Boxem, & Zhao (2015), focus on the different existing occupancy sensors, by identifying some benefits and drawbacks. Furthermore, as previously described in Chapter 2.3, Labeodan, Zeiler, Boxem, & Zhao (2015), illustrate some properties for real-time occupancy data collection. These properties are the spatial-temporal properties which typically can be obtained from occupancy detection systems found in commercial buildings. These systems can be grouped according to the method, function, and infrastructure. For describing the existing sensors used in detecting real-time occupancy, different sensor types were classified. For the present study, CO_2 and PIR are used in detecting real-time occupancy, therefore the classification made by Labeodan, Zeiler, Boxem, & Zhao (2015) was adapted for these sensor types. The classification also includes the Electromagnetic signals which are considered to be one of the most accurate detection methods by the mentioned study.

Table 2: Classification of occupancy detection systems by (Labeodan, Zeiler, Boxem, & Zhao, 2015)

	Method		**Function**		**Infrastructure**	
Sensor	Terminal	Non-terminal	Individualized	Non-individualized	Implicit	Explicit
CO_2		x		x		x
PIR		x		x		x
Electromagnetic signals (Wi-Fi, Bluetooth, RFID)	x			x	x	

Table 3: Classification of occupancy detection adapted from (Labeodan, Zeiler, Boxem, & Zhao, 2015)

Sensor	**Location**	**Presence**	**Count**	**Activity**	**Identity**	**Track**
CO_2	x	x	x	x		
PIR	x	x				
Electromagnetic signals (Wi-Fi, Bluetooth, RFID)	x	x	x	x	x	x

Additionally, Chen, Masood, & Soh (2016), also emphasize that the estimations of occupancy can be accomplished with different sensors and by collecting diverse sensor data: binary detection-presence /absence (usually the PIR sensors) or actual occupancy numbers (RFID tags, cameras or wearable sensors). The latter ones, however, cannot detect occupants who disengage from the activities that are monitored by the sensor. Moreover, Chen, Masood, & Soh (2016) focus on recent developments, on non-intrusive for users and low-cost sensors, specifically environmental sensors such as CO_2, humidity, temperature and air pressure. These sensors are influenced by the change of occupancy rather than counting the existing people in the room. In a similar research, Zhao, Zeiler, Boxem, & Labeodan (2015), found out that occupancy detection is unreliable when only based on individual physical sensors, due to uncertainties in the occupants activities and that occupants' complaints also play an important role in demand-driven applications, since based on their comfort, they will change back to manual operations.

Due to the availability of sensors provided, and the information concluded from previous research, this study uses PIR and CO_2, as occupancy detection systems. The literature review is further extended focusing on these two types as well as on the Wi-Fi sensors. The Wi-Fi sensors are not used in the current experiment, however, they are further explored in this study, due to their visible benefits in providing occupancy information.

CO_2 sensors for occupancy detection

CO_2 data provides both the estimate of people count and the information on the indoor air quality. However, an important drawback of using CO_2 sensors in detecting occupancy is the influence of external conditions such as the wind speed, space location, opening and closing doors or the pressure difference. CO_2 sensor data is also a measurement of the count and a lot of approximate values are used in previous studies (Labeodan, Zeiler, Boxem, & Zhao, 2015).

For detecting occupancy information, Chen, Masood, & Soh (2016) used CO_2, humidity, temperature and pressure levels measurements, with the sampling time of one minute. For recording the ground truth data, three Internet Protocol (IP) cameras were installed. Based on the results, Chen, Masood, & Soh (2016) concluded that environmental sensors alone can provide a high detection accuracy of 93% when combined with a fusion framework that uses data driven models such as the Extreme Learning Machine (ELM).

For a future study, Zikos, Tsolakis, Meskos, Tryferidis, & Tzovaras (2016), also used CO_2 sensors in a network of sensors consisting of double-beam sensors, pressure mat sensors, composite door counter sensor, an acoustic sensor, PIR motion sensor and a CO_2 sensor. The reported output of the CO_2 sensor was not always connected to the real occupancy level because factors such as open windows or previously trapped CO_2 are influencing the CO_2 value. Therefore, the study introduces an additional feature showing the difference between the current value and the value from ten minutes ago (Zikos, Tsolakis, Meskos, Tryferidis, & Tzovaras, 2016).

For addressing the mentioned drawbacks of CO_2 based occupancy detection, Cali, Matthes, Huchtemann, Streblow, & Muller (2014), investigate how well the CO_2 concentration in the indoor air can be used by an algorithm for detecting real-time occupancy. Their study analyzes which other information; such as window position, door position, mechanical ventilation or CO_2 concentration of the air; is needed for detecting accurate occupancy information. The used algorithm is based on the mass balance equation. This equation can simulate the level of CO_2 if the presence profile is known, or detect the occupancy if the concentration is known. Furthermore, the possibility to use information about the door or window position was also included in the algorithm. For this study, Cali, Matthes, Huchtemann, Streblow, & Muller (2014) formulated three scenarios based on the presence detection and the actual number of occupants. The first scenario includes information about window and door position, the second ignores air exchange with adjacent rooms, while the third ignores air changes through the windows and the door. Their results show that the algorithm generates high levels of presence detection when information about the door and window is included, however for the detection of the exact number of people present in the room, the results are imprecise, the algorithm being inadequate as an occupancy detection tool.

From the previously mentioned results, we can conclude that CO_2 based occupancy detection is accurate if combined and compared with occupancy data coming from other sensors and ground truth data, or when additional information about the external conditions are added to the study.

Furthermore, the present study approaches the CO_2 data as occupancy data, but more importantly as an indoor quality control variable, for improving the indoor living and working conditions.

PIR sensors for occupancy detection

The PIR sensors detect the energy given by objects within its view. As stated by Labeodan, Zeiler, Boxem, & Zhao (2015), PIR sensors are especially used for the control of lighting systems and provide fine-grained occupancy information on user presence and location. Drawbacks of PIR sensors and limitations to the occupancy driven control of lighting systems are, however, the binary output and not being able to provide information on the count. PIR sensors also require a direct line of sight and continuous motion to function properly. They can provide false registration of occupancy due to heat generation (Labeodan, Zeiler, Boxem, & Zhao, 2015).

As practice for using PIR sensors in occupancy detection, Duarte, Wymelenberg, & Rieger (2013), collected data from an 11 story commercial office building. A total of 629 ceiling mounted PIR sensors were placed, the majority being located in private offices, followed by open offices. The sensors used were infrared, ultrasonic and a hybrid between the two. The infrared sensors detect temperature changes, while the latter ones detect movement. According to Duarte, Wymelenberg, & Rieger (2013), it is important to note that the used sensors do not count people, but they only report time-stamped changes of state. However, in their study PIR sensors were used successfully for detecting occupancy patterns and creating profiles according to the time, day and month.

Furthermore, Wahl, Despenic, & Amft (2012), place PIR sensors in pairs, at gateways, for identifying the movement direction of the occupants. Physical gateways include doors, passages or hallways sections interconnecting the office rooms, while virtual gateways include sections in open office spaces or dividers in larger hallways. Even with pair placement, according to Wahl, Despenic, & Amft (2012), PIR sensors can, however, produce errors in movement detection. More specifically, motion events can be deleted or can be wrongly included in the system, due to heat fluctuations at windows and radiators. For compensating these errors, the applied algorithm uses sensor location details, e.g.: node distance compensates for motion errors and deals with concurrent activities. Similarly, the present research intends to remove the inconsistencies recorded by the PIR sensors, by using strategic placement in the office space as well as a combination of PIR and CO_2 sensors.

Electromagnetic signals for occupancy detection

The electromagnetic signals can be present in the form of Wi-Fi, Bluetooth, and radio frequency identification tags (RFID). Even though the signals provide a very high accuracy in occupancy detection, according to Labeodan, Zeiler, Boxem, & Zhao (2015) there are also a few drawbacks for using this type of occupancy detection. The drawbacks consist of privacy concerns due to constant monitoring of people or the multiple devices belonging to one occupant denoting false registration of presence, location, and count. Furthermore, the devices are battery powered so this method is not sustainable for long-term data acquisition.

In a recent research study, two wireless sensor technologies for building occupancy estimation profile were used. Specifically, Masoudifar, Hammad, & Rezaee (2014) used the Ultra-Wideband Real-Time Location System (RTLS) in determining the location of occupants and monitored the ZigBee wireless energy for energy consumption of IT equipment. The RTLSs are a type of ultra-wideband technology (UWB) and have an accuracy of 15 cm, being able to carry out signals passing through doors, or other obstacles. In the study, Masoudifar, Hammad, & Rezaee (2014) aim to obtain occupants detection, identification, and localization, and how much energy is used by every IT device. As a result, the study managed to accurately use the RTLs and energy meters by obtaining occupancy charts illustrating the daily presence, and occupancy profiles for daily and weekly occupancy.

According to Labeodan, Maaijen, & Zeiler (2013), Radio Frequency (RF) systems consist of wireless local area network (WLAN), Ultra-wideband (UWB), an indoor global positioning system (GPS), Radio-frequency identification (RFID), and Global System for Mobile Communications (GSM) systems. The RF systems can provide occupants coordinates with a higher degree of accuracy than any other systems. In choosing the right RF system for occupancy measurements, Labeodan, Maaijen, & Zeiler (2013) presented criteria for selection, based on the cost of deployment, the accuracy of each type, type of environment in which they are deployed (indoor/outdoor) and effects on energy consumption. The criteria for four types of RF systems described by Labeodan, Maaijen, & Zeiler (2013) can be found in Table 4. As seen in the table, the best RF system chosen can be considered the RFID or WLAN system, by having a good or moderate impact on all the parameters. The choice of RF system is made, however, based on the need and problem researched for each study.

Also, in the same paper on occupancy estimation, Labeodan, Maaijen, & Zeiler (2013) mention that WLAN systems have low infrastructure and deployment costs, but considerable high localization uncertainty. This is the case as they do not detect the location of the device connected to the network.

Table 4: Selection criteria adapted from Labeodan, Maaijen, & Zeiler (2013)

	Cost	**Accuracy**	**Environment**	**Energy**
Indoor GPS	××	✓✓	✓	✓×
RFID	✓	✓	✓✓	✓✓
WLAN	✓	✓	✓✓	✓
UWB	××	✓✓	✓✓	✓

✓✓ - **GOOD**, ✓ - **MODERATE**, ✓× - **FAIR**, ×× - **POOR**

Labeodan, de Bakker, Rosemann, & Zeiler (2016) prove the usefulness of using low-cost wireless sensors, in occupancy detection and energy reduction. They mention that the refurbishment process of older buildings for improving energy efficiency presents a number of challenges, however, the advancements in the application of low-cost Wireless Sensors and Actuators Network (WSAN) provide the opportunity to help in the energy reduction of this buildings. In their experiment, Labeodan, de Bakker, Rosemann, & Zeiler (2016) evaluate WSAN for occupancy driven control. Plug-in switching wireless nodes with an average range of 30 m, a maximum range of 100 m in the line of sight of the gateway were installed, and additionally, also motion sensors were placed above the occupants` workspace. As both sensor types have drawbacks, the study ensures that the fusion of sensors facilitates the availability of fine-grained occupancy information.

The results of the experiment introduced by Labeodan, de Bakker, Rosemann, & Zeiler (2016), show that wireless systems are highly prone to interference with other wireless sources within a building. A two-way acknowledgment algorithm was implemented for preventing the errors caused by the wireless system. Because the experiment was conducted for lighting energy reduction, the algorithm ensures there is a communication between the gateway and the luminaire before any control instruction is given. The experiment gave a 28% - 20% reduction in electrical power consumption between the first week without occupancy based lighting and the last two weeks with the sensors incorporated for lighting control.

The data collected for the present graduation thesis does not contain Electromagnetic Signals, due to its unavailability and privacy restriction which is given in the analyzed offices. However, this literature study shows that it is an accurate method for gathering occupancy information. All the authors recommend, however, that the data is collected from more sensor types, rather than only one, for improving the accuracy given by different external factors or by system errors.

Sensor and information fusion for occupancy detection

Information fusion is presented by Zhao, Zeiler, Boxem, & Labeodan (2015) as a technique combining the use of independent measurements and information from multiple sources, for improving accuracy. On the same topic, de Bakker, Aries, Kort, & Rosemann (2016) mention that a person-based technology (such as RFID tags) would, for example, fit the purpose of identifying activities which are not performed behind the desk. A motion detection sensor would also be helpful in sensing the presence of the occupant next to the luminaire, for better lighting control. At the same time, variance in the arrival and departure times of the employees in the office environment can be easily tracked with chair sensors. Therefore, a fusion of sensors can provide the needed information fusion for increasing the accuracy.

The current study proposes a fusion of sensors for occupancy detection (PIR, CO_2), precisely for the purpose of improving the accuracy and for providing fine-grained information. Furthermore, the drawbacks of each type of sensor used in this study will be mitigated in this way.

2.4. Bayesian Belief Network (BBN) for determining occupancy patterns

Occupancy information can be analyzed with various techniques, however, a rather novel approach is the application of the BBN method for determining occupancy patterns and estimating occupancy profiles. BBN was successfully applied in the domain of information fusion, knowledge discovery or probabilistic interference. The applications are in the HVAC systems, for fusing diagnostic information for fault detection (Zhao, Zeiler, Boxem, & Labeodan, 2015).

However, before the BBN method was considered appropriate for this type of data, Dodier, Henze, Tiller, & Guo (2005) firstly introduced, new sensing and data analysis techniques appropriate at the time for detecting occupancy. Two approaches were used at that time for data collection and analysis:

- Development of low-cost distributed and redundant sensor networks (multiple distributed occupancy detectors are better than relying on single points)
- Development of new analysis methods to treat data arising from distributed sensor networks

The method applied to analyze occupancy detection was through the belief network paradigm to the problem of energy management and security. The main goal of the project was to develop models of sensor networks and generate interferences for a control system. Interferences were expressed as the computation of the probability distribution (Dodier, Henze, Tiller, & Guo, 2005).

For occupancy detection, it is possible to develop a generic BBN. The study proposed by Zhao, Zeiler, Boxem, & Labeodan (2015) is a reliable, low-cost solution for demand-driven applications in HVAC. The used parameters can be learned from the historical data. The variables proposed in the study are shown in Table 5. In this table, a variable 'InRoom?' is introduced based on the Expectation-Maximization (EM) algorithm. If there is enough historical data, the parameters can be determined by using this algorithm. The study reveals that Wi-Fi connection information is effective to estimate the occupants' location and that it is unnecessary to record ground truth values as training data, due to the used algorithm. The node 'TimeOfDay' (showing the actual time when the data was collected) was useful for providing occupancy patterns.

Table 5: BBN Input for occupancy detection adapted from Zhao, Zeiler, Boxem, & Labeodan (2015)

Child Node	States	Parent Node	Sensor Input
InRoom?	Occupied	TimeOfDay	PIR2
	Vacant	DayOfWeek	Keyboard&Mouse
		Holiday	
TimeOfDay	Prevailing log time		
DayOfWeek	Prevailing weekdays		
Occupancy detection sensor (PIR)	Detected	InRoom?	
	Undetected		
Occupancy reaction information (Keyboard&Mouse)	Touched	InRoom?	
	Untouched		
	PowerOff		

Furthermore, Zikos, Tsolakis, Meskos, Tryferidis, & Tzovaras (2016) proposed a novel approach to estimating real-time occupancy in buildings from various sensor data, based on Conditional Random Field (CRF) probabilistic. The paper illustrates three types of occupancy information: presence/absence, occupancy density and the actual number of occupants. CRFs are represented as undirected graphs, with random variables and dependencies between them. The time used for this method is divided into constant length, transitions to a new state are performed at every time step. The used sensor network is a distributed sensor network forming a multi-sensorial cloud. Different sensor combination was proposed for four spaces. In what concerns occupancy density estimation, the proposed method achieves from 80% to 93% accuracy and for estimating the exact number of occupants, the accuracy is up to 78% in a Meeting Room when there are up to 13 occupants estimated.

This literature review shows that probabilistic models and the Bayesian Belief Network can be used successfully to analyze sensor data for occupancy detection. Furthermore, the present graduation project uses the study of Zeiler, Boxem, & Labeodan (2015) as a reference for the variables in the BBN network; some of the existing variables are kept: 'TimeOfDay', 'DayOfWeek', Occupancy detection sensors (PIR and CO_2). In the section 3.2.2, the method used in this study will be further explored and explained.

2.5. Conclusion

Poor indoor air quality and thermal comfort reduces performance and the motivation to perform in the working environment. IAQ, thermal comfort, the temperature and the RH in the office are interrelated and are influenced by the HVAC. Therefore, a balanced ventilation and indoor temperature can enhance the performance and can drastically improve the working environment. Furthermore, CO_2 relates to human metabolism and it usually increases in unventilated spaces. Hence, careful consideration should be given to these three parameters: CO_2, RH and temperature. This literature review shows that air sensors can successfully register these parameters and that a patterns recognition system can be implemented for future predictions of IAQ.

Real-time occupancy improves the control of indoor appliances and of the HVAC system based on human activity. Occupancy profiles and patterns generated in previous studies differ in parameters such as the weekday, and are similar for the time of the day. For generating occupancy profiles, a series of virtual sensors are used, and this literature review focuses on identifying the benefits and drawbacks of three main sensors: PIR, CO_2 and Electromagnetic Signals. While some Electromagnetic Signals such as the RFID or WLAN are very accurate in detecting occupancy, they come with privacy concerns and with the concern that the user needs to have a transmitting device which is usually battery powered. Furthermore, false registrations can be detected due to one user having multiple devices, or due to interferences. The PIR sensors produce big errors in occupancy detection, and different authors recommend using this method together with another sensor, or by using PIR pair placement. CO_2 sensors can detect presence in proportion of around 93%, and usually ground truth data is needed for checking the accuracy.

This literature review concludes, on one hand, that real-time occupancy detection can be very accurate if Electromagnetic Signals are used. On the other hand, this review also shows that accurate experiments can be performed by using a fusion of sensors, and that a non-intrusive method for detection can be obtained by avoiding the Electromagnetic Signals.

Indoor air quality and real-time occupancy are the main parts investigated by this literature review. A bridge between them is given by the CO_2, which is both an occupancy and an IAQ sensor and variable. Furthermore, as shown by this literature review, the HVAC operation influences the IAQ, while the occupancy plays an important role in the activation of the HVAC system. The impact of occupancy on the IAQ parameters is further explored in this thesis and constitutes the answer to RQ3.

3

RESEARCH APPROACH

3.1. Introduction

For answering the research questions, quantitative research is conducted, by collecting data in numerical form, which will be analyzed with the help of statistics. The objective of this quantitative research is to develop hypothesis coming from patterns (occupancy, CO_2, humidity and temperature) that can be used as input for improving the indoor air quality. Hence, this quantitative research, consists of the data collection, processing and analysis.

In this section the research approach used is thoroughly explained. Firstly, the method is presented together with the motivation for using the Bayesian Belief Network (BBN) approach for analyzing the sensor data. Secondly, the conducted experiment is explained, together with the data collection. Finally, the data analysis and preparation is presented together with the final conclusions on the research approach that was used.

3.2. Method

The method used for analyzing sensor data is primarily the BBN. For creating the BBN, the raw data is cleaned and processed in IBM SPSS Statistics. Descriptive analyses is initially performed in SPSS for complementing the BBN.

3.2.1. Descriptive analyses

For the purpose of the research and for the type of data collected, the *frequencies procedure*, *cross-tabulation*, *bivariate correlation* and *Linear Regression* are used.

The most common summary measure is the number of percentage of cases in each category known as the frequencies procedure. Through this method, tables that display both the number and percentage of cases for a variable are made. The variable can be graphically displayed in a frequency table with a bar chart or a pie chart. At the same time, the cross tabulation procedure is used for examining the relationship between two categorical variables, by creating a two (or more) dimensional table (Statistics, IBM SPSS, 2015).

Furthermore, the Pearson, Spearman, Kendall`s tau-b correlations are computed to measure how variables are related. A Pearson correlation coefficient is measuring a linear association, while the Spearman and Kendall`s tau-b are used for variables with ordered categories (Statistics, IBM SPSS, 2015). Lastly, the Linear Regression estimates how one dependent variable (Y) can be predicted based on an independent variable (X), based on the simple regression equation: $y = a + b * x$ (Division of information technolology services, CAL State LA, 2016).

IBM SPSS Statistics

For cleaning and processing the data as well as for performing the descriptive analyses IBM SPSS Statistics is used. SPSS Statistics is a leading software package used for statistical analysis by a variety of stakeholders such as: market researchers, health researchers, survey companies, government, data miners and others. The tool provides a range of analysis techniques for hypothesis testing and reporting, managing and accessing data or selecting and performing data analysis (IBM, 2017). Furthermore, SPSS offers the possibility to create customized tables, enabling an easy method for interpreting and summarizing the result.

SPSS data sets have two dimensional table structures, where the rows represent the cases, and the columns show the variables with the respective measurements. The data types defined in the present collected data are numeric, string and date, while the level of measurement can be scale, nominal, or ordinal. The scale measurement is given by the data measured on interval or scale, and it is also referred to as continuous or quantitative data. The nominal and ordinal values are the categorical values, which is the data with a limited number of distinct values or categories, and is referred to as qualitative data. In the nominal measurement there is no permanent order in the categories, while in the ordinal measurement there is a meaningful order of categories (Statistics, IBM SPSS, 2015). SPSS allows the continuous data to be re-coded into categorical data, according to the research questions, and the desired categories. Most algorithms for BBNs are designed for discrete variables, therefore the continuous variables have to be discretized prior to the creation of the network.

3.2.2. Bayesian Belief Networks

Belief Networks are graphical representations of models, that illustrate the relationships between the model`s variables. The networks identify the variables that interact directly with each other and simplify the belief updating by limiting each variable to its local neighborhood. According to Krieg (2001), a Bayesian Belief Network (BBN) is a specific type of causal belief network and is a graphical representation of a probabilistic dependency model. A dependency model is a probabilistic model where the dependencies between various events are captured. A BBN consists of nodes representing stochastic variables and connecting arcs showing the causal relationship between the variables (Krieg, 2001).

For estimating a BBN, we want to obtain estimates for events that are unobservable or observable at an unacceptable cost, for hypothesizing the occurrence of these events, reflecting the goals of the model. According to Krieg, (2001) it is necessary to determine the causal structure between variables. Hence, the appropriate conditional probabilities and dependencies need to be established between the information and hypothesis variables. This aspect is done by placing intermediate nodes between the information (evidence) and the hypothesis nodes.

According to (Kemperman & Timmermans, 2014), a BBN is an acyclic graph and can be written as:

$$BBN = (V, E),$$

In this equation V is a set of variables (X, Y) and E is a set of links(X, Y). If there is a link $X \rightarrow Y$, X is a parent of Y and Y is a child of X, and for each variable a conditional probability table (CPT) is given, which shows how much a variable depends on the parents. In accordance with Krieg (2001), the conditional probabilities in a Bayesian network can be Bayesian or physical. The Bayesian probabilities are derived from prior knowledge, whereas the physical ones are learnt from the data. Determining the degree of belief is known as a probability assessment.

The BBN has parents and child nodes and for reducing the number of combinations for a variable with multiple parents the divorcing technique can be applied. "Some of the parents of a variable are removed or divorced from that variable by introducing a mediating variable and making it a child of the divorced parents and a parent of the original child" (Krieg, 2001).

A simple BBN is illustrated by Verhoeven, Arentze, Waerden, & Timmermans (2017) which is adapted in Figure 3. In this model the 'appearance' is the child node of 'color' and 'coating', and a CTP will be generated for finding out how much the 'appearance' depends on the parents.

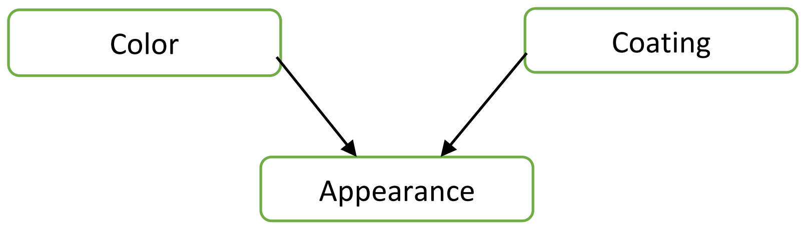


Figure 3: A simple BBN model retrieved from Verhoeven et.al. (2017)

In the present research, the BBN is constructed by using the program GeNIe Modeler, https://www.bayesfusion.com/genie-modeler, by BayesFusion, LLC. The CPT are estimated when creating the BBN based on the same data set, by using an expectation maximization algorithm (EM). The data files are generated in SPSS and saved in .csv format for importing in the GeNIe Modeler (GeNIe 2.1 Academic).

GeNIe Modeler

GeNIe Modeler was created for building graphical decision-theoretic problems, being tested extensively in many research and commercial environments (BayesFusion, LLC, 2017).

The software can 'learn' a new network by generating the graphical data structure from a data set. According to (Zong & Wang, 2015) the learning process can be described as follows: the variable S is the structure of the BBN, and P(S) is its prior probability distribution. The data-set D is represented by nodes in the BBN. Based on Bayes theorem, the posterior probability of S can be calculated:

$$P(S|D) = \frac{P(S,D)}{P(D)} = \frac{P(S)P(D|S)}{P(D)};$$

where $P(S|D)$ is the posterior probability, $P(S)$ is the prior probability of S, and $P(D)$ is the probability distribution of the data-set D.

The new network is generated based on a learning algorithm. The learning algorithm applied for this thesis is the Greedy Thick Thinning (GTT), which is based on the Bayesian Search approach. "GTT starts with an empty graph and repeatedly adds the arc (without creating a cycle) that maximally increases the marginal likelihood P(D|S) until no arc addition will result in a positive increase (this is the thickening phase). Then, it repeatedly removes arcs until no arc deletion will result in a positive increase in P(D|S) (this is the thinning phase)" (BayesFusion, LLC, 2017).

The priors used for creating the BBNs are the K2 choices as introduced by Cooper & Herskovits (1992). The Max Parent Count is set by default to 8, which limits the number of parents a node can have. The knowledge editor in GeNIe allows entering knowledge that will help in learning the structure of the network. This editor can assign variables to temporal tiers, force arcs (assurance of having those arcs in the structure) and forbid arcs (guarantee of excluding those arcs from the structure). The rule for the temporal tiers is that in the resulting network there will be no arcs from nodes in higher tiers to nodes in the lower tiers (BayesFusion, LLC, 2017).

After estimating the model, evidence can be entered, as one of the basic operations for a probabilistic model. This evidence makes the model adjust to a new situation, by showing new posterior probabilities. According to (Fenton, 1995), hard evidence for a node X is evidence that the state of X is definitely a certain value. For example, if X is the result of a football match for example, (win, lose or draw), hard evidence would show a definite victory.

3.2.3. Motivation of analysis choice

For the development of this research project, descriptive analyses and BBN (GeNIe Modeler) were used as research tools. These methods were selected because they were the most appropriate for the collected data and the research objectives.

In terms of data processing and cleaning, SPSS was used, because it allows to operate over large data files and to re-code the continuous variables into discrete variables. In this way, a customized data set was created, which facilitates the analysis of information and corresponds to the research goals. Furthermore, through descriptive analyses, specifically by providing tabular and graphical indications of frequencies, and through linear regression and bivariate correlation, the behavior and dependency of some variables could be analyzed, which contributed to the end results.

The BBN is a suitable method of analysis for the purpose of this research, because it provides the ability to make connections between variables in the model, and to estimate the occurrence of certain events based on posterior probabilities. Therefore, hypothesis and predictions of future patterns can be estimated. Moreover, the BBN method was chosen based on previous research developed on real-time occupancy patterns. Zhao, Zeiler, Boxem, & Labeodan (2015) proposed a sensor information fusion method for integrating multiple occupancy measurements, by using a BBN network. The network they illustrate is effective in estimating information of individual occupants, and the parameters could be used for assessing the performance of sensors. Therefore, the present thesis considers the parameters and network created by Zhao, Zeiler, Boxem, & Labeodan (2015) and adapts the BBN for a new occupancy experiment. Furthermore, the present thesis focuses primarily on the indoor air quality improvement based on sensor data and occupancy.

3.3. Case Selection

The case selected for achieving the research objective was done in consultation with the company Inteliments (https://inteliments.com/en), for which the graduation thesis was developed. This was done based on the offices where the company is monitoring the indoor air quality.

The research area is the Strijp-S area in Eindhoven, Netherlands. The former industrial Philips complex dates back from 1928. The revitalization process of the former Philips Company`s industrial complex started in 2000 when the company left the city of Eindhoven. Since 2006 it became a serious plan, with the aim of transforming the area into a new urban domain to live, work and play. The workspaces located in this area focus on entrepreneurs and creative industries. Apparatenfabriek, located on Torenallee 22, was redeveloped in 2011 into a business complex with over 2000 m^2 of commercial space and office space for creative industries. Similarly, the VideoLab (Office-S) on Torenallee 20 was turned into office space for entrepreneurs and creative industries (Strijp-S, 2017).

The first office space analyzed is located at the 6th floor of Office-S on Torenallee 20 and it is a private office (Credo Room) accommodating a maximum of six people. The second office space is Seats2Meet (S2M) located at the first floor of the building on Torenallee 24, and consists of two rooms: an open-space plan where 30 to 40 people work daily, and a meeting/conference room that can be used for approximately 150 people. The space used for analysis was the conference room.

The two office spaces are located in redeveloped buildings, constructed in 1928. An older building will affect the indoor air quality and thermal comfort differently than a modern building. Specifically, the heating system in Credo Room cannot be adjusted/switched on-off separately, only at the same time with other adjacent rooms. Furthermore, the HVAC system is an old system operating based on the adjustments made by office occupants; no automatic adjustments are currently implemented. Therefore, an adequate thermal comfort cannot be obtained for one particular room, depending on the people inside it. An automated climate control system based on occupancy can be created with difficulty, without affecting neighboring rooms. The corrective actions that will be suggested for improving the indoor air quality have to be suitable for this type of old building.

3.4. Data Collection

For achieving the aim of this research, two data types were collected. Firstly, historical data was gathered for answering the first research question: **How can future states of CO_2, humidity and temperature be predicted based on historical sensor data, for improving the indoor air quality.** Secondly, an experiment was conducted for answering the sub-question: **How accurate is the information fusion given by CO_2 and PIR sensors in detecting real-time occupancy.**

The historical data was collected in two office spaces in the Strijp-S area in Eindhoven, Netherlands.

In Credo Room, one IQRF sensor was installed for collecting CO_2, temperature and humidity data over a period of five months, starting with 1st of September 2016. IQRF is a platform for wireless connectivity ranging tens or hundreds of meters, used in the automation of buildings and cities. (Microrisc, 2017). The sensor was installed strategically, next to the door, at 1.20m above the ground. The position of the sensor is important for the accuracy of the data. The sensor recorded data every five minutes, on a 24 hours basis.

For the second office space (S2M), historical data on CO_2, temperature and humidity over a period of five months was collected from the conference room. One IQRF sensor was placed in the center of the room. As the IQRF sensor can have a range of tens to hundreds of meters, one sensor was sufficient for collecting the needed data. The sensor in this room recorded data also every five minutes, for 24 hours a day.

The experiment was conducted in the Credo Room, where, with the same IQRF sensor, CO_2, humidity, temperature and PIR data was collected. Ground truth data was collected for the period of 06.03.2017 until 07.04.2017, period in which all the sensor data should have been collected. Due to a data outage registered in this period, the sensor collected CO_2, temperature and humidity data only for the period 22.03 – 07.04, and no PIR data was retained. The PIR data was, however, collected at a later time for a period of five days between 16.04.2017 until 20.04.2017. Therefore, the experiment was conducted along two time periods. Firstly, the ground-truth data was compared to the CO_2 data for identifying CO_2 levels belonging to the actual presence in the room. Secondly, the complete data collected from the sensors over a period of five days was used for processing and constructing the BBN.

An overview of the previously mentioned information, in regards to the data collected, rooms and sensors is presented in Table 6. The collected data was collected in the InteliGlue© Platform. The mentioned platform acquires, stores and analyses data from sensors and field gateways. It is a highly scalable platform, built on connectivity and open communication standards, with the possibility to have it configured for any type of solution.

Table 6: Data collection overview

Room	Office	People	Sensor	Data
Conference Room	Seats2Meet	180 – fixed amount	1 IQRF Sensor: CO_2, Humidity, Temperature	Historical
Private Office - Floor 6 Credo Room 6003	Office S	6	1 IQRF Sensor: CO_2, Humidity, Temperature, PIR	Historical Experiment Ground Truth Data

3.4.1. Ground truth data

Ground truth data is collected for the experiment for the Credo Room, for the period 06.03.2017 – 07.04.2017. The number of people in the room are collected along six time periods:

- Going-to-work period: 07:00-09:00
- Morning period: 09:00-12:00
- Lunch period: 12:00-14:00

- Afternoon period: 14:00-17:00
- Going-home period: 17:00-19:00
- Night period: 19:00-07:00

The time frames were selected based on the research conducted by Liang, Hong, & Shen (2016), where a change in the number of occupants was observed along those periods. Liang, Hong, & Shen (2016) concluded that the first time frame is represented by the increasing number of people, followed by the morning where the occupancy is constant. The lunch period indicates a decrease in the number of people present in the office, and the afternoon shows an increase compared to the lunch. In the afternoon period, however, less people are present compared to the morning, as some people do not return after the lunch break. The going-home period is the period of constant decreasing, until reaching zero occupancy, followed by the night which also shows no occupancy in the office.

The ground-truth data was collected by counting the number of people and registering it in a form. The completed forms are found in Appendix 3: Questionnaires for ground-truth data.

3.5. Data Processing

3.5.1. Missing cases

After the data is gathered, the preparation phase starts and is performed in SPSS. For the historical data, the files with monthly data are merged into two big joint data files (one for each room). For the experiment data the daily files containing sensor data with all the four variables included (CO_2, humidity, temperature and PIR), is also merged into one joint file.

The first step in preparing the final files for creating the variables was to investigate the missing cases and data patterns. In this respect not all entries showed all the three or four parameters (CO_2, temperature, humidity and PIR) for a specific data and time. Therefore, all the cases missing one of the three, or four (in the experimental data), were excluded from the data file.

3.5.2. Variables

After removing the missing values, the historical data files consist of 30138 cases for the Credo Room and 21778 cases for the S2M conference room. The file gathering data for the experiment consists of 1134 cases. The ground truth data together with the CO_2, humidity and temperature for the period 22.03 – 07.04, is joined into a file consisting of 4179 cases. Subsequently, the variables re-coded in SPSS for both the historical data and the data for the experiment are explained below.

Historical Data Variables

The choice of variables was done for fulfilling the research objective, primarily for assessing the change in CO_2, temperature and humidity based on recorded sensor data. The initial values of CO_2, humidity and temperature were re-coded based on the levels indicated by the company, in terms of indoor air quality. A choice of eight re-coded variables was made for using in the future BBN: 'Month', 'Weekday', 'Timeshift', 'NewCO2', 'NewTemperature', 'NewHumidity', 'NewOutsideHumidity', 'NewOutsideTemperature'.

Credo Room

The variable 'month' is the indication of the month in which the indicated case was recorded. There are five values for this variable: 9 = 'September 2016', 10 = 'October 2016', 11 = 'November 2016', 12 = 'December 2016' and 1= 'January 2017'. This variable was extracted from the original variable 'date', which represents each specific date (mm/dd/yyy format) for the cases.

The 'Weekday' represents the different days of the week in which the case was registered. This variable was similarly extracted as the 'Month', from the 'Date'. For the 'Weekday' seven values were created, from 1 = 'Sunday' to 7 = 'Saturday'.

The variable 'Timeshift' was made based on the original variable 'time' where the actual time in hh:mm:ss format was registered. The 'Timeshift' variable contains five values: 1 = 'Going-to-work', 2 = 'Morning', 3 = 'Lunch', 4 = 'Afternoon', 5 = 'Going-home' and 6 = 'Night'. These five periods were constructed for matching the periods presented in the ground truth data, similarly, based on the research performed by Liang, Hong, & Shen (2016).

The indoor air quality parameters were re-coded in SPSS into 'NewCO2', 'NewTemperature' and 'NewHumidity' from the values identified in the initial variables 'CO2', 'Temperature' and 'Humidity'. The variables were re-coded from continuous into discrete data, based on the expertise from the company Inteliments. The CO_2 was re-coded in three levels: 1 = 'Healthy Level', 2 = 'Acceptable Level' and 3 = 'Stiffness/odors'. The healthy level is situated between 350 to 450 parts per million (ppm), and can be compared to a normal outdoor healthy level. The acceptable level for working is between 450 and 700 ppm, while for more than 700 ppm potential odors are present in the room. There is no recorded data outside the mentioned levels.

The variable 'NewTemperature' also consists of three levels: 1 = 'Level 1', 2 = 'Level 2', 3 = 'Uncomfortable High'. Between 19 and 22.9°C is the first comfortable working level, between 23 and 25.9° C the second level, while the highest level is the uncomfortable level, with temperature between 26 and 30°C. No additional data was registered outside these levels, meaning lower than 19 or higher than 30°C.

The variable 'NewHumidity' consists of only two values: 1 = 'Minimal uncomfortable humidity', 2 = 'Optimal Humidity'. The minimal uncomfortable humidity, suggests that the working space is too dry and it is represented by the humidity below 40%. The second level is the optimal humidity, which is between 40 and 60%. The last level, higher than 60% and representing the maximal uncomfortable humidity, was not shown in the collected data. An overview of the re-coded variables for the Credo Room, can be found in Table 7.

Table 7: Credo Room main variables

NewCO2	NewHumidity	NewTemperature	Timeshift	Weekday	Month
1 : 350…450 ppm – 'Healthy Level'	1 : >40% – 'Minimal uncomfortable humidity'	1 : 19 – 22.9°C – 'Level 1'	1 : 07:01:00 – 09:00:00– 'Going-to-work'	1 - 'Sunday	9 - 'September _2016'
2 : 451…700 ppm – 'Acceptable Level'	2 : 40% – 60% - Optimal Humidity	2 : 23 – 25.9 °C –'Level 2'	2 : 09:01:00 – 12:00:00 – 'Morning'	2 - 'Monday'	10 - 'October_2 016'
3 : 701…1000 ppm – 'Stiffness/odors'		3 : 26 – 29.9 °C – 'Uncomfortable High'	3 : 12:01:00 – 14:00:00 PM – 'Lunch'	3 - 'Tuesday'	11 - 'November _2016'
			4 : 14:01:00 – 17:00:00 – 'Afternoon'	4 - Wednesday'	12 - 'December _2016'
			5 : 17:01:00 – 19:00:00 – 'Going-home'	5 - 'Thursday'	1 - 'January_2 017'
			6 : 19:01:00 – 07:00:00 – 'Night'	6 - 'Friday'	
				7 - 'Saturday'	

In what concerns the outdoor variables they were re-coded into 'NewOutsideTemperature' and 'NewOutsideHumidity'. The original variables are the 'OutsideAverageTemperature' and 'OutsideAverageRelativeHumidity', given in the data set for each hour.

For re-coding the variables into different categories the decision tree procedure is used. The decision tree is a tree-based classification model created by putting cases into groups and predicting values of dependent (target) variables based on independent (predictor) variables (IBM SPSS Statistics, 2015). The outdoor variables (dependent variables) were re-coded according to this decision tree made in SPSS based on the indoor variables (independent variables). The splits present in Table 8 gave a significant difference in temperature and humidity percentages for the indoor variables. For predicting the growth of the decision tree, the Exhaustive CHAID (Chi-squared Automatic Interaction Detection) procedure is applied in SPSS. The Exhaustive CHAID procedure investigates all possible splits for each independent variable.

The classification trees for the Credo Room outside variables can be found in Appendix 4: SPSS Classification tree for weather conditions.

Table 8: Credo Room outside variables

NewOutsideTemperature (°C)	NewOutsideHumidity (%)
≤ 3.7 : Category 1	≤78 : Category 1
3.8 – 6.3 : Category 2	79 – 83 : Category 2
6.4 – 8.9 : Category 3	84 – 88 : Category 3

9 –13.3: Category 4	89 – 93 : Category 4
>13.4 : Category 5	>93 : Category 5

S2M Conference Room

The variables 'Weekday' and 'Timeshift' present the same categories as for the Credo Room. The 'Month' variable has slightly different values: 10 = 'October 2016', 11 = 'November 2016', 12 = 'December 2016', 1 = 'January 2017', 2='February 2017'. These categories are created due to the 'date' when the data was collected.

The variable 'NewCO2' presents two additional levels compared to the Credo Room: 1 = 'Minimum Level', 2 = 'Healthy Level', 3 = 'Acceptable Level', 4 = 'Stiffness/Odors', 5 = 'Drowsiness/Bad Air'. The additional categories are made due to the data registered below 350 ppm and higher than 1000 ppm. Therefore the 'Minimum Level' is found for values below 350 ppm, while the 'Drowsiness/Bar Air' is represented by values higher than 1000 ppm.

The 'NewTemperature' also shows one additional category: 1 = 'Uncomfortable low', 2 = 'Level 1', 3 = 'Level 2', 4 ='Uncomfortable High'. The ''Uncomfortable low' category consists of cases of 'Temperature' below 18.9°C.

Lastly, the variable 'NewHumidity' consists of the same cases as described in the previous room. No data above 60% was registered within the variable 'Humidity'.

The variables discussed for the S2M Conference Room are illustrated in Table 9.

Table 9: S2M Room main variables

NewCO2	NewHumidity	NewTemperature	Timeshift	Weekday	Month
1 : <350 ppm – 'Minimum Level'	1 : <40% – 'Minimal uncomfortable humidity'	1 : <19 – 'Uncomfortable low'	1 : 07:01:00 – 09:00:00– 'Going-to-work'	1 : 'Sunday	10 : 'October_ 2016'
2 : 350…450 ppm – 'Healthy Level'		2 : 19 – 22.9°C – 'Level 1'	2 : 09:01:00 – 12:00:00 – 'Morning'	2 : 'Monday'	11 : 'Novembe r_2016'
3 : 451…700 ppm – 'Acceptable Level'	2 : 40% – 60% - Optimal Humidity	3 : 23 – 25.9 °C –'Level 2'	3 : 12:01:00 – 14:00:00 PM – 'Lunch'	3 : 'Tuesday'	12 : 'Decembe r_2016'
4 : 701…1000 ppm – 'Stiffness/odors'		4 : 26 – 29.9 °C – 'Uncomfortable High'	4 : 14:01:00 – 17:00:00 – 'Afternoon'	4 : Wednesday'	1: 'January_ 2017'
			5 : 17:01:00 – 19:00:00 – 'Going-home'	5 : 'Thursday'	2: 'February _2017'
5 : >1000 ppm – 'Drowsiness/Bad Air'			6 : 19:01:00 – 07:00:00 – 'Night'	6 : 'Friday'	
				7 : 'Saturday'	

The outside variables were similarly re-coded into 'NewOutsideTemperature' and 'NewOutsideHumidity'. The categories (splits) were created based on the classification tree in SPSS, which can be seen in Appendix 4: SPSS Classification tree for weather conditions. The categories are illustrated in Table 10.

Table 10: S2M outside variables

NewOutsideTemperature (°C)	NewOutsideHumidity (%)
≤0.5 : Category 1	≤90 : Category 1
0.6 – 4.1 : Category 2	91 – 93 : Category 2
4.2 – 6.6 : Category 3	>93 : Category 3
6.7 –8.5: Category 4	
>8.5 : Category 5	

Experiment Data Variables

For the experiment two data files were created. The first one contains the PIR variable, and the second one contains the ground truth data. The first file is used for creating the BBN, and the second is used for comparing the CO_2 with the ground truth data. Both data files represent the data collected in the Credo Room. The variables for each of the two files are explained below.

BBN Variables

The variables 'Weekday' and 'Timeshift' are re-coded similarly as in the historical data. The 'Weekday' has five categories: 1 = 'Sunday', 2 = 'Monday', 3 = 'Tuesday', 4 = 'Wednesday', 5 = 'Thursday'. The sensor recorded no data during Friday and Saturday. The 'Timeshift' is re-coded into the same six categories: 1 = 'Going-to-work, 2 = 'Morning', 3 ='Lunch', 4 = 'Afternoon' 5 = 'Going-home', 6 = 'Night'.

The initial values present in the variable 'CO2' registered in ppm are re-coded into the variable 'NewCO2' similarly to the historical data. Because the data for the experiment is collected on a much shorter time frame than the historical data and due to the high amount of cases registered between 350 and 700 ppm, more categories are created. The variable 'NewCO2' presents the following levels: 1 = 'Healthy 1', 2 = Healthy 2', 3 = 'Acceptable 1', 4 = 'Acceptable 2', 5 = 'Stiffness/Odors. The first healthy level is between 350 – 400 ppm, the second between 401 – 450 ppm, the acceptable level 1 is between 451 – 550 ppm, while the second acceptable level is between 551 – 700 ppm. The 'stiffness/odors' are registered for values higher than 700 ppm.

For the variable 'NewTemperature' more levels than in the historical data are constructed: 1 = '19 to 19.9', 2 ='20 to 20.9', 3 = '21 to 21.9', 4 = '22 to 22.9', 5 = '23 to 23.9' and 6 ='24 to 25'. This is the case as the temperature is present only between 19 and 25°C, therefore the comfortable temperature had to be re-coded more than in the historical data. There is no presence of uncomfortable temperature in the registered data.

The variable 'Humidity' is re-coded into 'NewHumidity' in the following categories: 1 = 'Min uncomfortable 1', 2 = 'Min uncomfortable 2', 3 = 'Min uncomfortable 3'. The first category includes values of humidity between 14 and 20%, the second contains values between 21 – 25% and the third one contains values between 26 and 31%. The recorded humidity is registered between 14% until 31%, therefore only uncomfortable humidity is present, and no indication of optimal humidity was present in this data set.

The PIR sensor records in the 'MotionCount' variable the number of movements registered every five minutes along the day. The motion variates from 0 to up to 17 movements. The PIR sensor cannot actually detect the number of people present in the room, only how many times one or more people passed in front of the sensor. Of course, a higher number of movements along the 5 minutes period will indicate more than 1 person in the room, however the actual number is not provided by the PIR sensor. Based on the research performed by (Zhao, Zeiler, Boxem, & Labeodan (2015), the 'MotionCount' variable is re-coded into 'NewMotionCount' with two possible states: 0 = 'Undetected' and 1 = 'Detected'. Therefore, the detected state will indicate presence in the room, while the undetected indicates no presence. Only the 'NewMotionCount' variable is used in learning the BBN.

Apart from the CO_2 levels illustrated by the variable 'NewCO2', the 'CO2' variable is also re-coded into 'CO2Motion'. The 'CO2Motion' also has two states: 0 = 'Undetected' and 1 = 'Detected'. The undetected states include all the CO_2 values under 500 ppm, while the detected state includes all the values higher than 500 ppm. This implies that once 500 ppm was reached, there is presence detected in the room. This number was selected as a threshold between the undetected and detected state, based on the comparison made between the CO_2 and the ground truth data on the actual number of people present in the room. This comparison is illustrated in section 4.3.1. Ground truth data analysis.

An overview of the variables used for creating the BBN for the experiment are presented in Table 11.

Table 11: Experiment data variables

NewCO2	NewHumidity	NewTemperature	Timeshift	Weekday	NewMotion Count	CO2Motion
1 : 350...400 ppm – 'Healthy Level 1'	1 : 14% – 20 % 'Minimal uncomfortable humidity 1'	1 : 19 to 19.9° C – 'T19'	1 : 07:01:00 – 09:00:00– 'Going-to-work'	1 : 'Sunday'	0 = 'Undetected	0 = 'Undetected
2 : 401...450 ppm – 'Healthy Level 2'		2 : 20 to 20.9° C – 'T20'	2 : 09:01:00 – 12:00:00 – 'Morning'	2 : 'Monday'	1 = 'Detected'	1 = 'Detected'
3 : 451...550 ppm – 'Acceptable Level 1'	2 : 21% – 25% - 'Minimal uncomfortable humidity 2'	3 : 21 to 21.9° C – 'T21'	3 : 12:01:00 – 14:00:00 PM – 'Lunch'	3 : 'Tuesday'		
4 : 551...700 ppm – 'Acceptable Level 2'	3 : 25% – 31% - 'Minimal uncomfortable humidity 3'	4 : 22 to 22.9° C – 'T22'	4 : 14:01:00 – 17:00:00 – 'Afternoon'	4 : Wednesd ay'		
		5 : 23 to 23.9° C – 'T23'	5 : 17:01:00 – 19:00:00 – 'Going-home'	5 : 'Thursday'		
5 : 701...1000 ppm – 'Stiffness/odors '		6 : 24 to 25 – 'T24'	6 : 19:01:00 – 07:00:00 – 'Night'			

Ground truth data variables

For the data set collected along with the ground truth data (the actual people count in the room), the same variables were re-coded: 'NewCO2', 'NewTemperature', 'NewHumidity'. The same categories as for the historical data were created. The variable 'NewCO2' was split in: 1 = 'Healthy Level', 2 = 'Acceptable Level', 3 = 'Stiffness/Odors'. The temperature was re-coded into: 1 = 'Comfortable 1', 2 = 'Comfortable 2', 3 = 'Uncomfortable High'. No temperature outside these ranges was reflected in the data set. Lastly, the 'NewHumidity' has the following categories: 1 = 'Min uncomfortable humidity', 2 = 'Optimal Humidity'.

The ground truth data is represented in the data file under the variable PeopleCount and can take 5 states: 0 = 'No people', 1…4 = '1 to 4 people'. The ground truth data is used for associating the amount of people present in the room to the initial CO_2, temperature and humidity, as well as to the new re-coded variables. The results of this association are found in chapter 4.3.1. Ground truth data analysis.

An overview of the variables used in the ground truth data set is found in Table 12. The original ground truth data files (as collected in the validation phase) are found in Appendix 3: Questionnaires for ground-truth data.

Table 12: Ground truth data variables

NewCO2	NewHumidity	NewTemperature	PeopleCount
1 : 350…450 ppm – 'Healthy Level'	1 : <40% – 'Minimal uncomfortable humidity'	1 : 19 – 22.9°C – 'Level 1'	0 = 'No people'
2 : 451…700 ppm – 'Acceptable Level'		2 : 23 – 25.9 °C –'Level 2'	1 = '1 person'
3 : 701…1000 ppm – 'Stiffness/odors'	2 : 40% – 60% - Optimal Humidity	3 : 26 – 29.9 °C – 'Uncomfortable High'	2 = '2 people'
			3 = '3 people'
			4 = '4 people'

3.6. Data Analysis (Bayesian Model Set-up)

3.6.1. Construction of the Bayesian networks

For creating the structure of the networks the program GeNIe 2.1 Academic, https://www.bayesfusion.com/genie-modeler, is used. The program uses the data files processed previously in SPSS, where the variables were recoded from continuous variables into discrete data. For generating the network, the GTT learning algorithm was applied. Temporal tiers influence the arcs in such a way that no arcs will result in the network from a higher tier to a lower tier (e.g.: from tier 2 to tier 1 there will be no arcs). The variables were assigned to tiers in the following manner for the historical data:

Tier 1: 'Timeshift', 'Weekday', 'Month', '*NewOutsideTemperature', 'NewOutsideHumidity'*

Tier 2: NewCO2', 'NewTemperature', 'NewHumidity'

This distinction was made as the inside humidity, temperature and CO_2 normally should not have any impact on the time of the day, month, day of the week or outside conditions. Rather, the network is learned in causation terms, where variables in the second temporal tier occur later in time than the ones in the first tier. The outside conditions are additional factors, which were added for seeing their importance on the indoor conditions. Two different networks are estimated, one without the outdoor variables and another one with the outdoor variables, for testing the impact of the outdoor variables on the indoor variables as well as the strength of their influence.

At the same time, the program gives the possibility to forbid arcs between variables, apart from the categorization in different tiers. For this network, arcs were forbidden between Time, Weekday and Month, in the following combinations: 'Weekday' -> 'Month', 'Timeshift' -> 'Month', 'Timeshift' -> 'Weekday', 'Month' -> 'Weekday'. This arcs were forbidden as the impact of these variables on each other is beyond the scope of the research, as well as a complete data set for four months and a half was taken, so altering the probabilities of one of the variables will not have a major impact on the other two variables.

For the second network, the impact of the Month, Weekday and Time variables on both the 'NewOutsideTemperature' and 'NewOutsideHumidity' were restricted, as the changes in outside conditions depending on the date is out of the interest of this research. The reversed (outside variables on the date) was also restricted based on the same reasoning.

For the experiment the variables were assigned as follows:

Tier 1: 'Timeshift', 'Weekday', 'NewMotionCount', 'CO2Motion'

Tier 2: 'NewCO2', 'NewTemperature', 'NewHumidity'

The variables were assigned in this manner because the variables representing motion in the room, should impact the indoor controllers. The two variables symbolizing movement based on PIR and CO_2 are used for assessing the change in the 'NewTemperature', 'NewHumidity' and 'NewCO2'.

Only two arcs are forbidden, between the WeekDay and the TimeShift and the other way around. Similarly to the historical data, the connection between the two variables is outside the scope of this research.

3.6.2. Reasoning within the network

The BBN is used for making predictions about certain events conditioned upon entered evidence. According to (Fenton, 1995), there are a number of reasoning types considered as valid conditional propositions in a BBN. The different types are:

- Causal determination – it reflects the causal structure of the real world and provides an explanatory framework for prediction. It helps express theories about how the world operates and encoding uncertainties about its operation. Models structured by using causal ideas is a natural process for the expert as these are easily explained by comparison to the reality

- Statistical determination – BBNs are used as representation of statistical determination. The probabilities of events are determined by chance, or by selecting a particular event from a list of possible events or from a stochastic process.
- Structural determination – is modelled by employing logical or probabilistic relations. This type can be deterministic (relations between nodes are logical), definitional (nodes define the meaning of other nodes), architectural (nodes are related to a certain pattern), analogical (nodes inherit attributes of other nodes).

There are, of course, overlaps between the casual and statistical models of determination, statistical models being included in causal models, since the causal models also admit to chance.

The reasoning within this network incorporates various characteristics from the three types of determination. Firstly, the network is constructed by means of statistical determination, because the probabilities of each event are allocated based on real-time data and by following a statistical learning algorithm. Secondly, a purely causal determination is used when considering how the variables are assigned in different tiers. At the same time, the evidence is entered based on a causal determination, according to potential outcomes and the need of answering the research questions. Lastly, a structural determination is used when forcing or restricting certain arcs, as the intended relation between the arcs needs to be a logical one.

3.7. Conclusions

This chapter introduces the research approach. Explanation of the method and motivation behind the choice for this method is provided, together with the tools used for the analysis. Descriptive analyses (cross-tabulation, frequencies, linear regression and bivariate correlations) and BBN are used as methods for analysis in this research project. These methods can be used for detecting dependencies between variables, and for estimating certain events or the probability of occurrence of certain events. This is needed for predicting future patterns and ultimately for achieving the research objective.

Moreover, the way in which data is collected and processed is explained. Historical data on CO_2, humidity and temperature is collected for predicting future patterns in the IAQ. An experiment is conducted, where CO_2, humidity, temperature and PIR data is collected along with ground truth data, for checking the impact of occupancy on the IAQ. In terms of data processing, the inconsistencies and missing cases are removed and variables are re-coded based on the information provided by the company Inteliments, and based on the decision tree procedure.

Finally, an introduction into the analysis of the results (the model set-up) is provided in this section. The last part of the model set-up serves as reference for the following chapter.

4

RESULTS AND DATA ANALYSIS

4.1. Introduction

This study aims to answer two research questions:

1. ***How can future states of CO_2, humidity and temperature be predicted based on historical sensor data, for improving the indoor air quality?***
2. ***A) Literature: What sensor data can provide the most accurate information in detecting real-time occupancy information?***
   B) Case study: **How accurate is the information fusion given by CO_2 and PIR sensors in detecting real-time occupancy?**

For answering both of these questions, the results of the analyses will be discussed further in this chapter. Firstly, the historical data is analyzed, consisting of data gathered from two office spaces, over a period of five months. Both the descriptive analyses resulting graphics, and the construction of the Bayesian Belief Network with the proposed scenarios are displayed and discussed. Furthermore, the experimental data is analyzed, which consists of data gathered over the period of a week including, information from occupancy sensors (PIR). The experimental data is compared to the ground truth data and displayed through charts and tables made in SPSS.

4.2. Historical Data

Historical data on CO_2, humidity and temperature is collected over a period of five months, for answering the first research question. Furthermore, information about outside temperature is gathered, and illustrated in a separate Bayesian Belief Network, for checking the impact of the meteorological factors on the indoor characteristics. Next, the analyses for two different office spaces will be presented.

4.2.1. Office 1: Credo Room

Descriptive Analyses

This section presents the most important graphics and tables based on historical data collected from the first office space. Table 13 shows the cases for the three interesting variables for this analysis: the CO_2, the humidity and the temperature. There are over 30000 cases used in this analysis, where the CO_2 variates from 371 to 946 parts per million (ppm), the humidity is from 16.7% to 49.6% and the temperature has a minimum of 20.5°C and a maximum of 29.8°C.

Figure 4 shows how the CO_2 levels variate, a big amount of cases being concentrated between 371 ppm and 451 ppm. The result is caused by the big amount of cases registered during the night, when the level of CO_2 is maintained quite low, due to the lack of activity in the room. Figure 5 shows the frequency of temperature cases, where the biggest amount of these is observed between 21.1°C and 24.1°C. This result might be caused by the indoor temperature being maintained at an adequate level for working. Figure 6 emphasizes the variation of humidity, where most cases registered are between 32.8% and 37.8%. Up to 40% humidity, the space is considered to be dry for the working environment. The presence of low humidity cases can be justified by the temperature inside and outside the room, in combination with maintaining the space closed during the night, where most of the cases are registered.

Table 13: Cases for historical data in Credo Room

Statistics

		co2	humi	temp
N	Valid	30138	30138	30138
	Missing	0	0	0
Mean		477.63	34.529	23.268
Median		454.00	34.600	22.900
Std. Deviation		81.086	6.0616	1.7318
Minimum		371	16.7	20.5
Maximum		946	49.6	29.8

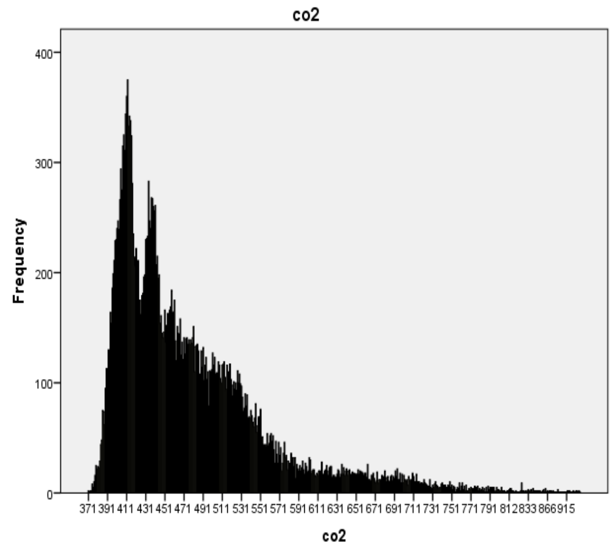


Figure 4: CO_2 levels (ppm) Credo Room

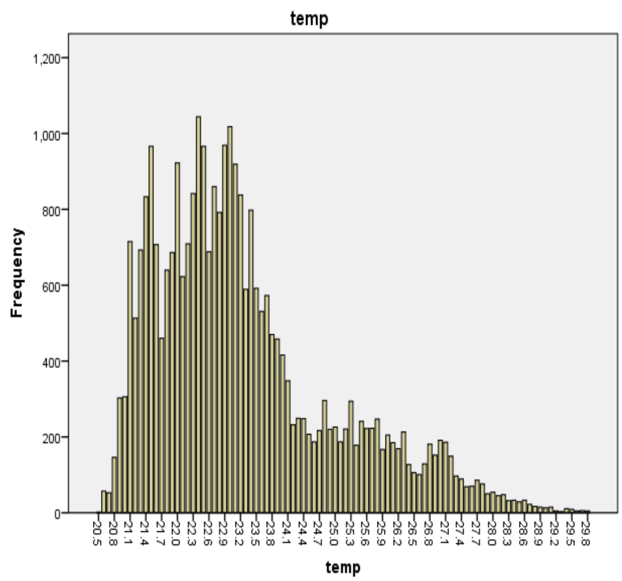


Figure 5: Temperature Levels (°C) Credo Room

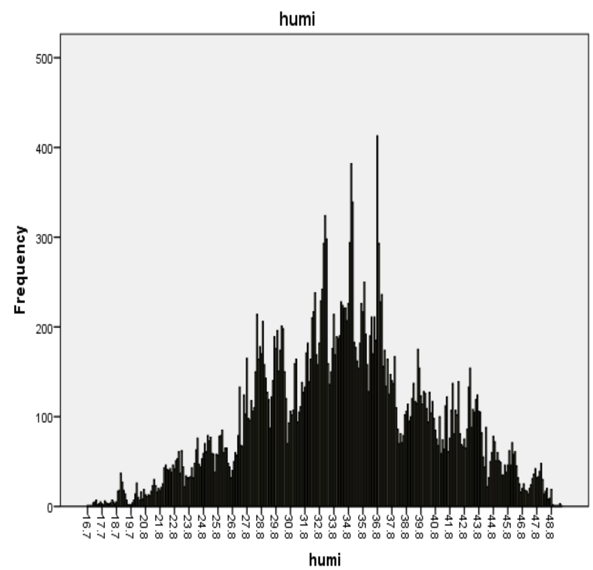


Figure 6: Humidity Levels (%) Credo Room

Further on, the variables were re-coded from continuous into discrete data, based on the expertise from the company Inteliments (which provided the data). The three variables are: 'NewCO2', 'NewTemperature', 'NewHumidity'. Different levels for each variable were made, as explained in the section 3.5.2.

Table 14 illustrates the number of cases for 'NewCO2' registered for each of the re-coded levels (healthy, acceptable and stiffness), while Figure 7 complements the table graphically. It can be observed that the healthy and the acceptable levels are the most present in the data set, while the stiffness is represented by only 791 of the cases. This is connected to the amount of people present at different times during the day. As the night is represented by most of the cases, a predominant healthy level should be registered by the data. During the day, when the space is populated, the CO_2 levels will increase to an acceptable level, until a corrective action is taken (such as opening the window, or the people leave the room during breaks). The last level is represented by the exceptional cases, when a big amount of people is present, for a longer period of time than in a normal working day.

Table 14: New CO2 levels after recoding in SPSS

NewCO2	Frequency	Valid Percent
Healthy Level	14471	48.0
Acceptable Level	14876	49.4
Stiffness/odors	791	2.6
Total	30138	100

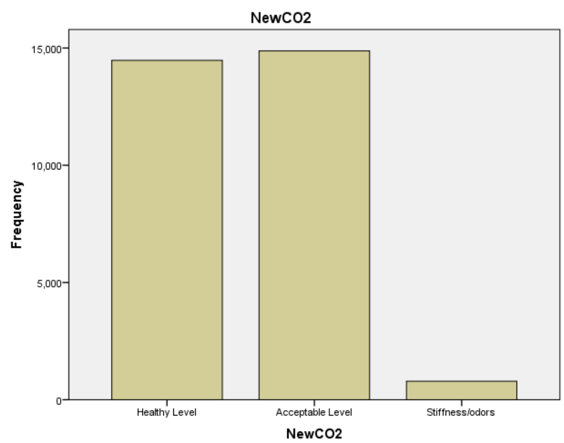


Figure 7: Levels of CO2 after re-coding in SPSS - chart

The temperature cases are re-coded into three different levels (comfortable 1, comfortable 2, uncomfortable high), as described in section 3.5.2. It is observed that there is mostly comfortable temperature present, with only 3031 cases of highly uncomfortable temperatures for working. As previously mentioned, this can be caused by the desire to keep an adequate level in the room by the employees, which impacts the operation of the HVAC system and the frequency of opening the windows and doors. Table 15 and Figure 8 illustrate the 'NewTemperature' variable where the variation in temperature levels in presented.

Table 15: New temperature levels after recoding in SPSS

NewTemperature	Frequency	Valid Percent
Level 1	15495	51.4
Level 2	11612	38.5
Uncomfortable High	3031	10.1
Total	30138	100

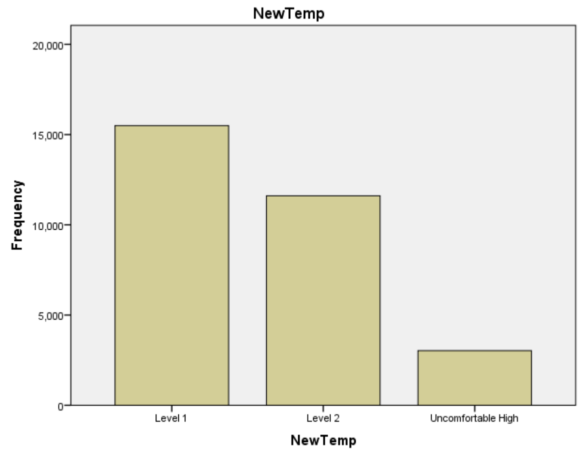


Figure 8: Levels of temperature after re-coding - chart

Lastly, two humidity levels are re-coded in SPSS, for matching the levels indicated by the company as adequate levels for the working environment. The cases within the two levels can be visualized in Table 16 and Figure 9. As suggested by the table and its resulting bar chart, the minimal humidity is present in the majority of the cases, with a percent of 80.4, while the optimal humidity is present only in 19.6 of the cases. The result can be connected to the season in which the data was collected, which is autumn and winter (September to January), when the air is mostly dry, especially due to the operation of the HVAC system inside the space.

Table 16: New humidity levels after re-coding in SPSS

NewHumidity	Frequency	Valid Percent
Minimal uncomfortable humidity	24245	80.4
Optimal Humidity	5893	19.6
Total	30138	100.0

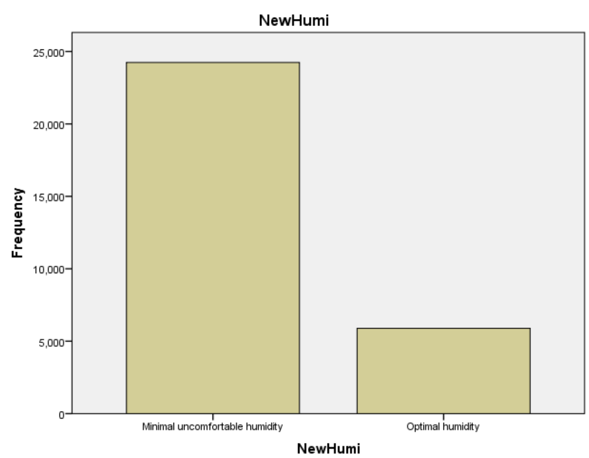


Figure 9: Levels of humidity after recoding - chart

To complement the future Bayesian Belief Network, and for answering the first research question: "*How can future states of CO_2, humidity and temperature be predicted based on historical sensor data, for improving the indoor air quality?*", the three variables previously described ('NewCO_2', 'NewTemperature', 'NewHumidity'), are presented in comparison to the weekdays, months and time shifts, in a cross-tab. A cross-tab is a procedure in SPSS that cross-tabulates two or more variables. It creates contingency tables from the multivariate frequency distribution of variables, presented in a matrix format (Optimus, 2017). The main results of the cross-tabulation process are illustrated subsequently. The chi-square tests are also performed for testing the strength of the cross-tabs, and all of the subsequent cross-tabs illustrate a positive chi-square test, with a low chi-square value below 0.005. The degree of freedom and the significance levels are illustrated after each cross-table.

Table 17 shows the cross-tab between the CO_2 levels and the weekdays with the assigned percentages, while Figure 10 is included for visually representing how the CO_2 levels change based on each day of the week. On one hand, when looking at the healthy levels, the biggest amount of cases is registered during the weekends, especially on Sunday (26.8%). If the weekend is excluded, the healthy level is the highest on Friday (15.7%), followed by Thursday (13.3%) and Monday (11.6%). Wednesday and Tuesday present a low percentage favorable towards a healthy level being maintained in the room. On the other hand, the acceptable level is the highest on Wednesday (19.8%), followed by Tuesday (18.6%) and Thursday (16.5%). Lastly, the stiffness level can be found primarily on Tuesday (50.7%), followed by Wednesday (27.9%) and Monday (17.7%). Overall, this results indicate the highest amount of emitted CO_2 on Tuesday and Wednesday.

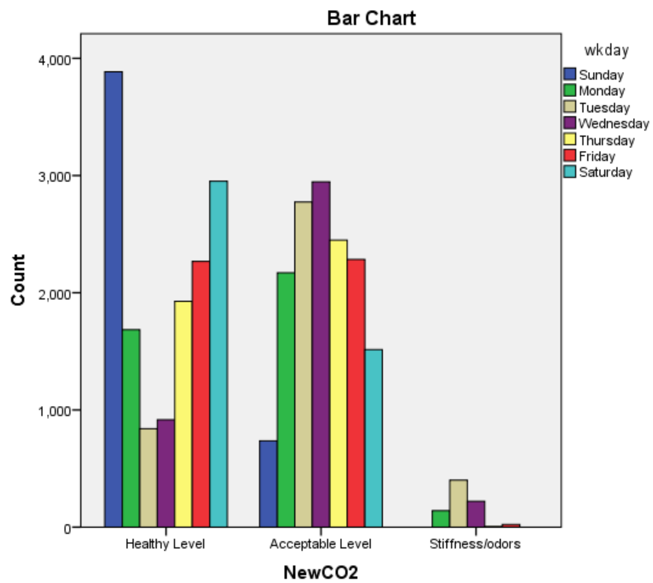


Figure 10: Bar chart: NewCO2 with the week days

The result can be connected to the room occupancy, which indicates that this office space is mostly populated during the indicated days, when a corrective action for maintaining the air at a healthy and acceptable level is needed.

Table 17: Adapted crosstab NewCO2 with the week days

		WeekDay						
		Sun	Mon	Tue	Wed	Thu	Fri	Sat
Healthy Level	%within NewCO2	***26.8%**	**11.6%**	5.8%	6.3%	**13.3%**	**15.7%**	20.4%
	%within WeekDay	84.1%	42.2%	20.9%	22.5%	44%	49.6%	66.1%
	% of Total	12.9%	5.6%	2.8%	3.0%	6.4%	7.5%	9.8%
Acceptable Level	%within NewCO2	4.9%	14.6%	**18.6%**	**19.8%**	**16.5%**	15.4%	10.2%
	%within WeekDay	15.9%	54.3%	69.1%	72.1%	55.9%	49.9%	33.9%
	% of Total	2.4%	7.2%	9.2%	9.8%	8.1%	7.6%	5%
Stiffness/odors	%within NewCO2	0%	**17.7%**	**50.7%**	**27.9%**	0.8%	2.9%	0%

	%within WeekDay	0%	3.5%	10%	5.4%	0.1%	0.5%	0%
	% of Total	0%	0.5%	1.3%	0.7%	0%	0.1%	0%
Pearson Chi-Square								
Df (degrees of freedom) = 12			**Asymp. Sig. (2 sided)= 0.000					

* Numbers in bold are explained in the text and are representative for the analysis

**Asymp. Sig. (2 sided) = asymptotic significance, the significance of the chi-square test

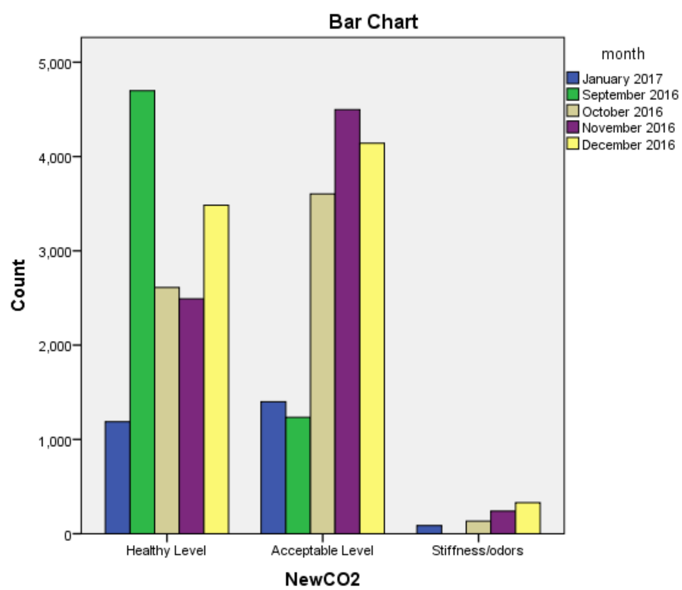


Figure 11: Bar chart: NewCO2 with the months

Furthermore, the CO_2 is compared to the months in which data was collected (September – January). Figure 11 illustrates how each CO_2 level variates across the months, while Table 18 is a cross-tab showing the exact percentages and number of cases for the three levels and for each month. When looking at the healthy level, the biggest percent is present in September (32.5%), followed by December (24.1%) and October (18%). The acceptable level is mostly found in November (30.2%), followed by December (27.8%) and October (24.2%). Lastly, the stiffness level is mostly present in December (41.6%), followed by November (30.5%) and October (16.9%).

When looking at each month, September illustrates mostly a healthy level (79.2%), while October, November, December and January all illustrate a higher percent of acceptable level present.

Overall, the mentioned results suggest that a healthy level is primarily found in September, which might be caused by a decreased occupancy in the room, either due to the holiday season or due to opening the windows in the room. November and December are the months when most CO_2 is emitted, hence, overall the highest amount of people can be found in the office in these months.

Table 18: Crosstab NewCO2 with the months

		Month				
		January 2017	September 2016	October 2016	November 2016	December 2016
Healthy Level	%within NewCO2	8.2%	**32.5%**	**18%**	17.2%	**24.1%**

	%within Month	44.4%	**79.2%**	41.1%	34.5%	43.8%
	% of Total	3.9%	15.6%	8.7%	8.3%	11.6%
Acceptable Level	%within NewCO2	9.4%	8.3%	**24.2%**	**30.2%**	**27.8%**
	%within Month	52.4%	20.8%	**56.8%**	**62.2%**	**52.1%**
	% of Total	4.6%	4.1%	12%	14.9%	13.7%
Stiffness/odors	%within NewCO2	11%	0%	**16.9%**	**30.5%**	**41.6%**
	%within Month	3.3%	0%	2.1%	3.3.%	4.1%
	% of Total	0.3%	0%	0.4%	0.8%	1.1%
Pearson Chi-Square						
Df = 8			Asymp. Sig. (2 sided)= 0.000			

The CO_2 is also compared to the time shifts. Following the same procedure as for the week and month, the three CO_2 levels are presented in Figure 12, based on six time periods, as previously described. Table 19 is a crosstab showing which percentage is allocated to each CO_2 level according to the time period. It also shows the percentages connected to the bar chart from Figure 12, emphasizing the change in the three levels along the time frames.

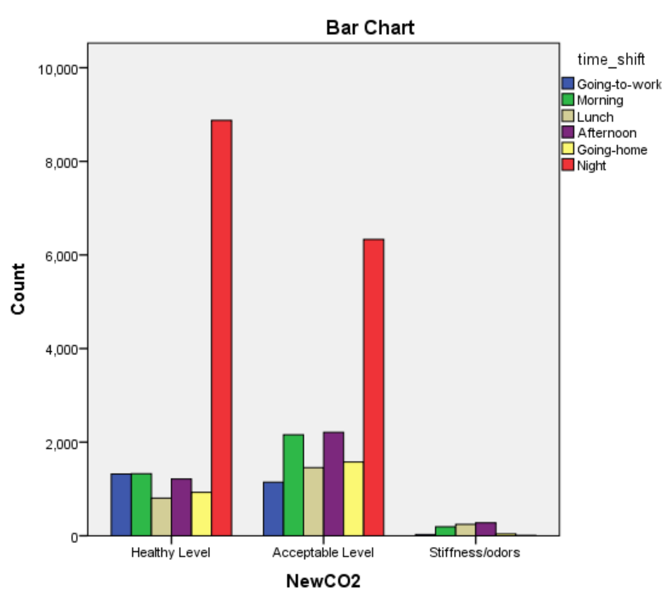


Figure 12: Bar chart: NewCO2 with the time shifts

The results of the comparison between the CO_2 levels and the time shifts indicate that the healthy level is highly present during the night (61.3%). This is followed by the going-to-work and morning periods with much smaller percentages (9.1% and 9.2%). The acceptable level is also found mostly during the night (42.6%), however it is followed by the afternoon and morning periods, with higher percentages of CO_2 levels than in the previous case (14.8% and 14.5%).

The level of stiffness is present frequently during the afternoon (35.1%), followed by the lunch and the morning period (31% and 24.3%). When assessing how the levels behave across the day, the healthy level is present mostly during the night and in the going-to-work period, while during all the other periods, an acceptable level can be found in the office. This result suggests that the healthy level is maintained only during the night and in the going-to-work period, which is the time frame when there are no people in the room. An increase in CO_2 emission is mostly present during the afternoon, lunch and morning periods, which is actually indicated by the time when people are inside the room. Attention should be especially given to the particular cases of stiffness level (above 700 ppm), and a corrective action should be taken for keeping a good indoor air quality. Specifically, as the afternoon is the most inclined towards reaching this level, windows and doors should be open in the lunch period. Also, a suggestion could be given to the employees to leave the room during the lunch period, for re-establishing the desired CO_2 level.

Table 19: Crosstab NewCO2 with the time shifts

		Timeshift					
		Going-to-work	Morning	Lunch	Afternoon	Going-home	Night
Healthy Level	%within NewCO2	**9.1%**	**9.2%**	5.6%	8.4%	6.4%	**61.3%**
	%within Timeshift	**53%**	36.1%	32.1%	32.8%	36.5%	**58.3%**
	% of Total	4.4%	4.4%	2.7%	4%	3.1%	29.5%
Acceptable Level	%within NewCO2	7.7%	**14.5%**	9.8%	**14.8%**	10.6%	**42.6%**
	%within Timeshift	45.9%	**58.1%**	**58.1%**	**59.7%**	**61.9%**	41.6%
	% of Total	3.8%	7.2%	4.8%	7.3%	5.2%	21%
Stiffness/odors	%within NewCO2	3.5%	**24.3%**	**31%**	**35.1%**	5.1%	1%
	%within Timeshift	1.1%	5.2%	9.8%	7.5%	1.6%	0.1%
	% of Total	0.1%	0.6%	0.8%	0.9%	0.1%	0%
Pearson Chi-Square							
Df = 10			Asymp. Sig. (2 sided)= 0.000				

Similar analysis, as for the CO_2, is performed for the other two variables: the temperature and the humidity. The change in temperature levels according to the week days is found in Figure 13 while Table 20 is the crosstab for these variables. For this analysis, it is interesting to see how the temperature is maintained during each day of the week. It can be observed that the temperature of Level 1 is present during almost all the days of the week. The exception is Thursday (47.1%) and Friday (51%), when Level 2 can be found the most in the room. In contrast, when looking at how each level performs, it is also interesting to see that an increased percent of high uncomfortable temperature is found during the weekend (Sunday - 21.2% and Saturdays –

18.7%). During the working days the biggest uncomfortable level is observed on Tuesday (14.6%) and Monday (13.5%).

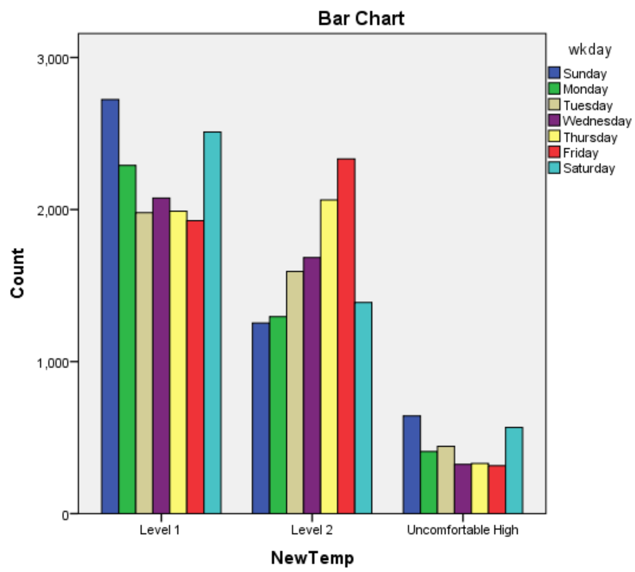


Figure 13: Bar chart: NewTemperature with the week day

From these results it might be concluded that, as during the weekends there are no people present in the room, the temperature is not maintained at an adequate level for working. The results also show that, during all the other days, the temperature is kept at Level 1 (between 19 – 22.9°C), except for the last two working days (Thursday and Friday).

On one hand this might suggest that in the last two working days the temperature is less adjusted, and therefore, the office is less populated. On the other hand, this result might also suggest that the increased temperature is the effect of more people present in the room, without any corrective action being taken for regulating the temperature level. When looking at the results from the CO_2 levels according to the week day, it can be seen that the healthy level is mostly found on Thursday and Friday, so the occupancy is low. If the CO_2 result is compared to the temperature one, it can be assumed that the first conclusion (less occupancy and less regulation of the HVAC system) is likely to be correct.

Table 20: Crosstab NewTemperature with the week day

		WeekDay						
		Sun	Mon	Tue	Wed	Thu	Fri	Sat
Level 1	%within NewTemperature	17.6%	14.8%	12.8%	13.4%	12.8%	12.4%	16.2%
	%within WeekDay	58.9%	57.3%	49.3%	50.8%	45.4%	42.1%	56.2%
	% of Total	9%	7.6%	6.6%	6.9%	6.6%	6.4%	8.3%
Level 2	%within NewTemperature	10.8%	11.2%	13.7%	14.5%	17.8%	20.1%	12%
	%within WeekDay	27.1%	32.4%	39.7%	41.2%	**47.1%**	**51%**	31.1%
	% of Total	4.2%	4.3%	5.3%	5.6%	6.8%	7.7%	4.6%
Uncomfortable High	%within NewTemperature	**21.2%**	**13.5%**	**14.6%**	10.7%	10.9%	10.4%	**18.7%**

	%within WeekDay	13.9%	10.2%	11%	7.9%	7.5%	6.9%	12.7%
	% of Total	2.1%	1.4%	1.5%	1.1%	1.1%	1%	1.9%
Pearson Chi-Square								
Df = 12	Asymp. Sig. (2 sided)= 0.000							

The temperature is presented compared to the months, in Figure 14 and Table 21. Figure 14 shows how the three levels variate along the months, while the cross-tab in Table 21 illustrates the allocated percentages for each level. When looking along the months, September shows mostly an increased temperature inside the room (52.1% - Level 2 and 47.9 - Uncomfortable High). In October, Level 2 is mostly present with 80.7%, while in November, December and January, Level 1 is predominant. In November, December and January there are actually zero cases illustrating an Uncomfortable High temperature.

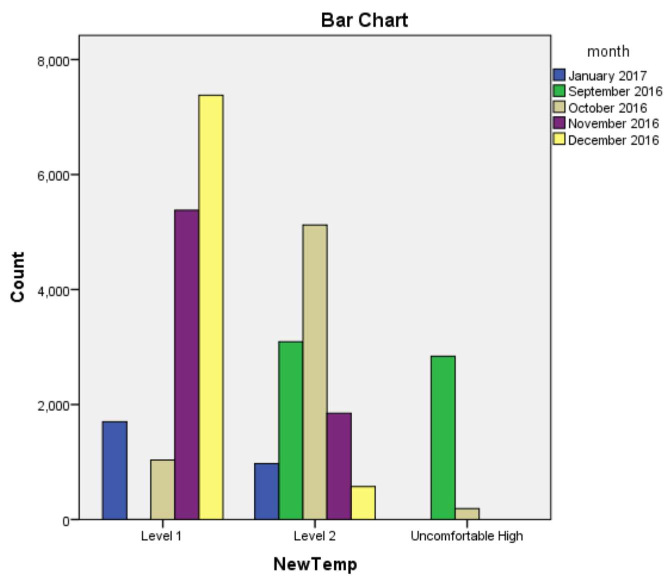


Figure 14: Bar chart: NewTemp with the month

This results indicate that the indoor temperature is related to the season, and that during the autumn period a higher indoor temperature is maintained, compared to the winter period. For checking how the outdoor temperature impacts the indoor temperature an analysis is performed, by adding the corresponding outside temperature in the data sets. The analysis of the weather impact is further explained in this chapter, in the Bayesian Belief Network results.

Table 21: Crosstab NewTemp with the month

		Month				
		January 2017	September 2016	October 2016	November 2016	December 2016
Level 1	%within NewTemperature	11%	0%	6.7%	34.7%	47.6%
	%within Month	**63.6%**	0%	16.3%	**74.4%**	**92.8%**
	% of Total	5.6%	0%	3.4%	17.9%	24. 5%
Level 2	%within NewTemperature	8.4%	26.6%	44.1%	15.9%	5%
	%within Month	36.4%	**52.1%**	**80.7%**	25.6%	7.2%
	% of Total	3.2%	10.3%	17%	6.1%	1.9%

Uncomfortable High	%within NewTemperature	0%	93.7%	6.3%	0%	0%
	%within Month	0%	**47.9%**	3%	0%	0%
	% of Total	0%	9.4%	0.6%	0%	0%
Pearson Chi-Square						
Df = 8		*Asymp. Sig. (2 sided)= 0.000				

Lastly, the temperature is also presented across the time periods. In Figure 15, all three levels (Level 1, 2 and Uncomfortable High), are present mostly during the night. As Table 22 indicates, during the day, Level 1 is mostly registered during the morning (11.2%), while Level 2 and the uncomfortable level are mostly present during the afternoon (16.7% and 20.4%).

However, if each period is individually observed, in the going-home, night, going-to-work and morning periods, Level 1 is primarily found in the room, while during lunch and afternoon Level 2 is found in a bigger ratio.

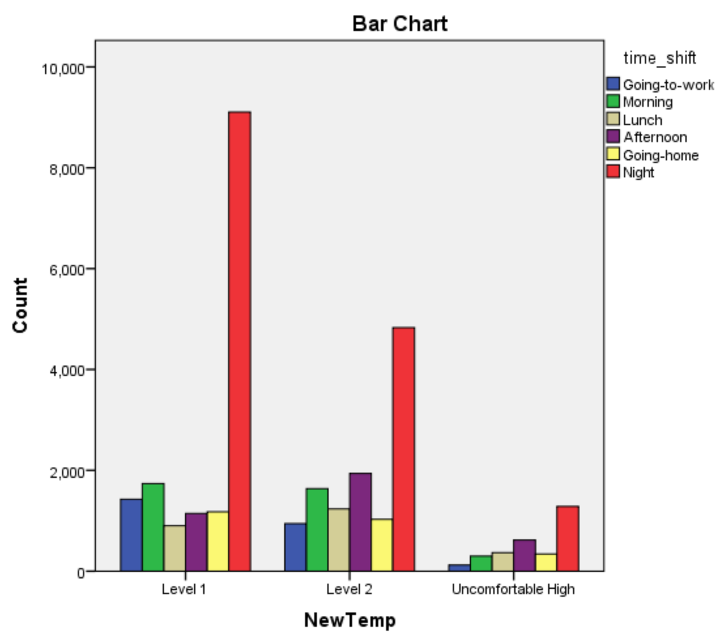


Figure 15: Bar chart: NewTemp with the time shift

Comparing these results to the conclusions drawn in the CO_2 analysis, the high percent of increased temperature during the afternoon and lunch period can be connected to the high stiffness level during the same period. A conclusion on occupancy cannot be, however, drawn based on temperature, as this is highly dependent on other factors such as: people`s preference for temperature, outside temperature, or the rooms situated next to the analyzed office, which are connected to the HVAC system.

Table 22: Crosstab NewTemp with the time shift

		Timeshift					
		Going-to-work	Morning	Lunch	Afternoon	Going-home	Night
Level 1	%within NewTemperature	9.2%	**11.2%**	5.8%	7.4%	7.6%	58.8%

	%within Timeshift	**57.3%**	**47.3%**	36%	30.9%	**46.2%**	**59.8%**
	% of Total	4.7%	5.8%	3%	3.8%	3.9%	30.2%
Level 2	%within NewTemperature	8.1%	14.1%	10.7%	**16.7%**	8.8%	41.6%
	%within Timeshift	37.9%	44.5%	**49.3%**	**52.4%**	40.4%	31.7%
	% of Total	3.1%	5.4%	4.1%	6.4%	3.4%	16%
Uncomfortable High	%within NewTemperature	4%	9.9%	12.2%	**20.4%**	11.3%	42.4%
	%within Timeshift	4.9%	8.1%	14.7%	16.7%	13.4%	8.4%
	% of Total	0.4%	1%	1.2%	2%	1.1%	4.3%
Pearson Chi-Square							
Df = 10			Asymp. Sig. (2 sided)= 0.000				

The last variable analyzed is the humidity, compared again to the weekday, the month and the time period. For checking how the humidity changes based on the days of the week, Figure 16 is created for the two levels present in the data set: the optimal humidity and the minimal uncomfortable humidity. Table 23 is the crosstab showing the change in humidity level as well as the percentages for each level according to the day of the week.

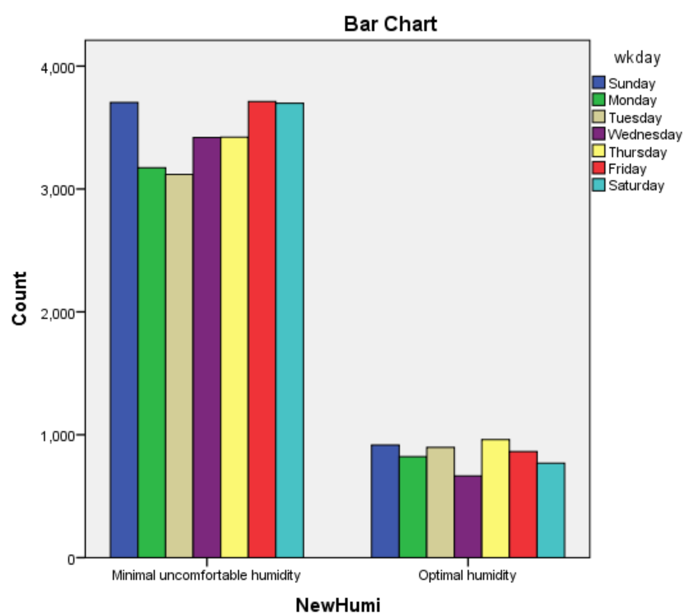


Figure 16: Bar chart: NewHumi with the week day

It can be observed that both levels perform similarly, with small differences in percentages when looking at each level individually. In the minimal humidity category the highest percentages are observed on Saturday and Sunday (15.2% and 15.3%), while in the optimal humidity category the highest percentage is observed on Thursday (16.3%). When looking at each day individually, the biggest percentage is given by the minimal humidity for all the days. Overall, these results suggest increased percentages of undesirable humidity for the working environment and a corrective action should be taken for all the days of the week, independent on the amount of people present in the room.

For improving the working comfort the introduction of a humidifier and the maintenance of temperature within the Level 1 is recommended.

Table 23: Crosstab NewHumi with the week day

		WeekDay						
		Sun	Mon	Tue	Wed	Thu	Fri	Sat
Minimal uncomfortable humidity	%within NewHumidity	15.3%	13.1%	12.9%	14.1%	14.1%	**15.3%**	**15.2%**
	%within WeekDay	**80.2%**	**79.4%**	**77.7%**	**83.7%**	**78.1%**	**81.1%**	**82.8%**
	% of Total	12.3%	10.5%	10.3%	11.3%	11.4%	12.3%	12.3%
Optimal humidity	%within NewHumidity	15.6%	13.9%	15.2%	11.3%	**16.3%**	14.6%	13%
	%within WeekDay	19.8%	20.6%	22.3%	16.3%	21.9%	18.9%	17.2%
	% of Total	3%	2.7%	3%	2.2%	3.2%	2.9%	2.6%
Pearson Chi-Square								
Df = 6			Asymp. Sig. (2 sided)= 0.000					

Following, the humidity is also illustrated within the month. Figure 17 shows how each level performs during five months. Table 24 is the cross-table illustrating the amounts of humidity along the months. Figure 17 shows that the biggest amount of minimal humidity is present during December and the lowest in September, while the highest level for optimal humidity is in September while in January there is no optimal humidity.

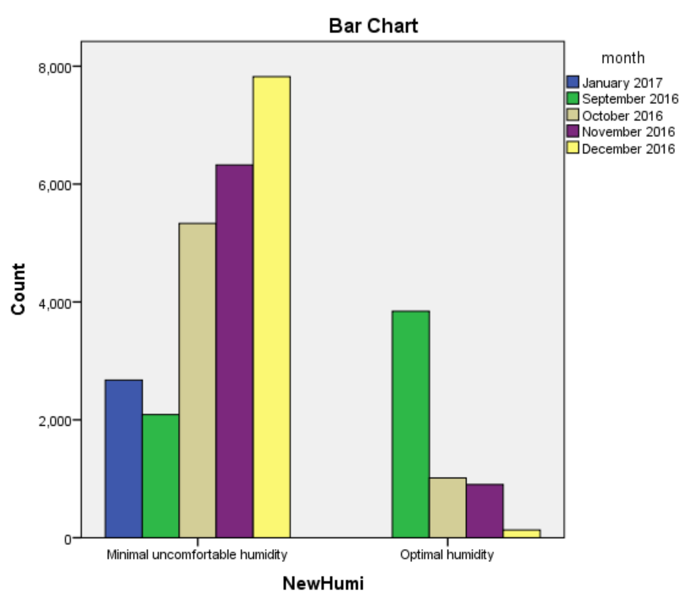


Figure 17: Bar chart: NewHumi with the month

Table 24 emphasizes that during the months October, November, December and January there is mostly minimal humidity in the office, and only September presents a higher percent in the optimal humidity category.

This result indicates that the season plays an important role in the increase of humidity. This assumption is connected to the operation of the heating system during winter (months November-January and partially October), due to the outside low temperature. Therefore, the heating system plays an important role in keeping the humidity at an optimal level in the office. An action for improving the humidity in the room needs to be taken, therefore, during the winter months.

Table 24: Crosstab NewHumi with the month

		Month				
		January 2017	September 2016	October 2016	November 2016	December 2016
Minimal uncomfortable humidity	%within NewHumidity	11%	**8.6%**	22%	26.1%	**32.3%**
	%within WeekDay	**100%**	35.2%	**84%**	**87.5%**	**98.3%**
	% of Total	8.9%	6.9%	17.7%	21%	26%
Optimal humidity	%within NewHumidity	**0%**	**65.2%**	17.2%	15.3%	2.2%
	%within WeekDay	0%	**64.8%**	16%	12.5%	1.7%
	% of Total	0%	12.8%	3.4%	3%	0.4%
Pearson Chi-Square						
Df = 4		Asymp. Sig. (2 sided)= 0.000				

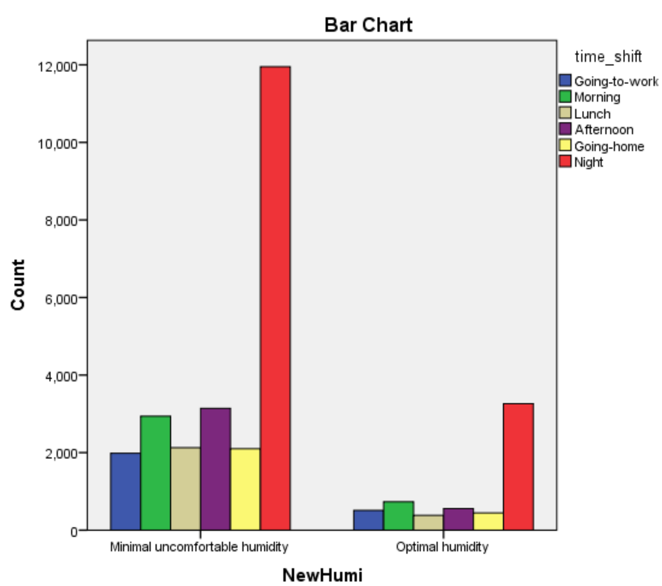


Figure 18: Bar chart: NewHumi with the time period

Lastly, the humidity variable is associated to the time periods and this association can be found in Figure 18. To complement the bar chart, Table 25 gives the percentages for humidity levels within time shifts and the percentages for each of the two levels within the humidity variable. Similarly to the other two variables (CO_2 and temperature) the highest percentages for all the levels are found during the night, due to the majority of the cases being registered then. However, if the night cases are excluded, the most uncomfortable humidity is found during the afternoon (13%) and morning (12.1%), while the optimal humidity is mostly found during the morning (12.5%). Similarly, for all the parts of the day an uncomfortable humidity is present, which indicates that the air is too dry.

Table 25: Crosstab NewHumi with the time period

		Timeshift					
		Going-to-work	Morning	Lunch	Afternoon	Going-home	Night
Minimal uncomfortable humidity	%within NewHumidity	8.2%	**12.1%**	8.8%	**13%**	8.7%	**49.3%**
	%within WeekDay	**79.5%**	**80%**	**84.8%**	**84.9%**	**82.6%**	**78.6%**
	% of Total	6.6%	9.8%	7.1%	10.4%	7%	39.7%
Optimal humidity	%within NewHumidity	8.7%	**12.5%**	6.5%	9.5%	7.5%	**55.4%**
	%within WeekDay	20.5%	20%	15.2%	15.1%	17.4%	21.4%
	% of Total	1.7%	2.4%	1.3%	1.9%	1.5%	10.8%
Pearson Chi-Square							
Df = 5		Asymp. Sig. (2 sided)= 0.000					

Bayesian Belief Network Results

The two original networks with their strength of influence are displayed in Figure 19 and Figure 20.

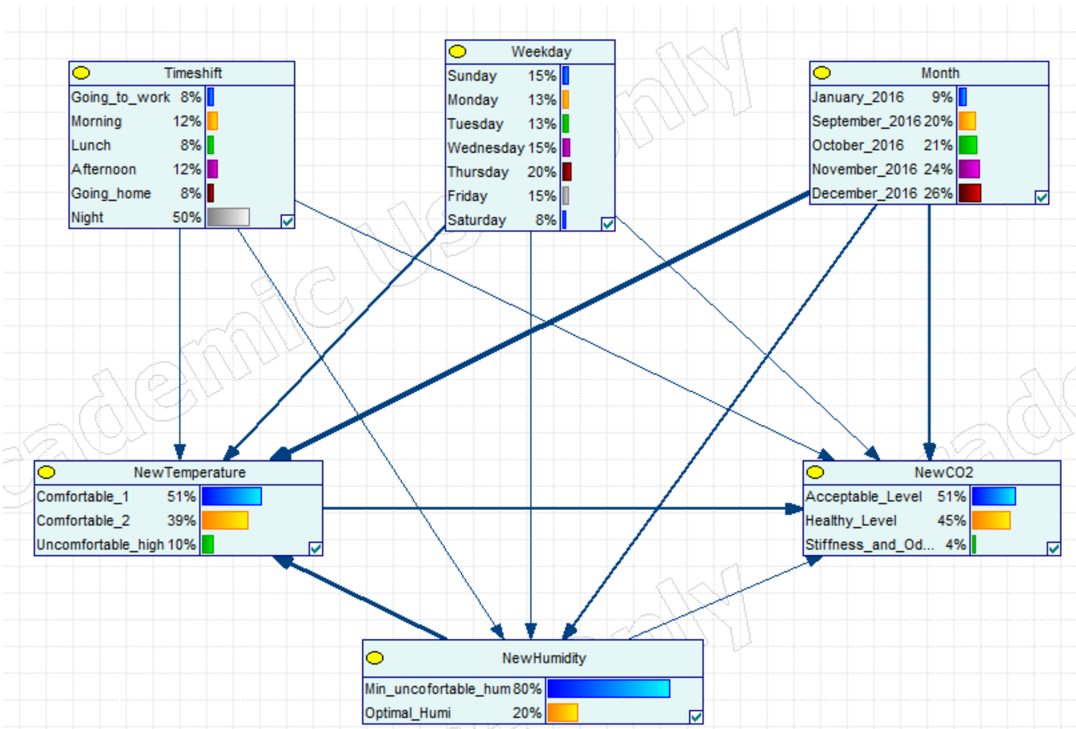


Figure 19: Original Bayesian Belief Network– Credo Room historical data

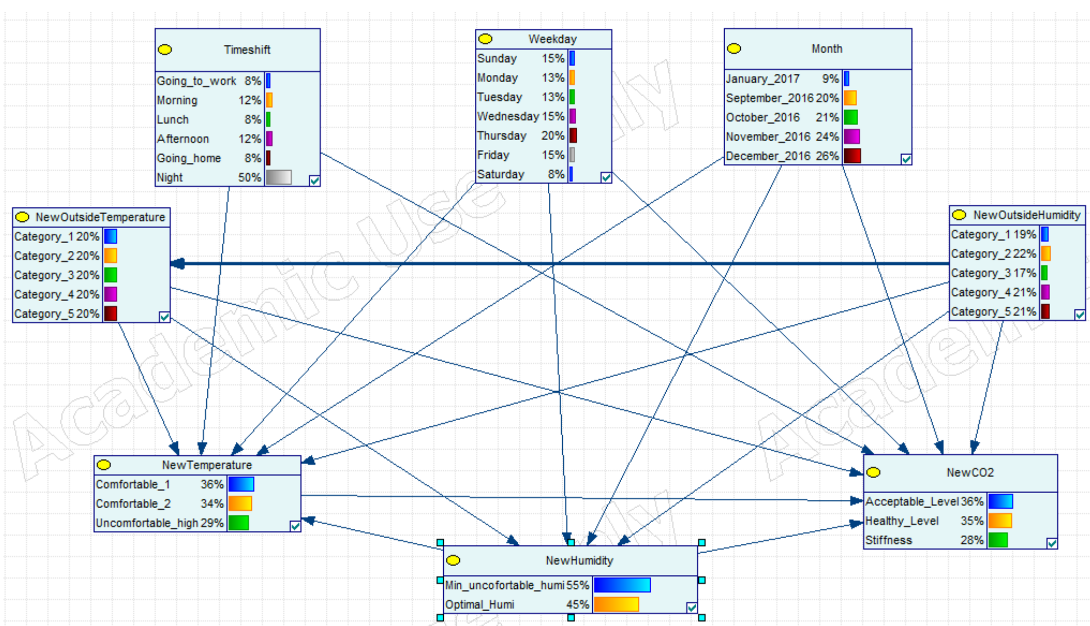


Figure 20: Bayesian Belief Network with outdoor variables – Credo Room historical data

Strength of Influence

Table 26 and Table 27 show the strength of influence between each arcs in the two networks illustrated in Figure 19 and Figure 20. According to the user manual (BayesFusion, 2017), the arcs have different thickness, depending on the strength of influence between the child and parent node. This strength is calculated based on the CPT of the child node and expresses the distance between the probability distribution of the child node dependent on the state of the parent node.

In Table 26, it can be seen that the strongest influence is between the month and the temperature inside the room, which means that if a change is applied in the variable 'Month', the temperature will change the most significantly, compared to other variables. The second strongest influence is given by the humidity on the temperature, while the third is again given by the variable 'Month' on the humidity inside the room. This indicates that the 'Month' is an important variable in the analysis, and its impact is the strongest on the indoor conditions, compared to the week day and the time shift. Furthermore, the influence of the time period on the humidity and the CO_2 is marked by the lowest strength of influence, which implies that its effect is limited compared to the effect of the month or the week day. The influence of the month might suggest that the season plays a more important role on the indoor conditions, compared to the other two variables. Therefore, for possible predictions of CO_2, humidity and temperature, the month should be primarily taken into consideration.

The high influence of the humidity on the temperature is also an interesting result, which suggests that the temperature and the humidity are highly related. This is the case, as the heating impacts how humid the air in the room will be, therefore the more the temperature increases, the higher the humidity will be.

Table 26: Strength of arcs for BBN without external variables – Credo Room

Origin (Parent Node)	Destination (Child Node)	Strength of influence
Month	NewTemperature	0.4223
NewHumidity	NewTemperature	0.3907
Month	NewHumidity	0.2886
NewTemperature	NewCO2	0.2409
Weekday	NewTemperature	0.2298
Month	NewCO2	0.2234
NewHumi	NewCO2	0.1973
TimeShift	NewTemperature	0.1787
Weekday	NewCO2	0.1697
Weekday	NewHumidity	0.1262
TimeShift	NewCO2	0.1176
TimeShift	NewHumidity	0.0733

In Table 27, the biggest strength of influence is between the outdoor variables (average relative humidity on average temperature), however analyzing this is outside the scope of the thesis. The highest strength of influence on the indoor variables, is given by the outside temperature on the humidity inside the room, followed by the month on the humidity, and the average relative humidity on the indoor humidity. This result suggests that the outdoor variables play an important role in the behavior of humidity in the room. This is not the case however for the CO_2, where the strength is much lower. Hence, the highest influence of the outdoor variables is on the humidity, followed by the temperature and lastly by the CO_2. This impact might be the reason of the increased number of cases registered within the minimal uncomfortable level of humidity. For predicting future levels of humidity an important role is played by the weather conditions, which should be taken into account.

Table 27: Strength of arcs for BBN with external variables – Credo Room

Origin (Parent Node)	**Destination (Child Node)**	**Strength of influence**
NewOutsideHumidity	*NewOutsideTemperature*	*0.3431*
NewOutsideTemperature	NewHumidity	0.1
Month	NewHumidity	0.0995
NewOutsideHumidity	NewHumidity	0.0944
Weekday	NewHumidity	0.0845
Month	NewTemperature	0.0579
NewOutsideTemperature	NewTemperature	0.0576
NewOutsideHumidity	NewTemperature	0.0544
NewHumidity	NewTemperature	0.0528
wkday	NewTemp	0.0491
FinalOutAVGTemp	NewCO2	0.0227
month	NewCO2	0.0227
FinalOutAVGHumi	NewCO2	0.0215
NewHumi	NewCO2	0.0205
NewTemp	NewCO2	0.0204
wkday	NewCO2	0.02
time_shift	NewTemp	0.0168
time_shift	NewCO2	0.0096

Scenarios

With the Bayesian Belief Network generated, it is possible to test different scenarios, in order to see how the variables interact with each other, and how the probabilities change when certain scenario is applied. The desired scenario is entered in the network, and the model predicts the respective percentages in regards to the entered change. The scenarios proposed for testing the validity of the second research question are:

- Scenario 1: Evidence for the variable "Month" – Monthly levels of CO_2 , humidity and temperature

- Scenario 2: Evidence for the variable "Weekday" – Weekly levels of CO_2 , humidity and temperature
- Scenario 3: Evidence for the variable "Timeshift" - Levels of CO_2 , humidity and temperature according to the time period

The first scenario tests the change in the variables CO_2, humidity and temperature when the evidence for each month is changed to 100%. In Table 28 the changes in percentages for September and December can be found. These months were selected and interpreted as they present the most interesting results and can be seen as the representative months for each season (autumn and winter). The networks generated for each month can be found in Appendix 5: Credo Room – BBN Scenarios.

Table 28: Scenario 1: Monthly levels

Variable	No Evidence	September – Evidence 100%	December – Evidence 100%
NewCO₂			
Healthy Level	45%	**74%**	42%
Acceptable Level	**51%**	23%	**53%**
Stiffness	4%	3%	5%
NewTemperature			
Comfortable1	**51%**	1%	**91%**
Comfortable2	39%	**52%**	8%
Uncomfortable_high	10%	46%	1%
NewHumidity			
Min_uncomfortable	**80%**	36%	**98%**
Optimal_Humidity	20%	**64%**	2%

In the first scenario, it can be observed that, based on the CO_2 emission, a healthy level is maintained mostly in September, while December presents an increase in both acceptable and stiffness levels. Similarly as suggested by the results given in the Descriptive Analysis, the increased healthy level in September (in the autumn season) suggests that there is less CO_2 emission then. Therefore corrective actions for decreasing the CO_2 should be taken less in September-October and more in the winter season (November-January). However, more data, for a full year should be taken into account, for being able to generalize this pattern.

In the temperature variable, there is a big difference between the months September and December. In September it is much more difficult to maintain an adequate temperature, therefore there is always a temperature higher than 23° C, while December presents in proportion of 91% a temperature between 19 and 22.9° C.

As illustrated in the strength of influence section, the season plays an important role on the indoor humidity and temperature, which is most probably also the cause of this big difference between

September and December. Furthermore, the humidity is also very different between the two seasons, therefore the same explanation can also be applied in this case.

The second scenario emphasizes the differences in indoor quality controllers based on the day of the week. Table 29 shows the changes in percentages for every day of the week. The networks generated for each weekday can be found in Appendix 5: Credo Room – BBN Scenarios.

Table 29: Scenario 2 - weekly levels

	No Evidence	**Mon – 100% Evidence**	**Tue – 100% Evidence**	**Wed – 100% Evidence**	**Thu – 100% Evidence**	**Fri – 100% Evidence**	**Sat – 100% Evidence**	**Sun – 100% Evidence**
NewCO_2								
Healthy Level	45%	42%	22%	24%	42%	**51%**	**65%**	**82%**
Acceptable Level	**51%**	**53%**	**67%**	**69%**	**56%**	47%	33%	16%
Stiffness	4%	5%	11%	7%	2%	2%	2%	2%
NewTemperature								
Comfortable1	**51%**	**57%**	**47%**	**49%**	45%	43%	**57%**	**62%**
Comfortable2	39%	31%	42%	41%	**47%**	**51%**	29%	23%
Uncomfortable_high	10%	12%	12%	10%	8%	6%	14%	15%
NewHumidity								
Min_uncomfortable	**80%**	**78%**	**75%**	**83%**	**78%**	**84%**	**82%**	**79%**
Optimal_Humi	20%	22%	25%	17%	22%	16%	18%	21%

The CO_2 is maintained at an acceptable level for all the working days (Monday – Friday), while during the weekend a healthy level is present. From these results it is important to point out that during Tuesday and Wednesday, there is a higher stiffness level than during the other days, which should be taken into account when adjusting the climate control system, or taking corrective actions such as opening the windows and doors.

The temperature is mostly maintained at a comfortable level, the only days when the uncomfortable high temperature percentage increases are during the weekends and Monday/Tuesday. Also, Thursday and Friday present an increase, by having mostly Level 2 of temperature. During those days, careful consideration should be given to the adjustment of a comfortable temperature, with emphasis on Monday and Tuesday.

As previously mentioned in the descriptive analysis, the humidity is constantly at an uncomfortable level, so a single corrective action might not be sufficient for increasing the humidity in this office. Consequently, an investigation on the comfort time compared to the productivity of people in a dry working environment should be performed.

The third scenario is represented by the change in the variable "Timeshift". This variable represents the different periods of time during the day, when certain changes in the indoor air quality controllers might occur. The validity of this assumption is tested in this scenario, and represented in Table 30. The networks generated for each time period can be found in Appendix

5: Credo Room – BBN Scenarios. Furthermore, according to Liang, Hong, & Shen (2016), there are certain occupancy rates associated with each time periods.

This aspect is tested by the indicated scenario, by observing the level of CO_2 associated with each period. The CO_2 level is correlated to the amount of people present in the room, therefore an increase in CO_2 would indicate a bigger amount of people present in the room.

Table 30: Scenario 3 - levels based on time

	No Evidence	1 – 100% Evidence	2 - 100% Evidence	3 - 100% Evidence	4 - 100% Evidence	5 - 100% Evidence	6 - 100% Evidence
NewCO_2							
Healthy Level	47%	**49%**	30%	28%	29%	33%	**57%**
Acceptable Level	**48%**	46%	**62%**	**60%**	**61%**	**62%**	42%
Stiffness	4%	5%	8%	13%	10%	5%	1%
NewTemperature							
Comfortable1	**51%**	**55%**	**47%**	36%	31%	**45%**	**59%**
Comfortable2	38%	38%	44%	**49%**	**53%**	41%	33%
Uncomfortable_high	11%	7%	9%	15%	16%	14%	9%
NewHumidity							
Min_uncomfortable	**80%**	**78%**	**80%**	**84%**	**84%**	**82%**	**78%**
Optimal_Humidity	20%	22%	20%	16%	16%	18%	22%

1 = Going-to-work ; 2 = Morning ; 3 = Lunch ; 4 = Afternoon ; 5 = Going-home ; 6 = Night

For this scenario, the behavior of the CO_2 is the most important as it relates the most to the presence of people in the room. It can be observed that the healthy level increases during the going-to-work and night period. The added percentages for the acceptable and stiffness levels increase during the morning and lunch period compared to the going-to-work period, while in the afternoon and going-home period they start to decrease compared to the lunch period. When translating this into CO_2 emission this result symbolizes a level of up to 450 ppm during the night and in the going-to-work period, which can be compared to zero or low occupancy in the room. The increase in recorded CO_2 during the morning and lunch periods denotes presence during the morning, but not necessarily also during the lunch period. This happens because it takes some hours for the CO_2 level to decrease when the number of people in the room decreases, and the other way around. Furthermore, the decrease in percentages in the afternoon means a decrease in the number of people present in the room during the lunch period, while the decrease in the going-home period symbolizes the decrease in the number of people during the afternoon period. Overall, the behavior of the CO_2 emission suggests that employees come during the going-to-work and morning shifts in the room, a part of them leave for lunch. The lunch time is, however, not sufficient, for the level to drop to a healthy level (below 450 ppm), before the employees return in the afternoon. During the afternoon some employees leave earlier, therefore another decrease in the going-home period is observed. The night period is maintained at a healthy level, as there are few or no people in the room, who are still in the office in the going-home period or during the evening.

Conclusions Credo Room

The first space, from where historical data on CO_2, humidity and temperature was collected, is an office space situated in the Strijp-S area in Eindhoven, Netherlands. The purpose of this analysis was primarily to answer the second research question: "***How can future states of CO_2, humidity and temperature be predicted based on historical sensor data, for improving the indoor air quality?***". Several patterns were identified from the data collected during a period of five months, every five minutes.

Firstly, in regards to the months when data was collected, September resulted to be the month when an adequate healthy level of CO_2 emission was registered the most, an optimal humidity but also the highest temperature. November and December are the months with the biggest CO_2 emission, the most uncomfortable minimal humidity, but a comfortable temperature of Level 1 (between 19 and 22.9°C). This result seems valid, when compared to the seasons autumn and winter (for the increase and decrease of temperature), or the holiday and working period (for the increase or decrease in CO_2 suggesting presence or absence). At the same time, this result also suggests that the indoor parameters are strongly related to the room occupancy, and implicitly to the impact of the people on the indoor environment.

Secondly, with regards to the day of the week, Tuesday and Wednesday present the highest CO_2 emission, while Monday and Tuesday show the biggest percent when uncomfortable temperature is registered. The minimal uncomfortable humidity is found in most cases throughout the whole week. This result suggest that Tuesday is the day when the occupancy is the highest, however this result should be further validated after the analysis of more rooms and the comparison with the ground truth data.

Thirdly, the time of the day, is more helpful for predicting CO_2 emission, rather than any specific pattern in temperature and humidity. The biggest increase in CO_2 is observed during the lunch period which is the result of people being present in the room in the morning and partially lunch. The decrease of CO_2 during the afternoon period suggests some people are going out of the office during the lunch break, however, the decrease is not really significant to reach a healthy level inside the room. The patterns of occupancy resulted throughout the day, could serve as a good framework for formulating improvement opportunities for each time period.

Finally, the estimated Bayesian Belief Network also shows that the month has the strongest influence on the three variables, and therefore is the most useful for future predictions. The time period was the least helpful in predicting temperature and humidity patterns, and this was also visible in the strength of influence of the network. Also, the outside variables and the season have an impact on the indoor humidity variable, and should also be taken into consideration when establishing future patterns.

4.2.2. Office 2: Seats 2 Meet Conference Room

Descriptive Analyses

The descriptive analyses results for the S2M conference room, present the frequencies for each variable and the derived graphics. Firstly, Table 31 displays the 21777 valid cases for each of the three variables recorded in this space. The CO_2 variates between 282 ppm and 2000 ppm, the humidity is between 16.9% and 59.8%, and the temperature is between 16.5°C and 27.3° C.

Secondly, Figure 21 presents how the CO_2 data variates, with the majority of the cases being concentrated in the first part of the image, which shows a high frequency of over 250 cases for the low levels. In the same figure, a strange relatively high frequency is observed in the right side of the graph, with over 50 cases for the maximum value of 2000.

Thirdly, in Figure 22 the humidity levels are distributed uniformly, however the maximum amount of cases is around the humidity of 24%. Lastly, temperature is focused in the center of the graph, as shown in Figure 23 , around 22°C.

Table 31: Cases for historical data in S2M Conference Room

Statistics

		co2	humi	temp
N	Valid	21777	21777	21777
	Missing	0	0	0
Mean		501.69	30.539	22.291
Median		451.00	30.900	22.400
Std. Deviation		164.057	6.5634	1.1624
Minimum		282	16.9	16.5
Maximum		2000	59.8	27.3

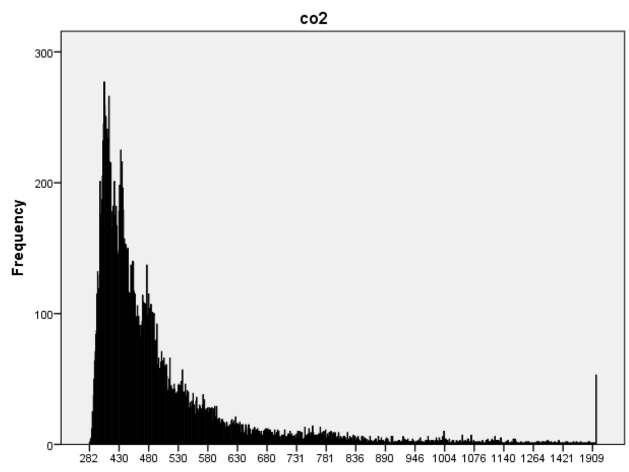
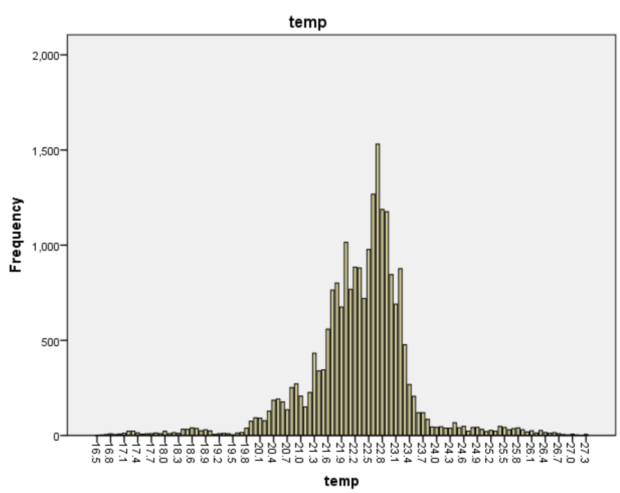


Figure 21: CO2 levels (ppm) S2M Conference Room



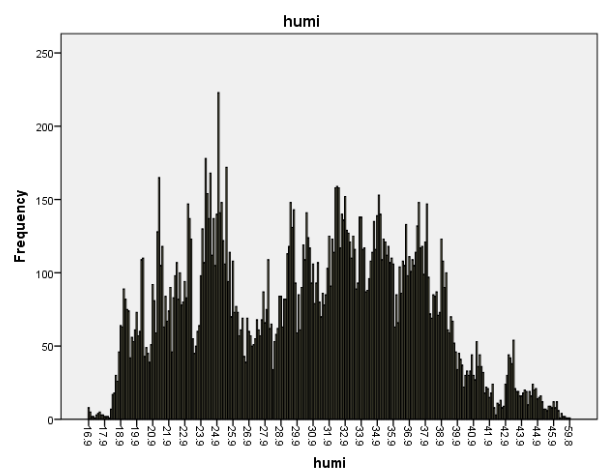


Figure 22: Humidity levels (%) S2M Conference Room

Figure 23: Temperature levels (°C) S2M Conference Room

Moving further, similarly to the previous room (the Credo Room), the CO_2 data is recoded in SPSS into the variable 'NewCO2', in five different levels. Apart from the three levels identified in the previous room (healthy, acceptable and stiffness), this room also presents a minimum level (below 349 ppm) and a drowsiness/bad air level (over 1001 ppm). Table 32 indicates only 1 case of minimum level, while the biggest frequency is still given by the healthy level (10819 cases), followed by the acceptable level (9351 cases) and finally by the stiffness (1128) and drowsiness (478).

Table 32: S2M - New CO2 levels after recoding in SPSS

NewCO2	Frequency	Valid Percent
Minimum Level	1	0
Healthy Level	10819	49.7
Acceptable Level	9351	42.9
Stiffness/odors	1128	5.2
Drowsiness/Bad air	478	2.2
Total	21777	100

The levels of humidity were re-coded in the variable 'NewHumidity', as emphasized in section 3.5. Data Processing. No cases above 60% were registered. Table 33 shows that the uncomfortable level is the highest, with 20381 cases, while the optimal humidity is given by 6.4% of the cases (1396 cases).

Table 33: S2M - New humidity levels after recoding in SPSS

NewHumidity	Frequency	Valid Percent
Minimal uncomfortable humidity	20381	93.6
Optimal Humidity	1396	6.4
Total	21777	100

Finally, the temperature is re-coded similarly to the previous room in the variable 'NewTemperature', with the four levels described in chapter *3.5*. Level 1 is found in the majority of the cases (16717), level 2 is found for 4490 cases, while the uncomfortable levels are found in 408 (low) and 162 (high) of the cases.

Table 34: S2M - New temperature levels after recoding in SPSS

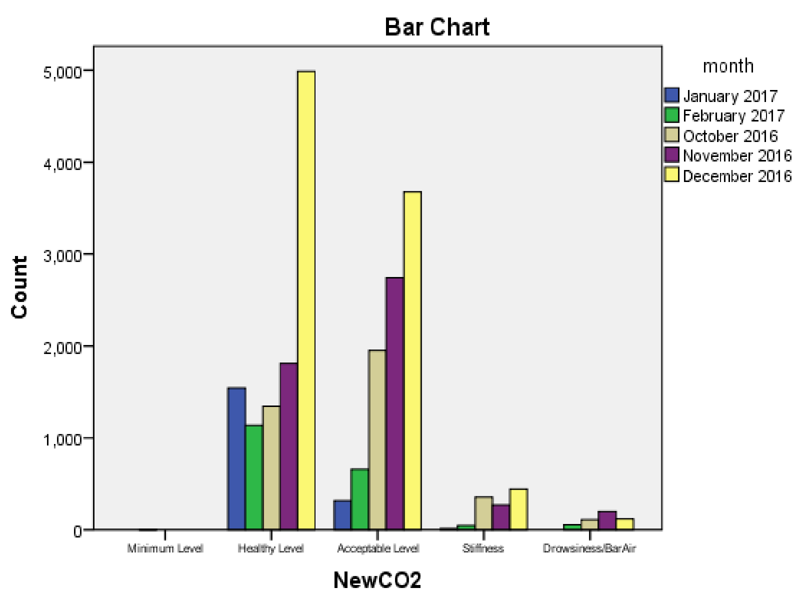
NewTemperature	Frequency	Valid Percent
Uncomfortable low	408	1.9
Level 1	16717	76.8
Level 2	4490	20.6
Uncomfortable High	162	0.7

Total	21777	100

For answering the second research question, the historical data collected for this space is presented by month, weekday and time period, in matrix formats, and in the associated bar charts. The following analysis complements the results that will be further presented in the Bayesian Belief Network scenarios.

Firstly the CO_2 is presented compared to the months when the data was collected. Table 35 shows that the minimum level is present in just 1 case, therefore, it will not be considered in the analysis, or in any further analysis, as it probably refers to an error registered by the sensor. The healthy level is present mostly in December (46.1% of all the months), and the same applies for the Acceptable Level (39.3%). In the latter case November is also present in a big percentage (29.4%) or October (20.9%).

The stiffness level is found the most in December (39.2%), followed by October (31.6%) and November (24%). The drowsiness level is predominant in November (42.1%), followed by December (24.2%) and October (22.6%).



The months January and February are not really predominant, due to the data outage registered then. However, when looking at those month individually, the biggest percentage in the healthy level is present, 82.3% - January, 60% - February. Figure 24 complements the indicated percentages, by showing the frequency of each level by month.

Figure 24: S2M - Bar chart: NewCO2 with the months

Generally, out of these results we cannot conclude that one of the months has the biggest emission of CO_2. All three months (December, November and October) are very high in the low CO_2 emission (healthy level) but also in the higher emission (acceptable and stiffness). The drowsiness level has, however, a high presence in November (42.1%), therefore, this month can be considered the most affected by poor indoor air quality.

Table 35: S2M - Crosstab NewCO2 with the month

		Month				
		January 2017	February 2017	October 2016	November 2016	December 2016
Minimum Level	%within NewCO2	0%	100%	0%	0%	0%
	%within Month	0%	**0.1**%	0%	0%	0%
	% of Total	0%	0.1%	0%	0%	0%
Healthy Level	%within NewCO2	14. 3%	10.5%	12.4%	16.7%	**46.1%**
	%within Month	**82.3%**	**60%**	35.7%	36%	**54.1%**
	% of Total	7.1%	5.2%	6.2%	8.3%	22.9%
Acceptable Level	%within NewCO2	3.4%	7%	**20.9%**	**29.4%**	**39.3%**
	%within Month	17%	34. 8%	**52%**	**54.6%**	39.9%
	% of Total	1.5%	3%	9%	12.6%	16.9%
Stiffness/odors	%within NewCO2	1.2%	4%	**31.6%**	**24%**	**39.2%**
	%within Month	0.7%	2.4%	9.5%	5.4%	4.8%
	% of Total	0.1%	0.2%	1.6%	1.2%	2%
Drowsiness/Bad air	%within NewCO2	0%	11.1%	**22.6%**	**42.1%**	**24. 3%**
	%within Month	0%	2.8%	2.9%	4%	1.3%
	% of Total	0%	0.2%	0.5%	0.9%	0.5%
Pearson Chi-Square						
Df = 16			Asymp. Sig. (2 sided)= 0.000			

The CO_2 is also illustrated according to the days of the week, for emphasizing the differences occurring between the four re-coded levels. The minimum level is also not analyzed in this case.

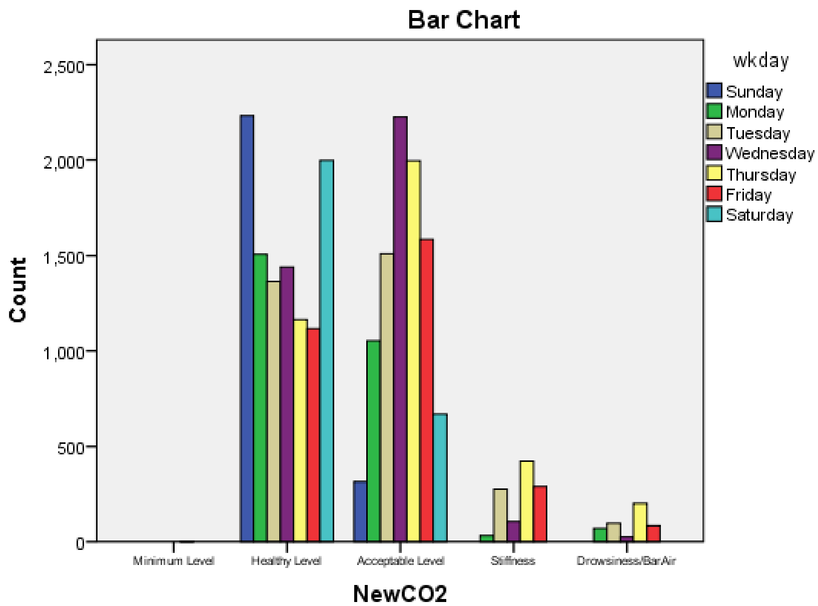


Figure 25: S2M - Bar chart: NewCO2 with the week days

As illustrated in Figure 25, the healthy level is present mostly during the weekend. If only the working days are analyzed, Monday has the highest healthy level, followed by Wednesday and Tuesday. The acceptable level is present the most on Wednesday, Thursday and Friday, while the stiffness is mostly seen on Thursday and Friday. The drowsiness level is also registered the most on Thursday followed by Tuesday and Friday. Table 36 also shows per day the percentages for each CO_2 level. Monday has the biggest percentage in healthy level (56.5%), while Tuesday presents a good balance between the acceptable (46.5%) and healthy level (42%). All the other working days have the highest percent in acceptable level.

Table 36: S2M - Crosstab NewCO2 with the week day

		WeekDay						
		Sun	Mon	Tue	Wed	Thu	Fri	Sat
Minimum Level	%within NewCO2	0%	0%	0%	0%	100%	0%	0%
	%within Week Day	0%	0%	0%	0%	0%	0%	0%
	% of Total	0%	0%	0%	0%	0%	0%	0%
Healthy Level	%within NewCO2	**20.6%**	**13.9%**	**12.6%**	**13.3%**	10.8%	10.3%	**18.5%**
	%within Week Day	**87.6%**	**56.5%**	**42%**	37.9%	30.8%	36.3%	**74.9%**
	% of Total	10.2%	6.9%	6.3%	6.6%	5.3%	5.1%	9.2%
Acceptable Level	%within NewCO2	3.4%	11.3%	16.1%	**23.8%**	**21.3%**	**16.9%**	7.1%
	%within Week Day	12.4%	39.5%	**46.5%**	58.6%	52.7%	51.5%	25.1%
	% of Total	1.4%	4.8%	6.9%	10.2%	9.2%	7.3%	3.1%
Stiffness/odors	%within NewCO2	0%	2.9%	24.6%	9.3%	**37.4%**	**25.8%**	0%

	%within Week Day	0%	1.2%	8.5%	2.8%	11.2%	9.5%	0%
	% of Total	0%	0.2%	1.3%	0.5%	1.9%	1.3%	0%
Drowsiness/Bad air	%within NewCO2	0%	14.9%	**20.1%**	5.4%	**42.1%**	**17.6%**	0%
	%within Week Day	0%	2.7%	3%	0.7%	5.3%	2.7%	0%
	% of Total	0%	0.3%	0.4%	0.1%	0.9%	0.4%	0%
Pearson Chi-Square								
Df = 24			Asymp. Sig. (2 sided)= 0.000					

From this analysis, on one hand, we can conclude that apart from the weekends (when the lowest CO_2 emission is registered), Monday is also predominantly healthy and therefore the least occupied day of the week. On the other hand, because there are big percentages of acceptable, stiffness and drowsiness levels, Thursday is the most populated day and therefore, the day with the biggest CO_2 emission. Friday also presents relatively high levels of emitted CO_2, therefore it should also be taken into account for future adjustments of the climate control system.

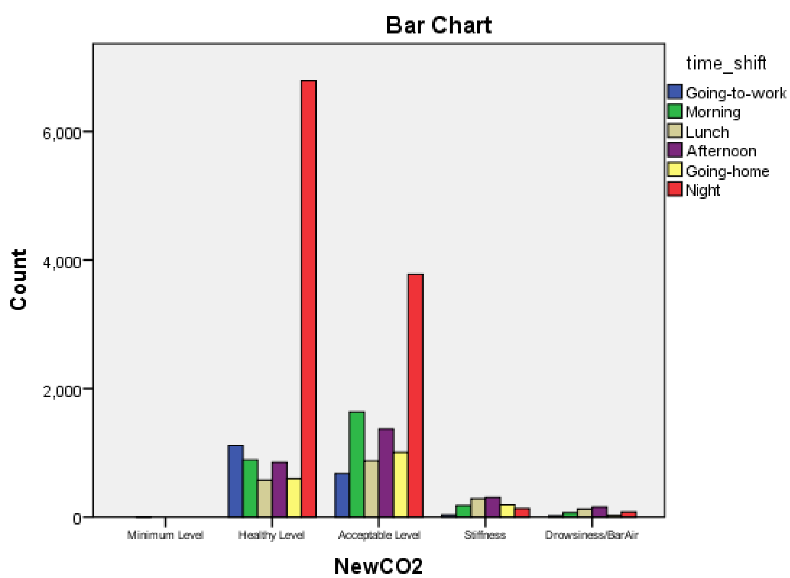


Figure 26: S2M - Bar chart: NewCO2 with time

When looking at how each CO_2 level changes according to the time period, from Figure 26 we can conclude that the healthy and acceptable levels (up to 700 ppm) are found mostly during the night period. Similarly to the previous room, the number of cases registered is high during the night (along 12 hours), while all the other periods show only the results for a maximum of 3 hours, hence there is a big difference in the percentages illustrated during the night compared to other time periods. However, the high number of cases within the healthy and acceptable levels during the night, can indicate that a normal level is maintained, due to the absence of people in that period. If we exclude the night cases, the healthy level shows the biggest percentage within this level during the going-to-work period (10.3%), followed by the morning (8.2%), as indicated in Table 37. The same table also indicates that the acceptable level is found in 17.5% of the cases in the morning period, followed by 14.7% of the cases in the afternoon. The stiffness is mostly found in the afternoon (27.2%) and lunch period (25.4%). Finally, the drowsiness is present in percentage of 32% during the afternoon and 25.5% during lunch. When looking at each time period individually, during the night and in the going-to-work period a healthy level is predominant (63% and 60.1%), while during all the other periods an acceptable level is the main level encountered.

From this results, it can be concluded that there are lower emissions than 450 ppm present mostly during the night and in the going-to-work period, while in the morning, the level is up to 700 ppm (partially healthy, partially acceptable). The acceptable level is predominant in the other periods, however, the stiffness and drowsiness periods are mostly found during the lunch period and afternoon, therefore a corrective action should be taken during these periods, for reducing the CO_2 and preventing the appearance of, especially, the drowsiness level.

Table 37: S2M - Crosstab NewCO2 with time

		Timeshift					
		Going-to-work	Morning	Lunch	Afternoon	Going-home	Night
Minimum Level	%within NewCO2	0%	100%	0%	0.0%	0.0%	0.0%
	%within Timeshift	0%	0%	0%	0.0%	0.0%	0.0%
	% of Total	0%	0%	0%	0.0%	0.0%	0.0%
Healthy Level	%within NewCO2	**10.3%**	**8.2%**	5.3%	7.9%	5.5%	**62.8%**
	%within Timeshift	**60.1%**	32.1%	30.9%	31.7%	32.7%	**63.0%**
	% of Total	5.1%	4.1%	2.6%	3.9%	2.7%	31.2%
Acceptable Level	%within NewCO2	7.3%	**17.5%**	9.4%	**14.7%**	10.8%	**40.4%**
	%within Timeshift	36.7%	58.9%	47.2%	51.2%	55.4%	35.0%
	% of Total	3.1%	7.5%	4%	6.3%	4.6%	17.3%
Stiffness/odors	%within NewCO2	3.3%	15.7%	**25.4%**	**27.2%**	16.8%	11.7%
	%within Timeshift	2%	6.4%	15.4%	11.4%	10.4%	1.2%
	% of Total	0.2%	0.8%	1.3%	1.4%	0.9%	0.6%
Drowsiness/Bad Air	%within NewCO2	4.8%	14.6%	**25.5%**	**32.0%**	5.9%	17.2%
	%within Timeshift	1.2%	2.5%	6.6%	5.7%	1.5%	0.8%
	% of Total	0.1%	0.3%	0.6%	0.7%	0.1%	0.4%
Pearson Chi-Square							
Df = 20				Asymp. Sig. (2 sided)= 0.000			

The humidity is analyzed compared to the months, days and time similarly to the CO_2. Figure 27 and Table 38 show that the minimal uncomfortable humidity is mostly present in December (44.3% within the 'NewHumidity' and 97.8% within the 'Months'). The optimal humidity is mostly present in October (61.6% within 'NewHumidity' and 22.9% within the Months). Starting with October when there is the most optimal humidity, November presents a decrease in optimal humidity and increase in uncomfortable humidity, while December shows the biggest percent in uncomfortable level. January and February show a percentage of 100% uncomfortable humidity within the month variable, therefore no optimal humidity was registered in those periods. This can be, however, due to the data outage registered in this period, when a lower number of cases was registered. Similarly to the previous room, the winter months present mostly an uncomfortable humidity, compared to the autumn months (only October and partially November in this case).

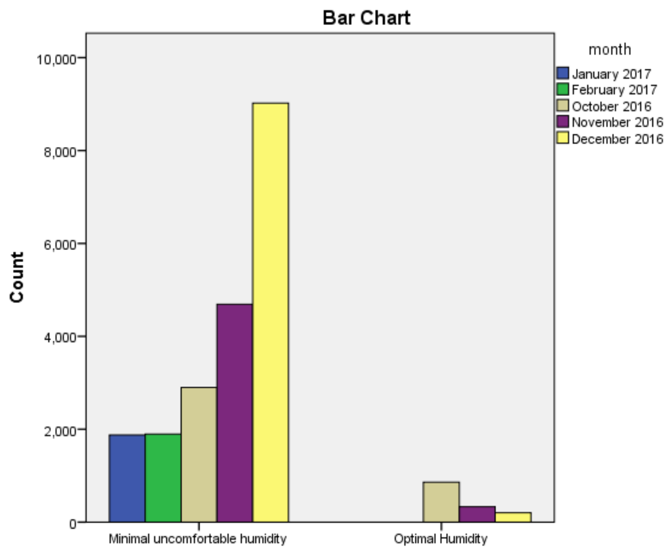


Figure 27: S2M - Bar chart: NewHumidity with the months

Table 38:S2M - Crosstab NewHumidity with the months

		Month				
		January 2017	February 2017	October 2016	November 2016	December 2016
Minimal uncomfortable humidity	%within NewHumidity	9.2%	9.3%	14.2%	23.0%	**44.3%**
	%within Month	**100.0%**	**100.0%**	77.1%	93.4%	**97.8%**
	% of Total	8.6%	8.7%	13.3%	21.5%	41.4%
Optimal humidity	%within NewHumidity	0.0%	0.0%	**61.6%**	23.9%	14.5%
	%within Month	0.0%	0.0%	**22.9%**	6.6%	2.2%
	% of Total	0.0%	0.0%	3.9%	1.5%	0.9%
Pearson Chi-Square						
Df = 4			Asymp. Sig. (2 sided)= 0.000			

Following, for the same indoor controller, the different levels are displayed according to the days of the week. The minimal uncomfortable humidity is present mostly on Wednesday (18.1%), followed by Thursday (17.2%), however, the percentages of uncomfortable humidity are well distributed along the week days. The cases of optimal humidity, are present on Tuesday (33.2%), Thursday (20.6%) and Friday (19.2%). When looking at each day, the whole week illustrates primarily an uncomfortable level present in the room, similar to the previous room.

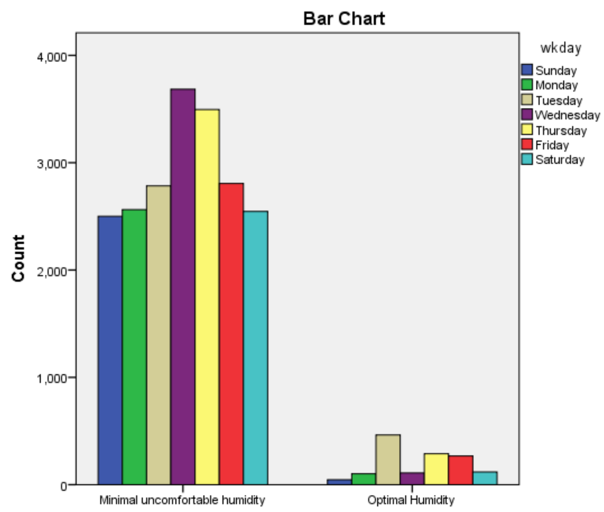


Figure 28: S2M - Bar chart: NewHumidity with the weeks

This analysis suggests a similar conclusion to the conclusion drawn for the previous room (Credo Room): both levels are uniformly distributed among the days, while all of the days present a predominant uncomfortable level. Therefore, a correction should be made for all the days of the week for maintaining an adequate humidity level, not for one particular day.

The analysis of the humidity per day does not play an important role in predicting future patterns. The mentioned results can be found in Figure 28 and Table 39.

Table 39: S2M - Crosstab NewHumidity with the weeks

		Weekday						
		Sun	Mon	Tue	Wed	Thu	Fri	Sat
Minimal uncomfortable humidity	%within NewHumidity	12.3%	12.6%	13.7%	**18.1%**	**17.2%**	13.8%	12.5%
	%within WeekDay	98.2%	96.2%	85.7%	97.1%	92.4%	91.3%	95.5%
	% of Total	11.5%	11.8%	12.8%	16.9%	16.1%	12.9%	11.7%
Optimal humidity	%within NewHumidity	3.3%	7.3%	**33.2%**	7.9%	**20.6%**	**19.2%**	8.5%
	%within WeekDay	1.8%	3.8%	14.3%	2.9%	7.6%	8.7%	4.5%
	% of Total	0.2%	0.5%	2.1%	0.5%	1.3%	1.2%	0.5%
Pearson Chi-Square								
Df = 6				Asymp. Sig. (2 sided)= 0.000				

Lastly, this variable is compared to the time shifts and the results are illustrated in Figure 29. We can see a similar result to the CO_2 in both cases of humidity (minimal uncomfortable humidity and optimal humidity), which are the highest during the night, period represented by the highest number of cases. Excluding the night cases, both the uncomfortable humidity and the optimal one are found mostly during the morning and afternoon. However, during all the periods there is predominantly uncomfortable humidity in the room.

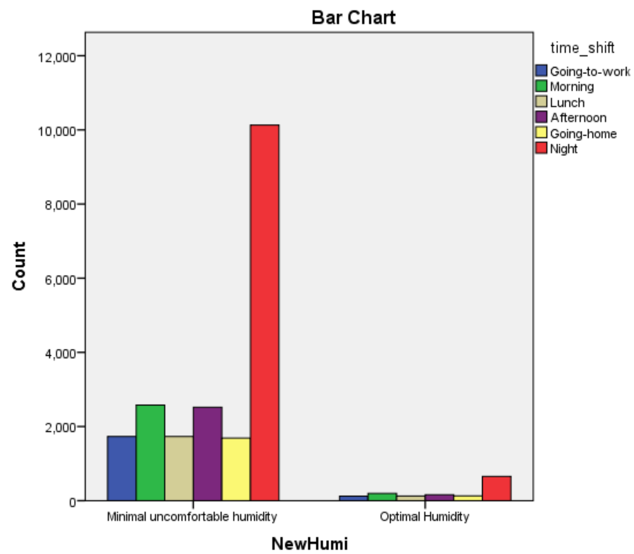


Figure 29: S2M - Bar chart: NewHumidity with time

Table 40 actually shows that the minimal uncomfortable humidity is present in more than 90% for all the time periods. This is an obvious result, as in the analysis according to the week days and months, we could also observe predominantly uncomfortable humidity.

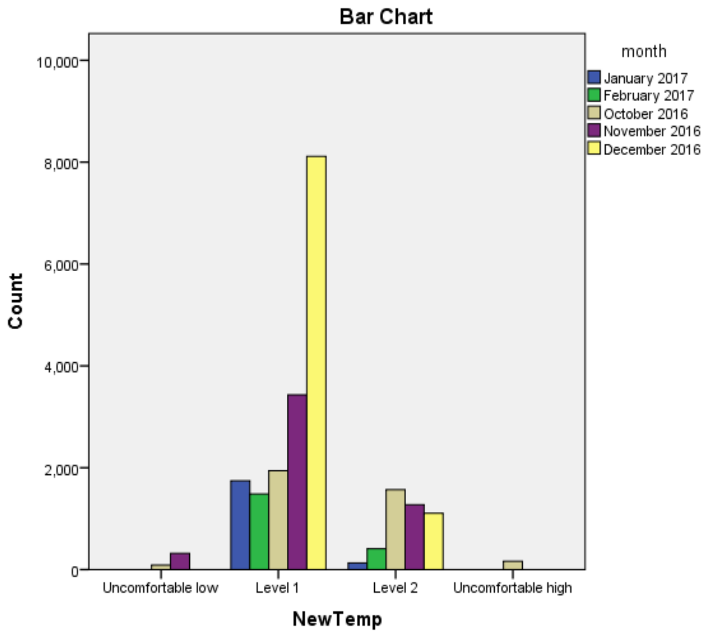
Based on these results, the time period also does not play an important role in determining humidity patterns. Furthermore, the Chi-Square test shows a significance level of 0.192.

This means that there is no statistical significant relationship between the two variables. We can conclude that the humidity depends more on the season and the month that on the time of the day or the amount of people present in the room.

Table 40: S2M - Crosstab NewHumidity with the time

		Timeshift					
		Going-to-work	Morning	Lunch	Afternoon	Going-home	Night
Minimal uncomfortable humidity	%within NewHumidity	8.5%	**12.6%**	8.5%	**12.4%**	8.3%	**49.7%**
	%within Time Shift	**93.4%**	**92.9%**	**93.3%**	**93.9%**	**92.8%**	**93.9%**
	% of Total	7.9%	11.8%	8.0%	11.6%	7.8%	46.5%
Optimal humidity	%within NewHumidity	8.7%	**14.2%**	9.0%	**11.7%**	9.5%	**47.0%**
	%within Time Shift	6.6%	7.1%	6.7%	6.1%	7.2%	6.1%
	% of Total	0.6%	0.9%	0.6%	0.7%	0.6%	3.0%
Pearson Chi-Square							

Df = 5	Asymp. Sig. (2 sided)= 0.192



The temperature is analyzed based on the same factors: the months when the data was collected, the days of the week and the time period. Firstly, the indoor temperature is illustrated according to the months. Figure 30 presents the temperature according to the months, where the uncomfortable low temperature is generally found in November, level 1 in December, while level 2 and the uncomfortable high are seen primarily in October.

Figure 30: S2M - Bar chart: NewTemperature with months

When looking along the months, Table 41 shows that all the months have typically level 1 temperature, followed by level 2, which means that an adequate working temperature is maintained in the room for all the periods. These results suggest that a higher temperature is found in the room in October, while for the other months it decreases and it is relatively kept at a comfortable level. Similarly to the previously analyzed room, the higher temperatures are found in the warmer months. Because in the current analysis there is data available only starting with October, this is the only month that can be considered as warmer than the winter months (November – February). Therefore, the temperature inside the room will be influenced primarily by the season in which the data is collected, and it will be adjusted for a comfortable level for working by the employees. Table 41 shows the percentages for each of the indicated months, and supplements the mentioned results.

Table 41: S2M - Crosstab NewTemperature with the months

		Month				
		January 2017	September 2016	October 2016	November 2016	December 2016
Uncomfortable low	%within NewTemperature	0.0%	0.0%	21.6%	**78.4%**	0.0%
	%within Month	0.0%	0.0%	2.3%	6.4%	0.0%
	% of Total	0.0%	0.0%	0.4%	1.5%	0.0%

Level 1	%within NewTemperature	10.4%	8.9%	11.6%	20.5%	**48.5%**
	%within Month	**93.0%**	**78.4%**	**51.6%**	**68.3%**	**88.0%**
	% of Total	8.0%	6.8%	8.9%	15.8%	37.3%
Level 2	%within NewTemperature	2.9%	9.1%	**34.9%**	28.4%	24.7%
	%within Month	7.0%	21.6%	41.7%	25.3%	12.0%
	% of Total	0.6%	1.9%	7.2%	5.8%	5.1%
Uncomfortable High	%within NewTemperature	0.0%	0.0%	**100.0%**	0.0%	0.0%
	%within Month	0.0%	0.0%	4.3%	0.0%	0.0%
	% of Total	0.0%	0.0%	0.7%	0.0%	0.0%
Pearson Chi-Square						
Df = 12			Asymp. Sig. (2 sided)= 0.000			

Indoor temperature is also analyzed based on the days, and Figure 31 presents the behavior of the temperature levels accordingly. According to Table 42, the uncomfortable low temperature is present on Wednesday (67.6%) and level 1 is uniformly distributed among the week days with the biggest amount of cases registered on Thursday (17.6%). Similarly, level 2 has a uniform distribution along the days, however Wednesday presents the highest percentage (21%), while the uncomfortable high temperature is mostly found on Monday (40.7%).

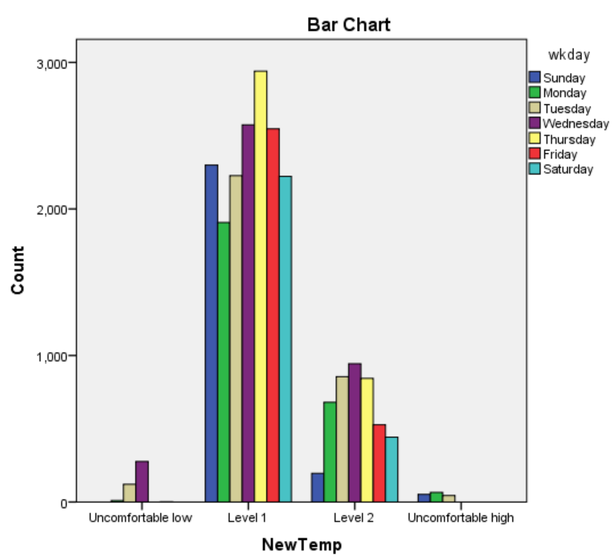


Figure 31: S2M - Bar chart: NewTemperature with week days

When looking at each week day, level 1 is again found for all of the days, in majority of the cases (more than 67% for all of the days). By comparing these results to the ones indicated in Credo Room, we see a difference in how the temperature is maintained along the days. In S2M conference room we observe lower temperatures (more level 1) during the weekends, Thursday and Friday, while in the Credo Room the analysis indicates higher temperatures (level 2 and uncomfortable high) in the same period.

When looking at the CO_2 emission for the S2M room, we observe an increase on Thursday and Friday. Therefore keeping the temperature at a lower level, can be connected to the amount of people in the room, and the desire to keep an adequate working temperature for the indicated presence. The CO_2 emission for this period does not suggest a lot of presence, during the weekends, therefore the low temperature cannot be justified by the occupancy, in this case. The result, might suggest that either the HVAC system is overused during the weekends (for keeping a comfortable level), either that level 1 is the normal temperature registered inside the S2M room, when no action is monitored.

Table 42: S2M - Crosstab NewTemperature with the days

		WeekDay						
		Sun	Mon	Tue	Wed	Thu	Fri	Sat
Uncomfortable low	%within NewTemperature	0.0%	2.5%	29.7%	**67.6%**	0.0%	0.2%	0.0%
	%within WeekDay	0.0%	0.4%	3.7%	7.3%	0.0%	0.0%	0.0%
	% of Total	0.0%	0.0%	0.6%	1.3%	0.0%	0.0%	0.0%
Level 1	%within NewTemperature	13.8%	11.4%	13.3%	15.4%	**17.6%**	**15.2%**	13.3%
	%within WeekDay	**90.3%**	**71.6%**	**68.6%**	**67.8%**	**77.7%**	**82.8%**	**83.4%**
	% of Total	10.6%	8.8%	10.2%	11.8%	13.5%	11.7%	10.2%
Level 2	%within NewTemperature	4.4%	15.2%	19.0%	**21.0%**	18.8%	11.7%	9.9%
	%within WeekDay	7.7%	25.6%	26.3%	24.9%	22.3%	17.1%	16.6%
	% of Total	0.9%	3.1%	3.9%	4.3%	3.9%	2.4%	2.0%
Uncomfortable High	%within NewTemperature	32.1%	**40.7%**	27.2%	0.0%	0.0%	0.0%	0.0%
	%within WeekDay	2.0%	2.5%	1.4%	0.0%	0.0%	0.0%	0.0%
	% of Total	0.2%	0.3%	0.2%	0.0%	0.0%	0.0%	0.0%
Pearson Chi-Square								
Df = 18			Asymp. Sig. (2 sided)= 0.000					

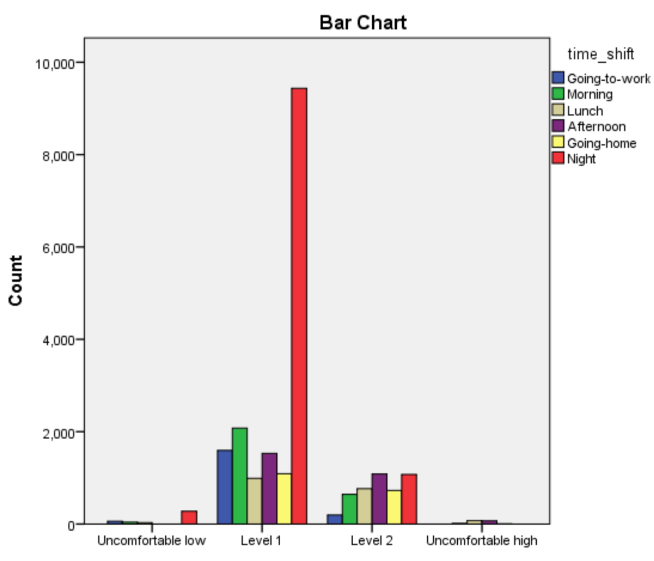


Figure 32: S2M - Bar chart: NewTemperature with the time

The temperature is ultimately compared to the time shifts and presented in Figure 32, while the associated percentages can be visualized in Table 43. According to these figures, apart from the night cases (which are represented the most in the levels uncomfortable low, level 1 and 2), the uncomfortable low level is distributed primarily, in the going-to-work and morning period. Level 1 is present in the morning and going-to-work period, as well as afternoon period. Level 2 is found the most during afternoon and lunch, while the uncomfortable level is similarly found the most during lunch and afternoon. When looking at all the time periods individually, all the periods show a predominant temperature Level 1. If the CO_2 levels are compared with the temperature, the afternoon and lunch periods are the periods when the most stiffness and drowsiness is registered, as well as the biggest temperature. Therefore, these two periods suggest a greater occupancy and an associated higher temperature caused by the increased occupancy.

Table 43: S2M - Crosstab NewTemperature with the time

		Time Shift					
		Going-to-work	Morning	Lunch	Afternoon	Going-home	Night
Uncomfortable low	%within NewTemperature	**14.7%**	**10.3%**	7.4%	0.0%	0.0%	67.6%
	%within TimeShift	3.2%	1.5%	1.6%	0.0%	0.0%	2.6%
	% of Total	0.3%	0.2%	0.1%	0.0%	0.0%	1.3%
Level 1	%within NewTemperature	**9.5%**	**12.4%**	5.9%	**9.2%**	6.5%	56.4%
	%within TimeShift	**86.1%**	**74.8%**	**53.3%**	**57.0%**	**59.8%**	**87.5%**
	% of Total	7.3%	9.5%	4.5%	7.0%	5.0%	43.3%
Level 2	%within NewTemperature	4.4%	14.3%	**17.1%**	**24.1%**	16.2%	23.9%
	%within TimeShift	10.6%	23.1%	41.2%	40.4%	40.0%	10.0%
	% of Total	0.9%	2.9%	3.5%	5.0%	3.3%	4.9%
Uncomfortable High	%within NewTemperature	0.0%	9.9%	**44.4%**	**43.2%**	2.5%	0.0%
	%within TimeShift	0.0%	0.6%	3.9%	2.6%	0.2%	0.0%
	% of Total	0.0%	0.1%	0.3%	0.3%	0.0%	0.0%
Pearson Chi-Square							
Df = 15			Asymp. Sig. (2 sided)= 0.000				

Bayesian Belief Network Results

For generating the results, two Bayesian Belief Networks are estimated and different scenarios are analyzed for the first network, similarly to the previous room. The same learning algorithm is applied: the Greedy Thick Technique. Also, variables are assigned to temporal tiers in the same way:

1. 'Timeshift', 'Weekday', 'Month', '*NewOutsideHumidity', 'NewOutsideTemperature*'
2. 'FinalTemperature', 'FinalHumidity', 'FinalCO_2'

Two networks are estimated, the first one illustrates only the main variables (CO_2, humidity and temperature), while the second one also presents the outside parameters (temperature and humidity). These networks can be found in Figure 33 and Figure 34, where also the strength of influence for each arc is presented. Further, the strength of influence is discussed for the two networks.

Apart from temporal tiers, the restriction between variables is also assigned in the same manner as for the Credo Room. For this network, arcs were forbidden between 'Timeshift', 'Weekday' and 'Month', in the following combinations: 'Weekday' -> 'Month', Timeshift' -> 'Month', 'Timeshift' -> 'Weekday', 'Month' -> 'Weekday'. These arcs were forbidden as the influence between the mentioned variables is beyond the scope of the research, so the impact on each other is not interesting for answering the research questions. For the second network, the impact of the 'Month', 'Weekday' and 'Timeshift' variables on both the 'NewOutsideTemperature' and 'NewOutsideHumidity' as well as the reserved were restricted, based on the same reasoning.

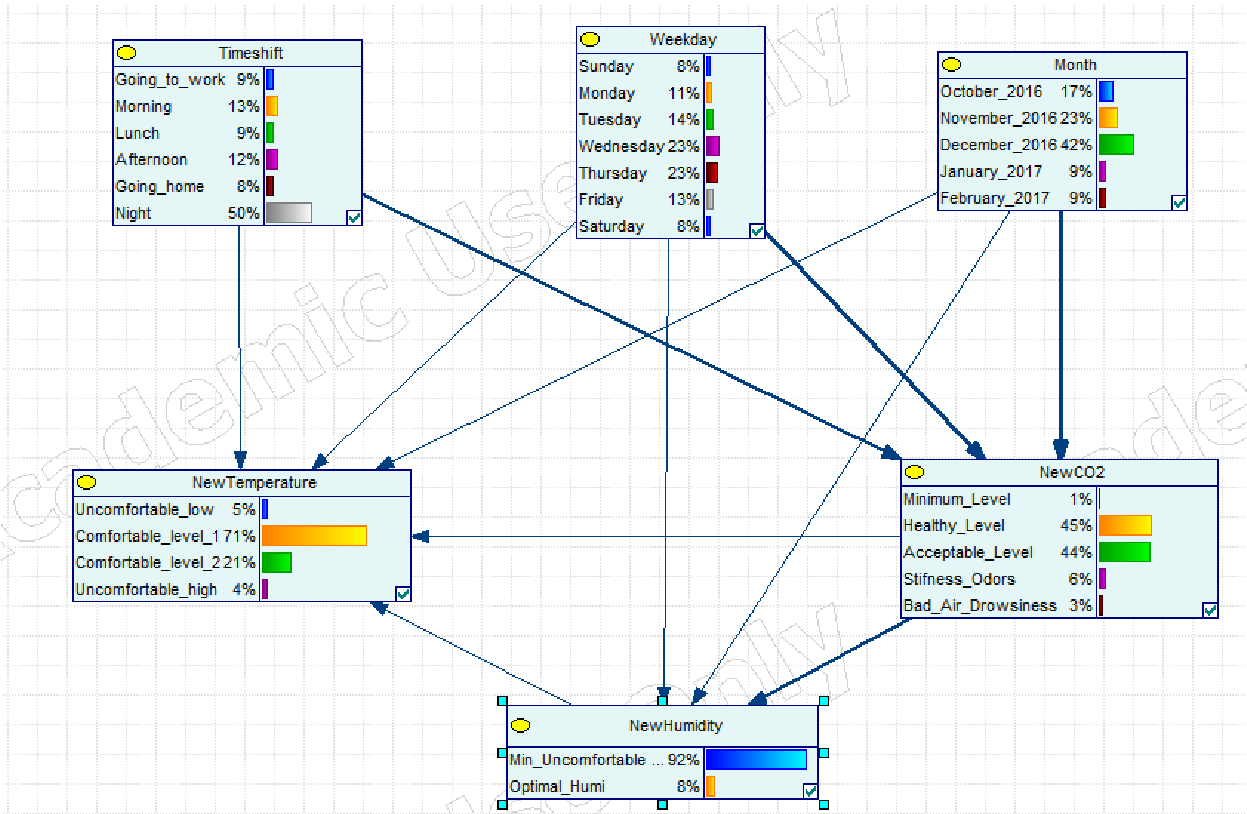


Figure 33: Original Bayesian Belief Network– S2M historical data

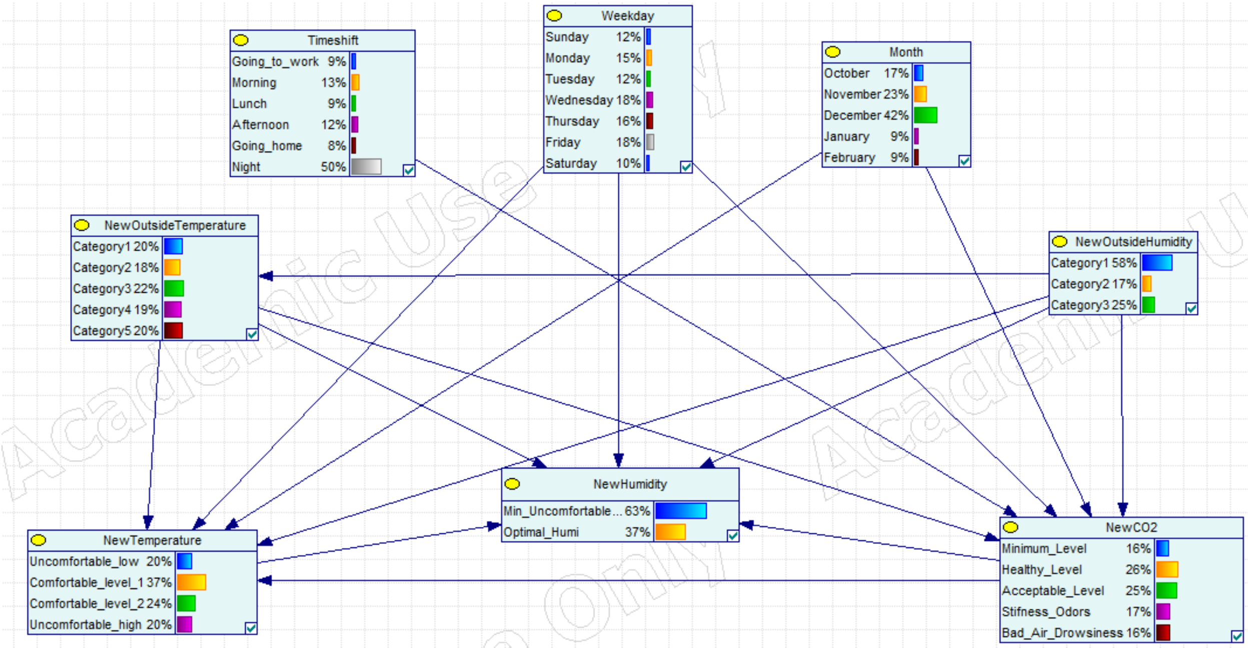


Figure 34: Bayesian Belief Network with outdoor variables – S2M historical data

Strength of Influence

The strength of influence shows the strength of relationship between the variables, included in the Bayesian Belief Networks. In the same way as for the previous room, the strength is calculated based on the CPT of the child node and is illustrated in Table 44 and Table 45.

Table 44 shows that the CO_2 variable is the most affected by the parent nodes (week day, month and finally time shift), the humidity is the second in the ranking and lastly comes the temperature. Compared to the previous room, where a strong influence of the month is observed, in this case, the month has a strong influence only on the CO_2 levels. Similarly to the previous case, the time still has a low influence, by having the lowest strength of influence on the temperature, and furthermore there is no arc between the variables 'Timeshift' and the 'NewHumidity'. These results suggest that the CO_2 is the variable that will show the biggest changes, and therefore the most accurate patterns will be generated based on the CO_2. Similarly, as suggested by the results for the Credo Room, the time period shows a low influence on most of the variables (with exception of the CO_2), therefore patterns based on the time period might not be entirely accurate.

Table 44: Strength of arcs for BBN without external variables – S2M Room

Origin (Parent Node)	Destination (Child Node)	Strength of influence
Weekday	NewCO2	0.3658
Month	NewCO2	0.338
NewCO2	NewHumidity	0.2536
TimeShift	NewCO2	0.2435

NewHumi	NewTemperature	0.1568
NewCO2	NewTemperature	0.150
Weekday	NewHumidity	0.1457
Month	NewHumidity	0.1417
Month	NewTemperature	0.1132
Weekday	NewTemperature	0.0992
TimeShift	NewTemperature	0.076

The biggest strength of influence estimated in the second BBN, as illustrated in Table 45, is between the two outdoor variables: humidity and temperature. If this relationship is excluded, as it is outside the scope of the current research, the month, outdoor humidity and week day present a big influence on the CO_2. This means that the CO_2 is the most affected child node. The following affected child nodes are the humidity and lastly the temperature. When looking at the influence of the outside humidity and temperature, the humidity seems to have a bigger influence on the indoor variables.

Table 45: Strength of arcs for BBN with external variables – S2M Room

Origin (Parent Node)	**Destination (Child Node)**	**Strength of influence**
NewOutsideHumidity	*NewOutsideTemperature*	*0.1691*
Month	NewCO2	0.1242
NewOutsideHumidity	NewCO2	0.1217
Weekday	NewCO2	0.1141
NewOutsideHumidity	NewHumidity	0.0683
NewTemperature	NewHumidity	0.067
NewCO2	NewHumidity	0.0611
Weekday	NewHumidity	0.059
NewOutsideTemperature	NewHumidity	0.0588
Month	NewTemperature	0.0557
NewOutsideTemperature	NewTemperature	0.0553
NewOutsideHumidity	NewTemperature	0.0538
Weekday	NewTemperature	0.05
NewCO2	NewTemperature	0.0441
Timeshift	NewCO2	0.04
NewOutsideTemperature	NewCO2	0.0229

Scenarios

For the network estimated and presented in Figure 33, the same scenarios as for the Credo Room are applied.

- Scenario 1: Evidence for the variable 'Month' – Monthly levels of CO_2 , humidity and temperature

- Scenario 2: Evidence for the variable 'Weekday' – Weekly levels of CO_2 , humidity and temperature
- Scenario 3: Evidence for the variable 'Timeshift' - Levels of CO_2 , humidity and temperature according to the time period

The first scenario illustrates evidence in the months of October, December and February. These months were selected as representative for the analysis, because the most important changes are noticed then. November presents similar results to October, and January similar to February, therefore only the evidence for October and February is considered. The evidence for the mentioned months is illustrated in Table 46.

Table 46: Scenario 1: Monthly levels S2M Room

Variable	No Evidence	October – 100% Evidence	December – 100% Evidence	February – 100% Evidence
NewCO$_2$				
Minimum Level	1%	1%	0%	3%
Healthy Level	**45%**	35%	**49%**	**50%**
Acceptable Level	44%	**50%**	43%	37%
Stiffness	6%	10%	6%	5%
Drowsiness	3%	4%	2%	6%
NewTemperature				
Uncomfortable_low	5%	6%	2%	5%
Comfortable1	**71%**	**49%**	**84%**	**60%**
Comfortable2	21%	37%	13%	30%
Uncomfortable_high	4%	8%	2%	5%
NewHumidity				
Min_uncomfortable	**92%**	**76%**	**97%**	**96%**
Optimal_Humidity	8%	24%	3%	4%

Based on the suggested evidence, the CO_2 healthy level is present mostly during February, however there is only 5% increase compared to the case without any evidence present. The acceptable level is predominant in October, with 50%, however, similarly to February, only a really small increase is observed (only 6%). At the same time, the stiffness and drowsiness levels increase during October, therefore this month presents the biggest CO_2 emission. Overall, the small differences in the original case and the cases with evidence, suggest that the evidence is not sufficient for making a coherent assumption on the behavior of the CO_2 level according to the months.

In what concerns the temperature, the evidence for all the months suggests a level 1 temperature in the room. October, however, presents a higher percentage in the comfortable level 2 and uncomfortable high. When comparing the CO_2 acceptable level predominating in this period and the level 2 temperature, it might be assumed that there is a relationship between the two. This relationship can be the reason of an increased occupancy during that month.

The humidity is maintained at an uncomfortable level for all the months. The only increase in optimal humidity is seen in October, which strengthens the conclusion made in the Credo room analysis, specifically that the season plays a role in the indoor humidity. More concretely, the warmer months will influence the increase of optimal humidity in the office.

The second scenario illustrates the weekly levels, by changing the percentages of each week day to 100% in the BBN. The given evidence and the behavior of each variable is given by Table 47. Each day of the week is included as a complete picture is needed for registering patterns during all the week days.

Table 47: Scenario 2: Weekly levels S2M Room

	No Evidence	**Mon – 100% Evidence**	**Tue – 100% Evidence**	**Wed – 100% Evidence**	**Thu – 100% Evidence**	**Fri – 100% Evidence**	**Sat – 100% Evidence**	**Sun – 100% Evidence**
NewCO_2								
Minimum Level	1%	1%	1%	1%	1%	1%	3%	1%
Healthy Level	**45%**	**54%**	**44%**	40%	31%	35%	**69%**	**84%**
Acceptable Level	44%	39%	43%	**55%**	**51%**	**49%**	24%	13%
Stiffness	6%	2%	8%	3%	11%	10%	3%	11%
Drowsiness	3%	3%	4%	1%	6%	4%	3%	1%
NewTemperature								
Uncomfortable_low	5%	3%	6%	7%	3%	4%	5%	3%
Comfortable1	**71%**	**67%**	**67%**	**67%**	**76%**	**76%**	**69%**	**81%**
Comfortable2	21%	25%	23%	24%	19%	17%	21%	12%
Uncomfortable_high	4%	5%	4%	2%	3%	4%	5%	5%
NewHumidity								
Min_uncomfortable	**92%**	**94%**	**87%**	**96%**	**92%**	**89%**	**89%**	**96%**
Optimal_Humi	8%	6%	13%	4%	8%	11%	11%	4%

The second scenario, illustrates an increased healthy level on Monday, and a higher CO_2 emission during Thursday and Friday (the acceptable, stiffness and drowsiness levels all increase). Tuesday presents similar levels of healthy and acceptable, while Wednesday is higher on acceptable level. The BBN suggests that an increased attention should be given to the last days of the week, Thursday and Friday, as those are the most inclined for unhealthy CO_2.

In terms of temperature, the evidence does not suggest big differences in the behavior of the four levels, with a comfortable level being maintained for all the weekdays. The humidity also does not show significant differences when the evidence is present, the uncomfortable humidity being present for the majority of the days. The biggest increase in optimal humidity is observed during Tuesday, which is the day when the least attention should be given to increasing the overall humidity present in the room.

The last scenario, scenario 3, presents evidence for the time periods. The results are presented in Table 48. The time periods are split in six categories similar to the previous room, according to the potential change of occupancy based on literature.

Table 48: Levels based on time S2M Room

	No Evidence	1 - 100% Evidence	2 - 100% Evidence	3 - 100% Evidence	4 - 100% Evidence	5 - 100% Evidence	6 - 100% Evidence
NewCO_2							
Minimum Level	1%	2%	1%	2%	2%	2%	0%
Healthy Level	45%	**55%**	27%	24%	24%	28%	**60%**
Acceptable Level	**44%**	37%	**61%**	**47%**	**52%**	**55%**	36%
Stiffness	6%	4%	7%	18%	14%	11%	2%
Drowsiness	3%	3%	3%	9%	8%	4%	1%
NewTemperature							
Uncomfortable_low	5%	7%	5%	7%	4%	5%	4%
Comfortable1	**71%**	**75%**	**69%**	**48%**	**51%**	**53%**	**84%**
Comfortable2	21%	13%	22%	38%	39%	36%	11%
Uncomfortable_high	4%	5%	4%	8%	6%	6%	1%
NewHumidity							
Min_uncomfortable	**92%**	**93%**	**92%**	**91%**	**90%**	**91%**	**93%**
Optimal_Humidity	8%	7%	8%	9%	10%	9%	7%

1 = Going-to-work ; 2 = Morning ; 3 = Lunch ; 4 = Afternoon ; 5 = Going-home ; 6 = Night

The last scenario presents evidence for each time period, and illustrates that the CO_2 has a similar behavior to the previously discussed room (the Credo Room). The evidence suggests increased percentages for the healthy level for the night and going-to-work periods. The morning period shows an increase in acceptable and stiffness level, compared to the going-to-work period, while the lunch period presents a further increase in the stiffness level and drowsiness level compared to the morning. The added percentages for the three levels denoting presence (acceptable, stiffness and drowsiness) in the lunch period is equal to the added percentages of the three levels during the afternoon. During afternoon, however, the stiffness and drowsiness levels decrease. The going-home period illustrates a decrease in $CO_{2,}$ illustrated by the decrease in stiffness and drowsiness levels, as well as the increase in the healthy and acceptable level.

Similarly to the previous room, the variation of CO_2 along the day can be compared to the room occupancy. An increase in CO_2 is marked by human activity that occurs slightly before the actual increased CO_2 emission is registered by the sensor. Therefore, the results suggest low or no presence during the night and going-to-work periods and significant presence during the morning. The lunch period illustrates an increase in stiffness and drowsiness compared to the morning, however this impact might be influenced by the increased amount of people and emitted CO_2 from the morning period. Therefore, it does not mean that more people were present during the lunch period compared to the morning period. The afternoon actually shows a decrease in stiffness and drowsiness, hence this is symbolized by a decline in occupancy during this period but also during the end of the lunch period. The similar levels registered during the afternoon and lunch might suggest that the same number of people was inside the room during these periods. A much bigger decrease is observed in the going-home period, therefore by the end of the afternoon period occupancy will drop, which impacts the going-home period. Finally, the healthy level present during the night indicates a constant drop in occupancy starting with the going-home period.

The temperature is maintained again at a comfortable level 1 for all the time periods. The only interesting situation observed is during the lunch period, when the comfortable 2 and uncomfortable high levels increase. This increase might be connected to the increase in stiffness and drowsiness during the same period, which indicates a higher occupancy.

The humidity does not present big changes along the time frames. This is an obvious result, because the humidity recorded is more than 90% below the optimal level (below 40%). Therefore, evidence in any time frame will still indicate a humidity below average.

Conclusions S2M Room

In this section the historical data for the conference room Seats2Meet collected over a period of five months was presented. The three indoor quality controllers were analyzed based on the month, day, time and weather conditions with the purpose of answering the first research question. The main conclusions will be indicated below.

Firstly, in regards to the CO_2, October and November seemed to be the most affected months by the increased emission, however no coherent difference or pattern was identified for the months when the data was collected. This is the case as the selected months have similar characteristics (similar outdoor conditions, similar working occupancy, not a lot of days off work etc.). Compared to the Credo Room, when a significant difference was observed between September and the rest of the months, no major differences were seen for the selected months for this room. In what concerns the week days, Thursday and Friday are the most affected by increased CO_2 emission and increased occupancy. The time periods illustrate a similar CO_2 emission behavior as for the previous room, where the levels increase up to the lunch period, and start decreasing in the afternoon. Due to the increased stiffness and drowsiness registered in the lunch and afternoon periods, a corrective action should be taken during the lunch period and at the end of the morning period, for avoiding potential decrease in the indoor air quality.

Secondly, it can be concluded that the humidity depends the most on the season and the months, with October illustrating the biggest humidity and January and February the lowest. No specific patterns were identified according to the time and the weeks days, which strengthens the conclusion previously suggested in the Credo Room, that the patterns for humidity can be mostly predicted according to the season.

Lastly, the indoor temperature is also influenced the most by the season and most of the patterns can be drawn according to the months. October illustrates the highest temperature, while the rest of the months when data is collected are maintained at the first comfortable working temperature. Thursday and Friday present the lowest temperatures during the week, but no big difference is observed in the variation along the days. Lastly, the lunch and afternoon period suggest also an increase in temperature, similarly to the increase of stiffness and drowsiness in the CO_2 levels. It can be concluded that the most important patterns in temperature are illustrated along the months, and likewise for the humidity, the days and time, do not play such an important role in the analysis for this room.

All in all, October is the month presenting the highest CO_2, highest humidity and temperature, Thursday and Friday show the highest CO_2, and the lowest temperature cases are registered. Lastly, the lunch and afternoon periods show the biggest stiffness and drowsiness in terms of CO_2 and the biggest amount of cases of high temperatures.

4.3. Experimental Data

As described in section 3.5. Data Processing, the experimental data includes data collected from the indoor quality controllers (CO_2, humidity and temperature) and from a PIR sensor. The data is collected over the period 16.04.2017-20.04.2017 and consists of over 1134 cases, each case showing the values for the quality controllers, every five minutes. The ground truth data, together with CO_2, humidity and temperature data is also collected for the period 22.03 – 07.04, for checking the suitability of the CO_2 sensors in regards to real-time occupancy detection. Subsequently, the collected data is presented and the correlations between the variables are explained. A BBN is estimated on the data collected during the experiment and evidence is set for answering the research questions.

4.3.1. Ground truth data analysis

Data for two weeks is collected, illustrating more than 4000 cases of humidity, temperature and CO_2. Table 49 shows the frequency data for all the parameters, including the ground truth data (the people count). The valid and missing number of cases are illustrated, as well statistics about the parameters, such as the minimum, maximum or the mean for each of the parameters. The table illustrates that there is a maximum of 4 people present throughout the whole period in the office, and a CO_2 variating between 391 ppm (parts per million) to 766 ppm, a temperature variating between 21 and 27° C and the humidity is between 17.3% to 42.5%.

Table 49: Frequency table for ground truth data

Statistics					
		Humidity	Temperature	CO2	PeopleCount
N	Valid	4178	4178	4178	4178
	Missing	250	250	250	250
Mean		29.252	24.07	473.04	.50
Median		28.600	24.00	439.00	.00
Std. Deviation		5.1169	1.256	75.852	.893
Minimum		17.3	21	391	0
Maximum		42.5	27	766	4

When looking at the ground truth data on its own (Table 50), it is observed that the most frequent category is represented by the category 'no people' with 2959 cases and a valid percent of 70.8. One person is found in the room in 595 of the cases (14.2%), two people are found in 442 of the cases (10.6%), three people are found for 130 of the cases (3.1%), while four people are present only in 52 of the cases (1.2%). The high number of cases in the 'no people' category, is related to the night cases when the data was collected, when there are no people present in the room, hence the 'no people' category is the highest.

Table 50: Ground truth data frequency

PeopleCount

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No People	2959	66.8	70.8	70.8
	1 person	595	13.4	14.2	85.1
	2 people	442	10.0	10.6	95.6
	3 people	130	2.9	3.1	98.8
	4 people	52	1.2	1.2	100.0
	Total	4178	94.4	100.0	
Missing	System	250	5.6		
Total		4428	100.0		

For understanding the impact of different variables on each other, in the ground truth data set, a Pearson correlation is applied. Pearson correlation is a technique which shows the linear relation between two quantitative, continuous variables. The Pearson coefficient is a measure of strength between the two variables. The Pearson technique only works if the relationship is linear and the coefficient is in range [-1,1]. A positive correlation and therefore a positive coefficient indicates that the variables either decrease or increase together, and if one variable increases the other one will have the same behavior. Negative correlation indicates that as one variable increases the other will decrease and the other way around (University of the West of England, 2017).

In the current analysis, a Pearson Correlation is performed for the continuous variables CO_2 and temperature, respectively between CO_2 and humidity. The purpose of this analysis is to see how the CO_2 impacts the two parameters. Table 51 shows the correlation between temperature and CO_2, and reveals that the coefficient has a positive value of 0.523, which is a value closer to 1 than 0, inclining therefore towards a positive correlation. This means that while the CO_2 in the room will go up or down, the temperature will perform similarly.

As the growth of CO_2 is connected to the increase of people in the room, this results indicate an increase of temperature in the room when an increase in the number of people occurs. The correlation of CO_2 and the number of people is based on previous literature as indicated in the section 0 CO_2 sensors for occupancy detection.

Table 51: Correlation between temperature and CO2

Correlations			
		Temperature	CO2
Temperature	Pearson Correlation	1	**.523****
	Sig. (2-tailed)		.000
	N	4178	4178
CO2	Pearson Correlation	**.523****	1
	Sig. (2-tailed)	.000	
	N	4178	4178
**. Correlation is significant at the 0.01 level (2-tailed).			

*** = 0.01 level = 1%, this percentage should be bigger than 0 for the evidence to be significant*

Similarly, a Pearson Correlation is performed for the initial continuous variables CO_2 and humidity. Table 52 shows the correlation between humidity and $CO_{2,}$ and reveals that the coefficient has a negative value of -0.233. This is called a negative correlation meaning that if the CO_2 goes up, the humidity tends to go down, and the other way around. However, as the coefficient is closer to 0 than -1, the relation is not so strong. This symbolizes that a significant negative correlation is not present. This type of correlation is called a moderate negative correlation. This results indicate that the connection between CO_2 and humidity is not so strong, and therefore good predictions of CO_2 will not necessarily contribute to the assessment of humidity in the room. We can conclude that the humidity is independent of the CO_2 increase (there is not enough evidence suggesting a strong relation).

Table 52: Correlation between CO2 and humidity

Correlations			
		CO2	Humidity
CO2	Pearson Correlation	1	**-.233****
	Sig. (2-tailed)		.000
	N	4178	4178
Humidity	Pearson Correlation	**-.233****	1
	Sig. (2-tailed)	.000	
	N	4178	4178
**. Correlation is significant at the 0.01 level (2-tailed).			

*** = 0.01 level = 1%, this percentage should be bigger than 0 for the evidence to be significant*

For determining the correlation between the CO_2 and the PeopleCount a Pearson correlation is again used, together with a Spearman and Kendall`s tau-b method. According to Laerd (2013), Spearman`s rand-order correlation is the non-parametric version of the Pearson correlation, and its coefficient measures the strength and direction between two ranked variables. The variables used in this case are ordinal, interval or ratio, and the coefficient determines the monotonic relationship between the desired variables, rather than strength and direction of the linear relationship. A monotonic relationship means that the variable can go in one direction only (up or down), it cannot go up and then down.

Furthermore, Laerd (2013) also mention the Kendall`s tau-b method as a non-parametric measure of the strength and direction of the association between two variables measured on ordinal scale. This method is considered to be a replacement of the Pearson`s correlation, when the data fails one or more assumptions for the test, however it is also considered an alternative to the non-parametric Spearman correlation (especially useful when there are many tied ranks in the data). All three types of correlations are applied between the CO_2 continuous variable and the PeopleCount variable, which variates between 0 to 4 people present in the room.

Due to the low amount of cases registered for the PeopleCount, this variable might arguably not be considered a continuous variable. Therefore, for strengthening the results given by the Pearson correlation technique, the other two (Spearman and Kendall`s tau-b method) will also be applied.

Table 53 shows the results given by the Pearson correlation which indicates a coefficient of 0.649, therefore a positive strength relation between the two variables. Table 54 indicates the results for the Spearman and Kendall`s tau-b method which also indicate a positive correlation between the PeopleCount and the CO_2 levels. The coefficients 0.488 and 0.615 indicate a strong positive correlation, by being closer to 1 than 0. Hence, this result is very important as it suggests that CO_2 in the room will increase with an increased amount of people present in the room, and similarly decrease if there are no people present in the room. The result emphasized that the CO_2 sensor can be used as a mean for detecting occupancy in the room.

Table 53: Correlation between CO2 and PeopleCount

Correlations			
		CO2	PeopleCount
CO2	Pearson Correlation	1	**.649****
	Sig. (2-tailed)		.000
	N	4178	4178
PeopleCount	Pearson Correlation	**.649****	1
	Sig. (2-tailed)	.000	
	N	4178	4178
**. Correlation is significant at the 0.01 level (2-tailed).			

*** = 0.01 level = 1%, this percentage should be bigger than 0 for the evidence to be significant*

Table 54: Non-parametric correlation between PeopleCount and CO2

Correlations				
			CO2	PeopleCount
Kendall's tau_b	CO2	Correlation Coefficient	1.000	**.488****
		Sig. (2-tailed)	.	.000
		N	4178	4178
	PeopleCount	Correlation Coefficient	**.488****	1.000
		Sig. (2-tailed)	.000	.
		N	4178	4178
Spearman's rho	CO2	Correlation Coefficient	1.000	**.615****
		Sig. (2-tailed)	.	.000
		N	4178	4178
	PeopleCount	Correlation Coefficient	**.615****	1.000
		Sig. (2-tailed)	.000	.
		N	4178	4178
**. Correlation is significant at the 0.01 level (2-tailed).				

*** = 0.01 level = 1%, this percentage should be bigger than 0 for the evidence to be significant*

The correlations method show a linear relationship between the CO_2 and the PeopleCount. For further investigating this relationship, and which amount of CO_2 corresponds to which amount of people, the Linear Regression model is applied. The Linear Regression analysis can be performed in SPSS, for identifying supported evidence in the relationship between two variables.

Before the actual analysis there are two hypothesis formulated:

1. Null hypothesis (H_0= 0): There is **no** supported evidence of relationship between CO_2 and the people count (ground truth data).
2. Alternative hypothesis (H_a≠ 0): There is supported evidence of relationship between CO_2 and the people count (ground truth data).

For checking the hypothesis, the regression analysis is performed between the dependent variable CO_2 and the independent variable PeopleCount. The most important result given by the Regression Analysis is the R Square, which results from the model summary. R Square in this model equals **0.421**. This value suggests that the PeopleCount accounts for 42.1% of the variation in CO_2.

The Regression equation is: $y = a + b * x + e$. This equation helps predicting the CO_2 for different values of the PeopleCount. The most important part of the regression analysis is given by the coefficients table (Table 55), where the Constant is the interceptor **'a'** in the regression equation. This value tells that when the PeopleCount is 0, the CO_2 is 445.651. The other important coefficient is the slope, which in this case is represented by the value: 55.094. The slope symbolizes the increase in CO_2 for each unit of PeopleCount.

Moving to the significance of the model, it is represented by the value 0. This value is smaller than 0.05 (the alpha level), which shows the possibility to reject the Null Hypothesis (H_0= 0). With H_0= 0 rejected, the Alternative Hypothesis is correct and therefore there is supported evidence of relationship between CO_2 and the people count (ground truth data).

Table 55: Regression analysis coefficients

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	**445.651**	1.022		435.989	**.000**
	PeopleCount	**55.094**	1.000	.649	55.093	.000
a. Dependent Variable: CO2						

After finding the coefficients, the regression equation: $y = a + b * x + e$ can be used for predicting CO_2 based on the number of people present in the room.

- For 0 people in the room the predicted CO_2 is: y = 445.651
- For 1 person in the room the predicted CO_2 is: $y = 445.651 + 55.094 * 1 = 500.742ppm$
- For 2 people in the room the predicted CO_2 is: $y = 445.651 + 55.094 * 2 = 555.839ppm$
- For 3 people in the room the predicted CO_2 is: $y = 445.651 + 55.094 * 3 = 610.933ppm$
- For 4 people in the room the predicted CO_2 is: $y = 445.651 + 55.094 * 4 = 666.027ppm$

The applied method is a linear regression, therefore there is a linear increase in the CO_2 values based on the amount of people. No external factors or error that might impact this linear effect is considered in this model. Therefore, the only result that will be considered is the difference between the CO_2 for no people present in the room (445.651 ppm) and the illustrated CO_2 for one person (500.742 ppm). A value of 500 ppm will be considered for the creation of the future BBN with the second data set including PIR data.

4.3.2. BBN Analysis

With the PIR sensor included, the 'NewMotionCount' variable was introduced. This variable shows the 'Undetected' status whenever there is no count registered for the certain amount of time, whereas the 'Detected' status is registered for any of the states from 1 to 17 movements. With the data described in 3.5.2, the Bayesian Belief Network present in Figure 35 is estimated. The strength of influence can also be visualized in the same figure.

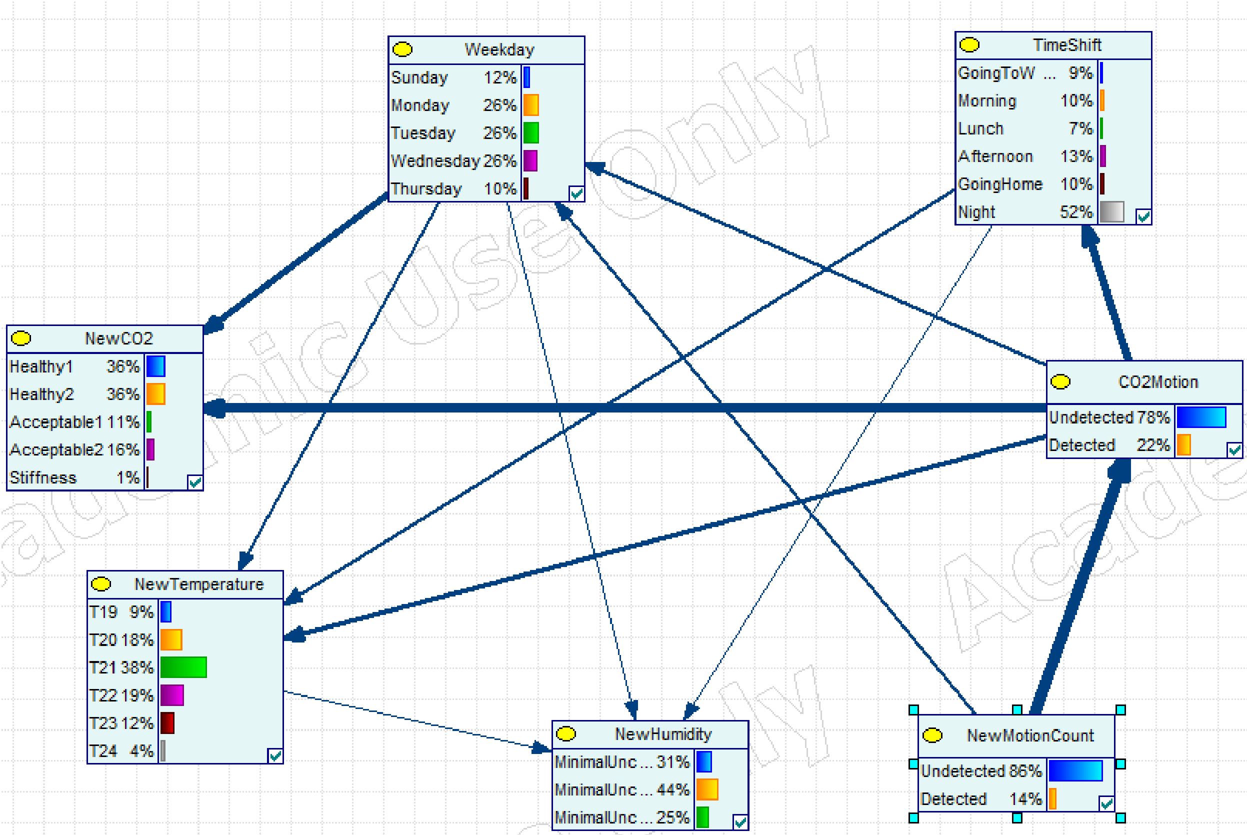


Figure 35: Bayesian Belief Network with PIR sensor

The strength of influence plays an important role in this analysis, as it shows how the motion variables are connected to each other and how do they relate to the other variables. This correlation gives information on how accurate future predictions of real-time occupancy are.

The strength of influence is calculated based on the CPT of each child node and it is illustrated in Table 56. The biggest strength of influence is given by the 'NewMotionCount' variable on the 'CO2Motion' variable. This shows that the motion will impact the most the CO_2 level below or above 500 ppm. The variable 'CO2Motion' was made based on the ground truth data denoting presence. Hence, the strong relation between the 'CO2Motion' and the 'NewMotionCount' given by the PIR sensor, is consequently an indication of good predictions that can result from the PIR sensor.

The second biggest strength is given by the 'CO2Motion' on the 'NewCO2', which is an obvious result, since both variables were re-coded based on the CO_2 emission. The strong impact of the 'CO2Motion' on the 'TimeShift' (0.5455) indicates that the time of the day can be predicted based on the presence in the room, which is given by CO_2 emission higher or lower than 500 ppm.

The same table also shows that the 'NewHumidity' is the weakest variable when it comes to correlations with other variables. Therefore, it is particularly difficult to predict humidity based on the time of the day, week day, or occupancy.

This connects to the results presented in the historical data analysis, based on which it is easier to predict CO_2 levels and temperature according to time and week day, whereas the humidity and temperature are more connected to the month (variable that is not present in the current analysis).

Table 56: Strength of influence BBN with motion count

Origin (Parent Node)	Destination (Child Node)	Strength of influence
NewMotionCount	**CO2Motion**	**0.8209**
CO2Motion	NewCO2	0.6193
CO2Motion	TimeShift	0.5456
Weekday	NewCO2	0.4035
CO2Motion	NewTemperature	0.3364
TimeShift	NewTemperature	0.2854
Weekday	NewTemperature	0.2731
NewMotionCount	Weekday	0.2531
CO2Motion	Weekday	0.2269
NewTemperature	NewHumidity	0.1887
Weekday	NewHumidity	0.1845
TimeShift	NewHumidity	0.1612

Scenarios

The possibility to test different scenarios is also explored with the experimental data. For the BBN estimated with the current data set, the following scenarios were tested:

- Scenario 1: Evidence for 'CO2Motion' (100% Undetected) and 'NewMotionCount' (100% Detected)
- Scenario 2: Evidence for 'CO2Motion' (100% Detected) and 'NewMotionCount' (100% Undetected)
- Scenario 3: Evidence for 'CO2Motion' and 'NewMotionCount' (100% Detected)
- Scenario 4: Evidence for 'TimeShifts' (change in the 'NewMotionCount' and CO2Motion variable according to the time of the day)

The first scenario presents evidence for the 'CO2Motion', where the 'Undetected' state is set to 100%, and evidence for the 'NewMotionCount', where the 'Detected' state is set to 100%. This means that the CO_2 level will always be under 500 ppm, while the PIR sensor ('MotionCount') will be higher than 0. Therefore, only the PIR sensor will be indicate presence in the room.

The second scenario illustrates the opposite compared to scenario 1. The evidence for 'CO2motion' is set to 100% in the 'Detected' state, while the evidence for 'NewMotionCount' is set to 100% in the 'Undetected' state.

The third scenario presents evidence for both the 'NewMotionCount' and the 'CO2Motion'. Both variables will be set to 100% in the 'Detected' state.

The three scenarios are illustrated in Table 57, where the percentages allocated to each state in the variables 'NewCO2', 'NewTemperature' and 'NewHumidity' are presented. These scenarios are presented in parallel to exemplify the differences between having activated only the PIR sensor and separately only the CO_2 sensor.

Table 57: Scenario 1, 2 and 3 for the BBN experimental data

Variable	No Evidence	Evidence CO2Motion (Undetected 100%) and Evidence NewMotionCount (Detected 100%)	Evidence CO2Motion (Detected 100%) and Evidence NewMotionCount (Undetected 100%)	Evidence CO2Motion (Detected 100%) and Evidence NewMotionCount (Detected 100%)
NewCO2				
Healthy1	36%	20%	2%	1%
Healthy2	36%	71%	2%	1%
Acceptable1	11%	8%	19%	22%
Acceptable2	16%	1%	74%	72%
Stiffness	1%	1%	4%	3%
NewTemperature				
T19	9%	15%	5%	4%
T20	18%	31%	5%	4%
T21	38%	31%	20%	22%
T22	19%	11%	36%	35%
T23	12%	8%	22%	25%
T24	4%	3%	12%	10%
NewHumidity				
MinimalUncomfortable1	31%	15%	30%	24%
MinimalUncomfortable2	44%	41%	46%	44%
MinimalUncomfortable3	25%	44%	25%	32%

The 'NewCO2' variable will present in Scenario 1 a predominately Healthy Level 1 and 2. The scenario suggests that even if motion is detected by the PIR sensor, the CO_2 emission will stay under 500 ppm. These situations are represented by 'Afternoon', 'GoingHome' and 'Night' periods, when only a few people are still in the office, therefore movement will be recorded, however the CO_2 level already dropped. In Scenario 2, the 'NewCO2' variable presents mostly cases of Acceptable Level and an increased Stiffness level. These are the cases where no movement is recorded for more than five minutes (throughout the whole day), however the level stays at a high level since people are still present in the room. The sensor does not present any movement because the employees might be sitting at their office or not passing in front of the sensor. Lastly, scenario 3 emphasizes the best scenario, with both sensors active and 100% detection. This case presents very similar results to scenario 2, when the 'CO2Motion' predicts the level of occupancy (an increased CO_2 acceptable level and decreased healthy level indicates presence). Furthermore, the PIR sensor applies an additional increase in the Acceptable Level and decrease in the Healthy level from Scenario 2.

This scenario strengthens therefore, the relationship between the CO_2 and the PIR sensor, showing additional increase in the CO_2 levels when the status of the PIR sensor is 'Detected'.

The 'NewTemperature' will increase in Scenario 1 in states T19 and T20, and decrease in the others. In Scenario 2 the same variable will increase in T22, T23 and T24 and decrease in T19, T20 and T21. This result only shows that the temperature is connected to the CO_2 increase and a low temperature will be associated with a low CO_2 emission, while a higher temperature will be associated to a higher emission. As this data set is limited at only one week of data, this outcome is not sufficient for analyzing this connection. The results on how the temperature changes based on the CO_2 is, however, presented in the historical data analysis. The third scenario shows again similar results to the Second scenario, which a further increase in temperature, which shows that an increase in occupancy will actually contribute to increasing the temperature in the working environment.

Scenario 1 identifies an increase in humidity the variable 'NewHumidity', therefore the motion count is associated with the humidity increase. However, when looking at scenario 2, the same humidity levels are obtained as in the initial case with no evidence present. Furthermore, scenario 3 also keeps the same patterns in humidity as in the initial case without evidence, with a predominant minimal uncomfortable humidity of level 2 present in the room. There is however, not enough background information for concluding that the movement detected with the PIR sensor can be directly connected to the increase in humidity, or that the increase in CO_2 does not impact the humidity in any way.

Scenario 3 shows a high level of presence in the room, due to the fact that both sensors are active. This scenario indicates a high level of CO_2 (72% Acceptable Level 2 and increase in Stiffness Level), a temperature higher than 21°C (22% for T21, 35% for T22, 25% for T23), and primarily an uncomfortable humidity level 2. This indicates overall a good maintenance of the indoor air quality when people are present in the room, as the CO_2 should be maintained under 700 ppm. However, the temperature can be adjusted to lower temperature (19° to 21°C), for avoiding over usage of the heating system and increase in energy consumption. At the same time, the humidity is primarily uncomfortable and too low for the working environment, and an increased consideration should be given to this aspect, as previously emphasized in the historical data analysis. Real-time occupancy does not play an important role in the change of humidity, similar levels will be present in the room with or without people.

The fourth scenario emphasizes the changes in the 'CO2Motion' and 'NewMotionCount' variables based on the time of the day selected. This scenario is illustrated for seeing how the presence is reflected throughout the day for potentially creating a room occupancy schedule for the Credo Room. Table 58 shows the allocated evidence for each time period.

Table 58: Scenario 4: CO2Motion and NewMotionCount variables according to the time of the day

	No Evidence	1 – 100% Evidence	2 - 100% Evidence	3 - 100% Evidence	4 - 100% Evidence	5 - 100% Evidence	6 - 100% Evidence
CO2Motion							
Undetected	**78%**	29%	33%	39%	**66%**	**99%**	**99%**
Detected	22%	**71%**	**67%**	**61%**	34%	1%	1%
NewMotionCount							
Undetected	**86%**	**57%**	**60%**	**64%**	**79%**	**98%**	**98%**
Detected	14%	43%	40%	36%	21%	2%	2%

1 = Going-to-work ; 2 = Morning ; 3 = Lunch ; 4 = Afternoon ; 5 = Going-home ; 6 = Night

Scenario 4 shows that the state 'Detected' in the room increases in the going-to-work period and decreases as progressing in the day, in both variables. The 'CO2Motion' shows 71% 'Detected' state and the 'NewMotionCount' shows 43%, in the 'Going-to-work' period. This gives a high indication that there will be people present in this period in the office. There is only a really small decrease in the 'Detected' state between the 'Going-to-work' period and the 'Morning' period (4% in 'CO2Motion' and 3% in 'NewMotionCount'), and between the 'Morning' and 'Lunch' period (6% in 'CO2Motion' and 4% in 'NewMotionCount'). However, a big difference between the 'Lunch' and 'Afternoon' period (27% and 15%) and between the 'Afternoon' and 'Going-home'/'Night' periods (33% and 19%) is observed. This suggests that people arrive in the 'Going-to-work' and 'Morning' periods, but mostly in the 'Going-to-work' due to the slightly higher percent. Occupancy drops in the 'Lunch' period, when some employees will go for lunch outside the working room. The 'Afternoon' period is less populated than the 'Morning', which means that either employees leave after 2pm, either that they do not return to the office after lunch. The 'Detected' state of 1 and 2% in the going-home and evening periods indicates no presence in the room.

4.3.3. Conclusions for the experiment

In Chapter 4.3 Experimental Data, the results of the conducted experiment were explained. The experiment consisted of two data sets: one set with the ground truth data and one set with the PIR sensor included. In the ground truth data analysis, the connection between the CO_2 emission and the number of people was explored, results that were used further for analyzing the second data set with the PIR included. The main results of the experiment are emphasized below.

Firstly, from the ground truth data analysis it can be concluded that there is a correlation between the CO_2 emission and the number of people present in the room. Furthermore, there is also a correlation between the temperature and the CO_2, but no relevant results show correlation between the humidity and the CO_2. Hence, the temperature and CO_2 are dependent on the number of people and presence in the office room. By means of linear regression, the CO_2 level corresponding to the number of people was calculated. The emission of CO_2 corresponding to presence (therefore 1 person in the room) was calculated to be 500 ppm. With this value the 'CO2Motion' was introduced in the second data set.

Secondly, the experiment data consists of PIR data and the re-coded 'CO2Motion' variable denoting presence at 500 ppm. Based on the new BBN network, it was concluded that the 'CO2Motion' and the PIR (the variable 'NewMotionCount') present the strongest connection. The 'CO2Motion' is a variable introduced based on the actual amount of people present in the room. Therefore, the strong connection between the CO2Motion and 'NewMotionCount' shows that the 'NewMotionCount' can predict accurately presence in the room. Furthermore, scenario 3 also emphasizes the added value of having both variables CO2Motion and NewMotionCount for predicting future CO_2, humidity and temperature levels.

Thirdly, the 'NewTemperature' is also visibly connected to the presence, as an increase in temperature is observed when applying scenario 2 and 3. The same cannot be said about the humidity in the room, the variable 'NewHumidity' is maintained at a similar level before any scenarios are applied. Therefore, humidity cannot be predicted based on presence in the room.

Lastly, the 'CO2Motion' and the 'TimeShift' show a high strength of influence. Furthermore, Scenario 4 shows how the presence changes along the time of the day. The main conclusion drawn is that the afternoon presents a lower rate of occupancy than the morning which means that either people do not return after lunch, or that they leave the office around 2pm.

4.4. Data Analysis conclusions

This chapter introduces the results of the descriptive analysis and BBN scenarios. For answering RQ1, the results for historical data analysis of two office spaces are displayed. For both Credo Room and S2M Room, the study concludes that the month is the most significant variable for detecting all three indoor parameters: CO_2, humidity and temperature. The humidity and temperature can be the best predicted according to the month, season and outdoor conditions. Based on the time of the day, the CO_2 variation can be analyzed for both office spaces. The conclusions for each of the two rooms also show specific patterns for the months, weekdays and times, which are formulated into recommendations and explained in the next chapter.

For approaching the second part of the research, an experiment is conducted, where correlation between CO_2 and the people count is observed, or between CO_2 and temperature; but no significant correlation is present between the CO_2 and the humidity. Furthermore, the BBN also shows that while CO_2 and temperature are connected to the occupancy in the room the humidity is the only variable showing the weakest correlation.

5

CONCLUSION AND DISCUSSION

Introduction

Research Questions Answers

Comparison Among Literature and Results

Scientific Relevance

Societal Relevance

Recommendations

Limitations of the research

Future research and personal reflection

5.1. Introduction

The last chapter of this report provides a final conclusion and the discussion on the developed research, for the topic of indoor air quality in the context of real-time occupancy and sensor data. Firstly, the answer to the research questions is summarized based on the mentioned results in the previous chapter, and a comparison between the existing literature and these results is presented. Next, the scientific and societal relevance of the performed study is presented, compared to the importance of this study from section 1.5. Moreover, recommendations for actions to be developed by office occupants, and Inteliments (the company providing the indoor air quality platform) are provided. Lastly, the limitations and weaknesses of the current research are illustrated. Furthermore a small section exemplifies a personal reflection on the topic and process for developing the thesis, as well as recommendations for future research.

5.2. Research question answers

Part 1: Indoor air quality improvement (IAQ)

The first research question was: **How can future states of CO_2, humidity and temperature in an office room be predicted based on historical sensor data for improving the indoor air quality?** Based on the developed research the future states for two rooms were predicted by (1) Assessing future states based on the period, (2) Comparing the different states between multiple rooms and, (3) Taking into account the weather conditions.

Firstly, CO_2, humidity and temperature were compared to the time, weekday and month, with the objective of foreseeing their future increase or decrease. For determining future states in an office room, the most important variable that should be considered is the month, based on which the biggest differences were observed for all three parameters. Furthermore, future states of CO_2 can be predicted based on the time of the day and day of the week, especially for the time of the day. In terms of humidity, the recorded humidity was primarily under 40%, which is the minimal desired humidity in an office environment. The monthly humidity predictions are very important because they are the only ones influencing the drastic change in humidity. The temperature levels can be predicted together with the CO_2, primarily a high emission of CO_2 will also indicate an increase in temperature, and this is dependent on the actions the people take in regards to the HVAC system. Temperature cannot be predicted based on the time of the day, however a prediction of CO_2 for the week days can also come with a prediction of temperature.

Secondly, the analysis performed in the second space mostly confirmed the assumptions generated from the first space, however some patterns are really connected to the place of research. An example here would be the CO_2 predicted based on the weekdays, which, in the first space is the highest on Tuesday and Wednesday, while in the second is the highest is on Thursday and Friday. Similarly, the temperature in the first space increases with the CO_2 emission, however this is not the case in the second room, where there are certain variations.

Lastly, in what concerns the impact of the weather on the future predictions, this should be taken into account when predicting temperature and especially the humidity. The humidity is strongly connected to the outdoor humidity so its change will not depend only on the month but also on the outside temperature and humidity.

The research sub-question was: *How do the indoor CO_2, humidity and temperature levels change according to the season, month, day of the week and time of the day?* This sub-question was used as basis for providing the answer to the main RQ, which is described above. For the two rooms, two similar predictions were made. The exact patterns identified in answering this sub-question can be found in the section Conclusions of the chapter 4.2.1 Office 1: Credo Room and in the section Conclusions of the chapter 4.2.2 Office 2: Seats 2 Meet Conference Room.

Part 2: Real-time occupancy detection

The second research question was: **What sensor data can provide the most accurate information in detecting real-time occupancy information?** For answering this question, a literature review was conducted, on the existing types of sensors, as well as on identifying the advantages and shortcomings of three sensor types: PIR, CO_2 and Electromagnetic Signals. Overall, most authors suggest and conduct experiments with a fusion of sensor, specifically a combination of two or more sensors. A PIR sensor would indicate the presence in the room, but is not useful in detecting the amount of people present in the room, similarly for the CO_2. Different types of electromagnetic signals have proven to be useful in occupancy detection (such as RFID or WLAN), however privacy concerns are raised in multiple occasions when using such systems. Therefore, a non-intrusive approach are the environmental sensors, such as the CO_2 sensors. Chair, mouse of keyboard sensors are also useful in occupancy detection, but contrasting the PIR sensors, they detect presence only when the person is seated or working at his/her desk. Therefore a combination of PIR/chair/keyboard sensors is suggested by some authors. All in all, an accurate prediction of real-time occupancy can be given by (1) A fusion of sensors, and (2) The Electromagnetic Signals provide the biggest accuracy however they come with the privacy disadvantage.

For exploring further this research question, a case study was built on the PIR and CO_2 sensors, for answering the sub-question: *How accurate is the information fusion given by the CO_2 and PIR sensors in detecting real-time occupancy?* The experiment concludes that (1) the CO_2 sensor is strongly connected to the real-time occupancy and, (2) the combination of CO_2 and PIR sensors strengthens the changes in the BBN compared to having just the CO_2 sensor alone detecting occupancy.

The conducted experiment fails to give a direct correlation between the PIR sensor and the real-time occupancy, due to the lack of ground truth data in the same period when the PIR sensor collected data. Therefore, the only conclusion that can be drawn is related to the correlation between the CO_2 sensor and the PIR sensor, which is very strong. The strong correlation between the three variables gives a positive indication on the accuracy of this information fusion, however another experiment should be conducted for verifying this result.

Finally the last research question was: **How do the indoor CO_2, humidity and temperature levels change according to the real-time occupancy?** After the experiment was conducted, it can be concluded that (1) CO_2 and Temperature increase with the increase of people in the room and (2) Humidity does not depend on the occupancy. In terms of CO_2 level, the number of people affects the CO_2 level, however the PIR sensor only detects presence and not the exact number of people. Therefore, an indication of presence will indicate a level higher than 500 ppm. In terms of temperature, presence in the room relates to a temperature higher than 22°C. The humidity is the variable, which, does not depend on the amount of people in the room, but depends on other factors as emphasized in the first RQ.

5.3. Comparison among literature and results

Part 1: Indoor Air Quality improvement (IAQ)

In what concerns the IAQ, literature shows that indoor air pollution and temperature are interrelated, and they can be controlled by making adjustments to the HVAC system. In this regards, the results show that indoor temperature and CO_2 are indeed interrelated, and recommendations are formulated for making adjustments to the HVAC system. Based on the results, humidity is however not connected to the temperature, and is considered to be a separate indoor parameter, controlled more by the weather conditions.

The literature review also shows that, according to Stazi, Naspi, & D`Orazio (2017), the indoor temperature is the second driver to intervene on the opening or closing of windows. In the current study, the way the users intervene on opening and closing the windows in the office space, is not specifically researched. However, in one room the temperature goes up with the increase of occupancy, while in the second, the opposite happens. This means that the occupants intervene on the indoor temperature for achieving thermal comfort, and either make adjustments to the HVAC system, either open/close the windows.

Finally, the literature review also reveals the fact that CO_2 relates to metabolism, which is shown by the results of the thesis, by having a connection between the CO_2 and the amount of people in the ground truth data. Furthermore, CO_2 also increases in inadequately ventilated spaces, hence, the recommendations concerning the decrease in CO_2 or temperature also focus on increasing the ventilation in the office space.

Part 2: Real-time occupancy detection

This section focuses on the comparison between literature and the results in terms of occupancy patters.

Firstly, in terms of months, literature identifies the occupancy peaks during January, March till June, September and October, medium occupancy in December, February and July, and low occupancy in November and August. The results of this thesis illustrate the peaks for two different rooms, showing an increased CO_2 from October to December, and the lowest CO_2 and therefore occupancy in October.

Secondly, a similar difference in patterns is observed for the weekdays. In the literature review, different sources point out different days for occupancy peaks. One study identifies the peak in occupancy occurring on Monday, and the lowest occupancy being on Friday, while another study shows that Tuesday is the most occupied room and Thursday the least. The results of this study show that during Tuesday the highest amount of people is present in Credo Room, while for the S2M Room the peak occurs on Thursday and Friday.

Lastly, in terms of hours, the literature and the results show similar results. Based on literature, peaks occur before lunch and during lunch, while the results show the same. However, literature also shows that afternoon presents higher occupancy than the morning period, which is not the case for this study.

Generally, the results do not correspond to the results illustrated in literature, and the reason for this can be the fact that the experiments are performed in different countries, and in different type of offices. In this thesis, while the first office analyzed is a private room, the second one is a conference room, where occupancy schedule does not correspond to a normal working schedule.

5.4. Scientific Relevance

From a scientific perspective this research was performed for contributing to existing academic literature, on the topic of real-time occupancy in the context of virtual sensors and on the correlation between real-time occupancy and the indoor air quality.

The real-time occupancy detection is presented from the point of view of multiple authors in a literature review, and a variety of experiments were completed for improving this field of research. This study adds value to this field, by providing an experiment focused on the detection of real-time occupancy by using both environmental (CO_2) and virtual sensors (PIR) and by estimating a model which connects the indoor air quality system with the occupancy.

Furthermore, the model used during the experiment is a BBN model, which adds value to existing literature by presenting a less explored method of analyzing real-time occupancy patterns. The BBN model proves to be successful not only into detecting patterns in the CO_2, humidity and temperature levels, but also into connecting and analyzing the relationship between the occupancy and the indoor parameters.

5.5. Societal Relevance

This research was carried out with the purpose of contributing to improving the quality of life of employees in an office space, through the decrease in air quality risks existent in their office. Furthermore, the study has the potential of reducing energy costs in the offices where the study was analyzed, based on the observations made in terms of HVAC over usage. This study also presents the opportunity to include the explored patterns for each individual room in the climate control system, for better monitoring and management of energy resources.

Firstly, the societal aspect of the thesis is determined by the air quality improvements, which can be made based on identified CO_2, humidity and temperature patterns existent in each room individually. The literature review shows that air quality has a strong contribution on the health and productivity of the employees, and therefore it can potentially affect their quality of life. The study reveals that certain months, days and times in a day are prone to increased exposure to non-optimal levels of CO_2, humidity and temperature. Therefore, the societal relevance of this study consists of the potential actions that will be taken by the employees, and climate control managers into reducing the existent risks in the identified periods.

Secondly, the energy costs reduction was not directly approached by this thesis. However, the results show that there are certain times when the temperature could be decreased in the office space, due to the high amount of people or excess of heating in the room. The energy costs reduction potential make this thesis relevant from the societal perspective, because they are strongly connected to the community of employees present in the researched areas.

Lastly, the climate control system is an interesting topic which is connected to both energy costs and the indoor air quality improvement. As described in the section 3.3 Case selection, the areas of research are located in old industrial buildings, where the climate control is not handled for each room individually but rather on an entire floor. Therefore, creating an automated climate control system is difficult in the study area, and an analysis for all the rooms connected to one HVAC system should be made. However, from a societal perspective, this thesis presents a good potential in generating some patterns in occupancy, or CO_2, humidity and temperature, that could be generalized for more rooms in future studies.

5.6. Recommendations

Based on the described results, several patterns both based on historical data and occupancy were identified. Recommendations for the climate control managers, office employees and Inteliments (company providing the air quality platform) were made, and are shown in Table 59.

Table 59: Recommendations for air quality improvement

Indoor Air Quality		
Pattern	**Recommendation**	**Who?**
September: highest temperature, lowest amount of CO_2, optimal humidity	No adjustments should be made to the HVAC system, as it presents the best levels of humidity and CO_2. Temperature is relatively high, however this contributes to having optimal humidity in the room.	Office employees Climate control managers
October, November and December: highest amount of CO_2, and minimal humidity	• Introduce humidifier • Increase ventilation in the room	Climate control managers
Tuesday: highest temperature, biggest amount of people in Credo Room	• Decrease heating • Open windows in the morning	Office employees

Lunch and Afternoon: highest CO_2, temperature	• Open windows at the end of the morning period and during lunch	Office employees
Minimal uncomfortable humidity all the winter months	• Introduce humidifier in winter	Climate control managers
	Real-time occupancy	
Pattern	**Recommendation**	**Who?**
Temperature and CO_2 are dependent on the number of people	• Include presence/absence in the dashboard (air quality platform) to be linked with PIR and increased amount of CO_2 • Increase ventilation before arrival of employees (in the going-to-work period) • Set adequate temperature before arrival in the room	Inteliments Climate control managers
Occupancy relates to CO_2 500 ppm	• Link presence to 500 ppm in the dashboard • Decrease temperature in the room for reducing the amount of fossil fuel generated, when the CO_2 exceeds 700 ppm with people in the room • When presence is indicated in the dashboard decrease temperature	Inteliments Office employees
Afternoon presents less occupancy than the morning	• Increase ventilation during the going-to-work and morning period	Office employees

5.7. Limitations of the research

In this section some limitations of the current research are discussed. One of the limitations is that this experiment lacked ground truth data corresponding to the period when the PIR data was collected. The ground truth data was collected for a certain period of time when the PIR sensor registered a data outage, therefore the ground truth data was only compared to the CO_2 information. Also, the PIR data was collected only for 5 days, therefore it was not possible to estimate daily or monthly patterns on occupancy.

Furthermore, another limitation is given by the lack of an additional type of sensor (such as a Wi-Fi sensor or chair sensor), that could complement the current information fusion. With a Wi-Fi sensor, the exact number of connected devices could be analyzed and compared to the results given by the regression model, or furthermore, included in the BBN for analyzing the occupancy throughout the day. Also, a chair sensor could provide additional information for when the CO_2 and PIR sensor do not register occupancy (one person sitting for a long period of time when PIR does not register movement and the CO_2 level dropped below 500 ppm).

Finally, the last limitation is related to the country and the place of the study. As the country is Netherlands, the results are dependent on the weather conditions and working schedule in this country, which makes it difficult for the results to be generalized. Furthermore, as described in section 5.3 there are differences in the occupancy patterns during the day and month, between this study and previous literature. These differences strengthen the conclusion that patterns in occupancy cannot be generalized according to the month or day of the week. Regarding the place of the study, the office space where the experiment is conducted is a private room, so more evidence is needed for the open plan offices which are becoming popular nowadays.

5.8. Future research and personal reflection

The last section of this report focuses on the future research recommendations and on illustrating the personal reflection in terms of process and topic. In terms of future research, some ideas are illustrated below.

Firstly, Section 5.7 illustrates the limitations which can be used as possible improvements for the future. Secondly, the study could be repeated and data could be collected over a period of a whole year. A complete data set for one full year will give an indication of different seasons and the weather impact on the indoor variables. Furthermore, as illustrated in the results, the month plays an important role in generating CO_2, humidity and temperature patterns, therefore a complete data set for one year would be ideal for a thorough analysis. Thirdly, another recommendation for future research is to perform the study in multiple rooms within the same floor, for comparison. As the HVAC system connects multiple rooms, ideally the study will be performed in those rooms, for seeing the differences between the indoor parameters and which automatic adjustments in the climate control system could be implemented. The last recommendation is to perform the study in different office types (open-space and private), with more occupants, for identifying how much the CO_2 levels change for more occupants than in this experiment.

Finally, this last paragraph illustrates the thoughts of the researcher on the topic of real-time occupancy and sensor data for improving the indoor air quality.

The evolution of sensors and IoT technology can play a significant role in improving the health and performance of a building and its occupants. The real-time occupancy detection with the help of IoT has proven to be a good method that can be used in many aspects of the built environment: better energy control, improvement of the IAQ, improving the health of the employees. However, the detection of real-time occupancy might not always be straight forward and can require additional efforts from the office managers for ensuring a non-invasive and non-intrusive monitoring. Therefore, there can be some challenges into collecting and analyzing data related to individual behavior or presence in an office building. This problem can be tackled by applying a non-intrusive method of occupancy detection. In this respect, the CO_2 sensors can have a double purpose. They can be further developed and explored as a mean for occupancy detection but also to contribute to maintaining a certain healthy level in the working environment. This study achieved a non-intrusive monitoring and inexpensive solution for occupancy detection, by making use of existing air quality data and easy to install sensors. Furthermore, this study does not only provide contribution in the field of occupancy detection in the office, but also incorporates improvements for indoor air quality in connection to the real-time occupancy.

References

Akkaya, K., Guvenc, I., Aygun, R., Pala, N., & Kadri, A. (2015). IoT-based Occupancy Monitoring Techniques for Energy-Efficient Smart Buildings. *2015 IEEE Wireless Communication and Networking Conference (WCNC)*, (pp. 58-63).

American Society of Heating, R. a.-C. (2007). *ASHRAE STANDARD* . 1791 Tullie Circle NE, Atlanta, GA 30329 .

Bakker, C. d., Aries, M., Kort, H., & Rosemann, A. (2017). Occupancy-based lighting control in open-plan office spaces: A state-of-the-art review . *Building and Environment 112*, 308-321.

BayesFusion, L. (2017). *GeNIe Modeler - USER Manual.*

BayesFusion, LLC. (2017). *GeNIe Modeler User Manual Version 2.1.1.*

Cali, D., Matthes, P., Huchtemann, K., Streblow, R., & Muller, D. (2014). CO2 Based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. *Building and Environment 86*, 39-49.

Chau, C., Hui, W., & Tse, M. (2006). Evaluation of health benefits for improving indoor air quality in workplace. *Environment International* , 186-198.

Chen, Z., Masood, M. K., & Soh, Y. C. (2016). A fusion framework for occupancy estimation in office buildings based on evironmental sensor data. *Elsevier: Energy and Buildings*, 790-798.

Cociorva, S., & Iftene, A. (2017). Indoor air quality evaluation in intelligent building . *Energy Procedia 112*, 261–268 .

COOPER, G. E., & HERSKOVITS, E. (1992). A Bayesian Method for the Induction of Probabilistic Networks from Data. *Kluwer Academic Publishers*, 309-347.

de Bakker, C., Aries, M., Kort, H., & Rosemann, A. (2016). Occupants' behaviour: an observation study for energy saving potential on lighting. *BEHAVE 2016 4th European Conference on Behaviour and Energy Efficiency*, (pp. 1-12). Coimbra.

Division of information technolology services, CAL State LA. (2016). *IBM SPSS Statistics 23, Part 3: Regression Analysis.* Los Angeles. Retrieved from http://www.calstatela.edu/its/training

Dodier, R. H., Henze, G. P., Tiller, D. K., & Guo, X. (2005). Building occupancy detection through sensor belief networks. *Elsevier: Energy and Buildings*, 1033-1043.

Duarte, C., Wymelenberg, K. V., & Rieger, C. (2013). Revealing occupancy patterns in an office building through the use of occupancy sensor data. *Elsevier: Energy and Buildings*, 587-595.

Duarte, C., Wymelenberg, K. V., & Rieger, C. (2013). Revealing occupancy patterns in an office building through the use of sensor data.

Erickson, V. L., Lin, Y., Kamthe, A., Brahme, R., Surana, A., Cerpa, A. E., . . . Narayanan, S. (2009). Energy Efficient Building Environment Control Strategies Using Real-time Occupancy Measurements. *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, (pp. 19-24). Berkeley, California.

Fenton, N. (1995). *Tutorial of SERENE (SafEty and Risk Evaluation using bayesian Nets).* Retrieved from EECS - School of Electronic Engineering and Computer Science, Queen Mary University of London: http://www.eecs.qmul.ac.uk/~norman/SERENE_Help/start_i.htm

Gruber, M., Truschel, A., & Dalenback, J.-O. (2014). CO2 sensors for occupancy estimations : Potential in building automation applications. *Elsevier: Energy and Buildings 84*, 548-556.

Gruber, M., Truschel, A., & Dalenback, J.-O. (2014). Model based controllers for indoor climate control in office buildings - complexity and performance evaluation. *Energy and Buildings 68*, 213-222.

Huizenga, C., Abbaszadeh, S., Zagreus, L., & Arens, E. A. (2006). Air quality and thermal comfort in office buildings: Results of a large indoor environmental quality survey. *Proceedings of Healthy Buildings Vol III* (pp. 393-397). Lisbon: eScholarship University of California.

IBM. (2017). *IBM SPSS Statistics*. Retrieved from IBM Official Website: https://www.ibm.com/us-en/marketplace/spss-statistics

IBM SPSS Statistics. (2015). *IBM SPSS Decision Trees 23.* Chicago.

Kemperman, A., & Timmermans, H. (2014). Environmental Correlates of Active Travel Behavior of Children. (S. Publications, Ed.) *Environment and Behavior Vol. 46*, 583 –608 .

Kik, S., Tang, J., Boxem, G., & Zeiler, W. (2016). Indoor air quality and thermal comfort of working squares in schools. *Paper presented at 12th REHVA World Congress (CLIMA 2016), May 22-25, 2016*, (pp. 1-10). Aalborg, Denmark.

Krieg, M. L. (2001). A Tutorial on Bayesian Belief Networks. *DSTO Electronics and Surveillance Research Laboratory*, 012–084.

Kumar, P., Skouloudis, A. N., Bell, M., Viana, M., Carotta, M. C., Biskos, G., & Morawska, L. (2016). Real-time sensors for indoor air monitoring and challenges ahead in deploying them to urban buildings. *Science of the Total Environment* , 150-159.

Labeodan, T., Aduda, K., Zeiler, W., & Hoving, F. (2015). Experimental evaluation of the performance of chair sensors in an office space for occupancy detection and occupancy-driven control. *Energy and Buildings*.

Labeodan, T., de Bakker, C., Rosemann, A., & Zeiler, W. (2016). On the application of wireless sensors and actuators network in existing buildings for occupancy detection and occupancy-driven lighting control . *Energy and Buildings, 127*, 75-83.

Labeodan, T., Maaijen, R., & Zeiler, W. (2013). The human behavior : a tracking system to follow the human occupancy. *Proceedings of the International Conference on Cleantech for Smart Cities and Buildings (CISBAT 2013), 4-6 September 2013, Lausanne, Switzerland* (pp. 513-518). Eindhoven: Technische Universiteit Eindhoven.

Labeodan, T., Zeiler, W., Boxem, G., & Zhao, Y. (2015). Occupancy measurement in commercial office buildings for demand-driven control applications - A survey and detection system evaluation. *Elsevier - Energy and Buildings*.

Laerd, S. (2013). *Kendall's Tau-b using SPSS Statistics.* Retrieved from Laerd Statistics: https://statistics.laerd.com/spss-tutorials/kendalls-tau-b-using-spss-statistics.php

Liang, X., Hong, T., & Shen, G. Q. (2016). Occupancy data analytics and prediction: A case study. *Building and Environment 102*, 179-192.

Marques, G., & Pitarma, R. (2016, November). An Indoor Monitoring System for Ambient Assisted Living Based on Internet of Things Architecture. *International Journal of Environmental Research and Public Health*.

Masoudifar, N., Hammad, A., & Rezaee, M. (2014). MONITORING OCCUPANCY AND OFFICE EQUIPMENT ENERGY CONSUMPTION USING REAL-TIME LOCATION SYSTEM AND WIRELESS ENERGY METERS. *Proceedings of the 2014 Winter Simulation Conference* (pp. 1108-1119). Montreal: IEEE.

Meerbeek, B., te Kulve, M., Gritti, T., Aarts, M., van Loenen, E., & Aarts, E. (2014). Building automation and perceived control : a field study on motorized exterior blinds in Dutch offices . *Building and Environment 79*, 66-77.

Microrisc. (2017). *About IQRF*. Retrieved from IQRF - Technology for wireless: http://www.iqrf.org/

Nguyen, T. A., & Aiello, M. (2012). Energy intelligent buildings based on user activity: A survey. *Elsevier: Energy and Buildings*, 244-257.

Optimus, R. (2017). *What is Cluster Analysis?* Retrieved from Research Optimus: https://www.researchoptimus.com/article/cross-tab.php

Oxford Dictionaries. (2017). Retrieved from Oxford Dictionaries: https://en.oxforddictionaries.com/definition/real_time

Rosiek, S., & Batlles, F. (2013). Reducing a solar-assisted air-conditioning system's energy consumption by applying real-time occupancy sensors and chilled water storage tanks throughout the summer: A case study. *Elsevier, Energy Conversion and Management*.

Shu, F., Halgamuge, M., & Chen, W. (2009). Building Automation Systems Using Wireless Sensor Networks: Radio Characteristics and Energy Efficient Communication Protocols. *Electronic Journal of Structural Engineering*, 66-73.

Statistics, IBM SPSS. (2015). *IBM SPSS Statistics 23 Brief Guide.* Chicago.

Statistics, IBM SPSS. (2015). *IBM SPSS Statistics Base 23.* Chicago.

Stazi, F., Naspi, F., & D`Orazio, M. (2017). A literature review on driving factors and contextual events influencing occupants` behaviours in buildings. *Building and Environment 118*, 40-66.

Strijp-S. (2017). *Strijp-S Projects*. Retrieved from Strijp-S: http://www.strijp-s.com/Projects/

Szczurek, A., Maciejewska, M., & Pietrucha, T. (2017). Occupancy determination based on time series of CO2 concetration, temperature and relative humidity. *Energy and Buildings 147*, 142-154.

Szczurek, A., Maciejewska, M., Wyłomanska, A., Zimroz, R., Zak, G., & Dolega, A. (2016). Detection of occupancy profile based on carbon dioxide concentration pattern matching. *Measurement 93*, 265 - 271.

University of the West of England, B. (2017). *Data Analysis.* Retrieved from LearnTech: http://learntech.uwe.ac.uk/da/Default.aspx?pageid=1442

Verhoeven, M., Arentze, T., Waerden, P. v., & Timmermans, H. (2017). Modelling Consumer Choice Behaviour with Bayesian Belief Networks. *Research Gate*.

Verschuren, P., & Doorewaard, H. (2010). *Designing a Research Project.* The Hague : Eleven International Publishing.

Wahl, F., Despenic, M., & Amft, O. (2012). A distributed PIR-based approach for estimating people count in office environments. *EUC 2012 : Proceedings of the 10th IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, 5-7 December, Cyprus* (pp. pp. 640-647). New York: Institute of Electrical and Electronics Engineers Inc.

Wargocki, P., & Wyon, D. P. (2016). Ten questions concerning thermal and indoor air quality effects on the performance of office work and schoolwork. *Building and Environment*, 359-366.

Webster, J., & Watson, R. T. (2002). ANALYZING THE PAST TO PREPARE FOR THE FUTURE: WRITING A LITERATURE REVIEW. *MIS Quarterly, Volume 26 Issue 2, June 2002* , 13-23.

Zhao, Y., Zeiler, W., Boxem, G., & Labeodan, T. (2015). Virtual occupancy sensors for real-time occupancy information in buildings. *Elsevier: Building and Environment*.

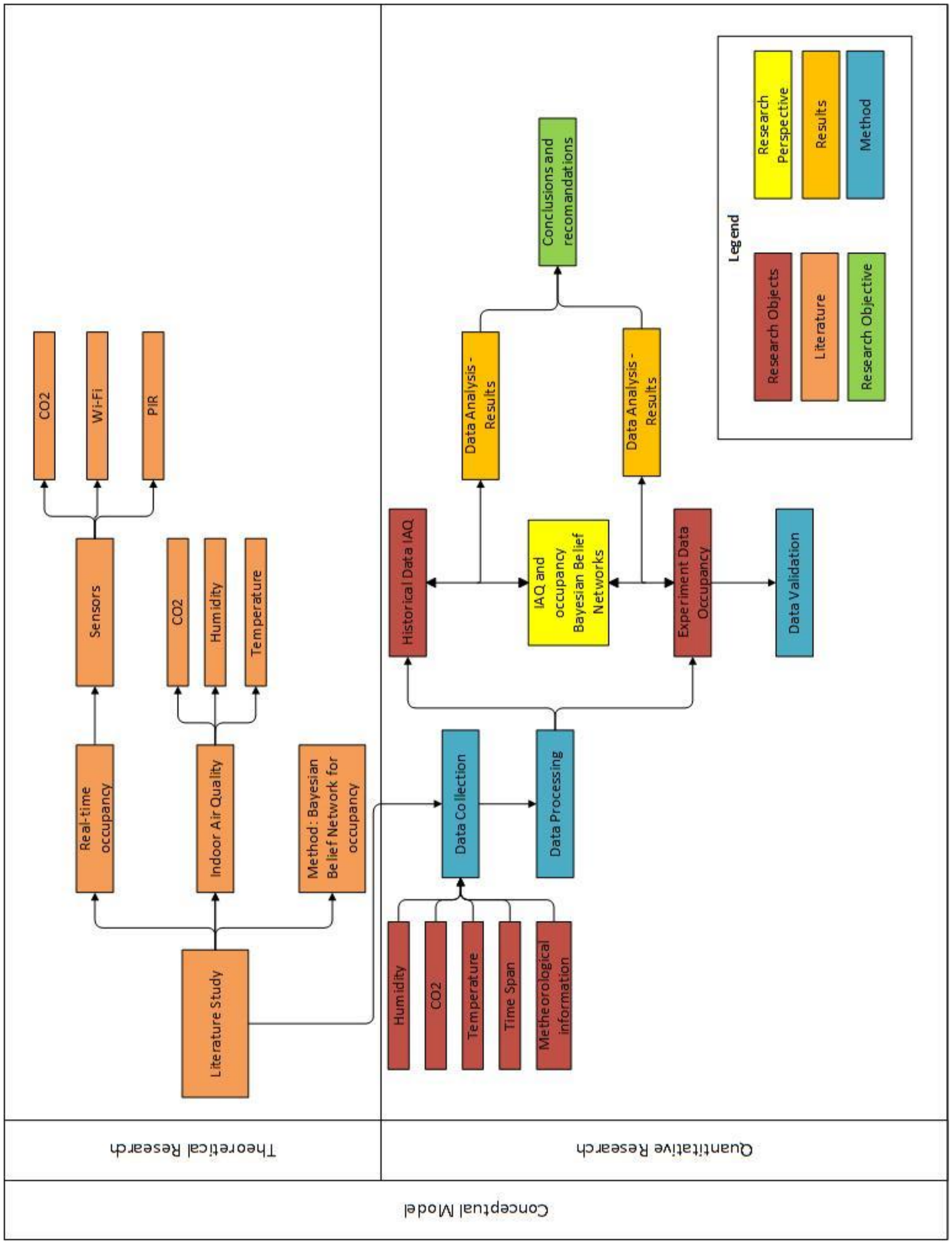
Zhao, Y., Zeiler, W., Boxem, G., & Labeodan, T. (2015). Virtual occupancy sensors for real-time occupancy information in buildings. *Building and Environment*, 11.

Zikos, S., Tsolakis, A., Meskos, D., Tryferidis, A., & Tzovaras, D. (2016). Conditional random fields - based approach for real-time building occupancy estimation with multi-sensory networks. *Elsevier: Automation in Construction*, 128-145.

Zong, F., & Wang, M. (2015, April 20). Understanding parking decisions with a Bayesian network. *Transportation Planning and Technology*, 585-600.

Appendices

Appendix 1: Theoretical Framework



Appendix 2: Research concepts

Articles		Concept							
		Real time occupancy							
	Occupancy patterns	CO2 Sensors	Temperature Sensors	Humidity Sensors	PIR sensors	EM signal detection sensors	Sensor Fusion	Bayesian Belief Network for occupancy detection	IAQ and thermal comfort
(Chau, Hui, & Tse, 2006)									x
(Kumar, et al. 2016)									x
(Wargocki & Wyon 2016)									x
(Cociorva & Iftene 2017)		x	x	x					x
(Kik, Tang, Boxem, & Zeiler 2016)									x
(Stazi, Naspi, & D`Orazio 2017)									x
(Nguyen & Aiello, 2012)		x	x						x
(Duarte, Wymelenberg, & Rieger, 2013)	x				x				
Labeodan, Maaijen, & Zeiler (2013)	x					x			
Liang, Hong, & Shen (2016)	x							x	
(Masoudifar, Hammad, & Rezaee, 2014)	x					x			
Wahl, Despenic, & Amft (2012)	x				x				
Akkaya, Guvenc, Aygun, Pala, & Kadri (2015)		x	x	x	x	x	x		
(Labeodan, Zeiler, Boxem, & Zhao, 2015)		x	x	x	x	x	x		
(Chen, Masood, & Soh, 2016)		x	x	x			x		
(Zhao, Zeiler, Boxem, & Labeodan, Virtual occupancy sensors for real-time occupancy information in buildings, 2015)					x	x	x	x	
(Zikos, Tsolakis, Meskos, Tryferidis, & Tzovaras, 2016)		x			x		x	x	
Cali, Matthes, Huchtemann, Streblow, & Muller, 2014)		x							x
(Labeodan, de Bakker, Rosemann, & Zeiler, 2016)						x			x
(de Bakker, Aries, Kort, & Rosemann, 2016)							x		
(Dodier, Henze, Tiller, & Guo, 2005)		x		x	x			x	

Appendix 3: Questionnaires for ground-truth data

Office-S - Floor 6 Credo Room 6003

Day	Date	Time Slot	Number of People	Remarks
Monday	06.03.2017	07 – 09:00	1	
		09 – 12.00	3	
		12 – 14.00	3	
		14 – 17.00	3	
		17 – 19.00	0	
		Later		
Tuesday	07.03.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	2	
		14 – 17.00	2	
		17 – 19.00	0	
		Later		
Wednesday	08.03.2017	07 – 09:00	4	
		09 – 12.00	5	
		12 – 14.00	5	
		14 – 17.00	5	
		17 – 19.00	1	
		Later		
Thursday	09.03.2017	07 – 09:00	1	
		09 – 12.00	1	
		12 – 14.00	1	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Friday	10.03.2017	07 – 09:00	1	
		09 – 12.00	1	
		12 – 14.00	1	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Saturday/Sunday	11-12.03.2017	All day		

Day	Date	Time Slot	Number of People	Remarks
Monday	13.03.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	2	
		14 – 17.00	2	
		17 – 19.00	0	
		Later		
Tuesday	14.03.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	2	
		14 – 17.00	2	
		17 – 19.00	0	
		Later		
Wednesday	15.03.2017	07 – 09:00	2	
		09 – 12.00	4	
		12 – 14.00	5	
		14 – 17.00	4	
		17 – 19.00	0	
		Later		
Thursday	16.03.2017	07 – 09:00	1	
		09 – 12.00	1	
		12 – 14.00	1	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Friday	17.03.2017	07 – 09:00	1	
		09 – 12.00	1	
		12 – 14.00	1	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Saturday/Sunday	18-19.03.2017	All day		

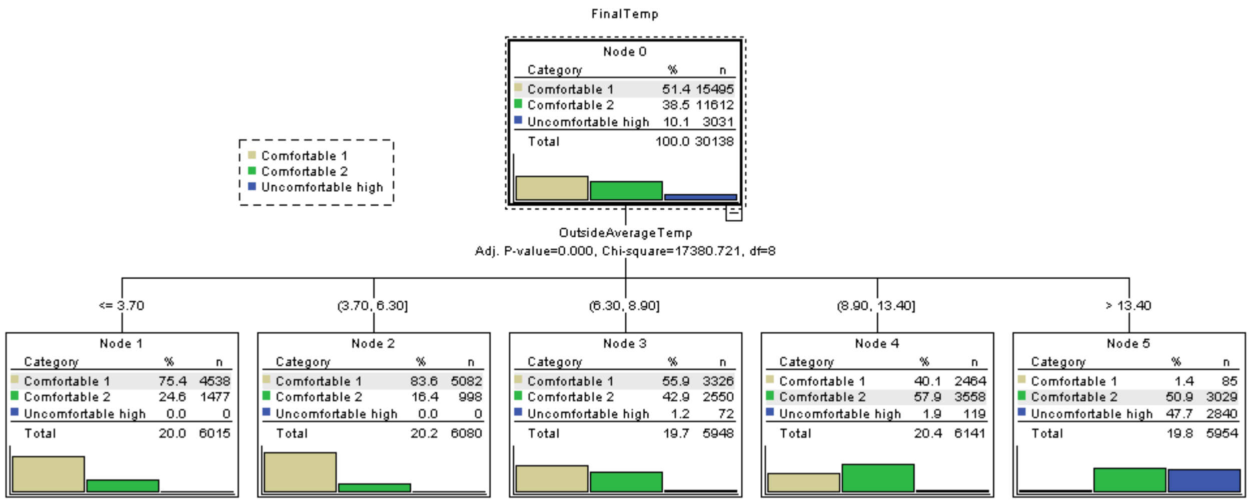
Day	Date	Time Slot	Number of People	Remarks
Monday	20.03.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	3	
		14 – 17.00	3	
		17 – 19.00	0	
		Later		
Tuesday	21.03.2017	07 – 09:00	2	
		09 – 12.00	4	
		12 – 14.00	4	
		14 – 17.00	4	
		17 – 19.00	0	
		Later		
Wednesday	22.03.2017	07 – 09:00	2	
		09 – 12.00	4	
		12 – 14.00	4	
		14 – 17.00	4	
		17 – 19.00	0	
		Later		
Thursday	23.03.2017	07 – 09:00	1	
		09 – 12.00	1	
		12 – 14.00	1	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Friday	24.03.2017	07 – 09:00	0	
		09 – 12.00	1	
		12 – 14.00	1	
		14 – 17.00	2	
		17 – 19.00	0	
		Later		
Saturday/Sunday	25-26.03.2017	All day		

Day	Date	Time Slot	Number of People	Remarks
Monday	27.03.2017	07 – 09:00	1	
		09 – 12.00	1	
		12 – 14.00	1	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Tuesday	28.03.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	2	
		14 – 17.00	2	
		17 – 19.00	0	
		Later		
Wednesday	29.03.2017	07 – 09:00	1	
		09 – 12.00	1	
		12 – 14.00	2	
		14 – 17.00	4	
		17 – 19.00	0	
		Later		
Thursday	30.03.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	2	
		14 – 17.00	2	
		17 – 19.00	0	
		Later		
Friday	31.03.2017	07 – 09:00	1	
		09 – 12.00	1	
		12 – 14.00	1	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Saturday/Sunday	1-2.04.2017	All day		

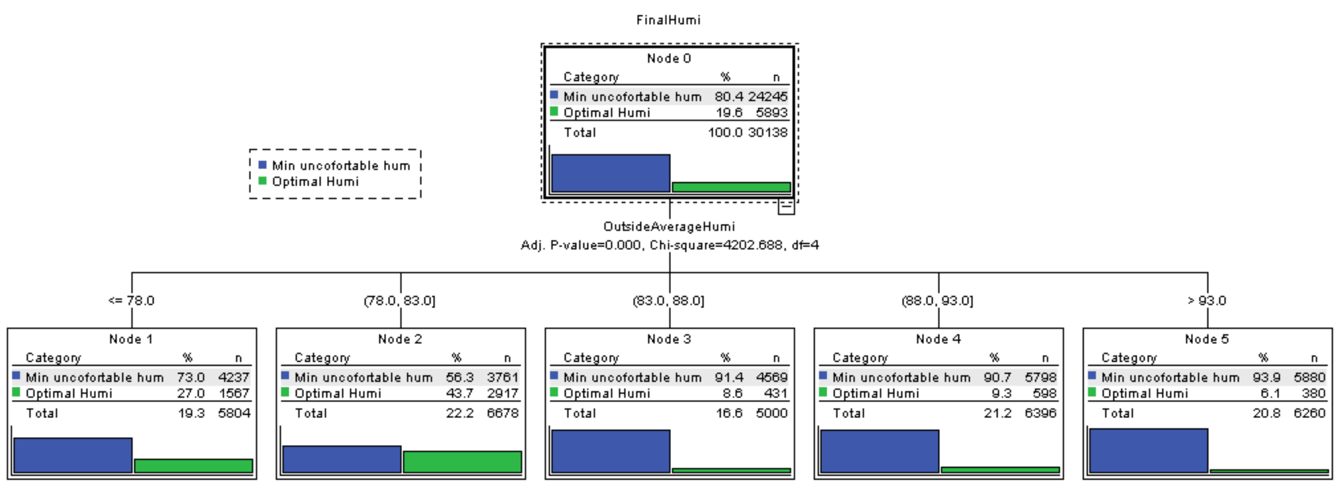
Day	Date	Time Slot	Number of People	Remarks
Monday	3.04.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	2	
		14 – 17.00	2	
		17 – 19.00	0	
		Later		
Tuesday	4.04.2017	07 – 09:00	2	
		09 – 12.00	3	
		12 – 14.00	3	
		14 – 17.00	3	
		17 – 19.00	0	
		Later		
Wednesday	5.04.2017	07 – 09:00	3	
		09 – 12.00	3	
		12 – 14.00	3	
		14 – 17.00	2	
		17 – 19.00	0	
		Later		
Thursday	6.04.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	2	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Friday	7.04.2017	07 – 09:00	1	
		09 – 12.00	2	
		12 – 14.00	1	
		14 – 17.00	1	
		17 – 19.00	0	
		Later		
Saturday/Sunday	8-9.04.2017	All day		

Appendix 4: SPSS Classification tree for weather conditions

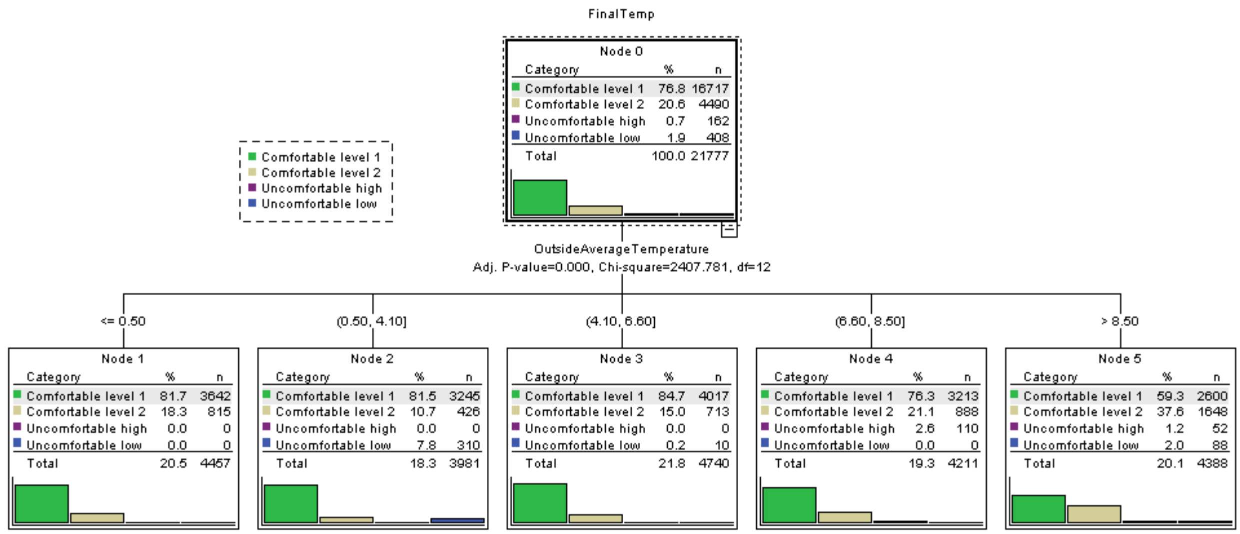
1. Credo Room: Impact of Temperature on the Outside Average Temperature splits



2. Credo Room: Impact of Inside Humidity on Outside Average Humidity splits

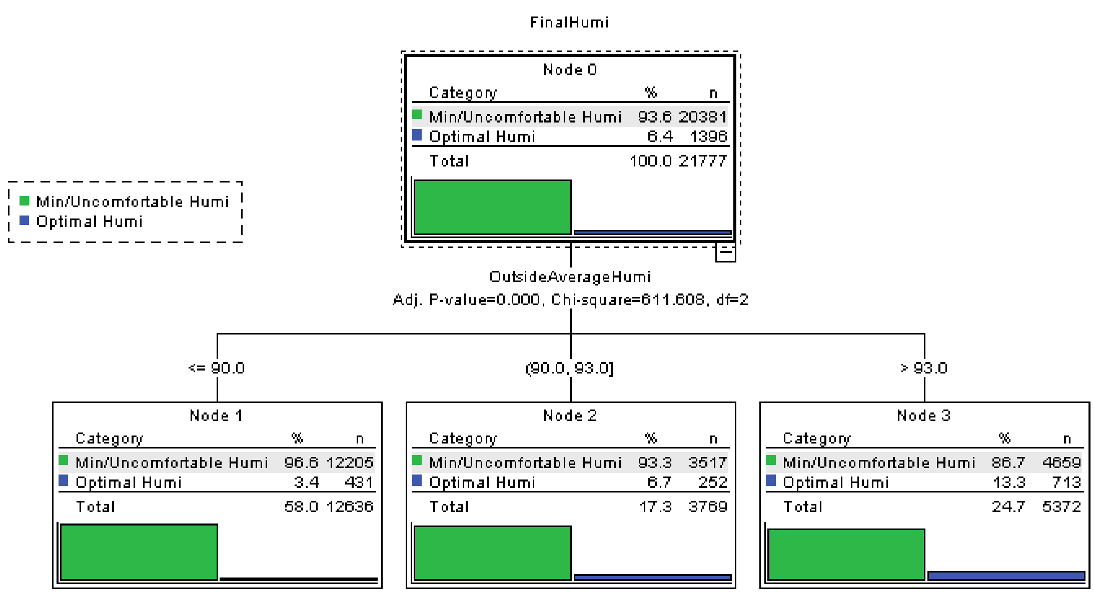


3. S2M Room: Impact of Temperature on the Outside Average Temperature splits



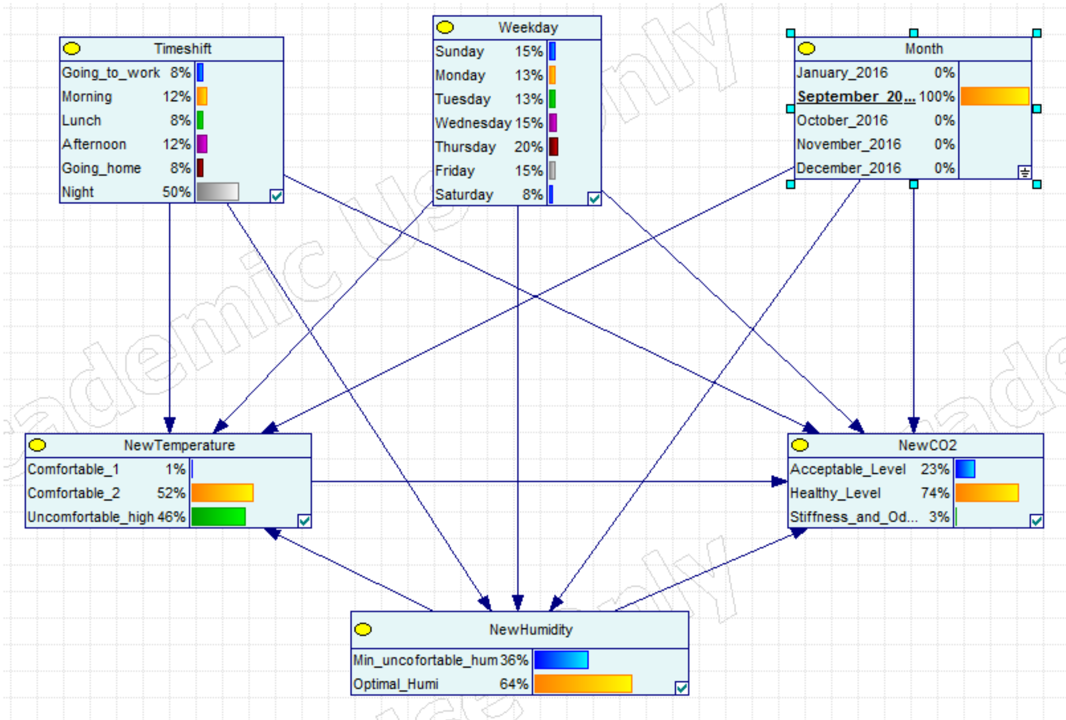
4. Impact of Inside Humidity on Outside Average Humidity splits

The last splits on humidity have only three categories, this is due to having too few observations for one of the categories in the variable. "Any category having too few observations (as compared with a user-specified minimum segment size) is merged with the most similar other category as measured by the largest p-value." (IBM SPSS Statistics, 2015)

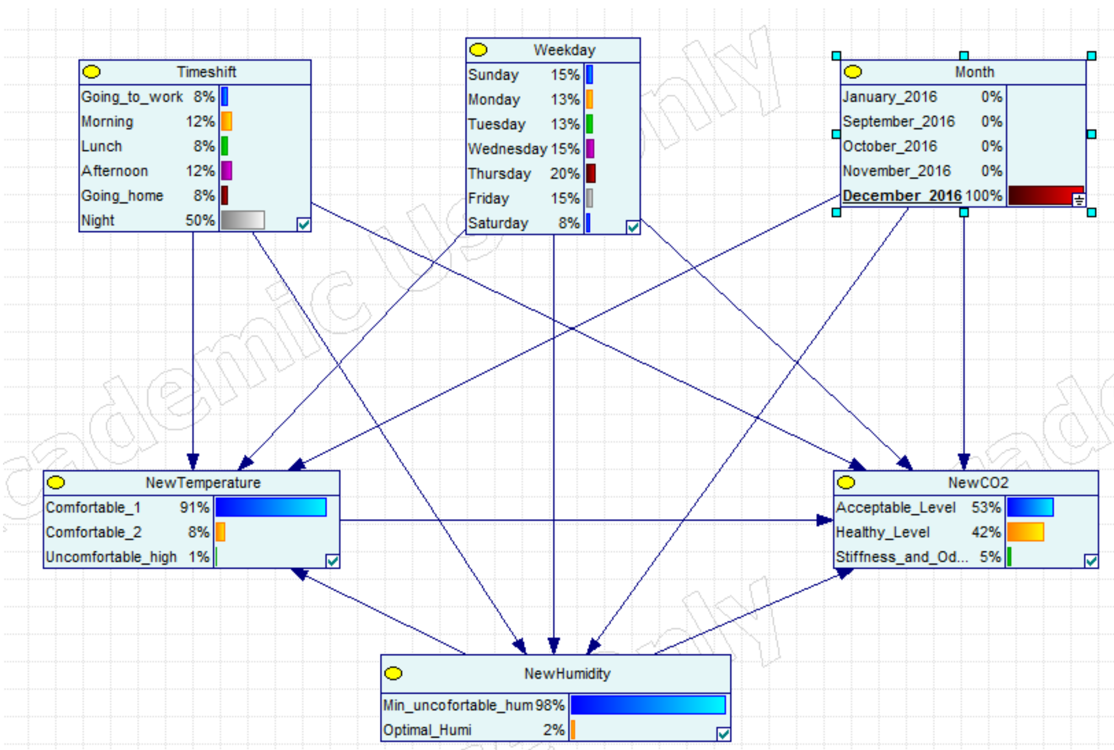


Appendix 5: Credo Room – BBN Scenarios

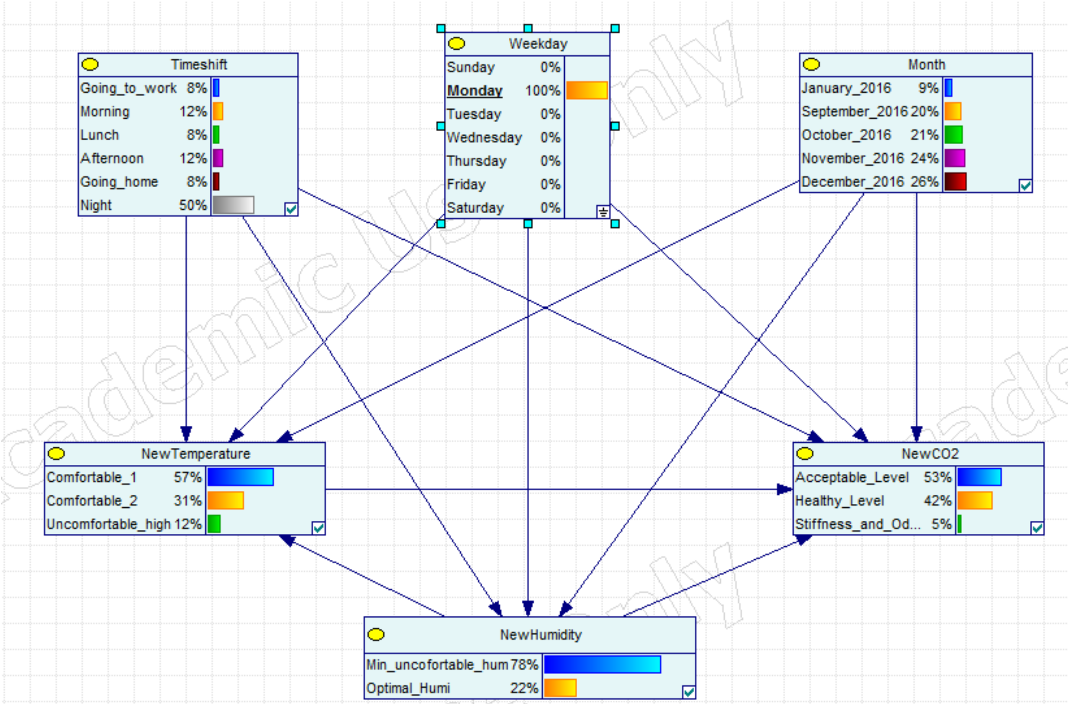
a. Scenario 1: evidence September 100%



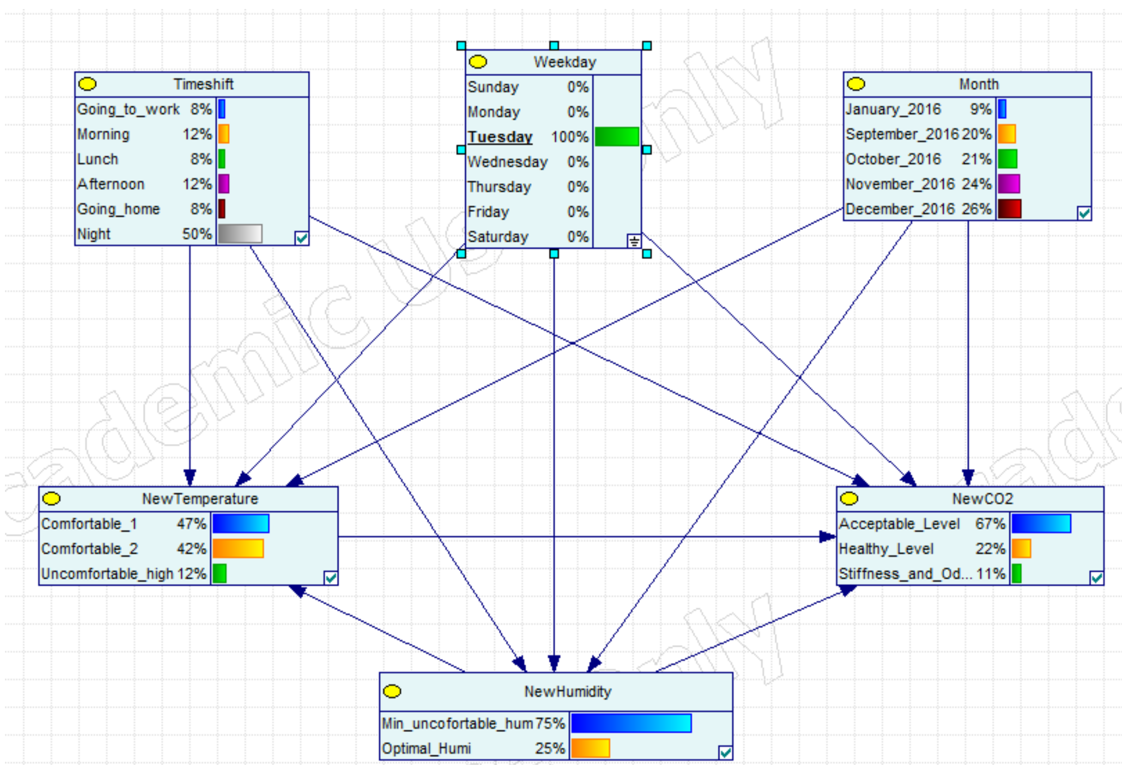
b. Scenario 1: evidence December 100%



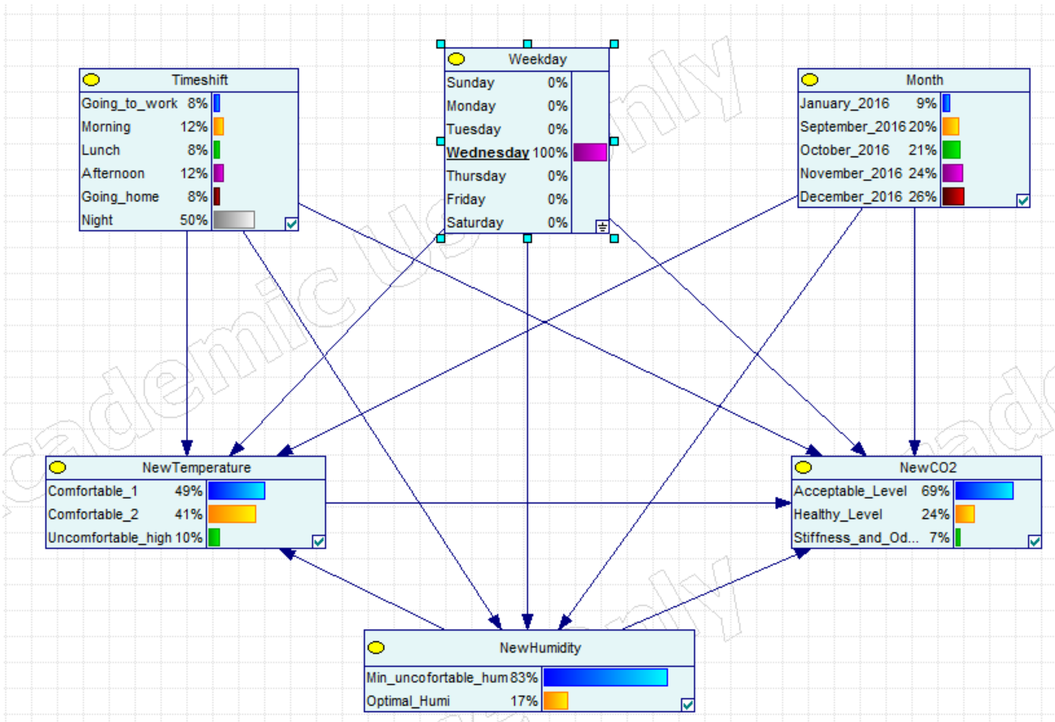
c. Scenario 2: evidence Monday 100%



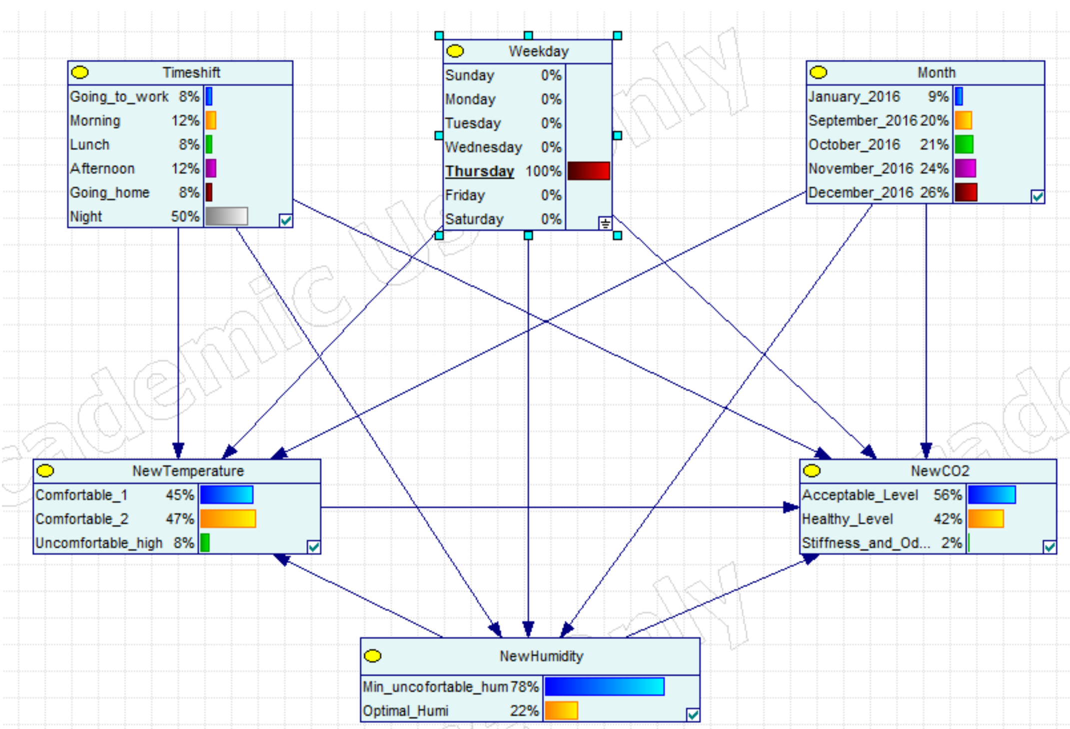
d. Scenario 2: evidence Tuesday 100%



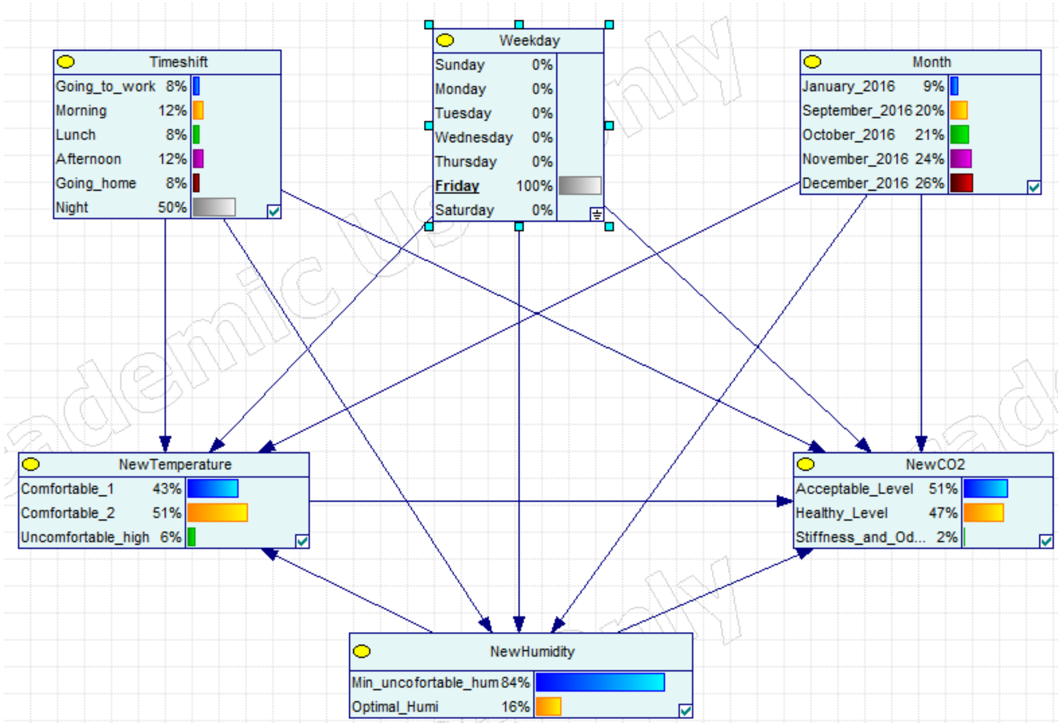
e. Scenario 2: evidence Wednesday 100%



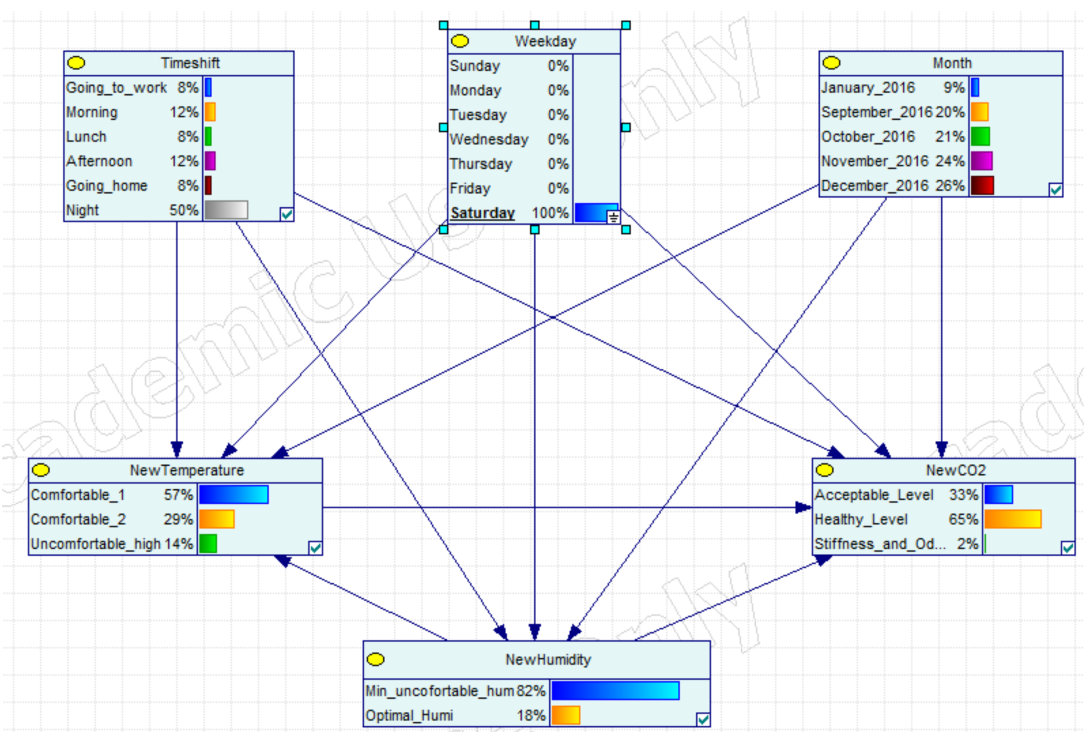
f. Scenario 2: evidence Thursday 100%



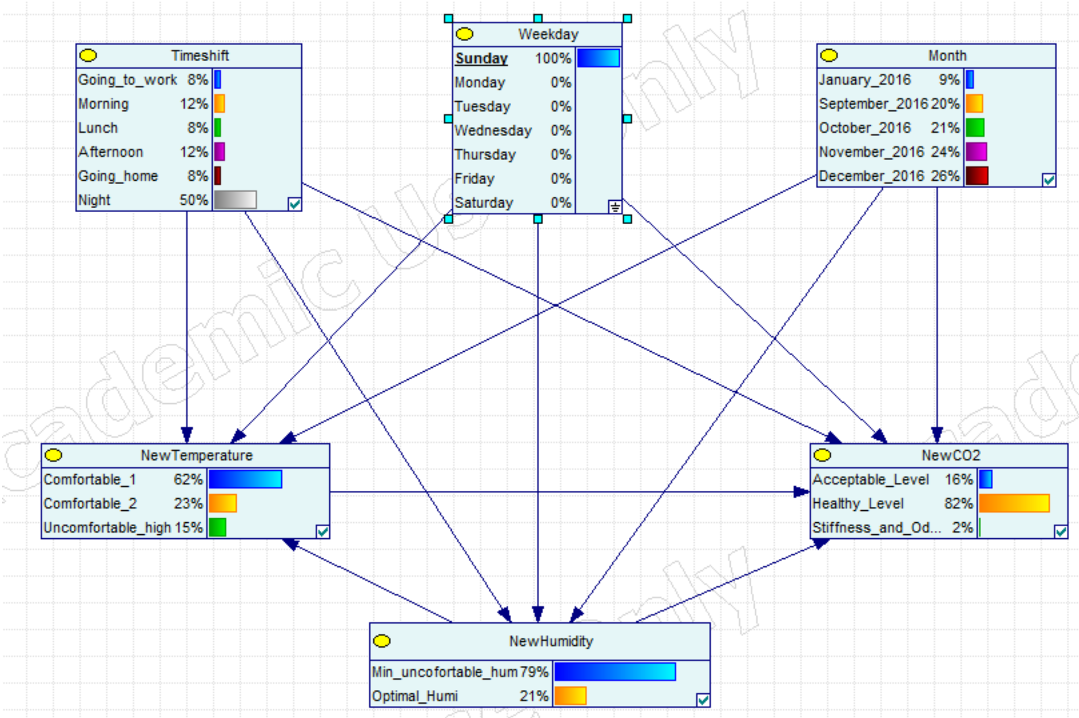
g. Scenario 2: evidence Friday 100%



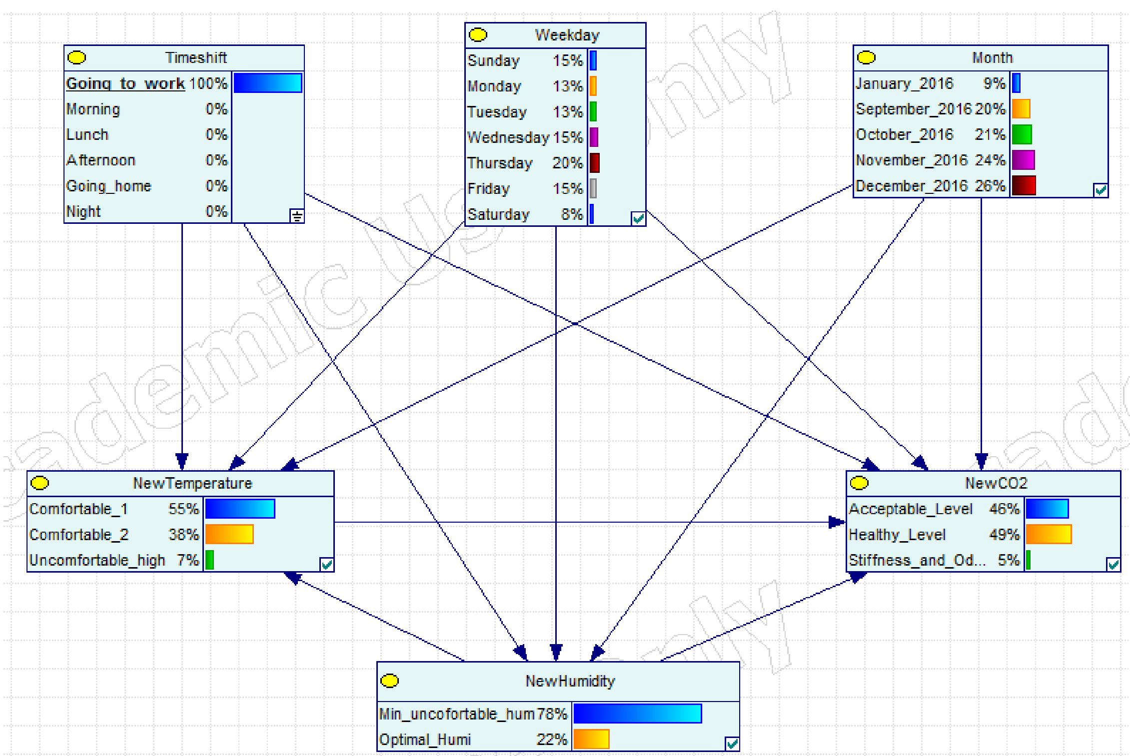
h. Scenario 2: evidence Saturday 100%



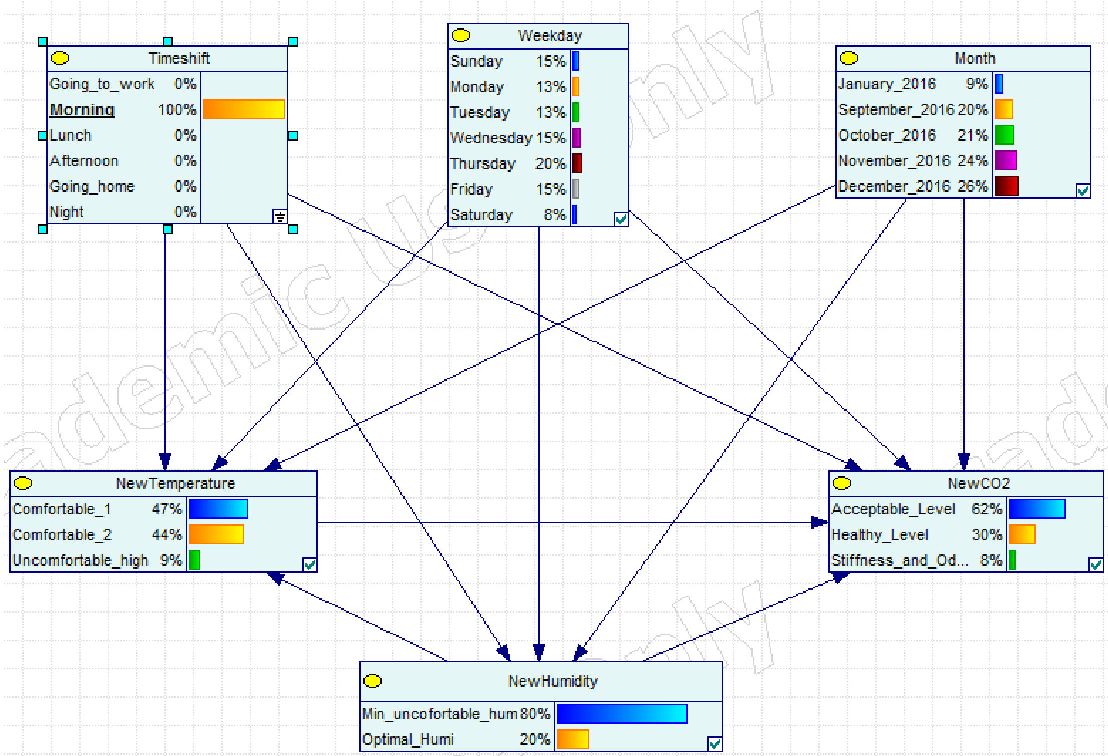
i. Scenario 2: evidence Sunday 100%



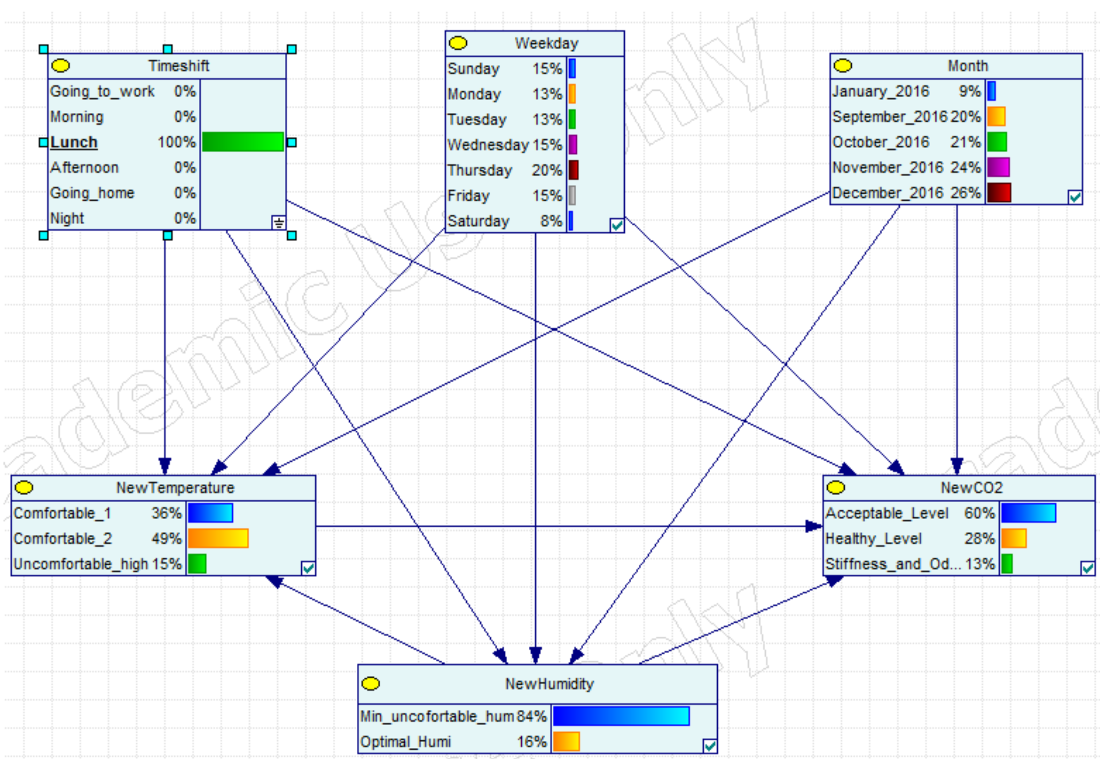
j. Scenario 3: evidence Going-to-work 100%



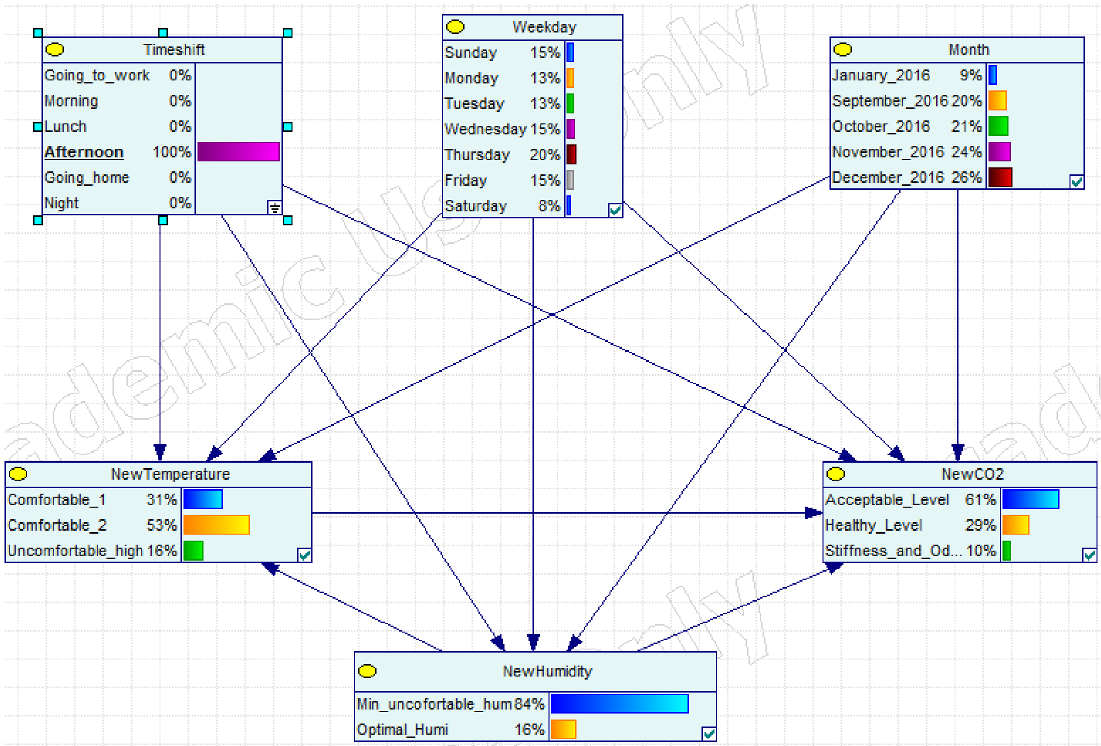
k. Scenario 3: evidence Morning 100%



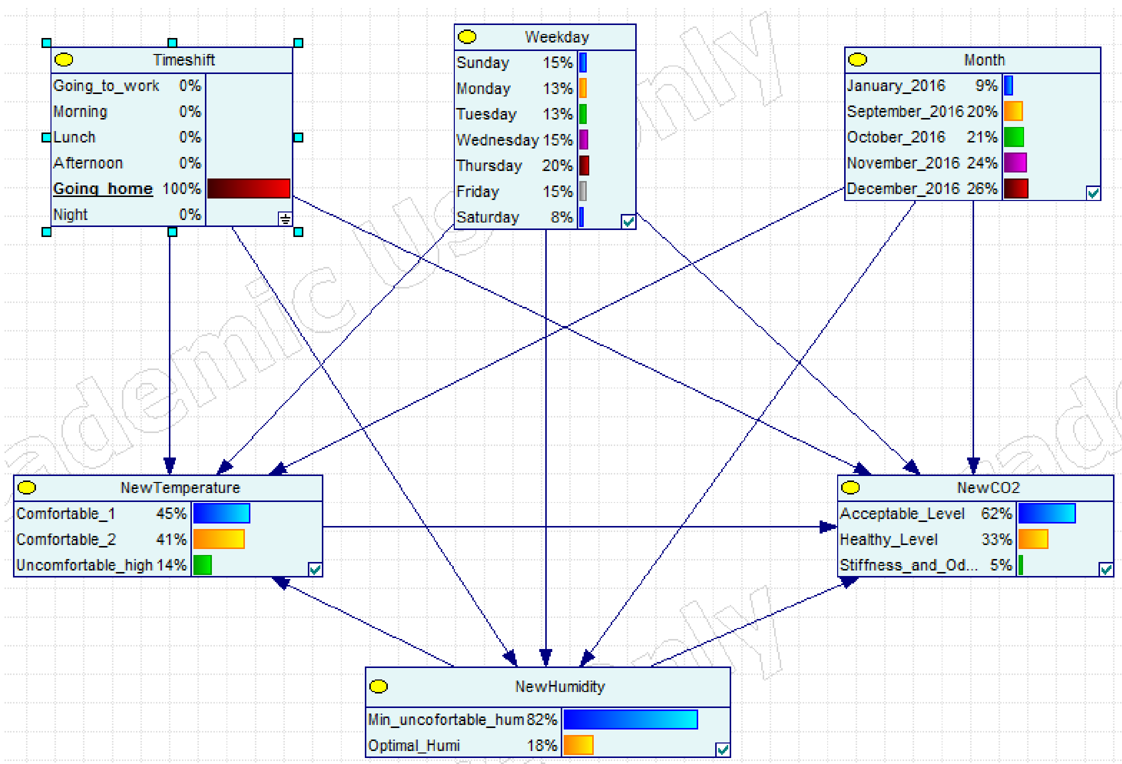
l. Scenario 3: evidence Lunch 100%



m. Scenario 3: evidence Afternoon 100%



n. Scenario 3: evidence Going-home 100%



o. Scenario 3: evidence Night 100%