Kåre Birger Lapstuen

Using historical data to predict future occupancy for improvement of indoor environmental factors

A case study at NTNU

Master's thesis in Datateknikk
Supervisor: John Krogstie
February 2019

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Kåre Birger Lapstuen

Using historical data to predict future occupancy for improvement of indoor environmental factors

A case study at NTNU

Master's thesis in Datateknikk
Supervisor: John Krogstie
February 2019

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Summary

This master's thesis explores the current indoor environmental conditions in selected rooms intended for teaching and learning at NTNU and suggests a neural network based model to predict future occupancy.

The first part of the study investigates how the current HVAC and temperature control systems behave under different circumstances. Based on a large number of sensors in different rooms, environmental variables are logged and compared across time and room types. The study shows that the environmental variables at NTNU are largely good, but that CO2 levels remain too high for certain times of the week, and that this may have a negative effect on the students' learning abilities.

In the second part of the thesis, a neural network based model to predict future occupancy is proposed, implemented and tested. The model is able to predict occupancy before it occurs with good accuracy and is able to adjust its predictions based on current sensor readings and historical data. This represents an promising alternative to the current system used at NTNU.

Sammendrag

Denne masteroppgaven undersøker innendørsklimaet på utvalgte undervisnings- og læringsarealer ved NTNU og foreslår en modell basert på nevrale nettverk for å forutse fremtidig rombruk.

Den første delen av studien undersøker hvordan det nåværende systemet for kontroll av ventilasjon og temperatur oppfører seg under forskjellige situasjoner. Basert på et stort antall sensorer i forskjellige rom blir variabler for innendørsmiljøet logget og sammenlignet på tvers av romtyper og over tid. Studien viser at innendørsmiljøet på NTNU stort sett er bra, men at CO2-nivået tidvis er for høyt på enkelte rom. CO2-konsentrasjonen er til tider så høy at det kan påvirke studentenes evne til å lære.

I den andre delen av studien foreslås, implementeres og testes en modell basert på nevrale nettverk for å forutse fremtidig rombruk. Modellen er i stand til å gjøre dette med god nøyaktighet og kan tilpasse sine prediksjoner basert på gjeldende sensoravlesninger og historiske data. En slik modell utgjør et lovende alternativ til det nåværende systemet som er i bruk ved NTNU.

Acknowledgments

I would like to thank the following persons for their great help during the project

- John Krogstie for general guidance, help with refining the scope as well as helpful pointers to relevant literature and persons to contact.
- Kristin Klubbo Brodahl for putting up with late night programming, less-than-interesting dinner table discussions on the finer details of loss functions and an ever growing tower of printed articles, excerptions and drafts slowly filling up the guest bedroom.
- Tor Leikvoll (Telenor) for providing me data, updates on the progression of the Telenor pilot project as well as allowing me to participate in the installation of the sensors. This gave a much needed context for the sensor data, as well as first hand knowledge of sensor and mounting specifics.

K.B.L.

Contents

Chapter 1

Introduction

1.1 Background and motivation

In the fall of 2018, Telenor started conducting a pilot study for IoT at the Norwegian University of Science and Technology (NTNU). Several different rooms were outfitted with IoT sensors from Yanzi measuring CO2, temperature, Volatile Organic Compound (VOC), sound level, humidity, human motion, occupancy and people's movement through identified choke-points. The data collection happens as a part of a project at NTNU, investigating how rooms can be retrofitted and changed to better suit modern ways of teaching students. Collecting data on indoor climate and room usage helps build a better understanding of how students interact with the rooms and how the indoor climate responds to students' presence and outside factors. The data collected from these senors are stored and were made available to the author of this thesis so that that new useful insights could be extracted from them.

1.2 Current HVAC and temperature control system at NTNU

NTNU has a central control system called *sentralt driftsanlegg* (SD). The SD controls lights, ventilation, cooling and heating. There is little public information on the system, but NTNU-internal documentation shows that it is made up of a combination of mainly Honeywell and Schneider based systems. [1] This system has access to a varying amount of sensors embedded in ventilation shafts, rotary encoders and other sensors. Telenor (2019)'s pilot study report points out some of the shortcomings of the current system. These include a lack of sensor data across different sub-systems and the fact that the system is hard to extend and program. In addition there is no unified dashboard to make sensor readings easy to see and act upon.

The system mainly relies on hard-coded on and off times and operates on a reduced setting during non-peak hours and weekend days.

[1] `https://fuglane.it.ntnu.no/pages/viewpage.action?pageId=722009#SD-anlegg(SentraltDriftsanlegg)-Forvaltning/administrasjon` - Internal documentation not available outside the IT department

1.3 Research goals

The two research goals (RG) presented below are the starting point of this thesis, and form the basis of the work carried out.

Two research goals (RG) were defined based on background and motivation as well as information about the current HVAC system. These are as follows:

- **RG1 - Investigate current indoor conditions.** How conducive to learning are the current indoors conditions at NTNU. Are the registered parameters within guidelines and in accordance with new research?
- **RG2 - Create predictive models based on historic data.** In what way can historic data be used to model future behaviour(s) and how can these models be used to improve the indoor learning space?

1.4 Thesis Outline

The outline of the rest of this report is as follows:

- **Chapter 2** defines a baseline for good indoor conditions based on available research.
- **Chapter 3** introduces the methodology used as a base for the thesis.
- **Chapter 4** details the particular sensors used in the setup, as well as their placement at NTNU.
- **Chapter 5** presents information on the dataset, the processing on data and any problems encountered with sensor faults.
- **Chapter 6** shows how NTNU performs in terms of indoor environmental factors as laid out in **Chapter 2** and points out some interesting trends in the data, both in terms of differences between week days, and some observations on the efficiency of the heating and HVAC system.
- **Chapter 7** details how a predictive model of occupancy was constructed using a convoluted neural network trained on the data presented in **Chapter 5**.
- **Chapter 8** sums up the findings for the two research goals and lays out some of the implications.
- **Appendix B** contains a short and simplified introduction to neural networks, in an effort to aid readers in understanding the solution presented in **chapter 7**.

Chapter 2

Indoor environmental factors, acceptable limits and their impact on learning

Indoor environmental factors such as temperature, humidity, CO2 (carbon dioxide), sound levels and VOC (volatile organic compound) affect the people using and staying inside buildings. Research has shown that certain factors such as CO2 can have detrimental effect on school performance (Myhrvold et al., 1996) and that high VOC levels can cause health issues within relatively short amounts of time (Pitten et al., 2000). It is therefore paramount that buildings are built and operated in ways that ensure that these factors are kept within acceptable ranges, and that readings are taken to ensure that actual real-life variables are within acceptable levels.

Today new buildings must be built to certain standards, Norway has a standard known as TEK17[1] governing among other things ventilation. However, less stringent standards are placed on old buildings still in use, or spaces in old buildings converted into new uses such as conversion from meeting rooms into office spaces.

This chapter aims to establish acceptable limits for the environmental variables collected at NTNU (**CO2, VOC, temperature and humidity** based on available research.

CO2, temperature and humidity will be measured based on their impact on student performance. Special focus will be given to studies measuring these variables impact on student's or pupils' performance since NTNU's main function is that of an university. Two different intervals/levels are defined; **acceptable** which entails that little to no negative impact on health or learning is found, **detrimental** where there are some proven detrimental effects on learning but no ill health effects.

Due to a lack of detail in Yanzi's manual as to exactly what VOC are measured, no safe or unsafe levels will be presented. VOC will be discussed in general terms.

The chapter is broken down into subsections, each pertaining to a specific variable among the sensor readings at NTNU.

[1] https://lovdata.no/dokument/SF/forskrift/2017-06-19-840

2.1 Temperature

Temperature and its effect on the human body is easily observed and felt by people. The human body performs relatively well over a large temperature interval, and there is little reason to believe that any real-life temperatures inside an office or university building will be outside safe levels in terms of low temperatures, excluding any specialized rooms such as walk-in freezers. As such, no lower limits for the **unsafe** level will be used for the temperature variable in the rest of this thesis.

One study (Park, 2017) found that students' performance on exams is detrimentally affected by high outside temperature. Park (2017)[p.37] found that an increase of 18 degrees Fahrenheit in outside temperature reduced test scores by 4.5 percent.

Wargocki et al. (2005) explored the link between decreased room temperature and air supply rate and their effects on school work performance on children. This was done by measuring the change in speed of mathematical operations (subtraction) as well as acoustic proof-reading and reading comprehension. By keeping the error rate constant, the change in speed could be measured while ensuring that speed did not affect the correctness of the students' work. Reducing the average room air temperature from 23.6°C to 20°C had a measurable and significant positive effect on study performance. Wargocki et al. (2005)[table 2] showed an increase in subtraction speed of 28 percent, a reduced error rate of 10 percent for proof-reading and an increase of 24 percent in speed of reading and comprehension.

Typing speed is also negatively impacted by heat stress. Wyon (1974) showed that both typing speed and work output was significantly increased at temperatures of 20 °C compared to 24 °C. This particular study was carried out on typewriters, but it is fair to assume that similar effects will be seen when using computers.

The studies show that high temperatures reduce student performance, both for tests with a varied content (Park, 2017), typing speed (Wyon, 1974) and mathematics and reading (Wargocki et al., 2005). Based on the results obtained across the cited studies, the following temperatures and their impact on students and workers was devised:

A temperature range between 20 °C and 22 °C will be deemed ideal, and a temperature range above 22, but below 24 °C will be classified as acceptable. Temperature ranges outside the ideal and acceptable ranges will be deemed as having a detrimental effect on student performance.

2.2 Air humidity

While air humidity poses few health effects by itself, it influences other variables directly and indirectly. High levels of air humidity drive spore growth, spread biotics (Baughman and Arens, 1996) and may induce problems for people with dust mite asthma (Andersen and Korsgaard, 1986). Elevated humidity levels may also cause long-term structural issues in school buildings, despite the air conditioning being operated correctly (Bayer et al., 2002).

Arundel et al. (1986) found that the majority of ill-health effects can be minimized or even removed completely by ensuring that air humidity is kept between 40 and 60 percent. Andersen and Korsgaard (1986) found that a relative humidity under 45 percent at 20-22 °C reduces the amount of airborne dust mite to very low levels.

Too low levels of relative air humidity can also cause problems. Uchiyama et al. (2007) shows that reduced relative air humidity causes increased evaporation of the tear film and may trigger the sensation of dry eyes for some people.

Since the available literature shows some ill effects at very low levels of relative humidity (sub 30 percent) and there are few apparent positive health effects and reduced airborne pollutants below 40 percent, a range of **30-50 percent relative humidity** will be used as a the preferable interval for relative air humidity. This interval should ensure that the humidity stays low enough to reduce the number of airborne irritants, and yet remain high enough to reduce the risk of dry eyes and other adverse physical reactions. This interval also allows for seasonal differences caused by differences in humidity of outdoor air in summer and wintertime.

A relative humidity outside the range of 35-50 is deemed unacceptable as it may cause respiratory problems in some students, cause irritation of the eyes as well as pose long term problems for the structure of the building.

2.3 CO2 concentration

The concentration of CO2 is a major driver when designing new buildings and their HVAC systems. The largest source of increased CO2 concentration in school rooms is the occupants and their exhaled breath with concentrations ranging from 35.000 PPM to 50.000 PPM (Prill, 2000). The CO2 concentration in school rooms has been shown to have a major effect on students' and pupils' performance at different mental tasks (Satish et al., 2012).

A study (Corsi et al., 2002) conducted at several Texas elementary schools showed median CO2 concentration of 1,268 ppm and a peak concentration of 2,062 ppm. Peak C02 concentration exceeded 1,000 ppm for 88 percent of the classrooms in the study.

As shown by Satish et al. (2012) in figure 2.1, the room CO2 concentration can have a dramatic effect on students' performance. The study was based on the subjects' performance in nine categories:

1. **Basic activity:** the numbers of actions taken within a period of time
2. **Applied activity:** the number of opportunistic actions
3. **Focused activity:** strategic actions
4. **Task orientation:** the focus on concurrent demanding tasks
5. **Initiative:** the development of new activities during the test
6. **Information search:** the ability and openness to search for information

7. **Breadth of approach:** the flexibility in approaching a task
8. **Basic strategy:** the number of strategic actions taken

Most tasks show a statistical significant reduction in performance under conditions where the room CO2 concentration is only about 600 ppm above the outdoor background concentration of about 400 ppm (Monastersky, 2013) . Increasing the CO2 concentration above 1000 ppm to 2500 ppm leads to further reductions in student's performance, and in some tests (initiative and basic strategy) the performance drops into dysfunctional levels (<25 percentile).

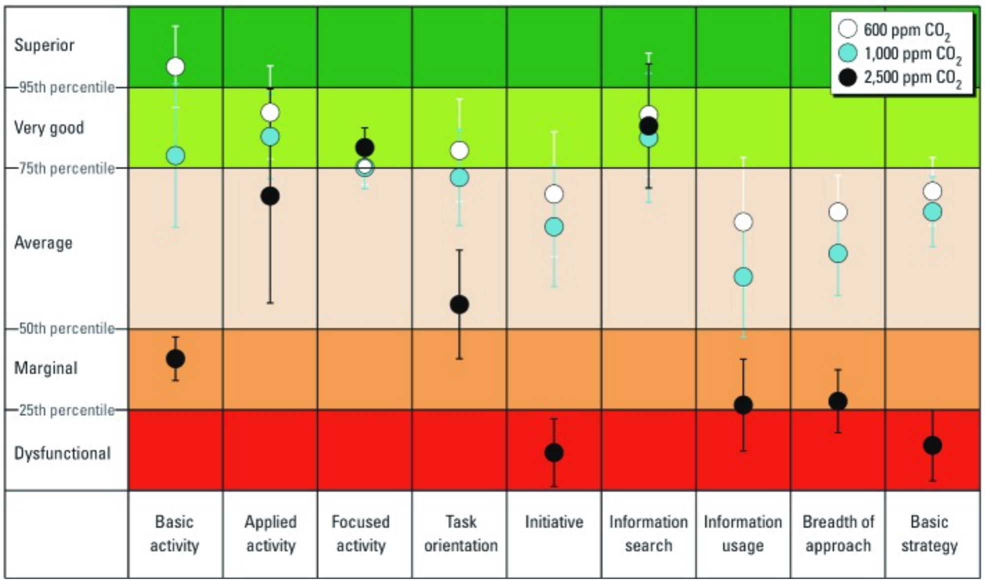

Figure 2.1: CO2 concentration in ppm and its affect on several mental tasks (Satish et al., 2012)

While there are statically significant reductions in student performance across certain tasks at as low levels as 1000 ppm, the reductions are relatively small and should not impact student's ability to learn the material.

Based on these studies, a **CO2 concentration of <1000 PPM** will be used as the preferable interval. CO2 concentrations between 1000 and 2000 will be deemed as having a detrimental effect on student performance, and even higher levels will be viewed as having unacceptable impact on student performance.

2.4 VOC concentration

Volatile Organic Compounds (VOC) are organic chemicals whose vapour pressure at ambient room temperatures makes them prone to turn into a gaseous state. VOCs in the indoor climate can come from many sources, i.e human breath (Phillips et al., 1999), newly painted surfaces (Wieslander et al., 1996), off-gassing from building materials (Wolkoff, 1995), other sources such as room fresheners, mothballs and consumer products Guo (2011) as well as from the outside air. VOC concentrations in rooms are therefore the sum of emmisions from outside the room entering through ventilation and windows, as well as the emissions produced within the room itself. Some VOCs are known to have detrimental effect on long term health if present in large enough concentrations (Jones, 1999).

The Yanzi user manual does not indicate what VOCs are measured by the COM sensor. It is therefore impossible to ascertain whether nor not the measured concentration poses any short- or long-term health problems to occupants of a given room. No detrimental and safe ranges will be given for VOC in this thesis, as it is not indicated what VOCs are measured, and the relative concentration needed to cause negative health impacts vary wildly between VOCs. The measurements of VOC are still included because they show changes over time, and such serve as a point of comparison between rooms. The relative concentrations between rooms thus become the interesting factor, and not the absolute readings.

Chapter 3

Methodology

This chapter will explain the reasons behind and elaborate on the methodology used in this thesis. The use of quantitative data vs qualitative data is discussed and the reader is presented to the execution of the project.

3.1 Research approach

To ensure that research was performed in a proper and productive way, the methodology proposed by Peffers et al. (2007) was used as a basis for the work carried out. This methodology can be broken into six sub-tasks, each defining a set of rules. In addition each sub-task describes possible research entry points. The problems to be solved were defined based on 1.3. Central in the methodology proposed by Peffers et al. (2007) is the concept of an *"artifact"*. An artifact is understood to mean IT products/solutions that can be applied to a known problem or well-defined set of problems. In the case of this thesis work, the artifact produced is a system that uses neural networks to predict future human occupancy in a room. The six activities are as follows:

- **Activity 1: Identify problem and motivate.**. The problem is defined; what are its boundaries, the nature of the problem. The output of this sub-task/activity is to show the importance of solving the given problem. Why is it worth spending time and resources attempting to solve or better describe this particular problem?
- **Activity 2: Define Objective of Solution.** How would the artifact help resolve the problem, what does it accomplish?
- **Activity 3: Design and development.** A novel IT system is developed based on the identified problems and motivations. The IT system is implemented to ensure that the objective of the solution is met. The output of this activity is the IT system itself, *artifact*.
- **Activity 4: Demonstration.** Use the created artifact to solve a problem, showing that the solution works in the given context.

- **Activity 5: Evaluation.** Observe the efficiency and efficacy of the solution. Does it work? Could it work better? Iterate on the design.
- **Activity 6: Communication.** Publish the results, in order to create and share disciplinary knowledge.

These steps can be visualized as shown below. Note the iterative process and the possible research entry points.

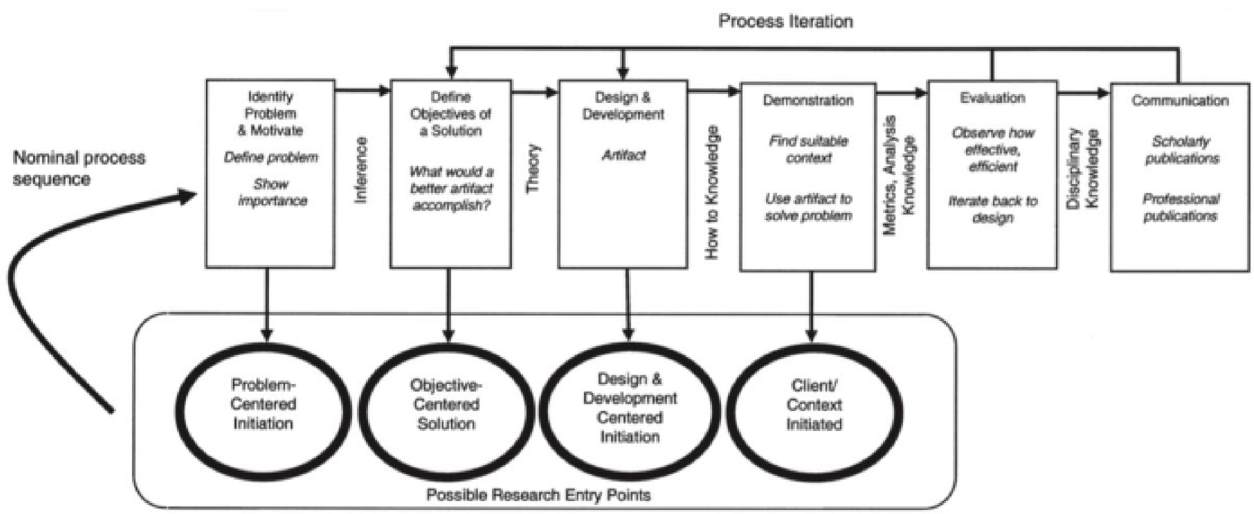

Figure 3.1: Design Process Model (Peffers et al., 2007)

In addition to using Peffers et al. (2007)s work to shape the research approach for designing and testing the predictive system, a data-driven methodology was used as a basis for the analysis of data.

3.2 A quantitative and data-driven methodology and design

Because the dataset this thesis is based on consists of sensor readings from different rooms over a period of time, it makes sense to approach the problems through a quantitative approach. The research referred to in 2 makes it possible to extrapolate qualitative insights based on quantitative data; namely the indoor air quality's, as measured by the sensors, effect on student performance and well-being.

McAfee et al. (2012) describes the mindset of a data-driven company motivated by "what do we know", "what data do we have", and not by "what do we think". This touches on perhaps the most important aspect of data-driven methodologies. The available data is used as a starting point to extract knowledge and new information as opposed to using preconceived notions, hunches and instincts as the starting point. Eriksson et al. (2010) describe how the gene linked to the photic-sneeze reflex [1] was uncovered based on 23AndMe's extensive database of customer-supplied genome and a web-questionnaire. The questionnaire allowed the participants to be divided into ever-expanding cohorts of self-selected individuals. By comparing the genome of these cohorts, traits that were common across the cohort could be correlated to specific genes only present in the cohort. This is one example of using existing data as a starting point to new discoveries.

This thesis aims to extract new information from the existing dataset, namely to establish how good the indoor climate is at NTNU. While the users of the building may have some hunches based on their own qualitative experience of using the building, the analysis of quantitative data may lead to a more robust understanding of the indoor climate, as well as how previous research indicates that these factors affect the occupants. The analysis of the dataset and the chosen methodology can therefore be said to constitute a data-driven approach.

Due to time-constraints and scope of this thesis, no qualitative data was collected from the occupants of the rooms included in the dataset. This leads to the loss of an important dimension, namely how the occupants subjectively rate and experience the indoor environment. While it would be both useful and interesting to correlate the qualitative data from occupants with the quantitative data, it would fall outside the scope of this thesis. It does however represent an interesting avenue for future work.

3.3 Data acquisition and research goals

Since the methodology is largely data-driven, it is paramount that enough high quality data is collected to ensure that there is a good dataset to draw conclusions from. Data was collected with high frequency, primarily to ensure a wide dataset, but this also enables different techniques in post-processing to remove noise, erroneous readings and other problems.

In some of the rooms, more than one sensor of each type were installed at different points in the room. This has several distinct advantages, such as increased total dataset size, redundancy in case of sensor failure. It sldo allows measuring environmental factor gradients such as temperature differences inside the room and gives a better understanding of how different parts of a room are used. The data acquisition from several different rooms within roughly the same time period enables a greater understanding of how different rooms differ in their environmental factors such as CO2 concentration, temperature and other relevant factors. All this is needed to create a better understanding of the indoor environment at NTNU, **RG1**

In order to create a model to predict future occupancy, **RG2**, a complete time-series of occupancy had to be recorded. This allows the creation of a model which accurately captures and predicts future occupancy based on historic trends, current occupancy and relevant factors such as time of day.

[1] A sudden increase in the light level may cause some people to sneeze in response to the light.

Chapter 4

Sensors and installation

4.1 Sensors and installation

This chapter details the sensors installed at NTNU and gives relevant information about the particular installation in the different rooms studied in this thesis.

4.1.1 Sensor background

All the sensors installed at NTNU are from the Swedish company Yanzi, which specializes in smart buildings and collecting sensor data using an IoT architecture [1] and post-processing

Telenor installed four different Yanzi sensors at NTNU. Comfort, Motion+, Footfall Camera and Presence Mini. In addition, a Yanzi gateway (GW) was installed in each room, working as an 4G internet gateway allowing sensor data collected from the sensors to be passed on to Yanzi's server-side architecture. The sensors communicate with the GW through a star-topology. In the event of a communication error between a sensor and the GW, the sensor itself can store a backlog of sensor reading to pass onto the GW once communication is resumed. The GW can also store a backlog of sensor readings if its internet connection is lost. Most sensors and the GW operate either solely on battery power, or have a built-in battery reserve to ensure that they can operate even when the electricity provided by the grid is turned off. This ensures that the data collection is robust, and can handle both intermittent power and communication loss.

[1] urlhttps://www.yanzi.se/solution/

4.1.2 Sensor details

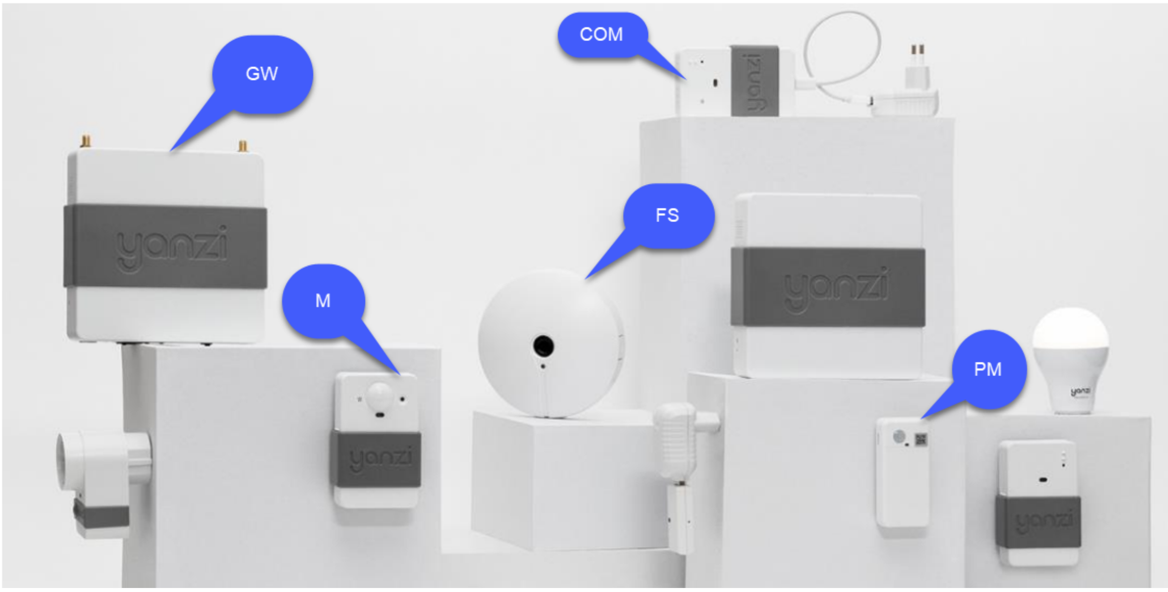

Figure 4.1: Sensors installed. GW: Gateway, PM: Presence Mini, Com: Comfort, M: Motion + and FS: Footfall sensor

Presence Mini

The presence mini (PM) sensor is a small rectangular box mounted on the underside of tables using a double-sided tape. The PM is battery operated, and has an expected running time of approximately 12 months [2]. The PM uses a built-in temperature sensor to measure the ambient room temperature, as well as to detect if a person is sitting directly below it. One PM has to be mounted for each seating place/chair that is to be monitored.

The PM transmits two data variables to the GW; **temperature** (in kelvin) and **occupancy** (binary, 0 or 1). If the occupancy bit is set to "1", the sensor is indicating that someone is seated under the sensor at the time of transmit. 0 denotes no detected person.

Due to time constraints and technical difficulties, no data from the PM sensors were used in this thesis.

Comfort

The Comfort (COM) sensor is a small rectangular box mounted on the side of a wall using double-sided tape. The COM is powered using a wall adapter providing 5V over a USB Type-C cable. The COM has no built-in battery backup and will seize collection and transmitting of data upon power-loss.

[2] Yanzi user manual

The COM records VOC (volatile organic compound, PPB), CO2 (PPM), temperature (Kelvin), sound pressure (decibels), barometric pressure (hPa) and humidity (relative humidity in percentage). The Comfort sensor automatically calibrates every 24 hours to ensure correct CO2 readings. In order to do this, the sensor assumes that the lowest measured CO2 concentration is 400 PPM and calibrates accordingly. The Yanzi user manual stresses that the CO2 readings may be incorrect if the sensor is placed in an area where the CO2 concentration remains elevated over the expected background concentration of 400 PPM over a whole 24 hour period.

The barometric pressure data was not used in this thesis, as most buildings and rooms are not built to keep constant indoor air pressure. The sound pressure readings were also disregarded for most rooms, as sound readings can vary wildly from second to second. Averaged sound readings are used as a proxy measure for motion event readings for in cases where no M+ sensor was installed.

Motion+

The Motion+ (M+) sensor is a small rectangular box mounted on the side of a wall using double sided tape. The sensor needs to be mounted in such a way that it has a clear line of sight to the area of movement to be recorded. The sensor has a 30 degree angle sensor, and thus has a wide field of view. The sensor is battery powered, and has a battery life of up to 5 years, depending on the amount of movement registered. The M+ sensor records the number of motion events that have occurred since the last time data was recorded.

Footfall sensor

The Footfall sensor (FS) is a round, roof mounted sensor equipped with a camera and is powered using Power-over-ethernet (PoE) Cat5 cable. The FS uses the camera and artificial intelligence (AI) to count the number of people entering or leaving a pre-defined area. The activation area is defined using a web interface, and the camera is used in combination with the AI to not only count the number of people passing the camera, but also what direction they are moving. This gives it superior accuracy (Kim et al., 2002) (Velipasalar et al., 2006) to conventional infrared systems usually employed to count the number of people passing through a section of space (Kajala, 2007). Using direction/vector information, the FS can record information on the direction to travel, and not just people crossing in general.

No data from the FS sensors were included in the data-set supplied by Telenor, and no data from the FS sensors are therefore used in this thesis.

4.2 Polling frequency and other details

The polling frequency varies depending on the sensor type. The table below details the polling frequency, as well as how the sensor is powered. All battery powered sensors have a running time of up to 5 years, but battery life depends on usage patterns.

Sensor Name	Polling frequency (s)	Power
Presence Mini	30-120	Built-in battery
Comfort	30-120	Wall outlet
Motion+	30-120	Built-in battery
Footfall	30-120 [a]	Power over ethernet (PoE)

[a]The footfall sensor sends two values, one for the number of entry events and one for the number of exit events

Most of the sensors in every room had their polling frequency set to 60 seconds, though the dataset indicates that there may be some differences in how the sensors were configured with regards to polling frequency. This could be caused by

4.3 Sensor installation and room configuration

This section describes the installation of each sensor in every room used in this thesis and the placement of individual sensors and the configurations of the rooms themselves. All the rooms examined a this thesis are part of a pilot project at NTNU that aims to investigate and evaluate alternatives to the old classroom and auditorium settings. Existing rooms and auditoriums were retrofitted with new audio and video equipment, new setting areas and the rooms were reconfigured.

The following symbols are used for sensors: Blue F - Footfall sensor, Green GW - Gateway, Red M - Motion+ sensor, Blue A - Comfort sensor, Purple P - Presence sensor.

4.3.1 Smia

Smia is located at the Gløshaugen campus, and consists of a remodelled room that was previously used for meetings, group assignments and other group-based activities [3]. One of the rooms' walls make up the facade of the building the room is in, and is therefore equipped with several windows. This may affect the temperature ranges measured for the room, as solar energy will heat the room, and opening of windows by occupants may decrease the temperature and affect the CO2, VOC and humidity readings.

Smia was equipped with two COM sensors, a single M+ sensor and one footfall camera to track occupant traffic in an out of the room. No data from the footfall sensor was included in the dataset used in this thesis.

[3]https://www.ntnu.no/laeringsarealer/smia

Figure 4.2: Smia after remodelling, picture courtesy of NTNU

Smia features several large oval-shaped tables, each with 6-8 chairs. Each table is equipped with audio and video equipment, allowing the users of the table to project images and video from their personal computers onto a whiteboard mounted at each table. The whiteboard can also be used as a conventional dry-erase board.

Smia is used for lectures, group assignments and by student's discussing the curriculum in study groups.

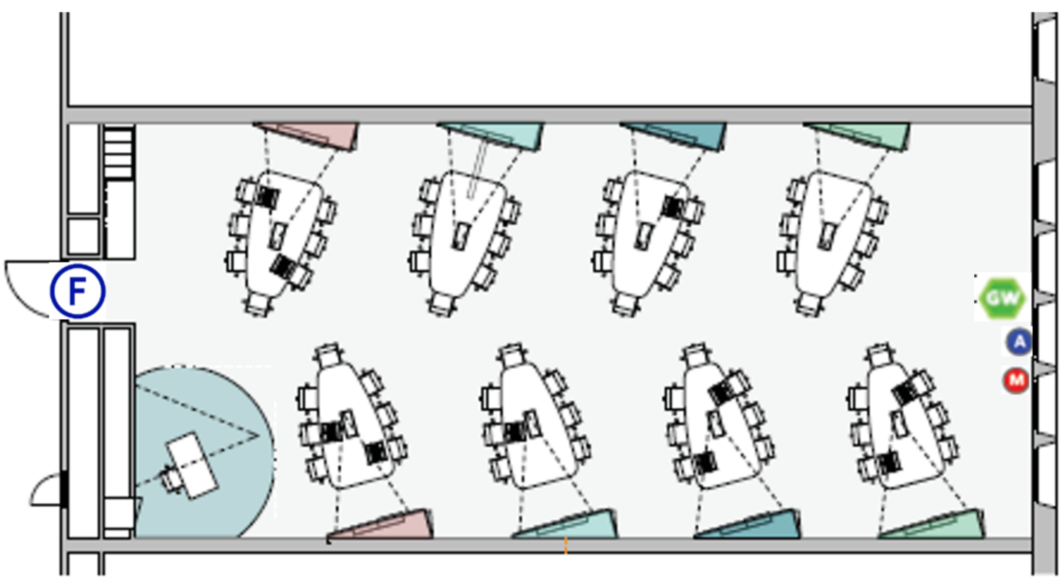

Figure 4.3: Room layout of Smia and approximate sensor placement

4.3.2 R2

R2 is used to investigate new takes on the old tried and tested auditoriums commonly seen in universities across the world. [4]

No data from R2 was included in the dataset used as a basis for this thesis. The room description is therefore included solely due to its inclusion in the pilot project to study innovative learning spaces.

Figure 4.4: R2 in use. Picture courtesy of NTNU

Seating has been changed from the old layout consisting of staggered rows of chairs into different levels consisting of large oval tables and chairs surrounding these. Video screens have been mounted at each table, allowing students to plug in their personal computers to display images and videos on these. In addition, the lecturer can control the screens and project the same image and video on all screens in the auditorium. This is commonly used during lectures, where the lecturer can draw and explain on his or her screen or tablet and have the pictures broadcast live to all screens. This replaces the old black/white boards usually seen in lecture halls and auditoriums.

R2 was fitted with a single footfall sensor, registering the number of people entering and leaving the room through the main door. In addition two COM sensors were installed at different points in the room, registering indoor climate data such as temperature, CO2 concentration, VOC concentration, sound pressure, barometric pressure, humidity and light levels.

[4] https://www.ntnu.no/laeringsarealer/r2

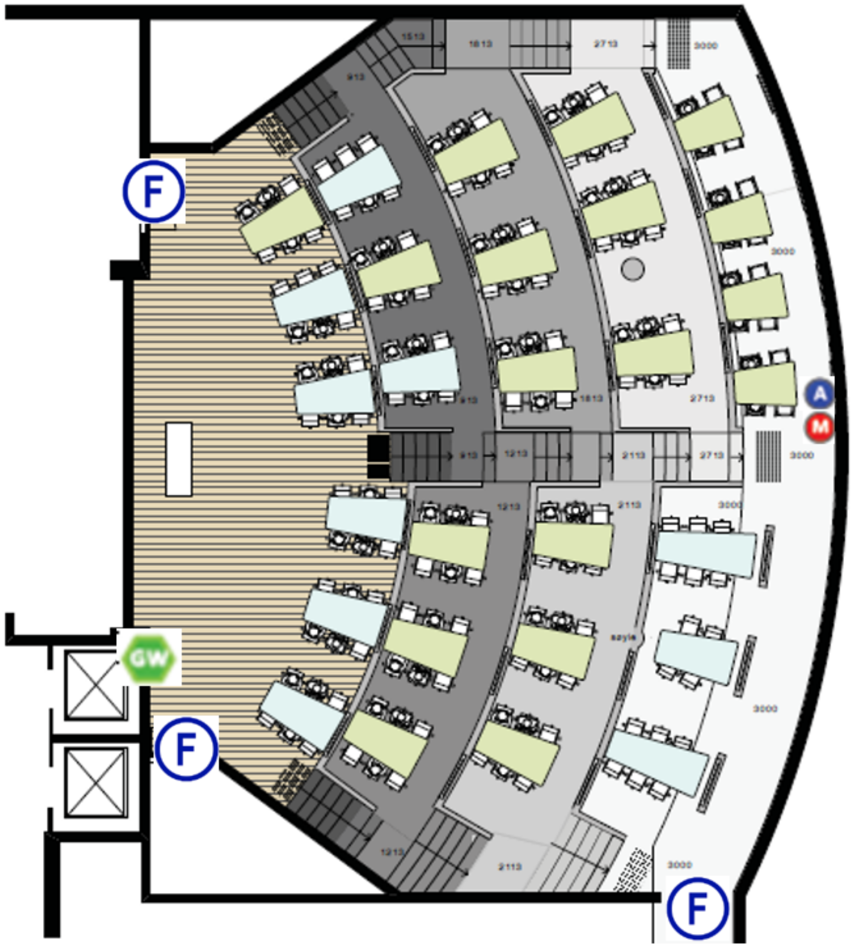

Figure 4.5: Room layout of R2 and approximate sensor placement

4.3.3 Sandkassa

Sandkassa was constructed to test the same layout and equipment as Smia, but was constructed at the Dragvoll campus [5]. Like Smia, Sandkassa was equipped with three COM sensors, three M+ and three footfall sensors. In addition, three motion sensors were mounted in the room.

No data from the footfall sensor was in the dataset used as a basis for this thesis, and the presence sensor data could not be used due to technical difficulties.

[5] https://www.ntnu.no/laeringsarealer/sandkassa

Figure 4.6: Sandkassa's present layout. Picture courtesy of NTNU

Like Smia, Sandkassa is equipped with large oval tables, each with 6-8 chairs. Students can project images and video from their personal computers onto the whiteboards present at each desk.

Sandkassa's areas of usage are similar to Smia, consisting of zones for group work and lectures.

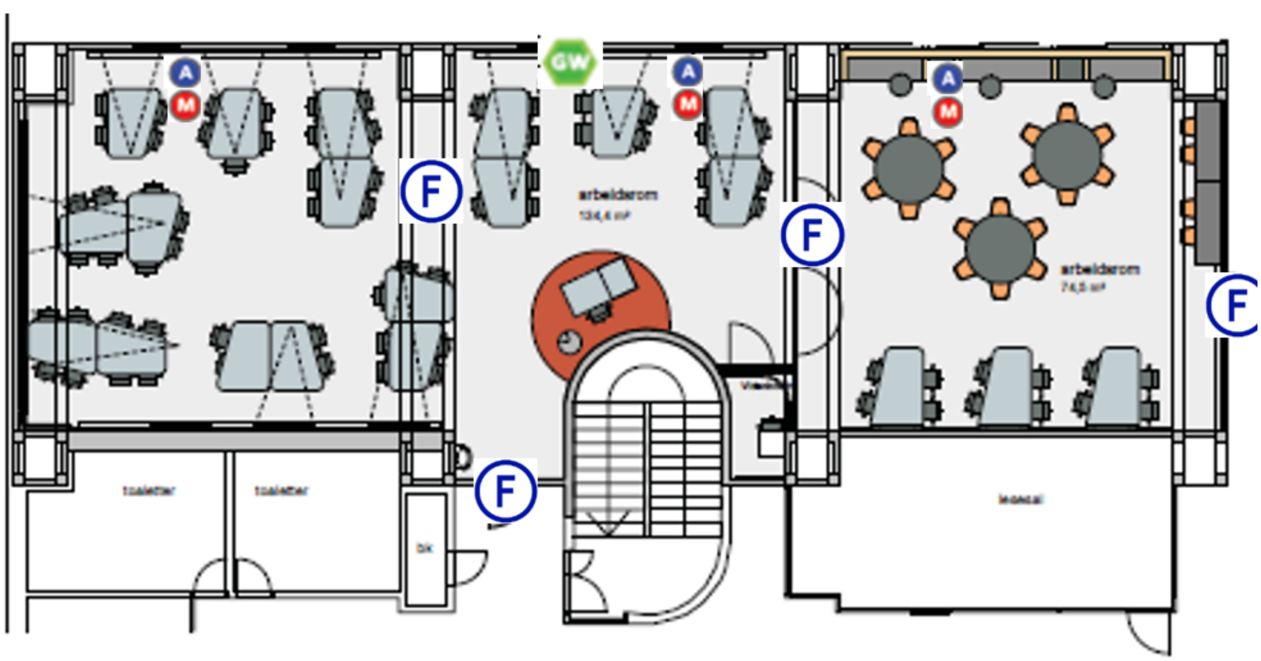

Figure 4.7: Room layout of Sandkassa and approximate sensor placement

4.3.4 Koopen

Koopen is a work area inside a large structure known as "Glassgården". Glassgården consists of a series of buildings linked together with large walls and roofs consisting of glass and metal framing, reminiscent of greenhouses.

Koopen is placed inside one of the atriums and is mainly used by students when performing project work and group assignments in subjects such as electronics and cybernetics. Koopen's seating area consists of several large tables, each with 8-10 chairs.

Figure 4.8: Seating areas in Koopen. Picture courtesy of NTNU

The large glass covered walls and roofs of glass allow large amount of sunlight to enter over the course of the day. This can cause some issues during the day, especially during the summer, as Koopen is prone to getting quite hot. The opposite problem is observed in the winter, as the large glass surface conducts heat better than standard walls, and heat loss occurs during hours of little to no sun.

Students have complained about cold temperatures during the winter, and additional heating ducts have been installed in an effort to improve the working conditions during the winter. Koopen was entered into the pilot study partly to evaluate the effectiveness of these measures.

Figure 4.9: Outside view of the atrium Koopen is located in. Note the large glass surfaces.

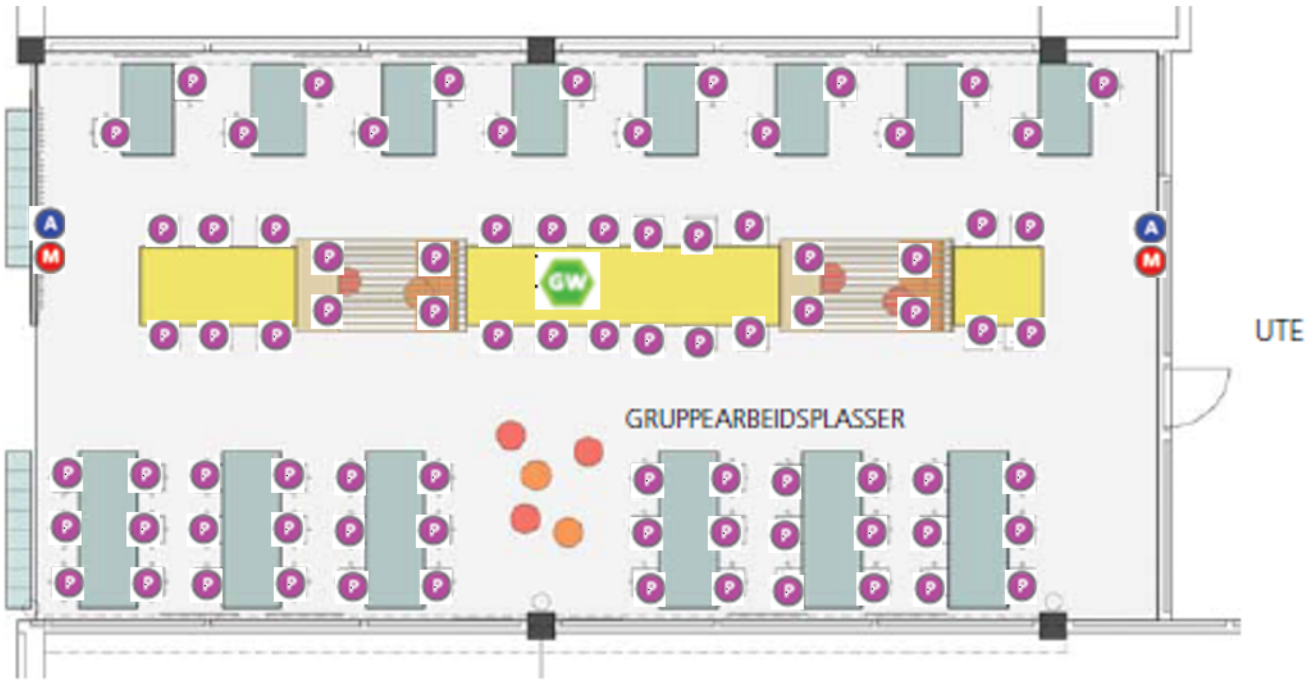

Figure 4.10: Room layout of Koopen and approximate sensor placement

4.3.5 U1

U1 consists of a combination of different room types and caters to a wide array of use cases. U1 covers approximately 1000 square meters and includes a maker workshop containing 3D printers and soldering stations, a large reading room consisting of a re-purposed library, four smaller rooms set aside for group work and a large lounge with couches and seating arrangements.

Since U1 hosts such a varied selection of work spaces, the analysis of sensor data is broken down into two sections: small group rooms and the maker space. The large reading room has been omitted as several other room types of similar layout has been covered in this thesis. The maker space and the smaller group rooms on the other hand present an unique type of room not studied elsewhere in this thesis. This allows identification of differences in indoor environmental variables depending on room layout, occupancy and the type of activity in the room.

Figure 4.11: Layout of U1 and the different rooms and work spaces

Chapter 5

Dataset and data processing

The data used in this thesis is comprised of a complete data dump obtained from Telenor, for the period 09.10.2018 to 29.11.2018. The dataset includes all sensor readings from all the installed sensors in every room in the pilot study. Due to some intermittent sensor problems and a gradual roll-out, the amount of data and the period of data collection vary from room to room.

All of the data manipulation was done using Python 3.6 and a series of commonly used Python frameworks, such as Pandas [1]; a framework designed to read, manipulate and extract data from formatted textfiles and SciPy [2]; a framework for mathematics, science and engineering. All plotting of data was done using the framework Matplotlib [3].

5.1 Data dump and formatting

The data dump was in the form of a single .CSV file, totalling 978 MB and 6.6 million lines of sensor readings. Each line represent a single sensor reading from a single sensor. Below is a sample of the first 4 lines of the data dump.

RowID	Sensor Value	Timestamp	PropertyID
0	294.03	2018-10-29T15:12:57.147Z	5bebf9df0bbbb500081e0ccf
1	294.03	2018-10-29T15:13:22.769Z	5bebf9df0bbbb500081e0ccf
2	1	2018-10-29T15:13:22.779Z	5bebf9df0bbbb500081e0cd0
3	294.2	2018-10-29T15:14:24.031Z	5bebf9df0bbbb500081e0ccf

The RowID is simply the zero-index row number of the data dump. The sensor value is the raw data from the sensor. RowID 0, 1 and 3 are CO2 readings (in PPM). RowID 2 is the motion counter for a M+ sensor, indicating that there has been a single motion event detected since the

[1] https://pandas.pydata.org/
[2] https://www.scipy.org/
[3] https://matplotlib.org/

last sensor reading. The Time stamp indicates the year, month, day, hour, minute and millisecond the reading was taken. The PropertyID represents the sensor's unique ID, which allows us to isolate data readings on a per-sensor basis.

Please note that the first rows are from the initial initialization of the sensors, so the amount of data and polling frequency are rather low.

5.2 Environmental variables and their inertia

The data in the dataset is comprised of what I refer to as slow/high inertia variables and fast/instantaneous variables. The temperature of a room is an example of a slow variable; turning up the heating setting on an electric radiator does not instantly raise the temperature of the room. Due to the thermal mass of the room and the air inside the room, the systems experiences a slow rise in temperature until the target temperature is reached. Turning on the light switch, on the other hand, instantly changes the light level in the room; the variable is instantaneous. Readings of such variables only represent a snapshot of the system's state with regards to that variable at a single point in time. A sensor reading of a slow variable is less dependent on the timing of the reading, as such variables change slowly over time.

Variable	**Slow/Fast**
Light	Fast
Temperature	Slow
CO2	Slow
Humidity	Slow
VOC	Slow
Sound	Fast

Table 5.1: Variables and their response time

It is important to keep this in mind when analyzing the dataset. Sensor readings of a fast variable may not be representative of the average or aggregate sum over a period of time.

Chapter 7.4 details how we can use predictive models to better adjust slow variables such as temperature and CO2, to not only increase comfort but also save energy and reduce running costs.

5.3 Data transformation

While most of the data from the data dump is usable in its raw form, some sensor data has to be transformed into a more useful state. The temperature output of the COM sensor is in °Kelvin, which is converted into °Celsius by adding 273.15 to the sensor reading and storing the new value. The sound-pressure data from the COM has to be divided by 10,000 to get the corresponding decibel reading.

Further inspection of some of the sensor readings indicates that there is some noise in some of the readings, especially for the PM sensor. The PM sensor has a tendency to indicate that there is someone sitting on a chair for a prolonged time, before suddenly indicating that the chair is empty for a single sensor reading, before going back to indicating that the seat is occupied. While it is possible that the occupant simply left the chair for a short amount of time, the number of such occurrences is so high that it seems to be mostly caused by a sensor glitch. This is discussed further and a fix is proposed in 5.4.

5.4 Resampling sensor data

Sensors are polled for data in a round-robin way, where each sensor sequentially transfers data to the GW. This causes some issues, as data may be registered at the GW somewhat later than the time the readings took place, and data that is temporally close may spread out due to differences in polling time. A temperature reading may for example arrive at the GW 1 minute later than the CO2 reading of the same room, but due to these variables being of a slow type, we can group them together in a temporal matter.

Grouping data with close temporal likeness can be very useful, especially when the data is run through neural networks. Since many of the variables are "slow"/have a high degree of inertia (as explained in 5.2), there is little reason to look at second-by-second changes.

Pandas' built-in resample function [4] was used to resample the dataset. All sensor variables were resampled into 10 minute bins, using either the *sum* or *avg* function.

The sum function was used on **motion count** data. This helps eliminate the problems regarding noise in PM sensor readings, as previously pointed out. A single non-correct reading's impact is reduced when the readings are summed up over a longer time period. The sum function works by adding all the readings seen during the resampling period, and combines these into a single number.

The *avg* function was used on the rest of the variables and helps reduce sensor noise by averaging the readings over a longer time period. This also makes the fast variable readings much more descriptive as to the long-term value of a given variable, as no single reading is used by itself and is instead averaged out with temporally close readings. The average function averages all the readings for the resampling period, producing a single value representing the average sensor reading for the time period.

Tables 5.2 and 5.3 show what the data looks like before and after resampling. Resampling the data into larger intervalls is one way of reducing the impact of sensor noise and erronous readings.

[4] https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.resample.html

Timestamp	Sensor reading
2000-01-01 00:00:00	1
2000-01-01 00:01:00	2
2000-01-01 00:02:00	3
2000-01-01 00:03:00	4
2000-01-01 00:04:00	5
2000-01-01 00:05:00	6
2000-01-01 00:06:00	7
2000-01-01 00:07:00	8
2000-01-01 00:08:00	9

Table 5.2: Sensor readings before resampling

Timestamp	Sensor reading
2000-01-01 00:00:00	3
2000-01-01 00:03:00	12
2000-01-01 00:06:00	21

Table 5.3: Sensor readings after resampling to 3 minutes using sum function

Chapter 6

NTNU indoor environmental quality

This chapter compares the data collected at a handful of rooms in the pilot study, and compares the results with the limits and ranges set out in **Chapter 2**. The data used has been processed as described in 5.4. First the results for each room will be given, and then these results will be compared and discussed in chapter 6.5.

All boxplots presented below consist of an orange line indicating the median reading value. The bounding box encompasses quartile 1 (Q1) to quartile 3 (Q3), 25-75 percent of the data observed. The whiskers encompass Q1 - 1.5IQR (interquartile range) and Q3 + 1.5IQR.

All data has been resampled on an hourly or daily basis, unless otherwise specified. All data was resampled using the built-in *avg* function of Pandas, unless otherwise specified.

The VOC readings from the dataset were very high, the minimum observed value was 12,500 PPB. This is probably not a correct reading, as the baseline level is expected to be lower.This Yanzi user manual does not contain enough information to correctly identify the source of this phenomenon. Every VOC reading has therefore been divided by 1,250 to produce a floor of a 100 PPB. This means that the VOC readings are not absolute, and every room's reading should only be used for comparison between relative concentrations.

6.1 Smia

The complete dataset from Smia consists of data collected by two sensors; a COM sensors and one M+ sensor. The readings from the COM sensor and the M+ sensor were resampled, as per 5.4. The dataset consists of 10665 readings, from 29-10-2018-15:24 to 29-11-2018-12:11.

6.1.1 CO2

Weekday	**Mean**	**Min**	**Max**
Monday	583.76	400	1103
Tuesday	525.84	399	854
Wednesday	562.77	399	1178
Thursday	541.01	400	1079
Friday	500.80	400	745
Saturday	774.21	400	1599
Sunday	1203.13	400	2431

Table 6.1: CO2 concentration data

Table 6.1 contains a brief summary of the CO2 data from Smia, broken down on a weekday basis. The min CO2 level is as expected approximately 400 PPM. An interesting thing to note is that the mean CO2 levels are severely elevated on Saturdays and Sundays compared to the rest of the week days. The max values are also the highest for Saturday and Sunday.

Weekday	Readings	Readings Above 1000	Readings above 1500	**%>1000**	**%>1500**
Monday	1855	14	0	0.75	0
Tuesday	2178	0	0	0	0
Wedensday	1416	1	0	0.07	0
Thursday	1618	12	0	0.74	0
Friday	1159	0	0	0	0
Saturday	1166	344	76	29.50	6.51
Sunday	1273	627	431	49.25	33.86
SUM	**10665**	**998**	**507**	**9.36%**	**4.75%**

Table 6.2: Readings relative to the limits set out in 2.3

CO2 readings for most days are excellent, but the CO2 levels remain elevated for Sundays and Saturdays. Almost 50 percent of the readings on Sunday are above 1000 PPM, a level shown to have detrimental effects on student performance 2.1

6.1.2 Temperature

The temperature data comes from the COM sensor, and is processed as described in 5.3. Please note that the temperature readings can be affected by sunlight falling on the sensor. While the COM sensor was mounted in such a way to minimize the risk of this happening, solar reflection off of other surfaces may cause sunlight to indirectly fall on the sensor, potentially skewing the measurements to above actual room temperature readings. No attempt has been made to correct such effects, as there are no other sensors in the room that can be used to attempt to filter out such variables. This could be mitigated by placing several COM sensors in a room, allowing the use of averages or median values to yield a better understanding of room temperature.

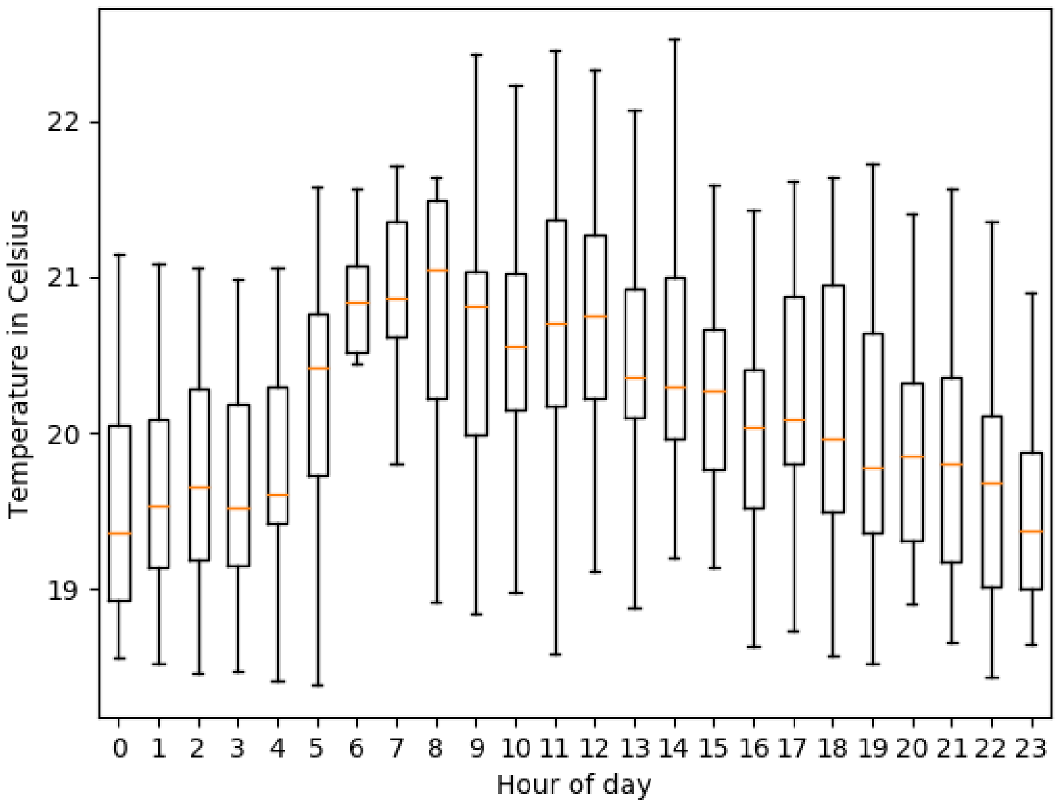

Figure 6.1: Temperature for Smia, re-sampled to 60 minutes and binned by the hour

One of the most noticeable data-points in this plot is the rapid increase in temperature for the hours 04.00 to 05.00. This increase in temperature is not caused by sunlight, as the sun rises above the horizon much later [1] than 05.00 in Trondheim for the the dates in the dataset. The rapid increase in temperature is thus most likely caused by a pre-programmed system increasing the temperature in anticipation of potential occupants. Also note the very short tail towards lower temperatures for the 06.00-06.59 readings.

Another interesting trend to note is that the temperature ranges remain relatively wide for the whole period, in fact 86 percent of all temperature readings for the period 07.00-17.00 fall within the range 19.2 °C and 21.8 °C. The wide range indicates that the system for regulating the temperature is unable to keep a constant temperature. This can be caused by a difference in outside temperature over the data collection period, or a sign of cycling, where the temperature regulation system works based on a duty cycle with an unknown ratio of on and off periods.

The occupants' actions may also cause changes in temperature. In fact, the mere presence of a person sitting still in a room will affect the room temperature due to heat dissipation, even at basal metabolic rates. Starner (1996) found that humans expend about 116 watts of energy

[1] `https://www.timeanddate.com/sun/norway/trondheim?month=10&year=2018`

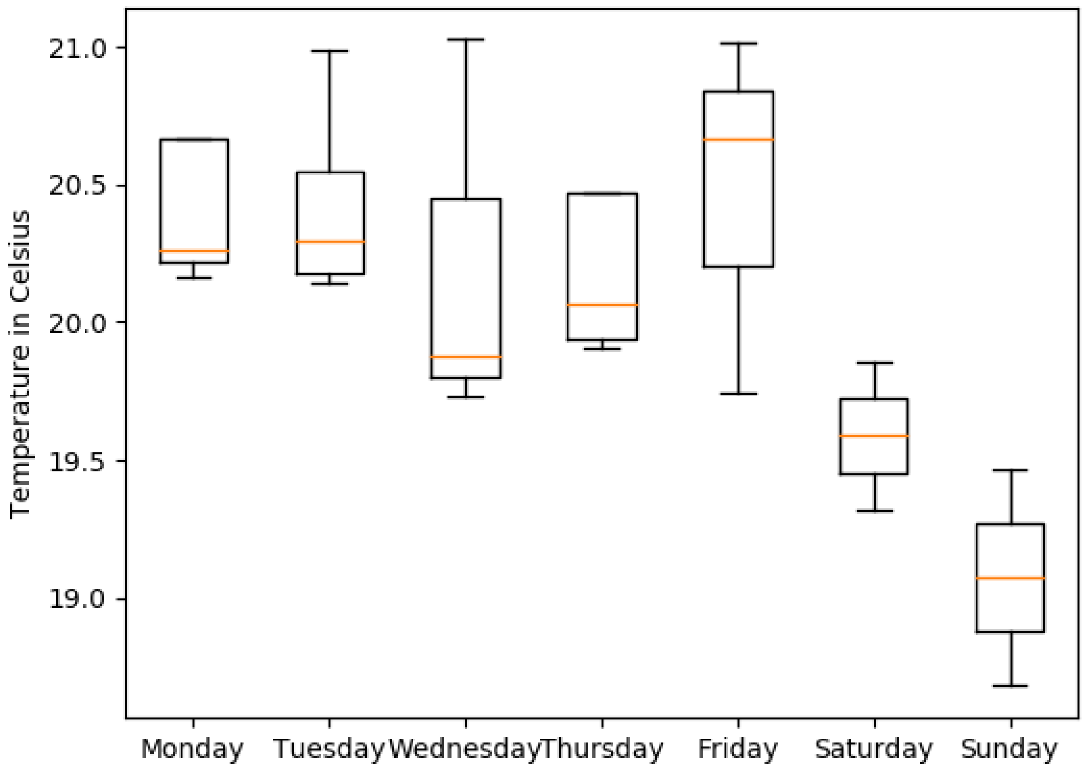

Figure 6.2: Temperature readings re-sampled to the days of the week.[2]

based on the work of Morton (1952) while sitting still. While there was no direct measurement of the occupants specific activity inside Smia, a fair assumption is that most of the activity involved sitting down and either discussing or reading silently.

While most human activity inside the room will lead to an increase in the system's total heat energy, some actions may reduce it. Opening a window and letting in outside air will reduce the temperature inside the room provided that the outside temperature is lower than that of the room. Since no sensors are installed to directly record this activity, it is hard to estimate the rate at which such events occur. The motion counter data may give us some indirect information as to the likelihood of an open window affecting the results. Provided that all windows are closed by the occupant(s) by the time the last person leaves the room, no open-window effect will be observed in the period 23.00-05.59, based on the motion counter.

6.1.3 Motion data

The M+ sensor registers movement in its field of view, in the form of movement events. Summing these events over a period of time, hour, day or week basis, allows us to gain an insight into the relative amount of movement inside the room. The footfall camera inside Smia had some problems due to bad mounting, and it has not been possible to extract useful data from this sensor. The M+ sensor's movement counter will therefore be used a a proxy measurement of the number of people inside the room at a given time. The data has been re-sampled on a daily, hourly and 30 minute basis, summing up every motion event within the time period. The aggregate counter data is assumed to be proportional to the number of people inside the room at a given time; larger counter means more movement which means more people. While this is not as precise as the data the footfall camera could have provided, it should help put other sensor readings into perspective.

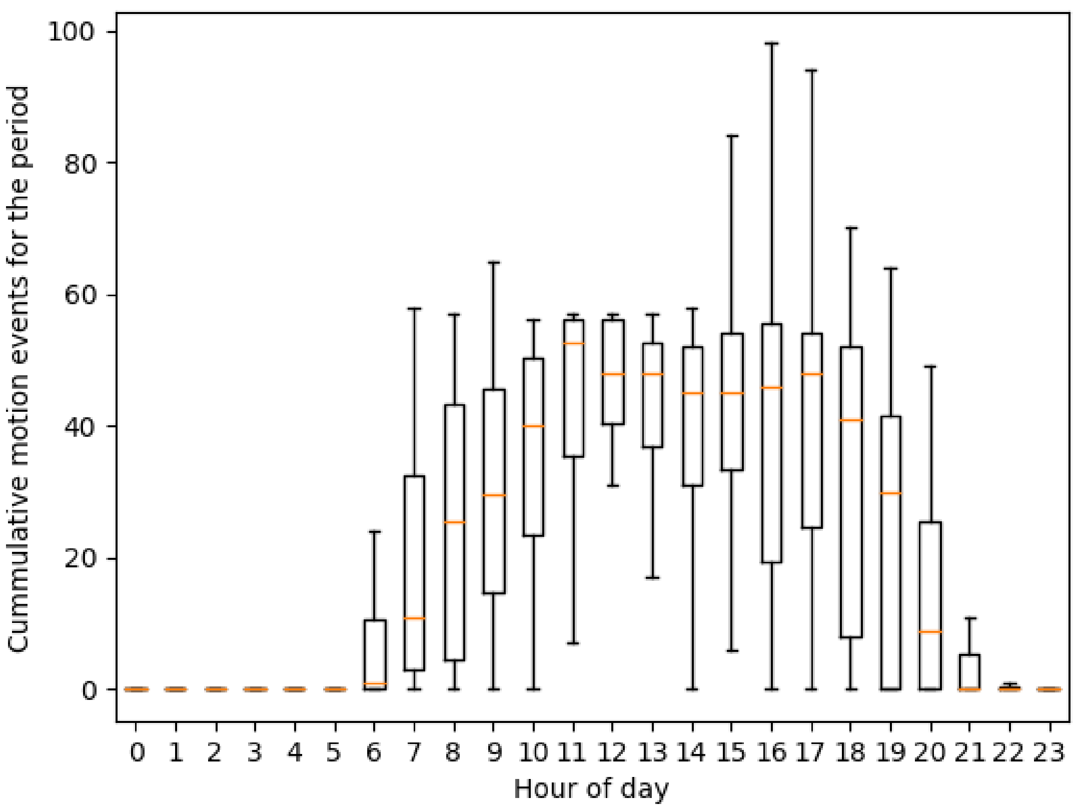

Figure 6.3: Motion counter for Smia, re-sampled to 1 hour slices and binned based on hour of day

The sensor data from the M+ sensor indicates that most of the activity inside Smia happens between 07.00 and 20.00. 11.00-12.00 has the highest median traffic, but some late-day traffic spikes are seen between 15.00 and 17.00. No activity was registered on the M+ sensor between 23.00 and 05.00, which corresponds well to the room's expected usage patterns. Smia is mostly used for group assignments and group-based lectures.

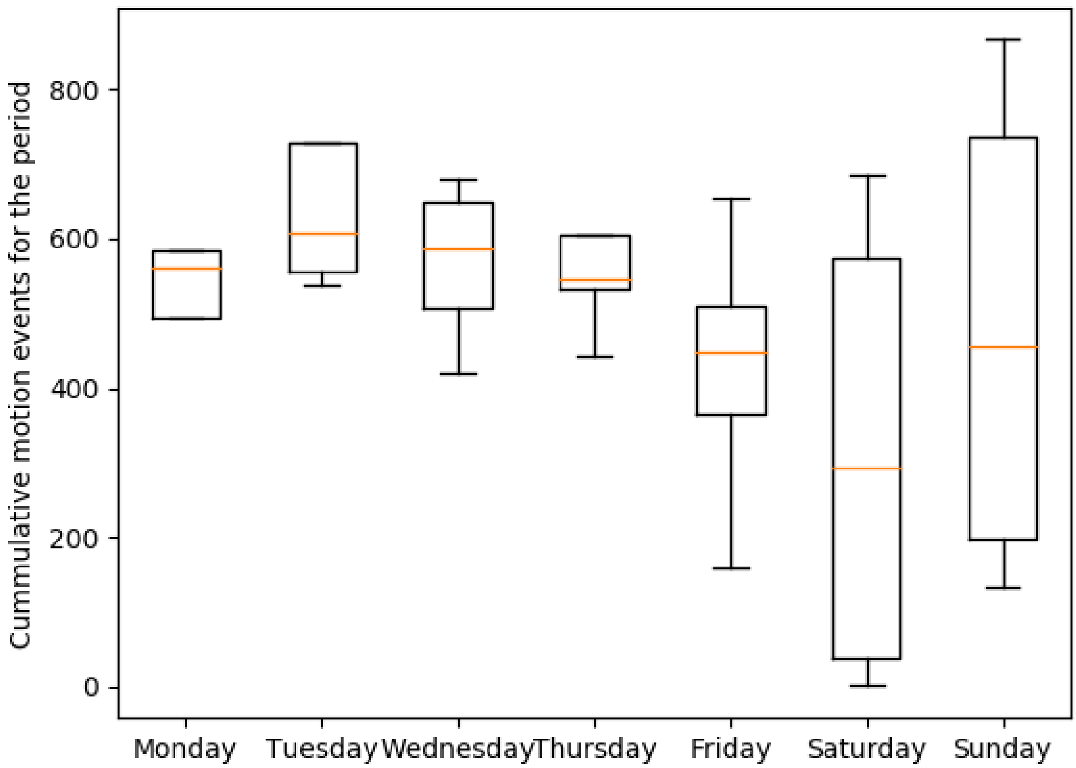

Figure 6.4: Motion counter for Smia, resampled to daily and binned based on weekday

Re-sampling the data to a daily basis, summing up all motion events for a given day, and binning the information into the seven days of the week shows the usage patterns for each of days of the week. Please note that the dataset for the M+ sensor at Smia only contains four week's worth of data, rendering it difficult to draw definitive conclusions. There is little difference in usage patterns between Mondays and Thursdays, the median values are relatively close. Both Saturdays and Sundays see a great deal of variance in the motion registered within the room. This could be caused by extra-curricular activity during the weekend taking place inside Smia. A single Sunday had the most activity out of any days in the data-set, and may help explain the elevated readings on Sundays, as pointed out in **Table 6.1** and **Table 6.10**.

6.1.4 VOC

VOC levels are the result of resampling and averaging the VOC readings across the COM sensors in the room. Smia has no equipment or installations expected to produce large amounts of VOC. Most of the VOC registered in the room should stem from the occupants' and the surfaces of the room itself. It is unknown how long it has been since the room has been painted, and whether or

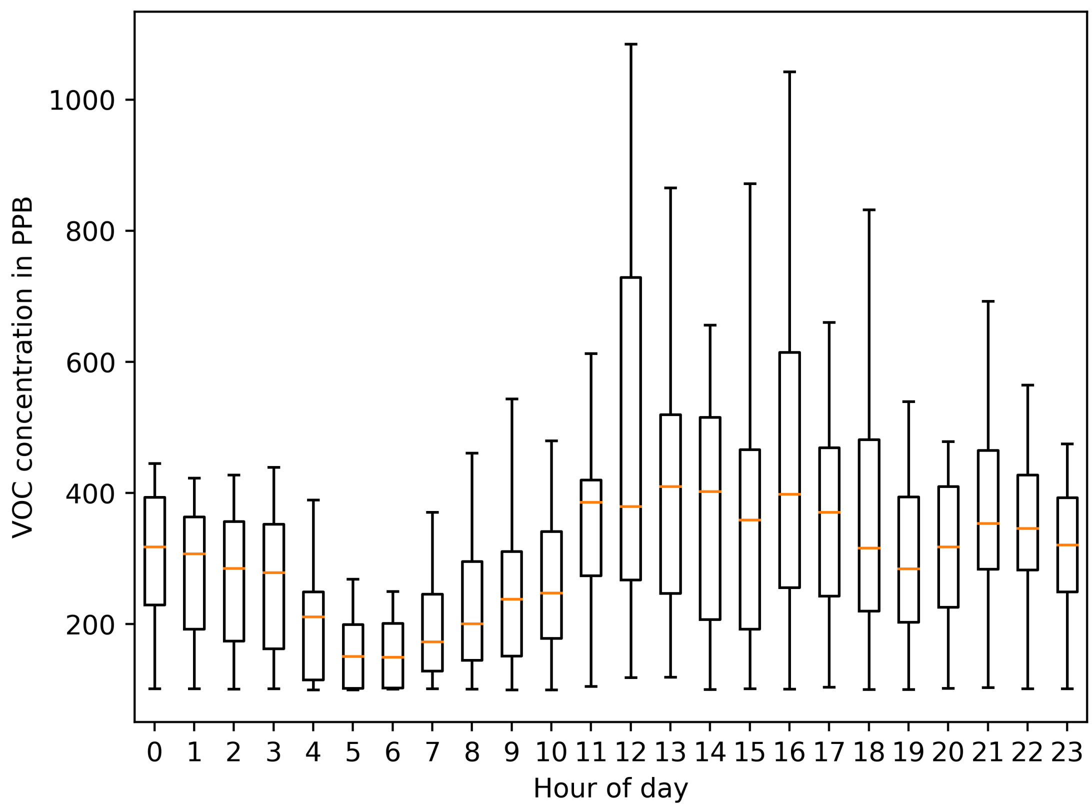

Figure 6.5: Hourly VOC readings for Smia

not this was done as a part of the remodelling of the room. Freshly painted surfaces will increase VOC production.

The hourly VOC readings indicate that the HVAC system is able to keep the concentration under good control. As expected the highest VOC readings are around peak traffic periods, but the reduction in VOC concentration is slow after 16.00, as the HVAC is turned over to a low duty cycle mode, reducing the ventilation rate.

As seen in the CO2 readings, the VOC readings are also elevated on Saturdays and Sundays compared to the working days. This is most likely caused by the redcued ventilation rate of these days.

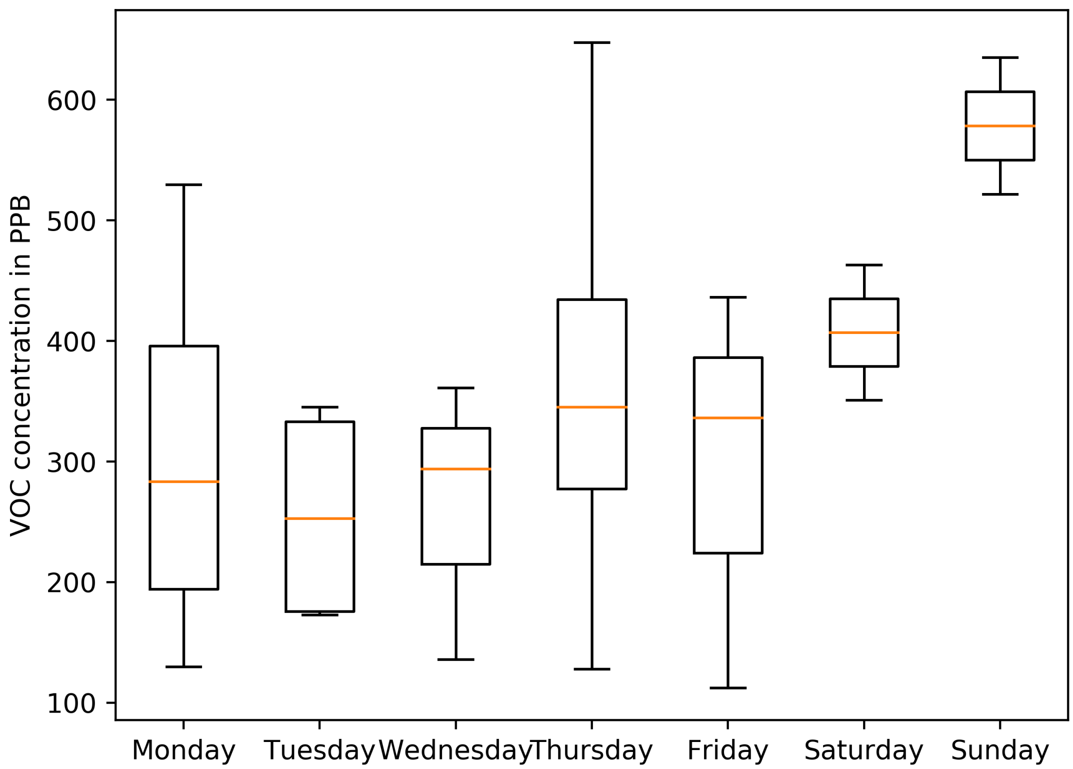

Figure 6.6: VOC readings binned by the day of the week

6.1.5 Humidity

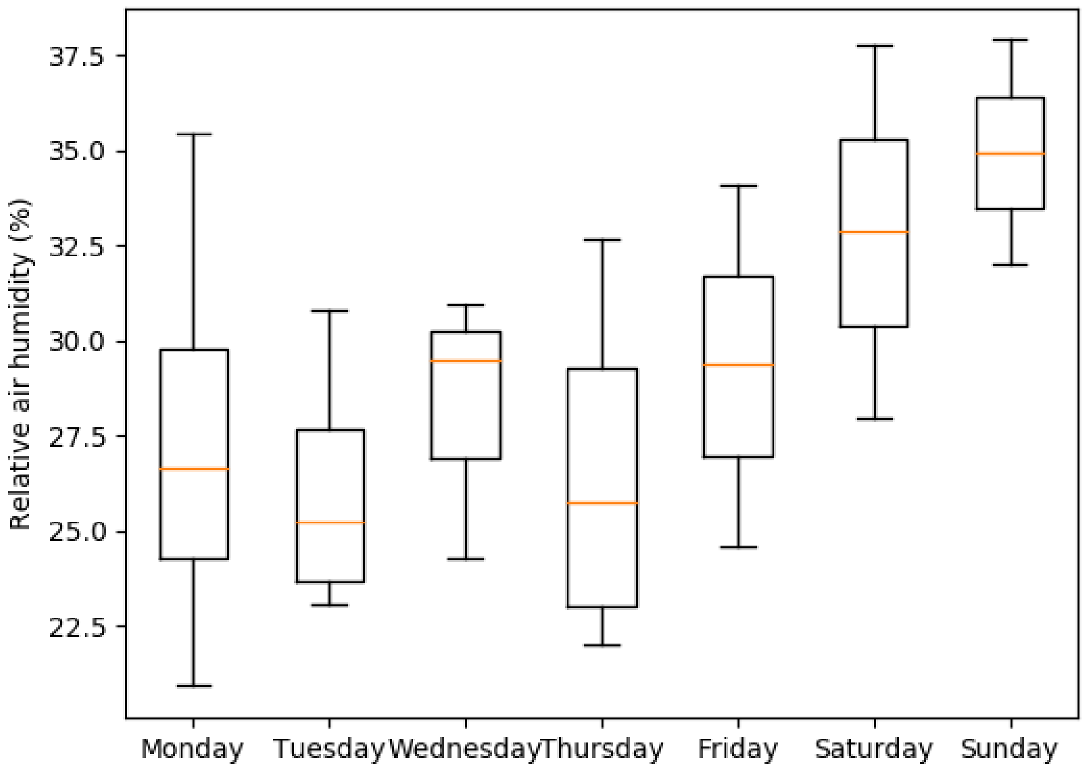

Figure 6.7: Relative humidity for Smia, binned to weekday intervals

The humidity data for Smia shows that most readings are between 35 and 23 percent relative humidity. There seems to be a trend towards higher relative air humidity in the weekends. This could be caused by a reduction in the HVAC system, as shown in the CO2 data.

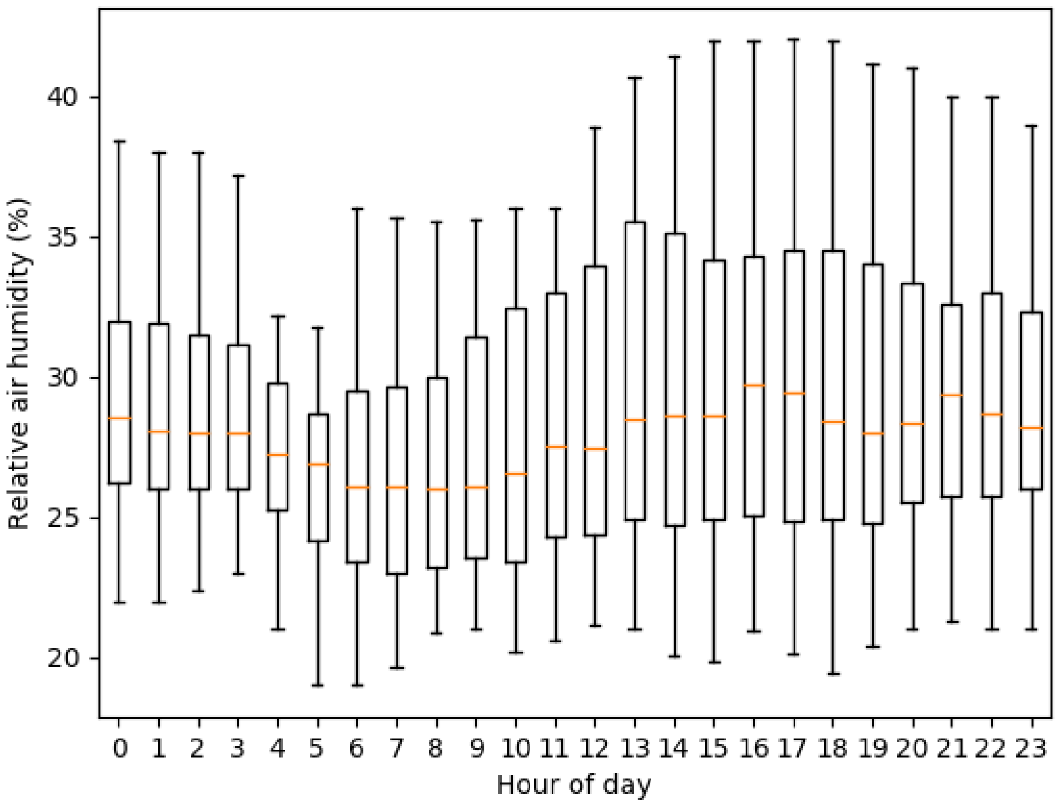

Figure 6.8: Relative humidity for Smia, binned to hourly intervals

67 percent of the readings are below 30 percent relative humidity when re-sampled to an hourly binning. This may cause irritation of the eye for some people, as noted in 2.2, but should ensure a reduction in many airborne pollutants. The difference between Q3 and Q1 remains about 10 percentage points for most hours, indicating the the HVAC system is able to keep the humidity levels relatively stable.

6.2 U1

Since U1 consists of several distinct zones with different activities and usage patterns, this section has been split into two subsections, each covering the two zones of U1 investigated in this thesis. U1 is outfitted with three footfall cameras, logging the number of people entering and leaving U1, in addition to tracking the usage pattern for each door covered. Footfall camera data was not present in the data-set used in this thesis, and is therefore not covered in this section.

Only two of the three zones in U1 are investigated and presented below. These are the maker space and the group rooms. This is because these two zones represent a type of room and setup that are unique in the dataset.

6.2.1 Maker space

The maker space section consists of a room containing workbenches outfitted with 3D printers, soldering stations and other equipment and tools used for electronics and mechanical construction.

The maker space is outfitted with three M+ sensors and three COM sensors. The readings from each of these sensors are averaged to produce an aggregate number of the room as a whole as opposed to looking a sub-sections of the room.

CO2

Weekday	**Mean**	**Min**	**Max**
Monday	571.05	395	886
Tuesday	603.60	395	996
Wednesday	588.72	399	883
Thursday	564.16	399	913
Friday	525.76	398	806
Saturday	546.70	399	1203
Sunday	575.96	440	1110

Table 6.3: CO2 concentration data in PPM

Weekday	Readings	Readings Above 1000	Readings above 1500	**%>1000**	**%>1500**
Monday	323	0	0	0	0
Tuesday	381	0	0	0	0
Wednesday	391	0	0	0	0
Thursday	327	0	0	0	0
Friday	316	0	0	0	0
Saturday	360	17	0	7.72	0
Sunday	339	17	0	5.01	0
SUM	**6255**	**34**	**0**	**0.54%**	**0%**

Table 6.4: Readings relative to the limits set out in 2.3

The CO2 readings of the maker space are excellent. The highest observed reading was 1203PPM. Median readings are only about 150-200PPM above baseline levels (400PPM). Little to no negative effects are expected at the CO2 levels recorded for the maker space.

Temperature

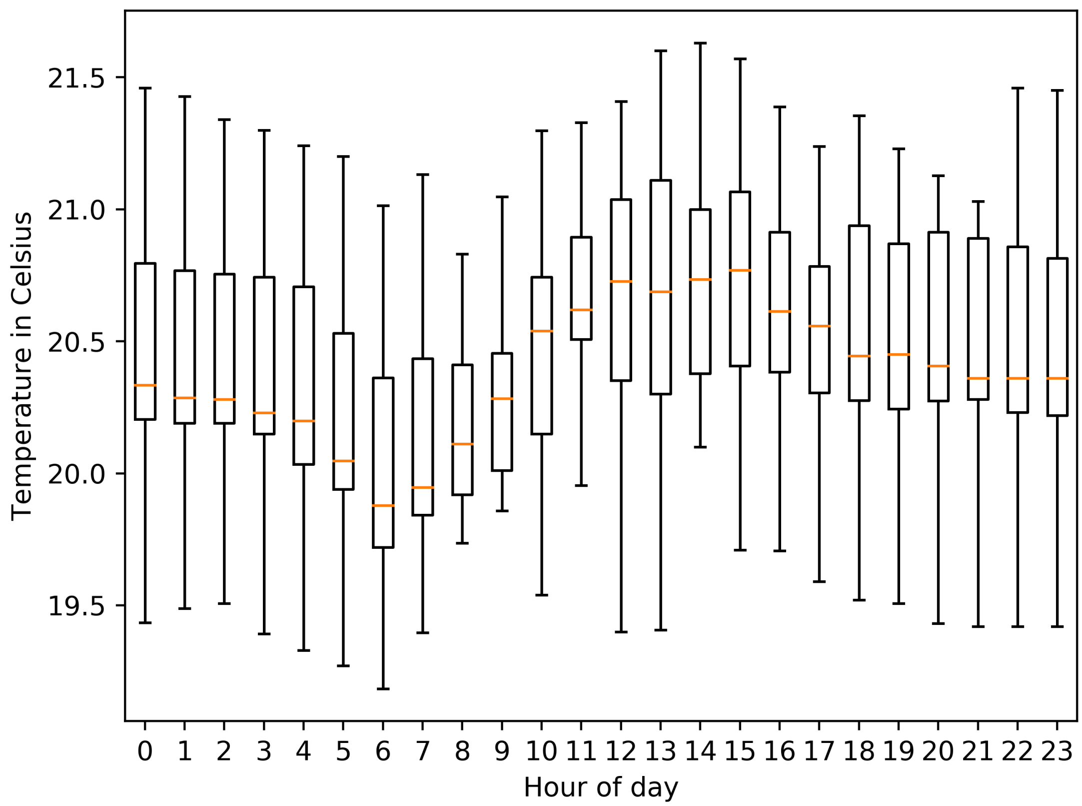

Figure 6.9: Temperature readings for the maker space, averaged across three sensors and binned by the hour

The hourly temperature readings shows that the HVAC and heating system is able to keep the maker space at a narrow temperature interval. For most of the hours in the dataset the range between maximum and minimum readings are about two degree Celsius. The pre-set heating time for the room seems to be set to somewhere between 06.00 and 07.00 based on the median temperature rise between these hours.

The hourly temperature range would probably be classified as comfortable by most occupants, but a minority would probably prefer the temperatures to be increase 0.5 to 1 degree Celsius.

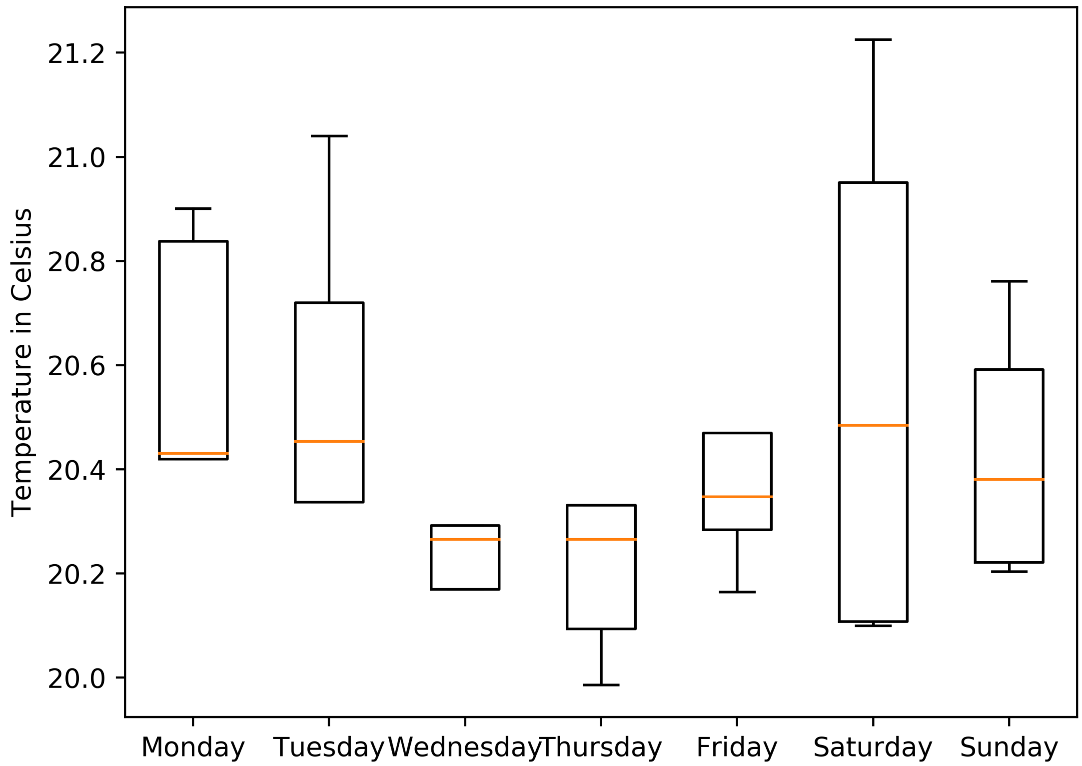

Figure 6.10: Temperature readings for the maker space, averaged across three sensors and binned by the day of the week

The dataset is relatively small for U1, making it difficult to draw clear conclusions for the temperature readings on a daily basis. The Q1 and Q3 ranges remain relatively low for most days, with Saturday having the biggest range of about 0.9 degree centigrade. This indicates that the temperature control system is well-tuned and able to keep the temperature within the set interval.

Motion data

The motion data indicates that the hours 11.00 to 18.00 are on average the busiest hours for the maker space. There is also relatively little variance in traffic for the hours 11.00 to 16.00, as indicated by the close grouping of Q1 and Q3. No activity has been registered in the period 0.00 to 05.00.

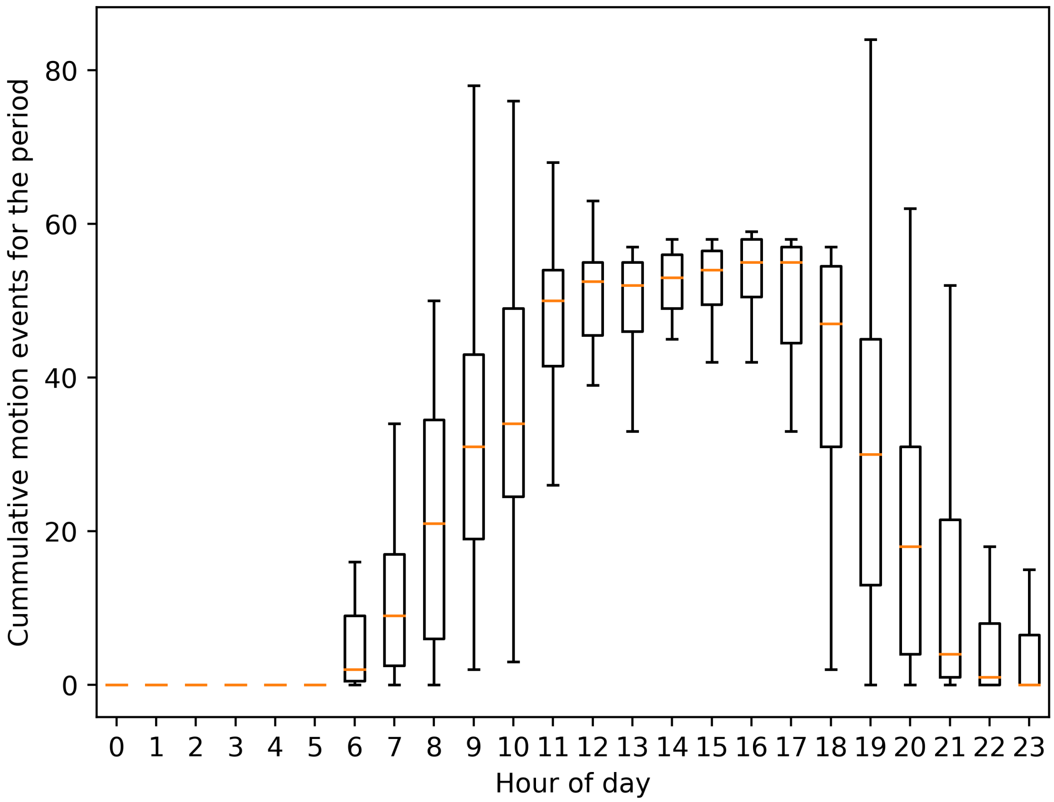

Figure 6.11: Motion data for the maker space, averaged across the M+ sensors and binned by the hour of the day

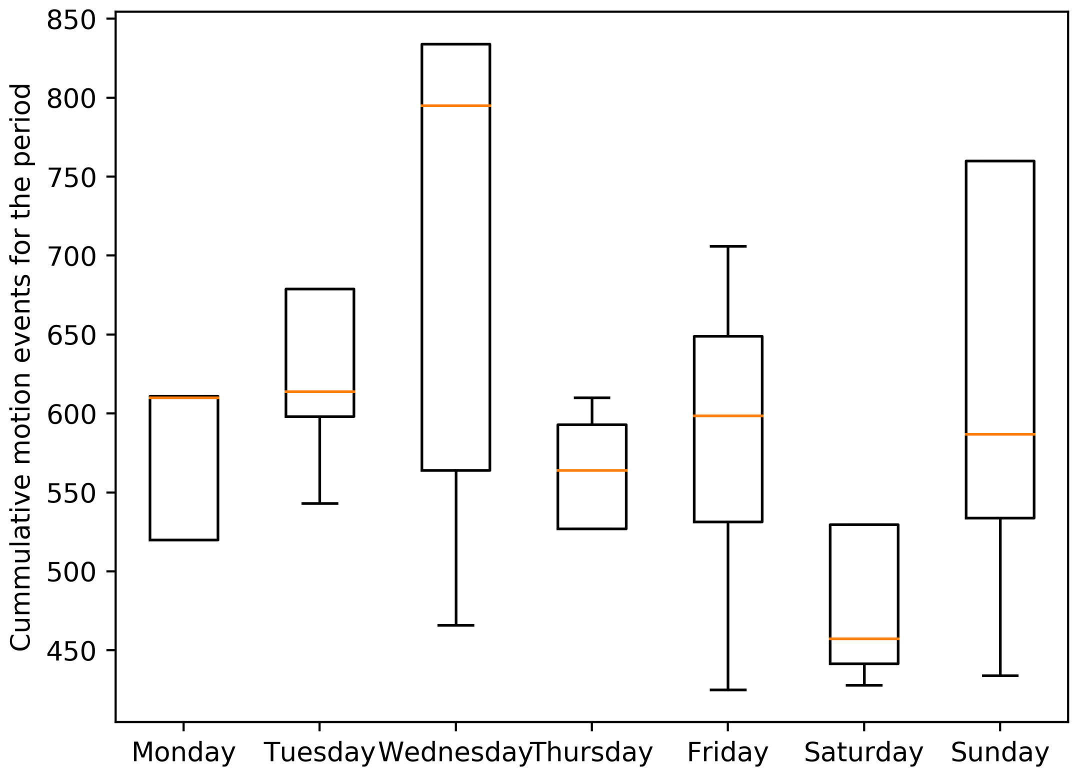

Figure 6.12: Motion data for the maker space, averaged across the M+ sensors and binned by the day of the week

The maker space sees the most activity Monday through Friday, with Wednesday being the busiest day, both in terms of median traffic as well as peak traffic. Sunday has a surprisingly large amount of traffic, especially compared to Saturday. This could be due to non-curricular activity, as the maker space is often used by students' for hobby projects and the like.

Humidity

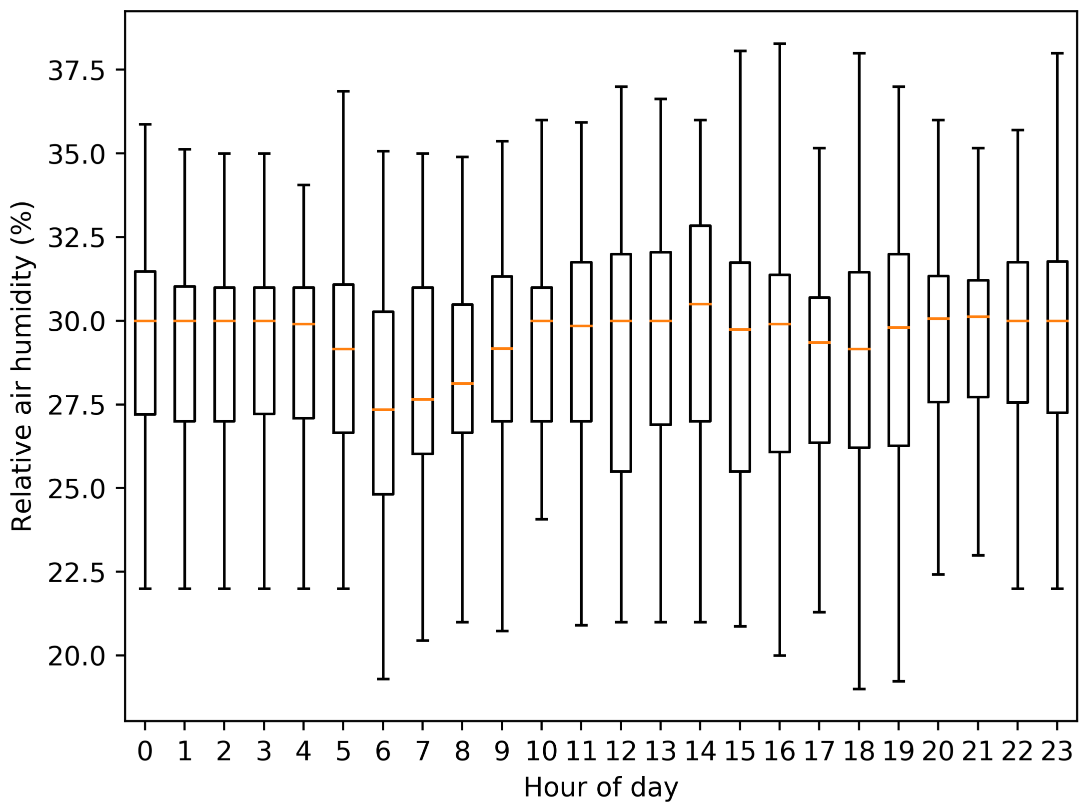

Figure 6.13: Relative air humidity binned by the hour of the day

The hourly relative humidity measurements show that the HVAC system is able to keep the humidity well regulated, with median values only differing by 0.5 percent relative humidity from the lowest to the highest median reading. There is a drop in the relative humidity between 05.00 and 06.00, indicating that the HVAC system has been pre-programmed to increase the ventilation rate, leading to a drop in the humidity.

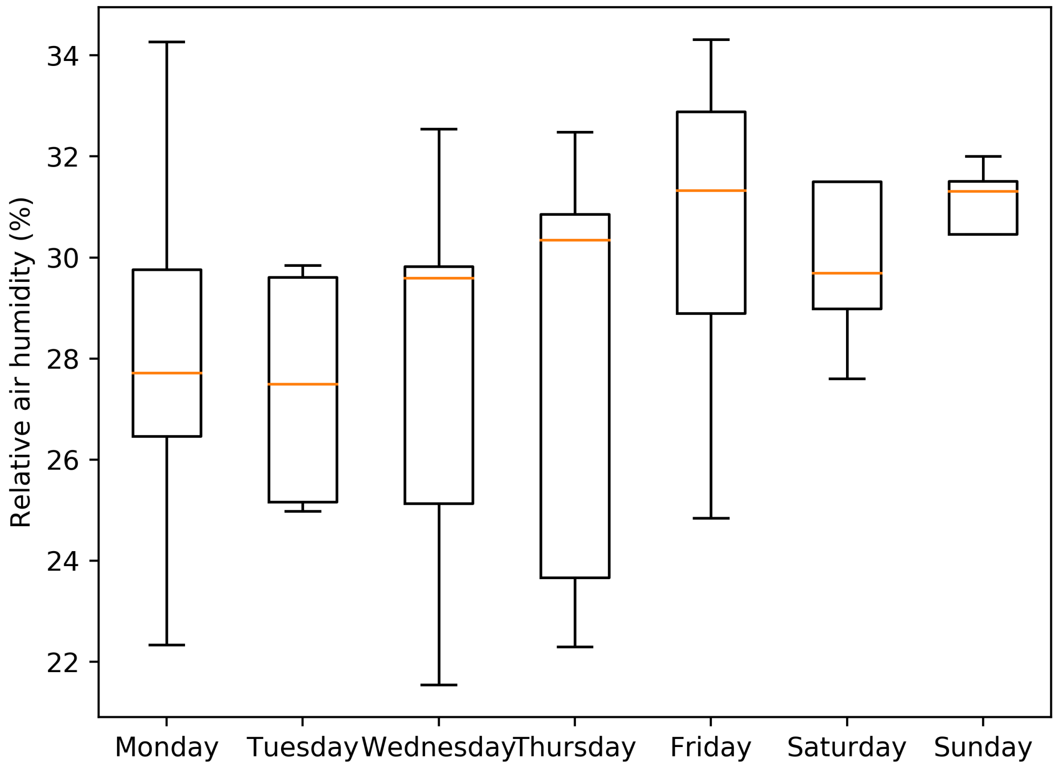

Figure 6.14: Relative air humidity binned by the day of the week

The relative humidity rates when binned by the day of the week shows that the Q1 and Q3 groupings are spaced wider apart than for the hourly readings. This may be due to the fact that there are few sensor readings available for U1 in the dataset, causing a wider grouping.

VOC

Since the maker space contains both soldering stations and 3D printers, the VOC levels are expected to be elevated. Note that the VOC data has been re-sampled and calibrated to ensure that 100 PPB is the baseline value.

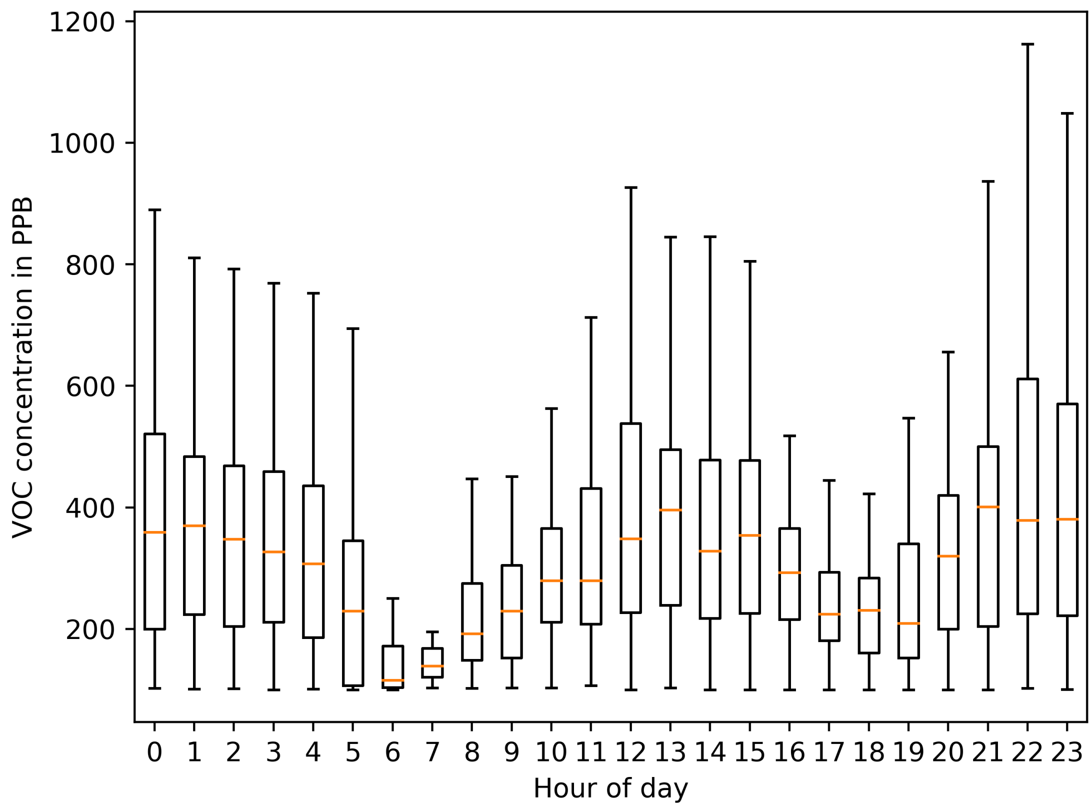

Figure 6.15: Hourly VOC concentrations for maker space

The hourly VOC readings have a wide separation between Q1 and Q3. Outliers also indicate that there is a great degree of variation between the readings for a given hour. This could be due to the aforementioned soldering and 3D printing activities, as these are expected to generate VOCs. There is also an increase in the median VOC levels between 19.00 and 01.00, indicating that the concentration is going up. The reasons for this is unknown. The motion counter data indicates that there is relatively little activity between 21.00 and 23.00. One possible explanation is that the reduced operating setting of the HVAC system allows the relative VOC concentration to increase due to the reduced ventilation rate.

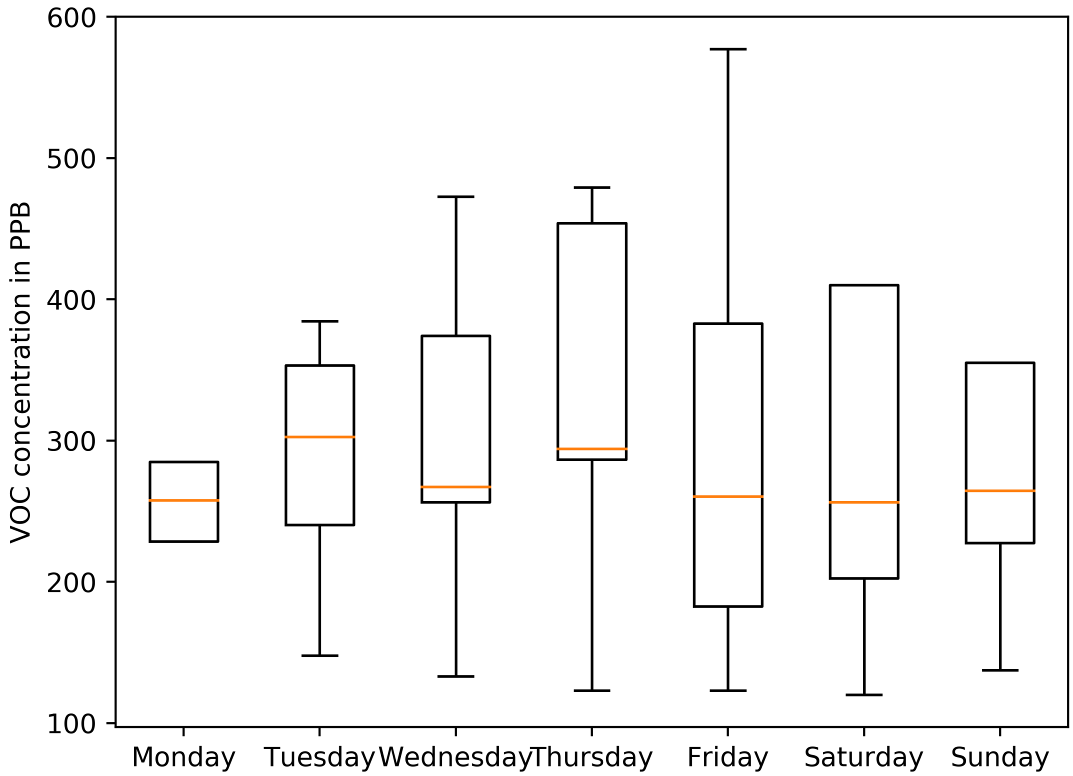

Figure 6.16: Dail VOC concentrations for maker space

The median VOC readings when resampled to a daily basis shows that the median value is more or less constant, with some variation. There is too little data available to make robust conclusions for the daily readings.

6.2.2 Group rooms

U1 contains four group rooms, each equipped with a table, 6-8 chairs and a whiteboard. Every group room is an enclosed individual room, with walls, a door and ceiling. New air is delivered into each room from the HVAC system via air-ducts in the ceiling.

Each group room is outfitted with a M+ sensor and COM sensor. This allows logging motion, CO2, VOC, temperature and humidity in addition to barometric pressure and sound levels. The latter two will not be covered in this section.

One out of the four group rooms was chosen at random. The rooms are identical so the readings from one room should be representative for the group rooms as a whole.

CO2

Weekday	Mean	Min	Max
Monday	803.31	400	1929
Tuesday	683.99	399	2022
Wednesday	752.37	400	2801
Thursday	683.71	400	2245
Friday	890.20	398	4078
Saturday	1460.64	399	3081
Sunday	1388.96	440	4508

Table 6.5: CO_2 concentration data in PPM

Weekday	Readings	Readings Above 1000	Readings above 1500	**%>1000**	**%>1500**
Monday	432	115	54	26.62	12.5
Tuesday	478	86	8	17.99	1.67
Wednesday	363	66	30	18.18	8.26
Thursday	374	40	17	10.70	4.54
Friday	381	73	63	19.16	16.54
Saturday	384	243	176	63.28	45.83
Sunday	383	213	121	55.61	31.59
SUM	**6255**	**836**	**469**	**13.36%**	**7.49%**

Table 6.6: Readings relative to the limits set out in 2.3

The group rooms show very high CO2 peak CO2 readings, with the maximum observed being more than 4000 ppm. Peak values for all days are close to or exceed 2000 PPM. This indicates that the HVAC system is unable to replace the air at sufficient rates to keep the CO2 levels at acceptable levels. The weekend days have by far the highest concentration of CO2, and more than 30 percent of the time the concentration is above what have been shown to reduce learning rate, coordination and problem solving.

Temperature

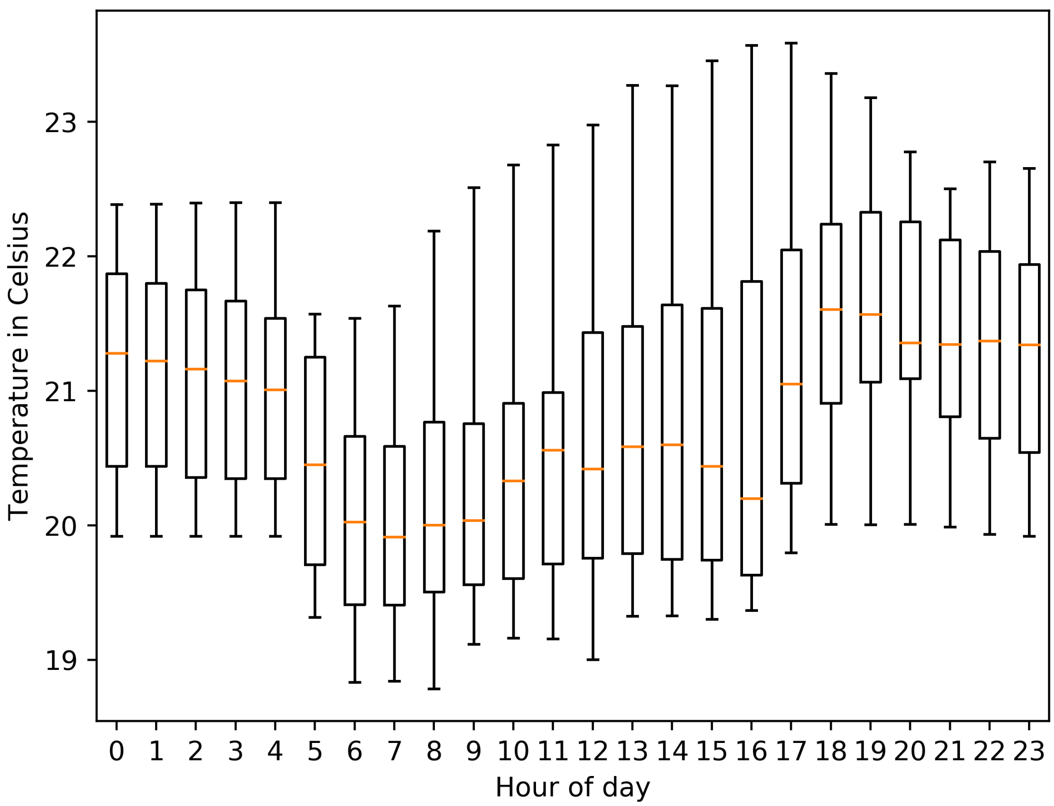

Figure 6.17: Average room temperature of the group rooms, binned by the hour of the day

The group rooms as U1 show some interesting behaviour for the temperature binned by the hour of the day; usually the temperature falls over the night time and then start increasing when the heating system turns on, usually around 06.00. This is not the case for the group rooms. The temperature slowly falls as indicated by the median temperature between 23.00 and 04.00, before it falls rapidly between 05.00 and 07.00. The reason for this in unknown. There is no active cooling installed in the room, so the most likely explanation is that HVAC system is removing hot air and replacing it with colder air, thus dropping the room-temperature. The median temperature then rises slowly until 11.00 when it begins fluctuating between a narrow interval until 16.00 when the temperature starts increasing quickly. This is probably when the HVAC system enters its reduced speed mode, and the rooms increase in temperature as air is not replaced as quickly.

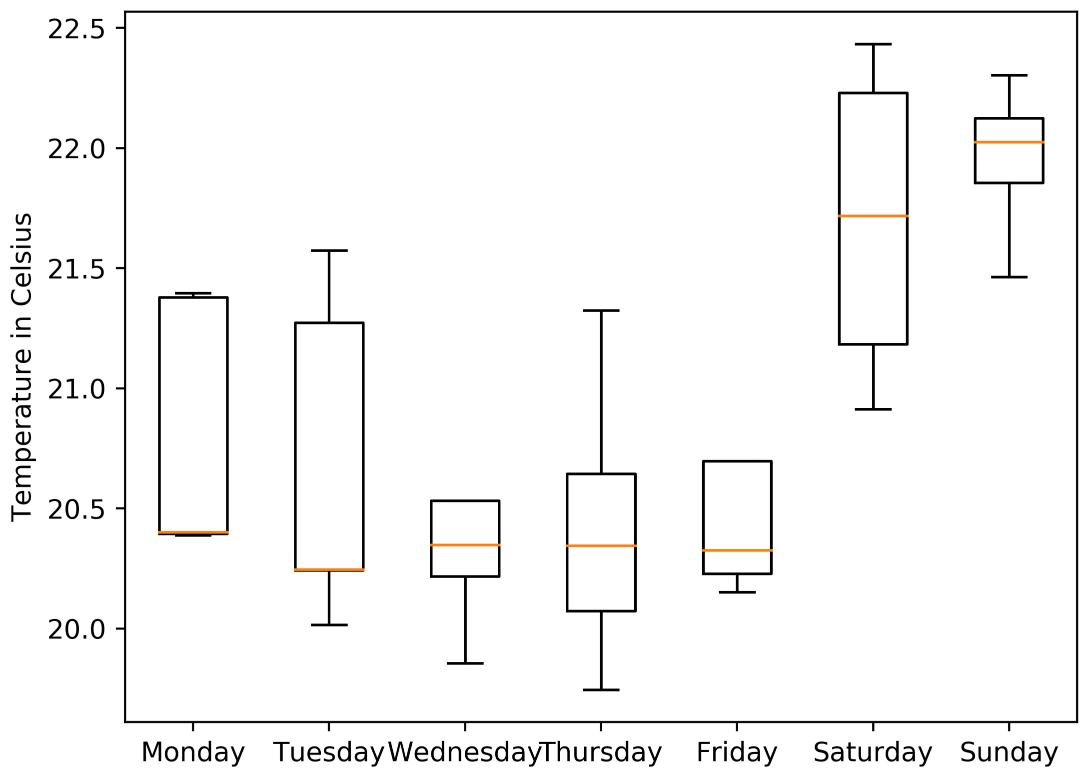

Figure 6.18: Average room temperature of the group rooms, binned day of the week

The median temperature for the days Monday to Friday remain more or less constant, before they increase dramatically for Saturday and Sunday. This is most likely caused by the reduced HVAC setting for the weekend.

Motion

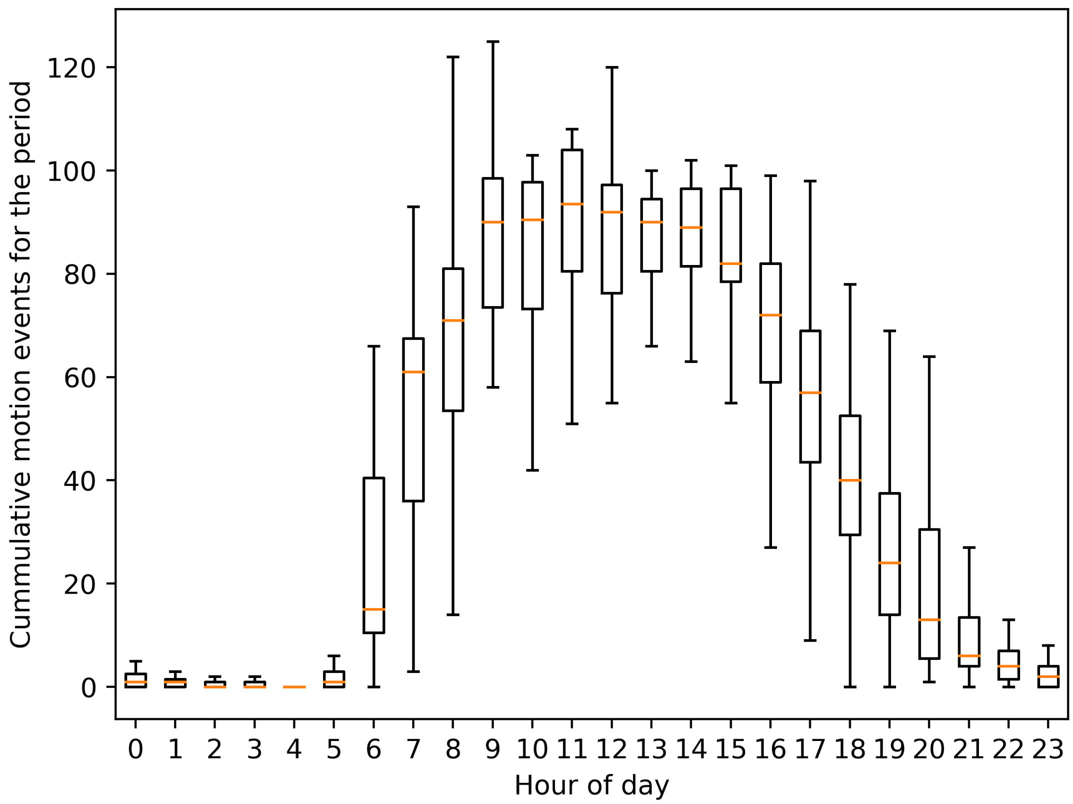

Figure 6.19: Average motion counter data for the group rooms, binned by the hour of the day

The hourly usage patterns for the group rooms are similar to those of the maker space. Most activity is seen between 09.00 and 16.00. Interestingly some activity has been observed between 0.00 and 03.00. Whether or not this is actually caused by occupants in the room or is due to sensor glitches or badly placed sensors is unknown.

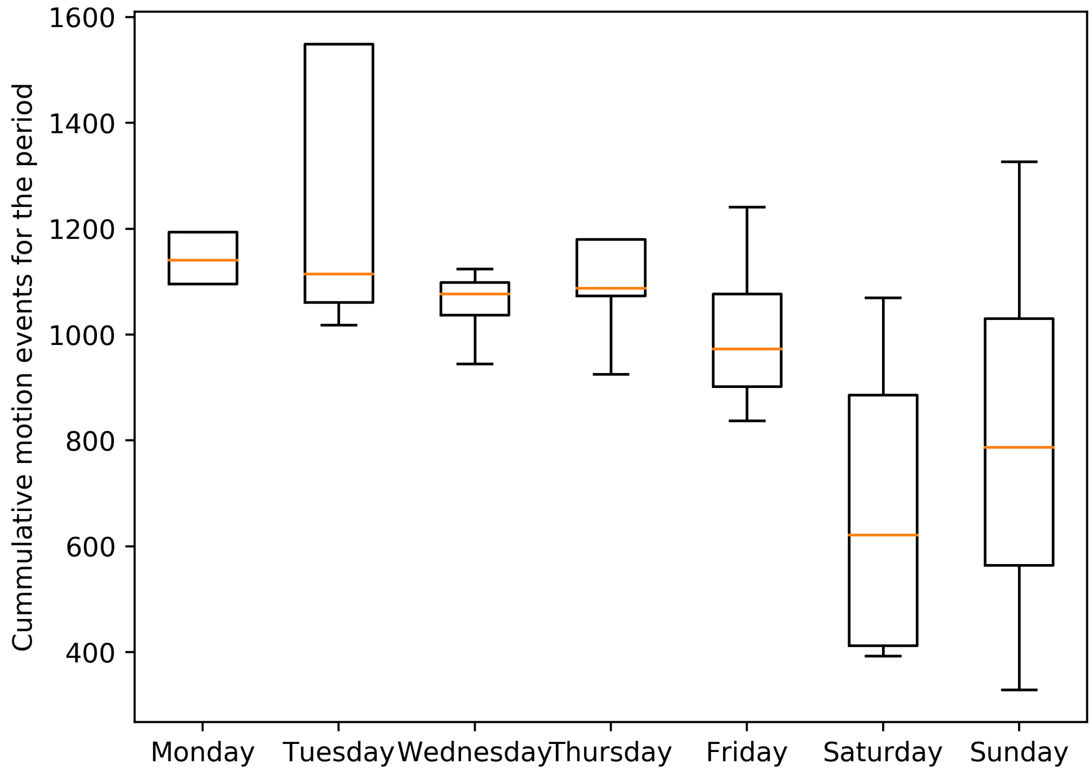

Figure 6.20: Average motion counter data for the group rooms, binned by the day of the week

Monday through Thursday show very similar median motion counter data. Tuesday has the highest Q3 reading, but this may be caused by sensor malfunctions. Friday sees somewhat reduced room usage, but not by a significant amount. Saturdays and Sundays show the lowest room usage, just as for the maker space. Interestingly Sundays have a higher median motion event readings than Saturdays, indicating that the rooms are more frequently used.

VOC

The group rooms are expected to have a lower concentration of VOCs than the maker space, as the group rooms contain no soldering stations or 3D printers, potentially emiting a large amount of VOCs when in use.

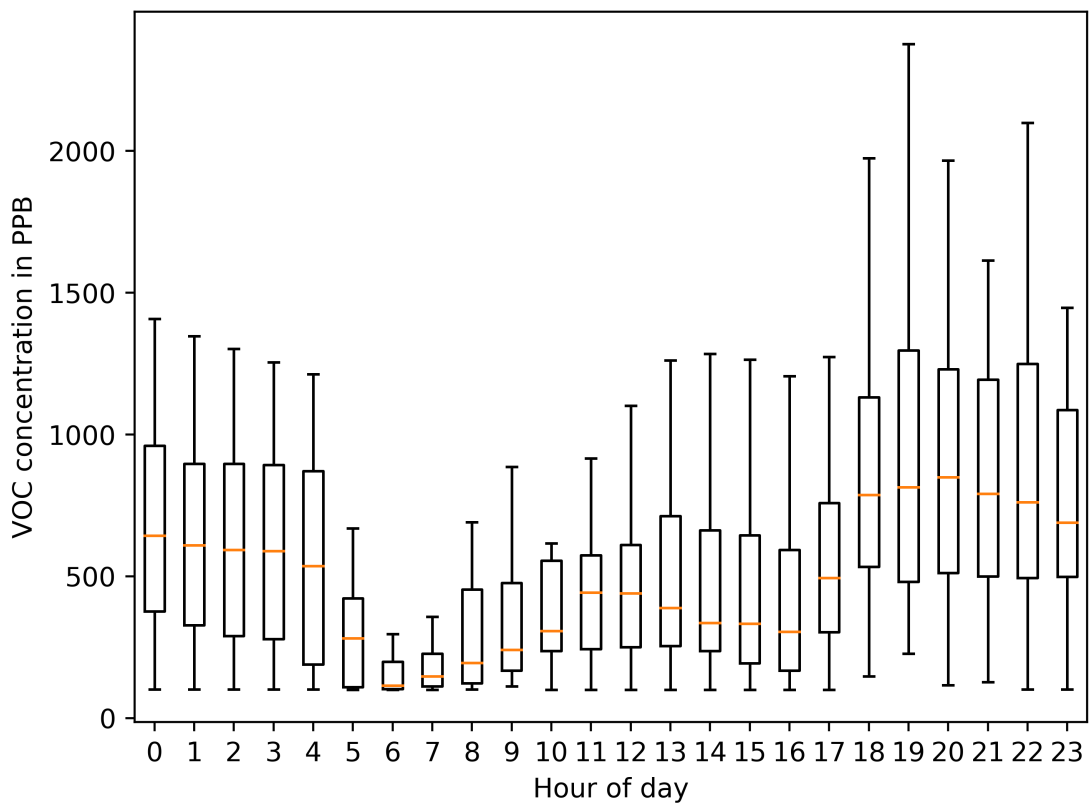

Figure 6.21: Average VOC concentration for the four group rooms, binned by the hour

The hourly VOC readings indicate that the HVAC system returns to it normal running mode somewhere between 04.00 and 05.00, as the VOC levels drop quickly. This happens around the same time as the median temperature of the rooms drop, indicating that the HVAC is indeed returning to normal run mode around those hours. A rapid increase is seen between 16.00 and 18.00, indicating that the HVAC is being put into its slower running night mode. During this period the median VOC concentration increases by about 400 PPB.

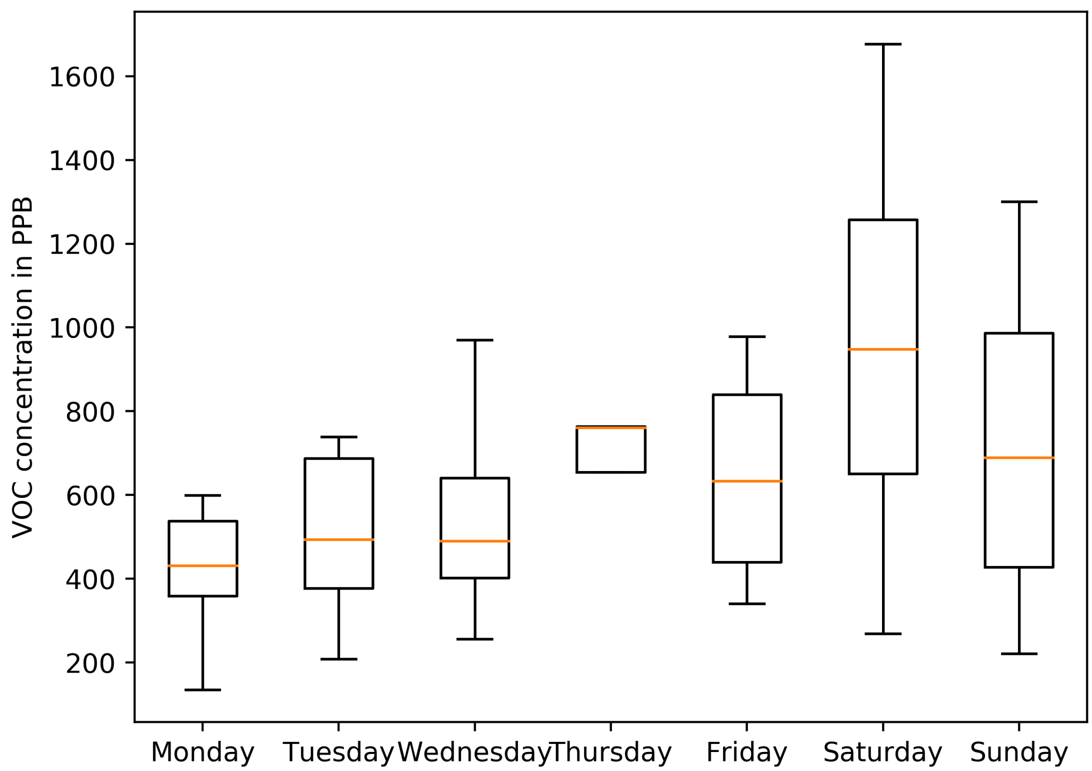

Figure 6.22: Average VOC concentration for the four group rooms, binned by the day of the week

Saturdays see the highest VOC concentration. This is somewhat unexpected as the room usage is usually lower than on Sundays, and yet the Sunday VOC concentrations are lower, despite both days seeing continued reduced HVAC settings. The reason for this is unknown, but it could simply be that the concentration continuously falls from Saturday to Sunday as the air is slowly exchanged over the course of the days.

6.3 Sandkassa

Sandkassa contains three COM sensors, but no motion counter sensors. Instead Sandkassa was equipped with three presence sensors, indicating whether a specific part of the workplace was occupied or not. Due to technical difficulties, the data from these sensors could not be extracted from the provided dataset. The data below is therefore presented without the added clarity of a relative motion count to put the data in perspective. The data from Sandkassa contained in the dataset only stretches over 23 days, making it difficult to draw any definitive conclusions as to the long-term indoor climate at Sandkassa. The data does however give a good indication of the indoor climate during the period the readings were recorded.

6.3.1 CO2

The CO2 data from the COM sensors was averaged, producing an average CO2 concentration for Sandkassa as a whole.

Weekday	**Mean**	**Min**	**Max**
Monday	942.33	400	1624
Tuesday	1039.75	399	1990
Wednesday	1149.28	400	2104
Thursday	756.15	400	1447
Friday	532.57	398	911
Saturday	915.29	399	2018
Sunday	1161.70	440	2286

Table 6.7: CO2 concentration data in PPM

Sandkassa has poor CO2 readings. Most days expect Thursday and Friday have mean CO2 readings close to or exceeding 1000 PPM, and that includes the whole 24 hour period of each day. This indicates that the ventilation is inadequate for the student volume using the room, or that ventilation control system is badly tuned. As in the other rooms analyzed in this study, the CO2 readings for Saturdays and Sundays are high, though the difference between weekend and week days are smaller than for the other rooms. The observed CO2 concentrations are so high that the levels set out in 2.3 are more than exceeded, and the concentration is expected to have a detrimental effect on students' ability to perform and may affect their ability to learn new material.

Weekday	Readings	Readings Above 1000 PPM	Readings above 1500	**%>1000**	**%>1500**
Monday	201	100	5	49.75	2.48
Tuesday	183	91	36	49.73	19.67
Wednesday	246	178	29	72.35	11.78
Thursday	242	48	0	19.83	0
Friday	190	0	0	0	0
Saturday	192	71	43	36.98	22.40
Sunday	192	107	45	55.73	23.44
SUM	**1446**	**505**	**158**	**34.92%**	**10.92%**

Table 6.8: Readings (in PPM) relative to the limits set out in 2.3

6.3.2 Temperature

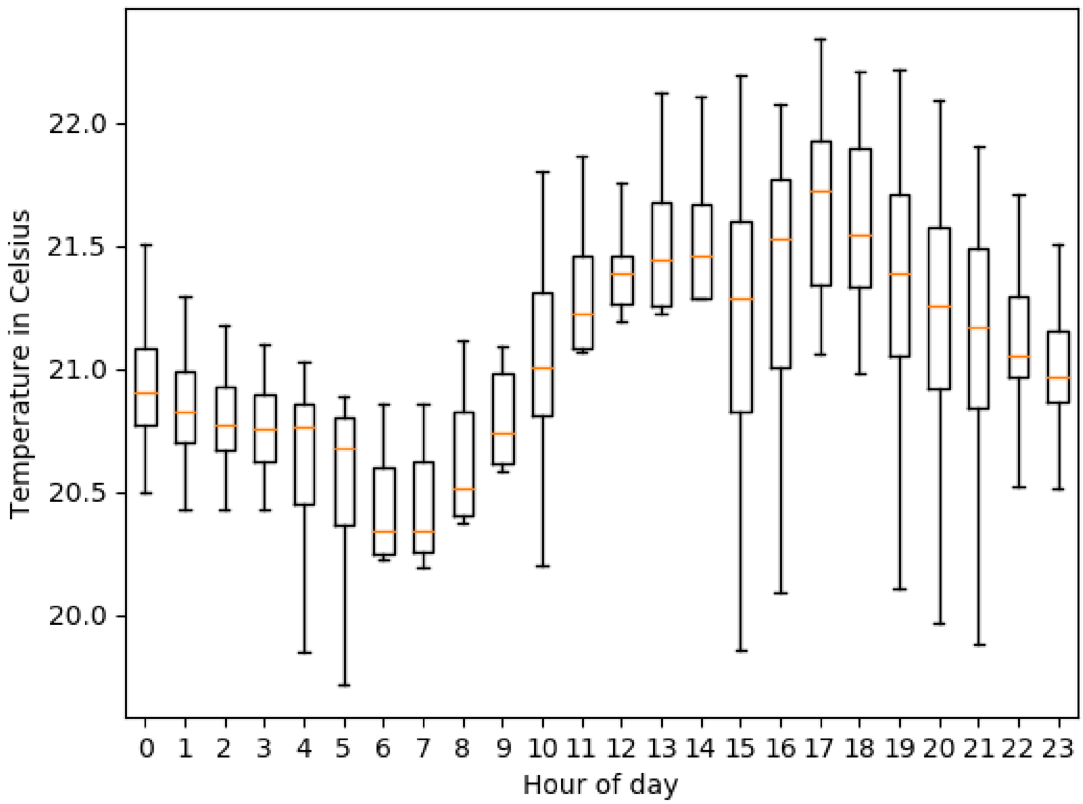

Figure 6.23: Temperature for Sandkassa, binned by the hour of the day

The hourly temperature readings show the same pattern as observed with Smia and Koopen; the temperature starts rising somewhere between 07.00 and 08.00 due to the automatic pre-programmed heating system kicking in. The temperature slowly rises before it starts dropping down again around 17.00-18.00, yet again due to the pre-programmed cycle. The temperature readings are good, and the boxplot indicates that the temperature control system is able to keep the temperature within acceptable ranges. Some of the tails are relatively long, indicating either sensor misreadings or outside factors affecting the temperatures. This could be caused by an open door or window, or other factors.

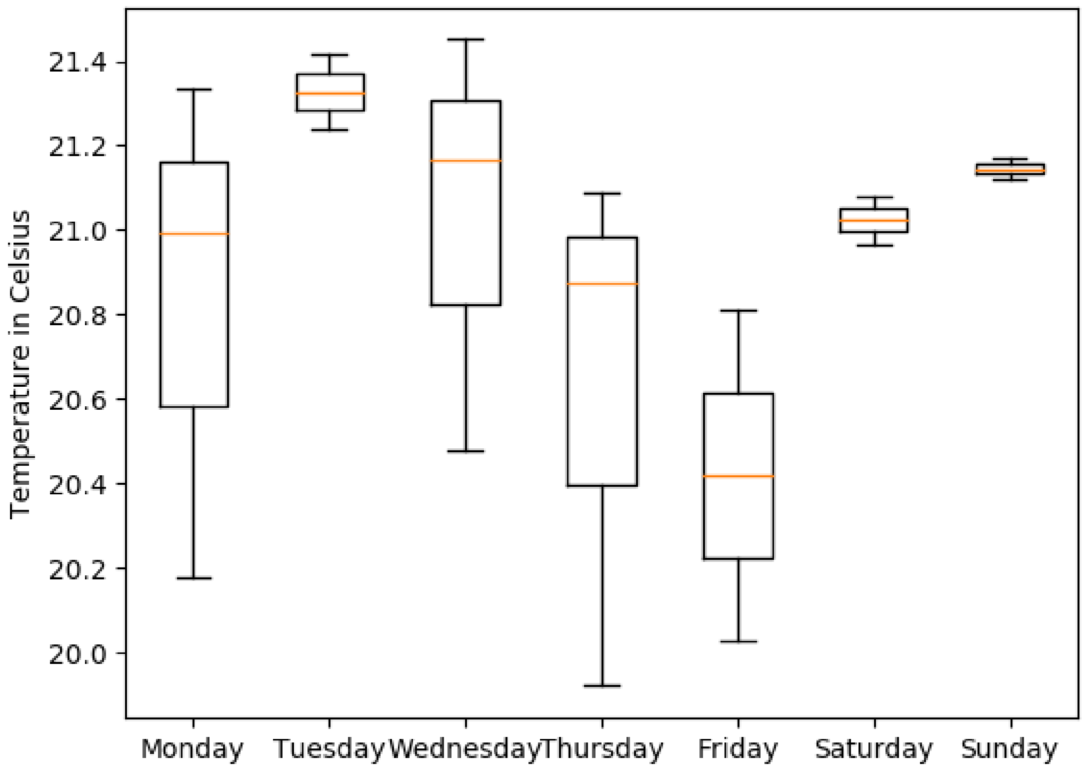

Figure 6.24: Temperature for Sandkassa, binned by the the day

The time frame for the data included for Sandkassa in the supplied dataset is relatively short, 23 days, thus reducing the number of data points available when binning by weekday. It is therefore not possible to draw robust conclusions as to the difference between temperatures across weekdays. Despite this, an interesting observation is the very narrow bands for the temperature ranges for Saturdays and Sundays.

Humidity

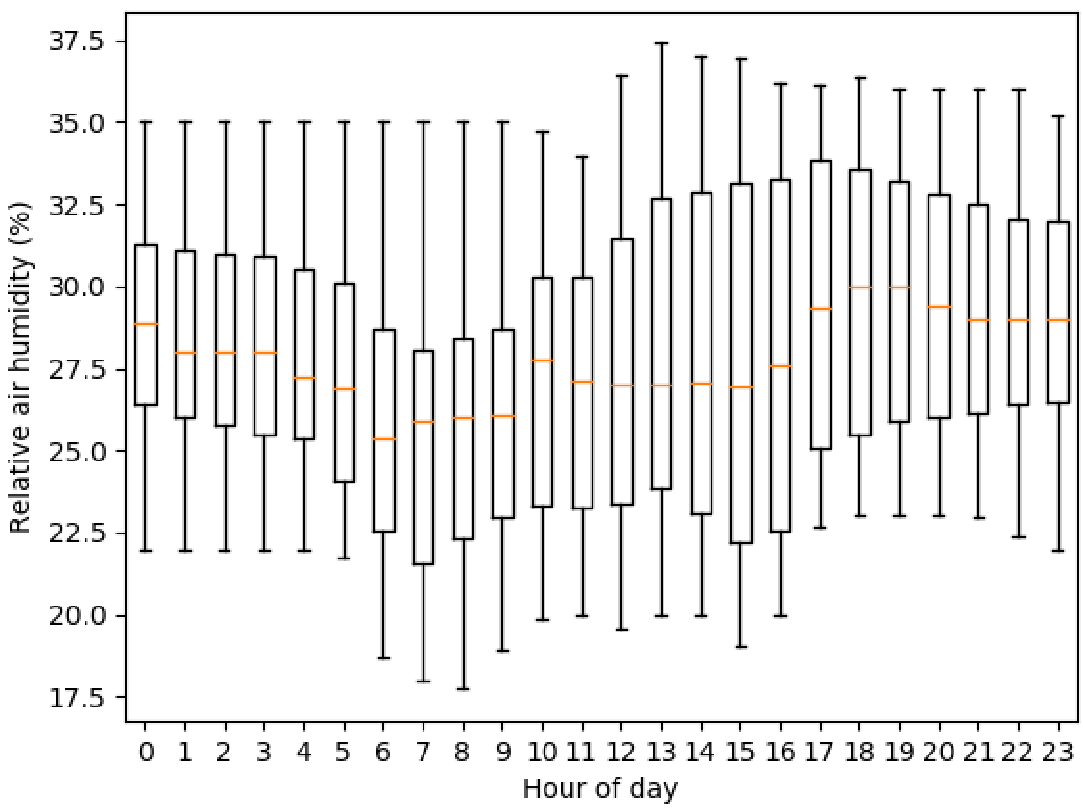

Figure 6.25: Relative air humidity re-sampled and binned by the hour.

The hourly humidity readings indicate that the HVAC system is able to maintain a relatively stable relative air humidity throughout the day. The median values are below 30 percent relative humidity for all hours, and the highest recorded relative humidity was 37.4 percent. While the humidity level is relatively stable, the median humidity of 27.7 percent is below what may be considered as optimal, according to the studies cited in 2.2. Some students may report sensation of dry eyes and mucous membranes at such levels.

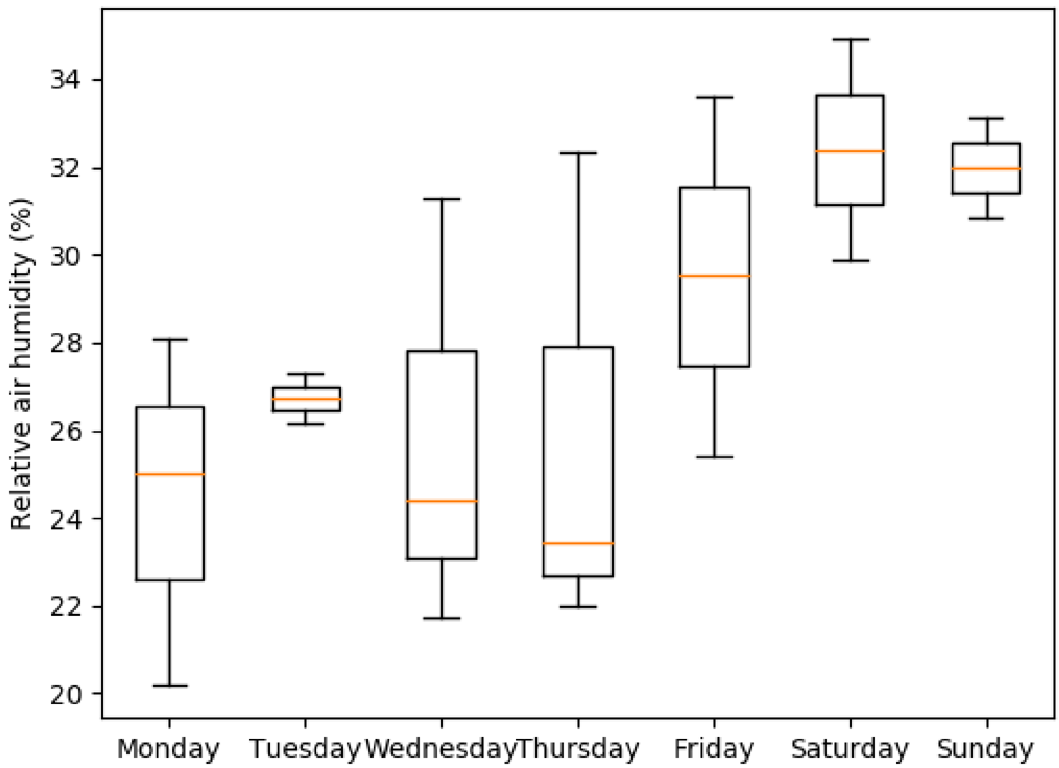

Figure 6.26: Relative air humidity re-sampled and binned by the day.

While the number of data points available when the data is binned according to weekday is low, there seems to be a trend towards higher relative rates of humidity towards the weekend. This may be caused by the pre-programmed reduction in the HVAC systems duty cycle. Further data recording is needed to verify this trend, though it would be in line with other findings.

6.3.3 Noise

The other rooms studied include one or more motion sensors, allowing us to put the other readings into perspective. The data dump from Sandkassa does not contain motion sensor data, meaning that it is difficult to gauge the relative activity an occupancy rate within the room. Using the average sound pressure level for 15 minute intervals, binned by the hour, should provide some of the same context. Higher median sound pressure indicates a higher baseline noise rate, and should correspond to the activity inside the room.

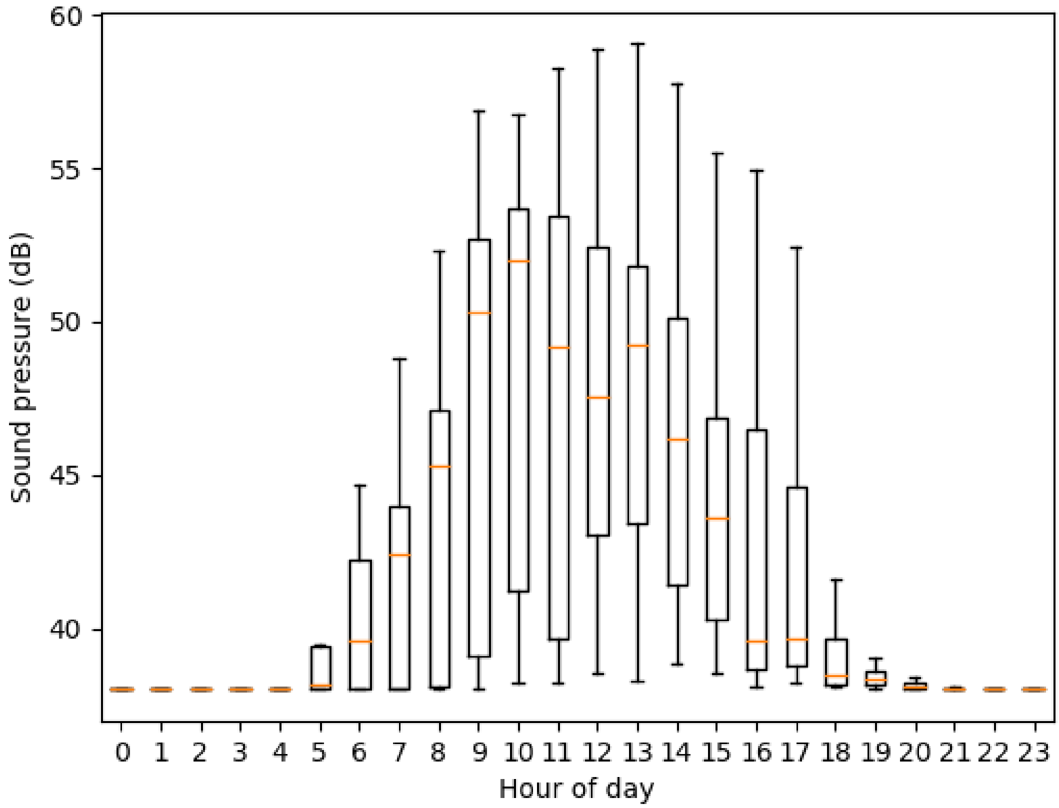

Figure 6.27: 15 minute average sound pressure in Sandkassa, binned by the hour

The noise readings indicate that the room sees little to no usage between 20.00 and 05.00. The activity quickly increases, before reaching a median maximum at 10.00 to 11.00. Activity slowly tapers off until 16.00. The noise readings from Sandkassa are roughly in line with the motion counter data seen in the other rooms equipped with the M+ sensor. While there may not be a perfect 1-to-1 relationship between the M+ and noise sensor readings, a fair assumption is that the average noise level should track the motion counter data closely.

6.3.4 VOC

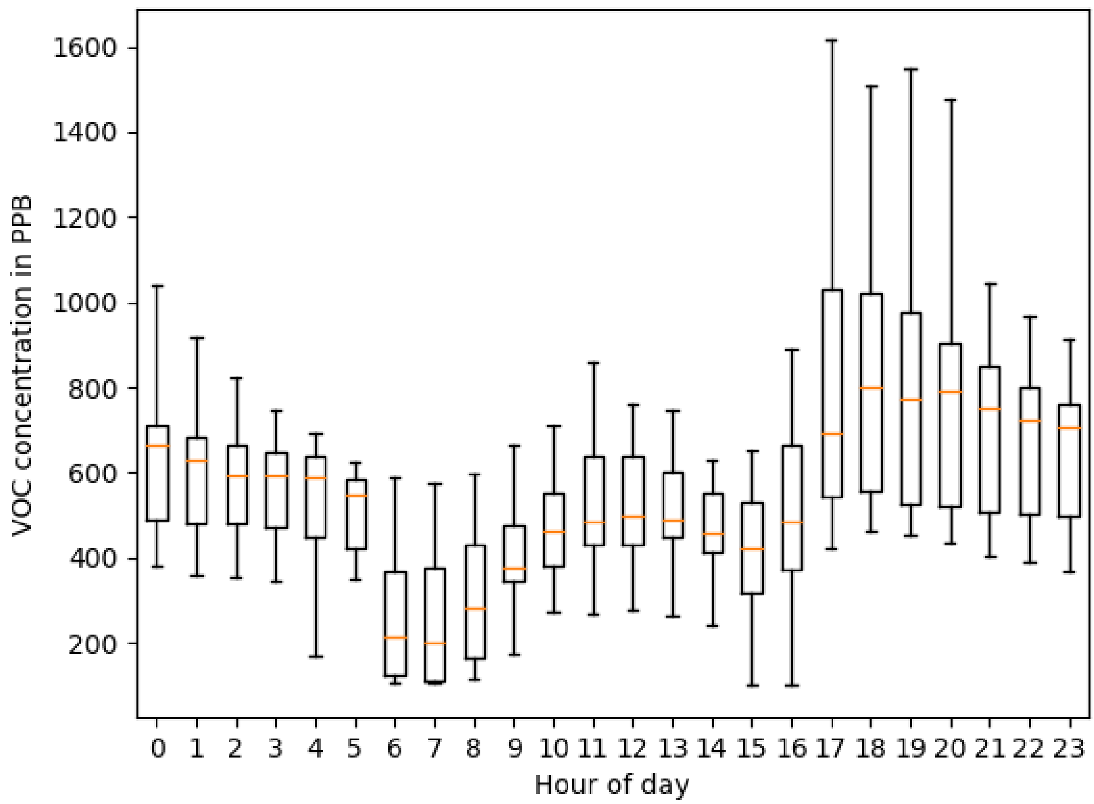

Figure 6.28: VOC concentration binned by the hour of the day

The abnormal readings from Monday are most likely caused by sensor low amount of data when binned by the day. Only two full days' worth of data were collected for Monday in the data-set used in this thesis. The median values of Saturday and Sunday exceed all other days, with the exception of Tuesday. This is most likely caused by the reduced HVAC ventilation rate during the weekend. More data needs to be collected before robust conclusions can be drawn.

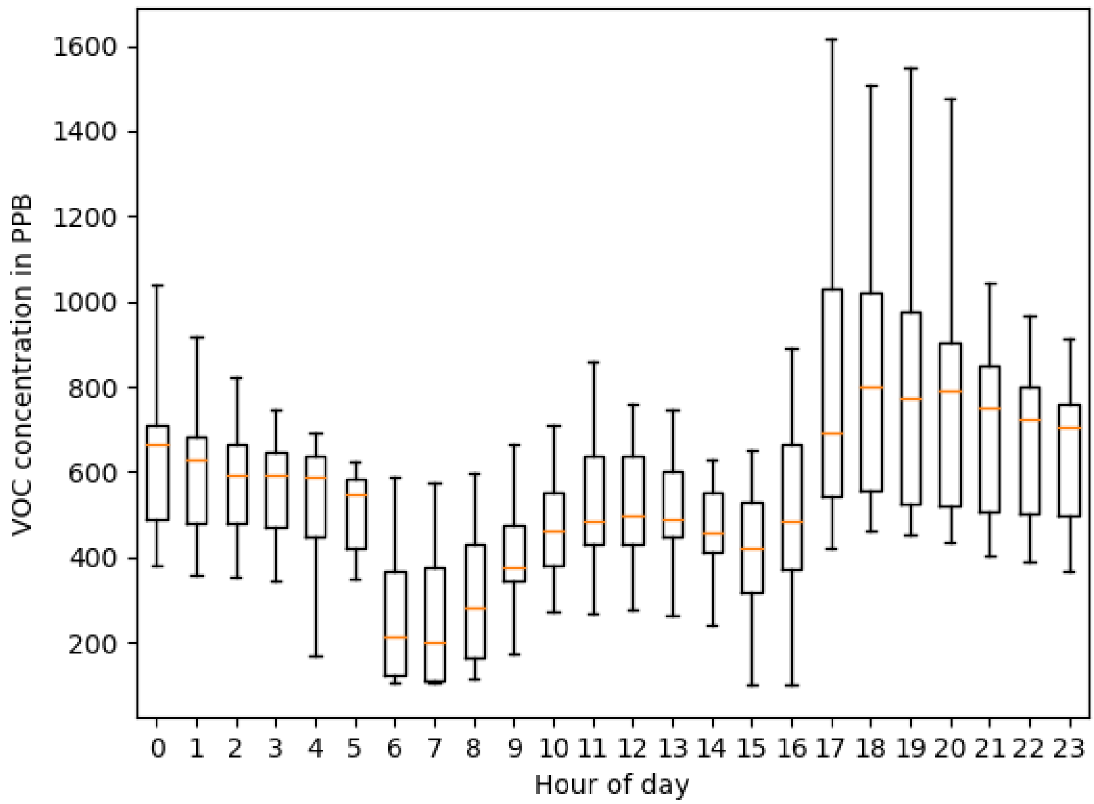

Figure 6.29: VOC concentration binned by the day of the week

A sharp rise in the hourly VOC values are observed from 16.00 to 17.00. This is most likely caused by the reduced operating speed of the HVAC system. A sharp decrease is also observed between 05.00 and 06.00, indicating that the HVAC system is programmed to increase airflow at those hours. The source of the VOCs are unclear. Sandkassa contains no lab equipment or other items that are expected to produce large amounts of VOCs, it is therefore most likely that the VOC originates either from the occupants themselves, the items the occupants bring into the room as well as any VOC produced by the surfaces of the room and its building materials. It is also possible for the HVAC system to distribute VOC from other rooms due cross-contamination via the air circulation.

6.4 Koopen

Koopen has two COM sensors installed, one in each corner of the room as well as three M+ sensors mounted at the entrance to the seating area and inside the seating area itself. Koopen consists of a working area made up of approximately 50 chairs, split over 12-15 tables. The working area is inside a large open building and the working area is sectioned off from the rest of the building by some bookcases and partitioning walls. This setup means that the working area's effective air volume to number of seats ratio is very high. This will affect both the CO2 readings, as the exhaled CO2 is diluted in a much larger air volume than in the other rooms examined, and the temperature readings, as the total amount of air to be heated is significantly larger than any of the other rooms examined in this thesis.

Since Koopen is equipped with two COM sensors, and thus logs the same environmental factors twice in a single time period, the sensor readings are averaged across the two sensors. This should reduce sensor noise as well as give a good understanding of the aggregate values of the environmental variables recorded. The motion data is also averaged across the three sensors, to get an understanding of the total use of the room, and not within the room itself.

6.4.1 CO2

Koopen contains two COM sensors, each measuring CO2 concentration. The data from these two sensors were combined, and the average values of the sensors were used in the tables and calculations below.

Weekday	**Mean**	**Min**	**Max**
Monday	590.87	400	1021
Tuesday	676.08	398	1080
Wednesday	662.17	397	1162
Thursday	576.39	398	1063
Friday	527.01	398	995
Saturday	459.47	399	789
Sunday	505.6	399	955

Table 6.9: CO2 concentration data in PPM

Koopen has very good CO2 readings, the maximum observed value was just 1163 PPM. While this level has shown to be detrimental to student performance, as per Chapter2.3, the effect should be relatively small. One potential explanation for the good numbers at Koopen stem from the fact that the air-volume of the building compared to the floor space occupied by the room ratio is very high, ensuring that the exhaled CO2 is distributed and diluted lowering the readings. While some values below the expected floor of 400 PPM were measured, the offset was relativity small, indicating that the sensors are well calibrated.

A closer examination of the daily readings show that Monday through Wednesday are the only days with CO2 concentration in excess of 1000 PPM. This is in line with the motion data shown in 6.33. Monday to Wednesday show the highest CO2 concentration, and the heaviest use. While the cumulative motion events for Thursday are close to those of Wednesday, a lower CO2 concentration is observed. This indicates that the combined occupancy of the room is similar for both days, but that the usage on Thursdays is distributed more evenly over the course of the day.

Weekday	Readings	Readings Above 1000 PPM	Readings above 1500 PPM	**%>1000**	**%>1500**
Monday	501	10	0	2.00	0
Tuesday	573	67	0	11.69	0
Wednesday	704	97	0	13.7	0
Thursday	631	0	0	0	0
Friday	576	0	0	0	0
Saturday	576	0	0	0	0
Sunday	576	0	0	0	0
SUM	**4137**	**174**	**0**	**4.21%**	**0%**

Table 6.10: Readings relative to the limits set out in 2.3

6.4.2 Temperature

Since the total air-volume of Koopen is so large, the inertia of the temperature is expected to be higher than in many of the other rooms in the pilot study. Koopen also heavily features glass roof and some glass walls. This should help the building absorb energy from the sun light and the concrete floor will act as a thermal mass during the later hours of the day. The glass features do come with a drawback; glass is a poorer insulator than non-glass walls and roofs and this should speed up energy loss during the hours of little to no sunlight.

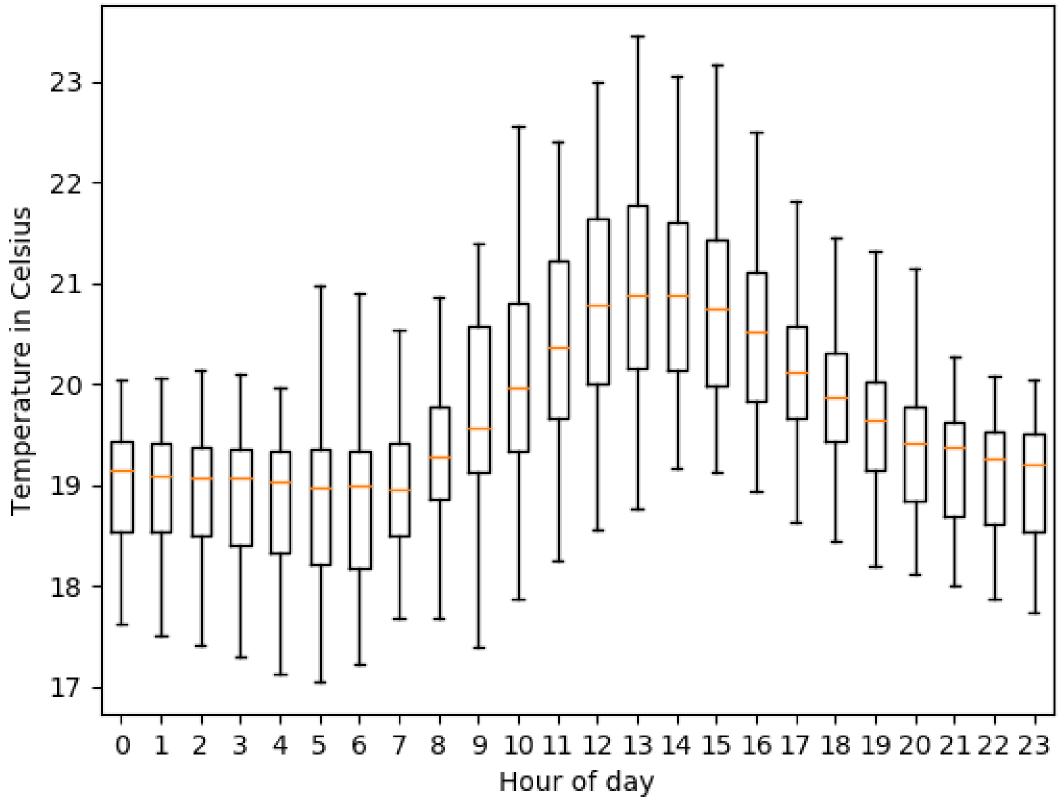

Figure 6.30: Koopen temperature re-sampled to 60 minutes and binned by the hour of the day

The hourly temperature readings show that the median temperature for the hours 21.00 to 07.00 remains more or less constant. The tail ends move towards lower temperatures as the night progresses, indicating that heat is lost from the system during night time. This is to be expected as the temperature control system is set for a lower ambient temperature in the evening and night hours to conserve energy. No outside temperature readings were taken during the time period, making it difficult to gauge the effect of outside temperature on the internal temperature readings.

The hourly temperature readings for Koopen are also much smoother and than those found in Smia, fig. 6.1. This may be caused by the increased total air volume, leading to higher temperature inertia, causing a smoother curve and a less sharp response curve.

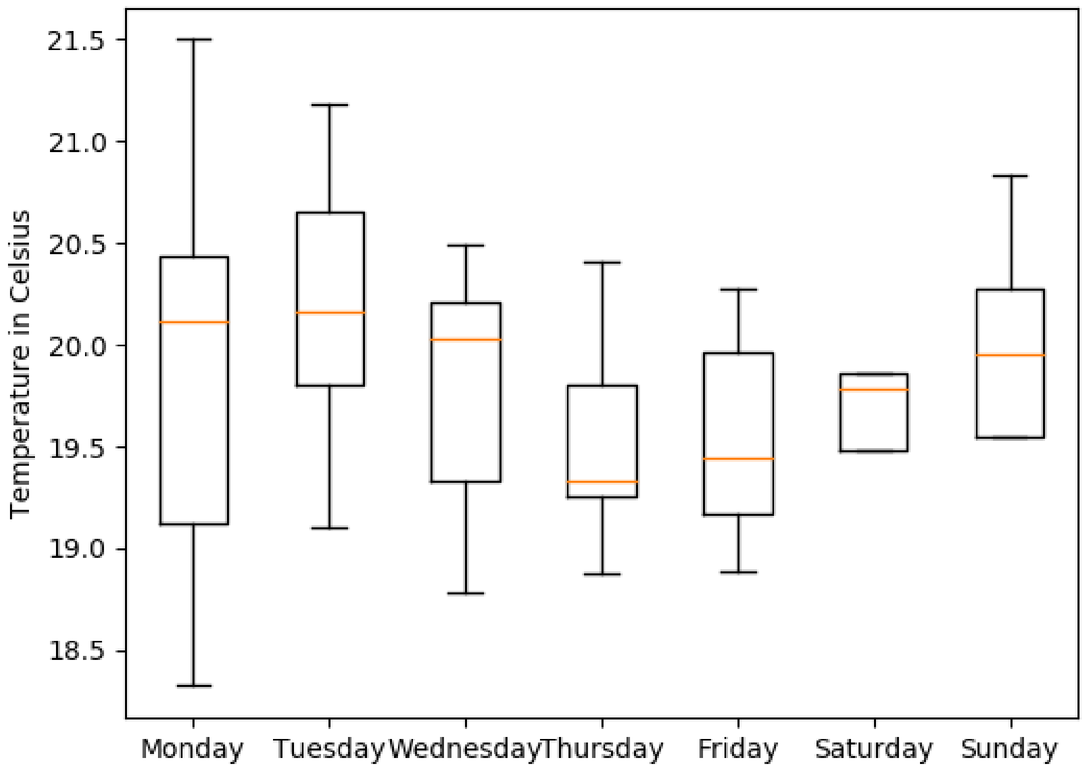

Figure 6.31: Koopen temperature re-sampled to 1 day intervals and binned by the day of the week.

Unlike the daily temperature readings of Smia, the median readings at Koopen show little to no sign of being lower during the weekend days. Since the period time period of sensor recordings only encompasses about 5 weeks, it is difficult to draw any clear conclusions based on such a limited dataset. It does however appear that the temperature control system operated at Koopen is not based on the day of the week, or that the temperature outside and the incoming sunlight is having a large effect on the system's ability to effectively control the temperature.

6.4.3 Motion data

All registered motion data from the 3 three motion sensors installed at Koopen was combined and re-sampled. This is to give a better indication of the room usage as a whole, rather than measuring smaller parts of the work space. One of the motion sensors was mounted in such a way that they could theoretically pick up someone walking close by the entrance to the work space, without actually entering. It is unknown how much this affected the results as there are no supplementary readings available to filter out such occurrences.

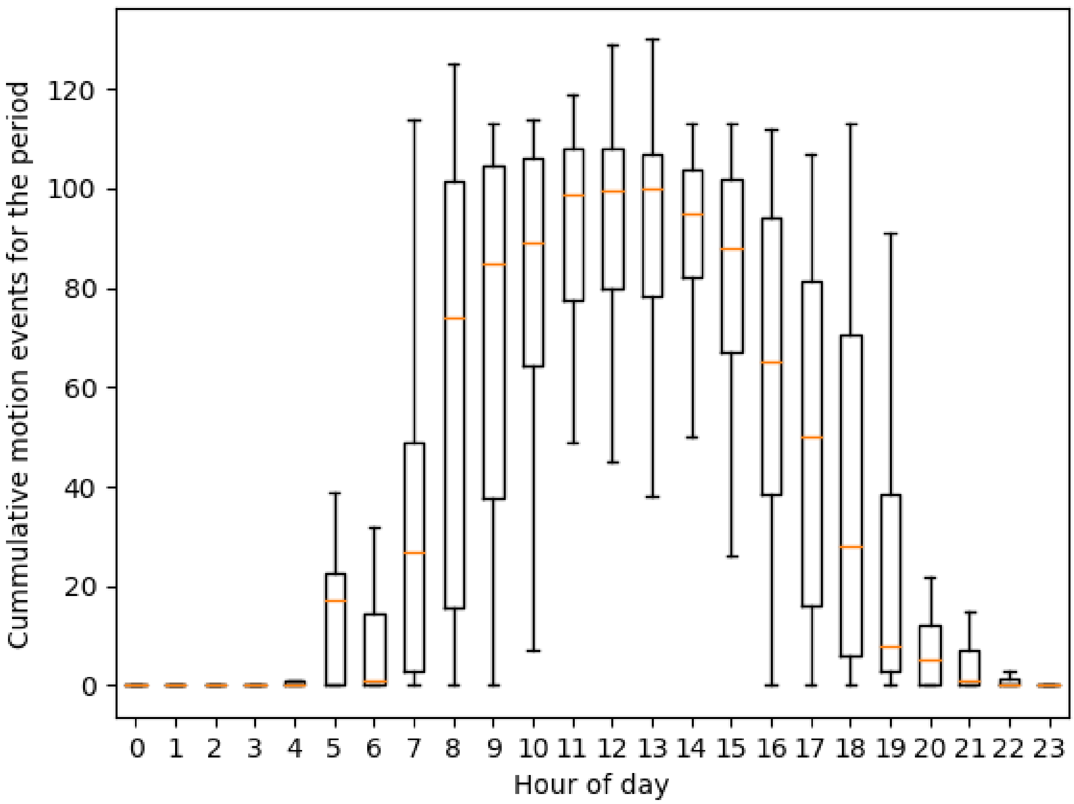

Figure 6.32: Motion counter for Koopen, re-sampled to 1 hour slices and binned based on hour of day

The hourly motion sensor data indicates that Koopen sees its most heavy use between 9.00 and 15.00. This is in line with the data registered at Smia, fig. 6.1.3, all though the traffic at Koopen drops off earlier than at Smia. An interesting observation is the registered movement between 05.00 and 06.00. While these movements may be caused by students, it may also be caused by people passing near the entrance and being registered despite not entering. The author does not have any information on when maintenance, cleaning and other facility services is carried out in the building Koopen is in, but it could very well be caused by maintenance operations such as cleaning.

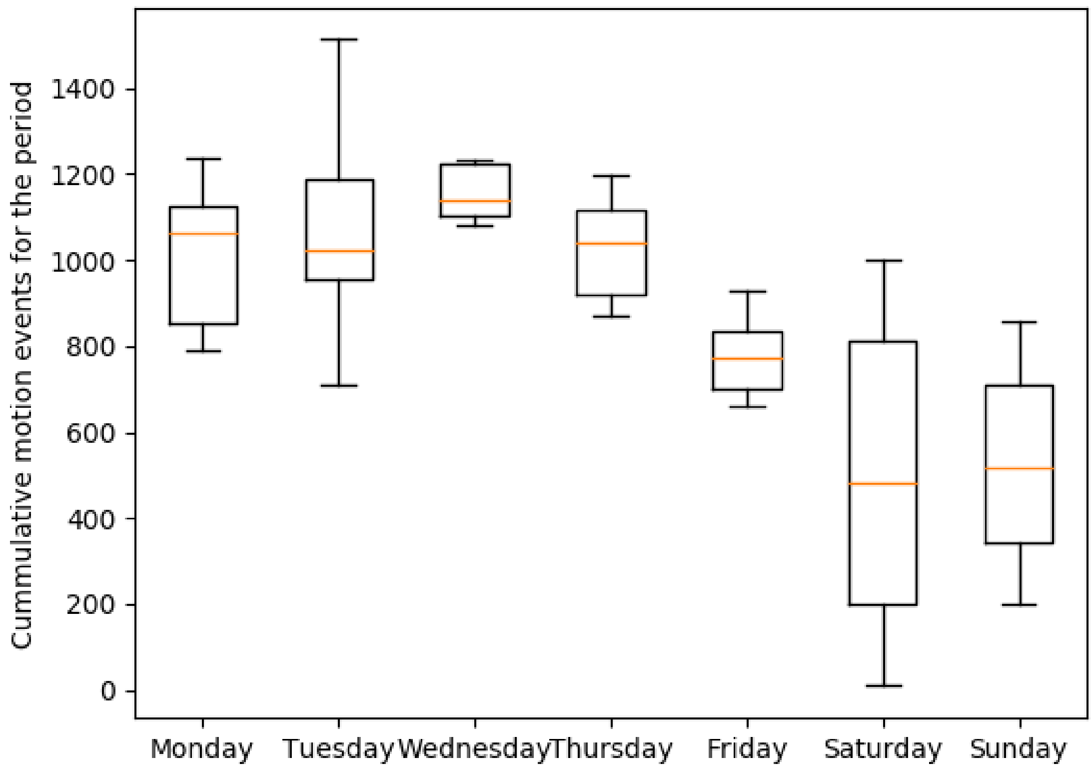

Figure 6.33: Motion counter for Koopen, re-sampled to daily and binned based on weekday

The daily motion sensor data shows that Monday through Thursday show very similar usage patterns, with median and average values grouping closely. The usage drops down on Fridays, and remain low on Saturdays and Sundays. The data shows that students make use of the work space even during the weekends. This may be due to the fact that most of the sensor data was collected late in the semester, and that weekend days see more use as the exams are approaching.

6.4.4 Humidity

The hourly humidity data for Koopen indicates that the HVAC system is able to regulate the humidity well. The largest difference between Q3 and Q1 for any given hour is about 30 percentage points. This indicates that the moisture level fluctuates over time, most likely caused by differences in the humidity of the outdoor air over time.

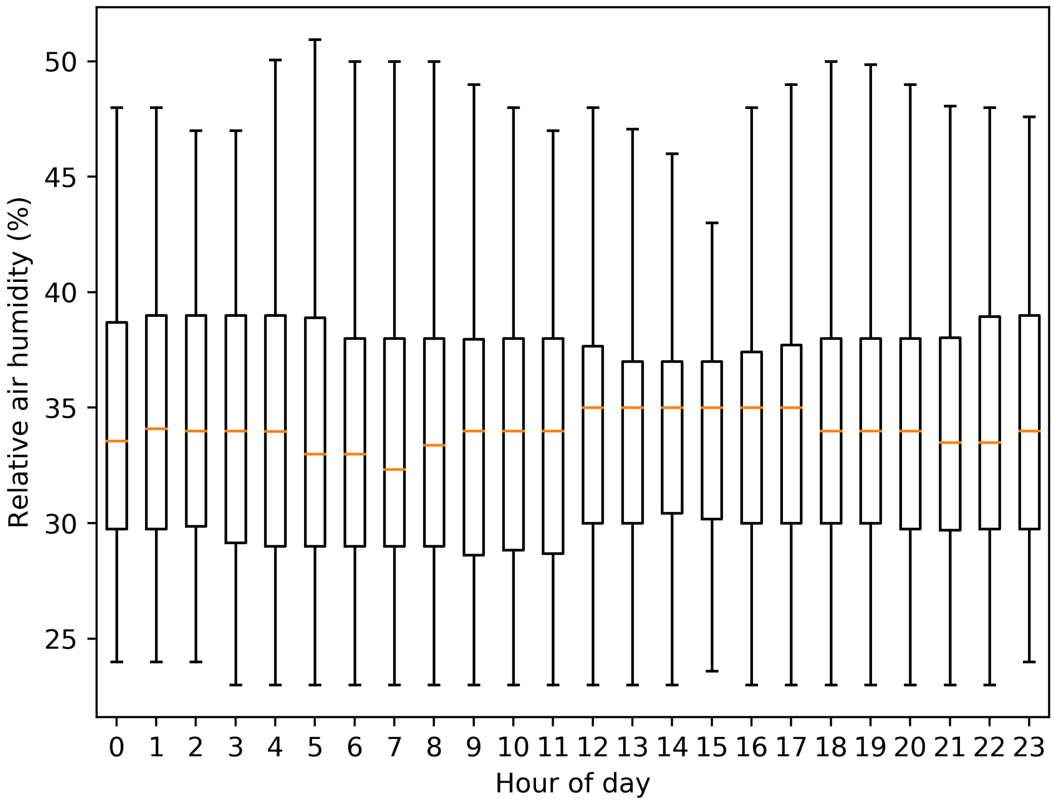

Figure 6.34: Relative humidity for Koopen, re-sampled to hourly intervals

The building Koopen is in also features some large climbing plants and trees planted inside in concrete beds sunk into the floor. These plants are regularly watered, and it is possible that this leads to some added air humidity through evaporation of the top layers of the soil. Plants can also release large amounts of water from their leaves during photosynthesis, this process is known as transpiration. The transpiration rates of plants vary wildly (Rawson et al., 1977), and the author did not attempt to measure or estimate transpiration rates for the particular plants and trees in the building. Any effects caused by the plants transpiration or watering of the plants is not visible in the dataset.

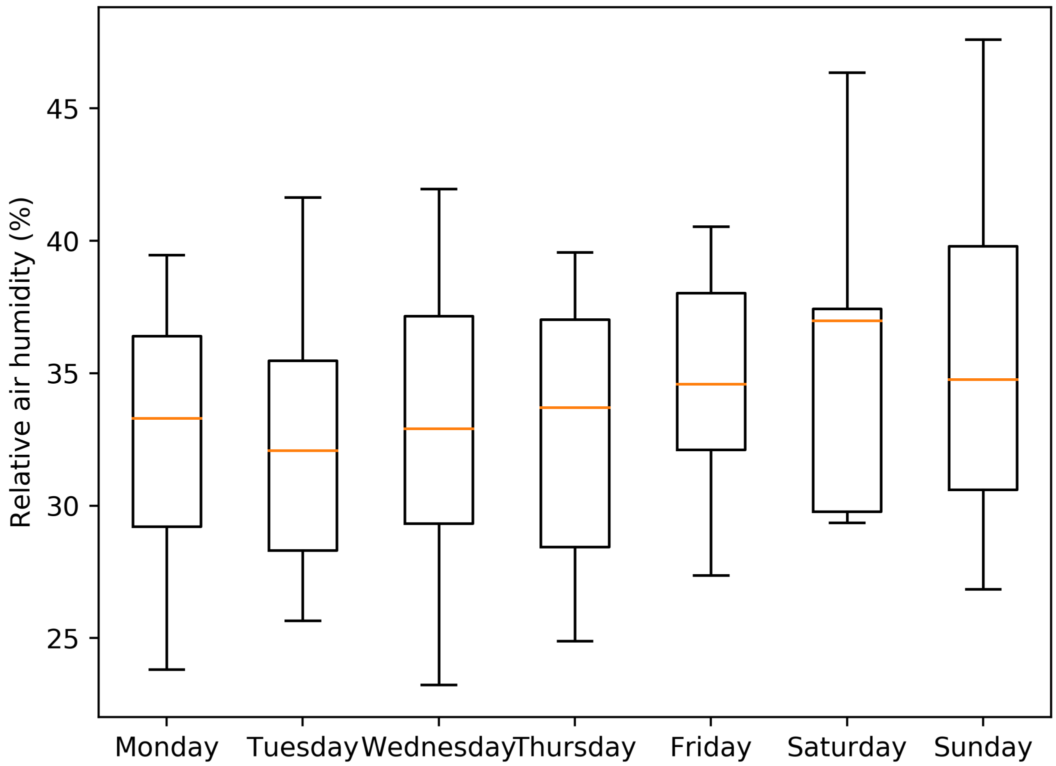

Figure 6.35: Relative humidity for Koopen, re-sampled to daily intervals

The daily humidity readings indicate a slow rise in median humidity from Tuesday to Saturday, before the median start falling again towards Monday. This may be caused by changes in the outside air humidity, and it is difficult to draw any definite conclusions as the dataset contains too few daily readings to draw a robust conclusion. The rise in humidity may also be caused by gradual build-up of humidity due to occupants' actions and presence, which slowly tapers off as the HVAC system is able to correct it during the slower weekend period.

6.5 Discussion of findings

The data collected at NTNU show some interesting findings with regard to the benchmarks set out in chapter 2. First The CO_2 concentration is generally speaking too high, especially during the weekends. This is most likely caused by the reduced HVAC ventilation rate during the weekend due to the pre-programmed schedule. The readings for the rest of the week also show elevated CO_2 readings, many of which are so high that they may impair students' ability to perform optimally. All the data was collected during fall and winter months, a period when doors and windows leading outside are less likely to be opened. It is possible that the CO_2 readings will be better during the summer, as students open the windows and doors to let in the fresh summer air. Despite this, the HVAC system should be designed, operated and sized to keep indoor CO2 levels acceptable without needing to rely on occupant' actions such as opening windows. Sandkassa saw very high CO2 readings, often over 1500 PPM, a level that has shown to be detrimental to learning.

Second, the temperature readings recorded across the rooms show that the temperature control is generally speaking effective for all the rooms across the dataset. Some rooms show a high degree of thermal inertia, taking a considerable amount of time to change the temperature in either direction. While this trait may be desirable as it creates stable and slowly changing temperatures, it also means that temperatures outside the desirable range are hard to correct within a reasonable amount of time. Koopen's placement means that the amount of sunlight collected pr square meter of seating area is very large, and this could create situations where the temperature is too high for comfort. This is probably most pronounced during the summer, when the days are long and the solar potential pr square meter is the largest. None of the data contained in the dataset encompasses the summer months, so no clear conclusion can be drawn as to how the indoor temperature responds to increased outside temperatures. The group rooms in U1 were interesting in that they behaved differently from the other rooms and work spaces studied. Instead of the temperature dropping during the evening and night hours, they stayed relatively high until the HVAC system was switched over to normal day-running speed and the temperature fell. The rooms largely had temperature readings that were within acceptable margins. None of the rooms had temperatures that were excessively high, but Koopen's temperatures were on the low side. This could be due to the heat loss occurring due to the glass walls and ceilings providing poor insulating capacities.

Third, humidity readings across the rooms contained in the dataset indicates that humidity levels are well regulated, although there seems to be an increase towards the weekend. This is probably due to the reduced HVAC setting. Even though the humidity readings are stable, the median readings are below the recommendations laid out in 2.2. It is possible that the studies cited are too pessimistic as to the effects of low relative humidity. Tracking subjective experiences of the indoor climate and any eye or mucosal membrane complaints and correlating it to the indoor humidity may shed some light on the status quo. This is however outside the scope of this thesis.

Fourth, VOC readings indicated that there is a great degree of difference in the concentration through-out the days. The maker space at U1 had surprisingly low VOC readings, despite the room having both 3D printers and soldering stations. It is possible that the VOC sensor do not pick up the gasses produced by the equipment, or that they were not operated extensively during the data collection. Due to the lack of clarity as to exactly which VOCs are measured, it is not possible to draw any conclusions as to the potential health or learning effect of the VOC levels observed.

6.5.1 Limits, constraints and future work

Due to a relatively short time of data recording, it is uncertain how representative the dataset actually is. Based on the author's personal impression of campus activity, there is a stark difference in room usage, working patterns and other occupancy patterns in different parts of the semester. The reading rooms are usually quite full close up to the exam period, and the traffic drops off steeply as students finish their exams and go home for winter or summer holidays. It is fair to assume that the environmental factors recorded and discussed in this chapter will be affected by this, but specifically in what way and by to what degree remains unclear until data can be collected for a longer period of time. The data collection period itself therefore represents a limit and constraint to the validity of the results.

There were also issues with some of the sensors. The author noticed during check-ups that some sensors were unplugged, especially the COM sensors, as students were using the wall outlets to charge their cellphone. Sometimes the sensors were not plugged back in once the student had finished using the outlet. This reduces the data collection, and happens during the vital period where the room is actually seeing occupant traffic. The footfall sensors were also problematic when it came to mounting them. The first mounting attempts involved adhering the sensors to the ceiling plates using double-sided tape. This fastening technique was not sufficient enough to hold the sensors in place, and they would eventually come loose and fall down. This meant that no detailed data on the rate of which people entered and left the room was recorded. This data could have added a much needed dimension to the motion counter data and allow for even better understanding of the movement patterns of occupants in different rooms.

Future work should include the acquisition of qualitative data from room occupants based on questionnaires and in-depth interviews. Correlating this with sensor readings can give important insights to which environmental factors affect people's experience of working in an indoor environment. It could also give important cues as to how the indoor environment can be changed to maximize occupant health, well-being and work performance. Another interesting avenue would be to install additional sensors. These can for example monitor the rate of which windows are opened, to better understand how this affects temperature and CO2 of rooms. Furthermore it would also be interesting to see whether or not the rate at which windows are opened correspond to high temperatures, high CO2 levels and subjective experiences such as "stale air". This could be used as a proxy measurement of which environmental factors reduce well-being among occupants.

Chapter 7

Predicting future occupancy

This chapter details the implementation and results of a feed-forward based neural network model to predict future occupancy in a room based on historic trends coupled with current sensor readings. Data was extracted on a room-by-room basis and re-sampled into different time periods. The optimal time period for the re-sample function was found through a simple trial-and-error approach. The model was continuously evaluated and iterated upon as shown in Peffers et al. (2007). Each tunable parameter of the model was changed individually and the outcome evaluated. The proposed model is by no means proven to be mathematically optimal, but the results are good enough to give a reasonable approximation of future occupancy.

The chapter assumes that the reader has some prior knowledge of neural networks, their implementation and concepts such as ground truth, dropout layers, epochs and so on. For a quick, and very basic introduction to neural networks, please consult **Appendix B**.

Due to time constraints, the model was only trained on the motion counter data from Smia, so the results may not be applicable to the rest of the rooms in the dataset. 25 percent of the motion counter data from Smia was used as validation data, and the model was never trained on this part of the dataset.

7.1 Background

Increasingly strict energy efficiency demands are placed on buildings in order to reduce greenhouse gas emissions and reduce the long-term impacts of global warming. The European Union has set a goal to reduce greenhouse gas emissions by 20 percent in 2020 compared to the emissions of year 1990 (Štreimikienė and Balezentis, 2016). Schools and universities are not exempt from the attempts made to lower energy usage, and ongoing efforts are made to reduce the environmental impacts of these institutions, both in terms of emissions from day-to-day activities as well as during construction and planning. Indoor environmental variables are not only about optimizing and reducing energy usage, but also maximizing the occupants' comfort levels.

In some cases, simple changes such as replacing the ballast of old fluorescent lighting tubes can yield significant energy savings and thus reduce emissions (Di Stefano, 2000). Installation of automatic light switches based on movement may also yield significant energy savings without reducing occupant comfort. Lighting is one area where energy savings can be obtained relatively easily, switching on and off lighting is instantaneous and optimal comfort level can be reached within seconds. However, as described in **chapter 5.2**, some of the variables making up the total indoor environment are "slow variables", whose system inertia means that it takes a considerable amount of time for peak occupant comfort to be reached. For those slow variables, it is therefore important to estimate indoor occupancy before it actually occurs, so that the systems can adjust and reach optimal comfort before the first occupants arrive. While there has been extensive research on present building occupancy based on a wide array of measurements, such as CO2 (Han et al., 2013), little research has been done on predicting future occupancy. The following sections will detail the implementation of a NN based solution that accurately predicts future occupancy before it occurs based on historic data and live sensor readings.

7.2 Neural network model

While there are many different neural network models and methods to deal with prediction of time-series, such as Hybrid ARIMA (Zhang, 2003), feed-forward and fuzzy networks (Gao and Er, 2005), recurring networks (Zhang and Xiao, 2000) and convoluted neural networks (CNN) (Yang et al., 2015), a simple sequential model using a combination of dense and dropout layers was chosen. This is due to two factors. First the initial prototypes showed that the sequential model using dense layers produced superior results compared to CNN and recurrent networks. Secondly the author is most experienced with using sequential dense models, allowing for faster prototyping and iterations. It is possible that other types of neural networks may provide similar or better performance than the proposed model, but time constraints meant that only a single model could be constructed and thoroughly tested.

The neural network consists of a series of input neurons. ten neurons are used as input for the ten latest historical motion counter readings, stretching back 480 minutes (12*40 minutes). 1 neuron is used as input for the current motion counter reading for of the last 40 minutes, and a single neuron whose input value is derived from the current time. The time neuron is described in more detail in 7.2.4.

7.2.1 Data processing

Since one of the methodologies used in this thesis is a data-driven approach, as detailed in 3.2, a natural starting point is to take a look at the data itself. Below is the plot of motion counter data from Smia over a period of 7 days. The data has been re-sampled as described in 5.4, and a re-sample period of 40 minutes was chosen. This re-sample period helps reduce sensor noise by reducing the impact of a single bad reading.

Visualizing the motion counter data immediately uncovers an interesting, yet not unexpected, cyclical pattern. As shown in fig. 6.3, the periods between 22.00 and 06.00 have very little to no

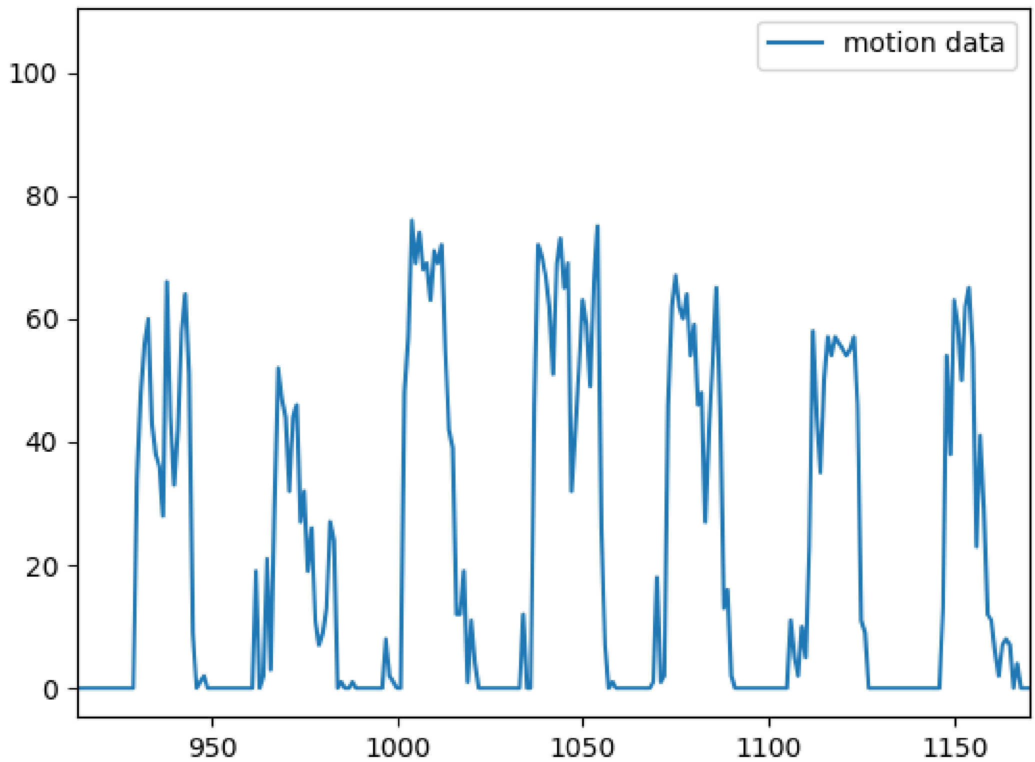

Figure 7.1: Motion data from Smia taken over a 7 day period, re-sampled to 40 minute intervals. Y axis is the 40 minute summed motion events and X axis is the line number in the data dump.

motion activity, and the periods between 10.00 and 18.00 show the most activity. This causes a cyclical pattern with lows between the periods of 22.00 and 06.00 an high between 10.00 and 18.00. The average period of the first recorded 0 motion counter for a 40 minute interval and the last recorded 0 motion counter for the 40 minute interval (falling and rising edge) is a about 13,27 periods, or about 8 hours and 50 minutes [1]. This corresponds well with fig. 6.3.

While the amplitude of this cyclical pattern changes from day to day, the time between high and low motions period seems to remain relatively constant. Any model trying to predict future motion and thus room occupancy should therefore incorporate time of day as an input. The finding of this cyclical pattern was used as a basis for the decision to use time as one of the input neurons. This is explained in more detail in chapter 7.2.4.

The initial dataset before re-sampling consisted of 98407 sensor readings, and re-sampling to 40 minute intervals leaves 1837 sensor readings for training and validation. While an even longer re-sample period may be chosen, it does reduce the amount of data left for training and evaluating the model. Since the dataset available for this thesis was rather limited, a 40 minute

[1] 13,27 periods of 40 minutes * 40 minutes = 530.8 minutes

re-sample interval was chosen to reduce sensor noise significantly, yet retain enough data for the training and validation of the neural network model. No other transformation or processing of the data was outside the previously mentioned re-sampling, the normalization detailed in chapter 7.2.2 and the time data to sine function described in chapter 7.2.4 was performed.

7.2.2 Normalizing data

While the normalization of data used in training of neural networks is optional, data normalization tends to increase learning speed and reduce errors (Sola and Sevilla, 1997). There are many different techniques for normalizing data, depending on the dataset, as well as the relative size of the inputs to be used. The input neuron for the time is normalized in such a way that the input value stays between 0.0 and 1.0, as given by the formula in chapter 7.2.4. It therefore makes sense to normalize the motion counter data to the same interval. A very simple and efficient way of normalizing input data is to divide each value in the input dataset by the maximum value observed, something which reduces the input space to the interval 0.0 to 1.0. This technique was used for the motion data used as input. The lowest observed motion counter data re-sampled to 40 minute intervals was 0. The largest observed value was 179. The normalized data can easily be converted back into a non-normalized scale by multiplying the normalized sensor readings by 179.

All data used in the training of the model was normalized, but all plots derived from the neural network model was de-normalized and represented in original scale.

7.2.3 Dealing with historical data using a sliding window approach

Using a dense model with input neurons for historical data necessitates transforming the historical sensor data into discrete inputs. There are many different ways of accomplishing this, and the author decided to split the historical data into partly overlapping vectors of a fixed length. This approach was selected based on initial prototyping attempts and personal experience.

A small python script was created to automatically split the historical data into a list of lists, each containing one set of inputs. The script was written with a single tunable parameter; **lookback** which controls how many historical data points are passed as the output of the function. The overlap between two consecutive lists is *(lookback)-1*. The lookback parameter was tuned based on a trail and error method using an iterative approach, as described in chapter 3.1. In addition to outputting a list of overlapping sensor data, the script also outputs the value used as the ground-truth value for training the neural network. The ground truth value is the n+1 reading, n denoting the index of the last reading in the sublist.

An example of a sample input and the output of the script is given below. In order to help with readability, sensor values have been replaced with the numbers 1 through 8. This should help make it more clear how the input is parsed and split into sublists using the sliding window approach. Please note that the example below is given in Python-like syntax, "`[]`" denotes a list.

The lookback parameter for the example table was set to 3.

Sensor data	[1 2 3 4 5 6 7 8]
Sensor data transformed	[[1 2 3] [2 3 4] [3 4 5] [4 5 6] [5 6 7]]
Ground truth	[4 5 6 7 8]

Table 7.1: Example of data transformation using a sliding window approach

This shows how historical sensor data can be split into smaller chunks suitable for feeding through a neural network, with one neuron for each element in the list.

7.2.4 Time as an input

Since the data used as the input in the model shows a cyclical nature, as detailed previously, it makes sense to incorporate time of day as an input in the model. Initial attempts used the number of minutes elapsed since midnight (0.00) as the input for the time neuron, though this seemed to cause less than optimal results. The reason(s) for this is not clear, it could be due to a relatively small dataset making it difficult to optimize the model across the spectrum of inputs. After some experimentation, it was found that using the number of hours passed since midnight transformed by a simple sine function yielded the best results:

$$x = \sin(2 * pi * \frac{x}{24})$$

Fig. 7.2 shows how the cyclical nature of the motion counter data aligns well with the sine transformed hour since midnight signal. Note that the output of the sine function has been multiplied by 40 to make the plot easier to read.

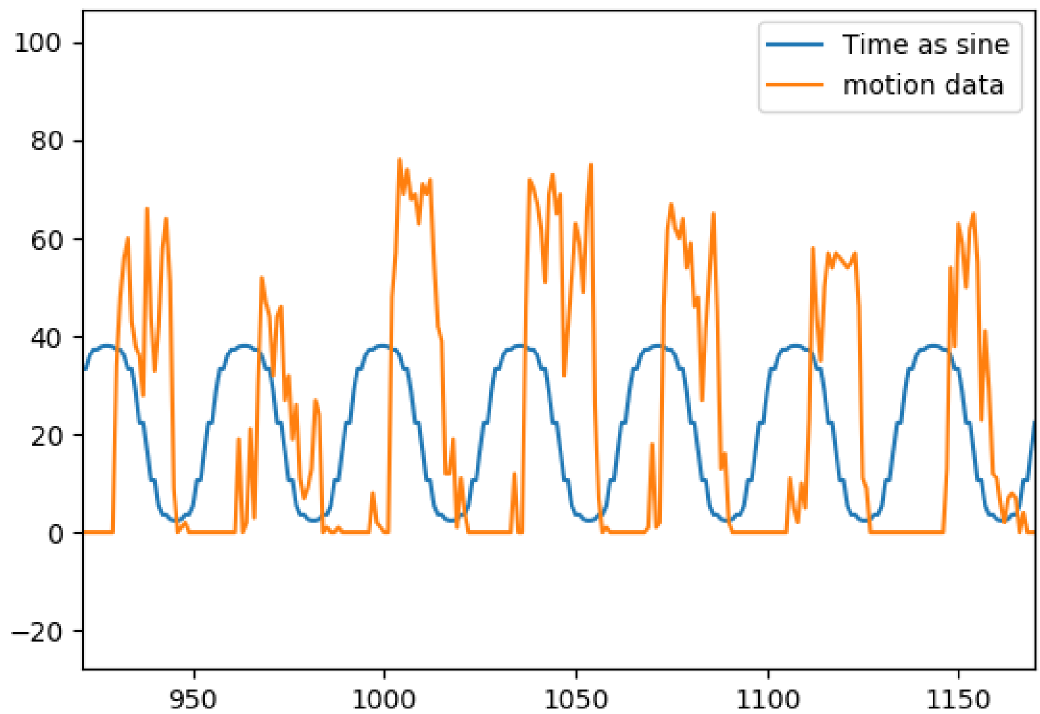

Figure 7.2: Motion data from Smia re-sampled to a 40 minute interval (orange) and the sine transformed hour of day (blue). Y axis is the motion sensor events resampled to 40 mins. X axis represents the line number in the data dump

While the sine signal is out of phase with the motion data, some interesting observations can be made without shifting the sine signal. The bottom of the sine signal aligns well with the steep decline observed for the motion data towards the end of the work day. The top part of the sine signal aligns well with the rising edge of the motion counter data. This sine signal can provide vital context to the prediction model by allowing the model to incorporate the time signal as a marker for the rise and fall of the motion counter data.

The sine signal by itself replicates the current time-based schedule of the HVAC system and the temperature control. In an effort to explore how useful the time signal is by itself, a simple model using only the sine transformed hour after midnight signal was created. This model uses a single input neuron; the sine transformed hour of day and has a single output neuron; the predicted motion counter data. The hidden layers consisted of a dense feed forward network consisting of 250, 100 and 30 neurons. In between each of these layers, a dropout layer with a dropout rate of 20 percent was applied.

The model was trained for 1000 epochs, the optimizer used was Kera's ADAM and the loss function used was Kera's built-in mean squared error. The ground truth the model was trained for was the actual 40 minute re-sampled motion data, without normalization. The input was split into a training and validation dataset with a ratio of 4 to 1, meaning that 25 percent of the total dataset was set aside for validation.

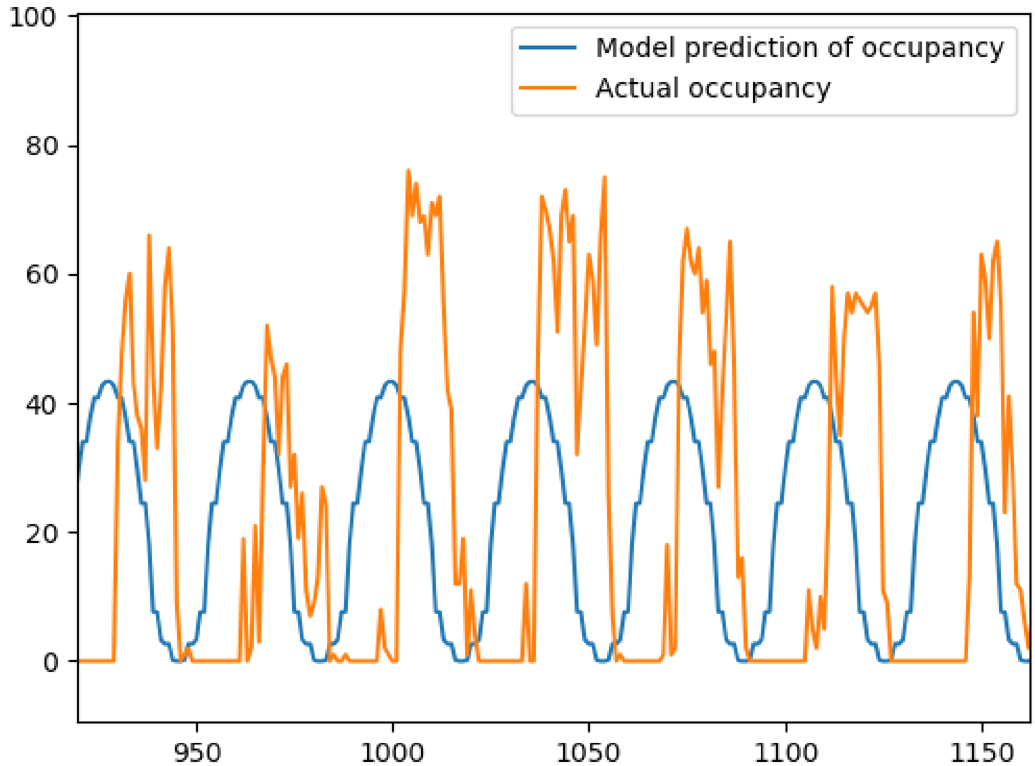

Figure 7.3: Model prediction of future occupancy with time of day as the only input. Y axis is the motion sensor events resampled to 40 mins. X axis represents the line number in the data dump

This simplistic model is obviously ill-suited to predict actual occupancy based solely on the time of day. Shifting the sine-signal to the left or right so that it better matches the cyclical nature of the actual occupancy data should help improve the accuracy, as the sine values better fit the actual data. No further exploration of this was done however, as the proposed model should depart from the current time of day and calendar controlled HVAC system in use at NTNU.

7.2.5 A sample of input data used in the model

The model relies on a combination of historical data, time of day and the current sensor reading.

7.3 Results

This section details the results of the model and shows its performance over a wide range of different parameters. The validation data was only used for validating the model, and it was not used in the training of the model. This is to ensure that the model does not try to predict on data it has already seen before, and thus has been able to fit in the training sessions.

7.3.1 How wide should the sliding window be?

A wide variety of window sized was tested, validated and scored. Using a large window size slows down training, and makes the input layer larger. While a larger window size may allow the model to learn long-term patterns in more detail, it may also lead to over-fitting, where the model learns to replicate specific patterns and does not learn the general shape and pattern of the motion counter data. Please see chapter 7.2.3 for more details on how the sliding window technique is applied.

All losses were computed using mean squared error, and the output activation function for the output neuron is Kera's built-in RELU function. The RELU function was used as it has been shown to produce excellent results across a variety of neural networks (Dahl et al., 2013), (Xu et al., 2015), and the RELU function is readily available within Keras. The model was trained for 10, 100, 250 and 1000 epochs at each window size setting. Each test was run 5 times, and average loss value was computed for each run. Note that the mean squared error measurement has been de-normalized, indicating the error rate for the actual sensor motion data and not the 0 to 1 normalized input.

Window Size	Loss at 10 epochs	Loss at 100 epochs	Loss at 250 epochs	Loss at 1000 epochs
4	20.194	6.169	3.071	11.521
6	12.814	5.051	3.12	5.351
8	12.13	3.205	5.38	7.501
10	9.903	4.99	2.412	5.13
12	17.821	6.301	4.68	9.7433
14	24.792	5.056	4.649	12.981

Table 7.2: Window size and epoch size and their impact on measured MSE for validation data

The data indicates that there is a U-shaped curve for the parameters window size and epoch size with regards to MSE. Increasing window size leads to a drop in MSE, but once the window size becomes too big, the error stops decreasing and starts increasing. The same trend is observed for epoch size. Based on these results, a window size of 10 and an epoch size of 250 will be used for the rest of the results. While it it possible that there exists an even better combination of both window size and epoch size, an MSE loss of 2-3 is low enough for this particular use-case. This combination should represent the local minima in terms of loss, but there may be other combinations that produce better loss values.

A window size of 10 periods and a period length of 40 minutes means that the model will take the last 400 minutes (6 hours and 40 minutes) of movement data into consideration when predicting the next periods motion data.

7.3.2 Predicting future occupancy

Below is a plot of the predicted motion data and the actual motion data (ground truth). The plot consists solely of validation data, and the model has not been trained with the validation data, and has thus had no opportunity to "memorize" this data. The model was given the last 10 sensor input readings, re-sampled to 40 minutes, as well as the current hour of day, as detailed in chapter 7.2.4. The model predicts a single 40 minute period into the future, producing one data point pr set of inputs.

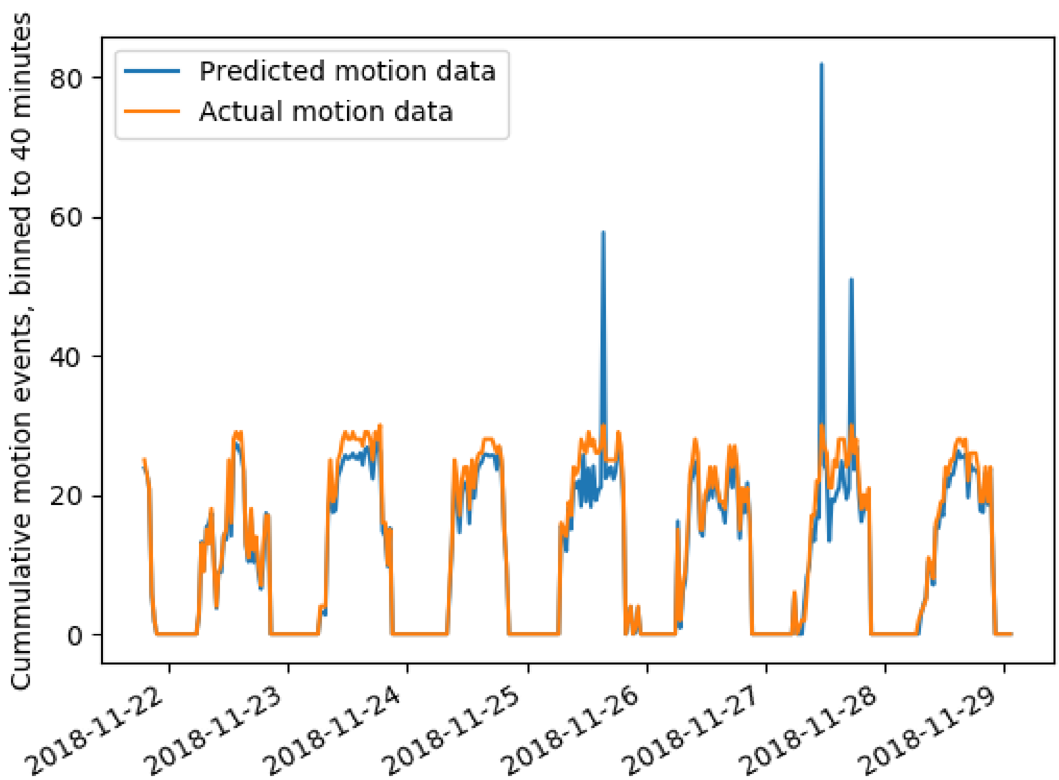

Figure 7.4: Motion prediction based on historical data. 10 window size and 250 epoch

The model does a good job of capturing the actual motion data as recorded by the sensor. It does however have a tendency to produce large spikes. One can be observed in the prediction for 25/11/2018 and two spikes in the prediction for 27/11/2018. The reasons for these spikes are unknown, but could be due to modelling errors, a small size of training inputs or other factors.

The model quickly recovers after making such errors, and the next prediction is much better once the model is fed the next set of input data, and "realizes" that it has made a mistake.

The interesting thing to note is that the predicted graph fits the shape of the actual motion data well. The phase looks correct, and the predicted period of activity aligns well with the actual data. While the model does at time misjudge the size of the cumulative motion events, the timing accuracy is good.

Since the aim of this master's thesis is to produce a model that accurately predicts when room occupancy is going to happen, and not the relative size of room usage, the model should be useful in accomplishing **RG2**.

7.3.3 Transforming the predictions into on/off events

While the model as shown above is able to predict the relative motion in the room with reasonable accuracy, this is not very useful when it comes to controlling the HVAC system or the temperature control of a room. By extending and making some small changes to the model, the output of the model can be used to turn on the temperature control system of a room before occupants are expected to arrive.

In order to achieve this, some changes were made to the input data and the ground truth data. Every reading above 5 motion events per period had its value set to 1, and every period with less than 5 motion events per period was set to 0. This causes the input data to represent people being present or not present, as opposed to the relative amount of activity happening inside the room.

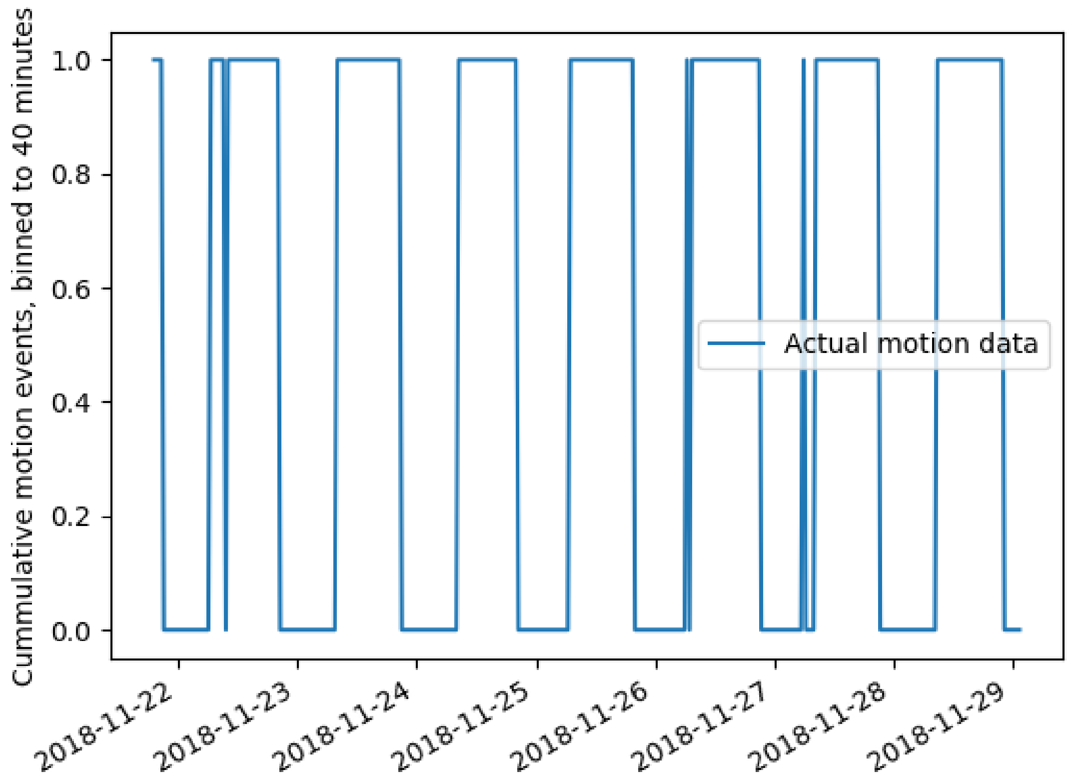

Figure 7.5: Room activity

This transformed data will be used as the ground truth when training the neural network. Instead of training the neural network to predict the relative size and frequency of motion, the network will instead be trained to predict future occupancy. As shown in fig. 7.5, there is some noise in the data, but this is inevitable when dealing with sensor readings.

In addition, the model is trained not to predict the occupancy for the next period, but for two period steps into the future. This means that the resultant model will attempt to predict future occupancy 80 minutes into the future, instead of the previous 40. This will allow slow-responding systems such as temperature control enough time to raise the temperature sufficiently before the first occupants arrive. A successful model will therefore indicate a rise in the predicted occupancy 40-80 minutes before the motion sensor indicates actual occupancy.

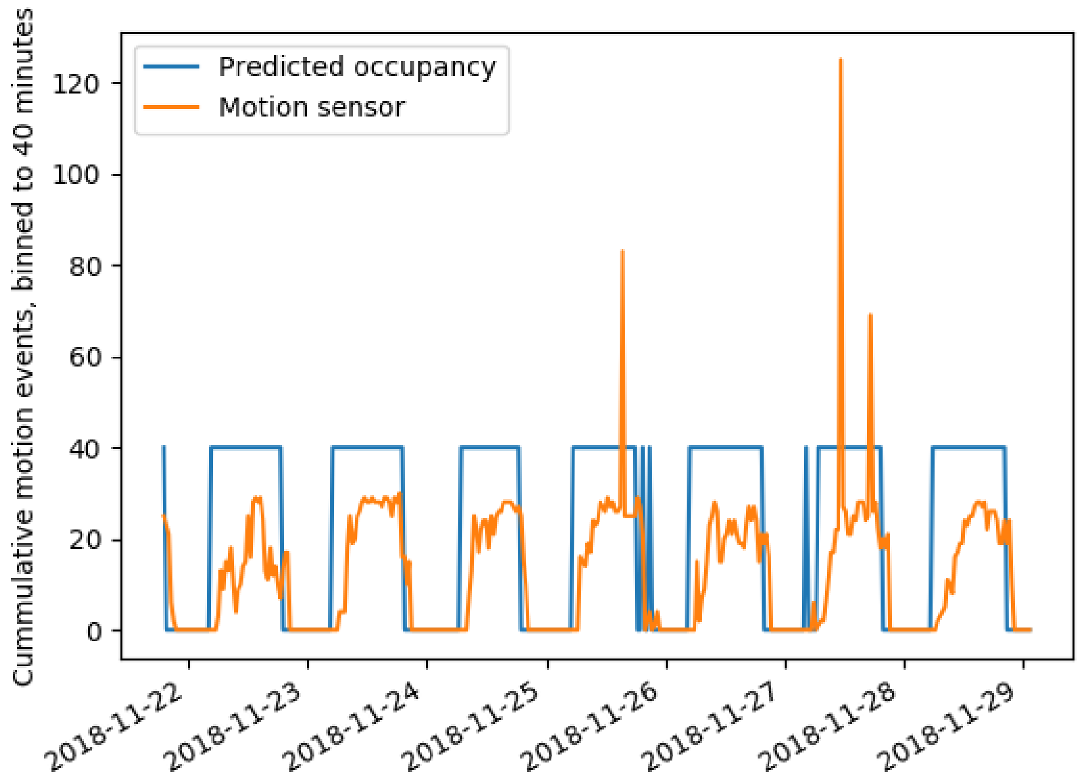

Figure 7.6: Predicted occupancy and sensor data

As the plot above shows, the model is able to predict future occupancy with a good degree of accuracy. The output of the model has been multiplied by 40 from the original 0 or 1 state, to enhance readability when plotted against the actual motion data. The data shown in the plot is from the validation dataset, and has not been used when training the model.

The model is able to correctly predict future occupancy at least 40 minutes before the actual occupancy occurs. While the model does incorrectly predict future occupancy both on 25th of November and the 28th of November, the model quickly adjusts its predictions as new motion sensor data is made available.

7.3.4 Adjusting for non-work days and other outlier days

Since the model makes predictions based on historical data, it has limited capability to predict future occupancy for days that are abnormal, or whose room usage pattern deviates significantly from what the model has previously observed and been trained to model. It is also possible to add additional input neurons to represent other relevant parameters. A single input neuron could for example be used to indicate to the network whether or not the current day is a national holiday, or how much of the semester has elapsed. No attempts were made to implement this, as the data pool available for Smia is too short to properly optimize and test such models.

7.4 Future occupancy and systems with high inertia, practical uses

As described in **chapter 5.2**, some of the environmental variables are slow to respond. This includes, among others, temperature, CO2 and humidity. By being able to predict future occupancy before it happens, systems with slow response times can be pre-adjusted so that they reach an optimal state before people enter and start using the room(s). The temperature in a room can for example be turned up 40-80 minutes before the first occupants arrive, and thus ensure that the room temperature reaches an optimal state before actual occupancy occurs. The current system at NTNU today tries to do this by using pre-programmed on and off times and days for different systems. While this may prove satisfactory most of the time, the system is very rigid and is unable to cope with and learn from new patterns of room usage.

The model shown in this thesis is able to learn and adapt based on historical trends and data, as well as the current sensor readings and can predict future occupancy before it occurs. The system also shows that it is able to adjust and correct itself it if makes a wrong prediction by using new sensor data as it becomes available. If a room sees unexpected usage based on historical trends, the model adjusts is output to reflect actual occupancy.

7.5 Discussion and future work

The proposed model shows positive results. The model is seemingly able to accurately predict future occupancy before it occurs, allowing it to be used for slow variables such as temperature. This would allow increasing the temperature to comfortable levels before the first occupants arrive. The model sometimes misjudges future occupancy, but quickly corrects itself once new sensor data is being made available. Incorporating time as in an input in addition to historical occupancy and current sensor readings allow the model to better simulate future occupancy. In addition the proposed model is less rigid than the current time and date based system, which is unable to respond well to occupancy behaviour that is outside the expected norm. Room usage probably varies over time, especially around holiday season and close to the exam. Having a system that is able to model this behaviour could lead to better indoor climate, reduced energy usage as well as increasing occupants' comfort.

The results of the model are promising; the model appears to accurately predict future occupancy based on historic trends and current sensor data. The model has only been tested an tuned for a specific room, Smia, so it is unknown how well it will perform in other rooms. As shown in **Chapter 6**, there is a great degree of difference in room usage for the different rooms in the pilot study. While most of the rooms have a similar occupancy curve as given by the motion counter data, the model may be more or less accurate for other rooms than Smia. A natural extension of the research presented in this thesis is to apply the proposed model to the complete dataset containing all rooms in the pilot study.

The data used to train the neural network for Smia consists of about 45 days of data, covering the later parts of the semester. A training set consisting of more data over a longer period of time may yield better results and allow the model to incorporate even more nuances of room usage. Future work on the prediction model should focus on data obtained over a longer period of time.

In addition, a long term study comparing rooms with the old system to rooms controlled by the prediction model could uncover any energy savings, as well as any change in indoor climate variables. Such a study would provide useful and robust insights into the practical uses of the proposed models, and allow the testing of the theorized improvements.

Chapter 8

Concluding remarks

In the following, the main findings in chapters 6 and 7 will be summarized and seen in context of the two research goals defined in chapter 1.3.

Research goal 1 was to investigate the current indoor conditions at NTNU and their effect on the students' learning environment. The current indoor climate at NTNU is relatively good. Temperature and humidity seem to be well regulated, but the humidity appears to be somewhat on the low side. This may be caused by the fact that the data was collected during the winter/fall period, where the outside air is naturally drier and helps dry out the inside air when the HVAC system brings in outside air to replace stale air inside the building. The temperature readings will by most occupants be seen as comfortable, but Koopen may see problems with low temperatures during winter, especially if the outside temperature drops to low levels.

The CO2 readings on the other hand ranged from good to very poor. The CO2 levels seen at Sandkassa were especially problematic. Such levels as those observed will have detrimental effect on the mental and physical capacities of the occupants potentially causing dizziness, complaints of stale air and poorer learning abilities. The latter is especially problematic as the rooms are used for both lectures and by students reading and discussing the curriculum. Greater care should be taken to adequately size the ventilation in rooms occupied by a great number of people. The motion counter data shows that most of the rooms see usage outside of the expected 07.00-17.00 time frame. This has implications for the programming of the HVAC system.

Research goal 2 aimed to build a predictive model for future occupancy in an effort to improve the indoor climate variables identified in research goal 1. The current control systems at NTNU are based on time of day and the day of the week. While this simplifies the system and configuration, it does present some major drawbacks. First of all the designers of the system have to make assumptions as to the usage patterns of the rooms, often without having a good understanding on how the rooms are used. This is especially problematic if the decisions as to the run times of these systems are decided upon designing or building the rooms. Unforeseen usage patterns may occur, the rooms' areas of usage may change and the usage patterns do not remain constant over the whole semester. One way of addressing this would be to log the actual room usage and the relevant indoor climate variables after the room has been used for a while and then make adjustments to the system. This does however require data collection equipment, manual processing of the data as well as trying to make generalized assumptions on the usage patterns. This method is also not able to cope with outlier days or periods with room usage that differs from the norm.

The neural network based system presented in this thesis solves some of those issues. The network can use historical data to make a projection about future room usage and act accordingly. The system also has access to current sensor readings and can update its predictions to better reflect actual room usage. This presents exiting possibilities; the system can dynamically adjust its operating to actual room usage and should be better able to cope with activity outside the expected norm. The system should also see its accuracy improve as more data is recorded and analyzed.

Obtaining a more diverse dataset covering a longer period of time will allow the model to better project future occupancy before it happens. Installing the proposed model as a part of a pilot study comparing the old control system to the new proposed model will make it possible to uncover any improvements both to indoor environmental variables, energy conservation as well as occupants comfort levels. The proposed model represents a promising and potentially useful addition to the systems already in use at NTNU as it may help improve both the indoor environment and the comfort levels of its occupants. Furthermore, it may prove to be a more robust and agile system that is able to adapt to changing occupancy patterns and unforeseen changes.

Appendix A

Acronyms

AI Artificial Intelligence

CNN Convoluted Neural Network

CO2 Carbon dioxide

FNN Feed-forward Neural Network

HVAC Heating, Ventilation and Air Conditioning systems

IoT Internet of Things

IQR Interquartile Range

NN Neural Network

NTNU Norwegian University of Science and Technology

PPB Parts Per Billion

PPM Parts Per Million

VOC Volatile Organic Compound

Q1 Quartile 1 (25 percent)

Q3 Quartile 3 (75 percent)

SD Sentralt driftsanlegg

Appendix B

Neural Networks: a short introduction

This appendix aims to give a short and succinct introduction to neural networks, and some of the terms used in **chapter 7**. The section is strictly intended for readers with no or very little prior knowledge of neural networks. In order to keep this section short, some aspects of neural networks have been simplified and other parts have been omitted because they are not relevant to understanding the neural networks described in the thesis. This appendix will omit activation functions, and only explain the fully-connected topology. Keywords are highlighted in bold face. The appendix will use present a practical example of how neural networks can be used.

Artificial neural networks take inspiration from biological neural networks, especially those found in the brains of mammals. Neural networks take one or more **inputs**, transmit it through 0 more **hidden layers** and finally produces one or more **outputs** in the output layer.

The figure below shows a simple neural network consisting of an **input layer** with three **neurons (nodes)**, a hidden layer with two neurons and a final output layer with a single neuron. Note the yellow arrows connecting the neurons in a directed acyclical graph. The information is passed from one layer to another, and is transformed along the way. Each node in the input layer is fed with a variable from the **input dataset**, and the output from the output layer is compared to the actual observed value, also known as **ground truth**.

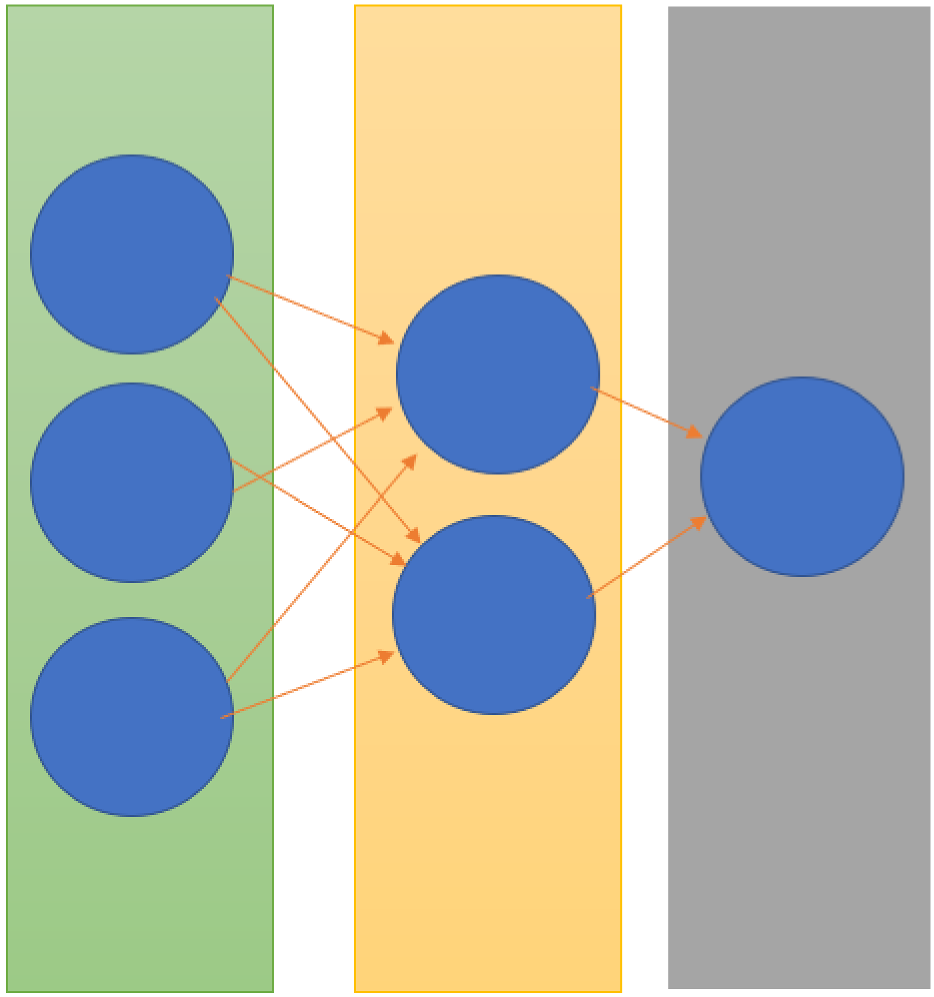

Figure B.1: Neural Network with Input layer (green), Hidden layer (yellow) and Output layer (gray)

To give a very simplified example of how this works in practice; assume that a bank wants to train a neural network to make a decision as to how much money to loan to a person. The neural network based system is to replace the old system based on look-up tables. The bank wants the neural network to take some factors into consideration: (1) income of the borrower, (2) age of the borrower, and (3) the amount of debt the borrower has already accrued. The bank settles on a very simple neural network topology: a single input layer with three neurons, one for each factor and an output layer consisting of a single neuron whose output value is the amount of money the bank is willing to lend to the particular customer. Each of the input neuron I_n is connected to the output neuron through a connection with a weight of W_n. One way to model this network mathematically, in a very simplified manner, is as a linear combination of the inputs and their respective weights: $\sum_{i=1}^{n} \mathrm{W}_i * I_n$. The input of of every input neuron has its respective values multiplied with the **weights** of the pathways connecting them to the output neuron, and the sum of these products is the value of the output neuron.

A practical example; a prospecting borrower has an income of 500,000 NOK (I_1), an age of 32 years (I_2) and existing debts amounting to 320,000 NOK (I_3). Initially all the weights (W_i) are set to 1. Running this through our neural network yields 820,032 NOK [1] as the sum of money the bank is willing to lend. The bank's current method of estimating willingness to loan indicates that the bank is willing to loan the customer 1,300,000 NOK.

All weights being set to one leads to some issues; first of all the amount of current debt is multiplied by a positive value of one, although pre-existing debt should lead to a reduction in the amount of money the bank is willing to lend. The weight of this factor thus needs to be negative to reflect this. The second problem is that the age of the borrower is multiplied by one, the same factor used for the income. This causes a distortion; the maximum value for income could be in millions of NOKs, whereas the age of the prospective borrower probably will never pass 100 years. The income of the borrower will always dwarf the age of the borrower in this equation when the weights are equal.

The neural network obviously needs to be **trained** in order to improve the results. One way of doing this, also known as **supervised learning**, is to give the neural network some inputs and compare the neural network's output to the expected value, also known as **ground truth**. In this case the bank has a pre-existing system for calculating the amount of money its willing to lend to a borrower. The bank decides to use the old system as the ground truth for the new neural network-based system. There are many ways to calculate the difference, also known as **loss**, between the output given by the neural network and the ground truth. The bank decides on the simplest form: the difference between the neural network and the old system's willingness to loan. The output of the neural network is subtracted from the expected output as dictated by the old system.

The difference between the produced and the expected output is used to change the weights of the neural network. After evaluating the first **iteration**, also known as **epoch**, the neural network is re-weighted, and the input is run through the neural network again. This time the weights have been adjusted to $W_1 = 3$, $W_2 = -70$ and $W_3 = -0.8$. This time the neural network's output is 1,241,760 NOK [2], an absolute difference (loss) of 58,240. The output of the neural network is now much closer to the ground truth than in the previous iteration. The network can improve its accuracy by going through even more iterations/epochs, slowly adjusting its weights until the neural network's loss relative to the ground truth diminishes.

This very simplified neural network is able to approximate the old system relatively well, but its performances in a real life situation would be poor. The network has only been trained for a single input, and it would probably give poor predictions if faced with a set of inputs it has not encountered previously. In practice neural networks are fed with large datasets, allowing them to learn based on a wide span of inputs, hopefully yielding a network that makes good general approximations instead of "memorizing" the inputs and expected outputs, also known as **overfitting**. Most neural networks used for real life applications also feature one or more hidden layers, this allow the networks to better model more advanced system. The network previously described in this section only consists of linear combinations, whereas many systems in real life

[1] $500,000 * 1 + 1 * 32 + 320,000 * 1$

[2] $500,000 * 3 - 70 * 32 - 320,000 * 0.8$

are far more complex and may require higher functions in order to get good approximations. A larger number of hidden layers allow the networks to construct more and more complex models of output approximations.

While this section contains a grossly simplified introduction to neural network, it has hopefully left the reader with a better understanding of how neural networks are constructed and how they are trained.

Bibliography

Andersen, I. and Korsgaard, J. (1986). Asthma and the indoor environment: assessment of the health implications of high indoor air humidity. *Environment International*, 12(1-4):121–127.

Arundel, A. V., Sterling, E. M., Biggin, J. H., and Sterling, T. D. (1986). Indirect health effects of relative humidity in indoor environments. *Environmental Health Perspectives*, 65:351.

Baughman, A. and Arens, E. A. (1996). Indoor humidity and human health–part i: Literature review of health effects of humidity-influenced indoor pollutants.

Bayer, C., Hendry, R., Crow, S., and Fischer, J. (2002). The relationship between humidity and indoor air quality in schools. *Indoor Air*, pages 818–812.

Corsi, R., Torres, V., Sanders, M., and Kinney, K. (2002). Carbon dioxide levels and dynamics in elementary schools: results of the tesias study. *Indoor Air*, 2:74–79.

Dahl, G. E., Sainath, T. N., and Hinton, G. E. (2013). Improving deep neural networks for lvcsr using rectified linear units and dropout. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 8609–8613. IEEE.

Di Stefano, J. (2000). Energy efficiency and the environment: the potential for energy efficient lighting to save energy and reduce carbon dioxide emissions at melbourne university, australia. *Energy*, 25(9):823–839.

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I., and Mountain, J. (2010). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS genetics*, 6(6):e1000993.

Gao, Y. and Er, M. J. (2005). Narmax time series model prediction: feedforward and recurrent fuzzy neural network approaches. *Fuzzy sets and systems*, 150(2):331–350.

Guo, H. (2011). Source apportionment of volatile organic compounds in hong kong homes. *Building and environment*, 46(11):2280–2286.

Han, H., Jang, K.-J., Han, C., and Lee, J. (2013). Occupancy estimation based on co2 concentration using dynamic neural network model. *Proc. AIVC*, 13.

Jones, A. P. (1999). Indoor air quality and health. *Atmospheric environment*, 33(28):4535–4564.

Kajala, L. (2007). *Visitor monitoring in nature areas: A manual based on experiences from the Nordic and Baltic countries*. Nordic Council of Ministers.

Kim, J.-W., Choi, K.-S., Choi, B.-D., and Ko, S.-J. (2002). Real-time vision-based people counting system for the security door. In *International Technical Conference on Circuits/Systems Computers and Communications*, volume 2002, pages 1416–1419.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., and Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10):60–68.

Monastersky, R. (2013). Global carbon dioxide levels near worrisome milestone. *Nature*, 497(7447):13.

Morton, D. J. (1952). *Human locomotion and body form: a study of gravity and man*. Williams & Wilkins.

Myhrvold, A., Olsen, E., and Lauridsen, O. (1996). Indoor environment in schools–pupils health and performance in regard to co2 concentrations. *Indoor Air*, 96(4):369–371.

Park, J. (2017). Temperature, test scores, and human capital production. Technical report, Harvard University.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.

Phillips, M., Herrera, J., Krishnan, S., Zain, M., Greenberg, J., and Cataneo, R. N. (1999). Variation in volatile organic compounds in the breath of normal humans. *Journal of Chromatography B: Biomedical Sciences and Applications*, 729(1-2):75–88.

Pitten, F., Bremer, J., and Kramer, A. (2000). Air pollution by volatile organic compounds (voc) and health complaints. *Deutsche Medizinische Wochenschrift (1946)*, 125(18):545–550.

Prill, R. (2000). Why measure carbon dioxide inside buildings. *Published by Washington State University Extension Energy Program WSUEEP07*, 3.

Rawson, H., Begg, J., and Woodward, R. (1977). The effect of atmospheric humidity on photosynthesis, transpiration and water use efficiency of leaves of several plant species. *Planta*, 134(1):5–10.

Satish, U., Mendell, M. J., Shekhar, K., Hotchi, T., Sullivan, D., Streufert, S., and Fisk, W. J. (2012). Is co2 an indoor pollutant? direct effects of low-to-moderate co2 concentrations on human decision-making performance. *Environmental health perspectives*, 120(12):1671.

Sola, J. and Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, 44(3):1464–1468.

Starner, T. (1996). Human-powered wearable computing. *IBM systems Journal*, 35(3/4):618–629.

Štreimikienė, D. and Balezentis, T. (2016). Kaya identity for analysis of the main drivers of ghg emissions and feasibility to implement eu "20–20–20" targets in the baltic states. *Renewable and Sustainable Energy Reviews*, 58:1108–1113.

Telenor (2019). Digitaliseringspilot - smarte bygg. Internal report intended for use by NTNU.

Uchiyama, E., Aronowicz, J. D., Butovich, I. A., and McCulley, J. P. (2007). Increased evaporative rates in laboratory testing conditions simulating airplane cabin relative humidity: an important factor for dry eye syndrome. *Eye & contact lens*, 33(4):174–176.

Velipasalar, S., Tian, Y.-L., and Hampapur, A. (2006). Automatic counting of interacting people by using a single uncalibrated camera. In *2006 IEEE International Conference on Multimedia and Expo*, pages 1265–1268. IEEE.

Wargocki, P., Wyon, D. P., Matysiak, B., and Irgens, S. (2005). The effects of classroom air temperature and outdoor air supply rate on the performance of school work by children. *Proceedings of indoor air*, 1(1):368–72.

Wieslander, G., Norbäck, D., Björnsson, E., Janson, C., and Boman, G. (1996). Asthma and the indoor environment: the significance of emission of formaldehyde and volatile organic compounds from newly painted indoor surfaces. *International archives of occupational and environmental health*, 69(2):115–124.

Wolkoff, P. (1995). Volatile organic compounds. *Indoor Air, Suppl*, 3:1–73.

Wyon, D. (1974). The effects of moderate heat stress on typewriting performance. *Ergonomics*, 17(3):309–317.

Xu, B., Wang, N., Chen, T., and Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.

Yang, J., Nguyen, M. N., San, P. P., Li, X., and Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. In *Ijcai*, volume 15, pages 3995–4001.

Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.

Zhang, J.-S. and Xiao, X.-C. (2000). Predicting chaotic time series using recurrent neural network. *Chinese Physics Letters*, 17(2):88.

Kåre Birger Lapstuen

Using historical data to predict future occupancy for improvement of indoor environmental factors

NTNU
Norwegian University of
Science and Technology