
Preface

This master thesis is written by Frédéric Poupart at the 2 year master of science program Energy and the Environment at the Norwegian University of Science and Technology, conducted in the spring of 2022. The thesis accounts for 30 ECTS credits and builds on the project work [1] "Modelling of air handling unit" which was conducted in the autumn of 2021.

The objective of this master thesis was to develop a mathematical model and a data-driven model capable of accurately predicting the behaviour of an air handling unit connected to a laboratory room, with various tests performed to evaluate the model accuracy. The data-driven model was an LSTM neural network model developed by Gaurav Chaudhary, PhD candidate.

I want to give my thanks to my supervisor Bjørn Austbø, associate Professor at NTNU's energy and process engineering department, and Gaurav Chaudhary, PhD candidate, for their expertise, consistent guidance and help throughout my time working with the master thesis. Thank you for being engaged in the assignment and being available throughout this process.

Frédéric Poupart Trondheim, 18th June 2022

Abstract

With the growth of climate awareness and the role energy usage plays in this respect, increasing the energy efficiency of buildings has steadily become a more important field of study. This can be done through the implementation of data-driven and mathematical models capable of optimizing heating, ventilation and air-conditioning systems.

This thesis attempts to develop a mathematical model and a data-driven model that can accurately predict the behaviour of an air handling unit connected to a laboratory room. These models can be used to predict the future indoor air temperature at different operating conditions for the air handling unit and thereby be useful for optimizing its operation. Tests performed with these models can also provide insight into the relevance of different input parameters.

In addition, this thesis provides a description of the components and operation of air handling units, followed by a literature review on the different modelling approaches to HVAC system behavior prediction. Some of the modelling types that will be discussed include mathematical models, hybrid models and data-driven models. HVAC system modelling applications in previous works are reviewed as well.

The literature review showed that there are multiple modelling types that are appropriate in describing HVAC system behavior accurately with different strengths and weaknesses spread between them. HVAC system modelling is also found to have several applications that can contribute to increasing the energy efficiency of buildings.

Development of the mathematical model was done in EnergyPlus. Unknown parameters in internal mass, air infiltration and internal gains from occupants and equipment were determined and calibrated against four scenarios where the air handling unit was operating under different conditions over a 79-day period. For the data-driven model, a long short-term memory recurrent neural network was used to predict the indoor air temperature. Two datasets were created for training, validation and testing, one generated by the mathematical model over a 1-year simulation and the other by measurements taken inside the laboratory room from 04.03 to 22.05.

After calibrating the mathematical model and identifying the missing parameters in the model, testing showed that the model can accurately predict the indoor air temperature, provided that the setpoint temperature and occupancy level is constant. For the mathematical model to properly model the thermal inertia of the room, it was concluded that further tests need to be performed.

Testing the data-driven model revealed that the model is able to predict the indoor air temperature to a sufficiently accurate degree when trained with the measured data, despite the limited time period the training dataset covered. It was also found that the model can also predict the measured indoor air temperature relatively accurate when trained with the simulated data. In terms of input parameter relevance, time index related parameters and the supply setpoint air temperature were the most impactful for the measured- and simulated datasets, respectively.

Sammendrag

Energieffektivisering av bygninger har blitt et stadig viktigere fagområde ettersom at klimabevissthet og energiforbruk har fått mer oppmerksomhet gjennom tidene. En måte å øke energieffektiviteten til bygninger er ved å benytte datadrevne og matematiske modeller som kan optimalisere varme-, ventilasjon- og sanitærsystemer.

I denne masteroppgaven gjøres det et forsøk i å utvikle en matematisk modell og en datadreven modell som kan forutsi virkemåten av et VVS-system koblet til et testrom i en laboratorie. Disse modellene kan brukes til å forutsi det termiske innklimaet i testrommet i forskjellige scenarier og kan dermed benyttes for å optimalisere drift av VVS-systemet. Forsøk gjennomført med modellene kan også gi et innblikk i hvordan ulike parametere har en innflytelse på det termiske innklimaet.

I tillegg gir denne oppgaven en beskrivelse av komponentene som inngår i ventilasjonsaggregater og deres virkemåte, etterfulgt av en litteraturstudie om forskjellige modelltyper av VVS systemer. Rapporten diskuterer bl.a. matematiske modeller, hybrid modeller, og datadrevne modeller. Anvendelser av VVS-systemmodeller er beskrevet i litteraturstudie også.

Fra litteraturstudien ble det funnet en rekke modelltyper som kan benyttes til VVS-systemer. Disse kan beskrive og forutsi virkemåten til VVS-systemer med god nøyaktighet og hver av de har sine egne styrker og svakheter. Det ble også funnet at VVS-systemmodeller kan bidra til energieffektiviseringen av byggesektoren.

Den matematiske modellen ble utviklet gjennom programmet EnergyPlus. Ukjente parametere i form av intern termisk masse, luftinfiltrasjon, og varme fra beboere og elektronisk utstyr ble funnet og kalibrert opp mot fire scenarier over en 79 dager lang periode. For den datadrevne modellen ble det benyttet en long short-term memory tilbakevendende nevralt nettverk for å forutsi innendørs lufttemperatur. To datasett ble benyttet for trening og testing av den datadrevne modellen. Ett av dem ble generert via en 365 dager lang simulering gjennomført av den matematiske modellen. Det andre datasettet ble hentet fra målinger tatt i perioden 04.03 til 22.05.

Tester gjennomført med den matematiske modellen peker mot at modellen kan beregne innnetemperaturen på nøyaktig vis dersom setpunktstemperaturen og mengden beboere holder seg konstant. Det argumenteres derimot at flere forsøk i testrommet må gjennomføres for at den matematiske modellen kan modellere testrommets termisk treghet.

Forsøk gjennomført med den datadrevne modellen viser at modellen kan forutsi målt innnetemperatur med tilstrekkelig nøyaktighet når den trenes opp med datasettet hentet fra målinger. Det ble også funnet at modellen kan forutsi målt innnetemperatur nøyaktig nok når den trenes opp med simulerte data. Av de 12 inngangsparameterene som ble benyttet var det tidsindeks-relaterte parametere og setpunktstemperaturen som hadde størst betydning for nøyaktigheten av modellen.

Table of Contents

List of Figures	vi
List of Tables	vii
1 Introduction	1
1.1 Background	1
1.2 Research question and objectives	1
1.3 Scope and limitations	1
1.4 Content structure	2
2 Literature Review	3
2.1 Indoor climate and the role of HVAC systems	3
2.2 Air handling units	4
2.3 Modeling types	6
2.3.1 Mathematical models	6
2.3.2 Data-driven models	9
2.3.3 Grey-box models	13
2.3.4 Validation criteria	13
2.3.5 Modelling applications	14
3 Methodology	15
3.1 System description	15
3.2 Mathematical model development	16
3.2.1 Simulation platform	16
3.2.2 Room construction	17
3.2.3 HVAC system	18
3.2.4 Experimental validation	20
3.2.5 Input data implementation	22
3.2.6 Model calibration	24
3.2.7 Year-long simulation	25
3.3 Black-box model development	28
3.3.1 Model architecture	28
3.3.2 Training and evaluation of models	28

3.3.3	Input parameters	29
3.3.4	Model testing	29
4	Results	31
4.1	Mathematical model	31
4.1.1	Model calibration	31
4.1.2	Year-long simulation	36
4.2	Black-box models	37
4.2.1	Model training	37
4.2.2	Feature relevance	40
4.2.3	Measured data prediction	42
5	Discussion	43
5.1	Mathematical model	43
5.1.1	Calibration results	43
5.1.2	Year-long simulation	44
5.2	Black-box models	44
5.2.1	Model accuracy	44
5.2.2	Dataset distribution	44
5.2.3	Timestep length	45
5.2.4	Testing periods	45
5.2.5	Feature relevance	45
5.2.6	Measured data prediction	46
5.3	Possible model improvements	46
6	Conclusion	47
	Bibliography	48
	Appendix	51
A	Cold water loops	51
B	Air handling unit component datasheet	52
C	Laboratory plan drawing	59
D	Room section drawing	60
E	Calibration code	61

List of Figures

1	<i>Example of an AHU and its components.</i>	4
2	<i>Schematic and psychrometric (dry-bulb temperature on x-axis, humidity ratio on y-axis) chart of dehumidification of air through a cooling- and heating coil. Adapted from [12].</i>	5
3	<i>Classification of data-driven modelling of HVAC systems, adapted from [14].</i>	9
4	<i>Schematic of an artificial neural network, adapted from [31]</i>	11
5	<i>Schematic of a support vector machine, adapted from [20].</i>	12
6	<i>Configuration of components in the air handling unit for the laboratory room.</i>	15
7	<i>Monitoring system of the air handling unit. Numbers in green are configurable setpoint values.</i>	16
8	<i>Sketch of the laboratory room geometry used in the model (not to scale). North direction is labelled at the bottom.</i>	17
9	<i>Node diagram of the air loop.</i>	19
10	<i>Node diagram of the hot water loop connected to heating coil. Each node is labelled with a name in the model.</i>	19
11	<i>Indoor air temperature for 4 different scenarios.</i>	21
12	<i>Outdoor dry bulb-, wet bulb temperature and relative humidity for year-long simulation.</i> 26	
13	<i>Supply setpoint air temperature for year-long simulation.</i>	27
14	<i>Indoor air temperature with constant setpoint temperature, 21.03-25.03.</i>	32
15	<i>Indoor air temperature with varying occupancy levels, 28.03-01.04.</i>	33
16	<i>Indoor air temperature at free-floating operation, 08.04-19.04.</i>	34
17	<i>Indoor air temperature with variable setpoint temperatures.</i>	35
18	<i>Simulated indoor air temperature over an entire year.</i>	36
19	<i>Error of black-box models EP-trained and Measured-trained.</i>	37
20	<i>Average RMSE for both black-box models with varying dataset distributions.</i>	38
21	<i>Average RMSE for both black-box models with varying timestep lengths.</i>	38
22	<i>Average RMSE for both black-box models with different testing periods.</i>	39
23	<i>Average RMSE for both black-box models trained with 1 missing feature.</i>	40
24	<i>Average RMSE for both black-box models trained with missing categories of data.</i> . .	40
25	<i>Simulated indoor- and supply setpoint air temperature over an entire year.</i>	41
26	<i>Measured indoor- and supply setpoint air temperature 04.03-22.05.</i>	41

27	<i>Error of prediction of measured indoor air temperature with EnergyPlus model and EP-trained black-box model.</i>	42
28	<i>Measured and predicted indoor air temperature with EnergyPlus model 04.03-22.05.</i>	42
29	<i>Node diagram of fluid loop providing tap water to the dehumidifier.</i>	51
30	<i>Node diagram of fluid loop providing chilled water to the second cooling coil.</i>	51
31	<i>Plan drawing of NTNU's Energy and Indoor Environment Laboratory, second floor.</i>	59
32	<i>Section drawing of the modelled laboratory room inside NTNU's Energy and Indoor Environment Laboratory with construction components.</i>	60

List of Tables

1	<i>Recommended operative temperature ranges for different physical activity levels according to the Norwegian building regulations TEK17 [8].</i>	3
2	<i>Advantages and disadvantages of mathematical HVAC modelling.</i>	6
3	<i>Advantages and disadvantages of data-driven HVAC modelling.</i>	9
4	<i>Advantages and disadvantages of grey-box HVAC modelling.</i>	13
5	<i>Recent modelling applications in connection with HVAC systems.</i>	14
6	<i>Properties of materials used in the model; roughness, thermal conductivity, density and specific heat.</i>	18
7	<i>Surface U-values of room model.</i>	18
8	<i>Design conditions of the coils in the air handler.</i>	20
9	<i>Final surface U-values (no film) after removal of fenestration surfaces.</i>	23
10	<i>Calibration parameter value ranges</i>	25
11	<i>Distribution of AHU availability modes in year-long simulation.</i>	26
12	<i>Distribution of air flow rate in year-long simulation.</i>	27
13	<i>Input parameters for black-box models</i>	29
14	<i>Dataset distributions</i>	30
15	<i>Parameter values for most accurately calibrated models ("model 1-3") and the original model ("previous model").</i>	31
16	<i>Average RMSE, MAE and MAPE of the most accurately calibrated model and the original model relative to measured indoor air temperature over 4 different scenarios.</i>	31
17	<i>RMSE, MAE and MAPE of the calibrated and old model when simulating a constant setpoint.</i>	32
18	<i>RMSE, MAE and MAPE of the calibrated and old model with varying occupancy levels.</i>	33

19	<i>RMSE, MAE and MAPE of the calibrated and old model when simulating free-floating operation.</i>	34
20	<i>RMSE, MAE and MAPE of the calibrated and old model with varying setpoint temperatures.</i>	35

1 Introduction

1.1 Background

Today, buildings account for 40% of the total energy usage in the world along with 36% of greenhouse emissions [2]. With the growth of climate awareness and the role energy usage plays in this respect, increasing the energy efficiency of buildings has steadily become a more important field of study. If we are to meet the United Nation's sustainable development goal for energy consumption, it is necessary to reduce the energy usage in the building sector. Over the past decades, numerous measures have been made in order to increase the sustainability of buildings. Building regulations such as TEK17 are steadily demanding greater requirements to energy efficiency. Environmental certification tools like BREEAM have been developed to motivate increased focus on the environment and sustainability.

Building heating, ventilation and air-conditioning system modelling is also relevant here, as better prediction of building energy usage allows for improved controlling of energy systems while maintaining an appropriate indoor climate. Newer building regulations have also led to a greater portion of the energy usage in buildings moving to the HVAC systems. In this respect, it is key to have an understanding of the processes that occur in these systems and how they can be modelled and eventually optimized.

1.2 Research question and objectives

The main objective of this thesis is to develop a mathematical model and a data-driven model capable of accurately predicting the behaviour of an air handling unit connected to a laboratory room with simulation data used as the training data for the data-driven model. The sub-objectives are listed as follows:

- Perform a literature review on modelling approaches to predicting the behaviour of air handling units and ventilation systems.
- Description of an air handling unit and ventilation system.
- Development, calibration and evaluation of mathematical model of air handling unit against multiple operation conditions. This model is then to be used to generate a large training dataset.
- Evaluation and testing of a black-box model with different training datasets.

1.3 Scope and limitations

Experiments conducted in this thesis project are limited to the spring of 2022. Length of experiments have also been limited by the availability of the case study space due to other activities occurring there.

As the case study space is located in the nordic climate of Trondheim, Norway, the paper focuses more on a nordic perspective in the literature review and utilizes common construction material in Norway during the model development process. Furthermore, the literature review of HVAC system modelling is mainly limited to mathematical- and data-driven modelling approaches.

1.4 Content structure

The thesis is split into six chapters:

Chapter 1 - Introduction

This chapter provides some background knowledge to the thesis topic as well as the motivation behind the thesis and its scope and limitations.

Chapter 2 - Literature review

This chapter provides a description of the components and operation of air handling units, followed by a literature review on the different modelling approaches to HVAC system behavior prediction and their applications. Specific focus is given to the data-driven and mathematical modelling approaches.

Chapter 3 - Methodology

This chapter gives a description of the cases-study HVAC system, followed by a detailed look into the development of a mathematical- and a data-driven model for the HVAC system.

Chapter 4 - Results

In this chapter, simulation results from the mathematical- and data-driven models are presented along with the measurement results. This includes different operating conditions for the HVAC system and different input parameters for the data-driven model.

Chapter 5 - Discussion

This chapter contains discussions on the content of chapter 4 in terms of what insight the results give and what may have affected the results.

Chapter 6 - Conclusion

In this chapter, conclusions from the discussion and observations from the results is presented in a holistic manner, referring back to the objective of the thesis.

2 Literature Review

2.1 Indoor climate and the role of HVAC systems

A key factor that is considered in building design and construction is the resulting climate in the interior spaces. 90% of people’s time is spent indoors in developed countries [3]. Poor climate conditions indoors have historically shown to have an adverse effect on the occupant’s health, resulting in decreased concentration levels and productivity, stress, headaches, allergic reactions and various other symptoms [4][5][6]. It is therefore of upmost importance to achieve a good indoor environment. This is reflected in current building regulations, where several requirements have been set to the usage of building materials, ventilation, lighting and more.

Indoor climate is generally described in terms of air quality, humidity levels, thermal comfort, acoustics, and lighting quality.

The quality of the air is often determined by measuring the concentration levels of harmful gases such as carbon dioxide in the air and finding how polluted the air is. According to the Norwegian Institute of Public Health’s indoor climate rapport, indoor air is required to have a CO₂-concentration level less than 1000 ppm to provide good indoor climate [7]. Achieving this usually entails having ways of ventilating indoor spaces.

The Norwegian building regulations TEK17 specifies demands for ventilation flow rates for different building types in Norway. Residential buildings for instance need to supply indoor spaces with at least $1.2m^3/h$ of fresh air per m^2 floor area as well as $26m^3/h$ for each bed space in bedrooms while the buildings are in use. For public- and working spaces, this increases to $2.5m^3/h$ per floor area in addition to a minimum air supply rate of $26m^3/h$ per person. Not only does air pollution play a role in determining appropriate ventilation rates, but also humidity levels - measured by relative humidity. At high humidity levels, mold and fungus can start to grow and have serious effects on the occupants’ health and the building construction. As a result of this, TEK17 demands increased ventilation in rooms where mold growth is more likely to occur (i.e. bathrooms and kitchens).[8]

When occupants experience ”thermal comfort”, they are in a satisfactory thermal environment. Whether or not an indoor space achieves thermal comfort is difficult to quantify, as it is a subjective experience and there are multiple relevant factors that affect this experience. Some of the most important factors to take into account are personal factors such as the occupant’s activity levels and use of clothing, and environmental variables like humidity and air temperature (usually given by the dry-bulb temperature). According to TEK17, the dry-bulb temperature should remain under $22^\circ C$ when there is need for heating [8]. The building regulations also provide different acceptable operative temperature ranges, which is a measure that combines the air temperature and thermal radiation to describe thermal comfort. These temperature ranges are tabulated in table 1.

Table 1: *Recommended operative temperature ranges for different physical activity levels according to the Norwegian building regulations TEK17 [8].*

Activity level	Low	Medium	High
Temperature $^\circ C$	19 – 26	16 – 26	10 – 26

TEK17 also provides technical requirements for radiation, sound, lighting and outside viewing.

When it comes to improving the air quality and thermal environment, heating, ventilation and air-conditioning (HVAC) systems play a significant part in this and are the main elements that control the indoor climate.

2.2 Air handling units

Most HVAC systems nowadays are equipped with an air handling unit (also known as an air handler and often abbreviated as the AHU). Air handlers are used to circulate air in a building and condition the supply air before it enters the ventilation ducting system and later returns to the AHU. A typical configuration of an air handler and its components are depicted in figure 1. The working principle behind an air handler is that outdoor air flows through various components that filter away particles and changes the relative humidity and temperature of the air before it is distributed to indoor zones. In the zones, the supply air mixes with the existing air and is later circulated into the air handling unit and leaves as exhaust air.

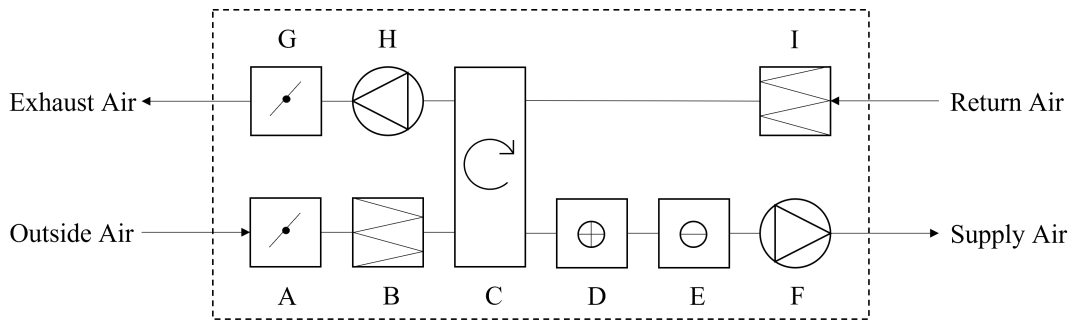


Figure 1: *Example of an AHU and its components.*

Air handling units are generally composed of the following components:

- Dampers (A & G). These are valves (or plates) that control the inlet and outlet openings of the air handler. Dampers can either have on/off functionality or variable opening levels.
- Filters (B & I). These filter out dust and other particles in order to improve the air quality. Filter types include fiberglass filters and High-Efficiency Particulate Air filters (HEPA). There is a relatively minor pressure drop across air filters, but this resistance has shown to increase as air filters get dirtier and thereby lead to higher energy consumption by fans [9][10]. Filters should therefore be cleaned on a regular basis.
- Heat recovery units (C). Heat recovery units utilize the thermal energy of the returning air to provide heating to the supply air. They can also theoretically provide cooling as well during cooling seasons. Common heat exchangers used for heat recovery include rotating heat exchangers (also known as thermal wheels) and fixed plate heat exchangers - both of which can recover heat at a sensible efficiency of over 80%. A disadvantage of using rotating heat exchangers is that they can also in certain conditions recover pollutants from the return air. [11]
- Heating- (D) and cooling (E) coils. These are placed after the heat recovery heat exchangers and heat up/cool down the supply air to the desired supply air temperature. Coils are either hot-/chilled water coils or direct expansion (DX) coils.
- Fans (F & H). These circulate air to- and from the indoor space through the supply- and return side of the air handler. Fans can be placed on both the supply and return side, in which case the fans need to be configured to work well together in terms of flow rate and pressure drop. For a constant air volume (CAV) system, the fans either run at a constant frequency or are shut off. For a variable air volume (VAV) system, the fan working frequency can change.
- Humidifiers. These increase the humidity of the supply air by spraying vaporized water into the duct. It is more common to use humidifiers in colder areas, as the heating process

can lead to dry air. However, one usually avoids utilizing humidifiers if it is not absolutely necessary, as humidifiers can contribute to the growth of microbial activity in the ducts. [11]

- Mixing boxes. In mixing boxes, the return air is mixed together with the outdoor air entering the air handler and is circulated back into the indoor space. Air mixing is not always necessary in air handlers and some air handlers do not have mixing boxes.

Dehumidification is also possible to achieve in air handling units. The most common method of dehumidifying the supply air is to use a chilled water cooling coil in combination with a heating coil (in this case called a reheat coil). The cooling coil is placed before the reheat coil, and cools down the air to the desired humidity level before it is heat up to an appropriate temperature by the reheat coil [12]. This process is illustrated in figure 2 with a psychrometric chart.

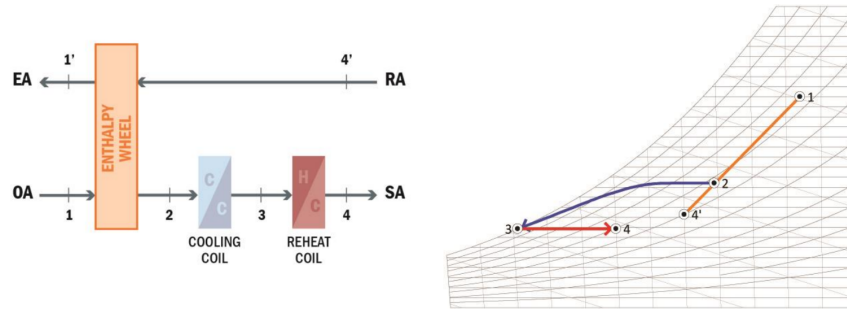


Figure 2: Schematic and psychrometric (dry-bulb temperature on x -axis, humidity ratio on y -axis) chart of dehumidification of air through a cooling- and heating coil. Adapted from [12].

Heating and cooling source

In the cooling/heating coils, the cold/hot water circulates through a pump and is cooled down/heated up by an energy source. There can also be a tanks that store cold/hot water. These energy sources can be in the form of electric chillers/boilers, central cooling/heating systems and district cooling/heating.[13] Central heating/cooling systems can contain heat pumps that transport cooling/heating energy from other outside or other processes in the system. Chillers are usually paired with a cooling tower that rejects the chiller condenser's heat [14].

HVAC system control

HVAC systems can have a control system for optimizing the operation of the various components in the air handler. Proper control of HVAC system operation is very important, since the conditions are constantly changing. These systems can be driven by changes manually made by users - for example with devices like wall thermostats and VAV boxes in rooms [15] - and/or they can be designed to regulate the HVAC systems automatically. Various controller strategies such as proportional-integral-derivative (PID) controllers are widely used for automatic regulation of HVAC systems [16]. For these strategies, a desired variable value (for example, air temperature) is continuously compared with the actual value measured by a sensor. From these values an error is computed, which passes through the controller. PID controllers deliver an output based on the integrated error, derivative of the error, and the error value itself, which is then used to change the operation of the components in the HVAC system. In a CAV system, variables such as air temperature and relative humidity are regulated by changing openings of valves connected to coils and heat exchangers. In a VAV system, the air flow rate can also be changed.

Strategies for HVAC system control has for a long time been a field of interest in the building automation industry. It is also one of the main applications of HVAC system modelling, as one of the most effective ways of reducing HVAC energy usage is through good control, which can be achieved with model-based control strategies [13][17].

2.3 Modeling types

This section will go through the different approaches of HVAC system modelling, with particular focus on mathematical- and data-driven models.

2.3.1 Mathematical models

Mathematical models, also known as white-box-, forward-, analytical- or physical models describe systems through knowledge of the physical principles that govern the processes that occur. These tend to be time-consuming to develop and require great expertise in numerous factors and their effect on the system behaviour. However, mathematical models share some unique advantages that make them a highly popular modelling approach for HVAC systems. Table 2 lists advantages and disadvantages of using the mathematical approach for HVAC modelling.

Table 2: *Advantages and disadvantages of mathematical HVAC modelling.*

Advantages	Disadvantages
<ul style="list-style-type: none">• High prediction accuracy [18]• Well suited for HVAC system design - can be used for HVAC systems that have not been built yet; empirical data is not required for model development [18][19]• Able to described heat transfer mechanisms and gain an understanding of the processes involved in air conditioning [20][21]• Model transparency leads to easier error detection [21]	<ul style="list-style-type: none">• Expertise is required for model development [20]• Time-consuming model development [20]• Complex mathematics lead to computationally expensive models [20][21]• Too many assumptions can lead to poor prediction accuracy [19]

Mathematical modelling approaches of HVAC systems is a well researched field of study with a history spanning several decades. Harish [18] suggests a 6-step procedure for mathematical modelling of HVAC systems and their energy usage based on previous models developed:

1. Acquire climate data for the location of the building to be studied
2. Acquire building design data, geographical location and construction data
3. Find type of HVAC system and characteristics of its subsystems
4. Acquire building operation schedules
5. Perform simulations over desired simulation periods
6. Predict energy consumption patterns [18]

Mathematical HVAC system models are usually split into two models that are later integrated: the zone (also known as building-) model and air handling unit model.

Zone modelling

According to Homod [21], the building modelling process is the most demanding. Here, the heat transfer through the building construction surfaces (roof, walls, windows, floor, etc.) need to be modelled. Moreover, internal heat gains from occupants, activity and lighting have to be considered, contributing to the model's complexity [21]. One method of modelling the indoor zone is to describe the heat transfer with a heat balance, in which conservation of energy (i.e. the first law of thermodynamics) is used to model the energy flow. This approach is used for each surface where heat is transferred as well as the air inside the zone. This leads to multiple equations that need to be solved simultaneously to find the air temperature.[14][22]

Another commonly used method is the thermal-network modelling approach. With this method, the building is represented by a network of nodes that intersect with each other with energy flowing through them. At each node (room, wall, floor) the building volume is assumed to be homogenous with uniform state variables - meaning each node is described by a unique state (temperature, humidity, etc.). For each node, heat balance is applied and the resulting equations are solved. This is a method utilized in many simulation programs, including EnergyPlus, IDA-ICE and TrnSys, and can be viewed as a continuation of the heat balance method described above [14]. Some key advantages of using this method are small computation times compared to other mathematical models and that it is well suited for energy consumption prediction. At the same time, the simplifications made makes it harder to investigate larger rooms and air quality amongst other things.[19][23]

Using computational fluid dynamics (CFD) is also a viable option in modelling building heat transfer. Here, the building zone is divided into a large number of volumes that form a mesh. Using this approach leads to a much greater computation time, but it also yields detailed information of the air- and pollutant flow inside buildings.[23] Other relevant means of modelling heat transfer in buildings are the heat conduction equation model and transfer function model [14][19].

Air handling unit modelling

AHU mathematical modelling can be categorized into two groups: complete- and component-based system models. Through complete system modelling, preconfigured HVAC systems are used. These can be modelled through various methods, such as a bilinear model. This approach of modelling air handlers can be found in software such as DOE-2 and BLAST, and is relatively inexpensive in terms of computer hardware requirements. However, the number of system configurations and control strategies available are limited, resulting in lower flexibility for the user with this approach. Moreover, the nonlinearity of HVAC systems makes this method not as accurate.[24][25]

In component-based system models, the components in HVAC systems are modelled individually and are then connected to each other to form a network of component models working with each other. Compared to complete models, this is a much more flexible solution. Components and their respective modelling approaches are listed below.

- Duct and pipe. As the air flows through the piping system, the air loses/gains energy from the pipe's surroundings. The heat transfer can be modelled with an energy balance approach based on the air temperature inside the duct and the surrounding environment, or on the inlet and outlet air temperature of the duct. [14]
- Heating- and cooling coil. Here, the energy balance approach on the water and air side of the coils can be used to model the various temperature, while a mass balance can be used to find the humidity of the air. [14][19]

-
- Mixing box. The temperature of the mixed air exiting the mixing box can be easily calculated based on the mass flow rate and air temperature (also the heat capacity) of the outside air and return air.
 - Fan. The temperature of the air after passing through a fan can be calculated based on the motor inefficiency. [26]
 - Humidifier. The humidification process can be modelled with energy and mass balance equations. [19]
 - Cooling tower. Heat rejection from towers and evaporation of water can be modelled physically through the Effectiveness-NTU method. [14][19]
 - Storage tank. The outlet temperature can be modelled based on heat transfer from the heat pump, with the surroundings and the water coming through the inlet. [14]
 - Boiler. The water temperature can be calculated based on the heating energy supplied to the boiler, the supply- and return water temperature. [14]
 - Chiller and heatpump. Chillers and heatpumps can be modelled as a collection of evaporators, condensers, compressors and expansion valves, where the evaporators and condensers can be treated as heat exchangers. [14]

Building and AHU model integration

After constructing the building- and AHU models, these can be integrated to fully describe the building's performance. This can be done sequentially or fully integrated. In the latter case, the models communicate and provide feedback to each other. Simulation programs that utilize a fully integrated solution scheme include EnergyPlus, IDA-ICE and TRNSYS. [25]

2.3.2 Data-driven models

Data-driven models, also known as, inverse- or black-box models, take a significantly different approach to modelling systems. Instead of looking at principles based on physics, data-driven models rely solely on empirical data. These receive data sets containing input data and output data from various experiments and compute a statistical relationship between the input- and output variables. There are a wide range of black-box model types such as statistical models and data-mining algorithms (also known as machine learning models). Figure 3 provides an overview of the different data-driven modelling approaches used in HVAC systems. In this paper, the literature review will go through the statistical- and machine learning methods.

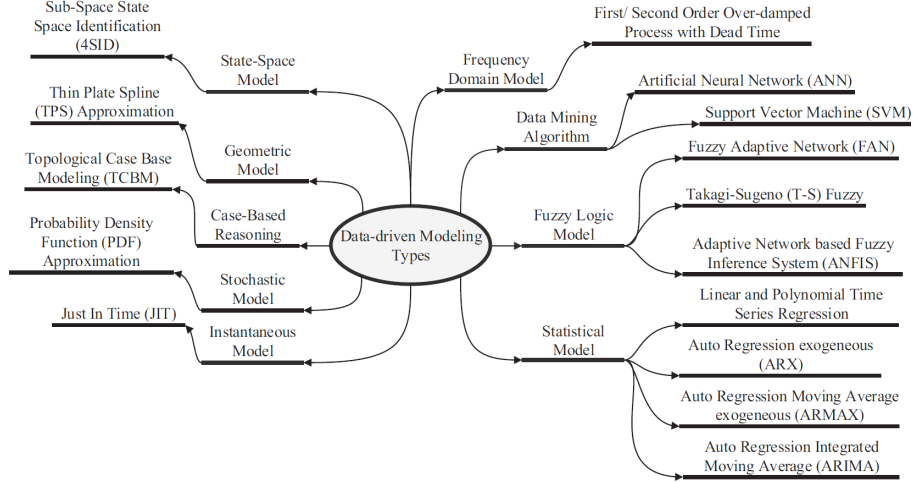


Figure 3: *Classification of data-driven modelling of HVAC systems, adapted from [14].*

By relying entirely on empirical data, data-driven models are able to overcome some of the main limitations to white-box models. This includes simpler model development procedures and less computationally expensive simulations. On the other hand, these models are completely lacking in transparency (hence "black-box"). Advantages and disadvantages of using data-driven modelling for HVAC systems are shown in table 3.

Table 3: *Advantages and disadvantages of data-driven HVAC modelling.*

Advantages	Disadvantages
<ul style="list-style-type: none"> • Can have high accuracy with enough training data [23][18] • Quick and less demanding calculations [20] • Easy model development [20][23] • Well suited when physical characteristics of the HVAC system are unknown [23] 	<ul style="list-style-type: none"> • Lack in model transparency makes it hard to interpret the model and detect errors • Not well suited for modelling HVAC systems that haven't been built yet • Requires significant amount of data for model training [18]

Data-driven modelling procedure

When developing a data-driven model, sufficient data needs to be gathered beforehand. This can be achieved through simulation programs, measurements or gathering data publicly available, with incorrect data either corrected or removed.[20]

After collecting the data, there may be a large amount of input variables. In these cases, data-driven simulations can be computationally expensive as there is a lot of data to go through, which eliminates one of the main advantages to using data-driven models. To alleviate this issue, feature selection plays a key role. If done properly, the computation time can be reduced significantly while maintaining a satisfactory prediction accuracy. Some common feature selection methods are variable ranking, filter and wrapper methods, and principal component analysis (PCA) [20].

After extracting the input variables to be considered, the model should be fed 3 datasets in order. The first dataset is used for training the model, the second for validating and making minor adjustments, and the third for testing the model in an unbiased evaluation. This is an iterative process in which the goal is to get a sufficiently decreasing error from the training set and onwards.[27] These datasets should all have extensive amounts of data, but do not necessarily need to be evenly distributed. In fact, most of the data tends to be used in the training stage [28], with at least 50% being used for training. Commonly used data-set distributions include 70-30%, 80-20% and 60-40% training-validation/testing [29].

Input parameters

A review of the most commonly used input parameters from 2015 to 2019 was performed in [20]. From this review, it was found that there are three types of input parameters that are most often used in modelling building systems: meteorological data, historical data and time-index.

Meteorological data consists of datapoints related to the outside climate, as in the outside dry bulb-, wet bulb-, dew point temperature, relative humidity, solar radiation and more. This group of input parameters was found to be used in over 70% of studies from 2015 to 2019. Historical data may include parameters such as the historical heating demand of the building. Historical data points such as the heating demand and electricity consumption in the previous hour have been found to be of high importance for predicting the HVAC energy consumption, making these parameters more attractive in recent years. When referring to time-index, it means the stano series for time. This can be represented by the time of day, day of the week, day type, etc. These parameters can often be used as a substitute to the occupancy when such data is not available and have been shown to be highly relevant in predicting future energy usage. [20]

Other input parameters such as the building characteristics and indoor climate are also occasionally utilized, but tend to remain relatively constant for most buildings and so is often discarded from the input dataset. The occupancy level has been shown to be of great importance, but it is unfortunately very difficult to gather reliable data on this data point [20].

Statistical regression

Statistical (linear) regression is an older statistical method that investigates the relationship between the input- and output features through a probabilistic perspective. It generally takes the form of a linear equation or a polynomial equation, shown in equations 1 and 2, respectively.[30]

$$Y_i = \alpha_i + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_m x_{i,m} + \epsilon_i \quad (1)$$

$$Y_i = \alpha_i + \beta_1 x_{i,1} + \beta_2 x_{i,2}^2 + \dots + \beta_m x_{i,m}^m + \epsilon_i \quad (2)$$

Where Y_i is the output prediction, ϵ_i is a random error that is assumed to be normally distributed, and the other coefficients are the weights for features x_i where $i = 1, \dots, m$. Linear regression can also take the form of a natural logarithmic function [20]. Estimation of the coefficients is done by using a finite number of datapoints [30].

Relative to other data-driven methods, linear regression is easy to use and understand, as the coefficients can be examined to find out how the different features contribute to the output calculation. However, it should be noted that feature selection is often necessary to avoid reduced prediction accuracy caused by high intercorrelations among independent variables. In addition, nonlinear input-output relationships cannot be determined through the use the general form described by equation 1. This issue can be avoided by using a polynomial or natural logarithm function instead. [20]

Machine learning methods

Artificial neural networks Artificial neural networks (ANNs) are machine learning models that take inspiration from- and attempt to mimic the functionality of the human brain. They are comprised of a network of basic processing units (modelling a neuron). These are fitted in layers and are connected to each other. Each processing unit receives signals from all the units in the layer "behind" it and plugs those signals into an activation function along with a bias set and weight specifically made for that process unit. The activation function then returns an output signal that is sent subsequently to the units in the next layer.[30]

Typically, an ANN contains three layers: an input layer, a hidden layer and an output layer. Processing units placed in the input layer receive the input values, while the output layer delivers the predicted output value(s) as illustrated in figure 4. ANNs can have multiple hidden layers, but the number of these should be carefully selected as to avoid over-fitting. Like other data-driven models, ANNs must be trained to estimate the needed weights and biases of each processing unit in order to provide accurate predictions.

In newer studies, recurrent neural networks (RNNs) are used more. These are a type of neural network that have memory and are better at processing sequences of inputs making them especially attractive when dealing with time series datasets.[30]

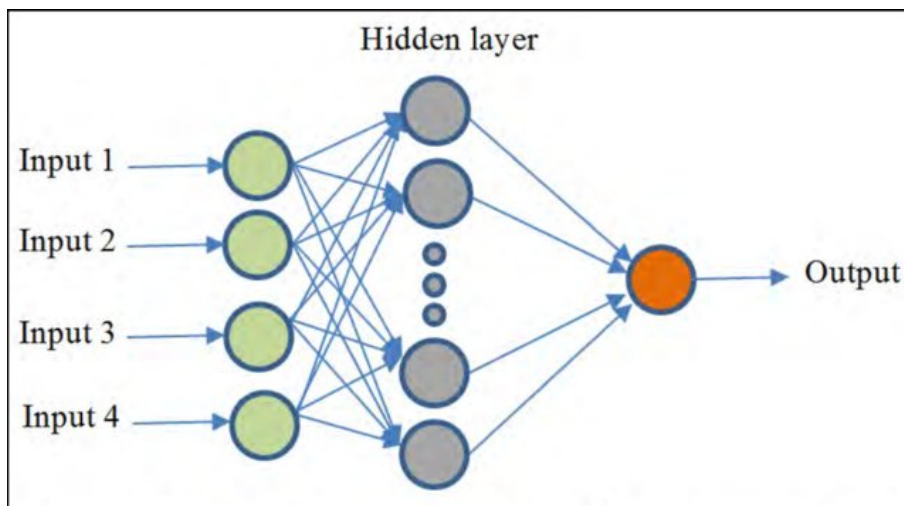


Figure 4: *Schematic of an artificial neural network, adapted from [31]*

Usage of ANNs is the currently the most common method in building energy consumption prediction [32][19]. ANNs have shown a strong level of robustness in tackling nonlinear problems without the need of extensive expertise of HVAC system operation [20]. They have also shown to possess

other advantages such as being able to continue running even when an element fails. On the other hand, ANNs can have long computation times if the network is large enough, and a new one has to be designed and trained each time a new modelling application is brought forth.[32][30]

Support vector machines

Support vector machines (SVMs) are another common machine learning method that is seeing continuous growth in usage in research and industry.[32] SVMs learn through historical data to construct a function that estimates output values based on input values. A threshold value is predefined during the model development, for which the error of the output estimation cannot exceed this value.[30]

Support vector machines are often used for pattern recognition and classifying data where they essentially draw a straight line that separates different groups of data, as shown in figure 5.

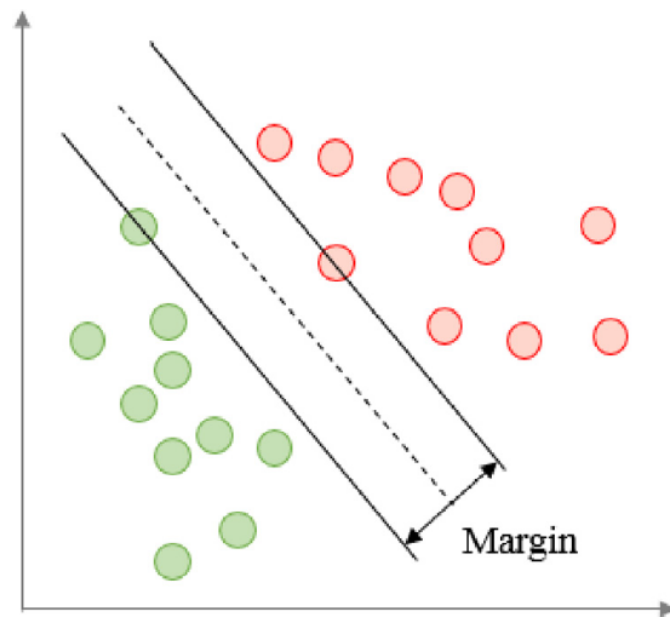


Figure 5: *Schematic of a support vector machine, adapted from [20].*

Although SVMs have historically been used for classification, they can also be used for regression purposes. Here, the machine is called support vector regression and uses the center line as the linear regression function. For non-linear problem applications, a kernel function can be used before applying support vector regression. When this is done, the input variables are mapped into a higher dimension space. This can also be done for non-linear classification.[20][30]

One of the main advantages to using SVMs is their ability to solve non-linear classification and regression problems with the help of kernel functions [32][30]. However, choosing the correct kernel function can be a challenging task at times [20][23] and modelling through SVMs can be pretty time-consuming for larger scale problems [30].

Other data-driven modelling types

Decision tree (DT) is a technique that divides data into different groups through the use of a tree-like flowchart. Such a flowchart consists of input data entering a root node, which is then split up into leaf nodes or internal nodes. At the latter, the data is split further into smaller groups, while at the former the output is presented. Similar to linear regression, DTs are easy to understand implement but aren't as accurate as the machine learning methods discussed above.[20][30]

Genetic algorithms are methods inspired by evolution (i.e. "survival of the fittest"). The main task

is to model an evolution process to identify the best solution. The user starts off with a collection of randomly chosen sets of weights used in a linear, quadratic and/or exponential equation, and applies the genetic algorithm, which yields the best set of weights and thereby most accurate equation(s).[30]

2.3.3 Grey-box models

Grey-box models, also known as hybrid models, can be viewed as the result of combining white- and black-box modelling principles [19]. The model uses physics based modelling as the basis for the model structure, and estimates its parameters with the help of empirical data.[14] They are commonly used for acquiring more accurate predictions from already existing white-box models [32]. Grey-box models differ from data-driven models in that they at least partly describe the physical processes that occur in HVAC systems. Parameter estimation also allows for identifying dynamics of the system that were previously not taken into account.[14] Advantages and disadvantages of using grey-box models are tabulated in table 4.

Table 4: *Advantages and disadvantages of grey-box HVAC modelling.*

Advantages	Disadvantages
<ul style="list-style-type: none"> • High accuracy [32] • Great potential for fault detection and diagnosis [18] • Provides insight on dynamics not previously covered in a model [14] • Can consider a limited amount of data [23] 	<ul style="list-style-type: none"> • Requires high level of user expertise in setting up equations and parameter estimation [18] • Requires data for training • High computation time [23]

2.3.4 Validation criteria

Common validation criteria for evaluating the accuracy of predictive models include the following equations:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - y_i| \quad (3)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - y_i}{y_i} \right| \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - y_i)^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - y_i)^2} \quad (6)$$

Where n is the number of output predictions, Y_i is the predicted output value and y_i s is the measured output value. MAE stands for mean absolute error. Mean absolute percentage error (MAPE) is the mean error in percentage, nondimensionalized. MSE stands for mean squared error and can be used to evaluate the variance and bias of simulation predictions. The root mean square

error has the same unit as the simulation output. [20]

2.3.5 Modelling applications

Application reviews show a variety of purposes of HVAC system modelling for all three main modelling approaches. Mathematical approaches comprising of thermal-network-, CFD- [33] and simple energy balance models [34] have been used to predict building indoor thermal environment, airflow distribution and working of various components in HVAC systems. One of the most common applications is building energy performance prediction, which has been done with multiple types of models, including physical multizone models [35], artificial neural networks and support vector machines [23]. Data-driven models have also been used for optimization of HVAC operations and minimizing building energy consumption [13].

Table 5 provides an overview of recent modelling applications with mathematical, data-driven and grey-box HVAC system models.

Table 5: Recent modelling applications in connection with HVAC systems.

Reference	Application	Model type	Methodology	Aim	Validation
Kusiak [13]	Full HVAC system	Black-box	Multilayer perceptron	Minimization of HVAC system energy consumption	Case study
Ciulla [36]	Full HVAC system	White-box and black-box	Multiple linear regression and TRNSYS model	Building energy demand prediction	Case study
Wang [34]	Cooling coil	White-box	Energy balance	Simplified modelling of cooling coil load	Experiments
Cui [37]	Cooling coil	White-box	White-box	Energy balance	Modelling and performance evaluation of air handler cooling coil
Experiment Ghiaius [38]	Cooling coil and humidifier	Grey-box	Heat exchange principles and parameter identification	Identification of air handler elements	Experiment
Hong [39]	Air handling unit	Black-box	Artificial neural network	Supply air temperature prediction	Case study
Razbann [40]	Air handling unit	White-box	Energy balance	Air handling unit subsystem energy usage optimization	Experiment
Yang [41]	Air handling unit	Grey-box	TRNSYS model and support vector regression	Fault detection strategy for air handling unit sensors	White-box simulations

3 Methodology

In this section, a description of a an HVAC system supplying ventilation to a laboratory room will be presented, followed by the development and validation of a mathematical model of this system, and lastly a description of the data-driven model development.

3.1 System description

The room in question is a testing facility within the Energy and Indoor Environment Laboratory at the Norwegian University of Science and Technology (Trondheim, Norway). The space surrounding the lab room is supplied with conditioned air (albeit slightly cooler than other zones in the building). As none of the outer layers of the smaller testing facility are exposed to outdoor air, the air conditioned space effectively functions as the outdoor climate for this system.

The lab room has an area of 27.5 m^2 , and the walls are well insulated (see chapter 3.2.2 for further details). Notably, there is an air gap within the wall construction for better sound insulation. Two of the walls are also fitted with double-pane windows with a 0.1 m air gap between the panels.

Although the room itself is not exposed to outdoor air, the space receives conditioned air from a separate air handling unit inside the outer laboratory space which takes in outdoor air, designed to deliver up to $2000 \text{ m}^3/h$ of air. The configuration of the air handling unit is shown in figure 6:

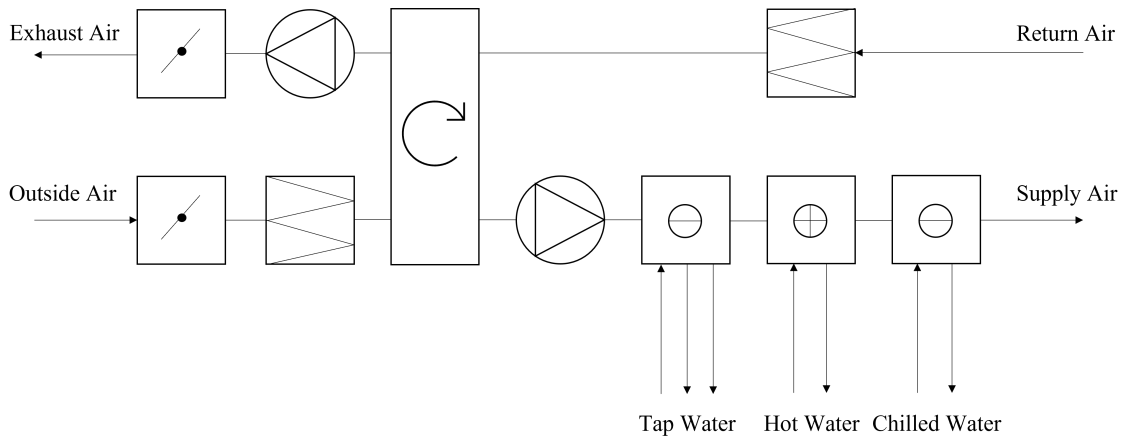


Figure 6: Configuration of components in the air handling unit for the laboratory room.

As can be seen above, the air handling unit is fitted with two dampers, two filters, a fan at both the supply and return side, a heat recovery unit, two cooling coils, and a heating coil. In principle, outdoor air enters the air handling unit on the supply side, is heated/cooled by the heat recovery unit and coils (if necessary), and is then blown by the supply fan into the ducts that distribute the supply air to the lab room. The lab room utilizes a mixed air ventilation system, meaning that the supply air enters the room from above and mixes with the existing air (ideally before reaching the occupant space). Mixed air is then sucked up through the return fan into the return side of the air handling unit and goes through the heat recovery unit before exiting.

The heat recovery unit is a rotating heat exchanger and primarily provides the supply air with heating energy from the exhaust air. It can also have a cooling function (if the exhaust air is cooler than the supply air), but this is generally not as relevant here with the cooler climate of Trondheim. As the HVAC system has both a supply- and return fan, it is important to avoid pressure discrepancies between the supply and return side. This means that the pressure rise over

the fans (and by extension, the air flow rate) should be about the same. The first cooling coil is supplied with tap water, the second with chilled water from a heat pump in the building. The heating coil is supplied with hot water in a circulating circuit from a central heating system in the building.

Even though the air handling unit is fitted with two cooling coils, only the second one (after the heating coil) is used for lowering the sensible energy of the supply air. The first cooling coil, on the other hand, functions as a dehumidifier that pre-cools the supply air before it enters the heating coil and thereby reduces the air humidity.

The air handling unit is also connected to a monitoring system (see figure 7) which provides real-time measurement data at various points. This includes air flow rate, -temperature, -relative humidity, pressure rise over components, and working frequencies of the heat recovery unit and fans. The monitoring system also allows for adjusting the setpoint pressure and temperature at the supply outlet and return inlet, as well as the fan and heat exchanger frequencies.

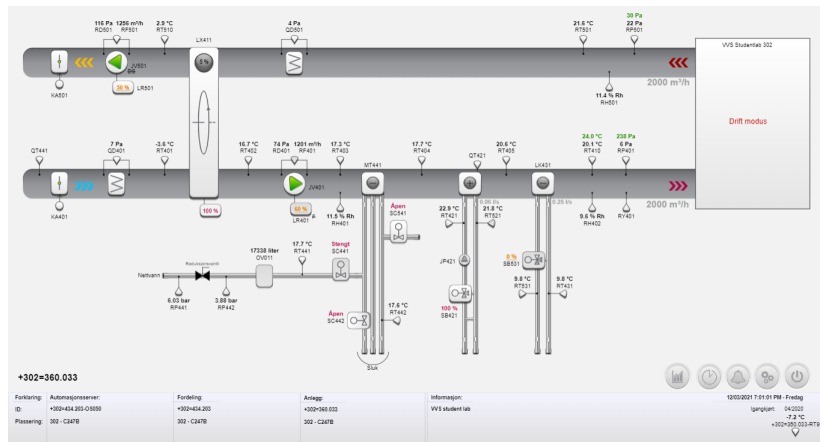


Figure 7: *Monitoring system of the air handling unit. Numbers in green are configurable setpoint values.*

In addition, the air handling unit is equipped with an outdoor sensor that provides it with weather compensation. This means that a compensation value for the supply air temperature is calculated based on the outdoor air temperature which is then added to that setpoint value. During colder days, for example, the outdoor compensation process yields an increase in the supply air temperature setpoint. The computing of this compensation value is done through the use of a compensation curve, which can be configured through the monitoring system.

3.2 Mathematical model development

3.2.1 Simulation platform

As described in section 2.3.1, there are several methodologies and platforms used for developing mathematical models of HVAC systems. For this project, EnergyPlus was chosen as the simulation platform. EnergyPlus is a building energy modelling engine designed by the U.S. Department of Energy and is a commonly used program at academic and commercial levels.[42] The simulation program is made to take into account interactions from most building components, including the building construction, internal loads and HVAC systems. It utilizes an integrated solution manager wherein the building, system and plant are solved simultaneously. In general, the program is represented by a series of elements connected by fluid loops. Further detail on these loops are discussed in section 3.2.3.

EnergyPlus models the air inside buildings by solving differential equations resulting from energy and moisture balances. It also uses conduction transfer function to compute heat fluxes on the building surfaces. Furthermore, EnergyPlus allows for a high degree of freedom when designing models with it and can easily be integrated with third party interfaces. Choosing EnergyPlus as the simulation platform was partly motivated by the latter point, since it will be easier to use this simulation engine in conjunction with a black-box model.

To use this software, the following system information was required to create a model accurate to a satisfactory degree:

- Building geometry, orientation and materials.
- Individual components in the air handling unit and their means of operation.
- Layout of component connections.

3.2.2 Room construction

As stated earlier, EnergyPlus requires the specification of the building envelope and its construction. In this case, the laboratory room is treated as the building. The geometry of the room was described based on plan- and section drawings of the space (see appendix C and D), and is illustrated in figure 8. EnergyPlus utilizes a three-dimensional coordinate system where all the coordinates of all the vertices for each surface need to be defined. In total, 6 building surface objects were created; 4 walls (North, South, West and East), a floor, and a roof. Moreover, two fenestration objects were created on the south and west walls in order to represent the windows on the lab room.

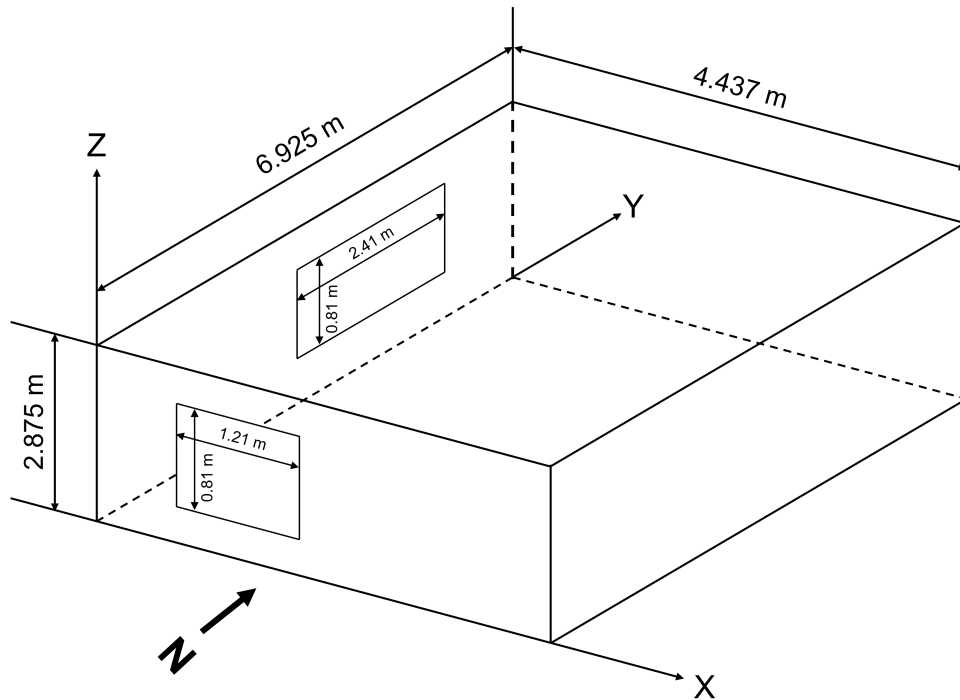


Figure 8: Sketch of the laboratory room geometry used in the model (not to scale). North direction is labelled at the bottom.

In addition, the construction of each surface had to be specified. This was also done based on the section drawings of the space (see appendix C and D). Here, the construction of the south- and

north walls were assumed to be the same as the west- and east walls, respectively. This was done as there was no documentation found regarding the construction of those two walls. Furthermore, the thermal properties of all the materials used had to be documented, which can be seen in table 6. Most of these properties were found based on similar materials used in other pre-built EnergyPlus models. GLAVA Proff 34 was assumed as the insulation material.

Table 6: *Properties of materials used in the model; roughness, thermal conductivity, density and specific heat.*

Material	Roughness	Thermal Conductivity (W/mK)	Density (kg/m^3)	Specific heat (J/kgK)
GLAVA Proff 34	Rough	0.034	20	840
Plasterboard	MediumSmooth	0.16	950	840
Soundboard	Rough	0.034	80	840
Wood	Rough	0.14	530	900
Chipboard	Rough	0.12	700	900
PVC	Smooth	0.19	900	900

The U-values (thermal transmittance) of each surface can be calculated by EnergyPlus and reported in an HTML report. These are tabulated in table 7.

Table 7: *Surface U-values of room model.*

Surface	U-value no film (W/m^2K)
North and East Walls	0.128
South and West Walls	0.177
Floor	0.146
Roof	0.126
Windows	0.150

3.2.3 HVAC system

EnergyPlus utilizes a node-branch system management scheme for modelling the HVAC systems of buildings. In essence, the HVAC system is represented by a network of fluid "loops" with connecting branches. Each branch consists of at least one component (e.g. a pipe, coil, boiler, etc...) and have nodes that connect components and also represent a point in a loop. There are generally two types of loops in EnergyPlus: air loops and plant loops. Air loops transport air (usually for representing air handling units) while plant loops transport liquids; chilled and hot water, for instance. In some cases, for example when adding cooling towers, a condenser loop is used. Each loop is also split into two half-loops representing the supply and demand side.

For this model, 4 loops were created: one air loop for the air going through the air handler and room, and one plant loop for each of the coils in the air handler. The air loop and its branches are shown in figure 9.

Here, the air entering and leaving the room represents the demand side of the loop, while the air flowing through the components in the air handler represents the supply side. As the air handler utilizes a rotary heat exchanger for heat recovery, an outdoor air subsystem within the air loop had to be created. This outdoor air system also had to include a mixing box. However, as there is no mixing of the outdoor air with the exhaust air, the mixed air fraction was set to 0 for this model. Moreover, the return fan had to be removed due to complications with the functionality of the heat recovery unit in the model. The supply fan was also moved to the end of the supply side of the loop, whereas in reality the fan is placed before the dehumidifier.

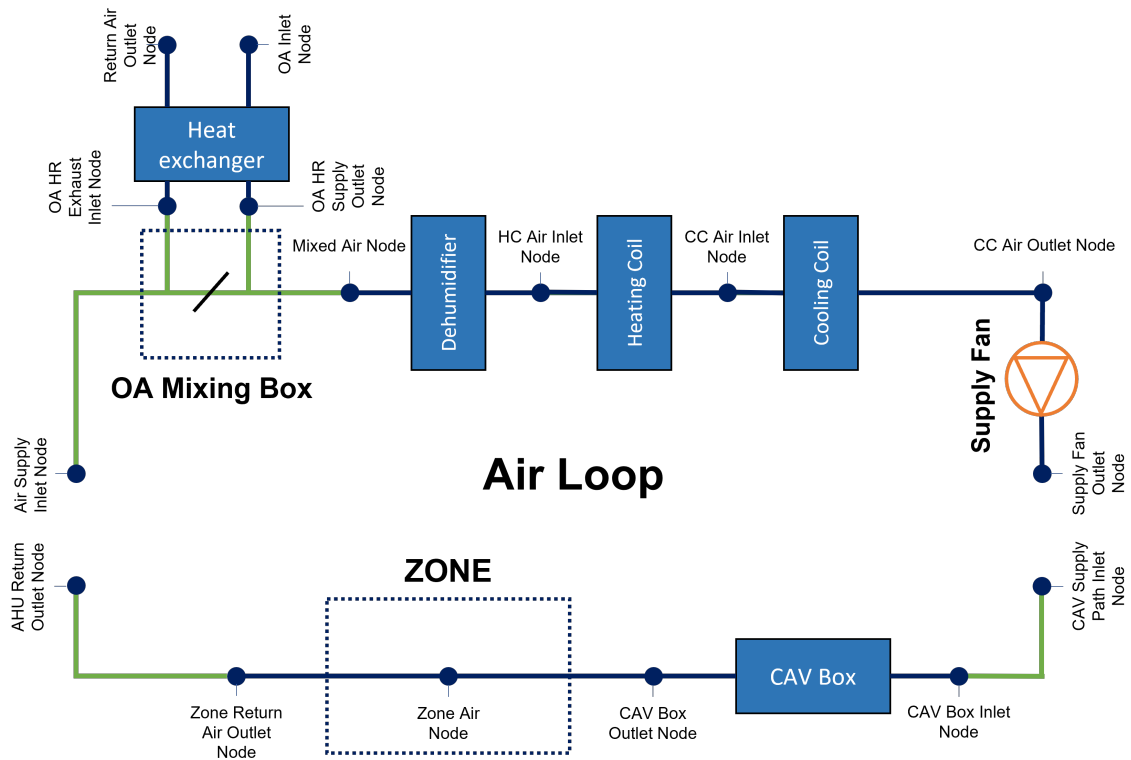


Figure 9: Node diagram of the air loop.

In the case of the plant loops, two chilled water loops were made for the cooling coils and one hot water loop for the heating coil. For the sake of simplicity, all the water loops were made with the same structure; a pump and district heating/cooling on the supply side (albeit a slightly higher supply temperature for the dehumidifier loop compared to the other cooling coil), and a bypass pipe on each side. A node diagram of the hot water loop connected to the heating coil is shown above in figure 10. The chilled water loops can be seen in appendix A.

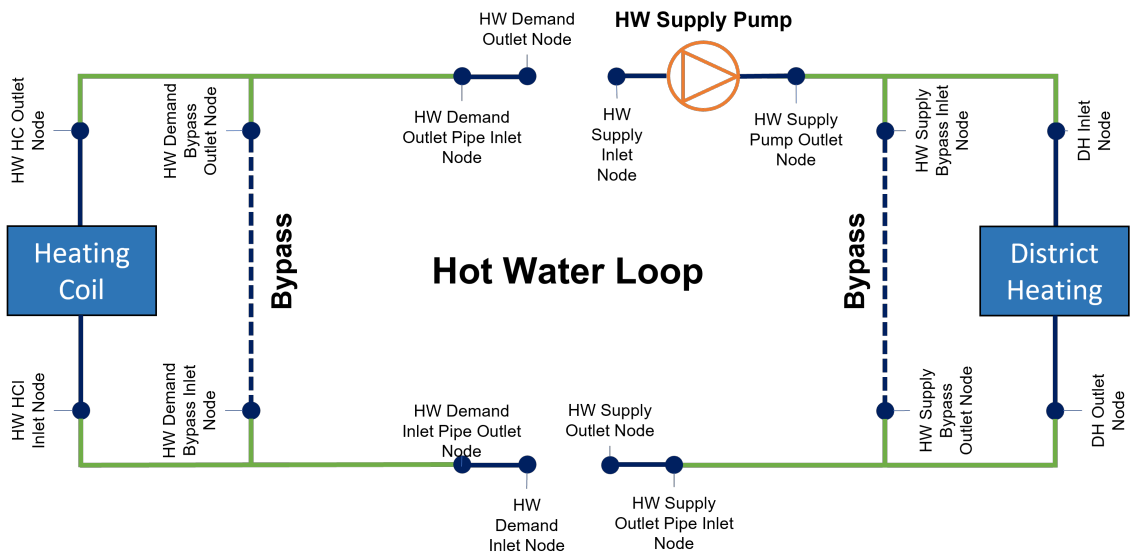


Figure 10: Node diagram of the hot water loop connected to heating coil. Each node is labelled with a name in the model.

In addition to creating these fluid loop structures, each component had their design conditions

defined as well. In the case of the rotating heat exchanger, the sensible and latent effectiveness had to be specified. The sensible effectiveness of a heat recovery unit can be calculated with the following equation:

$$\eta = \frac{T_2 - T_1}{T_3 - T_1} \quad (7)$$

Where T_1 and T_2 are the inlet and outlet temperatures on the supply side, and T_3 is the inlet temperature on the return side. From previous temperature readings on the monitoring system of the air handler, it was found that the heat exchanger has a sensible effectiveness of around 80% at the design flow rate of $2000 \text{ m}^3/\text{h}$, and only drops below this when it operates at a frequency below 60% (i.e. at low supply air temperature). For the latent effectiveness, it was assumed to be the same as the sensible effectiveness.

The design conditions of the coils are tabulated in table 8. For further details, see appendix A.

Table 8: *Design conditions of the coils in the air handler.*

Coil	$T_{air,in}$ ($^{\circ}C$)	$T_{air,out}$ ($^{\circ}C$)	$T_{water,in}$ ($^{\circ}C$)	$T_{water,out}$ ($^{\circ}C$)	Air flow rate (m^3/h)
Dehumidifier	24.0	11.0	7.0	12.0	2000
Heating coil	11.0	22.0	60	40	2000
Cooling coil	22.5	17.0	9.0	14.0	2000

3.2.4 Experimental validation

For the validation process of the mathematical model, measurements were held in the laboratory room with the air handling unit operating at different conditions.

Experimental setup

Two EasyLog temperature loggers were placed inside the laboratory room, one of them 1 meter away from the larger window on the west-side wall (EasyLog 6), and the other 1 meter away from the middle of the south-side wall (EasyLog 4) - both approximately 75 cm above the floor. In the time period where the temperature loggers were inside the laboratory room, multiple lab sessions were held inside the room, so they had to be placed in locations where they would not be of disturbance to occupants. Both of these loggers were set to measure the drybulb temperature (as well as the dewpoint temperature and relative humidity) at 10 minute intervals. The indoor temperature was then calculated by taking the average of them.

In the previous paper [1], thermocouples were also mounted outside the laboratory room by the air handling unit. These measured the air temperature outside the room as well as the air temperature at various points in the air handling unit in order to measure the performance of the heat recovery unit. For this project, these data points were no longer necessary for testing the mathematical models, but the results will be relevant for the input data implementation described later on.

Scenarios

The indoor air temperature was recorded over a 79 day period from March 4th 12:00 to May 21st 24:00. Over this 79 day period there were 4 time periods (scenarios) picked out for testing the accuracy of the mathematical model under different operating conditions. During the 79 day period the air flow rate along with the setpoint temperature were constantly changing. The highest and lowest air flow rate were $2300m^3/h$ and $350m^3/h$, respectively, and the highest and lowest setpoint temperatures were $26.5^{\circ}C$ and $6.6^{\circ}C$, respectively. No changes were made to the setpoint pressure

values, nor the opening levels of valves in the circuit. However, the air handler’s controller system could change the latter point if necessary to achieve the setpoint values.

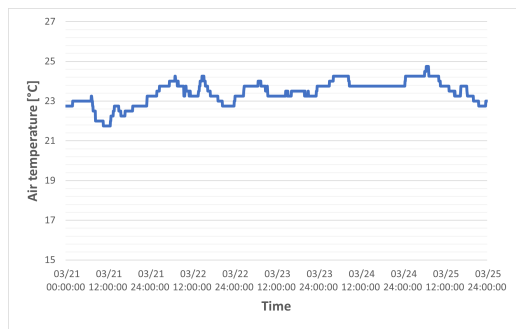
The first scenario took place from March 21st 00:00 to March 25th 24:00. In this scenario, the air handler was operating with a constant setpoint of $23^{\circ}C$. This scenario was taken into account in order to test if the mathematical model can predict a stable indoor air temperature.

The second scenario took place from March 28th 00:00 to April 1st 24:00. During this period, multiple lab sessions were held with upwards of 6 occupants inside the laboratory room. By having the mathematical model predict this time period, we could test the model’s ability to predict the effect of occupancy in the room.

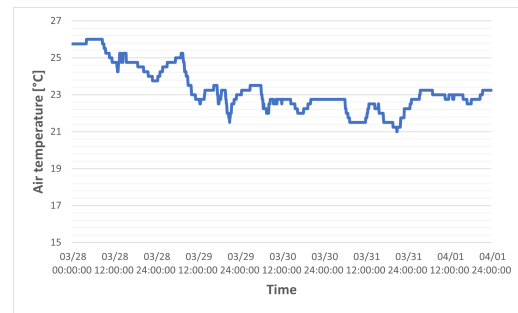
The third scenario took place from April 8th 00:00 to April 19th 24:00. During this period, the system was in free-floating operation (i.e. no ventilation from the air handler) due to the easter break, with minimal occupancy. This scenario will test how the model predicts the lack in ventilation affects the indoor air temperature.

The fourth scenario took place from May 10th 00:00 to May 18th 12:00. Here, the setpoint temperature was set to $15^{\circ}C$ from May 10th 09:00 to May 12th 09:00, then $26^{\circ}C$ from May 12th 09:00 to May 14th 09:00, and then $22^{\circ}C$ for the rest of the time period. This scenario was chosen in order to test if the model can represent the thermal inertia of the system accurately enough.

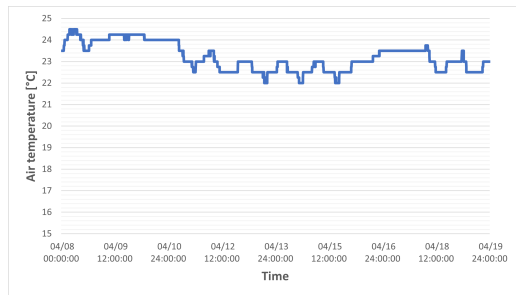
The indoor air temperature for each of these scenarios is illustrated in figure 11:



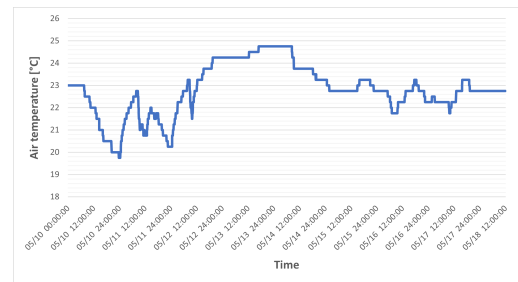
(a) Indoor air temperature constant setpoint 21.03-25.03.



(b) Indoor air temperature occupancy 28.03-01.04.



(c) Indoor air temperature free-floating 08.04-19.04.



(d) Indoor air temperature variable setpoint 10.05-18.05.

Figure 11: Indoor air temperature for 4 different scenarios.

3.2.5 Input data implementation

Weather file

In order to perform simulations in EnergyPlus, a weather file in the epw format (Energy Plus Weather) is required in addition to the model file. These weather files contain hourly values for a set period of time and need the following data:

- Dry bulb temperature of the outside air in celsius.
- Dew point temperature of the outside air in celsius.
- Air relative humidity in percentage.
- The atmospheric station pressure in pascal.
- Horizontal infrared radiation intensity in Wh/m^2 . If this is missing in the weather file, the Opaque Sky Cover value is used to calculate it.
- Direct normal radiation in Wh/m^2 . This represents the amount of solar radiation a surface perpendicular to the sun's rays receives directly from the solar disk (i.e. the visible surface of the sun).
- Diffuse horizontal radiation in Wh/m^2 . This represents the amount of solar radiation a horizontal surface receives from the sky.
- The wind direction in degrees, where North is 0° , East is 90° , etc.
- Wind speed in m/s .
- Total sky cover. This is the amount of the sky covered by clouds or other phenomena, given in tenths.
- Opaque sky cover. This is the amount of the sky covered by clouds or other phenomena preventing observation of the sky or higher cloud layers, given in tenths. Only used if the Horizontal infrared radiation intensity is not given in the file.
- Present weather (i.e. the weather at the time of measurement).
- Snow depth in cm .
- Liquid precipitation depth in mm .

There are also various other input fields that currently are not used in EnergyPlus calculations. For most of the data input listed above, a weather file from Værnes airport (2007-2021) was used to represent the outdoor climate conditions during the experiments. The outside dry bulb temperature was gathered through the temperature logger at the inlet to the AHU connected to the monitoring system described in section 3.1. However, there was no humidity logger at this location, only after the heat recovery unit. In order to find the outside dew point temperature and relative humidity, the dry bulb temperature and relative humidity after the heat recovery unit was gathered from the monitoring system. Using these 2 values, the dew point temperature after the heat recovery unit could be approximated through the use of the Magnus formula described in [43]:

$$T_{dew} = \frac{\beta * (\ln(\frac{RH}{100}) + \frac{\alpha T_{dry}}{\beta + T_{dry}})}{\alpha - (\ln(\frac{RH}{100}) + \frac{\alpha T_{dry}}{\beta + T_{dry}})} \quad (8)$$

Where β and α are $243.04^\circ C$ and 17.625 , respectively as recommended by [43]. An assumption was then made that there is no mass transfer through the heat recovery unit such that the dew point temperature remains the same (i.e. the dew point at the inlet is the same as after the heat recovery unit). With both the outside dew point- and dry bulb temperatures available, the outside relative humidity can be calculated by rearranging equation 8 as follows:

$$RH = 100 * \left[\frac{e^{\frac{\alpha T_{dew}}{\beta + T_{dew}}}}{e^{\frac{\alpha T_{dry}}{\beta + T_{dry}}}} \right] \quad (9)$$

Outside wall temperature

In most cases, the weather file would provide enough data for describing the outdoor climate. However, for this project the air surrounding the laboratory room is at a different condition from the air entering the air handling unit as described in section 3.1. To take this into consideration, the outside boundary condition of the room surfaces were changed to refer to a repeating hourly schedule, containing the drybulb temperature measurements from measurements made outside the air handling unit in the previous paper [1]. The roof and northern wall (See figure 8) were not referred to this schedule (as they were not exposed to air of the laboratory space) and were instead defined as adiabatic surfaces. This was because there was a lack in data of the air conditions surrounding the exterior of these surfaces and it was also likely that the difference in temperature across these surfaces were negligible in comparison to other heating/cooling loads.

Due to complications arising from the changes made to the outside boundary condition of the other surfaces, the fenestration surfaces had to be removed from the construction of the south and west walls (see figure 8). As was shown in table 7, the windows have much worse insulating capabilities in comparison to the rest of the surfaces. To compensate for this, the amount of insulating material inside the surface constructions were reduced in order to achieve the same average U-value of the room envelope. This corresponded to approximately a 25% decrease in the insulation thickness. The final U-values of each surface are presented in table 9.

Table 9: *Final surface U-values (no film) after removal of fenestration surfaces.*

Surface	New U-value no film (W/m^2K)
North and East Walls	0.168
South and West Walls	0.232
Floor	0.194
Roof	0.165

AHU operation

Regarding the operation of the HVAC systems, EnergyPlus allows for scheduling various operating schemes. For this model, the following parameters were made schedules for:

- Water loop supply temperatures. These were specified to match the design specifications.
- Supply setpoint air temperature. This was scheduled to match the supply setpoint air temperature recorded by the monitoring system.
- Air handling unit availability (on/off). This was scheduled based on the operating signal of the fans recorded by the monitoring system.
- Air flow rate. This was scheduled to match the supply air flow rate recorded by the monitoring system.

-
- Heat load from electrical equipment such as computers. This was set to be on from 07:00 to 17:00 with the initial heat load being 200W before calibrating the model (see section 3.2.6).
 - Heat load from lights. The lights were assumed to be on from 07:00 to 16:00 during weekdays and provide a heat load of approximately 350W.
 - Heat load from occupancy. The number of people was scheduled based on observations made throughout the measurement period (it is highly likely that there are multiple periods where there were undocumented occupants). The heat produced per person was initially assumed to be 140W before calibrating the model (see section 3.2.6).

After altering the schedules to match the input data, simulations were performed using the measuring data from each of the four scenarios with 12 time steps per hour (i.e. calculations every 5 minutes).

3.2.6 Model calibration

When performing the initial simulations, internal thermal masses and the air infiltration rate had not yet been defined. These are input data that are often hard to quantify and measure. As such, a hybrid model approach was used to calibrate the model to best fit the measured indoor air temperature.

There are two objects available in EnergyPlus that can be used to model internal thermal masses. The first object is the `internal_mass` object where one can define the construction of internal furnishing elements - essentially adding extra material inside the model of the room. How much of this material is added can be modified by changing the area of the `internal_mass` object. The second object that can be used is the `Zone capacitance multiplier`. By using this object, a multiplier is added to the calculation of the heat capacity of the air:

$$C_z = V \rho_{air} C_p C_T \quad (10)$$

Where V is the volume of the room, ρ_{air} is the air density, C_p is the air specific heat, and C_T is the heat capacity multiplier. A heat capacity multiplier greater than 1 will lead to an increase in the heat capacity of the air. The heat capacity of the air is used in the heat balance equation which forms the basis of EnergyPlus's simultaneous solution scheme:

$$C_z \frac{dT_z}{dt} = \sum_{i=1}^{N_{sl}} \dot{Q}_i + \sum_{i=1}^{N_{surfaces}} h_i A_i (T_{si} - T_z) + \sum_{i=1}^{N_{zones}} \dot{m}_i C_p (T_{zi} - T_z) + \dot{m}_{inf} C_p (T_{\infty} - T_z) + \dot{Q}_{sys} \quad (11)$$

Where $C_z \frac{dT_z}{dt}$ is the energy stored in the air and the other terms is the heat transfer due to convection, infiltration and air mixing. In addition to these parameters, the heat gain per person and the heat gain from electric equipment in the room were assumed as listed in the subsection above.

In the previous paper [1], a manual calibration procedure was utilized in identifying the heat capacity multiplier and air infiltration to be 15 and 0.3 air changes per hour, respectively. However, this paper calibrated the model against much shorter and less detailed datasets compared to what is available for this project. As such, the model was calibrated a second time, with the initial configuration being the one developed in [REF]. In addition to finding updated values to the heat

capacitance and air infiltration, this calibration procedure found updated values to the heat gain per person, heat gain from electric equipment and area of the internal mass object.

As the number of parameters to adjust increased from 2 to 5, a manual calibration process was not realistic due to time constraints. As such, a "brute-force" method was implemented where a number of different models with their own unique combination of the 5 parameters were generated and tested against the 4 scenarios described in section 3.2.4.

Generating the models was done in python by generating copies of the original model and replacing the 5 parameters with one of 4-6 values defined for each parameter. The range of values for each of these parameters were chosen based on what is realistically possible. These ranges are tabulated in the following table:

Table 10: Calibration parameter value ranges

Parameter	No. values	Value range
Heat capacitance multiplier	6	5-30
Internal mass area	5	0-100m ²
Heat gain per person	4	80-140W
Electrical equipment heat load	4	150-300W
Air infiltration	6	0-0.5/h

With this procedure 2880 models were generated. Evaluating these models was also done in python by using the subprocess package to run the simulation via python. For each model, the root mean squared error, mean absolute error and mean absolute percentage error of the predicted indoor air temperature were calculated and saved in an array. Finally, the 3 models with the best accuracy were extracted from this array. The code for this calibration process can be found in appendix E.

3.2.7 Year-long simulation

After calibrating the mathematical model, a year-long simulation was performed in order to generate training data for one of 2 black-box models to be trained. Here, it was key to create a dataset with varying outdoor air conditions, air handler availability, supply setpoint temperatures, air flow rate and internal gains. In addition, the value range of parameters such as the supply setpoint temperature had to be equal to or greater than the range of values from measurements made from 04.03 to 22.05. This was because the training and testing data needs to be normalized, and if the measured dataset has a larger range of values, the scaled values will exceed the range of what values the model is trained for.

Meteorological data

For the weather data, a weather file from Værnes airport (2007-2021) was used with a couple adjustments made to the relative humidity and dew point temperature. The measured dataset had a minimum relative humidity of 9.5% whereas the original weather file's relative humidity only went down to 19%. As such, the relative humidity was lowered in a 1-hour time period such that there was a greater range in relative humidity values. The same was also done with the dew point temperature. The outdoor dry bulb-, wet bulb temperature and relative humidity is illustrated in the following figure:

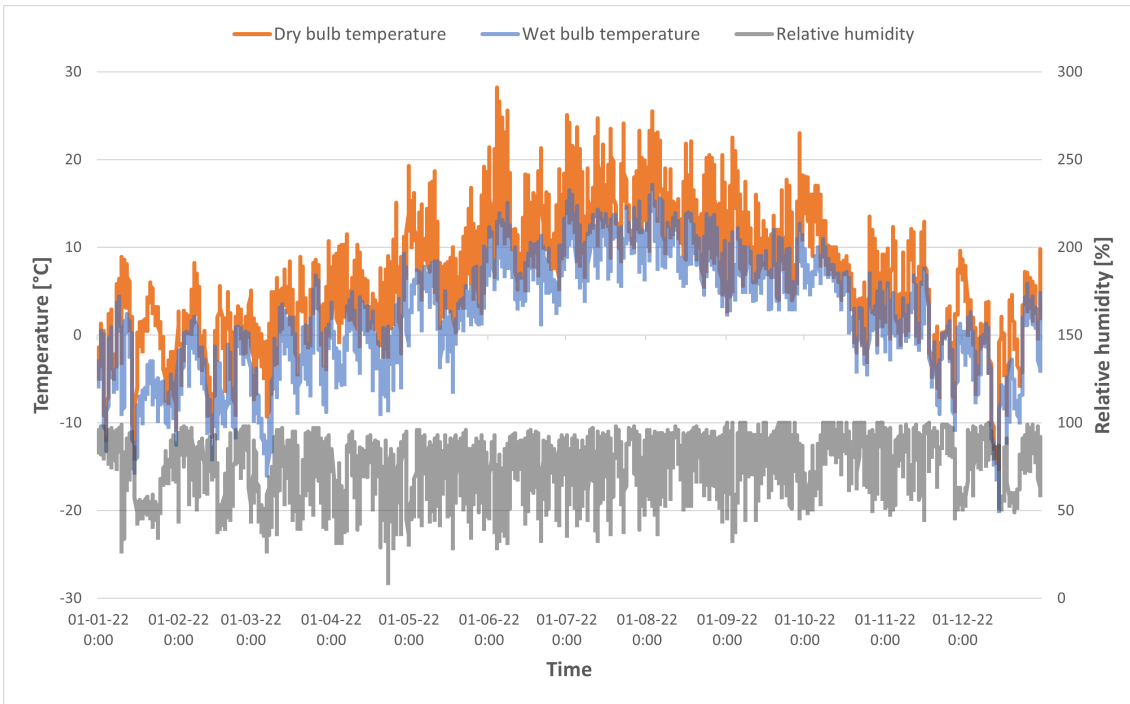


Figure 12: Outdoor dry bulb-, wet bulb temperature and relative humidity for year-long simulation.

AHU availability

Throughout the year-long simulation, the operation of the air handler swapped between 3 "modes". For the first mode ("standard operation"), the air handler was available from 07:00 to 18:00 for all weekdays. In the second mode ("longer operation") the air handler was available from 07:00 to 23:00 for all weekdays. In the third mode ("holiday") the air handler was shut off for all days. The distribution of these 3 modes is tabulated in table 11:

Table 11: Distribution of AHU availability modes in year-long simulation.

Period	Standard operation	Longer operation	Holiday
01/01 - 01/02	x		
01/02 - 14/02		x	
14/02 - 15/03	x		
15/03 - 31/03			x
31/03 - 15/05	x		
15/05 - 31/05		x	
31/05 - 30/06			x
30/06 - 30/09	x		
30/09 - 31/10		x	
31/10 - 15/12	x		
15/12 - 31/12			x

Supply setpoint air temperature

The supply setpoint air temperature was set to vary between 20°C and 26.5°C from winter to summer, but the simulation also had to include a setpoint temperature below 6.6°C to match the measured setpoint temperature range. As such, the period 14/01 - 21/01 was designated a supply setpoint air temperature of 6°C . The supply setpoint air temperature through the whole simulation is shown in figure 13:

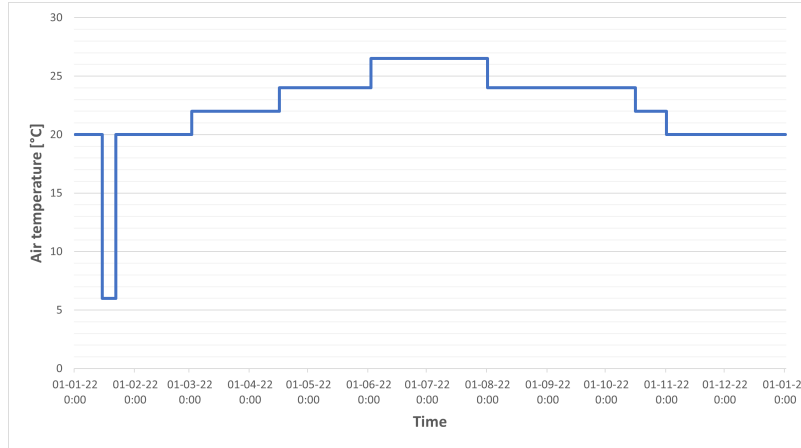


Figure 13: *Supply setpoint air temperature for year-long simulation.*

Air flow rate

For the year-long simulation, the air flow rate swapped between 3 flow rates (aside from no ventilation): $2400\text{m}^3/\text{h}$, $720\text{m}^3/\text{h}$ and $360\text{m}^3/\text{h}$. For most of the simulation the air flow rate was set to $720\text{m}^3/\text{h}$. The distribution of flow rate periods is tabulated in table 12:

Table 12: Distribution of air flow rate in year-long simulation.

Period	Air flow rate (m^3/h)
01/01 - 31/01	2400
31/01 - 31/03	360
31/03 - 30/06	720
30/06 - 31/07	2400
31/07 - 30/09	360
30/09 - 31/12	720

Internal gains

The electric equipment was set to be on from 07:00 to 17:00 for all days throughout the year, while the lights had the same schedule except that they were off during weekends and the holiday periods specified in table 11. The occupancy level was scheduled to vary between 0 occupants throughout the day to upwards of 6 occupants anytime between 07:00 and 16:00.

3.3 Black-box model development

3.3.1 Model architecture

In this thesis, a recurrent artificial neural network of the long short-term memory (LSTM) variant is utilized as the black-box model. RNNs have an advantage when it comes to predicting the behaviour of HVAC systems, since the air temperature is highly dependant on condition of the air before. The model is also capable of taking in multiple inputs and producing multiple outputs, but for this project only one output will be needed.

The development of the black-box model architecture was done by Gaurav Chaudhary, PhD candidate. The model was made through Python with the Keras library, which is a neural network API that runs on top of TensorFlow (a machine learning platform). The model uses multiple timesteps to predict the indoor air temperature multiple timesteps ahead. For this project the timesteps amounted to the model taking in 4 days worth of data and predicting the air temperature for following 24 hours. The model has a sequence-to-sequence architecture and uses multiple LSTMs for encoding and decoding the time series sequence.

In terms of the hyperparameters, the model has 100 neural network layers and uses a batch size (how many rows/samples to go through before updating the model parameters) of 64 and 50 epochs (how many times the algorithm goes through the training dataset). In addition, a dropout factor of 0.3 is added to prevent the model from being overfitted to the training data.

The code for the model can be seen in appendix F.

3.3.2 Training and evaluation of models

There are two datasets that can be used to train a black-box model in this project. The first is the measured dataset over a 79 day period and the second is the dataset produced by the year-long simulation in EnergyPlus. For each of these datasets, a black-box model was trained and tested with largely the same model architecture. The only difference between the two is that the black-box model trained with the simulation data utilizes a time step length of 30 minutes while the model trained with measured data utilizes a 15 minute length. This was chosen due to the simulation dataset being significantly larger than the measured dataset, and with initial tests showing that there was not a large difference between 15 minutes and 30 minutes for the simulation-data-trained model in terms of model accuracy. Moreover, training with the two datasets took roughly the same amount of time when using those two timestep lengths.

In terms of the distribution of the datasets, initial testing revealed that a distribution of 60% training, 20% validation and 20% testing was an attractive choice for both datasets. This is also a common distribution used in the field as found in the literature review. This results in the measured dataset being split into 47 days for training, and 16 for validation and testing each. For the simulation dataset, it's split into 219 days for training, and 73 days for validation and testing each. The testing periods are also the last days in the datasets unless the testing period is deliberately changed. This means that the model trained with the measured data will be testing for the period 07.05 to 22.05 and the model trained with simulated data will be testing for the period 20.10 to 31.12.

In order to determine the accuracy of the models, RMSE, MAE and MAPE of the air temperature predictions are calculated for the last timestep (i.e. the 24 hour prediction). When multiple models are compared with each other, only the RMSE is calculated as it is a commonly used error metric.

3.3.3 Input parameters

For the two black-box models to be comparable, the input parameters had to be the same. This put a significant restriction on the available input parameters. Common input parameters such as the heating demand can be generated for the simulated dataset, but not as easily for the measured dataset, for instance. As the building characteristics of the system are constant for both datasets, these are not included as input parameters. In terms of meteorological data, only the datapoints related to the air coming into the air handler is relevant, since the room itself does not interact with the outdoor environment. Internal gains through equipment and occupancy is included, as is the time index - which is split into the day of the week and hour of the day. In addition, a couple parameters related to the air handler operation are included as well. The input parameters and the data category they fall into are tabulated in table 13.

Table 13: Input parameters for black-box models

Input parameter	Data category	Unit
Day of the week	Time index	-
Hour of the day	Time index	-
Occupancy	Internal gains	-
Equipment	Internal gains	W
Lights	Internal gains	W
Outdoor dry bulb air temperature	Meteorological	$^{\circ}C$
Outdoor dew point air temperature	Meteorological	$^{\circ}C$
Outdoor relative humidity	Meteorological	%
Air flow rate	AHU	m^3/h
AHU availability	AHU	-
Supply setpoint air temperature	AHU	$^{\circ}C$
Indoor air temperature	-	$^{\circ}C$

In total, the black-box models use 12 input features for training. It's worth noting that the indoor air temperature is also the output parameter in addition to being an input parameter.

3.3.4 Model testing

For each of the tests described below, the models were trained and tested 3-5 times depending on how computationally expensive the tests are in order to get a more reliable picture of the model accuracy for each test.

Dataset distribution

The effect the ratio of the dataset distribution has on the black-box models was tested by comparing 4 dataset distributions. In the first case (80/10/10), 80% of the data is for training, with the rest split evenly for validation and testing. In the second case (60/20/20), 60% of the data is for training (this is the distribution used for all other tests described in this section). In the third case (50/25/25), 50% of the data is for training. In the fourth case, the dataset is distributed equally for training, validation and testing (i.e. 33% each).

Timestep length

For testing the effect of the timestep length, the black-box models were trained using datasets with a 5 minute, 10 minute, 15 minute, 30 minute, and 60 minute timestep. For the simulated dataset, the 5 minute case was omitted as it resulted in the model training requiring too much memory. Fewer simulations were also performed with the datasets with shorter timestep lengths due to time limitations.

Testing periods

3 different testing time periods were defined in order to test the model accuracy when the testing period is shifted. For each of these cases, the models were trained and tested 5 times with a dataset distribution of 60/20/20.

Table 14: Dataset distributions

Case	Training	Validation	Testing
Testing period 1	01.01 - 08.08	08.08 - 20.10	20.10 - 31.12
	04.03 - 20.04	20.04 - 07.05	07.05 - 22.05
Testing period 2	01.01 - 08.08	20.10 - 31.12	08.08 - 20.10
	04.03 - 20.04	07.05 - 22.05	20.04 - 07.05
Testing period 3	15.03 - 20.10	20.10 - 31.12	01.01 - 15.03
	21.03 - 07.05	07.05 - 22.05	04.03 - 21.03

Feature relevance

In this test the relevance of each input parameter is investigated by removing one of the features and training the black-box model without that feature. This means that for these cases, the number of input parameters goes down to 11. To see if the accuracy increases or decreases for each of these cases, the error metrics are compared with simulations performed with all the input parameters.

In addition, more input parameters were later removed to see how removing the different categories of data described in table 13 affect the model accuracy. For these cases, the input parameters were reduced to 9 or 10.

Measured data prediction

After training the black-box model with the simulated dataset, the model had to predict the measured indoor air temperature. This was done by applying the scalers used for normalizing the simulated dataset to the measured dataset and then using the scaled dataset as the input for prediction. This allows for evaluating the black-box model's ability to predict the measured indoor air temperature when it was originally trained with data produced by the mathematical model described in section 3.2.

4 Results

4.1 Mathematical model

In this section, the mathematical model before and after calibration will be tested against the measured data in 4 scenarios (detailed in section 3.2.4), followed by a look into the year-long simulation that generates the dataset for training one of the black-box models.

4.1.1 Model calibration

From the calibration process detailed in section 3.2.6, the 3 models generated with the least error were taken for further consideration. The parameters of each model, along with the manually calibrated model from [1] are tabulated in table 15:

Table 15: *Parameter values for most accurately calibrated models ("model 1-3") and the original model ("previous model").*

Model	Heat capacitance	Infiltration (/h)	Equipment (W)	Activity level (W)	Internal mass (m^2)
Previous model	15	0.3	200	140	0
Model 1	30	0	150	80	100
Model 2	30	0.3	150	80	100
Model 3	30	0.5	150	80	100

As can be seen in the table above, models 1 through 3 have nearly identical values for their parameters, with the only difference being the air infiltration value. Compared with the manually calibrated model, the values of the parameters related to thermal inertia (i.e. heat capacitance and internal mass) are far greater for the calibrated models. The opposite is true for the parameters related to internal gains, where both the activity level and equipment load have been decreased.

When it comes to the error of models 1 through 3 when simulating all the 4 scenarios described in section 3.2.4, they all had an RMSE, MAE and MAPE of around $2.02^\circ C$, $1.40^\circ C$ and 6.20%. The most accurate of these was model 1. Errors for model 1 and the manually calibrated model are tabulated in the following table:

Table 16: *Average RMSE, MAE and MAPE of the most accurately calibrated model and the original model relative to measured indoor air temperature over 4 different scenarios.*

Model	RMSE ($^\circ C$)	MAE ($^\circ C$)	MAPE (%)
Previous model	2.708	1.968	8.648
Model 1	2.017	1.402	6.161

Based on table 16, there is a clear decrease in the prediction error when going from the manually calibrated model to the new model. By using the new model, the RMSE, MAE and MAPE is decreased by 25.5%, 28.7% and 28.7%, respectively. As model 1 was the most accurately calibrated model from the calibration process, this model was used when performing further simulations in EnergyPlus.

The rest of this subsection will contain a further look into each of the 4 scenarios described in section [MODEL TESTING SECTION].

Constant setpoint

The predicted indoor drybulb air temperature from the old and new EnergyPlus models and the measured temperature during the constant setpoint scenario are depicted in figure 14:

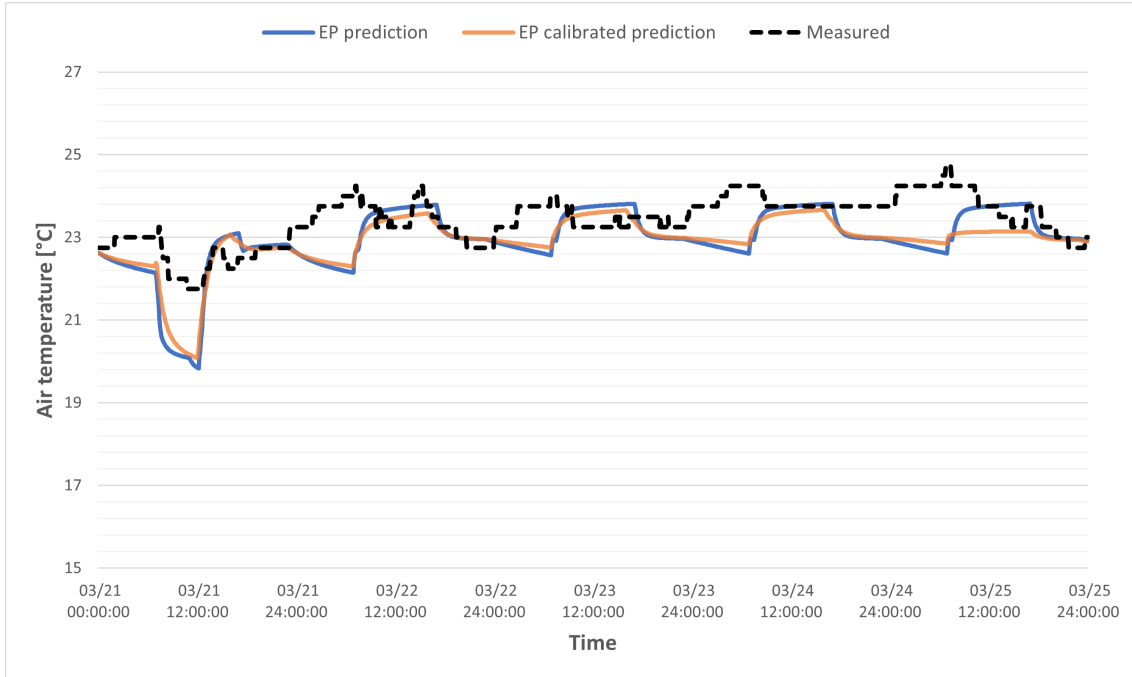


Figure 14: *Indoor air temperature with constant setpoint temperature, 21.03-25.03.*

From the figure above it can be seen that the predicted indoor air temperature from both EnergyPlus models are very similar throughout the simulation, with the calibrated model predicting slightly slower changes in the temperature. This is especially highlighted from 07:00 to 16:00 in the last day (25.03).

Compared with the measured indoor air temperature, the predicted air temperatures seem to match the most in the middle of the day as the AHU is in operation. When the AHU no longer provides conditioned air to the room, the predicted air temperature reduces gradually until the AHU turns on again - creating an oscillating predicted air temperature for both models. In actuality, the indoor air temperature seems to increase slightly overnight. There is also a large gap in the predicted and measured air temperature from 07:00 to 12:00 in the first day (21.03) - this is due to the setpoint temperature in this period being 18.5°C as opposed to 23°C for the rest of the scenario.

It is not evident which of the two EnergyPlus models yield the most accurate air temperature predictions from figure 14, but by calculating the RMSE, MAE and MAPE it was found that the calibrated model is slightly more accurate. These error values are tabulated in table 17:

Table 17: *RMSE, MAE and MAPE of the calibrated and old model when simulating a constant setpoint.*

Model	RMSE ($^{\circ}\text{C}$)	MAE ($^{\circ}\text{C}$)	MAPE (%)
Previous model	0.842	0.635	2.705
Model 1	0.767	0.601	2.553

Variable internal gains

The predicted indoor drybulb air temperature from the old and new EnergyPlus models and the measured temperature during the variable occupation scenario are illustrated in 15:

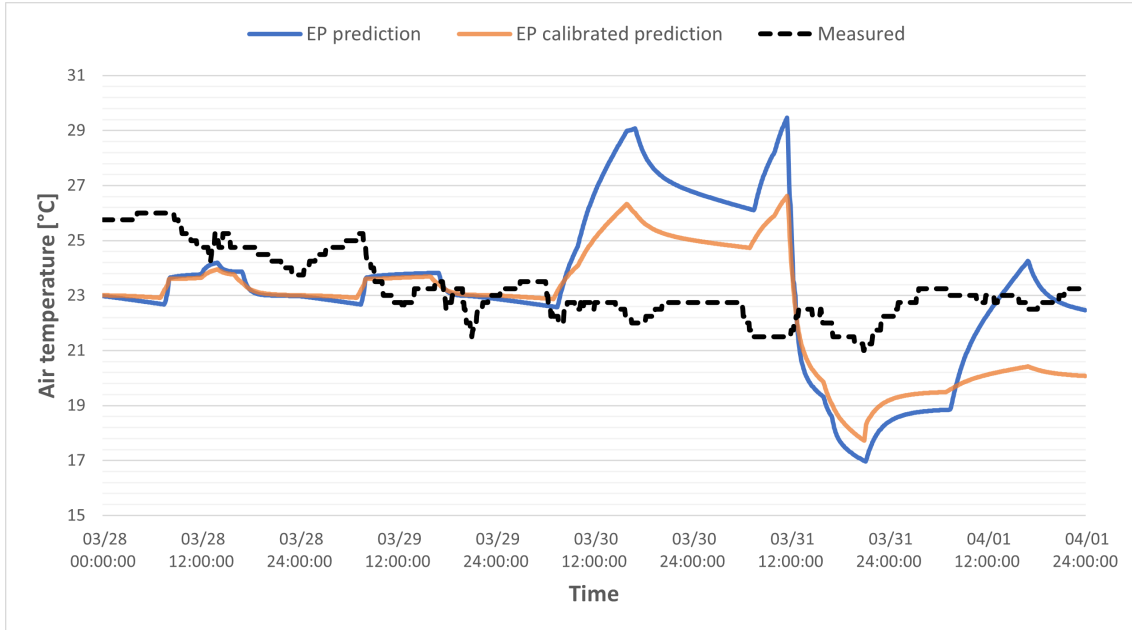


Figure 15: *Indoor air temperature with varying occupancy levels, 28.03-01.04.*

When looking at the predicted air temperatures, it is difficult to separate the two models until 08:00 in 30.03. From there, both models predict significant increases in the air temperature due to the increased occupancy levels, with the old- and calibrated models predicting an air temperature upwards of 29°C and 26.5°C , respectively. The fact that the calibrated model predicts a lower increase in the air temperature makes sense, given that the activity level specified is significantly lower for that model.

In the case of the measured air temperature, the increased occupancy from 30.03-31.03 does not seem to have an effect on the air temperature at all. This results in large differences between the predicted and measured air temperature. After the increased occupancy levels, the predicted air temperatures are significantly lower than the measured air temperature. Similar to the constant setpoint scenario, this was also a period where the setpoint temperature was reduced (in this case to 16°C).

Overall, neither of the EnergyPlus models seem able to predict the effect of occupancy levels well enough. It might even be the case that the predictions would be more accurate if the models omitted the occupancy levels. That being said, the calibrated model does produce more accurate results compared to the old model, as is evident from the figure above and the errors tabulated in table 18:

Table 18: *RMSE, MAE and MAPE of the calibrated and old model with varying occupancy levels.*

Model	RMSE ($^{\circ}\text{C}$)	MAE ($^{\circ}\text{C}$)	MAPE (%)
Previous model	3.034	2.404	10.5055
Model 1	2.388	2.089	9.069

Free-floating

The following figure shows the predicted and measured indoor drybulb air temperature from the old and new EnergyPlus models during free-floating operation (no ventilation):

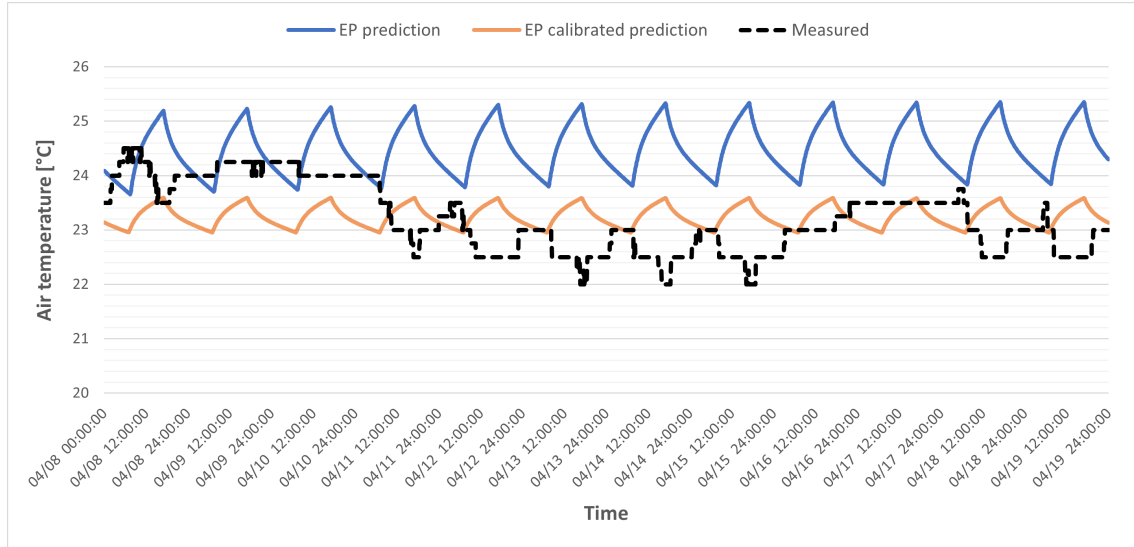


Figure 16: *Indoor air temperature at free-floating operation, 08.04-19.04.*

From figure 16 it can be seen that the EnergyPlus models predict relatively stable indoor air temperatures. For the old model, the air temperature oscillates between 23.6°C and 25.2°C , while for the calibrated model it oscillates between 23°C and 23.6°C . This oscillating behavior is most likely due to internal heat gains from the indoor equipment.

The measured air temperature is also pretty stable in this scenario, but does vary more than what is predicted by the EnergyPlus models. In general, both models predict the air temperature reasonably well during free-floating operation, but the lower air temperature predicted by the calibrated model is significantly more accurate. The computed errors are tabulated in table 19.

Table 19: *RMSE, MAE and MAPE of the calibrated and old model when simulating free-floating operation.*

Model	RMSE ($^{\circ}\text{C}$)	MAE ($^{\circ}\text{C}$)	MAPE (%)
Previous model	1.568	1.323	5.794
Model 1	0.724	0.606	1.288

Variable setpoint

Figure 17 depicts the predicted and measured indoor drybulb air temperature from the old and new EnergyPlus models at varying air supply setpoint temperatures:

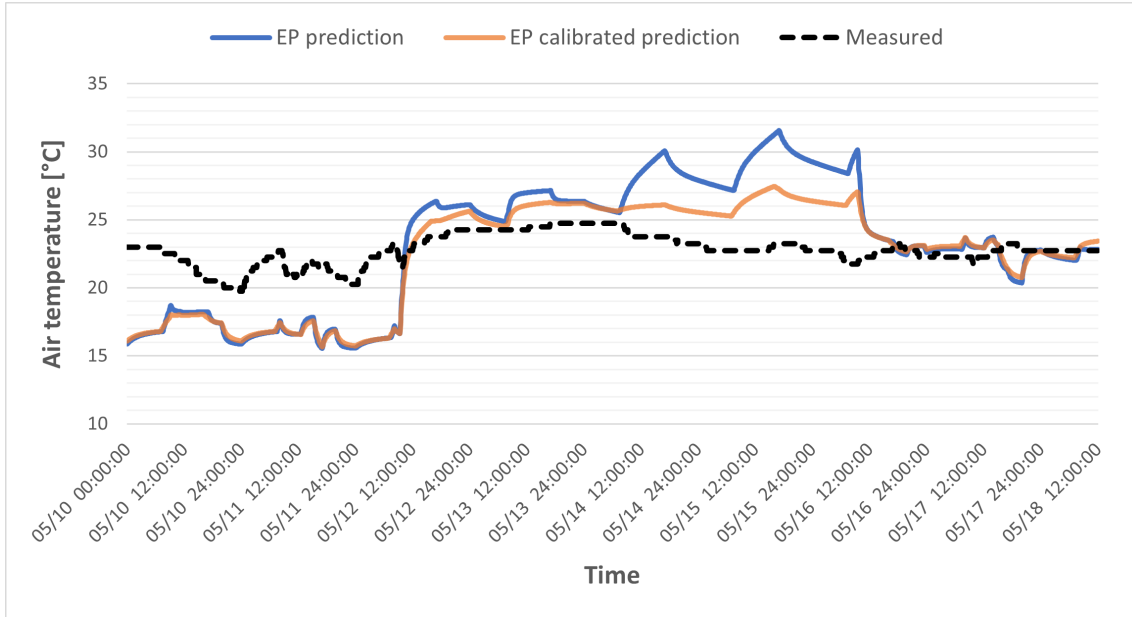


Figure 17: *Indoor air temperature with variable setpoint temperatures.*

From the figure above, it is clear that the predicted air temperatures are significantly more sensitive to changes in the setpoint temperature compared to the measured air temperature. In the first two days where the setpoint temperature is low ($14\text{--}17^{\circ}\text{C}$), the measured air temperature does gradually decrease from 23°C to 20°C , but not any further than that. The EnergyPlus models, however, predict an indoor air temperature that is slightly higher than the supply setpoint air temperature. The inverse is also the case when the setpoint temperature is high (26.5°C); the measured air temperature does increase over time, but not as much as the predicted air temperatures. However, the calibrated model is significantly more adept at predicting the air temperature during this period compared to the old model, whereas for the colder setpoint temperature period they are equally accurate.

Computed errors tabulated in table 20 do show that the calibrated model is more accurate than the old model, but the error is still significantly greater than in the other scenarios. The calibrated model does an ok job when the setpoint temperature is high, but the model shows a tendency to a colder air temperature than is actually the case. This can also be seen from the occupancy level and constant setpoint scenarios discussed before.

Table 20: *RMSE, MAE and MAPE of the calibrated and old model with varying setpoint temperatures.*

Model	RMSE ($^{\circ}\text{C}$)	MAE ($^{\circ}\text{C}$)	MAPE (%)
Previous model	4.125	3.404	15.079
Model 1	3.186	2.595	11.586

4.1.2 Year-long simulation

As described in section 3.2.7, a simulation was performed (with the calibrated model) over an entire year to produce a suitable dataset for training one of the black-box models. The simulated drybulb indoor air temperature over the entire simulation period is illustrated in the following figure:

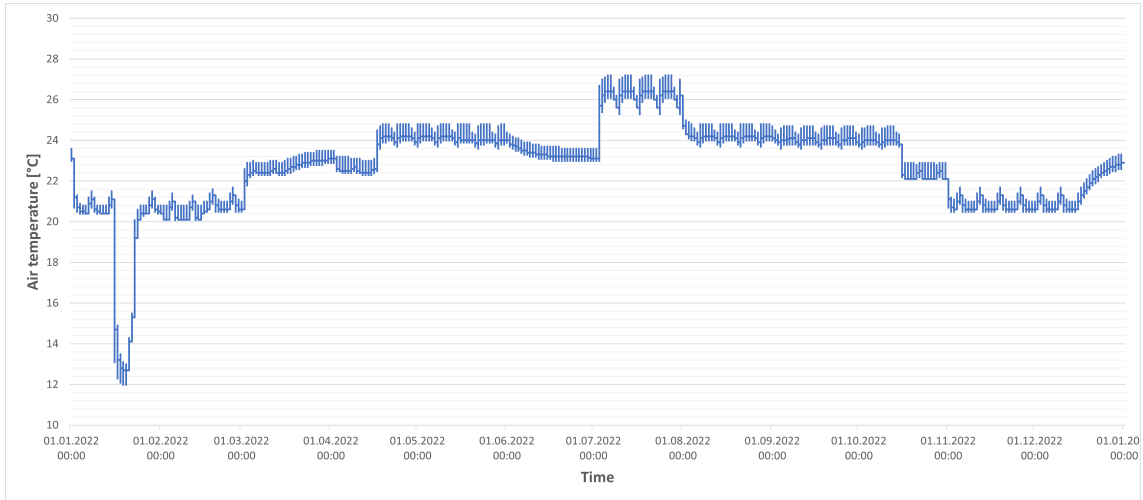


Figure 18: *Simulated indoor air temperature over an entire year.*

As shown above, the simulation produces a relatively stable air temperature throughout the simulation period, with the exception of the period 16.01-23.01 where the air temperature drops down to 12°C. This drop in air temperature is due to the supply setpoint air temperature being set to 6°C during this period. The reason why the air temperature doesn't drop all the way down to 12 degrees is due to heat gain through the walls, as the temperature outside the walls is higher than the indoor air temperature during this period. However, as has been discussed in the previous section, it is worth keeping in mind that the EnergyPlus model tends to overestimate how low the air temperature can get inside the room when the setpoint temperature is low.

For the rest of the simulation period, the setpoint temperature never drops below 20°C regardless of what the air flow rate is. In the periods where the AHU is shut off (i.e. 15.03-04.04, 01.06-01.07 and 15.12-31.12) gradually increase or decrease due to internal gains and heat losses through the walls.

4.2 Black-box models

In this section, 2 black-box models are trained with one dataset each. The first black-box model is trained by simulation data produced by a year-long simulation performed in EnergyPlus and the second black-box model is trained by the measured data of the actual system over a period of 79 days. From here, the former will be referred to as "EP-trained" and the latter "Measured-trained".

The section will start with evaluating the accuracy of the black-box models when trained and tested against their respective datasets. The effect of changing the dataset distribution, timestep length and testing periods will then be analyzed. Afterwards, features will be removed to test the importance of the different input parameters. Lastly, the EP-trained model's ability to predict the measured case will be tested.

As stated in section 3.3.4, training, validation and testing of the black-box models are done 3-5 times for each of these topics. When evaluating the black-box model accuracy (i.e. the first and last subsections here), the average RMSE, MAE and MAPE across the multiple predictions will be computed. For all other subsections, only the average RMSE will be shown. In addition, the EP-trained model will use a timestep length of 30 minutes, while the Measured-trained model will use 15 minutes for all predictions, except when the timestep length effect is tested.

4.2.1 Model training

Model accuracy

The average MAE, RMSE and MAPE of predictions made with the black-box models trained with EnergyPlus simulation data (EP-trained) and measured data (Measured-trained) are shown in figure 19:

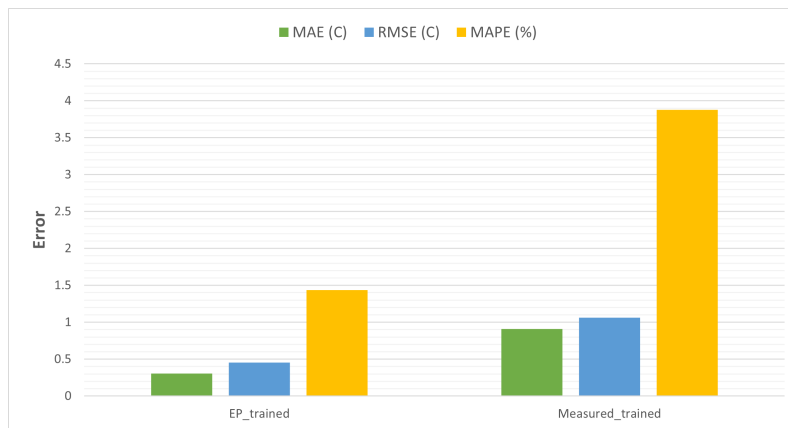


Figure 19: *Error of black-box models EP-trained and Measured-trained.*

With respect to all error metrics, the EP-trained model does a better job predicting the simulated air temperature than the Measured-trained model does predicting the measured data. In addition, the standard deviation of the indoor air temperature for the simulated and measured datasets were calculated to be $2.154^{\circ}C$ and $0.972^{\circ}C$, respectively. This makes the EP-trained model even more accurate in relative terms. It should also be noted that the EP-trained model is utilizing a longer timestep length.

Even though the EP-trained model is more accurate, the Measured-trained model has a satisfactory accuracy, as it has a percentage error of less than 5% and only a slightly higher root mean squared error than the air temperature standard deviation.

Dataset distributions

Figure 20 depicts the average RMSE for the black-box models EP-trained and Measured-trained with 4 different distributions of the datasets; 80/10/10, 60/20/20, 50/25/25, and 33/33/33 (training/validation/testing):

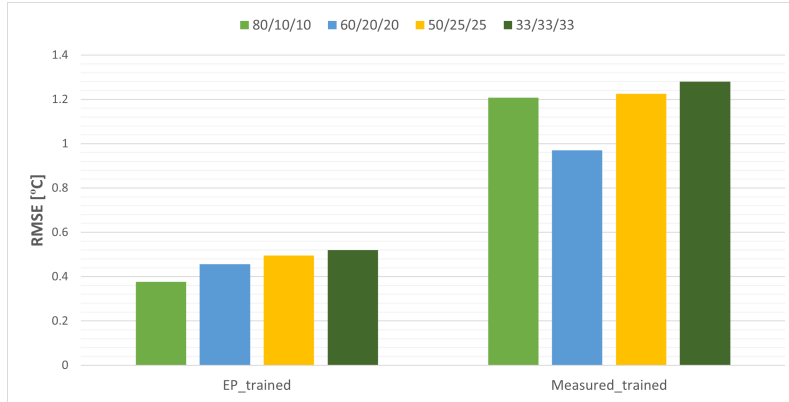


Figure 20: Average RMSE for both black-box models with varying dataset distributions.

For the EP-trained model, it can be seen that the RMSE increases as the training dataset gets smaller. The lowest RMSE of $0.377^{\circ}C$ occurs for a distribution of 80% training, 10% validation and 10% testing, a distribution of 33% for each of the steps yields an RMSE of $0.5192^{\circ}C$.

For the Measured-trained model, a 33% distribution also yields the highest RMSE. On the other hand, a distribution of 60% training, 20% validation and 20% testing gives a significantly better RMSE value compared to the other distributions.

Timestep length

The average RMSE for the black-box models EP-trained and Measured-trained with different timestep lengths are illustrated in figure 21:

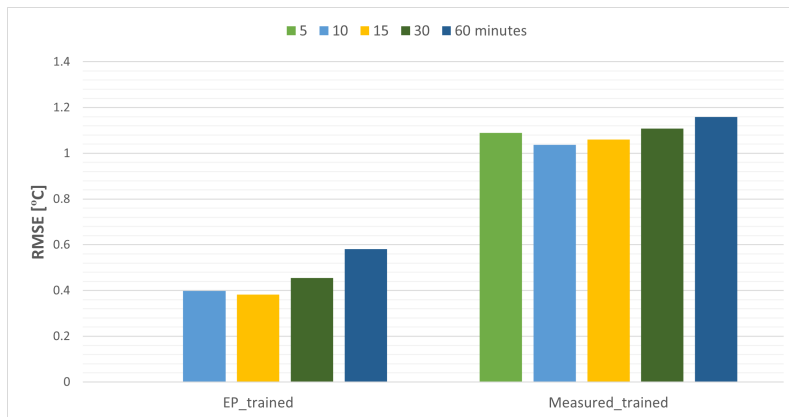


Figure 21: Average RMSE for both black-box models with varying timestep lengths.

Note that there was no testing done with a timestep length of 5 minutes for the EP-trained model due to it being too computationally expensive.

Both models produce their lowest RMSE-s when using a timestep length of 10-15 minutes. For the EP-trained model, the error increases slightly when increasing the timestep length to 30 minutes, and significantly more when using a timestep length of 60 minutes. For the Measured-trained model, however, the error increase isn't as large for a timestep length of 60 minutes. Moreover, the

Measured-trained model produces a similar RMSE when using a 5 minute and 30 minute timestep length.

The increase in error with timestep length in the case of the EP-trained model is to be expected. One would also expect the same for the Measured-trained model, but it may be the case that the indoor air temperature for that dataset is so stable that the timestep length does not have much of an effect on the model accuracy.

Testing periods

The average RMSE of the black-box models EP-trained and Measured-trained when using different testing time periods is shown in figure 22. See table 14 for a description of the testing time periods.

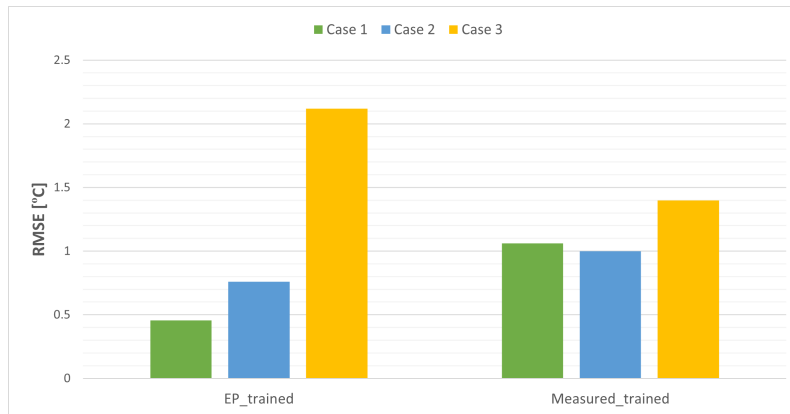


Figure 22: Average RMSE for both black-box models with different testing periods.

For the EP-trained model, case 1 (20.10-31.12) yields the lowest error compared to the other cases, with case 2 (8.08-20.10) yielding a 67% higher error and case 3 (1.01-15.03) the highest error by far. For the Measured-trained model, the highest error is also achieved with case 3 (04.03-21.03), but with not as large of a difference relative to the other cases. Furthermore, case 1 (07.05-22.05) has a slightly higher error than case 2 (20.04-07.05).

The most likely reason why case 3 for the EP-trained model yields such a high error is because that is the testing period where the simulated air temperature is much lower than in the rest of the simulation, and is therefore a testing dataset that is quite different from the training and validation datasets in this case.

4.2.2 Feature relevance

As noted earlier in this chapter, the relevance of each parameter used in the black-box models were tested by training a model with one feature missing as input data for each of the features. The average RMSE found when removing each of these features for both black-box models as well as when using all the parameters can be seen in the following figure:

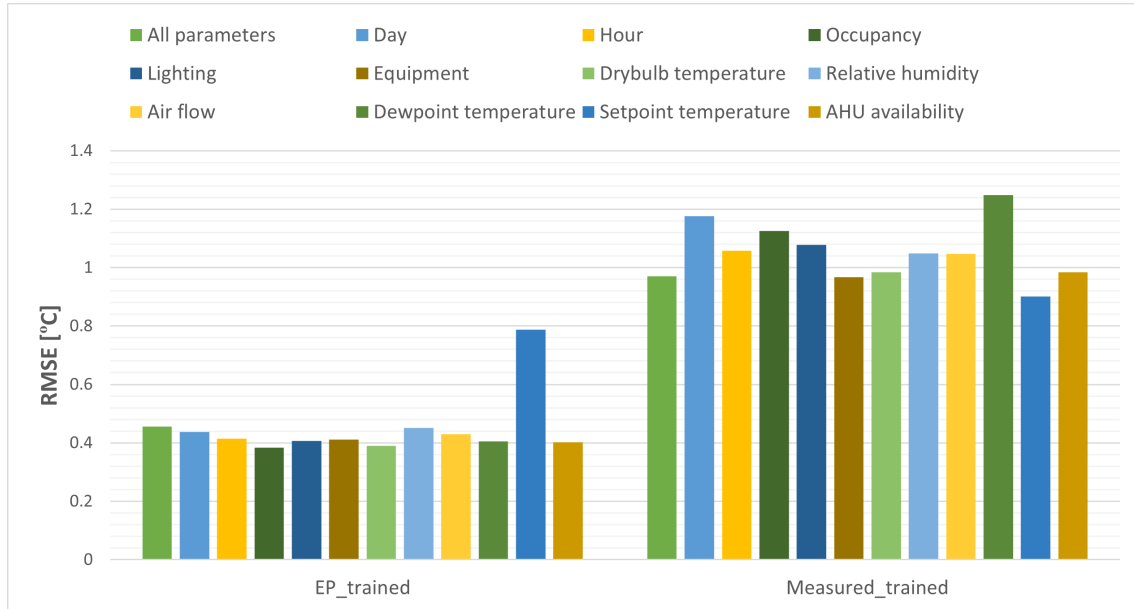


Figure 23: Average RMSE for both black-box models trained with 1 missing feature.

Looking at the RMSE-s for the EP-trained model, there is a clear increase in the error when removing the supply setpoint air temperature, meaning that model's accuracy is highly dependant on having access to the supply setpoint air temperature compared to ther parameters. For the Measured-trained model, it is not as evident what the most relevant parameters are. The features that yielded the highest errors when removed here were the outdoor dewpoint temperature, day of the week and occupancy level.

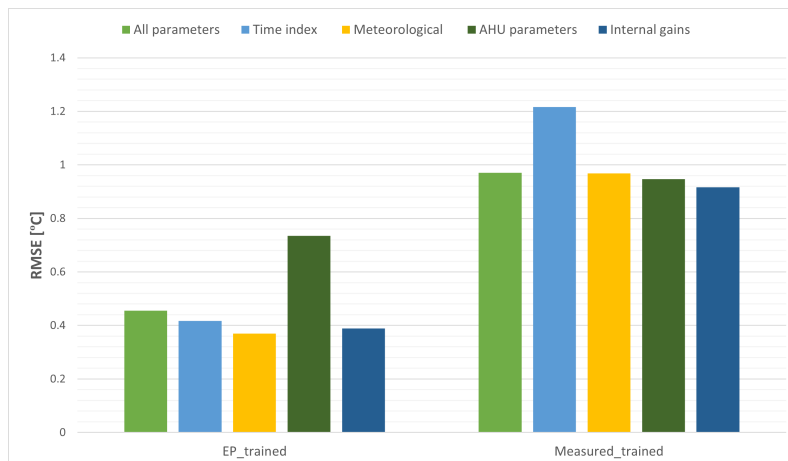


Figure 24: Average RMSE for both black-box models trained with missing categories of data.

To gain further insight into how relevant the different types of input parameters are, the input parameters were placed into categories as shown in table 13. Here, the same procedure as before was performed, but with removing categories of data instead of individual features. The average

RMSE for each of these categories are shown in figure 24.

For the EP-trained model, removing the AHU parameters gave the largest increase to the error. This makes sense, since the supply setpoint air temperature is in the AHU parameters category. There is also a category of data for the Measured-trained model that seems to be most relevant; the time index parameters. So despite the hour and day of the week not initially being the most relevant features based on figure 23, removing both of them have a large impact on the model accuracy.

One reason why the supply setpoint air temperature is of high relevance for the EP-trained model could be that there is a strong correlation between the output temperature and the setpoint temperature. This is clearly highlighted in the figure 25, where the indoor air temperature and setpoint temperature closely follow each other throughout the year-long simulation:

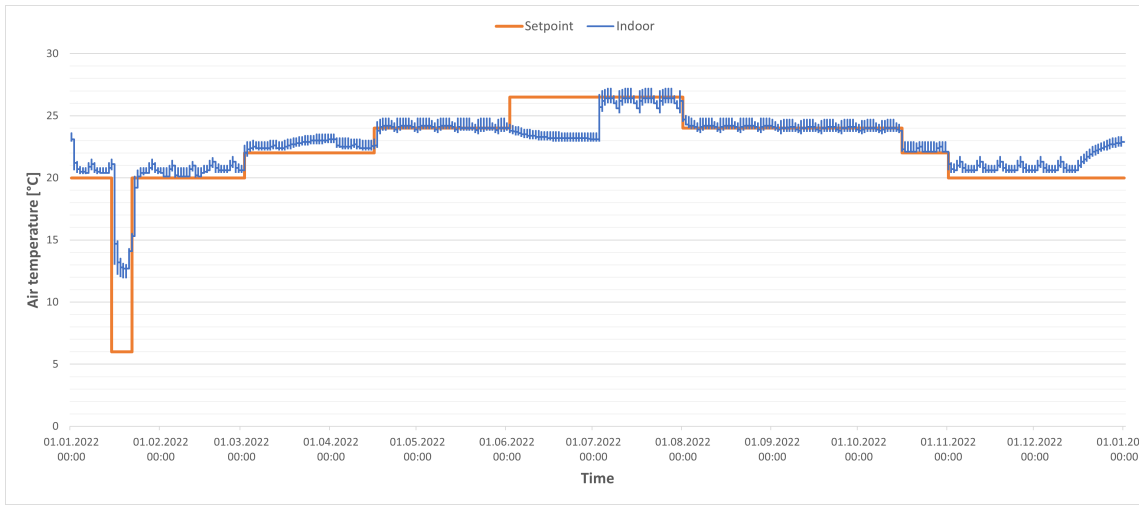


Figure 25: Simulated indoor- and supply setpoint air temperature over an entire year.

In the measured dataset, this is clearly not the case as shown below:

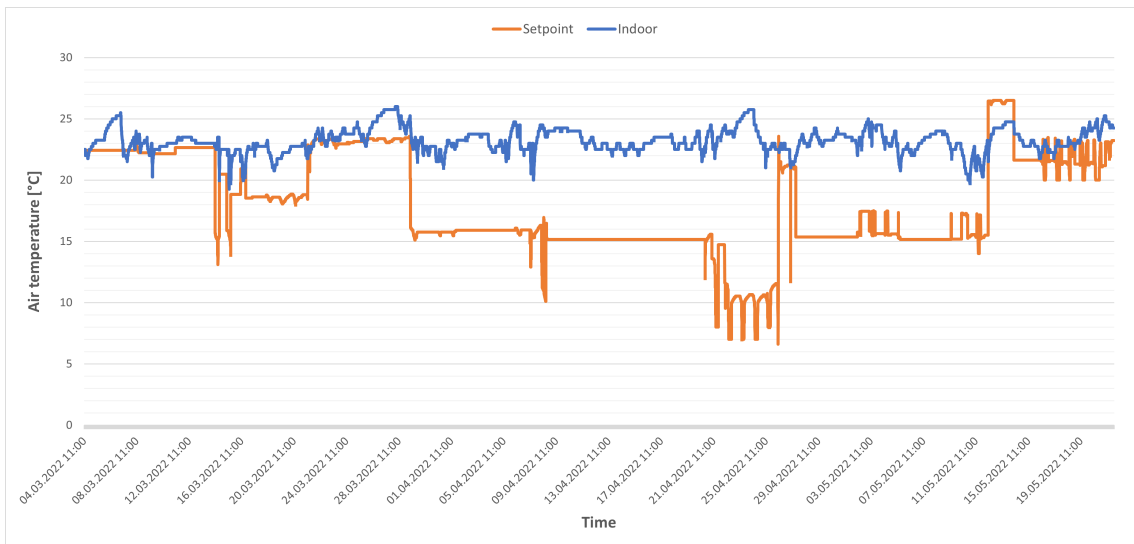


Figure 26: Measured indoor- and supply setpoint air temperature 04.03-22.05.

4.2.3 Measured data prediction

In this subsection, the EP-trained black-box model (i.e. a black-box model trained with simulation data from EnergyPlus) is used to predict the measured indoor air temperature. Note that the black-box model is still using a 60/20/20 distribution of the simulation data when training the model. In addition, the mathematical EnergyPlus model will also try to predict the measured indoor air temperature.

The average RMSE, MAE and MAPE for these predictions and the Measured-trained black-box model (same errors as in figure 19) are illustrated in figure 27:

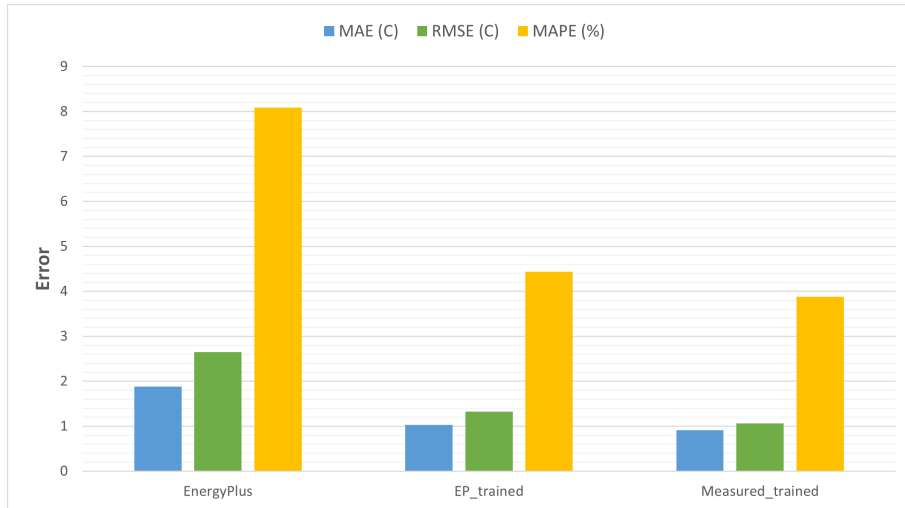


Figure 27: Error of prediction of measured indoor air temperature with EnergyPlus model and EP-trained black-box model.

Of the 3 models, the Measured-trained black-box model is the most accurate in predicting the measured indoor air temperature. This is to be expected as this model was trained with the measured data, while the EP-trained model was not. Figure 28 below shows the measured indoor air temperature and the temperature predicted by the EnergyPlus model. This shows multiple points in time where EnergyPlus predicts a very low air temperature due to decreases in the setpoint temperature, while the measured temperature remains relatively stable - resulting in large errors.

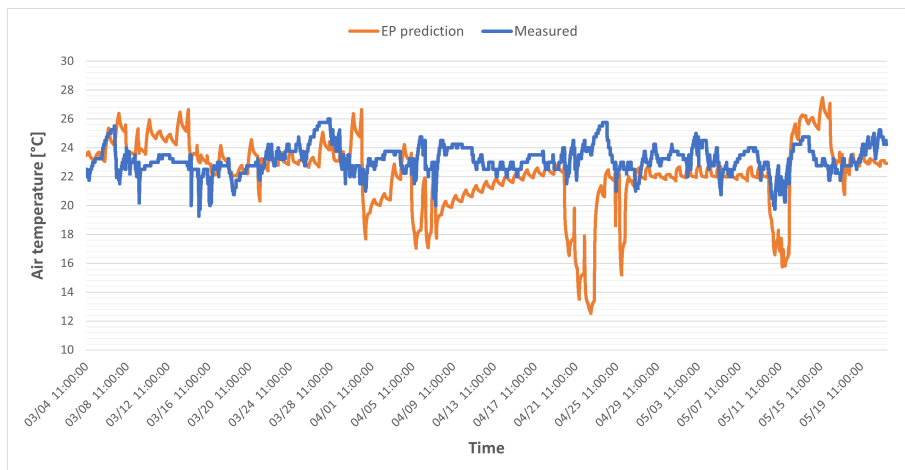


Figure 28: Measured and predicted indoor air temperature with EnergyPlus model 04.03-22.05.

5 Discussion

5.1 Mathematical model

5.1.1 Calibration results

When comparing the error metrics of the mathematical model before and after the "brute-force" calibration procedure, it is clear that the calibrated model has a significantly improved prediction accuracy. By using the calibrated model, the root mean squared error, mean absolute error and mean absolute percentage error are reduced by at least 25% each. That being said, there are still large errors in the predicted indoor air temperature even after the calibration process. Throughout the 4 testing scenarios, it can be observed that the predicted indoor air temperature is much lower than the measured air temperature whenever the setpoint temperature is low, with little difference between the original and calibrated model. The calibrated model does tend to predict a less steep temperature curve whenever the setpoint temperature drops, but the difference is marginal. For higher setpoint temperatures, however, the calibrated model does a better job of modelling the thermal inertia of the system as shown in the variable setpoint scenario.

When it comes to the internal heat gains, the mathematical model does seem to overestimate how much of an impact occupancy has on the indoor air temperature. During the variable internal gains scenario, the measured indoor air temperature remains relatively stable no matter the occupancy level, whereas the mathematical model predicts significant increases in the indoor air temperature. There are multiple factors that may have contributed to this difference in indoor air temperature. The first is that the occupancy level is assumed to be constant for several hours, when in reality it is likely that the number of occupants was continually changing during the lab sessions in the variable internal gains scenario. It is also possible that the door to the laboratory room was open for long periods, which would counteract the heat gain from the occupants since the air outside the laboratory room is slightly cooler than inside. The model also assumes a constant activity level all the time, which is unlikely to actually be the case. There being issues with modelling the internal heat gain from occupants is not particularly surprising though, as the literature has continuously stated that it is hard to accurately record the occupancy in buildings.

For the scenarios where the setpoint is stable or the air handling unit is shut off, the calibrated model does predict the indoor air temperature quite well. The model is found to be most accurate for predicting the indoor air temperature during the free-floating scenario, in which the RMSE, MAE and MAPE was found to be $0.724^{\circ}C$, $0.606^{\circ}C$ and 1.288%, respectively. This is also the scenario with the largest improvement in the model accuracy after calibration, since the old model predicted an indoor air temperature that was too high. What is interesting about this scenario is that it is the only period where there is a greater variance in the measured air temperature than the predicted air temperature. Why this is the case is not entirely clear, but it could be that the room was used over the easter break, as activities in the laboratory room was not recorded during this period. One other reason may have been that the temperature varied more than what was initially found from the previous paper and that it also should have been recorded for this project.

In the constant setpoint scenario, there is a pattern in the difference between the predicted and measured air temperature that can be observed. When the air handler is on, the predicted air temperature matches well with the measured air temperature, and then they start to slowly separate overnight when the air handler is shut off. From this it could be theorized that the mathematical model overestimates how much heat is lost through the walls when the air handler is shut off overnight. Overall, though, the mathematical model does a good job of predicting the air temperature when there is little change in the setpoint temperature and occupancy levels.

5.1.2 Year-long simulation

One concern from the previous paper was that the mathematical model was unable to produce a realistic indoor air temperature when the air flow rate was high. Fortunately, this issue was alleviated by changing how the supply air flow rate was calculated in EnergyPlus and by reducing the air infiltration. In the previous paper the simulated air flow rate was much higher than originally assumed as there was large amount of unwanted outdoor air that was added to the air flow provided by the fans in the air handler. By fixing this issue, the air handler in the mathematical model is much more capable of conditioning the supply air throughout the year. This is exemplified in the first month of the year-long simulation, where the system maintains a relatively stable air temperature of 20-21°C (except from the period where the setpoint temperature is very low) despite the outdoor air reaching temperature levels below -10°C and the air flow rate being over 2000m³/h.

As was discussed earlier, the mathematical model has shown to be more sensitive to lower setpoint temperatures than the actual system, and it is highlighted here as well during the time period 16.01 - 23.01. How the indoor air temperature would change in actuality in this period is hard to say, as there hasn't been a scenario where the setpoint temperature was as low as 6°C for an entire week. However, from the measurements gathered in this project it doesn't seem likely that the indoor air temperature would change as quickly and as far down as depicted by the year-long simulation.

5.2 Black-box models

5.2.1 Model accuracy

When training and testing the black-box model against the simulation data produced by the mathematical model, the black-box model does a good job of accurately predicting the simulated indoor air temperature. Comparing the predicted and simulated air temperature, the RMSE, MAE and MAPE were found to be 0.455°C, 0.306°C and 1.436%. Keeping in mind that the standard deviation of the simulated air temperature was 2.154°C, this clearly shows that the black-box model is able to predict the simulated indoor air temperature accurately, even without common input parameters such as the heat demand.

When training and testing the black-box model against the measured data, the black-box model yields less accurate results compared with the simulated data, but it is still accurate to a satisfactory degree. The RMSE value was found to only be slightly higher than the standard deviation of the measured air temperature. This is impressive considering that in the measured case, the model uses only 47 days worth of data for training, while in the simulated case, the model uses 219 days worth of data.

5.2.2 Dataset distribution

From the dataset distribution test results, when training the black-box model against the simulation data, we see a trend in the error increasing as the portion of the dataset distributed to the training phase gets smaller. When training the black-box model against measured data, the difference in error between a 80/10/10, 50/25/25 and 33/33/33 distribution is relatively small, whereas a 60/20/20 distribution yields by far the most accurate results.

For both datasets, using a distribution of 33/33/33 yields the highest error. This is to be expected, as most studies allocate at least 50% of the data to training of black-box models for building

simulations. The reason why a 60/20/20 distribution yields better predictions when training against measured data compared with other distributions is not entirely clear at first glance. It could have something to do with the last 16 days in the measured dataset being similar to the first 47 days, but this would imply that using another testing period (discussed later on in this section) would result in much higher error - which was not the case from further tests made.

5.2.3 Timestep length

From figure 21 we can see that the black-box model trained against measured data is less sensitive to changes made to the timestep length in the training data. This would imply that the measured indoor air temperature is more stable than the year-long simulated air temperature, which is indeed the case seeing as the standard deviation of the measured and simulated temperatures are $0.972^{\circ}C$ and $2.154^{\circ}C$, respectively. It can also be observed that a shorter timestep length yields a higher prediction accuracy, but at the cost of computational expensiveness. These results are to be expected, since more training data generally leads to more accurate models.

5.2.4 Testing periods

Figure 22 shows that utilizing different testing periods affects the black-box model trained against simulation data much more than the model trained with measured data. This makes sense since the measured indoor air temperature is much more stable than the simulated air temperature. When the testing period is from 1.01 to 15.03 for the simulated data, the error more than triples, exceeding even the error of the model trained with measured data. This is most likely because that is the specific period where simulated air temperature is very different from the rest of the simulation. Overall, it seems like the black-box model has the best test results when the training and testing datasets are the most similar to each other.

5.2.5 Feature relevance

When looking at the effect removing each feature has on the model accuracy compared to when all the features are used, it can be seen that the model trained with simulation data only decreases in accuracy when removing the supply setpoint air temperature. For the model trained with measured data, the opposite is almost the case; removing any of the features will result in roughly the same or higher error. What seems to have happened is that the model trained with simulation data has become overly reliant on the setpoint temperature when predicting the indoor air temperature, resulting in the other features being interpreted as noise. As discussed in section 4.2.2, this is most likely due to there being a strong correlation between the setpoint temperature and indoor air temperature as illustrated in 25. For the measured dataset, there is no such correlation as seen in figure 26, resulting in the model trained with measured data not viewing the setpoint temperature as an important feature.

When removing categories of data, the time-index stands out for the model trained with measured data, indicating that features related to the time-index are the most relevant for this model, while features related to internal gains have the lowest impact. This is not surprising, as the time-index is often cited as one of the most important input parameters for black-box models.

5.2.6 Measured data prediction

In section 4.2.3, the black-box model trained with simulation data has a more accurate prediction of the measured indoor air temperature than expected - only slightly lower than the black-box model trained with measured data. The RMSE, MAE and MAPE for these two cases were $1.2093^{\circ}C$ and $1.060^{\circ}C$, respectively. This is surprising, given that the feature relevance tests revealed that this model was highly reliant on the setpoint temperature after being trained with the simulation data, while there is little correlation between the measured indoor air temperature and setpoint temperature. One would also expect the black-box model's accuracy to be limited to how well the EnergyPlus model can predict the indoor air temperature, but that is evidently not the case when comparing the error metrics.

It could be the case that the simulation-data-trained model generally predicts a more stable indoor air temperature than the EnergyPlus model, which would then fit better with the measured data. This makes sense when thinking about the simulated indoor air temperature; there is only a 1-week period in a 52-week dataset that exhibits any extreme temperature levels. If the year-long simulation contained more periods with low setpoint temperatures, the black-box model may then have been more vulnerable to lower setpoint temperatures.

Overall, though, this section showed that a black-box model trained with a simulated dataset can be used to predict the actual indoor air temperature accurately enough. An advantage to using this kind of model is that it can be improved by making modifications to the mathematical model that produces the training dataset, and this does not necessarily require the need for further measurements to be made.

5.3 Possible model improvements

The feature relevance test revealed several parameters that could be removed in order to improve the accuracy of the black-box models. More specifically, it would be interesting to repeat the removal of each feature without the setpoint temperature as a feature to begin with. By doing this, the model trained with simulation data will most likely be less accurate, but it would be easier to evaluate the relevance of the other features for this model. A feature selection method such as variable ranking could then be used to find the best combination of features.

It could also be beneficial to perform a more detailed internal gains experiment in order to properly identify the impact occupation has on air heat balance - and thereby improving the accuracy of the mathematical model.

6 Conclusion

In this thesis, the objective was to develop a mathematical model and a data-driven model capable of accurately predicting the behaviour of an air handling unit connected to a laboratory room with simulation data used as the training data for the data-driven model. In addition, modelling approaches for the functionality of air handling units and ventilation systems were investigated in a literature review.

There are three types of HVAC system modelling approaches: mathematical (white-box) models, hybrid (grey-box) models and data-driven (black-box) models. Mathematical models tend to be complicated to make and computationally expensive, but can provide great accuracy levels and transparency in their working principles. The most applicable black-box modelling approaches for HVAC systems are artificial neural networks and support vector machines. Hybrid models use a combination of mathematical- and data-driven modelling principles and can also provide good accuracy levels in HVAC system behavior prediction.

HVAC system modelling has multiple applications. It can be used to predict indoor thermal environment, building energy usage and even for detecting and correcting faults in HVAC systems. Mathematical models are best suited for designing HVAC systems, while data-driven models can be used for fast and reliable prediction of existing systems.

A mathematical model of testing facility and its air handling unit was developed through the use of EnergyPlus. Unknown parameters in internal mass, air infiltration and internal gains from occupants and equipment were determined and calibrated against four scenarios where the air handling unit was operating under different conditions over a 79-day period. After the calibration procedure, it was found that the mathematical model can accurately predict the indoor air temperature when the supply setpoint air temperature and occupancy level is stable. However, the model was not able to accurately model the thermal inertia of the testing facility, resulting in a greater sensitivity to the setpoint temperature. It was found that further tests needs to be performed to properly locate the impact of the internal heat gains in the testing facility.

For the data-driven model, a long short-term memory recurrent neural network was used to predict the indoor air temperature. Two datasets were created for training and testing, one generated by the mathematical model over a 1-year simulation and the other by measurements taken inside the facility from 04.03 to 22.05. Two data-driven models were then trained and tested against each of these datasets.

From the tests performed with these data-driven models, it was found that the model trained with the simulated dataset was more accurate with predicting the simulated indoor air temperature, than the other model was with predicting the measured indoor air temperature. In the case of the former, the accuracy level was very high and for the latter the accuracy the accuracy level was good. Testing different dataset distributions, testing time periods and timestep lengths revealed that the measured dataset was much more stable than the simulated dataset.

The relevance of each of the input features was also evaluated. From this test, it was found that the strong correlation between the indoor air temperature and the setpoint temperature for the simulated dataset resulted in the model trained with that data being too reliant on the setpoint temperature when predicting the indoor air temperature. For the measured data case, time-index related parameters were found to be most relevant in predicting the measured indoor air temperature. Despite the model trained with simulated data being overly reliant on the setpoint temperature, it was able to predict the measured indoor air temperature to a satisfactory degree of accuracy. It was concluded that the accuracy of this model could be further increased by adjusting the mathematical model which does not necessarily require additional measurements.

Bibliography

- [1] F. Poupart, *Modelling of air handling unit*, 2021.
- [2] F. sambandet. (2017). ‘Fns bærekraftsmål’, [Online]. Available: <https://www.fn.no/Om-FN/FNs-baerekraftsmaal>. (Lastet ned: 29.04.2020).
- [3] P. Höppe and I. Martinac, ‘Indoor climate and air quality, review of current and future topics in the field of isb study group 10’, *International Journal of Biometeorology* 42, pp. 1–7, 1998.
- [4] H. Ajimotokan, L. Oloyede and M. Ismail, ‘Influence of indoor environment on health and productivity’, *New York Science Journal*, vol. 2, no. 4, pp. 46–49, 2009.
- [5] B. D. Hardin, B. J. Kelman and A. Saxon, ‘Adverse human health effects associated with molds in the indoor environment.’, *Journal of Occupational and Environmental medicine*, vol. 45, no. 5, pp. 470–478, 2003.
- [6] A. Myhrvold, E. Olsen and O. Lauridsen, ‘Indoor environment in schools–pupils health and performance in regard to co2 concentrations’, *Indoor Air*, vol. 96, no. 4, pp. 369–371, 1996.
- [7] Folkehelseinstituttet, *Karbondioksid (co2) og inneklima*, 2015. [Online]. Available: <https://www.fhi.no/ml/miljo/inneklima/artikler-inneklima-og-helseplager/karbondioksid-co2-og-inneklima/>.
- [8] D. for byggkvalitet, *Byggteknisk forskrift (tek17) med veiledning*. [Online]. Available: <https://dibk.no/regelverk/byggteknisk-forskrift-tek17/13/i/13-1/>.
- [9] N. Nassif, ‘The impact of air filter pressure drop on the performance of typical air-conditioning systems’, in *Building Simulation*, Springer, vol. 5, 2012, pp. 345–350.
- [10] C. Kanaoka and S. Hiragi, ‘Pressure drop of air filter with dust load’, *Journal of aerosol science*, vol. 21, no. 1, pp. 127–137, 1990.
- [11] Folkehelseinstituttet, *Ventilasjonsanlegg og inneklima*, 2013. [Online]. Available: <https://www.fhi.no/ml/miljo/inneklima/artikler-inneklima-og-helseplager/ventilasjonsanlegg-og-inneklima/>.
- [12] J. Pillai and R. Desai, ‘Dehumidification strategies and their applicability (c105)(p)’, in *2018 Building Performance Analysis Conference and SimBuild*, ASHRAE, 2018.
- [13] A. Kusiak, G. Xu and Z. Zhang, ‘Minimization of energy consumption in hvac systems with data-driven models and an interior-point method’, *Energy Conversion and Management*, vol. 85, pp. 146–153, 2014.
- [14] A. Afram and F. Janabi-Sharifi, ‘Review of modeling methods for hvac systems’, *Applied Thermal Engineering*, vol. 67, no. 1-2, pp. 507–519, 2014.
- [15] V. Rajkumar and P. SrihariDatta, ‘Optimization of ahu control strategy’, *IJITR [online]*, vol. 1, no. 2, 2013.
- [16] K.-S. Tang, K. F. Man, G. Chen and S. Kwong, ‘An optimal fuzzy pid controller’, *IEEE transactions on industrial electronics*, vol. 48, no. 4, pp. 757–765, 2001.
- [17] V. Vakiloroyaya, B. Samali, A. Fakhar and K. Pishghadam, ‘A review of different strategies for hvac energy saving’, *Energy conversion and management*, vol. 77, pp. 738–754, 2014.
- [18] V. Harish and A. Kumar, ‘A review on modeling and simulation of building energy systems’, *Renewable and sustainable energy reviews*, vol. 56, pp. 1272–1292, 2016.
- [19] Z. Afroz, G. Shafullah, T. Urmee and G. Higgins, ‘Modeling techniques used in building hvac control systems: A review’, *Renewable and sustainable energy reviews*, vol. 83, pp. 64–84, 2018.
- [20] Y. Sun, F. Haghighat and B. C. Fung, ‘A review of the-state-of-the-art in data-driven approaches for building energy prediction’, *Energy and Buildings*, vol. 221, p. 110 022, 2020.

-
- [21] R. Z. Homod, 'Review on the hvac system modeling types and the shortcomings of their application', *Journal of Energy*, vol. 2013, 2013.
- [22] R. American Society of Heating and A.-C. Engineers, *ASHRAE Handbook Fundamentals*. 2017.
- [23] A. Foucquier, S. Robert, F. Suard, L. Stéphan and A. Jay, 'State of the art in building modelling and energy performances prediction: A review', *Renewable and Sustainable Energy Reviews*, vol. 23, pp. 272–288, 2013.
- [24] R. Yusof, H. Selamat and S. Attaran, 'Short review on hvac components, mathematical model of hvac system and different pid controllers', *International Review of Automatic Control*, vol. 7, no. 3, pp. 263–270, 2014.
- [25] M. Trčka and J. L. Hensen, 'Overview of hvac system simulation', *Automation in Construction*, vol. 19, no. 2, pp. 93–99, 2010.
- [26] S. Klein and W. Beckman, *TRNSYS 16: A transient system simulation program: mathematical reference*. 2007.
- [27] D. P. Solomatine, 'Data-driven modelling: Paradigm, methods, experiences', in *Proc. 5th international conference on hydroinformatics*, 2002, pp. 1–5.
- [28] D. P. Solomatine and A. Ostfeld, 'Data-driven modelling: Some past experiences and new approaches', *Journal of hydroinformatics*, vol. 10, no. 1, pp. 3–22, 2008.
- [29] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo and P. Chatellier, 'Modeling and forecasting building energy consumption: A review of data-driven techniques', *Sustainable Cities and Society*, vol. 48, p. 101533, 2019, ISSN: 2210-6707. DOI: <https://doi.org/10.1016/j.scs.2019.101533>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2210670718323862>.
- [30] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han and X. Zhao, 'A review of data-driven approaches for prediction and classification of building energy consumption', *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1027–1047, 2018.
- [31] A. Kumar, D.-S. Kim, H. Omidvarborna, M. Yarlagadda, S. K. Kuppili and N. Sawtarie, 'Experimental modeling of nox and pm generation from combustion of various biodiesel blends for urban transport buses', 2016.
- [32] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah and R. Saidur, 'A review on applications of ann and svm for building electrical energy consumption forecasting', *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 102–109, 2014.
- [33] Z. Zhai, Q. Chen, P. Haves and J. H. Klems, 'On approaches to couple energy simulation and computational fluid dynamics programs', *Building and Environment*, vol. 37, no. 8-9, pp. 857–864, 2002.
- [34] Y.-W. Wang, W.-J. Cai, Y.-C. Soh, S.-J. Li, L. Lu and L. Xie, 'A simplified modeling of cooling coils for control and optimization of hvac systems', *Energy Conversion and Management*, vol. 45, no. 18-19, pp. 2915–2930, 2004.
- [35] S. A. Kalogirou, 'Use of trnsys for modelling and simulation of a hybrid pv–thermal solar system for cyprus', *Renewable energy*, vol. 23, no. 2, pp. 247–260, 2001.
- [36] G. Ciulla and A. D'Amico, 'Building energy performance forecasting: A multiple linear regression approach', *Applied Energy*, vol. 253, p. 113500, 2019.
- [37] X. Cui, B. Mohan, M. Islam and K. Chua, 'Modelling and performance evaluation of an air handling unit for an air treatment system with regulated outdoor-air fraction', *Energy Procedia*, vol. 105, pp. 4718–4723, 2017.
- [38] C. Ghiaus, A. Chicinas and C. Inard, 'Grey-box identification of air-handling unit elements', *Control Engineering Practice*, vol. 15, no. 4, pp. 421–433, 2007.
-

-
- [39] G. Hong and B. S. Kim, ‘Development of a data-driven predictive model of supply air temperature in an air-handling unit for conserving energy’, *Energies*, vol. 11, no. 2, p. 407, 2018.
- [40] A. Razban, A. Khatib, D. Goodman and J. Chen, ‘Modelling of air handling unit subsystem in a commercial building’, *Thermal Science and Engineering Progress*, vol. 11, pp. 231–238, 2019.
- [41] X.-B. Yang, X.-Q. Jin, Z.-M. Du, Y.-H. Zhu and Y.-B. Guo, ‘A hybrid model-based fault detection strategy for air handling unit sensors’, *Energy and buildings*, vol. 57, pp. 132–143, 2013.
- [42] V. Corrado and E. Fabrizio, ‘Steady-state and dynamic codes, critical review, advantages and disadvantages, accuracy, and reliability’, *Handbook of energy efficiency in buildings*, pp. 263–294, 2019.
- [43] M. G. Lawrence, ‘The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications’, *Bulletin of the American Meteorological Society*, vol. 86, no. 2, pp. 225–234, 2005. DOI: 10.1175/BAMS-86-2-225. [Online]. Available: <https://journals.ametsoc.org/view/journals/bams/86/2/bams-86-2-225.xml>.

Appendix

A Cold water loops

The following figures depict the fluid loops used for the cooling coils in the EnergyPlus model.

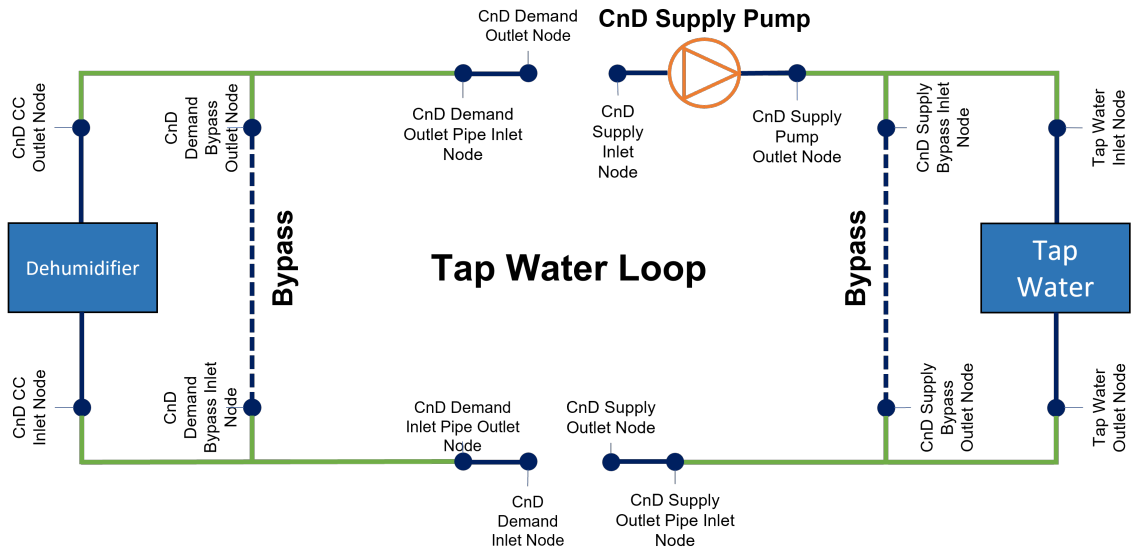


Figure 29: Node diagram of fluid loop providing tap water to the dehumidifier.

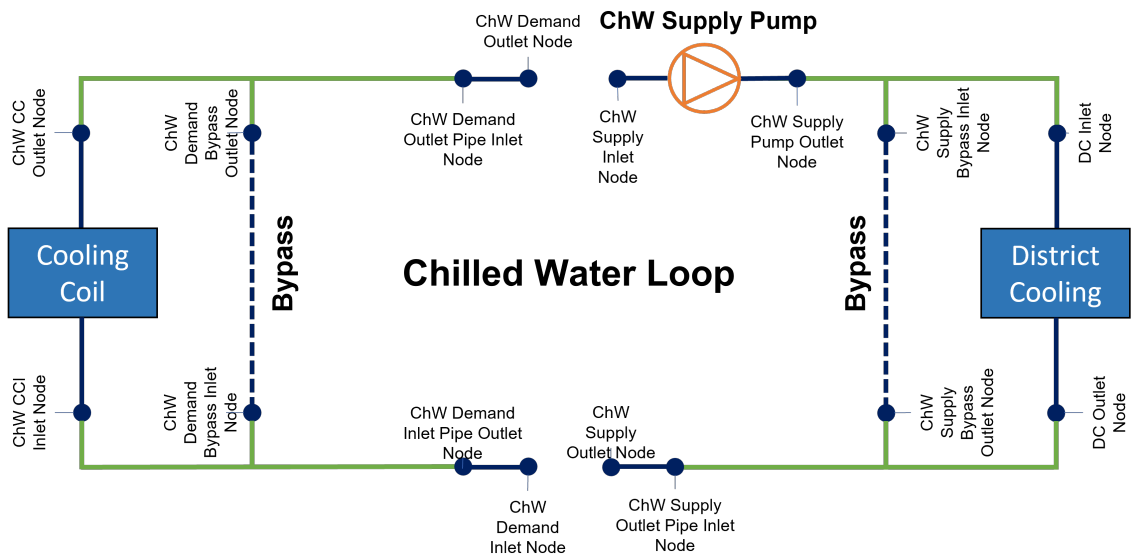

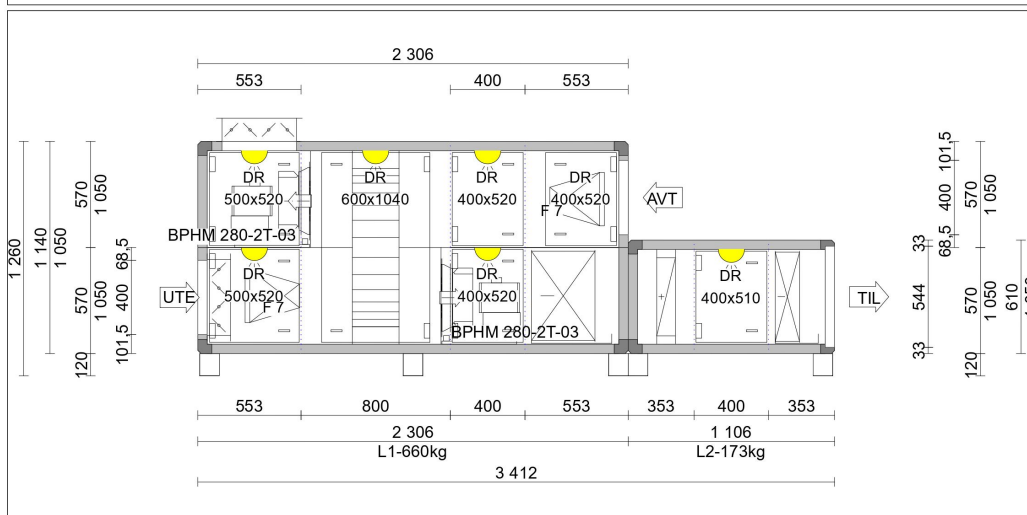


Figure 30: Node diagram of fluid loop providing chilled water to the second cooling coil.

B Air handling unit component datasheet

 <p>Covent Johan Jakobsen</p> <p>Røyslandsvegen 36 NO 4387 BJERKREIM Tel.: +47 51 45 96 00 www.covent.no post@covent.no</p>	Fra dato 08.09.2019 Utskrift dato 16.12.2019 Prosjekt NTNU - Undervisning	CoventAir v09/11
	Tilbud 101.49805 Systemnummer 02 System mark 36.xxx 2000m³/h AC Antall 1	Kunde GK Inneklima AS Kontakt Arild Frostad Revisjon 1 16.12.2019 Endret Vann temp på siste kj batt til 9/14

Serie	CK	Havnivå [m]	0
Utførelse	Stablet enhet	Spesifikk vekt [kg/m ³]	1,20
Aggregattype	Innendørs enhet	Spesifikk vitteffekt SFPv [w/(m ³ /s)]	1 671 SFP4
<i>Energimerke klasse designet for våte forhold</i>		Spesifikk vitteffekt SFPe [w/(m ³ /s)]	2 023 SFP4
		Spesifikk vitteffekt internt [w/(m ³ /s)]	737
		Virkningsgrad [%]	84,80



Installert strøm (hoved) [A]	11	Spenning [V]	3x400
		Total vekt [kg]	~835



Ansvarlig: **Johan Jakobsen**
 Dato: **16.12.2019**

Kunde: **GK Inneklima AS**
 Gate: **Baard Iversens veg 7**
 By: **Trondheim**
 Bygg prosjekt: **NTNU - Undervisning**
 Konstruksjons-gate:
 Konstruksjons-by:

CoventAirv09/11

Prosjektnr.: **101.49805**
 Systemnummer: **02**
 Tegning-ID: **02**
 System mark: **36.xxx 2000m3/h AC**
 Revisjon: **1**
 Side: **2 / 7**

Tilluft

Enhetsdefinisjon	NRVU/BVU	Aggregatthus: Energieffektivitetsklasse			
Aggregatstørrelse	CKAA 02				
Luftmengde [m ³ /h]	2 000	Lengde [mm]	3 412	Isolasjon	Mineralull 50 mm
Ekst. trykk [Pa]	200	Bredde [mm]	1 050	Innvendig	Galvanisert stål 0,9 mm
Tot. trykk [Pa]	642	Høyde [mm]	1 140	Utvendig	Galvanisert stål 0,9 mm
Lufthastighet [m/s]	1,12	Vekt [kg]	~833	Profiler	Galvanisert stål
Klasse DIN EN 13053	V1			Flenser	Galvanisert stål
Termisk overføring	T3	Aggregatthus lekkasje ved -400 Pa	L3R	Mekanisk stabilitet	D3M
Termisk brokkasse	TB4	Aggregatthus lekkasje ved +400 Pa	L2R	Filter bypasslekkasje	F9
Modell boks	Rockwool 50mm				

Filter	L1	Tilluft	553 mm	1,21 m2	74 kg	132 Pa
Produsent	Deltrian	Filterlengde [mm]	292,0			
Type	RPV85-SC	Filteroverflaten [m2]	9,00			
Klasse	F7	Celler stk x størrelse	1 x Fil-RPV85-5SC-C 592,0 x 490,0			
Start trykkfall [Pa]	52					
Slutt trykkfall [Pa]	200					
Design dP [Pa]	126					
Luftmengde [m ³ /h]	2 000	1,92 m/s				
Filterhåndtering	Uttrekk fra side					
Filter energieffektivitet	794 kWh/a					
Dør med hengsel og enkelt håndtak			Dimensjoner [mm] 500,0 x 520,0			
<u>Spjeld:</u>			Dimensjoner mm 500,0 x 410,0 x 125,0			
Tetthetsklasse (EN1751)	3	Luftmengde [m ³ /h]	2 000	Ramme	Aluminium	
Utløses av	Utløser	Lufthastighet [m/s]	2,71	Blad	Aluminium	
Antall spjeldarmer	1	Trykkfall [Pa]	6	Type	Covent	
Vridningsmoment [Nm]	2					
<u>Lampe</u>	Takarmatur LED-pære	Nominell data	230 V	5 W	0,03 A	IP44
Ja						



Ansvarlig: **Johan Jakobsen**
Dato: **16.12.2019**

Kunde: **GK Inneklima AS**
Gate: **Baard Iversens veg 7**
By: **Trondheim**
Bygg prosjekt: **NTNU - Undervisning**
Konstruksjons-gate:
Konstruksjons-by:

CoventAirv09/11
Prosjektnr.: **101.49805**
Systemnummer: **02**
Tegning-ID: **02**
System mark: **36.xxx 2000m3/h AC**
Revisjon: **1**
Side: **3 / 7**

Rotor i aggregatthus	L1	Tilluft	800 mm	3,5 m2	166 kg	113 Pa
Type	PL-852-A16-250-W			Energieffektivitetsklasse H1		
Varmemodus						
Tillførsel [m³/h]	2 000	Dp [Pa]	113			
Inngående [°C]	-19,0	Befuk. [%]	90	Φ t 84,9 %		
Temp ut [°C]	15,8	Befuk. [%]	34	Φ x 72,2 %		
Avkast [m³/h]	2 000	Dp [Pa]	118			
Inngående [°C]	22,0	Befuk. [%]	30			
Temp ut [°C]	-12,7	Befuk. [%]	99			
Tot. gjenv.kapa. [KW]			28,6			
Følb. gjennvinning kapasitet [KW]			23,3			
Drivdata						
R		Reimskive		26		
Nominell effekt [KW]		0,120		Nominell strøm [A]		0,97
						Nominell spenning [V] 3x230
Leveres med EMOTRON turtallsregulator på varmeveksler						
Dør med hengsel og enkelt håndtak				Dimensjoner [mm] 600,0 x 1040,0		
Lampe Takarmatur LED-pære						
Nominell data			230 V	5 W	0,03 A	IP44
Ja						

Kammervifte	L1	Tilluft	400 mm	0,88 m2	74 kg	Pa
Vifte	BPHM 280-2T-03			Motor HMA3 80 2-2		
Leverandør	Covent			Beskyttelse IP55		
Luftmengde [m³/h]	2 000			Isoleringsklasse F		
Eksternt trykk [Pa]	200			Effekt [KW] 1,10 0,00 0,00		
Inner PD AHU [Pa]	388			Hastighet [1/min] 2 910		
Statisk trykk [Pa]	605			Strøm +-5% [A] 2,4		
Total trykk [Pa]	642			Spenning 3x400 V / 50 Hz		
Hastighet [1/min]	2 679					
Virkningsgrad %	76,18					
Vifteakseffekt [KW]	0,467					
K-faktor	79					
Arbeidspunkt [Hz] 44,7						
Maksimal frekvens [Hz] 59,4						
Opptatt effekt [KW] 0,700						
Spesifikk vifteeffekt [W/(m3/s)] 807 SFP2						
Dør med hengsel og enkelt håndtak				Dimensjoner [mm] 400,0 x 520,0		
Åpning L				Dimensjoner [mm] 360,0 x 368,0		
Lampe Takarmatur LED-pære						
Nominell data			230 V	5 W	0,03 A	IP44
Ja						



Ansvarlig: **Johan Jakobsen**
 Dato: **16.12.2019**

Kunde: **GK Inneklima AS**
 Gate : **Baard Iversens veg 7**
 By: **Trondheim**
 Bygg prosjekt: **NTNU - Undervisning**
 Konstruksjons-gate:
 Kontruksjons-by:

CoventAirv09/11

Prosjektnr.: **101.49805**
 Systemnummer: **02**
 Tegning-ID: **02**
 System mark: **36.xxx 2000m3/h AC**
 Revisjon: **1**
 Side: **4 / 7**

Kjølebatteri	L1	Tilluft	553 mm	1,21 m2	99 kg	117 Pa
Luftmengde [m³/h]	2 000		Medium		Vann	
Lufthastighet [m/s]	1,48		Volum [l]		18,10	
Luft inn [°C]	24,0	Fuktighet [%] 80	Med. strømning [L/s]		0,98	
Luft ut [°C]	11,0	Fuktighet [%] 100	Med. hastighet [m/s]		0,90	
Total kapasitet [KW]	20,6		Med. inn [°C]		7,0	
Følb. kapasitet [KW]	8,8		Med. ut [°C]		12,0	
Luftrykkfall [Pa]	117/63	(vått/tørt)	Med. trykkfall [KPa]		29	
			SHR		0,43	
Cu-Al-FeZn P3012AR 10R-16T-780A-2.5pa 10C 1" (.11- .35- 1.			<u>Materialer:</u>			
Rør	10		Rør		Kobber	
Kretser	10		Finner		Aluminium	
Finneavstand [mm]	2,5		Samlestokk		Kobber	
Tilkobling inn	1 0/0"		Rammer		Galvanisert stål	
Tilkobling ut	1 0/0"		Finnebeskyttelse		-	
Utbygd batterikasse [mm]	Nei					
Dreneringskar		Kvalitet	Aluminium	Dreneringstilkobling	1 1/4"	

Varmebatteri	L2	Tilluft	353 mm	1,17 m2	60 kg	9 Pa
Luftmengde [m³/h]	2 000		Medium		Vann	
Lufthastighet [m/s]	1,62		Volum [l]		1,90	
Luft inn [°C]	11,0		Med. strømning [L/s]		0,09	
Luft ut [°C]	22,0		Med. hastighet [m/s]		0,83	
Kapasitet [KW]	7,4		Med. inn [°C]		60,0	
Luftrykkfall [Pa]	9		Med. ut [°C]		40,0	
			Med. trykkfall [KPa]		13	
Cu-Al-FeZn P3012AC 1R-14T-815A-2.0pa 1C 1/2" (.11- .35- 1.5			<u>Materialer:</u>			
Rør	1		Rør		Kobber	
Kretser	1		Finner		Aluminium	
Finneavstand [mm]	2,0		Samlestokk		Kobber	
Tilkobling inn	0 1/2"		Rammer		Galvanisert stål	
	0 1/2"		Finnebeskyttelse		-	

Tomdel	L2	Tilluft	400 mm	1,33 m2	44 kg	Pa
Dør med hengsel og enkelt håndtak			Dimensjoner [mm]		400,0 x 510,0	
<u>Lampe</u>	ETLED HIDDEN	Nominell data	12 V	3 W	0,35 A	IP44
		Ja				



Ansvarlig: **Johan Jakobsen**
Dato: **16.12.2019**

Kunde: **GK Inneklima AS**
Gate: **Baard Iversens veg 7**
By: **Trondheim**
Bygg prosjekt: **NTNU - Undervisning**
Konstruksjons-gate:
Konstruksjons-by:

CoventAirv09/11

Prosjektnr.: **101.49805**
Systemnummer: **02**
Tegning-ID: **02**
System mark: **36.xxx 2000m3/h AC**
Revisjon: **1**
Side: **5 / 7**

Kjølebatteri	L2	Tilluft	353 mm	1,17 m2	69 kg	17 Pa
Luftmengde [m³/h]	2 000		Medium		Vann	
Lufthastighet [m/s]	1,47		Volum [l]		5,80	
Luft inn [°C]	22,5	Fuktighet [%] 50	Med. strømming [L/s]		0,18	
Luft ut [°C]	17,0	Fuktighet [%] 71	Med. hastighet [m/s]		0,55	
Total kapasitet [KW]	3,8		Med. inn [°C]		9,0	
Følb. kapasitet [KW]	3,8		Med. ut [°C]		14,0	
Luftrykkfall [Pa]	17/17	(vått/tørt)	Med. trykkfall [KPa]		14	
			SHR		1	
Cu-Al-FeZn P3012AR 3R-16T-785A-3.0pa 3C 3/4" (.11- .35- 1.5			<u>Materialer:</u>			
Rør	3		Rør		Kobber	
Kretser	3		Finner		Aluminium	
Finneavstand [mm]	3,0		Samlestokk		Kobber	
Tilkobling inn	0 3/4"		Rammer		Galvanisert stål	
Tilkobling ut	0 3/4"		Finnebeskyttelse		-	
Utbygd batterikasse [mm]	Nei					
Åpning	E		Dimensjoner [mm]		984,0 x 544,0	
Dreneringskar		Kvalitet Aluminium		Dreneringstilkobling	1 1/4"	

Støyberegning									
Frkv. Hz	Lydeffekt [dB]								Sum [dB(A)]
	63	125	250	500	1000	2000	4000	8000	
Uteluft	72,3	60,4	68,5	61,3	49,7	56,4	48,2	40,1	63,9
Tilluft	72,3	57,4	73,5	70,3	63,7	58,4	43,2	32,1	70,6
Aggregathus	58,3	57,4	53,5	47,3	43,7	42,4	32,2	20,1	50,8
Tolerance +5 dB for 125 Hz, +3 db for 250-8000 Hz									

Avkast					
Enhetsdefinisjon	NRVU/BVU	Aggregathus: Energieffektivitetsklasse			
Aggregatstørrelse	CKAA 02				
Luftmengde [m³/h]	2 000	Lengde [mm]	2 306	Isolasjon	Mineralull 50 mm
Ekst. trykk [Pa]	200	Bredde [mm]	1 050	Innvendig	Galvanisert stål 0,9 mm
Tot. trykk [Pa]	504	Høyde [mm]	1 140	Utvendig	Galvanisert stål 0,9 mm
Lufthastighet [m/s]	1,12	Vekt [kg]	~0	Profil	Galvanisert stål
Klasse DIN EN 13053	V1			Flenser	Galvanisert stål
Termisk overføring	T3	Aggregathus lekkasje ved -400 Pa	L3R	Mekanisk stabilitet	D3M
Termisk brokkasse	TB4	Aggregathus lekkasje ved +400 Pa	L2R	Filter bypasslekkasje	F9
Modell boks	Rockwool 50mm				



Ansvarlig: **Johan Jakobsen**
Dato: **16.12.2019**

Kunde: **GK Inneklima AS**
Gate: **Baard Iversens veg 7**
By: **Trondheim**
Bygg prosjekt: **NTNU - Undervisning**
Konstruksjons-gate:
Konstruksjons-by:

CoventAirv09/11
Prosjektnr.: **101.49805**
Systemnummer: **02**
Tegning-ID: **02**
System mark: **36.xxx 2000m3/h AC**
Revisjon: **1**
Side: **6 / 7**

Filter	L1	Avka	553 mm	1,79 m2	70 kg	126 Pa
Produsent	Deltrian		Filterlengde [mm]	292,0		
Type	RPV85-SC		Filteroverflaten [m2]	9,00		
Klasse	F7		Celler stk x størrels 1 x Fil-RPV85-5SC-C 592,0 x 490,0			
Start trykkfall [Pa]	52					
Slutt trykkfall [Pa]	200					
Design dP [Pa]	126					
Luftmengde [m³/h]	2 000	1,92 m/s				
Filterhåndtering	Uttrekk fra side					
Filter energieffektivitet	794 kWh/a					
Dør med hengsel og enkelt håndtak			Dimensjoner [mm]	400,0 x 520,0		
Åpning	G		Dimensjoner [mm]	500,0 x 400,0		

Tomdel	L1	Avka	400 mm	1,3 m2	36 kg	Pa
Dør med hengsel og enkelt håndtak			Dimensjoner [mm]	400,0 x 520,0		
<u>Lampe</u>	Takarmatur LED-pære	Nominell data	230 V	5 W	0,03 A	IP44
Ja						

Rotor i aggregatthus	L1	Avka	800 mm	3,5 m2	166 kg	113 Pa
----------------------	----	------	--------	--------	--------	--------

Kammervifte	L1	Avka	553 mm	2,19 m2	141 kg	6 Pa
Vifte	BPHM 280-2T-03		Motor	HMA3 80 2-2		
Leverandør	Covent		Beskyttelse	IP55		
Luftmengde [m³/h]	2 000		Isoleringsklasse	F		
Eksternt trykk [Pa]	200		Effekt [KW]	1,10	0,00	0,00
Inner PD AHU [Pa]	250		Hastighet [1/min]	2 910		
Statisk trykk [Pa]	467		Strøm +-5% [A]	2,4		
Total trykk [Pa]	504		Spenning	3x400 V / 50 Hz		
Hastighet [1/min]	2 469					
Virkningsgrad %	77,87					
Vifteakseffekt [KW]	0,358					
K-faktor	79					
Arbeidspunkt [Hz]	41,2					
Maksimal frekvens [Hz]	59,8					
Opptatt effekt [KW]	0,570					
Spesifikk vifteeffekt [W/(m3/s)]	656 SFP1					
Dør med hengsel og enkelt håndtak			Dimensjoner [mm]	500,0 x 520,0		
<u>Spjeld:</u>			Dimensjoner mm	500,0 x 400,0 x 125,0		
Tetthetsklasse (EN1751)	3					
Utløses av	Utløser	Luftmengde [m³/h]	2 000	Ramme	Aluminium	
Antall spjeldarmer	1	Lufthastighet [m/s]	2,78	Blad	Aluminium	
Vridningsmoment [Nm]	2	Trykkfall [Pa]	6	Type	Covent	
Åpning	L		Dimensjoner [mm]	360,0 x 368,0		
<u>Lampe</u>	Takarmatur LED-pære	Nominell data	230 V	5 W	0,03 A	IP44
Ja						



Ansvarlig: **Johan Jakobsen**

Dato: **16.12.2019**

Kunde: **GK Inneklima AS**
Gate : **Baard Iversens veg 7**
By: **Trondheim**
Bygg prosjekt: **NTNU - Undervisning**
Konstruksjons-gate:
Konstruksjons-by:

CoventAirv09/11
Prosjektnr.: **101.49805**
Systemnummer: **02**
Tegning-ID: **02**
System mark: **36.xxx 2000m3/h AC**
Revisjon: **1**
Side: **7 / 7**

Støyberegning

Frkv. Hz	Lydeffekt [dB]								Sum [dB(A)]
	63	125	250	500	1000	2000	4000	8000	
Avtrekk	70,9	58,0	66,5	57,9	48,0	54,6	46,7	39,0	61,7
Avkast	71,9	62,0	67,5	69,9	69,0	66,6	57,7	48,0	73,1
Aggregathus	56,9	55,0	51,5	43,9	42,0	40,6	30,7	19,0	48,6

Tolerance +5 dB for 125 Hz, +3 db for 250-8000 Hz

**1 Sett Labber Standard feet 120,0 mm Fast høyde
FOR BRUK I UNDERVISNING - GJENOMSIKTIGE DØRER I PLEXI GLASS**

<u>Leveringsseksjoner</u>	Nr.	Bredde	Høyde	Lengde	Vekt
	1	1 050	1 140	2 306	660
	2	1 050	610	1 106	173

C Laboratory plan drawing

This section contains the second floor plan of the Energy and Indoor Environment Laboratory at NTNU. The smaller testing facility that was modelled in this project is located on the upper left corner - room C247A. The corresponding air handling unit can be found below this room, in C247B.

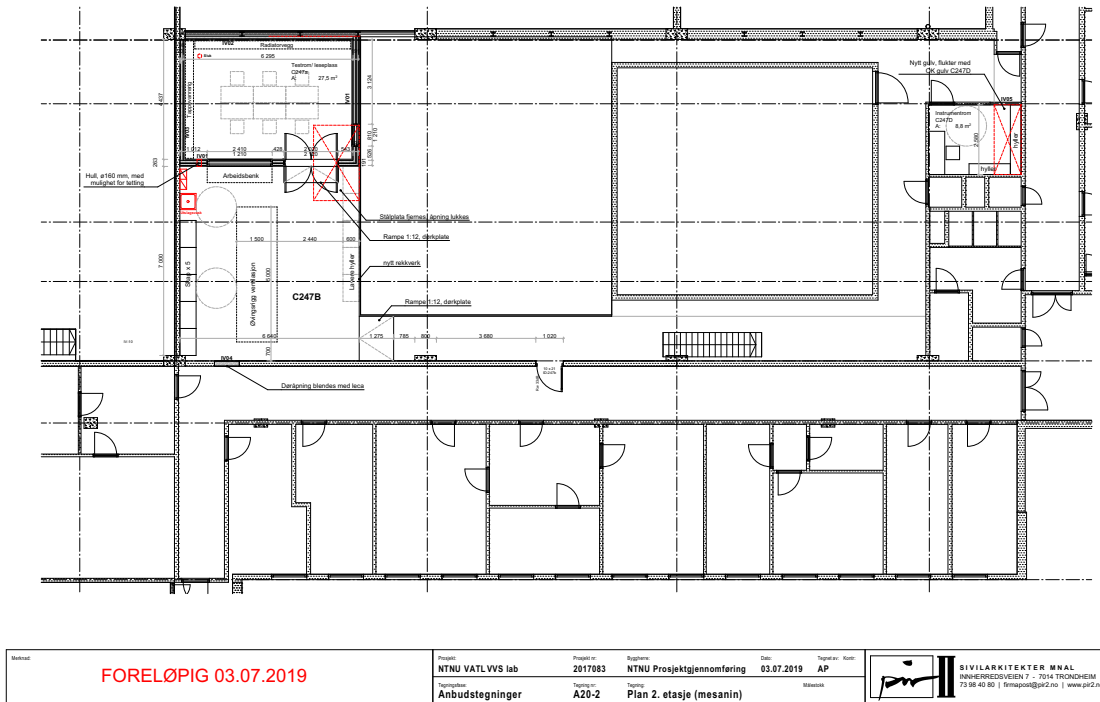
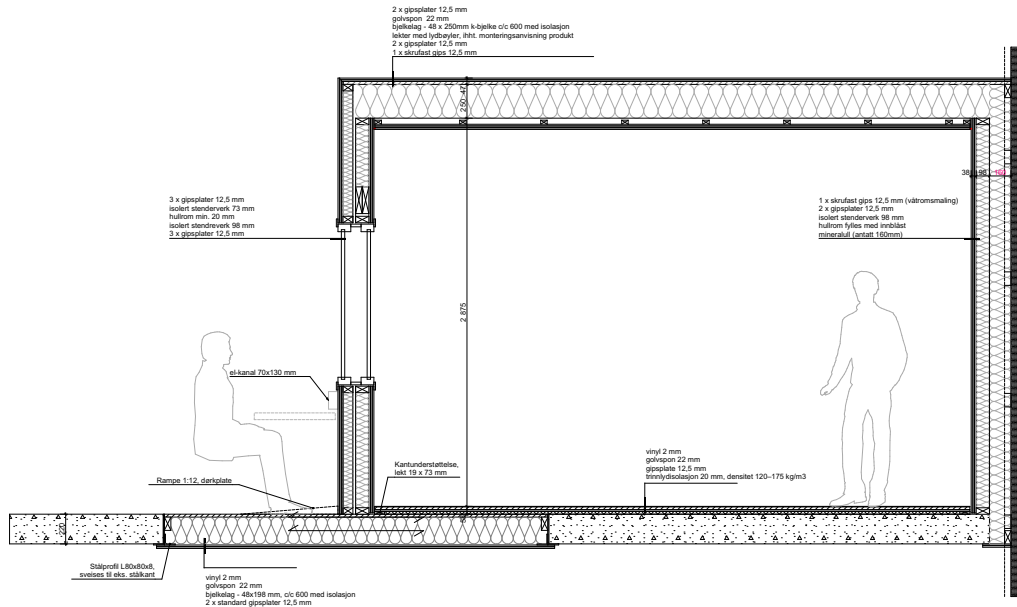


Figure 31: Plan drawing of NTNU's Energy and Indoor Environment Laboratory, second floor.

D Room section drawing

This section contains the section drawing of the modelled laboratory room, including a description of the components in the surface constructions.




Status	FORELØPIG 03.07.2019					 SIVILARKITEKTER MNAL INGENIØRSKOLEN 7 • 7014 TRONDHEIM 73 98 40 80 firmapost@pa2.no www.pa2.no
	Prosjekt	Prosjekt nr.	Begrunnelse	Dato	Tegnet av	
	NTNU VATLVS lab	2017083	NTNU Prosjektgjennomføring	03.07.2019	AP	
	Anbudstegninger	A60-01	Tegning	Snitt testrom		

Figure 32: Section drawing of the modelled laboratory room inside NTNU's Energy and Indoor Environment Laboratory with construction components.

E Calibration code

This section contains the code used for calibrating the mathematical model in EnergyPlus.

```
import subprocess
import pandas as pd
import sklearn
import numpy as np
from sklearn.metrics import mean_squared_error
import csv

#generating parameter combinations
cap = [5,10,15,20,25,30]
inf = [0,0.1,0.2,0.3,0.4,0.5]
equip = [150,200,250,300]
activ = [80,100,120,140]
intm = [1,25,50,75,100]
calibration_cases = ['Case number', 'Capacitance', 'Infiltration', 'Electric Equipment',
'Activity level', 'Internal Mass']
count = 0
for i in cap:
    for j in inf:
        for k in equip:
            for l in activ:
                for m in intm:
                    count += 1
                    row = [count,i,j,k,l,m]
                    calibration_cases = np.vstack([calibration_cases, row])
pd.DataFrame(calibration_cases).to_csv('calibration_cases.csv')

#generating idf files
with open('calibration_cases.csv') as f:
    reader = csv.reader(f)
    cases = []
    for row in reader:
        cases.append(row)
    num_rows = len(cases)
    for i in range(1,num_rows):
        case_num = cases[i][0]
        cap = cases[i][1]      # capacitance
        inf = cases[i][2]     # infiltration
        eet = cases[i][3]     # electric equipment
        als = cases[i][4]     # activity level
        intm = cases[i][5]    # internal mass area
        template_name = 'C:/Users/Frede/OneDrive - NTNU/Prosjektoppgave/EnergyPlus
model/Calibration_template.idf'
        with open(template_name, 'r') as template:
            idf = template.read()
            idf = idf.replace('$CAP@@", str(cap))
            idf = idf.replace('$INF@@", str(inf))
            idf = idf.replace('$EET@@", str(eet))
```

```

idf = idf.replace('$ALS@', str(als))
idf = idf.replace('$INM@', str(intm))
with open('idf_files/case_'+str(case_num)+'.idf','w') as new_file:
    new_file.writelines(idf)

#running simulations for each idf file
actual = pd.read_excel('Manually_calibrated_and_actual_Tdry.xlsx')
T_actual = actual['T_actual']
EP_path = 'F:\EnergyPlus\energyplus'
weather_file = r''C:\Users\Frede\OneDrive - NTNU\Prosjektoppgave\EP calibration files
\Trd_5min_new.epw'''
errors = ['Case','RMSE','MSE']
for i in range(1,2881):
    EP_file = 'C:/Users/Frede/OneDrive - NTNU/Prosjektoppgave/EP calibration
files/idf_files/case_'+str(i)+'.idf'
    output_directory = r''F:\EnergyPlus\EP_output'''
    text_input = (f'"{EP_path}" '
                  + '--readvars '
                  + f'--output-directory "{output_directory}" '
                  + f'--weather "{weather_file}" '
                  + f'"{EP_file}"'
                  )
    result=subprocess.run(text_input,
                           capture_output=True
                           )
    predicted = pd.read_csv('F:/EnergyPlus/EP_output/eplusout.csv')
    predicted.drop(range(2448,2592),inplace=True)
    T_predicted = predicted['ZONE:Zone Air Temperature [C](TimeStep)']
    rms = mean_squared_error(T_actual,T_predicted,squared=False)
    mse = mean_squared_error(T_actual,T_predicted,squared=True)
    row = [i,rms,mse]
    errors = np.vstack([errors,row])

#get 3 cases with the lowest error
print(errors[:,1].argsort()[:3])

```

F Data-driven model

This section contains the code for the data-driven model used in this thesis project, developed by Gaurav Chaudhary, PhD candidate.

```
from pandas import read_csv
from google.colab import drive
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
import tensorflow as tf
import os
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from math import sqrt

drive.mount('/content/drive')
importedcsv = r'/content/drive/My Drive/LSTM_DATA/EP_30_min.csv'

#hyperparameters
timestep = 2 #steps per hour
n_past = 96*timestep
n_future = 24*timestep
n_features_input = 12
n_features_output = 1

numlayer = 100
dropout_factor = 0.3
epoch_num = 50
batch_num = 64
nameofcase = 'EP_30_min_60-20-20'
filenameident = 'past-'+str(n_past)+'_future-'+str(n_future)+'_'+nameofcase+'_'+dis

#import CSV data
filenamecsv = importedcsv
data_df=pd.read_csv(filenamecsv, sep=',', header=0,)
data_df = data_df.drop(['Date/Time', 'RH_indoor', 'HR_heating', 'HR_cooling', 'T_supply'],axis=1)

#scaling
data_scaled = data_df
scalers={}
for i in data_df.columns:
    scaler = MinMaxScaler(feature_range=(-1,1))
    s_s = scaler.fit_transform(data_scaled[i].values.reshape(-1,1))
    s_s = np.reshape(s_s,len(s_s))
    scalers['scaler_'+ i] = scaler
    data_scaled[i] = s_s

#data splitting
multiplier = 24*timestep
```

```

mult1 = 219
mult2 = 73
T_arr1 = mult1 * multiplier
T_arr2 = mult2 * multiplier
T_arr3 = (365-(T_arr1 + T_arr2)) * multiplier
arr1, arr2, arr3 = np.split(data_scaled, [T_arr1, (T_arr1 + T_arr2)])
train = arr1
validate = arr2
test = arr3

#converting series to samples for supervised learning
def split_series(series, n_past, n_future):
    #
    # n_past ==> no of past observations
    #
    # n_future ==> no of future observations
    #
    X, y = list(), list()
    for window_start in range(len(series)):
        past_end = window_start + n_past
        future_end = past_end + n_future
        if future_end > len(series):
            break
        # slicing the past and future parts of the window
        past, future = series[window_start:past_end, :], series[past_end:future_end, :]
        X.append(past)
        y.append(future)
    return np.array(X), np.array(y)

X_train, y_train = split_series(train.values,n_past, n_future)
X_train = X_train.reshape((X_train.shape[0], X_train.shape[1],n_features_input))
y_train = y_train.reshape((y_train.shape[0], y_train.shape[1], n_features_input))
y_train=np.delete(y_train, range(0, n_features_input-n_features_output), 2) #assuming all output
#features are in end
X_validate, y_validate = split_series(validate.values,n_past, n_future)
X_validate = X_validate.reshape((X_validate.shape[0], X_validate.shape[1],n_features_input))
y_validate = y_validate.reshape((y_validate.shape[0], y_validate.shape[1], n_features_input))
y_validate=np.delete(y_validate, range(0, n_features_input-n_features_output), 2)
X_test, y_test = split_series(test.values,n_past, n_future)
X_test = X_test.reshape((X_test.shape[0], X_test.shape[1],n_features_input))
y_test = y_test.reshape((y_test.shape[0], y_test.shape[1], n_features_input))
y_test=np.delete(y_test, range(0, n_features_input-n_features_output), 2)
X_train.shape, y_train.shape, X_validate.shape, y_validate.shape, X_test.shape, y_test.shape

#model architecture, one encoder layer and one decoder layer
encoder_inputs = tf.keras.layers.Input(shape=(n_past, n_features_input))
encoder_l1 = tf.keras.layers.LSTM(numlayer, return_state=True, dropout=dropout_factor)
encoder_outputs1 = encoder_l1(encoder_inputs)
encoder_states1 = encoder_outputs1[1:]
decoder_inputs = tf.keras.layers.RepeatVector(n_future)(encoder_outputs1[0])
decoder_l1 = tf.keras.layers.LSTM(numlayer, return_sequences=True, dropout=dropout_factor)

```

```

(decoder_inputs, initial_state = encoder_states1)
decoder_outputs1 = tf.keras.layers.TimeDistributed(tf.keras.layers.Dense
(n_features_output))(decoder_l1)
model_e1d1 = tf.keras.models.Model(encoder_inputs, decoder_outputs1)
model_e1d1.summary()

#model training
reduce_lr = tf.keras.callbacks.LearningRateScheduler(lambda x: 1e-3 * 0.90 ** x)
reduce_lr1 = tf.keras.callbacks.EarlyStopping(monitor='val_loss', patience=7)
model_e1d1.compile(optimizer=tf.keras.optimizers.Adam
(lr=0.00001, beta_1=0.9, beta_2=0.999, epsilon=None, decay=0.0, amsgrad=False),
loss=tf.keras.losses.Huber())
history_e1d1=model_e1d1.fit(X_train,y_train,epochs=epoch_num,validation_data=
(X_validate, y_validate),batch_size=batch_num,verbose=1,callbacks=[reduce_lr,reduce_lr1])

#saving model
filename_h5 = 'Model_Epoch-'+filenameident
model_e1d1.save("/content/drive/My Drive/LSTM_DATA/"+filename_h5+".h5")

#predicting
pred_e1d1_ori=model_e1d1.predict(X_test)
pred_e1d1 = pred_e1d1_ori
y_test_scaled = y_test

#scaling back
scaler = scalers['scaler_T_dry_indoor']
pred_e1d1[:, :, 0]=scaler.inverse_transform(pred_e1d1[:, :, 0])
y_test_scaled[:, :, 0]=scaler.inverse_transform(y_test[:, :, 0])

#saving error metrics data
def MAPE(Y_actual,Y_Predicted):
    mape = np.mean(np.abs((Y_actual - Y_Predicted)/Y_actual))*100
    return mape

f = open('All Errors.CSV', "a")

for j in range(1,n_future+1):
    rmse1 = sqrt(mean_squared_error(y_test_scaled[:,j-1,0],pred_e1d1[:,j-1,0]))
    mape1= MAPE(y_test_scaled[:,j-1,0],pred_e1d1[:,j-1,0])
    mae1 = mean_absolute_error(y_test_scaled[:,j-1,0],pred_e1d1[:,j-1,0])

    print("Timestep ",j,end=", ")
    print('MAE : %.4f' % mae1,end=", ")
    print('RMSE : %.4f' % rmse1,end=", ")
    print('MAPE : %.4f' % mape1)
    f.write('Timestep '+str(j)+' '+str(round(mae1,4))+
    ", "+str(round(rmse1,4))+", "+str(round(mape1,4))+"\n")
f.close()

```
