Project acronym	B4B
Project full name	Brains for Building's Energy Systems
Grant No	M00I32004
Project duration	4 year (Starting date May 1, 2021)

Deliverable 2.1 - 2.2 - 2.3

Building and building systems energy prediction models to enhance energy flexibility control

Authors (in alphabetical order): Isa Dols (HHS), Robbert-Jan Dikken (Peutz), Behrouz Eslami Mossallam (TNO), David Fortini (O-Nexus), Sander van Gameren (Avans), Henri ter Hofte (Windesheim), Wim Kornaat (TNO), Yun Li (TU Delft), Ruud van der Linden (TNO), Bram Muller (HHS), Hossein Rahmani (Windesheim) Baldiri Salcedo (HHS), Nino Visser (HHS), Kees Wisse (DWA)

Work package	WP2
Result	3
Lead beneficiary	The Hague University of Applied sciences (HHS)
Due Date	1 August 2023
Deliverable Status	Final
File name	B4B WP2 D2.1-2.2-2.31 Energy Prediction Models.docx

This project received funding from Dutch Ministry of Economic Affairs and Climate Policy and Ministry of the Interior and Kingdom Relations under the MOOI program.

SUMMARY

The built environment requires energy-flexible buildings to reduce energy peak loads and to maximize the use of (decentralized) renewable energy sources. The challenge is to arrive at smart control strategies that respond to the increasing variations in both the energy demand as well as the variable energy supply. This enables grid integration in existing energy networks with limited capacity and maximises use of decentralized sustainable generation. Buildings can play a key role in the optimization of the grid capacity by applying demand-side management control. To adjust the grid energy demand profile of a building without compromising the user requirements, the building should acquire some energy flexibility capacity.

The main ambition of the Brains for Buildings Work Package 2 is to develop smart control strategies that use the operational flexibility of non-residential buildings to minimize energy costs, reduce emissions and avoid spikes in power network load, without compromising comfort levels. To realise this ambition the following key components will be developed within the B4B WP2: (A) Development of open-source HVAC and electric services models, (B) development of energy demand prediction models and (C) development of flexibility management control models. This report describes the developed first two key components, (A) and (B).

This report presents different prediction models covering various building components. The models are from three different types: white box models, grey-box models, and black-box models. Each model developed is presented in a different chapter. The chapters start with the goal of the prediction model, followed by the description of the model and the results obtained when applied to a case study. The models developed are two approaches based on white box models (1) White box models based on Modelica libraries for energy prediction of a building and its components and (2) Hybrid predictive digital twin based on white box building models to predict the dynamic energy response of the building and its components. (3) Using CO_2 monitoring data to derive either ventilation flow rate or occupancy. (4) Prediction of the heating demand of a building. (5) Feedforward neural network model to predict the building energy usage and its uncertainty. (6) Prediction of PV solar production.

The first model aims to predict the energy use and energy production pattern of different building configurations with open-source software, OpenModelica, and open-source libraries, IBPSA libraries. The white-box model simulation results are used to produce design and control advice for increasing the building energy flexibility. The use of the libraries for making a model has first been tested in a simple residential unit, and now is being tested in a non-residential unit, the Haagse Hogeschool building. The lessons learned show that it is possible to model a building by making use of a combination of libraries, however the development of the model is very time consuming. The test also highlighted the need for defining standard scenarios to test the energy flexibility and the need for a practical visualization if the simulation results are to be used to give advice about potential increase of the energy flexibility.

The goal of the hybrid model, which is based on a white based model for the building and systems and a data driven model for user behaviour, is to predict the energy demand and energy supply of a building. The model's application focuses on the use case of the TNO building at Stieltjesweg in Delft during a summer period, with a specific emphasis on cooling demand. Preliminary analysis shows that the monitoring results of the building behaviour is in line with the simulation results. Currently, development is in progress to improve the model predictions by including the solar shading from surrounding buildings, models of automatic shading devices, and model calibration including the energy use of the chiller.

The goal of the third model is to derive recent and current ventilation flow rate over time based on monitoring data on CO_2 concentration and occupancy, as well as deriving recent and current occupancy over time, based on monitoring data on CO_2 concentration and ventilation flow rate. The grey-box model used is based on the GEKKO python tool. The model was tested with the data of 6 Windesheim University of Applied Sciences office rooms. The model had low precision deriving the ventilation flow rate, especially at low CO_2 concentration rates. The model had a good precision deriving occupancy from CO_2 concentration and ventilation flow rate. Further research is needed to determine if these findings apply in different situations, such as meeting spaces and classrooms.

The goal of the fourth chapter is to compare the working of a simplified white box model and black-box model to predict the heating energy use of a building. The aim is to integrate these prediction models in the energy management system of SME buildings. The two models have been tested with data from a residential unit since at the time of the analysis the data of a SME building was not available. The prediction models developed have a low accuracy and in their current form cannot be integrated in an energy management system. In general, black-box model prediction obtained a higher accuracy than the white box model.

The goal of the fifth model is to predict the energy use in a building using a black-box model and measure the uncertainty in the prediction. The black-box model is based on a feed-forward neural network. The model has been tested with the data of two buildings: educational and commercial buildings. The strength of the model is in the ensemble prediction and the realization that uncertainty is intrinsically present in the data as an absolute deviation. Using a rolling window technique, the model can predict energy use and uncertainty, incorporating possible building-use changes. The testing in two different cases demonstrates the applicability of the model for different types of buildings.

The goal of the sixth and last model developed is to predict the energy production of PV panels in a building with the use of a black-box model. The choice for developing the model of the PV panels is based on the analysis of the main contributors of the peak energy demand and peak energy delivery in the case of the DWA office building. On a fault free test set, the model meets the requirements for a calibrated model according to the FEMP and ASHRAE criteria for the error metrics. According to the IPMVP criteria the model should be improved further. The results of the performance metrics agree in range with values as found in literature. For accurate peak prediction a year of training data is recommended in the given approach without lagged variables.

This report presents the results and lessons learned from implementing white-box, grey-box and black-box models to predict energy use and energy production of buildings or of variables directly related to them. Each of the models has its advantages and disadvantages. Further research in this line is needed to develop the potential of this approach.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Background and objectives

The built environment requires energy flexible buildings in order to prevent energy peak load and to maximize the use of (decentralized) renewable energy sources. The challenge is to arrive at smart control strategies that respond to the increasing variations in both the energy demand as well as the variable energy supply. This enables grid integration in existing energy networks with limited capacity and maximises use of decentralized sustainable generation.

The main ambition of the B4B project is to develop smart control strategies that use the operational flexibility of non-residential buildings to minimize energy costs, reduce emissions and avoid spikes in power network load, without compromising comfort levels. Furthermore the goal is to show that flexibility of energy production/demand of smart buildings can reduce operational constraints in the current power network.

To realise this ambition the following key components will be developed within the B4B project:

A. Development of open-source HVAC and electrical services models
   To lay the foundation for our research, we embark on the creation of open-source models for heating, ventilation, and air conditioning (HVAC) systems and electrical services within non-residential buildings. These models will serve as the building blocks for the development of the energy prediction models.
B. Development of energy prediction models
   Building upon the HVAC and electrical services models, we will proceed to construct advanced energy prediction models. These predictive models are essential for forecasting energy consumption patterns, enabling proactive and efficient energy management.
C. Development of flexibility management control methods
   To harness the inherent flexibility within non-residential buildings (due to thermal mass or local storage/generation options), we will pioneer innovative control methods. These methods will facilitate the dynamic allocation of energy resources, ensuring optimal/flexible energy utilization while minimizing wastage and ensuring indoor comfort. The most promising control methods will be tested in living labs, first in a digital environment and later on in real buildings.

This report describes the developed HVAC and electrical services models, and the energy prediction models within components (A) and (B) listed above.

1.2 Approach

A structured and collaborative approach has been defined to achieve the objectives of the B4B project. Before starting, the consortium partners discussed and decided about the distribution of tasks. This ensured that a wide variety of models have been developed in the B4B project and it has also allowed partners to exchange feedback about the different models, gaining a common insight on the pros and cons of each approach. The developed models have been tested using real data from living labs and use cases in order to validate their results and to demonstrate their possible implementation for the control of buildings' energy flexibility.

The work on the development of the three key components (A), (B) and (C) has been carried out simultaneously. As the learning outcomes of each key component can have a positive impact on the development of the other key components, regular outcome sharing meetings have been organized and collaboration between consortium partners have been established. As a consequence of the continuous feedback rounds the development of the models is still ongoing. This report covers the current results of key components (A) and (B) and will be updated with the last version of the models developed towards the end of the project.

1.3 Structure of the report

In Chapter 2, background information is given about energy flexibility in buildings, the different building energy models and building characteristics to be included in the building energy demand models. Furthermore, an overview is given of all models that are developed and described in this report. In the subsequent chapters the different modelling approaches are described in detail, where each chapter starts with a description of the model, followed by the results of their application in practice :

- Chapter 3 describes a white box modelling approach using Modelica libraries for energy prediction of the building and its components.
- Chapter 4 describes a hybrid modelling approach based on white box building models using a predictive twin for energy prediction of the building and its components.

- Chapter 5 describes an inverse grey-box modelling approach to use CO_2 monitoring data to derive either ventilation flow rate or occupancy.
- Chapter 6 describes a simple white box and a simple black box model to predict the heating demand of a building
- Chapter 7 describes a black box modelling approach using a feedforward neural network to predict the building energy usage and its uncertainty
- Chapter 8 describes a black-box modelling approach using neural network to predict the PV solar production.

2 BACKGROUND

2.1 Energy flexibility in buildings

The production Gap Report (SEII et al., 2021) proposes an average yearly reduction of 6% on the use of fossil fuels to achieve the CO2 reduction targets of the Paris Agreement (United Nations, 2015). As a consequence of the needed energy system change, there has been in recent years an increase on electricity demand, due to higher use of heat pumps and electric cars, as well as an important growth of renewable energy production, particularly solar and wind energy (OECD, 2021). Therefore, the electricity grid is working at a higher capacity rate causing grid congestion problems. Not only because of the increase of the total amount of electricity being transported but also because the increase in renewable energy production creates an imbalance in the electricity grid between energy production and consumption due to its dependence on natural elements (Zsiborács et al., 2021).

One way to rebalance the grid is to adjust consumption in relation to production that in the literature is referred to as demand-side management control. Buildings account for 40% of the total energy consumption in the EU (IEA, 2019). They can therefore play a key role in the balance of the grid and optimization of the grid capacity. To be able to adjust the grid energy demand profile of a building without compromising the user requirements, the building should acquire some energy flexibility (Reynders et al., 2018; Marszal-Pomianowska, 2019). The concept of building energy flexibility was addressed early on in (Six et al., 2011; Nuytten et al., 2013), among others and numerous definitions have been proposed since then. In this project building energy flexibility is defined as: *"The capacity of a building to deviate its grid energy supply and grid energy demand from the reference profile, reacting to a specific control request without impairing functionality or thermal comfort."*

The definition used is based on the definition outlined in Annex 67 report on 'Characterization of Energy Flexibility in Buildings' (EBC Annex 67, 2018). The main difference is the specification that the energy supply and energy demand patterns that are aimed to be modified are the grid energy demand and the grid energy supply. This allows to clearly identify that the exchange of energy is between the grid and the building. The other change is to make it explicit that the deviation is activated by a control request. A conceptual visualization of the definition is presented in Figure 2.1.

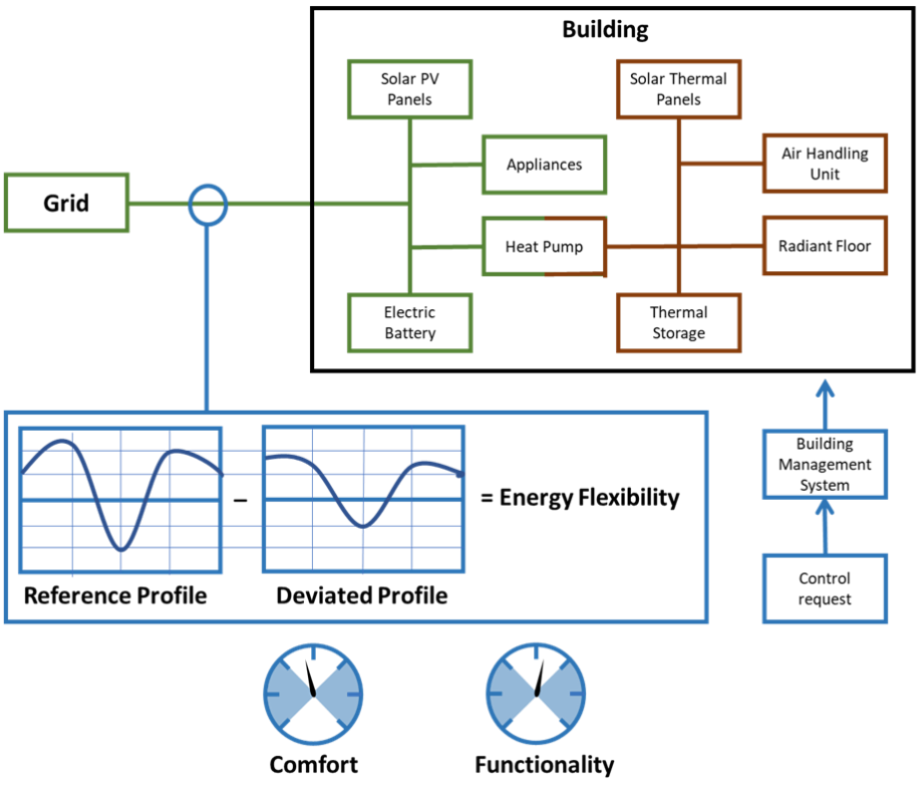


Figure 2.1. B4B conceptual visualization of building energy flexibility.

Numerous scientific publications have proposed KPIs to characterise energy flexibility. A complete list of energy flexibility KPIs that can be found in the literature are presented in Appendix A of the Annex 67 report on

'Characterization of Energy Flexibility in Buildings' (Marszal-Pomianowska, 2019). In the same report, a flexibility function is proposed to assess the energy flexibility. The function is described in Figure 2.2.

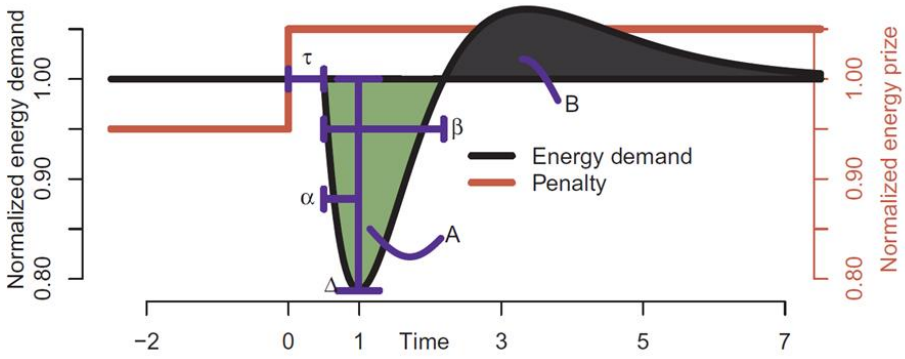


Figure 2.2. Energy Flexibility function. (Marszal-Pomianowska, 2019).

τ (Time): The delay of seeing an effect on the energy demand.

Δ (Power): The maximum change in demand following the penalty change.

α (Time): The time it takes from the start in change in demand until it reaches the lowest level.

β (Time): The total time of decreased energy demand.

A (Energy): The total amount of decreased energy demand.

B (Energy): The total amount of increased energy demand.

The flexibility function allows to numerically determine the characteristics of the difference between the reference energy profile and the deviated energy profile. To do so, it is only necessary to give as input the reference profile and the deviated profile. As it is difficult to measure in real life a reference and deviated energy building profile under similar circumstances, the energy flexibility is most likely to be assessed by making use of building simulation models. Therefore, the primary goal of this report is to develop models that allow to simulate the energy profile of a building when reacting to a specific control request.

This document describes the building models and the HVAC and electrical services models developed to generate a reliable building energy supply and demand prediction for non-residential buildings. The models could be used as a tool to assist in the building systems design process or as a digital twin to assist in flexibility control actions.

2.2 Building energy models

The techniques currently used to model energy-related systems in buildings, which include predicting, managing, and optimizing energy systems design and control, can be classified into three groups: physics-based modelling (also known as white-box models) including hybrid building models based on white box building models and data driven user models, data-driven modelling (also known as black-box models), and grey-box models (Yu et al., 2022; Habash,2022).

White-box models

White box models are based on the principles of physics. It is a modelling approach that explicitly represents the internal workings, physical properties, and energy flows within a building. The advantage of white box models is that they make use of known physical properties such as heat transfer and therefore need less informative data to calibrate the model. The downside is that detailed physical parameters of the building dimensions and material properties must be available and depending on the complexity of the white box model can take considerable time and effort to initiate the model. (Chen et al. 2022). TNO uses a white box model to describe the building model and uses data driven models for user behaviour (thermostat setpoint, shading etc)

Black-box models

Black-box models are data-driven models, as they are created by quantifying parameters from historical data and establishing correlations between building performance and data. The parameters of black-box models are typically adjusted automatically, which offers a notable advantage over white-box models. For a data driven model to be accurate it needs training data that has sufficient information. In contrast, white box models contain physical relationships that are already known. In many cases, the information and the quality of building data is low, which makes it difficult to train black box models. Nevertheless, the inner workings of

these models are mostly unknown, as the parameters do not have a physical meaning, which limits the control flexibility of the overall process. (Wilde, 2023)

Grey-box models

Grey-box models are a combination of white and black-box models. Grey-box models define a physical model to depict the building's physical layout and/or of the HVAC equipment. Subsequently, they employ statistical analysis to determine significant parameters that represent crucial aggregated physical characteristics and they may use black box models to estimate parts that are not described by physical models for instance user behaviour. Accomplishing this necessitates advanced user knowledge in formulating appropriate modelling equations and estimating these parameters. This approach has great potential, particularly in the domains of fault detection and diagnosis (FDD) as well as real-time control. Since the model contains known physical relations, it can be trained with building data that contains less information. (Harish & Kumar, 2016; Li,2021)

2.3 Building characteristics to be included in building energy models

To give insights into the most common non-residential building characteristics, a preliminary analysis has been carried out about the building characteristics of office buildings of The Central Government Real Estate Agency (CGREA), *Rijksvastgoedbedrijf*. The analysis is based on the available CGREA building information database data and the available building energy performance certificate reports. The results of this analysis are presented in appendix A of this report.

This analysis's initial goal was to identify the highest priority building component configurations to be modelled. However, in the course of this research project, it appeared that this information would not interfere with the decision of what building component configurations need to be modelled because the current most common building configurations offer reduced building energy flexibility capabilities (limited energy storage capacity, reduced energy production capacity and limited predictive energy control capacity). However, the analysis offers an overview of what are the average current office building characteristics and building component configurations. This is still a valuable piece of information when making building energy models of office buildings.

2.4 Building energy models developed

The partners involved in making this report have developed models using all of the three techniques described above to facilitate design decisions or the implementation of control algorithms that enhance the energy flexibility of buildings. The list of models, the use-case on which the model has been built, and the aim of the model are presented in Table 2.1. The table also refers to the report chapter in which the model is presented.

Table 2.1. Building and building system models developed

Model type	Living Lab/Use Case	Prediction goal	Comments	Chapter
White box model	HHS Delft	Using open-source software and libraries to simulate different building configurations to produce design and control advice for increased energy flexibility.	The higher education HHS Delft Building has 3 levels, it is well insulated and it is equipped with an aquifer thermal storage system and floor heating.	3
White box model, hybrid model	TNO Stieltjesweg Delft	Cooling demand of the building for use in predictive control	The TNO Stieltjesweg building is a 3 level office building with optical labs on the ground floor.	4
Inverse grey box	Building Windesheim	Derive current ventilation flow rates from recent CO_2 concentration and occupancy data and derive current occupancy data from recent CO_2 concentration and ventilation flow rate data.	Collected data from 6 office rooms at Windesheim for 3 weeks in the autumn of 2022, but had to reject the data of 3 office rooms due to data quality issues. Dataset of the 3 office rooms, 8 properties from 4 sources, totalling 0.4 million datapoints published as open dataset. Firmware for CO_2 and occupancy detection measurement device (€85 per room) and GEKKO Python based model and analysis software published as open source software.	5
White+ Black box model	O-nexus	White Box: Modelling heat demand with a physical equation.	The dataset comes from the existing O-Nexus database and is based on	6

Model type	Living Lab/Use Case	Prediction goal	Comments	Chapter
		Black box: Predict heating demand so that power steering control algorithms can be optimized.	different types of residential housing. It is a private dataset.	
Black box	Peutz	Indication of expected building energy usage and associated uncertainty in the expectation.	Tested with two datasets for an educational facility and a commercial facility.	7
Black box	DWA Gouda	Peak prediction of PV solar systems.	Tested with fault free and faulty data.	8

3 WHITE BOX MODELS: MODELICA LIBRARIES FOR ENERGY PREDICTION OF A BUILDING AND ITS COMPONENTS

Authors: Isa Dols, Gide de Ruijter, Baldiri Salcedo (HHS)

3.1 Goal of the prediction models

The main goal of this model is to predict the energy use and energy production pattern of different building configurations with open-source software and open-source libraries. The aim is to use this model to produce design and control advice for increasing the building energy flexibility. This chapter presents the discussion about the modelling software and libraries choice, the characteristics of the available libraries, the simulated case study to test the software and libraries, and the lessons learned. This chapter will be updated before the end of the B4B project to include the results of the HHS building case study.

3.2 Choice for Modelica and OpenModelica

The use of computer building energy simulations (BES) started in the 1960s and 1970s with the development of digital computers and the increasing demand for energy-efficient buildings. Since the first building energy simulation tool called BRIS was compiled by the Royal Institute of Technology, Stockholm in 1963, BES techniques have significantly advanced (Brown, 1990). In the following decade, powerful simulation engines such as BLAST, DOE-2, and ESP-r were developed. Later, with the advancement of computational techniques, more sophisticated BES programs such as DeST and EnergyPlus were developed in the 1990s, laying the foundation for current BES technology (Singh & Sharston, 2022). Thanks to continued development, BES tools can now accurately predict thermal loads and build energy performance based on fundamental heat balance equations. Today, BES is essential for architects, engineers, and building owners to optimize building energy use and reduce energy costs. A comprehensive list of BES tools is provided in Crawley et al. (2008) and IBPSA (n.d.).

The goal for the HHS case study is to develop a computer model based on open-source and freely available software to simulate the energy use of a building and the behaviour of the components of the energy systems of this building. The need to dynamically simulate the building components (such as the HVAC and renewable energy production systems) and define the control settings of these components reduce the amount of suitable BES tools. Atam (2017) compared the characteristics of the currently available software with these capabilities, see Table 3.1.

Table 3.1. Comparison of discussed major software/tools in terms of control-oriented and other closely-related aspects. The scales 0–5 indicate a property strength from lowest to the highest level. (Atam, 2017).

	Matlab	Dymola/Modelica	Trnsys	EnergyPlus	Revit	Hot2000	Esp-r	IDA-ICE
Open-source	Yes	Yes/Yes	Yes	Yes	No	No	Yes	Yes
Freely available	No	No/Yes	No	Yes	No	Yes	Yes	No
User-friendliness	5	3/2	1	1	3	3	1	3
Level of detail in building geometry modelling	0	1/1	4	4	5	2	4	3
Rich libraries for building and HVAC thermal models	1	4/5	5	4	3	2	3	4
Libraries for renewable energy systems	0	3/4	5	3	1	0	2	4
Advanced control design capability (built-in)	5	2/1	0	0	0	0	0	0
Allowing co-simulation	Yes	Yes	Yes	Yes	No	No	No	No
FMI support (built-in or through third-party tools)	Yes	Yes	Yes	Yes	No	No	No	No
Post processing capability	5	2/1	1	1	1	1	1	1

The goal of the case study is to develop a computer model based on open-source and freely available software that allows implementing advanced control design capability. Therefore, the choice has been made to use Modelica with OpenModelica software to develop the computer model for this case study as it is the only option

on the list that complies with the defined requirements. Moreover, Modelica has a rich library for building and HVAC components models.

3.3 Modelica libraries

The IBPSA project 1 executed between August 2017 and August 2022 created the IBPSA library composed of four Modelica libraries for building and district energy systems (IBPSA project 1,n.d.):

- AixLib, from RWTH Aachen University, Germany, (https://github.com/RWTH-EBC/AixLib)
- Buildings, from LBNL, Berkeley, CA, USA, (http://simulationresearch.lbl.gov/modelica)
- BuildingSystems, from UdK Berlin, Germany, (http://www.modelica-buildingsystems.de/)
- IDEAS, from KU Leuven, Belgium, (https://github.com/open-ideas/IDEAS)

IBPSA has recently created the IBPSA Modelica Working group to maintain and further develop the libraries. The library is being tested with three Modelica tools: OpenModelica, OPTIM-ICA/IMPACT and Dymola (Wetter, 2023).

From the four libraries, the libraries "Buildings" and "IDEAS" are selected to be used in the making of the model for the HHS because of their more advanced stage of development. The Buildings library developed by the Lawrence Berkeley National Laboratory has the support of the Department of Energy of the United State and is currently also developing a link to EnergyPlus called the Spawn of EnergyPlus (Spawn of EnergyPlus, n.d.). The IDEAS library is developed by KU Leuven. The IDEAS development group of KU Leuven has been contacted during the making of the model to get some support on using their library.

The Buildings and the IDEAS library can directly be activated in OpenModelica via the Manage Libraries menu. The Buildings library is fully compatible with OpenModelica while the current version of the IDEAS library is only partially compatible with OpenModelica. However, the developers of the IDEAS library that were contacted during this research expressed their aim to make their library fully compatible with OpenModelica in the close future.

3.4 Modelica library components

In this section there is an overview of the available model blocks in the Buildings 9.1.0 and IDEAS 3.0.0 library.

3.4.1 Thermal zones

The libraries offer 3 methods to define the thermal zones of a building, reduced order, detailed and EnergyPlus model. The reduced order are model blocks with different RC configurations, the detailed blocks are blocks implementing heat transfer equations for walls and windows. The IDEAS library are a bit more user friendly than the Building library blocks as it allows to define directly a thermal zone including the whole envelope of that thermal zone in one modelling block. However, the IDEAS library blocks for the thermal zones are not yet fully compatible with OpenModelica. The EnergyPlus model block is only available on the Buildings library and it is still in early phase of development within the project the Spawn of EnergyPlus (U.S. department of Energy, n.d.). However, it is promising as it could really increase the speed of defining thermal zones.

Table 3.2 Thermal zones model blocks IDEAS and Buildings libraries.

Model Blocks	IDEAS	Buildings	Comments
Reduced order	X	X	Similar in both libraries
Detailed	X	X	IDEAS more user friendly but not fully compatible with OpenModelica
EnergyPlus		X	In development in project the Spawn of EnergyPlus.

3.4.2 Heating and cooling generation

Both libraries have some thermal generation model blocks. See Table 3.3. The Buildings library has a few more blocks than the IDEAS library, like for example a boiler, or solar a collector block. For the heat pumps, there are different types of blocks, simplified versions with a COP depending on the temperature difference, Carnot heat pump, and more complex heat pump blocks that include the simulation of the compressor, condenser and evaporator. The IDEAS library has a few more heat pump types than the Buildings library. Both libraries include geothermal boreholes, however in both cases these are closed systems.

The availability of electricity generation blocks is limited. The Buildings library has solar panels and wind turbines blocks that model the generated power with a direct relationship with the solar irradiance for the solar

panel and with a direct relationship with the wind speed for the wind turbine. The solar panels block in the IDEAS library is still under validation.

Table 3.3 Generation model blocks in the IDEAS and Buildings libraries.

Model Blocks	IDEAS	Buildings	Comments
Boiler		X	Simple block model
Heat pump	X	X	Numerous model blocks with varying levels of complexity
Solar collector		X	Simple block model
Geothermal system	X	X	Only closed system.
PV panel	X	X	Simple block model
Wind turbine		X	Simple block model

3.4.3 Distribution

For the distribution of water and air both libraries have blocks to simulate pipes and mixing volumes taking into account moisture as well as heat exchange. See Table 3.4. There are also blocks of stratified storage tanks. There is not a block or example yet available of an Air Handling Unit. However, with the available blocks it is possible to configure one. For the distribution of electricity only the Buildings library has model blocks for AC and DC distribution. The default values of the blocks are according to the US electric system. The Buildings library includes an electric battery model block, but it is an extremely simplified model block.

Table 3.4 Distribution model blocks in the IDEAS and Buildings libraries.

Model Blocks	IDEAS	Buildings	Comments
Pipes	X	X	Simple block model
Heat exchangers	X	X	Numerous model blocks with varying levels of complexity
Mixing volumes	X	X	Including moisture
AHU			Could be modelled with the available blocks.
Storage tank	X	X	Stratified storage tank
AC/DC		X	Default values US electrical system
Electrical Battery		X	Extremely simple model

3.4.4 Delivery

For the delivery of heat and cold both libraries include several heat exchanger model blocks with different levels of complexity, the blocks include radiators, convectors and radiant floor. See Table 3.5.

Table 3.5 Delivery model blocks IDEAS and Buildings libraries.

Model Blocks	IDEAS	Buildings	Comments
Radiators	X	X	Numerous model blocks with varying levels of complexity
Convectors	X	X	Numerous model blocks with varying levels of complexity
Radiant floor	X	X	Numerous model blocks with varying levels of complexity

3.4.5 Boundary conditions

Both libraries have blocks related to climate boundary conditions and occupant boundary conditions. See Table 3.6. For the weather data both libraries have a block that loads information from TMY3 (Typical Meteorological Year) weather files. The files with this format can be downloaded at the EnergyPlus website. The block from the IDEAS library has also included the computing of the solar irradiance on the zone surfaces.

The occupant behaviour is also modelled for some blocks in both libraries. However, the Buildings library has numerous options to define the occupant behaviour in relation to presence, lighting, blinds use and settings in heating and cooling making use of adaptable schedules (Wang et al., 2019).

Table 3.6 Boundary conditions model blocks IDEAS and Buildings libraries.

Model Blocks	IDEAS	Buildings	Comments
Weather data	X	X	Load TMY3 files. IDEAS also computes irradiance in surfaces.
Occupants' behaviour	X	X	Buildings has more options to define occupants' behaviour.

3.4.6 Control

Both libraries have control model blocks. See library 3.7. IDEAS has a few basic control options: PID, off-timer and heating curve-based control. Buildings has plenty more control options, it also includes some energy demand prediction blocks and demand control blocks. The development of the control blocks is part of the Open Building Control project (U.S. department of Energy, n.d.).

Table 3.7 Control model blocks IDEAS and Buildings libraries.

Model Blocks	IDEAS	Buildings	Comments
Basic control	X	X	PID, Off timer, heating curves.
Demand response control		X	Developed in Open Building Control project.

3.4.7 Overall evaluation

The Buildings library is further developed and has more model blocks than the IDEAS library, however for some topics libraries can complement each other. It is also expected that the Buildings library will be quickly further developed because of the support of the Department of Energy of the United States and the spin-off projects to develop specific library components like the Spawn of EnergyPlus or the Open Building Control.

3.5 Case Study 1. Residential NZEB row house

The objective of the first case study is to evaluate how feasible it is to make a simple building model in OpenModelica making use of the existing model blocks to evaluate the energy flexibility of different building configurations and control strategies. The requirement of being a simple building for the first case study determined the choice for a residential building, in place of a non-residential building, even though the focus of the B4B project is non-residential buildings. The results presented in this report have been also published in a paper (Dols & Salcedo Rahola, 2023).

3.5.1 Buildings characteristics

The case study building is a Net Zero Energy Building (NZEB) row house located in the south-west of the Netherlands. See Figure 3.1. The house is equipped with an air-water heat pump used only for heating purposes with underfloor heating and a ventilation system. The house was chosen because of the availability of the house characteristics as well as the energy use of the heat pump in 120 similar houses. There was no information available about the occupants of the house.

Figure 3.1 Case study NZEB row house.

3.5.2 Computer model

The computer model created is based on the simple house example created by Michael Wetter from the IBPSA library, see Figure 3.2. This model is adapted to the envelope and installations of the case study. The building envelope is simulated with simple RC model blocks. At the start of this research, we made use of the IDEAS library, which has reduced order model blocks and detailed model blocks to define building thermal zones, but the detailed model blocks are now not compatible with OpenModelica. The choice for the IDEAS library was because we had the opportunity to have direct contact with the developers of the library what was of special interest given the lack of information and tutorials for making models with OpenModelica and specific with the IBPSA libraries.

The main components of the heating and ventilation system have been taken into account in the model. The heating system is controlled by a thermostat that is implemented by making use of OpenModelica standard library model blocks.

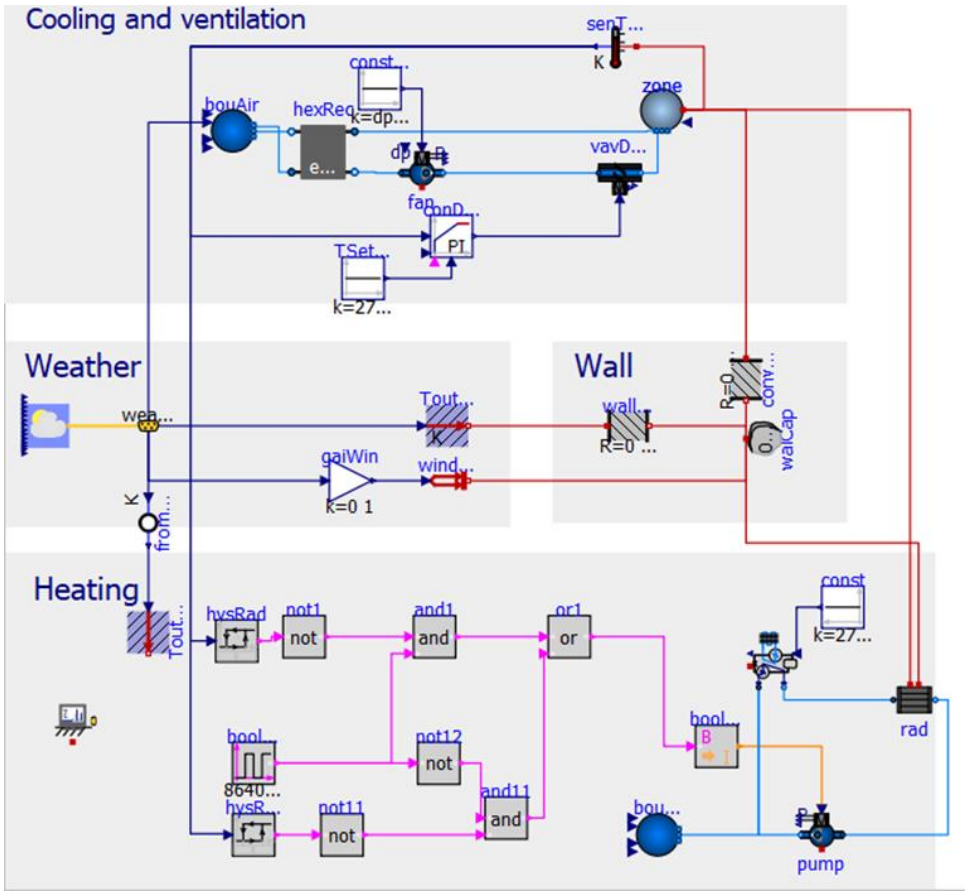
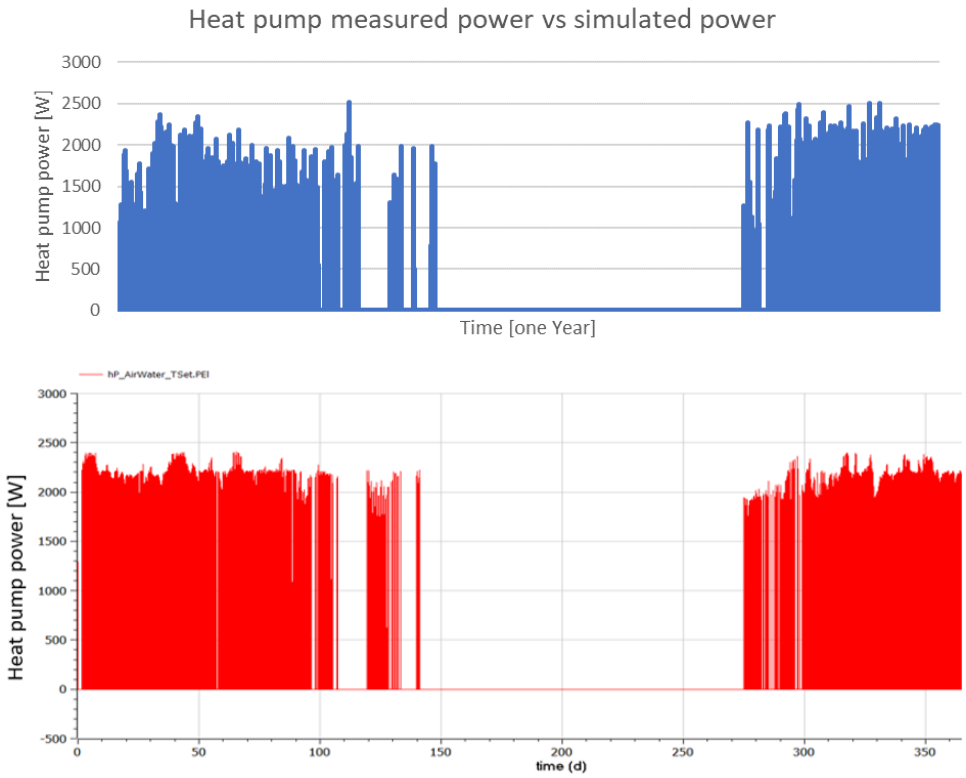


Figure 3.2 Reference model adjusted version from IBPSA example Simple model Michael Wetter.

The performance of this model has been verified by comparing its outputs with the data available from the case study. As there was no data available about the occupants and the occupant's behaviour, only the energy use trends and total energy use values have been used for the verification of the model. The model gave a total energy use for the heat pump 38% higher than the average measured heat pump energy use of the reference houses. A possible explanation for the difference in values is the fact that no internal heat gains were accounted in the model and that the simplicity of the RC model did not represent the houses accurately. However, the average actual energy use and the simulated energy use followed the same pattern, see Figure 3.3. The model result was close enough for the purpose of testing the feasibility of using OpenModelica to measure the energy flexibility.



	Yearly heat pump energy use [kWh/year]
Average measured	1190
Simulated	1506

Figure 3.3 Measured versus simulated heat pump power and energy comparison.

3.5.3 Energy flexibility test measurement

To assess the energy flexibility of the building, the computer model of the case study was used to generate the load curve of different building configurations. These building configurations were created by adjusting a parameter within the reference model to create a flexible model. The parameters considered by Annex 67 are: (IEA EBC Annex 67, 2017).

- Insulation level
- Thermal inertia
- Heating/cooling system
- Control strategy/penalty signal
- Building type
- Outdoor temperature
- Solar radiation

Flexible models were developed to assess the impact of the parameters on energy flexibility. When analyzing the sensitivity diagram of (Junker et al., 2018), three parameters were derived as the most applicable for this case study: insulation, thermal inertia and control. Therefore, these three parameters were analyzed within the case study. This resulted in three variant 'flexible' models, each with one adjustment.

Table 3.8 Flexible model parameters

Parameter	Adjustment	Reference model	Flexible model
Insulation	Rc_value	5 [m²*K/W]	7 [m²*K/W]
Thermal Inertia	Wall thickness (brick)	50 mm	100 mm
Control	Preheating	20-21°C	21-22°C

3.5.4 Setup

In the current literature there is no standard procedure for choosing under which scenario to measure the energy flexibility. Therefore, the penalty signal has been simulated on the moment is more likely to occur an energy shortage on the electricity grid. These models have been simulated in a worst-case scenario: a day with no sunshine and an outside temperature of -10 °C.

In the current literature there is neither a standard control response to the penalty signal. In the simulated case-study the only control component is the thermostat. Therefore, it was chosen that the control response of the model to the penalty signal would be to reduce the user requirement temperature to 19.5 °C from the initial 20 °C. As this research focuses on the response of the models, the initiation of this response was taken as a binary signal and was activated in the simulations at 8:00 am. Furthermore, the simulation was run for 14 hours to simulate a hypothetical time frame from 08:00 to 22:00, this was done to properly measure the rebound effect. The effect on the room temperature of the simulation of the three models compared to the reference model is shown in Figure 3.4.

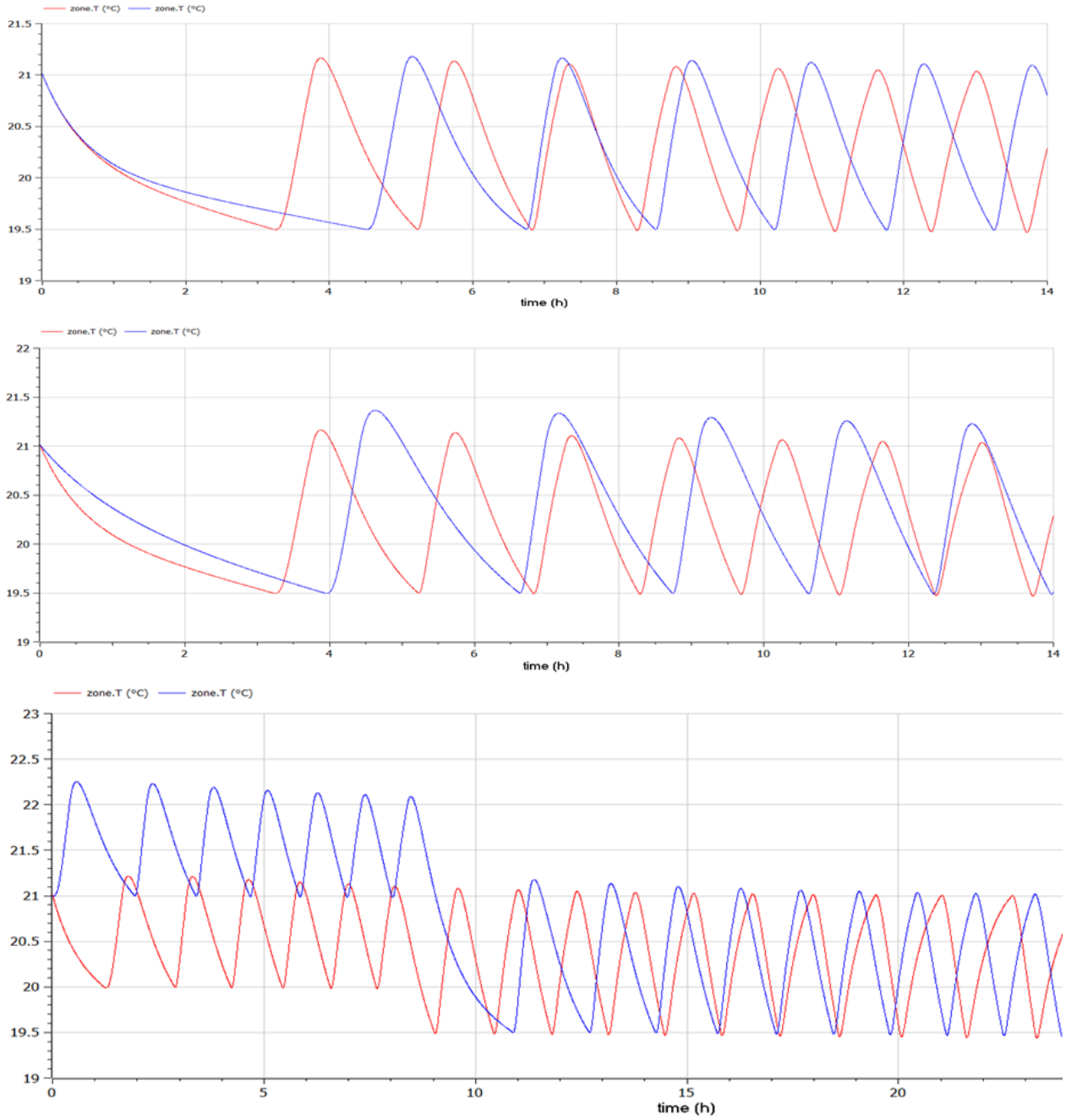


Figure 3.4 Indoor temperature for the insulation (top graph), thermal inertia (middle graph) and control (bottom graph) flexible model in comparison to the reference model.

3.5.5 Data Analysis

The energy load produced with the different computer models has been exported to Excel to produce the comparison graphs, see Figure 3.5. From these comparison graphs the value of the energy flexibility KPIs as defined by annex 67 could be extracted (Marszal-Pomianowska, 2019), for the KPIs list and definition see Figure 1.2.

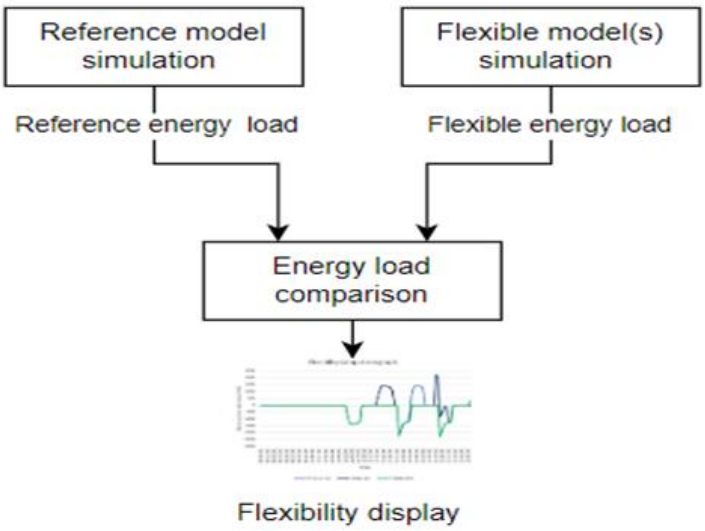
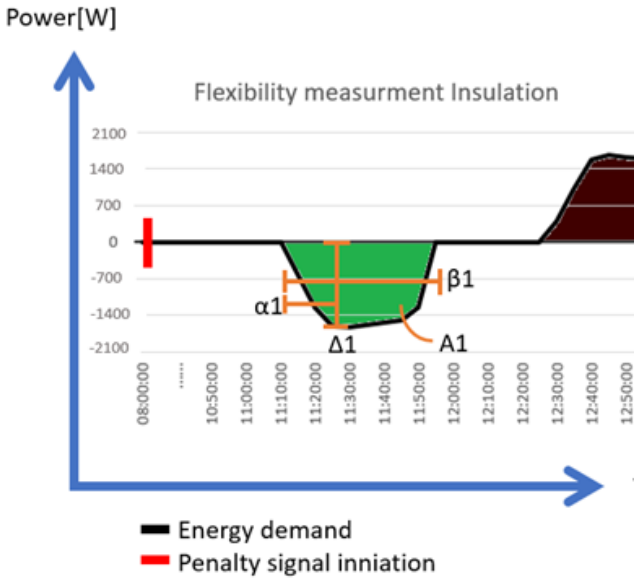


Figure 3.5 Data processing for the generation of the flexibility display.

The energy flexibility KPIs are measured in the period between the first initialization and the end of the rebound effect.

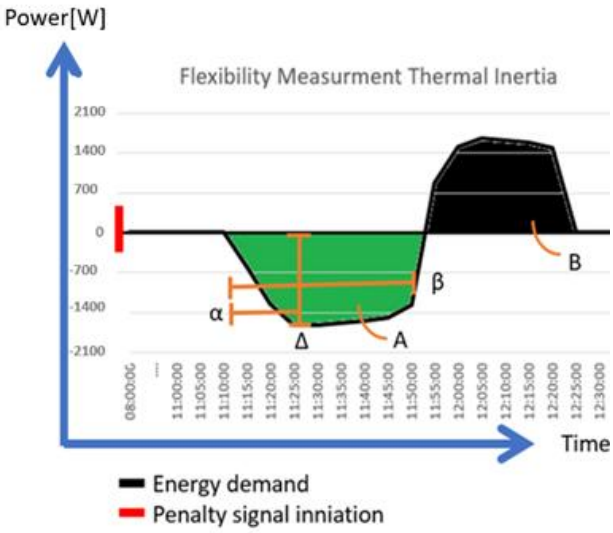
3.5.6 Results

The flexibility function and the values of the KPIs for the comparison of the three flexible models with the reference model can be seen in the Figures 3.6, 3.7 and 3.8. The green area is the shifted energy and the black area the energy of the rebound effect. In the case-study the heat pump is the only appliance consuming energy, therefore there are some periods in which the difference between the reference model and the flexible model is equal to zero, when heat pump is off in both models. Even though the simplicity of the model and of the control request there are clear differences on the energy flexibility function pattern. Specially in the case of the pre-heating test, Figure 3.8, in comparison the other two test, Figure 3.6 and 3.7. In the test of the pre-heating not only the energy shifted is higher but also the maximum power difference is higher.



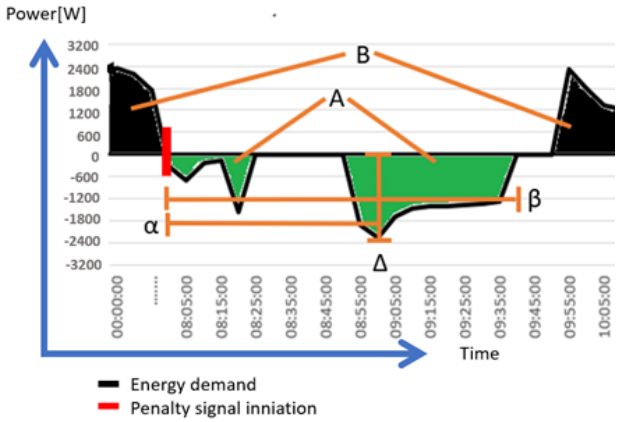
KPI	Ab.	Value	Unit
Duration time	β	3000	s
Duration peak time	α	1200	s
Shifted energy load	A	-0.78	kWh
Rebound effect	B	0.77	kWh
Maximum power shift	Δ	-1488	W

Figure 3.6 Insulation flexibility graph.



KPI	Ab.	Value	Unit
Duration time	β	2700	s
Duration peak time	α	1200	s
Shifted energy load	A	-0.77	kWh
Rebound effect	B	0.63	kWh
Maximum power shift	Δ	-1488	W

Figure 3.7 Thermal inertia flexibility graph.



KPI	Ab.	Value	Unit
Duration time	β	6300	s
Duration peak time	α	3600	s
Shifted energy load	A	-1.8	kWh
Rebound effect	B	6.5	kWh
Maximum power shift	Δ	-2420	W

Figure 3.8 Pre-heating flexibility graph.

3.5.7 Discussion

The graphs have been post-processed to facilitate their interpretation. However, when they were shown to a small group of engineers and architects to request for feedback, they expressed that it was difficult for them to interpret the graphs and to take design decisions based on the analysis of these graphs. From these conversations it appeared that a dashboard in which the values of certain building parameter, such as for example the concrete floor thickness or the building façade Rc value, could be modified with a slider could facilitate the interaction with the model and its understanding. A visualization of a future possible dashboard is presented in Figure 3.9.

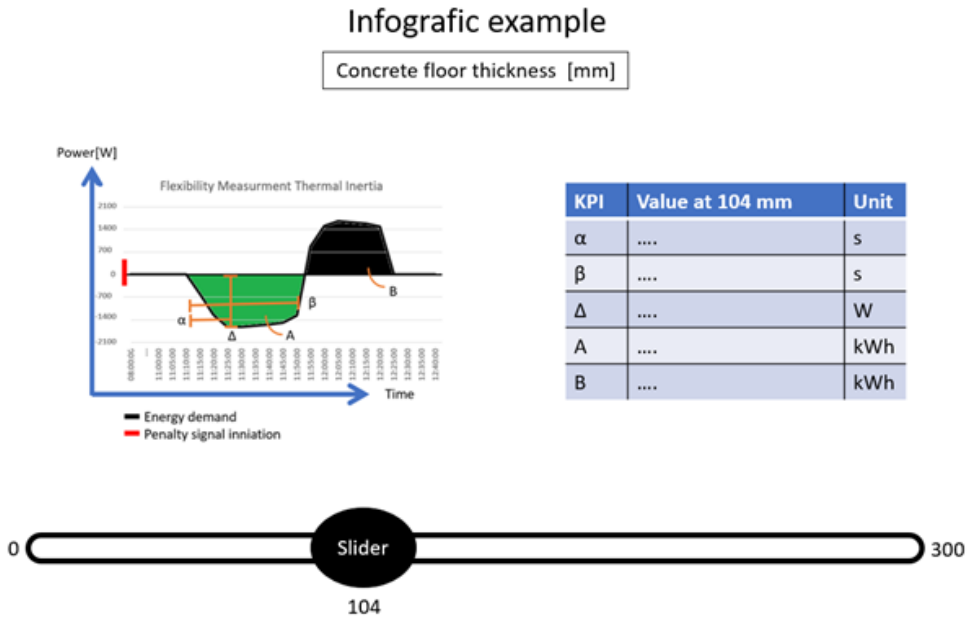


Figure 3.9 Possible future model interaction and visualization.

3.5.8 Conclusions

The developed case study has shown that it is possible to model a rowhouse with OpenModelica making use of the IDEAS library with the aim to analyse the energy flexibility of different house configurations. The energy use patterns of the reference house and the different alternative configurations can be compared to create a flexibility energy pattern. OpenModelica and the use of the IBPSA libraries has a steep learning curve. There is really limited documentation and tutorials available, and currently the IDEAS library is not fully compatible with OpenModelica.

It is necessary to define a standard situation in which the flexibility is analyzed to facilitate a fair comparison. In this case study it is proposed to compare the action of the penalty signal at an expected high grid energy demand moment like on a cold day (-10 degrees Celsius) at a peak hour time (8:00 in the morning). However, other standard situations could be defined, such as a peak summer demand, or a solar power production peak.

The comparison between energy patterns can be plotted on a graph and KPIs could be calculated to compare them. The graph generation and calculation of KPIs is easier to be implemented in software different than OpenModelica. In this case study Excel has been used with quite some post processing. It is necessary to limit the flexibility pattern analysis to the end of the rebound effect and to add a colour and labelling to facilitate the readability of the graphs. In the future it could be useful to automatise this process.

Moreover, to facilitate design choices regarding flexibility it is necessary to define a way to interact with specific building parameters, like for example the thickness of the floor concrete or the Rc value of the façade, and visualize how these building parameter changes influence the energy flexibility KPIs.

3.6 Case Study 2. Non-residential building: The Hague University of Applied Sciences Building.

The objective of the second case study is to evaluate how feasible it is to make a building model of a complex non-residential building in OpenModelica making use of the existing model blocks in order to evaluate the energy flexibility of different building configurations and control strategies. Unfortunately, at the time of the publication of this report the case study is not yet finished. In this chapter the work that has been done to create the model is presented, including the obstacles encountered during the process. It is expected that the model will be published in an updated version of this report.

3.6.1 Building characteristics

The case study building is the Delft building of the Hague University of Applied Sciences. The building designed by architecture office Syb van Breda was built in 2009, has a surface of 16270 m^2 and it was designed to host

2000 students. The building has an energy label A with an EPC value of 0,329. A visualization of the building characteristics is presented in Figure 3.10.

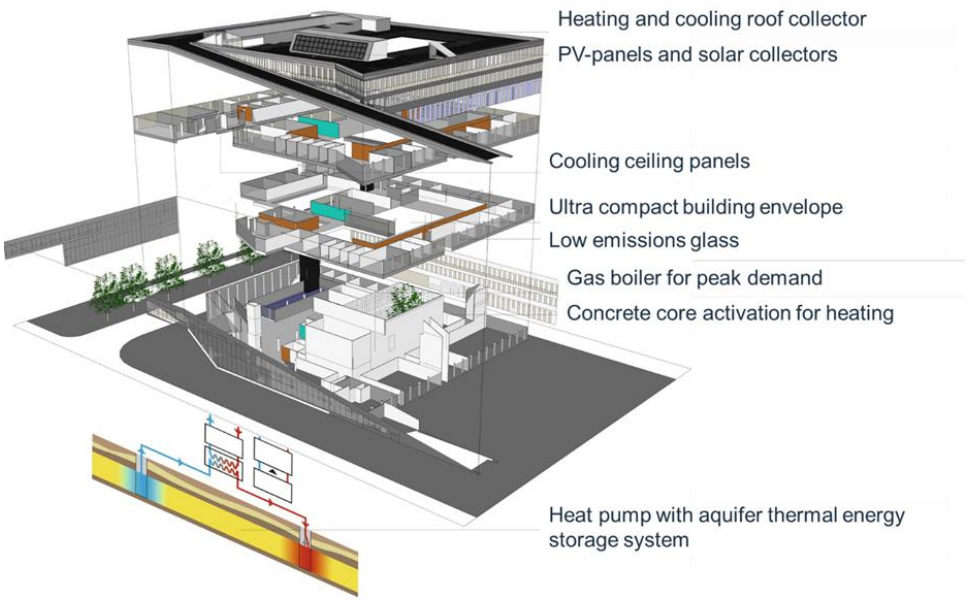
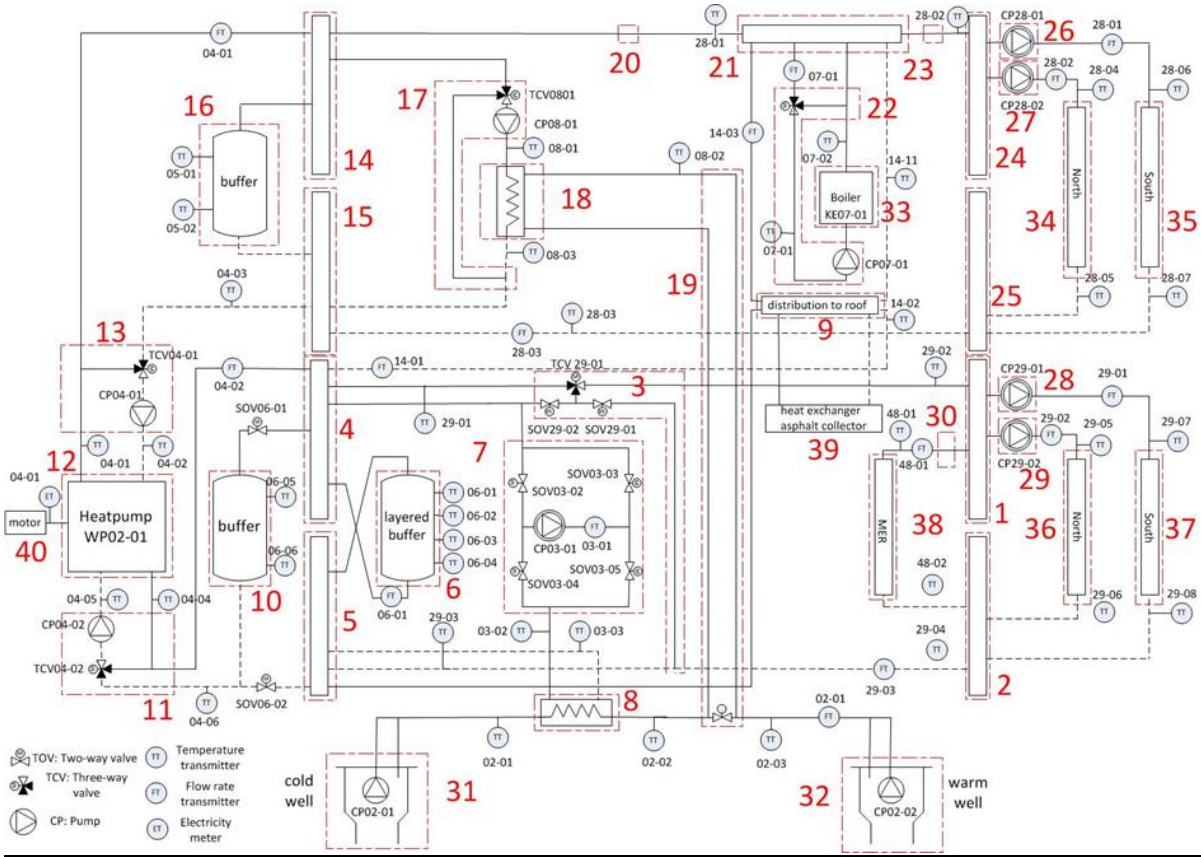


Figure 3.10 Delft THUAS building characteristics. (Syb van Breda, 2009)

The building's HVAC and energy generation system consists of a warm and cold well, heat pump, boiler, roof heat exchanger, PV solar cells and solar heat collectors. The roof heat exchanger is used to de-ice the roof and provide heat to the warm source for balance as it is required by law. A detailed visualization of the building's HVAC system is presented in Figure 3.11.



N.	Component	N.	Component	N.	Component
1	Cooling group header	15	Main return water collector (warm)	29	Pump (cooling North)
2	Return water collector (cooling)	16	Heat storage vessel	30	-

N.	Component	N.	Component	N.	Component
3	Valves	17	Heat regeneration unit	31	Cold well
4	Main cold water header	18	Heat regeneration	32	Warm well
5	Cold water header	19	Heat regeneration	33	Boiler
6	Buffer	20	-	34	Hot water group North
7	Valves/Pumps	21	Boiler Header	35	Hot water group South
8	Heat exchanger ground storage installation	22	Valves/Pumps	36	Cold water group North
9	Distribution system roof	23	-	37	Cold water group South
10	Buffer	24	Heating group header	38	Cold water group server room
11	Evaporator heat pump	25	Return water collector (cooling)	39	Roof
12	Heat pump	26	Pump (heating South)	40	Heat pump motor
13	Valves/Pumps	27	Pump (heating North)		
14	Main hot water header	28	Pump (cooling South)		

Figure 3.11 Delft THUAS HVAC system (Taal, 2021).

3.6.2 Building energy flows

Energy data is gathered from the Priva Building Management System (BMS) and the energy provider dashboard. The Priva BMS gives in depth information of building subsystems and the energy provider dashboard gives information about the total electricity and gas use. Making use of both data sources it was possible to make a yearly energy flow diagram, see Figure 3.12. The difficulty on making the diagram was that there are several gaps on the data collected. The data used is from the period 1-11-2020 to 1-11-2021, that is the period with the best available data quality. The data set was complemented with information about the average electricity production with PV panels of previous years because this information was not registered in the chosen period.

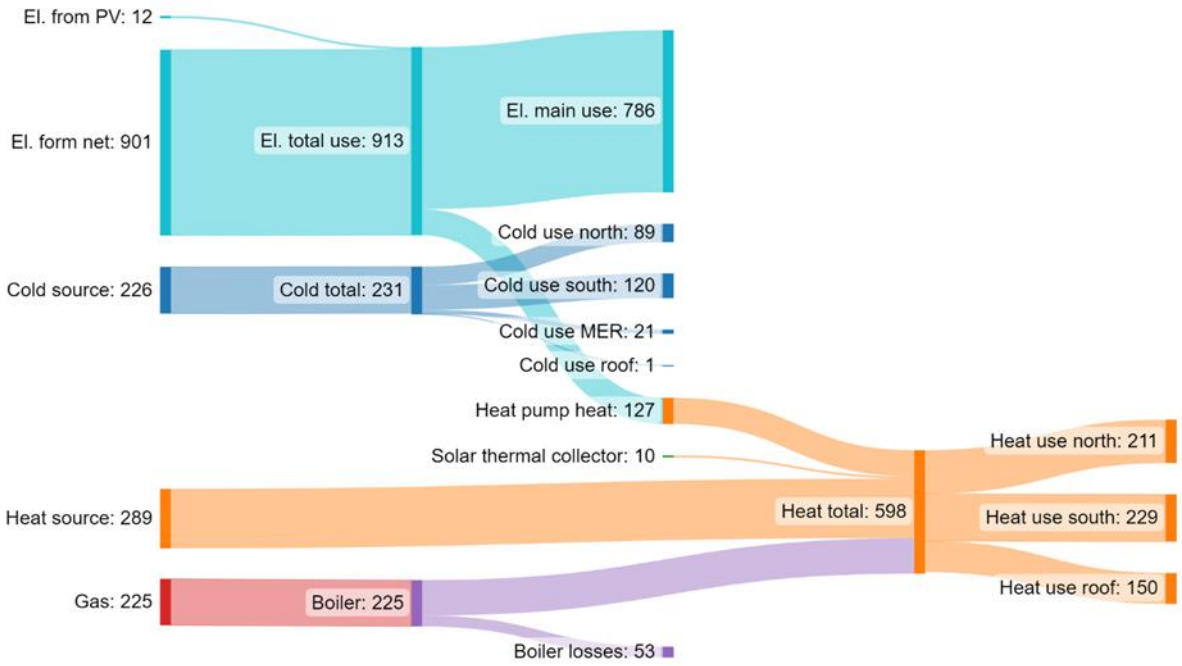


Figure 3.12. Energy flows in Delft THUAS building [MWh]

3.6.3 Conceptual description of the model

The model being developed contains two different building zones, the north and south side. The block used is the Rectangular Zone Template of the IDEAS library. This block allows to define the thermal physical properties of the building space as well as the internal heat gains caused by appliances and by occupancy and the

weather boundary conditions. Each of the zones has heating and cooling delivery via the floor heating system. The block used is the Single circuit Slab of the Buildings library. There is also cooling delivered via the ventilation system. The ventilation system consists of two identical air handling units. Each of the air handling units has a heat recovery system and a heating and cooling heat exchangers, but it has no humidification capabilities. The air handling unit block used is the GenericAHU from the AixLib library. The heating and cooling that is provided to the radiant floors and the air handlings units is generated by a water-water scroll heat pump, the used block is the ScrollWatertoWater from the IDEAS library. There is no block that models the Aquifer Thermal storage among the reviewed libraries. Therefore, the model being developed takes into account a predefined water temperature for the underground hot water source and another for the underground cold water source. A conceptual schematic from the developed model is presented in Figure 3.13.

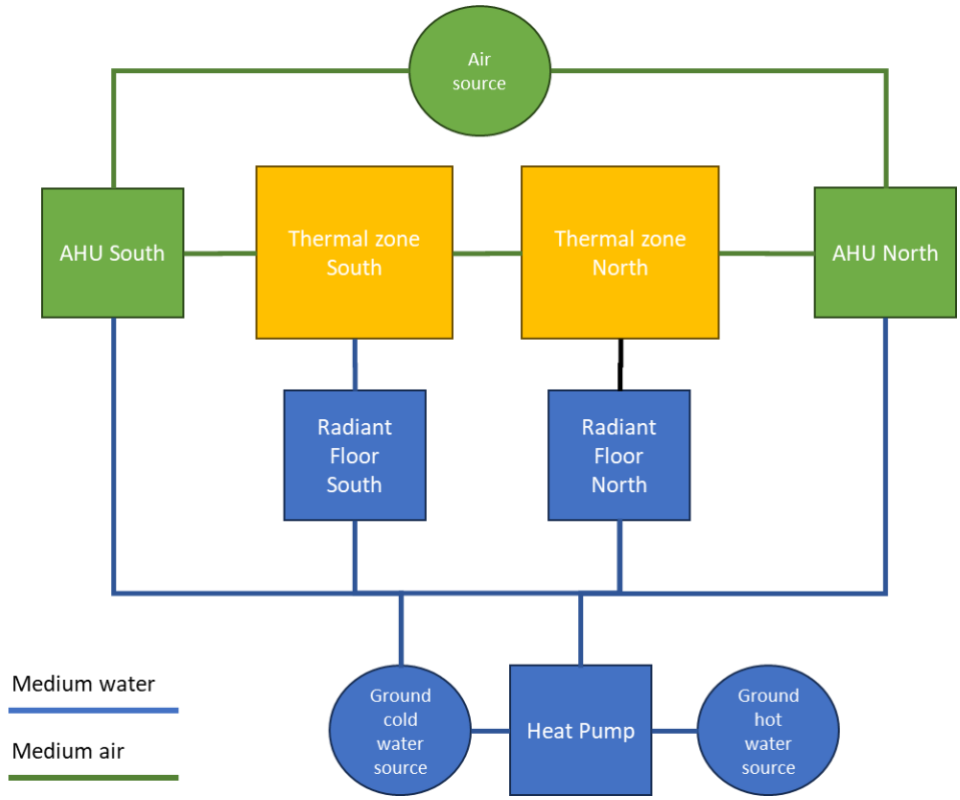


Figure 3.13. Conceptual schematic from the developed model

This model structure simplifies the building installation configuration described in section 3.6.1. However, it will allow to simulate the three most common working modes of the building HVAC system. Heating, heating and cooling simultaneously (heating one thermal zone and cooling the other), and free cooling (without the use of the heat pump).

3.6.4 Conclusions

The conclusions of the second case study have yet to be definitive as there is no computer model working. However, a couple of lessons have been learned on the process of making the model and validating its results with the measured energy flows:

- None of the IBPSA libraries contains all the needed model blocks to simulate a complex utility building. However, there is an active community of developers that continuously make new libraries and the model blocks of these libraries can be integrated in models developed with the IBPSA libraries.
- The possibilities of OpenModelica making use of existing libraries are endless, however when simulating complex systems, the models quickly become computationally demanding. It is necessary to decide in some case what components should be modelled in detail and which can be modelled by other software (like EnergyPlus) or as a blackbox model.

3.7 Conclusions

The possibilities of OpenModelica to model energy systems are endless. However, it has a steep learning curve. There are currently several libraries offering model blocks to simulate the building construction and the building energy systems elements. None of these libraries contains by itself all the model blocks to simulate a complex non-residential building, but with the combination of several libraries is it possible to do so.

Creating building models with OpenModelica to evaluate the energy flexibility of a building has been tested with the case study of a residential house but has not been tested yet with the case study of a non-residential building within this research. The development of the non-residential building model is still under development. In the current research the analysis of the flexibility function and the calculation of the energy flexibility KPIs have been done in Excel, what was laborious and time consuming.

Engineers and designers that evaluated the proposed energy flexibility visualization have addressed the issue that next-generation visualizations should include interaction with specific model parameters. This option will allow to use the tool for design decision making regarding energy flexibility.

Therefore, the next steps in this research are:

1. Finish and validate the model of the second case study.
2. Propose a method to visualize the energy flexibility that allows the interaction with specific building parameters. The initial plan is to use the python library OMpython to run the model with different parameter values and to also process the output in python to produce the visualizations.

The results of the next steps carried out within the Brains for Buildings projects will be published in an updated version of this report.

4 WHITE BOX MODELS: A HYBRID PREDICTIVE DIGITAL TWIN TO PREDICT THE DYNAMIC ENERGY RESPONSE OF A BUILDING AND ITS COMPONENTS

Authors: Wouter Borsboom, Behrouz Eslami Mossallam, Wim Kornaat, Ruud van der Linden (TNO)

4.1 Goal of the hybrid digital twin model

In the built environment, an imbalance exists between the sustainable energy production from solar and wind sources and the energy demand of building systems. Additionally, both the expansion of energy demand and supply are constrained by network congestion. To alleviate congestion and enhance energy balancing, it is crucial to shift energy demand, a concept referred to as "energy flexibility." Achieving energy flexibility requires forecasts for both energy demand and supply.

The research goal in this chapter is to employ a hybrid modelling approach for forecasting the energy demand and supply of buildings, thereby enabling energy flexibility to address balancing and congestion issues. The hybrid predictive digital twin model comprises a white-box model for analyzing heat gains and losses in the building and a black-box occupant model that simulates occupants' interactions with building components. This chapter outlines the white-box building, equipment, and black-box occupant models, all of which are integrated into a simulation and prediction model.

The model's application focuses on the use case of the TNO building at Stieltjesweg in Delft during a summer period, with a specific emphasis on cooling demand.

4.2 Introduction[1]

The hybrid predictive digital twin model consists of a white box building model based on physics which solves the heat flow balance equations, and a data-driven occupant model which models the interaction of the occupants with the building components (thermostats, electric appliances, etc.) and includes the effect of occupants actions in the heat flow balance equations.

The building model consists of a heat balance network that is automatically derived from the Building Information Model (BIM) that describes the geometric configuration and construction properties of the building (consisting of all spaces, walls, windows, doors, roofs, etc.) and the Building Energy Model (BEM) describing the building heating, cooling and ventilation equipment and its controllers. With the automatic generation of the heat network, the simulation model can be easily adapted to different building types, such as apartment buildings, row houses and office buildings. Furthermore, the data used to calibrate the model makes use of the standardized ontology of Haystack. The simulation model is therefore easily scalable for different building typologies and the initialization time is short.

A generic occupant module (framework) has been created which is responsible for reproducing the interaction of the occupants with the building. The occupant module contains distinct submodules, each associated with a certain occupant behavior such as occupancy or interaction with a thermostat setpoint. The implementation is quite flexible, in that each submodule could be connected to a set of various predictive models, ranging from simplistic approaches (e.g., fixed hourly profiles) to complex AI algorithms. Receiving the state of the building at each timestep from the building simulator, along with weather information (Figure 4.1), the occupant module predicts the occupant behavior for the next time step and sends it back to the building simulator. The data driven occupant module, in combination with the physics-based building simulator, makes it a hybrid digital twin.

[1] The text of this paragraph is taken from W A Borsboom et al. 2022 IOP Conf. Ser.: Earth Environ. Sci. 1085 012007

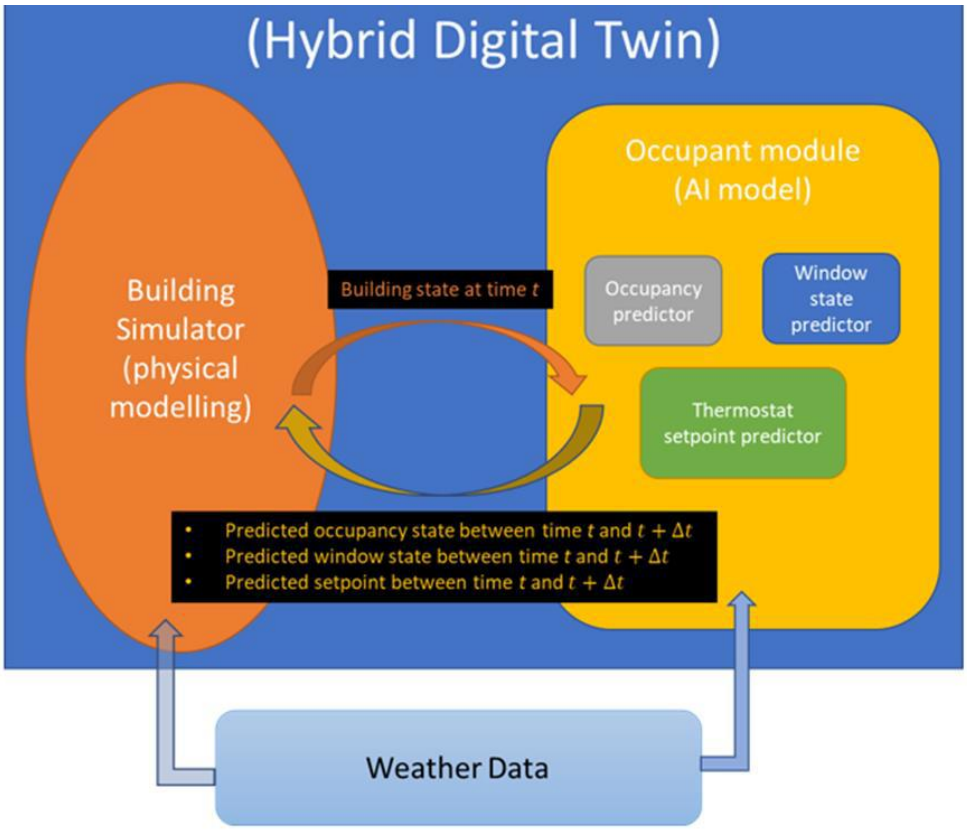


Figure 4.1: Interaction between the occupant model and the building simulation model.

4.3 Building model[2]

The building model is a multizone model. Each zone z_i is represented by a temperature node T_{zi} in the heat network. Each physical layer of boundary surfaces (i.e. walls, floor, ceiling and roofs) constitutes a temperature node in the heat network. For kth layer of the jth boundary surface $S_{j,k}$ ($k = 1$ corresponds to the innermost layer, $k = n$ to the outermost one), a temperature node $T_{Sj,k}$ is added to the heat network. In addition, all boundaries (outdoor environment, ground etc.) are represented by a temperature node. The heat flow balance equations can be summarized as follows:

For the zone z_i:

$$C_{zi}\frac{\partial}{\partial t}T_{zi} = \sum_{S_j \in z_i} A_j h_{\text{int}}^{\text{surf}}(T_{Sj,n} - T_{zi}) + Q_{zi,vent} + Q_{zi,int} + Q_{zi,sol} + Q_{zi,heat/cool} \quad (1)$$

For the innermost surface layer $S_{j,1}$ of the boundary surface S_j which is in direct contact with the zone z_i:

$$C_{Sj,1}\frac{\partial}{\partial t}T_{Sj,1} = A_j h_{\text{int}}^{\text{surf}}(T_{zi} - T_{Sj,1}) + A_j h_{2,1}^{(j)}(T_{Sj,2} - T_{Sj,1}) \quad (2)$$

For the internal layers of the surface S_j:

$$C_{Sj,k}\frac{\partial}{\partial t}T_{Sj,k} = A_j h_{k,k-1}^{(j)}(T_{Sj,k-1} - T_{Sj,k}) + A_j h_{k+1,k}^{(j)}(T_{Sj,k+1} - T_{Sj,k}) \quad (3)$$

For the outermost surface layer $S_{j,n}$ in contact with the outside environment:

$$C_{Sj,n}\frac{\partial}{\partial t}T_{Sj,n} = A_j h_{n,n-1}^{(j)}(T_{Sj,n-1} - T_{Sj,n}) + A_j h_{\text{ext-conv}}^{\text{surf}}(T_{out} - T_{Sj,n}) + A_j h_{\text{ext-rad}}^{\text{surf}}(T_{out} - T_{Sj,n}) + A_j F_{sky_j} h_{\text{ext-rad}}^{\text{surf}}(T_{Sky} - T_{out}) + Q_{Sj,sol} \quad (4)$$

[2] The text of this paragraph is taken from W A Borsboom et al. 2022 IOP Conf. Ser.: Earth Environ. Sci. 1085 012007

The parameters in the above equations are defined as:

- C_{zi}: thermal mass of zone z_i.
- $C_{Sj,\mathrm{k}}$: thermal mass of surface layer $S_{j,k}$ of the boundary surface S_j.
- A_j: the area of the boundary surface S_j.
- $h_{\mathrm{int}}^{\mathrm{surf}}$: the internal surface heat transmission coefficient, including both the convective and radiative transmissions.
- $h_{\mathrm{ext-conv}}^{\mathrm{surf}}$: the external convective surface heat transmission coefficient.
- $h_{\mathrm{ext-rad}}^{\mathrm{surf}}$: the external radiative surface heat transmission coefficient.
- $h_{k+1,k}^{(j)}$: conductive heat transmission coefficient between the kth and $(k+1)$th layers of the boundary surface S_j.
- T_{out}: outdoor temperature.
- T_{Sky}: apparent sky temperature.
- F_{sky_j}: view factor to the sky for the boundary surface S_j.
- $Q_{zi,vent}$: ventilation heat flow for the zone z_i.
- $Q_{zi,int}$: internal heat flow for the zone z_i due to occupants and household appliances.
- $Q_{zi,sol}$: solar heat flow for the zone z_i via the windows.
- $Q_{Sj,sol}$: absorbed solar power by the external boundary surface S_j.
- $Q_{zi,heat/cool}$: heating or cooling flow delivered to zone z_i via the floor heating system.

In addition to the building model, equipment models are developed to predict the ventilation, heating and cooling (e.g. $Q_{zi,vent}$ and $Q_{zi,heat/cool}$ in equation (1)). These models are described in Section 4.5.

4.4 Occupant module

The user behavior models that will be used in the occupant module are described in D3.8 of the B4B project (Eslami Mossalam et al., 2023).

4.5 Equipment models

In this paragraph a description is given of the equipment models that are developed by TNO. The modelled equipment components, in order as described in this paragraph, concern:

- an air handling unit;
- a chiller (only electricity use);
- induction units;
- solar panels (energy yield of PV);
- a buffer tank.

4.5.1 Air Handling Unit

The following generic models for components of an air handling unit are developed:

- a fan model
- a heat/cold recovery model for a heat recovery wheel
- a heating and cooling coil model

For the heating and cooling coils two types of models are developed:

- Type A: Type A is a more simplified model compared to type B with less input needed. In case only the design heating or cooling power is known, it is sufficient to use type A. This model determines if the heating or cooling setpoint can be reached plus the needed heating or cooling power using a Mollier-diagram (hx-diagram). Or better using the formulas from a Mollier-diagram to describe amongst others the dewpoint temperature, the wet bulb temperature, the enthalpy and the absolute moisture content. The valve position is simply calculated assuming a linear relation with the heating or cooling load. Furthermore, it is possible to determine the valve position based upon a given relation between valve position and the heating or cooling load.
- Type B: In this model the heating and cooling coil are modelled in detail. Besides the air temperature (after the coil) also the in/outgoing fluid temperatures of the coil are simulated plus the position of the control valve and the fluid flow. During the simulation an iterative process is performed to determine the

situation in which the heat transfer via the coil, based upon the logarithmic temperature difference over the coil and the transmission value of the coil (conductivity times surface), is in balance with the heat flow to or from the air and the heat flow to or from the fluid.

4.5.2 Fan model

A straightforward and simple model describes the air flow as a fraction of the design flow with the following formula:

flow[m³/h] = ctrl[%] * design_flow[m³/h]

with:

- ctrl[%] = actual rotations per minute(rpm) / maximum rotations per minute (max rpm)
- design_flow = flow at maximum rotations per minute

In combination with the fan characteristics, the fan model can more precisely determine the actual flow and the actual electricity demand depending on the control of the fan. An example of such a fan characteristic is given in the figure below. This concerns the fans used in the air handling unit (AHU) of the TNO office building at the Stieltjesweg. For different fan speeds (I,II,III,IV,V respectively 950,1200,1450,1600 and 1750 rpm) the pressure delivered by the fan (purple lines) and the electric power (blue lines) are given depending on the airflow. In the fan model this information is digitized. The fan model does the following:

1) Depending on the actual control of the fan (rpm/max rpm), the actual pressure curve and actual electric power curve is determined via interpolation of the given fan characteristics.
2) Based upon the design resistance of the air handling unit and the ducting of the air distribution system within the office building, the resistance characteristic of the total system is made. Or in other words: the resistance of the total system as function of the airflow is determined.
3) The intersection of the resistance characteristic (point 2) and the actual pressure curve of the fan (see point 1) is determined which gives the actual airflow that the fan will provide.
4) Based upon the actual airflow from the actual electric power curve (see point 1), the electric power of the fan is determined.

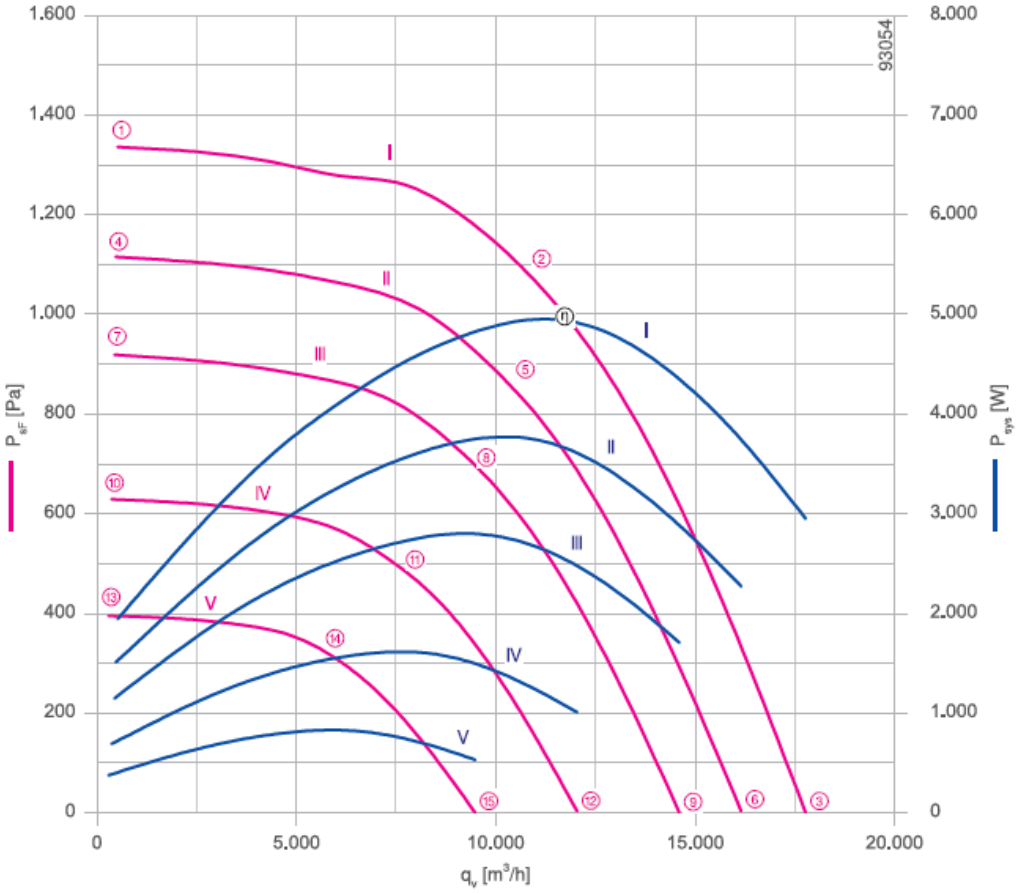


Figure 4.2: Fan characteristics, delivered pressure and electrical power consumption depending on the rotation speed of the fan as function of the airflow.

Specifications

- Fan characteristics: the pressure delivered by the fan and the electric power used by the fan depending on the airflow for different fan speeds (rpm).
- Design resistance of the air handling unit and the ducting of the air distribution with the related design airflow.

Input

- Fan speed, or better, ratio rpm/max rpm

Output

- Airflow delivered by the fan
- Electric power of the fan

4.5.3 Heat/cold recovery model for a heat recovery wheel

The heat/cold recovery model can be used for several heat/cold recovery units, e.g.: a counterflow, crossflow or heat wheel. The output temperature for the model (Thru) is based upon the following formula:

Thru[C]= efficiency * (Texhaust – Tambient)+Tambient

with:

- Efficiency = heat/cold recovery efficiency
- Texhaust = exhaust temperature from the building
- Tambient = outside air temperature

Besides the outgoing temperature (Thru), the model calculates the sensible heating or cooling load. The efficiency of the sensible heat/cold exchange is corrected in case the supply and exhaust flows are not in balance. For an enthalpy or heat wheel at this moment the exchange of moisture is not yet considered.
The model uses the maximum efficiency for a non-controlled heat/cold recovery unit for each moment. In case of a bypass the heat/cold recovery can be switched off. The model can work in two modes for a controlled heat recovery unit, e.g., a heat wheel with controllable speed.

Mode 1:

In this mode the model determines the speed of the heat recovery wheel (hrw) in a way in which the temperature after the hrw approaches the desired setpoint of the supply temperature to the building as much as possible. This can be regarded as ideal control of the hrw. For now, a linear relation is assumed between the speed of the heat recovery wheel and the efficiency. This should be adjusted based upon actual figures. Comparing the calculated temperature after the hrw with the measured value gives insight in how good the actual control is. For testing the performance of the hrw, mode 2 can be used.

Mode 2:

In this mode the efficiency of the heat recovery wheel is calculated depending on the actual speed of the heat recovery wheel at that time in combination with the design efficiency of the heat recovery and corresponding design speed. Based upon this efficiency the temperature after the hrw is simulated. Comparing this temperature with the measured value gives insight into the performance of the hrw. It is noted that the model can also be used in this way in case of a not-controlled heat/cold recovery unit.

Formulas

- Efficiency hru [%] = design efficiency hru [%] * rotation hru [%]/100 * (exhaust flow [m3/h] / supply flow [m3/h]) , limited at 100%, in case of no heat wheel: rotation hru [%]=100
- Rotation hru [%] = (current efficiency hru [%] / maximum efficiency hru [%]) * 100
- Temperature after hru [°C] = efficiency hru [%]/100 * (exhaust temperature [°C] – ambient temperature [°C]) + ambient temperature [°C]

Specifications heat recovery unit

- Thermal design efficiency of the heat/cold recovery [%]

  In case of heat wheel: efficiency [%] with corresponding rotation speed [%]
- Latent design efficiency of an enthalpy unit [%]

  In case of heat wheel: efficiency [%] with corresponding rotation speed [%]

Input

- Ambient temperature [°C]
- Measured exhaust temperature from the building [°C]

- Setpoint supply temperature to building [°C]
- Measured speed heat recovery wheel [percentage of max rpm %]
- Mode HRU [0/1 which corresponds to mode 2/mode 1 respectively]

Output

- Temperature after heat/cold recovery unit [°C]
- Cooling/heating power heat recovery unit [kW]

4.5.4 Heating and cooling coil model type A

This model determines to what extend the heating or cooling setpoint can be reached plus the needed heating or cooling power using a Mollier-diagram (hx-diagram). Or better using the formulas from a Mollier-diagram to describe amongst others the dewpoint temperature, the wet bulb temperature, the enthalpy and the absolute moisture content. This model can work in 2 modes.

Mode 1:

In this mode the model determines the heating/cooling power in such a way that the supply temperature to the building reaches the desired setpoint or approaches it as good as possible. This can be regarded as an ideal control. Depending on the setpoint and the air temperature before the coil (ambient temperature or temperature after a heat recovery unit) it is determined whether heating of cooling is needed. Taking into account the maximum available heating/cooling power, the model determines if the setpoint can be reached or not and when the setpoint can be reached how much heating/cooling power actually is needed. In case of cooling it is also considered whether or not dehumidification will occur. The end outcome is the sensible heating/cooling power, latent cooling power, the supply temperature to the building, the absolute moisture content of the supply air and the position of the valve of the coil. The valve position is calculated assuming a linear relation with the heating or cooling load. Furthermore, it is possible to determine the valve position based upon a given relation between valve position and the heating or cooling load (e.g. determined upon measured valve position and simulated load over a period in the past).

In case the measured supply temperature/moisture content deviates from the simulated value this can indicate that the control is not working properly or that the actual heating/cooling power deviates from the design.

Mode 2:

In this mode the measured position of the valve of the coil is used to determine the heating/cooling power. For now, a linear relation between valve position and power is assumed. Besides that it is also possible to use a given relation between valve position and the heating or cooling load (e.g. determined upon measured valve position and simulated load over a period in the past). Depending on the air temperature before the coil (ambient temperature or temperature after the heat recovery unit) and the setpoint it is determined whether heating or cooling is needed. Based upon the current heating/cooling power (depending on the valve position), the supply temperature to the building and the absolute moisture content of the supply are calculated. In case of cooling possible condensation is considered. Deviation between the measured and simulated supply air temperature and the moisture content gives insight into the performance of the coil (due to pollution) or the available heating/cooling power compared to the design.

Formulas

Heating power [kW] = design heating power [kW] * heating valve position [%]/100

Cooling power [kW] = design cooling power [kW] * cooling valve position [%]/100

Latent cooling power [kW] = airmass flow [kg/s] * (enthalpy out[kJ/kg] – enthalpy in[kJ/kg])

Sensible cooling power [kW] = cooling power[kW] – latent cooling power [kW]

Heating capacity sufficient to reach Tset:

Supply air temperature [°C] = Tset [°C]

Moisture content supply air [kg/kg] = ingoing moisture content [kg/kg]

Heating capacity insufficient to reach Tset:

Supply air temperature [°C] = air enthalpy in case of maximum heating capacity[kJ/kg] / (specific heat air[kJ/kg.K] + specific heat damp[kJ/kg.K] * moisture content supply air[kg/kg])

Moisture content supply air [kg/kg] = ingoing moisture content [kg/kg]

Cooling capacity sufficient to reach Tset and no condensation occurs:

Supply air temperature [°C] = Tset [°C]

Moisture content supply air [kg/kg] = ingoing moisture content [kg/kg]

Cooling capacity insufficient to reach Tset and no condensation occurs:

Supply air temperature [°C] = air enthalpy in case of maximum cooling capacity[kJ/kg] / (specific heat air[kJ/kg.K] + specific heat damp[kJ/kg.K] * absolute moisture content supply air[kg/kg])

Moisture content supply air [kg/kg] = ingoing moisture content [kg/kg]

Cooling capacity sufficient to reach Tset and condensation occurs:

Supply air temperature [°C] = Tset [°C]

Relative humidy supply air [%] =100 [%]

Moisture content supply air [kg/kg] = 6.112*2.71828^((17.67*Tset)/(Tset+243.5))*RH*18.02/((273.15+Tset)*100*0.08314)/1.2/1000

Cooling capacity insufficient to reach Tset and condensation occurs:

Supply air temperature [°C] = wetbulb temperature based upon air enthalpy in case of maximum cooling capacity [°C]

Moisture content supply air [kg/kg] = moisture enthalpy in case of maximum cooling capacity [kJ/kg] / (evaporation heat water [kJ/kg] + specific heat damp [kJ/kg.K] * supply air temperature)

Specifications heating/cooling coil

- Maximum heating power generation [kW]
- Maximum cooling power generation [kW]
- Maximum supply flow [m3/h]
- Maximum exhaust flow [m3/h]

Input

- Ambient temperature [°C]/ temperature after heat recovery unit[°C]
- Ambient absolute moisture content [kg/kg]
- Measured status Air Handling Unit [0/1, off/on]
- Measured supply fan speed [%]
- Measured exhaust fan speed [%]
- Setpoint supply temperature to building [°C]
- Measured position valve coil [%]
- Calc_LBK [0/1, mode 2/mode 1]

Output

- Simulated supply temperature to building [°C]
- Simulated absolute moisture content supply flow [kg/kg]
- Simulated sensible cooling/heating power coil [kW]
- Simulated latent cooling power coil [kW]

4.5.5 Heating and cooling coil model type B

In model type B the heating and cooling coil are modelled in detail. Besides the air temperature (after the coil) also the in/outgoing fluid temperatures of the coil are simulated plus the position of the control valve and the fluid flow. For that purpose, detailed information about the design is needed. Such as the design in/outgoing air temperatures, the design airflow, the design in/outgoing fluid temperatures and the design fluid flow. For

the air handling unit in the Stieltjesweg this information was easily obtained from the supplier of this unit based upon the name shield on the unit.

During the simulation an iterative process is performed to determine the situation in which:

- the heat transfer via the coil, based upon the logarithmic temperature difference over the coil and the transmission value of the coil (conductivity times surface), is in balance with
- the heat flow to or from the air and
- the heat flow to or from the fluid.

For the heating coil, in this way the air temperature (after the coil) and the in/outgoing fluid temperatures are determined. The heating fluid circuit is normally a mixing system. Thus, the fluid flow is not adjusted.

For the cooling coil, in this way the air temperature (after the coil), outgoing fluid temperature and the fluid flow are determined. The ingoing fluid temperature is normally constant while the capacity of the cooling system is controlled by the flow over the coil.

Based upon the simulated air temperatures after the coil (see above), the delivered heating/cooling power is known and from that the valve positions are determined. The valve positions are calculated assuming a linear relation with mixing percentage (heating) or splitting percentage (cooling). Based upon measured data of the valve positions and simulated heating/cooling load by the model, it should be checked to what extent a linear relation is correct. Otherwise, a fit between the measured valve position and simulated heating/cooling load should be made and used. For the cooling coil also situations with dehumidification are considered in the heat balance and the latent cooling power is calculated.

In case of deviations with the measured fluid and air temperatures and valve positions, these can be used to gain insight in the performance of the air handling unit.

Notice that the results of this model will depend on the accuracy of the information about the design and whether or not these design values are in line with the actual practice. Furthermore, assumptions need to be made about the heat transfer through the coil depending on the air and water flows. Right assumptions can be hard to make but are crucial for correct results.

Formulas

Heating/cooling power coil [kW] = conductivity fluid/air [kW/m2.K] * surface coil [m2] * logarithmic temperature difference [K]

DT1 = fluid ingoing temperature [°C] – air outgoing temperature [°C]

DT2 = fluid outgoing temperature [°C] – air ingoing temperature [°C]

Logarithmic temperature difference [K] = (DT1 – DT2) / log(DT1/DT2)

Heating/sensible cooling power air[kW] = capacity airflow [kW/K] * (air outgoing temperature [°C] – air ingoing temperature [°C])

Heating/sensible cooling power fluid[kW] = capacity fluid flow [kW/K] * (fluid ingoing temperature [°C] – fluid outgoing temperature [°C])

Heating valve position [%]= (fluid ingoing temperature [°C] - fluid outgoing temperature [°C]) / (heating system temperature [°C] - fluid outgoing temperature [°C]) * 100

Heating valve position [%]= (fluid ingoing temperature [°C] - fluid outgoing temperature [°C]) / (heating system temperature [°C] - fluid outgoing temperature [°C]) * 100

Specifications heating/cooling coil

- Maximum supply flow [m3/h]
- Maximum exhaust flow [m3/h]
- Temperature incoming air heater coil according design specifications [°C]
- Temperature outgoing air heater coil according design specifications [°C]
- Temperature incoming air cooler coil according design specifications [°C]
- Temperature outgoing air cooler coil according design specifications [°C]
- Maximum heating power delivered by heater coil according design specifications [kW]

- Liquid flow heater coil according design specifications [m3/h]
- Temperature of fluid to heater coil according design specifications [°C]
- Temperature of fluid from heater coil according design specifications [°C]
- Volume percentage glycol in heating system according design specifications [%]
- Maximum cooling power delivered by cooler coil according design specifications [kW]
- Liquid flow cooler coil according design specifications [m3/h]
- Temperature of fluid to cooler coil according design specifications [°C]
- Temperature of fluid from cooler coil according to design specifications [°C]
- Volume percentage glycol in cooling system according design specifications [%]
- Relative humidity after cooler coil in case of condensation [%]

Input

- Ambient temperature [°C] / temperature after heat recovery unit[°C]
- Ambient absolute moisture content [kg/kg]
- Measured status Air Handling Unit [0/1, off/on]
- Measured supply fan speed [%]
- Measured exhaust fan speed [%]
- Setpoint supply temperature to building [°C]
- Measured position valve heating/cooling coil [%]

Output

- Simulated supply temperature to building [°C]
- Simulated absolute moisture content supply flow [kg/kg]
- Simulated sensible cooling/heating power coil [kW]
- Simulated latent cooling power coil [kW]
- Simulated temperature of fluid to heater coil [°C]
- Simulated temperature of fluid from heater coil [°C]
- Simulated flow heater coil [m3/h]
- Simulated valve position heater coil [%]
- Simulated temperature of fluid to cooler coil [°C]
- Simulated temperature of fluid from cooler coil [°C]
- Simulated flow cooler coil [m3/h]
- Simulated valve position cooler coil [%]

4.5.6 Chiller model

The model consists of a chiller plus a dry cooler. The dry cooler removes the heat, subtracted from the building by the chiller, to the outside. The drycooler is a big heat exchanger from which the heat is removed with forced air. The aim of the model is to predict the electricity use of the compressor of the chiller and the fans of the drycooler based upon the actual cooling demand of the building. The electricity use of the circulation pumps on evaporator and condenser side of the chiller are not considered and are expected to be neglectable.

The efficiency of the chiller depends on the temperature lift between evaporator and condenser side. If the evaporator temperature drops or the condenser temperature increases the efficiency decreases. In general the model works as follows:

- assumptions are made about the expected temperature lift (which can vary in time depending on the actual situation),
- the efficiency for this temperature lift is determined based upon the specification of the chiller,
- the electricity use of the chiller now can be simulated based upon the efficiency and the actual cooling load of the building (load on evaporator side),
- the load on condenser side is simulated based upon the efficiency and the actual cooling load (or evaporator load),
- the airflow for the drycooler is simulated based upon the quotient of nominal airflow and nominal heat removal times the condenser load,
- the electricity use of the fans of the drycooler is simulated based upon the nominal electricity use times the quotient of the actual airflow and nominal airflow to the power 2.5 (see explanation below).

In the TNO Stieltjesweg building the chilled water circuit is controlled at 12/6°C. The temperatures on condenser side are for a large part controlled by the drycooler. The drycooler has 10 fans. The control of the glycol system of the drycooler is done with a bypass and speed-controlled pump. If this control is clear, it can be used in the model. However, in this case the control of the drycooler, as well as the chiller, is autonomously controlled by Carrier. Therefore based upon historical BMS data, the condenser ingoing temperature (outgoing temperature of drycooler) is described depending on the ambient temperature. This relation is shown in the figure below. This relation seems reasonable because the outside air temperature in relation to the control of the drycooler will determine the condensers ingoing temperature.

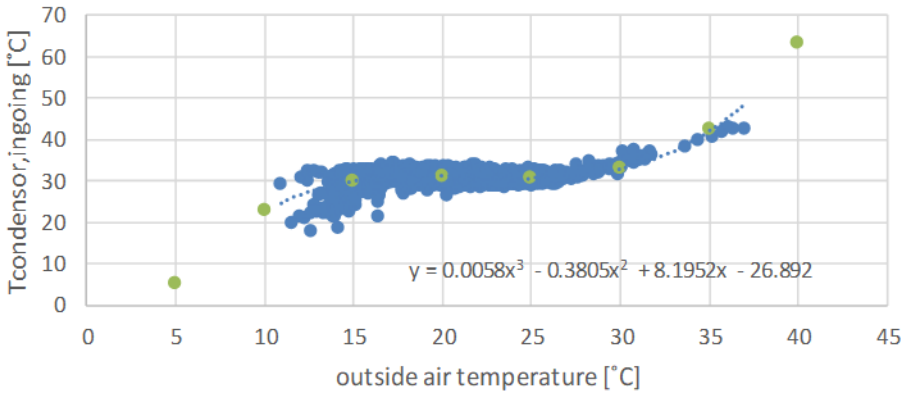


Figure 4.3: Condenser ingoing temperature depending on outside air temperature for the TNO Stieltjesweg building.

The chiller is a Carrier chiller type 30XW-0402, equipped with one compressor. The compressor can be controlled from 100-15%. Based upon documents of the chiller the energy efficiency ratio (EER=Energy Efficiency Ratio) is known for two situations (see table below). Those two situations mainly differ concerning the in- and outgoing temperature at condenser side (temperature from and to the drycooler respectively). On evaporator side the temperatures are controlled at 12/6°C (see before).

Table 4.1: Specifications Carrier chiller type 30XW-0402.

Situation	1	2
Tevaporator[°C], ingoing	12	12
Tevaporator[°C], outgoing	7	6
Tcondenser[°C], ingoing	30	41
Tcondenser[°C], outgoing	35	47
Pevaporator[kW]	461	380
Pcondenser[kW]	546	496
Pcompressor[kWe]	85	116
EER[-]	5.42	3.27
ESEER[-]	6.28	5.77
COP[-]	6.42	4.24
Flow evaporator[dm³/s]	-	15.1
Flow condenser[dm³/s]	-	20.99

The make of the drycooler is Gunter type GFHV FD 090.2NF.CAR. The known specifications are shown below.

Table 4.2: Specifications Gunter drycooler.

Nominal heat removal [kW]	494
Nominal electric fan power [kWe]	4.7
Nominal airflow [m3/h]	121325

To determine the electric use of the compressor and the fans of the drycooler, the following steps are incorporated in the model:

1) Calculate the expected ingoing condenser temperature depending on the actual outside air temperature according to the function given in Figure 4.3.
2) Calculate the EER for this actual situation via linear interpolation between situation 1 and 2 based upon the actual ingoing condenser temperature.

   EER= EER1 – { (EER1-EER2)/(Tcondenser1,ingoing-Tcondenser2,ingoing)*(Tcondenser1,ingoing-Tcondenser,actual) }

   EER= 5.42 – { (5.42-3.27)/(30-41)*(30-Tcondenser,actual) }
3) Calculate the electricity power of the compressor using the simulated cooling demand and the EER according to point 2.

   Ecompressor= simulated cooling demand / EER [kWe]
4) Calculate the heating power that needs to be removed via the condenser (or in other words via the drycooler)

   Heat removal condenser/drycooler = (EER+1)* Ecompressor [kW]
5) Calculate the airflow of the drycooler assuming a linear relation with the heat removal.

   Airflow_drycooler= Nominal airflow drycooler * (heat removal condenser/ nominal heat removal drycooler)
6) Calculate the electric power of the fans of the drycooler assuming this is relative to the airflow to the power 2.5. In theory the transport energy is linear with the airflow and the resistance is relative to the second power of the airflow, thus in total relative to the third power of the airflow. But due to changes in efficiency, this is assumed to be limited to 2.5.

   Edrycooler= nominal electric fan power drycooler * (airflow_drycooler/ nominal airflow drycooler)$^{2.5}$

4.5.7 Induction units

The model for induction units requires two inputs: the room temperature (T_{room}) and the local setpoint temperature (T_{set}). The generated cooling is then calculated based on the position of the cooling valve in percentage (f_{valve}) and the maximum cooling power of the equipment ($Q^{\text{Max}}_{\text{cool, induction unit}}$) as given below:

$$Q_{\text{cool, induction unit}} = 0.01 \times f_{\text{valve}} \times Q^{\text{Max}}_{\text{cool, induction unit}} \quad (5)$$

The valve position depends on the difference between the room and the setpoint temperatures. Inspection of the measured data from the TNO Stieltjesweg building revealed that, irrespective of the type of the induction unit, the cooling valve position follows a well-defined curve (Figure 4.4):

$$\begin{aligned} f_{\text{valve}} &= 0 && if\ T_{room} - T_{set} < 0.3 \\ f_{\text{valve}} &= \frac{(T_{\text{room}} - T_{set} - 0.3)}{0.02} && if\ 0.3 \le T_{room} - T_{set} < 2.3 \\ f_{\text{valve}} &= 100 && if\ T_{room} - T_{set} \ge 2.3 \end{aligned} \quad (6)$$

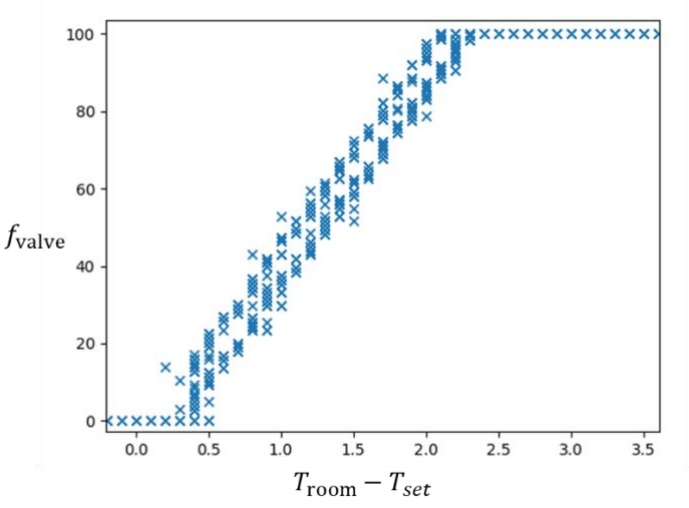


Figure 4.4: The position of the cooling valve as a function of the difference between the room temperature and setpoint temperature for the induction units of the TNO Stieltjesweg building.

The maximum power depends on the design specification of the equipment. For the units installed at the TNO Stieltjesweg building, the value is extracted from the table shown in Figure 4.5. These tables are valid for the case where the temperature of primary air is 16°C and the inlet water temperature is 15°C. Based on the measured data during summer 2022, both conditions are met for the Stieltjesweg building. In the digital twin of the TNO Stieltjesweg building, all individual units are modelled separately, assuming the water flow of 100 L/h. A data-driven method is used to predict the local temperature setpoint in each room, as described in D3.8 of the B4B project (Eslami Mossallam, Spiekman, and Hoes 2023). The room temperature for each induction unit is assumed to be equal to the temperature of the zone which the room belongs to, and is predicted by the building model in each simulation timestep. The position of the cooling valve for all the units in each room is then calculated from equation (6). The maximum cooling power of each unit is extracted from the table in Figure 4.5, based on the difference between the instantaneous room temperature and the inlet water temperature. Finally, equation (5) is used to calculate the generated cooling power of each unit and the output of all units in each zone is summed to obtain the total cooling per zone.

LUCHT						WATER												Snel selectie*	
Primair			Koelvermogen Lucht $t_{ruimte}-t_{pri}$ °C					Koelvermogen water $t_{ruimte}-t_{water\ in}$ °C										L_9 W_9	L_9 W_{10}
			8	9	10			6		7		8		9		10		11			
Vprim	Ps	Lw	Q_l	Q_l	Q_l	V_w	ΔP_w	Q_{wk}	Δt_w	Q_{wk}	Δt_w	Q_{wk}	Δt_w	Q_{wk}	Δt_w	Q_{wk}	Δt_w	Q_{wk}	Δt_w	Q_t	Q_t
Nozzle A 1																					
l/s	m³/h	Pa	dB(A)	W_8	W_9	W_{10}	l/h	kPa	W_6	°C	W_7	°C	W_8	°C	W_9	°C	W_{10}	°C	W_{11}	°C	$W_{9,9}$	$W_{9,10}$
6,9	25	66	-	67	76	84	50	1,0	132	2,3	154	2,7	176	3,0	198	3,4	220	3,8	242	4,2	274	296
							70	1,9	143	1,7	167	2,0	190	2,3	214	2,6	238	2,9	262	3,2	290	314
							100	3,6	152	1,3	178	1,5	203	1,8	229	2,0	254	2,2	279	2,4	305	330
							140	6,8	160	1,0	186	1,1	213	1,3	239	1,4	266	1,6	293	1,8	315	342
8,3	30	94	-	81	91	101	50	1,0	149	2,6	174	3,0	198	3,4	223	3,9	248	4,3	273	4,7	314	339
							70	1,9	163	2,0	190	2,3	217	2,6	244	3,0	271	3,3	298	3,6	335	362
							100	3,6	175	1,5	204	1,8	234	2,0	263	2,2	292	2,5	321	2,8	354	383
							140	6,8	185	1,1	216	1,3	246	1,5	277	1,7	308	1,9	339	2,1	368	399
9,7	35	128	16	94	105	117	50	1,0	163	2,8	190	3,3	217	3,8	244	4,2	271	4,7	298	5,2	349	376
							70	1,9	180	2,2	210	2,6	240	3,0	270	3,3	300	3,7	330	4,1	375	405
							100	3,6	196	1,7	228	2,0	261	2,2	293	2,5	326	2,8	359	3,1	398	431
							140	6,8	208	1,3	243	1,5	278	1,7	312	1,9	347	2,1	382	2,3	417	452
11,1	40	168	20	107	121	134	50	1,0	175	3,0	204	3,5	234	4,0	263	4,5	292	5,0	321	5,5	384	413
							70	1,9	196	2,4	228	2,8	261	3,2	293	3,6	326	4,0	359	4,4	414	447
							100	3,6	215	1,9	251	2,2	286	2,5	322	2,8	358	3,1	394	3,4	443	479
							140	6,8	229	1,4	267	1,6	306	1,8	344	2,1	382	2,3	420	2,5	465	503
12,5	45	212	23	121	136	151	50	1,0	186	3,2	217	3,7	248	4,2	279	4,8	310	5,3	341	5,8	415	446
							70	1,9	209	2,6	244	3,0	279	3,4	314	3,9	349	4,3	384	4,7	450	485
							100	3,6	232	2,0	270	2,3	309	2,6	347	3,0	386	3,3	425	3,6	483	522
							140	6,8	250	1,6	291	1,8	333	2,1	374	2,3	416	2,6	458	2,9	510	552

Figure 4.5: Design specifications of induction units installed at the TNO Stieltjesweg building. (the table shows the specifications for units with nozzle type A1. Similar tables are provided for other nozzle types).

4.5.8 Dynamic prediction PV supply

A photovoltaic array performance model, developed at Sandia National Laboratories (D L King, Boyson, and Kratochvil 2004) and implemented in PVLib library in MATLAB (Stein et al. 2016), is used to predict the power output of PV panels. In this model the IV curve of a PV module is determined by five characteristic points (Figure 4.6) which are calculated via the equations (7) - (16).

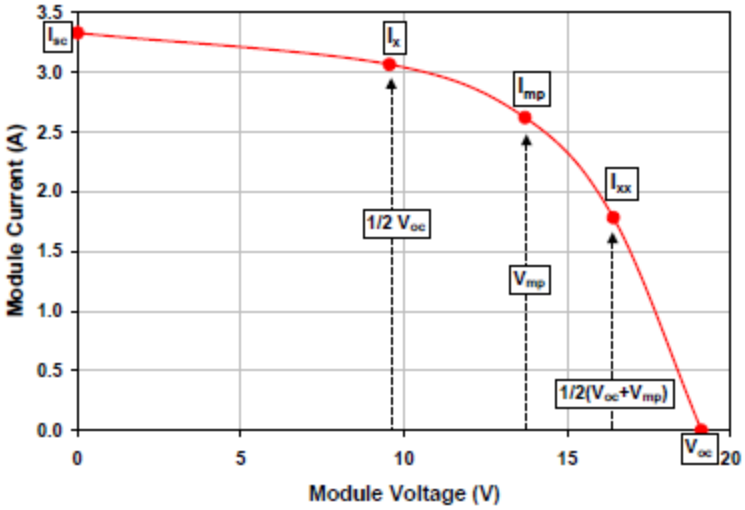


Figure 4.6: A schematic of the IV curve of a PV module, determined by five characteristic points in the Sandia model.

$$\delta(T_c) \equiv n \cdot k \cdot \frac{(T_c + 273.15)}{q} \tag{7}$$

$$E_e \equiv \frac{f_1(AM) \cdot \left(E_b \cdot f_2(AOI) + f_d \cdot E_{diff}\right)}{E_o} \tag{8}$$

$$I_{sc} = I_{sco} \cdot E_e \cdot \left(1 + \alpha_{Isc} \cdot (T_c - T_o)\right) \tag{9}$$

$$I_{mp} = I_{mpo} \cdot (C_0 \cdot E_e + C_1 \cdot E_e^2) \cdot \left(1 + \alpha_{Imp} \cdot (T_c - T_o)\right) \tag{10}$$

$$V_{oc} = V_{oco} + N_s \cdot \delta(T_c) \cdot \ln(E_e) + \beta_{Voc}(E_e) \cdot (T_c - T_o) \tag{11}$$

$$V_{mp} = V_{mpo} + C_2 \cdot N_s \cdot \delta(T_c) \cdot \ln(E_e) + C_3 \cdot N_s \cdot (\delta(T_c) \cdot \ln(E_e))^2 + \beta_{Vmp}(E_e) \cdot (T_c - T_o) \tag{12}$$

$$P_{mp} = I_{mp} \cdot V_{mp} \tag{13}$$

$$FF = \frac{P_{mp}}{I_{sc} \cdot V_{oc}} \tag{14}$$

$$I_x = I_{xo} \cdot (C_4 \cdot E_e + C_5 \cdot E_e^2) \cdot \left(1 + \alpha_{Isc} \cdot (T_c - T_o)\right) \tag{15}$$

$$I_{xx} = I_{xxo} \cdot (C_6 \cdot E_e + C_7 \cdot E_e^2) \cdot \left(1 + \alpha_{Imp} \cdot (T_c - T_o)\right) \tag{16}$$

The definition of the model variables and parameters are given in Table 4.3. The model has been parametrized for a wide range of PV systems, and the resultant parameters are stored in Sandia Modules Database[3].

Table 4.3: Definition of variables and parameters in the Sandia PV performance model.

Variable / Parameter	Definition
$\delta(T_c)$	'Thermal voltage' per cell at temperature Tc. For diode factor of unity (n=1) and a cell temperature of 25°C, the thermal voltage is about 26 mV per cell.
n	Empirically determined 'diode factor' associated with individual cells in the module, with a value typically near unity, (dimensionless). It is determined using measurements of open circuit voltage (Voc, see Figure 4.6) translated to a common temperature and plotted versus the natural logarithm of effective irradiance. This relationship is typically linear over a wide range of irradiance.
k	Boltzmann's constant, $1.38066 \times 10^{-23} \frac{J}{K}$.

[3] https://github.com/NREL/SAM/blob/develop/deploy/libraries/Sandia%20Modules.csv

Variable / Parameter	Definition
T_c	Cell temperature inside module (°C). Calculated from equations (11) and (12) in (D L King, Boyson, and Kratochvil 2004)
E_e	The 'effective' solar irradiance as previously defined by Equation (7). This value describes the fraction of the total solar irradiance incident on the module to which the cells inside actually respond.
AM	Absolute air mass, (dimensionless). This value is calculated from sun elevation angle and site altitude, and it provides a relative measure of the path length the sun must travel through the atmosphere, AMa=1 at sea level when the sun is directly overhead.
$f_1(AM)$	Empirically determined polynomial relating the solar spectral influence on Isc to air mass variation over the day.
AOI	Solar angle-of-incidence, (degrees). AOI is the angle between a line perpendicular (normal) to the module surface and the beam component of sunlight.
$f_2(AOI)$	Empirically determined polynomial relating optical influences on Isc to solar angle-of-incidence (AOI).
f_d	Fraction of diffuse irradiance used by module, typically assumed to be 1 for flat-plate modules. For point-focus concentrator modules, a value of zero is typically assumed, and for low-concentration modules a value between zero and 1 can be determined.
E_b	beam component of solar irradiance incident on the module surface, (W/m2).
E_{diff}	Diffuse component of solar irradiance incident on the module surface, (W/m2).
E_o	Reference solar irradiance, typically 1000 W/m2, with ASTM standard spectrum.
I_{sc}	Short-circuit current (A).
I_{sco}	Short-circuit current at the standard reference condition (A).
α_{Isc}	Normalized temperature coefficient for Isc, (1/°C). This parameter is 'normalized' by dividing the temperature dependence (A/°C) measured for a particular standard solar spectrum and irradiance level by the module short-circuit current at the standard reference condition, Isco.
T_o	Reference cell temperature, typically 25°C.
I_{mp}	Current at the maximum-power point (A).
I_{mpo}	Current at the maximum-power point at the standard reference condition (A).
C_0, C_1	Empirically determined coefficients relating Imp to effective irradiance, Ee. C0+C1 = 1, (dimensionless).
α_{Imp}	Normalized temperature coefficient for Imp, (1/°C). Normalized in the same manner as αIsc.
V_{oc}	Open-circuit voltage (V).
V_{oco}	Open-circuit voltage at the standard reference condition (V).
N_s	Number of cells in series in a module's cell-string.
$\beta_{Voc}(E_e)$	Temperature coefficient for module open-circuit-voltage as a function of the effective irradiance, Ee. Usually, the irradiance dependence can be neglected and βVoc is assumed to be a constant value.
V_{mp}	Voltage at maximum-power point (V).
V_{mpo}	Voltage at maximum-power point at the standard reference condition (V).
C_2, C_3	Empirically determined coefficients relating Vmp to effective irradiance (C2 is dimensionless, and C3 has units of 1/V)
$\beta_{Vmp}(E_e)$	Temperature coefficient for module maximum-power-voltage as a function of effective irradiance, E_e. Usually, the irradiance dependence can be neglected and β_{Vmp} is assumed to be a constant value.
P_{mp}	Power at maximum-power point (W)
FF	Fill Factor (dimensionless)
I_x	Current at module V = 0.5·V_{oc}, defines 4th point on I-V curve for modeling curve shape.
I_{xo}	I_x at the standard reference condition (A).
C_4, C_5	Empirically determined coefficients relating the current (I_x), to effective irradiance, E_e. C_4+C_5 = 1, (dimensionless).

Variable / Parameter	Definition
I_{xx}	Current at module V = 0.5·(Voc +Vmp), defines 5th point on I-V curve for modeling curve shape (A).
I_{xxo}	I_{xx} at the standard reference condition (A).
C_6, C_7	Empirically determined coefficients relating the current (Ixx) to effective irradiance, Ee. C6+C7 = 1, (dimensionless).

To calculate the resulting AC power output from the DC power produced by the PV it is assumed that the losses between the PV array and the max power point tracker are negligible, and the inverter (with MPPT capability) is able to perfectly track the max power point (MPP), so the module operates at the max power point at all times. Then the Sandia performance model for inverters is used for the calculation (David L. King et al. 2007). The schematic characterization of the model is shown in Figure 4.7. The AC power depends both on DC power and DC voltage and is calculated via equations (17) - (20).

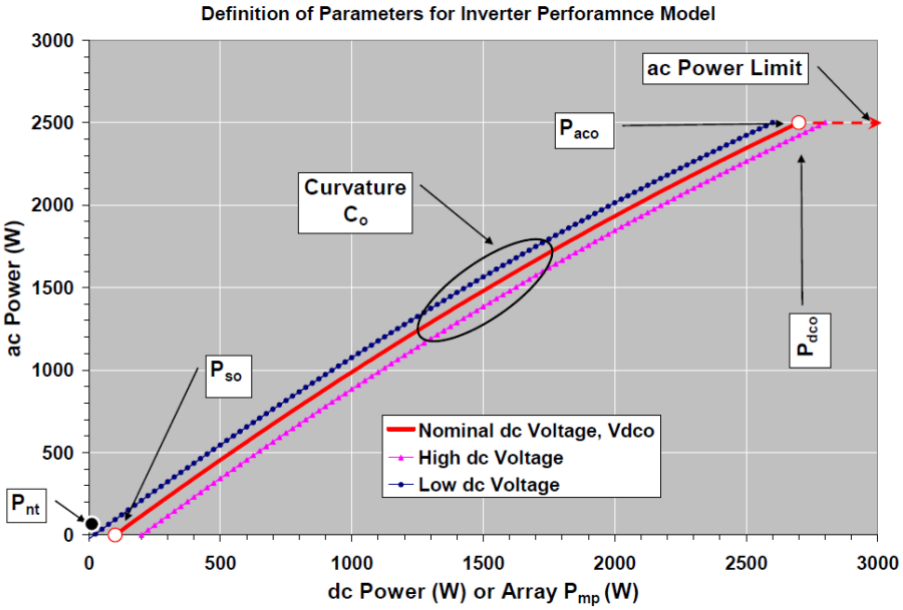


Figure 4.7: Illustration of the inverter performance model and the factors describing the relationship of ac-power output to both dc-power and dc-voltage input.

$$P_{ac} = \left(\frac{P_{aco}}{A-B} - C\cdot(A-B)\right)\cdot(P_{dc}-B) + C\cdot(P_{dc}-B)^2 \quad (17)$$

$$A = P_{dco}\cdot\left(1 + C_1\cdot(V_{dc}-V_{dco})\right) \quad (18)$$

$$B = P_{so}\cdot\left(1 + C_2\cdot(V_{dc}-V_{dco})\right) \quad (19)$$

$$C = C_o\cdot\left(1 + C_3\cdot(V_{dc}-V_{dco})\right) \quad (20)$$

The definition of variables and parameters for the inverter performance model are given in Table 4.4. The specifications of inverter systems is taken from the inverter database, maintained by NREL System Advisor Model[4].

Table 4.4: Definition of variables and parameters in the Sandia inverter performance model.

Variable / Parameter	Definition
P_{ac}	ac-power output from inverter based on input power and voltage, (W)
P_{aco}	maximum ac-power "rating" for inverter at reference or nominal operating condition, assumed to be an upper limit value, (W)
P_{dc}	dc-power input to inverter, typically assumed to be equal to the PV array maximum power, (W)

[4] https://sam.nrel.gov/

Variable / Parameter	Definition
V_{dc}	dc-voltage input, typically assumed to be equal to the PV array maximum power voltage, (V)
V_{dco}	dc-voltage level at which the ac-power rating is achieved at the reference operating condition, (V)
P_{dco}	dc-power level at which the ac-power rating is achieved at the reference operating condition, (W)
P_{so}	dc-power required to start the inversion process, or self-consumption by inverter, strongly influences inverter efficiency at low power levels, (W)
P_{nt}	ac-power consumed by inverter at night (night tare) to maintain circuitry required to sense PV array voltage, (W)
C_0	parameter defining the curvature (parabolic) of the relationship between ac-power and dc-power at the reference operating condition, default value of zero gives a linear relationship, (1/W)
C_1	empirical coefficient allowing Pdco to vary linearly with dc-voltage input, default value is zero, (1/V).
C_2	empirical coefficient allowing Pso to vary linearly with dc-voltage input, default value is zero, (1/V).
C_3	empirical coefficient allowing Co to vary linearly with dc-voltage input, default value is zero, (1/V)

4.5.9 Buffer tank (storage vessel)

A model of a simplified storage vessel for warm (heating) water is made (see figure below). In this model the storage vessel is split up into an upper part and lower part. Both parts have a fixed temperature. Furthermore, the simplification is made that both parts do not loss or gain energy from the environment and do not exchange energy (heat) between one and another. Also, no mixing of water between both parts is assumed.

In case the storage vessel is loaded (see figure below), warm water from the heating installation is supplied to the top of the vessel with a fixed temperature, e.g. the setpoint for the water temperature for heating (Tsupply). In case the storage vessel is unloaded, warm water from the top of the vessel is extracted and fed into the heating system, while return water (Treturn) from the heating system is supplied at the bottom of the vessel. Based upon the waterflow, the model calculates at each timestep the part of the storage vessel that is filled with warm water as fraction of the total volume (parameter fill, see figure below). The amount of warm water that can be delivered is limited to the available warm water content of the vessel, while on the other hand the amount of warm water that can be stored cannot exceed the cold water part that is present at that moment. Thus fill is limited between 0 and 1. The delivered energy by the storage vessel or stored energy in the storage vessel is determined each timestep.

Although the storage vessel is made for warm water storage, with the previous mentioned assumptions, it can also be used as a cold water storage. In that case in practice the supply and extraction of cool water should be from the bottom of the vessel.

Specifications

- Volume of the storage vessel [m3]
- Maximum load or unload flow [l/s]
- Heating/cooling supply temperature, Tsupply [°C]
- Heating/cooling return temperature, Treturn [°C]

Output

- Loaded fraction of the storage vessel, fill [-]
- Delivered or stored heating/cooling power per timestep [kW]

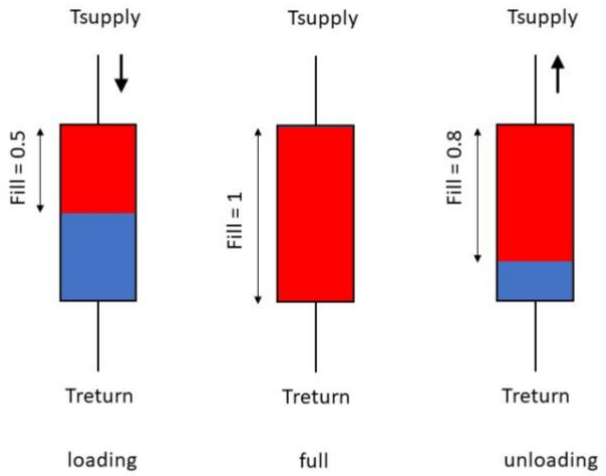


Figure 4.8: Simplified storage vessel.

4.6 Simulation and prediction models

In the previous sections the components of the hybrid model were described in detail. The overall building model (described in section 4.3) including the occupant model and equipment models can now be used to calculate the dynamic response of the building.

Conceptually, the set of all equations may be written as:

$$\frac{dx}{dt} = f\big(x(t), u(t), d(t)\big)$$
$$y(t) = g(x(t), u(t), d(t))$$

Here the first differential equation (ODE) is the change in time of the building state vector x(t), as a function of the current state x(t), the vector of non-manipulatable (external) influences d(t) (such as weather conditions) and the vector of manipulated inputs u(t) that may be used to control the building behaviour.

y(t) is a vector of all physical variables of interest that are calculated.

A simulation consists of running the model (integrating the ODE) from a given start time t0 to an end time t1 usually with a fixed timestep Δt.

If the simulation is conducted on historical data of the real building of applied inputs u(t) and d(t) and a known start state x(t0) the computed outputs y(t0..t1) may be compared with measurements of the same variables ym(t) measured in practice. This is used to assess the accuracy of the simulation compared to the measured behaviour.

An important goal of the model is to use it for **predicting** expected building behaviour over a near future horizon for a certain control scenario u(t0..t1).

To use the model in this way the following steps are performed:

1. Given current moment t0, determine the current state of the building x(t0) using the measured output y(t0) and the estimated state x(t0) from the previous prediction (at t0-Δt).
2. Determine the time sequences for the external inputs d(t) (e.g. weather forecast, expected occupancy) over the prediction horizon t0..t1.
3. Select a control strategy or control scenario u(t0..t1) for the prediction horizon.
4. Simulate the model over the prediction horizon t0..t1 and compute predicted outputs y(t0..t1).

This sequence of steps may be repeated each time step, resulting in a moving horizon prediction. The prediction model may be used to test the effect of different control scenarios u(t) on building performance.

Within the B4B project the prediction model (or a linearized approximation of it) is also used to find the control sequence u(t) that optimizes building performance. In this case the prediction model is used in a Model Predictive Control (MPC) context.

4.7 Application in use case TNO Stieltjesweg

4.7.1 Description use case TNO Stieltjesweg

The TNO Stieltjesweg building in Delft is an office building with optical labs in the basement, see Figure 4.9. This section gives an overview of the building itself, the HVAC components that are relevant to the prediction model and the monitoring set-up in the building. The component models themselves are described in section 4.5.

Figure 4.9: The TNO Stieltjesweg building.

Building

The TNO Stieltjesweg building is part of the Applied Physics building of the TU Delft. Year of construction is 1963, while the building is renovated in 2012. The building consists of 5 floors. The total gross internal floor area is 5400 m^2. In the basement are the optical labs and the (greater part of the) HVAC equipment. The first floor until the fourth floor contain offices and meeting rooms. Not much information is available about the building physics. The walls are heavy (roughly 20 cm concrete plus 20 cm lightweight concrete) and not insulated. Concrete floors and roofs are used. The roof is insulated on the outside (roughly 8 or 10 cm). The windows are equipped with double glazing. A central controlled outside solar shading is available on the south side of the building.

HVAC equipment

The HVAC equipment installed in Stieltjesweg is schematically represented in the diagram below.

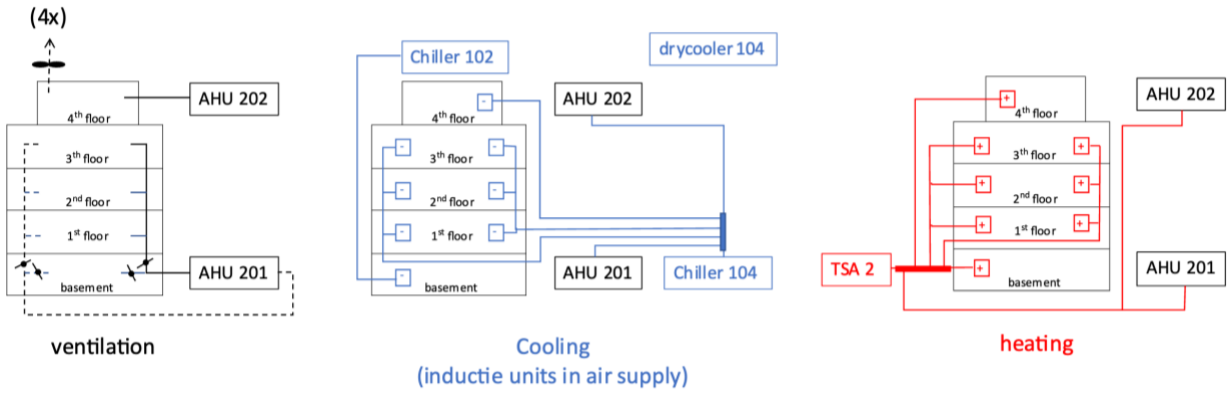


Figure 4.10: Diagram of the equipment in the TNO Stieltjesweg building.

There are two air handling units (AHUs) installed. AHU201 ventilates the labs and floors 1-2-3. The AHU202 ventilates floor 4 (extension of the building). A combination of a Chiller and Drycool unit supplies cold water to the AHUs and to the induction units. Chiller102 is dedicated to the induction units in the basement (labs).

The building has central heating (radiators). Heat is supplied by the district heating system.

Monitoring set-up

To analyze the prediction quality and uncertainty of the energy demand forecasting method a monitoring system has been set up for the use case TNO Stieltjesweg. The monitoring set-up consists of the following subsystems:

- A Johnson BMS. The data of the BMS is available through permanently installed PULSE core system software (of maintenance company SPIE) for monitoring.
- Additional sensors for temperature, relative humidity and CO_2 on room level
- Occupancy monitoring
- Electric submetering

PULSE Core system

The PULSE Core system continuously monitors all permanently installed sensors and manipulated variables in the building. The data is stored in a database which is on-line accessible to the prediction model. The data is tagged according to the standard “project Haystack” protocol, see the figure below.

id	absolute	air	airQuality	atmospheric	calc	calend
Stieltjesweg Naregeling E367 Setpoint instelling						
Stieltjesweg Naregeling E377 Ruimte temperatuur		✓				
Stieltjesweg Naregeling E265 Ruimtetemperatuur		✓				
Stieltjesweg Naregeling E152 Naregeling E152 Status						
Stieltjesweg Naregeling E282 Naregeling E282 Status						
Stieltjesweg 029 - Radiatoren en lab nieuw Aanvoertemp. warmw						
Stieltjesweg 104 - KM koude opwekking E3 CP condensor koelma						
Leeghwaterstraat 46 GKW groep Koeling luchtcompressoren KW:						
Stieltjesweg Naregeling E380 Werksetpoint		✓				
Leeghwaterstraat 46 LBK Kantoren Inblaas Absoluut vocht LBK Ka	✓	✓				
Stieltjesweg 104 - KM koude opwekking E3 Droge koeler bedrijf						
Leeghwaterstraat 46 Naregeling 080 nvoActiveSetPt		✓				
Stieltjesweg Naregeling E383 Uitgang verwarmer (0v=0%,10v=10						
Stieltjesweg Naregeling E280 - Uitgang verwarmer (0v=0%,10v=1						
Stieltjesweg 030 - Radiatorengroep 3de verd Werksetpoint						
Stieltjesweg 104 - KM koude opwekking E3 Intrede temp. conden						
Stieltjesweg Naregeling E369 Uitgang verwarmer						
Stieltjesweg 104 - KM koude opwekking E3 CP verdamper koelma						
Stieltjesweg Naregeling E265 Naregeling E265 Status						
Stieltjesweg Naregeling E272 Uitgang koeler (0v=0%,10v=100%)						
Stieltjesweg Naregeling E299 - Uitgang verwarmer (0v=0%,10v=1						
Stieltjesweg Naregeling E277 Setpoint instelling						
Leeghwaterstraat 44 Naregeling 0.010 Meting		✓				

Figure 4.11: Example of Haystack tagging in the TNO Stieltjesweg building.

Additional temperature, relative humidity and CO_2 sensors on room level

On each floor of the building 4 additional sensors are installed in the offices for temperature, relative humidity and CO_2. The purpose of those sensors is twofold, at first the temperature sensors are more accurate than the sensors of the thermostats. Second, the CO_2 and relative humidity are additional to the sensors of the BMS and give more information about the comfort and air quality in the offices. The 4 sensors per floor are distributed over the different zones. Additional to the sensors in the offices also one or two sensors per floor are installed in the meeting rooms. In total 16 additional sensors are installed in the building.

Occupancy sensors

To measure the number of people on each floor of the building people counters are installed throughout the building. Each sensor counts the number of people passing in both directions. By aggregating the sensors the number of people per floor can be calculated for each moment in time.

Electrical submetering

Additional kWh-meters are installed to monitor the electricity use of:

Chiller 104

Drycooler 104, including the circulation pumps on evaporator side (chilled water circuit) and condenser side (drycooler circuit)

AHU 201, including the circulation pumps of the induction groups 126, 127 and 128

One of the supply fans of AHU 201

4.7.2 AHU system simulation use case TNO Stieltjesweg

The air handling unit (AHU) used in the TNO Stieltjesweg building is given in Figure 4.12Figure 4.12: Air handling unit TNO Stieltjesweg building.

. A model is made of this air handling unit using the fan model (see section 4.5.2), the heat recovery model (see section 4.5.3) and the heating and cooling coil model type A (see section 4.5.4).

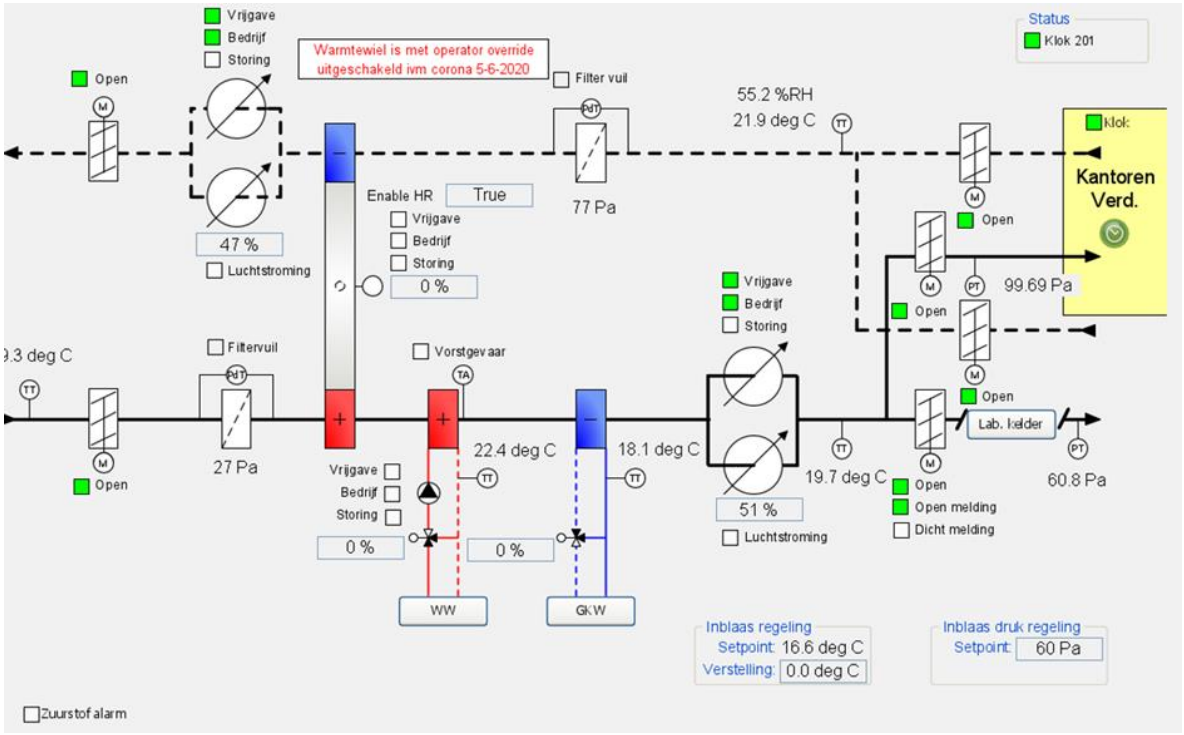


Figure 4.12: Air handling unit TNO Stieltjesweg building.

Simulations are performed over the period from May 1st until September 1st 2022. Measured data from that period is obtained from the building management system (BMS). This concerns:

- the setpoint for the air supply to the building (Tsetpoint)
- the measured air supply temperature (Tsupply-BMS)
- the measured exhaust air temperature (Texhaust-BMS)
- the control (rotation speed in % of max rotation speed) of the heat recovery unit (hruCtrl-BMS)
- the control of the heating valve (heatingValve-BMS)
- the control of the cooling valve (coolingValve-BMS)
- the control of the fans (fanCtrl-BMS)

Besides that, meteo data (Tambient) is obtained from a weather station nearby.
During the investigated period, the heat recovery unit (HRU) was controlled in two situations:

- Before June 21, the HRU was off. This was done due to COVID to prevent the spread of extracted air back to the building.
- After June 21, the HRU was controlled at 10% rotation speed.
- It is unclear why the HRU was set at this fixed position in this period.

To make a comparison possible between the simulated results and the measured values, the HRU is controlled during the simulations in the same manner.

Inputs used for the simulations (part of the measured items as given before) are the following:

- the setpoint for the air supply to the building (Tsetpoint)
- the measured exhaust air temperature (Texhaust-BMS)
- the control (rotation speed in % of max rotation speed) of the heat recovery unit (hruCtrl-BMS), being the two fixed values as indicated before
- the control of the fans (fanCtrl-BMS)
- Tambient as obtained from a weather station nearby

The outputs of the simulations and the items used for comparison with the measurements are:

- The supply air temperature to the building (Tsupply-sim),
- The valve position of the heating valve (heatingValve-sim),
- The valve position of the cooling valve (coolingValve-sim).

It is mentioned that also the air temperature after the HRU (ThruOut-sim) is an output of the simulations. However, this cannot be used for comparison because this temperature is not measured but is used as input for the cooling/heating coil in the simulations.

For a week in May (Sunday May 15 up until May 22) and a week in August (Sunday August 21 up until August 27), the results are shown in Figure 4.13 and Figure 4.14, respectively.

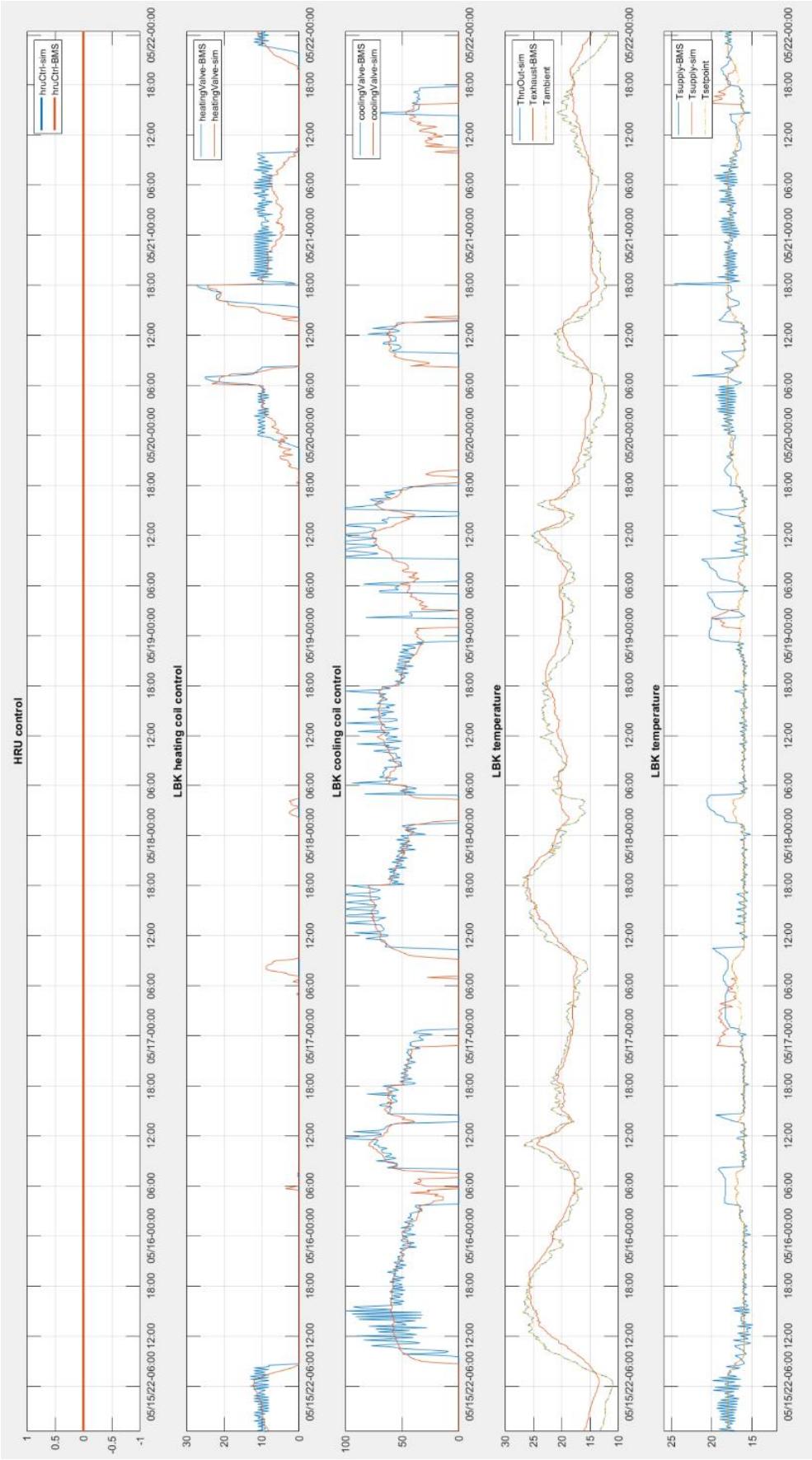


Figure 4.13: Simulated and measured results Sunday May 15 up until May 21 2022.

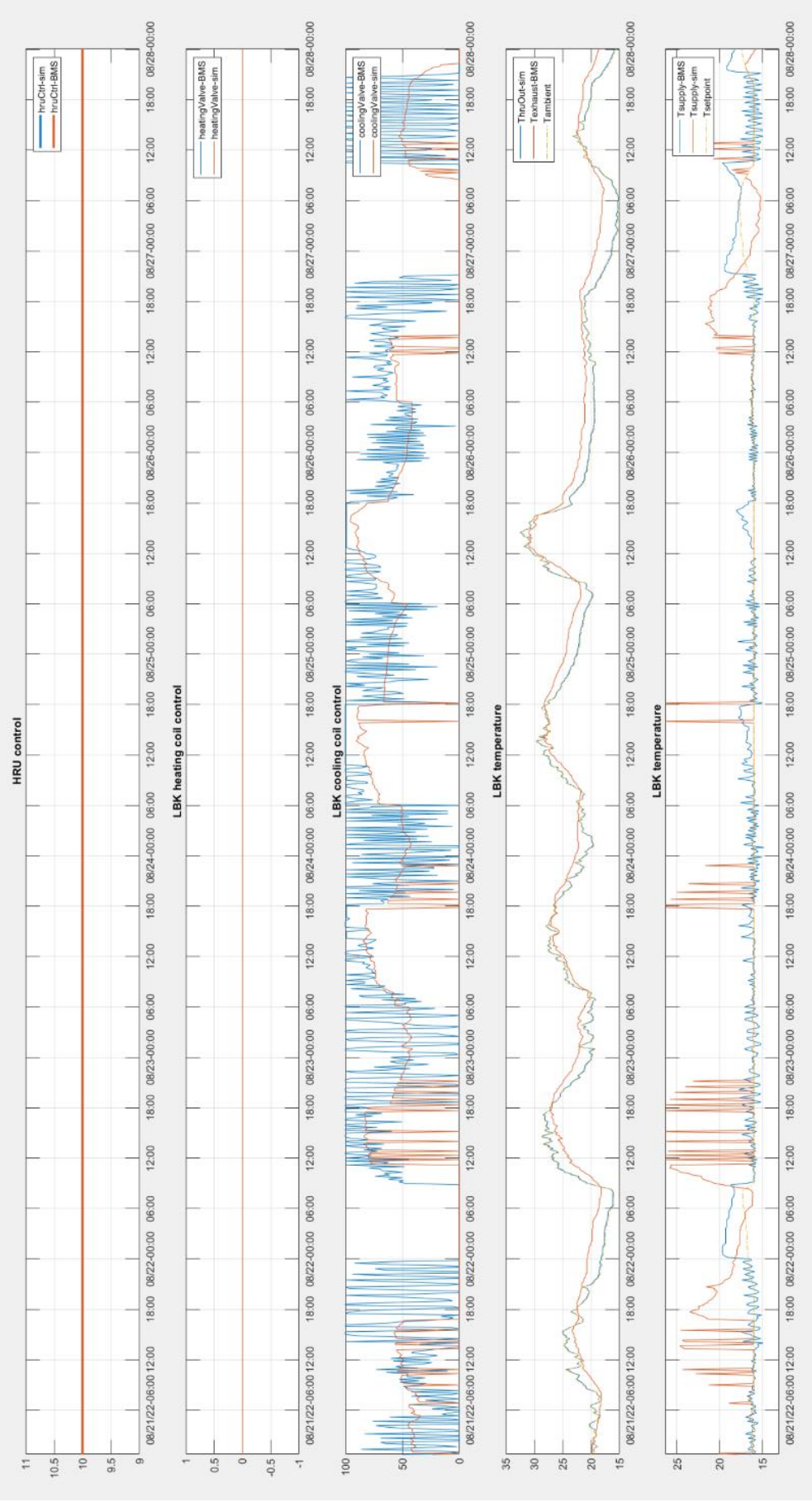


Figure 4.14: Simulated and measured results Sunday August 21 up until August 27 2022.

Both Figure 4.13 and Figure 4.14 show a good agreement between the simulated and measured supply air temperatures (Tsupply-sim and Tsupply-BMS). However, there are time periods in which the simulated supply air temperature and the supply air temperatures according to the BMS deviate. This is due to the fact that either the cooling in practice was working but in the simulations not, or the other way around. In the model, the cooling is related to the use-signal of the chiller (only when chiller is ON, cooling is provided), but for unknown reasons, this does not lead to a complete alignment of the use of cooling in practice and simulations.

Figure 4.13 shows quite a good agreement between the measured cooling valve position (coolingValve-BMS) and the simulated value (coolingValve-sim). It is noted that based upon the measured cooling valve position and the simulated cooling power from another period, the relation between valve position and the cooling power is derived. This relation is used to simulate the cooling valve position.

Figure 4.14 shows a great amount of instability concerning the measured cooling valve position (coolingValve-BMS). The chiller also switches frequently on and off (e.g. August 22 at 12:00 and 18:00 hours) as is shown by the simulated cooling (coolingValve-sim). Of course this disturbs the simulation. This problem with instability should be solved in the actual control of the AHU.

Figure 4.13 shows a good agreement between the simulated heating valve position (heatingValve-sim) and the measured heating valve position (heatingValve-BMS). To simulate the control of the heating valve the assumption is made that the valve position is given by the quotient between actual heating power and maximum heating power.

4.7.3 Chiller system simulation use case TNO Stieltjesweg

No information about the electricity use of the chiller 104 is available, because the kWh-meters still need to be installed. Therefore, it is not yet possible to test the electricity use predicted by the chiller model.

4.7.4 Total building simulation use case TNO Stieltjesweg

For the Stieltjesweg building a RC-model is constructed. The model of the building consists of 8 zones (covering ground floor up to and including the 3rd floor, with two zones per floor along the length axis of the building to discriminate between the south-east facing part and the north-west facing part). The basement where the labs are located is not modelled, but serves as a boundary condition for the ground floor. The dimensions, window sizes and material properties of walls and floors are taken from the available technical data and the "Energiepotentieel scan" of 2001[5].

Figure 4.15: Visualization of the BEM zone model of the TNO Stieltjesweg building.

The RC model describes the dynamic energy balance for each of the zones. The floors and roof are modelled, the separation walls between offices in a zone are not modelled. Each zone interacts with:

- Other zones (constituting a network model of the total building)
- Exterior conditions (outside temperature, wind, solar radiation, shading by other buildings, humidity)

[5] https://www.bepositief.nl/energie-potentieelscan/

- The Air handling unit (central unit + induction units)
- The cooling system in the basement
- The occupants (presence, internal heat loads)

The RC-model is generated directly from the building BEM file provided in standard gbXML format. The figure below shows the network heat transfer connections as an illustration.

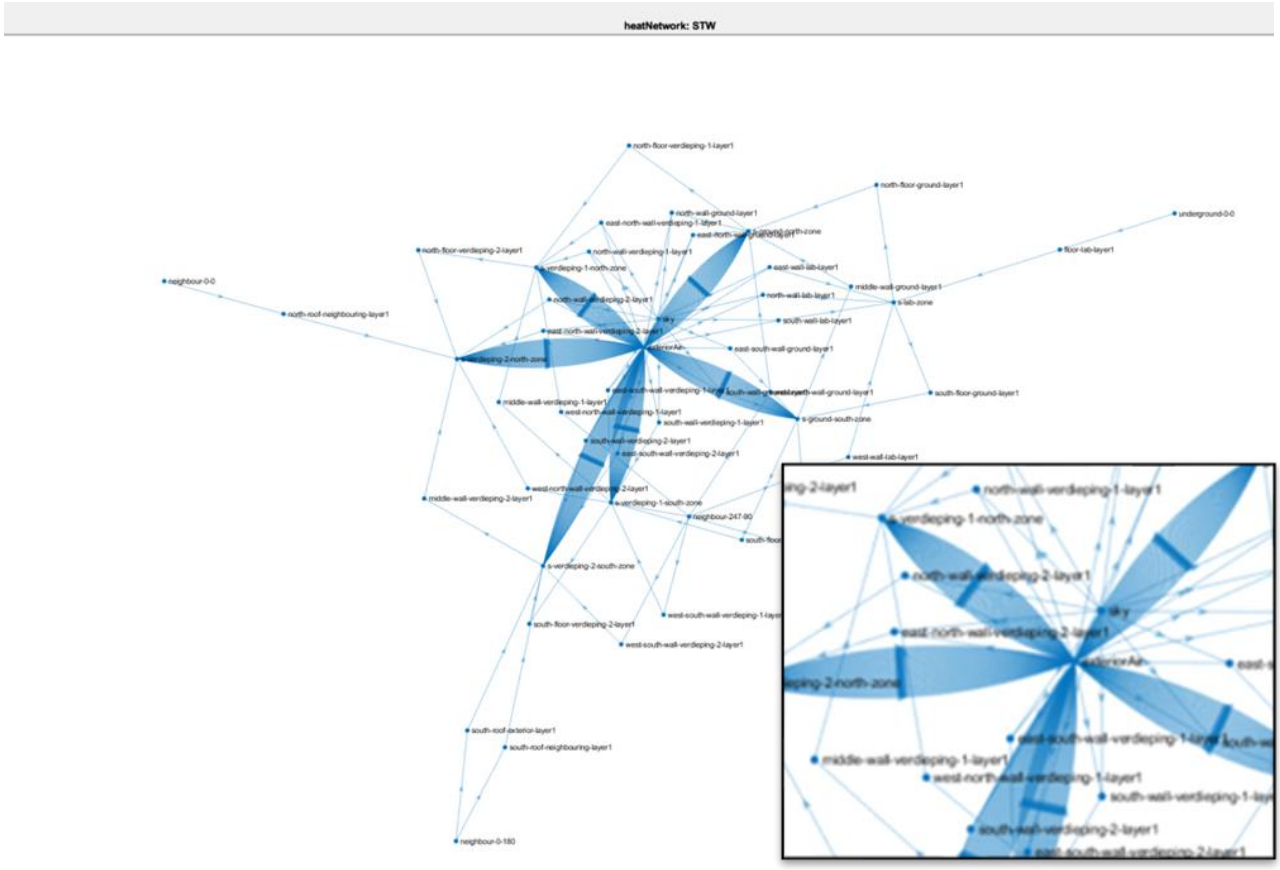


Figure 4.16: RC-heat network of the TNO Stieltjesweg building.

Each edge represents a heat transfer link between zones, walls, windows or roof. The network presented in the figure is a simplification of the network simulated, since that also contains multiple construction layers in the exterior walls.

Equipment models modules are developed from the basic building blocks described in Section 4.5 of this document. We implemented the following modules in the total building simulation model:

- AHU model type A (including models for the fan, heat recovery and cooling and heating coils)
- Induction unit model

In addition, the operation of the automatic blinds for the south-facing windows is simulated by a simple model, assuming two radiation thresholds: a threshold of 200 W/m^2 for closing the blinds, and a lower threshold of 170 W/m^2 for opening them.

The models for the chiller, buffer tank and PV will be added in the subsequent version of the model.

To compare the model with the measurement data, measured occupancy and the setpoint temperature have been used, instead of using models to predict them. Data-driven profiles have been used to take into account the heat gains from appliances.

Results

Primary simulation results are shown in Figure 4.17, where the predicted zone temperatures are compared with the averaged measured zone temperatures (obtained by averaging the measured temperatures for all the rooms within each zone).

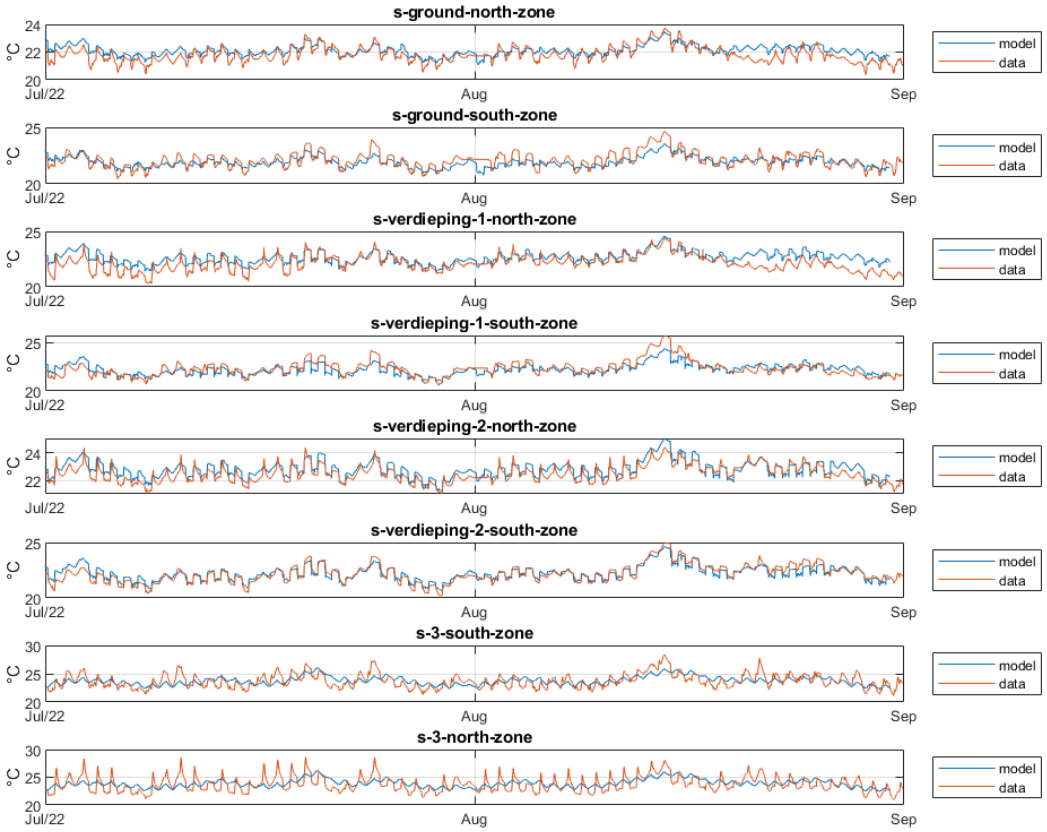


Figure 4.17: Predicted zone temperature (blue) vs the averaged measured zone temperature (red).

4.8 Conclusion

A hybrid model is constructed which combines a white box model physics based model for building and equipment and a black box model for occupant behaviour. This model is used in simulating and predicting dynamic building behaviour. The developed hybrid model is applied to the use case of the TNO building Stieltjesweg in Delft for a summer period focusing on the cooling demand. The real building behaviour is retrieved from several data sources and compared to the hybrid model. Preliminary analysis shows that the building behaviour is in line with the simulation results. Especially in the south-facing zones, the model is overestimating the zone temperatures, probably due to the overestimation of the solar radiation passing through windows. Currently development is in progress to improve the model predictions by including the solar shading from surrounded buildings and also include models of automatic shading devices. The updated results will be presented in the next update of the report.

At the moment no data on electricity consumption of individual HVAC components (such as the chiller) is available, but installation of submetering at Stieltjesweg building is in motion. Having access to measured electricity data, which facilitates calibration of the hybrid model, will improve prediction of real energy demand.

5 GREY BOX MODELS: USING CO_2 MONITORING DATA TO DERIVE VENTILATION FLOW RATE OR OCCUPANCY

Authors: Henri ter Hofte and Hossein Rahmani (Hogeschool Windesheim)

This chapter starts with a description of the goal of the models in section 5.1. Section 5.2 describes the purpose and research questions, followed by section 5.3 by the dynamic model we used in our inverse grey-box modelling approach. Section 5.4 describes the data collection, recruitment and measurement techniques used. In section 5.5, we describe the inverse grey-box analysis tool we used for our analysis (GEKKO Python), the error metrics we used in our analysis, and the results of the data analysis, both for synthetic room data and real room data. Finally, in section 5.6, we describe our conclusions and recommendations.

5.1 Goal of the models

The goal of the inverse grey box models we describe in this chapter, is to derive recent and current ventilation flow rate over time based on monitoring data on CO_2 concentration and occupancy, as well as deriving recent and current occupancy over time, based on monitoring data on CO_2 concentration and ventilation flow rate.

Section 5.2 describes the purpose and research questions, followed by section 5.3 by the dynamic model we used in our inverse grey-box modelling approach. Section 5.4 describes the data collection, recruitment and measurement techniques used. In section 5.5, we describe the inverse grey-box analysis tool we used for our analysis (GEKKO Python), the error metrics we used in our analysis, and the results of the data analysis, both for synthetic room data and real room data. Finally, in section 5.6, we describe our conclusions and recommendations.

Data was collected in 6 office rooms at Windesheim University of Applied Sciences, for 3 weeks during autumn 2022. Due unavailability of retrospective ventilation data from the building management systems, we had to reject data for 3 office rooms, leaving data from 3 rooms usable for analysis. This data is available as an open dataset (Ter Hofte et al., 2022/2023). In a final stage of the analysis, we discovered that Bluetooth-based presence detection had insufficient coverage compared to the capacity of 1 room, so, in this report, we analyze the data from the 2 remaining rooms. The firmware we used on our publicly available data collection hardware is available as open software (Twomes CO_2 and Bluetooth occupancy firmware for the M5CoreInk and SCD41, 2023), and the GEKKO Python models and analysis code we used, is available as open software as well (Twomes inverse grey-box modelling and analysis tools for homes and utility buildings, 2023).

5.2 Purpose & research questions

Knowing (patterns in) ventilation flow rates or occupancy can be important to assess the flexibility options of a building. However, depending on the building and building measurement system, it can be hard or costly to measure these features directly.

5.2.1 Research question A: ventilation flow rates

Ventilation, while necessary to maintain a healthy indoor air quality for occupants, not only consume an increasing part of the electrical energy of buildings, but often also cause a substantial fraction of heat loss in a building. Knowing the extent to which ventilation causes heat loss may help to assess the thermal characteristics of a building better, which in turn helps to assess the flexibility potential of a building better.

In many larger utility buildings (typically those that have a building management system) ventilation flow rates can be measured, and monitoring data can be obtained from a building management system. However, in many smaller SME buildings without a building management system, ventilation rates may be hard to obtain, thus hindering the assessment of opportunities for managing thermal flexibility of a building.

The research questions in this chapter are based on the observation that occupancy raises CO_2 concentration and ventilation lowers it, which is illustrated as an overly simplified equation in Figure 5.1.

Figure 5.1: Simplified relation between occupancy, ventilation and CO_2 concentration.

The challenge and observation described above gave rise to the first of two research questions that we address in this chapter:

A. To what extent can we reliably derive the ventilation flow rate in a room from data about CO_2 and occupancy?

5.2.2 Research question B: occupancy

In larger utility buildings with building management systems, obtaining ventilation flow rate data may be easier. However, a new challenge arose during and after the Covid pandemic when universities started experimenting with occupancy detection systems to optimize building usage. Some of these systems utilized camera-like sensors. These sensors typically captured images, often in the form of silhouettes, and processed them within the sensor device to determine the number of people present. Only the resulting numerical counts were transmitted, ensuring privacy. Nevertheless, the use of camera-like sensors frequently raised privacy concerns.

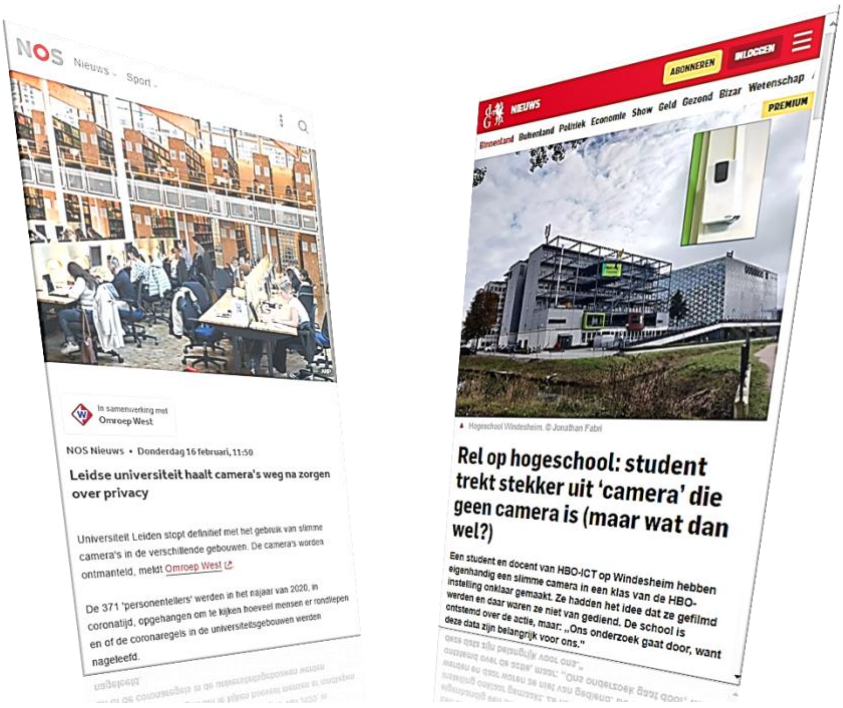


Figure 5.2: Concerns about privacy in the press about universities using 'cameras' for occupancy detection.

As can be observed from Figure 5.2, these concerns were not only present at Leiden University (Omroep West, 2023), but also at Windesheim (Redactie, 2021), where we intended to do the data collection for this study. This motivated research question B, below, which aims to assess the feasibility of a privacy-friendly way of detecting occupancy in utility buildings.

B. To what extent can we reliably derive occupancy in a room from data about CO_2 and ventilation?

Unlike some other research mentioned in this deliverable, the study reported in this chapter only intends to assess a current value of occupancy or ventilation rates from recent data; the purpose of this study was NOT to predict occupancy or ventilation rates in the near future; this would require further research.

5.3 Inverse grey box modelling of CO_2, ventilation, and occupancy

In this study, we used an inverse grey-box modelling approach to get an answer to the research questions. As its name suggests, inverse grey-box modelling is a modelling approach that combines characteristics of both white box modelling and black box modelling.

White box modelling is an approach that aims to represent building (energy) dynamics in a detailed manner, incorporating numerous physical parameters and equations. It is commonly utilized during the design phase of a building or installation, when the system being modelled has yet to be realized. By using white box modelling, one can comprehensively understand the system's behavior and performance. This approach allows for a thorough analysis and optimization of various aspects of the building or installation, considering factors such as energy efficiency and performance.

Black box modelling is an approach that relies on data to model a system. It is suitable when the system being modelled is already implemented, and there needs to be more knowledge about how it operates, such as the way inputs are transformed into outputs. The key requirements for a black box model are clean data and an appropriate machine learning algorithm. There are several advantages to using black box modelling, including the ability to handle nonlinearities without needing exact equations and adaptability to new data, as the model can improve itself with additional information. However, there are also some drawbacks. The performance of the model heavily relies on the quality of the data, and black box modelling requires a large dataset and significant computational power.

Inverse grey box modelling (Madsen et al., 1995; Bacher et al., 2011) is a useful approach when studying a system that is already realized, and there is both available data and some domain knowledge regarding how the system operates. This knowledge often involves a few equations that describe the system's underlying physics. In comparison to black box modelling, inverse grey box modelling provides computationally efficient results that are easier to interpret and explain. While it requires more domain knowledge than black box modelling, it necessitates less knowledge than white box modelling, which involves detailed physical parameters and equations. Inverse grey box modelling strikes a balance between computational efficiency and interpretability, making it a valuable tool for understanding and analyzing complex systems. In recent years, inverse grey-box modelling has been used in e.g., identifying model parameters of building heat dynamics, on a set of 247 Dutch residential buildings, using data from Toon smart-thermostats at granularities of 15 min intervals (Leprince, et al., 2022).

Below, in section 5.3.1, we describe the molar CO_2 concentration balance model that we used to describe the system dynamics and in section 5.3.2 we describe how we use this model in an 'inverse' way to learn parameters based on measured data.

5.3.1 Molar CO_2 concentration balance model

As illustrated graphically in Figure 5.3, we use a molar CO_2 balance model to describe the dynamics in the relation between CO_2 concentration, ventilation, infiltration and occupancy. This model assumes that CO_2 is spread evenly in the room and that infiltration scales linearly with wind speed.

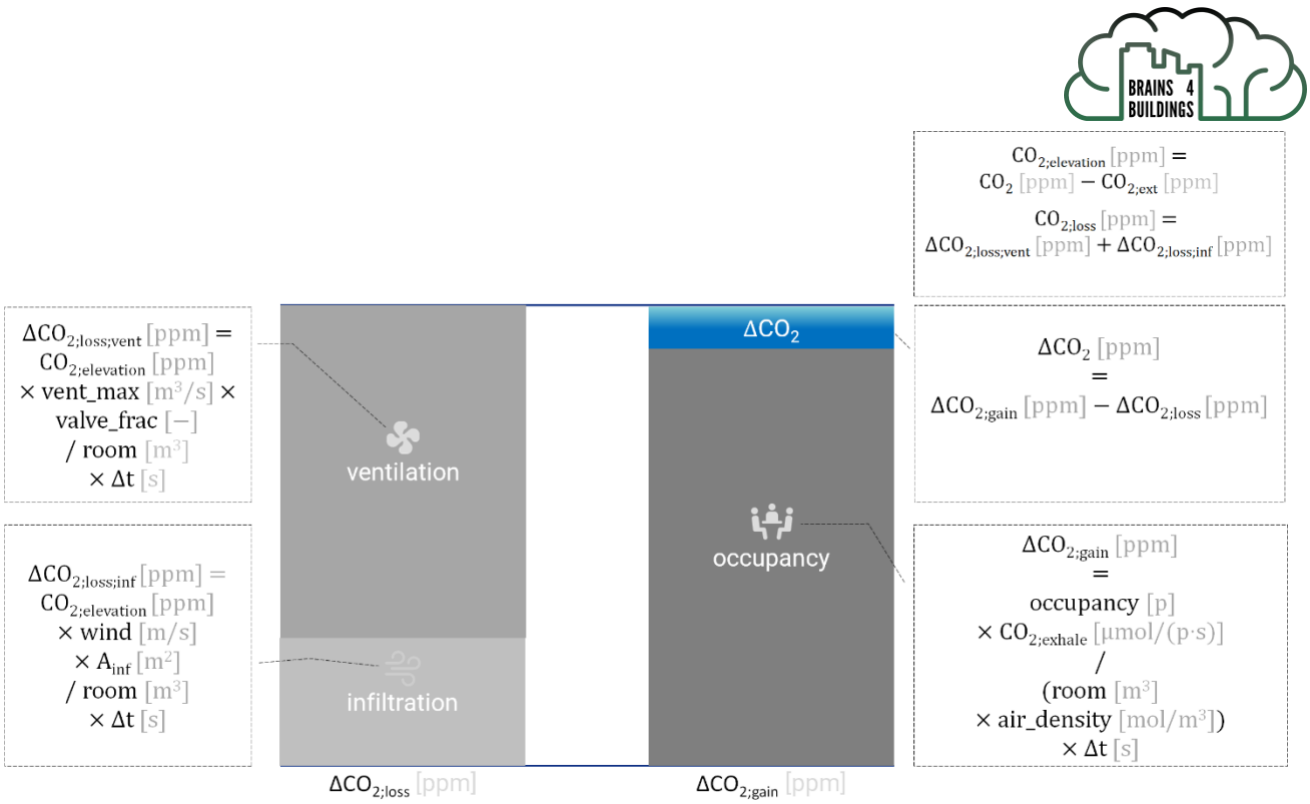


Figure 5.3: Molar CO_2 concentration balance model.

Expressed as a set of Ordinary Differential Equations (ODEs) with units added for clarity between square brackets and in a grey font, and with air change per hours metrics used as intermediate variables, the model can be expressed as follows:

- dCO_2/dt [ppm/s] = $CO_{2;gain}$ [ppm/s] − $CO_{2;loss}$ [ppm/s]
- $CO_{2;gain}$ [ppm/s] = occupancy [p] × $CO_{2;exhale}$ [µmol/(p·s)] / (room [m³] × air_density [mol/m³])
- $CO_{2;loss}$ [ppm/s] = $CO_{2;loss;vent}$ [ppm/s] + $CO_{2;loss;inf}$ [ppm/s]
- $CO_{2;loss;inf}$ [ppm/s] = $CO_{2;elevation}$ [ppm] × $air\_changes_{inf}$ [h^{-1}] / 3600 [s/h]
- $CO_{2;loss;vent}$ [ppm/s] = $CO_{2;elevation}$ [ppm] × $air\_changes_{vent}$ [h^{-1}] / 3600 [s/h]
- $air\_changes_{inf}$ [h^{-1}] = A_{inf} [m²] × wind [m/s] × 3600 [s/h] / room [m³]
- $air\_changes_{vent}$ [h^{-1}] = valve_frac [−] × $air\_changes_{max;vent}$ [h^{-1}]
- $CO_{2;elevation}$ [ppm] = CO_2 [ppm] − $CO_{2;ext}$ [ppm]

We used the following values for the constants in the model:

- $air\_changes_{max;vent}$ [h^{-1}] = vent_max [m³/h] / room [m³] (see also Table 5.1)
- $CO_{2;ext}$ [ppm] = 415 [ppm] (Lan, et al, 2022)
- $CO_{2;exhale}$ [µmol/(p·s)] = 270 [µmol/(p·s)], which we based on 28.7 [g/(p·h)] (Li et al., 2022) / (44.0095 [g/mol] × 3600 [s/h]) × 10^6 [µmol/mol/] × 1.5 [W^0] (Haskell et al., 2007, p. 1086), rounded to multiples of 10 [µmol/(p·s)]
- air_density [mol/m³] = 41,57 [mol/m³] (molar density of an ideal gas at 1 atm and 20 °C).

5.3.2 Inverse grey box modelling: learning parameters from data

Inverse grey box modelling tools employ models to transform input data into output data, as depicted in Figure 5.4 a and b. In our specific scenario, the simulated output data using the molar CO_2 concentration balance model is the concentration of CO_2, which is calculated based on the input data and the model. This approach is referred to as "inverse" because it iteratively adjusts the time-independent model parameters, and in our case, also one of the time-dependent input data series. The goal is to find the optimal configuration that closely aligns the measured CO_2 concentration values with the simulated concentration values. Through multiple iterations, the inverse grey box modelling process identifies the "best fit" between the measured and simulated CO_2 concentration values.

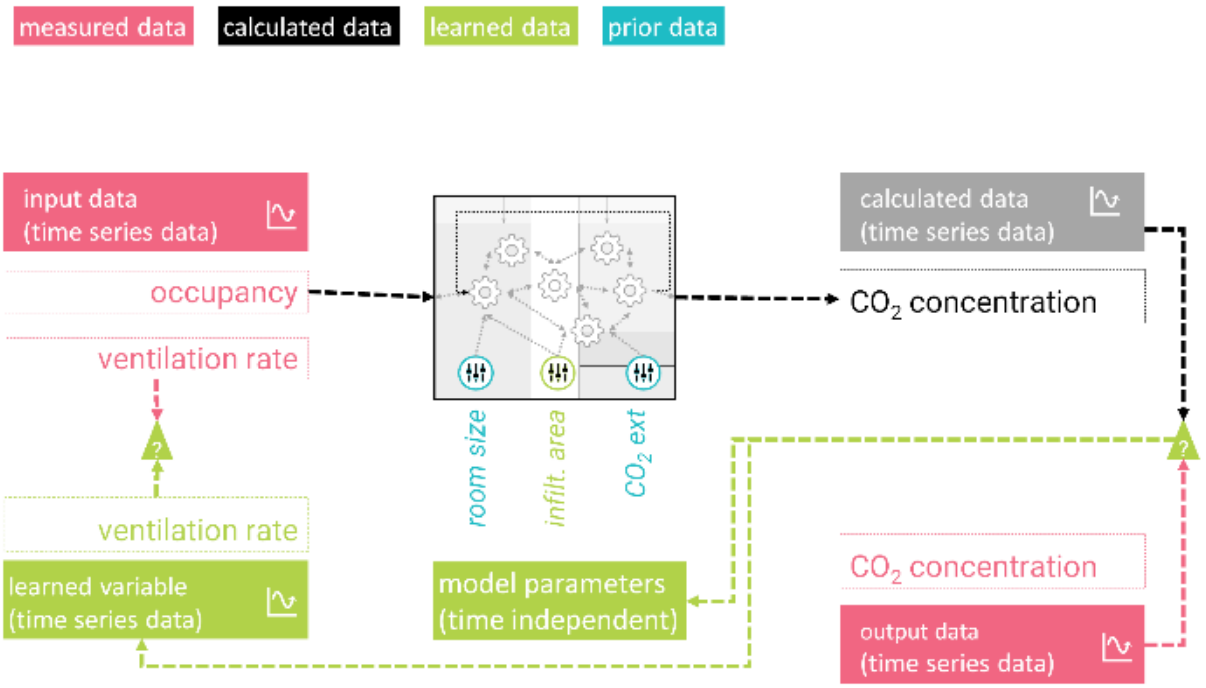


Figure 5.4a: Schematic illustration (for research question A) of the GEKKO estimation and errors that we calculated (green triangles) by comparing learned and observed data.

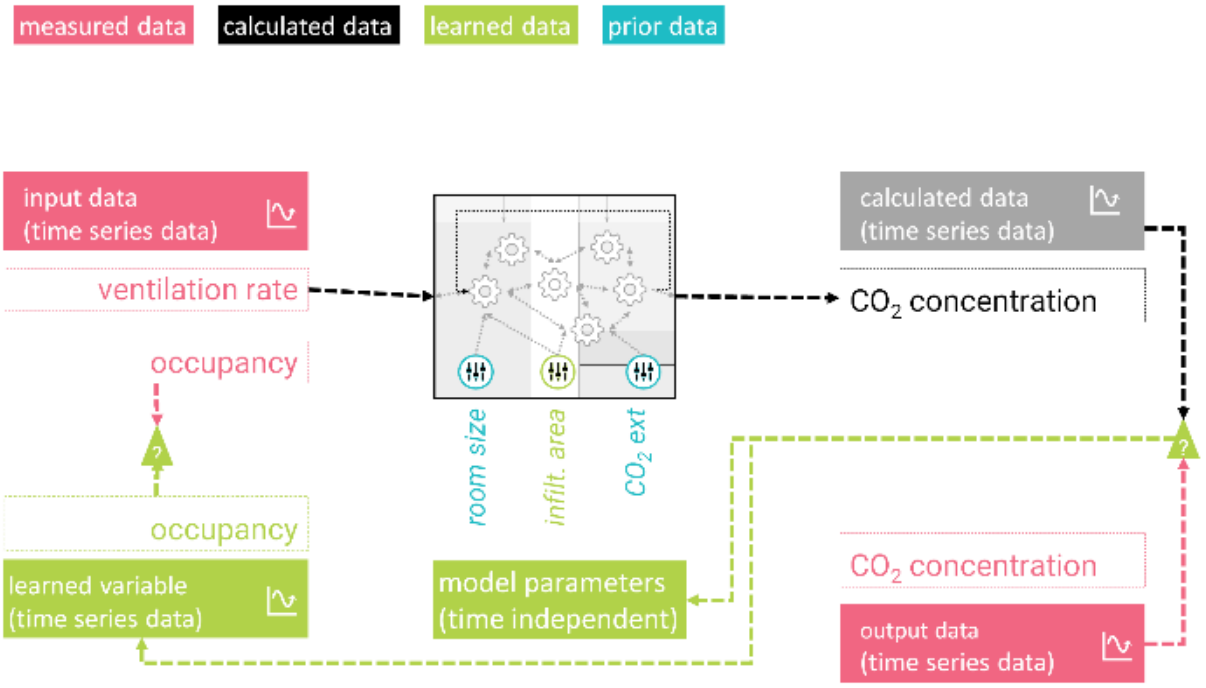


Figure 5.4b: Schematic illustration (for research question B) of the GEKKO estimation and errors that we calculated (green triangles) by comparing learned and observed data.

We allow the inverse grey-box modelling tool to manipulate one time-independent variable, A_{inf} [m²] (the infiltration area that mediates infiltration due to wind) and one time-dependent variable. We vary the time-dependent variable depending on the research question:

A. In this research question (derive ventilation flow rate from CO_2 concentration and occupancy), illustrated in Figure 5.4a, *ventilation flow rate* is the time-dependent variable we allow the inverse grey-box modelling tool to manipulate. To be more precise, the tool is allowed to manipulate the variable `valve_frac__0` (fraction that ventilation valve is open), which together with room-specific `vent_max__m3` determines the ventilation flow rate [m³/s]. The ventilation flow rate, in turn, together with `room__m3` (size of the room [m³]), determines the number of air changes per second [s^{-1}].

B. In this research question (derive occupancy from CO_2 concentration and ventilation flow rate), illustrated in Figure 5.4b, *occupancy* is the time-dependent variable that the inverse grey-box modelling tool is allowed to manipulate.

The GEKKO Python inverse grey box modelling tool first minimizes the CO_2 error, by manipulating the time-independent model parameter(s) and the time-dependent variable. After minimization of the CO_2 error, we call the optimal value for the time-independent variable A_{inf} [m^2] that was found the 'learned' A_{inf} [m^2] and we call the optimal values for the time-dependent variable that were found the 'learned' valve fraction (research question A) or 'learned' occupancy (research question B).

5.4 Data collection

In this section the method for data collection is described. In section 5.4.1, we describe subject recruitment, in section 5.4.2, we describe the measurement devices we used and in section 5.4.3., we describe the data that we collected.

5.4.1 Subject recruitment

Before we started recruitment, we requested and obtained approval for our study from Windesheim Research Ethics Committee, based on a description of the research, the privacy policy (Volledige Privacyverklaring Brains4Buildings, 2022) and Data Management Policy (Ter Hofte, 2022a).

Subjects were recruited via a recruitment e-mail targeted at people known to work in office rooms that satisfied the inclusion criteria for office rooms.

Inclusion criteria for office rooms at Windesheim were:

- Passive infrared (PIR) sensor-based occupancy data, CO_2 concentration and ventilation data was available via the existing Building Management System (BMS);
- 75% or more of the occupants of an office room consent to their presence being tracked.

Inclusion criteria for subjects were:

- subjects provided informed consent in the online recruitment survey (Ter Hofte, 2022b), which also referred to the privacy policy[Error! Bookmark not defined.] and verified the inclusion criteria below;
- subjects must work at Windesheim University of Applied Sciences in one of the eligible office rooms (we recruited among occupants of eligible office rooms according to facility management and asked in the survey to indicate in which of the eligible rooms they worked for more than one hour per week);
- subjects must have a smartphone running Android or an Apple iPhone;
- subjects must give the static Bluetooth MAC address of their smartphone to the researchers for the purpose of the research;
- subjects must be willing to turn / leave on their Bluetooth on their smartphone when they were at Windesheim.

We only installed a measurement device (see section 5.4.2) in an office room if more than 75% of the known number of regular office room occupants provided informed consent. We never tracked presence via Bluetooth without informed consent; the measurement devices we used are technically incapable of tracking Bluetooth based presence without a static Bluetooth MAC-address.

5.4.2 Measurement techniques

We measured data in 6 office rooms at Windesheim. As illustrated in Figure 5.5, we used the following measurement techniques to collect data (in square brackets we mention the source name we used in the open dataset to identify this technique and in subsequent sections, we describe what properties the various sources measure and how they measure it in more detail):

1) [`CO2-meter-SCD4x`] M5Stack CoreInk SCD41 CO_2 and occupancy measurement device for temperature, humidity, CO_2 and occupancy data;
2) [`bms`] Building management system for temperature, PIR, CO_2 concentration and valve fraction data;
3) [`xovis`] Xovis PC2SE 3D people counting sensor for occupancy data;
4) [`human_observer`] Human observers for occasional occupancy observations and observations of open/closed status of doors and windows;
5) [`knmi`] Royal Netherlands Meteorological Institute (KNMI) for weather data.

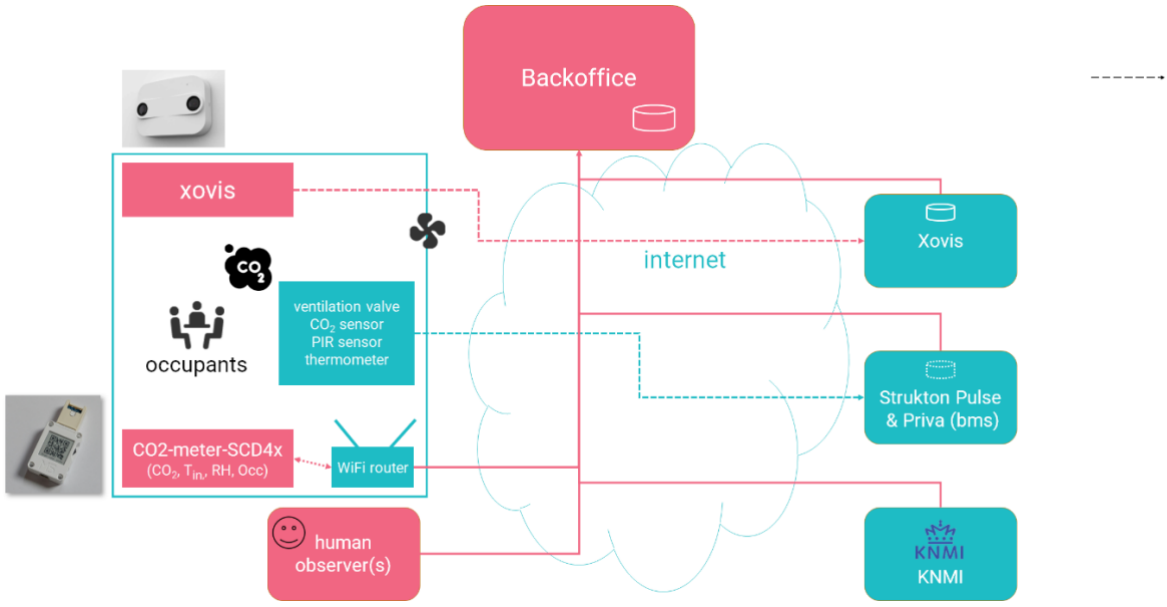


Figure 5.5: Measurement techniques used for data collection.

Ad 1) M5Stack CoreInk SCD41 CO_2 and occupancy measurement device

Based on the M5Stack CoreInk device and Seeed Grove SCD41 module mounted inside a an M5Stack Proto HAT, totalling ~ € 85 per room in hardware, we developed our own open source firmware (Twomes CO_2 and Bluetooth occupancy firmware for the M5CoreInk and SCD41, 2023) that could measure and upload CO_2 concentration and occupancy, as illustrated in Figure 5.6 on the left. The Sensirion SCD41 sensor is a novel, energy-efficient photo-acoustic CO_2 measurement sensor that has an accuracy ±(40 ppm + 5% MV) @400-5000 ppm.

The occupancy measurement is based on Bluetooth name requests and is inspired by the monitor program (Freyer, 2023). This technique requires the measurement device to know of the static Bluetooth MAC address of the smartphone of the occupants that are to be included in the occupancy count. The name request (LMP Layer Tutorial, n.d.) is a primitive of the Bluetooth GAP LWP-layer that all smartphones respond to automatically when Bluetooth is on (also when the device is NOT in discoverable mode). The original purpose of this primitive is after initiation of device discovery by a device, which is responded to only by devices in discoverable mode, with their static Bluetooth address. Subsequently, the initiator issues the name request. We only issue the name request to the static MAC address of registered smartphones, i.e. smartphones whose owners have provided us with informed consent to their presence being tracked and who have given us their static Bluetooth address as part of the recruitment procedure.

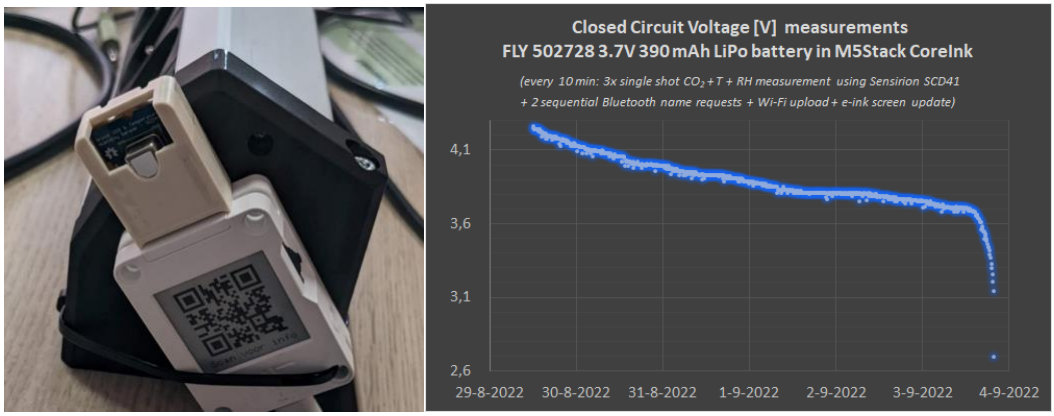


Figure 5.6: M5Stack CoreInk SCD41 measurement device for CO_2 and occupancy (~ € 85).

As illustrated in Figure 5.6 on the right, we designed the device to be very efficient in order to run autonomously on the embedded 390 mAh battery. Without a battery extension module, we could achieve slightly more than 5 days of runtime. Given the tight energy budget of 390 mAh (@3.7 V), 5 days of runtime is relatively long for a CO_2 sensor with a measurement interval of 10 minutes that also performs occupancy detection, but still not sufficient for our purposes. Therefore, we powered the measurement devices in this study using a 230V USB-

C power adapter. Based on our request, M5Stack now also has a CO2L unit available based on the SCD41 sensor, so if researchers need a battery operated solution that works for longer than 5 days, the open source software is available, and they would need to design their own battery extension module.

Ad 2) Building management system

We collected data from six rooms located in two different buildings, each equipped with a different building management system: Priva or Strukton Pulse. While we verified we earlier were able to obtain ventilation data for both buildings, we encountered an issue with one of the buildings. In this particular building, we discovered that valve fraction data could not be exported retrospectively, i.e., beyond 24 hours after it was recorded. As a result, we lacked retrospective ventilation data for four out of the six rooms we measured.

For one of these four rooms, we could reliably conclude[6] that the ventilation system was always running at maximum, so we included this room. So, only three rooms remained for our subsequent analysis.

Ad 3) Xovis PC2SE 3D people counting sensor

In all 6 rooms, we also installed the Xovis PC2SE 3D people counting sensor (see Figure 5.7), totalling ~ € 660 per room in hardware, mounted on the ceiling above the doorway entrance, with firmware set to privacy level 3, which guarantees that no traits or characteristics that could identify a person leave the device. Due to logistical issues, the Xovis devices were installed during the second week of the study period, yielding only slightly more than a week of Xovis data.

Figure 5.7: Xovis PC2SE People Counting Sensor for detecting occupancy (~ € 660).

Ad 4) Human observers

Several times per week, a human observer would walk by the studied rooms and observe and register in an Excel file:

- the room id;
- the time of the observation;
- the number of persons in the room;
- is at least 1 window in the room open? (yes/no);
- is at least 1 door in the room open? (yes/no).

Ad 5) Royal Netherlands Meteorological Institute (KNMI)

Average hourly weather data was collected from the Royal Netherlands Meteorological Institute (KNMI) and geospatially interpolated using Python code from the `HourlyHistoricWeather` repository (Peters, 2021/2022). For geospatial interpolation of weather data we used `lat = 52.499255`, and `lon =`

[6] Following the COVID-19 pandemic, ventilation systems at Windesheim were adjusted to maximize air exchange by setting the ventilation setpoint to 400 ppm. However, we encountered an issue with three out of the four rooms in the affected building. The room CO_2 sensors connected to the building management system in these rooms were not properly calibrated. As a result, these sensors frequently reported CO_2 concentration values below 400 ppm, which seemed highly unlikely when compared to the Keeling Curve data (Monroe, n.d.) from the last two years. Consequently, we were unable to obtain valve fraction data, except for one room where the CO_2 sensor was properly calibrated, allowing us to assume reliable and continuous ventilation.

6.0765167, the location of Windesheim in Zwolle, the Netherlands. Temporal interpolation was done as part of the preprocessing of the data.

5.4.3 Data collected

We measured data for 3 weeks between October 10 and November 2, 2022. This included (on purpose) the autumn holiday period, which lasted from 15-23 October 2022. During this period, typically very few people work in the office rooms of Windesheim.

In Table 5.1, we provide an overview of the data of the characteristics of the rooms for the analysis. In Table 5.2, we provide a catalogue of the data we collected. Finally, in Table 5.3, we provide an overview of the number of measurements we collected.

Table 5.1: Room data.

ID	#work-places	#occupants tracked via Bluetooth	Volume [m³] `room_m3`	Max ventilation rate [m³/h] `vent_max__m3_h_1`	Max Air Changes Per Hour (ACPH) [h⁻¹] `air_changes_max_vent__h_1`
999169	6	5	75	240	3,2
917810	6	2	75	240	3,2
925038	4	3	60	240	4,0

As can be observed from Table 5.1, for room with ID 917810 we tracked only 2 regular occupants via Bluetooth name requests using the `CO2-meter-SCD4x` measurement device. During the room selection and measurement rollout process, we included this room assuming it had only 3 workplaces. Only very late in the analysis process, we discovered that the room actually had 6 workplaces. This means we excluded this room for analysis using `CO2-meter-SCD4x` as source for occupancy data, leaving only 2 rooms for the analyses described in section 5.5.6.

Nevertheless, all data that we collected in these 3 rooms are available as open data (Ter Hofte et al., 2022/2023). To reduce the chance of room re-identification to an acceptable level and increase privacy, we rounded room volume to 5 m³ and the maximum ventilation rate to 30 m³/h, both in the table and in the analysis. We chose a level of privacy for the room that is equivalent to the level of privacy required for persons participating in medical research in the Netherlands, i.e., the chance of re-identification must be less than 9%.

Table 5.2: Measurement data catalogue.

Source	Property	Unit	Sensor	Interval [h:mm:ss]	Description
CO2-meter-SCD4x	co2__ppm	ppm	SCD41	0:10:00	CO_2 concentration
	temp_in__degC	˚C	SCD41	0:10:00	air temperature
	rel_humidity__0	-	SCD41	0:10:00	relative humidity
	occupancy__p	-	ESP32	0:10:00	number of smartphones responding to Bluetooth name request
	heartbeat	-	ESP32	0:10:00	measurement system heartbeat
bms	co2__ppm	ppm		0:01:00/1:00:00 [7]	CO_2 concentration
	temp_in__degC	˚C		0:01:00/1:00:00 [7]	air temperature
	rel_humidity__0	-		0:01:00/1:00:00 [7]	relative humidity
	occupancy__bool	0/1	PIR	0:01:00/1:00:00 [7]	is a person in the room?
	valve_frac__0	0		0:01:00/1:00:00 [7]	opening fraction of ventilation valve (0=fully closed, 1=fully open)
xovis	occupancy__p	-	PC2SE	0:05:00/0:15:00 [7]	number of persons in room
human_observer	occupancy__p	-		few per week	number of persons in room
	door_open__bool	0/1		few per week	is a door open?

[7] The measurement interval differed per room.

Source	Property	Unit	Sensor	Interval [h:mm:ss]	Description
	window_open__bool	0/1		few per week	is a window open?

Table 5.3: Measurement data catalogue: number of measurements per source (columns) and property (rows).

property\source	CO2-meter-SCD4x	bms	xovis	human_observer	Total
co2__ppm	8 989	69 563	-	-	78 552
occupancy__p	8 991	-	3 740	17	12 748
valve_frac__0	-	69 563	-	-	69 563
temp_in__degC	8 989	69 563	-	-	78 552
rel_humidity__0	8 989	68 892	-	-	77 881
occupancy__bool	-	68 892	-	-	68 892
window_open__bool	-	-	-	17	17
door_open__bool	-	-	-	17	17
Total	35 958	346 473	3 740	51	386 222

5.5 Analysis

In this section, we describe the analysis tool we used (5.5.1), the error metrics used in the analysis (5.5.2), synthetic room data generation (5.5.3), analysis of the synthetic room data (5.5.4), preprocessing of the real room data 5.5.5) and analysis of the real room data (5.5.6).

5.5.1 Analysis tool: GEKKO Python

We utilized GEKKO Python (Beal et al., 2018), which is an object-oriented Python interface that enables local execution of APMonitor (Hedengren et al., 2014). APMonitor is optimization software designed to solve a range of problems, including mixed-integer and differential algebraic equations, using different types of nonlinear solvers for constrained and unconstrained problems. It offers various operation modes (Hedengren & Beal, 2022) for linear programming, quadratic programming, integer programming, multi-object nonlinear optimization, as well as simulation, estimation, and control for different types of dynamic problems.

GEKKO Python is an open-source library available under the MIT license, suitable for use in both academic and commercial environments. In the context of inverse grey-box modelling, GEKKO proves to be highly versatile. With the provision of physics equations and a preprocessed dataset, GEKKO can determine the optimal value for a target parameter. In our research, the objective was to find the best value for occupancy/ventilation rate in utility buildings using time-series observed measurements. For this purpose, the Estimation for Simultaneous Dynamic mode, also known as Moving Horizon Estimation (`IMODE=5`), was selected as the most suitable. This mode is specifically designed for dynamic estimation, encompassing both states and parameter regression. Unlike deterministic approaches like the Kalman filter, Moving Horizon Estimation requires an iterative process that relies on linear programming or nonlinear programming solvers to find an optimal solution.

We implemented the molar CO_2 concentration balance model described in section 5.3.1 in GEKKO Python as:

- one ordinary differential equation,
- 7 intermediate variables,
- one control variable (`co2_ppm`),
- one fixed variable (`A_inf__m2`), and
- two manipulated variables (`valve_frac__0` and `occupancy__p`).

As illustrated in Figure 5.8, the GEKKO Python implementation of the ordinary differential equation and the 7 intermediate variables consists of a straightforward transformation of the mathematical equations described in section 5.2.

```
# GEKKO - Equations
co2_elevation__ppm = m.Intermediate(co2__ppm - co2_ext__ppm)
air_changes_vent__h_1 = m.Intermediate(valve_frac__0 * air_changes_max_vent__h_1)
air_changes_inf__h_1 = m.Intermediate(A_inf__m2 * wind__m_s_1 * s_h_1  / room__m3)
co2_loss_vent__ppm_s_1 =  m.Intermediate(co2_elevation__ppm * air_changes_vent__h_1 / s_h_1)
co2_loss_inf__ppm_s_1 =  m.Intermediate(co2_elevation__ppm * air_changes_inf__h_1 / s_h_1)
co2_loss__ppm_s_1 = m.Intermediate(co2_loss_vent__ppm_s_1 + co2_loss_inf__ppm_s_1)
co2_gain__ppm_s_1 = m.Intermediate(occupancy__p * co2_exhale__umol_p_1_s_1 / (room__m3 * air__mol_m_3))
m.Equation(co2__ppm.dt() == co2_gain__ppm_s_1 - co2_loss__ppm_s_1)

# GEKKO - Solver setting
m.options.IMODE = 5
m.options.SOLVER = 3
m.options.EV_TYPE = ev_type    # specific objective function (1 ~ MAE; 2 ~ RMSE)
m.options.NODES = 2
m.solve(disp = False)
```

Figure 5.8: Implementation of the core of the molar CO_2 concentration balance model in GEKKO Python.

The complete implementation of our analysis code is available under an open source license on GitHub (*Twomes Inverse Grey box Modelling and Analysis Tools for Homes and Utility Buildings*, 2021/2023).

5.5.2 Error metrics used in analysis

For all analyses, we calculate the mean absolute error (MAE) and Root Mean Squared Error (RMSE) between the measured and learned CO_2 concentration values.

For our synthetic room data analyses (see section 5.5.3.) we know the actual value for A_{inf} [m^2], so we can calculate the MAE and RMSE between the learned and actual values for A_{inf} over various time windows in our study. We cannot compute these error metrics for the real room data (see section 5.5.5), because we don't have an actual infiltration value (which might have been derived from a blower door test, which we did not do).

In our research setup, we can calculate the MAE and RMSE error metrics for the time-varying learned variables: not only for synthetic room data analyses, but also for real room data analyses. For ventilation, which can have jittery values, in particular for the synthetic room date, we calculate a Rolling Mean Absolute Error (RMAE) with a window size of 2 intervals of 15 minutes. Calculating error metrics allows us to express our results quantitatively for both research questions:

A. `rmae_valve_frac__0` and `rmse_valve_frac__0` are quantitative indicators for the extent to which ventilation rates can be derived reliably from CO_2 concentration and occupancy data;
B. `mae_occupancy__p` and `rmse_occupancy__p` are quantitative indicators for the extent to which occupancy can be derived reliably from CO_2 concentration and ventilation data.

To the quantitative (R)MAE and RMSE error metrics, we also added a qualitative judgement, as indicated in Table 5.4, below. We set norms and translate this to the following colors:

- good (colored green): deemed suitable for application, no further improvement needed;
- questionable (colored orange): might be suitable for application, further improvements desired;
- bad (colored red): not deemed suitable for application, improvement desperately desired.

Table 5.4: Qualitative judgement of mean absolute error metrics.

judgement\metric	mae_co2__ppm	mae_occupancy__p	rmae_valve_frac__0
good	<= 25	<= 1.0	<= 0.1
questionable	25 < mae_ <= 75	1.0 < mae_ <= 2.0	0..1 < mae_ <= 0.2
bad	>75	> 2.0	> 0.2

For the `mae_co2__ppm` error metric, we based our judgment on the accuracy of the Sensirion SCD41 sensor that we used: ±(40 ppm + 5% MV) @400-5000 ppm, which is fine for a regular office environment. Our raw measurement values ranged from 284-1221 ppm, hence the mean absolute error (MAE) between the measured value and the true CO_2 concentration in the room must have been in the range 52-99 ppm. We did not find model error norms, so we defined the norms relative to the sensor accuracy. We deemed trying an `mae_co2__ppm` error below 25 ppm to be a good fit (the model's error is definitely smaller than the measurement error of the sensor) and above 75 ppm to be a bad fit (the model's error is likely larger than the measurement error of the sensor).

The `mae_occupancy__p` error metric we used is motivated primarily by the application of occupancy monitoring in capacity planning and management of classrooms and meeting rooms and decisions about

building restructuring or renovation by facility managers. For such applications, achieving precision below 1.0 person might not be necessary. However, we considered being off by more than 2 persons as a potentially bad deviation warranting further improvements. It is important to note that we did not consult with Windesheim facility managers regarding their specific requirements or expectations for accuracy. Moreover, for practical purposes, we performed the measurements for this exploratory research in office rooms, not in classrooms and meeting rooms.

Concerning the `rmae_valve_frac__0` error metric, our focus was on understanding the role of `valve_frac__0` in distinguishing heat losses caused by ventilation in buildings from other forms of heat losses, such as infiltration and conduction. However, it is important to note that we did not gather specific data in this study to directly analyze heat losses themselves. Consequently, our primary objective was to assess the feasibility of estimating ventilation losses for this purpose, rather than directly evaluating the effectiveness of this approach in distinguishing ventilation-related heat losses from other sources. In terms of evaluating the accuracy of our estimations, we established criteria based on our judgment and expertise. We determined that a deviation of less than 10% would be satisfactory, indicating a reasonable estimation. Conversely, a deviation exceeding 20% would be deemed unsatisfactory, as it would suggest a significant discrepancy. It is crucial to acknowledge that these criteria are educated guesses and not based on a formal assessment of the effectiveness of the indirect estimation approach in achieving the goal of differentiating ventilation-related heat losses from other heat losses.

5.5.3 Synthetic room data generation

To test the implementation of our CO_2 concentration model and learning tool chain in GEKKO Python, we generated synthetic data for a set of imaginary rooms and ventilation systems that behave exactly according to the model. Subsequently, we analysed this synthetic room data and verified to what extent we could learn back the time-independent variable A_{inf} [m^2] and time-dependent variables (either ventilation flow rate or occupancy). The synthetic data is available on GitHub as part of our open source inverse grey box analysis software (*Twomes Inverse Grey box Modelling and Analysis Tools for Homes and Utility Buildings*, 2021/2023).

We first implemented the model described in section 5.3.1 in Microsoft Excel, with 5 rooms that varied in characteristics, which we also encoded in the room id (see Table 5.5).

The maximum number of Air Changes Per Hour of the synthetic rooms varied between 0.6 and 11.1 [h^{-1}].

To simulate the effect of wind-induced air infiltration, we set a minimum air infiltration rate (vent_min [m^3/h]) and assumed a fixed wind speed of 3.0 [m/s]. We calculated the infiltration area A_{inf} [cm^2] according to our simplified model that assumes a linear relation between wind speed and ventilation: A_{inf} [cm^2] = vent_min [m^3/h] / (wind [m/s] × 3600 [s/h]) × 10000 [cm^2/m^2]. The infiltration area of the synthetic rooms varied between 9 and 19 [cm^2].

Table 5.5: Synthetic room characteristics

id	room__m3	vent_max__m3_h_1	air_changes_max_vent__h_1	vent_min__m3_h_1	A_inf__cm2
100010045	45	100	2,2	10	9
100020045	45	100	2,2	20	19
100020180	180	100	0,6	20	19
500010045	45	500	11,1	10	9
500010090	90	500	5,6	10	9

Dynamics were defined by:

- a fixed weekday occupancy schedule as defined in Table 5.6. and no occupancy in weekends;
- a simple ventilation control algorithm that would operate at full capacity whenever the CO_2 concentration would exceed 800 [ppm], with a hysteresis of 50 [ppm], meaning ventilation would start when the CO_2 concentration surpasses 800+50 = 850 [ppm] and stop when it falls below 800 – 50 = 750 [ppm]); for each timeslot, we would register the fraction of time the ventilation operates at full capacity as the value for `valve_frac__0`.

Table 5.6: Synthetic weekday occupancy schedule.

starttime	endtime	occupancy
00:00	08:30	0

starttime	endtime	occupancy
08:30	09:00	1
09:00	12:00	3
12:00	12:30	2
12:30	13:00	0
13:00	13:30	2
13:30	16:00	3
16:00	17:00	2
17:00	00:00	0

In Figure 5.9, we plotted the resulting synthetic data for two of the five different synthetic rooms.

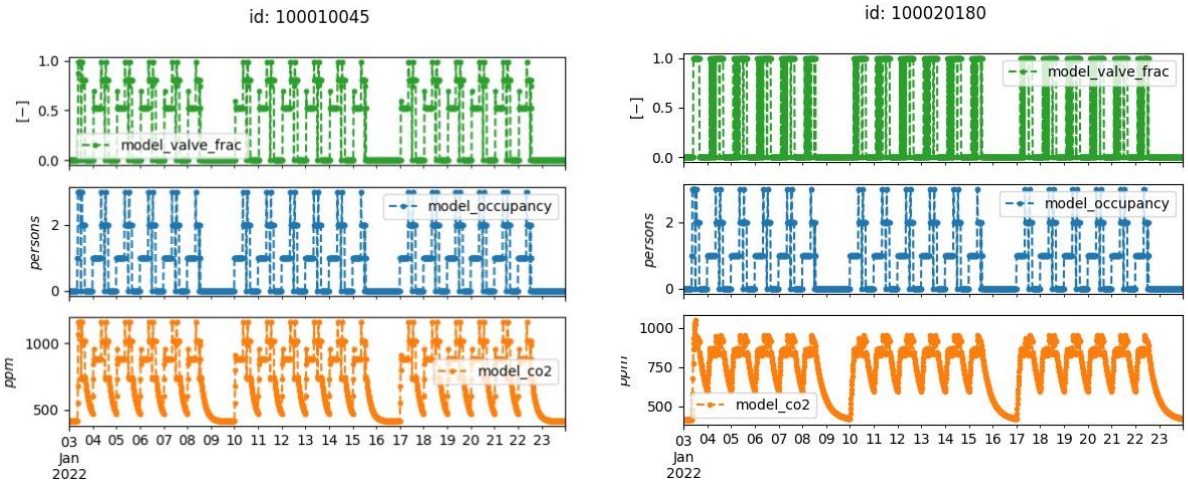


Figure 5.9: Plot of measurement data for two different synthetic rooms.

5.5.4 Analysis results of synthetic room data

Using our GEKKO Python inverse grey box model analysis tools, we learned the infiltration area as well as the ventilation flow rates, for all 5 synthetic rooms by fitting on the CO_2 concentration (see Figure 5.10 for an illustration of the fit). This takes about 35 seconds on a modern office laptop.

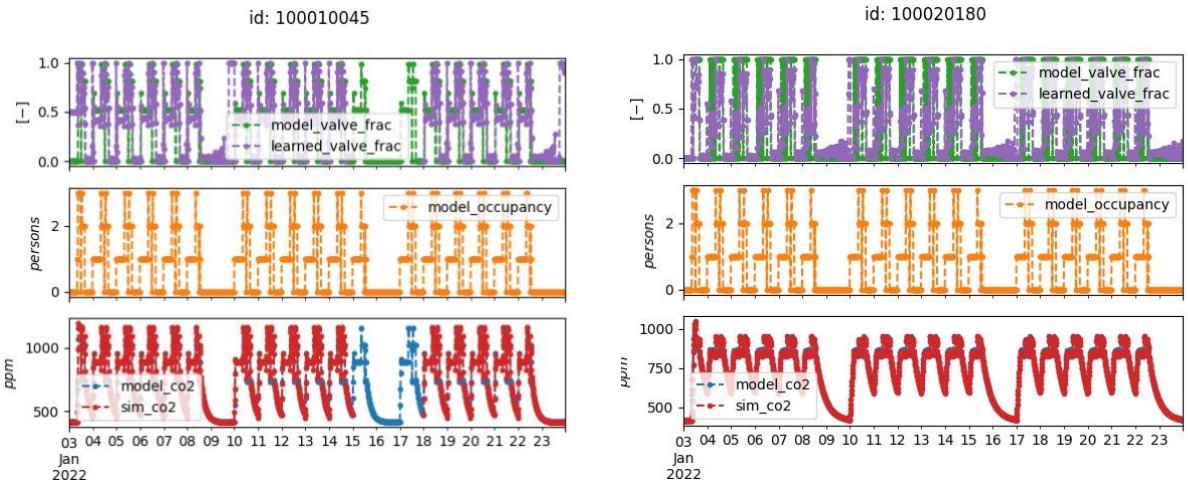


Figure 5.10: Plot of learned CO_2 concentration (`sim_co2`) versus 'measured' CO_2 concentration (`model_co2`), for two different synthetic rooms (subgraphs from top to bottom: valve fraction, occupancy, CO_2 concentration).

As can be observed from a closer look at two different parts of the synthetic room with id, 100020045 in Figure 5.11 (left), the `learned_valve_frac` resembles the `model_valve_frac` quite nicely, but starts to deteriorate once the CO_2 concentration nears the external CO_2 concentration, i.e. when the CO_2 elevation starts to near zero. In Figure 5.11 (right), we plotted the error between `learned_valve_frac` minus `model_valve_frac` for the same period. Apparently, the more the CO_2 elevation approaches zero, the less

the influence of ventilation is on the CO_2 concentration and the less the algorithm is punished for errors in its guess for `valve_frac`, i.e., the harder it becomes for the algorithm to derive the true `valve_frac` value.

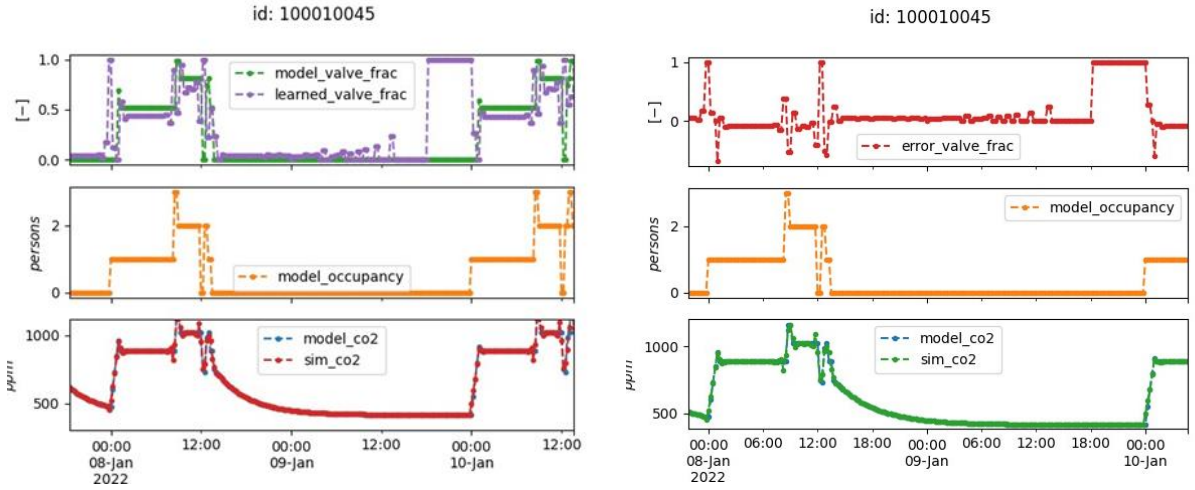
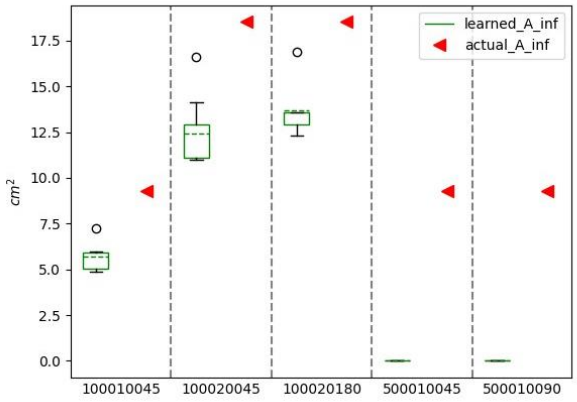


Figure 5.11: Close-up of three days: learned valve fraction versus measured (left); error between learned and measured (right).

In Figure 5.12 (top), we show a boxplot for the learned values for the infiltration area A_{inf} [cm^2], one per learning period of 3 days, compared to the actual values we used as input for the model that generated the synthetic data. For the three synthetic rooms with a maximum ventilation rate of 100 [m^3/h], the learned infiltration follows the actual infiltration area (with mean absolute errors of 3,6, 6,1 and 4.8 [cm^2], respectively). For the two synthetic rooms with a maximum ventilation rate of 500 [m^3/h], the learned infiltration rate is zero. We suspect that the lack of system excitation due to the assumed fixed wind speed, or a ventilation system that overpowers the effect of wind infiltration makes it harder to learn the infiltration area, but this is left for further study.

In Figure 5.12 (bottom), on the right, we show the error metrics, also for the learned valve fractions. The fit on CO_2 concentration is good, and the mean of the rolling mean absolute error on the valve fraction is questionable, except for the two rooms with a maximum ventilation rate of 500 [m^3/h], where it is good (the same absolute error in ventilation flow rate corresponds to a smaller error in valve fraction for these two rooms). Again, we suspect that the lack of system excitation due to the assumed fixed wind speed makes it hard to learn the infiltration flow rates, but we left this for further study.



id		mae_A_inf__cm2	mae_co2__ppm	rmse_co2__ppm	rmae_valve_frac__0	rmse_valve_frac__0
100010045	mean	0.378283	9	22	16%	36%
	min	0.025604	7	19	12%	31%
	max	0.947470	12	25	24%	46%
100020045	mean	2.034108	9	22	26%	45%
	min	1.769273	6	18	19%	38%
	max	2.320384	11	24	47%	66%
100020180	mean	0.547937	3	8	14%	32%
	min	0.513490	3	7	10%	27%
	max	0.592027	4	8	15%	34%
500010045	mean	8.975556	9	21	9%	24%
	min	8.825548	6	17	5%	13%
	max	9.259259	11	24	20%	42%
500010090	mean	9.259259	5	11	4%	9%
	min	9.259259	4	10	3%	8%
	max	9.259259	6	13	4%	12%

		mae_A_inf__cm2	mae_co2__ppm	rmse_co2__ppm	rmae_valve_frac__0	rmse_valve_frac__0
all	mean	4.303879	7	17	14%	29%
	min	0.025604	3	7	3%	8%
	max	9.259259	12	25	47%	66%

Figure 5.12: Actual and learned values for infiltration area and valve fraction (top), including error metrics (bottom).

We performed a similar analysis for learning (infiltration area and) occupancy, for all 5 synthetic rooms (see Figure 5.13 for two plots that illustrate the fit). Looking at the graphs, we can observe that the estimation algorithm during occupancy regularly estimates one more occupant based on CO_2 concentration and valve fraction data, with an occasional spike at the beginning.

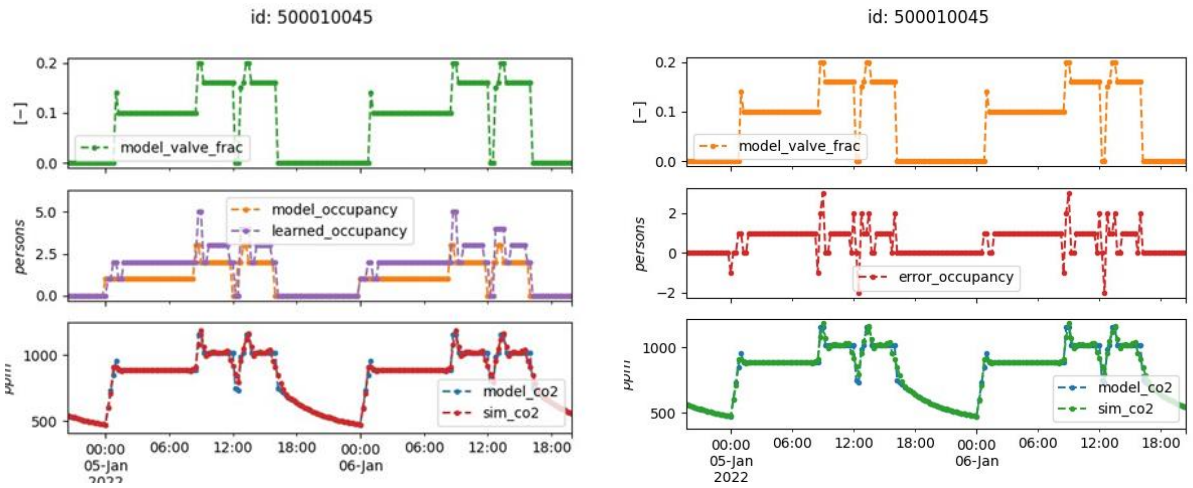
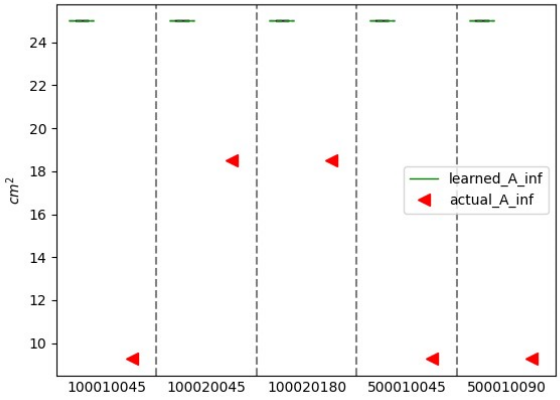


Figure 5.13: Close-up of 2 days: learned occupancy versus measured (left); error between learned and measured (right).

In Figure 5.14 (top), we show a boxplot for the learned values for the infiltration area A_{inf} [cm^2] for all 5 synthetic rooms and in Figure 5.14 (bottom), we show the error metrics, also for the learned occupancy. The infiltration areas were not learned well; the learning algorithm estimated all infiltration areas at the upper bound, which we set and 25 [cm^2]. Again, we suspect that the lack of system excitation due to the assumed fixed wind speed makes it hard for the system to learn the infiltration area properly. The fit on CO_2 concentration is good, and the mean of the mean absolute error on the occupancy is also good. Apparently, not learning infiltration areas well does not always interfere with learning occupancy.



id		mae_A_inf__cm2	mae_co2__ppm	rmse_co2__ppm	mae_occupancy__p	rmse_occupancy__p
100010045	mean	15.740741	7	19	0.3	0.6
	min	15.740741	5	16	0.2	0.5
	max	15.740741	9	22	0.4	0.6
100020045	mean	6.481481	7	21	0.3	0.6
	min	6.481481	5	16	0.2	0.5
	max	6.481481	9	23	0.3	0.6
100020180	mean	6.481481	3	7	0.2	0.5
	min	6.481481	2	6	0.2	0.4
	max	6.481481	3	8	0.3	0.6
500010045	mean	15.740741	7	19	0.3	0.6
	min	15.740741	5	16	0.2	0.5
	max	15.740741	9	22	0.4	0.6
500010090	mean	15.740741	5	13	0.3	0.5
	min	15.740741	4	11	0.2	0.4
	max	15.740741	7	16	0.3	0.5

		mae_A_inf__cm2	mae_co2__ppm	rmse_co2__ppm	mae_occupancy__p	rmse_occupancy__p
all	mean	12.037037	6	16	0.3	0.5
	min	6.481481	2	6	0.2	0.4
	max	15.740741	9	23	0.4	0.6

Figure 5.14: Actual and learned values for infiltration area and occupancy (top), including error metrics (bottom).

5.5.5 Preprocessing of real room data

As illustrated in Figure 5.15, the raw data we measured in real rooms contained various errors that required preprocessing.

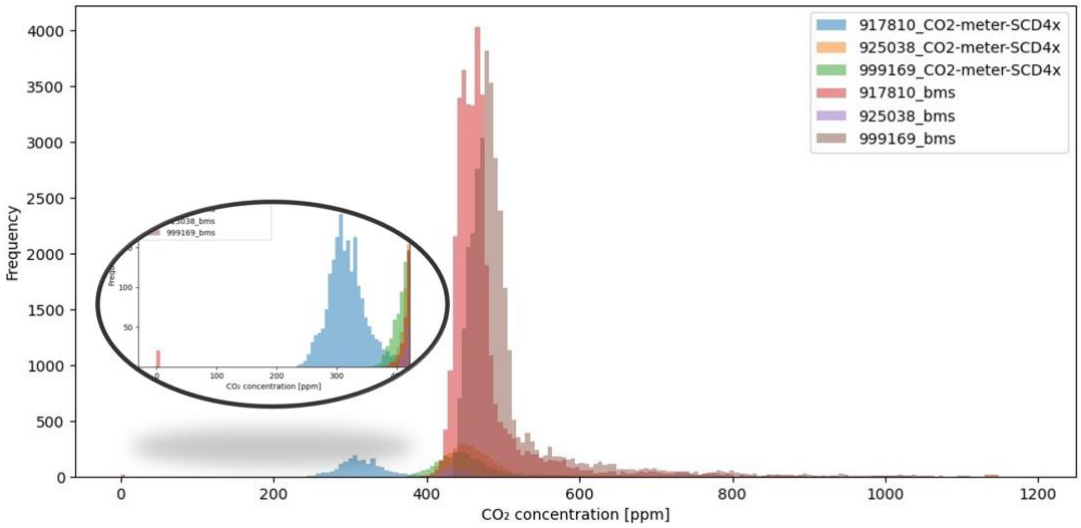


Figure 5.15: Histogram of CO_2 measurements for various rooms and measurement devices before preprocessing.

Preprocessing of measurements consisted of:

- Removal of duplicate measurements;
- Removal of CO_2 measurements that had no variation; this concerned one room that apparently had a faulty CO_2 sensor (the data for this room was already rejected for other reasons);
- Removal of all CO_2 concentration measurements with a value of less than 5 ppm (as illustrated in Figure 5.9, several CO_2 measurement values from the BMS came in as 0 ppm, which is clearly wrong);
- CO_2 baseline adjustment: per room and per measurement source, the minimum CO_2 measurement value was determined and subsequently all measurement values were raised by the same amount such that the minimum value would be to 415 ppm plus a margin (for which we chose 1 ppm). This preprocessing operation helps to counteract the effect of long-term drift that some CO_2 sensors are subject to. Some CO_2 sensors provide automatic occasional recalibration to a pre-determined CO_2 level. Not all CO_2 sensors used in this study may have had this feature, and some may have had this turned off (perhaps deliberately, to avoid sudden jumps). This baseline adjustment was also needed for the CO_2 sensors in our own `CO2-meter-SCD4x` measurement device, as we calibrated them, but not always in the same circumstances, unfortunately.
- Measurement interpolation to 15-minute intervals, except when they were 90 minutes or more apart.

Figure 5.16 provides a visual representation of the impact of preprocessing on the distribution of CO_2 measurement values, which seem much more reasonable than the values represented in in Figure 5.15.

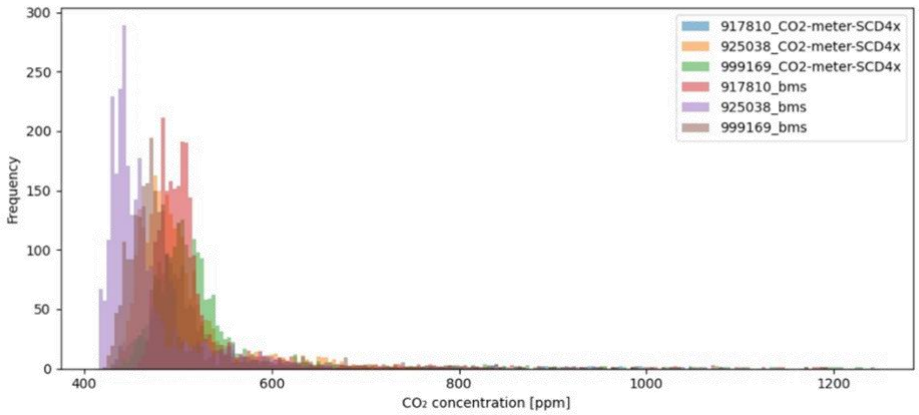


Figure 5.16: Histogram of CO_2 measurements for various rooms and measurement devices after preprocessing.

The plots of the most important preprocessed measurements for all rooms can be found in Figure 5.17. Note that for one room, we concluded that the ventilation was always set to the maximum value.

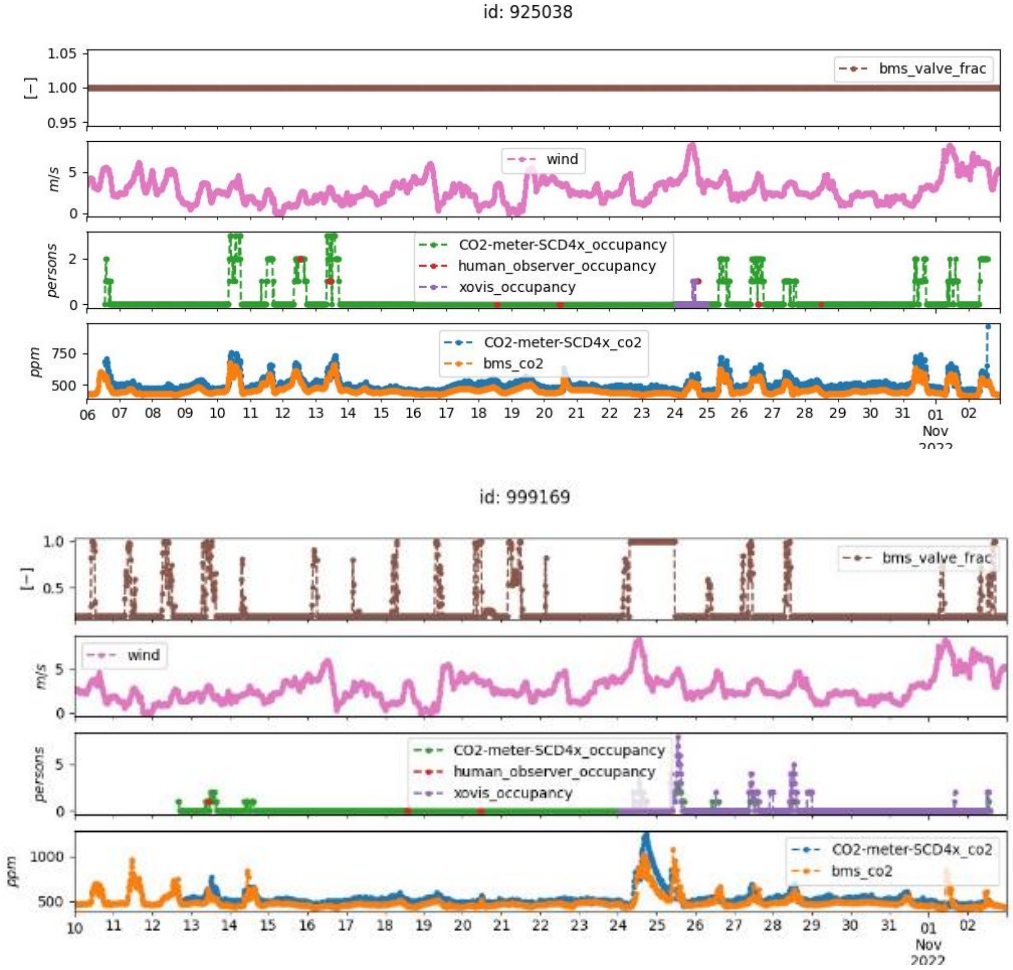


Figure 5.17: Plots of preprocessed measurements for various rooms.

5.5.6 Analysis results of real room data

The maximum number of Air Changes Per Hour of the real rooms, as calculated in Table 5.1, varied between 3.2 and 4.0 [h^{-1}]; the ventilation control algorithm and infiltration area A_{inf} [cm^2] of the rooms were unknown.

To get answers to research question A (Can we derive ventilation flow rate from CO_2 concentration and occupancy?) and research question B (Can we derive occupancy from CO_2 concentration and ventilation flow rate?), we used our GEKKO Python inverse grey box model analysis tools. We learned the infiltration area as well as the ventilation flow rates, for the 2 rooms by fitting on the CO_2 concentration (see Figure 5.18 for an illustration of the fit). This takes about 30 seconds on a modern office laptop.

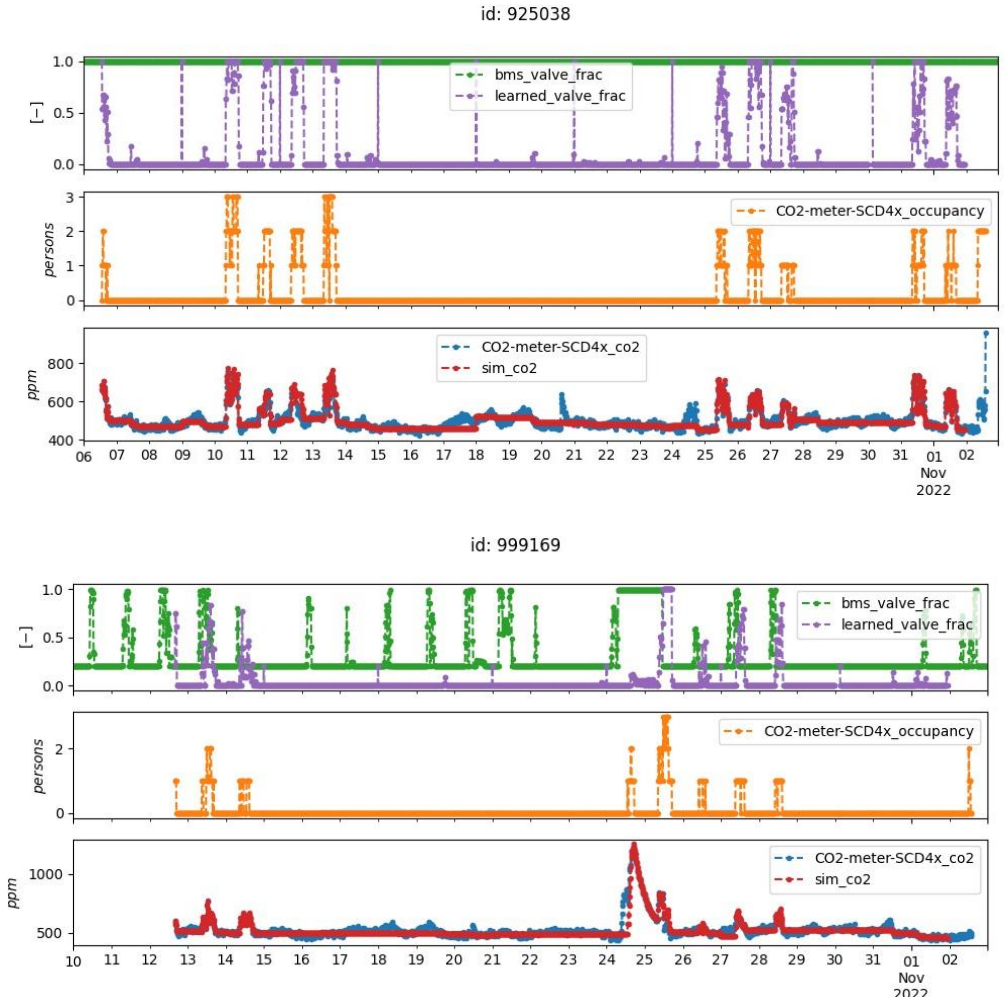


Figure 5.18: Plot of learned versus measured `valve_frac` for the two rooms studied.

A closer look at different timeframes of one of the rooms in Figure 5.19 reveals that the `learned_valve_frac` resembles the pattern observed `bms_valve_frac` at some times. However, the learned `learned_valve_frac` data seems to contradict the information we obtained that all rooms had a minimal ventilation flow rate that is 20% of the maximum air flow rate (valves were physically constrained; they could not close below 20%), Somewhat similar to the synthetic data, at very low CO_2 elevations, i.e. near the area where 'any `valve_frac` will do', the learning algorithm tends to set `valve_frac` to zero, whereas the ventilation flow rate should never go below 20% in both rooms, and never below 100% in one other room.

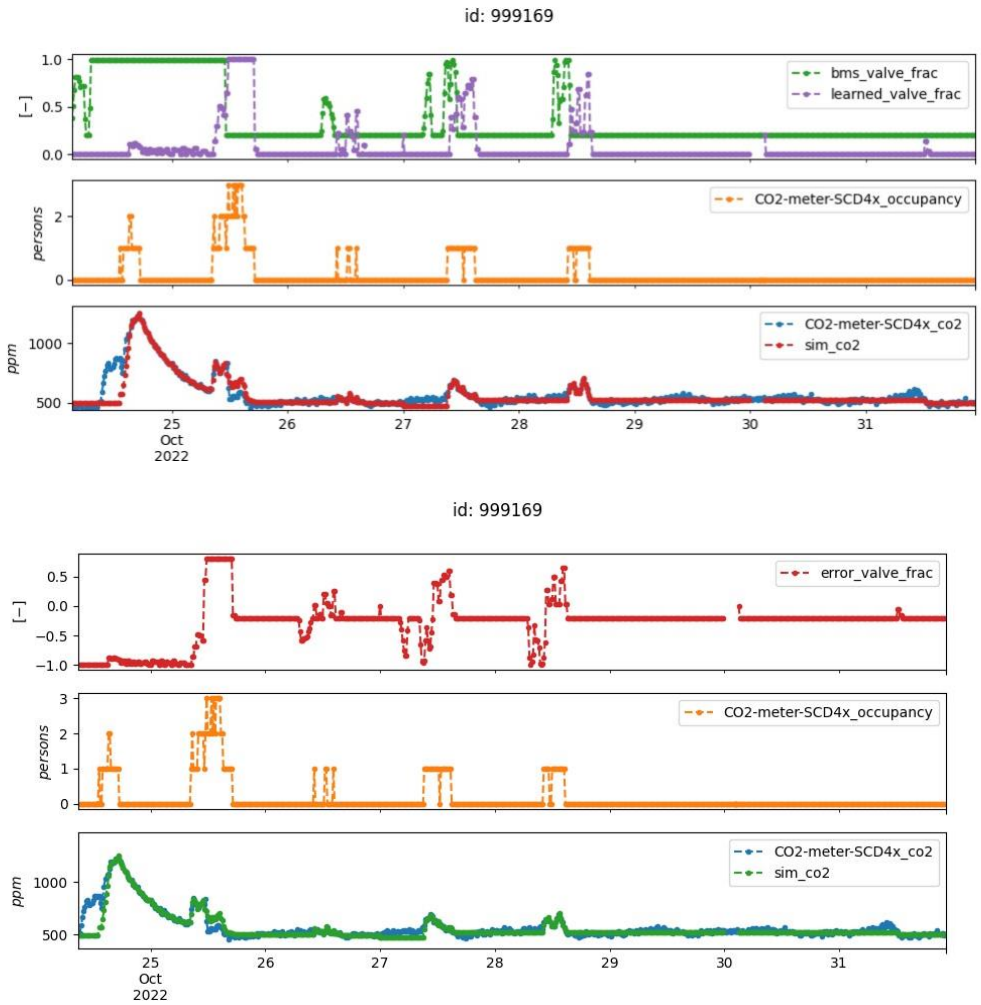
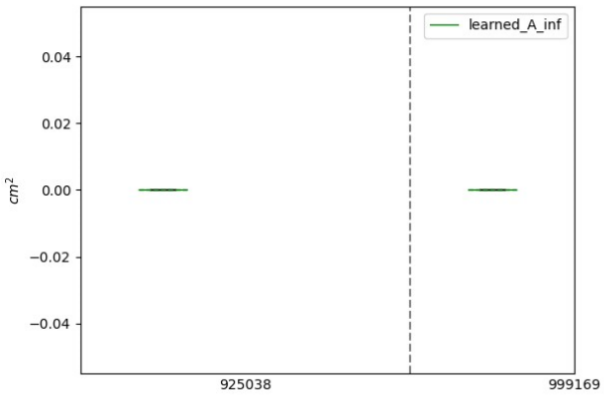


Figure 5.19: Close-up: learned vs. measured `valve_frac` (top); error between learned and measured (bottom).

As can be observed from the boxplot in Figure 5.20 (top), the learned values for the infiltration area A_{inf} [cm^2] of the real rooms studied were both zero. The cause of this phenomenon remains uncertain, and further investigation is required. One possible factor contributing to this outcome could be the absence of system excitation resulting from the assumed fixed wind speed. Alternatively, it is plausible that the ventilation system is overpowering the impact of wind infiltration. We left this issue for further study and incorporated geospatially interpolated wind speed data from KNMI into our dataset.

In Figure 5.20 (bottom), we show the error metrics for the learned valve fractions for the real rooms. The mean fit on CO_2 concentration is good, whereas the mean of the rolling mean absolute error on the valve fraction is bad.



id		mae_A_inf__cm2	mae_co2__ppm	rmse_co2__ppm	rmae_valve_frac__0	rmse_valve_frac__0
925038	mean	nan	15	21	90%	93%
	min	nan	8	10	75%	86%
	max	nan	22	33	100%	100%
999169	mean	nan	20	30	29%	35%
	min	nan	11	14	21%	23%
	max	nan	39	84	56%	66%

		mae_A_inf__cm2	mae_co2__ppm	rmse_co2__ppm	rmae_valve_frac__0	rmse_valve_frac__0
all	mean	nan	17	25	63%	68%
	min	nan	8	10	21%	23%
	max	nan	39	84	100%	100%

We tried several strategies to improve the `rmae_valve_frac__0` error metric, but none of these strategies provided a substantial improvement in the error metric (some even made it worse):

1. increasing the CO_2 baseline adjustment margin from 1 ppm to 50 ppm;
2. use occupancy data from `xovis` devices instead of from `CO2-meter-SCD4x` (which resulted in much less days worth of analyzable data);
3. use CO_2 concentration data from the `bms` data instead from `CO2-meter-SCD4x`.

We also attempted setting another lower boundary for `learned_valve_frac__0`. This improved the `rmae_valve_frac__0` error metric slightly, at the expense of a substantially worse `mae_co2__ppm`:

- applying a lower bound of 20%, the mean `rmae_valve_frac__0` improved from 55% to 37%, at the expense of the mean `mae_co2__ppm`, which increased from 18 ppm (good) to 80 ppm (bad);
- applying a lower bound of 10%, the mean `rmae_valve_frac__0` improved from 55% to 47%, at the expense of the mean `mae_co2__ppm`, which increased from 18 ppm (good) to 74 ppm (questionable).

A poor fit in ventilation flow rates is always undesirable. For example, *if* low CO_2 elevations mostly occur at with small temperature differences between indoor and outdoor environments (which may be the case at night), then a bad ventilation flow rate estimate is less problematic for estimating the contribution of ventilation to heat loss. Further research that links ventilation (derived from occupancy and CO_2 elevations) to heat loss explicitly would be required to determine this.

We performed a similar analysis when learning infiltration area and occupancy, for all rooms (see Figure 5.21 for two plots that illustrate the fit). Looking at the graphs, we can observe that the estimation algorithm occasionally takes the liberty to assume occupancy close to (5) or at (6) the upper bound that we defined for occupancy.

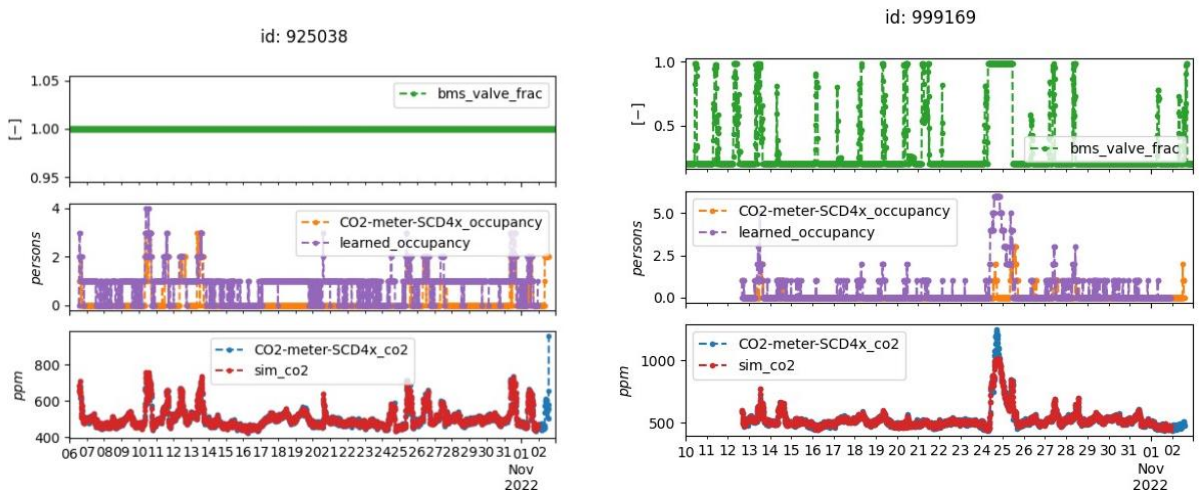


Figure 5.21: Plot of learned versus measured data for the two rooms studied after learning `occupancy__p`.

A closer look at two different timeframes of each of the rooms in Figure 5.22 reveals that `learned_occupancy` resembles the pattern observed by `CO2_meter_SCD4x_co2` at some times. However, whenever `bms_valve_frac` is high for a long time and the CO_2 elevation is low, the learning algorithm seems to make more errors. Another possible explanation is that the `CO2_meter_SCD4x_occupancy` in the two rooms studied were not allowed to track the one of the occupants (as listed in Table 5.1).

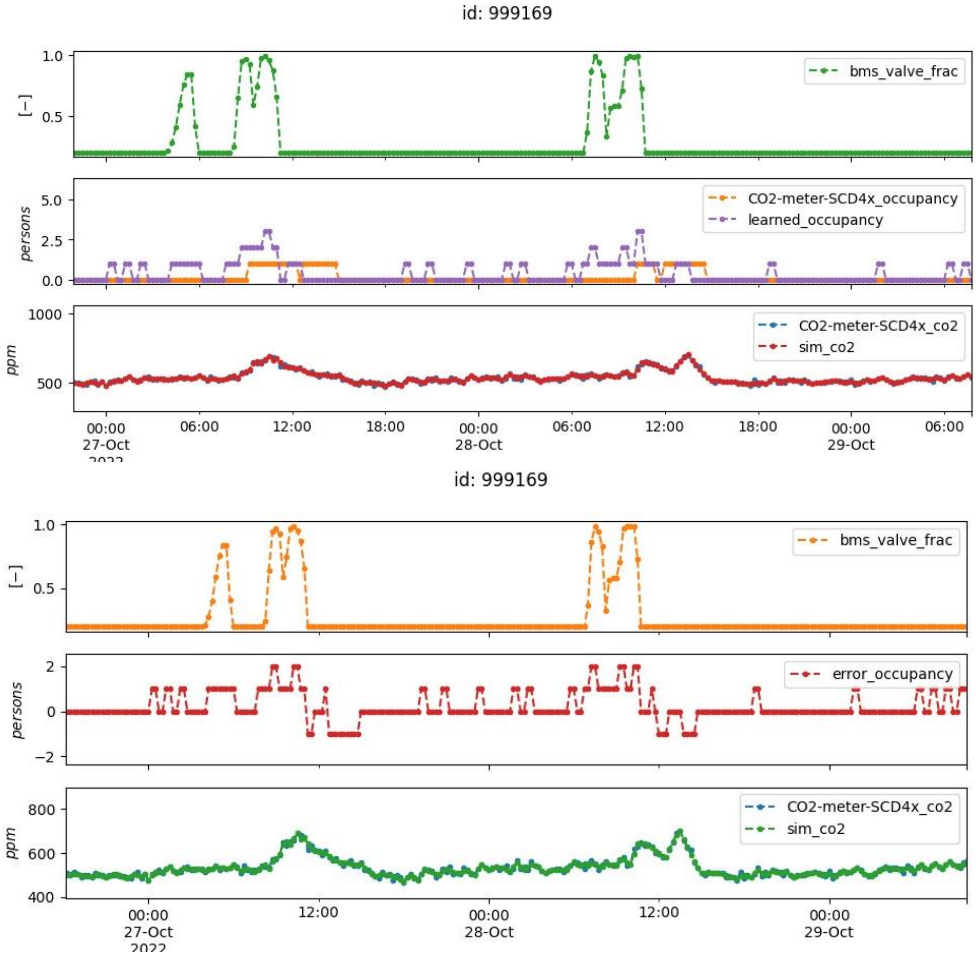
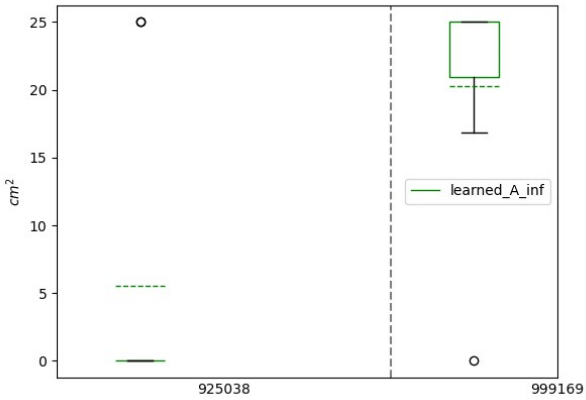


Figure 5.22: Close-up: learned vs. measured `occupancy` (top); error between learned and measured (bottom).

As can be observed from the boxplot in Figure 5.23 (top), the learned values for the infiltration area A_{inf} [cm^2] for both rooms are not zero, which seems realistic. We are unsure whether we can trust these numbers, since we assumed fixed wind speed in our analysis. We completed integrating geospatially interpolated wind speed

data from KNMI in our dataset. However, due to time constraints, this aspect could not be thoroughly investigated in the current study and warrants further examination.

In Figure 5.23 (bottom), we show the error metrics, also for the learned occupancy. The fit on CO_2 concentration is good, the mean of the mean absolute error on the occupancy is also good. With 4 – 6 regular occupants in the office rooms studied, a mean absolute error of 0.7 is considered good by the norms we set. How this scales to classrooms and meeting rooms would require further study, including asking more detailed norms from facility managers.



id		mae_A_inf__cm2	mae_co2__ppm	rmse_co2__ppm	mae_occupancy__p	rmse_occupancy__p
925038	mean	nan	5	6	0.7	0.7
	min	nan	4	5	0.5	0.6
	max	nan	6	8	0.8	0.9
999169	mean	nan	8	13	0.6	0.8
	min	nan	6	7	0.3	0.3
	max	nan	18	44	1.7	2.5

		mae_A_inf__cm2	mae_co2__ppm	rmse_co2__ppm	mae_occupancy__p	rmse_occupancy__p
all	mean	nan	6	9	0.6	0.7
	min	nan	4	5	0.3	0.3
	max	nan	18	44	1.7	2.5

Figure 5.23: Learned values for infiltration area and occupancy (top), including error metrics (bottom).

5.6 Summary, conclusions and recommendations

To answer the research questions, we performed various activities:

- We developed a molar CO_2 concentration balance model, which was implemented in GEKKO Python as one ordinary differential equation, 7 intermediate variables, one control variable (`co2_ppm`), one fixed variable (`A_inf__m2`) and two manipulated variables (`valve_frac__0` and `occupancy__p`). We made this implementation available under an open source license on GitHub (*Twomes Inverse Grey box Modelling and Analysis Tools for Homes and Utility Buildings*, 2021/2023).
- We built a data collection system to measure most features via multiple ways, comprising:
    a. an M5Stack CoreInk SCD41 CO_2 and occupancy measurement device (~€ 85 per room) running open source firmware we developed ourselves and made available on GitHub (*Twomes CO_2 and Bluetooth Occupancy Firmware for the M5CoreInk and SCD41*, 2021/2023), sending live data to our Twomes open source API and database server;

b. batch data import from Xovis PC2SE People Counting sensors for occupancy measurement (~ € 660 per room) running proprietary closed source firmware installed in the 6 offices specifically for our study;
c. batch data import from Priva BMS and Strukton Pulse BMS;
d. batch data import of observations made a few times per week by human observers that registered their observations in Excel;
e. batch data import from Royal Netherlands Meteorological Institute (KNMI).

- We collected data from 6 office rooms at Windesheim for 3 weeks in the autumn of 2022.
- We had to reject the data of 3 office rooms due to ventilation data not being available for export beyond 24 hours after recording, in combination with BMS CO_2 sensors not being calibrated properly.
- We published the dataset of the 3 office rooms, having 8 properties from 4 sources, totalling 0.4 million datapoints as open data under a Creative Commons license (Ter Hofte et al., 2022/2023).
- We had to exclude one more office room from analysis, due to lack of coverage (the smartphones of only 2 of the 6 occupants could be counted in Bluetooth-based occupancy measurement).
- We analysed the data of the remaining 2 rooms using an inverse grey box modelling approach in GEKKO Python.

Based on our analysis, we obtained the following answers to our research questions:

A. *To what extent can we reliably derive the ventilation flow rate in a room from data about CO_2 and occupancy?*
Using a molar CO_2 concentration balance model implemented in GEKKO Python to perform an inverse grey box analysis, we were able to learn ventilation flow rates from measured data on CO_2 concentration and occupancy. Learning ventilation flow rates often had bad results: the rolling mean absolute error in the learned valve fractions, which are indications of ventilation flow rate, was bad for both rooms (i.e., larger than 20%). (This, despite the fact that the fit of the model on CO_2 concentration was good i.e., the mean absolute error was less than 25 ppm). In particular, the error in the valve fraction seems to deteriorate when the CO_2 concentration approached the external CO_2 concentration, i.e., when CO_2 elevation was low. This is something that we also observed in the preliminary analyses we did with synthetic room data, where it led to rolling mean absolute errors that we deem questionable. As explained below, we recommend studying the relation between CO_2, occupancy and heat loss via ventilation more directly.

B. *To what extent can we reliably derive occupancy in a room from data about CO_2 and ventilation?*
Using a molar CO_2 concentration balance model implemented in GEKKO Python to perform an inverse grey box analysis, we were able to learn occupancy from measured data on CO_2 concentration and ventilation flow rates . The fit of the model on CO_2 concentration was good (i.e., the mean absolute error was less than 25 ppm). The mean of the absolute error for the learned occupancy was also good (i.e., less than 1), which means that we were able to derive occupancy data reliably from the measured data about CO_2 and ventilation. To determine if these findings are applicable in different situations and error criteria, such as in meeting spaces and classrooms, additional research would be necessary in these specific settings. This research would involve assessing them against standards established through discussions with relevant stakeholders.

Based on our experiences we have the following recommendations for further study:

– We recommend further study that determines the significance and timing of errors in learning valve fractions and ventilation rates. Specifically, when evaluating the impact of ventilation on heat loss independently of infiltration and conduction (which was one of the initial inspirations for our study), ventilation rate errors may not necessarily be problematic. These errors, which are more prevalent at lower CO_2 elevations, may not have adverse effects, e.g. *if* they mostly occur when the temperature difference between the indoor and outdoor environments is low (e.g., at night), thus potentially dampening the impact of ventilation estimation errors on ventilation heat loss estimation errors. We left this for further study.
– We recommend further study to investigate the potential improvements in the reliability of learning the wind sensitivity by incorporating a variable wind speed. In our study, we represented wind sensitivity as an infiltration area, which moderates the impact of wind speed on the rate of infiltration. However, we encountered challenges in learning the infiltration area when assuming a fixed wind speed in our models. Although we integrated geospatially interpolated wind speed data from KNMI into our dataset, time constraints prevented us from completing the analyses. Additionally, we recommend exploring whether a linear relationship between wind speed and infiltration flow rate is adequate or if a more complex model incorporating wind pressure on buildings is necessary.

- We recommend expanding the measurement of ventilation and infiltration in future experiments in a few ways. We recommend using additional ventilation and infiltration sensors, enabling triangulation of data, allowing identifying and discarding faulty sensor readings. We also recommend using additional sensors that detect things like open doors and windows automatically. This helps to assess how important these disturbances are. If important, they help to whether these additional sensors are also needed applications or the algorithms can be extended to detect things like opening and closing doors and windows without needing additional sensors.
- Furthermore, we advise thoroughly verifying the retrospective availability of data exported from building management systems (BMSs). Simply confirming the availability of recent data may not be sufficient. It is important to ensure the availability of historical data to support comprehensive analysis.

6 WHITE & BLACK BOX MODELS: PREDICTION OF THE HEATING DEMAND OF A BUILDING

Authors: David Fortini (O-Nexus)

6.1 Goal of the prediction models

The objective of O-Nexus is to develop a functional prototype for energy management in SME buildings. In pursuit of this goal, O-Nexus explored two distinct methods for heating forecast modeling: a physical equation-based approach (white box model) and a data-driven approach (black box model).

The models are firstly developed and tested on residential data, as O-Nexus did not have measurement data from SME buildings at the time of the analysis. In the final report, the models will be tested live in SME buildings.

With the physical equation-based approach the goal is to understand very well the relationship between outside temperature and the heating demand. In this way, we can learn about the dynamics happening in a dwelling related to the heating demand.

With the data-driven approach, the goal is to have a good heating prediction for the next day. This enables the O-Nexus system to create a smart strategy for energy steering so that it can optimize how much energy to store for the next day.

Both methods discussed in the current report are work in progress. We expect that their final form in the final report will be different from what it is now, as they will be enriched with more knowledge and also more data will be available to test them.

6.2 O-Nexus context

O-Nexus mission is to save 30% energy without changing comfort in buildings. To realize this goal O-Nexus is working on power steering to optimize energy flows in buildings. A residential house is quite similar to a small SME office as far as electricity usage is concerned. In fact, both buildings when having PV installed on the roof, can produce a lot of return energy at times and at the same time have energy shortages later on in the same day.

This creates many problems such as energy imbalance and putting the electricity grid under stress, as the peak moment for return energy or for using energy is mostly the same one for everyone. The current O-Nexus approach is to steer the excess energy in a thermal storage and use the thermal storage later to supply energy when needed.

Power steering is, among other technologies, an enabler to increase energy flexibility in buildings (Lizana et al., 2018) and it is often combined with a storage medium. Most residential dwellings can dramatically increase their energy flexibility with only a power steering device and a storage medium. In our setup, PV panels are also available.

The following research, analysis and results and project is carried out on residential houses as their energy related data was already available. The similarity of a dwelling with a small SME building makes it an ideal use case to test this new application of the O-Nexus power steering technology.

Our research has two main parts:

1. First a data analysis part, executed on a fitting subset the existing houses in the O-Nexus data. In this part the goal is to understand and forecast heating demand.
2. Second an implementation part, related to power steering and executed in the house of 2 friendly users. In this part the focus is on implementing power steering.

6.3 Purpose & research questions

The goal of O-Nexus is to have a working prototype for energy steering in SME buildings. To achieve this there is a main research question for each part.

For the first data analysis part:

– How can we predict the heat demand of a building given historical data? Is a physical strategy better or should we use Machine Learning?

For the second implementation part, the question is as follows:

– What is necessary to implement power steering on the O-Nexus Gateway? What are the main bottlenecks in doing so?

This second part includes connecting the Gateway to a storage device and steering it.

6.4 Measurement devices and hardware

The first step to implement power steering is to measure energy flows such as heating demand in a building and to check the PV production, energy from and to the grid etc. As O-Nexus has many hundreds of installations in residential houses, there are many different hardware implementations.

What they all share in common is that:

– Each measurement is taken every 5 minutes
– Most of the energy from and to the grid comes from a 3rd party (due to cost/implementation constraints)
– All the meters use Mbus or Modbus to communicate with the O-Nexus Gateway

For the implementation part, the requirements are slightly different as there are only 2 friendly users and we used our own Gateway to implement power steering:

– The shadow meter is communicating with the O-Nexus Gateway real time (it is not a 3rd party application) as it is necessary to do power steering
– There is a solid stare relay that enables electricity to flow to the storage device
– There is a storage device, a Flamco 6E Flextherm which can store 5 kWh of thermal energy (Flamco, n.d.)
– There is the O-Nexus Gateway
– The heat meters are all from Sharkey WMS (Wms, n.d.)
– The electricity meters are all from Finder (Findernet, n.d.)

Figgure 6.1 O-Nexus Gateway next to a watch as a reference point.

O-Nexus Gateway is purpose built to accommodate the following requirements:

1. Small, as it needs to fit in the fuse box of residential housing
2. Able to do power steering
3. Able to connect to many different devices, through ModBus, MBus or Solid-state relays.

The reasons to build, instead of buying, the O-Nexus Gateway were the following:

1. To have control over the hardware, its functionalities, its flexibility and its costs.
2. To be able to produce it cheaply on scale. Gateways with similar functionality (mbus, modbus, steering, LTE-network, Bluetooth, Wi-Fi) are priced easily around € 700, which is not competitive on the residentials market for this use case. O-Nexus is aiming at a much lower cost price.
3. To have a competitive edge against competitors.

Finally, the choice for the Flamco 6E Flextherm was made due mostly to space constraints. The Flamco 6E Flextherm is reasonably small in size, so it can fit in most dwellings (or SME offices to that extent). It can be cascaded with more Flextherms to obtain more storage and it is “simple” to steer with a relay. Also, it has high energy density, low heat loss and long lifetime compared for example to an electric battery. Also, the PCM buffer accommodates partial charging (say 30%) and still provide 60 °C water supply.

Other possible options would include a CV boiler storage, which is much cheaper and bigger. On the other hand, that has more heat loss and was not made to work always with the space heating system.

Another option would be an electric battery. This is the most expensive option, still big in size and a limited amount of charge/discharge moments compared to the other options.

In conclusion, for the O-Nexus use case the Flamco 6E Flextherm was found to be the ideal candidate as thermal heating medium for its power steering technology.

6.5 Data analysis - Overview

For the data analysis part, we aim to give a quick overview of the dataset used in the analysis.

In total there are 125 dwellings, where O-Nexus measures water and space heating, as well as energy from PV, energy from and to the net. The year of analysis is 2022 as it is a complete year for the dataset and contains clean data without interruptions or other data related problems. 121 dwellings were built in 2020, except for 4 that were built in 1975 but recently renovated.

All the dwellings have at least an A energy label, otherwise they have an A++.

The types of dwellings are well distributed (Figure 6.2), with each category having some representation. The dwellings footage (Figure 6.3) has a double peak one in around 80 and one around 180 square meters.

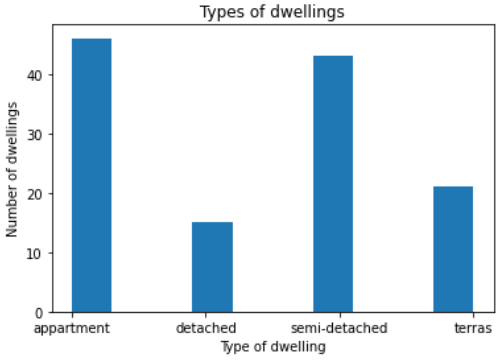


Figure 6.2 Dwelling types in the dataset

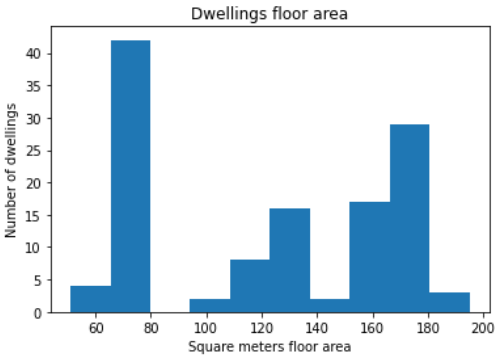


Figure 6.3 Dwellings's surface distribution in the dataset

To show the potential of power steering, a simple algorithm steering only excess PV power (return energy) into a thermal storage (and no further knowledge) on a daily basis saves on average 50% of the total heating consumption (fig 7.4). On the sample dataset, a dwelling consumes about 2000 kWh per year for space heating. This percentage can be easily improved upon, for example by incorporating more sophisticated algorithms using weather and heating forecast.

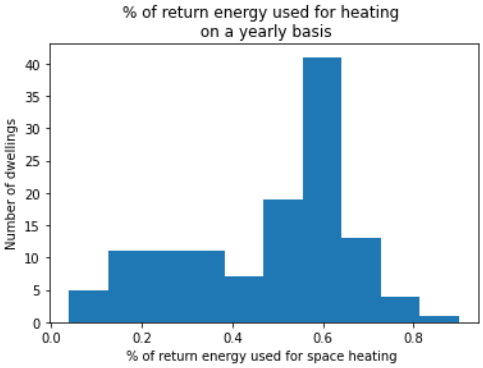


Figure 6.4 Percentage of excess PV that could be used for space heating.

6.6 Data analysis - Heating forecast

6.6.1 Physical modelling

For predicting heat demand in a building there is a wide array of models available. Mostly we can categorize them in either physical, statistical models or a hybrid approach. As a start, O-Nexus tried to use a physical model, in order to gain knowledge about the thermodynamics in a residential building and mostly to learn how to estimate a heat loss factor per building.

The advantage of using a physical model is that the model is white-box and understandable. It is known that for some buildings their heat demand dynamics could be too complex to model in this way meaning that a hybrid or black-box approach might be more suitable (Amara et al., 2015). Especially the flexibility of adapting to several different conditions is appealing when using a statistical model (Eseye & Lehtonen, 2020).

Building a physical model includes relying on very detailed information at scale (Frayssinet et al., 2018) and this leads often to the usage of simplified models as there are either too many factors, making it computationally expensive, or the information available is not as detailed as needed.

6.6.2 Physical modelling results

To come up with the final equation, we consulted several papers and tried to incorporate a mix of the information O-Nexus had available and information that papers recommend in 4 categories (Berisha & Dragusha, 2015):

- Climate data was easy to include as it is publicly available. Unfortunately, the data resolution is around 10 Km for publicly available sources, so it can introduce some noise.
- Material data is not present.
- Design data is limited to building footage, height in our dataset.
- Building usage data is not present, aside from the heat demand itself.

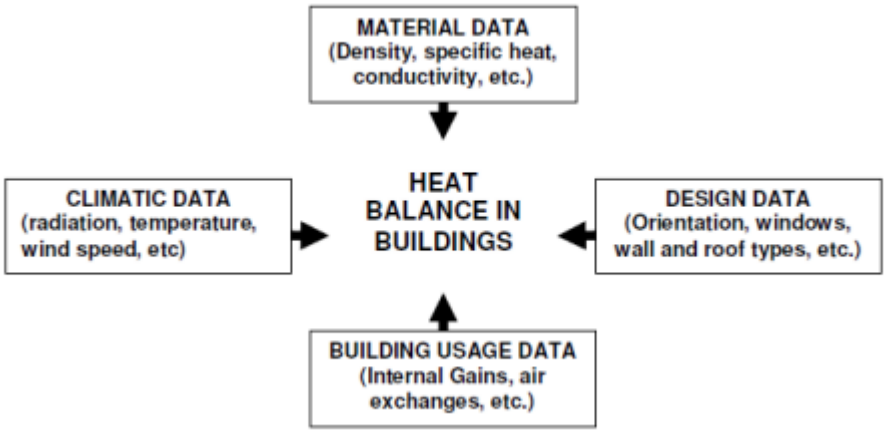


Figure 6.5 from (Berisha & Dragusha, 2015).

We could not come up with an equation that satisfied the situation all the time. Yet, the basic form was as follows:

Heat demand = (Surface of heat exchange / R value) * (Temperature Inside - Temperature Outside)

Where:

- R value = Approximated from energy label
- Surface of heat exchange = Estimated with footage and building type
- Heat demand = (Surface of heat exchange / R value) * (Temperature Inside - Temperature Outside)

Since all the values in the equations besides the temperature outside and inside are actually static, this equation assumes a very strong correlation between temperature and heat demand.

In Figure 6.6 we report the results of applying such an equation to a house. This dwelling had the lowest error of the dataset, which implies it also had the highest correlation between heating consumption and external temperature. Nonetheless, the real pattern (in violet) is not captured by the heating prediction (red), which means that the same model would perform very poorly in dwellings with more noisy data.

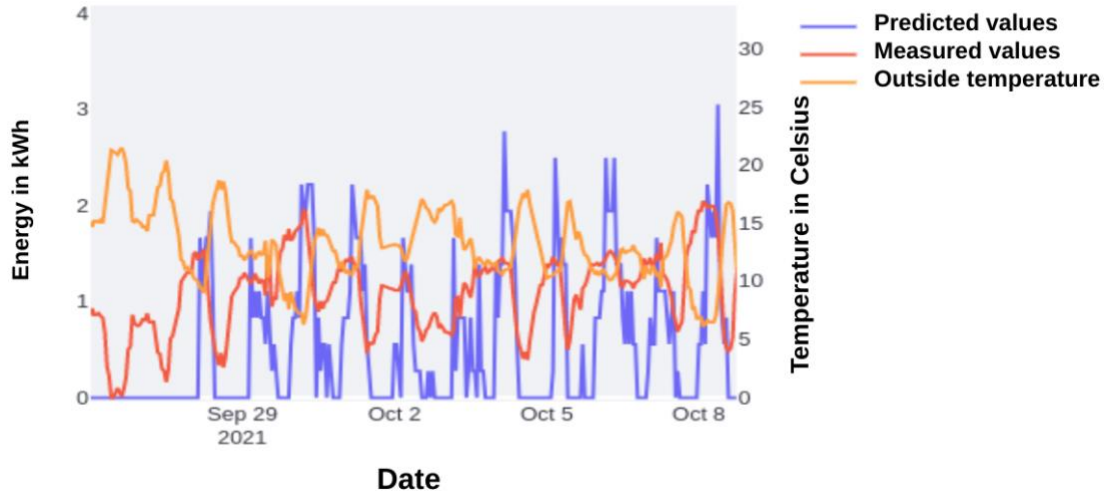


Figure 6.6 shows the heat forecast equation applied to a real dwelling consumption.

We do not report results on the whole dataset as this technique had too many limitations, discussed in the next section, which made it unsuitable for this specific use case.

6.6.3 Limitation and next steps

First of all, this physical modelling and equation based approach requires Internal temperature to work. O-Nexus does not have this measurement, as it is an extra cost to install the sensor in the building. Even for those addresses that measure it, we see that most people are not keeping a constant temperature in their houses, so that makes it hard to guess what the temperature inside someone's house can be.

Second, most of the heat loss comes from ventilation and openings like doors or windows (Open, n.d.), which are not present in our dataset and yet they create much noise in the data. This is the main hypothesis we could come up with for explaining noise. Yet there could be other factors that play a role: for example, tenants might choose to heat up their dwellings only at certain times, or maybe the internal temperature is not constant in the dwellings. Furthermore, even if we knew static information about ventilation and doors/windows, this still does not provide any clue about tenant behavior.

Third, a static equation cannot adapt to tenant behavior or to any dynamic trends at all. This makes it hard to scale this equation to several hundreds of houses. In general, the lack of flexibility is the biggest problem in our use case as the approach would be useful only if we can use it in a generalized fashion, not on a one-by-one analysis basis.

A next step would be to learn the parameters from the data and use physics-based linear regression. In this way, the different patterns for each household can be learnt in an independent and scalable way. The current result is just an intermediate model and a stepping stone towards the model that will be discussed in the final report.

6.6.4 Data driven modelling

Data driven or ML approaches have in the last years shown promising results for predicting heat demand in buildings (Amasyali & El-Gohary, 2018). For O-Nexus the goal is to obtain a prediction which is accurate (about 5% of daily usage), simple (using only heat demand and weather data) and fast to run (avoiding long fitting times). As a consequence, the explainability of the model is not relevant at this point.

The main model used was the Prophet model (Taylor & Letham, 2019). Prophet is a model created by Facebook that can take care of outliers and recognize historical trends in time-series data. It can also incorporate many factors such as holidays automatically.

To have a benchmark, we used few more models:

- A Naïve model, using the last value to predict the next value (last observation carried forward).
- A Naïve model, using the mean value of the whole dataset.

A Holt-Winters model with additive component.

6.6.5 Data driven modelling – results

The Prophet model made reasonable predictions in most cases. In the figure below we can see an example of multiple day prediction. The red line represents the estimated values, the blue line represents the actual ones, the orange line represents the outside temperature.

As we can see the model can incorporate the correlation between outside temperatures and heat consumption. It should be noted that some noise in the data could be due to tenant's behavior, so a perfect model would not be possible.

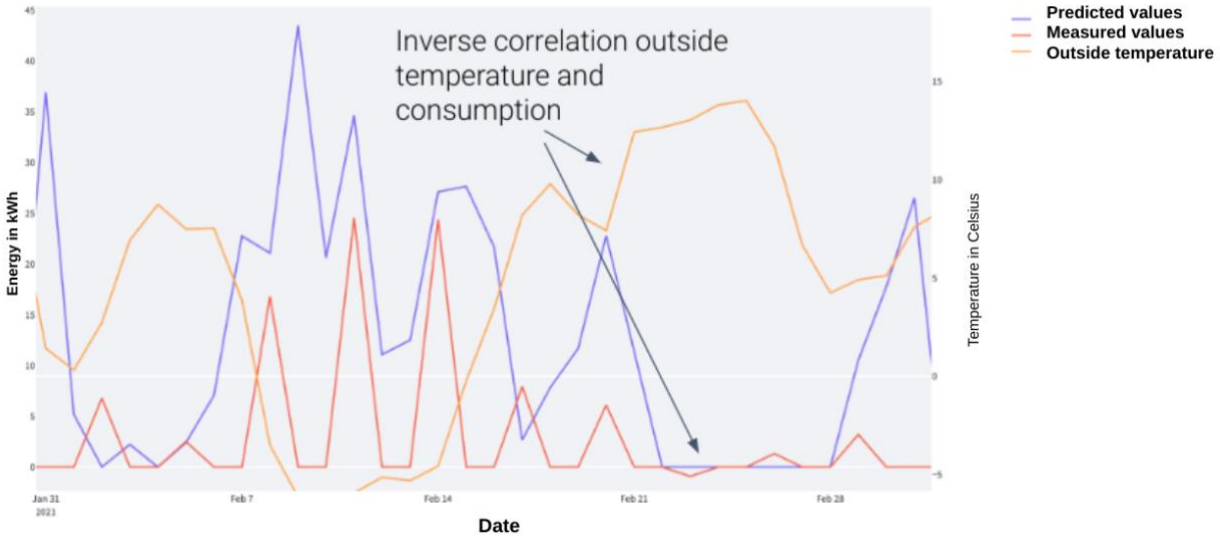


Figure 6.7 Prophet heat prediction vs actual data over several days. Red line: predicted values, blue line: measured values, orange line: outside temperature.

O-Nexus acceptability threshold for a model was that if the error on the prediction was 20% or less of the actual amount, this was acceptable for our use case. The rationale was that more standardized error metrics such as RMSE, MAE are not appropriate for our use case. In fact, in the winter an error of 1 kWh when the daily usage is 20 kWh would be fine, while in the summer, when a lot of dwellings have a near 0 kWh consumption a 1 kWh would be not acceptable. We wanted a metric that could be reliable both in the winter and in the summer, although the seasons have different patterns. Consequently, standard error metrics would not be a good choice.

Percentage of predictions with less than 20% of the actual value					
Site*	Prophet	Naive (Last)	Holt Winters	Naive (mean)	Physical equation
Site1	0.4	0.3	0.4	0.2	0.1
Site2	0.4	0.3	0.3	0.2	0.0
Site3	0.3	0.3	0.2	0.2	0.0
Site4	0.3	0.3	0.2	0.2	0.0

Table 6.1 Prediction performance overview over different sites.

**Site names are anonymized*

As we can see, the performance is different per site as each site has different building characteristics and HVAC installations and Prophet model performs always well compared to the benchmarks model (see section 6.6.4 for explanation). Sites with lower accuracy are in general sites with higher variance where it is not clear what is causing the heating demand from the data at hand.

The Prophet model experiences also some overfitting problems, meaning that the most recent temporal data has more weight in the prediction rather than the outside temperature.

Other times, maybe due to seemingly chaotic tenant behavior, the model is not able to fit the data properly and can even predict below 0 consumption as Figure 6.9 shows.

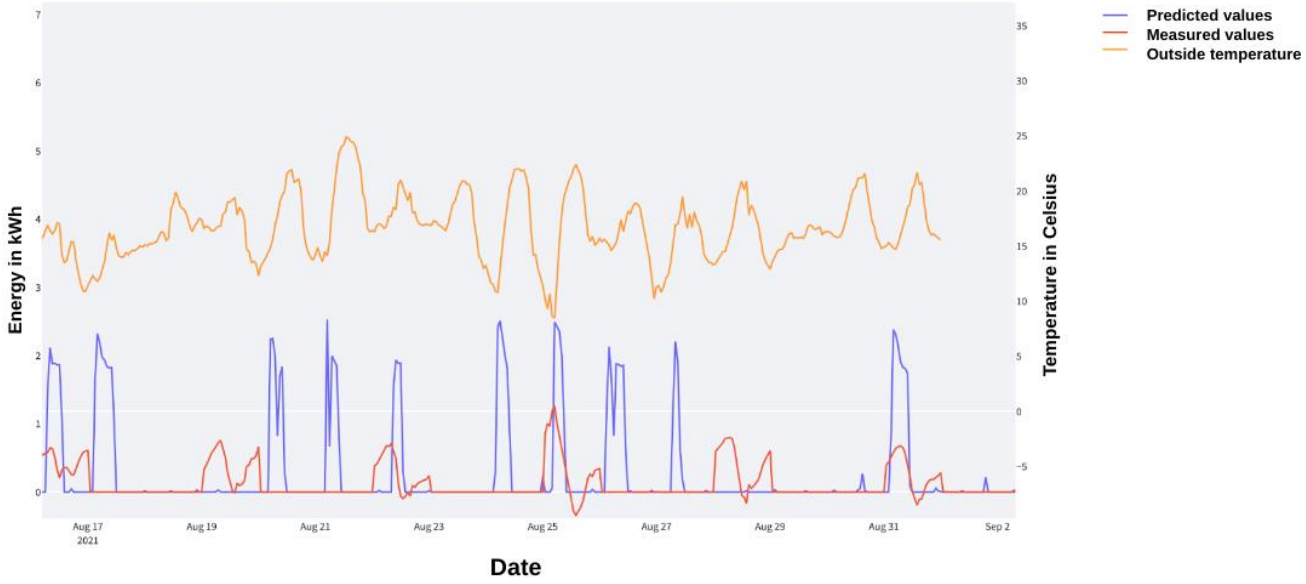


Figure 6.9 Prophet heat prediction vs actual data over several days: overfitted predictions.

Underfitting was also present, as sometimes the model learnt a spiky behavior different than the real consumption (red line = prediction, violet = actuals, orange = temperature) as Figure 6.10.

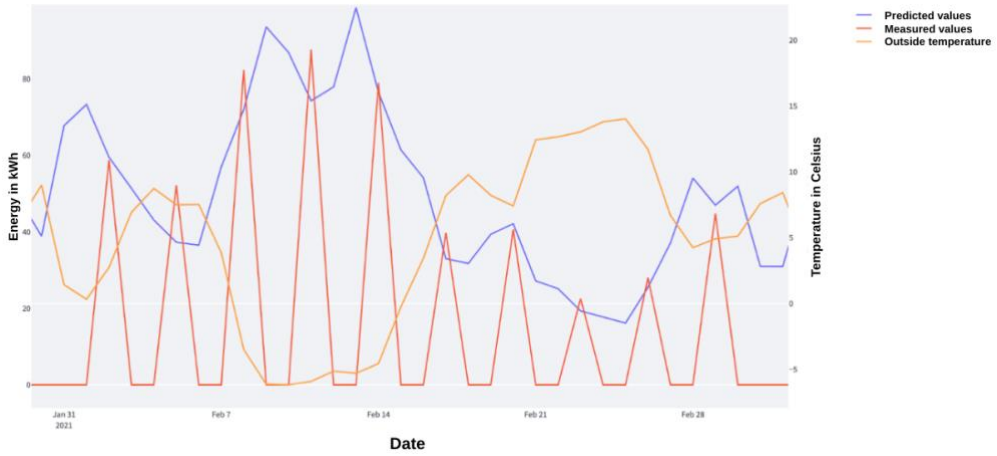


Figure 6.10 Prophet heat prediction vs actual data over several days: underfitted predictions.

In fact, although the Prophet model performs better than the rest, having only a good prediction in 40% or 30% of the days on a year basis, it is not sufficient for a real production usage.

6.6.6 Data driven modelling – limitations and next steps

As reported, the main limitation of this study is that more models could have been used especially tree-based models like random forest (Karasu & Altan 2019), xgboost (Zhang et al., 2023)(Wei et al., 2021), SVM or LSTM (Wei et al., 2021) that have shown very good results in predicting time series data, also in the energy domain. Prophet had a very slow fitting time, around 7 minutes for each unit, resulting in a non-acceptable fitting times for production usage. A much faster fitting procedure or algorithm should be used.

In the next steps, it is likely that the tree-based models will be implemented and benchmarked. The current results are therefore an intermediate result. In the final version of the deliverable the updated models will be discussed.

6.6.7 Conclusion on heating forecast modelling

In this unit, we explored two different ways of doing heating forecast modelling. Through the physical equation-based approach we have learnt how heat loss works for buildings and when its limitation when facing noisy data. In the data driven approach, we obtained slightly better results, yet they are not usable for live production environment due to little accuracy and slow fitting times.

In general, we can say that a data driven approach is more suitable when the goal consists of making accurate predictions, while using a physical approach can be more useful when data is not available and as a general learning experience for the domain.

6.7 Steering implementation – experimental setup

6.7.1 Necessary software

To be able to steer power at the right moment, the Gateway (GW) should be able to send 'write' commands to a physical device, such as a heating medium. For this prototype, the Gateway did not have any algorithm itself, so it needed to listen to incoming commandos from the O-Nexus cloud and forward them to the physical device. In other words, the Gateway makes a read from the shadow meter, sends it to the cloud, and the cloud algorithm decides what to do next, sends it to the Gateway and the GW converts that commando into a voltage for the heat storage. As a consequence, on the cloud side you need also a Database that can retrieve all the messages and an online server, that can continuously communicate with the Gateway.

This approach has the advantage to use the cloud for each decision, but it is very data heavy for a IoT connection. Plus, if the connectivity is lost, the heating storage might be fully on or off when it is not meant to be. Furthermore, the GW could discard an incoming commando if it was already overloaded with other tasks, such as reading out other sensors.

6.7.2 Necessary hardware

On the hardware side, it is necessary to read out the electricity from- and to- the grid real time with a meter as well as having a relay or another way (i.e. Modbus) that can "write" commandos to another device. This choice is very dependent on the chosen storage device. In this case, the Flamco 6E Flextherm has just a resistance so it can easily be steered based on the voltage it receives.

6.7.3 Algorithms at work

We implemented 2 simple algorithms:

1. PV export. In other words, when there is excess PV energy, steer that into the storage device.
2. Epex based algorithms. When electricity is cheap, load up the storage device.

In this stage, we did not want to implement sophisticated algorithms to avoid complexity. We wanted simple algorithms so that we can see how the hardware and software combination was behaving. What we learnt by implementing those is in the next section.

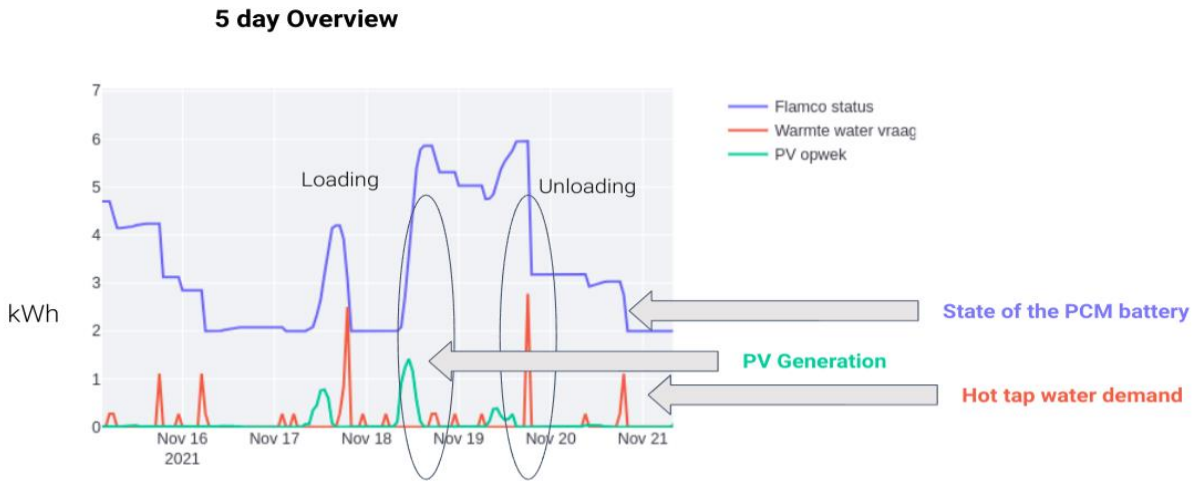


Figure 6.11 Loading a thermal battery with excess PV energy.

In Figure 6.11, we see what happens when steering is activated. The state of the PCM (thermal) battery such as the Flamco Flextherm, loads with the available PV generation (in green) and diminishes with the hot tap water demand (in red). As a consequence, the consumption related to the heat demand is shifted to convenient times instead of being equal to the consumption times as it would happen without steering.

6.7.4 Conclusion on steering setup

Although this setup can seem trivial, there are a lot of takeaways:

- First, having a fully cloud based power steering is neither financially nor technically convenient. Sending messages for reading electricity and sending commandos to react on the reading costs a lot of data traffic. O-Nexus goal is to keep the subscription costs low, and that means a low internet traffic. The learning is therefore to offload some calculations to the Gateway. The cloud-based approach is still valid for quick prototyping.
- Second, the strategy for the steering needs to be logged. During our experiments, the GW could have unexpected behaviors (i.e., loading when it should not). In the initial setup, it was hard to isolate if the error came from the cloud side or the hardware side. When we started logging the strategy, it became clear where the error came from.
- Third, mapping the cloud commando to a hardware action it is not trivial, as the hardware needs to be very transparent on how many watts it can handle to function properly. Each storage device can have several operating modes. Some of them work only on/off, others have more options. Regardless of the brand, the storage device needs to have quite transparent instructions on how to be steered. This is unfortunately not always the case, so it is advisable to pick the storage device first and check whether it is easy to integrate with a different technology.
- Fourth, even a seemingly simple storage device, can have its own logic as they are not programmed to be steered. For example, the Flamco Flextherm might not be loaded if it is almost full. The problem with these behaviors is that they are not documented anywhere, as a consequence, reverse engineering is the only option to explain them.
- Fifth, the state of charge of the storage device is not easy to calculate. Similar to the previous point, most storage devices do not have an automated way to know how charged they are. They do have an "expectation" from the manufacturer, which is based on a lot of assumptions. In theory, the energy going in is close to the energy going out minus some heat and conversion loss. In practice, there are many factors, such as the installation, insulation and material of the pipes, plus the combination with the internal device logic (see previous point) that can affect the state of charge.
- Sixth, the installation quality can determine the storage device performance. In our experiment, we noticed that a bad installation can cause very big amounts of heat losses and lower performance by a big margin. The advice is to check the installation very well, monitor its performance and making sure that its performance and the manufacturer "expectations" align. This should be done before starting with steering.

7 BLACK BOX MODELS: A FEEDFORWARD NEURAL NETWORK MODEL TO PREDICT THE BUILDING ENERGY USAGE AND ITS UNCERTAINTY

Authors: Robbert-Jan Dikken (Peutz)

7.1 Goal of the prediction model

The goal of the model developed in this chapter is to predict energy usage in a building together with a measure of the uncertainty in the prediction. Such model can be used to set up robust strategies to balance energy supply and demand in order to avoid congestion in the energy grid.

7.2 Introduction

The world undergoes an energy transition towards the use of sustainable energy sources to fight climate change induced by human action. Sustainable energy sources, like wind and solar power, are inherently stochastic. This can become a problem in the energy networks when a large part of the energy usage has to come from these sources. To balance energy supply and demand, it is important to be able to predict the expected energy supply and demand. Because of the stochastic nature of the problem, it is also important to have a measure of robustness in the energy prediction. In this chapter the focus is on a data-driven approach for energy usage prediction and uncertainty quantification as a margin on the prediction.

7.2.1 Concepts

Technique

In this chapter we focus on a black box model with the assumption that this enables and supports scalability of energy prediction methods for building energy usage. For energy usage forecasting models like recurrent neural networks and long short-term memory (LSTM) models are used (Maarif et al., 2023). However, to stimulate the scalability of the solution, for both energy usage prediction as the uncertainty prediction, we choose here as the main machine learning concepts applied to be feedforward neural networks and committee machines. Feed forward neural networks are networks that consist of connected (artificial) neurons (Mitchell, 1997). Each neuron has a mathematical function that relates an input substate to an output substate, as shown schematically in Figure 7.1.

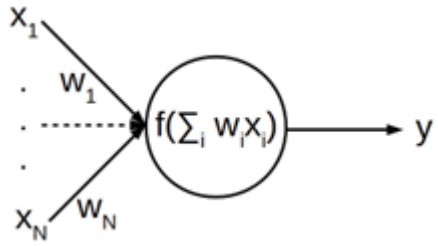


Figure 7.1: Schematic representation of a neuron.

The weights *w* of the connections between the neurons determines the collective action of a neural network to relate the full input state with the full output state.

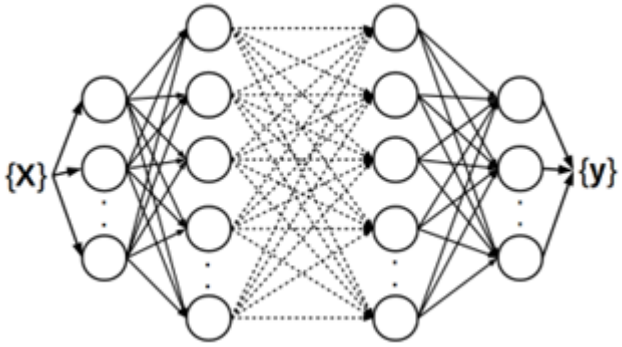


Figure 7.2: Schematic representation of a neural network.

Such a network is trained on known data in a supervised manner using a stochastic gradient descent algorithm. During training the weights of the neuronal connections are optimized to minimize the prediction error. The resulting model is tested on an independent test dataset, to verify that the model does not overfit on the training data, as that would mean a loss of capacity to generalize.

Neural networks are known to perform best when trained with large datasets. Because for problems involving energy usage prediction in buildings datasets are more often small to intermediate in size than large, we work with committee machines. These committee machines make use of ensembles of neural networks, and through the power of ensemble averaging the lack of data is handled (Fort, 2020).

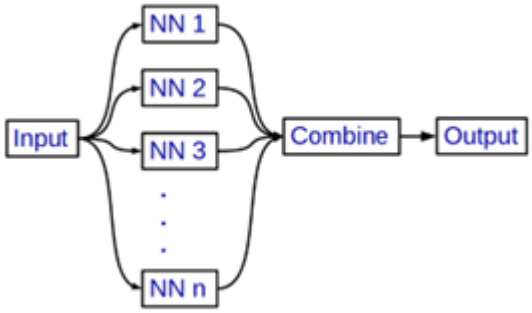


Figure 7.3: Schematic representation of a committee machine.

In this case we are not only interested in the prediction of energy usage itself, but also in the uncertainty that is associated with specific input in the model. Various methods for uncertainty prediction exist (Scher et al., 2018; Rahmati et al., 2019; Harakeh et al., 2021). Building on existing methods a new method is introduced which exists of two parts. The first part models the energy prediction. The second part models the uncertainty where it uses the energy prediction model outcome for training. The absolute error between the energy usage prediction and the ground truth of the energy usage at specified input is used to train the model to predict the absolute deviation which is a measure of the prediction uncertainty. This assumes that the uncertainty is intrinsically represented in the data (Scher et al., 2018). Important to note is that the absolute error is predicted, since it is impossible to predict the real error. The training of both the predictor model and the uncertainty model is performed on the same training dataset, before being evaluated on an independent test set.

For both the predictor model and the uncertainty model a committee machine is used. The full model is schematically represented in Figure 7.4.

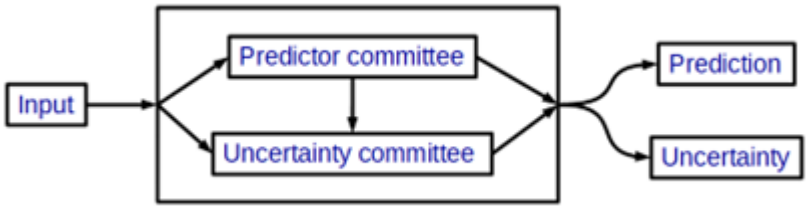


Figure 7.4: Schematic representation of the system to obtain a prediction and the uncertainty in the prediction.

Contextual data-driven modelling

As mentioned, a black box model enables and supports a scalable solution for predicting the energy usage in a building. In general, black box models are assumed to be unfathomable due to the complex structure that emerges from training the model on data. However, when the desired output of the model is considered, it is possible to define input parameters that are required, purely based on expert knowledge of the problem at hand. Through selection of relevant physical parameters for specific application, energy usage prediction in this case, it is possible to get a model that is more likely to give causal relations instead of correlations without physical meaning. When it comes to the application of HVAC energy usage in buildings, parameters of which it is known that they are relevant include time of day, day of the week, month, expected outside temperature, expected solar radiation, expected humidity, inside air quality (CO_2), inside temperature. The comfort requirements are taken as input to the model, such that the model can differentiate between adaptable levels of comfort versus external influences in energy usage.

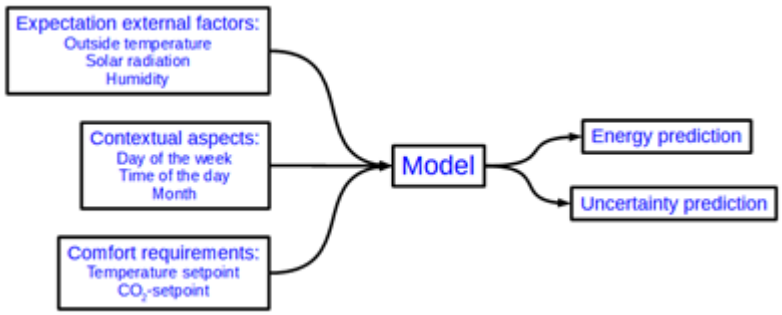


Figure 7.5: Schematic representation of parameters that are of importance for energy usage prediction

The method in this chapter for prediction of energy usage and associated uncertainty is not limited to the parameters mentioned in the schematic representation. When the predictions for a specific situation ask for additional information, this can simply be implemented by adapting the model to it, illustrating the scalability of the method.

To incorporate possible fundamental changes in the contextual input such as change in building use, a rolling window prediction technique is applied. This entails that a model is trained to perform a forecast, and this process is iterated repeatedly. In this way the forecast is continuously finetuned and changes are inherently implemented.

7.3 Testing on two cases

To test the conceptual approach given in the previous section to determine a prediction in energy usage and a measure of uncertainty, we use datasets of two different buildings. In the rolling window prediction technique the split in the dataset between training data and test data is naturally time-based. In this way the evaluation of the model is performed independently from the training, to verify the generalizability of the method. The first case to test the method is of an educational facility (Metsä-Eerola et al., 2022) and the second case is of a commercial facility (Pîrjan et al., 2017). The first case is concerned with the prediction of heating energy, while the second case is concerned with the prediction of electricity consumption. These examples will show that the approach for predicting energy usages and uncertainty is rather generic, both in type of use case and in the parameters involved.

7.3.1 Case 1

Case 1 is concerned with the prediction of heating energy of an educational facility. The available parameters in the dataset include weather conditions such as temperature, solar radiation and humidity, the required inside temperature and CO_2-level, and the time and date from which day of the week can be derived. To test the prediction the (normalized) weather conditions that are used for the prediction are adjusted randomly using a gaussian distribution with zero mean and a standard deviation of 0.1. For case 1 in Figure 7.6, the absolute deviation is given versus the predicted absolute deviation, giving a measure of the uncertainty in the prediction. The dashed line has a slope of 1. A regression analysis gives for the relation between predicted absolute deviation and the absolute deviation a slope of 0.89, with a coefficient of determination of 0.48. Although the coefficient of determination is not impressively high, the fact that the slope of the relation approaches 1 is a good enough indication that the method to predict absolute deviation works.

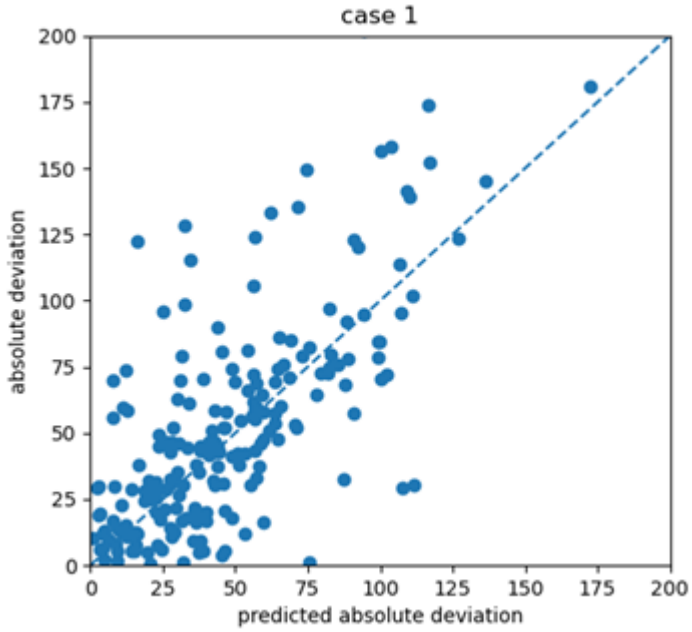


Figure 7.6: The predicted absolute deviation as a measure of the uncertainty in the energy prediction.

In Figure 7.7 the prediction of energy usage for heating together with the predicted uncertainty is shown for a situation where the rolling window forecast is done for four consecutive days. The prediction and uncertainty margin coincide rather nicely with the ground truth.

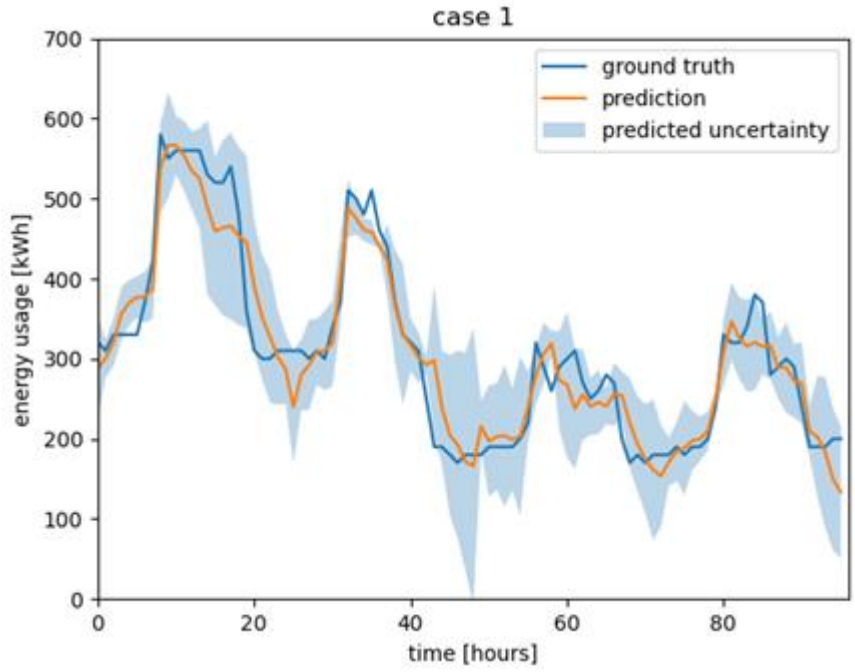


Figure 7.7: The prediction and uncertainty margin for four consecutive days for case 1, showing good agreement with the ground truth.

7.3.2 Case 2

Case 2 is concerned with the prediction of electricity usage in a large commercial facility. The available parameters in the dataset include weather conditions and time and date and day of the week, providing the context. Like for case 1, to test the prediction the (normalized) weather conditions that are used for the prediction are adjusted randomly using a gaussian distribution with zero mean and a standard deviation of 0.1. For case 2 in Figure 7.8, the absolute deviation is given versus the predicted absolute deviation, giving a measure of the uncertainty in the prediction. The dashed line has a slope of 1. A regression analysis gives for the relation between predicted absolute deviation and the absolute deviation a slope of 1.0, with a coefficient

of determination of 0.55. The method is clearly able to provide information about the uncertainty in the prediction, although it is still a rough measure.

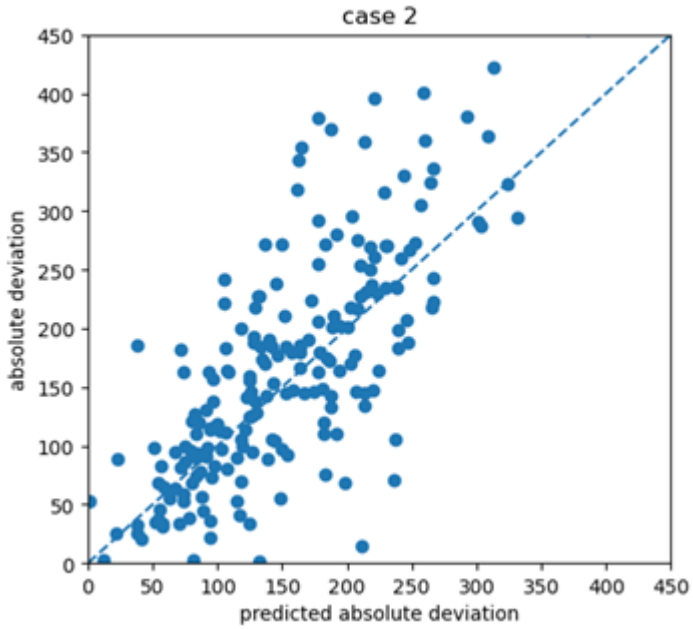


Figure 7.8: The predicted absolute deviation versus the absolute deviation of the prediction.

In Figure 7.9 the predicted electricity usage together with the predicted uncertainty is shown for a situation for case 2 where the rolling window forecast is done for four consecutive days. The prediction and uncertainty margin coincide rather nicely with the ground truth.

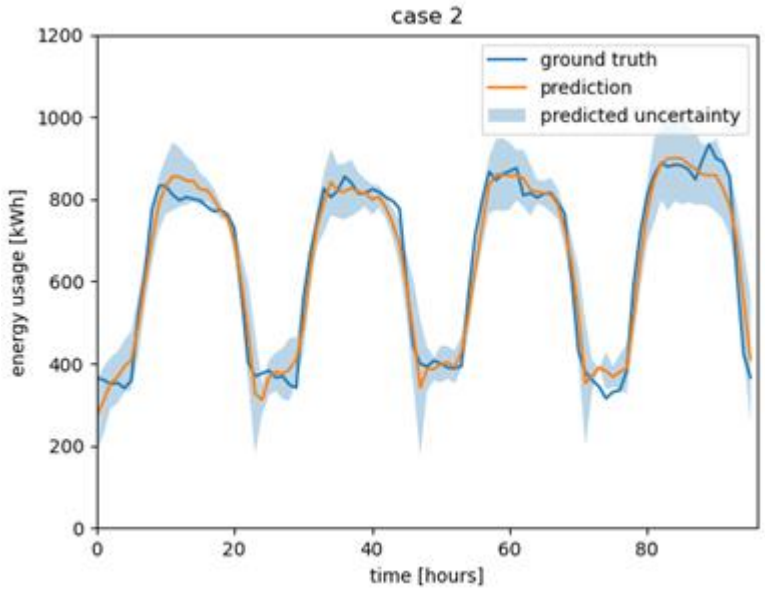


Figure 7.9: The prediction and uncertainty margin for four consecutive days for case 2, showing good agreement with the ground truth.

7.4 Concluding remarks

In this chapter a method is discussed to predict energy usage and associated uncertainty. The method is a purely data driven black box approach. This enables a wide range of application possibilities and supports scalability. The method consists of a stacked sequence of committee machines of neural networks that all have their own use. The power of the method is in the ensemble prediction and the realization that uncertainty is intrinsically present in the data as an absolute deviation. Using a rolling window technique, the method can predict both energy usage and uncertainty, incorporating possible changes in building use. The two test cases differ enough from each other to make the careful observation that the method indeed has a wide range of applications. Further tests for different applications should validate this observation.

8 BLACK BOX MODELS: PRIORITIZING MODELING AND AN EXAMPLE: PV SOLAR

Author: Kees Wisse (DWA)

8.1 Goal of the prediction models

The goal of the work is to develop black box models which can be used to optimize the demand and supply to the grid in the context of net congestion. To rank priorities for modeling, the demand and supply profile to the grid will be analyzed. In the current Dutch context of net congestion, energy supply to the grid is an important bottleneck. PV solar systems play an important role in energy supply to the grid. Therefore, as a first step, the goal of the second part of this chapter is to develop a black box model which is accurate enough to predict the peak energy production of PV solar systems. This is part of ongoing work, following the ranking of priorities as given in this chapter, other black box models will be developed in future.

8.2 Introduction

In this section the DWA office in Gouda (The Netherlands) will be used to discuss the following research questions:

- What are the main contributors to the peak demand from the grid and peak supply to the grid of this case study building? Can we obtain a ranking which models we should develop first and with a high accuracy?
- Based on the main contributors to the peak supply of the office building, the PV production was selected for black box modelling as a first step. The following research question was defined: using black box modeling strategies as given in recent literature, can we meet the requirements for a calibrated model according to common benchmark values for error metrics?

The DWA Office comprises the 3rd and 4th floor of a renovated building in Gouda (The Netherlands). The entire building is equipped with heat pump, aquifer thermal energy storage, PV solar and EV chargers (Figure 8.1). Inside the building, heating and cooling is delivered by so-called climate ceilings (Figure 8.2).

Figure 8.1 Office building DWA.

Figure 8.2 Room supply heating, cooling, and ventilation DWA Office.

8.3 Ranking priorities

8.3.1 An inventory of required models

Understanding the demand profiles is important before starting the development of prediction models.

To prioritize the modelling efforts, it is important to know what the main contributors are to the peak demand for electricity and peak supply to the grid. Based on expert knowledge and given the systems as given in Figure 8.1 and Figure 8.2, the inventory list for the DWA building contains the following 11 sub-models:

- EV chargers
- PV solar panels
- Lighting
- Office equipment
- Air handling units (electricity)
- Efficiency of the power plant, heating part (heat pumps + buffer)
- Efficiency of the power plant, cooling part (heat pumps + aquifer + buffer)
- Air handling units (thermal)
    1. Heating demand
    2. Cooling demand
- Room supply systems
    1. Heating demand
    2. Cooling demand

Please note this list is specific for this type of HVAC installation. For all-air systems for example, a simpler approach may possibly suffice.

8.3.2 Energy demand profiles

Starting from the public power grid electricity meter it is possible to determine which systems contribute the most to the peak demand and which situation is the most urgent driving force: heating or cooling and which period (summer or winter).

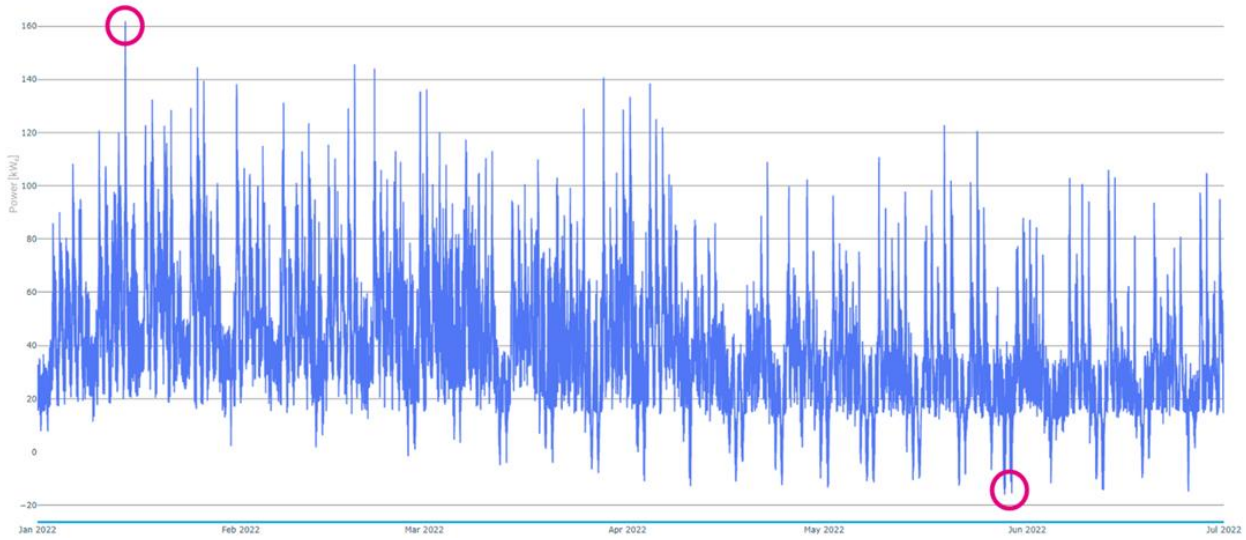


Figure 8.3 Electricity demand/supply public grid. Time resolution: 3'. EV-charging and PV included.

Figure 8.3 shows a half year of data from the main meter of the public power grid of the building with a time resolution of 3'. The data is available from the electricity meter in the Building Management System (BMS).

The peak demand from the grid during the first half year of 2022 occurs on 14 January 2022. The peak delivery to the grid takes place during the weekend, due to PV electricity production which is larger than the base load of the building during the weekend, for example as marked at the end of May in Figure 8.3. Due to the presence of the aquifer, the peak demand of the cooling system has limited impact on the peak demand during May - June. This 'peak demand day' is investigated in further detail.

To assess the key contributors to the peak demand, the submetering of the building was investigated in more detail.

The submetering is available on a time resolution of 15'. Figure 8.4 shows the sum of the available submeters, together with the results of the main meter of the power grid for one specific day (January 14 2022). The difference is mainly due to the (unmetered) charging stations of the electrical vehicles and some other minor unmetered processes.

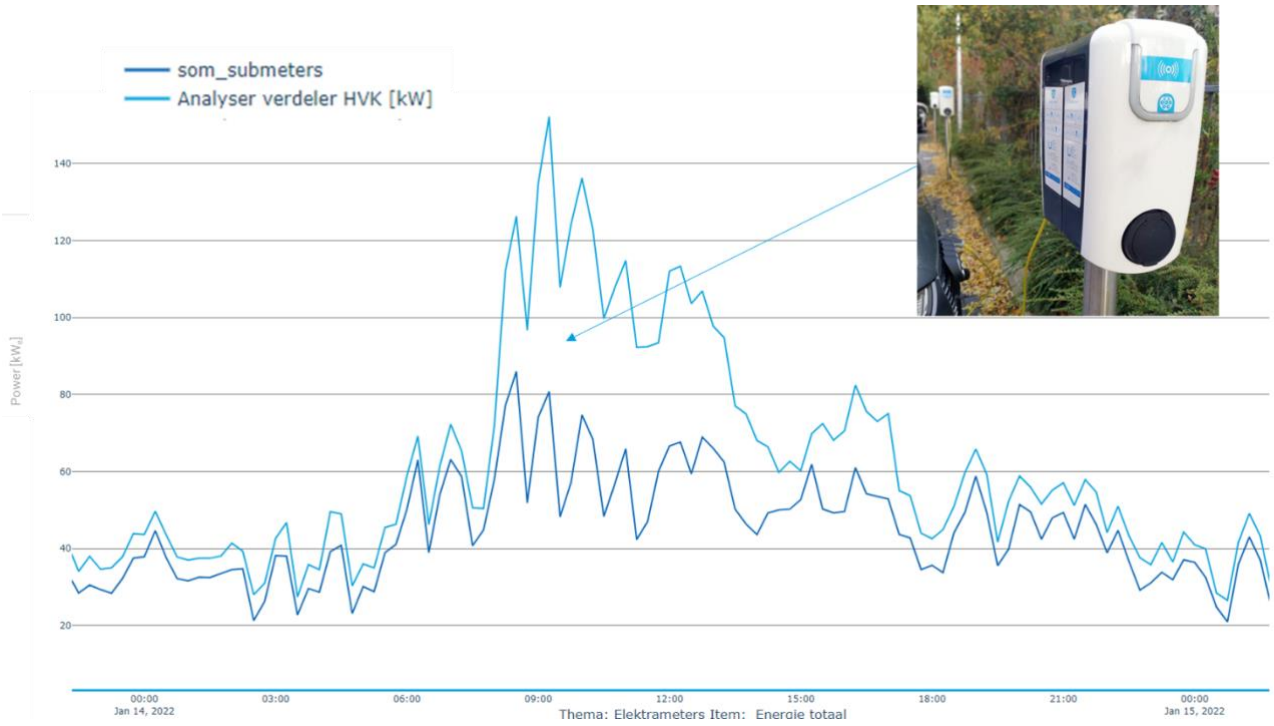


Figure 8.4 Sum submetering versus the main electricity meter on January 14. Time resolution: 15'.

Due to confidentiality not all details can be shared. It is a multi-tenant building, where DWA is one of the tenants (3rd and 4th floor). Therefore, the results of the submetering have been analyzed in further detail for the 3rd and 4th floor. Please note that the central heat pump energy and aquifer energy are scaled based on the floor area. Figure 8.5 provides the results for the first half of 2022.

From the submetering as given in Figure 8.4 and Figure 8.5, it can be concluded that the heat pumps are the main contributors to peak demand of the building, together with the EV charging system. During the summer season the peak delivery to the grid is mainly determined by the sum of the PV production and electricity demand for office equipment, lighting, the air handling units, and heat pumps (Figure 8.6), additional to the electricity demand of EV charging. The net delivery to the grid is occurring during the weekends when the air handling units are switched off and EV charging does not take place. During weekends the office equipment and the heat pump are the main contributors to the electricity demand of the building.

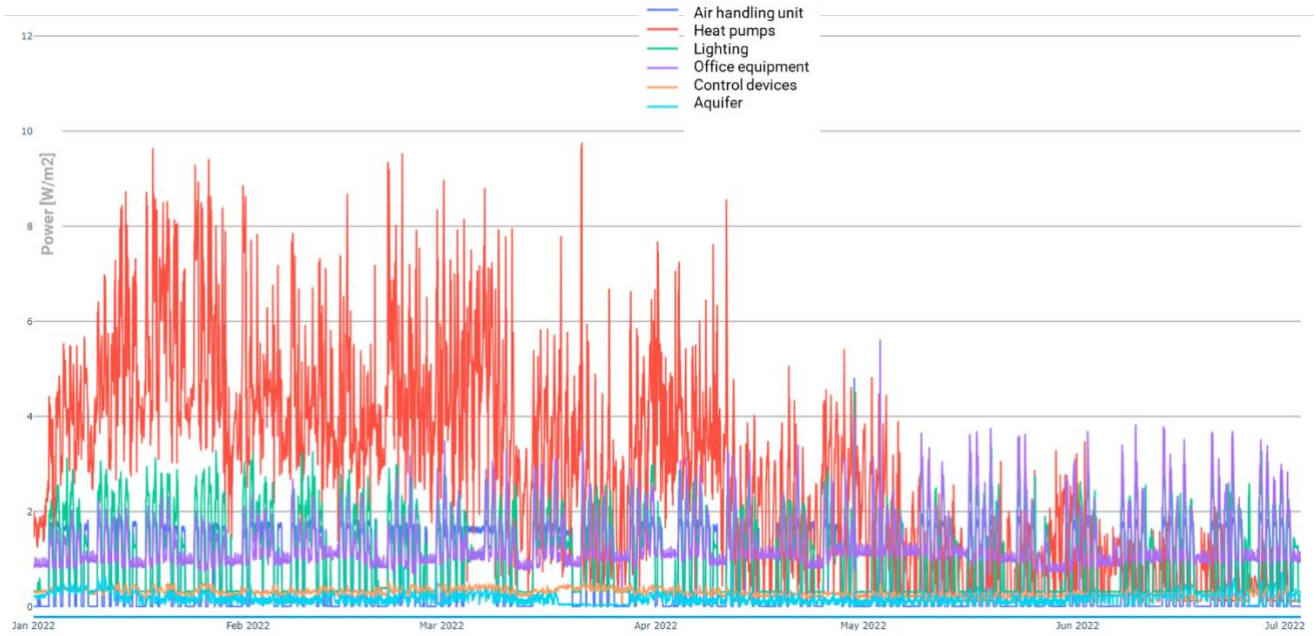


Figure 8.5 Results submetering Jan-June 2022 (3rd and 4th floor, excl. EV charging, excl. PV solar). Time resolution: 60'.

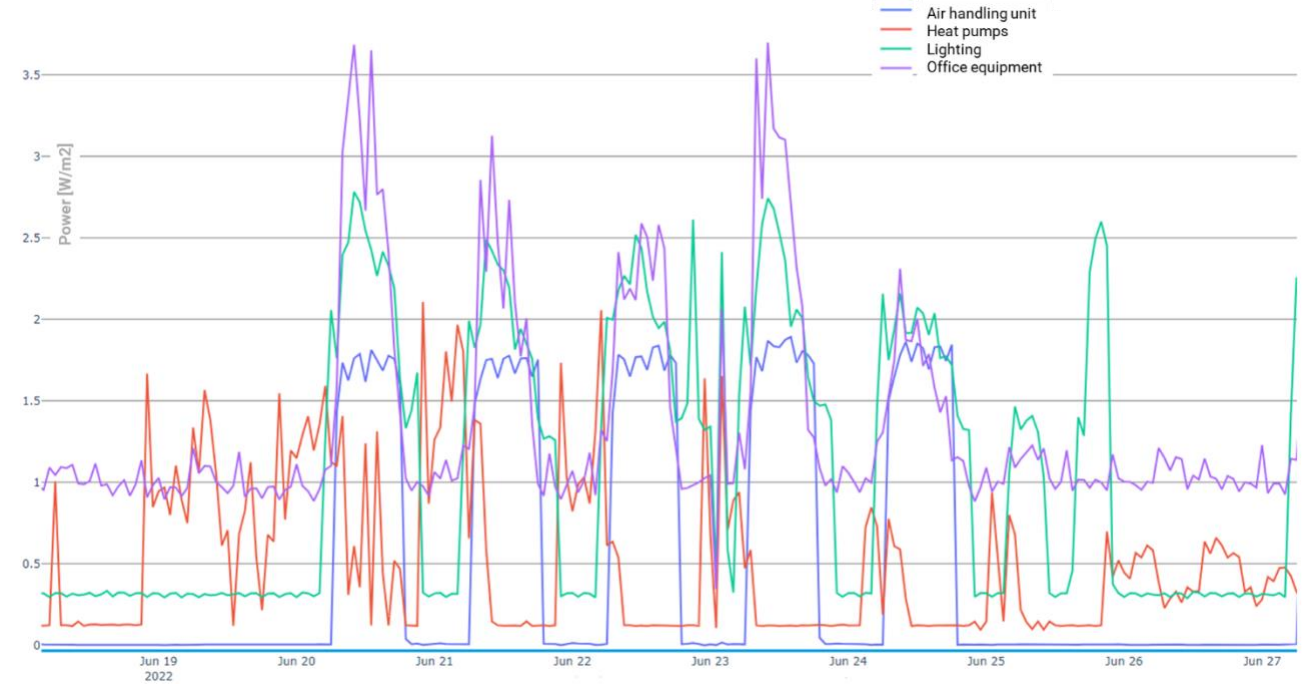


Figure 8.6 Results submetering 19-26 June 2022 (3rd and 4th floor, excl. EV charging, excl. PV solar). Time res.: 60'.

8.3.3 Lessons learned from demand profiles in the case study

- EV charging stations and the heat pumps are the key contributors to the peak demand from the grid. Due to the presence of heat pumps and an aquifer in the DWA office building, the heating situation is dominating the electricity demand profile from the grid. Due to the presence of an aquifer, the cooling situation during summer can be neglected in this case. Please note this can be different for other specific cases like:
    1. chillers are required due to the lack of possibilities for aquifer cooling
    2. buildings which utilize district heating and/or district cooling networks.
- The PV solar system is the key contributor to the peak delivery to the grid. The net delivery to the grid is occurring during the weekends. During weekends the office equipment and the heat pump are the main contributors to the electricity demand of the building.

Summarizing the ranking, the following sub-models with a high priority can be identified for this case study:

- Energy demand from the grid
    - EV charging
    - Heating demand to be delivered by the heat pump:
        - Room supply systems (thermal part, heating demand)
        - Air handling units (thermal part, heating demand)
    - Efficiency of the power plant, heating part (heat pumps + buffer)
- Net energy supply to the grid
    - PV solar
    - Office equipment

For lower priority processes, it is worthwhile to investigate the feasibility of aggregating them into one black box model (category 'other'). Following the ranking of priorities, other black box models will be developed in future.
The next section presents the results of a black box model for the PV solar electricity production.

8.4 Prediction model (PV solar)

8.4.1 Approach and results

Considerable research has been done in the field of PV solar load forecasting using black box models (see for example Kahn et al. (2022B), Markovics and Mayer (2022) and their references).

For this case study we have selected two ensemble methods (Random Forest and Extra Trees) and one boosting method (Gradient Boosting) and one deep learning algorithm (Artificial Neural Network (ANN)).

The features include KNMI weather data (Global horizontal irradiation, outdoor air temperature and dew point temperature, wind speed) as well as calculated solar incidence angles and calculated percentage of diffuse irradiation using the Perez model (Perez et al. (1987)). Calculated solar positions have been reported to increase the accuracy of the prediction models (Markovics and Mayer (2022)), and provide the opportunity to create more generic models, instead of creating one specific black box model for each PV solar system.

Please note that only actual weather data were used in this study, no predicted weather data. This enables to assess the performance of the algorithms first, before adding the uncertainty of the weather data itself. Furthermore, only actual values were used, and no lagged variables have been applied (features at time t, no features at time t-1, t-2 etc., prediction at time t).

To prevent the influence of the large number of zero-production hours during night, only those hours were considered when the global horizontal irradiation was larger than zero (see also Markovics and Mayer (2022)).

In future research the approach can be extended to the application of lagged variables and multi-step forecasting. The benefit of using only actual values is that it makes the model prediction less sensitive to faults occurring in the data of the past hours or days. In this way the current model can also be used for fault detection (co-benefit for workpackage 1).

The occurrence of faults is not a hypothetical situation for this dataset. From expert observations it appeared that up to half July of 2022 inverter faults were regularly occurring.

Therefore, the size of the fault-free dataset was limited. Table 8.1 provides the training and test set for the fault free dataset (training and test set 1). For assessing the potential of fault detection the full fault free dataset was used for training together with a test set with faults (dataset 2 in Table 8.1).

Table 8.1 Test and training sets.

	Training set		Test set		Remarks
Number	Start date (y-mm-dd)	End date (y-mm-dd)	Start date (y-mm-dd)	End date (y-mm-dd)	
1	2022-07-12	2023-01-31	2023-02-01	2023-05-08	15' resolution, without faults
2	2022-07-12	2023-05-08	2022-05-01	2021-07-11	15' resolution Training set without faults Test set with faults

Table 8.2 shows the performance metrics on test set number 1. For the evaluation metrics we used the R2, the normalized Root Mean Square Error (N_RMSE), the normalized Mean Absolute Error (N_MAE) and the normalized Mean Bias Error. These are common metrics to evaluate the performance of PV forecasting models (Kahn et al. (2022B), Markovics and Mayer (2022)).

For the normalization of the errors, we followed the approach of Markovics and Mayer (2022). All metrics were divided by the mean value of the actual PV production over the entire test set. A drawback of this approach is that it can be sensitive to seasonal effects.

The results of the performance metrics of Table 8.2 agree in range with Markovics and Mayer (2022). They found N_RMSE values in the range of 45 – 50%, N_MAE values in the range of 30-37%, for the N_MBE values in the range of -2% to 6%. The N_RMSE and N_MAE perform better with lower values in current approach. Please note however that the approach of Markovics and Mayer (2022) is different in some ways: their training period comprises a whole year instead of part of the year, they apply also lagged variables and use numerical weather predictions instead of actual weather data.

Following the criteria of FEMP and ASHRAE, a N_MBE should fit a value within +/-10% for hourly values and the N_RMSE should fit a value lower than 30% (Ruis and Bandera (2017)). For 15' data no criteria are given. Using a trained algorithm on 15' data resolution and evaluating it on hourly data, we found a N_RMSE of 0.27 and a N_MBE of -0.02 for the model with the lowest N_RMSE: ANN. Given that values, the ANN model meets the required criteria of a 'calibrated model' according to FEMP and ASHRAE. Following the stricter criterium of IPMVP the model should be improved further as they require a maximum N_RMSE of 20% (Ruis and Bandera (2017)).

For further information Figure 8.7 provides a comparison of part of the time series for the actual and predicted values, for visual clarity on hourly basis. At some days the prediction model underestimates the maximum peak production of the PV solar system.

Taking a longer period of training it can be expected that the model performance will improve the peak prediction further. Solar incidence angels and shadow effects on the PV solar system will change during the year and in the presented training set 1 only part of the year is present. Testing on months which are not present in the training set also challenges the extrapolation qualities of the algorithms. A training set of at least one year of data is therefore recommended in this approach without lagged variables.

Table 8.2 Performance metrics on test set 1: R2, normalized Root Mean Square Error and normalized Mean Absolute Error and normalized Mean Bias Error. Based on the time resolution of 15'.

Regressor	R2	N_RMSE	N_MAE	N_MBE
Random Forest	0.81	0.42	0.26	0.05
Extra trees	0.86	0.36	0.23	0.04
Gradient Boosting	0.86	0.36	0.23	0.04
ANN	0.88	0.35	0.24	-0.02

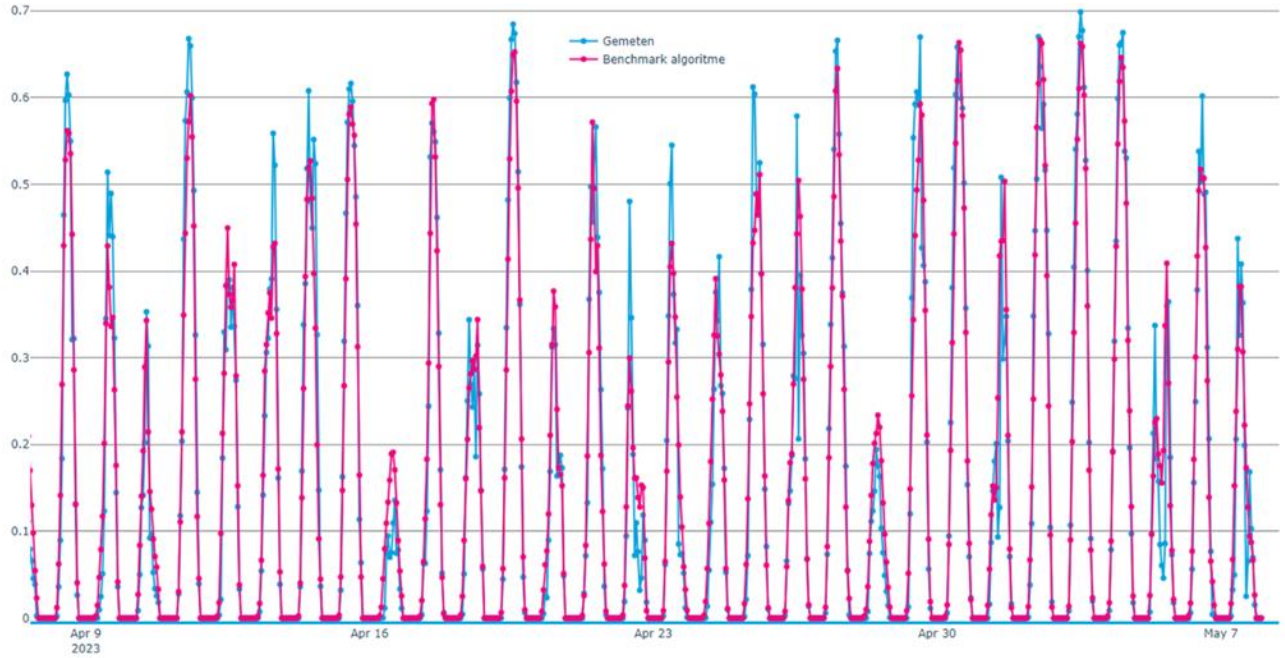


Figure 8.7 Comparison actual (gemeten) and predicted values (benchmark algoritme) of the ANN model (Time resolution: hourly). Y-axes: normalized PV power by the installed peak power of the PV system.

Given the certain limitations in predicting the maximum peak power, the model however can detect the failures in the second test set. Examples of the failing inverter are given in Figure 8.8 (blue arrows). Due to the failing inverter only part of the PV system provides power. With a certain threshold on the deviation between the actual and predicted value, the algorithm can be used to detect these faults. The thresholds can be based on error metrics of fault free training sets. In future research the threshold values can be specified in further detail. For a similar machine learning approach for fault detection for air handling units, see also Workpackage 1 (Chitkara et al. (2022)).

Apart from the interesting results related to fault detection, Figure 8.8 also shows the challenges for load forecasting in buildings with faulty operation. Models without faults, even black box models or white box models don't provide accurate predictions during faulty operation of the building. Buildings without any faulty operation seem to belong to an ideal world. Including faulty operation in white box models seems to be quite challenging. Therefore, it's valuable to evaluate in future research the performance of black box and hybrid models which are trained or calibrated on faulty datasets.

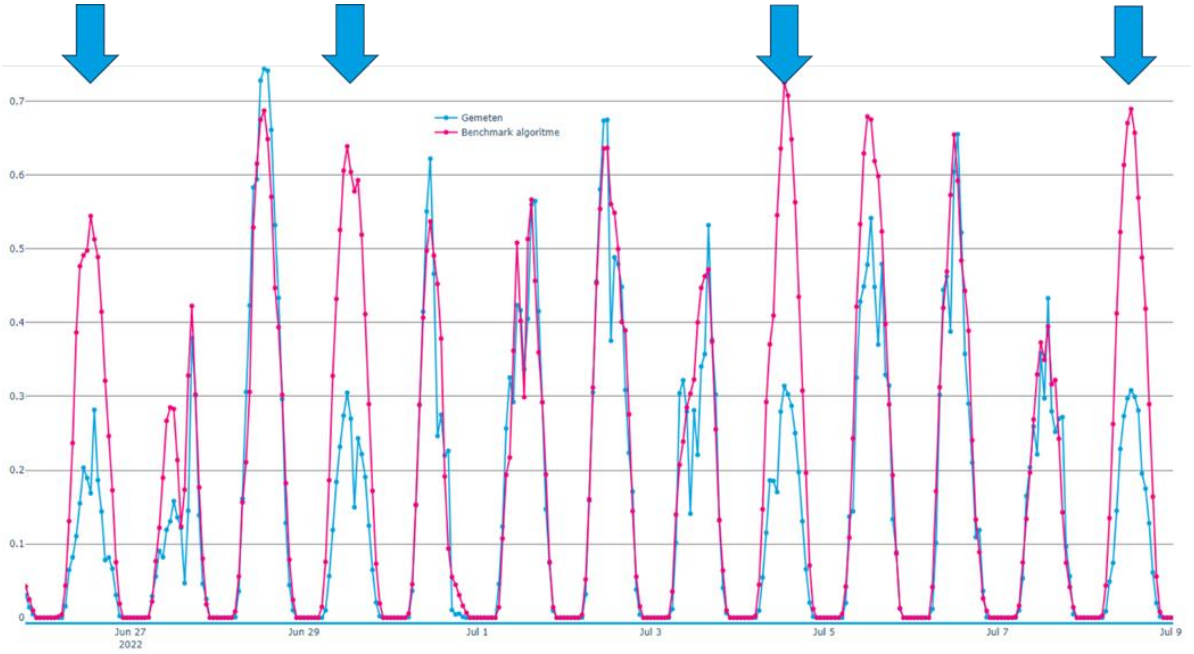


Figure 8.8 Examples of a failing inverter (blue arrows); ANN based on training set 2. Data from test set 2.

8.4.2 Lessons learned from PV prediction model

- The developed prediction model has been successfully demonstrated as model for fault detection.
- On a fault free test set, the model meets the requirements for a calibrated model according to the FEMP and ASHRAE criteria for the error metrics. According to the IPMVP criteria the model should be improved further. The results of the performance metrics agree in range with values as found in literature.
- For accurate peak prediction a year of training data is recommended in the given approach without lagged variables.
- Within the context of load forecasting the model can be developed further using lagged variables and uncertainty in the weather data.
- Faulty operation provides extra challenges to load forecasting models. This is both the case for black box models which are trained on fault free data and even for white box models if they don't model the faults explicitly. It's valuable to evaluate in future research the performance of black box and hybrid models which are trained or calibrated on faulty datasets.

9 CONCLUSIONS

In the built environment, there exists an imbalance between the sustainable energy production from solar and wind sources and the energy demand of building systems. Additionally, both the expansion of energy demand and supply are constrained by network congestion. To alleviate congestion and enhance energy balancing, it is crucial to shift energy demand, a concept referred to as "energy flexibility." Achieving energy flexibility requires forecasts for both energy demand and supply. As consortium partners simultaneously explored various modeling prediction approaches, we will gather and summarize the findings for different model types: white box models, grey box models, , and black box models. Prior agreements were made among the partners to assign responsibility for specific models. This approach ensures that the B4B project covers a wide range of model types and facilitates the sharing of their respective strengths and weaknesses. We conducted thorough testing of all models and methodologies using real-world data, including living labs and practical use cases. As the project approaches its conclusion, we will provide an updated version of this report.

9.1 White box modelling approaches based on Modelica

The objective of the HHS case study is to create a computer model using open-source and freely available software for simulating the energy consumption of a building and the behavior of its energy system components. As a result, the decision was made to employ Modelica with OpenModelica software to develop the computer model for this case study, as it is the only option on the list that aligns with the defined requirements. Additionally, Modelica boasts a comprehensive library of building and HVAC component models.

At the time of publishing this report, the building model remains a work in progress. The report will be updated at a later stage to include the completed model. Several factors contributed to the delay in model construction. The following conclusions can be drawn:

- OpenModelica presents a steep learning curve, primarily due to the lack of comprehensive documentation and tutorials.
- Complex buildings, such as the one in this case study, result in thousands of equations in the Modelica model, demanding substantial computational power. Therefore, it becomes necessary to determine which components should be detailed in the model and which can be represented by other software (such as EnergyPlus) or as black-box models.
- None of the IBPSA libraries contain all the required model blocks for simulating a complex utility building. However, there is an active community of developers continuously creating new libraries, and the model blocks from these libraries can be integrated into models developed with the IBPSA libraries. Additionally, libraries beyond IBPSA are available. For instance, the AixLib library includes an Air Handling Unit model, and the Technical University of Darmstadt has developed a borehole thermal model for open systems (Formhals et al., 2020). The feasibility of integrating these libraries into the case study model should be assessed.
- The quality of the case study data available was lower than expected, which also contributed to the delay in validating the parts of the model already constructed.

The objective of O-Nexus is to develop a functional prototype for energy management in SME buildings. In pursuit of this goal, O-Nexus explored two distinct methods for heating forecast modeling. In the physical equation-based approach, they gained insights into how heat loss operates in buildings and recognized its limitations in the presence of noisy data. On the other hand, in the data-driven approach, O-NEXUS achieved slightly superior results, albeit unsuitable for real-time production environments due to limited accuracy and slow fitting times. The key takeaways are the following:

- A fully cloud-based power management system is neither cost-effective nor technically practical. Transmitting messages for electricity readings and sending commands to respond to those readings consumes a significant amount of data traffic. O-Nexus aims to keep subscription costs low, which necessitates minimizing internet traffic. Therefore, the learning here is to offload certain calculations to the Gateway, while the cloud-based approach remains valuable for rapid prototyping.
- It's crucial to log the power management strategy. During our experiments, the Gateway exhibited unexpected behaviors, such as activating when it shouldn't. In the initial setup, it was challenging to determine whether the error originated from the cloud or the hardware side. However, once we began logging the strategy, it became evident where the error was coming from.
- Mapping cloud commands to hardware actions is not straightforward, as the hardware must transparently indicate its capacity in watts to function correctly. Each storage device can have multiple operating modes, some offering only on/off functionality, while others provide more options. Regardless of the brand, the storage device should offer clear instructions for integration with different technologies, although this is

not always the case. Therefore, it's advisable to select the storage device first and confirm its compatibility with the chosen technology.

- Even seemingly simple storage devices may have their unique logic, as they are not originally designed to be controlled externally. For example, the Flamco Flextherm may refuse to charge if it is nearly full. The challenge with such behaviors is that they are often undocumented, necessitating reverse engineering to understand them.
- Determining the state of charge of a storage device is not straightforward. Most storage devices lack automated methods to gauge their charge level and instead rely on manufacturer 'expectations' based on various assumptions. In theory, the energy input should closely match the energy output minus some heat and conversion losses. However, in practice, multiple factors, including installation quality, insulation, pipe material, and the device's internal logic (as mentioned earlier), can influence the state of charge.
- The quality of the installation can significantly impact the performance of the storage device. During our experiments, we observed that poor installation can result in substantial heat losses and significantly lower performance. It is advisable to thoroughly inspect the installation, monitor its performance, and ensure alignment between its performance and the manufacturer's 'expectations' before commencing with power management

9.2 White box model approach: Hybrid building model

TNO has developed a hybrid modeling approach to forecast energy demand and supply in buildings, enabling energy flexibility solutions to address balancing and congestion challenges. This hybrid predictive digital twin model combines a white-box model for analyzing heat gains and losses in the building with a black-box occupant model that simulates interactions between occupants and building components. TNO has developed the white-box building model, equipment models, and black-box occupant models, which are all integrated into a simulation and prediction model.

The constructed hybrid model by TNO incorporates both a physical model for the building and equipment, as well as a black-box model for occupant behavior. This model is utilized to simulate and predict dynamic building behavior. It has been applied to the use case of the TNO building at Stieltjesweg in Delft, focusing on cooling demand during a summer period.

Real building behavior data have been collected from various sources and compared to the hybrid model. Initial analysis indicates that the building behavior aligns with the simulation results.. Ongoing development efforts aim to enhance model predictions by considering solar shading from surrounding buildings and incorporating models of automatic shading devices. Updated results will be presented in the next report update.

Currently, there is no data available on the electricity consumption of individual HVAC components in the use case TNO Stieltjesweg (such as the chiller). However, efforts are underway to install submetering at the Stieltjesweg building. Access to measured electricity data will facilitate the calibration of the hybrid model, leading to improved predictions of actual energy demand.

9.3 Grey box modeling approaches

At Windesheim, inverse grey box models were employed to deduce recent and ongoing ventilation flow rates over time using monitoring data on CO_2 concentration and occupancy. Additionally, recent and current occupancy over time was derived based on monitoring data on CO_2 concentration and ventilation flow rate. Occupancy is one of the most crucial parameters of user behavior. When historical occupancy patterns can be extracted from a building dataset, these patterns can serve as training data for creating an occupancy prediction model. Such occupancy prediction is valuable input for energy forecasting. In this study, Windesheim utilized an inverse grey-box modelling approach to address the research questions. As the name suggests, inverse grey-box modelling combines characteristics of both white-box modelling and black-box modelling.

The following conclusions can be drawn regarding the main research questions:

A. To what extent can we accurately derive the ventilation flow rate in a room from CO_2 and occupancy data?

   Using a molar CO_2 concentration balance model implemented in GEKKO Python for an inverse grey box analysis, we managed to learn ventilation flow rates from the recorded data on CO_2 concentration and occupancy. However, the results for learning ventilation flow rates were often unsatisfactory, with the rolling mean absolute error in the learned valve fractions (indications of ventilation flow rate) exceeding 20% for both rooms. This was the case despite the model's good fit to CO_2 concentration data (mean absolute error less than 25 ppm). Notably, the error in the valve fraction seemed to worsen as the CO_2

concentration approached external levels, i.e., when CO_2 elevation was low. This observation was consistent with our preliminary analyses on synthetic room data, raising concerns about the reliability of the results. Therefore we recommend a more direct study of the relationship between CO_2, occupancy, and heat loss through ventilation.

B. To what extent can we accurately deduce occupancy in a room from CO_2 and ventilation data?

   Using a molar CO_2 concentration balance model implemented in GEKKO Python for an inverse grey box analysis, we succeeded in learning occupancy from recorded data on CO_2 concentration and ventilation flow rates. The model exhibited a good fit to CO_2 concentration data (mean absolute error less than 25 ppm). Furthermore, the mean absolute error for learned occupancy was also low (less than 1), indicating that we could reliably derive occupancy data from the measured CO_2 and ventilation data. To assess the applicability of these findings in different scenarios and with different error criteria, such as in meeting spaces and classrooms, additional research specific to these settings would be necessary. This research would involve comparing the results against standards established through discussions with relevant stakeholders.

The objective of O-Nexus is to develop a functional prototype for energy management in SME buildings. In pursuit of this goal, O-Nexus explored two distinct methods for heating forecast modeling. In the physical equation-based approach, they gained insights into how heat loss operates in buildings and recognized its limitations in the presence of noisy data. On the other hand, in the data-driven approach, O-NEXUS achieved slightly superior results, albeit unsuitable for real-time production environments due to limited accuracy and slow fitting times. The key takeaways are the following:

- A fully cloud-based power management system is neither cost-effective nor technically practical. Transmitting messages for electricity readings and sending commands to respond to those readings consumes a significant amount of data traffic. O-Nexus aims to keep subscription costs low, which necessitates minimizing internet traffic. Therefore, the learning here is to offload certain calculations to the Gateway, while the cloud-based approach remains valuable for rapid prototyping.
- It's crucial to log the power management strategy. During our experiments, the Gateway exhibited unexpected behaviors, such as activating when it shouldn't. In the initial setup, it was challenging to determine whether the error originated from the cloud or the hardware side. However, once we began logging the strategy, it became evident where the error was coming from.
- Mapping cloud commands to hardware actions is not straightforward, as the hardware must transparently indicate its capacity in watts to function correctly. Each storage device can have multiple operating modes, some offering only on/off functionality, while others provide more options. Regardless of the brand, the storage device should offer clear instructions for integration with different technologies, although this is not always the case. Therefore, it's advisable to select the storage device first and confirm its compatibility with the chosen technology.
- Even seemingly simple storage devices may have their unique logic, as they are not originally designed to be controlled externally. For example, the Flamco Flextherm may refuse to charge if it is nearly full. The challenge with such behaviors is that they are often undocumented, necessitating reverse engineering to understand them.
- Determining the state of charge of a storage device is not straightforward. Most storage devices lack automated methods to gauge their charge level and instead rely on manufacturer 'expectations' based on various assumptions. In theory, the energy input should closely match the energy output minus some heat and conversion losses. However, in practice, multiple factors, including installation quality, insulation, pipe material, and the device's internal logic (as mentioned earlier), can influence the state of charge.
- The quality of the installation can significantly impact the performance of the storage device. During our experiments, we observed that poor installation can result in substantial heat losses and significantly lower performance. It is advisable to thoroughly inspect the installation, monitor its performance, and ensure alignment between its performance and the manufacturer's 'expectations' before commencing with power management.

9.4 Black box modelling approaches

O-Nexus explored a method for heating forecast modelling based on black box models to develop a functional prototype for energy management in SME buildings. In this data-driven approach, O-Nexus achieved slightly superior results compared to the white box modelling approach, albeit unsuitable for real-time production environments due to limited accuracy and slow fitting times, In general, O-Nexus found the data-driven approach more suitable when striving for precise predictions, while the physical approach can be valuable in

situations where data availability is limited. Additionally, it serves as a valuable learning experience within the domain.

The objective of Peutz's black model approach is to predict energy usage in buildings while also providing a measure of prediction uncertainty. To enhance scalability in both energy usage prediction and uncertainty estimation, Peutz has chosen feedforward neural networks and committee machines as the primary machine learning concepts. This approach, demonstrated in two test cases, adopts a purely data-driven black-box methodology, offering versatility and supporting scalability. The method involves a stacked sequence of committee machines comprised of neural networks, each serving a distinct purpose. The strength of this approach lies in its ensemble prediction capabilities and the recognition that data inherently contains uncertainty in the form of absolute deviation. Utilizing a rolling window technique, the method can predict both energy usage and uncertainty, accommodating potential changes in building usage. The two test cases exhibit sufficient diversity to warrant the careful observation that the method indeed possesses a broad range of potential applications. Further testing in various scenarios and applications will be necessary to validate this observation thoroughly.

DWA develops black-box models for optimizing grid demand and supply, especially in the context of network congestion. To prioritize modeling efforts, an analysis of the grid's demand and supply profiles is conducted. In the current Dutch landscape, network congestion poses a significant challenge, particularly in terms of energy supply to the grid. PV solar systems play a pivotal role in this context, making it essential to develop a highly accurate black-box model capable of predicting peak energy production from PV solar systems. This effort is an ongoing project, and following the outlined priorities in this chapter, additional black-box models will be developed in the future. Here are some key points:

1. For the development of load forecasting models, it is advisable to begin by evaluating the smart meter of the power grid and submeters within the building. This approach allows for an energy flexibility assessment based on dynamic data, facilitating the selection of modeling strategies, such as determining which models should be prioritized for high accuracy and which loads can be aggregated into more generalized models.
2. Load forecasting methods can also serve as effective tools for fault detection.
3. Faulty operations present additional challenges to load forecasting models, whether they are black-box models trained on fault-free data or white-box models that do not explicitly account for faults. In future research, it is valuable to assess the performance of black-box and hybrid models trained or calibrated using faulty datasets. However, it's essential to recognize that while fault detection and diagnosis offer advantages, the requirement for a fault-free building can potentially hinder the practical application of load forecasting.
4. On a fault-free test set, the developed black-box model for PV solar systems meets the requirements for a calibrated model according to the FEMP and ASHRAE criteria for error metrics, using 6.5 months of training data.

REFERENCES

References Chapter 1

Chen, Y., Guo, M., Chen, Z., Chen, Z., & Ji, Y. (2022). Physical energy and data-driven models in building energy prediction: A review. Energy Reports, 8, 2656–2671. https://doi.org/10.1016/j.egyr.2022.01.162

Habash, R. (2022). 6 - Building as a control system. In R. Habash (Ed.), Sustainability and Health in Intelligent Buildings (pp. 161–189). Woodhead Publishing. https://doi.org/10.1016/B978-0-323-98826-1.00006-5

Harish, V. S. K. V., & Kumar, A. (2016). A review on modeling and simulation of building energy systems. Renewable and Sustainable Energy Reviews, 56, 1272–1292. https://doi.org/10.1016/j.rser.2015.12.040

IEA. (2019). Global Status Report for Buildings and Construction 2019. https://www.iea.org/reports/global-status-report-for-buildings-and-construction-2019

Li, Y., O'Neill, Z., Zhang, L., Chen, J., Im, P., & DeGraw, J. (2021). Grey box modeling and application for building energy simulations - A critical review. Renewable and Sustainable Energy Reviews, 146, 111174. https://doi.org/10.1016/j.rser.2021.111174

Marszal-Pomianowska, A. J. (2019). Characterization of energy flexibility in buildings (Armin Knotzer, Pernetti Roberta, Søren Østergaard Jensen, AEE INTEC, EURAC, & DTI, Eds.). Danish Technological Institute. https://www.annex67.org/publications/reports/

Nuytten, T., Claessens, B., Paredis, K., Van Bael, J., & Six, D. (2013). Flexibility of a combined heat and power system with thermal energy storage for district heating. Applied Energy, 104, 583–591. https://doi.org/10.1016/j.apenergy.2012.11.029

OECD. (2021). Renewables 2021: Analysis and forecast to 2026. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/energy/renewables-2021_6dcd2e15-en

Reynders, G., Amaral Lopes, R., Marszal-Pomianowska, A., Aelenei, D., Martins, J., & Saelens, D. (2018). Energy flexible buildings: An evaluation of definitions and quantification methodologies applied to thermal storage. Energy and Buildings, 166, 372–390. https://doi.org/10.1016/j.enbuild.2018.02.040

SEII, IISD, ODI, E3G, & UNEP. (2021). The Production Gap Report 2021. http://productiongap.org/2021report

Six, D., Desmedt, J., Houdt, D. van, & Bael, J. van. (2011). Exploring the flexibility potential of residential heat pumps combined with thermal energy storage for smart grids. Paper No 0442.

United Nations. (2015). The Paris Agreement. https://www.un.org/en/climatechange/paris-agreement

Wilde, P. de (2023). Building performance simulation in the brave new world of artificial intelligence and digital twins: A systematic review. Energy and Buildings, 292, 113171. https://doi.org/10.1016/j.enbuild.2023.113171

Yu, J., Chang, W.-S., & Dong, Y. (2022). Building Energy Prediction Models and Related Uncertainties: A Review. Buildings, 12(8), 1284. https://doi.org/10.3390/buildings12081284

Zsiborács, H., Pintér, G., Vincze, A., Birkner, Z., & Baranyai, N. H. (2021). Grid balancing challenges illustrated by two European examples: Interactions of electric grids, photovoltaic power generation, energy storage and power generation forecasting. Energy Reports, 7, 3805–3818. https://doi.org/10.1016/j.egyr.2021.06.007

References Chapter 2

Boogaard, S., Cheung, D., & Salcedo Rahola, T. B. (2023). Building energy flexibility assessment with static data. 2023 International Conference on Future Energy Solutions (FES), Vaasa, Finland. https://doi.org/10.1109/FES57669.2023.10183044

CBS. (2021). Voorraad woningen en niet-woningen; mutaties, gebruiksfunctie, regio. https://opendata.cbs.nl/#/CBS/nl/dataset/81955NED/table?ts=1687590639439

Eck, H. van. (2016). EPBD Implementation in The Netherlands, Status in December 2016. Concerted action energy performance of buildings. https://epbd-ca.eu/wp-content/uploads/2018/08/CA-EPBD-IV-The-Netherlands-

2018.pdf

Johra, H., Energy in Buildings and Communities Programme, & Annex 67 Energy Flexible Buildings, T. I., Technology Collaboration Programme. (2019). Examples of energy flexibility in buildings. Danish Technological Institute.

Junker, R. G., Azar, A. G., Lopes, R. A., Lindberg, K. B., Reynders, G., Relan, R., & Madsen, H. (2018). Characterizing the energy flexibility of buildings and districts. Applied Energy, 225, 175–182. https://doi.org/10.1016/j.apenergy.2018.05.037

Kremer, A. (2022). Gebouwenmatrix energie 2020 op 1 januari 2020 en 1 januari 2021. CBS. https://www.cbs.nl/-/media/_excel/2022/13/gebouwenmatrix-energie.xls

Lopes, R. A., Chambel, A., Neves, J., Aelenei, D., & Martins, J. (2016). A Literature Review of Methodologies Used to Assess the Energy Flexibility of Buildings. Energy Procedia, 91, 1053–1058. https://doi.org/10.1016/j.egypro.2016.06.274

Marszal-Pomianowska, A. J., Energy in Buildings and Communities Programme, & Annex 67 Energy Flexible Buildings, T. I., Technology Collaboration Programme. (2019). Characterization of energy flexibility in buildings. Danish Technological Institute.

Ministry of Interior and Kingdom Relations. (2018). Decree of October 17, 2018, amending the 2012 Building Code regarding the label requirement for office buildings.

NEN. (2021). NEN Whitepaper 'Highlights NTA 8800. https://www.nen.nl/media/wysiwyg/NEN_Whitepaper_Highlights_NTA_8800.pdf

References Chapter 3

Atam, E. (2017). Current software barriers to advanced model-based control design for energy-efficient buildings. Renewable and Sustainable Energy Reviews, 73, 1031–1040. https://doi.org/10.1016/j.rser.2017.02.015

Brown, G. (1990). The BRIS simulation program for thermal design of buildings and their services. Energy and Buildings, 14(4), 385–400. https://doi.org/10.1016/0378-7788(90)90100-W

Crawley, D. B., Hand, J. W., Kummert, M., & Griffith, B. T. (2008). Contrasting the capabilities of building energy performance simulation programs. Building and Environment, 43(4), 661–673. https://doi.org/10.1016/j.buildenv.2006.10.027

Dols, I., & Salcedo Rahola, T. B. (2023). From energy flexibility to design choice. 2023 International Conference on Future Energy Solutions (FES), Vaasa, Finland, https://doi.org/10.1109/FES57669.2023.10182994 .

Formhals, J., Hemmatabady, H., Welsch, B., Schulte, D. O., & Sass, I. (2020). A Modelica Toolbox for the Simulation of Borehole Thermal Energy Storage Systems. Energies, 13(9), 2327. https://doi.org/10.3390/en13092327

IBPSA Project 1. (n.d.). Retrieved June 19, 2023, from https://github.com/ibpsa/modelica-ibpsa

IBPSA. (n.d.). BEST directory. Retrieved May 5, 2023, from https://www.buildingenergysoftwaretools.com/

IEA EBC Annex 67. (2017). Energy Flexibility as a key asset in a smart building future. https://www.annex67.org/publications/position-paper/

Junker, R. G., Azar, A. G., Lopes, R. A., Lindberg, K. B., Reynders, G., Relan, R., & Madsen, H. (2018). Characterizing the energy flexibility of buildings and districts. Applied Energy, 225, 175–182. https://doi.org/10.1016/j.apenergy.2018.05.037

Singh, M., & Sharston, R. (2022). A literature review of building energy simulation and computational fluid dynamics co-simulation strategies and its implications on the accuracy of energy predictions. Building Services Engineering Research and Technology, 43(1), 113–138. https://doi.org/10.1177/01436244211020465

Syb van Breda. (2009). The Hague University building Delft. https://sybvanbreda.com/en/project/the-hague-university-building-delft/

Taal, A. C. (2021). A new approach to automated energy performance and fault detection and diagnosis of HVAC

systems: Development of the 4S3F method [Phd Thesis 1 (Research TU/e / Graduation TU/e)]. Eindhoven University of Technology.

U.S. department of Energy. (n.d.). OpenBuildingControl. https://www.energy.gov/eere/buildings/openbuildingcontrol

U.S. department of Energy. (n.d.). Spawn of EnergyPlus. https://www.energy.gov/eere/buildings/articles/spawn-energyplus-spawn

Wang, Z., Hong, T., & Jia, R. (2019). Buildings.Occupants : a Modelica package for modelling occupant behaviour in buildings. Journal of Building Performance Simulation, 12(4), 433–444. https://doi.org/10.1080/19401493.2018.1543352

Wetter, M. (2023). IBPSA Modelica Working Group. IBPSA.

References Chapter 4

Behrouz, E. M. , Spiekman, M. and Hoes, L. (2023). Occupant Models for Use in Hybrid Building Models. Deliverable 3.08. Brains for Buildings.

King, D. L., Boyson,W.E. and Kratochvil, J.A. (2004). Photovoltaic Array Performance Model. Sandia Report No. 2004-3535. http://www.osti.gov/bridge/product.biblio.jsp?osti_id=919131.

King, D L., Gonzalez, S., Galbraith, G. M. and Boyson, W. E.. (2007). Performance Model for Grid-Connected Photovoltaic Inverters. Sandia Report No. 2007-5036.

Stein, J. S., Holmgren, W. F., Forbess, J. and Hansen, C. W. (2016). PVLIB: Open Source Photovoltaic Performance Modeling Functions for Matlab and Python. Conference Record of the IEEE Photovoltaic Specialists Conference 2016-Novem(June): 3425–30.

References Chapter 5

Bacher, P., & Madsen, H. (2011). Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings, 43*(7), 1511–1522. https://doi.org/10.1016/j.enbuild.2011.02.005.

Beal, L. D. R., Hill, D. C., Martin, R. A., & Hedengren, J. D. (2018). GEKKO Optimization Suite. *Processes*, *6*(8), Article 8. https://doi.org/10.3390/pr6080106

Brains4Buildings2022 dataset. (2023). [CSV, Parquet]. Research group Energy Transition at Windesheim. https://github.com/energietransitie/twomes-dataset-windesheim-brains4buildings2022 (Original work published 2022)

Freyer, A. J. (2023). *Monitor* [Shell]. https://github.com/andrewjfreyer/monitor (Original work published 2018)

Haskell, W. L., Lee, I.-M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., Macera, C. A., Heath, G. W., Thompson, P. D., & Bauman, A. (2007). Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Medicine and Science in Sports and Exercise*, *39*(8), 1423–1434. https://doi.org/10.1249/mss.0b013e3180616b27

Hedengren, J. D., & Beal, L. D. R. (2022). *Modes—GEKKO 1.0.5 documentation*. https://gekko.readthedocs.io/en/latest/imode.html

Hedengren, J. D., Shishavan, R. A., Powell, K. M., & Edgar, T. F. (2014). Nonlinear modeling, estimation and predictive control in APMonitor. *Computers & Chemical Engineering*, *70*, 133–148. https://doi.org/10.1016/j.compchemeng.2014.04.013

Lan, X., E.J. Dlugokencky, J.W. Mund, A.M. Crotwell, M.J. Crotwell, E. Moglia, M. Madronich, D. Neff and K.W. Thoning (2022), *Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA GML Carbon Cycle Cooperative Global Air Sampling Network*, 1968-2021, Version: 2022-11-21, https://doi.org/10.15138/wkgj-f215 \

Leprince, J., Madsen, H., Miller, C., Real, J. P., van der Vlist, R., Basu, K., & Zeiler, W. (2022). Fifty shades of grey: Automated stochastic model identification of building heat dynamics. *Energy and Buildings, 266*, 112095. https://doi.org/10.1016/j.enbuild.2022.112095.

Li, M., Bekö, G., Zannoni, N., Pugliese, G., Carrito, M., Cera, N., Moura, C., Wargocki, P., Vasconcelos, P., Nobre, P., Wang, N., Ernle, L., & Williams, J. (2022). Human metabolic emissions of carbon dioxide and methane and their implications for carbon emissions. *Science of The Total Environment*, *833*, 155241. https://doi.org/10.1016/j.scitotenv.2022.155241

LMP Layer Tutorial. (n.d.). Retrieved July 2, 2023, from https://www.amd.e-technik.uni-rostock.de/ma/gol/lectures/wirlec/bluetooth_info/lmp.html#Name%20Request

Madsen, H., & Holst, J. (1995). Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings, 22*(1), 67–79. https://doi.org/10.1016/0378-7788(94)00904-X.

Monroe, R. (n.d.). *The Keeling Curve*. The Keeling Curve. Retrieved July 2, 2023, from https://keelingcurve.ucsd.edu

Omroep West. (2023, February 16). *Leidse universiteit haalt camera's weg na zorgen over privacy*. https://nos.nl/artikel/2464026-leidse-universiteit-haalt-camera-s-weg-na-zorgen-over-privacy

Peters, S. (2022). *Stephanpcpeters/HourlyHistoricWeather* [Python]. https://github.com/stephanpcpeters/HourlyHistoricWeather (Original work published 2021)

Redactie. (2021, December 4). *Rel op hogeschool: Student trekt stekker uit 'camera' die geen camera is (maar wat dan wel?)*. gelderlander.nl. https://www.gelderlander.nl/binnenland/rel-op-hogeschool-student-trekt-stekker-uit-camera-die-geen-camera-is-maar-wat-dan-wel~a54a8005/

Ter Hofte, G. H. (2022a). *Brains4Buildings data collection Windesheim winter 2022- 2023*. https://github.com/energietransitie/twomes-dataset-windesheim-brains4buildings2022/blob/main/data_management/DMP_online_Brains4Buildings_data_collection_Windesheim_winter_2022-2023.pdf

Ter Hofte, G. H. (2022b). *Vragenlijst_B4B_Bluetooth-extra*. https://github.com/energietransitie/twomes-dataset-windesheim-brains4buildings2022/blob/main/data_management/Vragenlijst_B4B_Bluetooth-extra.pdf

Ter Hofte, G. H., Van Ravenzwaaij, N., & Nijboer, E. (2023). *Brains4Buildings2022 dataset* [CSV, parquet]. Research group Energy Transition at Windesheim. https://github.com/energietransitie/twomes-dataset-windesheim-brains4buildings2022 (Original work published 2022)

Twomes CO_2 and bluetooth occupancy firmware for the M5CoreInk and SCD41. (2023). [C]. Research group Energy Transition at Windesheim. https://github.com/energietransitie/twomes-co2-occupancy-scd41-m5coreink-firmware (Original work published 2021)

Twomes inverse grey box modelling and analysis tools for homes and utility buildings. (2023). [Python]. Research group Energy Transition at Windesheim. https://github.com/energietransitie/twomes-inverse-grey-box-analysis (Original work published 2021)

Volledige Privacyverklaring Brains4Buildings. (2022, September 21). https://www.energietransitiewindesheim.nl/brains4buildings2022/privacy/privacy-full.html

References Chapter 6

Amara, F. , Agbossou, K. , Cardenas, A. , Dubé, Y. and Kelouwani, S. (2015) Comparison and Simulation of Building Thermal Models for Effective Energy Management. Smart Grid and Renewable Energy, 6, 95-112. doi: 10.4236/sgre.2015.64009.

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, 1192-1205.

Berisha, X., & Dragusha, B. (2015, September). The influence of the outside temperature during the design of a heating system. In *International Conference &Workshop REMOO-2015, Budva*.

Eseye, A. T., & Lehtonen, M. (2020). Short-term forecasting of heat demand of buildings for efficient and optimal energy management based on integrated machine learning models. IEEE Transactions on Industrial Informatics, 16(12),7743-7755.

Findernet. Energy meters. https://www.findernet.com/en/worldwide/residential-and-commercial/power/energy-

meters/

Flamco. 18201 - FlexTherm Eco 6E. https://flamco.aalberts-hfc.com/nl/catalog/buffervaten-en-boilers/thermische-batterijen/flextherm-eco/flextherm-eco/18201/groups/g+c+p+a+nr+view

Frayssinet, L., Merlier, L., Kuznik, F., Hubert, J. L., Milliez, M., & Roux, J. J. (2018). Modeling the heating and cooling energy demand of urban buildings at city scale. *Renewable and Sustainable Energy Reviews*, *81*, 2318-2327.

Karasu, S., & Altan, A. (2019, November). Recognition model for solar radiation time series based on random forest with feature selection approach. In 2019 11th international conference on electrical and electronics engineering (ELECO) (pp. 8-11). IEEE.

Lizana, J., Friedrich, D., Renaldi, R., & Chacartegui, R. (2018). Energy flexible building through smart demand-side management and latent heat storage. *Applied energy*, *230*, 471-485.

López, C. J. (2017). Data-driven Predictive Control for Heating Demand in Buildings. TU Delft. https://repository.tudelft.nl/islandora/object/uuid:15ee9126-0c03-46fd-988f-b7d583cc2398

Open. Calculating the total heat loss of a house. https://www.open.edu/openlearn/nature-environment/energy-buildings/content-section-2.4.1

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37-45.

Wei, Z., Zhang, T., Yue, B., Ding, Y., Xiao, R., Wang, R., & Zhai, X. (2021). Prediction of residential district heating load based on machine learning: A case study. Energy, 231, 120950.

Wms. Warmtemeters. https://www.wms.nl/producten/warmtemeters

Zhang, T., Zhang, X., Liu, Y., Chow, Y., Iu, H., & Fernando, T. (2023). Long-term Energy and Peak Power Demand Forecasting based on Sequential-XGBoost. *IEEE Transactions on Power Systems*.

References Chapter 7

Maarif, M.R., et al. (2023) Energy Usage Forecasting Model Based on Long Short-Term Memory (LSTM) and eXplainable Artificial Intelligence (XAI), *Information*, *14*(5), 265, https://doi.org/10.3390/info14050265

Mitchell, T.M. (1997) *Machine Learning*, The McGraw-Hill Companies, Inc., ISBN 978-0-07-042807-2

Fort, S., Hu, H., Lakshminarayanan, B. (2020), Deep Ensembles: A Loss Landscape Perspective, *arXiv:1912.02757*, https://doi.org/10.48550/arXiv.1912.02757

Scher, S., Messori, G. (2018) Predicting weather forecast uncertainty with machine learning, *Quarterly Journal of the Royal Meteorological Society*, *144*(717), 2830-2841, https://doi.org/10.1002/qj.3410

Rahmati, O., et al. (2019) Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods, *Science of The Total Environment*, *688*, 855-866, https://doi.org/10.1016/j.scitotenv.2019.06.320

Harakeh, A., Waslander, S.L. (2021), Estimating and Evaluating Regression Predictive Uncertainty in Deep Object Detectors*, arXiv:2101.05036*, https://doi.org/10.48550/arXiv.2101.05036

Metsä-Eerola, I., et al. (2022), Hourly Forecasting Heating Energy Consumption of HVAC with Recurrent Neural Networks, *Energies*, *15*, 5084, https://doi.org/10.3390/en15145084

Pîrjan, A., et al. (2017), Devising Hourly Forecasting Solutions Regarding Electricity Consumption in the Case of Commercial Center Type Consumers, *Energies*, *10*, 1727, http://dx.doi.org/10.3390/en10111727

References Chapter 8

Knotzer, A., Pernetti, R., Jensen, S.O., (2019). Characterization of Energy Flexibility in Buildings, *Energy in Building and Communities Programme Annex 67 Energy flexible buildings.* ISBN: 978-87-93250-09-3

Telegraaf (13 juni 2023). Stroomnet in grote delen van Nederland op slot.

Ke, Y., S. Mumma. (1997). Using carbon dioxide measurements to determine occupancy for ventilation controls. ASHRAE Transaction 103(2):365–374

Wisse, C.J., (2023), Comfort-apps, hoe vaak gebruiken we ze? TVVL-magazine nr. 01 / februari 2023

Markovics, D., Mayer, M.J., (2022). Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction, *Renewable and Sustainable Energy Reviews 161 (2022) 112364* *https://doi.org/10.1016/j.rser.2022.112364*

Khan, W., Liao, J.Y., Walker, S., Zeiler, W., (2022a). Impact assessment of varied data granularities from commercial buildings on exploration and learning mechanism. Applied Energy 319 (2022) 119281 https://doi.org/10.1016/j.apenergy.2022.119281

Khan, W., Walker, S., Zeiler, W., (2022b). Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. Energy 240 (2022) 122812, https://doi.org/10.1016/j.energy.2021.122812

R. Perez, R. Seals, P. Ineichen, R. Stewart en D. Menicucci, (1987) "A new simplified version of the Perez diffuse irradiance model for tilted surfaces," *Solar Energy*, vol. 39, No. 3, pp. 221-231. https://doi.org/10.1016/S0038-092X(87)80031-2

Ruiz, G.R., Bandera, C.F., (2017) Validation of Calibrated Energy Models: Common Errors. Energies 2017, 10, 1587, https://doi:10.3390/en10101587

Chitkara, S., Thamban, A., Zeiler, W., (2022) Approaches to ML & system-based diagnose and conclusions on the beta design of a software module. B4B Deliverable D1.8a.

10 APPENDIX A: OFFICE BUILDING CHARACTERISTICS

Authors: Bram Muller (HHS), Baldiri Salcedo (HHS), Nino Visser (HHS)

10.1 Goal of this appendix

The initial goal for characterizing office buildings was to prioritize the building configurations to be modelled in this report. The prioritization of the building configurations to be modelled has not occurred as the average office building has currently limited energy flexibility capabilities and therefore is not a good reference model for the control of energy flexibility. However, the overview of the current office building characteristics offers interesting insights that can be used for developing energy building models.

10.2 Office Buildings in the Netherlands

The Brains for Buildings project focuses on non-residential buildings' energy building management capabilities. Non-residential real estate objects in the Netherlands account for 14% of the total real estate objects and represent a floor surface area of 38% of the total real state objects surface area in the Netherlands. See Table A.1. There is a great variety of non-residential real estate objects. In a recent request to the Central Agency for Statistics of the Netherlands information about the different typologies of non-residential buildings have been compiled in one matrix table (Kremer, 2022). In this matrix 26 different categories have been defined. See Table A.2.

Table A.1 Residential and non-residential stock (CBS, 2021).

	Real state objects	Floor surface in m2
Residential	7 966 331	947 993 389
Non-residential	1 302 433	590 937 000

Table A.2 Non-residential buildings per type building (Kremer, 2022).

	Real Stage Objects	Floor surface in 1000 m2
House	298572	61963
Office	95237	62663
Shop without refrigeration	128430	44087
Supermarket	5968	4199
Cafe, restaurant	31780	7584
Event venue	28589	19968
Day care	4329	2158
Hotel	7683	7833
Lodging, other	59358	5732
School, primary	8914	12182
School, secondary	2068	10850
School, vocational	287	1410
School, tertiary	237	2085
School, other/unknown	3035	7532
Practice	18216	7695
Laboratory	179	583
Hospital	494	6007
Nursing home	65998	12105
Penitentiary	73	860
Indoor sports	11347	10307
Outdoor sports	10476	5519
Swimming pool/Sauna	1657	2617
Car company	27888	20905
Warehouse	479452	240450
Agriculture	11983	32603

	Real Stage Objects	Floor surface in 1000 m2
Data centre	183	1040
Total	1302433	590937

The case studies used in the B4B project are mainly offices and educational buildings. Offices account for the 7.3% of the real state objects and represent 10.6% of the floor surface and educational buildings account for 1.1% and represent 5.7% of the floor surface of the non-residential real state objects total.

In the current literature there is a lack of information about the non-residential building characteristics, from the characteristics of the building envelope as well as the characteristics of the building installations. This chapter aims to give some insights into the office building energy characteristics.

This chapter focuses only on office buildings because the analysis used the data from the building portfolio of The Central Government Real Estate Agency (CGREA), *Rijksvastgoedbedrijf*. The main question of the case study is:

- What are the energy characteristics of the office buildings of the Central Government Real Estate Agency?

This information can be especially of interest when aiming to develop energy building computer simulations to analyse building energy flexibility. In the next section there is a description of the data sources used in the case study, followed by a general description of the building portfolio characteristics. Afterwards there is a discussion of the possible use of this analysis and the chapter is finalized with the analysis conclusions, limitations of the study and recommendations for further research.

10.3 Method and data sources

The information collected to carry out the analysis has been extracted from two databases of the CGREA. The first data base is the Government Real State Portal that contains some information about all the building in the building portfolio of CGREA. The second database is the Filenet, that is the shared storage drive at CGREA.

Data Government Real State Portal

The information available on the portal have been used to make the overview of the general building portfolio characteristics presented in section A.4. Moreover, it was possible to make a selection of the office buildings of the central government. The list contains 454 buildings and contains among others the following information fields.

- Reference Code
- Address
- Gross Floor Area
- Energy Label
- Construction Year
- Building Type

The Government Real State portal needs to contain information about the specific characteristics of the building construction type or about the building services. It is also important to note that the current portal has numerous empty fields on the general building information. For example, for 26% of the analysed office buildings there is no information about the gross floor area and for 50% of the analysed office buildings there is no information about the energy label.

EPAU files stored at Filenet

In the Netherlands, as in other European countries, an energy certificate is required for buildings. Since 2008, the energy certificate is mandatory for new buildings and for the sale of existing buildings (Eck, 2016). This means that not all existing buildings currently have an energy certificate. In the non-residential sector, however, almost all office buildings should have an energy certificate, as the Dutch government required all office buildings to obtain a certificate with at least energy class C by 1 January 2023 in order to continue operating (Ministry of Interior and Kingdom Relations, 2018). To obtain a building certificate, a calculation is made using a standard assessment method. Until 2021, the procedure for non-residential buildings was the NEN 7120 standard, after which a new procedure, the NTA8800, has been in use. The new NTA8800 standard uses even more complete data input than the previous version (NEN, 2021).

Vabi EPA-U is the most common software used to make calculations following these standards. The software stores the information on a file format .epau_project. Using the file format, it was possible to identify the Vabi EPA-U files in the Filenet database. In total 321 EPA-U files have been collected from the Filenet, from which

69 files are from office buildings. What means that only 16% of the CGREA offices have an EPAU file stored on Filenet. Most of the building's files analysed were produced according to the NEN 7120 standard.

The EPA-U files are the source of information that has been used to make an overview of the office building characteristics presented in section A.5. The extraction of the parameter's values from the EPA-U files has been done manual, what has been highly time-consuming. There has been an attempt to automatise the extraction of the parameter values. Unfortunately, it has not been successful as there is no API for the VABI EPA-U software and the export capabilities of the software are limited. The two available export functions deliver files with a slightly different structure and mainly with a labelling system that is not unique making the automatization of the extraction process really complex.

10.4 General building portfolio characteristics

The total number of buildings owned or managed by the CGREA in the Netherlands is 12.471 as of April 2023. The portfolio is divided into two big groups, the defence buildings, 81.5%, and the central government buildings, 18.5%. See Figure A.1. In this case study, the defence part of the portfolio has not been taken into account. The central government buildings can be further divided by function, where offices represent 19.3% of the central government buildings. See Figure A.2.

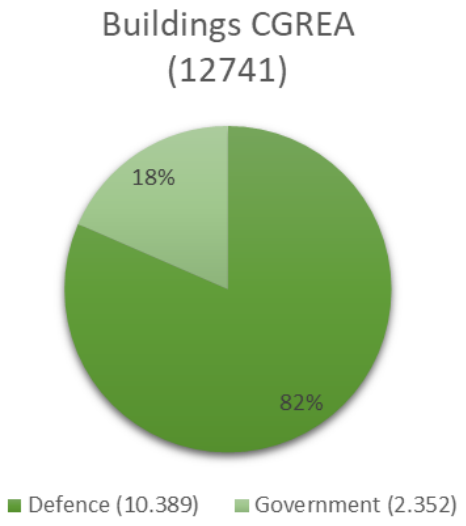


Figure A.1 Total building portfolio CGREA in the Netherlands.

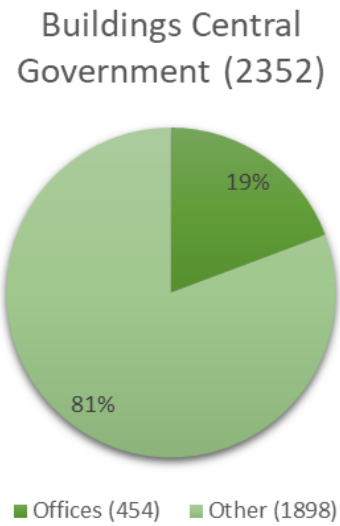


Figure A.2 Office buildings central government.

Most of the office buildings of the central government were built in the period between 1976 and 2010, 57.7%. However, the central government account with quite some historical buildings, 11% of the offices were built before 1900. See Figure A.3.

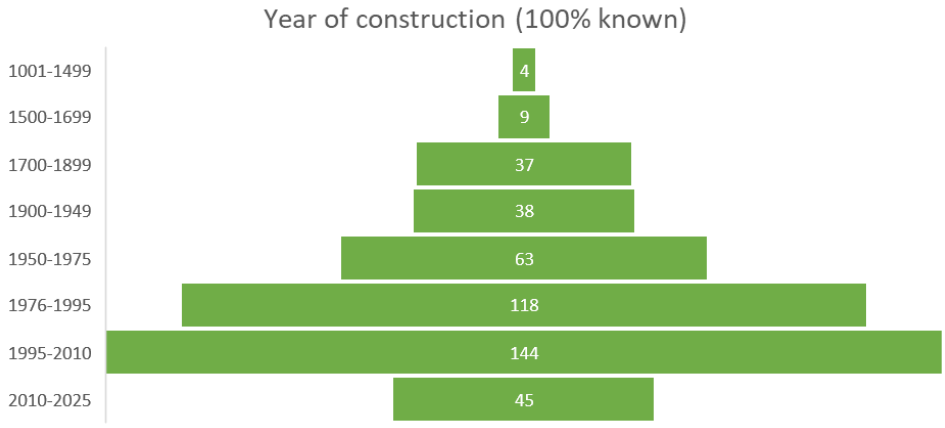


Figure A.3 Year of construction of the Central Government office buildings.

The median gross floor area of the office building of the central government is approximately 9000 m². However, there is a large variety of sizes in gross floor area among the Central Government offices. See Figure A.4. There is also a slight positive relation between the construction year and the gross floor area. The first

office building above 50000 m^2 of the Central Government was built in 1958. See Figure A.5. It should be also taken into account that in the Government Real State Portal the construction year was registered only for 74% of the office buildings.

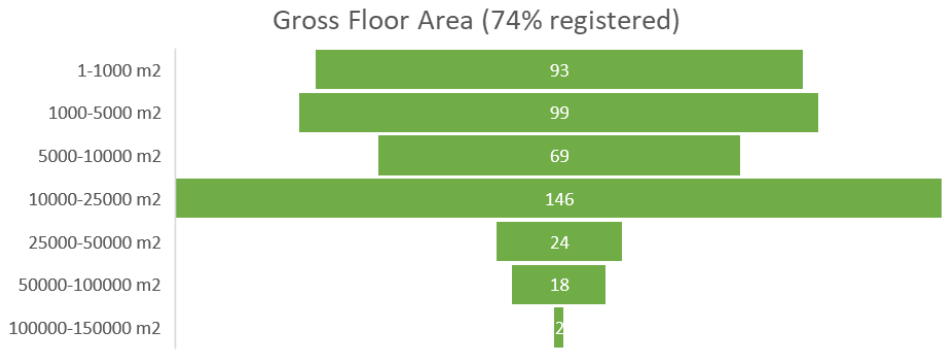


Figure A.4 Gross floor area of the Central Government office buildings.

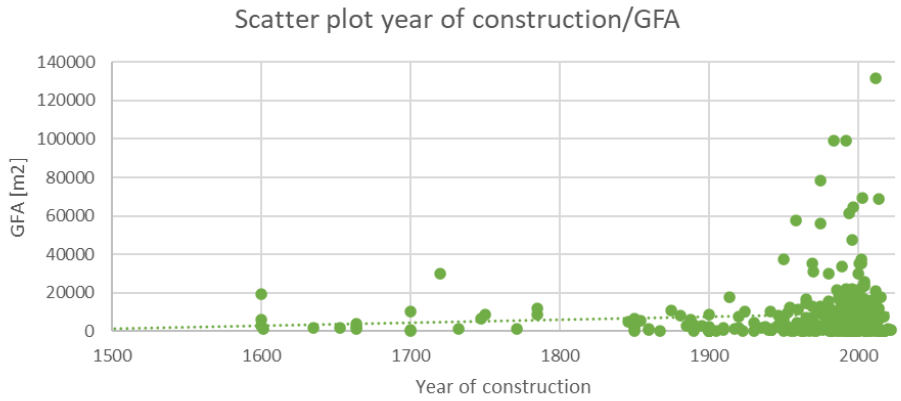


Figure A.5 Scatter plot Year of construction / Gross floor area of the Central Government office buildings.

Around 50% of the Central Government office buildings have an energy label A or higher. However, there are still a considerable amount of building registered as having a label D or lower, 22%. See Figure A.6. It should be taken into account that in the Government Real State Portal the energy label was registered only for 50% of the office buildings and that it was not possible during the analysis period to verify if the energy label values registered in the portal were the current energy label values. There is a slight positive relation between the construction year and the energy label. See Figure A.7. The reason that the relation is not more evident could be explained by the renovation works that have been carried out in older buildings.

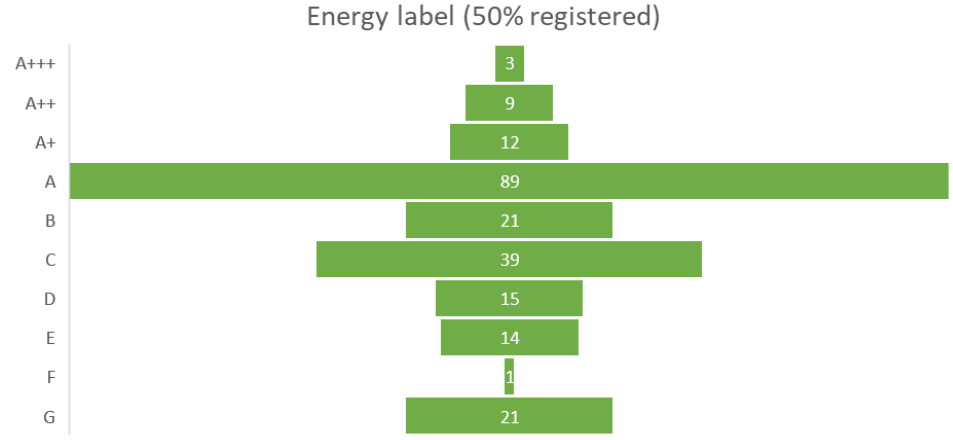
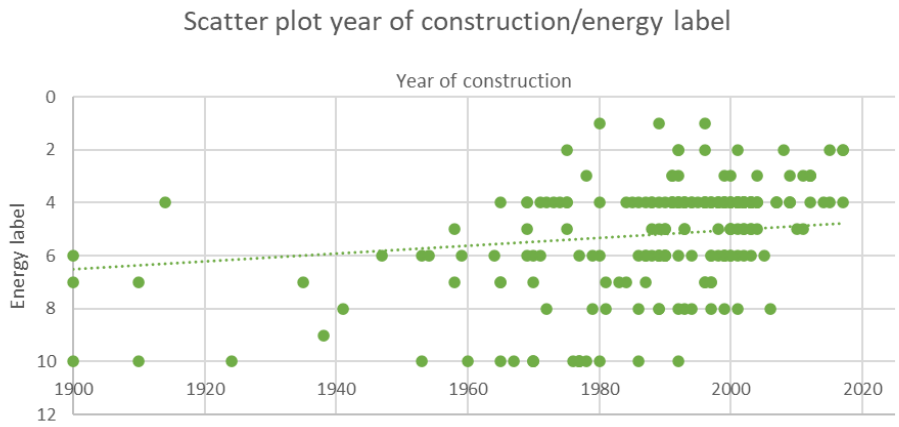


Figure A.6 Energy label of the Central Government office buildings.



Energy label A+++ =1, A++=2, A+=3, A=4, B=5, C=6, D=7, E=8, F=9, G=10.

Figure 9.7 Scatter plot Year of construction / Energy label of the Central Government office buildings

10.5 Office building characteristics

A selection of the building characteristics of the 69 office buildings analysed are presented in this section. The 69 buildings analysed have not been randomly selected but are the ones that had an EPA-U file stored on the CGREA Filenet. Figure A.8, A.9 and A.10 shows the relation of the construction year, gross floor area and energy label between the office portfolio characteristics presented in the previous section and the ones from the 69 analysed buildings, to visualize how representative is the sample regarding the total office building portfolio.

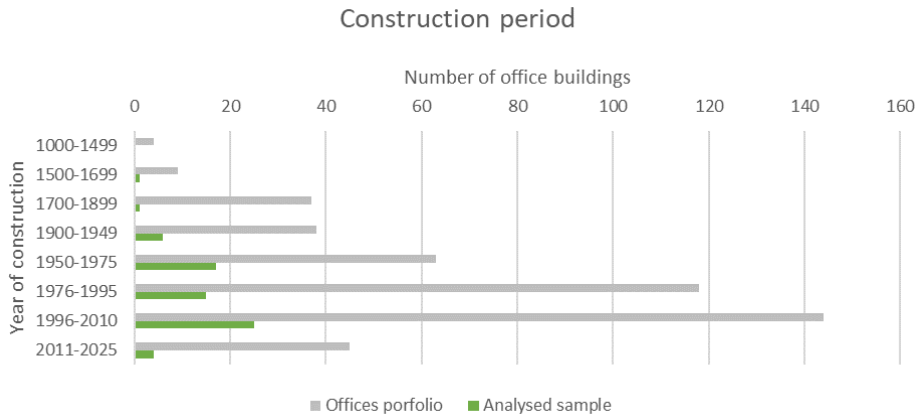


Figure A.8 Relation between the office CGREA portfolio and the analysed sample construction year distribution

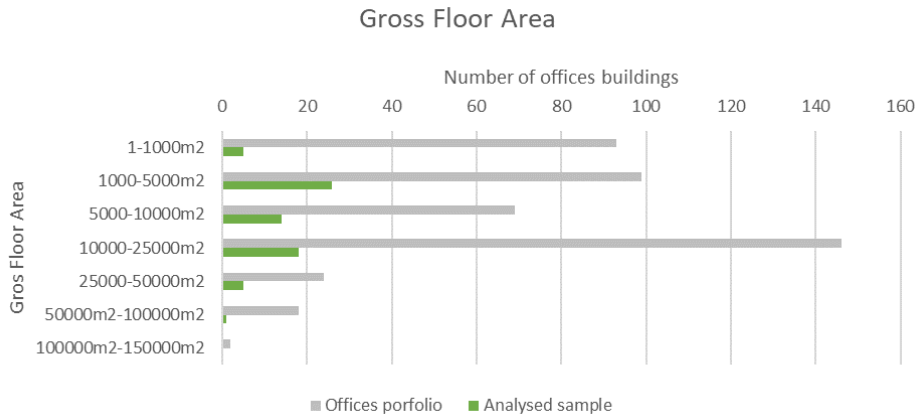


Figure A.9 Relation between the office CGREA portfolio and the analysed sample gross floor area distribution.

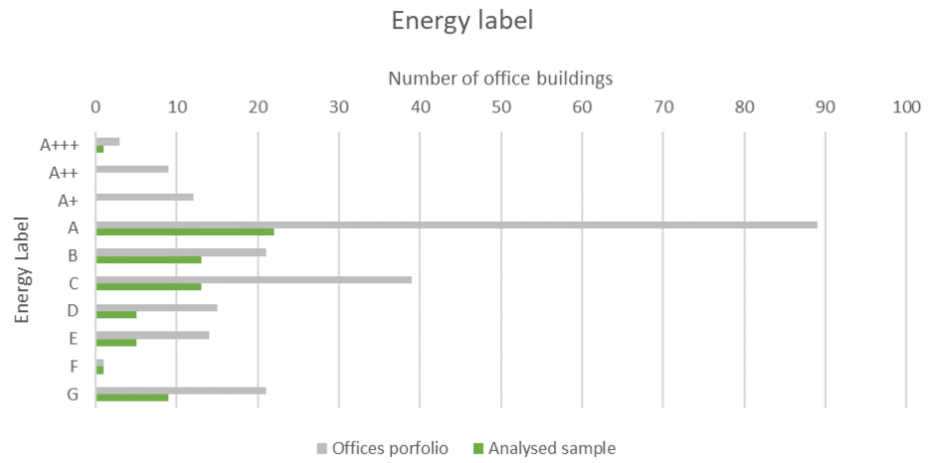


Figure A.10 Relation between the office CGREA portfolio and the analysed sample energy label distribution

The building characteristics analysed are:

- Average envelope RC value
- Construction mass
- Heating systems
- Cooling systems
- Temperature control strategy
- Solar energy

Average envelope Rc value

The average Rc value for the building envelope of the analysed sample is 1,76 [m2K/W]. The sample for the Rc value contained 64 buildings, because 5 of the buildings contained faulty data in regard to surfaces and Rc values and therefore it was not possible to calculate and average Rc value for the whole envelope. The distribution of the Rc values is presented in Figure A.11.

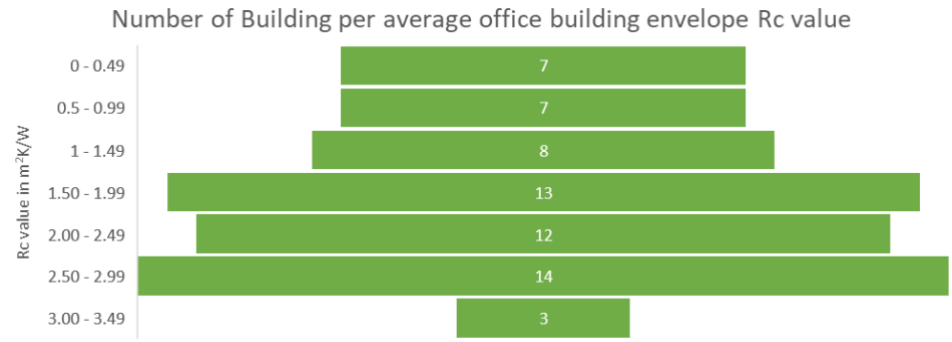


Figure A.11 Average Rc values in [m²K/W] of the analysed sample

Construction mass

On the EPA-U software the thermal mass of a building can be classified in four categories. All the office buildings analysed are above 500kg/m² and %16 of them are above 750kg//m². See Figure A.12

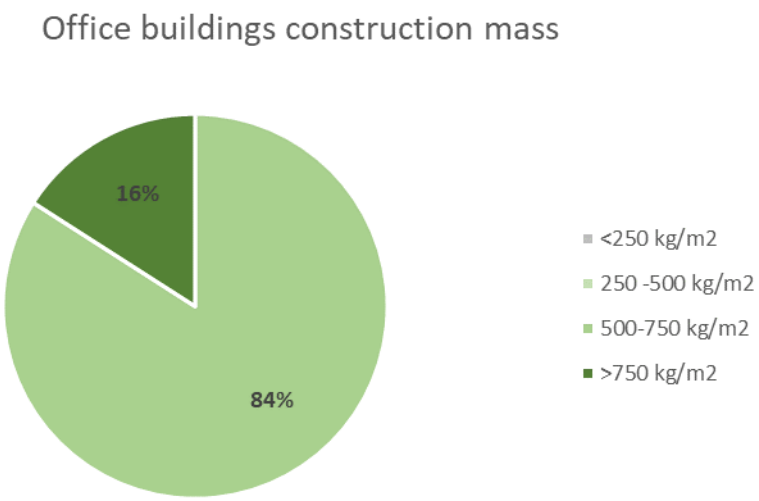


Figure A.12 Percentage of buildings per construction mass of the analysed sample

Heating system

The main heating generation system for more than the half of the office buildings analysed is a gas boiler, and the second most common option, 25%, is third party heat supply (district heating). Electric heat pumps are used in 16% of the office buildings. On the EPA-U software is it possible to differentiate between 4 different types of gas boilers, but for the electric heat pumps there is only one category. See Figure 9.13.

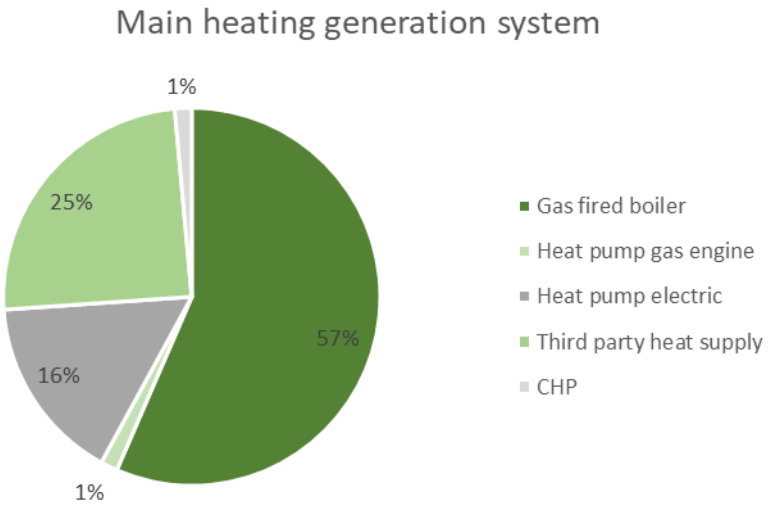


Figure 9.13 Percentage of buildings per main heating generation system.

On the EPA-U software there is also the option to register a secondary heating generation system. Among the buildings equipped with a heat pump 64% have also a secondary heating generation system. See Figure A.14.

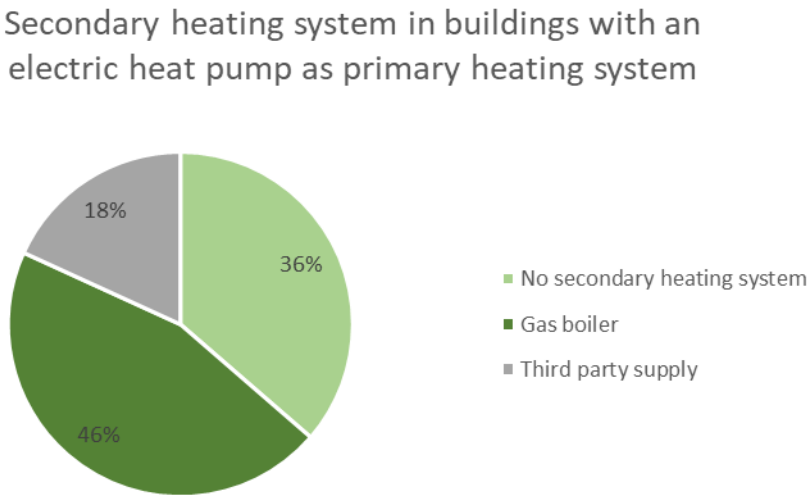


Figure A.14 Percentage of buildings per secondary heating generation system of buildings with an electric heat pump as main heating generation system.

On the EPA-U software is it possible to register 7 different heating distribution systems. (1) Local (Radiant heating), (2) Radiators/convectors, (3) Fan convector, (4) Floor heating, (5) Ceiling heating, (6) Wall heating, (7) Air heating. Radiators/convectors is the heating generation system in 97% of the analysed buildings. See Figure 15. It is interesting to note, that 16% of the buildings are equipped with an electric heat pump, what is often combined with other heating distribution systems. In this analysis it was not possible to verify if the information registered in the EPA-U files is correct.

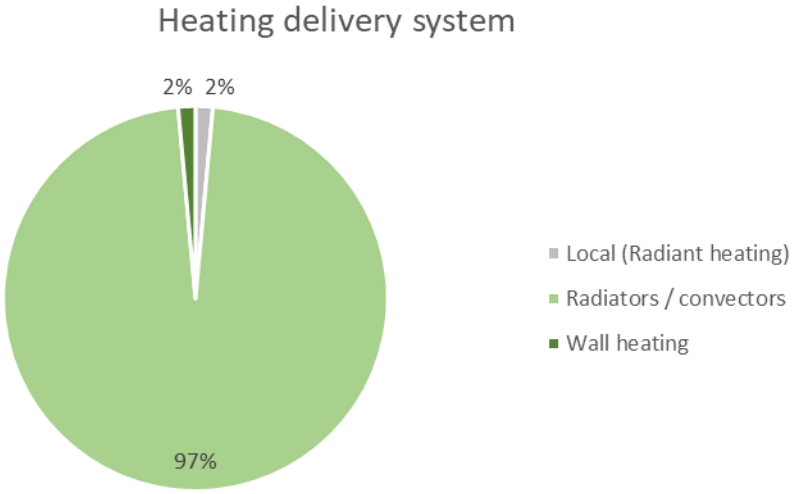


Figure A.15 Percentage of buildings per heating distribution system.

Cooling system

On the EPA-U software there are three possible cooling generation systems that can be registered: (1) absorption cooling, (2) compression cooling and (3) passive or free cooling. In the analysed sample none of the buildings has absorption cooling as the main cooling generation system. For 15% of the buildings there is no cooling system, or it has not been registered in the EPA-U files. See Figure A.16. All the buildings that have passive or free cooling as main cooling generation system have also a compression cooling system as secondary cooling system.

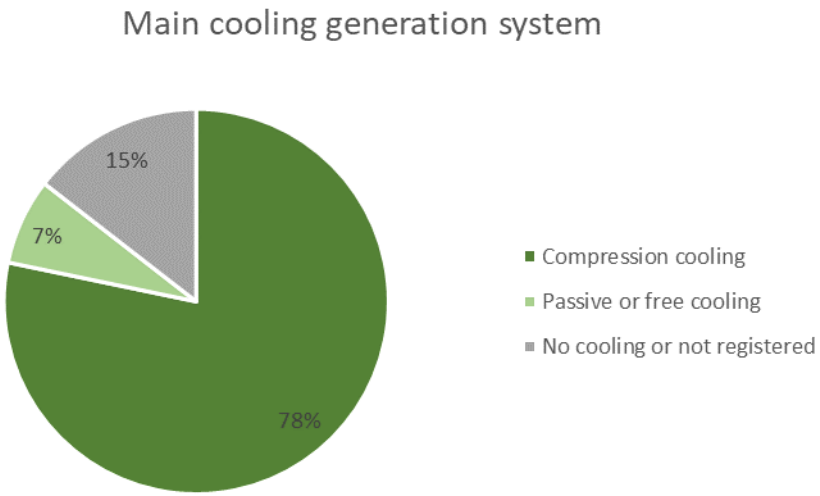


Figure A.16 Percentage of buildings per main cooling generation system.

The most common cooling delivery system among the buildings with cooling generation is the fancoil unit on the ceiling, 56%, the other most common cooling delivery system is air cooling. See Figure A.17. It is interesting to note that 3% of the buildings have cooling delivery via the floor, a system that appears outside the heating distribution system presented in Figure A.15.

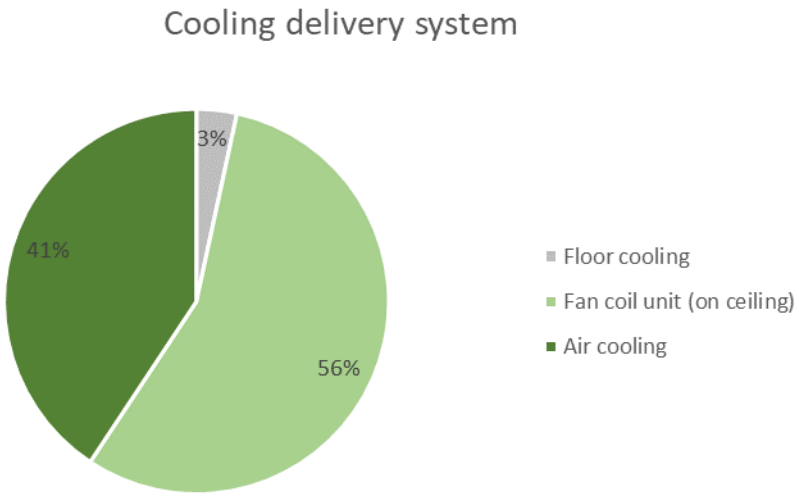


Figure A.17 Percentage of buildings per cooling delivery system.

Control system

On the EPA-U software there are ten possible control systems that can be registered: (1) control in main room (room thermostat), (2) automatic control per room, (3) automatic regulation per room + manual override (on/off), (4) automatic regulation per room + manual override (on/off) + adaptive control, (5) automatic control with unknown control, (6) supply flow temperature control, (7) retour flow temperature control, (8) Control certified according to EN215/EN15500, (9) Quality declaration (EN215/EN15500) and (10) unknown regulation. Most analysed buildings have automatic control per room, 97%. See Figure A.18.

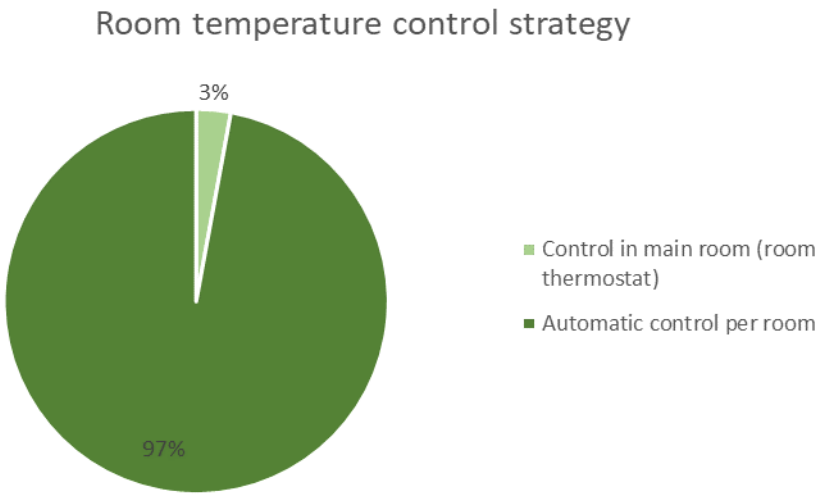


Figure A.18 Percentage of buildings per room temperature control strategy.

Solar energy systems

Only two of the 69 analysed building have solar energy generation systems. One building has solar collectors for production of hot tap water and another has solar panels. It is to be noted that we could not identify the date in which the information was registered in the EPA-U file, and therefore is it possible that solar energy generation systems have been installed after the registration of the building characteristics in the EPA-U.

10.6 Conclusions

From the analysis of 69 EPA-U files it was possible to describe the main energy characteristics of the of CGREA office buildings. The average CGREA office building was constructed in the period between 1976 and 2010, has a gross floor area of 9000 m^2 and has energy label C with an average building envelope Rc value of 1,76 [m2K/W]. The average building heating is generated by a gas boiler and distributed by radiators, the cooling is generated with a compression heat pump and distributed by fancoil units in the ceiling. The room temperature is automatically controlled at room level and the building do not have any type of solar energy system. Not all buildings have exactly the same characteristics as the average building, but there is not a great variety of building configurations, as from the analysis it appears that a few of the options for each system are the ones in place in each building.

This analysis has however a few limitations. First, only office buildings of CGREA were analysed and even only a part of the office buildings of the CGREA were analysed, 69 of 454. Second, the level of detail of the analysis was limited by the available information on the CGREA databases, the EPA-U files offered the best possible source of information, however these files contain only general descriptions of the building characteristics and building systems. Third, the quality of the information of the EPA-U files could not be verified. From the analysis it appears that not all the results seem to be logic. For example, no building has floor heating as heating distribution system, but a couple of buildings have floor cooling a cooling distribution system. Moreover, some characteristics of the buildings can have changed since the registration of the building characteristics in the EPA-U files, a plausible change is the implementation of solar energy systems.

In any case, the analysis of the main energy characteristics of office buildings offers a more accurate description of the average office building energy characteristics than the ones currently available. This information could be of special interest when developing computer models to predict the energy use and production of buildings. In this regard it also appears from this analysis, that when developing a building computer model aiming to control energy flexibility, the average office building is not the best-case study, as it offers limited energy flexibility possibilities. The building has limited energy storage capacity, reduced energy production capacity and limited predictive energy control capacity.