Effects of Sensor Selection and Additional State Measurements on HVAC Predictive Control

Fons ten Klooster

Master of Science Thesis

TU Delft Delft University of Technology

Delft Center for Systems and Control

Effects of Sensor Selection and Additional State Measurements on HVAC Predictive Control

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

Fons ten Klooster

April 8, 2022

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

Abstract

Research in passive Heating, Ventilation, and Air Conditioning (HVAC) systems has gained traction over the last few years. Although passive HVAC is not a new concept, advances in environment sensing, control methods, and hardware have made it a more viable method. Some difficulties still exist, such as optimal sensor placement and optimal control strategies. Sensor selection is an important aspect of HVAC design. The system can become difficult to control with incorrect placement of sensors, resulting in higher energy consumption, lower comfort levels, or poor air quality.

There are essentially three methods to determine the optimal sensor location: model-driven, data-driven, and simulation-driven. The model-driven methods use mathematical models to maximize the observability of the system but are mostly used for simplified simulated rooms. Data-driven methods often use clustering algorithms, or maximize metrics such as entropy or mutual information. These methods focus on estimating the indoor air temperature distribution. Simulation-driven methods use simulations to determine the airflow or temperature fields, often with CFD. These are used to find local hot spots or locations for fast detection of contaminants. No research was found that used sensor data of additional building components besides of the indoor air temperature.

In this work, the sensors are selected based on model prediction accuracy and the overall control performance to determine the effect of addition state measurements. A model is constructed to simulate the building, together with an MPC and an extended Kalman filter for state estimation. These are combined to run the optimization and determine the control performance. The sensor set average of each measured state is considered the true temperature. For all possible sensor combinations, the error of the combination average w.r.t. the true temperature is assumed Gaussian. The fitted Gaussian error distributions are then used as measurement noise in the model. The building and control response is simulated with the measurement error over multiple days. Two algorithms are implemented to find the optimal sensor set: a predictive method and greedy method. The results are compared to each other and both methods showed that the indoor air temperature measurements have the largest effect on performance. Measuring additional states only resulted in a small increase in performance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1-1 Motivation

Reduction of greenhouse gas emissions is becoming increasingly important. Climate objectives set by the Dutch government, in accordance with the Paris Agreement, aim to reduce the emission of greenhouse gases by 49% in 2030 compared to 1990 [53]. Households and the service sector combined account for 36% of the total energy consumption and 25% of the carbon emissions [53]. For households as well as the service sector about two-thirds of the energy use is committed to heating, cooling and ventilation [53]. Therefore significant progress towards the reduction of energy consumption can be made in this sector.

Reducing energy consumption can be accomplished by using more efficient active systems or implementing passive systems in buildings. The latter has been researched over the past three decades and is seeing a renewed interest [62]. Passive technologies are of interest, as they use solar and wind energy directly without converting it to electrical energy first. This keeps the electrical energy available for other uses and thereby complements the installation of renewables to reach the climate objectives. Furthermore, they use few to no mechanical parts that could break down or need maintenance. Using passive techniques for heating, cooling, and ventilation still has some major challenges regarding controllability. Supply and demand do not always correspond with each other due to uncertainties in weather conditions and load. Therefore, an integrated and intelligent design in building, control, and prediction methods is required.

For thermal comfort and control, temperature measurements are essential. The type of sensors used and their placement both play a role in determining the current room air temperature. Different methods for determining sensor positions have been used, but there is no standard methodology to approach this problem. Most research focuses on sensor placement in large open spaces, such as offices or greenhouses, to accurately determine the temperature distribution. There is, however, little research that investigates the potential of measuring additional building components, *e.g.* walls and floor. These measurements might provide a better temperature estimates and could result in better temperature predictions.

1-2 Research Goals and Contributions

In building Heating, Ventilation, and Air Conditioning (HVAC) systems design no standardized procedures for optimal sensor placement and control exist. Conventional procedures rely on intuition or expert knowledge when sensor locations are chosen. This is because buildings vary widely and have complex interactions with their environments. Situations where sensor selection or placement is optimized mainly focus on dealing with temperature distribution within a space, such as office buildings or greenhouses. These papers use sensors placed throughout the room all measuring the air temperature. No relevant research was found that investigated the potential of measuring additional building components for temperature control. These additional measurements could provide a better temperature estimate and prediction. The goal of this thesis can be summarized by the following research question:

What is the optimal sensor selection for building temperature control?

To study such effects, this work draws on the CONVERGE building, since it provides measurement of additional building components, *i.e.* walls, floor, and ceiling. One part of this question is of particular interest, and the question can be rephrased as:

Is there a benefit for control performance if additional building elements are measured?

The optimal in optimal sensor selection and optimal control, refers to minimizing energy consumption while maintaining thermal comfort. To answer these questions a combination of sensor data and modelling is used. The building model, as used by [6], is slightly modified to increase computation speed. To determine the effect of sensor selection on control performance, a state estimator and controller were implemented. After a literature search, for state estimation and control was chosen for an extended Kalman filter and a nonlinear Model Predictive Control (MPC). Two optimization methods for sensor selection are constructed in Section 3-4 and their results are compared in Chapter 4.

1-3 Thesis Outline

The remainder of this report is divided into 4 chapters. In Chapter 2 some preliminary information is given on HVAC design, comfort definition, first-principles of thermodynamics, and sensor placement strategies. This chapter is a summary of the conducted literature review. The following Chapter 3 sets up the optimization methodology used in the research. The model used to simulate the building is described first, followed by the state estimator and predictive model controller. Lastly, the designed algorithms are given, together with the method to simulate the effect of the different sensor locations. Chapter 4 contains the results and discussion of the optimization methods. A conclusion and recommendations for future work are given in Chapter 5.

Chapter 2

Background on Building Indoor Climate Control

This chapter contains the background information needed for building modeling and sensor placements. A general description of active and passive Heating, Ventilation, and Air Conditioning (HVAC) systems, and comfort definitions is given in Section 2-1 first, followed by a section about the building considered for the case study in this thesis. Some of the fundamentals in thermal modelling are given in Section 2-3. This is followed by a section on sensor placement methods in Section 2-4. Finally, a summary of the main methods used for control is given in Section 2-5.

2-1 Heating, Ventilation and Air Conditioning

The task of HVAC systems is to regulate the indoor building environment. Besides temperature, other variables controlled can be humidity, light, CO_2, etc. Generally, HVAC systems consist of both active and passive components.

2-1-1 Active Systems

Active systems, such as fans, pumps, chillers, boilers, etc., are still the most widely used methods for HVAC. A major advantage is the controllability of these systems, as they can be easily designed to fit the requirements of the building. This comes at the cost of requiring energy, *e.g.* electricity or natural gas, and having more mechanical parts. An overview of a general active HVAC system can be seen in Figure 2-1. In this figure, VFD stands for variable frequency drive, which means the speed of the motor is adjustable and with it the flow rate. The left-hand side of the system circulates heated or cooled air through the building. Fresh air is mixed with air from the building to maintain air quality and comfort while conserving energy. The right-hand side consists of a water boiler and chiller controlling the temperature of the air conditioning and ventilation system.

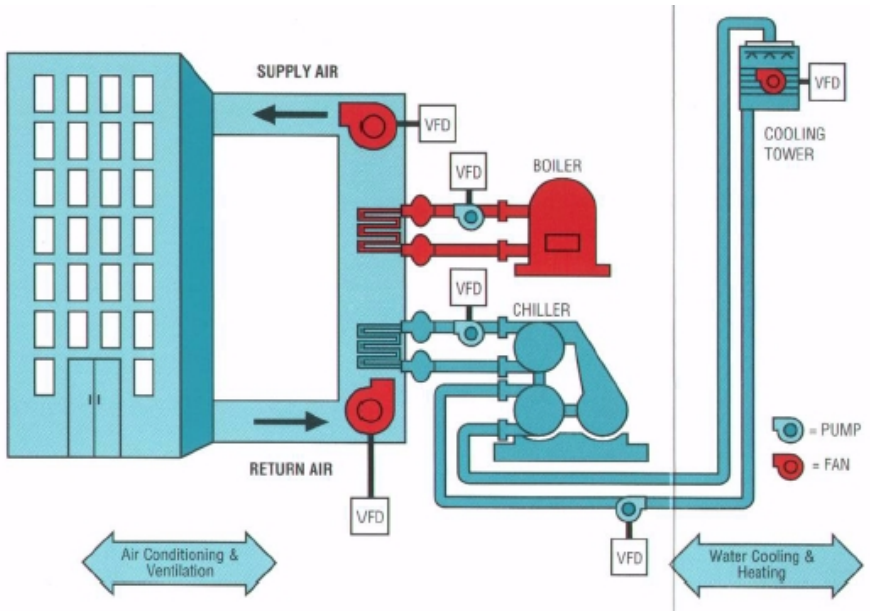


Figure 2-1: Schematic of a classical HVAC system from [42] with a Variable Frequency Drive (VFD).

HVAC units can come in constant air volume or variable air volume variants, either being single zone or multi-zone. Usually, a constant air volume system is implemented in residential buildings, assuming a relatively uniform distribution of heat [23]. Constant air volume provides a constant airflow with a variable temperature. Contrary to variable air volume, where the airflow is variable and the temperature is generally constant. In a study by Lu and Warsinger [31], the retrofitting of buildings with multi-zone variable air volume is investigated. These systems can have beneficial energy savings over single constant air volume, especially for cooling. Besides the energy savings, using variable air volume also results in more precise temperature control.

2-1-2 Passive Systems

Passive systems, contrary to active systems, do not require energy for heating or cooling. Some actuators might still be required though, *e.g.* to open and close windows or to adjust solar shading. One of the main sources of heat that can be used passively is solar irradiation. Building orientation and placement of windows can be used, to block or absorb incoming sunlight when needed. Automated shading systems can be installed to control the incoming amount. Providing shade for cooling or letting in sunlight to heat the walls and floor of a room. Ventilation can also be used for cooling when the air temperature outside is lower than the temperature inside. Air currents can be induced by differences in air temperature or pressure, using the sun or wind respectively. Passive ventilation systems using sunlight come in multiple forms, using the same principle of natural convection. The main idea is to heat the air and use the buoyancy to drive ventilation. The air can be heated in a solar chimney or transpired solar collectors [63], the latter being essentially a double wall. A simple schematic is shown in Figure 2-2. To extend the operation time of a solar chimney materials can be added to store the heat energy and release it over a longer period, for instance overnight.

These materials, usually phase changing materials, can be beneficial in heating as well as cooling through buffering. Solar radiation can be used to preheat the incoming air in winter, and ventilation can be used for cooling in summer. Pre-ventilation can not only reduce the cost of cooling in summer, but also improve air quality and aid with heating in winter [62]. The effectiveness of these principles depends on the climate in which they are positioned. Especially for temperate climates such as Germany and the UK, passive systems can add to better air quality and improved energy efficiency [62]. Letan et al. [30] showed the feasibility of ventilation in summer and heating in winter for a five-story building. This was done by simulations and testing on a lab-scale model. Their model showed good performance even with low solar radiation intensity. Besides solar energy, wind can also be used for ventilation. Using a chimney effect or simply opening windows to create a pressure difference and drive air currents in or out of the building.

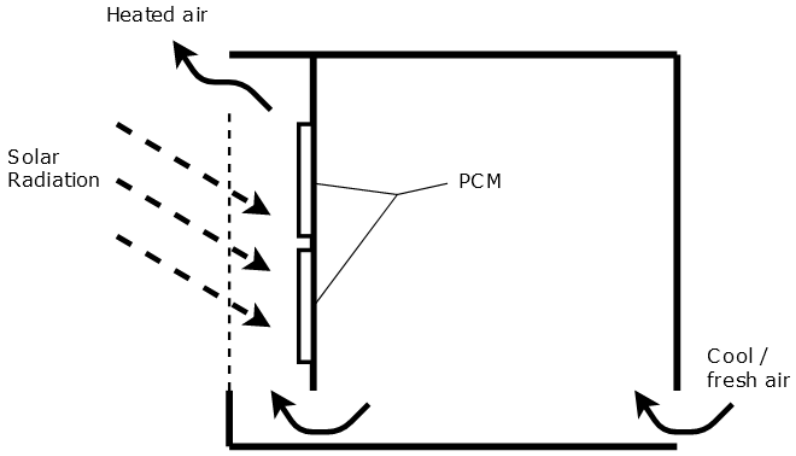


Figure 2-2: Schematic of a simple passive ventilation system with Phase Changing Materials (PCM) using solar energy.

2-1-3 Comfort Levels and Air Quality

When it comes to determining comfort levels and indoor air quality, multiple standards exist. Standards such as ANSI/ASHRAE Standard 62.1 for ventilation rates, and thermal comfort ASHRAE 55 and EN ISO 7730. While the ASHRAE is more popular in the United States, the ISO is more internationally orientated [64]. These are originally more focused on, but not limited to, traditional active HVAC systems. Natural ventilated systems are included in the European standards EN 15251 [15] and EN 13799 [56][1] for thermal comfort and air quality, respectively, and hold slightly different recommendations. This is due to different levels of expectations for active, compared to passive ventilation [54]. In all of the standards, multiple categories of buildings are defined, each having different recommendations.

Comfort levels are a subjective matter. To objectively quantify comfort, two models were introduced: Predicted Mean Vote (PMV) and Predicted Percentage Dissatisfied (PPD). Both of these are used in ASHRAE as well as the ISO standard. The PMV ranges from -3 (cold) to $+3$ (hot) and depends on factors like metabolic rate, clothing insulation, ambient temperature,

[1] EN 13799 has had some small revisions since EN 16798 [40].

air velocity, mean radiant temperature, and relative humidity [64]. PMV being zero means that 95% of the occupants are satisfied [14]. EN 15251 uses a running weighted mean of ambient air temperature as a basis for indoor temperature recommendations. The mean is generally taken over the preceding week. Different weights are applied to the preceding days and for different seasons. The reason is the different clothing and expectations of indoor temperature in winter, compared to summer conditions. This results in an adaptive thermal comfort level. A more in-depth evaluation of the standard can be found in [48]. Most of the papers reviewed assume some fixed temperature range, without calculating the PMV and PPD values. In [33], a temperature comfort level is set between 20 °C and 22 °C during the day for an unspecified location, while [19] uses a set-point range between 20 °C and 26 °C for the climate of Portugal. For a winter case, [62] uses a set-point between 18 °C and 24 °C with relative humidity between 30% and 70%. All of the parameters and comfort ranges depend on a multitude of factors, such as climate, HVAC system, activity level, and clothing.

Taking all these parameters into account would make the overall problem more complicated and they are not necessarily important to meet the research goals of this thesis. To simplify the overall problem, discomfort is measured by the degree hours criteria method [47] with fixed comfort bounds. The time spent outside of the specified comfort bounds is multiplied by which the temperature exceeds that bound. These summed together form the discomfort index D_{ind} in degrees hour [°Ch]. The allowed discomfort index is set to $D_{\text{ind}} \leq 0.5$ °Ch per day, similar to [33]. This means that the temperature is allowed 0.5 °C outside of the bounds for one hour, or 1 °C for half a hour, when the building is occupied.

2-2 The CONVERGE Building

The CONVERGE building, located in The Green Village in Delft[2], is a test-bed for passive HVAC systems. This is a cooperation project of the Delft University of Technology (TU Delft) and multiple companies from the building sector. The project has set multiple goals to test the feasibility of implementing passive HVAC systems and provide a basis for thesis and academic research. The main focus points are research and design of a thermal tower, sensor placement, developing smart shading systems and, testing the use of heat and cold buffering materials. Keeping the most important aspect in mind, making it realizable and attractive to use in practice.

The building is equipped with an array of sensors and actuators. All of these sensors are fitted to test different components and new concepts in passive HVAC design. A picture of the CONVERGE building is shown in Figure 2-3. Besides passive methods, some active systems are incorporated in this building as well. The CONVERGE building uses a combination of the two, which include a thermal tower with phase changing materials, solar shades, a heat pump, and solar energy-driven fans. The active systems are installed to aid the passive methods when needed, but the main idea is to avoid the use of the active systems as much as possible.

The approximately 80 sensors in the building consist of multiple types, including CO_2, sound, movement, and NTC temperature sensors. Of these, 35 are temperature sensors positioned on different building components to give a precise measurement of the indoor environment,

[2]More information on The Green Village can be found on: https://www.thegreenvillage.org/

Figure 2-3: Picture of the CONVERGE building with the thermal tower.

including air, floor, interior walls, and ceiling temperatures. The building also has its own local weather station, to give more accurate external environment measurements compared to a nearby station. This local station provides data on humidity, air temperature, wind speed and light conditions. Solar global horizontal irradiance is also are measured using a pyranometer.

2-3 Thermal Transport and Solar Radiation

Before the system modelling is discussed in Section 3-1, it is helpful to understand the thermodynamic concepts involved. The model is based on first-principles. For each of the building elements and room air temperature, an energy balance is considered. This is based on the first law of thermodynamics, where the energy gain of an element without work done is equal to the sum of heat transfers, given by:

$$\mathrm{mc_p}\frac{\mathrm{d}T}{\mathrm{dt}} = \sum \dot{q} \tag{2-1}$$

where m is the components mass, $\mathrm{c_p}$ is the specific heat, $\mathrm{d}T$ is the temperature change, dt the time step and $\dot{q}$ the energy flow. The left-hand side of this equation shows heat storage. This modelling is also called the lumped capacitance approach, where the change in internal energy of a component is equal to the net energy flow through its boundaries, and the temperature is considered spatially uniform. The energy flows $\dot{q}$ consist of conductive, convection, radiation,

and internally generated heat. For more detailed information on heat transfer and solar radiation is referred to [24] and [13] respectively.

Conduction

Conduction is the heat flow through a material, *e.g.* insulation layer. The rate at which heat transfer occurs can be described by Fourier's law of thermal conduction. In the building model, this type of energy transfer occurs between roof and ceiling, and between the ground and the concrete floor. The 1-D heat flow through a insulation layer between two building components over a surface area, A, is given by:

$$\dot{q}_{\text{cond}} = -\kappa \text{A} \frac{\text{d}T}{\text{d}x} \tag{2-2}$$

Here, κ is the thermal conductivity and $\text{d}T/\text{d}x$ is the temperature gradient in the direction of the heat flow. The temperature distribution within the material, or insulation, becomes linear when steady-state conditions are assumed. If this condition is assumed, (2-2) can be approximated by

$$\dot{q}_{\text{cond}} \approx \frac{\kappa}{\text{L}} \text{A}(T_2 - T_1) \tag{2-3}$$

where T_1 and T_2 are the temperatures on each side of the insulation layer, and L is the thickness of the insulation. Although the components are generally not under steady-state conditions, this approximation describes the heat transfer reasonably well when L is small in relation to the surface area A.

Convection

Heat transfer between a solid and a fluid or gas can occur through different mechanisms. Common method is to consider that the surface is in contact with a bulk fluid, as the temperature of the fluid near the surface is generally unknown. A temperature some distance from the surface is used. The energy transfer through convection from a surface can be described by Newton's law of cooling as:

$$\dot{q}_{\text{conv}} = h\text{A}(T_{\text{s}} - T_\infty) \tag{2-4}$$

where h is the convective heat transfer coefficient, T_s the temperature of the surface and T_∞ is the temperature of the fluid some distance away from the surface. The convective heat transfer coefficient relates to the Nusselt number by:

$$h = \frac{Nu\text{k}}{\text{L}_{\text{c}}} \tag{2-5}$$

where Nu is the Nusselt number, k is the thermal conductivity of the fluid, and L_{c} the characteristic length. Different correlations have been established for the Nusselt number under certain conditions. These conditions depend for instance on fluid velocity and thermophysical properties.

Radiation

Energy between materials is exchanged continuously through radiation. Different materials have different efficiencies in absorbing and emitting radiation. The radiation between two surfaces can be described by :

$$\dot{q}_{\text{rad}} = \frac{\sigma(T_2^4 - T_1^4)}{\frac{1-\epsilon_1}{\epsilon_1 \text{A}_1} + \frac{1}{\text{A}_1 \text{F}_{12}} + \frac{1-\epsilon_2}{\epsilon_2 \text{A}_2}} \tag{2-6}$$

where ϵ_1 and ϵ_2 are the surface emissivity of the two layers, σ is the Stefan-Boltzmann constant, and F_{12} is the view factor. Emissivity is the fraction of energy emitted compared the that of a black body at the same temperature. For the building model, two cases are especially useful: 1) where the view factor F_{12} is unity and A_1 is equal to A_2; 2) one small surface is enclosed by a larger surface, such that A_1/A_2 approaches zero. The former is the case for the small insulation cavity between the glass layers in the facade, and the latter is an approximation for the exterior building components with the sky. Applying the conditions for the surfaces between the glass layers on (2-6) results in the following equation:

$$\dot{q}_{\text{rad,cav}} = \frac{\sigma \text{A}(T_2^4 - T_1^4)}{\frac{1}{\epsilon_1} + \frac{1}{\epsilon_2} + 1} \tag{2-7}$$

In the other special case where the radiation energy transfer between the exterior and surroundings is considered, (2-6) can be written as:

$$\dot{q}_{\text{rad,sky}} = \epsilon \sigma \text{A}(T_{\text{ext}}^4 - T_{\text{sky}}^4) \tag{2-8}$$

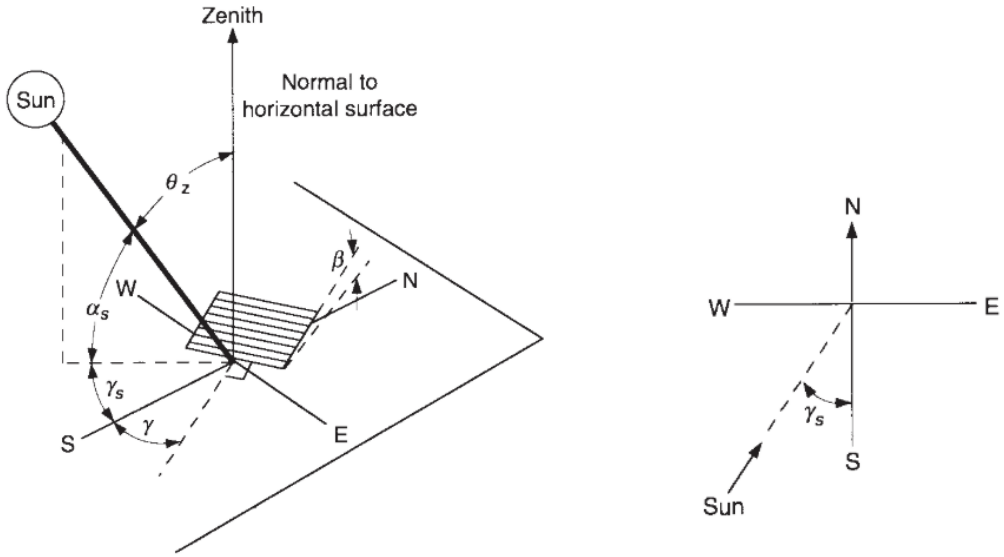
where T_{ext} an exterior building component temperature and T_{sky} the reference sky temperature. This equation applies to are the wall exterior and the roof. The equations above describe the radiative interactions between building components and between building components and surroundings. The input energy from the sun is described by:

$$\dot{q}_{\text{sol}} = \text{IA}\tau^{\text{n}}\alpha \tag{2-9}$$

where I is the incident solar irradiation, τ is the transmittance, n is the number of layers the radiation has to pass through before reaching the surface, and α is the absorptance. The transmittance, or transmissivity, is the fraction of radiant energy passing through a material. Absorptance is the fraction of the energy absorbed by a material. The incident solar irradiation I is the sum of three different components, I_b, I_d, I_g, which are the beam, diffuse and ground reflected irradiation, respectively. The beam irradiation is the radiation coming directly from the sun, whereas the diffuse irradiation are the incoming rays that have been scattered by the atmosphere. The beam radiation is calculated by the following relation:

$$I_\text{b} = I_{\text{bn}} \cos\theta \tag{2-10}$$

where I_{bn} is the solar irradiation normal to a surface and the angle of incidence θ. The angle of incidence is depicted in Figure 2-4. The calculation of the angle is as follows:



(a) Angles relative to a surface **(b)** Overview of the sun azimuth angle γ_s.

Figure 2-4: Diagram of solar angles relative to a surface [25].

$$\begin{aligned}\cos\theta = &\sin\phi(\sin\delta\cos\beta + \cos\delta\cos\gamma\cos\omega\sin\beta)\\ &+ \cos\phi(\cos\delta\cos\omega\cos\beta - \sin\delta\cos\gamma\sin\beta)\\ &+ \cos\delta\sin\gamma\sin\omega\sin\beta\end{aligned} \tag{2-11}$$

where ϕ is the latitude, δ is the solar declination, β is the slope of the surface, γ the surface azimuth angle and ω the hour angle. An overview of these angles can be seen in Figure 2-4.

The diffuse irradiation depends on a large amount of factors, *e.g.* temperature, humidity, cloud levels, etc. To approximate the diffuse irradiation multiple models have been constructed. A widely used model that shows good performance is the Perez model. The Perez model is a sky diffuse model that combines the horizon, isotropic and circumsolar components[50]. The coefficients are derived from empirical data and show a good approximation in most conditions:

$$I_\mathrm{d} = I_\mathrm{dh}\left[(1-F_1)\left(\frac{1+\cos\beta}{2}\right) + F_1\left(\frac{a}{b}\right) + F_2\sin\beta\right] \tag{2-12}$$

with diffuse horizontal irradiance I_dh, β is the surface inclination from the horizontal. The coefficients a and b to account for the angle of incidence. F_1 and F_2 are coefficients for circumsolar and horizon brightness, derived from empirical data. This leaves the ground reflected irradiation, which can be described by:

$$I_\mathrm{g} = I_\mathrm{h}\rho_\mathrm{g}\left(\frac{1-\cos\beta}{2}\right) \tag{2-13}$$

where I_h is the horizontal surface irradiation and ρ_g the albedo. The albedo is the reflectiveness of a surface. This value depends on the material and conditions, such as rainwater or

snow.

2-4 Sensor Placement Methods

Sensor types and placement are important aspects of HVAC design. With incorrect placement the system can become difficult to control, resulting in higher energy consumption, lower comfort levels, or poor air quality. Since cost is usually important in the design process, there is a trade-off between cost and the number of sensors and the amount of information available for control. There are essentially three methods to determine the optimal sensor location: model-driven, data-driven, and a combination of these two which can be seen as a third method. The model-driven method, also called analytical or principle-based, relies on knowledge of the system in advance to determine the optimal sensor location. A mathematical model of the building is constructed and the sensor position is determined by using a control theory-based approach. In the data-driven approach, multiple sensors are installed and a selection of sensors is made afterwards. The initial placement is usually done by placing sensors at a fixed distance, heuristic manner, or relying on the expertise of the engineer. The third also uses these mathematical models, but uses them to obtain data from simulations. Usually, a combination is used.

Model-driven Methods

This section uses a control theory or dynamical system-based approach in determining the optimal sensor positioning. Using a model to carry out simulations could also be seen as model-driven, but this method is discussed separately in the next section. The dynamical system approach has the advantage over the purely computational approach to provide some insight into the solution. The idea here is to determine the optimal placement by optimizing certain system criteria, such as criteria for controllability and observability Gramian.

The use of linear advection partial differential equations is considered by Vaidya et al. [57], to describe the dispersion of scalar quantities such as temperature and contaminants. The equations are used to construct an explicit controllability and observability Gramian. Finite approximations for sensor placement do not work in this case, because of the transport properties. A theoretical analysis for optimal sensor placement is constructed based on the observability Gramian. [57] mention that the simplifications and assumptions made should be reduced and relaxed to make the results applicable in practice. Fontanini et al. [20] describe a dynamic system approach in determining sensor locations in a given environment, with the objective of tracing contaminants in the air. The sensors must be placed to ensure speed and full coverage for contaminants detection in for example an aircraft. An air velocity vector field obtained through Computational Fluid Dynamics (CFD) simulations is transformed into a Markov matrix, describing the diffusion of contaminants. This makes the calculations of large sets more efficient compared to solving partial differential equations [20]. The sensor placement is optimized using the observability Gramian constructed from the Markov matrix. An advantage is that the sensor accuracy and location restrictions are also incorporated in the Gramian. The contaminants could refer to CO_2, although [20] focuses on extreme events, where detection speed important. Fang et al. [16] approach was to maximize the trace of the controllability and observability Gramian. This is done as maximizing the trace generally leads to high-rank matrices. The observability Gramian is relatable to the amount of

output energy. Both sensor and actuator locations are determined using the Gramian. The optimization problem was computationally simplified by converting it from a maximization problem to a sorting problem [16]. Some practical limitations were incorporated in the optimization, such as unsuitable locations and sensor clustering avoidance. Simulations of a 2D data center showed improved state estimation over random deployment. It should be noted that the location of the sensors is only optimized with respect to the Gramian metric. It is argued that this gives a good first indication, nonetheless, and its closed-form solution and low computational cost make it an attractive method.

Data-driven Methods

When enough sensor data is available, an optimal sensor configuration can be obtained by analysis such as classification methods, selecting those data points containing the most essential information about the system and removing redundancy.

Clustering algorithms were used by Yoganathan et al. [66] to determine the optimal sensor placement. The study is carried out for an office floor containing cubicles. Sensors used are for temperature, relative humidity, and light. The initial placement is done in a heuristic manner, placing a sensor at each cubicle so that each sensor covers approximately 10m^2. Their goal is to remove 80% of the sensors, basing this percentage on the Pareto Principle. This is, however, somewhat of an arbitrary statement, as it would depend on the initial number of sensors. The two partition-based clustering algorithms tested, X-means and Clustering for Large Applications (CLARA), produce similar results. Partition-based algorithms cluster the data based on a distance measure. X-means is an extension of K-means clustering, where the number of clusters is automatically determined based on the Bayesian Information Criterion [49]. Both X-means and CLARA are supervised clustering algorithms.

Fu et al. [21] used clustering techniques to obtain a simpler model for HVAC controller design. The same principles apply for optimizing sensor location and number, *i.e.* selecting the least amount of sensors that provide a sufficient amount of information. In this case, only the temperature sensors were optimized with clustering techniques. Other sensors include CO_2, humidity, and airflow at the variable air volume output. Sensors chosen to best represent a cluster were chosen in a top-down approach. This has the advantage of not making any assumptions of the sensors following a certain Gaussian process [21]. Two clustering methods are used: stratified near-mean selection and stratified random selection, which are both spectral based algorithms. They have the advantage over partition-based clustering methods, such as K-means, to often have better performance and can be solved efficiently by standard linear algebra methods [60]. These are tested with correlation-based and Euclidean distance criteria. Correlation-based clustering showed better results compared to Euclidean distance. The best results were obtained with stratified near-mean selection and stratified random selection outperforming the Gaussian process and simple random selection clustering techniques. The latter two functioned as a comparison if sensor locations are chosen at random. Yun and Kim [67] tested multiple clustering methods based on similarity measures to reduce the number of sensors. The measurements are temperature, humidity, and illumination. The Support Vector Machine algorithm showed the best classifier results in terms of mean and standard deviation, compared to Bayes net, Decision tree, Decision table, Instance-based learning, Multi-layer perceptron, and Naïve Bayes. Support Vector Machine learning performed best, followed by decision tree. The similarity measures that are compared are Euclidean distance,

complexity invariance distance, and dynamic time warping distance. However, the main goal of this paper was to reduce the number of sensors, while still accurately measuring the indoor environment and classify comfort.

Maasoumy et al. [34] determined the best sensor configuration by taking the set of sensors with the least Root Mean Square Error (RMSE), essentially taking the configuration with the least temperature deviation with respect to the mean temperature value of all the sensors. The was carried out for a relatively small-sized setup, namely for two connected 20-foot containers. The goal was to accurately measure the average temperature for a given number of sensors. Error-based and entropy-based sensor placement methods were conducted by Lee et al [29] for a greenhouse. Besides the mean temperature, for instance, areas with large deviations also need to be monitored. The selected sensor locations needed to be able to detect rapid changes in local temperatures, caused by for example external factors. In this study, only the air temperature was considered. The error-based method, using the RMSE and absolute percentage error, performed best in estimating the overall climate conditions inside the greenhouse. The information entropy-based method was added to find locations with large temperature variations.

Spatial phenomena can often be modelled as Gaussian processes [28]. This is the approach Krause et al. [28] used to determine the optimal sensor placement for an office floor. To define what a good location is, multiple criteria have been mentioned. Krause et al. used the information-theoretic mutual information criterion of Caselton and Zidek, and showed it outperforms the entropy criteria. The biggest advantage of information-theory mutual information criterion over entropy is that the former also considers the uncertainty of the unsensed location. Entropy tends to place the sensors at the edges of the areas of interest where the uncertainty is highest [28]. They created an efficient algorithm with a polynomial-time approximation, where no other Gaussian Process method or criteria provides such guarantee [28]. The placement is also robust against node failures and model uncertainties [28] grounded with theoretical approximation guarantees. The approach is tested with two real-world data sets.

By doing an extensive literature review and questionnaires, Mahyuddin et al. [35] found that most professionals and researchers preferred a location for CO_2 sensors in the middle of the room at a height between 1.0 m and 1.2 m. This also corresponds to the recommended height by multiple industry standards [35]. Furthermore, the sample location should be some distance away from the occupants for a correct reading. A 2.0 m distance was mentioned by some to be sufficient [35]. From the databases investigated and results from questionnaires a p-value analysis was carried out to look at the number of CO_2 sensors used and the total area. For a room with less than 100 m^2, one sensor was used, and for a room between 200 and 300 m^2, three sensors. Factors with the most significant effect on CO_2 monitoring were the number of occupants, type of ventilation, and the effect of windows and doors.

Simulation-driven Methods

Simulation-driven approach is essentially a combination of the previous two. Here the data is generated through simulations. The advantage compared to the data-driven approach is that there is no need to install various sensors first. Different locations can be tested through simulation. The downside is the need for a numerical model, which has a trade-off between

computation time and accuracy. The construction of this model itself also requires expert knowledge.

Chen and Wen [7] tested multiple airflow models to identify suitable applications. The goal of the sensor placement is harmful contaminant detection and ensuring the safety of the occupants. The different model types tested were CFD, zonal, and multizone. The CFD served as a benchmark as it was the most detailed of the methods. The environments for which tested are a small office, large hall, and an office suite. The multizone approach seemed to work just as well as the more complex CFD model.,though the impact of different characteristics, such as obstacles and layout, on the model agreements have not been fully investigated. A Genetic Algorithm (GA) was used for the sensor locations, minimizing the detection time and occupant exposure time of contaminants. McGibney et al. [41] used evolutionary strategies to find the optimal sensors locations using a CFD model of a living room. As not every point was calculated and/or measured, an inverse weighted distance interpolation was used during the optimization process.

The CFD and building energy simulation strategies can be combined to provide optimal control for HVAC systems, which is done by Du et al. [12]. An small office space was simulated and measurement data is used for validation. The multiple sensor positions have been simulated to determine the effects on energy consumption and thermal comfort. The conclusion was that conventional sensor locations, near the inlet and outlet of the HVAC, are not always optimal. Feng et al. [18] used CFD simulations for a greenhouse to determine the optimal sensor locations.The best locations were the ones where no large variations occur, *i.e.* avoiding the hot spots. It was argued that the sensor locations could be chosen intuitively by examining the simulation results.

2-5 Control Methods

Control methods used for passive HVAC systems, as well as for active HVAC systems are reviewed in this section. Active systems are by far used the most and have been in development for a long time. Most papers consider a hybrid HVAC system, where the passive ventilation is to assist the active systems. This again, is mainly for controllability purposes as the effectiveness of passive systems is highly dependent on weather conditions. Belic et al. [4] have done a review on control methods for active systems. They found that generally the approach can be considered to be in one of three categories: improving classical control methods, using predictive methods, and the use of intelligent control methods.

Classical Control Methods

Classical control methods do not rely on models to make predictions of state evolutions. PID control with feedback and feedforward are also considered, together with ON/OFF types of control based on rules or thresholds. These are the most basic types of control, while also used the most in practice [4]. These methods are made more advanced over the years, for example by combining them with more advanced optimization techniques or expanding on PID control with MIMO.

Fanti et al. [17] use a particle swarm optimization algorithm for ON/OFF control to open windows for natural ventilation. The study is carried out for a residential building in the

Mediterranean. The temperature set-point is determined by the adaptive comfort levels for passive ventilated buildings. A particle swarm optimization is an evolutionary optimization algorithm, suitable for nonlinear problems. Compared to GA, particle swarm is robust and generally requires fewer function evaluations to obtain better or similar results [17, 51]. The windows are opened by *if-then* rules, where the comparison is made between the case where the *if* set-points are fixed and with the conditions optimized with particle swarm. These conditions are based on indoor and outdoor temperatures. Simulations showed a significant reduction in the number of discomfort hours with automated windows. Furthermore showed a reduction in hours of thermal discomfort for overheating compared to fixed rules. For undercooling the results for both cases were similar. Goyal et al. [22] compared a feedback controller with Model Predictive Control (MPC). Their findings showed that MPC had only a little benefit over a simpler rule-based feedback control at the cost of increased complexity. Although it should be noted that this was done for a small-sized room. Maasoumy et al. [33] compared a PID controller with ON/OFF and MPC. The ON/OFF controller, in this case, opens and closes the air ventilation valve if certain thresholds are met. Their results showed vast improvements in energy consumption with the ON/OFF and MPC control methods.

Classical control methods have the advantage of being relatively simple and computation inexpensive. Even though MIMO provides better theoretical results, decoupling it into multiple SISO systems can be useful in practice [4]. Making it easier to tune the different subsystems. These methods are widely implemented in practice, but most of these systems are active HVAC systems. These are less dependent on external disturbances, as the actuators can be designed with sufficient control capabilities.

Predictive Control Methods

For passive HVAC systems, MPC is a promising method. To utilize changing weather conditions and occupancy for HVAC, predictions have to be made about the near-future conditions. Control actions that are possible now are not necessarily possible with passive control in the future and vice versa. Deterministic as well as stochastic MPC types are used for control, either linear or nonlinear. While MPC is used in most cases, stochastic MPC showed increased performance when uncertainties in weather forecasts and occupancy are considered.

Mantovani et al. [37] applied an MPC strategy for temperature control to a shopping mall. Here the problem lies with vertical stratification in the large open spaces, and the coupled convection between the zones [37]. It also addresses some extensions with regard to renewables and electricity pricing, as these can be fairly easily included in the objective function. A nonlinear mathematical model of a real shopping mall is used to carry out the simulations. Included are external inputs and disturbances, such as solar radiation, external temperature, and occupancy. The MPC is combined with a Kalman filter as a state observer. Modifications with regard to the heating system are mentioned to improve the control and energy efficiency of the shopping mall. The MPC strategy was not able to mitigate the stratification problem, although their strategy showed promising results with regard to energy savings. Maasoumy et al. [33] compare a PID, ON/OFF, and MPC strategy for temperature and airflow in a university library located in California. Compared to the original PID control strategy, significant reductions in total airflow rate and energy consumption were achieved with ON/OFF control and MPC. The MPC showed only a relatively small improvement compared to the simpler ON/OFF control, although the MPC was able to reduce the maximum airflow. An MPC was

also designed in [34] in the case of accurate sensors, and a robust MPC in the case of less accurate sensors. For robust MPC the worst-case scenario is considered while satisfying the input and temperature constraints. These control algorithms are combined with an unscented Kalman filter and an extended Kalman filter state observer. With fewer sensors used and a lower sensing accuracy, the best control approach was unscented Kalman filter-robust MPC. This resulted in the best combined performance considering thermal discomfort and energy cost. Vidrih et al. [59] Generalized Model-based Predictive Weather Control for night-time ventilation of an office room. A prediction of the next 24 hours is made, to utilize the ambient air at night for cooling. Generalized Model-based Predictive Weather Control is in the form of control matrices. A control matrix summarizes the dynamics and control. These matrices are used to determine if and to what extent it is efficient to use night-time ventilation in conjunction with mechanical ventilation. The algorithm was not implemented into the control system but functioned as an aid to the monitors. The developed control matrix was rather static but reasonably accurate and easy to implement [59].

Oldewurtel et al. [46] investigate the potential of stochastic MPC with weather forecasts for energy-efficient building climate control. Multiple cases from other papers are considered, not only looking at temperature control, but also illuminance and CO_2 levels. To simplify the calculations disturbances are assumed Gaussian and occupancy is considered known. This makes it possible to solve a sequence of linearized subproblems. stochastic MPC was found to be superior compared to deterministic MPC in terms of energy savings and thermal comfort. Furthermore, it can be made such that easy tuning is possible by a single parameter. [46] mention that performance is highly dependent on the model and weather prediction used. Ma et al. [32] used a stochastic MPC approach to minimize energy consumption and maintain thermal comfort levels. The library building is constructed of multiple thermal zones, each controlled by an active variable air volume HVAC system. To make the problem computationally viable a feedback linearization scheme is proposed. Guarantees are provided on the probability of comfort violations. The uncertainties in load and weather conditions are modelled by finitely supported probability density functions, which are continuously updated as new data is obtained. Uncertainties in load and occupancy are non-Gaussian in practice, and exact solutions of stochastic MPC with these disturbances are generally computational intractable [32]. To deal with the non-Gaussian disturbances, multiple scenarios are considered where the chance constraints are made deterministic. It has the advantage not depending on the linearized system dynamics [32], as in [46]. A total of five control methods, Perfect (P), Certainty equivalent (C), Explicit (E) Stochastic, Gaussian (G) stochastic MPC and Sample-based (S) stochastic MPC, are all compared to the current implemented control method. The PMPC has a perfect weather forecast, which of course is not possible in practice. This serves as a comparison of the different uncertainty prediction methods. The CMPC uses the mean predicted disturbance as a nominal value. This disturbance is essentially an estimated average of scenarios or average likeliness. The ESMPC uses discrete convolution to approximate the chance constraints. Explicit MPC uses offline optimization with lookup tables. These methods are however less suited for large-scale problems [10]. The SSMPC uses a sample-based method for this approximation, using multiple realizations. To reduce conservatism some data points are omitted. If many realizations are used this would basically result in a worst-case scenario in terms of disturbance. Lastly, the GSMPC, which does assume a Gaussian distribution of the uncertainties. The ESMPC and SSMPC performed almost as well as the MPC with perfect knowledge of the disturbances.

Sun et al. [55] combined active and passive HVAC, shading, and lighting for a two-room building in Beijing. To overcome computational difficulties with regard to optimization, GA or decomposition methods such as Lagrangian relaxation are used. [55] have chosen for the latter. Lagrangian relaxation is useful when the problem is separable. In [55] some modifications had to be made to deal with the nonlinear couplings between rooms. To ease computation costs two heuristic strategies are developed based on the stochastic dynamic programming and rollout technique. The new control method reduced the energy cost by approximately 10% compared to the existing strategy with good computation costs. The Lagrangian relaxation dynamic programming drastically reduced computation time compared to dynamic programming and the rollout technique was able to reduce it further at a low cost of increased energy consumption.

To deal with the increased complexity and computation time of nonlinear MPC, different variations such as distributed MPC, explicit MPC, hybrid MPC, and stochastic MPC have been tested with different levels of success [10]. The difficulty still lies with the complexity when all the factors are considered. A simple model for a medium-sized building can consist of hundreds of states and control inputs [32]. While disturbances are often modelled as Gaussians [46], Ma et al [32] used non-Gaussian disturbances by transforming chance constraints into deterministic ones using discrete convolution integrals and sample-based techniques. Lagrangian relaxation methods and heuristics can be used to further simplify computations [55].

Intelligent Control Methods

Obtaining an accurate model can be a cumbersome task, due to time lag, uncertainties, and complex interactions of indoor climate variables. This is where intelligent control methods can be useful. Methods such as fuzzy logic, neural networks, or other machine learning methods. These are able to deal with the nonlinear dynamics and MIMO systems.

Mirinejad et al. [43] did a review on intelligent control methods in HVAC systems for thermal comfort improvement and energy consumption reduction. For fuzzy logic controllers, the main advantage is that there is no need for a complex mathematical model for controller design [43]. Fuzzy logic is suitable for complex nonlinear systems. A downside is the need for expert knowledge and usually a trial-and-error design process which can be time-consuming. Methods, such as neural networks or GA, can be used to tune or learn either rule-base or membership functions, or both. When a neural network is used this is considered a neuro-fuzzy system [43]. In the case of GA, there is genetic learning and genetic tuning. With genetic learning, the membership functions or rule-base, or both at the same time, can be automatically learned without prior knowledge [9]. For genetic tuning, the general structure is already defined, and the parameters are optimized automatically. Tuning is less difficult than learning the whole controller, but for both, there are no guarantees of finding the optimum control due to complexity. Dounis et al. [11] proposed a fuzzy logic control structure to control HVAC, shading, and lighting. The building used passive solar energy to decrease energy consumption. The controller used is hierarchical, with a master outputting new set-points based on energy consumption prediction and comfort levels. The slave is comprised of two negotiation machines and a decision-making unit. One negotiation machine looks at the illuminance difference, and the other at temperature and comfort levels. Tuning the membership functions can be a difficult task. To optimize the parameters, tuning of the

membership functions is done with a GA.

Manjarres et al. [36] used machine learning algorithms for an active ON/OFF HVAC system. Random forest regression techniques are used to make 24-hour predictions for an office building in Spain. Using black-box models of buildings has some advantages over principle-based or grey-box models. When data is available, they have higher generalization capabilities and do not require the time-consuming process of developing physical models [36]. The ON/OFF scheduling showed good performance. Xu et al. [65] used neural networks with Lagrangian relaxation and stochastic dynamic programming to find near-optimal solutions. It was effective in dealing with uncertainties and reducing peak demands of the HVAC system.

Fuzzy logic does not need a model of the building. Tuning can be done through expert knowledge or learning methods. The mentioned control methods are by no means all there is. Each method has its variations and different methods can be combined to supplement another, such as fuzzy logic combined with MPC in [45]. Black-box methods such as machine learning techniques have the advantage of being adaptable to system changes. Although this is only the case when enough data is available. Furthermore, it removes the need for an intensive modelling process involved with principle-based methods. Over the past few years, there is an increased interest in using neural networks for energy prediction models over other machine learning methods, such as support vector machines, random forests, and decision trees. This is mainly due to advantages in reliability, interchangeability with other building software, and nonlinear input-output handling capabilities [44].

2-6 Summary

Passive systems for HVAC in buildings are of interest because of their low energy requirements and mechanical simplicity compared to active systems. Studies have shown that passive systems can reduce energy consumption in warm as well as in cold climates, through ventilation and heating, and improve air quality. Though the systems perform best in tempered climates, where the surrounding temperature is relatively close to the desired indoor temperature. A downside of these systems is the controllability, as their effectiveness is highly dependent on local weather conditions. Nevertheless, passive systems are a good supplement for active systems in HVAC. For the indoor environment, multiple standards can be used for determining comfort levels. Some models have been introduced to give an objective measure for comfort, taking into account metabolic rate, clothing insulation, temperature, air velocity, mean radiant temperature, and relative humidity. Although these models are used the most, different methods of calculating recommended comfort levels exist. The acceptable values depend on the standard used, building category, and climate conditions.

Models based on first-principles have been able to accurately describe the most important system dynamics. Some fundamentals of thermal transport effects of solar irradiation have been given, that form the basis of the building model. The model should also account for enough of the system dynamics. The basic interactions between building components and other external factors needed in building modelling have been described.

For sensor placement, data-driven methods such as hidden Markov models and neural networks are widely used and show good results. Model-based sensor placement has more potential than geometric-based sensor placement. The main aspect of model-based sensor place-

ment is the objective function, deciding which metric is used for optimization, for example, entropy, maximal mutual information, Gramian, or RMSE. With model-driven sensor placement most popular is observability Gramian. This is an insightful metric to optimize. When CFD is used sensor locations are often chosen based on intuition. Choosing the position which best resembles the average temperature in the enclosure. Trial-and-error is also common, essentially randomly generating the sensor placements with simulation-driven methods and choosing the configuration with the least amount of error with respect to obtained data. This method can be effective but is less ideal. Common practice is to place sensors near HVAC exhaust and in-take, but this only works well for small-sized rooms. When the optimality criteria such as entropy or maximal mutual information are used, the goal is to obtain as much information as possible about the temperature distribution within the room. Most papers focus on either sensor placement or control, but rarely consider the effect of the sensor positioning on the control performance. No other research was found that included additional state measurements. An overview of the methods used the most can be found in Table 2-6.

Classical control methods have the advantage of being relatively simple and computation inexpensive. These methods are widely implemented in practice, but most of these are used for active HVAC systems. These are less dependent on external disturbances, as the actuators can be designed with sufficient control capabilities. When active systems are supposed to complement the passive system instead of the other way around, predictive or intelligent methods are better suited. These control methods show improved performance compared to the classical methods, at the cost of some increase in complexity. Major downsides for intelligent systems are the need for training with large amounts of data and the complex choice for the model structure. A summary of the different control methods can be found in Table 2-6.

Table 2-1: Summary sensor placement methods most used as described in Section 2-4.

Method	Approach	Notes	Ref.
Model-driven	Maximizing support or trace of observability Gramian	Simulated empty rooms; assumptions not applicable in practice.	[57], [20], [16]
Data-driven	Clustering: K-means, X-means, and clustering for large Applications, criteria: minimal information loss.	Maximize spatial temperature information, assumes measuring the same variable.	[66]
	Minimize error, RMSE	Simple, used to estimate temperature distribution.	[34], [29]
	GP-modelling, maximizing mutual information, maximizing entropy.	Advanced, no model needed, accurately describing temperature distribution.	[28]
Simulation-driven	Simulate control with CFD coupled with Building Energy Simulation software.	One sensor is used at a time, CFD high computation time for a small room.	[12]
	CFD and zonal-model to minimize detection time contaminants released.	CFD is not much better than a zonal model, goal: fast detection of contaminants.	[7]

Table 2-2: Compact overview of HVAC control methods described in Section 2-5.

Control	Complexity	Advantage	Disadvantage	Ref.
PID	Simple	Easy tuneable, intuitive, low computation cost.	Not optimal, less efficient dealing with uncertainty.	[33]
Rule-based	Simple	No model needed	Generally less accurate and efficient.	[22], [17], [33], [34].
MPC	Moderate	Predictive control	Uncertainty can affect control.	[22], [33], [34], [37], [46], [55].
Stochastic MPC	Complex	Deal with uncertainty and noise	Limited use in practice.	[46], [32].
robust MPC	Complex	Robust properties.	Mostly theoretical.	[34]
Machine-learning	Complex	Generalizable	Black-box, no guarantees, overfitting risks	[36], [43], [65].
Fuzzy	Moderate	Intuitive	Expert knowledge or trial-and-error.	[43], [11].

Chapter 3

Building Model and Controller Design

This chapter contains the general setup for this research. To simulate the building temperature, a transient nonlinear model is used. The model is based on first-principles and later fitted to sensor data from the CONVERGE building. This is described in Section 3-1. For state estimation, an Extended Kalman Filter is chosen as described in Section 3-2. This is followed description of the predictive control system in Section 3-3. The methods used for optimal sensor selection are given in Section 3-4.

3-1 Nonlinear Model

The thermal behavior of the building is simulated through transient modelling in MATLAB, which is similar to the model used in [6] and has shown good accuracy compared to sensor data from the CONVERGE building. The temperature response of the building elements is based on the thermal energy transfer between the elements and the surrounding. With the building elements is referred to the walls, floor, ceiling, and roof. For simplicity, some assumptions are made. The model uses a lumped capacitance for each element. This implies that the spatial temperature gradients are neglected, *i.e.* temperatures of the building elements are considered homogeneous. Often in building simulations, spatial temperature gradients are disregarded as it drastically simplifies the model, while still providing accurate temperature approximations. This is furthermore motivated by the building design. All the walls are made entirely out of glass and the layout is essentially a one-room rectangle box with a flat roof. A schematic of the building design is shown in Figure 3-1. The air is considered well-mixed with a homogeneous temperature. While a temperature stratification will be present, due to buoyancy forces and sunlight hitting only certain parts of the floor, the air is supplied at low speed over a large area through the tiles. This reduces the local hot or cold spots found near the exhaust of traditional Heating, Ventilation, and Air Conditioning (HVAC) systems, where the air is supplied through a small opening. Temperature stratification is not that significant considering the relatively low ceiling. Another assumption is that the radiation energy transfer between the indoor building elements is negligible, *e.g.* between walls, floor, and ceiling. The building elements and indoor air will have roughly the same temperature

and convective heat transfer is dominant. The main dimensions of the building are shown in Table 3-1. For a complete overview of all the parameters used is referred to Table A.

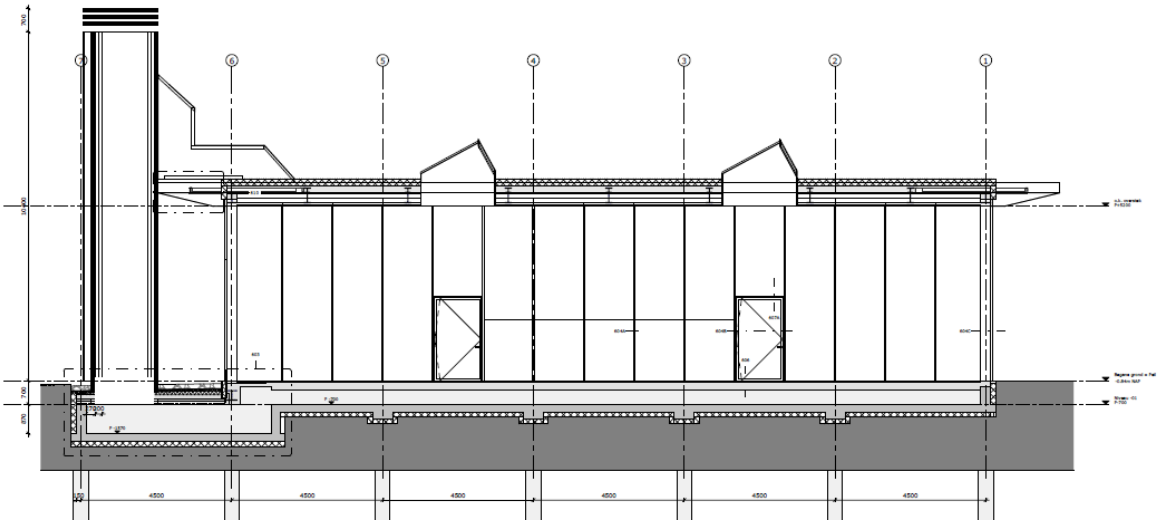


Figure 3-1: Schematic of the CONVERGE building, with the climate tower on the left-hand side. Schematic from the CONVERGE project in The Green Village.

Table 3-1: Main dimensions of the building. Thickness refers to one glass panel.

Component	State	Length [m]	Height [m]	Thickness [m]	Orientation (γ) [°]
South wall	T_1-T_3	13.5	5.2	0.01	0
East wall	T_4-T_6	22.5	5.2	0.01	-90
West wall	T_7-T_9	22.5	5.2	0.01	90
North wall	T_{10}-T_{12}	13.5	5.2	0.01	180

For each building element and the indoor air temperature, an energy balance is considered. This is based on the first law of thermodynamics, where the energy gain of an element is equal to the sum of heat flows, given by

$$\mathrm{m}_i \mathrm{c}_{\mathrm{p},i} \frac{\mathrm{d}T_i}{\mathrm{dt}} = \sum_j \dot{q}_{j-i} + \dot{q}_{\mathrm{int}} \tag{3-1}$$

where $\mathrm{m_i}$ is the mass of component "i", $\mathrm{c_p}$ is the specific heat, T the temperature, $\dot{q}_{j-i}$ is the energy flow from "j"to "i", and $\dot{q}_{\mathrm{int}}$ is the internal heat generation. For a complete overview of all the building parameters and material properties is referred to Table A-1.

State Equations

The building is divided into 17 states, each with an energy balance equation as (3-1). Each wall is modelled as three separate components. The floor, which consists of a base floor and raised tiles, is also modelled as two distinct components. In total, the 17 states are made up of 12 glass facade layers, an external roof, an internal ceiling, a ground floor, raised floor, and the air. A description of all the states can be found in Table 3-2. The walls of the

building are constructed from triple-glazed panels. Layered windows mainly lose heat through radiation. Low emissivity glass, or low-ϵ, is manufactured by coating a glass surface with a low emissivity material. This reduces the radiative heat loss through the windows or facade. To further improve the insulation, each cavity between the layers is filled with argon gas. Compared to air, argon has low thermal conductivity. The conductive heat transfer between the layers is neglected, leaving only the radiative and convective heat transfer. Equation (3-1) is rewritten to obtain the temperature gradient for each layer. This results in the equations for a wall as shown by:

Table 3-2: Description of the 17 state temperatures.

State	Variable	Description
x_1	T_1	South facade external temperature
x_2	T_2	South facade intermediate temperature
x_3	T_3	South facade interior temperature
x_4	T_4	East facade external temperature
x_5	T_5	East facade intermediate temperature
x_6	T_6	East facade interior temperature
x_7	T_7	West facade external temperature
x_8	T_8	West facade intermediate temperature
x_9	T_9	West facade interior temperature
x_{10}	T_{10}	North facade external temperature
x_{11}	T_{11}	North facade intermediate temperature
x_{12}	T_{12}	North facade interior temperature
x_{13}	T_{13}	Roof temperature
x_{14}	T_{14}	Ceiling temperature
x_{15}	T_{15}	Room air temperature
x_{16}	T_{16}	Raised floor temperature
x_{17}	T_{17}	Ground concrete floor temperature

$$\mathrm{d}T_1 = (\dot{q}_{\mathrm{sol}-1} - \dot{q}_{\mathrm{rad,1-sky}} - \dot{q}_{\mathrm{rad},1-2} - \dot{q}_{\mathrm{conv},1-\mathrm{amb}} - \dot{q}_{\mathrm{conv},1-2})\frac{\mathrm{dt}}{\mathrm{m}_1\mathrm{c}_{\mathrm{p,gl}}} \tag{3-2a}$$

$$\mathrm{d}T_2 = (\dot{q}_{\mathrm{sol}-2} + \dot{q}_{\mathrm{rad},1-2} - \dot{q}_{\mathrm{rad},2-3} + \dot{q}_{\mathrm{conv},1-2} - \dot{q}_{\mathrm{conv},2-3})\frac{\mathrm{dt}}{\mathrm{m}_2\mathrm{c}_{\mathrm{p,gl}}} \tag{3-2b}$$

$$\mathrm{d}T_3 = (\dot{q}_{\mathrm{sol}-3} + \dot{q}_{\mathrm{rad},2-3} + \dot{q}_{\mathrm{conv},2-3} - \dot{q}_{\mathrm{conv},3-15})\frac{\mathrm{dt}}{\mathrm{m}_3\mathrm{c}_{\mathrm{p,gl}}} \tag{3-2c}$$

where 1,2,3 refers to the exterior, intermediate, and interior glass layer respectively. $\dot{q}_{\mathrm{i-j}}$ is the heat flow from component "i"to "j". The masses m_1, m_2, m_3 refer to the masses, and $\mathrm{c}_{\mathrm{p,gl}}$ is the specific heat of the glass. The layers for the east-facing wall, $\{T_4, T_5, T_6\}$, west-facing wall, $\{T_7, T_8, T_9\}$, and north-facing wall, $\{T_{10}, T_{11}, T_{12}\}$, are modeled similar to (3-2). The solar radiation, $\dot{q}_{\mathrm{sol-i}}$ with $i = 1, 2, 3$, absorbed by each layer is given by:

$$\dot{q}_{\mathrm{sol}-1} = \mathrm{A}(I_{\mathrm{b}}\alpha_{\mathrm{b}} + (I_{\mathrm{g}} + I_{\mathrm{d}})\alpha_{\mathrm{d}})\beta_{\mathrm{blinds}} \tag{3-3a}$$

$$\dot{q}_{\mathrm{sol}-2} = \mathrm{A}(I_{\mathrm{b}}\alpha_{\mathrm{b}}\tau_{\mathrm{b}} + (I_{\mathrm{g}} + I_{\mathrm{d}})\alpha_{\mathrm{d}}\tau_{\mathrm{d}})\beta_{\mathrm{blinds}} \tag{3-3b}$$

$$\dot{q}_{\mathrm{sol}-3} = \mathrm{A}(I_{\mathrm{b}}\alpha_{\mathrm{b}}{\tau_{\mathrm{b}}}^2 + (I_{\mathrm{g}} + I_{\mathrm{d}})\alpha_{\mathrm{d}}{\tau_{\mathrm{d}}}^2)\beta_{\mathrm{blinds}} \tag{3-3c}$$

which are derived from (2-9), with the beam irradiation separated from the ground reflected and diffused irradiation. Here A is the surface area of the considered building facade. τ is the transmittance and α is the absorptance, with the subscripts "b"and "d"referring to the relative beam and diffusive components. β_{blinds} is the control input, determining the fraction of irradiance let through the blinds, ranging from zero to one. All blinds are controlled by one input. This reduces the number of optimization variables for control, without affecting the results. Independent control of each facade is useful when light and glare comfort is considered inside the building. This is not taken into account in this research. The energy flow from the exterior layer to the ambient air, and the flow from the interior layer to the indoor air through convection are provided by (2-4), as:

$$\dot{q}_{\text{conv},1-\text{amb}} = h_{1-\text{amb}}\text{A}(T_1 - T_{\text{amb}}) \tag{3-4}$$

$$\dot{q}_{\text{conv},3-15} = h_{3-15}\text{A}(T_3 - T_{15}) \tag{3-5}$$

where T_{amb} is the ambient air temperature. The energy flows inside the cavities between the glass layers, and from the exterior layer to the surroundings through radiation are provided by (2-7) and (2-8), respectively. These result in the following equations for the radiation inside the cavities:

$$\dot{q}_{\text{rad},1-2} = \frac{\sigma\text{A}(T_1^4 - T_2^4)}{\frac{1}{\epsilon} + \frac{1}{\epsilon_{\text{low}}} + 1} \tag{3-6}$$

$$\dot{q}_{\text{rad},2-3} = \frac{\sigma\text{A}(T_2^4 - T_3^4)}{\frac{1}{\epsilon} + \frac{1}{\epsilon_{\text{low}}} + 1} \tag{3-7}$$

and from the exterior layer to the surroundings as:

$$\dot{q}_{\text{rad},1-\text{sky}} = \epsilon\sigma\text{A}(T_1^4 - T_{\text{sky}}^4) \tag{3-8}$$

where σ is the Stefan-Boltzmann constant, ϵ the emissivity of the glass, ϵ_{low} the emissivity of the glass with the low-ϵ coating, and T_{sky} is the reference sky temperature. Different correlations are used for the convection coefficients of the exterior, $\text{h}_{1-\text{amb}}$, in the cavities, h_{1-2}, h_{2-3} and for the interior, h_{3-15}. A summary of the correlations used in this section is shown in Table 3-3. The heat transfer coefficient depends on the Nusselt correlation, defined by (2-5). The Nusselt correlation of the exterior surface, $\text{h}_{1-\text{amb}}$, is provided by the Zhukauskas's correlation for windy conditions, given by:

$$\text{Nu} = 0.037Re^{0.8}Pr_{amb}^{0.3334} \tag{3-9}$$

In the vertical cavities between the glass layers, h_{1-2}, h_{2-3} the MacGregor correlation for natural convection (3-10) is used.

$$\mathrm{Nu} = 0.42Ra^{0.25}Pr^{0.012}\left(\frac{\mathrm{L_C}}{\mathrm{S}}\right)^{-0.3} \tag{3-10}$$

where $\mathrm{L_C}$ and S are the dimensions along the gravity direction and perpendicular to gravity for the vertical surface respectively. For the interior, $\mathrm{h_{3-air}}$, with no wind the Churchill & Chu's correlation is used, given by:

$$\mathrm{Nu} = \begin{cases} 0.68 + \left(\frac{0.67Ra^{1/4}}{[1+(0.492/Pr)^{9/16}]^{4/9}}\right), & \text{if } Ra < 10^9 \\ \left(0.825 + \frac{0.387Ra^{1/6}}{[1+(0.492/Pr)^{9/16}]^{8/27}}\right)^2, & \text{if } Ra \geq 10^9 \end{cases} \tag{3-11}$$

More information on the Zhukauskas, MacGregor and Churchill & Chu's correlations can be found in [8] and [24]. The roof and ceiling temperatures, T_{13} and T_{14} respectively, are given by:

$$\mathrm{d}T_{13} = (\dot{q}_{\mathrm{sol-13}} - \dot{q}_{\mathrm{rad,13-sky}} - \dot{q}_{\mathrm{conv,13-amb}} - \dot{q}_{\mathrm{cond,13-14}})\frac{\mathrm{dt}}{\mathrm{m_{roof}c_{p,roof}}} \tag{3-12}$$

$$\mathrm{d}T_{14} = (\dot{q}_{\mathrm{cond,13-14}} - \dot{q}_{\mathrm{conv,14-15}})\frac{\mathrm{dt}}{\mathrm{m_{ceil}c_{p,ceil}}} \tag{3-13}$$

with the masses $\mathrm{m_{roof}}$, $\mathrm{m_{ceil}}$ and specific heat $\mathrm{c_{p,roof}}$, $\mathrm{c_{p,ceil}}$ relate to the roof and ceiling respectively. The absorbed solar radiation by the roof, $\dot{q}_{\mathrm{sol-13}}$, and conductive heat transfer from the roof to the ceiling, $\dot{q}_{\mathrm{cond,13-14}}$, are modelled with (2-9) and (2-3), respectively. These are given by:

$$\dot{q}_{\mathrm{sol-13}} = I\mathrm{A}\alpha \tag{3-14}$$

$$\dot{q}_{\mathrm{cond,13-14}} = \frac{\kappa}{\mathrm{L}}\mathrm{A}(T_{13} - T_{14}) \tag{3-15}$$

where A is the surface area of the roof and ceiling, α is the absorptivity of the roof, and κ and L are the thermal conductivity and thickness of the insulation layer, respectively. Instead of a vertical surface, as the case for the wall, now the convection correlations for a horizontal surface are applied. The Nusselt correlation for the roof in windy conditions, Zhukauskas's correlations, is the same as (3-9). This correlation is not dependent on the surface slope. For the ceiling, a no wind condition is considered. Here, McAdams's correlations for horizontal surfaces is applied, given by:

$$\mathrm{Nu} = \begin{cases} 0.54Ra^{0.25}, & \text{if } T_{14} < T_{15} \wedge Ra \leq 10^7 \\ 0.15Ra^{0.3334}, & \text{if } T_{14} < T_{15} \wedge Ra > 10^7 \\ 0.27Ra^{0.25}, & \text{if } T_{14} > T_{15} \end{cases} \tag{3-16}$$

The room air temperature, T_{15}, has besides the convective heat transfers from the walls,

floor, and ceiling, also an energy input from the occupants and HVAC system. The latter is modelled as a direct input in the energy balance equation for the air, without considering airflow speed or regenerative heat exchange methods. This results in the following equation for the air temperature:

$$\begin{aligned} \mathrm{d}T_{15} =&(\dot{q}_{\mathrm{hvac}} + \dot{q}_{\mathrm{occ}} + \dot{q}_{\mathrm{conv},3-15} + \dot{q}_{\mathrm{conv},6-15} + \dot{q}_{\mathrm{conv},9-15} \\ &+ \dot{q}_{\mathrm{conv},12-15} + \dot{q}_{\mathrm{conv},14-15} - \dot{q}_{\mathrm{conv},15-16})\frac{\mathrm{dt}}{\mathrm{m_{air}c_{p,air}}} \end{aligned} \tag{3-17}$$

where $\dot{q}_{\mathrm{hvac}}$ is the energy supplied by the HVAC system and $\dot{q}_{\mathrm{occ}} = \mathrm{P_{occ}}n_{\mathrm{occ}}$ is the internal heat gain generated by the occupants. $\mathrm{P_{occ}}$ is the average thermal power generated per person and n_{occ} is the total number of occupants. Occupants are one of the main sources of internal heat generation. The average heat generation per person, $\mathrm{P_{occ}}$, is set to 120W. This is a typical value used for office buildings [2].

The remaining two states are the raised floor temperature, T_{16}, and the ground floor temperature, T_{17}. The raised floor exchanges energy by convection from the air, conduction from the ground floor and is heated by incoming sunlight. The ground floor exchanges energy through conduction with the ground and the raised floor. The temperature differences for the raised floor and ground floor are given by:

$$\mathrm{d}T_{16} = (\dot{q}_{\mathrm{sol}-16} + \dot{q}_{\mathrm{conv},15-16} - \dot{q}_{\mathrm{cond},16-17})\frac{\mathrm{dt}}{\mathrm{m_{rf}c_{p,rf}}} \tag{3-18a}$$

$$\mathrm{d}T_{17} = (\dot{q}_{\mathrm{cond},16-17} - \dot{q}_{\mathrm{cond},17-\mathrm{grd}})\frac{\mathrm{dt}}{\mathrm{m_{gf}c_{p,gf}}} \tag{3-18b}$$

where the masses $\mathrm{m_{rf}}$, $\mathrm{m_{gf}}$ and specific heat $\mathrm{c_{p,rf}}$, $\mathrm{c_{p,gf}}$ relate to the raised floor and ground floor respectively. $\dot{q}_{\mathrm{cond},17-\mathrm{grd}}$ is the heat transfer from the ground floor the to ground underneath. For the convection correlation between the air and the raised floor, $\dot{q}_{\mathrm{conv},15-16}$, the McAdams from (3-16) is used. The incoming sunlight hitting the floor depends on the position of the sun, the geometry of the building, and state of the blinds. The relation is given by:

$$\dot{q}_{\mathrm{sol}-16} = IA_u\tau^3\alpha\beta_{\mathrm{blinds}} \tag{3-19}$$

where A_u is the unshaded floor area. The light has to pass through three layers of glass, hence the τ^3 factor. All these temperature increments $\mathrm{d}T$ are only valid for a small dt, as $\mathrm{d}T/\mathrm{dt}$ in (3-1) is the temperature gradient at a specific moment. In the simulations, a time step dt of 300 s is used. This is the interval at which the sensors log the temperature data and the model has shown a good fit to the sensor data. The overall nonlinear model is described by:

$$\begin{aligned} x(k+1) &= f(x(k), u(k), d(k)) \\ y(k) &= \mathrm{C}x(k) \end{aligned} \tag{3-20}$$

with $x \in \mathbb{R}^{17}$, $u \in \mathbb{R}^2$ and $d \in \mathbb{R}^5$. The dimensions of y depend on the choice of sensors. The

states, inputs and disturbances are defined as:

$$x(k) = \begin{bmatrix} T_1(k) \\ T_2(k) \\ \vdots \\ T_{17}(k) \end{bmatrix}, \quad u(k) = \begin{bmatrix} \dot{q}_{\text{hvac}}(k) \\ \beta_{\text{blinds}}(k) \end{bmatrix}, \quad d(k) = \begin{bmatrix} I(k) \\ n_{\text{occ}}(k) \\ T_{\text{amb}}(k) \\ T_{\text{sky}}(k) \\ T_{\text{grd}}(k) \end{bmatrix} \tag{3-21}$$

solar irradiation, I, relies on a combination of measurement data and model approximation as described in Section 2-3. The other disturbances, with the exception of T_{grd}, can be measured. The ground temperature is therefore approximated by a model, from [6], as follows:

$$T_{\text{grd}} = 5.3 \cos\left(\frac{2\pi(day - 43.9)}{365}\right) + 15 \tag{3-22}$$

where *day* is the day of the year. The parameters have been fitted to ground temperature data from the Royal Netherlands Meteorological Institute [26].

Table 3-3: Convection correlations. H is the height, A is the surface area, and P is its perimeter.

Correlation	Condition	Applies to	Characteristic length L_C	Equation
Zhukauskas	Forced; vertical and horizontal surface	External components with surrounding: walls and roof.	vert. H horz. A/P	(3-9)
MacGregor	Natural; vertical cavity	Insulation layer glass facades.	H	(3-10)
Churchill & Chu	Natural; vertical surface	Interior walls to air.	H	(3-11)
McAdams	Natural; horizontal surface	Ceiling and floor to air.	A/P	(3-16)

Computation Time

Initially, the *fsolve* function was used to solve the energy balance in (3-1) while the CoolProp library [5] is used to determine all the Rayleigh, Reynolds, and Prandtl numbers for the convection correlations. Removing the *fsolve* function reduces the computation time by approximately 10%. Using average values for air properties and removing CoolProp library further reduces the computation time by 97.6% while producing nearly identical temperature results. All of the parameters and values used can be found in Table A-1. There is one downside to removing the *fsolve* function. The model starts to deviate from the sensor data if the time interval dt is set to large. The discretization only works well for small time steps between 300 and 900 seconds. Therefore, the model is simulated with a time step dt of 300 seconds. This is the same interval at which the sensors log the temperature data in the CONVERGE building. The predictive model uses a time step of 600 seconds with control inputs applied over a 3600 seconds interval. This reduces the computation time significantly as the function is evaluated numerous times during the optimization.

Optimization and Validation of Model Parameters

While the obtained model could be used to simulate the building temperatures, estimating the properties of the building components is generally difficult in practice. In order to provide a more accurate simulation of the building, some of the parameters of the model have been fine-tuned to fit measurement data. These parameters include, α, τ, ϵ, ϵ_{low}, and κ. The parameters are fitted by minimizing the Root Mean Square Error (RMSE) between the measured states and sensor data. The *fmincon* function in MATLAB is used with the *sqp* algorithm for optimization. To prevent overfitting, tight bounds of $\pm 10\%$ relative to the initial parameter have been applied over the A complete overview of the parameters and other properties of the building can be found in Table A. To analyze the fit, the RMSE and Variance Accounted For (VAF) are used as performance metrics. The VAF is a widely used measure to verify a model's accuracy and is calculated as [58]:

$$\text{VAF} = \max\left(0, \left(1 - \frac{\frac{1}{N}\sum_{k=1}^{N} ||y(k) - \hat{y}(k)||_2^2}{\frac{1}{N}\sum_{k=1}^{N} ||y(k)||_2^2}\right) \cdot 100\%\right) \tag{3-23}$$

where N is the number of data points, and $y(k)$ and $\hat{y}(k)$ are the measured and estimated output respectively. The higher the VAF, to closer the model dynamics follow the data. A VAF around 90% and above is generally considered a good fit. The RMSE is also calculated, as the model can still have a 100% fit, but also contain a bias. The ten days of available sensor data, collected from April 2nd till April 12th in 2020, has been split into two sets: an optimization data set containing the first five days, and a validation data set with the remaining five days. The results of the 5 day validation period can be seen in Figure 3-2.

Table 3-4: Optimization and validation values of the building model.

	Optimization		Validation	
State	VAF	RMSE	VAF	RMSE
3	92	1.6	90	1.6
6	94	1.3	89	1.3
9	96	1	95	1.1
12	97	0.6	96	0.7
14	95	0.6	94	0.7
15	94	0.6	93	0.6
16	84	1	88	0.8
17	92	0.2	97	0.2

The model shows a good fit to the validation data, although not a 100%. There are a number of reasons why there will always be some misfit. As mentioned, obtaining accurate parameters of the building materials is very difficult in practice. The actual values can vary from component to component, *e.g.* one insulation layer may be slightly different from another. The assumptions and simplifications made during the modelling process, discussed at the start of this section, also play an important part. The uniform temperature assumption of the building components for instance does not hold in reality. Another possible reason is the model used to determine the irradiation disturbances. It is difficult to calculate, and even measure, the actual direct beam and diffuse irradiation. The goal was to obtain a model that is able to describe the general dynamics of the building temperatures. In this regard, the model is able to simulate the average measured temperatures within reasonable accuracy.

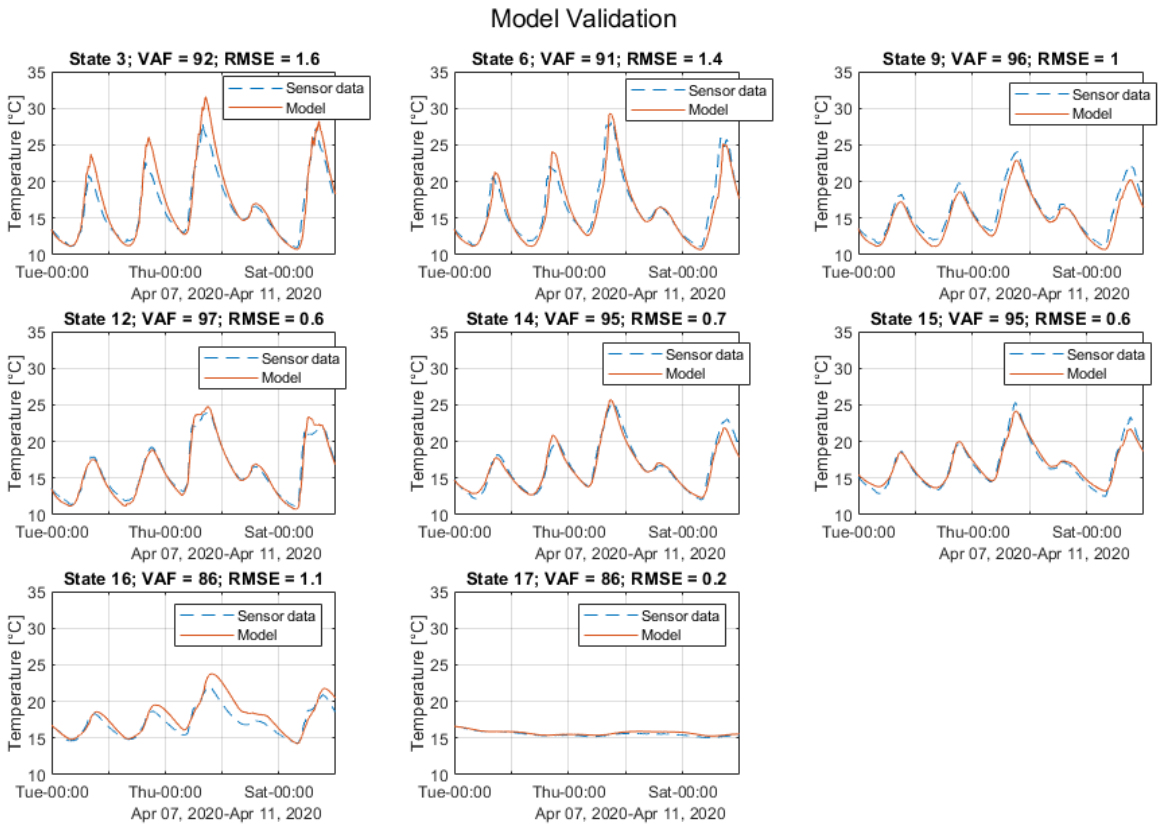


Figure 3-2: Model validation over 5 days in April. The solid red line is the model output and the dashed blue line is the averaged sensor data.

3-2 Extended Kalman Filter

The extended Kalman filter is a widely used estimator for nonlinear models [61]. The extended Kalman filter, as its name implies, is an extension of the regular Kalman filter. It uses a linear approximation of the nonlinear function and then applies Kalman filtering. Another often used extension of the Kalman filter is the unscented Kalman filter. The unscented Kalman filter approximates the mean and covariance by first propagating a chosen set of sample points through the nonlinear system. More information about the unscented Kalman filter can be found in [61]. Multiple studies have compared the performance of the extended Kalman filter and the unscented Kalman filter for building temperature estimation [27, 34]. Both the extended and unscented Kalman filters were able to accurately estimate the states. The unscented performed only slightly better and has some other advantages, such as the ability to deal with large estimate deviations, but this comes at the cost of a higher computation time. A summary of the most common state estimators is given in Table 3-5. The extended Kalman filter is chosen, as it performed sufficiently well and has a low computation cost. Even with large initial errors in the estimate, the filter was able to converge. A graphical description of an extended Kalman filter is shown in Figure 3-3.

Table 3-5: Nonlinear state estimators overview [38].

Estimator	Model	Distribution	Computation cost
Kalman filter	Linear	Gaussian	Low
Extended Kalman filter	Locally linear	Gaussian	Low (analytical Jacobians) Medium (numerically computed Jacobians)
Unscented Kalman filter	Nonlinear	Gaussian	Medium
Particle filter	Nonlinear	Non-Gaussian	High

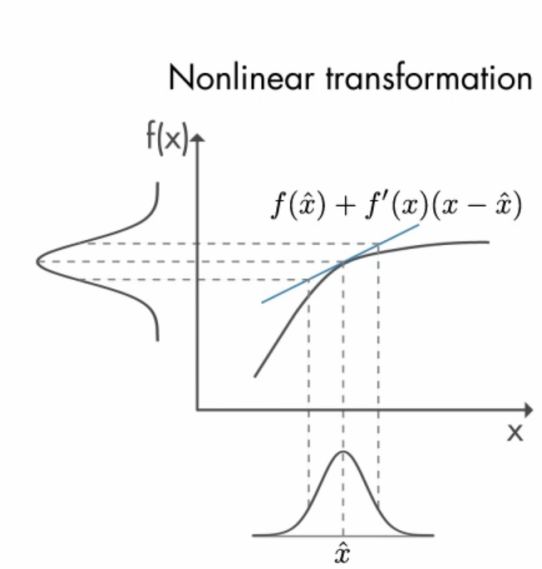


Figure 3-3: General depiction of an extended Kalman filter [38], where the nonlinear model is locally approximated by a linear function.

The temperature gradient with respect to time, $\mathrm{d}T/\mathrm{dt}$, at each time step is known from the equations described in the previous section. A local linear approximation of the nonlinear system, $\dot{x} = f(x, u, d)$, is constructed and discretized for the extended Kalman filter to estimate the states. The nonlinear terms come from the radiation heat transfer. The radiation terms from (2-6) can be rewritten as [13]:

$$\dot{q}_{\mathrm{rad}} = \mathrm{A} h_r (T_2 - T_1) \tag{3-24}$$

where h_r is the radiation transfer coefficient. This is given by:

$$h_r = \frac{\sigma(T_2^2 + T_1^2)(T_2 + T_1)}{\frac{1-\epsilon_1}{\epsilon_1} + \frac{1}{\mathrm{F}_{12}} + \frac{1-\epsilon_2}{\epsilon_2}} = \frac{4\sigma \bar{T}^3}{\frac{1-\epsilon_1}{\epsilon_1} + \frac{1}{\mathrm{F}_{12}} + \frac{1-\epsilon_2}{\epsilon_2}} \tag{3-25}$$

where $\bar{T}$ is the average temperature:

$$\bar{T}^3 = \frac{1}{4}(T_2^2 + T_1^2)(T_2 + T_1) \tag{3-26}$$

This is done for both the radiation heat transfer between the glass layers (2-7) and the radiation heat transfer with the surroundings (2-8). The approximation above holds for small temperature differences ($10° - 20°$K) [3]. The linearized model is discretized using the *c2d* command in Matlab with zero-order hold method. The resulting linear model is given by:

$$\begin{aligned} x(k+1) &= A_k x(k) + B_k u(k) + E_k d(k) \\ &= A_k x(k) + B_k^{'} u^{'}(k) \end{aligned} \tag{3-27}$$

where the control inputs and disturbances are grouped into $B_k^{'} = [B_k \quad E_k]$ and $u^{'}(k) = [u(k) \quad d(k)]^{\mathrm{T}}$. The state matrices are now dependent on time, as they depend on the location of the nonlinear system, *i.e.* states, inputs, and disturbances. The Kalman filter with the prediction and update step is formulated as:

$$\hat{x}(k) = \underbrace{A_{k-1}\hat{x}(k-1) + B_{k-1}^{'} u^{'}(k-1)}_{\text{Prediction}} + \underbrace{K_k[y(k) - C\hat{x}(k)]}_{\text{Update}} \tag{3-28}$$

where $\hat{x}$ is the state estimate and K_k is the time-varying Kalman gain. The prediction step is the estimated state based on the model and previous estimation, while the update step is the correction factor for the measurements. The Kalman gain is defined as:

$$K_k = (\mathrm{S} + A_k P_k \mathrm{C}^{\mathrm{T}})(\mathrm{C} P_k \mathrm{C}^{\mathrm{T}} + \mathrm{R})^{-1} \tag{3-29}$$

where P_k is the solution to the discrete-time algebraic Riccati equation at time k, given by (3-30). Q, R, and S are the correlations matrices of the noise, and C is the output matrix of the system.

$$P_k = A_k P_k A_k^{\mathrm{T}} + \mathrm{Q} - (\mathrm{S} + A_k P_k \mathrm{C}^{\mathrm{T}})(\mathrm{C} P_k \mathrm{C}^{\mathrm{T}} + \mathrm{R})^{-1}(\mathrm{S} + A_k P_k \mathrm{C}^{\mathrm{T}})^{\mathrm{T}} \tag{3-30}$$

The Q, R, and S matrices are the correlation matrices of the normally distributed measurement and process noise, with mean and variance

$$\begin{bmatrix} v(k) \\ w(k) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathrm{R} & \mathrm{S} \\ \mathrm{S} & \mathrm{Q} \end{bmatrix} \geq 0 \right) \tag{3-31}$$

The noise v and w in (3-31) are zero-mean white noise and uncorrelated, *i.e.* $\mathrm{S} = 0$. In reality, the noise w is unknown and comes from some uncertainty in the model. For simulations, the value for w is added to cause some discrepancy between the plant model and model used for predictions and control, as would occur in the actual building. For the Kalman filter, the ratio between R and Q essentially determines the amount of trust in the measurement versus the model. The values of Q are chosen based on the response of the model and Q is a diagonal matrix with all zeros off-diagonal:

$$\mathrm{Q} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_{17}^2 \end{bmatrix} \tag{3-32}$$

where σ_i corresponds to the i-th state. The values for the walls, *i.e.* T_1-T_{12}, are $\sigma_{1-12} = 0.2$. The other values are chosen to be $\sigma_{13} = 0.8$, $\sigma_{14} = 0.2$, $\sigma_{15} = 0.2$, $\sigma_{16} = 0.15$ and $\sigma_{17} = 0.03$. The choice for the measurement σ values in R and the noise v are discussed in Section 3-4. With the additive process and measurement noise, the nonlinear plant model is described by:

$$\begin{aligned} x(k+1) &= f(x(k), u(k), d(k)) + w(k) \\ y(k) &= \mathrm{C}x(k) + v(k) \end{aligned} \tag{3-33}$$

This model will represent the actual building during the simulations. This concludes the plant model and Kalman filter. To determine the effect on the energy consumption a predictive controller is implemented. This is discussed in the next section.

3-3 Model Predictive Controller Design

In practice, buildings are often controlled by a simple PID or ON/OFF controller. These are generally simple to implement, but also lack in performance and are less suitable for passive buildings. For passive HVAC systems, Model Predictive Control (MPC) is a promising method. To utilize changing weather conditions and occupancy for HVAC, predictions have to be made about the near-future conditions. Control actions that are possible now are not necessarily possible with passive control in the future and vice versa. The main goal is to investigate the potential benefit of additional state measurements on prediction accuracy and control performance. For this research, a nonlinear predictive controller is used. This is more advanced than a linear controller, while still computational feasible with some simplifications. For large-scale buildings, or when considering more control inputs, this would be a less attractive choice. The number of control inputs is reduced to limit the number of optimization variables. All the blinds are controlled by one input. The actual building has independent control of the blinds for each facade. This is useful when lighting conditions and glare prevention are considered inside the building, but this is not the case in this research. Controlling all the blinds by one input will not affect the control performance in terms of energy consumption. Another simplification is the use of $\dot{q}_{\mathrm{hvac}}$ as a direct control input in (3-17). In practice, window opening control, supply air velocity, and heat recovery from the room would also be considered.

An MPC uses, as its name suggests, a model of a plant to predict future evolution. It then uses an optimization algorithm to optimize the set of control inputs, according to the defined objective function and constraints. At each current time step, the set of optimal control inputs over a horizon N is calculated. The first control input is applied to the system and this process then repeats itself. The predicted optimal control sequence and model output at a time step is depicted in Figure 3-4. The optimization has a general structure given as:

$$J_{\mathrm{opt}} = \underset{u_0, \ldots, u_{N-1}}{\text{minimize}} \quad \sum_{k=0}^{N-1} l_k(x_k, u_k) + l_f(x_k, u_k) \tag{3-34a}$$

$$\text{subject to} \quad (x_k, u_k) \in \mathcal{X}_k \times \mathcal{U}_k, \tag{3-34b}$$

$$x_0 = x, \tag{3-34c}$$

$$x_{k+1} = f(x_k, u_k) \tag{3-34d}$$

where J_{opt} is the optimal cost function, l_k and l_f are the intermediate and final cost function, and N is the prediction horizon. In the optimization above, (3-34b) is the state and input constraints, (3-34c) is the initial condition, and (3-34d) is the system dynamics. The allowable sets $\mathcal{X}$ and $\mathcal{U}$ are defined by some (non)linear (in)equality constraint functions.

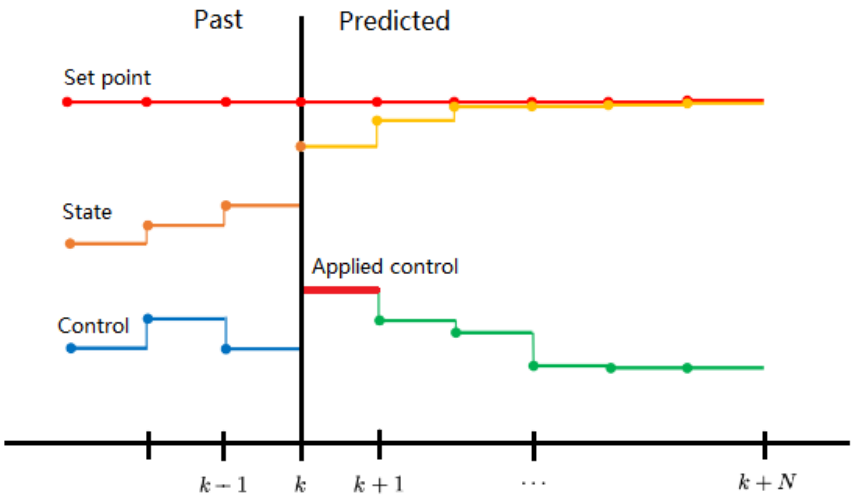


Figure 3-4: General description of MPC, where the control input sequence is optimized over horizon N and the first input is applied.

Control Structure and Algorithm

The nonlinear MPC is based on the building model (3-20). Important to note is that all the future disturbances $d(k)$ are assumed to be known. With the goal of minimizing energy consumption while ensuring thermal comfort the following objective function and constraints are constructed:

$$J_{\text{opt}} = \min_{\substack{\mathbf{u}(0),\dots,\\ \mathbf{u}(N-1)}} \quad \sum_{k=0}^{N-1} |u_1(k)| + \lambda \sum_{k\in\mathcal{S}_{\text{occ}}} (x_{15}(k) - \mathrm{T}_{\text{ref}})^2 \tag{3-35a}$$

$$\text{s.t.} \quad x(k+1) = f(x(k), u(k), d(k)), \tag{3-35b}$$

$$T_{lb}(k) \leq x_{15}(k) \leq T_{ub}(k), \qquad k = \{1,\dots,N\}, \tag{3-35c}$$

$$\dot{\mathrm{q}}_{\min} \leq u_1(k) \leq \dot{\mathrm{q}}_{\max}, \qquad k = \{0,\dots,N-1\}, \tag{3-35d}$$

$$0 \leq u_2(k) \leq 1, \qquad k = \{0,\dots,N-1\}, \tag{3-35e}$$

$$x(0) = x_0 \tag{3-35f}$$

The control inputs u_1 and u_2 are the HVAC supplied power and blinds position respectively. With a prediction horizon N of 24-hour and a control time step of one hour, there are a total of 48 optimization variables. The weight λ is a tuning parameter to adjust the relative dominance of the two objectives. The choice for this parameter is described down below. The set $\mathcal{S}_{\text{occ}}$ contains the time steps k for which the building is occupied. When the building is

unoccupied, no weight is applied to the temperature deviation from the reference temperature T_{ref}. The lower and upper room temperature bound T_{lb} and T_{ub} are time-dependent, *i.e.* a different bound is considered for night and day. The cooling and heating capacity of the HVAC system is limited to $\dot{q}_{min} = -7$ [kW] and $\dot{q}_{max} = 7$ [kW], respectively.

The Matlab function *fmincon* is used to solve the optimization in (3-35a). The optimization problem is nonlinear, constrained, and nonconvex. Based on these properties, the sequential quadratic programming algorithm, *sqp*, is chosen. Other algorithms include *interior-point*, *active-set*, and *trust-region-reflective*. The *trust-region-reflective* needs to be supplied with gradient functions for the constraints and objective, which is not available. The *interior-point* uses a barrier function and needs all the solutions to be feasible. This could be done by implementing soft constraints on the state instead of hard constraints. The two algorithms, *sqp*, and *interior-point*, showed comparable results. Therefore, the hard constraint was kept and the *sqp* algorithm was used instead. The *active-set* algorithm is essentially the same as *sqp*, but the former only considers the currently violated, also called active, constraints. The *sqp* method uses a more efficient linear algebra routine, that is faster than the ones used in *active-set* [1]. More information on the different algorithms can be found in [39].

Normally, when dealing with nonlinear nonconvex optimization a multi-start is applied to find the global optimum. This, however, increases the computation of the optimization. The solutions were not affected much when using a multi-start. All the local minima lie close together, indicating a near-convex optimization problem. A convexity analysis was done by [6] and the multi-start did indeed not improve the solution by much. To reduce computation time, a multi-start is not considered and the maximum number of iterations is reduced. The first maximum number of iterations is set to 50, and to 25 thereafter. These values are chosen based on observations of the objective cost during iterations. The value of the objective function did not increase significantly after 25 iterations. The first optimization uses more iterations, given the initial guess. For each new calculation of the optimal control inputs, the previously found solution is shifted one time step and used as the starting point.

Multi-Objective Optimization

The objective of the controller is to minimize energy consumption, while also satisfying thermal comfort. This results in two competing objective functions in (3-35a) that need to be minimized. There are several methods to deal with multi-objective optimization, such as the weighted sum strategy, goal attainment method, and ϵ-constraint method [39]. The main difficulty with the weighted sum strategy is the choice of the relative weight between the objectives. It is not necessarily clear what the relation, in this case, between the sum of the squared temperature error and power consumption should be. Also, not all the Pareto optimal points can be found if the solutions are non-convex. The goal attainment method is more intuitive in the sense that it minimizes the error in the objectives relative to some predefined objective goals. But again, the goal for both objectives is not directly clear. A simpler and more intuitive method is the ϵ-constraint method, where a boundary is put on one of the objectives. For the objectives, the standard deviation from the reference temperature and the energy consumption are chosen. The constraint for the standard deviation is set to $\sigma = 0.333$. This is a somewhat arbitrary value, but means that during control the temperature stays within $\pm 1^{\circ}$C of the reference temperature 99.7% of the time. The constraint on the temperature in the remainder of this research is also set to ± 1 °C within the

reference temperature of 21 °C during occupied hours. This leaves some room for when the Kalman filter is implemented together with the process and measurement noise. To see the effect of the weight on the two objective functions, the constraint (3-35c) has temporarily been removed. The balance between reference tracing and the energy consumption is fully determined by the two objective costs. The results for different weights λ over a 3 day period are shown in Figure 3-5. A weight of $\lambda = 400$ is chosen for the MPC.

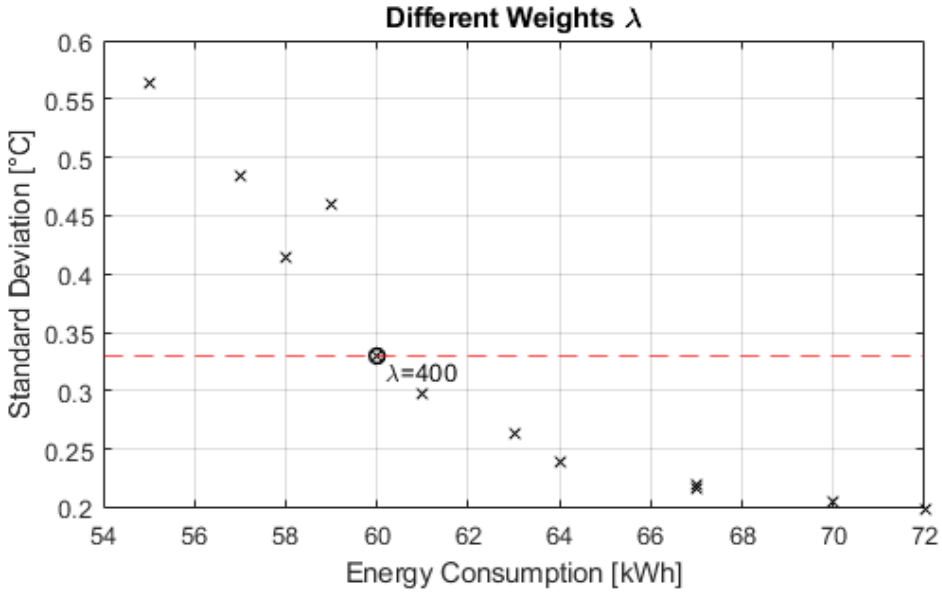


Figure 3-5: Effect of different weights on the standard deviation and energy consumption over a 3 day period.

3-4 Sensor Selection and Optimization

As mentioned in Section 2-4, there are numerous ways to select the optimal sensor locations. Most methods use the RMSE or entropy of the selected set average compared to the average of all sensors as a measure for optimally. However, these studies focus on a collection of sensors measuring indoor air temperature. The sensors are located throughout an office space or large room with large spatial temperature differences. Clustering algorithms are employed to group sensors with similar temperature readings and filter out the redundant sensor locations. The CONVERGE building is different in this aspect, in that it is one medium-sized room, and equipped with a lot of sensors, that measure, besides the air temperature, also different building components, *e.g.* walls, ceiling, and floor. Therefore, these methods cannot be directly applied to this research. Two different methods are used to determine the sensor positions: one method where the sensors are selected before control is applied, and the other where the sensors are selected based on control performance. The methods are referred to as the *prediction method* and the *greedy method* respectively. Both methods rely on an assumption that was used by [34] to compare different controllers. They assumed that the error of a subset of sensors compared to the average of the whole can be described by a Gaussian distribution. An example is depicted in Figure 3-6, where the number of temperature error occurrences over a 10 day period are shown. The error is between the average of a sensor

selection compared to the average of the whole set. The whole set, here, refers to all the sensors positioned on one building element, *i.e.* measuring one state. The average of all the sensors measuring the same state is considered the true temperature of that state. This is done for all of the measured states and all the sensor combinations.

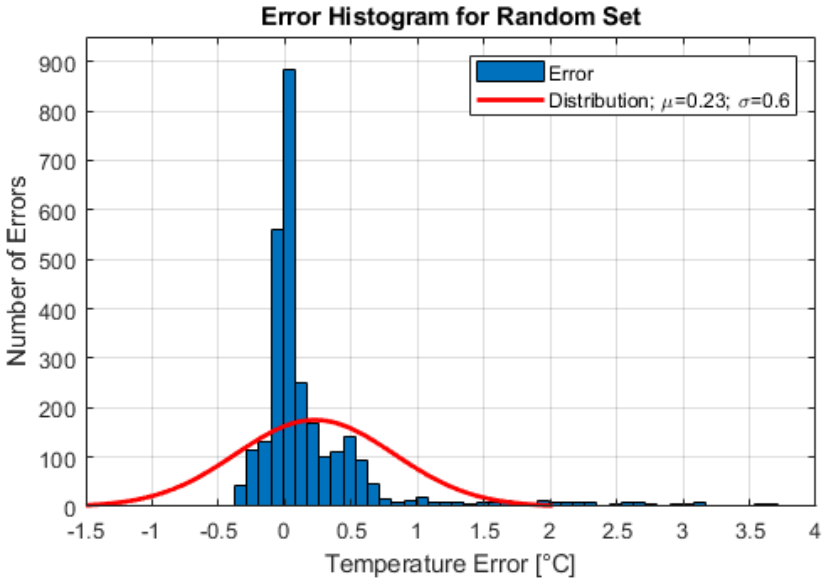


Figure 3-6: Error histogram of a sensor subset average compared to the whole set over 10 days with a fitted normal distribution.

The error distributions obtained from different sensor sets are used in (3-33) as an added sensor measurement error v_k. Normally, this error represents the measurement error arising from the physical properties of the sensors, *i.e.* measurement noise, but now it is used as a general uncertainty of the sensor measurement. With the measurement error due to sensor location assumed to be Gaussian, both the sensor measurement noise, v_m, and sensor set error, $v_{\mathcal{S}}$, can be added together. The sum of two normally distributed random variables, v_m and $v_{\mathcal{S}}$, results in another normally distributed random variable v [52], as shown by:

$$v_m \sim \mathcal{N}(\mu_{v_m}, \sigma^2_{v_m}), \tag{3-36a}$$
$$v_{\mathcal{S}} \sim \mathcal{N}(\mu_{v_{\mathcal{S}}}, \sigma^2_{v_{\mathcal{S}}}), \tag{3-36b}$$
$$v = v_m + v_{\mathcal{S}}, \tag{3-36c}$$
$$v \sim \mathcal{N}(\mu_{v_m} + \mu_{v_{\mathcal{S}}}, \sigma^2_{v_m} + \sigma^2_{v_{\mathcal{S}}}) \tag{3-36d}$$

The measurement noise is set to $\sigma_{v_m} = 0.1$, meaning that 95% of the temperature readings lie within a 0.2°C deviation. This is value is chosen arbitrarily, but it is a reasonable value for the temperature sensors. The measurement noise is zero-mean, $\mu_{v_m} = 0$. The sensor set noise parameters, $\mu_{v_{\mathcal{S}}}$ and $\sigma_{v_{\mathcal{S}}}$, are chosen to represent the temperature error due to the sensor position and have the values described above. With this setup, the different sensor selections are compared. The first method to assess the optimally of a given set is done without considering the control performance. The potential advantage of measuring

more states is that better near-future temperature predictions can be made. With better predictions, control input lack can be reduced, or in other words, the system can respond faster to changing temperature conditions. This also leads to fewer temperature violations and possibly more optimal control, *i.e.* less power consumption. The second method considers the control action and uses a greedy algorithm to select the best sensors.

Prediction Method

The first method uses the Kalman filter and a temperature prediction to assess the effect of the sensor selection. This method is similar to methods used in other research, in the sense that it uses the RMSE and maximum error of the model compared to the actual data. However, there are some differences. In general, adding more sensors results in a better temperature estimate. In this case, the sensors are not only measuring the air temperature, but also some additional states. To take into account the effect of additional state measurements, the predicted model output is used as a performance metric instead of only the model accuracy at the current time step. The accuracy of the current temperature estimate is largely determined by the sensor reading of the air temperature. Using additional state measurements, *e.g.* of the walls, does little to improve the accuracy of the current estimate. The predicted temperature evolution, however, is affected by the temperature of all the building components. Thus, it is reasoned that having a better estimate for these states will result in a more accurate temperature prediction in the near future. When making temperature predictions over a longer period of time, *e.g.* multiple hours, other uncertainties in the model and disturbance estimates will become more dominant. Therefore, a prediction period of one hour is used. This is the same time step used for the control update, for which the control input remains constant. The RMSE, between the prediction and actual temperatures, is used to compare and evaluate the performance of the different sensor selections. The calculation of the RMSE is given by:

$$RMSE = \sqrt{\frac{1}{\mathrm{N_p}} \sum_{k=1}^{\mathrm{N_p}} (T_{\mathrm{pred}}(k_0 + k) - T_{15}(k_0 + k))^2} \tag{3-37}$$

where $\mathrm{N_p}$ is the number of prediction steps, T_{pred} and T_{15} are the predicted and actual air temperature, and k_0 is the initial time step. The maximum error over the prediction horizon is also used to compare the different sensor sets. The infinity-norm of the error, *i.e.* maximum absolute error, is given by:

$$||e||_\infty = \max_{k \in \{1,\dots,\mathrm{N_p}\}} |T_{\mathrm{pred}}(k_0 + k) - T_{15}(k_0 + k)| \tag{3-38}$$

No control in applied when optimizing using the prediction method, but only real sensor data or simulated sensor data. The simulated sensor data is generated with the plant model, (3-33), and measurement error v, (3-36d), corresponding to the selected sensor set. The prediction method with the real sensor data uses the measured sensor data directly, instead of $y(k) = \mathrm{C}x(k) + v(k)$, to estimate the current state using the Kalman filter. This makes the method computationally more attractive compared to the other method. The performance of all the sensor combinations up to six sensors can be evaluated within a reasonable amount of time. A description is shown in Algorithm 1.

Algorithm 1 Prediction Method Algorithm

Input: Maximum number of sensors, weather data, sensor data, number of steps n, and prediction horizon N_p.

while number of sensors $\leq$ maximum number of sensors
  Chose sensor sets with a specified number of sensors; get noise covariance matrix R
  Simulated the plant and y over n steps, or use sensor data
  Run Kalman filter: estimate state temperatures, $\hat{x}$, over n steps
  for $t = 1 : n$
    for $k = 1 : N_p$
      Predict the temperature $x(t + k)$ with $x_0 = \hat{x}(t)$
    end for
    Calculate RMSE and $||e||_\infty$ over prediction horizon N_p
  end for
  Select the next sensor set
end while

Output: Average RMSE and $||e||_\infty$ of n_{est} predictions for each sensor set.

Greedy Method

Simulating all the different possible combinations of sensors with control would be extremely computationally intensive. With 35 sensors, there are $2^{35} - 1 = 34,359,738,367$ different sensor combinations. Even if only one sensor is used for each state there would be $2^8 - 1 = 255$ combinations possible. While optimization methods such as Genetic Algorithm (GA) and particle swarm could be used for this problem, they are generally less suitable when the evaluation of the objective function has a high computation time. Therefore, the method used to determine the effect of sensor selection on control performance is based on a greedy algorithm. As for the predictive method, it is reasoned that the accuracy of the temperature predictions increases when more sensors are added. This gives the possibility to better utilize the passive systems, in this case, the solar blinds, to make temperature adjustments with minimal energy consumption. The algorithm starts with one sensor. There are eight different states measured, so there are a maximum of eight optimization points at each new iteration. The sensor with the lowest σ_v value to the corresponding state is chosen as described at the start of this section. The plant model with the Kalman filter and MPC is simulated over a period of multiple days. The resulting energy consumptions are compared and the sensor set with the lowest energy consumption is chosen to initiate the next iteration. In the next iteration, another sensor is added. This leads, again, to a maximum of eight new options. The algorithm can choose to add another state measurement, or add another sensor to a state that is already measured if more sensors are available for that state. In the case where another sensor is added to one of the already measured states, the set of sensors with the lowest σ_v is chosen. This process is repeated until the specified maximum number of iterations, *i.e.* number of sensors is reached. A description is shown in Figure 3-7 and Algorithm 2. Only sensor sets that satisfy the discomfort bound of $D_{ind} \leq 0.3$, are considered when choosing the minimal energy consumption.

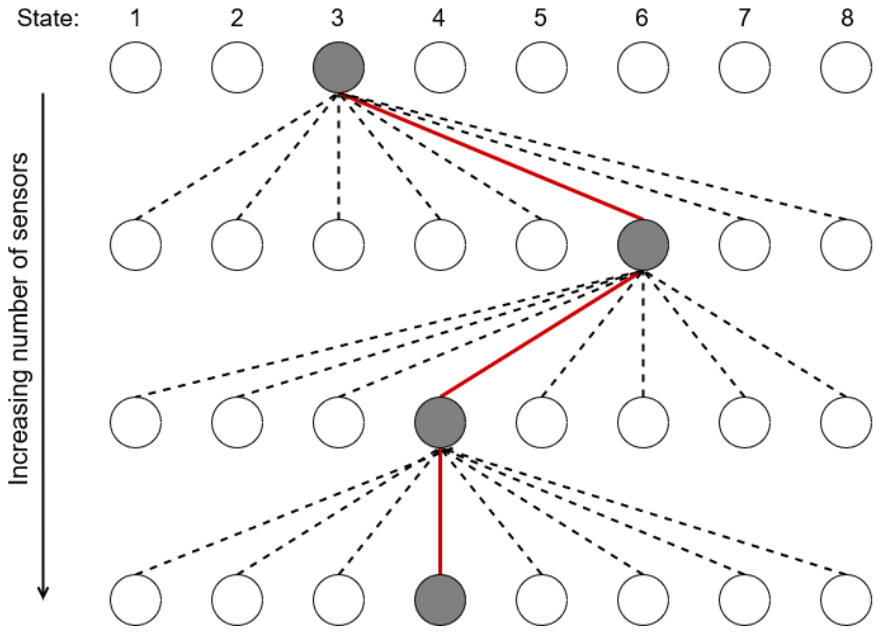


Figure 3-7: Graphical depiction of the greedy algorithm, where the grey nodes have the minimum values. Each layer adds another sensor and the node values represent the power consumption. The algorithm selects the next sensors to be added in a greedy fashion by choosing the sensor with the lowest energy consumption. The resulting combination of four sensors in this example would be one state 3 sensor, two state 4 sensors, and one state 6 sensor.

Algorithm 2 Greedy Algorithm

Input: Maximum number of sensors, weather data, sensor data, simulation steps t_{sim}, and control horizon N.

Start with zero sensors for each measured state: options $= \emptyset$
while number of sensors $\leq$ maximum number of sensors
  Use previously found optimal option if available
  for each state with sensors
    option(state) $\Leftarrow$ add 1 sensor
  end for
  for size(option)
    Select sensor set with minimal RMSE according to the number of sensors for each state
    Create covariance matrix R and set op Kalman filter
    Simulate control over t_{sim} period with control horizon N
    Determine control performance: Q_{HVAC} and discomfort index
  end for
  Select option with minimal Q_{HVAC} that satisfies discomfort bound
end while

Output: Optimal sensor set for given number of sensors, Q_{HVAC} and discomfort index.

3-5 Conclusions

In this chapter, the general setup that is used to determine the effect of sensor selection in Chapter 4 has been discussed. In Section 3-1 a white-box model has been established using the first-principles as described in Section 2-3. Some of the parameters have been optimized by fitting the model outputs to the sensor data. The results of the model on the validation data have been shown. Although there is still some discrepancy between the model and sensor data, the model is able to accurately describe the overall temperature dynamics. For state estimation, an extended Kalman filter has been implemented in Section 3-2. This filter uses a linearized version of the building model at each time step to estimate the states based on the prediction from the model and received sensor data. For the controller, an MPC has been used as described in Section 3-3. With selected parameters, the controller is able to control the room temperature within the specified bounds.

Section 3-4 describes the method used to simulate and compare different sensor selections, and two algorithms constructed to optimize the selection. The goal of the constructed methods is to determine if additional state measurements lead to improved model and control performance. One method is based on the prediction accuracy of the model, the prediction method, and the other is used to optimize the sensors based on control performance, the greedy method. The former uses the RMSE and infinity-norm of the error of the predicted temperature versus the actual temperature. The reasoning is that better estimates of the current states, *i.e.* temperatures of all the building elements, will lead to a more accurate predicted temperature evolution for the near future. With this method, both the simulated sensor data, using the noise $v(k)$ as described in Section 3-4, and actual sensor data can be used during the state estimation. The estimated states serve as the starting points for the temperature prediction over the next hour. The greedy method directly selects the optimal sensors based on the control performance. Starting with one indoor air temperature sensor, and added a sensor one at a time by selecting the sensor set with the lowest energy consumption that satisfies the thermal comfort bounds. The different models used and their use cases are summarized in Table 3-6.

Table 3-6: Overview of the different models used.

Use case	Model	Reference
Plant	$x_{k+1} = f(x_k, u_k, d_k) + w_k$ $y_k = \mathrm{C}x_k + v_k$	(3-20)
Kalman filter	$\hat{x}_k = A_{k-1}\hat{x}_{k-1} + B'_{k-1}u'_{k-1} + K_k[y_k - C\hat{x_k}]$	(3-27)
Prediction & Control	$x_{k+1} = f(x_k, u_k, d_k)$ $y_k = \mathrm{C}x_k$	(3-33)

Chapter 4

Simulations and Evaluation

With the model, observer, and controller complete, the methods designed for the sensor selection in Section 3-4 can be tested. First, the general setup of the two optimization methods is described in Section 4-1. This is followed by a section on the results and a discussion of the methods in Section 4-2. Lastly, this chapter is closed with a conclusion of the results obtained, in Section 4-3.

4-1 Case Study Setup

Interesting periods for passive Heating, Ventilation, and Air Conditioning (HVAC) systems are spring and autumn. This is due to the variability in weather conditions. To test the different methods for sensor selection, data is used that has been collected over a period of 10 days in April. This is also the same period the model has been validated for. In Figure 4-1 the indoor air temperature is plotted with and without control applied. During this period the building needs to be heated, as well as cooled. As can be seen, the temperature would violate both the lower and upper temperature bounds when no control is applied. This makes it an interesting choice for the greedy method, to see the effect of the prediction accuracy on the control performance.

Since considering all of the different sensor combinations would be too computationally expensive, both methods only consider the sensor sets with the minimal Root Mean Square Error (RMSE) for a given number of sensors. This reduces the number of combinations for one state, *e.g.* the air temperature, with three sensors from eight to three options. The values used for the air temperature sensors are given in Table 4-1. In this case, if one sensor is used to measure the air temperature, sensor {2} is selected, and when two sensors are used, the set {1,3} is selected. The numbers refer to a particular sensor, *i.e.* sensor 1, sensor 2 and sensor 3. Even with this reduction, considering all of the sensors would still result in a lot of possible combinations. Both the predictive and greedy method only consider combinations of up to six sensors. It is unlikely that more than six sensors would be used in practice. A complete overview of the sensor sets used can be found in Table B-1. Both methods are discussed in

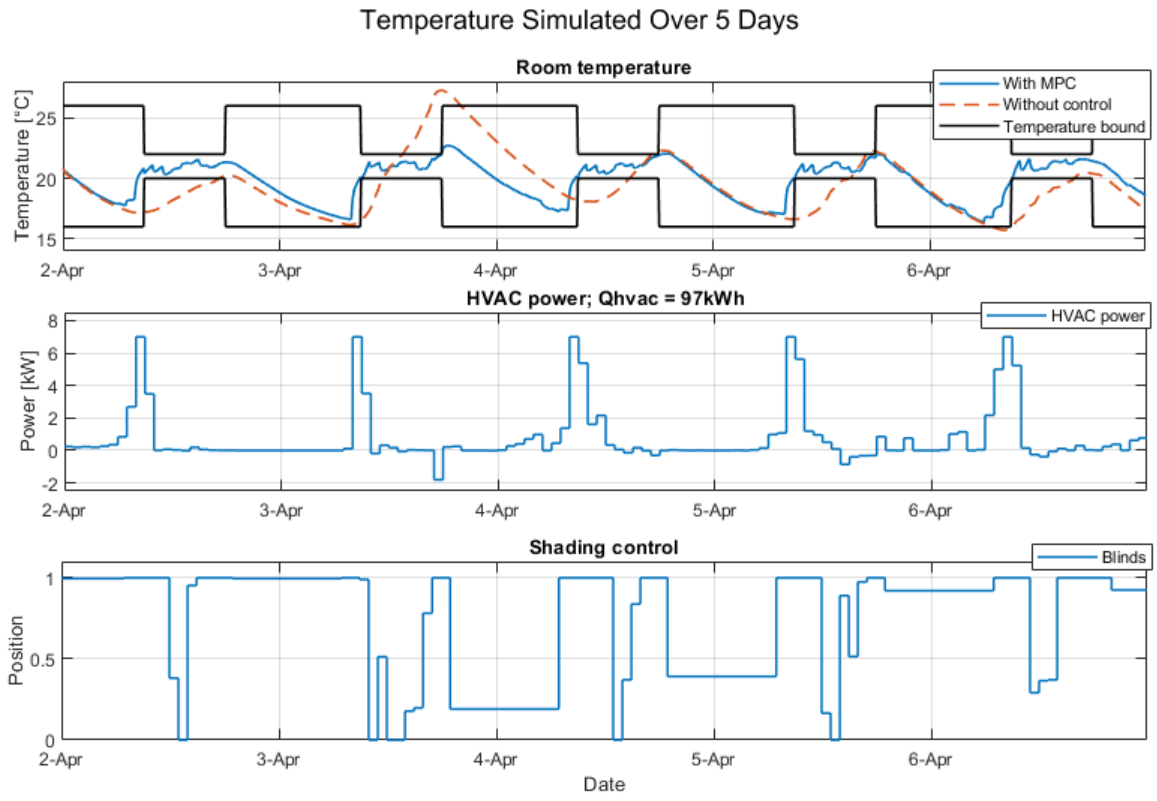


Figure 4-1: Indoor temperature with and without control simulated over 5 days.

more detail in the next two paragraphs.

Table 4-1: Example of the sensor combinations for the three air temperature sensors.

Sensor set $\mathcal{S}$:	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}
RMSE	0.41	0.24	0.45	0.22	0.12	0.21	0
$\mu_{v_\mathcal{S}}$	-0.37	0.35	0.02	-0.18	-0.01	0.18	0
$\sigma_{v_\mathcal{S}}$	0.18	0.24	0.27	0.14	0.12	0.09	0

All of the simulations are carried out using MATLAB 2021a, with the Optimization Toolbox for the optimizations.

Predictive Method

For the predictive algorithm, no occupancy is considered. This is not relevant, since this method does not include temperature control. But more importantly, with no occupants, the simulated plant model can be compared to the actual sensor data. In the predictive method, both the simulated sensor and sensor data are used. With simulated sensor data is referred to the situation where the plant model is:

$$\begin{aligned} x(k+1) &= f(x(k), u(k), d(k)) + w(k) \\ y(k) &= \mathrm{C}x(k) + v(k) \end{aligned} \tag{4-1}$$

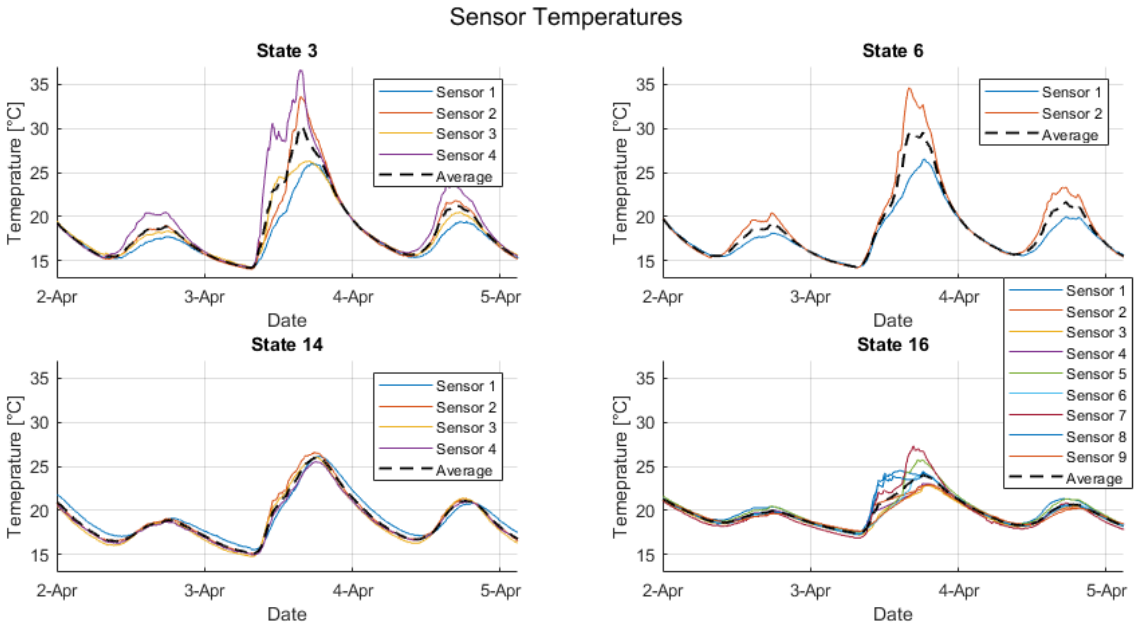


Figure 4-2: Sensor data for some of the states.

with the measurement noise (3-36d) as described in Section 3-4. The general structure, when the simulated data is used, is shown in Figure 4-3. The prediction model is similar to the plant model, but without the measurement and process noise added:

$$x_{\text{pred}}(k+1) = f(x_{\text{pred}}(k), d(k)) \tag{4-2}$$

where the initial starting point for x_{pred} is equal to the estimated state x_{est} at the start of the predictions over the prediction horizon. All of the models used, observer, plant, and prediction consider the same time step of 300 seconds. This is the same interval at which the sensor data is logged.

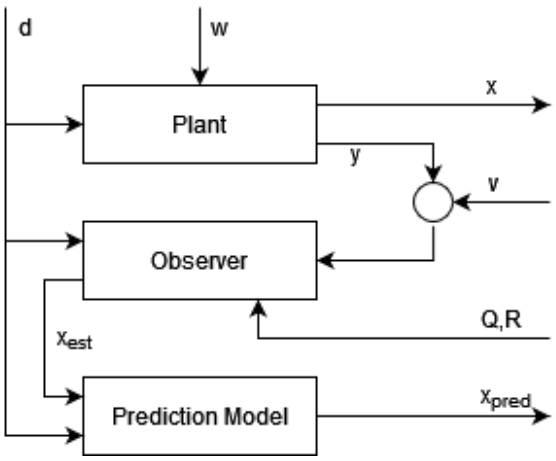


Figure 4-3: Structure used in the predictive method.

When the sensor data is used, the same structure as in Figure 4-3 applies. The only difference is that the plant model and output y are replaced with the sensor data. As not all of the states are measured, the missing states are added by running the plant model without process noise and using the averaged sensor data for the measured states. The output y is the sensor data of the selected sensor set. Both the simulated data and sensor data use the same model, (4-2), for the predictions.

Greedy Method

To take the control performance into account, a greedy optimization method is also implemented. This method simulated the building over a period of three days with control. The method uses the same model for the plant and for prediction as in the predictive method but also incorporates the Model Predictive Control (MPC) designed in Section 3-3. An overview of the system structure with the model, Kalman filter, and the controller is shown in Figure 4-4. More details about the algorithm itself can be found in Section 3-4.

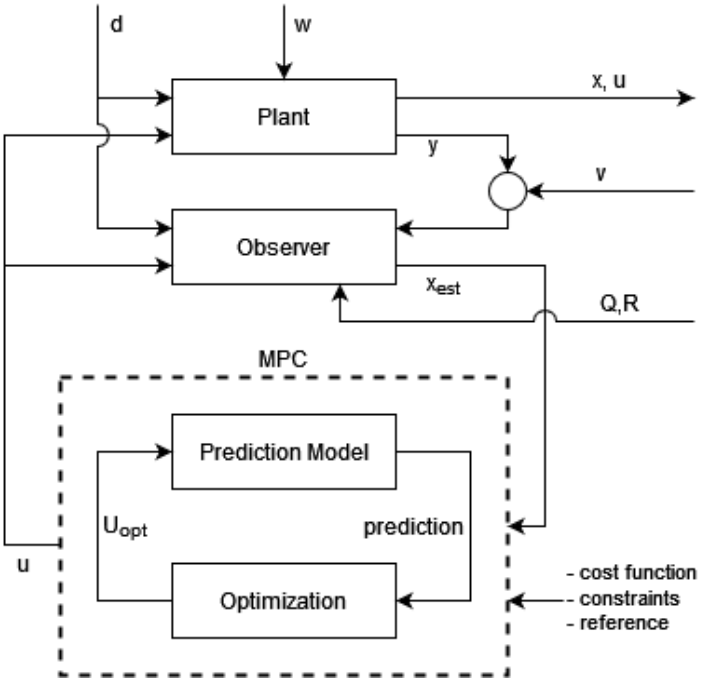


Figure 4-4: Overall system structure used in the greedy method.

The greedy method does include occupancy during the morning and afternoon, as it affects the control performance. A typical occupancy of between 10 and 30 occupants is used for each of the three days. The occupancy schedule is shown in Figure 4-5. The occupied hours are set to be from 8 a.m. till 17 p.m. Occupants are one of the main sources of internal heat generation and need to be included when considering control performance. The occupancy schedule used is shown in Figure 4-5. For thermal comfort, the lower and upper temperature bounds are set to 16 °C and 26 °C during unoccupied hours, and to 20 °C and 22 °C during occupied hours, respectively. In the latter, the reference temperature is set to 21 °C, and no reference temperature is used during the unoccupied hours.

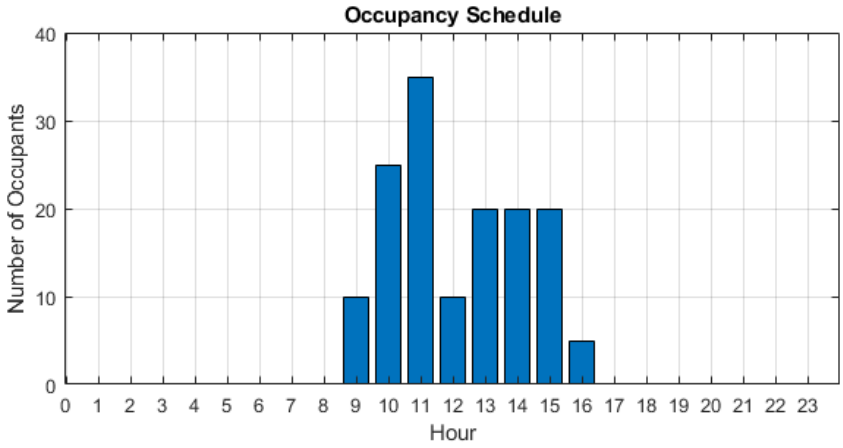


Figure 4-5: Occupancy schedule during the day used in the greedy method.

The time step for the plant and observer is with 300 seconds the same as for the predictive method. The control input is kept constant over a period of one hour to reduce computation time. With two control inputs and a control horizon of 24 hours, this results in a total of 48 optimization variables. The objective function is evaluated a considerable amount of times during each optimization. To reduce the computation time further, the prediction model uses a time step of 600 seconds. While a larger time step could be used, there is a compromise between model accuracy and computation time. Using a value of 600 for the time step still results in an accurate prediction with half the number of simulation steps.

4-2 Case Study Results

Predictive method

The results for the sensor sets up to six sensors are shown in Figure 4-6a and Figure 4-6b for the simulated and sensor data, respectively. Both the simulated data and sensor data sets show improved accuracy when more sensors are added, although the difference is quite small. The sensor sets with minimal RMSE and minimal $||e||_\infty$ in this figure are shown in Table 4-2 and Table 4-3 for the simulated and real sensor data, respectively.

Different results are obtained for the simulated and real sensor data in the worst-case predictions. The simulated worst-case sensor sets vary slightly in prediction accuracy, but no general trend can be seen when more sensors are added. When using the predictive method with real sensor data, however, average RMSE decreases after an initial increase for two sensors and the maximum absolute error increases. The increase in the average RMSE for two sensors using the sensors can be explained by a state that is poorly estimated with one sensor, caused by large local temperature differences within that state. Adding a poorly positioned sensor will decrease the accuracy of the state observer, and with it the temperature predictions made. This can be better observed by comparing Table 4-2 and Table 4-3. With the simulated data, other states are added besides the air temperature when the number of sensors is increased, whereas in the case of the real sensor data air temperature sensors are added first. The Gaussian measurement noise and process noise, added in the simulated case, are more easily approximated by the observer. The actual noise, or error, in the measurement

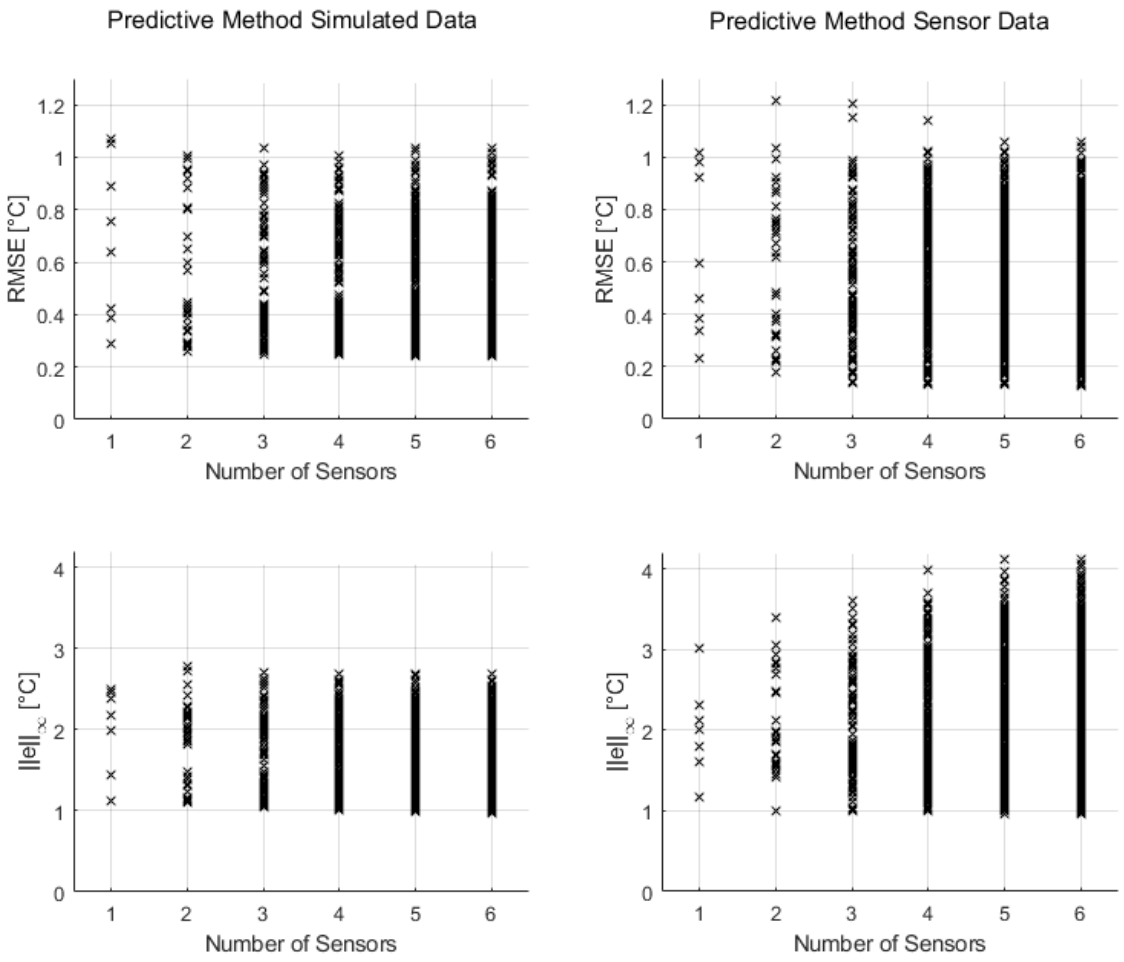


Figure 4-6: Predictive method using simulated sensor data in (a) and using real sensor data in (b). Each black cross is a different sensor combination, simulated over three days with an one hour prediction. The x-axis displays the number of sensors used in total. The total number of sensors used ranges from one to six.

is not Gaussian. More sensors added to the air temperature results impact the accuracy of the state estimate more in the case when the sensor data is used. Note that multiple combinations of sensor sets are very close in terms of minimal RMSE and $||e||_\infty$, so there are other combinations nearly as optimal as the ones shown in Table 4-2 and Table 4-3.

Another difference between the simulated and real sensor data is the maximum absolute error. The maximum temperature error increases for the worst sensors sets when more sensors are used. Here, again, the reason is assumed to be due to the Gaussian distribution. In Figure 3-6, such Gaussian distribution fitted to the measurement error has been shown. The majority of the errors are spread around the mean, but there are a few larger errors visible to the far right. When simulating the sensor data, such errors are unlikely to occur. Using more states with large temperature deviations results in a higher chance of a larger prediction error over the entire period the optimization is carried out. Some the sensor measurements are shown in Figure 4-2. Some states have a large temperature deviation between the sensors. This could, for example, be caused by sensors being located in the shade part of the time. The third state, indoor south wall temperature, is added only by the real data sensor set. This is also the state with the largest misfit between the model and sensors data, as seen in Figure 3-2. It could be the reason why these states are chosen for additional measurement, although the difference in accuracy with other sets still remains small.

Table 4-2: Twelve sensor sets with minimal RMSE or maximum absolute error with the simulated sensors for the number of sensors used and minimization criteria.

Minimization criteria	Total number of sensors	Number of sensors per state: 3	6	9	12	14	15	16	17	RMSE	Max absolute error: $\|\|e\|\|_\infty$
RMSE	1	0	0	0	0	0	1	0	0	0.29	1.12
$\|\|e\|\|_\infty$	1	0	0	0	0	0	1	0	0	0.29	1.12
RMSE	2	0	0	0	0	0	2	0	0	0.26	1.11
$\|\|e\|\|_\infty$	2	0	0	0	0	0	2	0	0	0.29	1.1
RMSE	3	0	0	0	0	0	2	1	0	0.25	1.05
$\|\|e\|\|_\infty$	3	0	0	0	0	1	1	1	0	0.25	1.05
RMSE	4	0	0	0	1	1	1	1	0	0.25	1.02
$\|\|e\|\|_\infty$	4	0	0	0	1	1	1	1	0	0.25	1.02
RMSE	5	0	1	0	1	1	1	1	0	0.24	1.02
$\|\|e\|\|_\infty$	5	0	0	0	1	1	1	2	0	0.24	0.98
RMSE	6	0	1	0	1	1	1	2	0	0.24	0.99
$\|\|e\|\|_\infty$	6	0	0	1	1	1	1	2	0	0.24	0.96

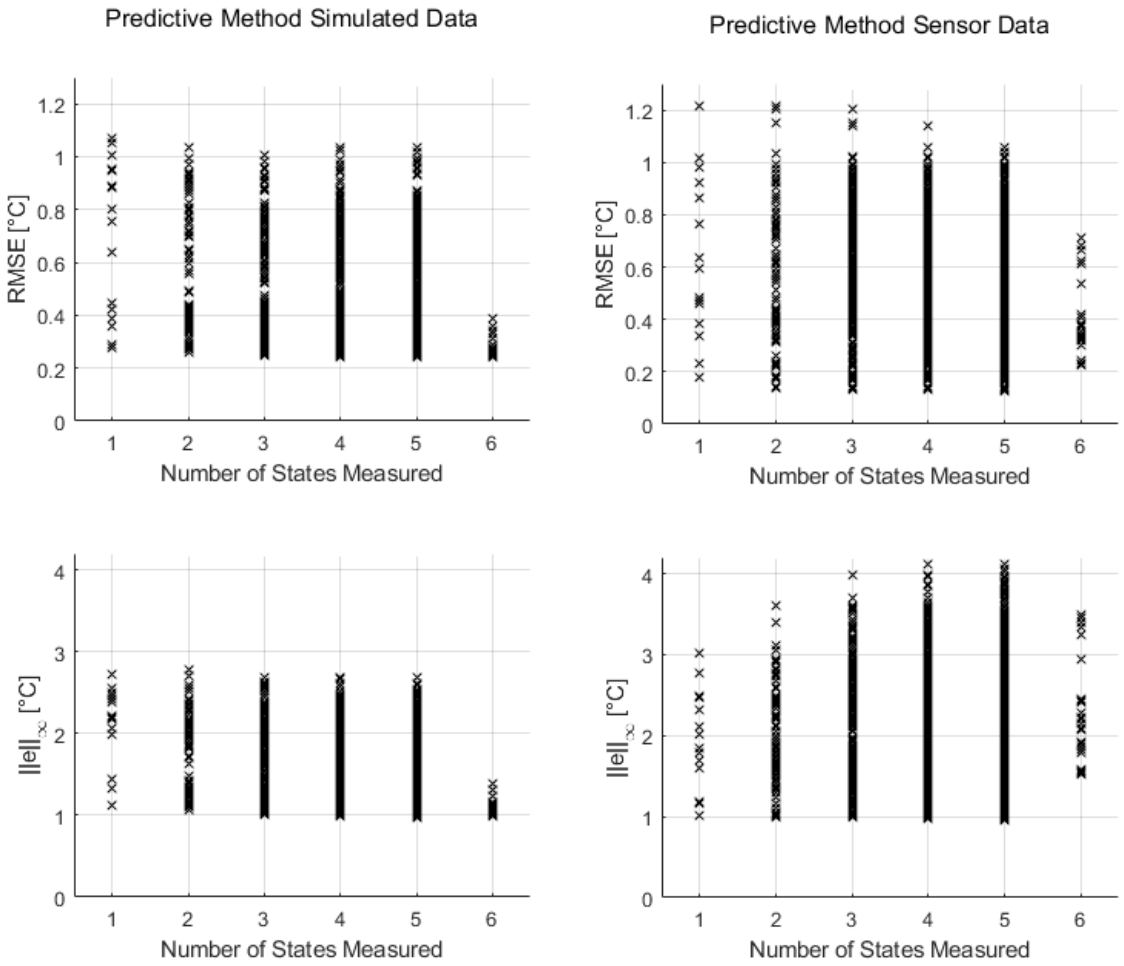
To better visualize the effect of the additional state measurements, Figure 4-7a and Figure 4-7b show the prediction error for the number of states measured. The number of sensors used for each data point ranges from one to six. The maximum for both the average RMSE and absolute error did not improve with multiple states measured for the simulated sensors in Figure 4-7a. This is likely due to the fact that not measuring an additional state is better than using a sensor that gives an inaccurate reading of that state. Some states have a wide range of sensor readings, *i.e.* one sensor might not be sufficient to approximate the average temperature value. The accuracy of the measurement for those state improves when multiple sensors are used. Since in this case, the maximum number of sensors used is limited to six, with more states measured fewer sensors are used for each state. When six states are measured, for instance, only one sensor is used per state. The accuracy of the best predictions, with minimal RMSE and absolute error, does improve when more states are measured. Indicating that there

Table 4-3: Twelve sensor sets with minimal RMSE or error based on real sensor data for the number of sensors used and minimization criteria.

Minimization criteria	Total number of sensors	Number of sensors per state: 3	6	9	12	14	15	16	17	RMSE	Max absolute error: $\|\|e\|\|_\infty$
RMSE	1	0	0	0	0	0	1	0	0	0.23	1.6
$\|\|e\|\|_\infty$	1	0	0	0	0	0	1	0	0	0.46	1.17
RMSE	2	0	0	0	0	0	2	0	0	0.18	1
$\|\|e\|\|_\infty$	2	0	0	0	0	0	2	0	0	0.18	1
RMSE	3	0	0	0	0	0	3	0	0	0.14	1.16
$\|\|e\|\|_\infty$	3	1	0	0	0	0	2	0	0	0.17	1
RMSE	4	0	1	0	0	0	3	0	0	0.13	1.16
$\|\|e\|\|_\infty$	4	1	0	1	0	0	2	0	0	0.17	1
RMSE	5	1	1	0	0	0	3	0	0	0.13	1.16
$\|\|e\|\|_\infty$	5	3	0	0	0	0	2	0	0	0.17	0.97
RMSE	6	1	1	0	1	0	3	0	0	0.13	1.17
$\|\|e\|\|_\infty$	6	4	0	0	0	0	2	0	0	0.17	0.96

is a benefit in using multiple measured states. The increase in performance, however, is only small.

The results when using the actual sensor data are shown in Figure 4-7b. The results are similar, with the exception of the maximum values for RMSE, *i.e.* worst sensor sets. An explanation for this difference is the Gaussian distribution assumption made for the error, as shown in Figure 3-6. Some errors of the sensor set displayed in this figure have a large error in the positive direction, resulting in a wider Gaussian distribution in the negative direction as well. The averaged values for the RMSE in the case of the real data used will average out, while the larger errors that occur less often affect the maximum error that occurs when those states are considered. More states measured with the worst sensor sets lead to a higher maximum absolute error. Using the simulated data set with Gaussian errors assumed will average these effects out. Similar to the simulated data, the best cases show a slight improvement in accuracy when additional states are measured with an exception in case when six states are measured. As mentioned, one sensor is used per measured state when six states are measured. Some the the states can not accurately measure the state average with one sensor. The states are in those cases poorly estimated.



(a) **(b)**

Figure 4-7: Effect of additional state measurements on the prediction accuracy of the model. Using the predictive method with simulated sensor data in (a) and using real sensor data in (b). Each black cross is with a different sensor combination, simulated over three days with an one hour prediction. The x-axis displays the number of states that are measured by at least one sensor. The total number of sensors used ranges from one to six.

Greedy method

The results for a greedy optimization run are shown in Figure 4-8. The results are comparable with the previous method, in the sense that the additional state measurements have little effect. The total energy consumption goes down when multiple sensors are added, although this effect is very small. The different sensors selected can be more clearly seen in Table 4-4. Similar to the predictive method, a sensor is added to the air temperature first. This also shows the biggest decrease in energy consumption. The decrease in energy consumption coincides with increased discomfort, although the discomfort bound of 0.5 is not violated for any of the cases. This is due to the hard constraint set on the temperature in (3-35c), which is incorporated into the objective function by the *sqp* algorithm when violated. The decreased energy consumption could be more related to the increased discomfort levels. The greedy algorithm is only able to the solutions on its current path, so other more optimal solutions could exist.

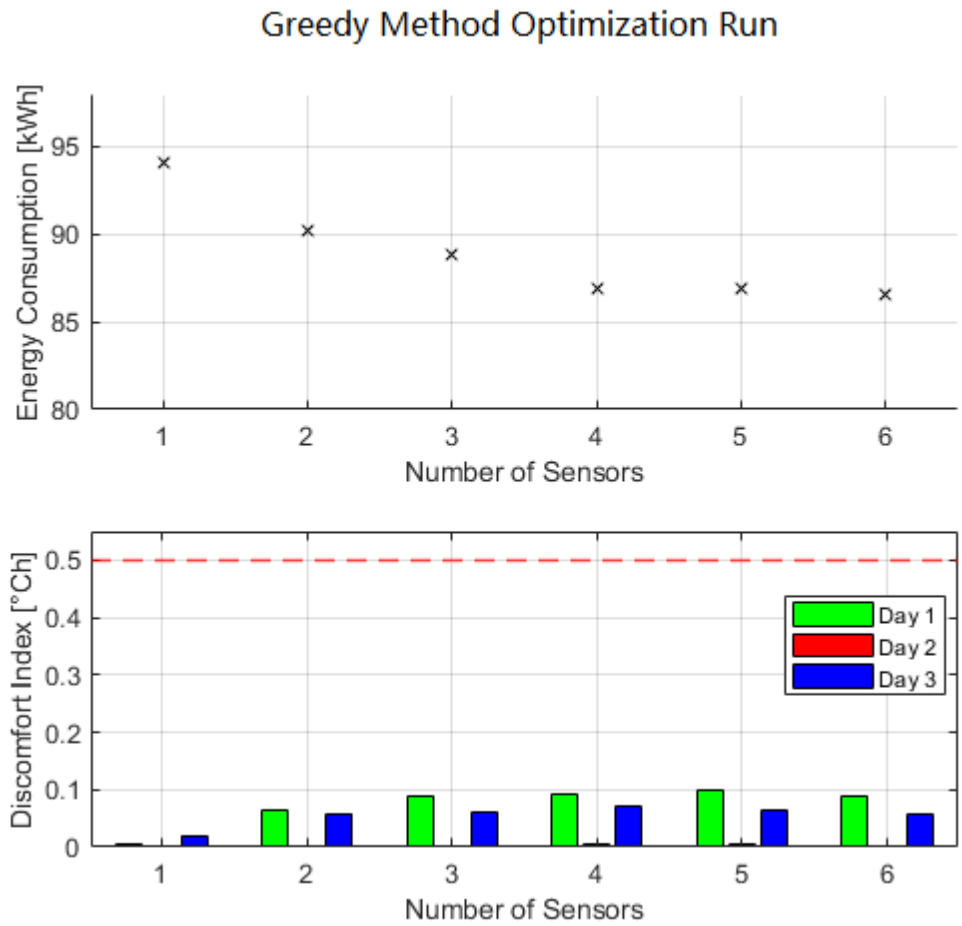


Figure 4-8: Minimum energy consumption for different amounts of sensors added by the greedy method over a three-day simulation period. The discomfort index for each day is shown.

Table 4-4: Greedy method results with hard constraints on the temperature.

	Number of sensors for each state:									
Iteration	3	6	9	12	14	15	16	17	Energy [kWh]	Difference
1	0	0	0	0	0	1	0	0	94.1	-
2	0	0	0	0	0	2	0	0	90.2	-4.14%
3	0	0	0	0	1	2	0	0	88.8	-1.55%
4	0	1	0	0	1	2	0	0	86.9	-2.14%
5	0	2	0	0	1	2	0	0	86.8	-0.12%
6	0	2	0	0	1	2	0	1	86.5	-0.36%

4-3 Conclusions

The goal of this case study is to determine the effect of additional state measurements on the control performance. Multiple methods have been constructed, to be able to compare the results. When the predictive method results are considered, the effect of sensors in the worst cases average out in the simulated data, which assumes a Gaussian distribution in the data. Sensor placement within a building element is important, adding a sensor that is not able to accurately approximate the average temperature makes the predictions worse. It is interesting to see that both using the simulated data and sensor data only show a slight improvement in prediction accuracy when additional states are measured. Using state measurements with a sensor set that is not sufficient to approximate the average value leads to worse prediction accuracy. The improvement in prediction accuracy is negligible after three sensors. Adding more does not result in better performance. The impact on energy consumption is limited. The level of discomfort has a larger effect on the total energy consumption. The greedy method, in its current form, might not be sufficient to determine the optimal sensor selection. The algorithm chooses its next sensor only based on the energy consumption and could choose a sensor location that prioritizes energy consumption over thermal comfort. A stricter bound, or more comparative discomfort levels should be incorporated during each iteration to fully compare the effect of the sensors on the energy consumption.

Chapter 5

Conclusion and Recommendations

5-1 Conclusion

The research in this thesis was carried out based on the CONVERGE building, located in The Green Village in Delft, which is a test-bed for energy-saving building systems and design. To fill in the research gap on using multiple building component measurements, this research set out to answer two questions:

1. *What is the optimal sensor selection for building temperature control?*
2. *Is there a benefit for control performance if additional building elements are measured?*

To answer these questions, a nonlinear building model has been implemented to simulate the temperature evolution of the different states. The overall model showed a good match to collected sensor data of the building. For the state observer, is chosen for an extended Kalman filter. This filter was chosen for its good estimation accuracy compared to its relatively low computation cost. To use the additional state information to its maximum potential, a nonlinear Model Predictive Control (MPC) has been chosen to simulate the effect on control performance. The number of control inputs was limited to two and a control update step of one hour is used to reduce computation time. The overall system was able to control the temperature to within the specified bounds.

Two methods have been adopted to determine the effect: a predictive and a greedy method. The first method used the RMSE and maximum absolute error of the predicted temperature one hour ahead compared to the actual temperature. The second method incorporated the MPC to add additional sensors in a greedy manner, based on the total energy consumption over a three-day simulation period. Since there is no direct comparison with previous work, the results of these methods can only be compared to each other. With the setup and case study complete, the research questions could be answered.

The optimal sensor selection for building temperature control

The air temperature sensors had the largest impact on both the prediction accuracy and

energy consumption. If one or two sensors are used, sensors measuring the air temperature would be chosen first. The improvement in prediction accuracy is negligible after three sensors. The greedy method showed similar results for the sensor selection. The air temperature sensors have the most added benefit. When more than two or three sensors are used, sensors placed on the ceiling or floor showed some improvement in performance.

The added benefit of additional state measurements for control performance

The current work shows some increase in performance with addition state measurements. The results of the prediction method indicate a higher temperature prediction accuracy over the next hour, although the relatively small increase might not be worth the cost of additional sensors. Similar results are obtained when considering energy consumption. If additional states are added, might be opted for the ceiling or floor. For all of the methods, little improvement is gained after adding more than three sensors. To fully test the potential of multiple state measurements, the greedy method would need to be modified. There is a chance that the greedy method chooses sensors that actually perform worst in terms of temperature sensing but that do result in lower energy consumption. The results showed an increase in thermal discomfort simultaneously with a decrease in energy consumption. All the building temperatures remained well within the specified comfort bounds, but this should be taken into account during potential next steps for this research.

5-2 Recommendations and Future Work

During this research, several assumptions and simplifications have been made. To validate the results of this work, these would need to be addressed. There are a number of areas in which this method could be extended in the following research. These have not been implemented in this research, due to time considerations. Some of the possible next steps for this work are listed down below.

- **Extend simulation period:** Both methods only optimize the sensor set over a period of three days in April. This has been done to reduce the computation time, but such a short window might not be representative for the whole year. More sensor data would need to be collected over a whole year, also to determine the validity of the model under different circumstances. The effect of the selected sensors could change the results of the used optimization methods if more days are considered.
- **Changes to the building model:** Some simplifications have been made to the building model to reduce complexity and computation time. One control input, for example, is the energy added directly added or removed from the air. In reality, supply airflow rate, supply air temperature, heat recovery, and outside air temperature also need to be considered. The building elements could also be divided into smaller sections, relaxing the assumption of uniform temperature in each component, *i.e.* using a zonal model. As seen from the sensor data, local temperature differences do exist.
- **Different control parameters:** In this research, a nonlinear predictive controller has been used, with one set of parameters. Increasing the control update step, changing weights or relaxing constraints all have an effect on the energy consumption of the system. The changes with the largest effect on energy consumption seems to be the constraint. With the greedy algorithm used in this research, the decrease in energy consumption seems to be related to the increase in thermal discomfort. A next step would be to incorporate the bound on thermal discomfort in the MPC constraints. This would result in a better comparison between the different sensor selections, as all of the set would adhere to the same maximum level of discomfort.
- **Adding disturbance uncertainties:** All the disturbances, such as solar intensity, ambient temperature, and occupancy, have been considered to be known of the prediction horizon. In reality, this will not be the case. The uncertainties in the disturbances will have an effect on the prediction accuracy, as well as the control performance.
- **Different control methods:** Besides adjusting the parameters of the current controller, an entirely different method for control could also be used. When the disturbance uncertainties are added, for instance, a stochastic model predictive controller might be more suitable. Linear or a combination of linear and nonlinear, model predictive control could be implemented to reduce computation time.
- **Validation with other methods:** One of the biggest assumptions made during this research, is the Gaussian distribution assumption in the sensor error. As shown by the predictive method in Section 4-2, some discrepancy exists in the results when simulated data or real sensor data is used. Using a method that does not require a model of the building is particularly interesting. Of these, GP model predictive control based on [28] extended with control and sensor selection, and SINDYc, Sparse Identification

of Nonlinear Dynamics with Control, are noted as interesting possible candidates for future research. The former has been used before, to select sensors with the largest amount of information about the temperature distribution within a large room. It has not been applied to sensor selection considering additional states, or include the control performance. The latter has, to the best of the author's knowledge, not been applied to a building, and by extension the sensor selection in buildings. It has shown great interest over the past few years and promising results in other applications.

Appendix A

Building Dimensions and Properties

Table A-1: Dimensions and properties of the CONVERGE building described in Chapter 3

	Abbreviation	Description	Value
Dimension	L	Building length	22.5 m
	W	Building width	13.5 m
	H	Building height	5.2 m
Orientation	γ_{south}	South wall	0°
	γ_{west}	West wall	90°
	γ_{north}	North wall	180°
	γ_{east}	East wall	-90°
	β_{walls}	Inclination walls	90°
	β_{top}	Inclination roof	0°
	latitude	-	51.996°
	longitude	-	4.378°
Walls	t_{glass}	Thickness	0.01 m
	ρ	Density	2470 kg/m^3
	c_p	Heat capacity	800 J/g· K
	ϵ	Emissivity	0.78
	ϵ_{low}	Emissivity of coated low-ϵ glass	0.14
	α	Absorptance	0.085
	τ	Transmissivity	0.81
Raised floor	t_{rf}	Thickness	0.038 m
	c_p	Heat capacity	707 J/K
	ρ	Density	1550 kg/m^3
	α	Absorptance	0.1
Concrete floor	t_{gf}	Thickness	0.225 m
	c_p	Heat capacity	840 J/g· K
	ρ	Density	2000 kg/m^3
	α	Absorptance	0.1
	$t_{\text{insulation}}$	Insulation thickness	0.008 m
	$\kappa_{\text{insulation}}$	Thermal conductivity	0.0213 W/m· K
Roof	t	Thickness	0.01 m
	c_p	Heat capacity	1800 J/g· K
	ρ	Density	1050 kg/m^3
	α	Absorptance	0.87
	ϵ	Emissivity	0.93
	$t_{\text{insulation}}$	Insulation thickness	0.18 m
	$\kappa_{\text{insulation}}$	Thermal conductivity	0.21 W/m· K
Ceiling	t	Thickness	0.003 m
	ρ	Density	7850 kg/m^3
	c_p	Heat capacity	500 J/g· K
Air	μ	Dynamic viscosity	1.80×10^{-5} kg/m· s
	c_p	Heat capacity	1006 kJ/kg· K
	κ	Thermal conductivity	0.0255 W/m· K
	ρ	Density	1.229 kg/m^3
Argon	μ	Dynamic viscosity	2.18×10^{-5} kg/m· s
	c_p	Heat capacity	521.7 kJ/kg· K
	κ	Thermal conductivity	0.017 W/m· K
	ρ	Density	1.7125 kg/m^3

Appendix B

Table with Sensor Set Data

Only the sensor sets with the minimal Root Mean Square Error (RMSE) are shown, as the table would be to large otherwise. These are also the sets used in the optimization methods.

Table B-1: Error data for the sensor sets with the lowest RMSE for a given number of sensors, in [°C].

		Number of sensors in a set:								
		1	2	3	4	5	6	7	8	9
State 3	RMSE	0.98	0.32	0.33	0					
	μ_{v_S}	0.17	-0.14	-0.058	0					
	σ_{v_S}	0.96	0.29	0.32	0					
State 6	RMSE	1.29	0							
	μ_{v_S}	-0.70	0							
	σ_{v_S}	1.09	0							
State 9	RMSE	0.42	0							
	μ_{v_S}	-0.28	0							
	σ_{v_S}	0.32	0							
State 12	RMSE	0.64	0							
	μ_{v_S}	0.23	0							
	σ_{v_S}	0.60	0							
State 14	RMSE	0.27	0.14	0.09	0					
	μ_{v_S}	-0.22	0.027	0.072	0					
	σ_{v_S}	0.17	0.13	0.056	0					
State 15	RMSE	0.24	0.12	0						
	μ_{v_S}	0.016	-0.0081	0						
	σ_{v_S}	0.24	0.12	0						
State 16	RMSE	0.37	0.17	0.085	0.064	0.051	0.042	0.050	0.047	0
	μ_{v_S}	-0.18	0.0011	0.027	0.019	-0.016	-0.013	0.00032	0.022	0
	σ_{v_S}	0.33	0.17	0.081	0.061	0.048	0.040	0.049	0.041	0
State 17	RMSE	0.47	0.15	0.042	0.058	0.046	0.021	0.044	0.059	0
	μ_{v_S}	-0.41	-0.10	0.019	-0.042	0.034	-0.0098	0.030	0.052	0
	σ_{v_S}	0.22	0.11	0.037	0.040	0.032	0.018	0.032	0.028	0

Bibliography

[1] J. Agnarsson, M. Sunde, and I. Ermilova. *Parallel Optimization in Matlab*. PhD thesis, Uppsala University, 01 2013.

[2] K. Ahmed, J. Kurnitski, and Olesen B. Data for occupancy internal heat gain calculation in main building categories. *Data in Brief*, 15:1030–1034, 12 2017.

[3] G.S. Bakken and D.M. Gates. Linearized heat transfer relations in biology. *Science*, 183:976–977, 3 1974.

[4] F. Belic, Z. Hocenski, and D. Sliskovic. HVAC control methods - A review. In *2015 19th International Conference on System Theory, Control and Computing, ICSTCC 2015 - Joint Conference SINTES 19, SACCS 15, SIMSIS 19*, pages 679–686. Institute of Electrical and Electronics Engineers Inc., nov 2015.

[5] I.H. Bell, J. Wronski, S. Quoilin, and Lemort V. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library coolprop. *Industrial and Engineering Chemistry Research*, 53:2498–2508, 2 2014.

[6] T.J. Ceha. Hierarchical model predictive control in building climate systems for passive energy sources. Master's thesis, TU Delft Center for Systems and Control, 2021. Available at: http://resolver.tudelft.nl/uuid:8e221205-139f-48ea-9d9a-575396dc3901.

[7] Y.L. Chen and J. Wen. The selection of the most appropriate airflow model for designing indoor air sensor systems. *Building and Environment*, 50:34–43, apr 2012.

[8] S.W. Churchill and H.H.S. Chu. Correlating equations for laminar and turbulent free convection from a vertical plate. *International Journal of Heat and Mass Transfer*, 18(11):1323–1329, 1975.

[9] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena. Ten years of genetic fuzzy systems: Current framework and new trends. In *Fuzzy Sets and Systems*, volume 141, pages 5–31. North-Holland, jan 2004.

[10] Y. Ding, L. Wang, Y. Li, and D. Li. Model predictive control and its application in agriculture: A review, 2018.

[11] A. I. Dounis, P. Tiropanis, A. Argiriou, and A. Diamantis. Intelligent control system for reconciliation of the energy savings with comfort in buildings using soft computing techniques. *Energy and Buildings*, 43(1):66–74, jan 2011.

[12] Z. Du, P. Xu, X. Jin, and Q. Liu. Temperature sensor placement optimization for VAV control using CFD-BES co-simulation strategy. *Building and Environment*, 85:104–113, 2015.

[13] J.A. Duffie and W.A. Beckman. *Solar engineering of thermal processes*, volume 4. John Wiley and Sons, Inc., 2013.

[14] C. Ekici and I. Atilgan. A comparison of suit dresses and summer clothes in the terms of thermal comfort. *Journal of Environmental Health Science and Engineering*, 11(1), dec 2013.

[15] European standards committee. EN 16798-1:2015 Indoor Environmental Input Parameters for Design and Assessment of Energy Performance of Buildings Addressing Indoor Air Quality, Thermal Environment, Lighting and Acoustics. Technical report, EPB standard, 2015.

[16] H. Fang, R. Sharma, and R. Patil. Optimal sensor and actuator deployment for HVAC control system design. *Proceedings of the American Control Conference*, pages 2240–2246, 2014.

[17] M.P. Fanti, A.M. Mangini, M. Roccotelli, F. Iannone, and A. Rinaldi. A natural ventilation control in buildings based on co-simulation architecture and Particle Swarm Optimization. *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*, pages 2621–2626, 2017.

[18] L. Feng, H. Li, and Y. Zhi. Greenhouse cfd simulation for searching the sensors optimal placements. pages 504–507, 08 2013.

[19] A. Figueiredo, J. Figueira, R. Vicente, and R. Maio. Thermal comfort and energy performance: Sensitivity analysis to apply the Passive House concept to the Portuguese climate. *Building and Environment*, 103:276–288, 2016.

[20] A.D. Fontanini, U. Vaidya, and B. Ganapathysubramanian. A methodology for optimal placement of sensors in enclosed environments: A dynamical systems approach. *Building and Environment*, 100:145–161, 2016.

[21] Y. Fu, M. Sha, C. Wu, A. Kutta, A. Leavey, C. Lu, H. Gonzalez, W. Wang, B. Drake, Y. Chen, and P. Biswas. Thermal modeling for a HVAC controlled real-life auditorium. *Proceedings - International Conference on Distributed Computing Systems*, pages 73–82, 2014.

[22] S. Goyal, P. Barooah, and T. Middelkoop. Experimental study of occupancy-based control of HVAC zones. *Applied Energy*, 140:75–84, feb 2015.

[23] A. Hesaraki and S. Holmberg. Demand-controlled ventilation in new residential buildings: Consequences on indoor air quality and energy savings. *Indoor and Built Environment*, 24(2):162–173, apr 2015.

[24] F.P. Incropera and D.P. DeWitt. *Fundamentals of Heat and Mass Transfer*, volume 7. John Wiley and Sons, Inc., 2002.

[25] S.A. Kalogirou. Chapter two - environmental characteristics. In S.A. Kalogirou, editor, *Solar Energy Engineering*, pages 49–762. Academic Press, Boston, 2009.

[26] KNMI. Knmi - bodemtemperaturen. Accessed on 2022-02-05.

[27] J. Koeln, B. Keating, A. Alleyne, C. Price, and B.P. Rasmussen. Multi-zone temperature modeling and control. In *Advances in Industrial Control*, number 9783319684611 in Advances in Industrial Control, pages 139–166. Springer International Publishing, 2018.

[28] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(February 2008):235–284, 2008.

[29] S. Lee, I. Lee, U. Yeo, R. Kim, and J. Kim. Optimal sensor placement for monitoring and controlling greenhouse internal environments. *Biosystems Engineering*, 188:190–206, 2019.

[30] R. Letan, V. Dubovsky, and G. Ziskind. Passive ventilation and heating by natural convection in a multi-storey building. *Building and Environment*, 38(2):197–208, 2003.

[31] D.B. Lu and D.M. Warsinger. Energy savings of retrofitting residential buildings with variable air volume systems across different climates. *Journal of Building Engineering*, 30(March):101223, 2020.

[32] Y. Ma, J. Matusko, and F. Borrelli. Stochastic model predictive control for building HVAC systems: Complexity and conservatism. *IEEE Transactions on Control Systems Technology*, 23(1):101–116, 2015.

[33] M. Maasoumy and A. Sangiovanni-Vincentelli. Total and peak energy consumption minimization of building hvac systems using model predictive control. *IEEE Design and Test of Computers*, 29(4):26–35, 2012.

[34] M. Maasoumy, Q. Zhu, C. Li, F. Meggers, and A. Sangiovanni-Vincentelli. Co-design of control algorithm and embedded platform for building HVAC systems. *2013 ACM/IEEE International Conference on Cyber-Physical Systems, ICCPS 2013*, pages 61–70, 2013.

[35] N. Mahyuddin and H.B. Awbi. A review of CO2 measurement procedures in ventilation research. *International Journal of Ventilation*, 10(4):353–370, 2012.

[36] D. Manjarres, A. Mera, E. Perea, A. Lejarazu, and S. Gil-Lopez. An energy-efficient predictive control for HVAC systems applied to tertiary buildings based on regression techniques. *Energy and Buildings*, 152:409–417, oct 2017.

[37] G. Mantovani, L. Ferrarini, and S. Member. Temperature Control of a Commercial Building With Model Predictive Control Techniques. *IEEE Transaction on industrial electronic*, 62(4):2651–2660, 2015.

[38] MathWorks. Understanding kalman filters, part 5: Nonlinear state estimators video - matlab. Accessed on 2022-02-25.

[39] The MathWorks. Optimization toolbox™ user's guide - version 9.3. Natick, Massachusetts, 2022.

[40] L. Mazzarella and J. Hogeling. REHVA Journal 02/2018 - CEN Standard EN 16798-3:2017 on ventilation for non-residential buildings: performance requirements.

[41] A. McGibney, D. Pusceddu, S. Rea, D. Pesch, M. Geron, and M. Keane. A methodology for sensor modeling and placement optimization to support temperature monitoring. *BuildSys 2012 - Proceedings of the 4th ACM Workshop on Embedded Systems for Energy Efficiency in Buildings*, pages 88–90, 2012.

[42] Mckinnon. Components of Your HVAC System | mackinnon heating and cooling. Accessed on 2020-08-27.

[43] H. Mirinejad, K.C. Welch, and L. Spicer. A review of intelligent control techniques in HVAC systems. In *2012 IEEE Energytech, Energytech 2012*, 2012.

[44] S.R. Mohandes, X. Zhang, and A. Mahdiyar. A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing*, 340:55–75, may 2019.

[45] M. Nowak and A. Urbaniak. Utilization of intelligent control algorithms for thermal comfort optimization and energy saving. In *Proceedings of the 2011 12th International Carpathian Control Conference, ICCC'2011*, pages 270–274, 2011.

[46] F. Oldewurtel, A. Parisio, C.N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, 45:15–27, feb 2012.

[47] B.W. Olesen. Indoor environmental input parameters for design and assessment of energy performance of buildings-addressing indoor air quality, thermal environment, lighting and acoustics. 2006.

[48] B.W. Olesen. The philosophy behind EN15251: Indoor environmental criteria for design and calculation of energy performance of buildings. *Energy and Buildings*, 39(7):740–749, 2007.

[49] D. Pelleg and A. Moore. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *In Proceedings of the 17th International Conference on Machine Learning*, pages 727—-734, 2000.

[50] R. Perez, P. Ineichen, R. Seals, J. Michalsky, and R. Stewart. Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy*, 44(5):271–289, 1990.

[51] R. E. Perez and K. Behdinan. Particle swarm approach for structural design optimization. *Computers and Structures*, 85(19-20):1579–1588, oct 2007.

[52] M.I. Ribeiro. Gaussian probability density functions: Properties and error characterization. *Instituto Superior Tcnico, Lisboa, Portugal, Tech. Rep*, 2004.

[53] Rijksdienst voor Ondernemend Nederland. Monitor energiebesparing gebouwde omgeving 2017, 2018.

[54] T. Schulze, D. Gürlich, and U. Eicker. Performance assessment of controlled natural ventilation for air quality control and passive cooling in existing and new office type buildings. *Energy and Buildings*, 172:265–278, aug 2018.

[55] B. Sun, P.B. Luh, Q.S. Jia, Z. Jiang, F. Wang, and C. Song. Building energy management: Integrated control of active and passive heating, cooling, lighting, shading, and ventilation systems. *IEEE Transactions on Automation Science and Engineering*, 10(3):588–602, jul 2013.

[56] Technical Committee CEN/TC 156. EN 13779: Ventilation for non-residential buildings - Performance requirements for ventilation and room-conditioning systems. *European Standard*, page 5, 2007.

[57] U. Vaidya, R. Rajaram, and S. Dasgupta. Actuator and sensor placement in linear advection PDE with building system application. *Journal of Mathematical Analysis and Applications*, 394(1):213–224, 2012.

[58] V. Verhaegen and M. Verdult. *Filtering and System Identification : A Least Squares Approach*. Cambridge University Press, 2007.

[59] B. Vidrih, C. Arkar, and S. Medved. Generalized model-based predictive weather control for the control of free cooling by enhanced night-time ventilation. *Applied Energy*, 168:482–492, apr 2016.

[60] U. Von Luxburg. A Tutorial on Spectral Clustering. Technical Report 4, Max Planck Institute for Biological Cybernetics, 2007.

[61] E. A. Wan and R. Van Der Merwe. The unscented Kalman filter for nonlinear estimation. *IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, AS-SPCC 2000*, pages 153–158, 2000.

[62] Y. Wang, J. Kuckelkorn, F.Y. Zhao, H. Spliethoff, and W. Lang. A state of art of review on interactions between energy performance and indoor environment quality in Passive House buildings. *Renewable and Sustainable Energy Reviews*, 72(October 2016):1303–1319, 2017.

[63] Y. Wang, A. Shukla, and S. Liu. A state of art review on methodologies for heat transfer and energy flow characteristics of the active building envelopes. *Renewable and Sustainable Energy Reviews*, 78(September 2016):1102–1116, 2017.

[64] J. Wilde. Indoor Thermal Standards: ASHRAE 55 vs. ISO 7730 I SimScale. Accesed on 2021-07-10.

[65] J. Xu, P.B. Luh, W.E. Blankson, R. Jerdonek, and K. Shaikh. An optimization-based approach for facility energy management with uncertainties. *HVAC and R Research*, 11(2):215–237, 2005.

[66] D. Yoganathan, S. Kondepudi, B. Kalluri, and S. Manthapuri. Optimal sensor placement strategy for office buildings using clustering algorithms. *Energy and Buildings*, 158:1206–1225, 2018.

[67] J. Yun and J. Kim. Deployment support for sensor networks in indoor climate monitoring. *International Journal of Distributed Sensor Networks*, 2013, 2013.

Glossary

List of Acronyms

TU Delft	Delft University of Technology
HVAC	Heating, Ventilation, and Air Conditioning
VAF	Variance Accounted For
PPD	Predicted Percentage Dissatisfied
PMV	Predicted Mean Vote
MPC	Model Predictive Control
RMSE	Root Mean Square Error
CFD	Computational Fluid Dynamics
GA	Genetic Algorithm

Nomenclature

$\dot{q}_{max}$	Maximum heating power [7 kW]
$\dot{q}_{min}$	Maximum cooling power [-7 kW]
$\dot{q}_{j-i}$	Energy flow from component "j"to component "i"[W]
$c_{p,ceil}$	specific heat capacity ceiling [J/kg· K]
$c_{p,gf}$	specific heat capacity concrete ground floor [J/kg· K]
$c_{p,gl}$	specific heat capacity glass [J/kg· K]
$c_{p,rf}$	specific heat capacity raised floor [J/kg· K]
$c_{p,roof}$	specific heat capacity roof [J/kg· K]
m_{ceil}	Mass ceiling [kg]
m_{gf}	Mass concrete ground floor [kg]
m_{gl}	Mass glass [kg]
m_{rf}	Mass raised floor [kg]
m_{roof}	Mass roof [kg]
α	Absorptance
β	Slope of surface
β_{blinds}	Solar blinds, open-closed position from one to zero
δ	Solar declination, angle equator to sun plane

$\dot{q}_{\text{conv}}$	Convection energy flow [W]
$\dot{q}_{\text{hvac}}$	Energy supplied by HVAC [W]
$\dot{q}_{\text{occ}}$	Energy generation from occupants [W]
$\dot{q}_{\text{rad}}$	Radiation energy flow [W]
$\dot{q}_{\text{sol-i}}$	Energy flow from the sun to component "i"[W]
ϵ	Emissivity
ϵ_{low}	Emissivity of low-ϵ coating
γ	Azimuth surface
κ	Thermal conductivity [W/m $\cdot$ K]
λ	Objective function weight for MPC
ω	Hour angle
ϕ	Latitude
σ	Stefan-Boltzmann constant [5.67×10^{-8} W/m^{2}K^4]
τ	Transmissivity
θ_z	Zenith angle
A	Surface area [m^2]
A_u	Unshaded floor area [m^2]
D_{ind}	Discomfort index [$^\circ Ch$]
H	Component height [m]
h	Convective heat transfer coefficient [W/m$^2 \cdot$ K]
I	Solar incidences [W/m^2]
I_b	Direct beam irradiation [W/m^2]
I_d	Diffusive irradiation [W/m^2]
I_g	Ground reflected irradiation [W/m^2]
I_{bn}	Irradiance normal to surface [W/m^2]
I_{dh}	diffuse horizontal irradiation [W/m^2]
L	Component length [m]
L_C	Characteristic length
N	Prediction horizon
n_{occ}	Number of occupants
Nu	Nusselt number
P	Surface perimeter [m]
p_{air}	Air pressure [Pa]
P_{occ}	Average thermal heat generation per occupant [120W]
Pr	Prandtl number
Ra	Rayleigh number
Re	Reynolds number
T_{ref}	Reference temperature [°C]
T_{amb}	Ambient air temperature, related to convection [°C]
T_{grd}	Ground temperature [°C]

T_i	Temperature of component "i"[°C]
T_{sky}	Sky temperature, related to radiation heat transfer [°C]
v	Total measurement noise [$°C$]
$v_{\mathcal{S}}$	Sensor set error [$°C$]
v_m	Sensor measurement noise [$°C$]
W	Component width [m]
w	Process noise [$°C$]
m	mass [kg]