

# Linear State-Space System Identification for Automatic Greenhouse Climate Control

L. van der Lely

Master of Science Thesis



# **Linear State-Space System Identification for Automatic Greenhouse Climate Control**

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Robotics at Delft University of  
Technology

L. van der Lely

December 7, 2022

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



DELFT UNIVERSITY OF TECHNOLOGY  
DEPARTMENT OF  
COGNITIVE ROBOTICS (CoR)

The undersigned hereby certify that they have read and recommend to the Faculty of  
Mechanical, Maritime and Materials Engineering (3mE) for acceptance a thesis  
entitled

LINEAR STATE-SPACE SYSTEM  
IDENTIFICATION FOR AUTOMATIC  
GREENHOUSE CLIMATE CONTROL

by

L. VAN DER LELY

in partial fulfillment of the requirements for the degree of  
MASTER OF SCIENCE ROBOTICS

Dated: December 7, 2022

Supervisor(s):

---

Dr. M. Mazo Espinosa (DCSC)

---

Dr.-ing J. Kober (COR)

Reader(s):

---

Prof. Dr. Ir. T. Keviczky (DCSC)

---

Dr. L. Laurenti (DCSC)



---

# Abstract

Over nine billion people have to be fed fresh and healthy food in 2050 according to United Nations. This puts pressure on the horticulture sector, responsible for a large portion of the world's food production. Greenhouse crop growing has many benefits, but it also comes with its own challenges. There is a decreasing number of knowledgeable growers who are responsible for resource-efficient climate control. Therefore, the need for more automation in climate control arises. Additionally, the greenhouse horticultural sector is still a large consumer of gas and electricity, creating a demand for more sustainable growth. The literature has shown that automatic optimal control algorithms are able to make better use of resources and can even outperform growers in terms of resource efficiency.

These automatic optimal control methods either make use of prediction models based on first principles, such as mass and energy balances, or make use of data-driven control techniques, excluding model assumptions. While applications of both methods show promising results, few are made adaptive to the greenhouse-crop system, assuming the system to be time-invariant. That is, the prediction models or techniques used in the controllers are not updated to adapt better to the evolving greenhouse-crop system. This thesis sets a first step towards this on-line system adaptation: it investigates the possibility to use a linear state-space model to estimate the non-linear greenhouse-crop systems dynamics around an operating point, for the purpose of automatic optimal climate control. The main reason to use this linear estimation of the greenhouse-crop dynamics, is that corresponding non-linear models from the literature are not suitable for optimal control, because of computational costs and because solvers can only give local optima. Also, the linear state-space system can easily be updated using data, which makes it suitable for on-line system adaptation. In this work, a linear state-space model is identified and is implemented in an automatic optimal controller based on Model Predictive Control (MPC).

To create the input-output data for the system identification and to test control methods, a ground-truth simulation based on a fine-grained non-linear model derived from first principles is built. This ground-truth simulator is calibrated using data from a real-life tomato-growing experiment, namely the Autonomous Greenhouse Challenge.

A linear state-space system with model order 10 is specified and is identified using the parameter estimation method (PEM). The identified system can be used as a prediction model

in the MPC-based controller for making predictions 1 day ahead. This is concluded after the validation of simulation experiments where different input frequencies and amplitudes are included. Moreover, the system can only be used around a specific operating point. This operating point is chosen to be around the period where the crop states reach steady-state (day 78), because here the non-linear dynamics are better represented by a linear approximation than in other periods where the dynamics evolve more.

The proposed automatic optimal control architecture is composed of an MPC-based controller with a linear state-space model as a predictor. This proposed control method is different from earlier control attempts, such as DeePC, as it does not split the control problem into two layers, but finds the optimal control inputs directly from the predicted crop growth. Also, it does not use the input-output trajectories directly for predictions, but it first finds an appropriate linear system representation and uses that to make predictions. The proposed controller is built with a prediction horizon of 1 day and a control horizon of 30 minutes. The included objective is profit; yield from selling harvested fruits and costs from using resources based on applied inputs. When the profit is based on the harvested dry weight and a relatively short prediction horizon is used, there is only an indirect incentive to grow new fruits and ripen the more mature fruits. For this, a hard constraint is added to make sure that the total dry weight is at least maintained for the complete prediction horizon. Moreover, a Luenberger state observer is added for state estimation. A simulation is performed for 14 days from day 78 onward.

The results of the deployed controller are compared to that of a baseline growing strategy, based on the inputs a grower applied in the Autonomous Greenhouse Challenge. The proposed controller did not outperform the baseline, but matched it in terms of profit ( $1.22 \text{ €/m}^2$ ). The attained total costs ( $0.29 \text{ €/m}^2$ ) were lower than that of the baseline ( $1.23 \text{ €/m}^2$ ), as was the yield ( $1.51 \text{ €/m}^2$  vs  $2.45 \text{ €/m}^2$ ) for the simulation period of 14 days. However, these numbers are deceiving; for longer growing periods the baseline will definitely outperform the controller as the controller did not suffice in maintaining the total fruit weight. The defined constraint on maintaining the total fruit weight was not achieved by the controller in the actual simulation. Therefore, it can be concluded that the identified linear system does not have sufficient prediction performance. The reason for this bad prediction performance could be that either the model order (10) was too low or that the experiments did not include enough variation of inputs and disturbances. A proposed on-line system identification control architecture could potentially improve this and is given as a recommendation.

---

# Table of Contents

<b>Acknowledgements</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1-1 Motivation . . . . .	1
1-2 Thesis research questions . . . . .	3
1-3 Thesis contributions . . . . .	4
1-4 Thesis outline . . . . .	4
<b>2 Greenhouse-Crop System</b>	<b>5</b>
2-1 The Greenhouse-Crop System . . . . .	5
2-1-1 The modern-day greenhouse . . . . .	5
2-1-2 The greenhouse-crop system . . . . .	6
2-2 Greenhouse climate . . . . .	7
2-2-1 Greenhouse climate: fundamentals . . . . .	7
2-2-2 Greenhouse climate: control inputs . . . . .	9
2-2-3 Greenhouse climate: feedback . . . . .	11
2-2-4 Greenhouse climate models: literature summary . . . . .	12
2-3 Crop growth . . . . .	12
2-3-1 Crop growth: fundamentals . . . . .	12
2-3-2 Crop growth: feedback and measurements . . . . .	13
2-3-3 Crop growth models: literature summary . . . . .	14
2-4 Model selection . . . . .	14

2-4-1	Climate model granularity . . . . .	14
2-4-2	Crop model granularity . . . . .	14
2-4-3	Model selection . . . . .	15
2-5	Chapter summary and conclusion . . . . .	15
<b>3</b>	<b>Greenhouse-Crop Ground-truth Simulation Model</b>	<b>17</b>
3-1	System description . . . . .	17
3-2	Crop mathematical model . . . . .	20
3-2-1	Crop growing dynamics . . . . .	20
3-2-2	Crop development stages . . . . .	21
3-3	Greenhouse climate mathematical model . . . . .	22
3-3-1	Greenhouse air temperature . . . . .	22
3-3-2	Greenhouse air absolute humidity . . . . .	24
3-3-3	Greenhouse air $CO_2$ concentration . . . . .	25
3-4	Greenhouse-crop model parameters . . . . .	25
3-5	Calibration of the greenhouse-crop ground-truth simulation model . . . . .	25
3-5-1	AGC dataset and data pre-processing . . . . .	26
3-5-2	Calibration of the ground-truth crop model . . . . .	28
3-5-3	Calibration of the ground-truth greenhouse climate model . . . . .	28
3-5-4	Calibration interval . . . . .	29
3-5-5	Calibration parameter vector and bounds . . . . .	30
3-5-6	Solver . . . . .	31
3-5-7	Calibration results . . . . .	31
3-5-8	$CO_2$ -concentration calibration . . . . .	33
3-5-9	Physical meaning of the parameters . . . . .	35
3-6	Steady-state fruit weight . . . . .	35
3-7	Leaf and stem weight . . . . .	36
3-8	Sampling time . . . . .	37
3-9	Chapter summary and conclusion . . . . .	38

---

<b>4</b>	<b>Greenhouse Climate Control</b>	<b>39</b>
4-1	System identification of the linear state-space system . . . . .	39
4-1-1	Linear state-space system . . . . .	39
4-1-2	Data processing for system identification and making prediction . . . . .	40
4-1-3	Prediction-Error Method (PEM) . . . . .	41
4-1-4	Performance definition and validation of the identified system . . . . .	42
4-2	Problem definition . . . . .	43
4-2-1	Grower's objective . . . . .	43
4-2-2	Literature on greenhouse climate control . . . . .	43
4-2-3	Formal problem definition . . . . .	46
4-3	Proposed control architecture . . . . .	48
4-3-1	Overview . . . . .	49
4-3-2	MPC-based controller . . . . .	49
4-3-3	Luenberger state observer . . . . .	50
4-3-4	Control architecture algorithm . . . . .	51
4-4	Performance test setup . . . . .	51
4-5	Chapter summary and conclusion . . . . .	51
<b>5</b>	<b>Results</b>	<b>53</b>
5-1	System identification . . . . .	53
5-1-1	Fit of the identified system on the training dataset . . . . .	54
5-1-2	Fit of the identified system on the validation dataset . . . . .	55
5-1-3	Fit of the identified system on validation dataset with enhanced inputs . . . . .	56
5-1-4	Discussion of the system identification results . . . . .	57
5-1-5	Fit of the identified system on different operating points . . . . .	59
5-1-6	Final remarks on the system identification results . . . . .	59
5-2	Results of Greenhouse crop control . . . . .	60
5-2-1	Simulated output trajectories . . . . .	60
5-2-2	Inputs applied by the controller . . . . .	64
5-2-3	Resulting costs and profit . . . . .	66
5-3	Chapter summary and conclusion . . . . .	67

<b>6 Conclusion and Recommendations</b>	<b>69</b>
6-1 Conclusion . . . . .	69
6-2 Recommendations . . . . .	71
6-2-1 Limitations of the ground-truth simulator . . . . .	71
6-2-2 Boundaries of the system and input constraints . . . . .	71
6-2-3 Time-varying cost and price coefficients . . . . .	72
6-2-4 On-line system identification . . . . .	72
<b>Appendices</b>	<b>74</b>
<b>A Greenhouse-crop model for simulation purposes</b>	<b>75</b>
A-1 Mass flows crop model . . . . .	75
A-2 Ground-truth: energy and mass fluxes . . . . .	76
A-3 Ground-truth: model parameters . . . . .	77
A-4 Ground-truth calibration data windows . . . . .	81
A-5 Ground-truth simulation figures . . . . .	83
A-5-1 Ground-truth fruit dry weight per development stage . . . . .	83
A-5-2 Ground-truth season-wide simulations . . . . .	84
A-5-3 Ground-truth 3-day simulations . . . . .	85
<b>B System Identification Results</b>	<b>89</b>
B-1 Training dataset fits . . . . .	90
B-2 Validation dataset fits . . . . .	96
<b>Bibliography</b>	<b>98</b>
<b>Bibliography</b>	<b>99</b>
<b>Glossary</b>	<b>103</b>
List of Acronyms . . . . .	103

---

## List of Figures

2-1	A schematic example of a Venlo-type greenhouse [1, Fig. 1.1]. . . . .	6
2-2	Energy and mass fluxes in greenhouse-crop systems [1, Fig. 1.3]. . . . .	7
2-3	Block diagram representation of the common crop growing processes. 1: photosynthesis, 2: growth respiration, 3: maintenance respiration, 4: accumulating assimilates, 5: assimilate partitioning, 6: accumulating biomass, 7: tomato harvest, 8: leaf harvest. $p$ represents the assimilates generated through photosynthesis, $g_r$ the assimilates used for growth respiration and $g$ the amount of assimilates converted to biomass. A part of the biomass is used for maintenance respiration $m$ . $h_1$ and $h_2$ describe the harvest of fruit and leaves, respectively [2]. . . . .	13
3-1	Block diagram representation of the greenhouse control problem [3, Fig. 1.]. . .	18
3-2	Flowchart of the fruit development model by Vanthoor (2011), based on Rabbinge and Ward (1989) [4]. . . . .	22
3-3	The measured harvested fruit dry weight and the simulated harvested fruit dry weight from the calibrated non-linear crop model, using the measured climate states.	29
3-4	Window on the 12th day. $Q_{srd}$ and $Q_{ven}$ are scaled down. . . . .	30
3-5	Season-wide temperature simulation of the calibrated ground-truth model, compared to the measured and outside temperature. . . . .	31
3-6	The measured harvested fruit dry weight and the simulated harvested fruit dry weight from the calibrated non-linear crop model, with the calibrated ground-truth model. . . . .	32
3-7	3-day temperature simulation of the calibrated ground-truth model starting from day 12, compared to the measured and outside temperature. . . . .	33
3-8	Season-wide $CO_2$ -concentration simulation of the calibrated ground-truth model including the adjusted $CO_2$ -dynamics, compared to the measured and outside $CO_2$ -concentration. . . . .	34

3-9	The measured harvested fruit dry weight and the simulated harvested fruit dry weight from the calibrated non-linear crop model, with the calibrated ground-truth model and with the adjusted $CO_2$ -dynamics. . . . .	35
3-10	The simulated leaf weight. . . . .	36
3-11	The air temperature for different sampling times, compared to the measured AGC data. The difference between the simulated temperature and the measured temperature in the AGC can be ignored. . . . .	37
4-1	A general MPC scheme, only the first input of $\mathbf{u}(t + k)$ is applied [5, Fig. 2.] . .	44
4-2	Proposed control architecture and simulation setup. . . . .	49
5-1	Fit of identified linear state-space system on the training dataset from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	54
5-2	Fit of identified linear state-space system on the validation dataset from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	55
5-3	Fit of identified linear state-space system on the validation dataset from day 76 onward, with higher frequency. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	56
5-4	Fit of identified linear state-space system on the validation dataset from day 76 onward, with larger amplitude. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	57
5-5	Simulated climate trajectories, generated by the profit controller. . . . .	61
5-6	Simulated and baseline mass trajectories for day 76 till day 92. . . . .	63
5-7	Inputs applied to the simulator by the controller. Note that both heating pipe temperature completely overlap, the same applies to the inputs of both windows. Also note that a screen input of zero means that it is completely opened. . . . .	64
5-8	Disturbances from the outside weather. . . . .	65
6-1	Proposed control architecture for including on-line system identification. . . . .	73
A-1	Window on the 72th day. $Q_{srd}$ and $Q_{ven}$ are scaled down. . . . .	81
A-2	Window on the 130th day. $Q_{srd}$ and $Q_{ven}$ are scaled down. . . . .	82
A-3	Window on the 149th day. $Q_{srd}$ and $Q_{ven}$ are scaled down. . . . .	82
A-4	Season-wide fruit dry weight per development stage simulation of the calibrated ground-truth model. . . . .	83
A-5	Season-wide absolute humidity simulation of the calibrated ground-truth model, compared to the measured and outside absolute humidity. . . . .	84

A-6	Season-wide $CO_2$ -concentration simulation of the calibrated ground-truth model, compared to the measured and outside $CO_2$ -concentration. . . . .	84
A-7	3-day absolute humidity simulation of the calibrated ground-truth model starting from day 12, compared to the measured and outside absolute humidity. . . . .	85
A-8	3-day $CO_2$ -concentration simulation of the calibrated ground-truth model starting from day 12, compared to the measured and outside $CO_2$ -concentration. . . . .	85
A-9	3-day temperature simulation of the calibrated ground-truth model starting from day 130, compared to the measured and outside temperature. . . . .	86
A-10	3-day absolute humidity simulation of the calibrated ground-truth model starting from day 130, compared to the measured and outside absolute humidity. . . . .	86
A-11	3-day $CO_2$ -concentration simulation of the calibrated ground-truth model starting from day 130, compared to the measured and outside $CO_2$ -concentration. . . . .	87
B-1	Fit of identified linear state space system on the training dataset of a second experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	90
B-2	Fit of identified linear state space system on the training dataset of a third experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	91
B-3	Fit of identified linear state space system on the training dataset of a fourth experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	92
B-4	Fit of identified linear state space system on the training dataset of a fifth experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	93
B-5	Fit of identified linear state space system on the training dataset of a sixth experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	94
B-6	Fit of identified linear state space system on the validation dataset from day 79 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	95
B-7	Fit of identified linear state space system on the validation dataset from day 85 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	96
B-8	Fit of identified linear state space system on the validation dataset from day 4 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	97
B-9	Fit of identified linear state space system on the validation dataset from day 145 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit. . . . .	98



---

## List of Tables

2-1	Description of the information flows for the greenhouse-crop system. . . . .	8
3-1	Greenhouse-crop states $\mathbf{x}$ . . . . .	19
3-2	Greenhouse-crop control inputs $\mathbf{u}$ . . . . .	19
3-3	Greenhouse-crop outputs $\mathbf{y}$ . . . . .	20
3-4	Greenhouse disturbances $\mathbf{d}$ . . . . .	20
4-1	Cost and price coefficients [6] . . . . .	48
5-1	Comparison of the resource usage and costs . . . . .	66
5-2	Comparison of the profit . . . . .	67
A-1	Overview of the mass flow rates for crop growth. The dynamics and model parameters are well described in [7] and [7, Tab. 9.1.], respectively. $*rg_X$ is linearly dependent on $T_{c24}$ [7, Eq. 9.28]. $**h_Y$ represents a non-instantaneous temperature-dependent inhibition, i.e. sub-optimal growth, see [7, Fig. 3.3.]. Physical constants are not considered model parameters. . . . .	76
A-2	Energy and mass fluxes of the greenhouse-crop model for simulation purposes. . . . .	76
A-3	Model parameters of the greenhouse-crop model for simulation purposes. . . . .	77
A-4	Calibrated Model parameters of the ground-truth crop model . . . . .	77
A-5	Calibrated Model parameters of the ground-truth greenhouse climate model . . . . .	80



---

# Notations and definitions

In this thesis the following notations and definitions are used:

1. Units are denoted in italic between square brackets [*unit*].
2. Matrices and vectors are denoted in bold, e.g. **A**.
3. An instance of a trajectory in continuous time  $t$  is denoted with brackets, e.g.  $x(t)$ . An instance in discrete time  $i$  is denoted with square brackets, e.g.  $x[i]$ .
4. All trajectories, i.e. a set of instances  $x(t)$  or  $x[i]$ , are denoted in bold, e.g. the trajectory of state  $x \in \mathbb{R}^n$  is denoted as **x**.
5. States **x**, inputs **u**, disturbances **d** and outputs **y** that belong to the linear state-space system are denoted with a  $L$  in subscript, e.g. **x<sub>L</sub>**, while the ones of the non-linear ground-truth model are denoted without a subscript.
6. Unless stated otherwise, all systems described are Linear Time-Invariant (LTI).



---

# Acknowledgements

For this thesis, I made the cross-over to the domain of Systems and Control. As an MSc. Robotics student, I had to learn Model Predictive Control and System Identification for practical use from scratch. This could not have been achieved without the help of my daily supervisor Dr. Manuel Mazo Espinosa and I would like to thank him for his help and patience during this learning process.

Especially, I wish to thank my parents for always putting my brother and me first throughout our childhood and for running the ‘hotel’, as we like to call it at home. Next, I’m grateful for the financial support that my grandmother and grandfather provided for my studies. Finally, I would like to thank Nienke for supporting me through thick and thin during this mentally demanding project.

The thesis you are about to read concludes my studies at Delft University of Technology. I could have not completed my studies without all the project-mates, mentors and professors; a big thank you to everybody I’ve worked with!

Delft University of Technology  
December 7, 2022

L. van der Lely



---

# Chapter 1

---

## Introduction

This first chapter explains the motivation behind this thesis work, the research questions, the thesis contributions and the thesis outline.

### 1-1 Motivation

The global population is increasing at a fast pace. United Nations projects that in 2050, more than nine billion people have to be fed with fresh and healthy food [8] [9]. Greenhouse horticulture will play a large role in meeting this demand, as it enables controlled, resource-efficient and sustainable crop growing in climates wherein it is impossible to grow outdoors. However, scaling up the number of greenhouses is a difficult challenge, because of the lack of skilled laborers and the decreasing number of knowledgeable growers. The growers can be seen as the brains of the greenhouse; they have to make very complex well-thought-out decisions to maximize production and minimize resource consumption of the greenhouse [6]. Today's growers determine the climate, irrigation and crop management strategies based on experience and define the setpoints for climate and irrigation control manually. Actuators, e.g. heating, irrigation and ventilation systems, then operate based on climate setpoints configured in a process computer, while sensors give feedback on measured climate data for the control loop. The most important climate states are the temperature, humidity,  $CO_2$ -concentration and solar radiation. The fruit and leaf weight are the states used to describe the growth of the tomato plant, which is the selected crop for this thesis, as it is most important for The Netherlands in terms of export size [10].

The scarcity of these so-called 'brains' of the greenhouse sparks the need for automated systems, in the form of automatic climate control. While actuation equipment is already quite advanced in terms of automation, automatic control algorithms in practice are lacking [1] [11]. One of the main reasons for this is a mismatch between reality and simulation. The processes that describe the greenhouse-crop system are complex and all attempts at greenhouse-crop

modeling involve non-linearities, making accurate modeling a difficult task. A. More complexity is introduced by the different time scales and time delays of the greenhouse-crop system [12]. This means that the greenhouse sub-system acts on faster (*minutes to hours*) dynamics than the much slower crop sub-system (*days to weeks*). A second reason for more automation in greenhouses is that greenhouses use large amounts of energy and gas, and emit  $CO_2$  gasses. The Dutch horticultural sector was responsible for 8% of the electricity usage [13] in 2017, from which only 7% comes from sustainable energy sources [14]. The newest climate policies of the Netherlands demand the greenhouse industry to reduce its gas emissions [15] by at least 25%. For the grower to still meet the production demand and do this as sustainably as possible, he or she must make smarter use of his or her resources. Automatic optimal climate control might provide a solution to more sustainable production.

Automatic optimal climate control found in literature can be divided into two types. First, model-based optimal control methods rely, as the name suggests, on an underlying model. An example is Model Predictive Control (MPC), which uses the model of the greenhouse-crop system to predict crop yield and resource use, and gives the optimal control inputs according to the specified objective function [16]. MPC was one of the strategies applied during the Second Edition of the Autonomous Greenhouse Challenge 2020, organized by the Wageningen University and Research Institute. The climate strategy created by the MPC controller was shown to be more efficient than a rule-based strategy provided by a professional grower [6]. Second, on the other side of the spectrum, there are data-driven methods. Data-Enabled Predictive Control (DeePC) is one of the relevant methods that has been applied to greenhouse-crop control [17] [18]. It does not rely on an explicit parametric description of the system, rather it uses previously recorded input-output trajectories to predict the future behavior of the system. For this, the greenhouse-crop control problem is split into two layers, i.e. an upper layer to generate climate trajectories based on optimizing crop growth and a lower layer that tracks the reference climate trajectories. While it is computationally more expensive, it was shown to be more efficient in terms of resource use.

Representing the complex greenhouse-crop system with an accurate model is difficult. Even if there would be an accurate description of the calibrated systems' dynamics, the real system can be time-variant. This is, the system's dynamics depend on the current operating point, which may change because of changing outside weather conditions and the development stage of the crop. Also, there already exist various highly non-linear greenhouse-crop models, but these are not suitable for optimal control because of computational costs and because solvers can only give local optima [19]. Therefore, it may be beneficial to let the automatic control algorithms, including more computationally suitable linear models, have the ability to adapt themselves on-line [18] [20]. This thesis sets a first step towards this on-line system adaptation: it investigates the possibility to use a linear state-space model to estimate the non-linear greenhouse-crop systems dynamics around an operating point, for the purpose of automatic optimal control. The automatic optimal control is composed of an MPC-based controller with a linear state-space model as a predictor, which is identified with the Parameter Estimation Method using collected input-output data. This proposed control method is different from earlier attempts, such as DeePC, as it does not split the control problem into two layers, but finds the optimal control inputs directly from the predicted crop growth. Also, it does not use the input-output data directly for predictions, rather it first finds an appropriate linear

system representation and uses that to make predictions.

## 1-2 Thesis research questions

The main research question for this thesis is

*To what extent can a greenhouse-crop system be estimated with a linear state-space representation for the purpose of automating the decision-making in greenhouse climate control?*

This main research question can be answered by answering the following sub-questions

a. *How does a modern-day greenhouse for tomato growing work and how can the greenhouse-crop system be modeled into a ground-truth simulation?*

The modern-day greenhouse and its actuators will be studied, alongside the theory behind tomato crop growth. Then, the dynamics of the greenhouse-crop system will be represented as a mathematical non-linear model using first principles from the literature. Finally, the equations will be used to create a ground-truth simulator that is calibrated with input-output data from a real-life growing experiment. The simulator helps answer the research question as it can be used to create artificial input-output data and it can be used for experimenting with the proposed control method.

b. *Under what conditions can the greenhouse-crop system be estimated with a linear state-space representation by using system identification?*

In order to obtain a linear state-space representation of the greenhouse-crop system, first a linear state-space model, including the states, inputs, outputs and disturbances, will be selected. Then, the conditions under which the ground-truth simulator dynamics can be represented in a linear fashion will be studied. For this, system identification is needed to estimate the linear state-space model. The answer to this sub-question will determine the operating point, prediction horizon and constraints for the controller responsible for automatically making the decisions for greenhouse climate control.

c. *How can an identified state-space model be used in an MPC-based controller to automate the decision-making in greenhouse climate control?*

Once the conditions are known under which the linear state-space model can be used for making predictions of the non-linear greenhouse-crop system, the model can be implemented as a predictor in an MPC-based controller. This controller will make use of an objective, composed of obtained yield and resource costs. Its performance will be compared to that of an actual grower, which will act as a baseline.

## 1-3 Thesis contributions

The contributions of this thesis project are stated as follows

- *A calibrated greenhouse-crop system simulator that contains all modern-day types of actuators and that will be used as a ground-truth.*
- *A Model Predictive Control based controller that automatically makes decisions for greenhouse climate control with the objective to make a profit, defined as the yield from selling harvested fruits minus the costs of resources.*
- *Identification of a linear state-space model, specifically for the inputs and outputs of the non-linear greenhouse-crop system.*
- *A performance comparison between the proposed controller and the baseline, in terms of profit.*

## 1-4 Thesis outline

To answer the research questions, this thesis first presents the modern-day greenhouse-crop system, including the dynamics of the greenhouse climate and the tomato crop in general terms. In Chapter 3, an appropriate greenhouse-crop model is composed of literature and it is converted into a simulator that will be used as the ground truth. This simulator is calibrated using data from real-life growing experiments. Subsequently, Chapter 4 first presents the system identification method used to estimate the defined linear state-space model. The chapter continues by presenting the grower's objective in formal terms. This objective is then implemented in the proposed MPC-based controller, where the identified linear state-space model is used as the prediction model. Next, in Chapter 5 the results are presented, including the realized climate and mass trajectories. The controller will also be compared with the baseline. In the final chapter, the system identification method and the performance of the proposed controller are discussed, followed by the conclusions and recommendations.

# Greenhouse-Crop System

The greenhouse-crop system must be understood before constructing a ground-truth simulator. For this, the modern-day greenhouse, including its actuators, energy and mass fluxes, states and disturbances, will be studied. Additionally, the influences of the greenhouse climate and disturbances on crop growth and development will be explained. The various kinds of available models of the greenhouse and crop will be set out, followed by model selection. After this chapter all information is gathered that can be used for setting up a formal greenhouse-crop model and corresponding simulator.

## 2-1 The Greenhouse-Crop System

This first section explains the characteristics and general dynamics of the modern-day greenhouse.

### 2-1-1 The modern-day greenhouse

Greenhouse horticulture can be found all over the globe. The many types of climates in which the greenhouses are located, require different kinds of shapes, properties and technologies. The buildings can range from very simple arch-shaped one-layer PE roofs found in warmer climates, to more advanced Venlo-type greenhouses with single glass layers mostly found in marine climates [21]. Technologies also greatly differ per greenhouse. Some greenhouses only use simple fans and ventilation windows, while others have complex HVAC-, heating-pipe-,  $CO_2$ -dosing and dehumidification- systems.

There is at least one thing all greenhouses have in common: transparent roofs and sidewalls to let through as much light as possible. Light is one of the four main driving factors for photosynthesis, besides Carbon Dioxide ( $CO_2$ ), temperature and water ( $H_2O$ ). Photosynthesis is the process where  $CO_2$  and  $H_2O$  received by the plant are converted to sugars and oxygen ( $O_2$ ), enabled with light energy. The more photosynthesis, the more sugars are created that the plant uses for fruit- and crop growth [22]. More on crop growth can be read in Section 2-3.

Artificial lighting systems, such as HPS- and LED lights, can also be installed to compensate for the lack of solar radiation [3]. Moreover, to keep warmth inside and thus save energy, special screens can be installed, e.g. thermal and shade screens. In agriculture, crops are grown in natural soil. In horticulture, the natural soil is replicated by placing the roots of the plants inside substrates, which are placed onto slabs organized in rows. To optimize plant growth, irrigation systems supply the substrate and plant with water filled with nutrients. For this thesis only the commonly used Dutch Venlo-type greenhouse is considered, installed with modern-day equipment. Also, the main focus of this project is the influence of the greenhouse climate on crop growth, excluding irrigation and fertilization, i.e. these two processes are considered to be controlled sufficiently and do not limit the growth. Figure 2-1 shows a schematic example of a Venlo-type greenhouse. More on the climate control equipment and inputs of the greenhouse climate system is explained in Subsection 2-2-2.

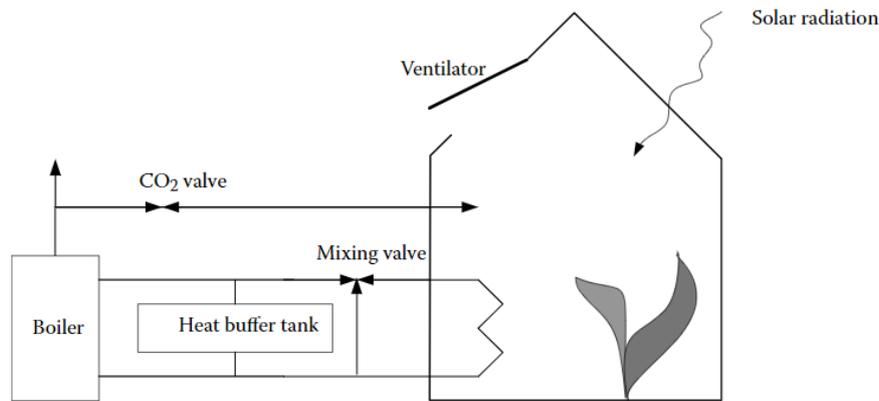


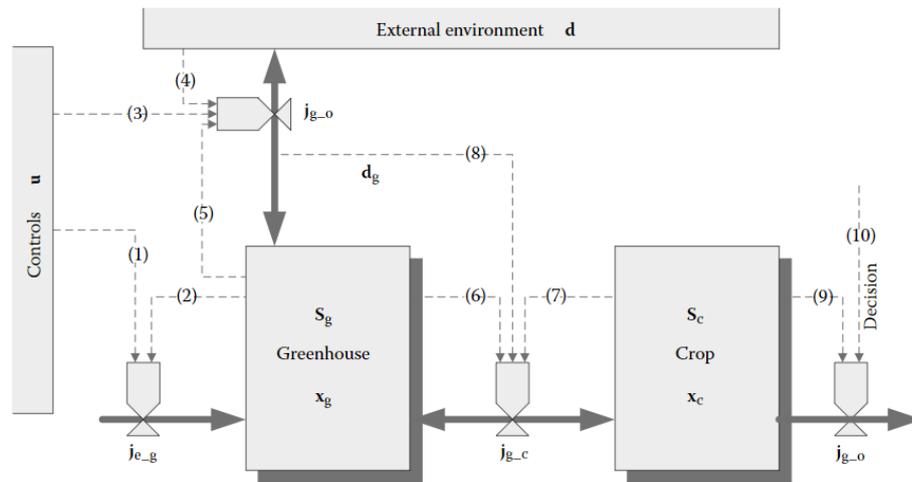
Figure 2-1: A schematic example of a Venlo-type greenhouse [1, Fig. 1.1].

## 2-1-2 The greenhouse-crop system

It is important to understand the fundamentals of the greenhouse-crop system that needs to be controlled. Figure 2-2 shows the energy and mass fluxes influencing the greenhouse climate and the information flows between the different components of the greenhouse model. The following summary of the greenhouse-crop system is mostly drawn from van Straten et al. [1].

First of all, the greenhouse compartment, the crop compartment, the external environment and the supplied resources are indicated by  $g$ ,  $c$ ,  $o$  and  $e$ , respectively. The arrows stand for flows of energy, water or carbonaceous material and the vectors associated with them are  $\mathbf{j}_{e_g}$ ,  $\mathbf{j}_{c_o}$ ,  $\mathbf{j}_{g_o}$  and  $\mathbf{j}_{g_c}$ . Energies and masses can be stored in the greenhouse compartment  $\mathbf{S}_g$  and crop compartment  $\mathbf{S}_c$ , per unit projected greenhouse area. These storages depend on the size of the greenhouse and the number of crops.

The greenhouse state  $\mathbf{x}_g$  and crop state  $\mathbf{x}_c$  denote the intensive variables, such as temperature and  $CO_2$ -concentration, that are coupled to the extensive variables of the storages. One reason for using the intensive variables is that those are the ones being measured in the greenhouses, i.e. heat fluxes cannot be measured directly, but temperature can be. Another reason for



**Figure 2-2:** Energy and mass fluxes in greenhouse-crop systems [1, Fig. 1.3].

using intensive variables besides the extensive variables is that the fluxes depend strongly on the intensive variables. For example, heat flux between the greenhouse and the external environment  $\dot{j}_{g-o}$  depends on the inside and outside temperature (intensive), while the heat flux is also related to the shape of the greenhouse building (extensive). Information flows, i.e. factors that influence the fluxes, are represented by dashed arrows. These flows represent two inputs: control inputs, denoted by the control variable  $\mathbf{u}$ , and environmental inputs, denoted by the external environment variable  $\mathbf{d}$  [1]. Table 2-1 gives a description of the ten information flows of the greenhouse-crop system.

It should be noted that there are constraints to the different fluxes and influences, that are not always in the hands of the controller. For example, a controller is able to control the ventilation in some sense by opening the windows, but the actual ventilation strongly depends on the uncontrollable wind speed. These complicated time-varying constraints require accurate predictions of the external environment and realistic models of the greenhouse and the crop. While there are models that show good approximations, there is still a gap to find the one-to-one representation of reality. The next section will show in more detail the greenhouse climate, where after the growth of the crop is shown.

## 2-2 Greenhouse climate

Next, the greenhouse climate fundamentals, energy and mass balances, control inputs and climate feedback will be presented.

### 2-2-1 Greenhouse climate: fundamentals

The greenhouse climate can be described by three important states, which are: air temperature, absolute or relative humidity and  $CO_2$  concentration. The temperature is considered the most important climate factor, as it has a strong influence on the humidity and the crop

**Table 2-1:** Description of the information flows for the greenhouse-crop system.

Information flow	Description
(1)	$\mathbf{j}_{e_g}$ depends on the control variable $\mathbf{u}$ , e.g. opening of the heating valve or the $CO_2$ dosing.
(2)	$\mathbf{j}_{e_g}$ also depends on the greenhouse state $\mathbf{x}_g$ , e.g. radiation received from the heating pipes and greenhouse temperature.
(3)	$\mathbf{j}_{g_o}$ depends on the control variable $\mathbf{u}$ , e.g. opening of the windows to exchange water and $CO_2$ , and to exchange heat through radiation and ventilation.
(4)	$\mathbf{j}_{g_o}$ also depends on the state of the external environment itself, e.g. wind speed to influence the ventilation.
(5)	$\mathbf{j}_{g_o}$ also depends on the state of the greenhouse, e.g. moisture and $CO_2$ flows are obviously influenced by the concentrations of the respective masses.
(6)	$\mathbf{j}_{g_c}$ depends on the state of the greenhouse, e.g. $CO_2$ uptake by crop photosynthesis is influenced by the concentration of $CO_2$ of the internal climate.
(7)	$\mathbf{j}_{g_c}$ also depends on the state of the crop, e.g. $CO_2$ uptake by crop photosynthesis is influenced by growth stage of the crop.
(8)	$\mathbf{j}_{g_c}$ also indirectly depends on the external environment, e.g. solar radiation influences the photosynthesis process, which influences the $CO_2$ uptake of the crop. This information flow is denoted by an extra variable $d_g$ , that can be influenced by control $\mathbf{u}$ , i.e. artificial lighting and climate screens can manipulate the greenhouse-crop system's fluxes.
(9)	$\mathbf{j}_{c_o}$ depends on the state of the crop, e.g. if the crop state is sufficient enough for harvesting, the grower will make the decision to harvest.
(10)	$\mathbf{j}_{c_o}$ also depends on discrete higher level decisions, that are not part of any automatic control loop, e.g. the grower deciding to harvest the crop, which will influence the state of the crop in the form of decreased crop weight.

growth processes. The temperature increases through the use of heating systems, artificial lighting in the form of lamps and uncontrollable solar radiation. The temperature decreases through opening windows, condensation and heat losses at the greenhouse cover or by transpiration by the crop. These processes and actuations also strongly influence the  $CO_2$  and humidity levels inside the greenhouse.

The dynamics of the three climate states are described by energy and mass balances. In the

literature, most physical models, also known as *mechanistic models*, are based on the same first principles but show some deviations, e.g. some include the soil or exclude a buffer or a heat exchanger. The different models also each have their own way of modeling the fluxes [23]. How they are filled in, depends on the level of detail and the control inputs that are used. The general balances, as presented next in (2-1), (2-2) and (2-3), are expressed in the form of Ordinary Differential Equation (ODE)'s and are based on the work of Van Beveren (2015), Kerkhof (2020) and Kuijpers (2021) [24] [17] [20]. The next chapter will give a more detailed model of the greenhouse-crop system.

### *Air temperature*

The energy balance of the air temperature for the greenhouse climate can be given by:

$$\frac{dT_{air}}{dt} = \frac{1}{c_{cap}}(Q_{srd} - Q_{cov} - Q_{trans} + Q_{pipe} + Q_{lamp} - Q_{vent}), \quad (2-1)$$

with the following heat fluxes: incoming radiation  $Q_{srd}$ , heat losses through the cover  $Q_{cov}$ , transpiration by the crop  $Q_{trans}$ , heating by the pipe rail system  $Q_{pipe}$ , artificial lighting  $Q_{lamp}$ , natural ventilation  $Q_{ven}$  all in  $[Wm^{-2}]$ .  $c_{cap}$  denotes the specific heat capacity of the air.

### *Humidity*

The mass balance of the humidity for the greenhouse climate can be given by:

$$\frac{dH_{air}}{dt} = \frac{1}{h}(\phi_{trans} - \phi_{cov} - \phi_{vent}), \quad (2-2)$$

with the following vapor fluxes: crop transpiration ( $\phi_{trans}$ ), condensation on the cover ( $\phi_{cov}$ ) and vapor exchange with outside air by natural ventilation ( $\phi_{vent}$ ), all in  $[gm^{-2}s^{-1}]$ . The average height of the greenhouse is denoted by  $h$  [m].

### *CO<sub>2</sub>-concentration*

The mass balance of the  $CO_2$  concentration for the greenhouse climate can be given by:

$$\frac{dCO_{2,air}}{dt} = \frac{1}{h}(\psi_{c,inj} - \psi_{c,ass} - \psi_{c,vent}), \quad (2-3)$$

with the following  $CO_2$  mass fluxes: injection of  $CO_2$   $\psi_{c,inj}$ , assimilation of  $CO_2$  by the crop  $\psi_{c,ass}$  and the  $CO_2$  exchange with the outside air due to ventilation  $\psi_{c,vent}$ , all in  $[gm^{-2}s^{-1}]$ . The average height of the greenhouse is denoted by  $h$  [m].

All energy, mass and vapor fluxes are summarised in Table A-2.

## **2-2-2 Greenhouse climate: control inputs**

The exact properties of the installed greenhouse equipment and infrastructure depend on the type of crop that is grown. Leafy greens, e.g. lettuce, are grown on slabs that need to be separated throughout the growing process and they require different climate conditions than

bulk crops that need to be harvested from a plant. These factors all play a role in what types of equipment are installed and how the climate is controlled with the different available system inputs. This thesis only considers crops that grow fruits that have to be harvested, in specific tomatoes.

The following inputs are considered to be installed in the greenhouse described in this thesis: heating,  $CO_2$ -dosing, ventilation, climate screening and artificial illumination [6].

## Heating

The most obvious system used in greenhouse climate control is the heating system. In most Venlo-type greenhouses heating is performed by running heated water through heating pipes laid down between the rows containing the crops [25]. The heating pipes also serve as a transportation system, i.e. as rails for carts that carry the harvested fruits and the laborers harvesting the fruit. To control the temperature inside the greenhouse, growers must give the climate computer a temperature setpoint. Then the heating system will automatically control the water temperature, based on a heating model of the greenhouse. Heat can be added by a boiler or by a Combined Heat and Power (CHP) system. Excessive heat can be stored in a buffer storage, to be drawn out again later.

Besides the control inputs, the grower has the ability to change settings in this low-level control system, e.g. the minimum and maximum pipe temperature, and the radiation influence [26]. First, setting the minimum pipe temperature too low would result in costly high temperature gains when heating is necessary. Setting the temperature too high causes a minimum temperature inside the greenhouse that may be unwanted. Second, some temperatures may never be reached by setting the maximum pipe temperature too low. Increasing the maximum can be beneficial as the greenhouse can heat up in a shorter amount of time, but it will also result in more evaporation from the crop locally, as the pipes lie within the canopy [25]. The radiation influence can be used to automatically adjust the pipe temperatures, based on the outside solar radiation.

## $CO_2$ -dosing

The outside air contains about 400ppm of  $CO_2$ , which is enough for plant growth. However, in cases of relatively high (solar) radiation, higher  $CO_2$  concentrations are desirable to optimize growth [25]. To maintain optimal photosynthesis and resource usage, the rate of supply of  $CO_2$  should follow the rate of absorption of the plant. There are three sources of  $CO_2$ : outside air, storage tanks and generators e.g. a Combined Heat and Power system (CHP). The climate computer can perform  $CO_2$  dosing by controlling the opening of the windows, see Section 2-2-2, or by injecting  $CO_2$ .

A point of attention is the distribution of  $CO_2$  throughout the indoor greenhouse structure, both vertically and horizontally. Sub-optimal distribution of  $CO_2$  can be caused by the plants using  $CO_2$  in the higher layers of the greenhouse and the greenhouse not being ventilated sufficiently. The same goes for the temperature inside the greenhouse. However, in most greenhouse climate models it is assumed that the building is a perfectly stirred tank, i.e. distribution of heat and gasses is uniform [27].

## Ventilation

A Venlo-type greenhouse is typed by its windows on the lee side and wind side to induce natural ventilation. These windows are mainly used for exchanging air, for multiple reasons. First, they can be opened to let fresh air containing  $CO_2$  in. Second, by opening the windows humid and warm air can leave the indoor system. Last, the windows can be opened to create air movement inside the canopy. The windows on the windward side can let air in through the wind. The leeward windows create a pressure difference across the greenhouse, the so-called *Venturi-effect*. While cooling by opening the windows is a cost-efficient method, it also results in costly humid and  $CO_2$ -filled air leaving the greenhouse, which can be unfavorable in some cases. Natural ventilation is not effective when the external climate is similar to the internal climate, thus not creating any gradient. An alternative to cooling the greenhouse by opening windows is using a mechanical cooling system, but these are quite costly as they have to cool huge amounts of air.

For this thesis, only natural ventilation is considered, i.e. opening windows. To control the ventilation, the climate controller has to specify the relative opening of the windows, on the leeward and windward sides.

## Climate screening

In most greenhouses, climate screens are installed. There are at least two types of screens. First, there are blackout screens, that prevent artificial light from polluting the sky and prevent high solar radiation. Second, there are (transparent) climate screens that reduce energy consumption during cold days [26]. They are made transparent so that they still allow solar radiation into the greenhouse. The climate controller has to specify the relative opening of the screens, where a fully opened screen means that it is not being used and a fully closed screen means that is used to its full extent.

## Artificial illumination

When there is not enough solar radiation, for instance during early morning or late afternoon, artificial illumination can be turned on to still achieve optimal plant growth. There are two types of lights that can be installed: High Pressure Sodium (HPS) lighting to Light Emitting Diode (LED) lighting. The former creates the typical orange light and it has the advantage that it also emits heat when turned on, which is useful as the crop needs light and heat at the same moment. However, these lamps do not have efficient energy usage ( $1.85 [\mu mol J^{-2}]$  [3]). A more energy-efficient arising solution is LED lighting. LED lights outperform HPS lights in terms of energy efficiency ( $3.0 [\mu mol J^{-2}]$  [3]) and are shown to be a viable alternative for tomato growing [28]. Lights are controlled by the climate controller by specifying the percentage of lights that is turned on.

### 2-2-3 Greenhouse climate: feedback

To control the greenhouse-crop system, the received feedback from the greenhouse climate is also important. The necessary climate measurements at the canopy level are inside air

temperature, humidity and  $CO_2$  concentration, as those are the driving forces for the photosynthesis of the crop. Outside measurements are also essential, as these are the disturbances that the climate control system must counteract as efficiently as possible. Relevant outside climate measurements are wind velocity and direction, solar radiation, humidity, temperature and  $CO_2$ -concentration. These do not depend on the greenhouse-crop states and are therefore considered as disturbances.

Sensors measuring the greenhouse and outside climate are quite accurate as of now. Measuring the crop is more difficult, as will be discussed in Subsection 2-3-2. Therefore, in literature, it is often assumed that the sensors are perfect, i.e. the sensor data does not contain measurement noise or any other faults. By Bontsema et al. (2011) [29] it was shown that greenhouse climate sensor inaccuracies can amount to an increase or decrease of 5% of the expected performance. Measurement and state estimation might help in this case, but are out of the scope of this thesis.

#### **2-2-4 Greenhouse climate models: literature summary**

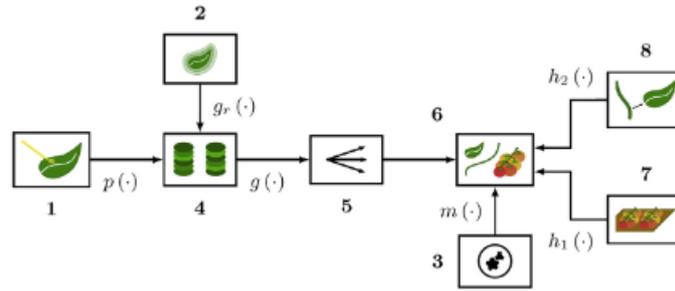
Many have researched greenhouse climate models [12]. They can be divided into two main groups: first-principle mechanistic dynamical models, i.e. white-box models, and black-box models without any a priori knowledge built in [30]. A third group is a fusion of the two and those are the grey box models, which combines the a priori knowledge of the first principles in the form of structures and the experimental data characteristics of the black-box model [31]. Another form of modeling that is of interest is the hybrid form, where predictions of both a black box and a mechanistic dynamical model are used to model the system [32]. The reader is referred to the preceding literature survey for a complete overview of all the greenhouse climate models in the literature.

### **2-3 Crop growth**

The goal of greenhouse horticulture is to grow salable crops or parts of the crops, e.g. fruits. To be able to do this efficiently in terms of resource usage, it is important to understand a crop and its growth. A considerable understanding of the crop-growing processes enables models to be created that can predict crop development.

#### **2-3-1 Crop growth: fundamentals**

The enabling process behind crop growth is photosynthesis. There are four main driving factors of photosynthesis: light, carbon dioxide ( $CO_2$ ), temperature ( $T$ ) and water ( $H_2O$ ).  $CO_2$  and  $H_2O$  received by the plant are converted to assimilates (sugar) and oxygen ( $O_2$ ), enabled with light energy. The more photosynthesis, the more assimilates are created that the plant uses for fruit- and crop growth [22]. The assimilates have two roles. First, they are used as a building material, and, second, they are used by the plant as an energy source. The plant will always balance its consumption and production of assimilates. If there is an assimilate shortage, the plant has to cut down on consumption, at the cost of fruit development or



**Figure 2-3:** Block diagram representation of the common crop growing processes. 1: photosynthesis, 2: growth respiration, 3: maintenance respiration, 4: accumulating assimilates, 5: assimilate partitioning, 6: accumulating biomass, 7: tomato harvest, 8: leaf harvest.  $p$  represents the assimilates generated through photosynthesis,  $g_r$  the assimilates used for growth respiration and  $g$  the amount of assimilates converted to biomass. A part of the biomass is used for maintenance respiration  $m$ .  $h_1$  and  $h_2$  describe the harvest of fruit and leaves, respectively [2].

quality [25]. There exist numerous other processes that influence and enable the growth of the plant, summarised in Figure 2-3.

Most tomato crop growth models provide a relationship between crop yield and indoor climate factors, i.e. temperature ( $T_{in}$ ), global radiation from the sun ( $Q_{sun}$ ) converted to actually received radiation by the plant canopy,  $CO_2$ -concentration ( $CO_{2,in}$ ) and the biomass production  $y_b$  and/or the total yield of the crop ( $y_f$ ) [2]. The tomato crop growth model ( $f_c$ ) can be mathematically represented as:

$$\begin{bmatrix} y_b \\ y_f \end{bmatrix} = f_c(T_{in}, Q_{sun}, CO_{2,in}), \quad (2-4)$$

where  $f_c$  represents the relation between the inputs and the outputs.

The most important state variables in crop models are fruit weight and leaf weight. Additionally, buffer weight and stem weight can be added to complete the description of the crop's state. Some models also divide fruit weight into multiple fruit development stages, see Subsection 2-3-3 for more details on the different kinds of crop models.

### 2-3-2 Crop growth: feedback and measurements

To be able to accurately control the system, it is important that measurements of the outputs of the crop system are available in the form of feedback. Not all of the aforementioned crop states, e.g. leaf, fruit and buffer weight, can be measured at all times. For example, the amount of assimilates currently in the buffer cannot be measured directly, but have to be estimated by integrating the airflow and  $CO_2$  exchange rates [33]. It is also infeasible to measure the weight of all crops and fruits at once; it is still a physically heavy manual task. While new computer vision techniques allow the counting and size estimation of tomatoes [34], they are not yet used in practice. For crop growth modeling and validation, literature often assumes that the crop-relevant states are measurable in some fashion, meaning that the output of the model is equal to the state  $y_C = x_C$ . Others try to improve state estimation

with data assimilation, e.g using Kalman filtering [35]. Measuring the weight of the leaves is something that is uncommon in current greenhouse operations, but it is also important for accurate photosynthesis estimation. Good estimations can be achieved by determining the Leaf Area Index (LAI), defined as the fraction of the greenhouse that is covered by the canopy [26].

### **2-3-3 Crop growth models: literature summary**

As with greenhouse models, a distinction can be made between first-principle models and black-box models for crop models. In the literature, the first-principle models are mostly used and are the main source of prediction in practice. The reader is referred to the preceding literature survey for a complete overview of all the crop models in the literature. The next section will present differences between greenhouse and crop model characteristics and when to select what type.

## **2-4 Model selection**

The dynamics of both the greenhouse and the crop can be modeled in various ways, with different granularities, as will be explained below.

### **2-4-1 Climate model granularity**

As mentioned above, climate models are often based on first-principles and the climate states evolve according to energy and mass balances. How the balances and their fluxes are filled in determine the complexity and granularity of the models. In some models the fluxes are considered as direct control inputs. For example, Kuijpers et al. (2021) uses the heat flux and ventilation flux as inputs without making use of complicated non-linear heating pipe temperature and window ventilation relations [20]. On the other hand Tap (2000) gathers all kinds of non-linear ODE's describing how ventilation depends on the window opening and the outside wind-speed [36].

### **2-4-2 Crop model granularity**

In literature, numerous models exist with different granularity for different application purposes. Coarse crop models describe the total biomass of the plant as a single weight, leading to only one to four states to be modeled. Examples are the Greenhouse Technology Application (GTA) by De Wit (2006) and the model by Seginer et al. (1994) [2] [37]. Crop models with medium granularity model the biomass with three weights: the combined mass of the stem and roots, the mass of the fruits and the mass of the leaves. An example of a medium coarse model is the model of Vanthoor (2011), also referred to as the big leaf, big fruit model. It uses eight states of the crop, plus two per truss development stage [7]. A third level of granularity is fine granularity, also referred to as small leaf, small fruit models. These models, like TOMGRO or TOMSIM [38] [39], model the weight of the crop for separate fruits, leaves and stem segments, taking into account the physiological age of these parts. The detailed

model TOMSIM by Jones (1991) contains the most model parameters: twelve, plus seven per truss and four per vegetative unit (the part of the stem and leaf corresponding to a truss). Kuijpers et al. (2019) studied the aforementioned models and separated the eight earlier mentioned crop growing processes, defining the level of granularity for each of them [2]. Then, he studied 27 different combinations of the four original models. The best combination showed a decrease of 13% in RSME, while allowing a 7.5% increase in model computation time compared to one of the original models.

### 2-4-3 Model selection

Not every greenhouse or crop model is useful for the same application. Intuitively, the more detailed the modeled is, the more states, the larger the computational time. For control purposes the coarser crop models are more suitable, e.g. model from Tap (2000), but for real-life application they might not give good results because of model mismatch. A more detailed crop model can be used for simulation purposes and to create data sets, e.g. from Vanthoor (2011) [7].

The model, a combination of Vanthoor (2011) and Kuijpers (2021) [7] [20], that is used for the ground-truth simulation in this thesis is a mechanistic model based on first-principles, but will be calibrated using data. This choice was made because there is already a lot of research done on greenhouse climate modeling, and it would be wasteful not to use the gathered knowledge. The ground-truth simulation model can be found in Chapter 3, including its calibration.

## 2-5 Chapter summary and conclusion

This chapter showed that the greenhouse-crop system is build up of complex processes. The crop is influenced by temperature, humidity,  $CO_2$ -concentration and light of the sun and artificial lights, which compose the complete greenhouse climate. The crop also influences the greenhouse climate through processes such as transpiration and photosynthesis. The greenhouse climate can be controlled via heating,  $CO_2$ -dosing, climate screening, artificial illumination and ventilation. These control inputs will play an important role further on, as they compose the complete decision-making actions of the proposed controller. Crops grow through the process named photosynthesis; light, heat and  $CO_2$  are converted into sugars, also known as assimilates, which are the building blocks for plant and tomato fruit growth. For the crop, various models that describe their growth exist, each with different granularities. One of the more detailed models will be used for the ground-truth simulator and will be explained in its entirety in the next chapter. This choice is made for the main reason that as much detail and real-life representation is desired for the simulator to act as the ground truth, while computational costs are not a limiting factor.



# Greenhouse-Crop Ground-truth Simulation Model

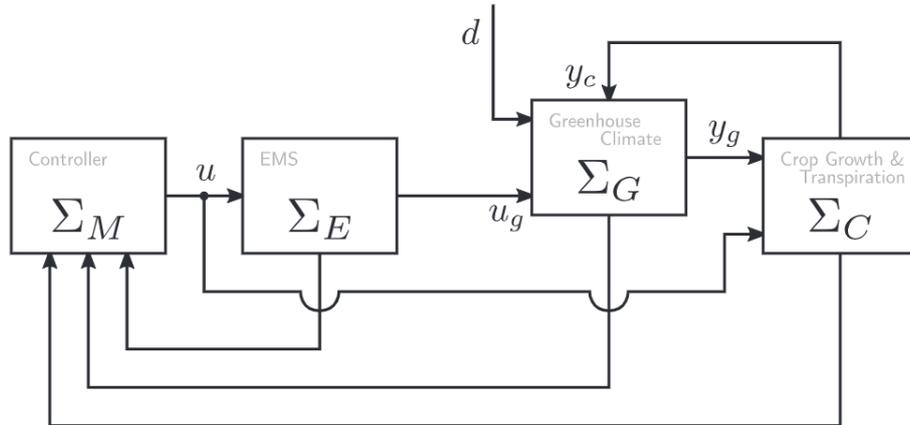
This chapter will present a complete overview of the fine-grained greenhouse-crop model that is used for a calibrated ground-truth simulator. The model of the greenhouse-crop system is mainly based on the model of Kuijpers et al. (2021), which is originally created by Vanthoor (2011) [7] [20]. First, a system description of the greenhouse and will be given, explaining all states, inputs, outputs and disturbances. Second, the interaction and dynamics of the system will be presented in two sub-subsequent sections. The remaining part of this chapter will describe the dataset, coming off the Autonomous Greenhouse Challenge, and how the simulator is calibrated with this input-output data of the real-life growing experiment.

### 3-1 System description

Figure 3-1 shows a block diagram representation of the selected greenhouse-crop system. The different elements are: a controller  $\Sigma_M$ , an energy management system  $\Sigma_E$ , a greenhouse climate and lighting system model  $\Sigma_G$ , and a crop growth and transpiration model  $\Sigma_C$ . Effects and controls on the different components of the system are: Control inputs  $\mathbf{u}_g$ , uncontrolled inputs or disturbances such as the outside weather  $\mathbf{d}$ , effects of the greenhouse climate on the crop (temperature,  $CO_2$  concentration, radiation and relative humidity)  $\mathbf{y}_g$  and the effect of the crop on the greenhouse (assimilation and transpiration)  $\mathbf{y}_c$ .

Note that the energy management system (EMS) is simplified for control and simulation purposes. Originally, the EMS consisted of a buffer and a combined heat and power (CHP) unit. These are replaced by more simple control inputs, i.e. a pipe temperature of which it is assumed that the temperature can be changed instantly and the supply of  $CO_2$  which can be applied to the greenhouse system directly without any generation via a CHP unit.

The states  $\mathbf{x}$  ( $\mathbf{x}_c$  and  $\mathbf{x}_g$ ), inputs  $\mathbf{u}$  ( $\mathbf{u}_c$  and  $\mathbf{u}_g$ ), outputs  $\mathbf{y}$  ( $\mathbf{y}_c$  and  $\mathbf{y}_g$ ) and disturbances  $\mathbf{d}$ ,



**Figure 3-1:** Block diagram representation of the greenhouse control problem [3, Fig. 1].

of the greenhouse-crop system are given by

$$\mathbf{x} = \begin{bmatrix} T_{air} \\ AH_{air} \\ CO_{2,air} \\ C_{buf} \\ C_{leaf} \\ \mathbf{C}_{Frt\{s\}} \\ C_{stem} \\ T_{c24} \\ \mathbf{N}_{Frt\{s\}} \\ H_F \\ H_L \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_{CO2} \\ u_{scr1} \\ u_{scr2} \\ T_{pipe,1} \\ T_{pipe,2} \\ u_{led} \\ \phi_{Lee} \\ \phi_{Win} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} T_{air} \\ AH_{air} \\ CO_{2,air} \\ T_{c24} \\ H_F \\ m_F \\ m_L \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} T_{out} \\ AH_{out} \\ CO_{2,out} \\ Q_{sun} \\ v_{wind} \end{bmatrix}. \quad (3-1)$$

The symbols are described below in Table 3-1, Table 3-2, Table 3-3 and Table 3-4, respectively.

**Table 3-1:** Greenhouse-crop states  $\mathbf{x}$ .

State	Unit	Description
$T_{air}$	$^{\circ}C$	Greenhouse indoor air temperature
$AH_{air}$	$gm^{-3}$	Greenhouse indoor absolute humidity
$CO_{2,air}$	$gm^{-3}$	Greenhouse indoor $CO_2$ concentration
$C_{buf}$	$gm^{-2}$	Assimilate buffer weight
$C_{leaf}$	$gm^{-2}$	Weights of the leaves
$C_{Frt\{i\}}$	$gm^{-2}$	Weight of the fruits in development stage $s$
$C_{stem}$	$gm^{-2}$	Weight of the stems
$T_{c24}$	$^{\circ}C$	24 Hour average of the greenhouse air temperature
$N_{Frt\{i\}}$	$m^{-2}$	Number of fruits in development stage $s$
$H_F$	$gm^{-2}$	Harvest fruit dry weight
$H_L$	$gm^{-2}$	Harvested leaf weight

**Table 3-2:** Greenhouse-crop control inputs  $\mathbf{u}$ .

Input	Unit	Description
$T_{pipe,1}$	$^{\circ}C$	Temperature of the heating pipes
$T_{pipe,2}$	$^{\circ}C$	Temperature of the heating pipes
$u_{led}$	$Wm^{-2}$	Electrical power to the artificial Light Emitting Diode (LED) lights
$u_{CO_2}$	$gm^{-2}s^{-1}$	Injecting $CO_2$
$\phi_{lee}$	%	Opening of the leeward side window
$\phi_{win}$	%	Opening of the windward side window
$u_{scr1}$	-	Opening the temperature screen
$u_{scr2}$	-	Opening the blackout screen

**Table 3-3:** Greenhouse-crop outputs  $y$ .

Output	Unit	Description
$T_{air}$	$^{\circ}C$	Greenhouse indoor air temperature
$AH_{air}$	$gm^{-3}$	Greenhouse indoor absolute humidity
$CO_{2,air}$	$gm^{-3}$	Greenhouse indoor $CO_2$ concentration
$T_{c24}$	$^{\circ}C$	24 Hour average of the greenhouse air temperature
$H_F$	$gm^{-2}$	Harvest fruit dry weight
$m_F$	$gm^{-2}$	Total fruit dry weight
$m_L$	$gm^{-2}$	Leaf weight

**Table 3-4:** Greenhouse disturbances  $d$ .

Disturbance	Unit	Description
$T_{out}$	$^{\circ}C$	Outside air temperature
$AH_{out}$	$gm^{-3}$	Outside absolute humidity
$CO_{2,out}$	$gm^{-3}$	Outside $CO_2$ concentration
$Q_{sun}$	$Wm^{-2}$	Outside solar radiation
$v_{wind}$	$ms^{-2}$	Outside wind speed

The greenhouse is equipped with heating pipes, windows that can be opened, a blackout screen, an energy screen, artificial lighting (LED), and a system capable of injecting  $CO_2$ .

## 3-2 Crop mathematical model

This section describes the crop model, including the states, dynamics and the crop development stages.

### 3-2-1 Crop growing dynamics

The states that belong to the crop subsystem are the assimilate buffer weight, the weights of the leaves, weight of the fruits in development stage  $s$ , the weight of the stem, 24 hour average of the greenhouse air temperature, and the number of fruits in development stage  $s$ , the fruit harvest weight and leaf harvest weight, the denoted by  $C_{buf}$ ,  $C_{leaf}$ ,  $\mathbf{C}_{Frt\{s\}}$ ,  $C_{stem}$  [ $gm^{-2}$ ],  $T_{c24}$  [ $^{\circ}C$ ],  $\mathbf{N}_{Frt\{s\}}$  [ $m^{-2}$ ],  $H_F$  [ $gm^{-2}$ ],  $H_L$  [ $gm^{-2}$ ], respectively. The evolution of the crop states is described by differential equations, that are composed of mass flow rates. The

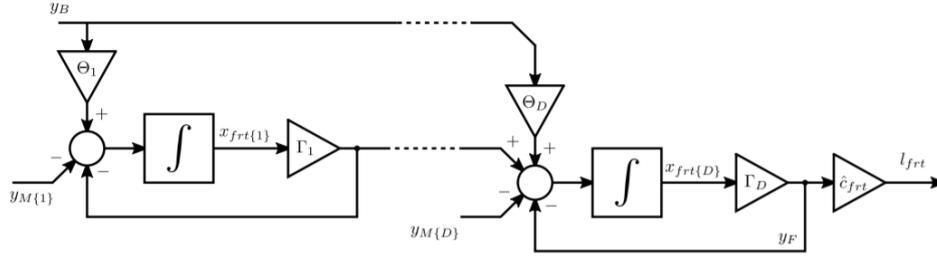
model can be summarised in the following state vector and dynamics:

$$\begin{bmatrix} \dot{C}_{buf} \\ \dot{C}_{leaf} \\ \dot{\mathbf{C}}_{Frt\{i\}} \\ \dot{C}_{stem} \\ \dot{T}_{c24} \\ \dot{N}_{Frt\{i\}} \\ \dot{H}_F \\ \dot{H}_L \end{bmatrix} = \begin{bmatrix} MC_{AirBuf} - MC_{BufFrt} - MC_{BufStem} - MC_{BufAir} \\ MC_{BufLeaf} - MC_{LeafAir} - MC_{LeafHar} \\ MC_{BufFrt\{i\}} + MC_{Frt\{i-1\}Frt\{i\}} - MC_{Frt\{i\}Frt\{i+1\}} - MC_{FrtAir\{i\}} \\ MC_{BufStem} - MC_{StemAir} \\ \tau_{Tcan}^{-1} (T_{air} - T_{c24}) \\ MN_{Frt\{i-1\}Frt\{i\}} - MN_{Frt\{i\}Frt\{i+1\}}, \\ MC_{FrtHar} \\ MC_{LeafHar} \end{bmatrix} \quad (3-2)$$

where  $\tau_{Tcan}$  denotes the number of seconds per day (86400s) and  $s$  denotes the fruit development stage (see next paragraph). The mass flow rates  $MC_{XY}$  are explained in Table A-1 in the appendix. Note that  $\dot{\mathbf{C}}_{Frt\{s\}}$  is a vector with length  $\mathbb{D}$ , denoting the number development stages, see next paragraph. Also note that in the last development stage, the assimilate outflow to the next stage is  $MC_{FrtHar}$ , representing fruit harvesting. The next paragraph will only give a part of the crop model, as it is important for the conditions for the system identification chapter of this thesis. The interested reader is referred to Vanthoor (2011) for all details on the mass flows and crop growing dynamics[7].

### 3-2-2 Crop development stages

Vanthoor (2011) based his fruit growth on the 'fixed boxcar train' model by Rabbinge and Ward (1989) [4]. This model implies that assimilates and the number of fruits flow from one fruit development stage to the next with a specific development rate. It is a way to take into account the time delay between fruit set and fruit harvest, also known as the fruit growth period (FGP). The basic structure of the fruit development stage model is shown in a diagram in Figure 3-2. The model has  $\mathbb{D}$  development stages per fruit. Each stage  $s$  contains an assimilate buffer  $C_{Frt\{s\}}$  that represents the assimilates of fruits in development stage  $s$ . The assimilates coming from the assimilate buffer to the fruit  $y_b$  (or  $MC_{BufFrt}$ ) are distributed over the  $\mathbb{D}$  development stages, through gains  $\Theta_s$ . The gains  $\Gamma_s$  control the flow from buffer to buffer. The gains  $\Theta_s$  and  $\Gamma_s$  are non-linear functions of the 24 hour average temperature  $T_{c24}$  and the number of fruits  $N_{Frt\{s\}}$  in development stage  $s$ . The assimilates flow through the subsequent stages until the fruit is harvested after stage  $\mathbb{D}$ . When harvested, it is assumed that the harvested assimilates  $y_F$  are sold immediately and create revenue  $l_{Frt}$ , calculated with price coefficient  $\hat{c}_{Frt}$ .



**Figure 3-2:** Flowchart of the fruit development model by Vanthoor (2011), based on Rabbinge and Ward (1989) [4].

### 3-3 Greenhouse climate mathematical model

The states that belong to the greenhouse subsystem are the greenhouse air temperature, absolute humidity and  $CO_2$ -concentration,  $T_{Air}$ ,  $AH_{air}$  and  $CO_{2,air}$  respectively. There are three balances that describe the evolution of these states: an energy balance, an absolute humidity mass balance and a  $CO_2$  mass balance, shown in (3-3), (3-17) and (3-23), respectively. All energy, mass and vapor fluxes are summarised in Table A-2. The model parameters are described in Table A-3, Table A-4 and Table A-5.

#### 3-3-1 Greenhouse air temperature

The energy balance is given by

$$\frac{T_{air}}{dt} = \frac{1}{c_{cap}}(Q_{srd} - Q_{cov} - Q_{trans} + Q_{lamp} - Q_{ven} + u_{hea}), \quad (3-3)$$

with the following heat fluxes: incoming radiation  $Q_{srd}$ , heat losses through the cover  $Q_{cov}$ , transpiration by the crop  $Q_{trans}$ , artificial lighting  $Q_{lamp}$ , natural ventilation  $Q_{ven}$  and heating by the pipe rail system  $u_{hea}$ , all in  $[Wm^{-2}]$ .  $c_{cap}$  denotes the specific heat capacity of the air.

##### *Solar Radiation*

The actual transmitted solar radiation through the greenhouse cover is given by

$$Q_{srd} = \tau_{cov}(1 - (1 - \tau_{scr1})(1 - u_{scr1}))(1 - (1 - \tau_{scr2})(1 - u_{scr2}))Q_{sun}, \quad (3-4)$$

where  $\tau_{cov}$  [-] is the transmittance of the cover,  $\tau_{scr1}$  [-] is the transmittance of the temperature screen and  $\tau_{scr2}$  [-] is the transmittance of the climate screen.

##### *Cover*

The screen and the greenhouse cover are modeled as two heat conductors in series:

$$Q_{cov} = \frac{1}{((1 - u_{scr1})\alpha_{scr1})^{-1} + ((1 - u_{scr2})\alpha_{scr2})^{-1} + (\alpha_{cov} \frac{A_{cov}}{A_{floor}})^{-1}}(T_{air} - T_{out}), \quad (3-5)$$

where  $\alpha_{cov} \frac{A_{cov}}{A_{floor}}$  denotes the conductance of the greenhouse cover with the cover area  $A_{cov}$  and floor area  $A_{floor}$ . The heat conductance of the greenhouse cover and the screens are denoted by  $\alpha_{cov}$  and  $\alpha_{scr1,2}$ , respectively.

#### Lamps

Heat generated and added to the greenhouse air by the lamps is given by

$$Q_{lamp} = \eta c_{lamp} u_{led}, \quad (3-6)$$

where  $c_{lamp}$  [-] is the Lamp heat coefficient, which is added for calibration purposes, see Section 3-5.

#### Ventilation

The heat flux by the ventilation is modeled as:

$$Q_{ven} = \rho_a c_p \Phi_v (T_{air} - T_{out}) \quad , \quad (3-7)$$

where  $\Phi_v$  [ $ms^{-1}$ ] ventilation flux from indoor to outdoor air and  $\rho_{air}$  [ $kgm^{-3}$ ] is the density of air.

#### Transpiration

The heat flux of crop transpiration is given by

$$Q_{trans} = L \phi_{trn}, \quad (3-8)$$

where  $\phi_{trn}$  [ $gm^{-2}s^{-1}$ ] represents the transpiration rate, given by

$$\phi_{trn} = \frac{2LAI}{(1 + \epsilon)r_b + r_s} (AH_{crop} - AH_{air}), \quad (3-9)$$

where  $\epsilon$  [-] is the ratio of latent to sensible heat content of saturated air and  $r_b$  [ $sm^{-1}$ ] is the boundary layer resistance. The water vapor concentration at crop level  $AH_{crop}$  is given by

$$AH_{crop} = AH_{air,sat} + \epsilon \frac{r_b}{2LAI} \frac{R_{tot}}{L} = 5.563e^{0.0572T_{air}} + \epsilon \frac{r_b}{2LAI} \frac{R_{tot}}{L}. \quad (3-10)$$

$L$  [ $Jg^{-1}$ ] represents the energy needed to evaporate water and is given by

$$L = l_1 - l_2 T_{Air}, \quad (3-11)$$

where  $l_1$  [ $Jg^{-1}$ ] and  $l_2$  [ $Jg^{-1}C^{-1}$ ] are water vaporisation energy coefficients.

The stomatal resistance  $r_s$  [ $sm^{-1}$ ] is given by

$$r_s = (82 + 570e^{-\gamma \frac{R_{tot}}{LAI}})(1 + 0.023(T_{air} - 24.5)^2). \quad (3-12)$$

The net absorbed radiation at crop level is given by  $R_{tot} = R_{PAR} + R_{NIR}$  [ $Wm^{-2}$ ] in which

$$R_{PAR} = (1 - e^{-k_{PAR} \cdot LAI})(a_{sunPAR} \eta_{sunPAR} Q_{srd} + a_{LEDPAR} R_{LEDPAR}), \quad (3-13)$$

and

$$R_{\text{NIR}} = (1 - e^{-k_{\text{NIR}} \cdot \text{LAI}})(a_{\text{sunNIR}} \eta_{\text{sunNIR}} Q_{\text{srd}} + a_{\text{LEDNIR}} R_{\text{LEDNIR}}) \quad (3-14)$$

where the extinction coefficient of the canopy is denoted by  $k$  [-], the leaf area index is denoted by LAI [ $m^2 m^{-2}$ ].  $\eta_{\text{sunPAR}}$  [-] denotes the fraction of PAR in the radiation. The absorption parameters are denoted by  $a_{\text{sunPAR}}$ ,  $a_{\text{LEDPAR}}$  and  $a_{\text{LEDNIR}}$  [-].  $R_{\text{LEDPAR}}$  and  $R_{\text{LEDNIR}}$  are given by

$$R_{\text{LEDPAR}} = \eta_{\text{LEDPAR}} u_{\text{led}}, \quad (3-15)$$

$$R_{\text{LEDNIR}} = \eta_{\text{LEDNIR}} u_{\text{led}} \quad (3-16)$$

where  $\eta_{\text{LEDPAR}}$  and  $\eta_{\text{LEDNIR}}$  [-] are the fractions of the electrical input to the lights that is converted to PAR and NIR, respectively.

### 3-3-2 Greenhouse air absolute humidity

$$\frac{dAH}{dt} = \frac{1}{h}(\phi_{\text{trn}} - \phi_{\text{cov}} - \Phi_v(AH_{\text{air}} - AH_{\text{out}})), \quad (3-17)$$

with the following vapor fluxes: crop transpiration  $\phi_{\text{trn}}$ , condensation on the cover  $\phi_{\text{cov}}$  in [ $gm^{-2}s^{-1}$ ] and natural ventilation  $\Phi_v$  in [ $ms^{-1}$ ]. The average height of the greenhouse is denoted by  $h$  [ $m$ ].

#### Ventilation

The total ventilation flux modeled by from De Jong 1990 [40] is given by

$$\Phi_v = (1 - (1 - P_{\text{scr1}})(1 - u_{\text{scr1}}))(1 - (1 - P_{\text{scr2}})(1 - u_{\text{scr2}}))\left(\frac{\sigma \phi_{\text{Lee}}}{1 + \chi \phi_{\text{Lee}}} + \zeta + \xi \phi_{\text{Wind}}\right)v_{\text{wind}} + \psi, \quad (3-18)$$

where  $P_{\text{scr1,2}}$ ,  $\sigma$  [ $\%^{-1}$ ],  $\chi$  [ $\%^{-1}$ ],  $\zeta$  [-],  $\xi$  [ $\%^{-1}$ ] and  $\psi$  [ $ms^{-1}$ ] are ventilation rate parameters, and  $\phi_{\text{Lee}}$  [%] and  $\phi_{\text{Wind}}$  [%] are the window openings on the leeward and windward side of the greenhouse.

#### Condensation

Condensation the cover of the greenhouse  $\phi_{\text{cov}}$  [ $gm^{-2}$ ] is given by

$$\phi_{\text{cov}} = g_c(0.2522e^{0.0485T_{\text{air}}}(T_{\text{air}} - T_{\text{out}}) - (AH_{\text{air,sat}} - AH_{\text{air}})), \quad (3-19)$$

with greenhouse air humidity ratio at vapour pressure  $p_{\text{Air}}$   $AH_{\text{air,sat}}$ , and the condensation conductance of the cover ( $g_c$ ) [ $ms^{-1}$ ] given by

$$g_c = p_{g_c} \cdot S((T_{\text{air}} - T_{\text{cov}}), -2, 1) \cdot P(T_{\text{air}} - T_{\text{cov}}), \quad (3-20)$$

where  $T_{\text{air}}$ ,  $T_{\text{cov}}$  and  $p_{g_c}$  denoted the greenhouse air temperature, the temperature of the cover and the specific properties of the cover [ $m^\circ C^{\frac{1}{3}}s^{-1}$ ], respectively. The temperature of the cover can be estimated as a *weighted mean* between the outside and inside air temperature:

two-thirds outside and one-third inside temperature [41].  $S(s_v, s_k^l, s_s)$  is a smoothed if-else function shown in (3-21) that approximates a max operator, presented in the original model of van Beveren et al. (2015) [24]. This smoothed function is used to make this part of the model differentiable.  $P(x)$  is a fourth-order polynomial shown in (3-22) that approximates the third-order root that was originally present in the model. In the paper, it was shown that this was a good approximation, with a maximum error of  $0.7 \text{ ms}^{-1}$ .

$$S(s_v, s_k^l, s_s) = \frac{1}{1 + e^{s_k^l(s_v - s_s)}}, \quad (3-21)$$

where the slope, switching parameter and value that determine the value of  $S$  are denoted by  $s_k^l$  [-],  $s_s$  [-] and  $s_v$  [-] ( $= T_{air} - T_{cov}$ ), respectively.  $l$  is the process that is influenced by the switch and  $k$  is the state that determines  $S$ .

$$P(x) = -4.03 \cdot 10^{-5} \cdot x^4 + 2.4 \cdot 10^{-3} \cdot x^3 - 0.05 \cdot x^2 + 0.49 \cdot x + 0.30 \quad (3-22)$$

### 3-3-3 Greenhouse air $CO_2$ concentration

$$\frac{dCO_{2,air}}{dt} = \frac{1}{h}(u_{CO_2} - \psi_{c,ass} - \psi_{c,vent}), \quad (3-23)$$

with the following  $CO_2$  mass fluxes: injection of  $CO_2$   $u_{CO_2}$ , assimilation of  $CO_2$  by the crop  $\psi_{c,ass}$  (equal to  $MC_{AirBuf}$ ) and the  $CO_2$  exchange with the outside air due to ventilation  $\psi_{c,vent}$ , all in  $[gm^{-2}s^{-1}]$ . The average height of the greenhouse is denoted by  $h$  [m].

The exchange of  $CO_2$  with the outside air, due to ventilation, is given by

$$\psi_{c,vent} = \Phi_v(CO_{2,air} - CO_{2,out}). \quad (3-24)$$

## 3-4 Greenhouse-crop model parameters

All model parameters, including the calibrated values, for the crop- and greenhouse system are described in, Table A-3, Table A-4 and Table A-5 in the appendix.

## 3-5 Calibration of the greenhouse-crop ground-truth simulation model

To make the ground-truth greenhouse-crop model represent reality, it was calibrated using the data from the Autonomous Greenhouse Challenge (AGC). This data set had to be pre-processed. Then, the non-linear crop model was calibrated using the fruit weight in the calibration cost function. Afterward, the greenhouse model was calibrated using the temperature,  $CO_2$ -concentration and absolute humidity.

### 3-5-1 AGC dataset and data pre-processing

For calibration, it is important to have data available that is of good quality, i.e. with sufficient measurement frequency and consistent measurements. The used dataset for calibration of the ground-truth model is the dataset created from the Autonomous Greenhouse Challenge (AGC) in 2020 [6]. Multiple measurements were conducted in a real-life greenhouse growing experiment where teams had to autonomously control a tomato greenhouse. Measured states of the crop were: stem elongation, stem thickness, number of trusses and fruits, the weight of the harvested fruits and plant density. Crop measurements were performed roughly every 5 days. Alongside these crop measurements, the dataset includes various parameters describing the greenhouse climate and outside disturbances including but not limited to: temperature, solar radiation, humidity,  $CO_2$ -concentration. These measurements were performed every 5 minutes. Finally, the dataset includes the control inputs applied every 5 minutes, allowing the dataset to be a good fit for calibration of the ground-truth model. For this thesis, the data of team *Automatoes* was selected.

#### Time-span

The dataset contains measurements within a time span of 165 days, starting in mid-December and ending in May. This promises to give a broad range of climate states, inputs and outside weather disturbances as the experiment is performed during cold winter and early summer.

#### NaN data points

The climate and input dataset contains NaN data points. These data points are filled in using the *Matlab fillmissing* function [42], where each NaN value was replaced by the previous non-NaN value. This was considered sufficient as measurements on the climate and the inputs were available for every 5 minutes. For simulation purposes, it could be that a lower sampling time than 5 minutes is necessary. Therefore, the measurements on the climate and inputs are re-sampled to a smaller sampling time. This means that for 5 minutes the climate state is assumed constant and then the current state is updated using the next measurement.

#### Relative Humidity

The dataset contains measurements of the relative humidity ( $RH_{air}$  [%]) and the humidity deficit ( $HD_{air}$  [ $gm^{-3}$ ]), while the ground-truth model used for calibration uses the absolute humidity ( $AH_{air}$  [ $gm^{-3}$ ]). The following conversion equation is used [18]

$$AH_{air} = \frac{RH_{air} HD_{air}}{100 - RH_{air}}. \quad (3-25)$$

#### $CO_2$ -injection

The injected  $CO_2$  ( $\tilde{u}_{CO_2}$ ) in the dataset is defined in [ $kgha^{-1}h^{-1}$ ], while the calibration model makes use of ( $u_{CO_2}$ ) in [ $gs^{-1}m^{-2}$ ]. The conversion is given by

$$u_{CO_2} = \frac{1}{3.6 * 10^4} \tilde{u}_{CO_2}. \quad (3-26)$$

### Disturbances of the weather

The dataset contains all sorts of measurements of the outside weather conditions. The measurements that are disturbances to the greenhouse-crop system can all be directly used for the models;  $T_{out}$ ,  $AH_{out}$ ,  $v_{wind}$  and  $Q_{sun}$ . The outside  $CO_2$ -concentration is assumed to be constant at 370 ppm, which is equal to  $0.688 \text{ gm}^{-3}$ , based on [7].

### Control Inputs

The dataset also contains inputs applied to the greenhouse. Some of them could be imported directly, namely: the window opening ( $\phi_{lee}$  and  $\phi_{win}$ ), the  $CO_2$ -dosing ( $u_{CO_2}$ ) and the screen opening ( $u_{scr1}$  and  $u_{scr2}$ ).

The heating ( $u_{hea}$ ) needs more calculation. The input of the heating in the dataset is in the form of a pipe temperature;  $T_{pipe\_1}$  and  $T_{pipe\_2}$ . The supplied heat through the pipes ( $u_{hea}$ ) is calculated using

$$u_{hea} = c_{pipe1}(T_{pipe\_1} - T_{air}) + c_{pipe2}(T_{pipe\_2} - T_{air}) \quad (3-27)$$

The two parameters  $c_{pipe1}$  and  $c_{pipe2}$  are used in the calibration of the ground-truth model. The pipes can only supply heat to the greenhouse air and cannot cool it down, e.g. if the air temperature is above the pipe temperature, the accompanying contribution of that pipe to  $u_{hea}$  is zero.

### Fruit weight

As mentioned above, the measurements of the fruit weight were only available every 5 days. Since the calibration, see the next subsections, will be performed with a higher frequency, the fruit weight needed to be interpolated. The *Matlab interp1* function with linear interpolation is used for this.

Also, the dataset only contains measurements of the harvested fruit weight from the 65th day onward. Both models, ground truth and control, need to have a fruit weight to be larger than 0. Therefore, the fruit weight is considered to be 5 grams at the beginning of the simulation and on the other days to the 65th day, the fruit weight is interpolated between 5 grams and the weight on the 65th day.

A final adjustment that was made on the fruit weight in the dataset has to do with the dry weight. The measured fruit weight is the fresh weight and not the dry weight which is used in the models from the literature. Therefore, the fresh weight is multiplied by the dry matter content of the fruit, which was on average 8.94%.

Figure 3-3 shows the harvested fruit dry weight of the calibrated simulation, see next subsections, and the measured harvested fruit dry weight from the pre-processed AGC dataset.

### 3-5-2 Calibration of the ground-truth crop model

The harvested fruit weight was used as calibration output, as it was the only measurement available of the crop in the dataset. The harvested fruit weight ( $H_F$ ) is assumed to represent the actual fruit weight in the growing experiment accurately, i.e. it is assumed that all fruits were harvested and that there were no significant losses at the harvesting moments. The harvested fruit dry weight from the dataset ( $H_F$ ) is compared to the simulated harvested fruit dry weight ( $\hat{H}_F$ ) and the parameters of the model are adjusted accordingly, taking into account the measured climate outputs. The parameters that will be used for calibration are shown in Table A-4. The used cost function is given as

$$J(H_F, p) = \sum_{h=1}^L (\hat{H}_F(t, p) - H_F(t))^2 \quad (3-28)$$

$$p^* = \arg \min J(H_F, p)$$

where the time of the last  $H_F$  measurement of the growing season (165 days) is denoted by  $L$ , the sampled time is denoted by  $h$ , the simulated  $H_F$  at time instance  $t$  for parameter vector  $p$  is denoted by  $\hat{H}_F(t, p)$ , the measured  $H_F$  at time instance  $t$  is denoted by  $H_F(t)$  and the calibrated parameter vector  $p$  that gives the lowest difference between the simulated and measured fruit dry weight is denoted by  $p^*$ . The solver that was used is the *Matlab fmincon* [43]. The measured harvested fruit dry weight and the simulated harvested fruit dry weight are shown in Figure 3-3.

As can be seen in the figure, the measured and calibrated harvested fruit dry weight are roughly of the same order of magnitude. Therefore, the calibration was considered successful and the non-linear crop model is considered sufficient.

The calibrated values of the crop model parameters can be found in Table A-4.

### 3-5-3 Calibration of the ground-truth greenhouse climate model

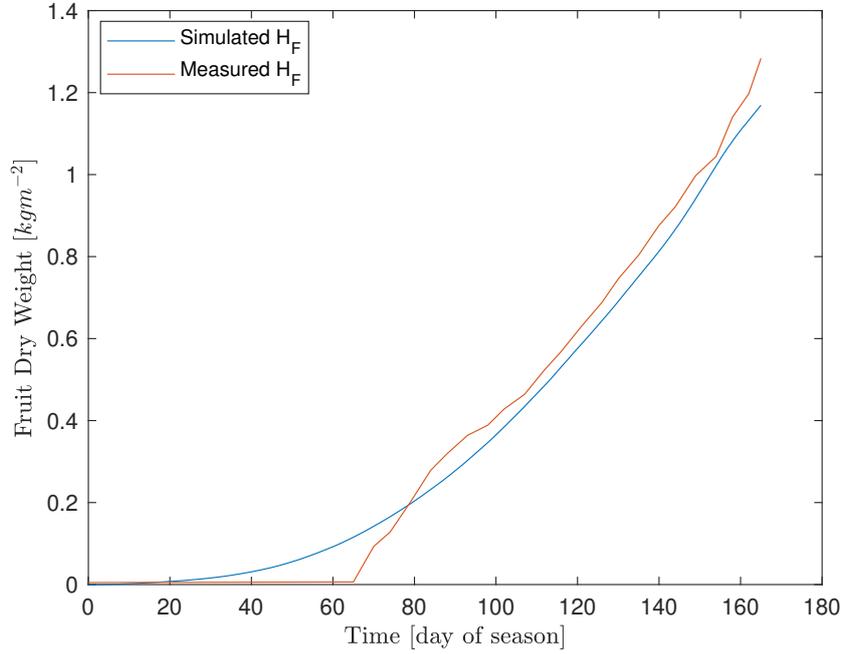
For the calibration of the ground-truth greenhouse model the states used for minimizing the state error are the air temperature ( $T_{Air}$ ),  $CO_2$ -concentration ( $CO_{2,Air}$ ) and the absolute humidity ( $AH_{Air}$ ). This means that the error between the measured climate from the AGC and the modeled climate is minimized while applying the same inputs and weather disturbances. The previously calibrated crop model parameters are also included and are not adjusted any further in this calibration process. The cost function is given by

$$J(\mathbf{T}_{Air}, \mathbf{AH}_{Air}, \mathbf{CO}_{2,Air}, p) = \sum_{h=1}^L w_1 * (\hat{T}_{Air}(t, p) - T_{Air}(t))^2 + (\hat{AH}_{Air}(t, p) - AH_{Air}(t))^2 +$$

$$(\hat{CO}_{2,Air}(t, p) - CO_{2,Air}(t))^2,$$

$$p^* = \arg \min J(\mathbf{T}_{Air}, \mathbf{AH}_{Air}, \mathbf{CO}_{2,Air}, p) \quad (3-29)$$

where the time of the last measurement of the selected calibration interval is denoted by  $L$ , the sampled time is denoted by  $h$ , the simulated  $\mathbf{T}_{Air}$ ,  $\mathbf{CO}_{2,Air}$  and  $\mathbf{AH}_{Air}$  at time instance



**Figure 3-3:** The measured harvested fruit dry weight and the simulated harvested fruit dry weight from the calibrated non-linear crop model, using the measured climate states.

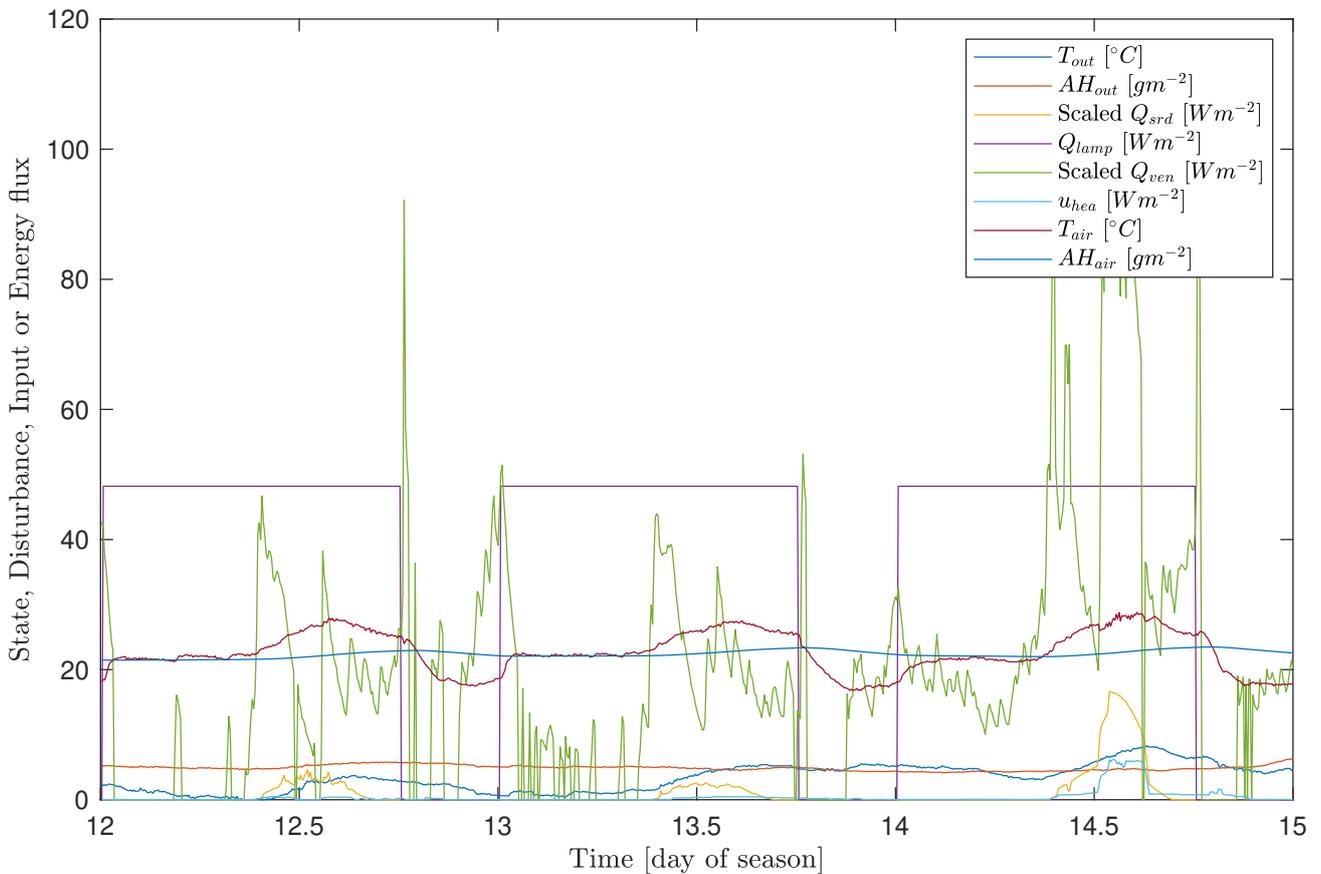
$t$  for parameter vector  $p$  are denoted by  $\hat{T}_{Air}(t, p)$ ,  $\hat{AH}_{Air}(t, p)$ ,  $\hat{CO}_{2,Air}(t, p)$ , the measured  $T_{Air}$ ,  $CO_{2,Air}$  and  $AH_{Air}$  at time instance  $t$  are denoted by  $T_{Air}(t)$ ,  $CO_{2,Air}(t)$  and  $AH_{Air}(t)$  and the calibrated parameter vector  $p$  that gives the lowest difference between the simulated and measured fruit dry weight is denoted by  $p^*$ . The temperature and absolute humidity are roughly of the same order of magnitude, so they have roughly the same effect on the cost. However, the temperature is considered to be the most important factor as the crop-growth parameters are directly influenced by peak and average temperatures [44], therefore different factors ( $w_1$ ) for the temperature were tried and this resulted in a weight of 1.3 to give the best calibration results.

### 3-5-4 Calibration interval

Throughout the season the weather conditions and the applied inputs to the greenhouse system strongly differ. During winter the peak radiation of the sun is around  $100 W/m^2$ , while during summer solar radiation can reach  $1000 W/m^2$ . Also, the average outside temperature and humidity fluctuates throughout the year. Using a 165-day simulation with a 5-minute sampling time would take too long to calibrate. Therefore a smaller calibration interval of 5 days from one moment in time onward was used for calibration. Choosing this moment to be during the beginning of the growing season, in mid-December at simulation day 12, resulted in an underestimation of the temperature later on in the season (May around simulation day 150). Choosing the start day of the calibration later on in the season, in May around simulation day 150, resulted in an overestimation of the temperature earlier on in the season. This

is can be concluded that for season-wide simulation, multiple calibration moments have to be selected. For this, four calibration windows throughout the growing season were selected that show representative values for the inputs and disturbances and that show enough fluctuations to capture a wide variety of greenhouse climate dynamics.

The calibration windows of the 3 days that were used for the calibration are: days 12-15, days 72-75, days 130-133 and days 149-152. This means that the total cost of one calibration round is the sum of the separate simulations of the four windows. The relevant climate states, disturbances, inputs and energy fluxes of the first window are plotted in Figure 3-4. The other three windows are plotted Figure A-1, Figure A-2 and Figure A-3 in the appendix.



**Figure 3-4:** Window on the 12th day.  $Q_{srd}$  and  $Q_{ven}$  are scaled down.

### 3-5-5 Calibration parameter vector and bounds

The parameter vector used for this calibration can be found in Table A-5. The parameter vector was scaled down according to the bounds, e.g. the upper bound was equal to one and the lower bound was equal to zero. This was done to improve the optimization algorithm used for calibration, which could become inaccurate when the parameter vector contained considerably large fluctuations of the order of magnitude between the values. Also, the bounds

were chosen such that the algorithm looked at values 100 times smaller and 100 larger than the original taken from the literature.

### 3-5-6 Solver

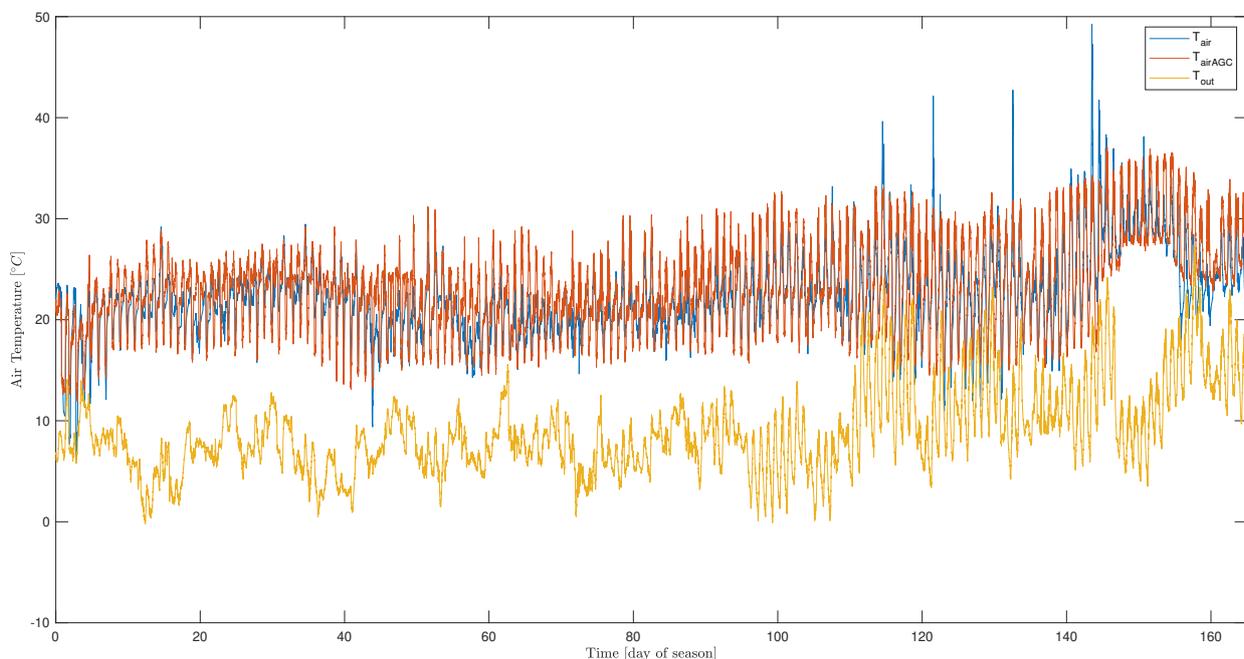
To attempt to find the global minimum of the cost function, the function *Matlab particleswarm* from the *Global Optimization Toolbox* was used [45].

### 3-5-7 Calibration results

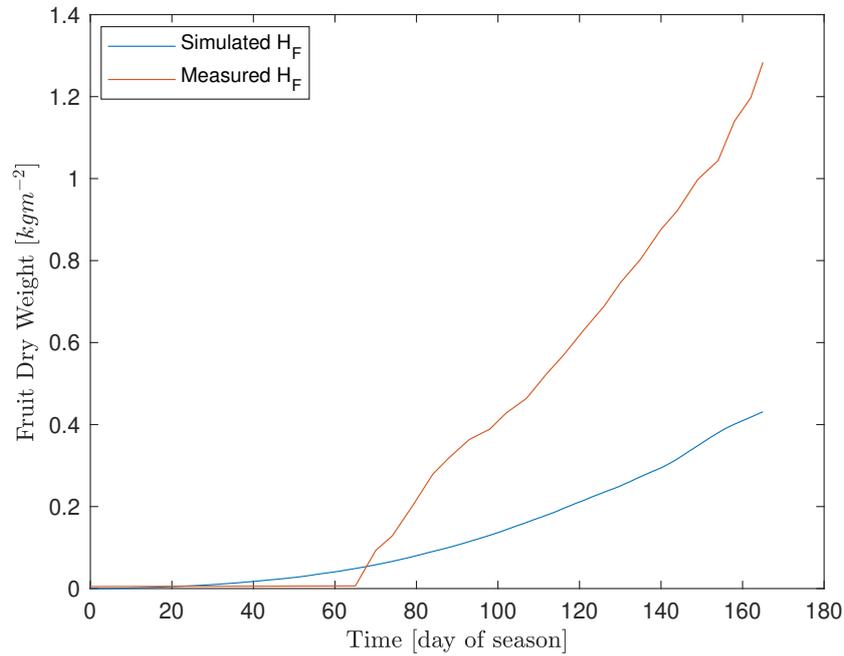
The calibrated values of the greenhouse climate model parameters can be found in Table A-5.

#### *Season-wide simulation*

The goal was to achieve accurate season-wide simulation with the same inputs and disturbances used for the calibration. Figures 3-5, A-5 and A-6 show season-wide simulations of the ground-truth model with the calibrated parameters. It can be seen that multiple high peaks in temperature are present. These peaks can be explained by the fixed inputs that were directly imported from the AGC dataset. It is assumed that these peaks will be prevented by the optimal controller, as it will optimize crop growth which is not enhanced with such high temperatures. Also, the simulated  $CO_2$ -concentration is inaccurate in comparison with the measured  $CO_2$ -concentration. The next paragraph will explain the potential causes and the adjustments to the  $CO_2$ -dynamics.



**Figure 3-5:** Season-wide temperature simulation of the calibrated ground-truth model, compared to the measured and outside temperature.

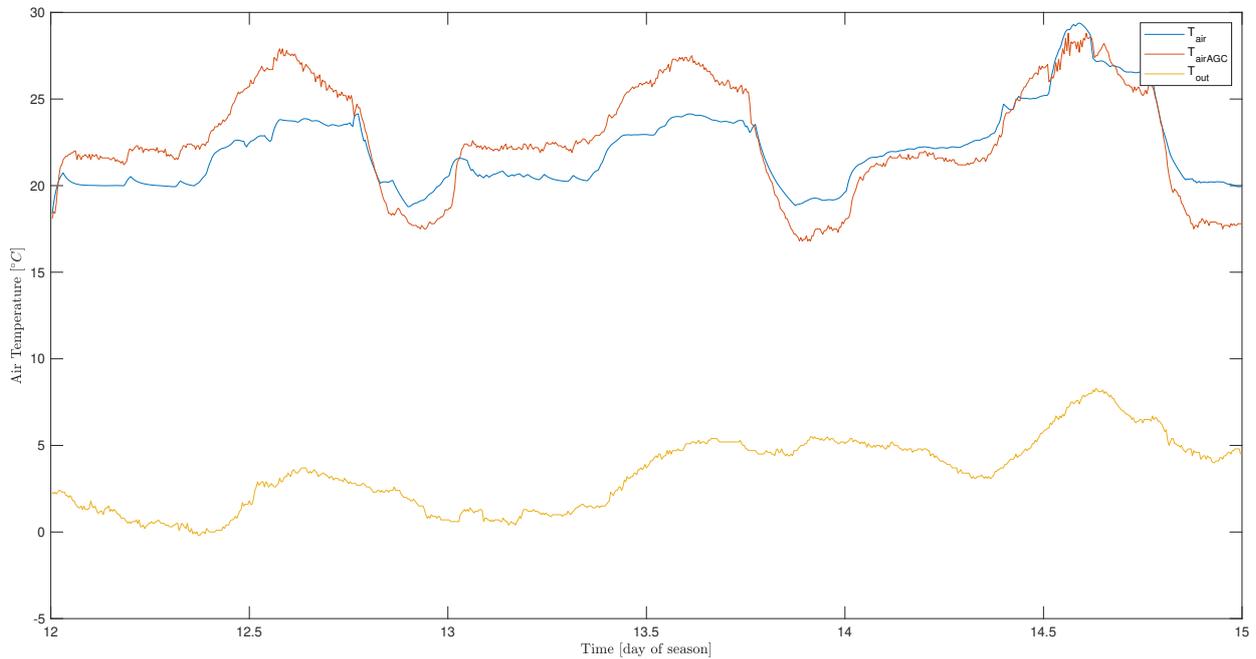


**Figure 3-6:** The measured harvested fruit dry weight and the simulated harvested fruit dry weight from the calibrated non-linear crop model, with the calibrated ground-truth model.

Figure 3-6 shows the resulting simulated harvested fruit dry weight against the fruit dry weight achieved during the AGC.

### *3-day simulation*

Most optimal controllers in the literature use a prediction horizon in the optimal controller that is much smaller than the full season-wide time span of 165 days. Figures 3-7, A-7, A-8, A-9, A-10 and A-11 show the simulation results of the climate states starting from the measured climate state and control inputs, at the beginning (day 12) and near the end of the growing season (day 130).



**Figure 3-7:** 3-day temperature simulation of the calibrated ground-truth model starting from day 12, compared to the measured and outside temperature.

### 3-5-8 $CO_2$ -concentration calibration

Figure A-6 shows the simulated and measured  $CO_2$ -concentration. It can be seen that after calibration the simulated  $CO_2$ -concentration does not represent the measured  $CO_2$ -concentration. Applying a heavier weight in the cost function to the  $CO_2$  state error also could not improve this.

The balance that describes the dynamics of the  $CO_2$ -concentration consists of three factors, see (3-23). The first factor, representing photosynthesis, is too important for the crop model and calibration, to be tuned again. The second factor, representing the ventilation, is neither a candidate for a new calibration, as the ventilation strongly influences the temperature and humidity, both of which are considered more important than the  $CO_2$ -concentration [44]. However, the third and final remaining factor, representing  $CO_2$ -injection, could be calibrated with a slight adjustment. A final calibration has been performed with the following cost function

$$J(CO_{2,Air}, CO2\_corr) = \sum_{h=1}^L (\hat{CO}_{2,Air}(t, CO2\_corr) - CO_{2,Air}(t))^2, \quad (3-30)$$

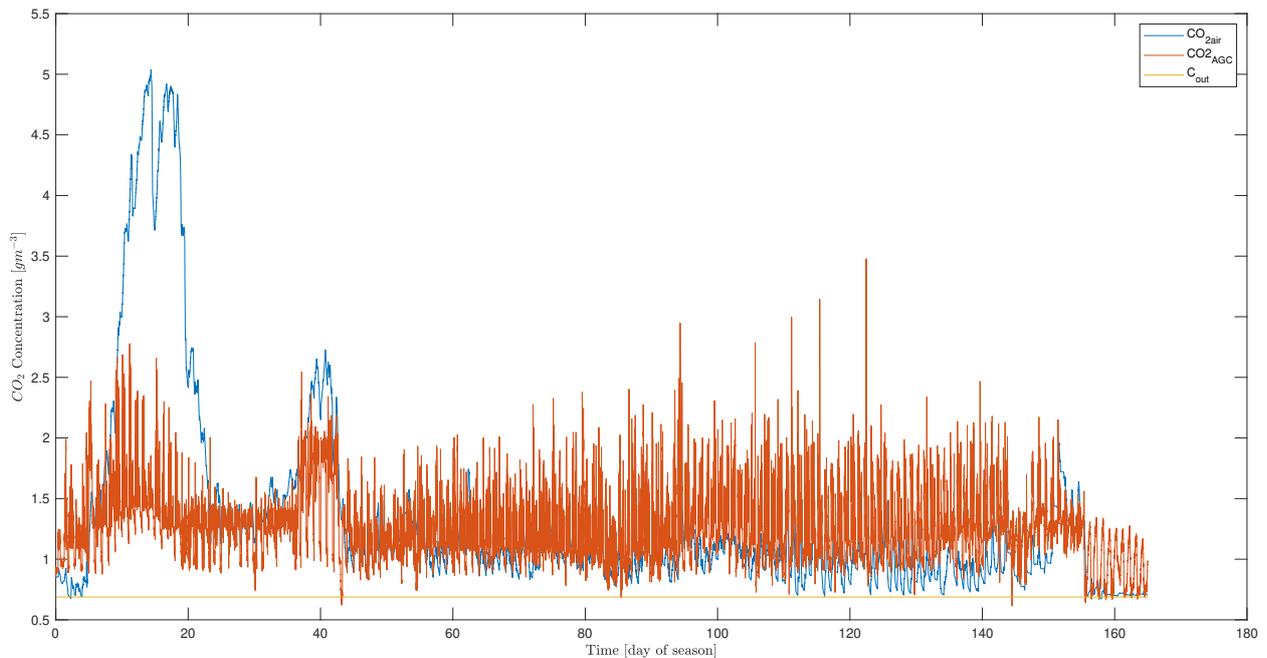
$$CO2\_corr^* = \arg \min J(CO_{2,Air}, CO2\_corr)$$

where  $CO2\_corr$  [-] denotes the correction factor for the  $CO_2$ -injection. This means that in (3-23) the  $CO_2$ -injection ( $u_{CO_2}$ ) is replaced by  $CO2\_corr * u_{CO_2}$ .

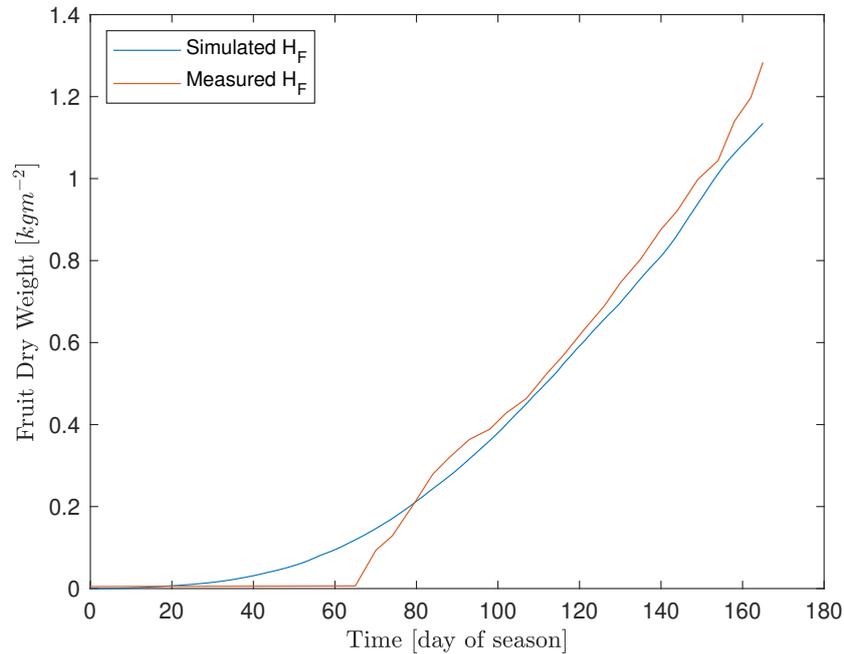
The calibration windows of 3 days that were used for the calibration are: days 0-3, days 72-75, days 130-133 and days 149-152. This means that the total cost of one calibration round is the sum of the separate simulations of the four windows. Previously, day 12 was used, but for the calibration of solely the  $CO_2$ -concentration day 0 was better as it showed more suitable inputs and state fluctuations.

The correction factor  $CO_2\_corr$  for the  $CO_2$ -injection that gave the best results according to the cost function is  $1.79 * 10^4$ . This value is quite large and therefore it is possible that there was an error in the reporting of the dataset of the AGC. It can be seen in Figure 3-8 that the simulated  $CO_2$ -concentration still does not accurately follow the measured  $CO_2$ -concentration. However, by looking solely at the fruit dry weight, see Figure 3-9, the adjusted  $CO_2$ -dynamics are considered sufficient.

The simulated fruit dry weight per development stage is shown in Figure A-4, it is similar to the results of the original model [20, Fig. 5.5].



**Figure 3-8:** Season-wide  $CO_2$ -concentration simulation of the calibrated ground-truth model including the adjusted  $CO_2$ -dynamics, compared to the measured and outside  $CO_2$ -concentration.



**Figure 3-9:** The measured harvested fruit dry weight and the simulated harvested fruit dry weight from the calibrated non-linear crop model, with the calibrated ground-truth model and with the adjusted  $CO_2$ -dynamics.

### 3-5-9 Physical meaning of the parameters

It has to be noted that some of the parameters that are tuned by the calibration process are not realistic any more. For example, the height of the greenhouse after calibration is  $86m$ , which is obviously wrong. However, since this simulation will be used as ground truth and only for giving realistic climate and crop states back, it is considered to be sufficient. It can not give accurate feedback on internal processes, but this is not of considerable interest as most internal processes are also not measurable in real-life growing operation.

## 3-6 Steady-state fruit weight

It is important to note from Figure A-4 that the states representing the fruit weight reach steady-state around day 76. This means that the crop model can provide meaningful predictions of the fruit weight from day 76 onward, indicating the start of the productive phase [20]. This is also the day on which the first profit can be made on an interval of one day, i.e. the yield is larger than the costs made. To prevent the controller from giving a conservative solution during the first 76 days, i.e. a purely profit-based controller does not induce any inputs to reduce costs because the yield is not high enough to make a profit. Therefore, the proposed profit controller, see Chapter 4, will only be deployed from day 76 onward. Other ways to deal with this will also be presented in the discussion, see Chapter 6.

Until the crop development reaches steady-state, the dynamics representing the fruit growth

are also evolving quickly. This means that the crop is showing more complex non-linear dynamics and these are, therefore, harder to represent with a linear state-space system. Selecting this steady-state growing stage will increase the chance to obtain a useful identification of a linear state-space system. Therefore, the research presented in this thesis on system identification will only take place in steady-state growing stages.

### 3-7 Leaf and stem weight

The leaf weight was not measured during the AGC and was, therefore, not present in the acquired dataset. This led to the leaf weight not being used in the calibration process. The simulated leaf weight growth is shown in Figure 3-10 and is as expected: it rises towards the value where the maximum Leaf Area Index (LAI) is reached and then it stays there. The simulation of the leaf weight is the same order of magnitude as other results from the literature show and therefore, the model of the leaf weight is considered sufficient [46].

The ground-truth simulation is only valid for the generative growth of the crop; the plant is already grown tall and is not a small plantlet. Therefore, the leaves and stem already have representative weight at the beginning of the simulation, respectively  $40g$  and  $30g$ , based on [7, Tab. 3.3].

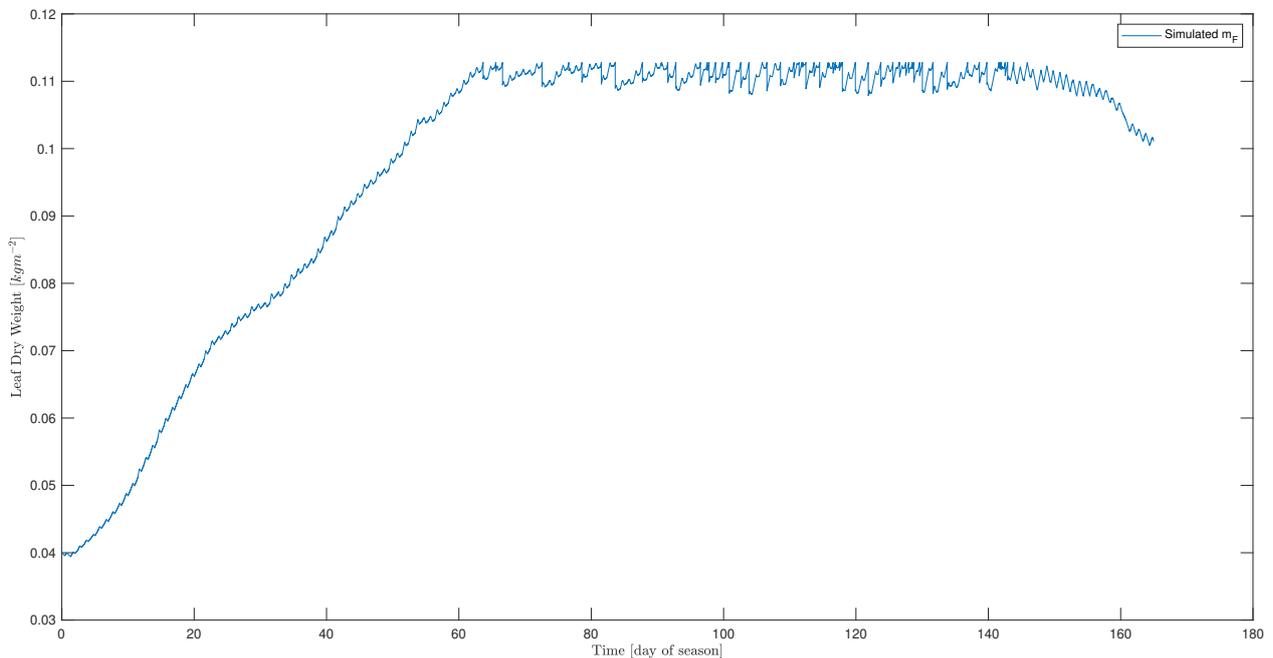
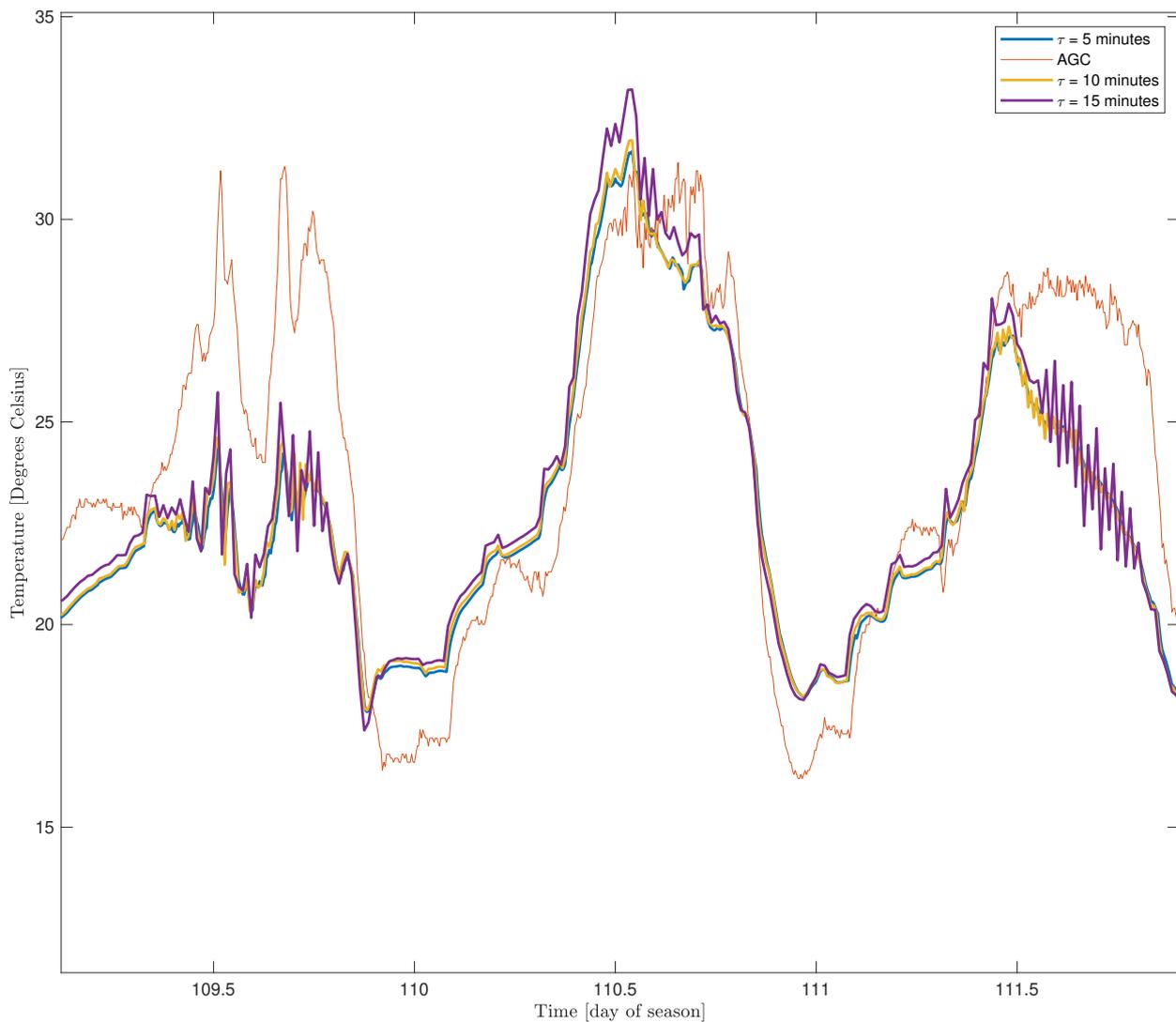


Figure 3-10: The simulated leaf weight.

### 3-8 Sampling time

Three simulations with different sampling times are performed, to check the limits of the sampling time. From Figure 3-11, it can be seen that the simulation becomes unstable for sampling times larger equal to and larger than 10 minutes and that the simulation does not follow the AGC data sufficiently. Therefore, the sampling time is chosen to be 5 minutes.



**Figure 3-11:** The air temperature for different sampling times, compared to the measured AGC data. The difference between the simulated temperature and the measured temperature in the AGC can be ignored.

### 3-9 Chapter summary and conclusion

This chapter first showed the formal system description that is needed to build a ground-truth simulator of the greenhouse-crop system. The system's states corresponding to the climate are the air temperature, humidity,  $CO_2$ -concentration and the 24-hour average temperature. The states that describe the crop are more extensive, they consist of the number of fruits per development stage and the assimilate weight in the buffer, in the 8 development stages of the fruit, in the stem and in the leaves. The disturbances are the outside temperature, humidity,  $CO_2$ -concentration, solar radiation and wind speed. Not all states can be measured; only the mass of the fruits, the climate factors and the harvested fruit and leaf weight can be measured. It is assumed that the disturbances are perfectly predictable and that the output measurements do not contain noise. Finally, the inputs to the system are the temperature of two heating pipes, the opening of the windows, the LED lighting, the usage of climate and blackout screens and the injection of  $CO_2$ .

The climate states evolve according to energy, vapor and mass fluxes, described by corresponding balances. The tomato crop grows by generating sugars via photosynthesis. These sugars, also known as assimilates, then spread by going to the stem, leaves and fruits. The development of fruits is modeled according to the fixed boxcar train'-model, i.e. the assimilates flow from one development stage to another [7]. If the assimilates move out of the last development stage they are considered harvested. The crop and climate dynamics are integrated together into one non-linear greenhouse-crop model, that will be used for a season wide (165 days) simulator. The parameters of the model in the simulator are calibrated. To make the simulator a realistic ground truth, the input-output data from a real-life growing experiment of the Autonomous Greenhouse Challenge are used [6]. The sampling time is chosen to be 5 minutes because larger sampling times cause the ground truth to become unstable. Finally, the research on system identification presented in this thesis on system identification will only take place in steady-state growing stages. The rationale behind this is twofold. First, the dynamics representing the fruit growth are evolving quickly until the fruit weight reaches steady state and are, therefore, harder to accurately represent as a useful linear state-space system. Second, profit, which will be the main objective element in the proposed controller, can only be made after the crop reaches steady-state around day 76 and, therefore, the controller will be deployed after day 76.

# Greenhouse Climate Control

The previously presented ground-truth simulator can be used to create output data and can be used to perform experiments with control techniques. While it is a model rich in information, it is also computationally expensive and non-linear. For these reasons, it is not suitable for optimal control techniques and conventional optimization solvers. Therefore, a Linear Time Invariant (LTI) state-space system will be specified and will be estimated using system identification. From now on a LTI system is denoted solely as ‘linear’. The linear state-space representation will be created when the crop is more mature, i.e. its dynamics reach steady-state, as explained in Section 3-6. The rationale behind this is that in this crop growth stage, the non-linear dynamics are evolving less and are, therefore, better to accurately represent in a linear fashion. After the system identification procedure is presented in this chapter, the control problem is defined. Moreover, an MPC-based controller which optimizes profit is proposed for this control problem. An additional constraint, e.g. the total fruit weight is maintained, is added to make sure that trivial solutions and conservative control strategies are avoided.

## 4-1 System identification of the linear state-space system

To identify the linear state-space system for the MPC-based controller, the Prediction-Error Method (PEM) is used, as described in this section. This section also presents the data processing steps and the definition of the performance of the identification process.

### 4-1-1 Linear state-space system

The model that will be used to locally represent the non-linear dynamics in the form of a linear state-space system and is given as

$$\begin{aligned} \mathbf{x}[i+1] &= \mathbf{A}\mathbf{x}_L[i] + \mathbf{B}\mathbf{u}_L[i] \\ \mathbf{y}[i] &= \mathbf{C}\mathbf{x}_L[i] + \mathbf{D}\mathbf{u}_L[i] \end{aligned} \tag{4-1}$$

where the states are denoted by  $\mathbf{x}_L \in \mathbb{R}^{n_{x_L}}$ , the outputs are denoted by  $\mathbf{y}_L \in \mathbb{R}^{n_{y_L}}$  and the inputs are denoted by  $\mathbf{u}_L \in \mathbb{R}^{n_{u_L}}$ , with  $n_{\{x_L, u_L, y_L\}}$  being the respective sizes of the vectors. Note that the states  $\mathbf{x}_L$  of the linear state-space system are not equal to the actual states of the system, rather they are arbitrary auxiliary variables of a chosen model order. The linear state-space matrices are denoted by  $\mathbf{A} \in \mathbb{R}^{n_{x_L} \times n_{x_L}}$ ,  $\mathbf{B} \in \mathbb{R}^{n_{x_L} \times n_{u_L}}$ ,  $\mathbf{C} \in \mathbb{R}^{n_{y_L} \times n_{x_L}}$  and  $\mathbf{D} \in \mathbb{R}^{n_{y_L} \times n_{u_L}}$ .

The input-vector  $\mathbf{u}_L$  used for the linear state-space system consists of controllable inputs  $\mathbf{u}_{con}$  and uncontrollable disturbances  $\mathbf{d}$ :

$$\mathbf{u}_{con} = \begin{bmatrix} T_{pipe,1} \\ T_{pipe,2} \\ \phi_{lee} \\ \phi_{win} \\ u_{CO2} \\ u_{led} \\ u_{scr1} \\ u_{scr2} \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} T_{out} \\ H_{out} \\ CO_{2,out} \\ Q_{sun} \\ v_{wind} \end{bmatrix}, \quad \mathbf{u}_L = \begin{bmatrix} \mathbf{u}_{con} \\ \mathbf{d} \end{bmatrix}, \quad (4-2)$$

where the inputs and disturbances are explained in Table 3-2 and Table 3-4, respectively.

The output vector does not contain all the outputs of the original non-linear system, because they are not all measurable in real life, as will be explained in Subsection 4-2-3. The output vector is defined as

$$\mathbf{y}_L = \begin{bmatrix} T_{Air} \\ AH_{air} \\ CO_{2,air} \\ T_{c24} \\ \Delta H_F \\ m_F \\ m_L \end{bmatrix}, \quad (4-3)$$

where the outputs are explained in Table 3-3. Instead of using the total harvested dry weight ( $H_F$ ) directly as a prediction, the change in this output is used, denoted by  $\Delta H_F$ . The total harvested dry weight will create a trajectory that is a linear upward line, representing the integration of the changes. It is hard to fit a state-space model on such a trajectory and, therefore, it is assumed that the resulting pattern of the alteration will increase the fit of the linear state-space system. This will also result in more accurate predictions.

#### 4-1-2 Data processing for system identification and making prediction

For the purpose of increasing the performance of the linear state-space prediction model two data processing steps are introduced, namely scaling and bias.

##### Scaling

The inputs and outputs of the greenhouse-crop models each have their own units and minimum and maximum value. For example, the temperature varies between 6 and 40, while the change

in fruit weight is in order of magnitude of  $10^{-5}$ . This may cause bad results, as the values are at different scales [47]. To prevent any scaling issues, both the input and output data are scaled between 0 and 1. This is achieved by taking each column of the input and output data separately and setting 0 for the minimum value and 1 for the maximum value, where every other value is scaled between 0 and 1. The following equation is used for this,

$$\mathbf{u}_{\mathbf{L}scaled}(z) = \frac{\mathbf{u}_{\mathbf{L}original}(z) - \min_{\mathbf{u}_{\mathbf{L}original}(z)}}{\max_{\mathbf{u}_{\mathbf{L}original}(z)} - \min_{\mathbf{u}_{\mathbf{L}original}(z)}} \quad \forall z \in \begin{bmatrix} T_{pipe,1} \\ T_{pipe,2} \\ \phi_{lee} \\ \phi_{win} \\ u_{CO2} \\ u_{led} \\ u_{scr1} \\ u_{scr2} \end{bmatrix} \quad (4-4)$$

where  $\mathbf{u}_{\mathbf{L}original}$  is the complete dataset of all measured inputs with each column containing measured data samples of the input variables,  $\mathbf{u}_{\mathbf{L}scaled}$  is the complete dataset of all scaled inputs,  $z$  denotes the input variable,  $\min_{\mathbf{u}_{\mathbf{L}original}(z)}$  denotes the minimum value of all measured input variable  $z$  and  $\max_{\mathbf{u}_{\mathbf{L}original}(z)}$  denotes the maximum value of all measured input variable  $z$ . The same equation is used for the output data, but  $\mathbf{u}_{\mathbf{L}}$  is replaced by  $\mathbf{y}_{\mathbf{L}}$ . It should be noted that for calculating the yield and costs, the outputs are scaled back.

## Bias

In order to even further improve system identification performance, bias is removed from the outputs. This means that the original output  $\mathbf{y}_{\mathbf{L}}[i]$  is replaced with a new output  $\delta\mathbf{y}_{\mathbf{L}}[i] = \mathbf{y}_{\mathbf{L}}[i] - \mathbf{y}_{L,bias}$ . The bias factor,  $\mathbf{y}_{L,bias}$  is equal to the first value of the output trajectory that is used for the system identification. The alteration makes the system give predictions around a specific operating point. It is chosen to first perform the scaling process, after which the scaled bias is removed. It should be noted that for plotting the output trajectories and for calculating the yield and costs, the bias is added back to the outputs.

### 4-1-3 Prediction-Error Method (PEM)

The Prediction-Error Method (PEM) [48] fits the inputs, including and measurable disturbances, and outputs onto a pre-defined system representation, which is in our case the linear state-space form as given in the previous paragraph.

To obtain estimations for the linear state-space matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{D}$ , the prediction-error method minimises the following objective function:

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}} = \sum_{k=1}^N \mathbf{e}[k]^T \mathbf{Q}_{\mathbf{r}}[k] \mathbf{e}[k], \quad (4-5)$$

where the error between the predicted output by the linear state-space system and the actual output is denoted by  $\mathbf{e}$  and the number of samples is denoted by  $N$ . Weight matrix  $\mathbf{Q}_{\mathbf{r}}$  is a diagonal matrix with equal weights on all errors.

### PEM: training data

For the system identification, 14 days of input and output training data are collected from 6 simulation experiment runs. Each experiment run has its own set of inputs. The reason for creating multiple experiment datasets is that this will excite the system properly within the same period. Using only a single dataset could result in over-fitting. In real-life growing operations, this can be considered as capturing data from separate growing years, where the crop was more or less equally developed. The input data is created by taking the baseline inputs of the grower and adding noise on top of it, where each experiment has its own noise component. Variation between the noise components lies in the frequency and amplitude, where large chattering input signals are avoided.

### PEM algorithm and model order

The algorithm used for this is the *Matlab System Identification Toolbox: ssest*-algorithm. The goal is to get a linear state-space system that performs sufficiently in predicting the future with new inputs, therefore, the *prediction*-option was used for the *ssest* process. The model order  $n_{x_L}$  was chosen to be 10, as it is a reasonable compromise between prediction accuracy and computational costs.

#### 4-1-4 Performance definition and validation of the identified system

The prediction performance fit is defined as 100% minus the normalized root mean squared error (NRMSE) between the predicted outputs and the measured outputs, and is given as:

$$fit(z) = 100\% - \frac{\|\mathbf{y}_L(z) - \hat{\mathbf{y}}_L(z)\|}{\|\mathbf{y}_L(z) - \bar{\mathbf{y}}_L(z)\|} 100\%, \quad (4-6)$$

where the realized output is denoted by  $\mathbf{y}_L$ , the predicted output  $\hat{\mathbf{y}}_L$  is denoted by and the mean of the realized output is denoted by  $\bar{\mathbf{y}}_L$ , for output variable index  $z$ . The closer this performance measure to 100 %, the better.

The results of the system identification will be presented in Section 5-1. For this, the system is identified on the training data and then validated on the validation data in a validation experiment. The validation dataset is different from the training dataset, but it is of similar pattern, amplitude and frequency.

Before the actual validation experiment is started, one day of simulation with the Luenberger state observer included, see Subsection 4-3-3. The reason for this is to make sure the validation experiment starts with a sufficient internal initial state, resulting from one day of state estimation. This first day will be excluded from the performance metric as presented in this subsection.

## 4-2 Problem definition

The linear state-space representation of the non-linear ground-truth model will be used in the proposed controller that will be presented next. In this section, the grower's objective, i.e. making a profit by selling harvested fruits and using costly resources, and a summary of the literature on greenhouse climate control will be presented. Then, the formal problem definition will be given, which is converted into an optimal control problem in the form of an MPC-based controller.

### 4-2-1 Grower's objective

The objective that growers want to optimize is profit. Revenue is made by selling the harvested yield. The main costs of a grower are labor and resource costs. Labor costs will not be considered for this problem, i.e. harvesting the tomatoes is free. The resource costs depend directly on the inputs applied to the system:

- heating costs from using gas to create a warm greenhouse climate
- $CO_2$ -costs from injecting  $CO_2$  into the greenhouse
- electricity costs from turning on the artificial lighting

In modern-day greenhouses, the grower selects temperature and humidity setpoints, and the inputs are adjusted automatically by a climate controller. The inputs are chosen such that they counteract the disturbances from outside, introduced in Chapter 3. This sounds like a simple tracking control problem, but the catch is that it takes years of training and experience to know what climate setpoints are best under which conditions. These climate setpoints, for example, depend on the state of the plant, and the outside weather patterns and it requires frequent checks to be performed. This makes it especially difficult to build greenhouses in unfamiliar weather patterns, e.g. in other remote parts of the world.

A second problem that the grower faces is that of efficient resource usage. The main source of heat in a modern-day greenhouse is gas, which is expensive and not good for the environment. Therefore we must come up with better, more efficient control techniques that make optimal use of the available resources, under any circumstance.

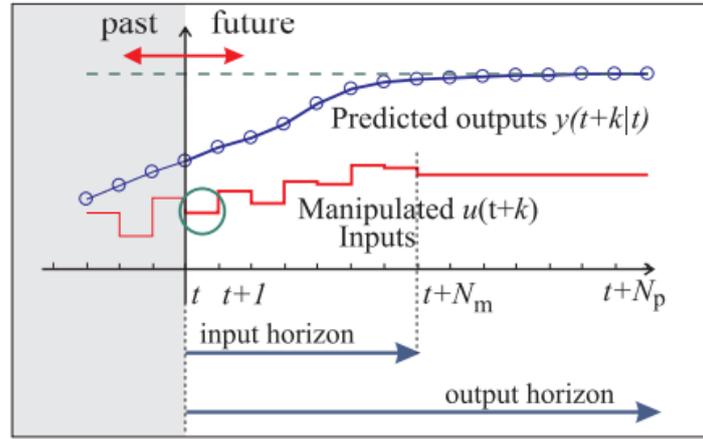
To overcome these difficulties, new alternatives to this non-optimal tracking-like control strategy are proposed and will be presented next.

### 4-2-2 Literature on greenhouse climate control

The reader is referred to the preceding literature survey for a complete overview of all the greenhouse climate control methods in the literature. The next paragraphs give a summary of the relevant greenhouse climate control techniques studied in the literature, ending with the shortcomings.

## Model Predictive Control (MPC)

Most proposed control techniques for greenhouse climate in the literature make use of Model Predictive Control (MPC). This control technique optimizes on a cost function while satisfying specified constraints on the system's inputs, outputs and states [5]. In MPC, the set of optimal control inputs ( $\mathbf{u}^*$ ) is calculated for a predefined control horizon ( $N$ ), but only the first control step of this set is actually applied to the real system. A model  $\mathbf{F}$  is used to predict the states of the system due to the manipulated inputs ( $\mathbf{u}$ ) and uncontrollable disturbances ( $\mathbf{d}$ ). This model can either be linear or non-linear, and additional stochastic or robustness properties can be added. The calculation of the optimal control inputs will then be performed in every update step. The updated state of the system is determined by using the output function  $\mathbf{G}$ .



**Figure 4-1:** A general MPC scheme, only the first input of  $\mathbf{u}(t+k)$  is applied [5, Fig. 2.]

A general MPC framework with an arbitrary cost function is shown in (4-7) and Figure 4-1 shows a general MPC scheme [49]. Note that here the system is discretized, i.e. time is not continuous but it is transformed into discrete timesteps.

$$\begin{aligned}
 \min_{\mathbf{u}} \quad J_N(\mathbf{x}, \mathbf{u}) &= \sum_{i=t}^{N-1} \ell(\mathbf{x}[i], \mathbf{u}[i]) + \mathbf{V}_f(\mathbf{x}[N], \mathbf{u}[N]) \\
 \text{s.t.} \quad \mathbf{x}[i+1] &= \mathbf{F}(\mathbf{x}[i], \mathbf{u}[i], \mathbf{d}[i]), \quad \forall i \in \{t, \dots, t+N-1\}, \\
 \mathbf{y}[i] &= \mathbf{G}(\mathbf{x}[i], \mathbf{u}[i], \mathbf{d}[i]), \quad \forall i \in \{t, \dots, t+N-1\}, \\
 \mathbf{x}_0 &= \hat{\mathbf{x}}[t] \\
 \mathbf{b}_{\mathbf{u}}^{low}[i] &\leq \mathbf{u}[i] \leq \mathbf{b}_{\mathbf{u}}^{upp}[i], \quad \forall i \in \{t, \dots, t+N-1\}, \\
 \mathbf{b}_{\mathbf{x}}^{low}[i] &\leq \mathbf{x}[i] \leq \mathbf{b}_{\mathbf{x}}^{upp}[i], \quad \forall i \in \{t, \dots, t+N-1\}, \\
 &\text{Additional constraints,}
 \end{aligned} \tag{4-7}$$

with states  $\mathbf{x} \in \mathbb{R}^{n_x}$ , inputs  $\mathbf{u} \in \mathbb{R}^{n_u}$ , outputs  $\mathbf{y} \in \mathbb{R}^{n_y}$  and disturbances  $\mathbf{d} \in \mathbb{R}^{n_d}$ , with  $n_x, n_u, n_y, n_d$  being the respective sizes of the vectors. The objective  $J_N(\mathbf{x}, \mathbf{u}) \in \mathbb{R}$  has to be minimized. The stage cost is presented as  $\ell$  and it is updated each prediction time step  $i \in \mathbb{Z}_{>0}$ . The terminal cost is presented as  $\mathbf{V}_f(\mathbf{x}[N], \mathbf{u}[N]) \in \mathbb{R}^{n_x}$ , and it is defined as the cost at the end of

the prediction horizon ( $i = N$ ), which is also recomputed every time step. Initial state  $\mathbf{x}_0$  is equal to the estimated state  $\hat{\mathbf{x}}[t]$ , on current timestep  $t \in \mathbb{Z}_{>0}$ .  $\mathbf{G}$  generates the output  $\mathbf{y}[i]$ , given  $\mathbf{x}[i]$ ,  $\mathbf{u}[i]$  and  $\mathbf{d}[i]$ . The dynamical model  $\mathbf{F}$  and the associated output function  $\mathbf{G}$  can either be linear or non-linear. If they are linear, they are in the form of:

$$\begin{aligned}\mathbf{F}(\mathbf{x}[i], \mathbf{u}[i], \mathbf{d}[i]) &= \mathbf{A}\mathbf{x}[i] + \mathbf{B}_1\mathbf{u}[i] + \mathbf{B}_2\mathbf{d}[i] \\ \mathbf{G}(\mathbf{x}[i], \mathbf{u}[i], \mathbf{d}[i]) &= \mathbf{C}\mathbf{x}[i] + \mathbf{D}_1\mathbf{u}[i] + \mathbf{D}_2\mathbf{d}[i],\end{aligned}\tag{4-8}$$

with  $\mathbf{A} \in \mathbb{R}^{n_x \times n_x}$ ,  $\mathbf{B}_1 \in \mathbb{R}^{n_x \times n_u}$ ,  $\mathbf{B}_2 \in \mathbb{R}^{n_x \times n_d}$ ,  $\mathbf{C} \in \mathbb{R}^{n_y \times n_x}$ ,  $\mathbf{D}_1 \in \mathbb{R}^{n_y \times n_u}$  and  $\mathbf{D}_2 \in \mathbb{R}^{n_y \times n_d}$ .

As explained before, in this thesis a linear prediction model will be used in the proposed control strategy. The reason for this is that for convex objective functions based on linear models, there are methods available that will guarantee that a local optimum is also a global optimum [19]. Also, using this linear model for optimization is less computationally expensive than using optimal-estimation methods for non-linear models. Note that in this work the controllable inputs ( $\mathbf{u}$ ) and uncontrollable disturbances ( $\mathbf{d}$ ) are concatenated into ( $\mathbf{u}_L$ ), see Subsection 4-1-1.

The general control algorithm for MPC is given in algorithm 1.

---

#### Algorithm 1: MPC Algorithm

---

**Input:** System function  $\mathbf{F}(\mathbf{x}[i], \mathbf{u}[i], \mathbf{d}[i])$ , prediction horizon  $N$  and constraint sets for  $\mathbf{x}[i]$  and  $\mathbf{u}[i]$ .

- 1) Obtain initial state estimate  $\hat{\mathbf{x}}[i]$ .
  - 2) Solve (4-7) for the optimal input sequence  $\mathbf{u}^*$ .
  - 3) Apply only the first input  $\mathbf{u}[i] = \mathbf{u}_0^*$ .
  - 4) Set  $k$  to  $k + 1$ .
  - 5) Return to 1).
- 

## Hierarchical Control

Another relevant concept for greenhouse climate control is the hierarchical control framework. It divides control problems into smaller control layers, to improve computational performance. This framework fits the greenhouse-crop system because of the different time scales and time delays of the greenhouse-crop system [12]. The greenhouse sub-system acts on faster (minutes to hours) dynamics, rather than the much slower crop sub-system (days to weeks). An example of hierarchical control in the scope of greenhouse-crop control, is the proposed control method of van Duijn [18]. She decomposed the control problem into two sub-problems. The upper layer solely creates the climate setpoints, acting on a larger time scale to optimize crop growth, incorporating the slow dynamics. The lower layer solely tracks the setpoints while optimizing resource usage on a smaller scale, incorporating the fast dynamics. The control strategy used for the upper layer was Data Enabled Predictive Control (DEEPC) [50]. DeePC takes measured input and output trajectories and based on that, tries to find the optimal input trajectory for the next period. The DeePC algorithm resulted in a higher net profit [ $\text{€}/\text{m}^2$ ] compared to the benchmark reference grower. The proposed control method in

this thesis work does not split the control problem into two, but rather tries to optimize the slow and fast dynamics simultaneously.

### Literature shortcomings

One of the findings of the preceding literature survey was that the greenhouse-crop system is time-invariant and that the current control methods, such as DeePC, do not take this into account. To overcome this, the introduction of this thesis proposed to update the model used for predicting the future and finding the optimal control inputs. The first step will be set towards this proposed method; it investigates the possibility to have a linear state-space model to estimate the non-linear greenhouse-crop systems dynamics around an operating point, for the purpose of automatic optimal control.

Another shortcoming is the price of resources. In the literature, prices are often assumed constant, while in reality they are time-variant [20]. For example, there exist on-peak and off-peak hours in electricity prices, and gas prices strongly depend on the season. The proposed control strategy in this thesis will be using time-variant electricity prices.

### 4-2-3 Formal problem definition

Next, the grower's problem is converted to a formal problem definition, that can be given as:

$$\begin{aligned}
 \max_{\mathbf{u}} \quad J(\mathbf{u}, \mathbf{x}) &= \sum_{i=1}^{N+1} \frac{\Delta H_F[i]}{DMC} c_F[i] - u_{led}[i] c_{led}[i] \\
 &\quad - u_{CO_2}[i] c_{CO_2}[i] - u_{hea}[i] c_{hea}[i] \\
 \text{s.t.} \quad \mathbf{x}[i+1] &= \mathbf{x}[i] + \mathbf{F}(\mathbf{x}[i], \mathbf{u}[i], \mathbf{d}[i]), \quad \forall i \in \{1, \dots, N+1\}, \\
 \mathbf{y}[i] &= \mathbf{G}(\mathbf{x}[i], \mathbf{u}[i]), \quad \forall i \in \{1, \dots, N+1\}, \\
 \mathbf{x}[1] &= \hat{\mathbf{x}}[1], \\
 \mathbf{b}_{\mathbf{u}}^{low} &\leq \mathbf{u}[i] \leq \mathbf{b}_{\mathbf{u}}^{upp}, \quad \forall i \in \{1, \dots, N+1\}, \\
 \mathbf{b}_{\mathbf{y}}^{low} &\leq \mathbf{y}[i] \leq \mathbf{b}_{\mathbf{y}}^{upp}, \quad \forall i \in \{1, \dots, N+1\}, \\
 u_{hea,pipe1}[j] &\geq 0 \quad u_{hea,pipe2}[j] \geq 0, \quad \forall j \in \{1, \dots, N+1\}, \\
 |T_{air}[i+1] - T_{air}[i]| &\leq 4^\circ C, \quad \forall i \in \{1, \dots, N+1\},
 \end{aligned} \tag{4-9}$$

where  $i$  denotes the timestep into the growing season length  $N$ ,  $c_{\{x,u\}}[i]$  denotes the time-variant cost and price coefficient and  $\mathbf{b}_{\{x,u\}}^{\{upp,low\}}$  denotes an upper or lower bound for the control states and inputs. The dry matter content, denoted by  $DMC$ , is used because the selling price  $c_F$  of the harvested fruits is given per unit of fresh weight and the simulated  $H_F$  is the dry weight. The last three constraints are added to the ground-truth simulation and are explained in the next paragraph. The dynamics of the system are denoted by  $\mathbf{F}$  and the output function is denoted by  $\mathbf{G}$ . The initial condition is given by  $\hat{\mathbf{x}}$ . The states  $\mathbf{x}$ , inputs  $\mathbf{u}$ , outputs  $\mathbf{y}$  and disturbances  $\mathbf{d}$ , of the greenhouse-crop system are

$$\mathbf{x} = \begin{bmatrix} T_{air} \\ AH_{air} \\ CO_{2,air} \\ C_{buf} \\ C_{leaf} \\ \mathbf{C}_{Frt\{i\}} \\ C_{stem} \\ T_{c24} \\ \mathbf{N}_{Frt\{i\}} \\ H_F \\ H_L \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_{CO2} \\ u_{scr1} \\ u_{scr2} \\ T_{pipe,1} \\ T_{pipe,2} \\ u_{led} \\ \phi_{Lee} \\ \phi_{Win} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} T_{air} \\ AH_{air} \\ CO_{2,air} \\ T_{c24} \\ \Delta H_F \\ m_F \\ m_L \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} T_{out} \\ AH_{out} \\ CO_{2,out} \\ Q_{sun} \\ v_{wind} \end{bmatrix}, \quad (4-10)$$

where  $\mathbf{x} \in \mathbb{R}^{n_x}$ ,  $\mathbf{u} \in \mathbb{R}^{n_u}$ , outputs  $\mathbf{y} \in \mathbb{R}^{n_y}$  and disturbances  $\mathbf{d} \in \mathbb{R}^{n_d}$ , with  $n_{x,u,y,d}$  being the respective sizes of the vectors. Note that it is assumed that the (harvested) fruit and leaf weights are continuously measurable. Currently, only estimates are made by taking samples through manual labor, but it is assumed that in the near future, automated measurements will be continuously performed on a larger scale, e.g. with image processing. Note that here the pipe temperature is included, which is not analogous with the heating input  $u_{hea}$  in the ground-truth simulation. However, with a simple conversion equation, both can be connected, see (3-27). The symbols are described below in Table 3-1, Table 3-2, Table 3-3 and Table 3-4, respectively. The upper and lower bounds of the inputs are based on the maxima and minima of the baseline inputs and are defined as

$$\mathbf{b}_u^{low} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{b}_u^{upp} = \begin{bmatrix} 0.0105 \\ 1 \\ 1 \\ 60 \\ 60 \\ 50 \\ 100 \\ 100 \end{bmatrix}, \quad (4-11)$$

### Additional constraints to the ground-truth simulation

In real-life greenhouse operations, there are no physical limits to the outputs  $\mathbf{y}$ . However, because some states are unwanted for plants, the simulation has additional constraints. For example, the air temperature must not reach higher than  $40^\circ C$ , as this will damage the plants. While this effect is captured in the simulation dynamics to some extent, it is a precaution to prevent unrealistic results. The values are based on the lowest and highest values of the states in the ground truth simulation with the inputs of the AGC reference data.

The constraints of the states in the form of lower and upper bounds:

$$\mathbf{b}_y^{low} = \begin{bmatrix} 6.1 \\ 8.2 \\ 0.67 \\ 6.1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{b}_y^{upp} = \begin{bmatrix} 40.3 \\ 38.9 \\ 5.03 \\ 40.3 \\ \infty \\ \infty \\ \infty \end{bmatrix}. \quad (4-12)$$

Second, a constraint is added to the ground-truth simulation to prevent unrealistic temperature drops. The original reference data of the AGC showed a maximum temperature drop of  $4^\circ C$  and a maximum temperature rise of  $4^\circ C$ , both during a time interval of five minutes. Therefore, the temperature drop and rise are clipped to these values, respectively.

Third,  $u_{hea}$  must be either equal to zero or be positive, i.e. the heating pipes cannot cool down the greenhouse. This will force the input of the heating pipe temperature to never be lower than the air temperature.

### Cost and coefficients

As mentioned, profit is made by selling the harvested fruits ( $H_F$ ), while using resources for heating,  $CO_2$ -injection and illumination. Table 4-1 shows the coefficients and their values. The price of electricity depends on whether the usage is during on-peak (7:00-23:00) or off-peak hours (23:00-7:00) [6]. Note that the coefficients for  $CO_2$  and heating are kept constant throughout the growing season.

**Table 4-1:** Cost and price coefficients [6]

Coefficient	Value	Unit	Description
$c_F$	2.0	$kg^{-1}$	Price per harvested fruit weight
$c_{led}$ on-peak	0.08	$kWh^{-1}$	Cost of electricity during on-peak hours (7:00-23:00)
$c_{led}$ off-peak	0.04	$kWh^{-1}$	Cost of electricity during off-peak hours (23:00-7:00)
$c_{CO_2}$	0.1	$kg^{-1}$	Cost of buying or generating $CO_2$
$c_{hea}$	0.03	$kWh^{-1}$	Cost of heating

## 4-3 Proposed control architecture

To control the greenhouse-crop system a control architecture including an MPC-based controller is proposed below. This controller uses the linear state-space model and a profit-based objective to predict future steps. Moreover, constraints are added to prevent instability and trivial solutions.

### 4-3-1 Overview

The proposed control architecture is shown in Figure 4-2. First, the MPC-based controller optimizes on the defined objective, over prediction horizon  $P_{hor}$  using an identified linear model. Then,  $\mathbf{u}_L[i \rightarrow i + C_{hor}]$ , is sent to the ground-truth simulator that represents reality. This simulation uses non-linear model dynamics  $F(\mathbf{x}[i], \mathbf{u}[i], \mathbf{d}[i])$ , multiplied by sampling time  $T_s$ . The output  $\mathbf{y}[i]$  is related to the state as  $G(\mathbf{x}[i])$ . The output  $\mathbf{y}[i]$  is used as feedback for the MPC-based controller. The objective is to make a profit, of which a more detailed description is given in the next subsection. The algorithm accompanying the control process is given in algorithm 2.

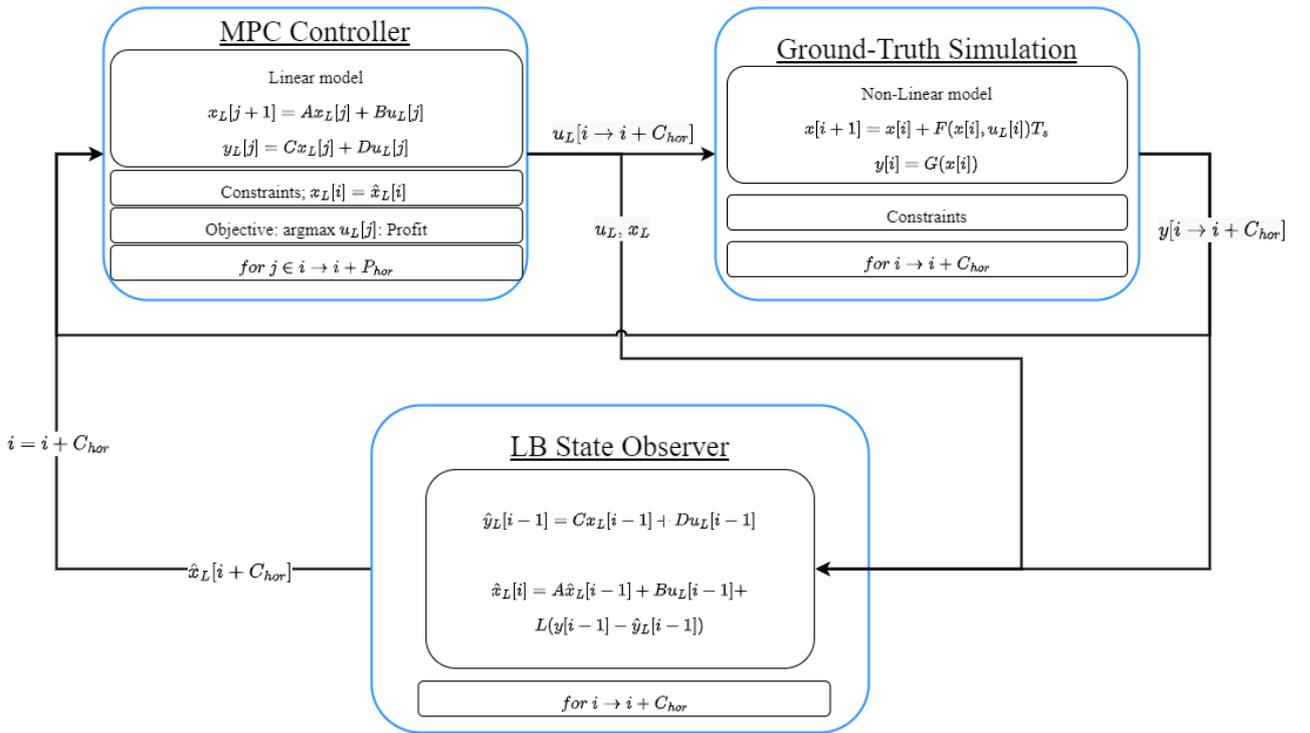


Figure 4-2: Proposed control architecture and simulation setup.

Where  $C_{hor}$  denotes the control horizon and it's assumed that the disturbances are predictable without any noise or uncertainty for prediction horizon  $P_{hor}$ . It is important to note that the state  $\mathbf{x}_L[i]$  is not equal to the state  $\mathbf{x}[i]$  of the simulation;  $\mathbf{x}_L[i]$  is an arbitrary internal state variable used for the linear state-space model and system identification.  $\mathbf{y}[i]$  represents the outputs of the simulation. The initial state  $\mathbf{x}_L[i]$  of the MPC controller is equal to the state estimate  $\hat{\mathbf{x}}_L[i]$  of the Luenberger State Observer, which will be covered in Subsection 4-3-3.

### 4-3-2 MPC-based controller

The control problem that has to be solved representing the MPC-based controller is shown in (4-13).

$$\begin{aligned}
\max_{\mathbf{u}} \quad J_L(\mathbf{u}, \mathbf{x}) &= \sum_{j=i}^{i+P_{hor}+1} \frac{\Delta H_F[j]}{DMC} c_F[j] - c_{led}[j] u_{led}[j] \\
&\quad - c_{CO_2} u_{CO_2}[j] - c_{hea} u_{hea}[j] \\
\text{s.t.} \quad \mathbf{x}_L[j+1] &= \mathbf{A}\mathbf{x}_L[j] + \mathbf{B}\mathbf{u}_L[j], \quad \forall j \in \{i, \dots, i+P_{hor}\}, \\
\mathbf{y}_L[j] &= \mathbf{C}\mathbf{x}_L[j] + \mathbf{D}\mathbf{u}_L[j], \quad \forall j \in \{i, \dots, i+P_{hor}\}, \\
\mathbf{x}_L[i] &= \hat{\mathbf{x}}_L[i], \\
\mathbf{b}_{\mathbf{u}_L}^{low} &\leq \mathbf{u}_L[j] \leq \mathbf{b}_{\mathbf{u}_L}^{upp}, \quad \forall j \in \{i, \dots, i+P_{hor}\}, \\
\mathbf{b}_{\mathbf{y}_L}^{low} &\leq \mathbf{y}_L[j] \leq \mathbf{b}_{\mathbf{y}_L}^{upp}, \quad \forall j \in \{i, \dots, i+P_{hor}\}, \\
u_{hea,pipe1}[j] &\geq 0 \quad u_{hea,pipe2}[j] \geq 0, \quad \forall j \in \{i, \dots, i+P_{hor}\}, \\
m_F[j] &\geq m_F[i], \quad \forall j \in \{i, \dots, i+P_{hor}\},
\end{aligned} \tag{4-13}$$

where the states are denoted by  $\mathbf{x}_L \in \mathbb{R}^{n_{x_L}}$ , the outputs are denoted by  $\mathbf{y}_L \in \mathbb{R}^{n_{y_L}}$  and the inputs are denoted by  $\mathbf{u}_L \in \mathbb{R}^{n_{u_L}}$ , with  $n_{x_L, u_L, y_L}$  being the respective sizes of the vectors. The reader is referred to Subsection 4-1-1 for the composition of the state-, output- and input vectors. The linear state-space matrices are denoted by  $\mathbf{A} \in \mathbb{R}^{n_{x_L} \times n_{x_L}}$ ,  $\mathbf{B} \in \mathbb{R}^{n_{x_L} \times n_{u_L}}$ ,  $\mathbf{C} \in \mathbb{R}^{n_{y_L} \times n_{x_L}}$  and  $\mathbf{D} \in \mathbb{R}^{n_{y_L} \times n_{u_L}}$ . All outputs are scaled and have their respective biases removed according to the processing steps from Subsection 4-1-2. The prediction horizon  $P_{hor}$  depends on the accuracy of the linear state-space model and will be substantiated in Section 5-1. The control horizon is chosen to be 30 minutes equal to 6 samples. There are also two new constraints introduced on the total dry weight of the fruits. The rationale behind these constraints is presented next.

### Rational behind the constraint on the total dry weight

When basing the profit on the harvested dry weight and using a relatively short prediction horizon (e.g. a few days), there is only an indirect incentive to grow new fruits and to ripen the more mature fruits. To prevent the controller from just draining the fruit stages and to also let it grow new fruits, the final constraint in (4-13) is added. These make sure that the total dry weight  $m_F[i]$  is at least maintained for the complete prediction horizon. Concretely, this gives the following mathematical constraint:  $m_F[j] \geq m_F[i]$ .

### 4-3-3 Luenberger state observer

For the MPC-based controller to make accurate predictions over the prediction horizon, it has to start with a good initial guess. Since the order of the model (10) is larger than the number of outputs (7), one can not simply use  $\mathbf{x}_L[i] = \mathbf{C}^{-1}\mathbf{y}[i]$ . Therefore, a Luenberger State Observer is implemented to produce the internal state estimates for the controller, as can be seen in Figure 4-2.

The state observer determines the state estimate  $\hat{\mathbf{x}}_L[i]$  by also taking into account the error between the actual measured output  $\mathbf{y}[i]$  and the output  $\hat{\mathbf{y}}_L[i]$  estimated by the predictor. This error converges to zero when the eigenvalues of  $\mathbf{A} - \mathbf{L}\mathbf{C}$  are inside the unit circle and

that is achieved by the configuration of matrix  $\mathbf{L}$  [51]. This gives the following updated state system including state estimation, that will be used alongside the proposed controller:

$$\begin{aligned}\hat{\mathbf{x}}_{\mathbf{L}}[i+1] &= \mathbf{A}\hat{\mathbf{x}}_{\mathbf{L}}[i] + \mathbf{B}\mathbf{u}_{\mathbf{L}}[i] + \mathbf{L}(\mathbf{y}[i]) - \hat{\mathbf{y}}_{\mathbf{L}}[i] \\ \hat{\mathbf{y}}_{\mathbf{L}}[i] &= \mathbf{C}\hat{\mathbf{x}}_{\mathbf{L}}[i] + \mathbf{D}\mathbf{u}_{\mathbf{L}}[i]\end{aligned}\quad (4-14)$$

The predicted state is used for the initial state each time a new prediction is started by the prediction model in the controller.

#### 4-3-4 Control architecture algorithm

The control algorithm for the Control Architecture is given in algorithm 2.

---

#### Algorithm 2: Control Architecture Algorithm

---

**Input:** System function with  $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ , prediction horizon  $P_{hor}$ , control horizon  $C_{hor}$ , measured disturbances  $\mathbf{d}[i \rightarrow i + P_{hor}]$  and constraint sets for  $\mathbf{y}_{\mathbf{L}}$  and  $\mathbf{u}_{\mathbf{L}}$ .

**Objective:**  $J_P(\mathbf{x}, \mathbf{u})$

- 1) Obtain initial state estimate  $\mathbf{x}_{\mathbf{L}}[i]$ :  
 $\mathbf{x}_{\mathbf{L}}[i] = \hat{\mathbf{x}}_{\mathbf{L}}[i] = \mathbf{A}\hat{\mathbf{x}}_{\mathbf{L}}[i-1] + \mathbf{B}\mathbf{u}_{\mathbf{L}}[i-1] + \mathbf{L}(\mathbf{y}[i-1]) - \hat{\mathbf{y}}_{\mathbf{L}}[i-1]$ .
  - 2) Solve (4-13) for the optimal input sequence  $\mathbf{u}_{\mathbf{L}}^*$  for  $[i \rightarrow i + P_{hor}]$ , for the Objective:  
 $J_P(\mathbf{x}, \mathbf{u})$
  - 3) Apply inputs top ground-truth simulation for control horizon:  
 $\mathbf{u}[i] = \mathbf{u}_{\mathbf{L}}^*[i \rightarrow i + C_{hor}]$ .
  - 8) Set  $i$  to  $i + C_{hor}$ .
  - 9) Return to 1).
- 

## 4-4 Performance test setup

In the next chapter, the performance of the will be tested using the following setup. First, the linear state-space system for the MPC-based controller will be identified for the period between day 76 till day 90 with the methods described in the previous section and its fit on validation experiments will be presented. Second, the MPC-based controller will apply the inputs of the baseline grower for the first 2 days. This is done to let the state estimation estimate the states, to eventually have a good initial state guess before the deployment of the controller on day 78. Third, the controller will be used to control the ground-truth simulation for 14 days from day 78 onward. Fourth, the inputs and outputs will be collected and will be evaluated on the realized climate trajectories, applied inputs and gained profit. To solve the optimization problem the Gurobi solver is used [52].

## 4-5 Chapter summary and conclusion

In this chapter, a linear state-space system that will act as the prediction model is identified. The output data used for the system identification of the model is scaled and bias is removed

to improve the performance of the system identification. A training dataset of multiple experiments is artificially created with the ground-truth simulator, alongside a validation dataset that will be used to measure the performance of the identified linear state-space system. The performance measure of the system identification is the fit or 100% minus the Normalised Root Mean Squared Error (NMRSE) between the realized outputs and predicted outputs, see (4-6).

The objective of a grower is to optimize its profit by selling harvested tomatoes while minimizing its costs of  $CO_2$ -dosing, heating and electricity usage. Attempts from the literature on solving this problem are mostly using Model Predictive Control (MPC) based optimal controllers, that optimize an objective while satisfying specified constraints on the system's inputs, outputs and states. One relevant control technique that was studied for solving the grower's problem is that of Kerkhof (2020) and van Duijn (2021) [17] [18]. The hierarchical DeePC method that was proposed uses two layers, separating the slow crop dynamics and fast greenhouse climate dynamics. The DeePC controller in the upper layer solely focuses on creating reference climate trajectories, that is then tracked by a lower-layer controller. The proposed control method in this thesis work does not split the control problem into two, it rather tries to optimize the slow and fast dynamics simultaneously.

Constraints are added to prevent the controller to give trivial solutions. Concretely this means that instead of only harvesting the fruits and draining the total fruit weight, the controller must also maintain this weight by growing new ones.

---

# Chapter 5

---

## Results

The results of the simulation will be presented and discussed in this chapter. First, the performance of the system identification of the linear state-space model will be presented, followed by a conclusion on the prediction horizon along which the predictor can be used. Next, the predictor is implemented in the proposed MPC-based controller and a simulation run will be performed. Of this simulation run, the output climate and mass trajectories are shown. The performance of the proposed controller in terms of profit will be shown. Each result is also compared to that of the baseline grower. The performance metric is the obtained profit, composed of the yield from selling the fruits and the costs of using resources, for a predetermined period of 14 days.

### 5-1 System identification

The prediction performance of the identified system will be discussed on four evaluation points. First, and foremost, the prediction performance fit will be compared, as defined in Subsection 4-1-4. Second, the pattern in the output trajectory will be evaluated, as that is influenced by the relatively slow day and night rhythm of the crop and greenhouse states. This means that there must be a day and night pattern present in the output trajectories; a more flat pattern indicates that the system is not sufficiently capturing the typical crop and greenhouse dynamics. The slopes of the relatively fast climate conditions, such as the temperature and  $CO_2$ -concentration of the greenhouse air, will be evaluated as a third. Finally, from all evaluations, a conclusion will be drawn on the appropriate prediction horizon for the MPC-based controller.

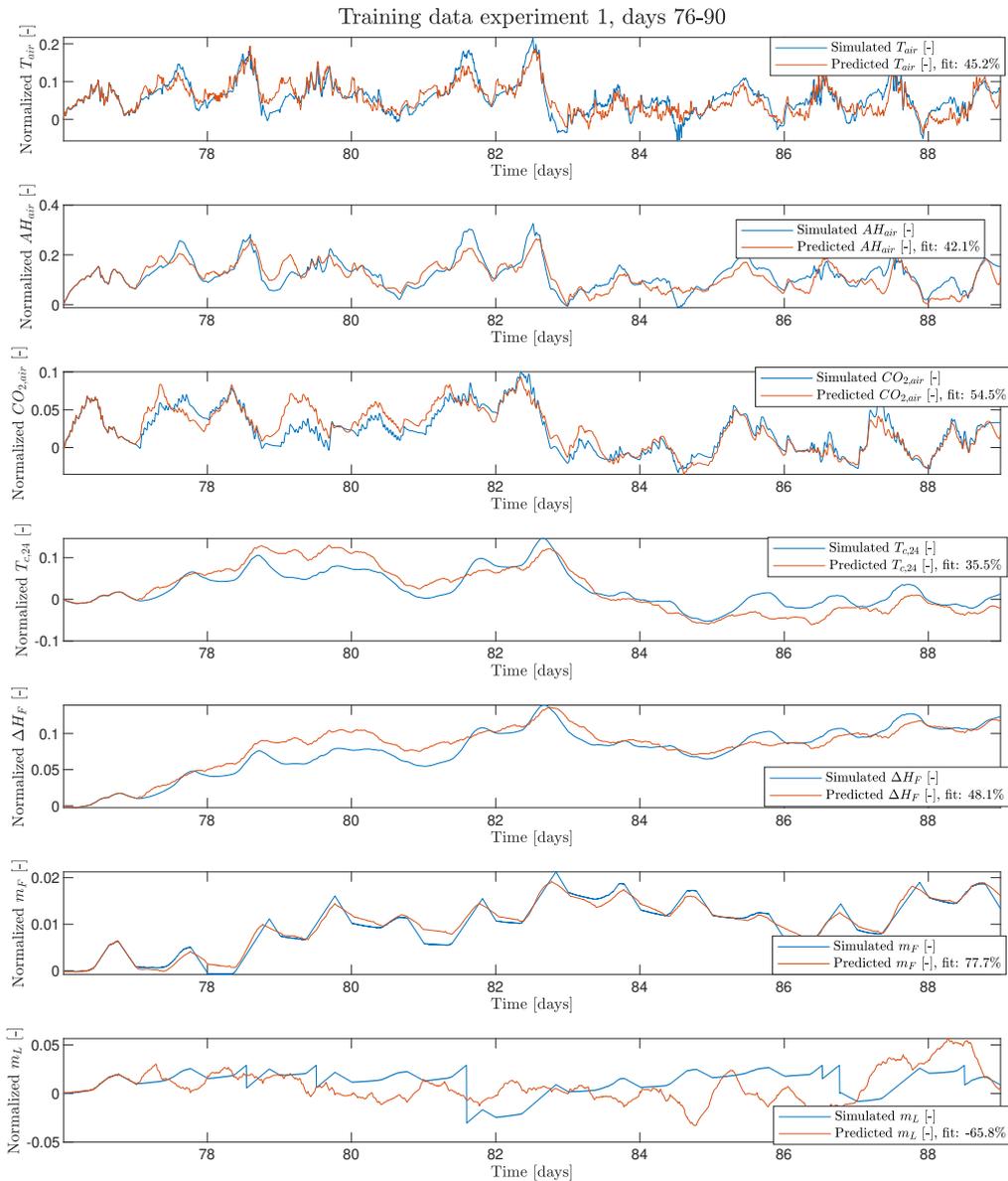
Comparisons of performance will be made between four types of experiments:

1. Applying inputs of the training dataset on the ground-truth simulator and the identified system.
2. Applying inputs of a validation dataset on the ground-truth simulator and the identified system.
3. Applying inputs of a validation dataset with inputs of a higher frequency on the ground-truth simulator and the identified system.

4. Applying inputs of a validation dataset with inputs of a higher amplitude on the ground-truth simulator and the identified system.

### 5-1-1 Fit of the identified system on the training dataset

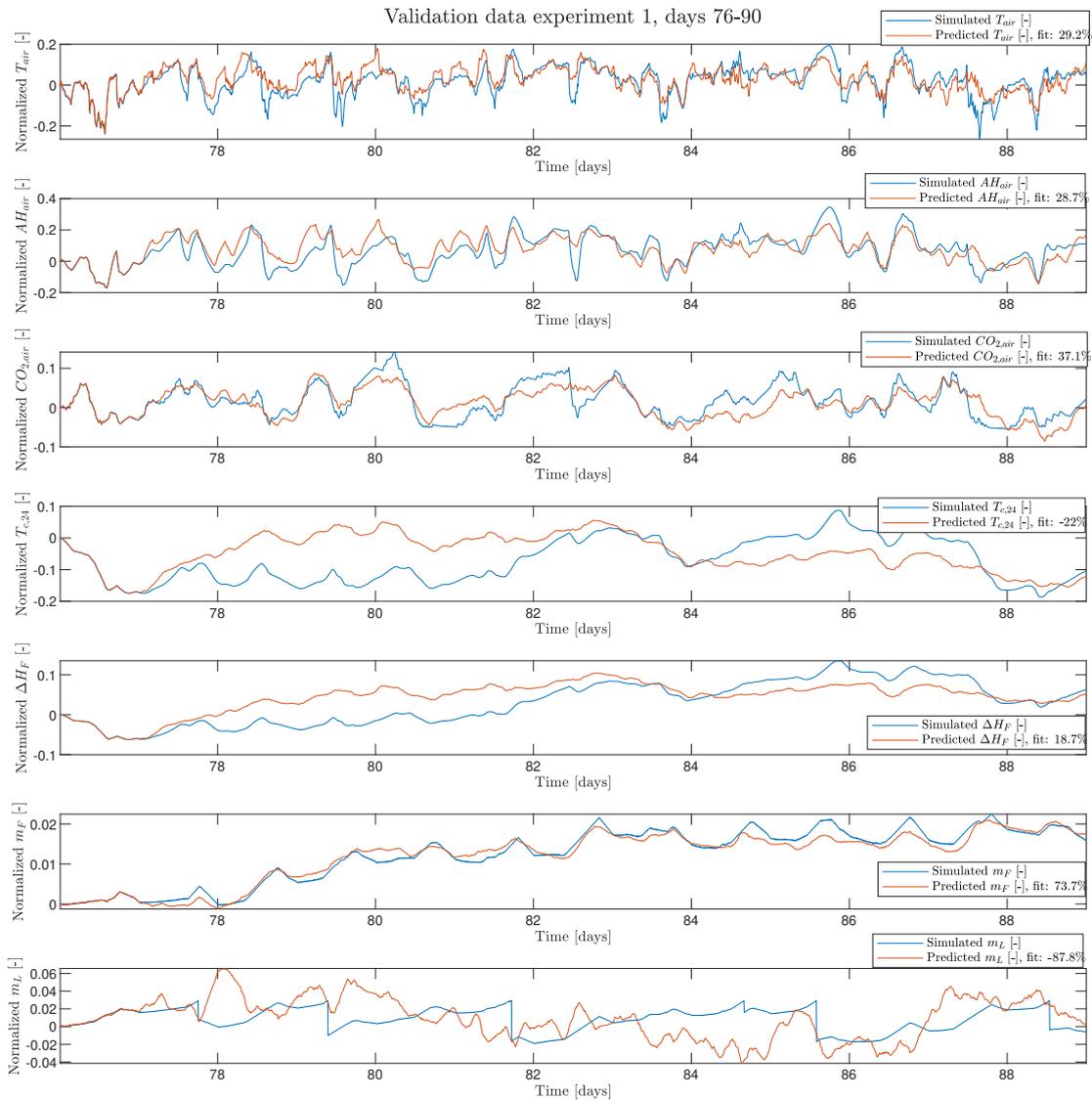
The fit of the identified linear state-space system on one of the training experiments, presented in Subsection 4-1-3, can be found in Figure 5-1. The fits on the other training datasets can be found in Section B-1 of the Appendix.



**Figure 5-1:** Fit of identified linear state-space system on the training dataset from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.

### 5-1-2 Fit of the identified system on the validation dataset

Next, Figure 5-2 shows the fit of the identified linear state-space system, but then with the inputs from the validation dataset for the validation experiment. The validation dataset is different from the training dataset, but it is of similar pattern, amplitude and frequency.

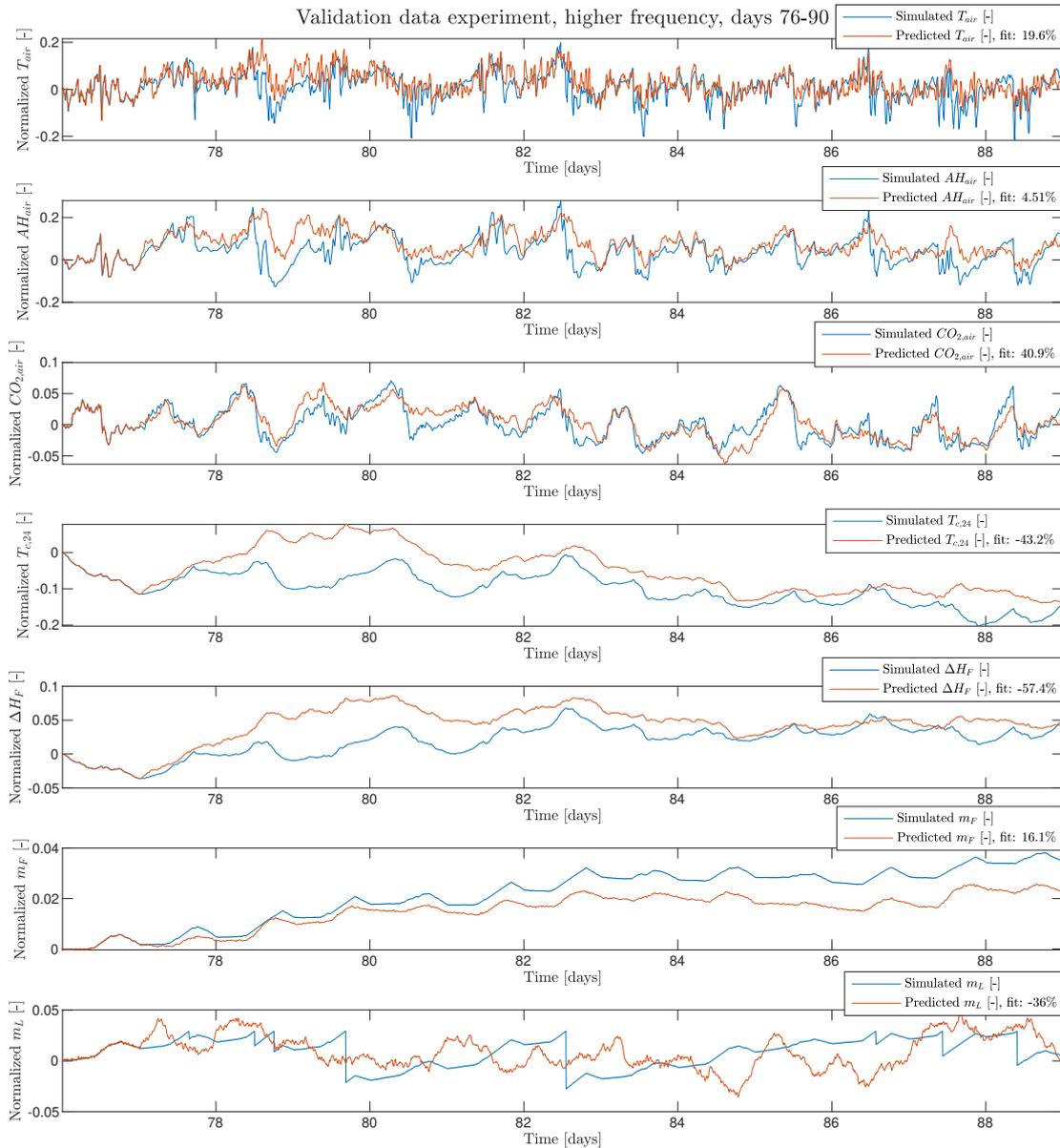


**Figure 5-2:** Fit of identified linear state-space system on the validation dataset from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.

The system's predictions of the temperature,  $CO_2$ -concentration, absolute humidity and the total fruit dry weight  $m_F$  with validation inputs, that can be seen in Figure 5-2, are considered sufficient for the complete 14 day period. It can also be seen that the system is less good at predicting the change in harvested fruit dry weight  $\Delta H_F$  for horizons longer than a day. Other aspects will be discussed in Subsection 5-1-4.

### 5-1-3 Fit of the identified system on validation dataset with enhanced inputs

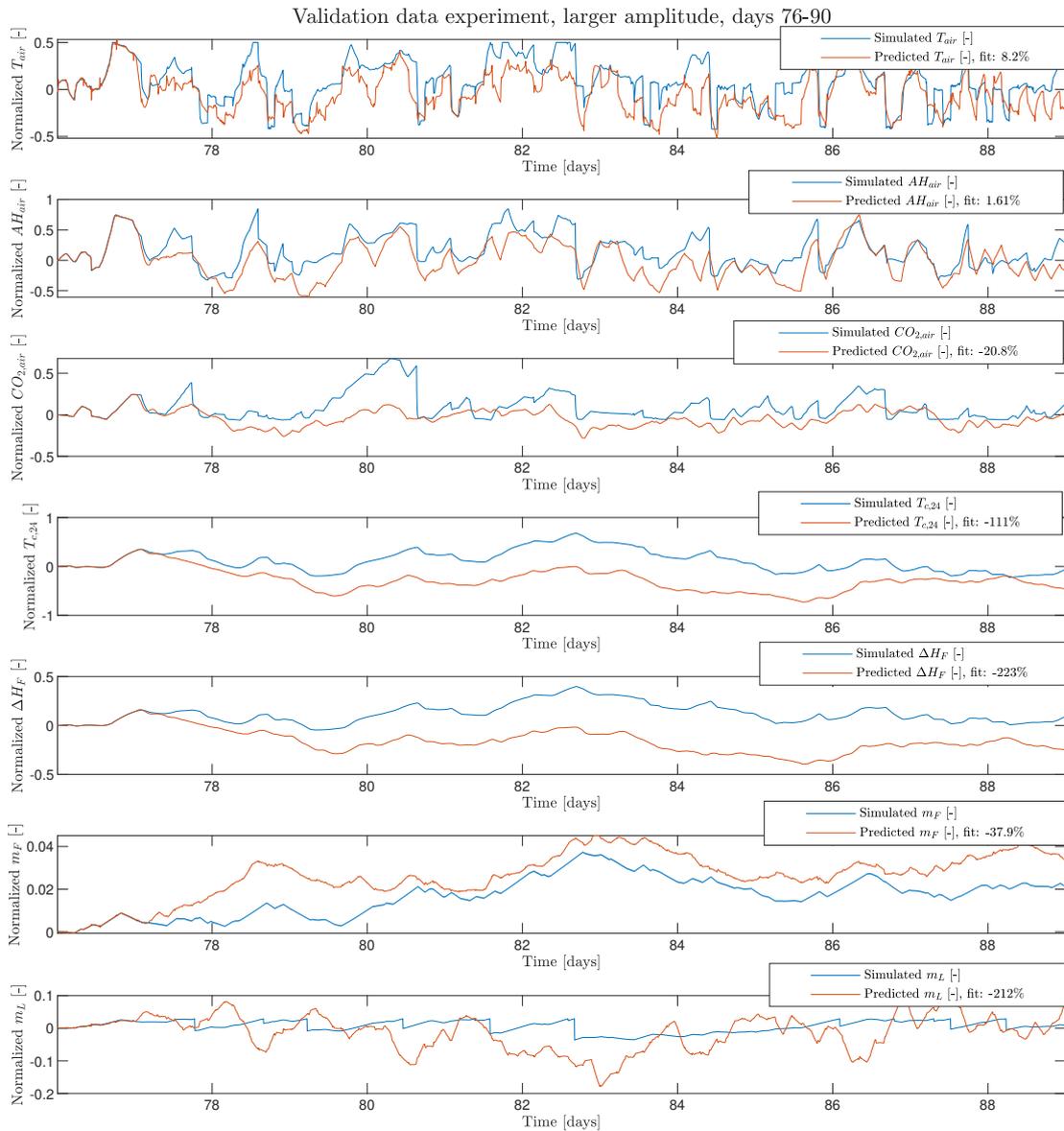
The next figure shows the fit of the system on a validation dataset composed of the baseline inputs enhanced with input noise with a higher frequency.



**Figure 5-3:** Fit of identified linear state-space system on the validation dataset from day 76 onward, with higher frequency. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.

This final Figure 5-4, representing the results of the final validation experiment, shows the fit

of the system on a validation dataset composed of the baseline inputs enhanced with input noise with a larger amplitude.



**Figure 5-4:** Fit of identified linear state-space system on the validation dataset from day 76 onward, with larger amplitude. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.

#### 5-1-4 Discussion of the system identification results

It can be noted from the two alternative validation experiments with higher frequency and amplitude, that under these more extreme input conditions the system performance worsens in terms of predicting the outputs. On the other hand, the first validation experiment with

similar inputs to that of the training data showed reasonable prediction performance by the identified system. In the next paragraphs, all separate outputs generated in the validation experiments will be discussed on evaluation points that stand out.

### **Temperature**

First of all, rapid changes in greenhouse air temperature are more difficult to predict by the linear state-space system. The system reacts in such a way that it predicts that the temperature changes more frequently than it actually does; it gives priority to the higher frequencies. This also comes back in the natural day-night pattern of the temperature that is much less present in the predicted outputs from the corresponding validation experiment.

Applying inputs with higher amplitudes generates a fit that is worse than that on the validation inputs with lower amplitudes. There is a clear day-night pattern visible, but the temperature rises are not sufficiently predicted, as they are mostly predicted less steep than simulated.

### **$CO_2$ -concentration**

The output predicted trajectory of the  $CO_2$ -concentration is considered to be sufficient for inputs with similar frequency to that of the training set and for inputs with higher frequency. The plot shows that, at least for the first days, the slopes of the predicted and simulated output are similar. The same cannot be said about the output from simulation with inputs of higher amplitude. The simulated values of the  $CO_2$  levels are higher than those predicted by the linear system and the pattern is not the same.

### **Harvested fruit dry weight**

The fit of the harvested fruit dry weight in the first validation experiment without enhanced inputs, see Figure 5-2, roughly follows the slow pattern. It is this slower pattern that is important, as it was concluded in the detailed crop model that crop growth is influenced by slow dynamics based on daily averages. When the frequency or the amplitude of the inputs is raised, the performance of predicting the harvested fruit dry weight worsens.

It can also be seen that the pattern in the change in harvested fruit dry weight follows the pattern of the average 24 hours temperature in all three experiments. This is as expected, as the non-linear crop model dynamics describe a correlation between the dry weight growth and the average temperature-related states.

### **Total fruit dry weight**

The identified linear system performs quite well in predicting the total fruit dry weight in the validation experiment with and without raised frequencies; the output pattern is followed, the fit is relatively high and the slopes are similar. However, when the amplitude of the inputs is increased, the prediction performance worsens a lot. Both the total fruit dry weight and the harvested fruit dry weight exclude influences of higher frequencies; they are less influenced by the faster dynamics.

### **Total leaf weight**

The total leaf weight is predicted the worst in both the training and the validation simulations. This is not an issue, as the controller does not use this output to compute the optimal control inputs in any form.

#### **5-1-5 Fit of the identified system on different operating points**

It was expected that the identified system would serve as a good representation of the non-linear ground truth only around a specific operating point, corresponding to the surrounding period the former system was identified on. To investigate the influence of the operating point, two new experiments were conducted: 1. a simulation run during the beginning of the growing season from day 4 onward and 2. a simulation run near the end of the growing season from day 145 onward. Figure B-8 and Figure B-9 show the resulting prediction performance. It can be seen that indeed simulations on different operating points with the same system identified on another period of the growing season, result in worse prediction fit. However, the prediction performance further on in the growing season is better compared to that of earlier in the growing season using the same linear system. This can be explained by how the operating point for the system identification was chosen. That is, it was chosen to be in the growing period where the crop growth reached and maintained a steady state. Therefore, there is less difference between that chosen period and the future, where the crop states are roughly equal. However, the prediction performance is still not similar and thus the system probably still needs to be re-identified, see Section 6-2. Looking at a period earlier on in the growing period the prediction performance is poor; no output is predicted with sufficient accuracy. A possible explanation can be, again, found in the operating point of the crop states. That is, earlier on in the season the crop states have definitely not reached a steady state yet and their non-linear and quickly evolving dynamics are, therefore, harder to predict with a system that was identified in the steady state period.

#### **5-1-6 Final remarks on the system identification results**

From the experiments, it became clear that inputs that have raised frequencies and amplitude result in worsened performance. A method to prevent this is to add constraints to the MPC-based controller that describe limits on the slopes, duration and frequencies of the system's inputs. These types of constraints were not included in this thesis research because of time constraints and it is, therefore, recommended to study these in future work.

From the experiments earlier and later on in the growing season using the identified system from the middle of the growing season, it can be seen that the prediction performance depends on the starting operating point. This confirms the assumption that the system needs to be updated throughout the growing period. Section 6-2 in the Recommendations will elaborate more on this and will propose a solution to this problem.

The decision is made to use a prediction horizon for the profit controller of 1 day, or 288 samples. The control horizon is chosen to be 30 minutes, or 6 samples. For this, the computation time is also taken into consideration, as the control problem with a horizon longer than a day takes a considerable amount of time to solve for the available processing unit.

## 5-2 Results of Greenhouse crop control

With the identified system and the chosen control and prediction horizons, the proposed controller can be built and deployed. A simulation will be run to evaluate the performance of the controller. The controller will be deployed from day 78 onward to day 92. The days before the deployment moment, wherein the baseline inputs are applied, are solely used for state observation to estimate the initial state. The results of this simulation run will be presented and discussed on at least the following aspects:

1. The simulated output trajectories;
2. The inputs applied by the controller and the weather disturbances;
3. The resulting costs and profit.

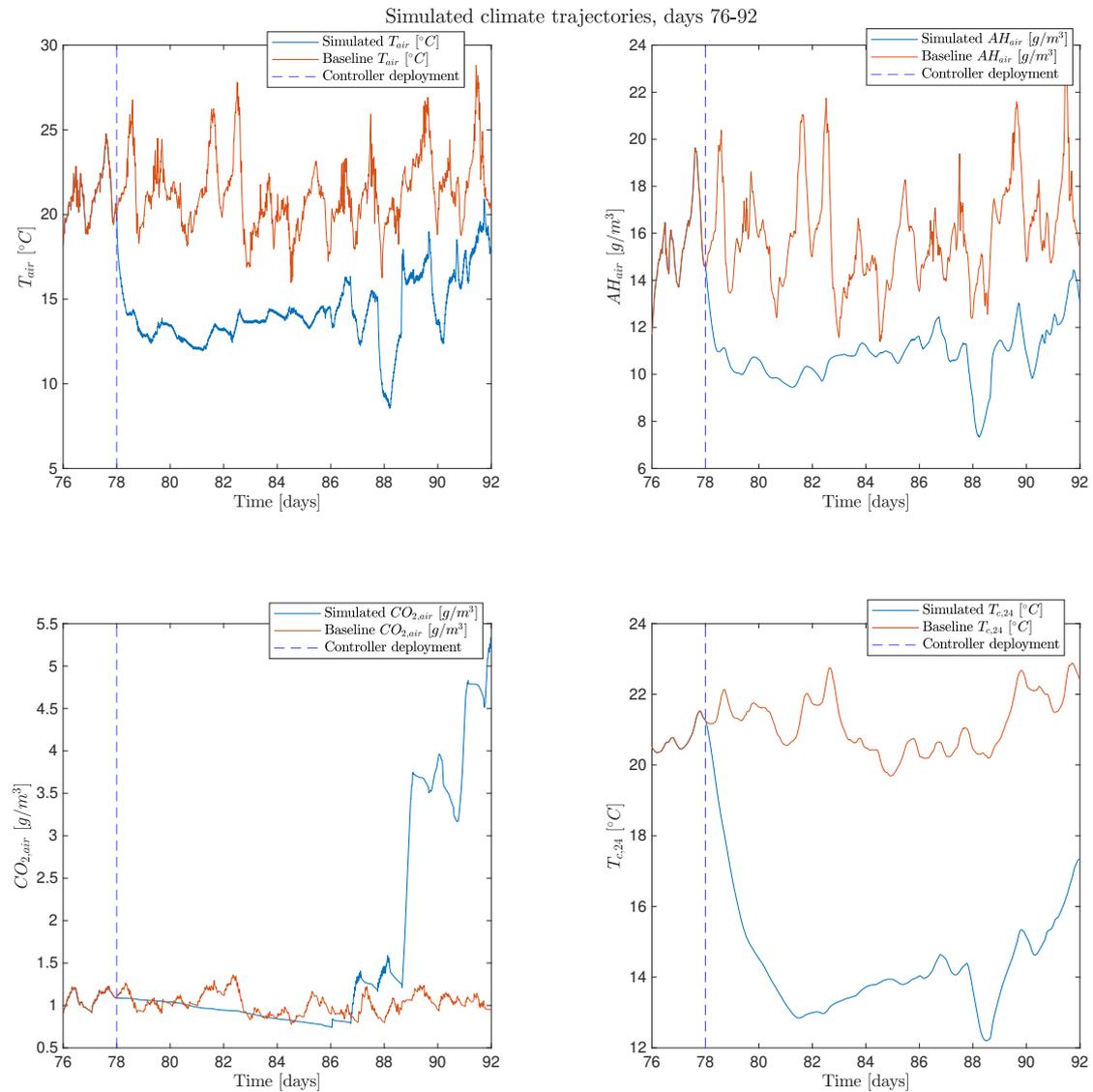
The results will be compared to the *baseline*, which is defined as the results of the grower's inputs applied to the ground-truth simulator.

### 5-2-1 Simulated output trajectories

The simulated output trajectories will be presented next, starting with the climate outputs, followed by the outputs describing the crop.

#### Simulated climate trajectories

The trajectories of the temperature, absolute humidity,  $CO_2$ -concentration and 24h-average temperature are shown in Figure 5-5. The figures have two plotted trajectories; the simulated output trajectory and the original baseline outputs with the inputs of the grower. This last trajectory is solely presented in the plots to show what a representative output would look like in a real greenhouse operation.



**Figure 5-5:** Simulated climate trajectories, generated by the profit controller.

From the simulated climate trajectories it can be seen that the air temperature is very low in comparison to the baseline temperature. Subsection 5-2-2 will present and discuss the inputs applied to the system and will give a possible explanation for the low air temperatures. One conclusion that can be drawn from comparing the greenhouse air temperature to that of the outside air, see Figure 5-8, is that they are not equal. The inputs on the opening of the windows actually show that the windows are never opened and that, thus, the heat is stored in the greenhouse air.

The absolute humidity follows the same pattern as the temperature, which is as expected as the absolute humidity is mainly determined by the air temperature, see Subsection 3-3-2.

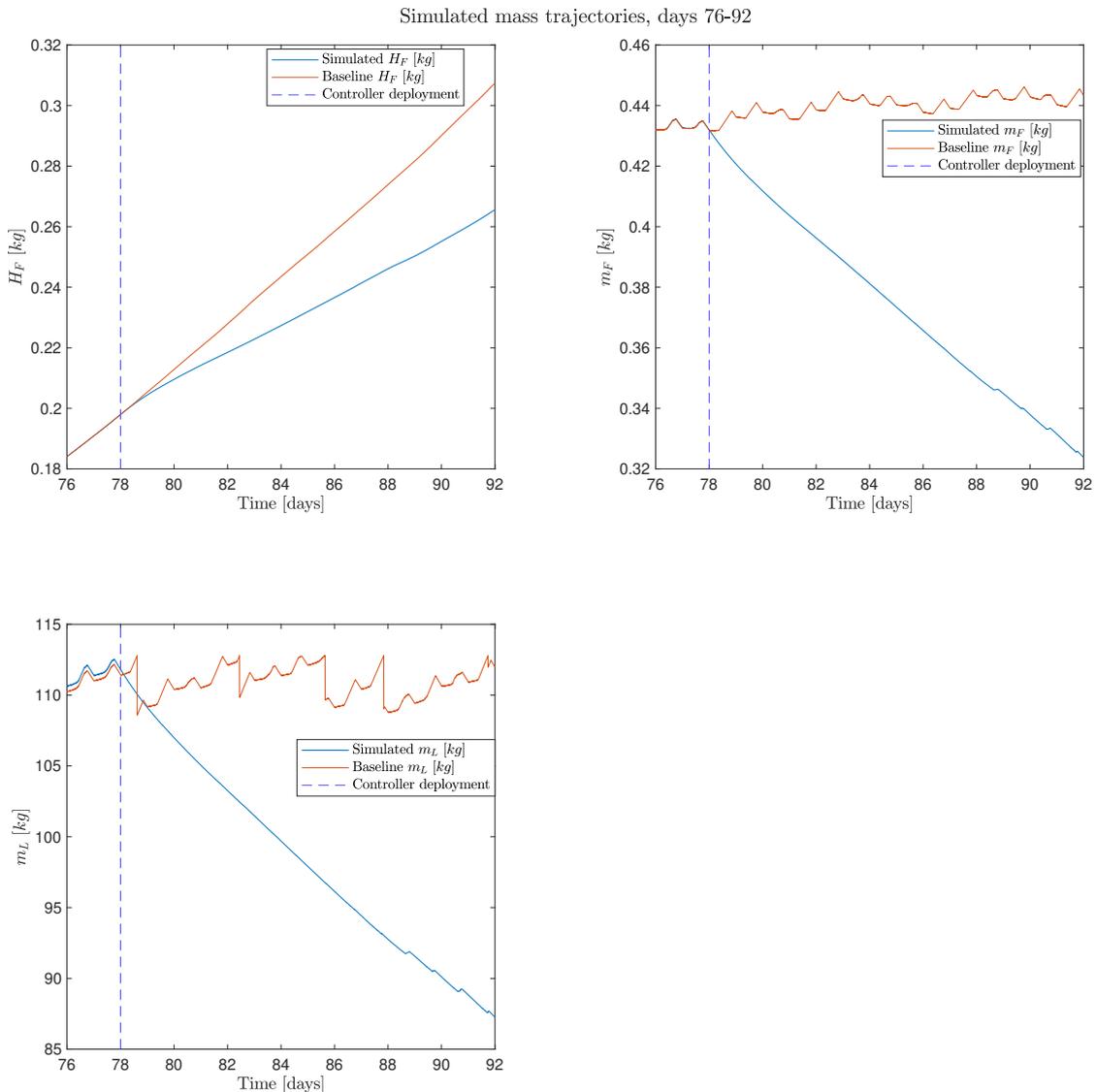
The plot also shows that the  $CO_2$ -concentration decreases with a gentle slope till day 86. This can be explained by the  $CO_2$ -injection being turned off and the windows being shut, see

Subsection 5-2-2. The decrease is then solely caused by the plants slowly extracting  $CO_2$  from the greenhouse air. From day 86 onward relatively much  $CO_2$  is injected the days before the controller deployment, which is the period till day 78 and is shown in the plot as a realistic baseline. This causes the  $CO_2$  to rise above atmospheric levels. It is also from day 86 onward four small rises in total fruit dry weight are visible. A possible explanation for this is could be that there the average 24-hour temperature,  $CO_2$ -concentration and outside solar radiation (see Figure 5-8) have reached higher levels. More importantly, the blackout screens are not opened and solar radiation is let through to the canopy. It is this combination that creates somewhat better growing conditions than at other simulated periods.

### **Simulated mass trajectories**

The trajectories of the harvested fruit dry weight  $H_F$ , the total fruit dry weight  $m_F$  and the total leaf weight  $m_L$  are shown in Figure 5-6.

The figures have two plotted trajectories; the simulated output trajectory and the original baseline outputs with the inputs of the grower. This last trajectory is solely presented in the plots to show what representative output would look like in a real greenhouse operation.

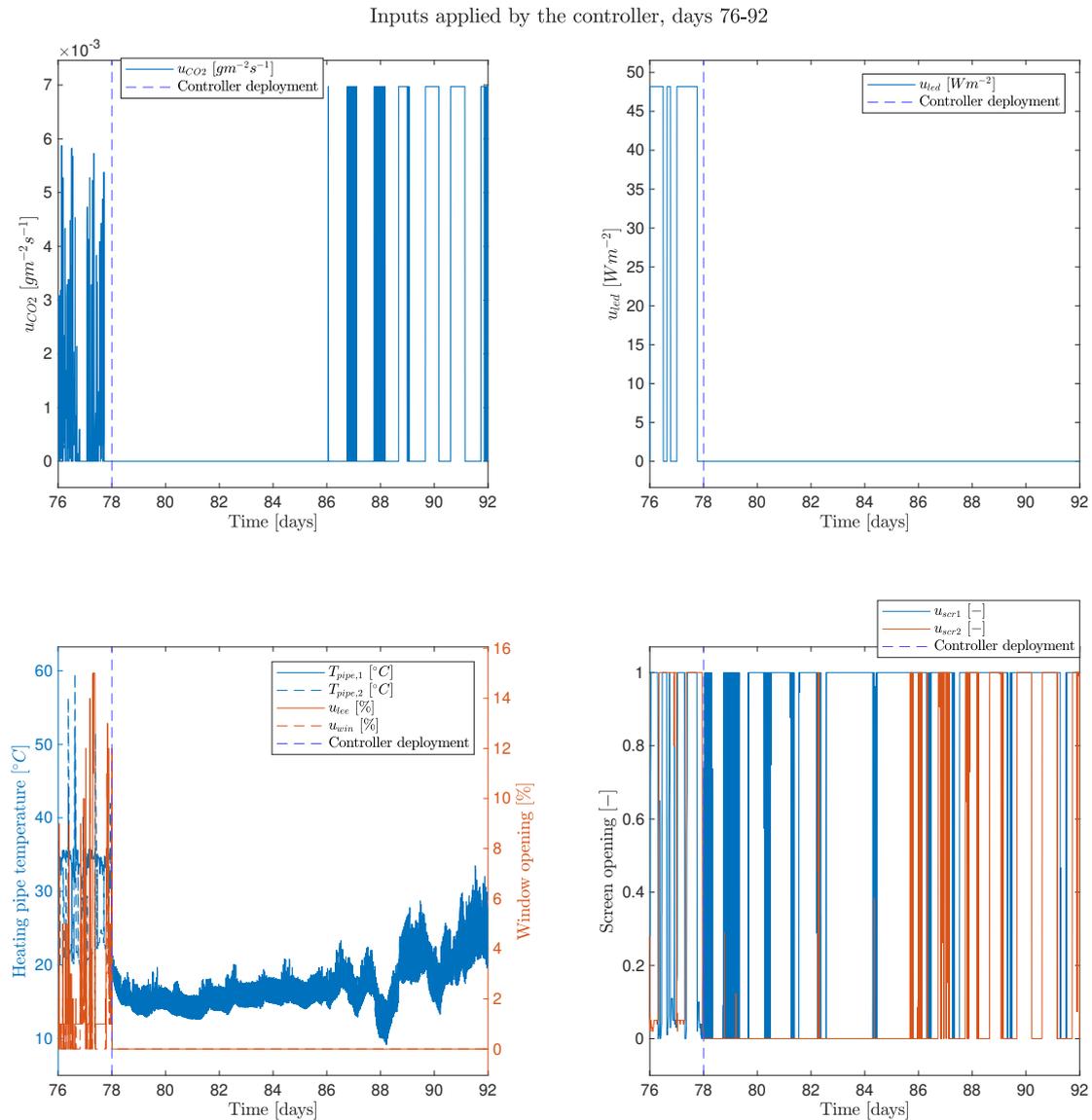


**Figure 5-6:** Simulated and baseline mass trajectories for day 76 till day 92.

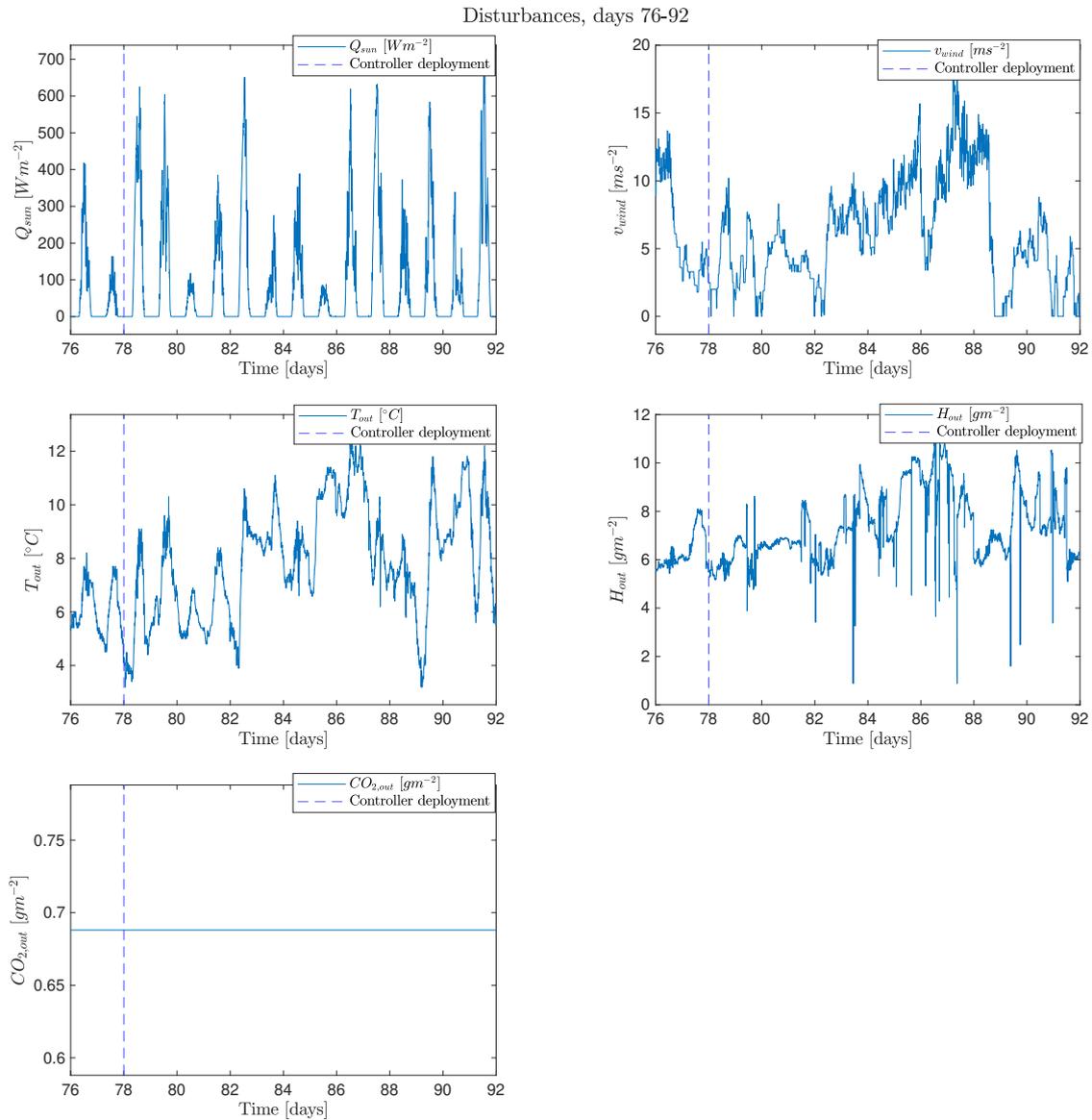
It can be observed from the plots that the performance in terms of fruit growth and harvest is worse. The slope of the harvested fruit dry weight is less steep than that of the baseline. The explanation for this can be found in the top right plot which shows that the total fruit dry weight of the fruits is not maintained and decreases. The interesting aspect of these observations is that despite the constraints defined for the control problem on maintaining the total fruit dry weight of the fruits are met, the actual simulation does not display a constant total fruit dry weight output trajectory. In other words, the predictions made by the linear state-space system model are not accurate enough to let the constraints also be satisfied in the ground-truth simulation. Just as the total fruit dry weight, the total leaf weight decreases, which is in line as they both are driven by the same assimilate generation principles. The explanation of why the total fruit dry weight decreases will be given in the following subsection, as it strongly relates to the inputs and disturbances to the system.

### 5-2-2 Inputs applied by the controller

The inputs applied to the system are shown in Figure 5-7, followed by the disturbances from the outside weather in Figure 5-8.



**Figure 5-7:** Inputs applied to the simulator by the controller. Note that both heating pipe temperature completely overlap, the same applies to the inputs of both windows. Also note that a screen input of zero means that it is completely opened.



**Figure 5-8:** Disturbances from the outside weather.

As was mentioned in the previous subsection, the  $CO_2$ -injector was only turned on after a few days of simulation, which made the  $CO_2$  levels rise quickly. The total usage of  $CO_2$  is roughly equal to that of the baseline, see Table 5-1, but the controller injected  $CO_2$  for longer consecutive periods and more towards the end of the simulation period. The exact reason for this has yet to be discovered. Just having sufficient  $CO_2$ -levels will not make the plants grow; they also need heat and light, as explained in Section 2-3. The controller was modest with turning on the heating; Table 5-1 shows that only a third of the baseline heating energy usage was used. Moreover, Figure 5-7 shows that the controller had no clear preference in using one of the two heating pipes, that both have different heating characteristics.

After closer investigation of predictions of the prediction model in the controller, it became

clear that the air temperature was predicted to be going up and down quickly. This chattering behavior also explains the chattering of the heating pipe temperatures as inputs, caused by the constraint that made sure that the heating pipe temperature was at least as large as the air temperature. It can also be observed that the controller never opens the windows; the assumed explanation for this is that the controller tries to save heating energy by doing so. Also, the air temperature is already relatively low, and opening the windows will not heat up the greenhouse as the outside air temperature was lower than the simulated greenhouse air temperature. Therefore, keeping the windows shut is the natural action to perform for the sake of energy saving.

Finally, the manipulable lighting conditions are also far from conventional. Even on bright days with plenty of received solar radiation, it is typical to turn on the artificial illumination. The climate control strategy produced by the controller did not include any artificial illumination. One of the possible reasons for this is that turning on the lights is relatively expensive, i.e. the controller tries to save costs instead of using the resources to grow plants. The other reason could be that the prediction model does not sufficiently capture the positive effect of turning on the lights on crop growth. The blackout screens ( $u_{scr2}$ ) are opened during the complete growing period till day 86. After, they are deployed only when solar radiation is larger than zero. This is the complete opposite of what is desired and what the baseline strategy showed. A reason for this chosen strategy could be that those screens do actually keep a lot of heat inside the greenhouse and so the controller tries to save energy. The climate screens ( $u_{scr1}$ ) are barely opened in the simulated period. The exact reason why the controller chose the blackout screens over the temperature screens has yet to be discovered.

### 5-2-3 Resulting costs and profit

In this section, the performance of the proposed controller is compared to that of the baseline grower for the specified period of 14 days. This is done by looking at the achieved yield, the resource usage and costs, and the profit.

#### Costs

The inputs are shown in Figure 5-7. Table 5-1 shows the resource usage and costs of heating,  $CO_2$ -injection and the artificial lighting. The cost and price coefficients can be found back in Table 4-1.

**Table 5-1:** Comparison of the resource usage and costs

Controller	Heating MJ/m <sup>2</sup>	$u_{hea}$ €/m <sup>2</sup>	CO <sub>2</sub> kg/m <sup>2</sup>	$u_{CO2}$ €/m <sup>2</sup>	Electricity MJ/m <sup>2</sup>	$u_{led}$ €/m <sup>2</sup>	Total €/m <sup>2</sup>
Baseline days 78 to 92	39.2	0.33	1.26	0.18	39.0	0.72	1.23
Controller days 78 to 92	14.7	0.12	1.22	0.17	0	0	0.29

The previous subsection showed that the controller did not apply costly inputs and, therefore, the resource usage and corresponding costs are lower. Only the CO<sub>2</sub> input is roughly equal, while the input trajectories also showed that the  $CO_2$  usage of the baseline was more moderate

and consistent throughout the simulation period. The total costs made by the controller for the simulated growing period of 14 days are 4 times lower than the baseline.

### Yield and profit

The simulated total fruit dry weight and harvested fruit dry weight of the fruits are shown in Figure 5-6. Yield is generated from selling harvested fruits against a fixed price. Table 5-2 shows the yield made and the resulting total costs and profit. Notice that the yield is in terms of fresh weight, rather than dry weight. The conversion between these two numbers can be achieved by dividing the dry weight by the Dry Matter Content which is taken constant as 8.94%, in line with Subsection 3-5-1.

**Table 5-2:** Comparison of the profit

Controller	Yield €/m <sup>2</sup>	Costs €/m <sup>2</sup>	Profit €/m <sup>2</sup>
Baseline days 78 to 92	2.45	1.23	1.22
Controller days 78 to 92	1.51	0.29	1.22

The lack of resources put into the system by the controller comes back in the resulting harvested and total fruit dry weight. The yield gained from selling the fruits is almost two times higher in the baseline case. Figure 5-6 already showed that the controller is bad at growing new fruits and that the harvest mainly comes from draining the already existing highest fruit stage. Solely looking at the profit of the two gives a distorted view; they are equal (1.22 €/m<sup>2</sup>) for the growing period of just 14 days, but for longer growing periods the baseline will definitely outperform the controller as the controller did not suffice in maintaining the total fruit weight.

## 5-3 Chapter summary and conclusion

The first part of this chapter showed the results of the system identification on the training dataset and the validation dataset, composed of multiple experiments with different but similar input trajectories. The prediction performances on the validation experiments were judged to be sufficient. However, two other experiments with inputs of higher frequency and larger amplitude also showed that the prediction performance diminished with enhanced input signals. From the experiments earlier and later on in the growing season using the identified system from the middle of the growing season, it can be concluded that the prediction performance depends on the operating point. This confirms the assumption that the system needs to be updated throughout the growing period. Section 6-2 in the Recommendations will elaborate more on this and will propose a solution to this problem. The experiments showed that the prediction of the harvested fruit dry weight is the limiting factor, as it only showed reasonable results for 1 day ahead. Therefore, a prediction horizon of 1 day was chosen for the proposed MPC-based controller with a control horizon of 30 minutes.

The main conclusion that can be drawn from the resulting outputs and inputs of the deployed controller on a simulation period of 14 days, is that the controller tries to save resources, rather

than spending resources to improve crop growth. The combined absence of sufficient heat,  $CO_2$  and light cause the plants to barely grow. The constraint on maintaining the total fruit dry weight was introduced to counteract this passive input strategy, but the prediction performance of the prediction model was not good enough to actually satisfy the constraint by applying the right inputs to the simulator. The reason for the bad prediction performance of the prediction model in the deployed controller could be that either the model order (10) was too low or that the experiments did not include enough variation of inputs and disturbances. Moreover, if the model would be updated over time, it could become more representative of the ground truth and have increased prediction performance.

Following the lower resource usage of the proposed controller, the costs were also lower ( $0.29 \text{ €/m}^2$ ), compared to those of the baseline ( $1.23 \text{ €/m}^2$ ). Solely looking at the profit of the two gives a distorted view; they are equal ( $1.22 \text{ €/m}^2$ ) for the growing period of just 14 days, but for longer growing periods the baseline will definitely outperform the controller as the controller did not suffice in maintaining the total fruit weight.

The next chapter will answer the research questions and will give recommendations for future work, based on the previously presented results.

# Conclusion and Recommendations

In the previous chapter, the results of the system identification and the performance of the proposed MPC-based optimal controller were presented and discussed. This chapter will draw the main conclusions of the thesis work, resulting in answers to the research questions. Moreover, recommendations for future work will be given.

## 6-1 Conclusion

The initial objective of this thesis was to investigate to what extent can a greenhouse-crop system be estimated with a linear state-space representation for the purpose of automating the decision-making in greenhouse climate control. In this work a ground-truth simulation of the greenhouse-crop dynamics was successfully created, that can be used for testing climate controllers. This simulation was composed of dynamics based on first principles, for which the mathematical representation was calibrated using data from the real-life growing experiment named the Autonomous Greenhouse Challenge. Next, system identification was investigated for optimal profit control. A linear state-space system was specified and was identified using the Prediction-Error Method (PEM). This system was used as a prediction model in the MPC-based controller for a maximum prediction horizon of 1 day and a control horizon of 30 minutes. The objective in the controller was chosen to be the profit, defined as the yield minus the costs of resource usage. One crucial constraint was added that forced the total amount of fruit weight to be maintained, to prevent trivial solutions and to prevent the controller from not growing any new fruits. The proposed controller did not outperform the baseline grower, but matched it in terms of profit ( $1.22 \text{ €/m}^2$ ). The attained total costs ( $0.29 \text{ €/m}^2$ ) were lower than that of the baseline ( $1.23 \text{ €/m}^2$ ), as was the yield ( $1.51 \text{ €/m}^2$  vs  $2.45 \text{ €/m}^2$ ) for the simulation period of 14 days. However, these numbers are deceiving; for longer growing periods the baseline will definitely outperform the controller as the controller did not suffice in maintaining the total fruit weight. The defined constraint on maintaining the total fruit weight was not met by the controller in the actual simulation. Therefore, it can be concluded that the identified linear system does not have sufficient prediction performance. The reason

for this bad prediction performance could be that either the model order (10) was too low or that the experiments did not include enough variation of inputs and disturbances. The research sub-questions can be answered as follows.

- **How does a modern-day greenhouse for tomato growing work and how can the greenhouse-crop system be modeled into a ground-truth simulation?**

A modern-day greenhouse contains an internal climate that influences the crops growing inside it. As described in Chapter 2, the greenhouse climate can be controlled with various inputs: heating, ventilation through opened windows, injection of  $CO_2$  and opening of screens for shading and keeping warmth inside. The most important greenhouse climate states are the air temperature, humidity and  $CO_2$ -concentration. Additionally, received light by solar radiation through the glass roof and by artificial lighting controlled as an input, play an important role in plant growth. The greenhouse climate influences plant and tomato-fruit growth. This growth can be described by non-linear dynamics including various fruit stages and assimilate weights, as presented in Chapter 3. In Chapter 3 a ground-truth simulation of the greenhouse-crop system was built, from a combination of fine-grained non-linear models from literature [20] [7] [24]. The simulation has been calibrated with data from a real-life tomato growing experiment, namely the Autonomous Greenhouse Challenge [6].

- **Under what conditions can the greenhouse-crop system be estimated with a linear state-space representation by using system identification?**

In Chapter 4 a linear state-space system with model order 10 is specified, after which it is identified for the greenhouse-crop system by using system identification. This so-called linearization is performed in an operating point around the time of the growing period where the plant growth reaches a steady-state, namely days 78. Here the dynamics of plant growth are evolving less. Baseline input-output data is used for the system identification procedure, where multiple experiment datasets are composed by altering the frequency and amplitude of the inputs. The total harvested dry weight is replaced with the change in harvested dry weight as the output of the prediction model to improve the system identification performance. The linear state-space model provides sufficient predictions on the air temperature,  $CO_2$ -concentration, humidity and the total dry weight for a period of at least 14 days. The model is less good at predicting the change in harvested dry weight for periods longer than 1 day. Therefore, the identified linear state-space model was used in the MPC-based profit controller for a prediction horizon of 1 day.

- **How can an identified state-space model be used in an MPC-based controller to automate the decision-making in greenhouse climate control?**

The identified system has been implemented as a predictor in an MPC-based controller to optimize on profit. Profit is achieved when the sold harvested dry weight is larger than the increase in resource costs. The objective in the proposed controller was based on the harvested dry weight, the costs of injecting  $CO_2$ , costs of heating and costs of using artificial lighting, all predicted with the identified linear state-space system over the prediction horizon. To ensure that the proposed controller not just harvested the already grown fruits, a new constraint is added that enforced new fruits to be grown.

This answers the research question as follows.

**To what extent can a greenhouse-crop system be estimated with a linear state-space representation for the purpose of automating the decision-making in greenhouse climate control?**

After deploying the proposed MPC-based controller for a simulation period of 14 days it became clear that the identified linear state-space representation of the greenhouse-crop system cannot be used to automate the decision-making with the objective to make a profit in greenhouse climate control, as the prediction model implemented in the proposed controller was not sufficiently accurate to produce the right system inputs to actually grow new fruits. The proposed controller tried to save resources, rather than spending resources to realize crop growth; the combined absence of sufficient heat,  $CO_2$  and light caused the plants to barely grow.

## 6-2 Recommendations

There are four main recommendations for future work on linear state-space system identification for automatic greenhouse climate control. The first one regards the limitations of the ground-truth simulation. The second is about the bounds of the inputs. The third recommendation is about the price coefficients and the last recommendation is on a proposed control architecture, namely on-line system identification.

### 6-2-1 Limitations of the ground-truth simulator

The ground-truth simulator has its own limitations regarding the modeling of the boundaries of the climate. One of which is that there are no plant diseases or other plant stresses modeled, while these are certainly present in real life when the greenhouse climate reaches its boundaries. Another limitation is the absence of dynamics and states describing the dry matter content and the quality of the tomatoes. The latter determines the selling price of the tomato, which is a very important aspect of running a real-life greenhouse operation. To make an even more realistic simulator these would be points of improvement.

### 6-2-2 Boundaries of the system and input constraints

From the results, it became clear the proposed controller it generates inputs that are either on the lower or upper boundary values, see  $CO_2$ -injection and the opening of screens in Subsection 5-2-2. As was shown in Section 5-1, the prediction performance of the identified state-space model decreases with inputs of this kind. Also, there are no bounds given on the outputs the controller can steer the climate to, while reaching those bounds may cause inaccurate predictions and bad fruit quality. Therefore, it is recommended to let the bounds of the climate and the inputs vary with time and the operating point. An attempt at this was made by van Beveren et al. (2015), where the controller had to take into account the grower' defined bounds [24].

### 6-2-3 Time-varying cost and price coefficients

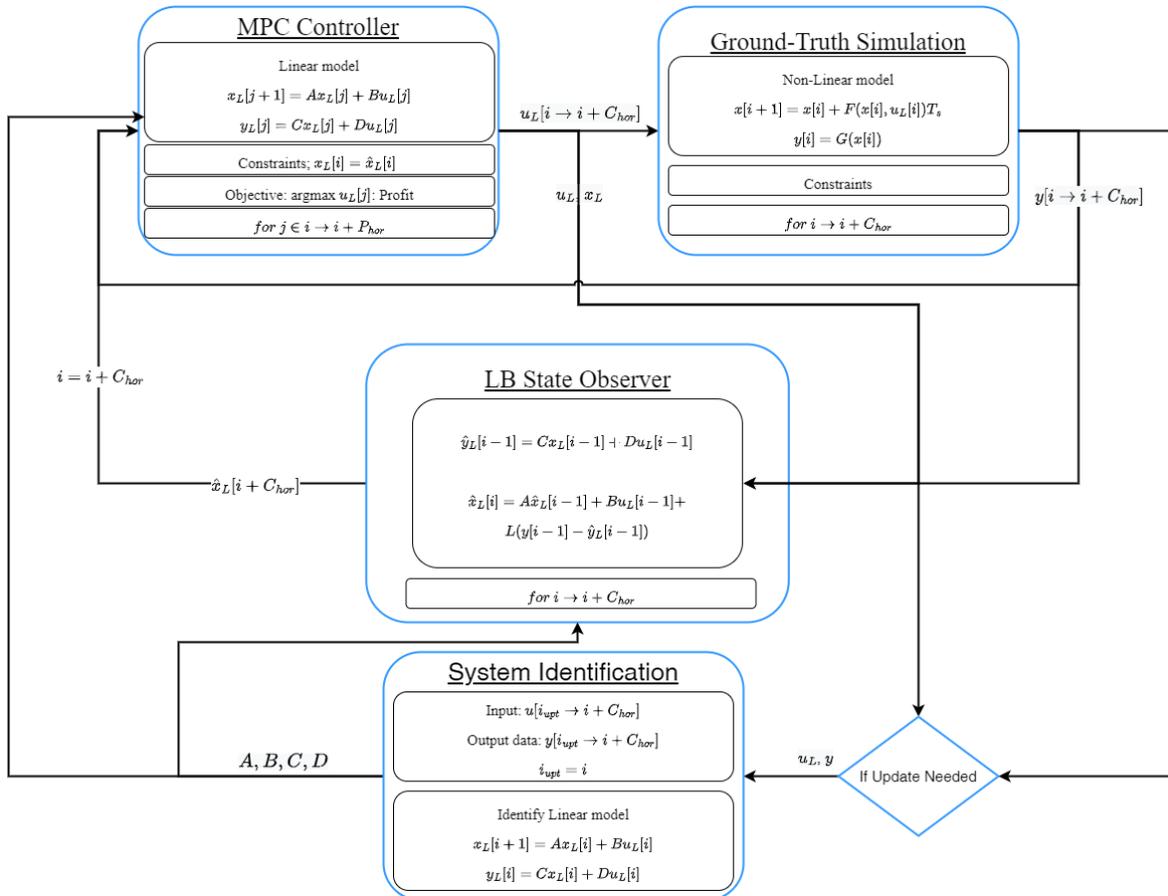
Originally, the plan was to also include time-varying cost and price coefficients. The preceding literature study showed that all previous attempts to control the greenhouse-crop system included fixed cost and price coefficients, which was not realistic as these coefficients fluctuate throughout the growing season. Therefore, for future work, it is recommended to include the factor of time in these coefficients. This thesis excluded this because the prediction horizon was not long enough to properly capture the dynamics of the changing coefficients.

### 6-2-4 On-line system identification

From the validation experiments, it became clear that the predictions made by the identified linear state-space model were only successful for the middle of the growing season, as the data used for the identification procedure was captured around that time. The model performed much worse at the beginning of the growing season and the rationale behind this is that in this crop growth stage the non-linear dynamics are evolving more and are, therefore, more difficult to accurately represent in a linear fashion. Moreover, the implementation of the prediction model in the proposed controller was less successful as constraints could not be met. There is a potential solution to improve prediction capabilities over the complete growing seasons that is recommended to be investigated further: *on-line system identification*. We propose the following control method that updates the linear state-space prediction model when the predictions are not sufficiently accurate anymore.

The control architecture of the proposed method is shown in Figure 6-1, and can be formulated as follows. First, the MPC-based controller optimizes on the defined objective, over prediction horizon  $P_{hor}$  using the identified linear model. Then,  $\mathbf{u}_L[i \rightarrow i + C_{hor}]$ , is sent to the ground-truth simulation that represents reality. This ground-truth simulation, presented in Chapter 3, uses non-linear model dynamics  $\mathbf{F}(\mathbf{x}[i], \mathbf{u}_L[i])$ , multiplied by sampling time  $T_s$ . The output  $\mathbf{y}[i]$  is related to the state as  $\mathbf{G}(\mathbf{x}[i])$ . The output  $\mathbf{y}[i]$  is used as feedback for the MPC-based controller and is translated to an estimate state by the state observer. The objective, see Subsection 4-3-2 is making a profit while maintaining the total fruit weight.

The linear model is updated via system identification when a system update protocol indicates so. This system update protocol can be, for example, based on the fit, see (4-6), exceeding a threshold. For this, the output data is compared with the input data, from timestep  $i_{upt}$  till the current timestep  $i$ . The outputs of this process are the linear system matrices  $\mathbf{A} \in \mathbb{R}^{n_{xL} \times n_{xL}}$ ,  $\mathbf{B} \in \mathbb{R}^{n_{xL} \times n_{uL}}$ ,  $\mathbf{C} \in \mathbb{R}^{n_{yL} \times n_{xL}}$  and  $\mathbf{D} \in \mathbb{R}^{n_{yL} \times n_{uL}}$ .



**Figure 6-1:** Proposed control architecture for including on-line system identification.

With this method, one must be careful with closed-loop identification, as explained next.

### Closed-loop system identification

The goal of the identification process is to find a good fit of the open-loop system onto the input-output data. However, since the input-output data is generated by a closed-loop optimal controller, there is a possibility that the input data is correlated with the disturbances and outputs. This typical problem of closed-loop identification is also explained by P. Van den Hof [53]. One of the solutions given by Van den Hof to this problem is to identify the system without taking into account the presence of a feedback controller. The interested reader is referred to the corresponding paper.

Another solution that could in particular work for greenhouse climate control is to build in a period where the system's dynamics are sufficiently excited, to create better datasets. For this to work in a real operational setting, the inputs have to be chosen such that plant growth is not significantly obstructed. One way to do this is to apply an input signal that would still give acceptable climate trajectories within bounds, but with extra noise to the input.



---

# Appendix A

---

## Greenhouse-crop model for simulation purposes

The following model of the greenhouse-crop system is mainly based on the model of Kuijpers et al. (2021), which is originally created by Vanthoor (2011) [7] [20].

### A-1 Mass flows crop model

Mass Flow Rate [ $gm^{-2}s^{-1}$ ]	Dynamics	Description
$MC_{AirBuf}$	Non-Linear	The photosynthesis rate; generating assimilates.
$MC_{BufFrt}$	Non-Linear	The flow of assimilates from assimilate buffer to the fruits.
$MC_{BufLeaf}$	Non-Linear	The flow of assimilates from assimilate buffer to the leaves.
$MC_{BufStem}$	Non-Linear	The flow of assimilates from assimilate buffer to the stems.

$MC_{BufAir}$	Linear	The growth respiration of the assimilate buffer. Summation of growth respiration of different crop organs. Linearly related to assimilate flow to the corresponding organ.
$MC_{FrtAir}$	Non-Linear	The maintenance respiration of the fruits.
$MC_{LeafAir}$	Non-Linear	The maintenance respiration of the leaves.
$MC_{StemAir}$	Non-Linear	The maintenance respiration of the stems.
$MN_{Frt\{i\}Frt\{i+1\}}$	Non-Linear	Fruit flow from one development stage to the next.
$MC_{BufFrt\{i\}}$	Non-Linear	The fruit assimilate flow from the buffer to a development stage.
$MC_{Frt\{i\}Frt\{i+1\}}$	Non-Linear	The fruit assimilate flow from one development stage to the next.
$MC_{LeafHar}$	Non-Linear	The leaf harvest rate.

**Table A-1:** Overview of the mass flow rates for crop growth. The dynamics and model parameters are well described in [7] and [7, Tab. 9.1.], respectively.  $*rg_X$  is linearly dependent on  $T_{c24}$  [7, Eq. 9.28].  $**h_Y$  represents a non-instantaneous temperature-dependent inhibition, i.e. sub-optimal growth, see [7, Fig. 3.3.]. Physical constants are not considered model parameters.

## A-2 Ground-truth: energy and mass fluxes

**Table A-2:** Energy and mass fluxes of the greenhouse-crop model for simulation purposes.

Model Fluxes	Unit	Description
$Q_{srd}$	$Wm^{-2}$	Incoming solar radiation
$Q_{cov}$	$Wm^{-2}$	Heat loss through the cover
$Q_{trans}$	$Wm^{-2}$	Transpiration by the crop
$Q_{pipe}$	$Wm^{-2}$	Heating by the pipe rail system

$Q_{lamp}$	$Wm^{-2}$	Heating from artificial Lighting
$Q_{vent}$	$Wm^{-2}$	Natural ventilation
$\phi_{trans}$	$gm^{-2}s^{-1}$	Crop transpiration
$\phi_{cov}$	$gm^{-2}s^{-1}$	Condensation on the cover
$\phi_{trans}$	$gm^{-2}s^{-1}$	Vapor exchange with outside air due to natural ventilation
$\psi_{c,inj}$	$gm^{-2}s^{-1}$	Injection of $CO_2$
$\phi_{c,ass}$	$gm^{-2}s^{-1}$	Assimilation of $CO_2$ by the crop
$\phi_{c,vent}$	$gm^{-2}s^{-1}$	$CO_2$ exchange with outside air due to natural ventilation

### A-3 Ground-truth: model parameters

**Table A-3:** Model parameters of the greenhouse-crop model for simulation purposes.

Model Parameters	Unit	Value	Description
$\rho_{air}$	$kgm^{-3}$	1.29	density of air
$L$	$Jkg^{-1}$	$2.27 * 10^3$	energy needed to evaporate water
$p_{atm}$	$kpa$	101	Atmospheric pressure
$c_p$	$J^{\circ}C^{-1}kg^{-1}$	1010	Specific heat of air at constant pressure
$C_p$	$J^{\circ}C^{-1}kg^{-1}$	4180	Specific heat of water at constant pressure
$\eta_{mg\_ppm}$	$ppmmg^{-1}m^3$	554	$CO_2$ conversion factor from $mg$ to $ppm$
$M_{CH_2O}$	$mg \{CH_2O\} mol^{-1} \{CO_2\}$	$30 * 10^{-3}$	Molar mass of $CH_2O$
$SLA$	$m^2 \{leaf\} mg^{-1} \{CH_2O\}$	$2.66 * 10^{-5}$	Specific Leaf Area Index
$R$	$Jmol^{-1}K^{-1}$	8.314	Molar gas constant
$G^{Max}$	$mgfruit^{-1}$	$mg \{CH_2O\} fruit^{-1}$	Potential fruit dry weight at harvest

**Table A-4:** Calibrated Model parameters of the ground-truth crop model

Model Parameters	Unit	Original value	Calibrated value	Description
$a_{sunPAR}$	—	0.89	0.74	fraction of PAR in the radiation
$a_{sunNIR}$	—	0.21	0.21	fraction of NIR in the radiation
$a_{LEDPAR}$	—	0.94	1.1	fraction of PAR in the Light Emitting Diode (LED) light
$a_{LEDNIR}$	—	0.92	0.92	fraction of NIR in the LED light
$k_{PAR}$	—	0.7	0.93	extinction coefficient for PAR of the canopy
$k_{NIR}$	—	0.27	0.27	extinction coefficient for NIR of the canopy
$C_{Buf}^{Max}$	$mgm^{-2}$	$20 * 10^3$	$20 * 10^3$	Maximum buffer capacity
$s_{C_{Buf}}^{MC_{AirBuf}}$	—	$5 * 10^{-4}$	$5.26 * 10^{-4}$	Switching parameter
$C_{Buf}^{Min}$	$mgm^{-2}$	$1 * 10^3$	$1 * 10^3$	Minimum amount of carbohydrates in the buffer
$s_{C_{Buf}}^{MC_{BufOrg}}$	—	$-5 * 10^3$	$-5 * 10^3$	Switching parameter
$\alpha$	—	0.385	0.577	The conversion factor from photons to electrons including an efficiency term
$\theta$	—	0.7	0.597	Degree of curvature of the electron transport rate
$E_j$	$Jmol^{-1}$	$37 * 10^3$	$37 * 10^3$	Activation energy for $J^{POT}$ calculation
$S$	$Jmol^{-1}K^{-1}$	710	710	Entropy term for $J^{POT}$ calculation
$J_{25,Leaf}^{MAX}$	$\mu mol \{e^{-}\} m^{-2} \{leaf\} s^{-1}$	210	210.1	Maximum rate of electron transport at $25^{\circ}C$ for the leaf
$\eta_{CO_2AirStom}$	$\mu mol \{CO_2\} mol^{-1} \{air\}$	0.67	0.53	Conversion factor from the greenhouse air $CO_2$ -concentration to the $CO_2$ -concentration in the stomata

$c_{Gamma}$	$\mu mol \{CO_2\} mol^{-1} \{air\} K^{-1.7}$		2.1		The effect of canopy temperature on the $CO_2$ compensation point
$H$	$Jmol^{-1}$	$22 * 10^4$	$22 * 10^4$		Deactivation energy for $J^{POT}$ calculation
$n_{Plants}$	—	2.5	2.9		Number of plants
$C_{BufFruit1}^{Max}$	$fruits plant^{-1} s^{-1}$	$-1.71 * 10^{-7}$	$-1.63 * 10^{-7}$		Max fruit set regression coefficient 1
$C_{BufFruit2}^{Max}$	$fruits plant^{-1} s^{-1} C^{-1}$	$7.31 * 10^{-7}$	$8.25 * 10^{-7}$		Max fruit set regression coefficient 2
$r_{BufFrt}^{MaxFrtSet}$	$mg \{CH_2O\} m^{-2} s^{-1}$	0.1	0.1		Carbohydrate flow from buffer to the fruits above which fruit set is maximal
$s_{MC_{BufFruit}}^{MN_{BufFruit}}$	—	-58.9	-58.9		Switching parameter
$MC_{BufFruit}^{Switch}$	$mgm^{-2} s^{-1}$	0.05	0.05		Switching parameter
$c_{Dev1}$	$s^{-1}$	$-7.64 * 10^{-9}$	$-8.61 * 10^{-9}$		Fruit Development rate coefficient 1
$c_{Dev2}$	$s^{-1} C^{-1}$	$1.16 * 10^{-8}$	$7.38 * 10^{-9}$		Fruit Development rate coefficient 2
$s_{T_{can}}^{Min}$	—	-0.8690	-0.7512		Switching parameter
$s_{T_{can}}^{Max}$	—	0.5793	0.5131		Switching parameter
$T_{Can}^{Min}$	$^{\circ}C$	10	10.6		Switching parameter
$T_{Can}^{Max}$	$^{\circ}C$	34	49.1		Switching parameter
$s_{T_{c24}}^{Min}$	—	-1.1587	-1.0497		Switching parameter
$s_{T_{c24}}^{Max}$	—	1.1904	1.5019		Switching parameter
$T_{Can24}^{Min}$	$^{\circ}C$	15	12.4		Switching parameter
$T_{Can24}^{Max}$	$^{\circ}C$	24.5	59		Switching parameter
$rg_{Fruit}$	$mg \{CH_2O\} m^{-2} s^{-1}$	0.328	0.2813		Potential fruit growth rate coefficient at $20^{\circ}C$
$rg_{Leaf}$	$mg \{CH_2O\} m^{-2} s^{-1}$	0.095	0.0820		Potential leaf growth rate coefficient at $20^{\circ}C$
$rg_{Stem}$	$mg \{CH_2O\} m^{-2} s^{-1}$	0.074	0.0875		Potential stem growth rate coefficient at $20^{\circ}C$
$c_{Fruit\_m}$	$mg \{CH_2O\} mg^{-1} \{CH_2O\} s^{-1}$	$1.16 * 10^{-7}$	$1.16 * 10^{-7}$		Fruit maintenance respiration coefficient
$c_{Leaf\_m}$	$mg \{CH_2O\} mg^{-1} \{CH_2O\} s^{-1}$	$3.47 * 10^{-7}$	$3.90 * 10^{-7}$		Leaf maintenance respiration coefficient
$c_{Stem\_m}$	$mg \{CH_2O\} mg^{-1} \{CH_2O\} s^{-1}$	$1.47 * 10^{-7}$	$1.80 * 10^{-7}$		Stem maintenance respiration coefficient
$c_{Stem\_g}$	—	0.3	0.3		Stem growth respiration coefficient

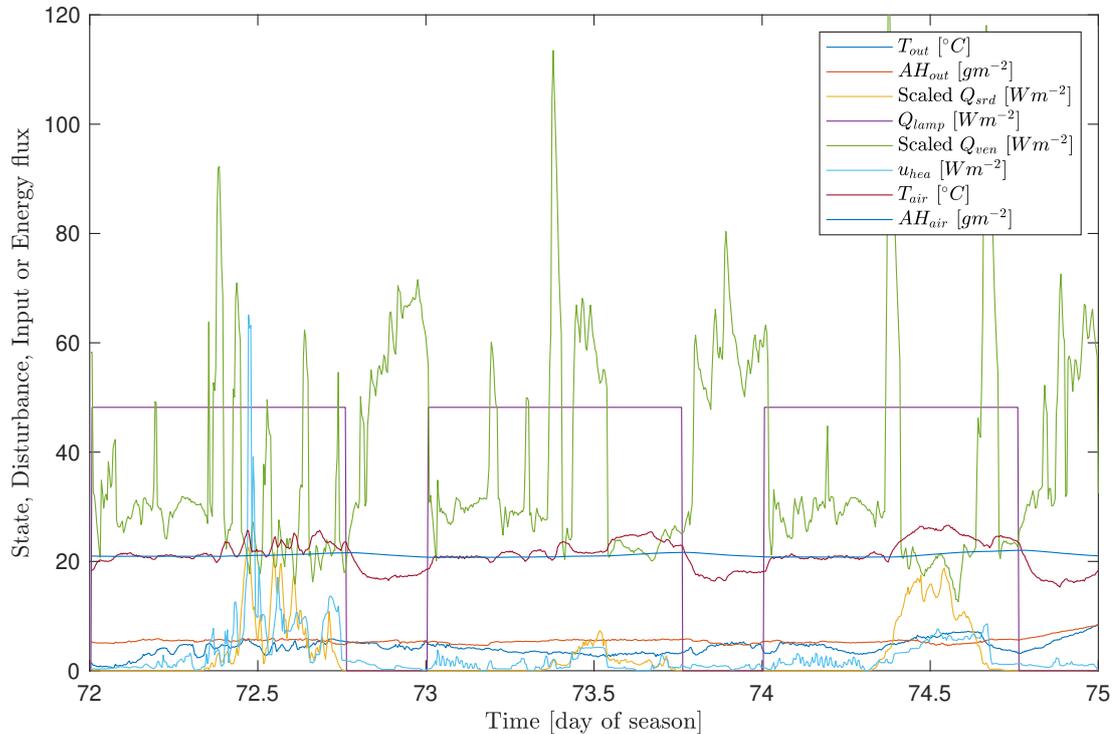
$c_{Leaf\_g}$	—	0.28	0.3176	Leaf growth respiration coefficient
$c_{Fruit\_g}$	—	0.27	0.3163	Fruit growth respiration coefficient
$Q_{10}$	—	2	1.7312	$Q_{10}$ value of temperature effect on maintenance respiration

**Table A-5:** Calibrated Model parameters of the ground-truth greenhouse climate model

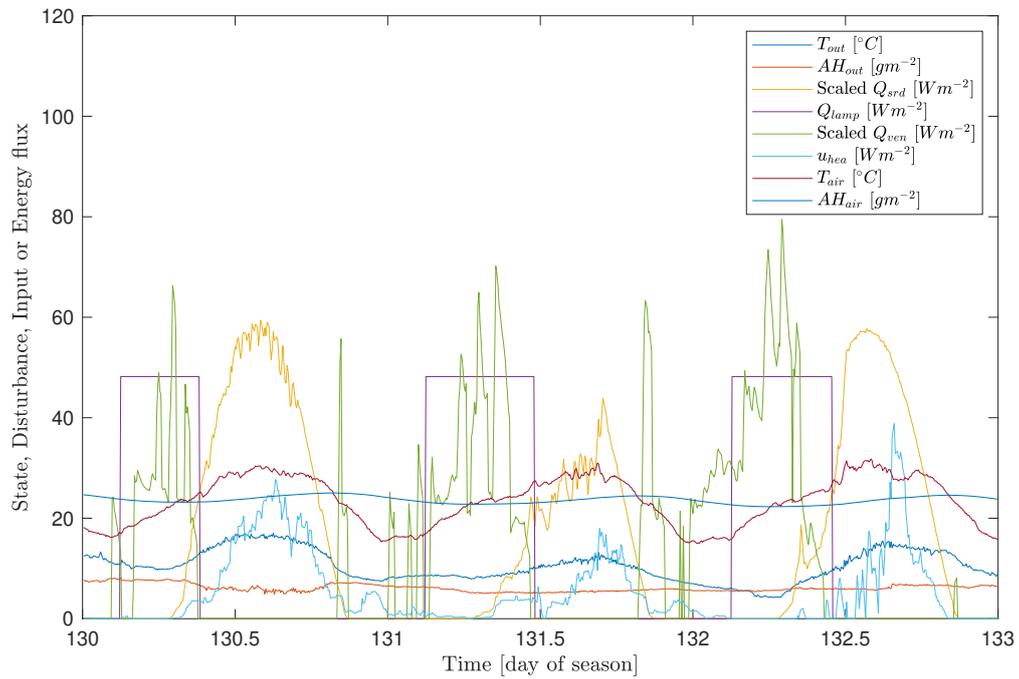
Model Parameters	Unit	Original value	Calibrated value	Description
$\alpha_{cov}$	—	5	499	heat conductance of the greenhouse cover
$\alpha_{scr1}$	—	9.33	0.0933	heat conductance of the temperature screen
$\alpha_{scr2}$	—	9.33	0.0933	heat conductance of the blackout screen
$\gamma$	—	0.4	0.0615	Apparent psychometric constant
$\epsilon$	—	3.0	0.03	ratio of latent to sensible heat content of saturated air
$\zeta$	—	$2.7060 * 10^{-5}$	$4.0920 * 10^{-4}$	Ventilation rate parameter
$\xi$	$\%^{-1}$	$6.3233 * 10^{-5}$	$2.1400 * 10^{-4}$	Ventilation rate parameter
$\sigma$	$\%^{-1}$	$7.1708 * 10^{-5}$	$7.1708 * 10^{-7}$	Ventilation rate parameter
$c_{lamp}$	—	0.5	1.0	Lamp heat coefficient
$\tau_{cov}$	—	0.7	0.7	transmittance of the cover
$\tau_{scr1}$	—	0.9	0.3053	transmittance of the temperature screen
$\tau_{scr2}$	—	0.3042	0.0101	transmittance of the climate screen
$P_{scr1}$	—	0.5	0.1544	Cover screen ventilation parameter
$P_{scr2}$	—	0.5	0.1	Cover screen ventilation parameter
$\chi$	$\%^{-1}$	0.0156	1.56	Ventilation rate parameter
$\psi$	$ms^{-1}$	$7.4 * 10^{-5}$	$7.4 * 10^{-7}$	Ventilation rate parameter
$c_{cap}$	$^{\circ}C^{-1}m^{-2}$	30000	31784	specific heat capacity of the greenhouse air
$pgc$	$m^{\circ}C^{-\frac{1}{3}}s^{-1}$	$1.8 * 10^{-3}$	0.011	specific properties of the cover
$\omega$	—	0.622	0.0011	Humidity ratio parameter

$\eta$	—	0.7	0.2846	fraction of PAR in the radiation
$h$	$m$	6.8	81.4549	average height of the greenhouse
$r_b$	$sm^{-1}$	150	123.17	boundary layer resistance
$c_{pipe1}$	—	0.6	2.1388	Pipe 1 heat radiation coefficient
$c_{pipe2}$	—	2.1	1.7305	Pipe 2 heat radiation coefficient
$CO2\_corr$	—	1	$1.79 * 10^4$	$CO_2$ -injection correction factor

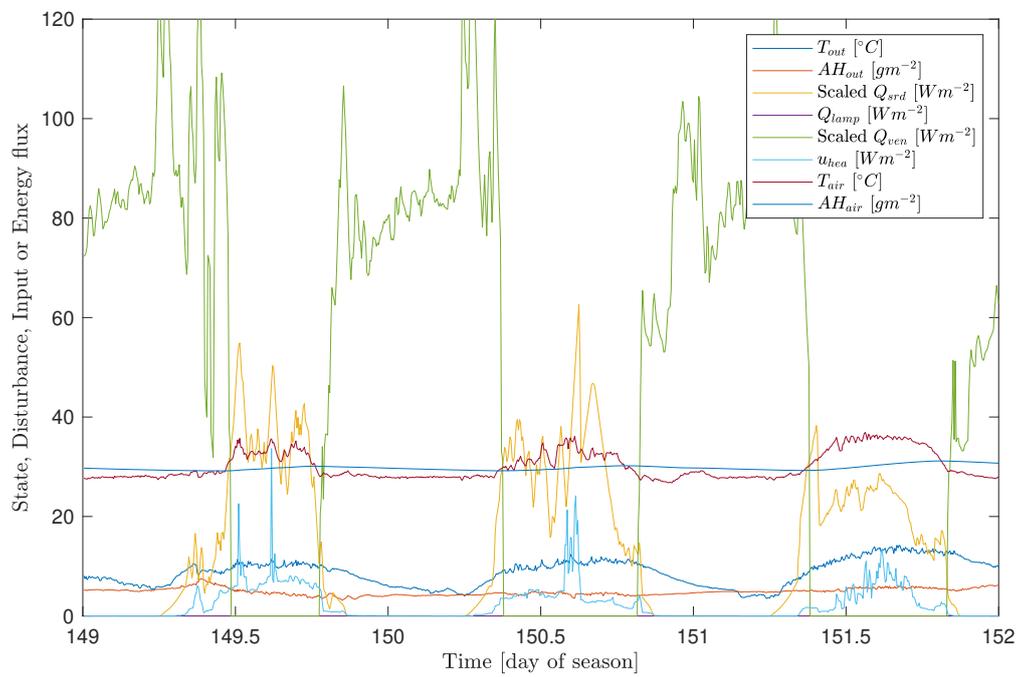
## A-4 Ground-truth calibration data windows



**Figure A-1:** Window on the 72th day.  $Q_{srd}$  and  $Q_{ven}$  are scaled down.



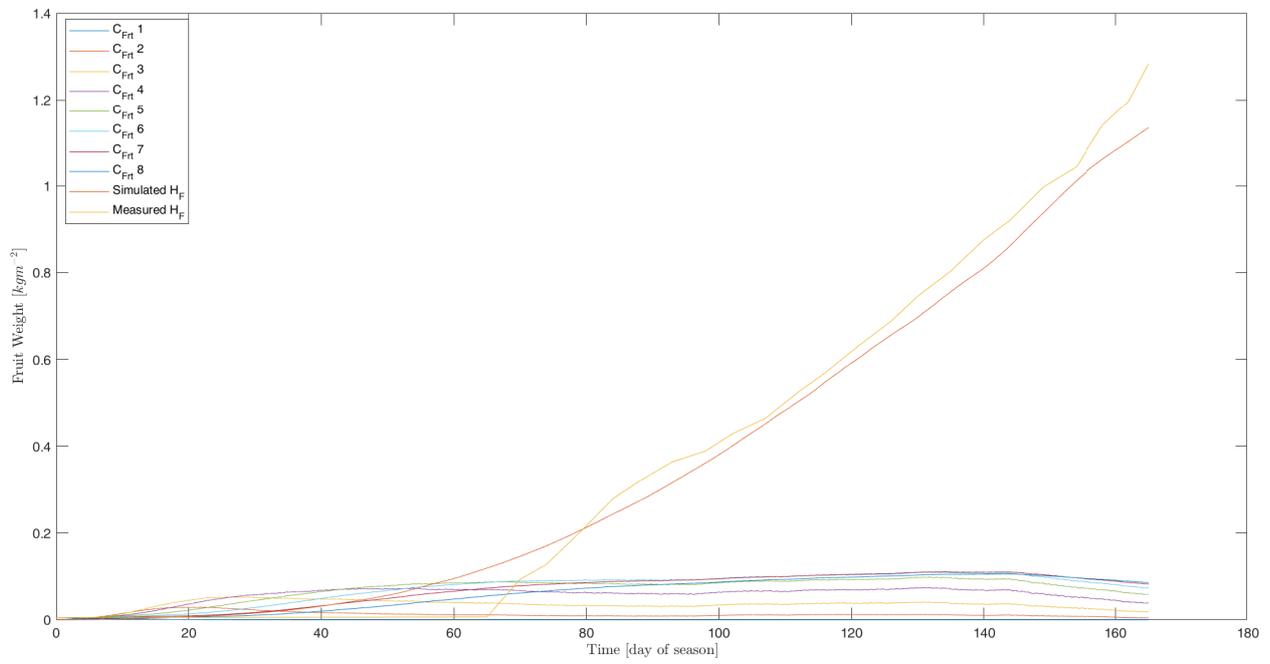
**Figure A-2:** Window on the 130th day.  $Q_{srd}$  and  $Q_{ven}$  are scaled down.



**Figure A-3:** Window on the 149th day.  $Q_{srd}$  and  $Q_{ven}$  are scaled down.

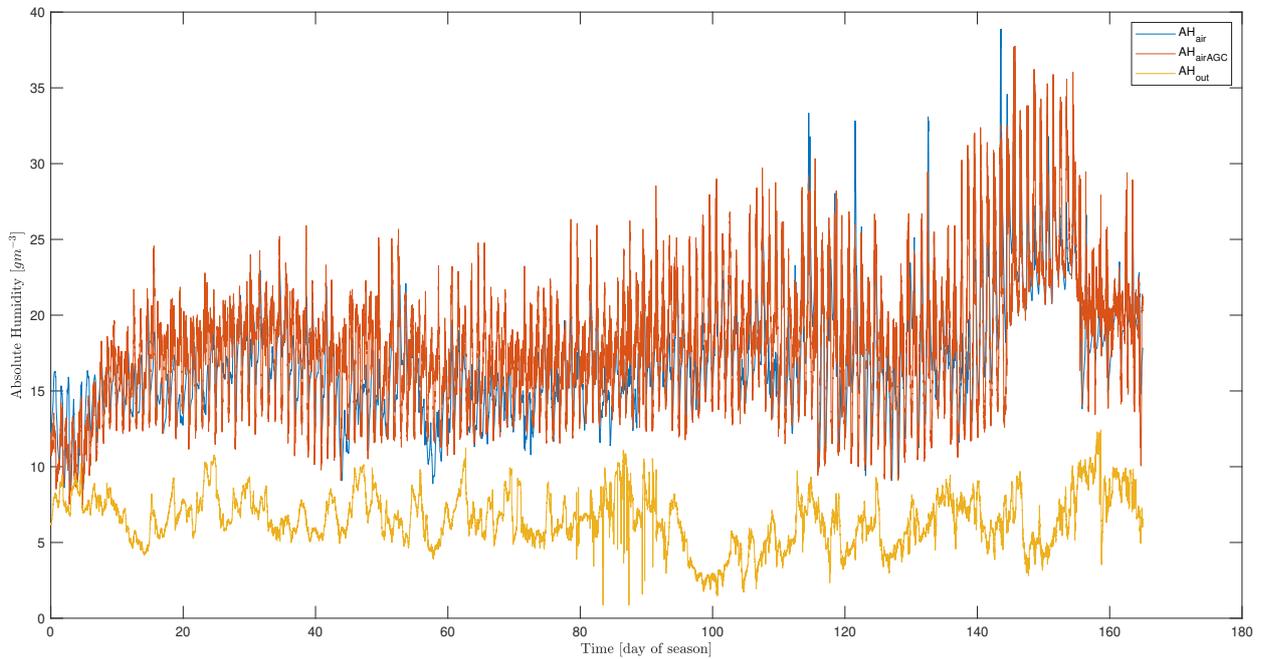
## A-5 Ground-truth simulation figures

### A-5-1 Ground-truth fruit dry weight per development stage

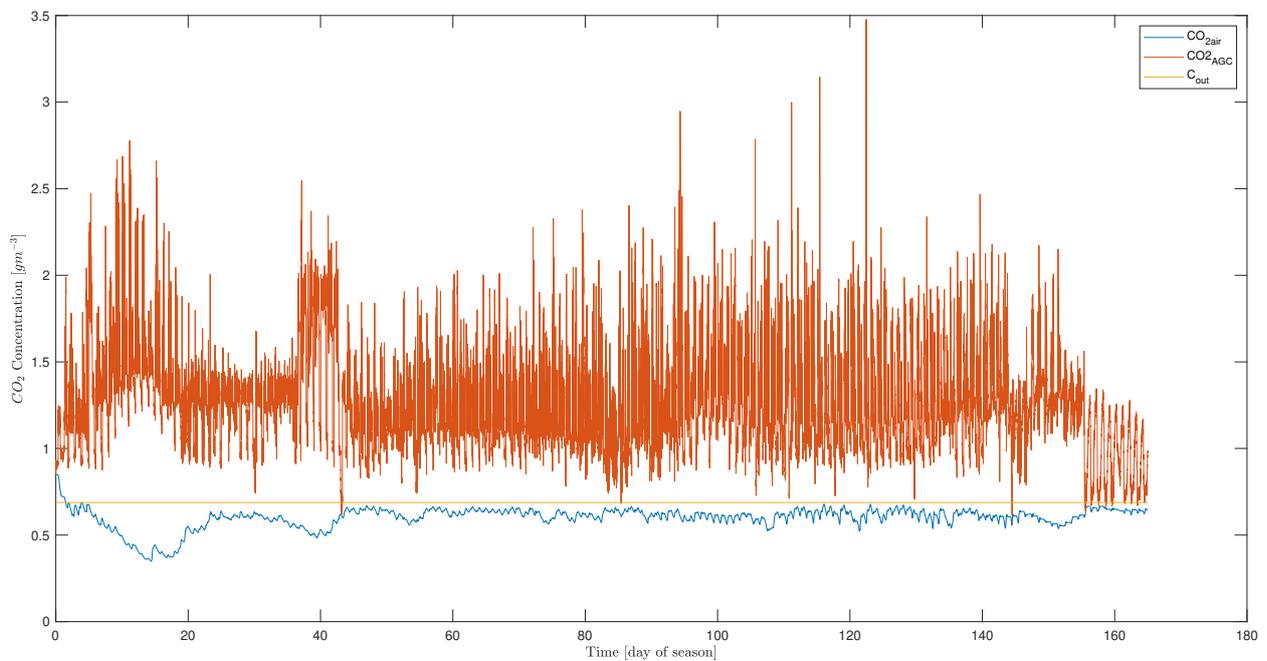


**Figure A-4:** Season-wide fruit dry weight per development stage simulation of the calibrated ground-truth model.

### A-5-2 Ground-truth season-wide simulations

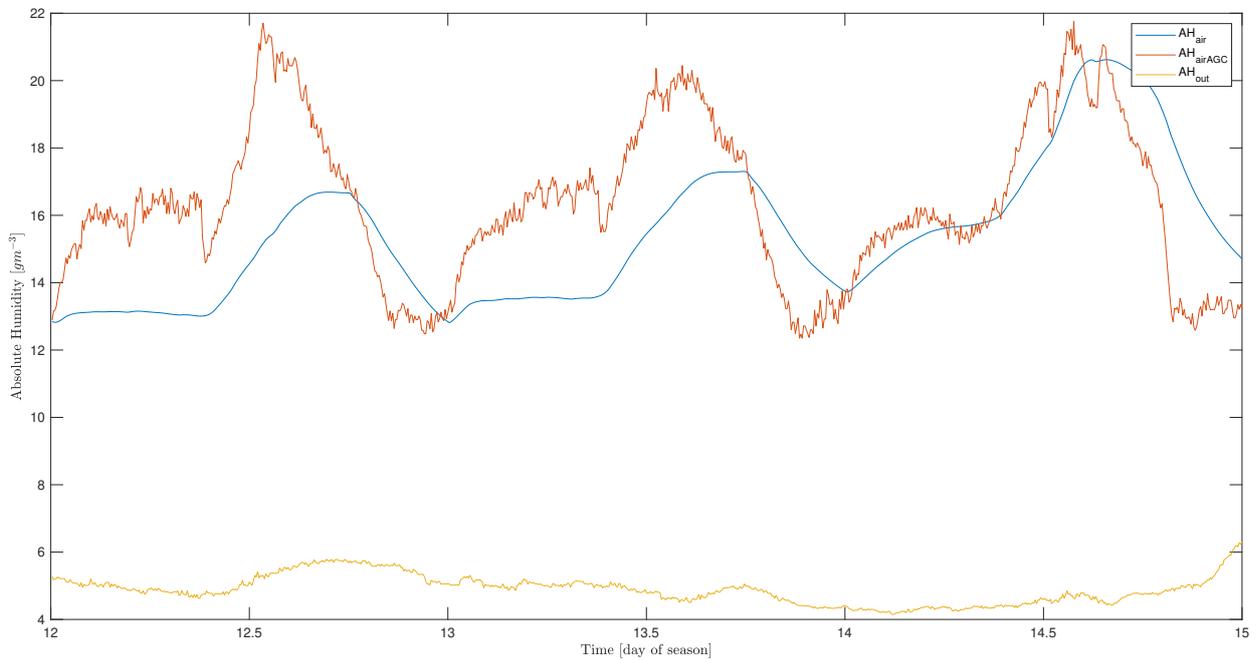


**Figure A-5:** Season-wide absolute humidity simulation of the calibrated ground-truth model, compared to the measured and outside absolute humidity.

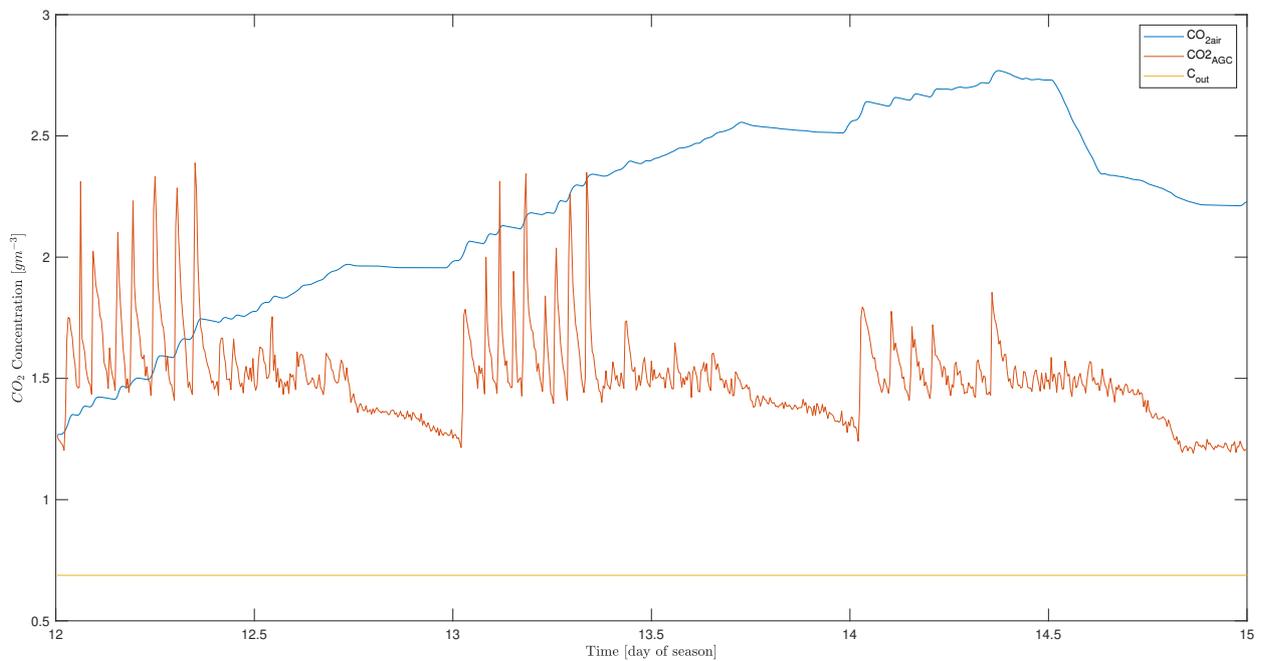


**Figure A-6:** Season-wide  $CO_2$ -concentration simulation of the calibrated ground-truth model, compared to the measured and outside  $CO_2$ -concentration.

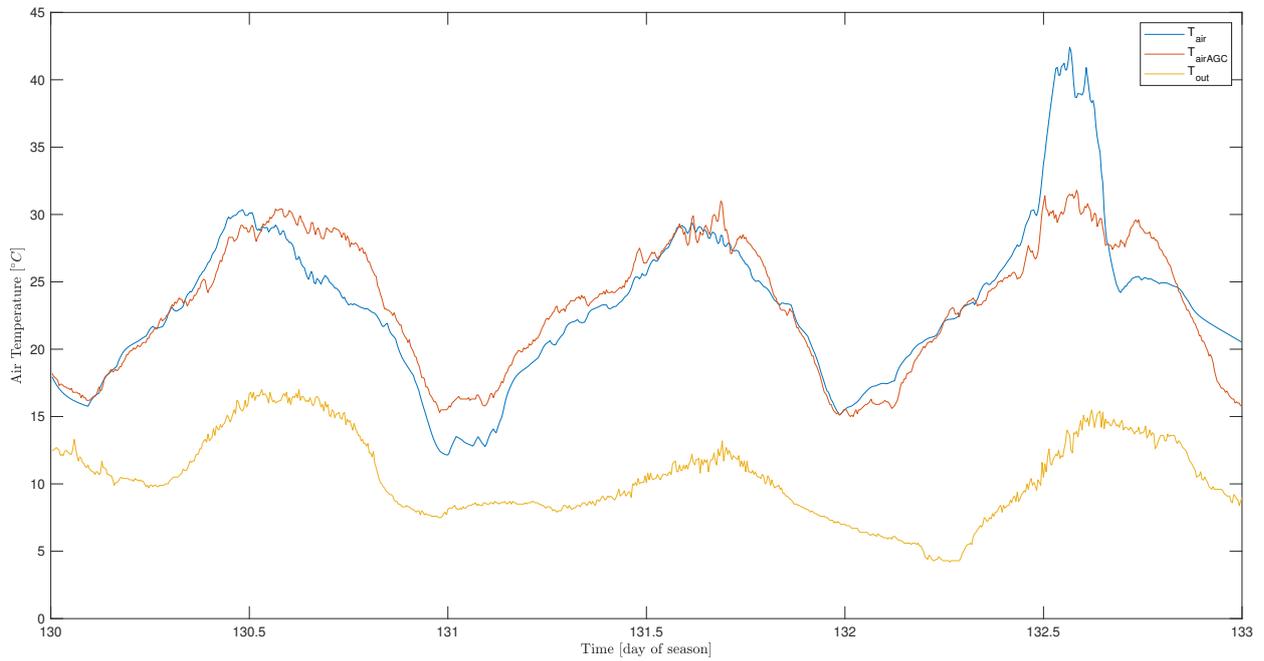
### A-5-3 Ground-truth 3-day simulations



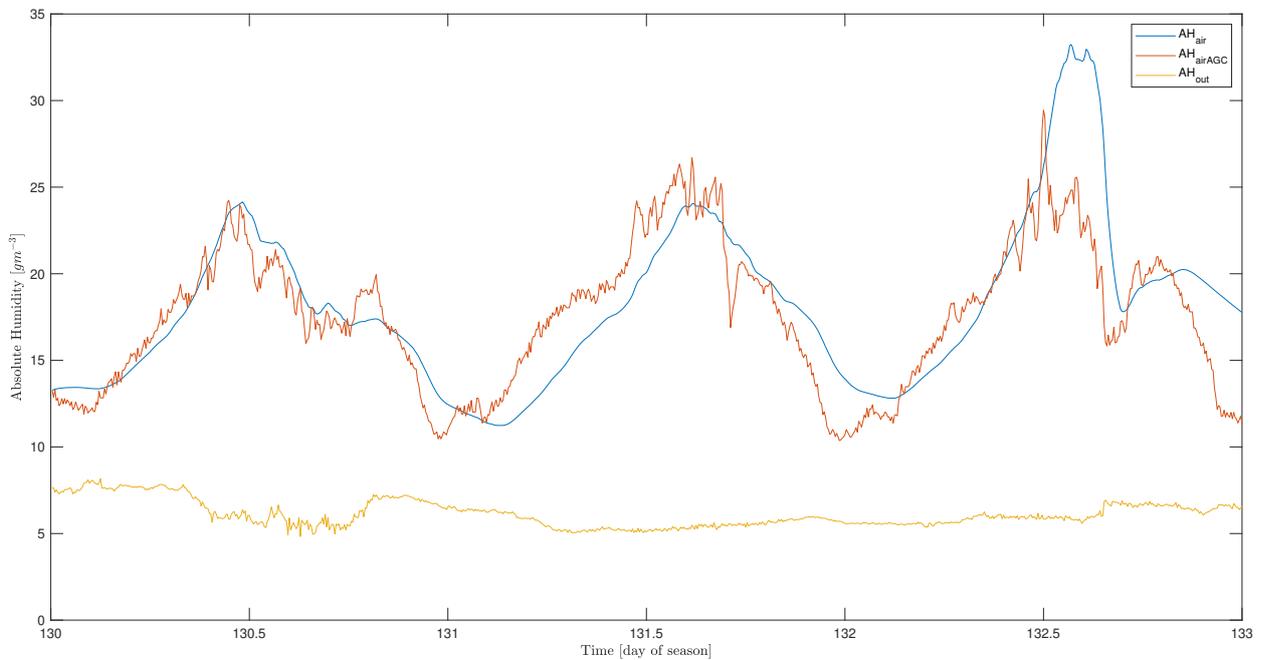
**Figure A-7:** 3-day absolute humidity simulation of the calibrated ground-truth model starting from day 12, compared to the measured and outside absolute humidity.



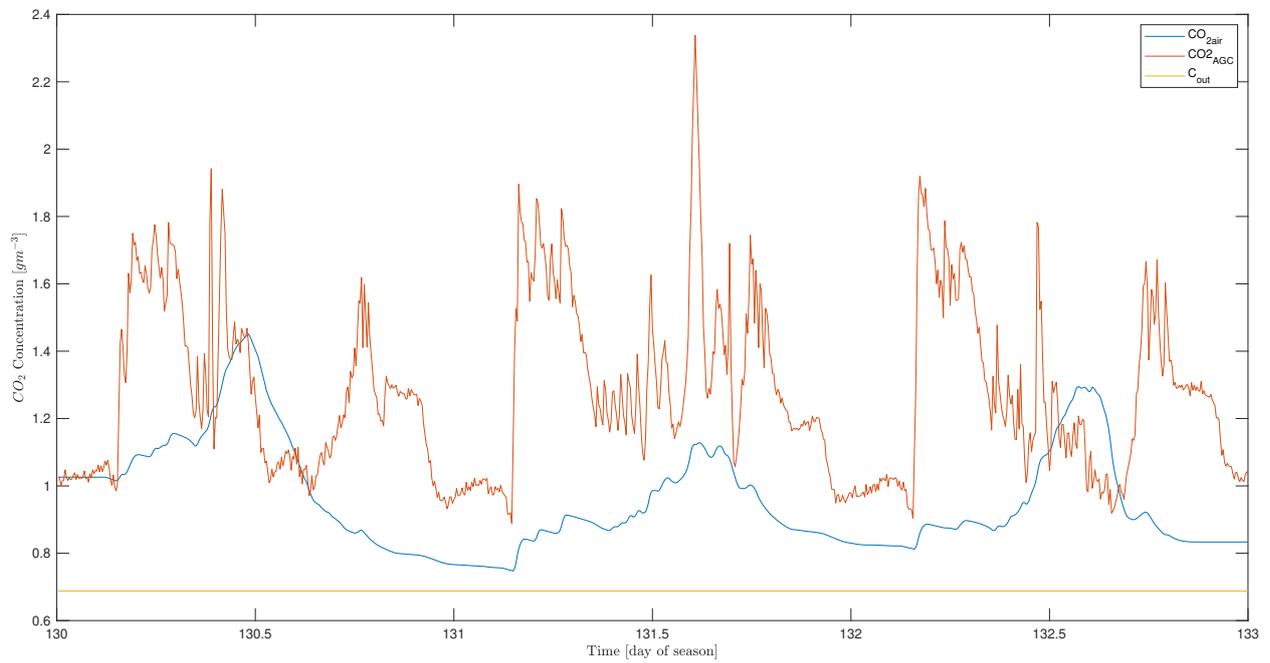
**Figure A-8:** 3-day  $CO_2$ -concentration simulation of the calibrated ground-truth model starting from day 12, compared to the measured and outside  $CO_2$ -concentration.



**Figure A-9:** 3-day temperature simulation of the calibrated ground-truth model starting from day 130, compared to the measured and outside temperature.



**Figure A-10:** 3-day absolute humidity simulation of the calibrated ground-truth model starting from day 130, compared to the measured and outside absolute humidity.



**Figure A-11:** 3-day  $CO_2$ -concentration simulation of the calibrated ground-truth model starting from day 130, compared to the measured and outside  $CO_2$ -concentration.



---

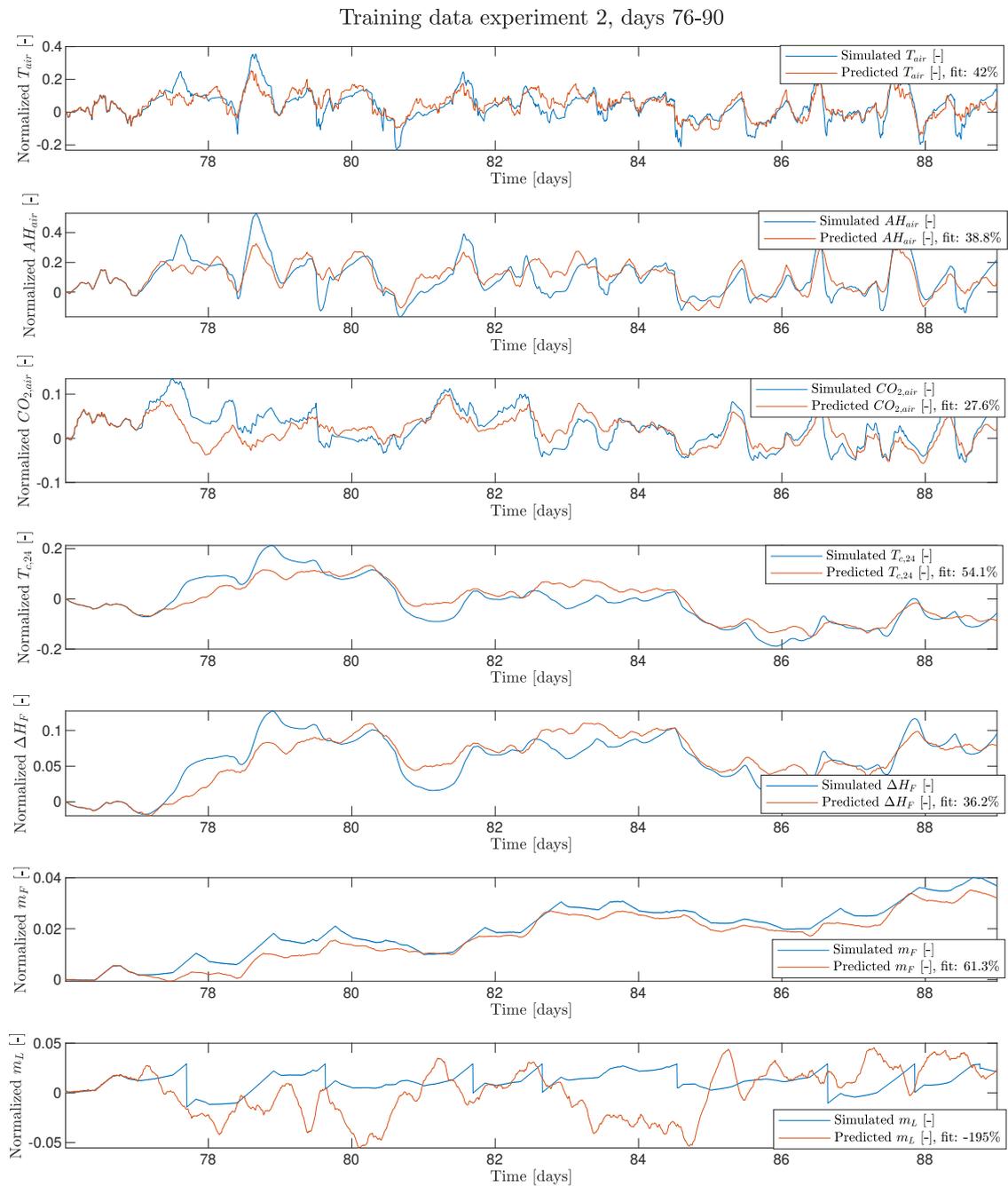
## Appendix B

---

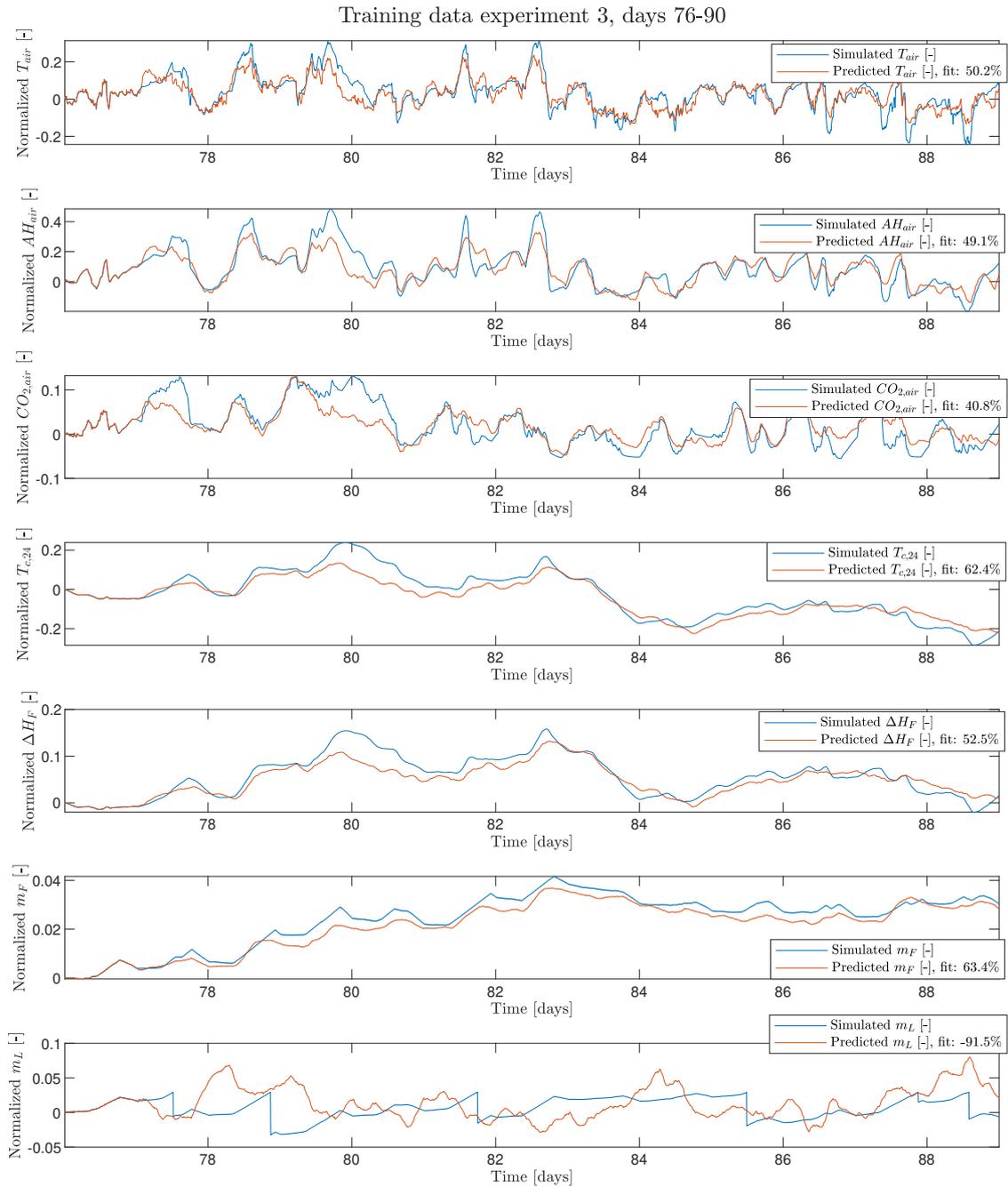
# System Identification Results

This appendix section shows the fits of the identified system in the training and validation experiments.

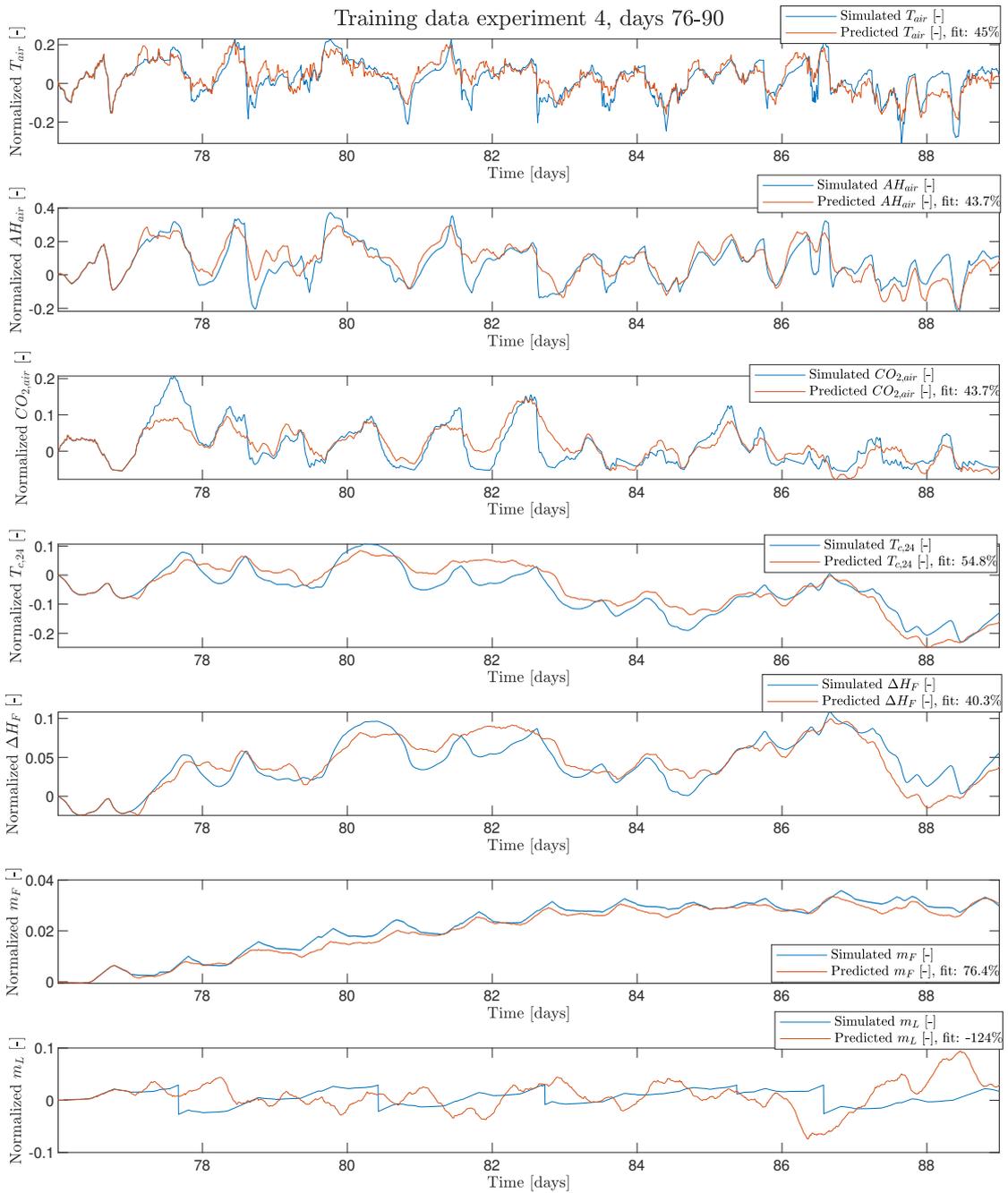
## B-1 Training dataset fits



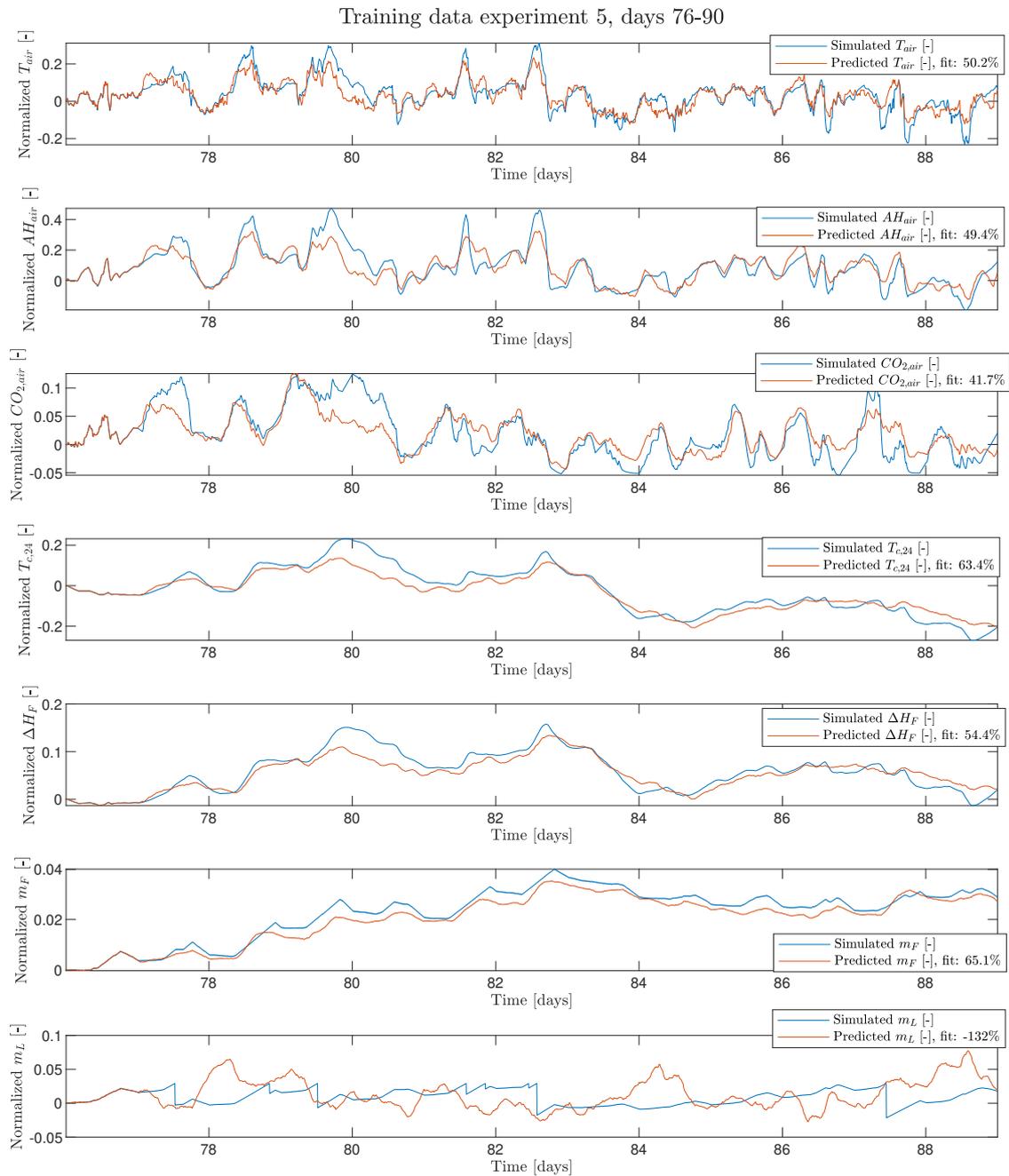
**Figure B-1:** Fit of identified linear state space system on the training dataset of a second experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.



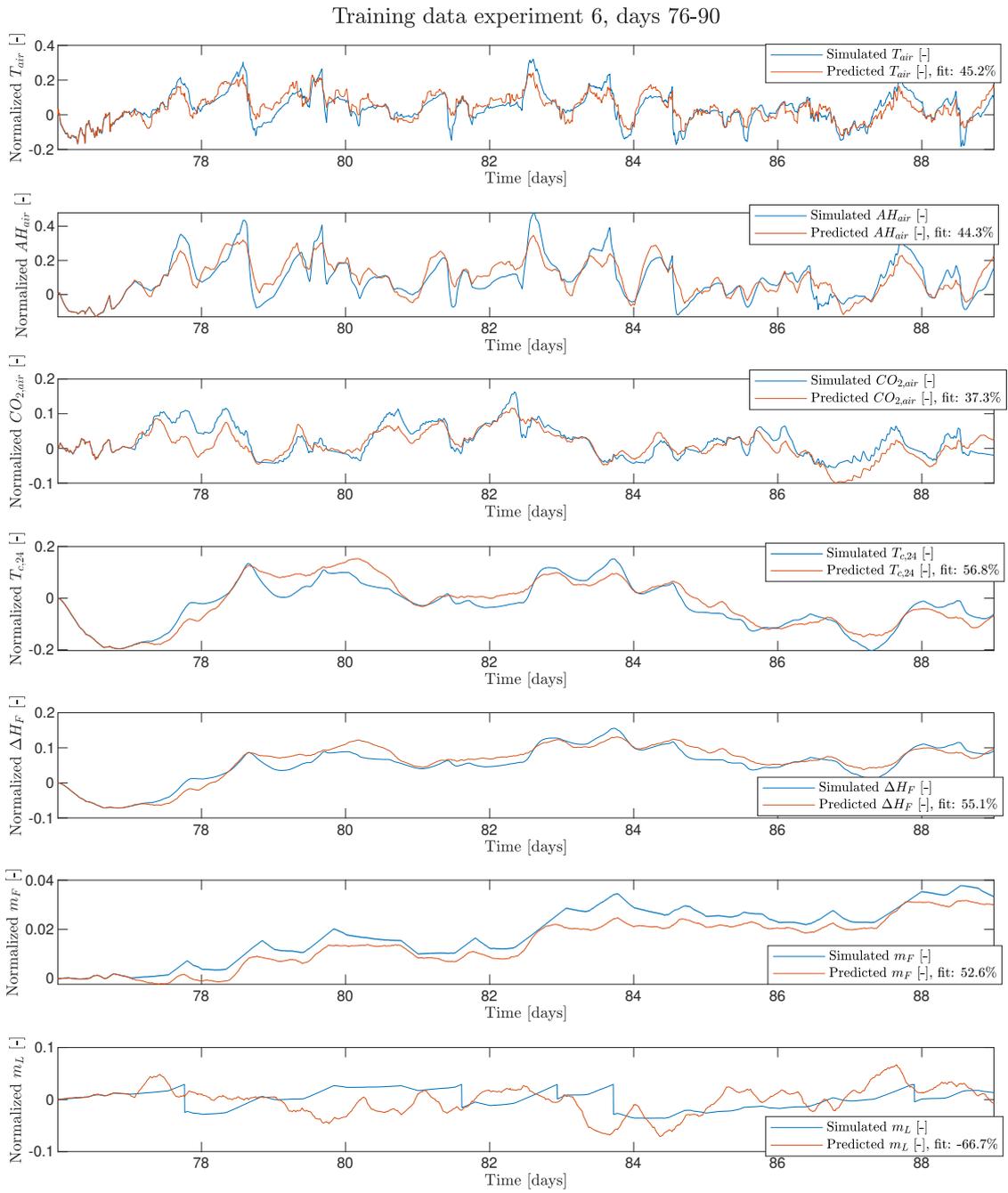
**Figure B-2:** Fit of identified linear state space system on the training dataset of a third experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.



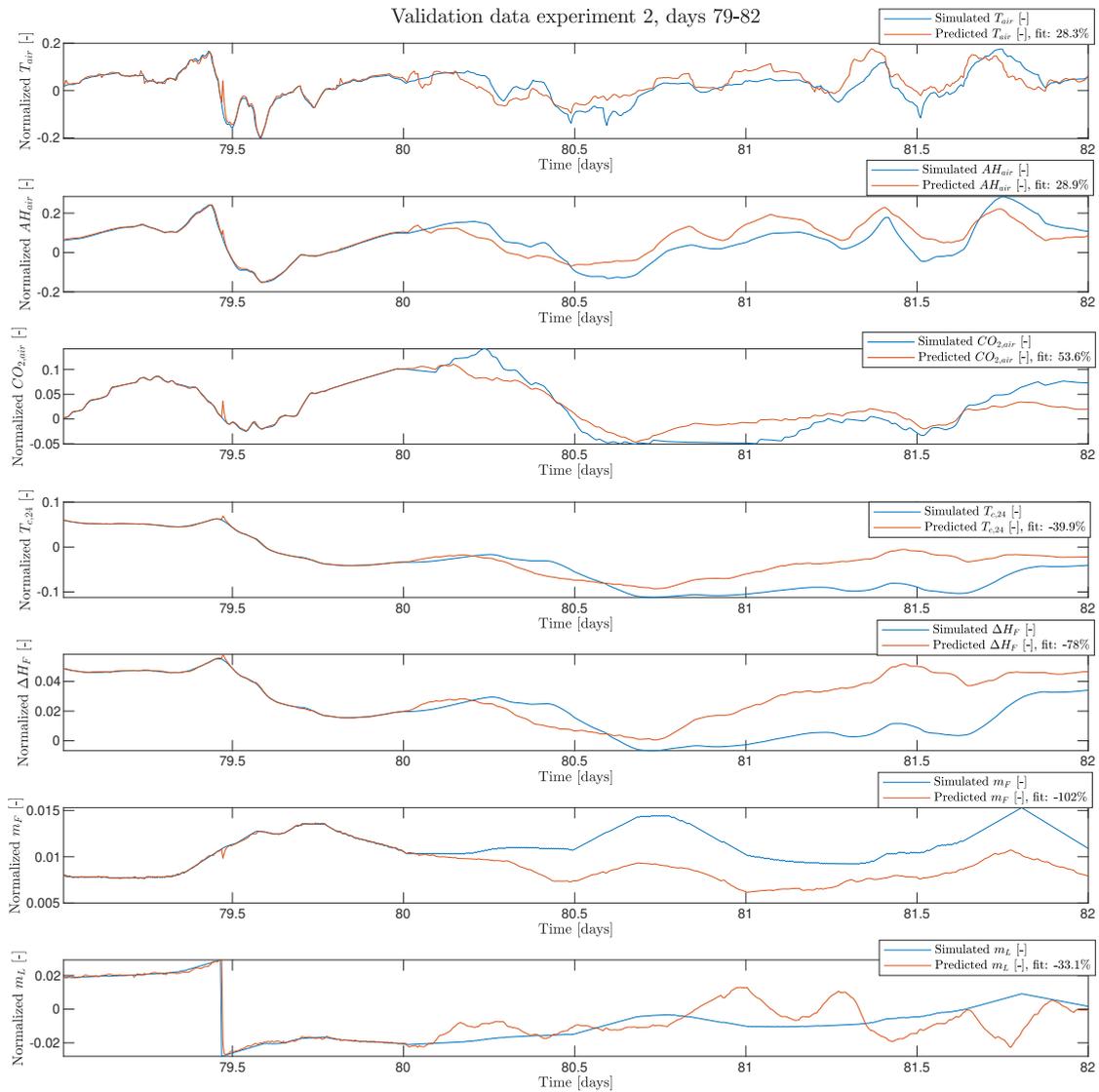
**Figure B-3:** Fit of identified linear state space system on the training dataset of a fourth experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.



**Figure B-4:** Fit of identified linear state space system on the training dataset of a fifth experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.

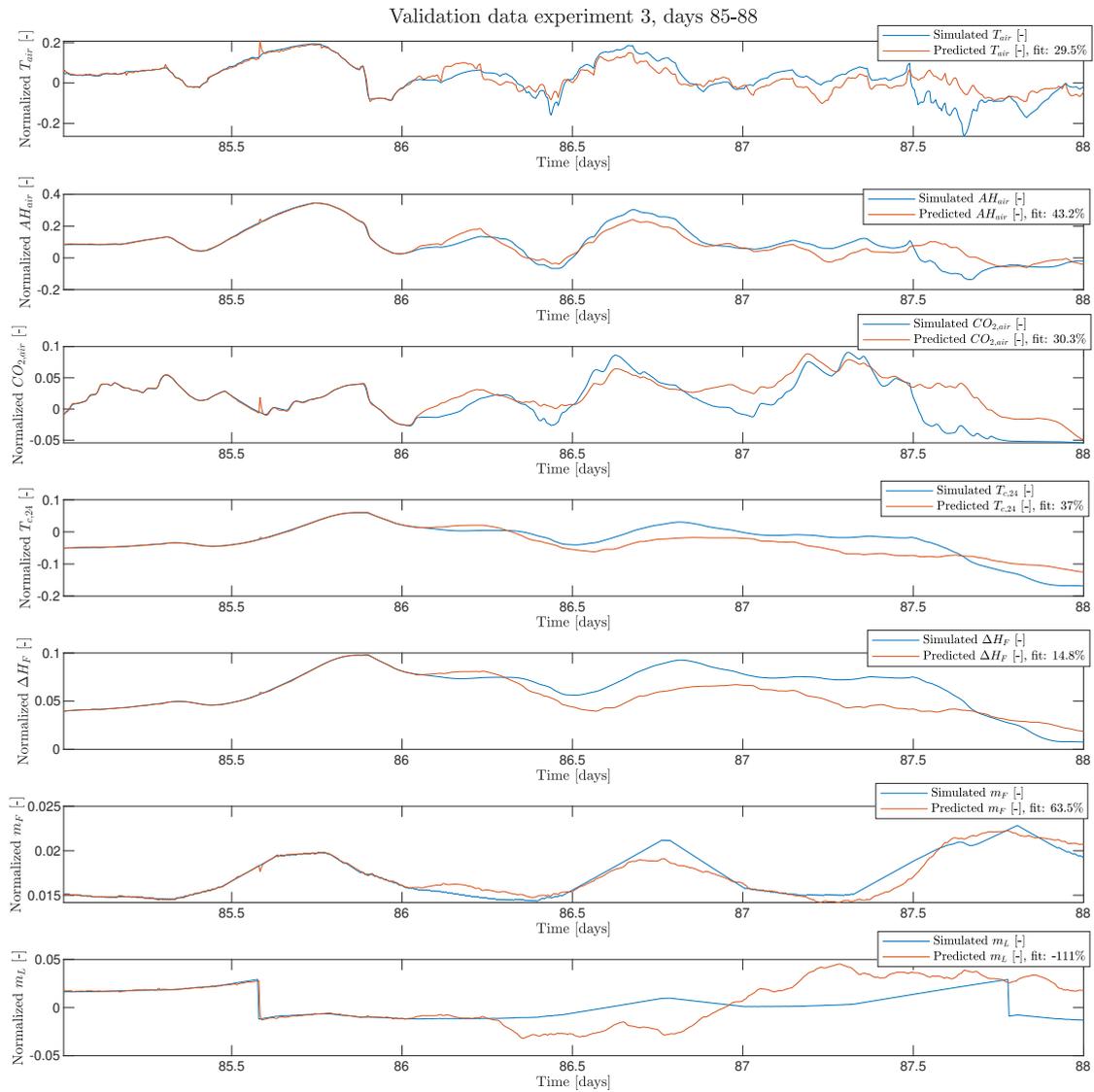


**Figure B-5:** Fit of identified linear state space system on the training dataset of a sixth experiment from day 76 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.

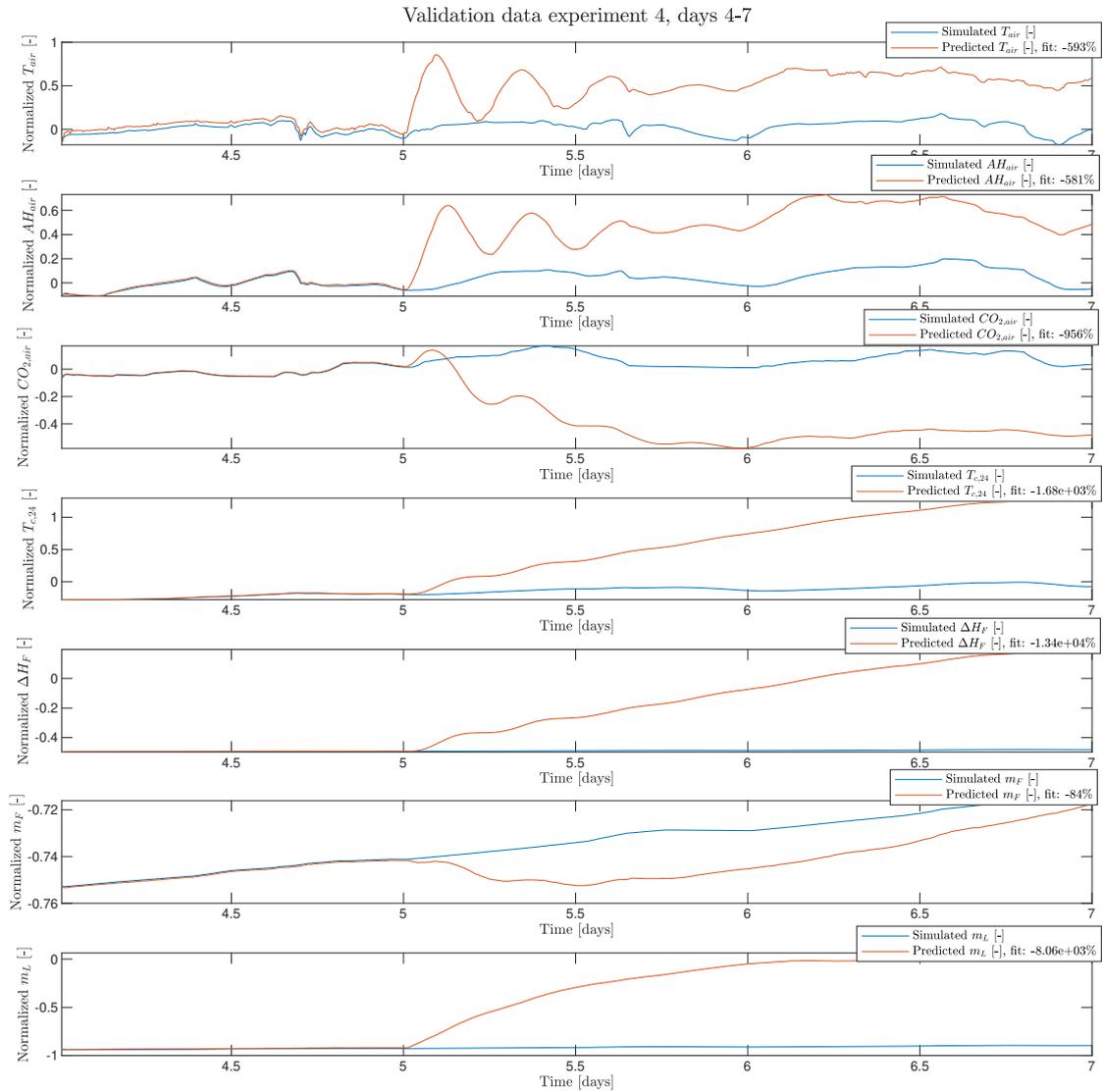


**Figure B-6:** Fit of identified linear state space system on the validation dataset from day 79 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.

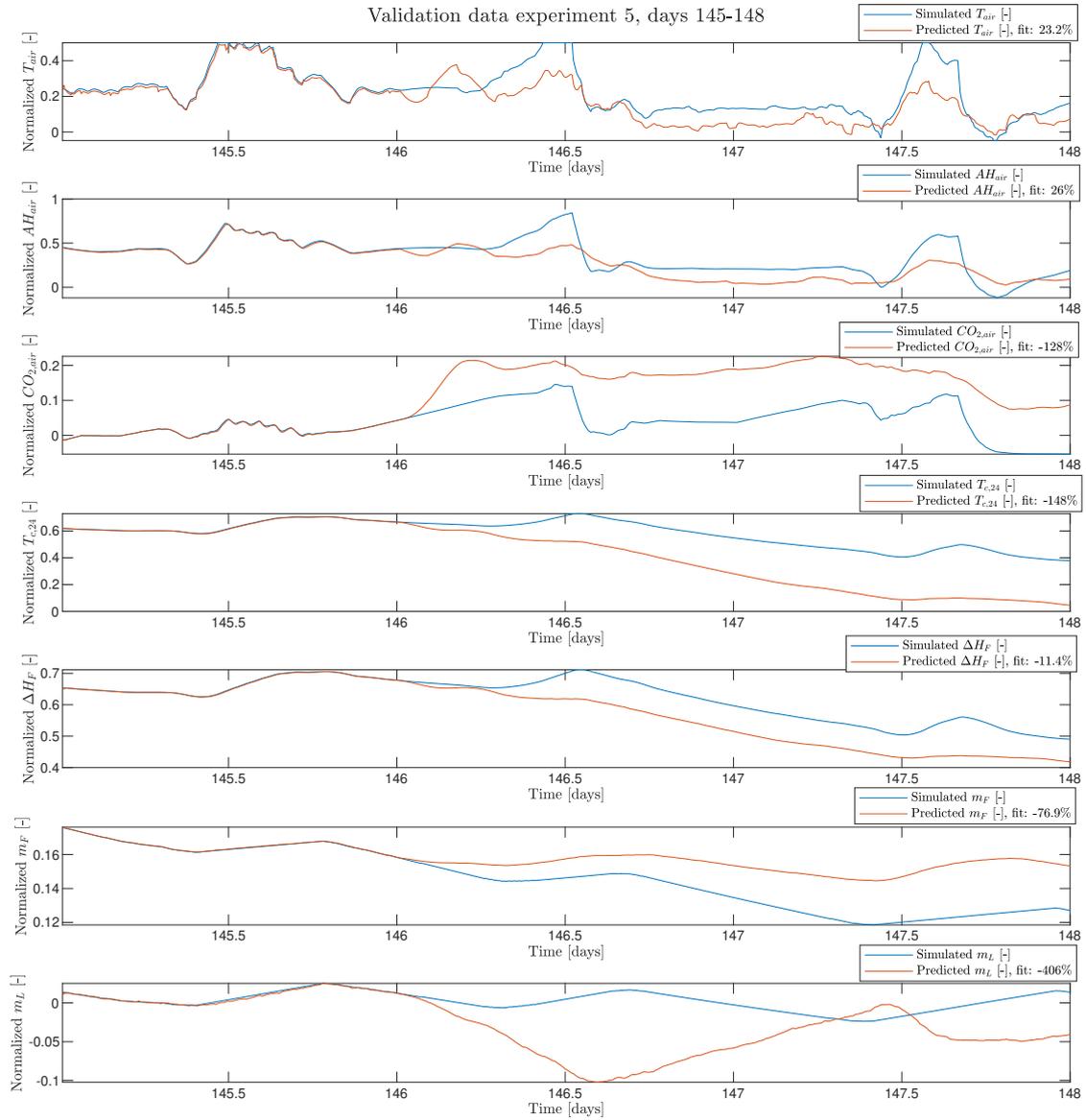
## B-2 Validation dataset fits



**Figure B-7:** Fit of identified linear state space system on the validation dataset from day 85 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.



**Figure B-8:** Fit of identified linear state space system on the validation dataset from day 4 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.



**Figure B-9:** Fit of identified linear state space system on the validation dataset from day 145 onward. The first day is solely used for estimating the internal state and is excluded from the calculation of the performance fit.

---

# Bibliography

- [1] G. van Straten, G. van Willigenburg, E. van Henten, and R. van Ooteghem, “Optimal control of greenhouse cultivation,” *Optimal Control of Greenhouse Cultivation*, pp. 1–297, 11 2010.
- [2] W. J. Kuijpers, M. J. van de Molengraft, S. van Mourik, A. van t Ooster, S. Hemming, and E. J. van Henten, “Model selection with a common structure: Tomato crop growth models,” *Biosystems Engineering*, vol. 187, pp. 247–257, 11 2019.
- [3] W. J. Kuijpers, D. Katzin, S. van Mourik, D. J. Antunes, S. Hemming, and M. J. van de Molengraft, “Lighting systems and strategies compared in an optimally controlled greenhouse,” *Biosystems Engineering*, vol. 202, pp. 195–216, 2 2021.
- [4] R. Rabbinge, H. H. van Laar, and S. A. Ward, “Simulation and systems management in crop protection,” *Agricultural Systems*, vol. 41, pp. 105–106, 1989.
- [5] A. Bemporad and M. Morari, “Robust model predictive control: A survey,” *Robustness in identification and control*, pp. 207–226, 10 1999.
- [6] S. Hemming, F. de Zwart, A. Elings, A. Petropoulou, and I. Righini, “Cherry tomato production in intelligent greenhousesensors and ai for control of climate, irrigation, crop yield, and quality,” *Sensors 2020*, vol. 20, p. 6430, 11 2020.
- [7] B. Vanthoor, “A model-based greenhouse design method,” 2011.
- [8] U. Nations, “2018 revision of world urbanization prospects.” <https://population.un.org/wup/>, 2018.
- [9] I. FAO, UNICEF, *et al.*, “The state of food security and nutrition in the world: building climate resilience for food security and nutrition,” *Food and agriculture Organization of the United Nations (FAO), Rome*, 2018.
- [10] M. G. M. Raaphorst and J. Benninga, “Kwantitatieve informatie voor de glastuinbouw 2019,” 2019.

- [11] G. V. Straten, H. Challa, and F. Buwalda, "Towards user accepted optimal control of greenhouse climate," *Computers and Electronics in Agriculture*, vol. 26, pp. 221–238, 5 2000.
- [12] I. Lopez-Cruz, E. Fitz-Rodríguez, S. Raquel, A. Rojano-Aguilar, and M. Kacira, "Development and analysis of dynamical mathematical models of greenhouse climate: A review," *European Journal of Horticultural Science*, vol. 83, pp. 269–279, 11 2018.
- [13] J. Verheul, "Glastuinbouw verduurzaamt niet snel genoeg," Nov 2019.
- [14] N. van der Velden, LEI Performance and Impact Agrosectors, and P. Smit, "Energiemonitor van de nederlandse glastuinbouw 2017," tech. rep., Wageningen, 2018.
- [15] "Climate policy." <https://www.government.nl/topics/climate-change/climate-policy>, Jan 2020.
- [16] Y. Ding, L. Wang, Y. Li, and D. Li, "Model predictive control and its application in agriculture: A review," *Computers and Electronics in Agriculture*, vol. 151, pp. 104–117, 8 2018.
- [17] L. G. J. Kerkhof, "Optimal control of autonomous greenhouses a data-driven approach," 2020.
- [18] M. van Duijn, "Hierarchical data-enabled predictive control: With application to greenhouse tomato crop production," 2021.
- [19] R. T. Rockafellar, "Lagrange multipliers and optimality," *Society for Industrial and Applied Mathematics*, vol. 35, 1993.
- [20] W. Kuijpers, "Model selection and optimal control design for automatic greenhouse climate control," 2021.
- [21] B. H. Vanthoor, C. Stanghellini, E. J. V. Henten, and P. H. D. Visser, "A methodology for model-based greenhouse design: Part 1, a greenhouse climate model for a broad range of designs and climates," *Biosystems Engineering*, vol. 110, pp. 363–377, 12 2011.
- [22] B. H. Vanthoor, P. H. de Visser, C. Stanghellini, and E. J. van Henten, "A methodology for model-based greenhouse design: Part 2, description and validation of a tomato yield model," *Biosystems Engineering*, vol. 110, pp. 378–395, 12 2011.
- [23] E. Iddio, L. Wang, Y. Thomas, G. McMorrow, and A. Denzer, "Energy efficient operation and modeling for greenhouses: A literature review," *Renewable and Sustainable Energy Reviews*, vol. 117, p. 109480, 2020.
- [24] P. J. van Beveren, J. Bontsema, G. van Straten, and E. J. van Henten, "Optimal control of greenhouse climate using minimal energy and grower defined bounds," *Applied Energy*, vol. 159, pp. 509–519, 12 2015.
- [25] P. V. Weels, P. Geelen, and J. Voogt, "Plant empowerment: The basic principles," *Plant Empowerment: The Basic Principles*, 2018.
- [26] F. d. Zwart, *Greenhouse climate control: In simulation and in reality*. 2021.

- [27] J. C. Roy, T. Boulard, C. Kittas, and S. Wang, “Paprecision agriculture: Convective and ventilation transfers in greenhouses, part 1: the greenhouse considered as a perfectly stirred tank,” *Biosystems Engineering*, vol. 83, pp. 1–20, 9 2002.
- [28] C. Gómez and C. A. Mitchell, “Supplemental lighting for greenhouse-grown tomatoes: Intracanopy led towers vs. overhead hps lamps,” *undefined*, vol. 1037, pp. 855–862, 2014.
- [29] J. Bontsema, E. J. V. Henten, T. H. Gieling, and G. L. Swinkels, “The effect of sensor errors on production and energy consumption in greenhouse horticulture,” *Computers and Electronics in Agriculture*, vol. 79, pp. 63–66, 10 2011.
- [30] X. Blasco, M. Martínez, J. M. Herrero, C. Ramos, and J. Sanchis, “Model-based predictive control of greenhouse climate for reducing energy and water consumption,” *undefined*, vol. 55, pp. 49–70, 1 2007.
- [31] A. K. Duun-Henriksen, S. Schmidt, R. M. Røge, J. B. Møller, K. Nørgaard, J. B. Jørgensen, and H. Madsen, “Model identification using stochastic differential equation grey-box models in diabetes,” *Journal of Diabetes Science and Technology*, vol. 7, pp. 431–440, 2013.
- [32] R. Linker and I. Seginer, “Greenhouse temperature modeling: a comparison between sigmoid neural networks and hybrid models,” *Mathematics and Computers in Simulation*, vol. 65, pp. 19–29, 4 2004.
- [33] L. M. Mortensen and F. Ringsevjen, “Semi-closed greenhouse photosynthesis measurements a future standard in intelligent climate control,” *European Journal of Horticultural Science*, vol. 85, pp. 219–225, 8 2020.
- [34] M. Afonso, H. Fonteijn, F. S. Fiorentin, D. Lensink, M. Mooij, N. Faber, G. Polder, and R. Wehrens, “Tomato fruit detection and counting in greenhouses using deep learning,” *Frontiers in Plant Science*, vol. 11, p. 1759, 11 2020.
- [35] J. C. Torres-Monsivais, I. L. López-Cruz, A. Ruíz-García, J. A. Ramírez-Arias, and R. D. Peña-Moreno, “Data assimilation to improve states estimation of a dynamic greenhouse tomatoes crop growth model,” *Acta Horticulturae*, vol. 1170, pp. 433–439, 7 2017.
- [36] F. Tap, “Economics-based optimal control of greenhouse tomato crop production,” 2000.
- [37] I. Seginer, C. Gary, and M. Tchamitchian, “Optimal temperature regimes for a greenhouse crop with a carbohydrate pool: A modelling study,” *Scientia Horticulturae*, vol. 60, pp. 55–80, 12 1994.
- [38] J. Jones, E. Dayan, L. Allen, H. van Keulen, and H. Challa, “A dynamic tomato growth and yield model (tomgro).,” *Transactions of the ASAE*, vol. 34, pp. 663–672, 1991.
- [39] E. Heuvelink, “Tomato growth and yield : quantitative analysis and synthesis,” 1996.
- [40] T. de Jong, “Natural ventilation of large multi-span greenhouses,” 1990.
- [41] C. Stanghellini and T. de Jong, “A model of humidity and its applications in a greenhouse,” *Agricultural and Forest Meteorology*, vol. 76, pp. 129–148, 9 1995.

- [42] Mathworks, “fillmissing version r2021b.”
- [43] Mathworks, “fmincon version r2021b.”
- [44] I. Seginer, “Optimizing greenhouse operation for best aerial environment,” in *Symposium on Computers in Greenhouse Climate Control 106*, pp. 169–178, 1979.
- [45] Mathworks, “Global optimization toolbox: particleswarm version r2021b.”
- [46] A. D. Koning, “Development and dry matter distribution in glasshouse tomato : a quantitative approach,” *undefined*, 1994.
- [47] J. Hoagg, S. Lacy, R. Erwin, and D. Bernstein, “First-order-hold sampling of positive real systems and subspace identification of positive real models,” vol. 1, pp. 861 – 866 vol.1, 01 2004.
- [48] L. Ljung, “System identification theory - for the user,” 1999.
- [49] R. Amrit, J. B. Rawlings, and D. Angeli, “Economic optimization using model predictive control with a terminal cost,” *Annual Reviews in Control*, vol. 35, pp. 178–186, 12 2011.
- [50] J. Coulson, J. Lygeros, and F. Dorfler, “Data-enabled predictive control: In the shallows of the deepc,” *2019 18th European Control Conference, ECC 2019*, pp. 307–312, 11 2018.
- [51] D. G. Luenberger, “Observing the state of a linear system,” *IEEE transactions on military electronics*, vol. 8, no. 2, pp. 74–80, 1964.
- [52] Gurobi Optimization, LLC, “Gurobi Optimizer Reference Manual,” 2022.
- [53] P. M. Van Den Hof and R. J. Schrama, “Identification and control closed-loop issues,” *Automatica*, vol. 31, no. 12, pp. 1751–1770, 1995. Trends in System Identification.

---

# Glossary

## List of Acronyms

**MPC** Model Predictive Control

**LAI** Leaf Area Index

**AGC** Autonomous Greenhouse Challenge

**LED** Light Emitting Diode

**HPS** High Pressure Sodium

**ODE** Ordinary Differential Equation

**LTI** Linear Time Invariant