Stomatal Conductance Patterns and Advanced Automatic Control

With Application to Small Scale Agriculture

Ioannis Panagopoulos

Master of Science Thesis

TU Delft Delft University of Technology

Stomatal Conductance Patterns and Advanced Automatic Control

With Application to Small Scale Agriculture

Master of Science Thesis

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

Ioannis Panagopoulos

April 11, 2023

Faculty of Electrical Engineering, Mathematics & Computer Science (EEMCS)
&
Faculty of Mechanical, Maritime and Materials Engineering (3mE)
Delft University of Technology

ABSTRACT

Humanitarian, environmental, and political concerns have contributed to the evolution of agricultural technology, also known as AgTech. Researchers from various scientific backgrounds are diving into AgTech to ensure the world's food security, create a sustainable future for agriculture, and pave the way for autonomous cultivation methods. This thesis project attempts to contribute to the aforementioned subjects from the control engineering perspective. The thesis objective is the design and testing of a novel predictive climate controller, for tomato greenhouses agriculture, able to conclude on the optimal yield-energy consumption ratio with limited intervention by the human factor. The main novelty introduced by this algorithm is the use of crop variables in the decision-making process according to the Speaking Plant Approach (SPA). However, no straightforward recipe indicates which crop signals could be used. In the context of this study, it is explored how crop variables, measurable by thermal imaging, can be used for the formulation of a SPA-based objective function. Specifically, the research focuses on stomatal conductance, the canopy, and the mean canopy temperature, for the SPA-based objective function formulation. The cost function generation entails the definition of the necessary state constraints. Except for the objective function definitions and the determination of constraints, a predictive controller requires a system representation that acts as a predictor. Nevertheless, the complex and non-linear nature of the climate-crop system complicates the system identification process. Concurrently, data science is blooming and new data-driven system representation techniques are breaching. Data-Enabled Predictive Control is a novel control policy based on systems behavioral theory which uses a non-parametric system representation enabling the omission of the system identification process. This approach has not been tested for the description of highly complex climate-crop systems. Therefore, another target of this project is to examine the capabilities of this data-driven predictor for the representation of the climate-crop model and evaluate if and how can be used as a predictor in a climate control regime.

PREFACE

This thesis project emerged from my personal fascination with control and agriculture, which ultimately led me to delve into the field of plant physiology. What began as an experiment for my 3D-printed greenhouse concluded in something bigger broadening my horizons. I'm thoroughly grateful to all those who were by my side during the whole process.

I extend my heartfelt appreciation to my TUD supervisor, Prof.dr.ir.Keviczky, for his continued guidance and support regarding all control-oriented challenges I encountered. I also extend my gratitude to Javier Lomas from Sigrow, for his invaluable input and guidance whenever I faced plant-physiology issues. Furthermore, I extend my thanks to Dr. Yorke-Smith for reviewing my work as a member of my thesis committee. I am indebted to my family for their love and support, without which the realization of this master's degree would not have been possible. I am grateful to all these beautiful people that helped me keep my sanity and stay motivated during the challenging moments of the research process. Lastly, I cannot thank enough my library cohabitants for the lovely study breaks, the rap sessions, and the drama that they added to my life during the last months of this thesis project.

Delft, University of Technology
March 29, 2023

Ioannis Panagopoulos

CONTENTS

LIST OF FIGURES

List of Tables

1

INTRODUCTION

1.1. MOTIVATION

This thesis project dives into the design and implementation of a novel climate controller exploiting crop-generated signals. From a high-level perspective, the project is under the umbrella of agricultural technology, also known as AgTech. This encompasses a wide range of technologies, including precision agriculture, smart irrigation systems, autonomous tractors and drones, sensors and monitoring systems, and more. The motives behind the occupation with AgTech are multidimensional.

People living in industrialized countries have faced a significant increase in their quality of life during the 21^{st} century. However, the standardization of people's better living standards in combination with the constantly increasing world population threatens the preservation of food security [1]. Furthermore, the number of experienced growers around the globe is declining rising the question of who is going to carry the required food production [2]. From a humanitarian and social point of view, engaging with AgTech increases food availability and ensures food security.

Environmental concerns are considered a motive of great importance as people and governments try to decelerate climate change. The Dutch horticultural industry can be considered a case in point. Dutch farmers achieve high yield production making the Netherlands one of the biggest exporters of vegetables in the world. But high yield comes with the cost of high fossil fuel consumption. To assure the continuous decrease of its environmental footprint, the Dutch horticultural industry signed an agreement with the Dutch government to reduce fossil fuel consumption and introduce innovations to improve the energy efficiency of cultivation [3].

Political concerns can also demonstrate the importance of AgTech. The Dutch horticulture sector consumed 9% of the total natural gas used in the Netherlands, corresponding to 3.15 billion cubic meters of gas for 2021. This amount of energy consumption

may have not been considered extravagant a few years ago. Nevertheless, the COVID-19 pandemic followed by major political events in eastern Europe has triggered a series of economic events that have led to the current energy crisis, causing extreme volatility in the prices of the European natural gas market.

From a scientific perspective, agriculture technology is interesting on its own. It is a multidisciplinary field where engineers and scientists cooperate for the exploration of new strategies exploiting today's technological equipment. Concerning control engineering, AgTech offers challenges and open problems in the fields of climate control, and climate-crop system representation. Novel climate control recipes are means able to increase yield production and reduce energy consumption even more. Moreover, the development of climate-control methods falling under the ample range of Artificial Intelligence (AI) can steer crop production in an autonomous direction. On the other hand, representative climate-crop models have great importance as they can be used to gain insight into the system, act as ground truth for the simulation of experiments, or become parts of predictive climate controllers. However, climate-crop modeling is not trivial as the complex and non-linear nature of the system complicates the system identification process. For these reasons, researchers have been charged to explore model-based and data-based methods for the development of mechanisms able to represent the climate-crop system behavior.

To conclude, occupation with AgTech and specifically with control engineering applications has several motives. First of all, food production has to increase for humanitarian reasons. Secondly, environmental and political motives indicate the imperative need to reduce the energy requirements of the agricultural sector. Finally, scientific curiosity combined with a temper for technological research can be the driving force for the exploration of new ideas and applications in the agricultural sector.

1.2. THESIS CONTRIBUTION

The Modern Agricultural industry faces three main issues. The first one is the increased demand for high crop yields able to ensure the world's food security. The second one is the necessity to decrease energy consumption to guarantee the sustainability of this industrial sector. Last but not least, the agricultural sector becomes understaffed over time, thus the transition from conventional agriculture to autonomous agriculture is unavoidable in order to insure its viability. This thesis project aims to contribute to the solution of the aforementioned problems from the control engineering perspective. The thesis objective is the design and testing of a novel predictive climate controller able to conclude on the optimal yield-energy consumption ratio with limited intervention by the human factor. Particularly, the design of such a controller translates to the definition of an optimization criterion, the denotation of constraints regarding the control inputs, the system's states, and outputs, and the implementation of the climate-crop system representation.

One important problem in autonomous agriculture is the definition of optimal environmental conditions able to achieve the best yield-cost ratio. A common approach is to

apply a set of reference values or trajectories for the day and night time accordingly. This approach is a non-optimal empirical method founding its efficiency on the grower's experience. Because of its empirical nature, this method includes a great issue in terms of autonomy, it considers the grower as an inextricable part of the climate control loop. Another option incorporates the use of an optimization-based approach. In this case, the generation of the environmental condition setpoints requires a yield representation where through the formulation, the calibration, and the solution of an optimization problem, the computer concludes on a set of optimal environmental conditions. A challenge of this approach is that, in some cases, yield data cannot get measured from day 0, complicating yield modeling even more. For example, in tomato crops, fruit production starts after the growing and blooming phase of the plant. Additionally, yield is hard and time-consuming to be measured on a live plant as it requires labor. Therefore, the yield model should be accurate enough as the climate controller would not be able to receive feedback on prediction errors.

Speaking Plant Approach (SPA) is a concept that could contribute to the solution of the aforementioned issues. SPA proposes that the optimal environmental conditions should not be indicated by yield increase but by the plant itself via a single or a set of signals correlated with the crop yield [4]. A great advantage of SPA is that the crop becomes a part of the control loop limiting the grower's role. Additionally, SPA includes feedback capabilities as the control system can measure the crop index and examine if the plant behaved as expected. In this way, the error measurements can be used to adjust the applied control policy. However, SPA has some major issues concerning applicability. The mechanisms ruling the yield production are not completely modeled yet and there is no straightforward recipe indicating which crop signals could be used. Another issue is data sensing, as the data should be measured in a reliable, fast, and cheap way. In terms of reliability, the measured index should be able to be measured with good accuracy to reduce the prediction errors occurring from measurement errors. The measurement procedure should be as fast and automated as possible, so it can get repeated multiple times per day or growing season. Finally, both the operational and the installation costs of the sensors have to be as low as possible, to avoid the introduction of extra costs in the growing process. A means able to satisfy the SPA applicability, reliability, and financial requirements are infrared thermal cameras. Thermal imaging is a non-invasive, non-contact, and non-destructive method to determine the temperature distribution of any surface of interest in a short time period. Moreover, it has great potential for agricultural applications as it has already been used for the prediction of water stress in crops, irrigation scheduling, disease and pathogen detection, bruise detection, and stomatal conductance estimation [5].

The contribution of this thesis project, regarding autonomous agriculture, is the definition of a SPA-based cost function, exploiting thermal imaging, able to be used in a predictive climate controller and limit the grower's role in the growing area. Stomatal conductance is a good candidate for this application as it can be measured through canopy temperature [6] and in most cases, it has a positive correlation with photosynthesis [7]. Furthermore, stomatal conductance is a signal suitable for an online control policy as it

can be measured right after the crop's germination. Finally, the thesis explores the introduction of crop biological processes such as respiration and assimilates partitioning as parts of the optimization criterion.

Historically, the field of Systems and Control has contributed to AgTech through the introduction of various predictive climate control policies able to treat Multiple Input Multiple Output (MIMO) systems like the climate-crop system. Predictive control policies could be categorized as model-based and data-based policies. Model-based policies such as Model Predictive Control (MPC) have proven their value in agricultural applications [8]. However, the complexity of crop systems complicates the system identification process and does not let optimal control become a standard in climate control. At the same time, data science is blooming and as a result, researchers' interest has been steered toward data-driven system representation techniques. Data-enabled predictive control (DeePC) is a novel control policy based on systems behavioral theory able to be used instead of MPC to solve an optimal control receding horizon problem. The advantage introduced by DeePC and its variants (See [9], [10]) is the use of a non-parametric system representation enabling the omission of the system identification process [11]. This non-parametric system representation has proven its value in the representation of Linear Time-Invariant (LTI) systems. Moreover, in some cases, this data-driven predictor copes with the representation of non-linear and time-variant systems, despite its linear nature [9]. However, it has not been tested for the representation of highly complex climate-crop systems. This thesis project contributes to the exploration of the method's potential in the context of AgTech. Specifically, it is tested if such a data-driven predictor is capable to be used for the representation of the non-linear climate-crop system presented in [12]. From a high-level perspective, the universal use of optimal predictive climate control could permit better yield-cost ratios assisting the insurance of food security and limiting energy consumption.

At this point, it should be noted that the testing of the algorithm takes place in a simulated greenhouse environment. Greenhouses are exceptional means to create controlled environmental conditions for crops growing far from their geographical origin or for off-season grows. Specifically, greenhouses permit control of the air temperature, air humidity, concentration of carbon dioxide, and lighting conditions. In the context of this thesis project, testing the control algorithm in a greenhouse environment has a dual meaning. Greenhouses are highly related to modern agriculture in the Netherlands and northern Europe in general. Therefore, the simulation of a greenhouse environment was chosen to ensure that the tested control algorithm can be easily transferred to a real-world industrial scenario, despite the algorithm's academic nature. Furthermore, the applicability of methods designed for small-scale agriculture is not always guaranteed in large-scale agricultural environments such as greenhouses. Thus, validating the control algorithm in a large-scale environment ensures its industrial potential.

To conclude, modern agriculture needs to ensure food security, minimize fossil fuel consumption and surpass the obstacles created by the diminishing number of growers. From the control engineering point of view, these objectives can be guaranteed by establishing

optimal climate control methods and contributing to the transition from conventional to autonomous agriculture. Regarding autonomous agriculture, the thesis participates in the transition process through the development and testing of a SPA-based objective function for climate control. This thesis project contributes to optimal predictive climate control by examining if the data-driven system representation, originally presented in [11], is capable to describe the behavior of a complex climate-crop system.

1.3. RESEARCH QUESTIONS

This thesis explores the design of a SPA-based predictive control algorithm. Moreover, the project aims to examine the capabilities of a linear data-driven predictor regarding the representation of a greenhouse system. Hence, the following research questions are defined:

Is the linear data-driven predictor, presented in [11], able to describe the behavior of the complex climate-crop system presented in [12] without persistently exciting training data?

How does the designed SPA-based climate control algorithm perform compared to a conventional rule-based climate controller and a yield maximization NMPC algorithm?

To answer this main research question, a set of sub-questions must be answered. The sub-questions are formulated as follows:

- *What greenhouse, climate, and crop system models should be selected for the purposes of ground truth, disturbance description, model-based, and/or data-based predictive controller design?*
- *Which of the available actuators are going to be controlled and why?*
- *Which crop processes should be incorporated in the SPA-based objective function?*
- *What modifications does the SPA-based objective function need in order to fit in the NMPC greenhouse control scheme?*
- *Which are the optimal performance metrics to become benchmarks for the intended comparisons?*

During the design process of the proposed SPA-based control approach, applicability was considered. Thus, it would be interesting to answer the following question:

- *Under which assumptions could the proposed approach be considered an applicable real-world application?*

1.4. THESIS OUTLINE

This thesis project is structured as follows: Chapter 2 introduces the reader to the concept of predictive control by highlighting its main components and demonstrating an overview of some predictive control variants. Chapter 3 presents the principles of a

greenhouse system and the climate-crop model that has been chosen as the ground truth for this project. The main focus of Chapter 4 is to present the research carried out in developing and testing a SPA-based objective function for predictive climate control techniques. Chapter 5 explores if and how a data-driven linear representation method is able to sufficiently represent the behavior of the mechanistic climate-crop model presented in Chapter 3. The necessary mathematical concepts and conditions governing the design of such a data-driven predictor are demonstrated. Additionally, Chapter 5 presents the results from the attempted data-driven representation of the mechanistic greenhouse model. In Chapter 6 a case study of three different climate control regimes takes place. Each controller is demonstrated and then all climate controllers are tested on a simulated greenhouse. Finally, the simulated results are compared. Last, answers to the research questions and recommendations for future work are given in Chapter 7.

2

PRELIMINARIES ON PREDICTIVE CONTROL

The main objective of this thesis project is the development and testing of a novel predictive climate control approach. The following chapters present the procedure behind the design of such a controller, thus the high-level presentation of the fundamental components of predictive control has great importance to ensure the proper understanding of the upcoming chapters. Therefore, Chapter 2 aims to introduce the reader to the essentials of predictive control and its variants.

2.1. INTRODUCTION TO PREDICTIVE CONTROL

Predictive control is a very general term expressing the control problem in a meaningful manner. As a term, it does not designate a specific control strategy but a substantial range of control methods. However, all these methods consist of some common characteristics and components. Precisely, predictive control uses a representation of the process to predict the process output at future time instants and obtain the control signal by minimizing an objective function [13]. It should be noted that different types of constraints can be incorporated into the control approach. Moreover, the number of future time instants is named the prediction horizon and it is usually noted as N. It should be underlined, that the finite horizon formulation complicates the system theoretic guarantees related to stability and robustness.
To sum up, the main components of predictive control are a cost function acting as an optimization criterion, a set of constraints regarding the control inputs, the system's states, and outputs, and a system representation.

2.1.1. OBJECTIVE FUNCTION

From the mathematical perspective, the objective or cost function is the function whose value is to be minimized or maximized, in the context of an optimization problem, over a set of feasible alternatives. Practically, the objective function is responsible to return

a scalar as a function of the system's states and inputs. Its formulation has great importance as it defines the goal of the predictive controller. The exact formulation of the objective function is dependent on the control engineer, however, Equation (2.1) displays a mathematical formation regularly presented in the literature.

$$J_N(x,u) := \sum_{k=0}^{N-1} l(x(k),u(k)) + V_f(x(N)) \tag{2.1}$$

Where $l : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ the stage cost function specifying the considered performance criterion, and $V_f : \mathbb{X}_f \to \mathbb{R}$ the terminal cost function, usually used to guarantee closed-loop stability. It is important to note that the terminal cost is often omitted.

2.1.2. CONSTRAINTS

Both model-based and data-based predictive control approaches can embed different types of constraints bounding the space where the optimal future control trajectories live. The constraints can be either soft meaning that their slight violation is acceptable, or hard where constraint violation leads to an infeasible solution. In practice, all sensors and actuators underlie some constraints. Usually, these constraints are generated because of safety, environmental, economic, or just physical reasons. For example, when tomato crops experience canopy temperatures below 12 ^oC destructive results occur, such as fruit's excessive softening, electrolyte leakage, and failure to ripen [14]. Typically, constraints regard the control action amplitude, the slew rate of the control signal, and limits on the system's states and output. The mathematical formulation of the commonly used constraints is given below:

$$\begin{aligned} u_l \le u(t) \le u_u,\ \forall t \\ \Delta u_l \le u(t) - u(t-1) \le \Delta u_u,\ \forall t \\ x_l \le x(t) \le x_u,\ \forall t \\ y_l \le y(t) \le y_u,\ \forall t \end{aligned} \tag{2.2}$$

Where subscripts l and u declare the lower and upper limit respectively.

2.1.3. SYSTEM REPRESENTATION

Predictive control needs a dynamical representation of the open-loop process to explicate the relationships between the system's inputs, measured outputs, and internal states. Ideally, the used representation should fully capture the process dynamics, allow the calculation of the output at future instants, and permit the system's theoretic analysis [13]. A great advantage of predictive control in general is that it can handle both Single Input Single Output (SISO) and Multiple Input Multiple Output (MIMO) processes. Moreover, processes with either simple or complex dynamics, including unstable systems, systems with long delays, or non-minimum phase can also be handled from a predictive control algorithm. The system's representation can be either parametric or non-parametric. Any form of process representation, either model-based or data-based such as state space, transfer function, impulse response, step response, neural networks, or fuzzy logic can be used for the construction of a predictive controller.

2.2. PREDICTIVE CONTROL VARIANTS

Model-Predictive Control (MPC) has its origins in the late seventies and is one of the most popular predictive control approaches in academia and industry. For these reasons, it was chosen to be used as the guiding example for the overview of the predictive control variants.

2.2.1. NOMINAL MPC

To introduce the general MPC formulation, the following state-space system representation will be considered.

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \tag{2.3}$$

Where $x(t) \in \mathbb{R}^{n_x}$ the system states, $u(t) \in \mathbb{R}^{n_u}$ the control inputs and $y(t) \in \mathbb{R}^{n_y}$ the system's outputs at time $t \in \mathbb{Z}_{\geq 0}$. $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, $C \in \mathbb{R}^{n_y \times n_x}$ and $D \in \mathbb{R}^{n_y \times n_u}$ the system matrices. At this point, it should be reminded that any form of modeling a process is capable of concluding on an MPC formulation. Although, it was chosen to use the state-space representation as it permits the use of multivariable processes in a straightforward manner [13].

The open-loop MPC optimization problem is given below:

$$\begin{aligned} \min_{u(t)} \quad & J_N(x,u) \\ \text{s.t.} \quad & x(k+1) = Ax(k) + Bu(k),\ k \in \{0,\dots,N-1\},, \\ & y(k) = Cx(k) + Du(k),\ k \in [0,\dots,N-1], \\ & x(0) = x(t), \\ & x(k) \in \mathbb{X},\ u(k) \in \mathbb{U},\ k \in \{0,\dots,N-1\}, \end{aligned} \tag{2.4}$$

Where J_N is the objective function to be minimized over the prediction horizon N. $\mathbb{X}$, $\mathbb{U}$ are the state and input constraint sets, respectively. t is the time at which the optimization is solved and k is the sampling instant. As mentioned, the controller is on an open-loop configuration. In general, nominal MPC is based on the assumption that the mismatch between the prediction model and the plant process is neglected as it is considered sufficiently small [15].

Algorithm 1 Model Predictive Control Algorithm

Input: Dynamic Model, Constraints and Costs
Output: Optimal Sequence of Future Control Moves
(1) Measure or estimate the current state of the system
(2) Solve the optimization problem 2.4 for the optimal input sequence
$u^* = (u_0^*, \dots, u_{N-1}^*)$
(3) Apply the first N_c control inputs
(4) Set t to $t + N_c$
(5) Return to 1

As shown in Algorithm 1, the output of the optimization problem is the optimal input sequence $u^* = (u_0^*, \ldots, u_{N-1}^*)$ as in fact the optimal control problem is formulated over a time horizon N that starts at the current time t and ends at time $t+N-1$. However, only the first N_c samples of the optimal input sequence are applied to the system, the rest are discarded. As N_c the control horizon is defined where $1 \leq N_c \leq N$. Then, at time $t+N_c$ a new optimization problem is solved over a shifted prediction horizon N. As a result of the aforementioned characteristic, literature usually refers to MPC control as *Receding-Horizon Control* [16].

PREDICTION HORIZON

The selection of the prediction horizon has a major role in MPC. When the control and the maximum prediction horizons approach infinity and no constraints are applied, the MPC problem becomes equivalent to the Linear Quadratic Regulator (LQR) problem [13]. In the case of a perfectly modeled system, where the exact initial conditions are known and the control and prediction horizons are infinite, the MPC approach can be implemented in an open-loop. However, this is not possible in general, because of disturbances and model-plant mismatches, the behavior of the real system differentiates from the predicted behavior [17]. In the case of the greenhouse control problem, wrong weather predictions act as disturbances. Further, model-plant mismatches occur due to the "stirred tank" assumption ruling the process-based greenhouse climate models and the lack of understanding of the complex biochemical crop processes. Thus, incorporating a feedback mechanism combined with the receding horizon approach frequently appears in greenhouse optimal control applications.

2.2.2. REFERENCE TRACKING MPC

Reference tracking MPC is a method suitable in cases where the future output y should follow a determined reference signal r with the minimum control effort. In our case, reference tracking MPC can be used to make greenhouse environmental conditions follow the desired trajectory. As an optimal control method, a properly set Tracking Model Predictive Controller can outperform conventional PID-type climate controllers under realistic external conditions [18]. Commonly, trajectory tracking MPC is based on an objective function of a quadratic form. The objective function consists of three main terms as shown in the equation below:

$$J(N) = \sum_{k=0}^{N-1} \left(||\hat{y}(k) - r(k)||_Q + ||u(k)||_{R_1} + ||\Delta u(k)||_{R_2} \right) \tag{2.5}$$

Where N the prediction horizon, Q, R_1, and R_2 the quadratic positive semi-definite tracking, control input and slew rate cost matrices accordingly.

On the right-hand side of Equation (2.5), the first term represents the minimization of the error between the predicted output $\hat{y}$ and the desired reference trajectory r. The second term denotes the minimization of the control inputs cost and the third represents

the minimization of the control input's slew rate. All terms can be penalized to tune the controller to behave according to the designer's requirements.

2.2.3. ECONOMIC MODEL PREDICTIVE CONTROL

Economic MPC (EMPC) is the integration of the economic optimization criterion in a receding horizon policy [19]. In the case of EMPC, the optimizer attempts to find the minimum actuator's cost that will lead to the maximization of economic revenue. A general formulation of the EMPC objective function is given below:

$$J_{eco} = ||q(\boldsymbol{u})|| - ||\boldsymbol{y}|| \tag{2.6}$$

Where $q(\boldsymbol{u})$ is an expression of the control input cost and $\boldsymbol{y}$ is an expression of the economic revenue related to the process outputs. The exact determination of the aforementioned terms is application-oriented and left to be a designer's option.

2.2.4. NON-LINEAR MODEL PREDICTIVE CONTROL

The climate-crop system is a complex, non-linear, time-varying process hence the climate controller should be able to treat these system characteristics. Based on that, it is considered necessary to introduce the concept of Non-Linear Model Predictive Control (NMPC). The non-linear, time-varying system is represented as:

$$\begin{cases} x(t+1) = f(x(t), u(t), d(t), t) \\ y(t) = h(x(t), u(t), d(t), t) \end{cases} \tag{2.7}$$

Where $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \times \mathbb{R} \to \mathbb{R}^{n_x}$ the nonlinear function mapping the current state and input to the next state under exogenous inputs, $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \times \mathbb{R} \to \mathbb{R}^{n_y}$ the nonlinear function mapping the current state, control and exogenous input to the current output, $x(t) \in \mathbb{R}^{n_x}$ the state vector, $u(t) \in \mathbb{R}^{n_u}$ the control input vector, $y(t) \in \mathbb{R}^{n_y}$ the output vector, $d(t) \in \mathbb{R}^{n_d}$ the vector of exogenous inputs, at time $t \in \mathbb{Z}_{\geq 0}$.

The open-loop optimization problem for the non-linear predictive controller is constructed as follows:

$$\begin{aligned} \min_{u(t)} \quad & J_N(x, u, d) \\ \text{s.t.} \quad & x(k+1) = f(x(k), u(k), d(k), k), \ k \in \{0, \dots, N-1\}, \\ & x(0) = x(t), \\ & x(k) \in \mathbb{X}, \ u(t) \in \mathbb{U}, \ k \in \{0, \dots, N-1\}, \\ & x(N) \in \mathbb{X}_N \end{aligned} \tag{2.8}$$

It becomes understandable that the extension of MPC to NMPC is straightforward. However, dealing with non-linearities is not a trivial case. Before anything else, the lack of identification techniques for non-linear processes complicates the generation of the system representation through data. Furthermore, modeling the non-linear system through first principles is not always feasible. In terms of optimization, the problem becomes

non-convex, as a result, it is not sure that a global minimum exists. Moreover, the solution of the non-linear optimization problem is computationally expensive in comparison with QP problems and finally, studying the stability and robustness of a non-linear system is an open research field [13].

3

THE GROUND TRUTH GREENHOUSE-CROP SYSTEM

3.1. INTRODUCTION

In the context of this thesis project, a climate-crop model is required in order to be used in various ways. First, some of its submodels will be used for the design and validation of the proposed objective function (See Chapter 4). Second, it will be exploited for the generation of the training and validation data needed in the implementation of the proposed data-driven representation. Afterward, it will be used as a baseline for the performance evaluation of the data-driven predictor (See Chapter 5). Moreover, it will be used as a testing platform for the comparison of the results that occurred by the proposed SPA-based climate controller (See Chapter 6). Therefore, this climate-crop model can be regarded as ground truth.

Numerous greenhouse-crop models have been developed to describe the behavior of the micro and macro climate in a greenhouse and their interaction with the crop. However, the present thesis project requires a greenhouse-crop model with specific characteristics. First, the pursued model should be able to describe the climate's thermodynamics, vapor pressure, carbon dioxide concentration, and artificial lighting. Further, crops grow by increasing their canopy surfaces and occupying more volume in the greenhouse. Canopy's heat capacity increases as the crop's volume increase, affecting heat exchanges in the greenhouse. Besides, increased crop volume results in increased transpiration and respiration, affecting carbon dioxide and vapor mass balances. As a result, climate dynamics change as crop evolves. Hence, the required greenhouse climate models have to take into account the crop's state. Air temperature has a major role in climate control and the climate controller should be able to allow both high and low extreme temperatures to efficiently manage energy resources [20]. Thus, the model should describe the effects of extremely high and low temperatures efficiently. Another requirement is the mechanistic nature of the model. In the content of this study, a process-based model will be used

as ground truth for the simulation of optimization-based experiments. The necessity for a mechanistic model arises from the need for a proper physical interpretation of the results. This set of requirements is satisfied by the Greenlight model presented in [12]. The Greenlight model is an exploratory, validated, process-based greenhouse model introduced as a modified version of the well-known KASPRO [21] and Vanthoor-Stangellini [22] models.

Concluding, this chapter encloses the presentation of the greenhouse system principles. An overview of the Greenlight climate-crop model which is used in this project and finally, a presentation of the assumptions that are introduced by the selected models.

3.2. GREENHOUSE SYSTEM PRINCIPLES

Despite the different types of greenhouses and crops, there are some fundamental principles regarding the mass and energy fluxes of every greenhouse. These principles are illustrated in Figure 3.1 as presented in [23].

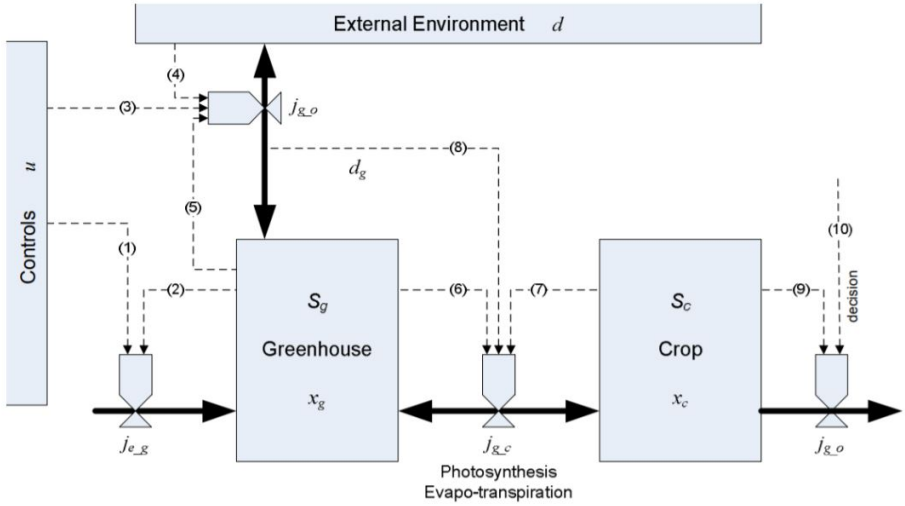

Figure 3.1: Greenhouse fundamental mass and energy fluxes as defined in [23]

In Figure 3.1, the grey rectangles denote both the stored energies and masses in the greenhouse (S_g) as well as the crop (S_c) with states x_g and x_c, respectively. Solid arrows indicate the fluxes, and the dashed arrows denote the variables that influence each flux as defined in Table 3.1. As information flows, the control inputs, the greenhouse states, the crop states, and the disturbances acting on the greenhouse are defined [23].

Observing Figure 3.1, it is understood that there is a high degree of interaction between the greenhouse, the crop, the external environment, and the control signals. As a result, the complexity in combination with the non-linear nature of the climate system complicates the modeling of the dynamical system. The next sections contain an overview of

Table 3.1: Notation Used in Figure 3.1

Stomatal Development	
j_{c_g}	Fluxes between the control equipment and the greenhouse
j_{g_o}	Fluxes between the greenhouse and the outdoor environment
j_{g_c}	Fluxes between the greenhouse and the crop
j_{c_o}	Fluxes between the crop and the outdoor environment
(1)	Greenhouse control inputs (i.e heating/cooling, irrigation, CO_2)
(2), (5), (6)	Greenhouse States (i.e. air temperature, CO_2 concentration, humidity level etc)
(3)	Control inputs permitting the interaction between the greenhouse and the environment (i.e. windows, fans)
(4)	Exogenous signals (i.e. external temperature, weather conditions, solar radiation)
(7), (9)	States of the crop (i.e. growth stage, canopy, amount of fruits)
(8)	Influence of solar radiation on the j_{g_o} fluxes
(10)	Decision actions determined by the grower(i.e. pruning, fruit harvesting)

the elements constituting the greenhouse system.

3.3. OVERVIEW OF GREENHOUSE CLIMATE PROCESSES

Numerous physical entities and processes in the greenhouse climate system can be categorized as greenhouse climate states, crop states, climate inputs, and exogenous inputs. The greenhouse climate states, denoted by x_g, consist of the variables of temperatures, vapor pressure, and carbon dioxide concentration state variables inside the greenhouse. The crop states x_c incorporate the variables defining the crop's stage and conditions in terms of temperature and plant matter. The exogenous inputs d consist of the variables that affect the climate state but cannot be controlled. In the greenhouse case, the exogenous inputs are defined by weather conditions. Finally, the climate control inputs u are the variables compensating or even exploiting the effect of exogenous inputs on climate states. In the context of greenhouse control, the control inputs consist of the actuators

that affect the climate variables.

The state variables of the greenhouse climate-crop system can all be described by differential equations. Regarding the greenhouse climate, the differential equations express three attributes: energy balance, carbon balance, and vapor balance. With respect to the crop, the differential equations express the energy balance and the carbohydrates balance of the tomato plant. The first principles are reviewed in the following sections using the notation introduced in [21]. The state variables of the model are denoted by names with capital letters followed by one subscript. The fluxes are notated with a capital letter followed by two subscripts. The first and second subscripts represent the source and the destination of the flux, respectively. Capital R, H, and L denote radiation, conductive or convective heat, and latent heat exchanges individually.

3.3.1. GREENHOUSE STATES

CLIMATE STATES

Greenhouse modeling is grounded on two assumptions. First, the greenhouse air is considered to be a "perfectly stirred tank", assuming no spatial differences in temperature, CO_2, and vapor concentration. This assumption allows all fluxes to be expressed in units of energy or mass per m^2 of the greenhouse floor. Second, the greenhouse is divided into two main compartments. The compartments above the blackout and thermal screens (top compartment) and the compartment below the screens (main compartment)[22]. The state vector for the greenhouse climate model is presented in Equation (3.1).

$$x_g = \begin{bmatrix} T_{Cov,e} \\ T_{Cov,in} \\ T_{Top} \\ T_{Air} \\ T_{BlScr} \\ T_{ShScr} \\ T_{Can} \\ T_{Flr} \\ T_{So(j)} \\ T_{Pipe} \\ T_{GroPipe} \\ T_{Lamp} \\ T_{IntLamp} \\ CO_{2_{Air}} \\ CO_{2_{Top}} \\ VP_{Air} \\ VP_{Top} \end{bmatrix} \tag{3.1}$$

Following the notation given in Table 3.2 it becomes clear that the state vector includes the temperature variables of all greenhouse elements in ^oC. Particularly, it incorporates the temperature of the outer and inner cover surfaces ($T_{Cov,e}$ and $T_{Cov,in}$), the air temperature in both compartments (T_{Top} and T_{Air}), the various screens (T_{BlScr} and T_{ThScr}),

the canopy (T_{Can}), floor (T_{Flr}), soil ($T_{So(j)}$), heating elements (T_{Pipe} and $T_{GroPipe}$) as well as the temperatures of the artificial lighting (T_{Lamps} and $T_{IntLamp}$). Further, the carbon dioxide concentration ($CO_{2_{Air}}$ and $CO_{2_{Top}}$), in $mg \cdot m^{-3}$, and the vapour pressure (VP_{Air} and VP_{Top}), in Pa, are enclosed for all greenhouse compartments. At this point, it should be noted that the canopy temperature is not exactly a climate state. Although, it has to be included in the climate state vector as it plays a major role in heat exchange processes and acts as a connective link between the crop and climate subsystems.

Table 3.2: Subscripts used in Energy, Carbon, and Vapour Balance Equations

Subscripts Used in Differential Equations			
Air	Greenhouse air compartment below thermal screen	Ind	Industrial source
Blow	Direct air heater	Mech	Mechanical cooling
Boil	Boiler	Out	Outside air
Can	Canopy	Pad	Pad and fan system
e	External side	Pas	Passive heat storage facility
Ext	External CO_2 source	Pipe	Pipe heating system
Flr	Floor	Sky	Sky
Fog	Fogging system	So(j)	The 'j'th the soil layer
Geo	Geothermal heat	Sun	Sun
Glob	Global Radiation	Top	Compartment above the thermal screen
in	Indoor side	ThScr	Thermal Screen
BlScr	Blackout Screen	Lamp	Top-lights
InLamp	Inter-lights		

CROP STATES

The crop growth systems' state variable vector is presented in Equation (3.2), where $Time$ presents the date and time.

$$x_c = \begin{bmatrix} Time \\ T_{Can_{24}} \\ C_{Buf} \\ C_{Leaf} \\ C_{Stem} \\ C_{Fruit} \\ T_{Can_{Sum}} \end{bmatrix} \tag{3.2}$$

$T_{Can_{24}}$ is the average canopy temperature in the last 24 hours in Celsius. $T_{Can_{24}}$ rate is described by the ordinary differential equation (3.3). The mean canopy temperature is

introduced as a state, as it has been shown that it affects the length of the fruit growth period and the size of the fruits [24], [25].

$$\dot{T}_{Can_{24}} = \frac{1}{\tau}(k \cdot T_{Can} - T_{Can_{24}}) \tag{3.3}$$

In Equation (3.3), τ represents the time constant of the process and k is the gain of the process.

C_{Buf} refers to the mass of the accumulating assimilates and C_{Leaf}, C_{Stem}, C_{Fruit} the amounts of carbohydrates in leaves, stems and fruit respectively. The mass of carbohydrates is expressed in units of CH_2O mass per unit area per time unit, in this case; $mg_{CH_2O} \cdot m^{-2} \cdot s^{-1}$. Finally, $T_{Can_{Sum}}$ is the integral of the historical canopy temperature experienced by the plant, and it is used to represent the crop's development stage.

The crop growth process is divided into two stages, the vegetative stage and the generative stage. The use of $T_{Can_{Sum}}$ (Eq. (3.4)) as a stage index may seem counter-intuitive, however, several studies have shown that sub-optimal canopy temperature during the vegetative stage, delay the switching to the generative stage [20]. Hence, the switching point between the two stages could be marked by the integration of the experienced canopy temperatures,

$$\dot{T}_{CanSum} = \frac{1}{t_{Day}} T_{Can} \tag{3.4}$$

where t_{Day} is the time interval for a whole day in appropriate units and T_{Can} is the simulated canopy temperature. Initially, T_{CanSum} is set to a negative value representing that the crop is in the vegetative stage. When T_{CanSum} reaches 0^oC, the crop is passing from the vegetative to the generative stage and the first fruits start appearing.

3.3.2. GREENHOUSE EXOGENOUS INPUTS

In general, exogenous variables are variables whose measures are imposed on the model as defined outside of the model. In the case of greenhouse climate modeling, external climate inputs are regarded as exogenous greenhouse input variables. The external climate inputs incorporated in the GreenLight are:

$$d = \begin{bmatrix} CO_{2_{out}} \\ I_{Glob} \\ T_{Out} \\ T_{Sky} \\ T_{SoOut} \\ VP_{Out} \\ u_{wind} \end{bmatrix} \tag{3.5}$$

where $CO_{2_{out}}$ the outdoor CO_2 concentration in $mg \cdot m^{-3}$, I_{Glob} the outside global radiation in $W \cdot m^{-2}$, T_{Out} the outdoor temperature in ^oC, T_{Sky} the sky temperature in ^oC, T_{SoOut} the soil temperature of outer soil layer in ^oC, VP_{Out} the outdoor vapour pressure in Pa and u_{wind} the outdoor wind speed in $m \cdot s^{-1}$.

3.3.3. GREENHOUSE CLIMATE CONTROL INPUTS

Modern greenhouses aim to influence all three types of greenhouse climatic variables T, CO_2 and VP [21]. To complete this, numerous actuators are required, such as natural ventilation through the window opening, thermal and blackout screens, heating systems, carbon dioxide enrichment systems, and humidifiers/dehumidifiers. In the context of the Greenlight model, the climate control input vector has the following form:

$$u = \begin{bmatrix} Boil \\ BoilGro \\ ExtCo_2 \\ Roof \\ ThScr \\ BlScr \\ Lamp \\ IntLamp \end{bmatrix} \tag{3.6}$$

Where $Boil$ and $BoilGro$ are the boiler values used to heat the heating pipes and the grow pipe system, respectively. $ExtCO_2$ the actuator of the CO_2 enrichment system and $Roof$ the roof opening acting as natural ventilation. All the aforementioned control inputs are continuous variables taking values in the range [0, 1] where 0 declares zero actuator action and 1 proclaims full capacity. $ThScr$ and $BlScr$ the deployment of the thermal and blackout screen accordingly. Screen deployment variables are continuous and their values are in the range [0, 1], expressing the percentage of the deployment. Lastly, $Lamp$ and $IntLamp$ are discrete, boolean variables declaring the state of the top and inter-lights.

3.3.4. CLIMATE'S MASS AND ENERGY FLUXES

The mass and energy fluxes of the greenhouse climate are divided into three main aspects, the energy balance, the carbon balance, and the vapor balance. The following three subsections present the differential equations describing the three greenhouse climate attributes. Moreover, the presented equations are based on the assumption that the greenhouse is a perfectly stirred tank.

ENERGY BALANCE

The energy balance is achieved through the exchange of radiation, convection, conduction, and latent heat between the different entities existing in the greenhouse system. In the context of this section, the radiative heat exchange is divided into two subcategories. The first is the radiative heat exchange because of the short-wave radiation (PAR and Near Infrared Radiation (NIR)). The second subcategory integrates the radiative heat exchange that takes place in the form of Far Infrared Radiation (FIR). It should be underlined that convection and conduction are quite different processes, however, in greenhouse climate modeling, they are lumped together. A schematic representation of the greenhouse climate energy balance is depicted in Figure 3.2.

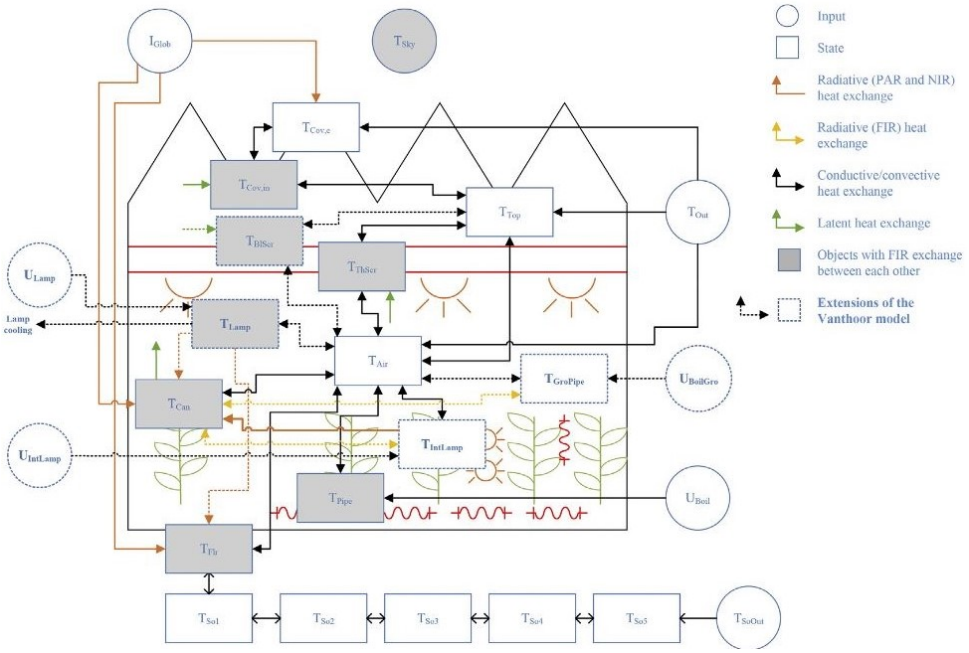

Figure 3.2: Scheme of the GreenLight model energy balance. Dashed and bold items are additions to the Vanthoor model. All the items in grey exchange FIR between each other [12]

An analytical representation of the thermal exchange differential equations governing the climate system is given below. The heat exchange terms are expressed in $W \cdot m^{-2}$. Generally, these differential equations consist of the left-hand side containing the capacity of the corresponding state variable and the time derivative of the state variable. On the right-hand side, the energy fluxes affecting the state variable are declared. At this point, should be mentioned that time derivatives of the state variables are indicated by a dot above the state symbol. Moreover, it is underlined that Table 3.2 presents all subscripts used by the following equations.

Covering Surfaces Covering surfaces play a major role in greenhouse heat conductivity as they consist of a greenhouse design parameter able to significantly induce the temperature gradient across the cover. The internal cover temperature $T_{Cov,in}$ and external cover temperature $T_{Cov,e}$ are described by:

$$cap_{Cov,e}\dot{T}_{Cov,e} = R_{Glob\_SunCov} + H_{Cov,inCov,e} - H_{Cov,eOut} - R_{Cov,eSky} \tag{3.7}$$

$$\begin{aligned} cap_{Cov,in}\dot{T}_{Cov,in} = {} & H_{TopCov,in} + L_{TopCov,in} + R_{CanCov,in} + R_{FlrCov,in} + R_{PipeCov,in} \\ & + R_{ThScrCov,in} - H_{Cov,inCov,e} + R_{BlScrCov,in} + R_{LampCov,in} \end{aligned} \tag{3.8}$$

Where $cap_{Cov,in}$ and $cap_{Cov,e}$ are the heat capacities of the internal and external cover layers respectively. Concerning the internal cover, the condensation of the greenhouse cover causes a latent heat flux affecting its temperature. Additionally, the inner cover exchanges energy, through FIR radiation, with the heating pipes, the screening, the canopy, and the floor. Conductive energy exchange takes place between the inner cover and the top compartment, as also between the inner and outer cover surfaces. The outer covering surface receives energy from the sun and the hotter inner cover. However, it is expected to lose energy due to the low external temperature and sky.

Air Temperature The air temperature above the blackout and thermal screens (top compartment) is defined separately from the greenhouse air below the screens. The reason is that these two temperatures behave differently. The following differential equations present the relationship between the top compartment's air temperature (T_{Top}) with the screenings, the internal covering surface, the external air, and the greenhouse air. The heat capacity of the top compartment's air temperature cannot be neglected.

$$cap_{Top}\dot{T}_{Top} = H_{ThScrTop} + H_{AirTop} - H_{TopCov,in} - H_{TopOut} + H_{BlScrTop} \tag{3.9}$$

The greenhouse air temperature T_{Air} is described by:

$$\begin{aligned} cap_{Air}\dot{T}_{Air} = {} & H_{CanAir} + H_{PipeAir} + R_{Glob\_SunAir} - H_{AirFlr} - H_{AirThScr} \\ & - H_{AirOut} - H_{AirTop} - H_{AirBlScr} + H_{LampAir} + R_{LampAir} \\ & + H_{intLampAir} + H_{GroPipeAir} \end{aligned} \tag{3.10}$$

Where cap_{Air} is the heat capacity of the greenhouse air. Greenhouse air is constantly exchanging sensible heat with the surrounding elements and surfaces. Air receives thermal energy from the canopy, the heating pipes, and the lamps in a conductive way. It also receives radiative energy from the sun in the form of global radiation. Although, all the surrounding surfaces and the air of the top compartment are receiving thermal energy from the greenhouse air. Finally, since greenhouse air has a higher temperature than the air outside of it, a conductive heat exchange takes place between the air inside and outside the greenhouse.

Screens The equations for the temperatures of the blackout (T_{BlScr}) and thermal (T_{ThScr}) screens are

$$\begin{aligned} cap_{BlScr}\dot{T}_{BlScr} = {} & H_{AirBlScr} + L_{AirBlScr} + R_{CanBlScr} + R_{FlrBlScr} + R_{PipeBlScr} \\ & - H_{BlScrTop} - R_{BlScrCov,in} \\ & - R_{BlScrSky} - R_{BlScrThScr} + R_{LampBlScr} \end{aligned} \tag{3.11}$$

$$\begin{aligned} cap_{ThScr}\dot{T}_{ThScr} = {} & H_{AirThScr} + L_{AirThScr} + R_{CanThScr} + R_{FlrThScr} + R_{PipeThScr} \\ & - H_{ThScrTop} - R_{ThScrCov,in} \\ & - R_{ThScrSky} + R_{BlScrThScr} + R_{LampThScr} \end{aligned} \tag{3.12}$$

where cap_{BlScr} and cap_{ThScr} are the heat capacities of the blackout and thermal screen respectively. Both screens are affected by surrounding elements in a similar way. Screens

receive conductive energy and latent heat from the air. In addition, they receive radiant energy from the canopy, the floor, the lamps, and the heating pipes. Both screens provide energy to the top compartment, the inner covering surface, and the sky. The main behavioral difference can be found in the heat exchange between the screens, where the blackout screen provides energy and the thermal screen receives it.

Canopy Temperature Canopy temperature T_{Can} is described by:

$$\begin{aligned} cap_{Can}\dot{T}_{Can} = {} & R_{PAR\_SunCan} + R_{NIR\_SunCan} + R_{PipeCan} - H_{CanAir} - L_{CanAir} \\ & - R_{CanCov,in} - R_{CanFlr} - R_{CanSky} - R_{CanThScr} + R_{PAR\_LampCan} \\ & + R_{NIR\_LampCan} + R_{FIR\_LampCan} + R_{PAR\_IntLampCan} \\ & + R_{NIR\_IntLampCan} + R_{FIR\_IntLampCan} + R_{GroPipeCan} \end{aligned} \tag{3.13}$$

where cap_{Can} is the heat capacity of the canopy. Canopy temperature is highly dependent on incident radiation. The radiant energy comes in multiple bands (PAR, NIR, FIR) emitted by the Sun, the Lamps, and the inter-lights that may be installed in the greenhouse. Moreover, the radiation from heating pipes $R_{PipeCan}$ is also affecting T_{Can}. However, it should be noted that the canopy emits radiation which is absorbed by the surrounding surfaces. Finally, the canopy exchanges thermal energy with the air in the form of conductive and latent heat.

Floor and Soil The temperature of the floor surface is dependent on the air, the incident short-wave radiation, and the long-wave radiation emitted by the surrounding elements. The floor temperature T_{Flr} is described by:

$$\begin{aligned} cap_{Flr}\dot{T}_{Flr} = {} & H_{AirFlr} + R_{PAR\_SunFlr} + R_{NIR\_SunFlr} + R_{CanFlr} + R_{PipeFlr} \\ & - H_{FlrSo1} - R_{FlrCov,in} - R_{FlrSky} - R_{FlrThScr} - R_{FlrBlScr} \\ & + R_{PAR\_LampFlr} + R_{NIR\_LampFlr} + R_{FIR\_LampFlr} \end{aligned} \tag{3.14}$$

where cap_{Flr} is the heat capacity of the floor. The floor temperature receives convective energy from the air and provides it to the first layer of soil. In terms of radiative heat exchanges, the floor receives energy from the sun, the canopy, the lamps, and the pipes, and provides energy to the inner cover, the sky, the thermal screen, and the blackout screen.

For modeling purposes, it is assumed that the soil is divided into five layers. Each layer is denoted by the subscript j where $j \in \{1,\dots,5\}$. As can be seen in the next equation, each layer exchanges thermal energy by convection with its neighboring layers, by receiving energy from the previous layer and providing it to the next one.

$$cap_{So(j)}\dot{T}_{So(j)} = H_{So(j-1)So(j)} - H_{So(j)So(J+1)}, \quad j = 1,\dots,5 \tag{3.15}$$

Heating Pipes As presented in Equation (3.16), the heating pipes receive energy from the boiler and the lamps in a convective and radiative way, accordingly. Then the heating

pipes emit energy to the sky, inner cover, canopy, floor, thermal and blackout screen in the form of radiation. Furthermore, the pipes heat the air by convection.

$$\begin{aligned} cap_{Pipe}\dot{T}_{Pipe} = H_{BoilPipe} - R_{PipeSky} - R_{PipeCov,in} - R_{PipeCan} - R_{PipeFlr} \\ - R_{PipeThScr} - H_{PipeAir} - R_{PipeBlScr} + R_{LampPipe} \end{aligned} \tag{3.16}$$

The differential equation describing the grow pipes temperatures $T_{GroPipe}$ is given below:

$$cap_{GroPipe}\dot{T}_{GroPipe} = H_{BoilGroPipe} - R_{GroPipeCan} - H_{GroPipeAir} \tag{3.17}$$

The main difference between the heating and grow pipes is the assumption that for a mature crop, the majority of the radiative heat from the grow pipes is absorbed by the canopy. Hence, the FIR exchange terms between the grow pipes and other greenhouse elements are neglected.

Lamps The lamp temperatures T_{Lamp} and $T_{IntLamp}$ are described by:

$$\begin{aligned} cap_{Lamp}\dot{T}_{Lamp} = Q_{LampIn} - R_{LampSky} - R_{LampCov,in} - R_{LAmpThScr} \\ - R_{LampBlScr} - H_{LampAir} - R_{PAR\_LampCan} - R_{NIR\_LampCan} \\ - R_{FIR\_LampCan} - R_{LampPipe} - R_{PAR\_LampFlr} - R_{NIR\_LampFLr} \\ - R_{FIR\_LampFLr} - R_{LampAir} - H_{LampCool} \end{aligned} \tag{3.18}$$

$$\begin{aligned} cap_{LampInt}\dot{T}_{LampInt} = Q_{IntLampIn} - H_{IntLampAir} - R_{PAR\_IntLampCan} \\ - R_{NIR\_IntLampCan} - R_{FIR\_IntLampCan} \end{aligned} \tag{3.19}$$

Where cap_{Lamp} and $camp_{LampInt}$ are the heat capacities of the top lights and the inter-lights. Lamp temperatures are affected by the electrical input Q_{LampIn}, the FIR exchange between the lamps and the sky, cover, thermal screen, blackout screen, canopy, heating pipes, and the floor. There are also radiative exchanges between the lamps and the canopy and the floor, which take place in the PAR and NIR bands. Short wave radiation not absorbed by the floor or canopy is assumed to be transferred to the greenhouse air ($R_{LampAir}$). Besides, between lamps and greenhouse air, the convective heat transfer $H_{LampAir}$ is taking place. Finally, the term $H_{LampCool}$ becomes nonzero for the case of lamps with active cooling systems.

CARBON BALANCE

The carbon balance is used for the definitions of two state variables, the CO_2 concentration in the greenhouse air and the concentration in the top compartment. The CO_2 concentration of the greenhouse air is denoted by $CO_{2_{Air}}$ and is described as follows:

$$cap_{CO_{2_{Air}}}\dot{CO}_{2_{Air}} = MC_{ExtAir} - MC_{AirCan} - MC_{AirTop} - MC_{AirOut} \ \ [mg \cdot m^{-2} \cdot s^{-1}] \tag{3.20}$$

Where $cap_{CO_{2_{Air}}}$ is the capacity of the air to store CO_2. The carbon dioxide exchange takes place between the greenhouse air and the surrounding elements. An external CO_2

source MC_{ExtAir}, the top compartment MC_{AirTop}, and the outdoor air MC_{AirOut} are considered as surrounding elements. The term MC_{AirCan} defines the CO_2 flux between the greenhouse air and the canopy. Respiration and photosynthesis are responsible for this flux.

The CO_2 concentration of the top compartment is expressed as $CO_{2_{Top}}$ and is described by the following equation:

$$cap_{CO_{2_{Top}}} \dot{CO}_{2_{Top}} = MC_{AirTop} - MC_{TopOut} \ [mg \cdot m^{-2} \cdot s^{-1}] \tag{3.21}$$

Where $cap_{CO_{2_{Top}}}$ is the capacity of the top compartment air to store CO_2. MC_{TopOut} recites the O_2 exchange between the top compartment air and the air outside the greenhouse.

VAPOUR BALANCE

Similarly to carbon balance, vapor balance defines the vapor pressure of the greenhouse air (VP_{Air}) and the vapor pressure of the air in the top compartment (VP_{Top}). VP_{Air} is presented in Equation (3.22) where $cap_{VP_{Air}}$ is the capacity of the air to store water vapor. Vapor is exchanged between the air and surrounding elements. Several surrounding elements can be applied to a greenhouse system such as fogging systems, direct air heaters, and pads. However, in our case, the elements determining the vapor balance are the canopy (MV_{CanAir}), the thermal screens ($MV_{AirThScr}$), and the blackout screens ($MV_{AirBlScr}$). Terms that should not be neglected are the exchanges with the top compartment (MV_{AirTop}) and the outside air (MV_{AirOut}).

$$\begin{aligned} cap_{VP_{Air}} \dot{VP}_{Air} = MV_{CanAir} - MV_{AirThScr} - MV_{AirTop} - MV_{AirOut} \\ - MV_{AirBlScr} \ [kg \cdot m^{-2} \cdot s^{-1}] \end{aligned} \tag{3.22}$$

VP_{Top} is described as below:

$$cap_{VP_{Top}} \dot{VP}_{Top} = MV_{AirTop} - MV_{TopCov,in} - MV_{TopOut} \ [kg \cdot m^{-2} \cdot s^{-1}] \tag{3.23}$$

Where $cap_{VP_{Top}}$ is the capacity of the top compartment to store water vapor, and $MV_{TopCov,in}$ is the vapor exchange between the top compartment and the internal cover layer. Finally, MV_{TopOut} portrays the vapor exchange between the top compartment and the outside air.

3.3.5. CROP'S MASS AND ENERGY FLUXES

Regarding the crop, the mass and energy fluxes are realized in the form of carbohydrate flow. The following paragraphs present the ordinary differential equations describing the mass and energy fluxes taking place in the tomato crop.

Assimilates in the Buffer The availability of carbohydrates in the buffer is ruled by photosynthesis and growth respiration. Photosynthesis produces carbohydrates, growth rate consumes a portion of the stored carbohydrates, and part of the remaining resources are distributed to crop organs. The rate of change of the stored accumulating assimilates is noted as $\dot{C}_{Buf}$ and is described as follows

$$\dot{C}_{Buf} = MC_{AirBuf} - MC_{BufFruit} - MC_{BufLeaf} - MC_{BufStem} - MC_{BufAir} \tag{3.24}$$

where MC_{AirBuf} the mass of produced assimilates by photosynthesis, MC_{BufAir} the mass of consumed carbohydrates by growth respiration and $MC_{BufFruit}$, $MC_{BufLeaf}$, $MC_{BufStem}$ the carbohydrate flows from buffer to fruits, leaves and stems respectively.

The carbohydrate buffer has limited capacity, and when the biomass approaches its lower limit, the carbohydrate flow to the organs stops. On the other hand, when the upper limit is reached, photosynthesis stops the production of biomass, as newly produced carbohydrates can not be stored.

Carbohydrates in Organs As presented in the previous paragraphs, each organ updates the number of carbohydrates through the assimilation partitioning process, and it consumes a part of the received carbohydrates through the maintenance respiration. However, leaves and fruits are also affected by the pruning and harvesting processes. In order to describe the amount of the stored carbohydrates in the different organs and also express their rate of change, the ordinary differential equations (3.25), (3.26) and (3.27) are introduced.

$$\dot{C}_{Fruit} = MC_{BufFruit} - MC_{FruitAir} - MC_{FruitHar} \tag{3.25}$$

where $MC_{BufFruit}$ the carbohydrates' flow from the buffer to the fruits, $MC_{FruitAir}$ the maintenance respiration of fruits, and $MC_{FruitHar}$ the fruit harvesting.

$$\dot{C}_{Leaf} = MC_{BufLeaf} - MC_{LeafAIr} - MC_{LeafHar} \tag{3.26}$$

where $MC_{BufLeaf}$ the carbohydrates' flow to the leaves, $MC_{LeafAir}$ the maintenance respiration of leaves, and $MC_{LeafHar}$ the leaf pruning.

$$\dot{C}_{Stem} = MC_{BufStem} - MC_{StemAir} \tag{3.27}$$

where $MC_{BufLeaf}$ the carbohydrates' flow to the stem and $MC_{LeafAir}$ the maintenance respiration of stems.

The presentation of the differential equations describing the climate-crop system does not contain all the analytical expressions required for the estimation of the individual terms. The extensive review of all needed analytical equations is not included for the sake of brevity. Readers interested in a more detailed description of the numerous terms appearing in the differential equations are referred to Appendix A of [12].

3.4. OVERVIEW OF THE MODELS USED

The previous section presented the coupled climate-crop model named Greenlight which is basically an extension of the well-known KASPRO [21] and Vanthoor-Stangellini [22] models. In the context of this thesis project, the Greenlight climate-crop model is extensively used to satisfy various requirements. However, some of the presented experiments do not require the use of the coupled model but only the crop model or a crop model variant. For this reason, it was considered appropriate to provide a high-level overview of the models used in the next chapters to prevent any confusion.

3.4.1. THE GREENLIGHT MODEL

The Greenlight greenhouse model occurs by coupling the greenhouse climate model and the greenhouse crop model. This model is used for data generation in Chapter 5, and as the ground truth and predictive model for the climate controller in Chapter 6. The coupled model can be expressed as:

$$\dot{x} = \begin{bmatrix} \dot{x}_g \\ \dot{x}_c \end{bmatrix} = f_{gh}(x, u, d, t) \tag{3.28}$$

Where x the greenhouse-crop state vector (3.2), u the control inputs (3.6), d the exogenous inputs (3.5), t the time and f_{gh} contains the state transition equations (3.3)-(3.27).

For simulation purposes, a weather dataset describing the weather conditions in Amsterdam will be used. It should be mentioned that the dataset containing the exogenous inputs is sampled every 5 minutes (300 seconds). Hence, the greenhouse model has to take the following non-linear discrete-time form:

$$x(k+1) = \begin{bmatrix} x_g(k+1) \\ x_c(k+1) \end{bmatrix} = f_{gh}(x(k), u(k), d(k), k) \tag{3.29}$$

A high-level, graphical demonstration of the Greenlight model is provided in Figure 3.3.

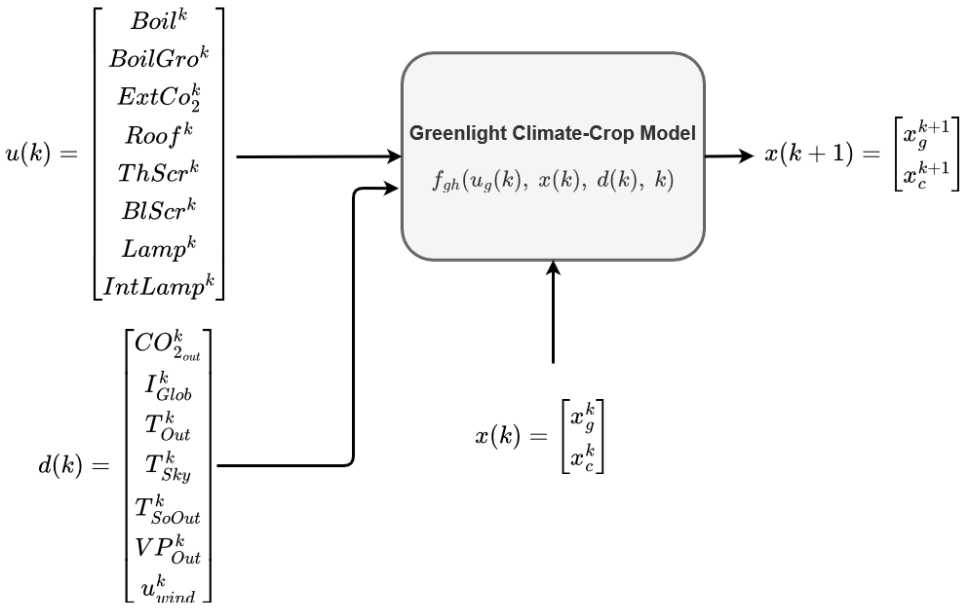

Figure 3.3: A visual representation of the Greenlight climate-crop model in a block diagram form. The climate-crop variables living in the state vector x can be found in Equation (3.2)

The non-linear discrete form occurs through Euler's discretization method where:

$$x(k+1) = x(k) + h \cdot f_{gh}(x(k), u(k), d(k), k) \tag{3.30}$$

Subsequently, the model is implemented in `MATLAB` where the solver `ode15s` is used for the calculation of discrete-time differential equations. The introduced model is parametric and as a result, calibration is required. The parameters used in this thesis project were taken from [12] as authors have already validated the GreenLight model using data from a greenhouse experiment realized from 16 October 2009 to 1 July 2010 in Bleiswijk, the Netherlands. The presentation of the parameters is out of the scope of the thesis projects. For this reason, the readers interested in the parameter's evaluation, are referred to [12] and the provided source code in the given link:

`https://github.com/davkat1/GreenLight`

For the sake of completeness, it is noted that the models presented in the following paragraphs are similarly discretized based on the same sampling time. The needed crop parameters were also taken from [12].

3.4.2. THE VANTHOOR TOMATO CROP MODEL

The Vanthoor tomato crop model is the original crop model incorporated in the Greenlight model. The crop model is described by the state vector presented in Equation (3.2)

and the differential equations (3.3), (3.4), (3.24), (3.25), (3.26), and (3.27). When used independently, the Vanthoor model requires an input vector containing the canopy temperature, the incident radiation above the canopy, the carbon dioxide concentration, and the vapor pressure in the air.
The tomato crop model can be visualized in the form of a block diagram as shown in Figure 3.5.

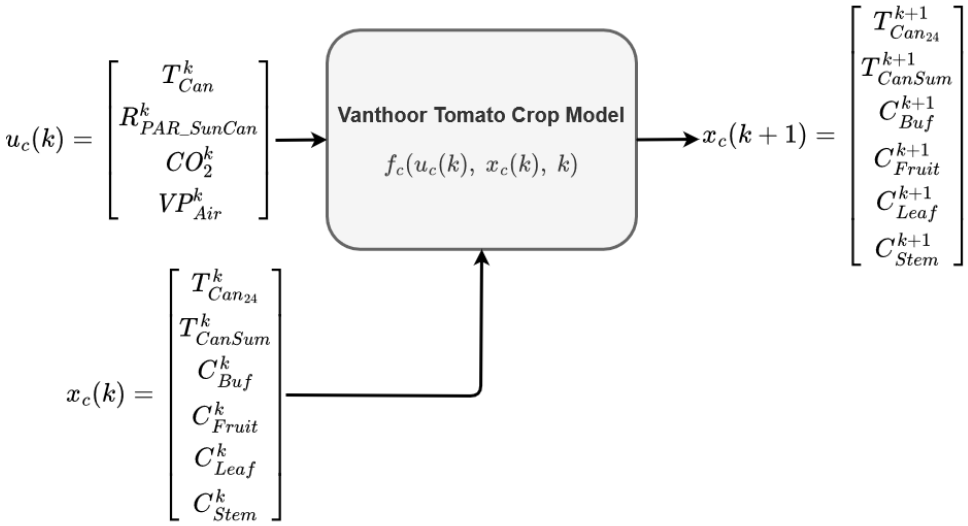

Figure 3.4: A visual representation of the Vanthoor tomato crop model in a block diagram form.

An application requiring the presented crop model can be found in Section 4.6 for the comparison between two controllers. The experiment is used as a proof of concept to show that a novel objective function suitable for climate control can produce comparable results with the well-tested fruit maximization NMPC approach. Particularly, the Vanthoor tomato crop model is used as the ground truth for both controllers and also as the predictive model of the reference controller.

3.4.3. CLIMATE-DEPENDENT TOMATO CROP MODEL

In Chapter 5, and specifically in Section 5.5, a crop model receiving only climate variables as inputs is required. The model has to receive as inputs the air temperature, the vapor pressure in the air, the carbon dioxide concentration, and the incident radiation above the canopy. The main difference with the Vanthoor model is that now the canopy temperature has become a state. Essentially, the following differential equation has been added in the Vanthoor tomato crop model:

$$\dot{T}_{Can} = \frac{1}{cap_{Can}}(R_{PAR\_SunCan} - H_{CanAir} - L_{CanAir}) \tag{3.31}$$

The climate-dependent tomato crop model can be visualized as:

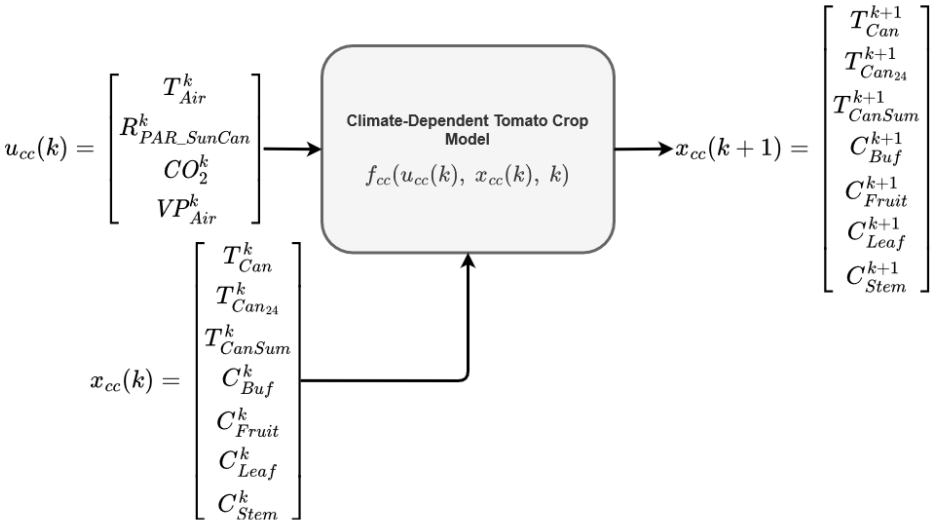

Figure 3.5: A visual representation of the climate-dependent, modified Vanthoor, tomato crop model in a block diagram form.

4

SPA Objective Function Definition

4.1. Introduction

Conventionally, the grower acts as the connective link between the climate computer and the crop. The grower's job is to periodically examine the crop in a greenhouse environment and receive information about the crop status based on his senses. Then, the sensed information in combination with the environmental data and the grower's experience help the grower decide the future climate setpoints that would be beneficial for the crop. These climate setpoints are fed into a reference tracking controller and they are eventually realized. Afterward, the grower repeats his periodical crop check, decides about future conditions, and informs the climate controller. Practically, the grower constitutes an important subsystem in the climate control loop. The problem, in this case, has to do with the fact that the efficiency of the method depends on the human factor. Consequently, the elimination of human interference from the decision-making process is a matter of great importance and it has to be achieved to steer the evolution to autonomous greenhouses.

A method capable to contribute in these changes is the SPA proposing that the optimal environmental conditions should be indicated by the plant itself via a single or a set of signals correlated with the crop yield [4]. As mentioned in Chapter 1, thermal cameras satisfy the SPA applicability, reliability, and financial requirements. Thus, it can be assumed that thermography can become a standard in greenhouse measurement equipment. The introduction of thermography permits the admission of the following assumption:

- *The climate controller is aware of the temperatures of the various greenhouse surfaces at all times.*

As a result, the temperature states presented in section 3.3.1 can now be considered measurable states of the greenhouse system. Canopy temperature is a state of great signif-

icance for the proposed SPA approach and its role will be analytically presented in the following sections.

The scope of this chapter is to present the research carried out regarding the formulation of a SPA-based objective function able to be used in predictive climate control techniques. The designation of the canopy's temperature importance is also in the context of the chapter's scope. Particularly, this chapter contains the presentation of the fundamentals behind the design of the objective function and the analytical overview of its components.

4.2. OBJECTIVE

The ingredients of the tomato crop have to be understood before diving into the targets of the proposed objective function. The Vanthoor tomato mechanistic model, presented in [20], describes the biochemical processes of the crop. The structural processes of the tomato crop can be summarized in the high-level block diagram description presented in Figure 4.1 introduced in [3]. For a better understanding of Figure 4.1, it should be noted that the direction of the arrows represents whether the signal is an input or an output signal to a specific process.

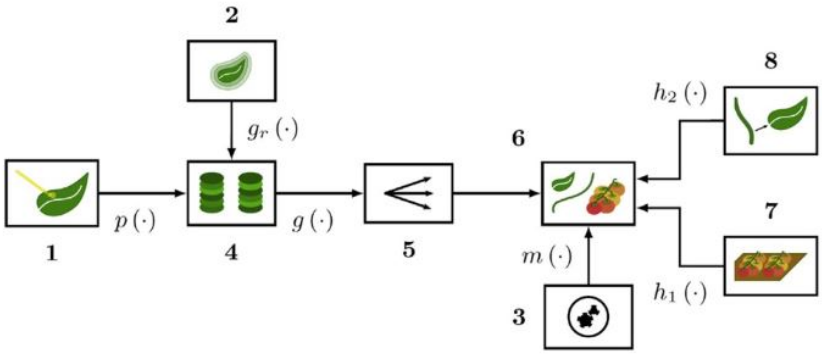

Figure 4.1: Common biological processes of tomato crop growth. The numbers and symbols represent the following processes: 1: photosynthesis, 2: growth respiration, 3: maintenance respiration, 4: accumulating assimilates, 5: assimilate partitioning, 6: accumulating biomass, 7: tomato harvest, 8: leaf harvest, p: assimilates generated through photosynthesis, g_r: the assimilates used for growth, g: the amount of assimilates converted to biomass, m: part of the biomass used for maintenance respiration, h1: harvest of fruits, h2: harvest of leaves [3]

From a high-level perspective, photosynthesis (Block 1 in Figure 4.1) is the process by which the plant converts light energy and CO_2 into carbohydrates [26]. The produced carbohydrates are stored in a carbohydrate buffer(Block 4 in Figure 4.1) and they have several fates. The largest fraction of the produced carbohydrates is used by the respiration process. Respiration can be considered reverse photosynthesis where the stored carbohydrates are oxidized. The oxidization process releases energy which is used for

the production of plant matter (Growth Respiration, Block 2 in Figure 4.1) and the maintenance of the already existing plant matter (Maintenance Respiration, Block 3 in Figure 4.1) [27]. Block 4 in Figure 4.1 acts as a buffer collecting the carbohydrates that will, eventually, be converted to biomass. The stored carbohydrates are distributed to the plant's organs (stems, leaves, fruits, etc) through Block 5 (Fig. 4.1), concluding on the accumulated biomass Block 6 (Fig. 4.1). Finally, the last labor processes, such as harvest and leaf picking, have to be defined. These processes are described by Blocks 7 and 8.

The objective function is going to be used in a predictive climate controller that has to define the future, optimal, actuator inputs leading to the optimal yield-cost ratio according to the crop's needs. In other words, the objective function should permit monitoring the carbohydrates production, ensuring a high degree of carbohydrates distribution from the buffer to the organs, and eliminating the carbohydrates losses by the crop while it takes into consideration the economic impact. The actuator's inputs are responsible to steer the canopy temperature, the mean canopy temperature, and VPD according to the decided trajectories. Concerning CO_2 a constant reference carbon dioxide level has to be tracked. The design of such an objective function requires a good understanding of the background processes running in Blocks 1, 2, 3, 4, and 5 of Figure 4.1.

4.3. MONITORING CARBOHYDRATES PRODUCTION

From the biological perspective, carbohydrate production in a tomato plant is achieved through photosynthesis. Photosynthesis is a complex biochemical process directly dependent on the canopy temperature, the absorbed photosynthetically active radiation (PAR), and the CO_2 concentration in the greenhouse. Vapor pressure deficit (VPD) indirectly affects photosynthesis through stomatal conductance. Moreover, photosynthesis has a positive correlation with crop yield [28]. Hence, the monitoring of a crop's carbohydrate production requires the introduction of a term related to photosynthesis.

Several photosynthesis models have been proposed in the literature with FvCB [29] and its variants ([30], [31] and [32]) being the most widely used. FvCB-based models describe the kinetic properties of the enzymes governing photosynthesis (Rubisco and RuBP), the electron transport process, the chemical reactions consuming CO_2, and the photorespiration. However, these photosynthesis models are complex as they summarize the biochemistry behind the process and they are highly dependent on crop-specific parameters. These characteristics make FvCB-based models a deficient and computationally expensive candidate for a general-purpose objective function. In consequence, carbohydrate production has to be monitored through a computationally cheaper index, able to be measured using thermography. This index should be capable to be measured right after the crop's germination and finally, it should maintain a positive correlation with photosynthesis.

Stomatal conductance is an index able to achieve indirect yield coordination while satisfying the aforementioned criteria. Leaf temperature is a variable with great importance for the photosynthesis process and it is strongly dependent on transpiration. Transpiration is dominated by stomatal conductance, and studies have shown that this dom-

inance is intensified under deficient fertirrigation [33]. Another important variable of photosynthesis is the intercellular CO_2 concentration, which depends on the ambient CO_2 concentration, the stomatal aperture, and the photosynthetic rate. Particularly, the stomatal aperture controls the gas exchange flows between the plant and the air, as a result, stomata can significantly affect the intercellular CO_2 concentration. Figure 4.2 displays the relationship between net photosynthesis and stomatal conductance as a function of light, where each black line presents the value of net photosynthesis for constant light intensity, air temperature, relative humidity, and variable stomatal conductance. Figure 4.2 validates the positive correlation between stomatal conductance and photosynthesis as for constant environmental conditions, increasing stomatal conductance triggers the increase of photosynthesis. Concluding, the objective function has to include a stomatal conductance maximization term or a stomatal resistance minimization term to ensure high levels of carbohydrate production.

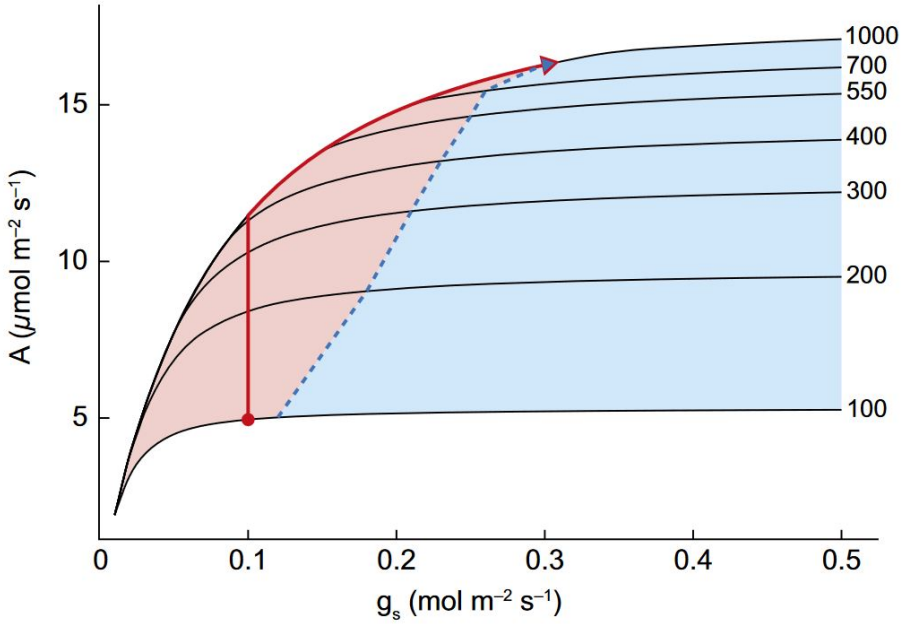

Figure 4.2: Relationship between net CO_2 assimilation A and stomatal conductance g_s, as a function of light intensity (indicated on each line, $\mu mol \cdot m^{-2} \cdot s^{-1}$) under constant temperature and air relative humidity. Each line represents the value of A if light intensity was kept constant and only g_s varied. Net CO_2 assimilation was calculated using the equations provided in [34]. The red line illustrates the trajectory of A after a step change in irradiance (from 100 to 1000$\mu mol \cdot m^{-2} \cdot s^{-1}$), showing the instantaneous increase in A followed by a slow increase of it limited by the g_s response. The blue dashed line represents the g_s value required to achieve 95% of maximum A(depending on light intensity), and represents the trajectory if A and g_s were fully synchronized. Values to the left of the trajectory(blue line) represent g_s limitation of A(red shading), whilst those to the right (blue shading) represent unnecessary water loss relative to CO_2 gain. Red dots represent the initial g_s and the arrow the final steady-state g_s. [7]

4.3.1. STOMATAL RESISTANCE MODEL ANALYSIS

The control of gas exchange between plants and the atmosphere is considered as one of the holy grails in plant physiology [35]. For this reason, more than 40 stomatal conductance models capable to account for multiple environmental factors such as air temperature, relative humidity, VPD, carbon assimilation, and lighting conditions have been developed. In the context of this project, the Stanghellini stomatal conductance model ([36]) is used as ground truth.

As shown in Equation (4.1), the Stanghellini model considers stomatal conductance a product of four factors.

$$r_S(T_{Can}, VP_{Air}, R_{Can}, CO2_{ppm}) = \frac{1}{g_S(T_{Can}, VP_{Air}, R_{Can}, CO2_{ppm})} = r_{S_{min}} \cdot r_{f_{rCan}}(R_{Can}) \cdot r_{f_{CO_2}}(R_{Can}, CO2_{ppm}) \cdot r_{f_{V_p}}(T_{Can}, VP_{Air}, R_{Can}) \quad (4.1)$$

Where r_S the stomatal resistance, g_S the stomatal conductance, $r_{S_{min}}$ the minimum canopy resistance for transpiration (constant), $r_{f_{rCan}}$ the influence of solar radiation, $r_{f_{CO_2}}$ the influence of carbon dioxide and $r_{f_{V_p}}$ the influence of vapor pressure on stomatal resistance. In Stanghellini model, stomatal resistance is a function of the canopy temperature T_{Can} in ^oC, the vapor pressure in air VP_{Air} in Pa, the PAR and NIR radiation above the canopy R_{Can} in $W \cdot m^{-2}$ and the carbon dioxide concentration $CO2_{ppm}$ in Ppm.

An analytical overview of all factors living in the Stanghellini model is out of the scope of this project. However, in order to understand the impact of each dependent variable on stomatal resistance, the conduction of the sensitivity analysis is a matter of high importance. As depicted in the upper left plot in Figure 4.3 the absence of incident radiation increases the stomatal resistance, restricting the gas exchange flows and transpiration. During the daytime, the increasing PAR and NIR radiation above the canopy does not have a significant effect on stomatal resistance. It should be noted, though, that the function of stomatal resistance to incident radiation is monotonically non-increasing. According to the lower left plot, increasing CO_2 monotonically increases stomatal resistance till it reaches a threshold value.

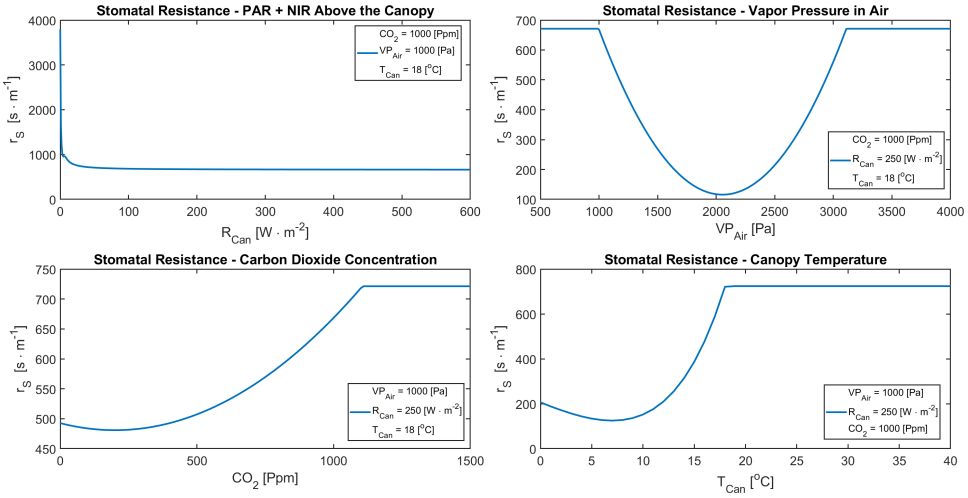

Figure 4.3: One variable sensitivity analysis on the Stanghellini stomatal conductance model

From the plots of the right column in Figure 4.3 becomes clear that the depicted functions are convex. As a result, it is understandable that for given CO_2 and R_{Can} values, a combination of VP_{Air} and T_{Can} values resulting in minimum stomatal resistance can be found. This observation is depicted in Figure 4.4 where the 3D plot of stomatal resistance to VP_{Air} and T_{Can} for constant CO_2 and R_{Can} values is presented.

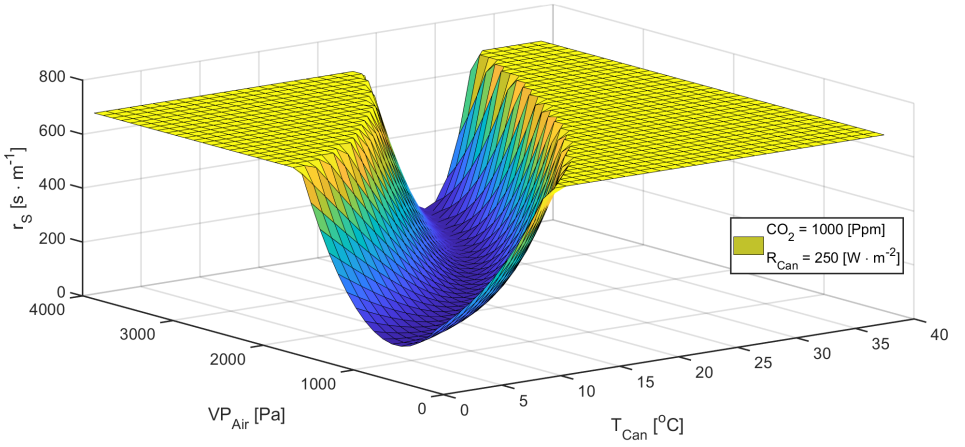

Figure 4.4: Two factor sensitivity analysis on the Stanghellini stomatal conductance model

Based on Figure 4.3 stomatal resistance becomes minimum for maximum R_{Can}, minimum CO_2 and a combination of VP_{Air} and T_{Can} falling on the bottom of the convex surface shown in Figure 4.4. However, low CO2 values result in limited intercellular CO2, restricting photosynthesis. For this reason, the stomatal resistance minimization has to

take place for a given level of CO2 concentration. Consequently, r_{fCO_2} does not need to be a part of the carbohydrates monitoring term. Furthermore, it has to be mentioned that the climate controller developed in this thesis project does not include the control of artificial lighting. Therefore, $r_{f_{rCan}}$ does not have to be a part of the objective function. Based on these constraints, the minimization of stomatal resistance can be achieved through the minimization of $r_{f_{Vp}}$.

4.3.2. TEMPERATURE INDUCED RESISTANCE CONDUCTANCE MINIMIZATION

The vapor pressure influence on stomatal resistance is described by Equation (4.2).

$$r_{f_{VP}}(T_{Can}, VP_{Air}, R_{Can}) = min\Big(5.8,\ 1 + p_{C_{Evap_4}} \cdot R_{Can} \cdot (p_1 \cdot e^{\frac{p_3 \cdot T_{Can}}{T_{Can}+p_2}} - VP_{Air})^2\Big) \quad (4.2)$$

Where $p_{C_{Evap_4}}$ the parameter for vapor pressure influence on stomatal resistance in Pa^{-2} as described in Equation (4.3) and p_i for $i = \{1,2,3\}$ parameters used for the calculation of the saturated vapor pressure at temperature T_{Can}.

$$p_{C_{Evap_4}}(R_{Can}) = \begin{cases} p_{EvapNight},\ R_{Can} \leq 0 \\ p_{EvapDay},\ R_{Can} \geq 0 \end{cases} \quad (4.3)$$

Where $p_{EvapNight}$ the coefficient of the vapor pressure effect on stomatal resistance at night in Pa^{-2}, $p_{EvapDay}$ the coefficient of the vapor pressure effect on stomatal resistance during the day time in Pa^{-2}.

Concluding, the minimization of stomatal resistance for the case of no lighting control in combination with a constant CO_2 reference level can be achieved through the minimization of the following term:

$$\min_u ||r_S||_2^2 \equiv \min_u ||p_1 \cdot e^{\frac{p_3 \cdot T_{Can}}{T_{Can}+p_2}} - VP_{Air}||_2^2 \quad (4.4)$$

4.4. MONITORING CARBOHYDRATES DISTRIBUTION

According to most mechanistic tomato crop models, the produced carbohydrates are temporarily stored in a buffer (Block 4, Figure 4.1). A portion of these carbohydrates will be consumed by the growth respiration process (Block 2) and another portion will be distributed to the organs (Block 5). However, the photosynthetic capacity is dependent on the capacity of the carbohydrate buffer. The more carbohydrates in the buffer, the lower the maximum photosynthesis. In the extreme scenario where the buffer is full, photosynthesis stops. Consequently, it has great importance to maintain the high flow of carbohydrates from the buffer to the organs to ensure high photosynthetic potential that will eventually lead to increased yields.

4.4.1. CARBOHYDRATES FLOW FROM BUFFER TO ORGANS

The carbohydrate flow from the buffer to the fruits is determined by the following equation

$$MC_{BufFruit} = h_{Buf}^{MC_{BufOrg}} \cdot h_{T_{Can}} \cdot h_{T_{Can_{24}}} \cdot h_{T_{CanSum}} \cdot g_{T_{Can_{24}}} \cdot rg_{Fruit} \tag{4.5}$$

where h are the inhibition factors. Inhibition factors are taking values in the range $[0,\ 1]$ and they define the impact of insufficient carbohydrate concentration in the buffer ($h_{Buf}^{MC_{BufOrg}}$), and of non-optimal instantaneous and mean canopy temperatures ($h_{T_{Can}}$, $h_{T_{Can_{24}}}$). At first, the crop is in the vegetative stage and all carbohydrates are used for stem and leaf growth, and when a given canopy temperature sum is reached, the crop switches to the generative stage. Specifically, when T_{CanSum} exceeds 0^oC, the carbohydrates' distribution to the fruits increases linearly from zero till its full potential is reached at the temperature sum T_{CanSum}^{End}. When T_{CanSum}^{End} is reached, the carbohydrates' distribution to the fruits remains at its potential value. This behavior is described through the inhibition factor of the development stage $h_{T_{CanSum}}$. The effect of temperature on the carbohydrate flow to fruits is described by $g_{T_{Can_{24}}}$, and rg_{Fruit} is the potential fruit growth rate coefficient for a canopy temperature at 20^oC. An experimental value for rg_{Fruit} has been proposed in [37].

The carbohydrates flow from the buffer to the organs is determined as follows

$$MC_{BufOrg(i)} = h_{Buf}^{MC_{BufOrg}} \cdot h_{T_{Can_{24}}} \cdot g_{T_{Can_{24}}} \cdot rg_{(i)} \tag{4.6}$$

where i represents the plant organ code for $Leaf$ and $Stem$, and $rg_{(i)}$ is the potential organ growth rate coefficient at 20^oC. It should be underlined that the flow to the organs is not affected by the instantaneous canopy temperature. Values for the potential growth rate coefficients $rg_{(i)}$ have been proposed in [38].

CARBOHYDRATES FLOW AND CANOPY TEMPERATURE

According to De Koning [37], the tomato's flowering rate per unit of time is linearly related to temperature. Equation (4.7) represents the temperature effect on structural carbon flow to organs, in other words, it expresses the growth dependency on temperature.

$$g_{T_{Can_{24}}} = 0.047 \cdot T_{Can_{24}} + 0.06 \tag{4.7}$$

The analytical formulation of the growth inhibition functions $h_{T_{Can}}$ and $h_{T_{Can_{24}}}$ is given in Equations (4.8) and (4.9) respectively.

$$h_{T_{Can}} = \frac{1}{1 + e^{-0.869 \cdot (T_{Can} - p_{T_{Can_{Min}}})}} \cdot \frac{1}{1 + e^{0.5793 \cdot (T_{Can} - p_{T_{Can_{Max}}})}} \tag{4.8}$$

$$h_{T_{Can_{24}}} = \frac{1}{(1 + e^{-1.1587 \cdot (T_{Can_{24}} - p_{T_{Can_{24min}}})})} \cdot \frac{1}{(1 + e^{1.3904 \cdot (T_{Can_{24}} - p_{T_{Can_{24Max}}})})} \tag{4.9}$$

Where $p_{T_{Can_{Min}}}$ and $p_{T_{Can_{Max}}}$ the inhibition of carbohydrate flows because of low and high instantaneous temperatures in ^oC accordingly, $p_{T_{Can_{24min}}}$ and $p_{T_{Can_{24Max}}}$ the inhibition of carbohydrate flow because of low and high temperatures in ^oC.

$$h_{T_{CanSum}} = \frac{1}{2}\cdot\Big(\frac{1}{T^{End}_{CanSum}}T_{CanSum} + \sqrt{\Big|\frac{1}{T^{End}_{CanSum}}T_{CanSum}\Big|^2 + 1\cdot 10^{-4}}\Big) - \frac{1}{2}\cdot\Big(\frac{1}{T^{End}_{CanSum}}(T_{CanSum} - T^{End}_{CanSum}) + \sqrt{\Big|\frac{1}{T^{End}_{CanSum}}(T_{CanSum} - T^{End}_{CanSum})\Big|^2 + 1\cdot 10^{-4}}\Big) \quad (4.10)$$

Where T_{SumEnd} the temperature sum where crop is fully generative in ^oC.

4

4.4.2. TEMPERATURE INDUCED MAXIMIZATION IN CARBOHYDRATES FLOW

One of the main objectives of the climate controller is to maintain the flow of carbohydrates from the buffer to the organs on a high level. The introduction of Equations (4.7)-(4.10) clarifies the impact of canopy temperature on carbohydrate distribution.

According to Figure 4.5 the maximum flow rate of carbohydrates can be achieved under a certain canopy and mean canopy temperature. Particularly, the canopy temperature has to be maintained in its optimal range between 14 and 27^oC degrees. In terms of mean canopy temperature, the maximum distribution rate is achieved for $T_{Can_{24}} = 19.88^oC$.

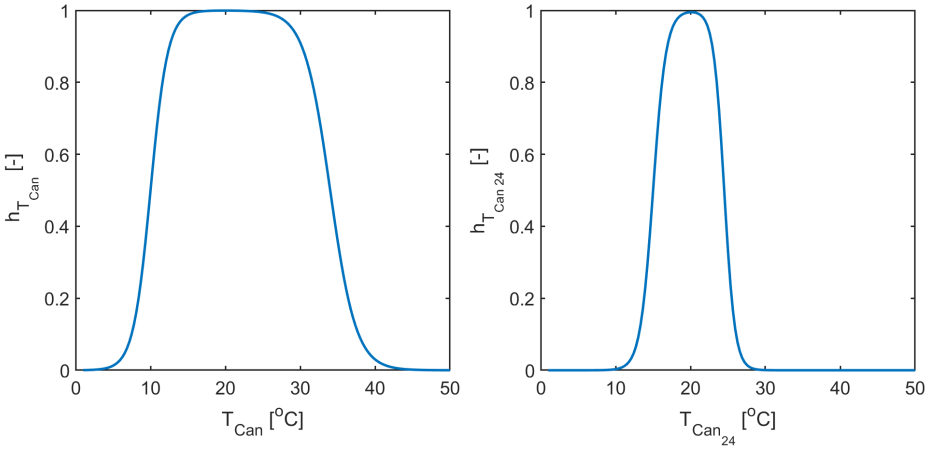

Figure 4.5: Growth inhibition function due to instantaneous canopy temperature ($h_{T_{Can}}$) and mean canopy temperature ($h_{T_{Can_{24}}}$)

The canopy temperature requirements can be incorporated into the dynamic optimization as inequality constraints. The optimal point of the mean canopy temperature can be ensured through the reference tracking term below:

$$\min_u ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 \equiv \min_u ||19.88 - T_{Can_{24}}||_2^2 \tag{4.11}$$

The inhibition due to the generative phase ($h_{T_{CanSum}}$) starts at zero and increases linearly to full potential with increasing temperature sum as presented in Equation (4.12). Therefore, it is a matter of time to become maximum and it was chosen to not be included in the objective function for the sake of simplicity.

$$h_{T_{CanSum}} = \begin{cases} 0 & \text{if } T_{CanSum} \leq 0 \\ \frac{T_{CanSum}}{T^{End}_{CanSum}} & \text{if } 0 \geq T_{CanSum} \leq T^{End}_{CanSum} \\ 1 & \text{if } T_{CanSum} \geq T^{End}_{CanSum} \end{cases} \tag{4.12}$$

4.5. MONITORING CARBOHYDRATES LOSSES

According to the common structure depicted in Figure 4.1, some of the produced carbohydrates are distributed to the organs and the rest are consumed through growth and maintenance respiration. Respiration is an extremely important process as it consumes carbohydrates to guarantee the survival of the crop. However, under non-optimal environmental conditions, respiration consumes more carbohydrates than needed for crop maintenance. In this way, respiration becomes a loss as it utilizes carbohydrates that could have been distributed to the organs and contributed to the final yield. This case can be prevented through proper climate control which demands a good understanding of both growth and maintenance respiration processes.

4.5.1. GROWTH RESPIRATION

The growth respiration MC_{BufAir} is the integration of the growth respiration of the individual plant organs:

$$MC_{BufAir} = MC_{FruitAir_g} + MC_{LeafAir_g} + MC_{StemAir_g} \tag{4.13}$$

In order to fit growth respiration in the common structure formulation (Figure 4.1), it has to be expressed in $mg_{CH_2O} \cdot m^2 \cdot s^{-1}$. The growth respiration of the individual organs is linearly related to the carbohydrate flow from the buffer to the corresponding organ as follows

$$M_{FruitAir_g} = c_{Fruit_g} MC_{BufFruit} \tag{4.14}$$

$$M_{LeafAir_g} = c_{Leaf_g} MC_{BufLeaf} \tag{4.15}$$

$$M_{StemAir_g} = c_{Stem_g} MC_{BufStem} \tag{4.16}$$

where c_{Fruit_g}, c_{Leaf_g} and c_{Stem_g} are unitless growth respiration coefficients for the fruits, leaves, and stems respectively. The coefficient values can be calculated based on the assimilate requirements for the formation of fruits, leaves, and stems presented in [38].

4.5.2. MAINTENANCE RESPIRATION

A detailed presentation of the individual maintenance respiration $MC_{FruitAir}$, $MC_{LeafAir}$ and $MC_{StemAir}$ is introduced in [38]. Maintenance respiration can be defined as

$$MC_{Org(i)Air} = c_{Org(i)_m} \cdot Q_{10_m}^{0.1(T_{can_{24}}-25)} \cdot C_{Org(i)} \cdot (1 - e^{-c_{RGR} \cdot RGR}) \tag{4.17}$$

where $Org(i) = \{Fruit,\ Leaf,\ Stem\}$ for $i = 1,2,3$ representing the different plant organs, $c_{Org(i)_m}$ is the maintenance respiration coefficient of the corresponding plant organ in $mg_{CH_2O} \cdot mg_{CH_2O}^{-1} \cdot s^{-1}$, Q_{10_m} is the Q_{10} value for temperature effect on maintenance respiration, $C_{Org(i)}$ the carbohydrate weight of plant organ, RGR the net relative growth rate in s^{-1} and c_{RGR} the regression coefficient for maintenance respiration in s.

4.5.3. TEMPERATURE INDUCED MINIMIZATION OF CARBOHYDRATES LOSSES

As presented in Equations (4.14) - (4.16) the growth respiration is linearly related to the carbohydrates flow from the buffer to the organs. The minimization of growth respiration entails the minimization of the distribution flow which is unacceptable because of the correlation between assimilates partitioning and yield. Consequently, the climate controller should not affect growth respiration.

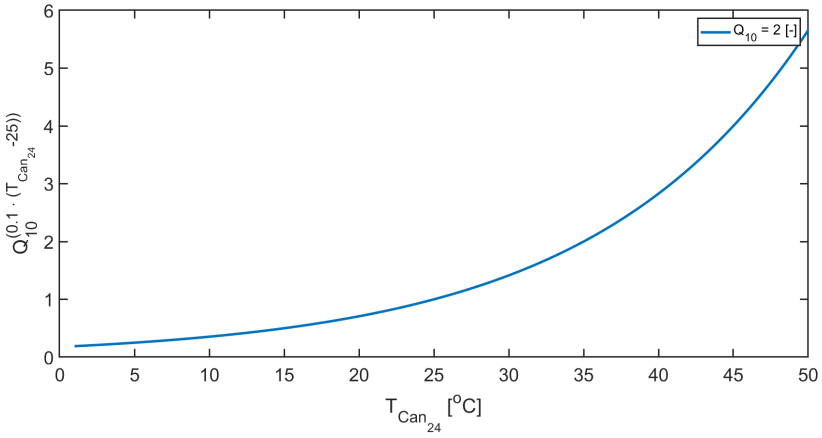

Figure 4.6: The role of the mean canopy temperature on maintenance respiration.

On the other hand, maintenance respiration can be affected through the exponential factor including the mean canopy temperature. As shown in Figure 4.6, the exponential factor $f(T_{Can_{24}}) = Q_{10}^{0.1 \cdot (T_{Can_{24}}-25)}$ is monotonically increasing, therefore maintenance respiration is minimized for minimum mean canopy temperature. The minimization of maintenance respiration can be introduced in the objective function as below:

$$\min_{u} ||T_{Can_{24}}||_2^2 \tag{4.18}$$

4.6. PROOF OF CONCEPT - CARBOHYDRATES MONITORING

According to the previous sections, canopy temperature plays a major role in the production, distribution, and consumption of carbohydrates by the crop. This section is introduced to prove the concept that these three processes can be manipulated through a SPA-based predictive controller. The proof of concept incorporates two simulated experiments where the inputs of a tomato crop model are defined by two predictive controllers. The first controller bases its objective function on SPA embedding the terms presented in the previous sections. On the other hand, the second controller acts as a reference for comparison. In this case, the controller defines the future crop inputs that would maximize the number of carbohydrates in the fruits.

The Vanthoor [20] tomato crop model is chosen to be used as ground truth. The tomato crop model receives the following input vector:

$$u = \begin{bmatrix} T_{Can} \\ R_{Par\_SunCan} \\ CO_{2_{Ppm}} \\ VP_{Air} \end{bmatrix} \tag{4.19}$$

Where T_{Can} the canopy temperature, R_{Par_SunCan} the PAR radiation from the sun reaching the canopy, $CO_{2_{Ppm}}$ the CO_2 concentration and VP_{Air} the vapor pressure in air. The state vector of the crop model is presented by Equation (3.2). The differential equations defining the model's states have been introduced in Sections 3.3.1 and 3.3.5.

For the sake of simplicity and in order to demonstrate the impact of canopy temperature on crop yield, the simulated controllers make decisions only on the future T_{Can} values. The rest inputs are considered exogenous and their values are given through a recorded dataset.

4.6.1. SPA PREDICTIVE CONTROLLER

In this case, a predictive controller defines the future canopy temperature trajectory based on the minimization of a switching objective function. During the daytime, the controller has to minimize the stomatal resistance and the mean canopy temperature while tracking the optimal mean canopy temperature. Revoking the role of stomatal resistance becomes clear that its minimization targets to increase carbohydrate production. However, during the night time, no photosynthesis is taking place and as a result, stomatal resistance can be removed from the night part of the objective function. During the nighttime, the controller has to maintain a slow rate of carbohydrate distribution ensuring that the buffer will never be full. In this way, the controller prevents the accumulation of non-distributed carbohydrates in the buffer that could result in pausing photosynthesis during sunrise. Moreover, the monitoring of maintenance respiration becomes more important during the night because of the lack of photosynthesis. Hence, the nighttime control requires different weights from the daytime control as its objectives differ. The mathematical formulation of the SPA predictive controller is given

below:

$$
\begin{aligned}
\min_{T_{Can}} \quad & (C_0 \cdot ||p_1 \cdot e^{\frac{p_3 \cdot T_{Can}}{T_{Can}+p_2}} - VP_{Air}||_2^2 + C_1 \cdot ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 + C_2 \cdot ||T_{Can_{24}}||_2^2) \cdot D + \\
& (C_3 \cdot ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 + C_4 \cdot ||T_{Can_{24}}||_2^2) \cdot (1 - D) \\
\text{s.t.} \quad & \\
& 14 \leq T_{Can}(k) \leq 27, \ \forall k \in \{1, \dots, N\} \\
& T_{Can_{24}}(k+1) = \frac{1}{\tau}(k \cdot T_{Can}(k) - T_{Can_{24}}(k)), \ \forall k \in \{1, \dots, N\} \\
& D = \begin{cases} 1, & \text{if } \exists i \in \{1, \dots, N\} \text{ for which } R_{Can_i} > 0 \\ 0, & \text{if } \forall i \in \{1, \dots, N\} \ R_{Can_i} = 0 \end{cases} \\
& 0.3 \leq \Delta T_{Can} \leq 0.3
\end{aligned}
\tag{4.20}
$$

Where C_i for $i \in \{0, \dots, 4\}$ the weights of the individual terms and specifically $C_0 = 10^{-10}$, $C_1 = 10^3$, $C_2 = 10^{-8}$, $C_3 = 4 \cdot 10^{-2}$ and $C_4 = 10^4$. N the prediction horizon, $T_{Can_{24}} = 19.88^o C$ the reference mean canopy temperature, D a boolean variable declaring the Day and Night time, $\tau = 86400 sec$ the time constant of the process and $k = 1$ the gain of the process.

4.6.2. MAXIMIZATION OF CARBOHYDRATES IN FRUITS

A reference experiment is required for the evaluation of the SPA controller. To ensure a fair comparison between the two controllers, the reference should express the potential of the final crop yield. For this reason, the optimization problem calculates the future canopy temperature trajectory concluding in the maximum value of carbohydrates in the fruits at the end of the prediction horizon.

$$
\begin{aligned}
\min_{T_{Can}} \quad & -C_{Fruit}(N) \\
\text{s.t.} \quad & \\
& 14 \leq T_{Can}(k) \leq 27, \ \forall k \in \{1, \dots, N\} \\
& x(k+1) = \begin{bmatrix} T_{Can_{24}}(k+1) \\ T_{Can_{Sum}}(k+1) \\ C_{Buf}(k+1) \\ C_{Fruit}(k+1) \\ C_{Leaf}(k+1) \\ C_{Stem}(k+1) \end{bmatrix} = \\
& \qquad f_c([T_{Can}(k) \ R_{Par\_SunCan}(k) \ CO_{2_{Ppm}}(k) \ VP_{Air}(k)], \ x(k), \ k) \\
& 0.3 \leq \Delta T_{Can} \leq 0.3
\end{aligned}
\tag{4.21}
$$

Where N is the prediction horizon and f_c denotes the tomato crop model incorporated in the Greenlight model. In other words, f_c is the Vanthoor tomato crop model also used as ground truth in this experiment.

4.6.3. SIMULATION

Both experiments were simulated in MATLAB using TOMLAB for the solution of the optimization problems. In particular, the snopt solver was used in both cases because of their nonlinear nature. The optimization problems were solved for a prediction horizon (N) of 1 hour and a control horizon (N_c) of 0.5 hours. The sampling time used was $5min$, thus $N = 12$ and $N_c = 6$ time intervals. Moreover, the experiments start with the same initial conditions of a mature crop and run for 60 days.

The SPA-based controller concluded with a final yield of 26.6772 kilos of tomatoes per square meter. On the other hand, the yield maximization controller achieved a harvest of 26.9558 $kg \cdot m^{-2}$ which can be considered as the maximum achievable yield. In comparison, the SPA approach concluded on 1.0335% less yield than the reference problem. This percentage can be justified by the fact that the reference problem incorporates the ground truth crop model, including the photosynthesis model. As a result, the reference controller can choose the canopy temperature trajectory maximizing the photosynthetic rate in contrast with the SPA-based controller which approximates photosynthesis through stomatal resistance.

In terms of performance, the mean execution time per iteration of the SPA controller is $0.1746sec$ with a minimum execution time of $0.0408sec$ and a maximum time of $0.8305sec$. The reference controller has a mean execution time per iteration equal to $1.7572sec$ with $0.4665sec$ for the fastest and $14.2035sec$ for the slowest case respectively. The increased execution time of the reference controller has to do with the fact that the crop model runs several times in each iteration for the prediction of the future canopy temperature trajectories. The overview of the simulation results in terms of carbohydrates production, distribution, losses, and execution time can be found in Table 4.1.

Table 4.1: Simulation results of the proof-of-concept experiment for the comparison of the SPA controller and the fruit maximization algorithm

	Fruit Max. Controller	SPA controller	Units
Produced CH_2O	4.10947	4.0526	$kg_{CH_2O} \cdot m^{-2}$
Distributed CH_2O to Fruits	2.1228	2.0906	$kg_{CH_2O} \cdot m^{-2}$
Losses CH_2O	2.2299	2.0735	$kg_{CH_2O} \cdot m^{-2}$
Yield (Wet)	26.9558	26.6772	$kg_{tomato} \cdot m^{-2}$
Mean Exec.Time	1.7572	0.1746	sec

4.7. OBJECTIVE FUNCTION - FINAL FORM

The previous section showed that the proposed SPA control approach is capable to monitor carbohydrates production, distribution, and losses achieving high fruit harvest. How-

ever, in a real-world scenario, the coupled climate-crop system has to be taken into consideration. According to this fact, the climate controller should be able to track the carbon dioxide concentration in the air. Moreover, high yield does not make sense if it comes with extremely high cost, hence growers require the monitoring of energy resources by the climate controller, especially in the context of the current energy crisis. Therefore, modern climate controllers should be capable to steer the system's production on the optimal yield-cost ratio. The fulfillment of the aforementioned requirements orders the introduction of more terms in the objective function.

As explained in Section 4.3, the proposed SPA controller is not able to generate meaningful CO_2 setpoint values. For this reason, it was decided to introduce a reference tracking term for CO_2 control where the CO_2 setpoints are defined by the grower. This reference tracking term is described in Equation (4.22) where CO_2^* represents the optimal carbon dioxide concentration in the greenhouse air.

$$\min_u ||CO_2^* - CO_2||_2^2 \tag{4.22}$$

Conventionally, greenhouse heating, cooling, shading, and thermal insulation systems are considered as the greenhouse actuators. The greatest energy load of a Dutch Venlo-type greenhouse is demanded by the heating system, thus the minimization of energy consumption entails the minimization of the heating system use. The greater the input of the heating system, the greater the energy consumption over time as the system's input can be deduced to the burning rate of fossil fuel. On the other hand, greenhouse actuators like windows and thermal screens do not require an energy load according to their opening/deployment. In other words, windows and thermal screen deployment is static as energy is required only when the level of opening/deployment has to change. Regarding these observations, the greenhouse energy consumption can be monitored through the application of proper weights on the terms demonstrated in Equation (4.23).

$$\min_u \sum_{k=0}^{N-1} ||u_k||_R + ||u_k - u_{k-1}||_{R_1} \tag{4.23}$$

Where u the control input vector, $R \succcurlyeq 0$ the quadratic cost of the control inputs and $R_1 \succcurlyeq 0$ the quadratic cost of the differential control inputs. A more analytical presentation of the energy consumption terms and their quadratic cost is carried out in Chapter 6.
Summarising, the proposed objective function able to monitor the carbohydrates flow in the crop while achieving the reference tracking of the CO_2 concentration and attaining a reasonable trade-off between the energy consumption and final yield is displayed below:

$$\begin{aligned}\min_u \quad & (\sum_{k=0}^{N-1} ||u_k||_{R_{\Delta_k}} + ||u_k - u_{k-1}||_{R_1} + C_0 \cdot ||p_1 \cdot e^{\frac{p_3 \cdot T_{Can}}{T_{Can}+p_2}} - VP_{Air}||_2^2 \\ & + C_1 \cdot ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 + C_2 \cdot ||T_{Can_{24}}||_2^2 + C_3 \cdot ||CO_2^* - CO_2||_2^2) \cdot D + P + \\ & (\sum_{k=0}^{N-1} ||u_k||_R + ||u_k - u_{k-1}||_{R_1} + C_4 \cdot ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 + C_5 \cdot ||T_{Can_{24}}||_2^2) \cdot (1 - D)\end{aligned} \tag{4.24}$$

Where $R_{\Delta_k} \succcurlyeq 0$ and $R \succcurlyeq 0$ the quadratic cost of the control inputs for the day and the night time respectively, and P a penalty function incorporating all the necessary inequality constraints regarding the climate conditions.

5

DATA-DRIVEN SYSTEM REPRESENTATION

5.1. INTRODUCTION

Greenhouses are complex, non-linear systems and their study in the physical world requires time and resources. For this reason, various greenhouse representation schemes have been introduced in the literature through the years. Greenhouse system representations are valuable tools for the analysis of greenhouse systems, the design of new climate control regimes, and the simulation of scientific concepts. They are means of permitting the preliminary validation of an idea before proceeding with an expensive real-world experiment. However, the complexity of such a system and the interconnectivity between its subsystems, make parametric greenhouse modeling a hard task. Therefore, it would be interesting to examine if and how the coupled climate-crop system could be represented using data.

This chapter explores if the mechanistic greenhouse model presented in Chapter 3 can be described adequately by a data-driven linear representation method originally introduced in [11]. As distinct from the classical systems theory, the method subjected in question does not rely on a particular parametric system representation but it aims to describe the system in terms of its behavior. The system's behavior can be learned by applying a sufficiently rich and long input signal, capable of exciting the system and producing a representative output sequence. Section 5.2 is introduced to clarify the needed mathematical concepts and the conditions that have to be met by the input-output dataset to efficiently describe the corresponding dynamical system. Section 5.3 presents the methodology behind the implementation of the system's non-parametric representation. Section 5.4 displays the results of the in question data-driven representation of the ground truth coupled climate-crop model. Finally, Section 5.5 tests if and how the data-driven predictor could be used for the representation of the canopy and the mean canopy temperature behavior.

5.2. PRELIMINARIES

5.2.1. HANKEL MATRIX

Hankel matrices are of great importance to the concept of non-parametric system representation as they contain the input-output dataset defining the system's behavior. The mathematical definition of a Hankel matrix is given below.

Definition 1: Assume a given signal $z : \mathbb{Z} \rightarrow \mathbb{R}^{\sigma}$, the vectorized form of the signal z to the interval $[k,\ k+T] \cap \mathbb{Z}$ where $k \in \mathbb{Z},\ T \in \mathbb{N}$ is

$$z_{[k,\ k+T]} = \begin{bmatrix} z(k) \\ \vdots \\ z(k+T) \end{bmatrix}$$

The Hankel matrix associated with z is denoted as

$$Z_{i,t,N} = \begin{bmatrix} z(i) & z(i+1) & \dots & z(i+N-1) \\ z(i+1) & z(i+2) & \dots & z(i+N) \\ \vdots & \vdots & \ddots & \vdots \\ z(i+t-1) & z(i+t) & \dots & z(i+t+N-2) \end{bmatrix}$$

Where $i \in \mathbb{Z}$ and $t, N \in \mathbb{N}$. Specifically, i denotes the time at which the first sample is taken, t is the number of samples per column and N is the number of samples per row.

5.2.2. PERSISTENCY OF EXCITATION

As will be shown in the following subsections, persistency of excitation (PE) is a strong condition, necessary for the unique identification of the underlying dynamical system from data.

Definition 2: The signal $z_{[0,T-1]} \in \mathbb{R}^{\sigma}$ is persistently exciting of order L if the matrix

$$Z_{0,L,T-L+1} = \begin{bmatrix} z(0) & z(1) & \dots & z(T-L) \\ z(1) & z(2) & \dots & z(T-L+1) \\ \vdots & \vdots & \ddots & \vdots \\ z(L-1) & z(L) & \dots & z(T-1) \end{bmatrix}$$

has full row rank $\sigma \cdot L$. Another condition is that the signal must be sufficiently long and specifically $T \geq (\sigma+1) \cdot L - 1$.

5.2.3. BEHAVIORAL THEORY

NON-PARAMETRIC SYSTEM REPRESENTATION

Behavioral Theory is a general and comprehensive way to define a dynamical system. In Willem's behavioral theory, the definition of a dynamical system has three main components determining the subspace of the signal space in which the system's trajectories live. Hence, behavioral system theory is able to describe the properties of a dynamical system independent of any particular parametric system representation. Following [39]

the dynamical system and its behavioral properties are defined as:

Definition 3: A dynamical system is a 3-tuple $(\mathbb{Z}_{\geq 0}, \mathbb{W}, \mathscr{B})$ where $\mathbb{Z}_{\geq 0}$ is the discrete-time axis, $\mathbb{W}$ is a signal space and $\mathscr{B} \subseteq \mathbb{W}^{\mathbb{Z}_{\geq 0}}$ is the behavior.

Definition 4: Let $(\mathbb{Z}_{\geq 0}, \mathbb{W}, \mathscr{B})$ be a dynamical system.
i) $(\mathbb{Z}_{\geq 0}, \mathbb{W}, \mathscr{B})$ is linear if $\mathbb{W}$ is a vector space and $\mathscr{B}$ is a linear subspace of $\mathbb{W}^{\mathbb{Z}_{\geq 0}}$.
ii) $(\mathbb{Z}_{\geq 0}, \mathbb{W}, \mathscr{B})$ is time invariant if $\mathscr{B} \subseteq \sigma\mathscr{B}$ where $\sigma : \mathbb{W}^{\mathbb{Z}_{\geq 0}} \to \mathbb{W}^{\mathbb{Z}_{\geq 0}}$ is the forward time shift defined by $(\sigma w)(t) = w(t+1)$ and $\sigma\mathscr{B} = \{\sigma w | w \in \mathscr{B}\}$
(iii) $(\mathbb{Z}_{\geq 0}, \mathbb{W}, \mathscr{B})$ is complete if $\mathscr{B}$ is closed in the topology of pointwise convergence.

The class of systems $(\mathbb{Z}_{>0}, \mathbb{R}^{n_u+n_y}, \mathscr{B})$ satisfying the conditions (i)-(iii) is denoted by $\mathscr{L}^{n_u+n_y}$ with $n_u, n_y \in \mathbb{Z}_{>0}$.

Definition 5: A system $\mathscr{B} \in \mathscr{L}^{n_u+n_y}$ is controllable if for every $T \in \mathbb{Z}_{>0}$, $w^1 \in \mathscr{B}_T$, $w^2 \in \mathscr{B}$ there exists $w \in \mathscr{B}$ and $T' \in \mathbb{Z}_{>0}$ such that $w_t = w'_t$ for $1 \leq t \leq T$ and $w_t = w^2_{t-T-T'}$ for $t > T + T'$.
Definition 6: Let $L, T \in \mathbb{Z}_{>0}$ such that $T \geq L$. The signal $u = \text{col}(u_1, \ldots, u_T) \in \mathbb{R}^{T \cdot n_u}$ is persistently exciting of order L if the Hankel matrix

$$\mathscr{H}_L(u) := \begin{bmatrix} u_1 & \ldots & u_{T-L+1} \\ \vdots & \ddots & \vdots \\ u_L & \ldots & u_T \end{bmatrix}$$

is of full row rank.

5

PARAMETRIC SYSTEM REPRESENTATION

A behavioral system could also be expressed in a parametric representation in multiple, equivalent ways. The classical state space representation (5.2) of a behavioural system $\mathscr{B} \in \mathscr{L}^{n_u+n_y}$ is denoted by $\mathscr{B}(A,B,C,D) = \{\text{col}(u,y) \in (\mathbb{R}^{n_u+n_y})^{\mathbb{Z}_{\geq 0}} | \exists\, x \in (\mathbb{R}^n)^{\mathbb{Z}_{\geq 0}}$ s.t. $\dot{x} = Ax + Bu,\ y = Cx + Du\}$. The state space representation of the smallest order is called a *minimal representation* of the system and the minimum order is denoted by $\mathbf{n}(\mathscr{B})$. An additional property that should be noticed is the *lag*. The *lag* $\mathbf{l} \in \mathbb{Z}_{>0}$ of a system $\mathscr{B} \in \mathscr{L}^{n_u+n_y}$ is the smallest non-negative integer for which the observability matrix $\mathscr{O}_l := \text{col}(C, CA, \ldots, CA^{l-1})$ has rank $\mathbf{n}(\mathscr{B})$. Henceforward, the lag is denoted as $\mathbf{l}(\mathscr{B})$. Eventually, the lower triangular Toeplitz matrix consisting of A, B, C, D is denoted by

$$\mathscr{T}_N(A,B,C,D) := \begin{bmatrix} D & 0 & \ldots & 0 \\ CB & D & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ CA^{N-2}B & \ldots & CB & D \end{bmatrix}$$

At this point, a uniqueness *Lemma* can be presented.

Lemma 1: Let $\mathscr{B} \in \mathscr{L}^{n_u+n_y}$ and $\mathscr{B}(A,B,C,D)$ a minimal input/output/state representation. Let $T_{ini}, N \in \mathbb{Z}_{>0}$ with $T_{ini} \geq \mathbf{l}(\mathscr{B})$ and $\text{col}(u_{ini}, u, y_{ini}, y) \in \mathscr{B}_{T_{ini}+N}$. Then there exists a unique $x_{ini} \in \mathbb{B}^{\mathbf{n}(\mathscr{B})}$ such that:

$$y = \mathscr{O}_N(A,C)x_{ini} + \mathscr{T}_N(A,B,C,D)u \tag{5.1}$$

Particularly, for a sufficiently long window of input-output data $\text{col}(u_{ini}, y_{ini})$, the state to which the system is driven by the u_{ini} sequence of inputs, is unique.

5.2.4. WILLEM'S FUNDAMENTAL LEMMA

Consider a system with the following representation.

$$\begin{aligned} x(k+1) &= A \cdot x(k) + B \cdot u(k) \\ y(k) &= C \cdot x(k) + D \cdot u(k) \end{aligned} \tag{5.2}$$

Where $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$, $y \in \mathbb{R}^{n_y}$ and $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, $C \in \mathbb{R}^{n_y \times n_x}$ and $D \in \mathbb{R}^{n_y \times n_u}$ and assume that the represented Linear Time-Invariant (LTI) system is controllable and observable.

Given an input-output dataset $(u_d[0,\ T-1],\ y_d[0,\ T-1])$ let the corresponding Hankel matrices be

$$\begin{bmatrix} U_{0,t,T-t+1} \\ Y_{0,t,T-t+1} \end{bmatrix} = \left[\begin{array}{cccc} u_d(0) & u_d(1) & \dots & u_d(T-t) \\ u_d(1) & u_d(2) & \dots & u_d(T-t+1) \\ \vdots & \vdots & \ddots & \vdots \\ u_d(t-1) & u_d(t) & \dots & u_d(T-1) \\ \hline y_d(0) & y_d(1) & \dots & y_d(T-t) \\ y_d(1) & y_d(2) & \dots & y_d(T-t+1) \\ \vdots & \vdots & \ddots & \vdots \\ y_d(t-1) & y_d(t) & \dots & y_d(T-1) \end{array}\right] \tag{5.3}$$

The matrix $X_{0,T-t+1} = [x_d(0)\ x_d(1)\ \dots\ x_d(T-t)]$ contains the state samples $x_d(i)$ produced by the system (5.2) for the input sequence $u_d[0, T-1]$. The rank condition presented in equation (5.4) plays a great role in the representation of the system through data. However, in general, it is difficult to examine the validity of the rank condition as the system's states may not be measurable and only the input-output data are accessible. To surpass this barrier, the next two *Lemmas* are introduced.

$$rank \begin{bmatrix} U_{0,t,T-t+1} \\ X_{0,T-t+1} \end{bmatrix} = n_x + t \cdot n_u \tag{5.4}$$

Lemma 2: ([40], Corollary 2)
If $u_d[0, T-1]$ is persistently exciting of order $n_x + t$ then the rank condition (Eq. (5.4))

holds.

Lemma 3: ([40], Theorem 1)
i) if $u_d[0, T-1]$ is persistently exciting of order $n_x + t$, then any t-long input-output trajectory of system (5.2) can be expressed as

$$\begin{bmatrix} u_{[0,t-1]} \\ y_{[0,t-1]} \end{bmatrix} = \begin{bmatrix} U_{0,t,T-t+1} \\ Y_{0,t,T-t+1} \end{bmatrix} \cdot g \tag{5.5}$$

where $g \in \mathbb{R}^{T-t+1}$
ii) Given a T-long input-output trajectory of the system (5.2), any linear combination of the columns of the matrix in (5.3) that is

$$\left[\begin{array}{c} U_{0,t,T-t+1} \\ \hline Y_{0,t,T-t+1} \end{array}\right] \cdot g$$

is a t-long input-output trajectory of (5.2).

The original proof of *Lemma 3* can be found in [40]. *Lemma 3* is referred as the *fundamental lemma* using a finite input-output dataset to describe a linear system.

5.3. NON-PARAMETRIC SYSTEM REPRESENTATION - METHODOLOGY

The composition of the non-parametric system representation is based on two main processes, the data collection and the construction of the data-driven predictor. The system's representation relies on the past input-output data of the original system. Particularly, the recorded input-output dataset captures the system's behavior. Then, the system's behavior is imprinted in the form of Hankel matrices. These Hankel matrices are then used in combination with the input data recorded right before time zero for the estimation of the initial conditions and the prediction of the future output trajectories. The following subsections guide the reader through the steps of data collection and the construction of the data-driven predictor concluding on the final non-parametric system representation.

DATA COLLECTION

Assume an unknown controllable LTI system $\mathscr{B} \in \mathscr{L}^{n_u+n_y}$ with minimal representation $\mathscr{B}(A,B,C,D)$. Generate an offline input sequence $u^d = col(u_1^d, \cdots, u_T^d) \in \mathbb{R}^{T \cdot n_u}$ of length $T \in \mathbb{Z}_{>0}$, apply it to the system and measure the outputs $y^d = col(y_1^d, \cdots, y_T^d) \in \mathbb{R}^{T \cdot n_y}$. It is of great importance to ensure that the recorded dataset is capable of describing the behavior of the system. For this reason, the input sequence u^d has to be persistently exciting of order $T + \mathbf{n}(\mathscr{B})$ to satisfy Willem's fundamental lemma. Furthermore, the condition of persistency of excitation requires a long enough data sequence which, in this case, should satisfy the inequality:

$$T \geq (n_u + 1)(T + \mathbf{n}(\mathscr{B})) - 1 \tag{5.6}$$

The input-output data have to be partitioned into past (subscript p) and future data (subscript f). The past data, of length $T_{ini} \in \mathbb{Z}_{>0}$, are needed for the estimation of the initial conditions, and the future data, of length $N \in \mathbb{Z}_{>0}$, for the estimation of the future control input as shown in Equation (5.7). Hence, inequality (5.6) can be reformulated as $T \geq (n_u + 1)(T_{ini} + N + \mathbf{n}(\mathscr{B})) - 1$. Finally, ensure that $T_{ini} \geq \mathbf{l}(\mathscr{B})$ in order to satisfy the *Lemma* 2 and be sure that a unique $x_{ini} \in \mathbb{R}^{n(\mathscr{B})}$ exists.

$$\begin{pmatrix} U_p \\ U_f \end{pmatrix} := \mathscr{H}_T(u^d) = \mathscr{H}_{T_{ini}+N}(u^d), \quad \begin{pmatrix} Y_p \\ Y_f \end{pmatrix} := \mathscr{H}_T(y^d) = \mathscr{H}_{T_{ini}+N}(y^d) \tag{5.7}$$

DATA-DRIVEN PREDICTOR

A recent input-output sequence $\text{col}(u_{ini}, y_{ini})$ of length T_{ini}, which will be used to estimate the initial conditions, has to be recorded online. From Willem's fundamental lemma, any trajectory of length $T_{ini} + N$, of the $\mathscr{B}_{T_{ini}+N}$ system can be constructed using the recorded data sequences. Specifically, a trajectory $\text{col}(u_{ini}, u, y_{ini}, y)$ belongs to $\mathscr{B}_{T_{ini}+N}$ if and only if there exists $g \in \mathbb{R}^{T-T_{ini}-N+1}$ such that:

$$\left[\begin{array}{c} U_p \\ Y_p \\ U_f \\ \hline Y_f \end{array}\right] \cdot g = \left[\begin{array}{c} u_{ini} \\ y_{ini} \\ u \\ \hline y \end{array}\right] \tag{5.8}$$

If $T_{ini} \geq \mathbf{l}(\mathscr{B})$ a unique $x_{ini} \in \mathbb{R}^{\mathbf{n}(\mathscr{B})}$ exists based on *Lemma* 1. Then, given an input trajectory u of length N, $g \in \mathbb{R}^{T-T_{ini}-N+1}$ can be calculated from the first three equations of (5.8). Finally, the future output trajectory y can be predicted based on the prerecorded data in Y_f and the estimated g as $y = Y_f \cdot g$. Vice versa, given a reference output trajectory y a feedforward control input u can be calculated.

5.4. EXPERIMENT: CLIMATE-CROP SYSTEM REPRESENTATION

Testing the hypothesis that the ground truth coupled climate-crop model can be approximated by the displayed non-parametric system representation requires the possession of input-output data. The Greenlight model, in its default code configuration, is delivered including a sigmoid P climate controller for the control of the simulated environmental conditions. By default, it is set up to maintain the relative humidity below 87%, the CO_2 at 1000 ppm during the daytime, and the air temperature at 18.5^oC and 19.5^oC during the night and day time accordingly. The analytical overview of the climate controller can be found in Section 2.3.2 of [41]. Based on this configuration and the introduction of a weather dataset representing the weather in Amsterdam (exogenous inputs) a 60days experiment was simulated. It should be noted that the simulation started with a mature crop. The data that occurred by this simulation were recorded in order to test the non-parametric system representation.

The first $T + 2 \cdot T_{ini} + N$ were isolated from the dataset as depicted in Figure 5.1. The selection of T, T_{ini} and N values was completed in an experimental way where the values producing the best prediction accuracy were kept. For the rest of this section, it can be considered that $T = 4320$, $T_{ini} = 5$ and $N = 12$ *entries* where the sampling time between entries is 5 *min*. Each data segment contains all input u, exogenous input d and output/state $x = [x_g^T \; x_x^T]^T$ data as described in Equations (3.6), (3.5), (3.1) and (3.2) accordingly.

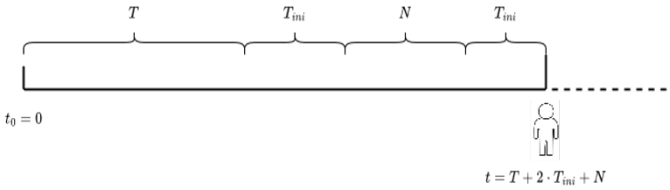

Figure 5.1: Initial data division in four segments with lengths T, T_{ini}, N and T_{ini} respectively

5

Before proceeding with the construction of Hankel Matrices, it should be ensured that the length of the data segment T is sufficiently long. Thus, the condition $T \geq (n_u + 1)(T_{ini} + N + \mathbf{n}(\mathscr{B})) - 1$ is checked for $n_u = 8$ and $\mathbf{n}(\mathscr{B}) = 23$ concluding on a positive result.

5.4.1. STATIC HANKEL MATRICES

The data living in the first data segment of length T were used for the construction of the Hankel matrices $U \in \mathbb{R}^{n_u \cdot (T_{ini}+N) \times (T-L)}$, $V \in \mathbb{R}^{n_d \cdot (T_{ini}+N) \times (T-L)}$ and $Y \in \mathbb{R}^{n_x \cdot (T_{ini}+N) \times (T-L)}$ corresponding to the inputs, exogenous inputs, and outputs/states. The next step is to examine the persistency of excitation (PE) of the recorded data by testing *Definition 6* on the Hankel matrix containing all inputs (control and exogenous). It should be mentioned that for any tested T value, the full input dataset is not PE. As a result, there is no theoretical validation that the ground truth climate-crop system can be linearly approximated.

The length of time segments T_{ini} and N were utilized for the proper partitioning of the Hankel matrices into past (subscript p) and future data (subscript f) as shown in Equation (5.7).

The next step was the definition of the initial data u_{ini}, v_{ini}, and y_{ini} for the calculation of the initial conditions. The data with the ini subscript live in the first T_{ini} time segment. Then the future data u_f, v_f and y_f were defined using the recorded data of the time segment N. The vector $[u_{ini}^T \; v_{ini}^T \; y_{ini}^T \; u_f^T \; v_f^T \; y_f^T]^T$ was then used for the calculation of the initial guess g_0 through the linear system of equations in (5.9). Initial guess g_0 is going to be used in the next steps of the data-driven predictor. The `MATLAB` function `linsolve` was used for the solution of the linear system of equations.

$$\begin{bmatrix} U_p \\ V_p \\ Y_p \\ U_f \\ V_f \\ Y_f \end{bmatrix} \cdot g_0 = \begin{bmatrix} u_{ini} \\ v_{ini} \\ y_{ini} \\ u_f \\ v_f \\ y_f \end{bmatrix} \tag{5.9}$$

At this point, the initialization of the data-driven predictor is over. Then, the past and future data were loaded (See Figure 5.2) updating the matrices u_{ini}, v_{ini}, y_{ini}, u_f, v_f and y_f.

Figure 5.2: Introduction of the future N data needed for the prediction of the future output trajectories

The construction of an optimization problem becomes necessary for the prediction of future output trajectories. In the context of the current experiment, it was decided to use the TOMLAB `lsei` solver, a solver suitable for dense linear least squares problems. Therefore, the under review system representation can be expressed in the following form:

$$\begin{aligned} \min_{g} \quad & ||u_f - U_f \cdot g|| \\ \text{s.t.} \quad & \\ & \begin{bmatrix} U_p \\ V_p \\ Y_p \\ V_f \end{bmatrix} \cdot g = \begin{bmatrix} u_{ini} \\ v_{ini} \\ y_{ini} \\ v_f \end{bmatrix} \end{aligned} \tag{5.10}$$

The optimization problem returns a vector g which is used for the prediction of the future output trajectories through the equation below:

$$y_{f_{pred}} = Y_f \cdot g \tag{5.11}$$

The first predicted output values are compared with the corresponding measured output values (y_f) and the rest are discarded as would happen in a receding horizon predictive approach with a control horizon equal to 1. Finally, the process starts over at time $T + 2 \cdot T_{ini} + N + 1$ as shown in Figure 5.3. The described process is also expressed in an algorithmic formulation presented in Algorithm 2.

Figure 5.3: Implementation of Receding Horizon in the data-driven predictor testing process

Algorithm 2 Algorithm describing the initialization and testing of the data-driven system representation

1: **Define** T, T_{ini} and N
2: **Collect** $T + 2 \cdot T_{ini} + N$ **data**
3: **if** T is sufficiently long **then**
4: Construct Hankel Matrices U, V and Y
5: Define u_{ini}, v_{ini}, y_{ini}, u_f, v_f and y_f
6: Solve the System (5.9)
7: **else**
8: Go to Step 1
9: **end if**
10: **while** True **do**
11: Define the new u_{ini}, v_{ini}, y_{ini}, u_f, v_f and y_f
12: Compute g^* (Solve the Optimization Problem (5.10))
13: Compute the optimal output sequence $y^* = Y_f \cdot g^*$
14: Compare the first predicted and measured outputs
15: Shift on time
16: **end while**

SIMULATION RESULTS

Figure 5.4 depicts the results of Algorithm 2 run for a period of 9 days. The left column shows the predicted versus the measured values with green and black colors respectively. The right column of the figure exhibits the evolution of the prediction errors over time. For the sake of brevity, only 5 out of 23 states/outputs have been included in Figure 5.4. The mean execution time of Algorithm 2 per iteration is 1.7895 seconds with the maximum time of 1.8552 and the minimum time of 1.7424 seconds.

It has great importance to note that despite the violation of the persistency of excitation condition, the predictions conclude on acceptable error values for the first 3000 timestamps. However, over time the absolute prediction errors increase. This increase is justified by the time-variant nature of the greenhouse system. The climate-crop dynamics change as the crop grows, the canopy surface increases, the air volume of the greenhouse decreases, and surfaces such as the floor and the soil start reacting differently. Nevertheless, the Hankel matrices U, V, and Y are static representing an older

version of the climate-crop behavior. A solution to this problem would be the online update of the Hankel matrices as demonstrated in Section 5.4.2.

5.4.2. DYNAMIC HANKEL MATRICES

As presented in the previous section, the time-variant nature of the climate-crop system does not permit the use of static Hankel matrices. The increase of the absolute prediction error over time can be surpassed by updating the Hankel matrices U, V, and Y in an online manner. In this way, it is guaranteed that the current system representation is up to date in every timestamp. This approach is summarized in the Algorithmic formulation 3.

Algorithm 3 Algorithm describing the initialization and testing of the online data-driven system representation

1: **Define** T, T_{ini} and N
2: **Collect** $T + 2 \cdot T_{ini} + N$ **data**
3: **if** T is sufficiently long **then**
4: Construct Hankel Matrices U, V and Y
5: Define u_{ini}, v_{ini}, y_{ini}, u_f, v_f and y_f
6: Solve the System (5.9)
7: **else**
8: Go to Step 1
9: **end if**
10: **while** True **do**
11: Define the new u_{ini}, v_{ini}, y_{ini}, u_f, v_f and y_f
12: Compute g^* (Solve the Optimization Problem (5.10))
13: Compute the optimal output sequence $y^* = Y_f \cdot g^*$
14: Compare the first predicted and measured outputs
15: Shift on time
16: **if** 24 Hours Passed **then**
17: Update U, V and Y from the Recently Recorded Data
18: **end if**
19: **end while**

SIMULATION RESULTS

As depicted in Figure 5.5 the frequent update of the Hankel matrices had the desired results. The left column of the figure clearly displays that the prediction errors do not increase over time as the data-driven system representation reflects the behavior of the ground truth system at every timestamp. For the sake of completeness, it should be mentioned that both experiments used the same dataset and T, T_{ini} and N values. Furthermore, as in the first experiment, the persistency of excitation condition was violated. Regarding the execution time of Algorithm 3 per iteration the mean execution time was 1.8201 with a maximum and a minimum value at 1.9748 and 1.7391 respectively.

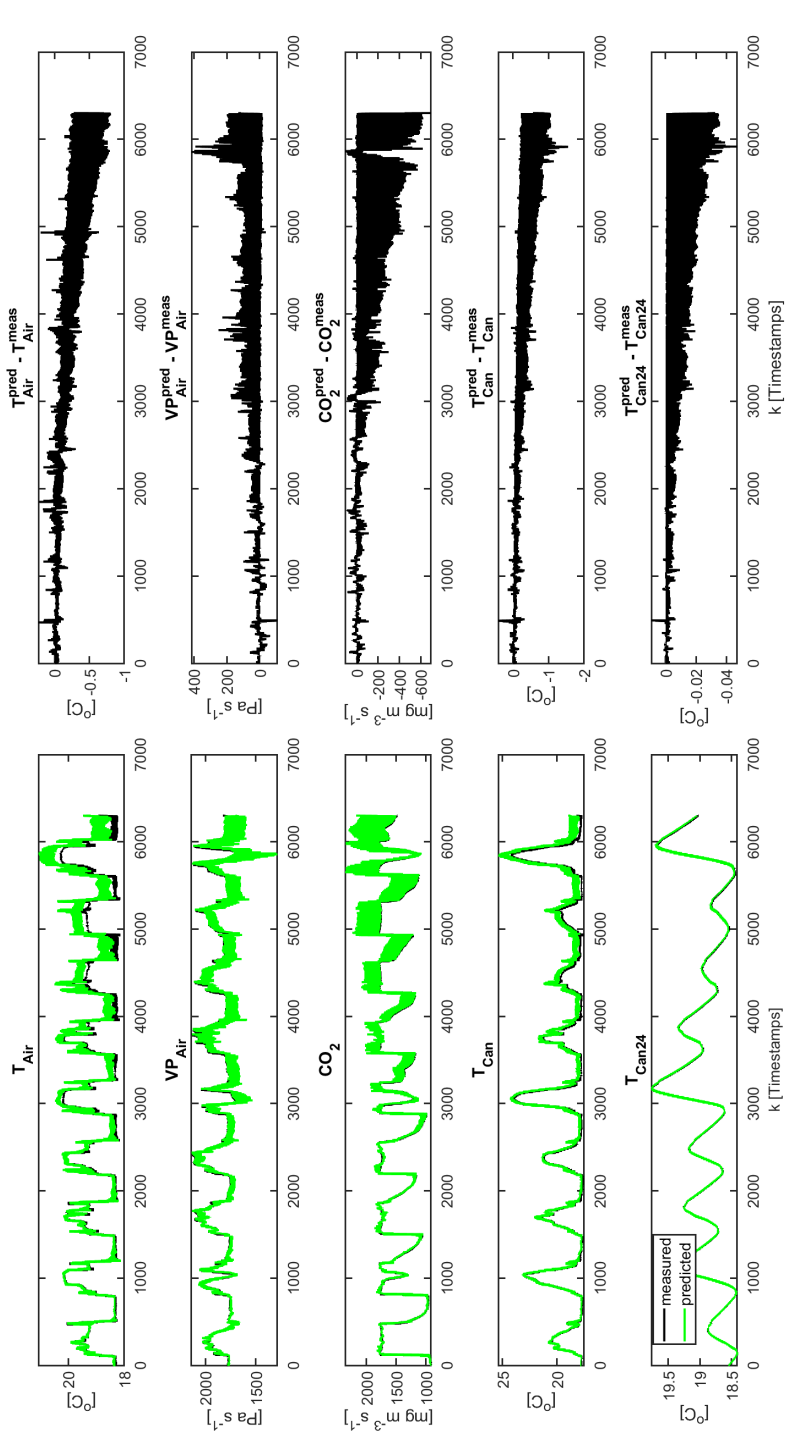

Figure 5.4: Implementation of Receding Horizon in the data-driven predictor testing process with $T = 4320$, $T_{ini} = 5$ and $N = 12$. The left column indicates the predicted values (green) versus the measured values (black). The right column displays the evolution of the prediction error over time.

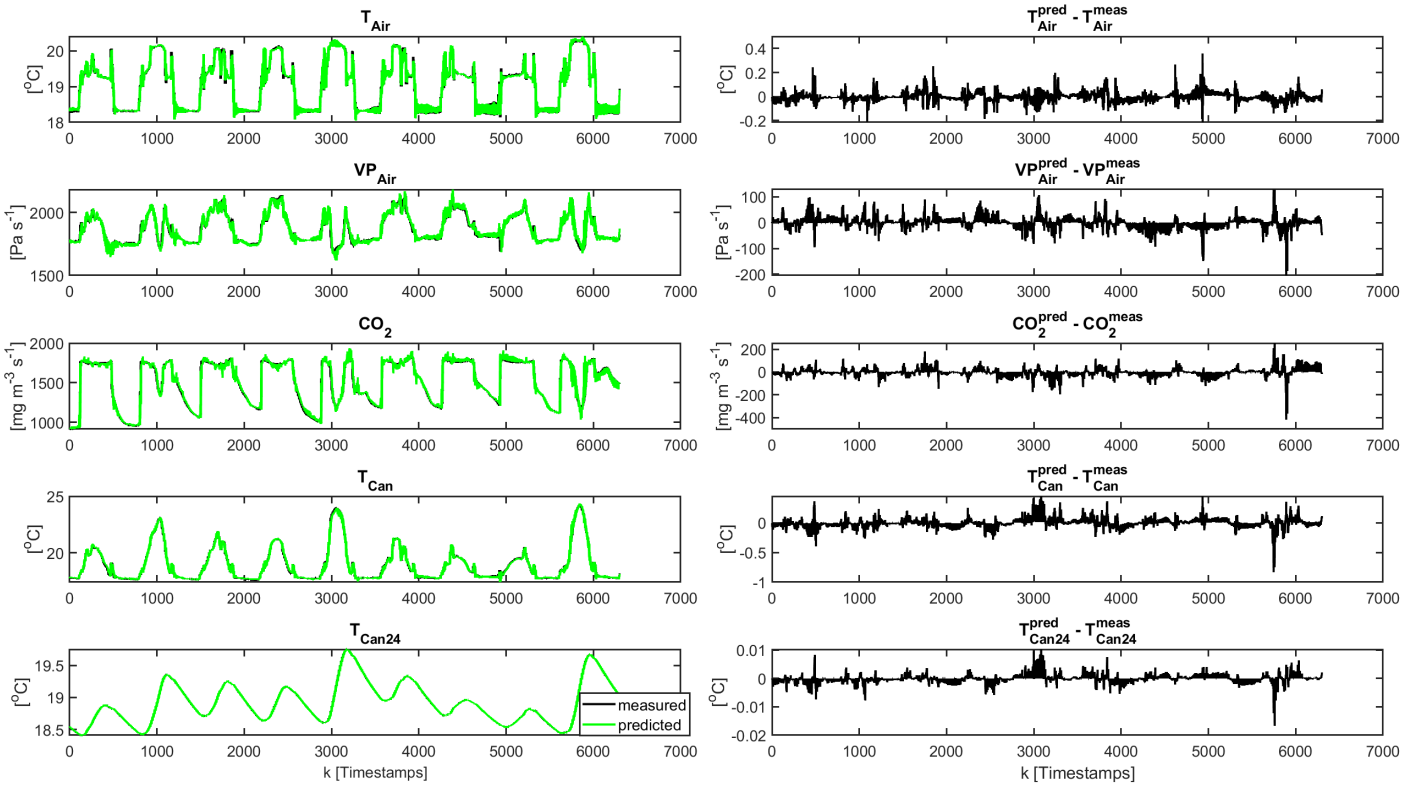

Figure 5.5: Implementation of Receding Horizon in the data-driven predictor testing process with online updating of Hankel matrices every 24 hours and $T = 4320$, $T_{ini} = 5$ and $N = 12$. The left column indicates the predicted values (green) versus the measured values (black). The right column displays the evolution of the prediction error over time.

5.4.3. DISCUSSION

This section is devoted to the presentation of some noteworthy remarks that occurred during the experimental process.

The selection of the exact T, T_{ini}, and N has great importance as it balances the trade-off between prediction accuracy and execution time. T was selected to be equal to 4320 corresponding to 15 days of data. For smaller T values, the data-driven representation concluded on greater prediction errors. On the other hand, when T is greater than 4320 the execution time of the optimization problem increases without providing any significant decrease in prediction errors. The exact same pattern holds for the selection of T_{ini}, and N. This behavior can be justified by the fact that the dimensions of the Hankel matrices are $n_u \cdot (T_{ini}+N) \times (T-L)$, $n_d \cdot (T_{ini}+N) \times (T-L)$, and $n_x \cdot (T_{ini}+N) \times (T-L)$ for the matrices U, V, and Y respectively. Precisely, the values of T_{ini} and N affect the number of linear equations in the linear system of equations (5.12). T impacts the number of columns of the Hankel matrices and the dimension of the control variable $g \in \mathbb{R}^{T-L}$. Unfortunately, there is no way to predict the execution time of the optimization problem as it is dependent on several factors such as the optimization criterion, the constraints, the solver, the initial guess and the used hardware. Moreover, because of the PE condition violation, prediction accuracy cannot be estimated. As a result, it is suggested to always use the lower acceptable T, T_{ini}, and N values.

The persistency of excitation condition was tested for the Hankel matrix $\mathscr{H}_L([u\ d])$ containing both the control and the exogenous inputs. Nonetheless, when testing PE for $\mathscr{H}_L(u)$, *Definition* 6 is satisfied. Fulfilling *Definition* 6 indicates that the control inputs used are exciting enough frequencies on the ground truth system, permitting the unique identification of the underlying dynamical system from data. On the other hand, the PE condition does not hold when tested only for the Hankel matrix $\mathscr{H}_L(d)$. Hence, exogenous inputs do not sufficiently excite the system.

A reason justifying this fact may have to do with the patterns existing in the exogenous inputs. For example, global radiation follows every day the same pattern with different intensities. Specifically, the global radiation becomes zero during the night time, increases till its peak point, and then decreases returning to zero. Similar behavior can be seen in the outdoor temperature where it oscillates between variable minimum night temperatures and the maximum daily temperatures. Figure 5.6 displays the global radiation and the outdoor temperatures used in the construction of matrix V.

The violation of the PE condition entails the absence of validation for *Lemma* 3. As the input data are not PE, it cannot be ensured that any linear combination of the columns of $[U^T\ V^T\ Y^T]^T$ is a t-long input-output trajectory of the ground truth system. Therefore, there are cases where the optimization problem returns a feasible solution that numerically validates the linear system of equations (5.12) but not semantically.

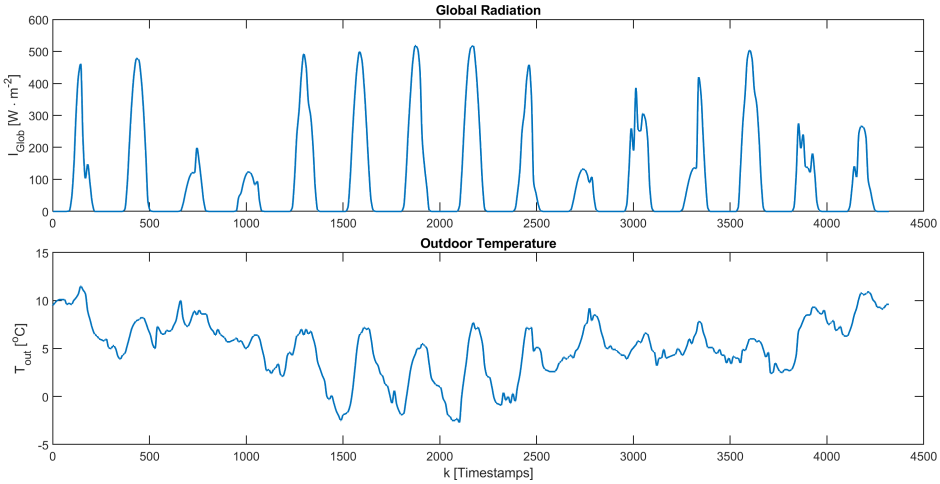

Figure 5.6: Global radiation and outdoor temperature trajectories utilized for the construction of the Hankel matrix V used in the experiments of Sections 5.4.1 and 5.4.2

$$\begin{bmatrix} U_p \\ V_p \\ Y_p \\ V_f \end{bmatrix} \cdot g = \begin{bmatrix} u_{ini} \\ v_{ini} \\ y_{ini} \\ v_f \end{bmatrix} \tag{5.12}$$

Such an example can be seen in Figure 5.7 where the data-driven predictor was trained using the dataset exploited in Sections 5.4.1 and 5.4.2 but it was not tested on the sequel of the training dataset. The testing dataset still included a mature crop but the weather conditions corresponded to a different time period. This experiment concluded on similar execution times where the mean, the maximum, and the minimum execution times were measured at 1.8103, 1.9109, and 1.7475 correspondingly.

Concluding, when the PE condition does not hold, the non-parametric system representation is capable to be used as a predictor if the input signals comply with the behavior of the input signals living in the Hankel matrices. In any other case, there is the risk to predict trajectories that do not comply with the ground truth behavior resulting in extreme prediction errors. This phenomenon could be compared with the overfitting phenomenon in machine learning.

5.5. EXPERIMENT: THE CROP'S SYSTEM PARTIAL REPRESENTATION

According to the previous experiment, the data-driven predictor can describe the behavior of the climate-crop system when the training and the validation dataset contain similar behavior in terms of inputs. Section 5.4.3 shows that when the PE condition is violated and the training and validation datasets contain different behavior, the data-driven

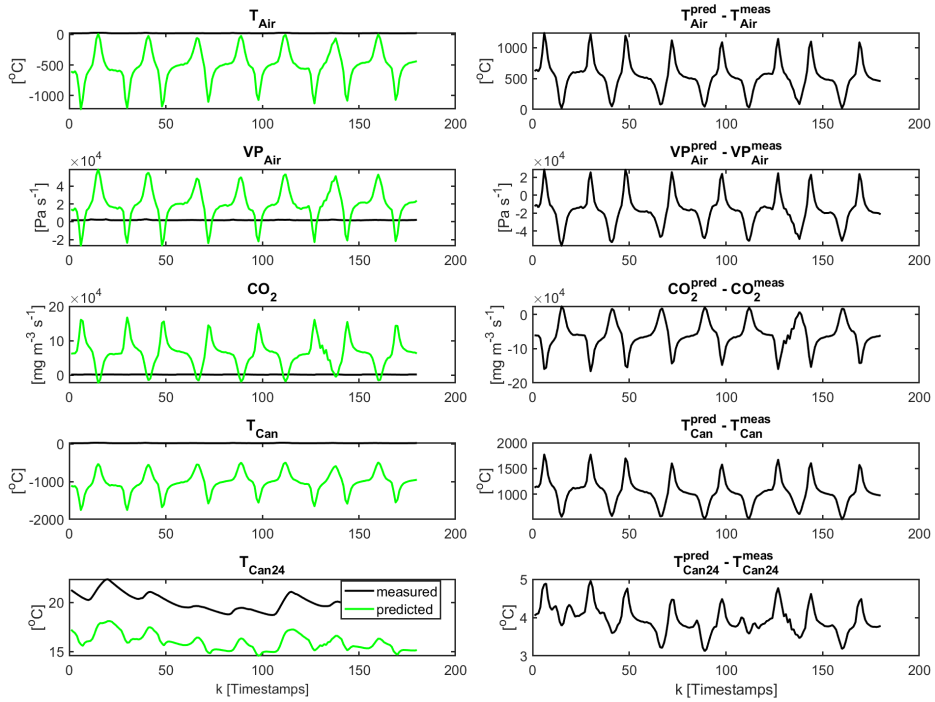

Figure 5.7: Presentation of a case where the training and the testing dataset do not follow the same patterns concluding in highly increased prediction errors. The used data-driven predictor was implemented for $T = 4320$, $T_{ini} = 5$ and $N = 12$.

predictor fails. At this point, it would be interesting to examine how the data-driven representation would behave as the system representation of a Data-Enabled Predictive Controller (DeePC), despite the PE violation. This experiment intends to provide some insight into the behavior of the data-driven predictor when the future exogenous inputs are similar to the training exogenous inputs but the control inputs differ. However, for T value close to 4320 the computational load of the DeePC becomes extremely high as the solver has to calculate 4304 entries of the control variable g. Thus, it was decided to implement the experiment on a sub-model of the crop model. Specifically, it is examined if DeePC can efficiently predict and control the canopy temperature T_{Can} based on the inputs T_{Air}, VP_{Air}, CO_2 and the exogenous input R_{Can}.

The climate-dependent tomato crop system described in Section 3.4.3 is used for the production of the ground truth data. The states of the sub-model, that DeePC is tested on, are expressed by the following differential equations:

$$\dot{T}_{Can} = \frac{1}{cap_{Can}}(R_{PAR\_SunCan} - H_{CanAir} - L_{CanAir}) \tag{5.13}$$

$$\dot{T}_{Can24} = \frac{1}{\tau}(k \cdot T_{Can} - T_{Can24}) \tag{5.14}$$

Non-linearities are introduced through the latent heat flux term L_{CanAir}. An analytical formulation of the term is given below:

$$L_{CanAir} = L \cdot (satVp(T_{Can}) - VP_{Air}) \cdot \frac{2 \cdot \rho_{Air} \cdot c_{P_{Air}} \cdot lai}{L \cdot \gamma \cdot (r_B + r_S)} \tag{5.15}$$

Where L is the coefficient of the latent heat of evaporation, ρ_{Air} the air density, $c_{P_{Air}}$ the air's specific capacity, γ the psychrometric constant, r_B the boundary layer resistance of the canopy for transpiration and *lai* the measured leaf area index. The non-linearities are introduced from the saturation function $satVp(T_{Can}) = p_1 \cdot e^{p_3 \cdot \frac{T_{Can}}{T_{Can}+p_2}}$ and the stomatal resistance model r_S. A more analytical overview of the stomatal resistance model can be found in Chapter 4.

For the sake of complicity, it should be noted that the experimental process was repeated as presented in the previous sections. However, in this experiment $T = 288$, $T_{ini} = 12$ and $N = 12$ (entries). The selected T, T_{ini}, and N are able to provide a sufficiently long input signal but the PE condition is still violated. Figure 5.8 depicts the results of Algorithm 2 run for a period of one day. The Algorithm's 2 measured mean execution time per iteration is 0.0079 seconds.

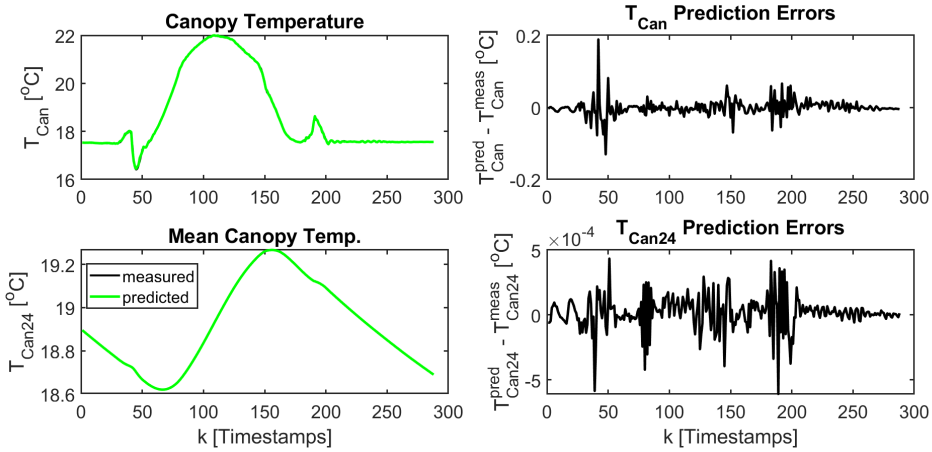

Figure 5.8: Implementation of Receding Horizon in the T_{Can} data-driven predictor testing process. The left column indicates the predicted values (green) versus the measured values (black). The experiment run for a period of one day for $T = 288$, $T_{ini} = 12$ and $N = 12$.

The DeePC formulation is given below:

$$
\begin{aligned}
\min_{g} \quad & ||T_{Can} - 22||_2^2 \cdot D + ||T_{Can} - 14||_2^2 \cdot (1 - D) + ||\Delta u||_R \\
\text{s.t.} \quad & \\
& \begin{bmatrix} U_p \\ V_p \\ Y_p \\ V_f \end{bmatrix} g = \begin{bmatrix} u_{ini} \\ v_{ini} \\ y_{ini} \\ v_f \end{bmatrix} \\
& u_l \leq U_f \cdot g \leq u_u \\
& D = \begin{cases} 1, & \text{if } \exists i \in \{1, \ldots, N\} \text{ for which } R_{Can} > 0 \\ 0 & \text{if } \forall i \in \{1, \ldots, N\}\ R_{Can} = 0 \end{cases}
\end{aligned}
\tag{5.16}
$$

5

The objective of the data-driven controller is to predict the future T_{Air} and VP_{Air} trajectories for given future CO_2 and R_{Can} trajectories that will steer T_{Can} to track the reference level of 22^oC and 14^oC during the day and night time respectively. The future, given trajectories have been extracted as the sequel of the training dataset. The controller predicts over a prediction horizon $N = 12$ and applies the first 3 entries (control horizon $N_c = 3$), the rest 9 predicted values are discarded. The control variables are T_{Air} and VP_{Air} properly formulated with the help of g vector and the input constraints indicate that $10 \leq T_{Air} \leq 25$ and $1000 \leq VP_{Air} \leq 2500$. Finally, the experiment was simulated in `MATLAB` using `TOMLAB` and the solver `conSolve` was used for the solution of the optimization problem. The mean execution time was calculated at 3.1982 seconds. The results of the simulation are presented in Figure 5.9. The first row of plots depicts the inputs and the exogenous input of the system and the second row displays the tracking of the reference trajectory and the prediction errors.

The DeePC can, efficiently, track the reference trajectory. However, it should be noted that the data-driven predictor overestimates the system's response time as increased prediction errors occur when the reference level changes. The lower plots in Figure 5.9 exhibit increased prediction errors during the transitions from the initial condition to the 14^oC reference level ($k \in [0,\ 12]$), from 14^oC to 22^oC reference ($k \in [36,\ 41]$), and from 22^oC to 14^oC ($k \in [260,\ 265]$). A noteworthy remark has to do with the applicability of the reference trajectory. It was known *a priori* that the system is capable to achieve $T_{Can} = 22^oC$ and $T_{Can} = 14^oC$ during the day and the night time accordingly. In consequence, the data-driven predictor only had to indicate how the system can be steered in those directions. Now, it would be interesting to examine the predictor's accuracy in a problem where the controller has to calculate the optimal canopy temperature and implement it. In other words, the DeePC does not only need to find how to steer the system but also where to steer it. The SPA objective function, introduced in Chapter 4 can be

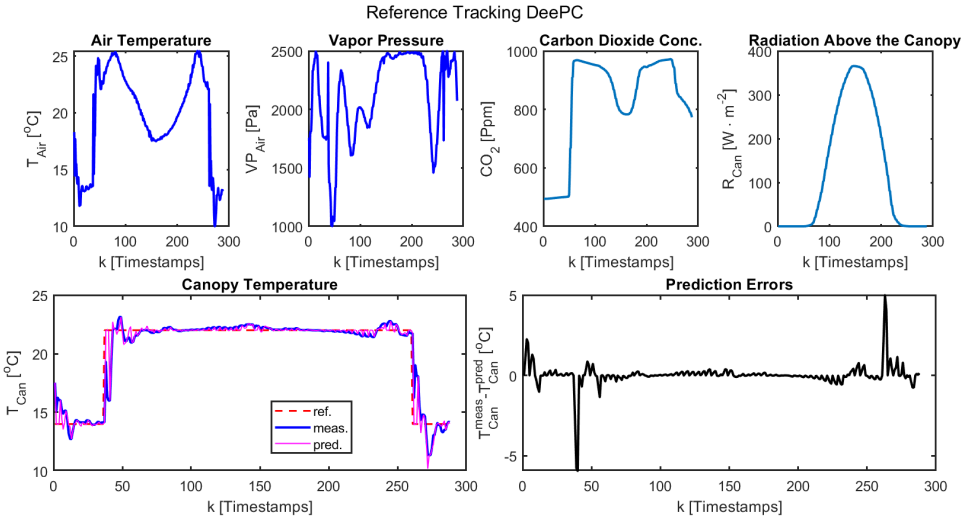

Figure 5.9: Simulation results of the T_{Can} reference tracking DeePC. The inputs and exogenous inputs are displayed in the first row of the plots. The lower left plot demonstrates the reference tracking capabilities of the controller where the dashed red line displays the reference trajectory, the solid blue line represents the measured output, and the magenta line shows the DeePC predictions. The lower right plot demonstrates the prediction errors that occurred during the simulation. The experiment was conducted for one day.

exploited to test this concept. The formulation of the SPA-DeePC is given below:

$$\begin{aligned}\min_{g} \quad & (C_0 \cdot ||p_1 \cdot e^{\frac{p_3 \cdot T_{Can}}{T_{Can}+p_2}} - VP_{Air}||_2^2 + C_1 \cdot ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 + C_2 \cdot ||T_{Can_{24}}||_2^2) \cdot D + \\ & (C_3 \cdot ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 + C_4 \cdot ||T_{Can_{24}}||_2^2) \cdot (1-D) + ||\Delta u||_R \\ \text{s.t.} \quad & \begin{bmatrix} U_p \\ V_p \\ Y_p \\ V_f \end{bmatrix} g = \begin{bmatrix} u_{ini} \\ v_{ini} \\ y_{ini} \\ v_f \end{bmatrix} \\ & u_l \leq U_f \cdot g \leq u_u \\ & y_l \leq Y_f \cdot g \leq u_u \\ & D = \begin{cases} 1, & \text{if } \exists i \in \{1,\dots,N\} \text{ for which } R_{PAR^{\,i}_{SunCan}} > 0 \\ 0 & \text{if } \forall i \in \{1,\dots,N\}\ R_{PAR^{\,i}_{SunCan}} = 0 \end{cases}\end{aligned} \tag{5.17}$$

Where $C_0 = 10^{-10}$, $C_1 = 10^3$, $C_2 = 10^{-8}$, $C_3 = 4 \cdot 10^{-2}$, $C_4 = 10^4$, $R = diag([10^5\ 10])$, and $T^*_{Can_{24}} = 19.88^o C$. Parameters p_1, p_2, and p_3 reflect the parameters of the vapor pressure saturation function. Regarding the constraints, it holds that $10 \leq T_{Air} \leq 25$, $1000 \leq VP_{Air} \leq 2500$, $14 \leq T_{Can} \leq 27$ and $14 \leq T_{Can_{24}} \leq 27$.

The SPA-DeePC algorithm uses the same predictor as also the same horizon values as the

reference tracking DeePC and no changes are applied regarding the simulation software. The results of the SPA-DeePC experiment can be found in Figure 5.10 where the first row of plots contains the inputs. The lower left plot demonstrates the system's output where the solid blue line represents the measures output, and the magenta line shows the DeePC predictions. Finally, the lower right plot shows the prediction errors that occurred during the simulation process. The simulated time corresponds to one day and the mean execution time per iteration is measured at 67.9881 seconds.

According to the SPA objective function, the DeePC is expected to steer the canopy temperature at a low level during the night to reduce sugar losses through maintenance respiration. On the other hand, high canopy temperatures are expected during day time as the increased T_{Can} values will increase photosynthesis and assimilates partitioning. Despite the known pattern, the exact canopy temperature values should be indicated by the system's permissible behavior. From the lower left plot of Figure 5.10 becomes clear that the controller achieves this pattern but not as predicted. The predictor still overestimates the system's response time concluding in high prediction errors. The magnitude of the prediction errors is a result of the PE condition violation entailing the *Lemma's* 3 dissatisfaction. This observation, also, justifies the fact that the solver returns feasible `exitFlags` in all iterations despite concluding on physically, non-achievable trajectories. For example, during the dark hours ($k \in [0,\ 45] \cup [240,\ 288]$), the R_{PAR_SunCan} term of Equation (5.13) becomes zero. Then the canopy exchanges energy only with the air. From an empirical perspective, the permissible system's behavior would be $0 \leq T_{Air} - T_{Can} \leq 1.5$ but the data-driven predictor insists that T_{Can} can go to 14^oC while $T_{Air} \geq 17.6^oC$.

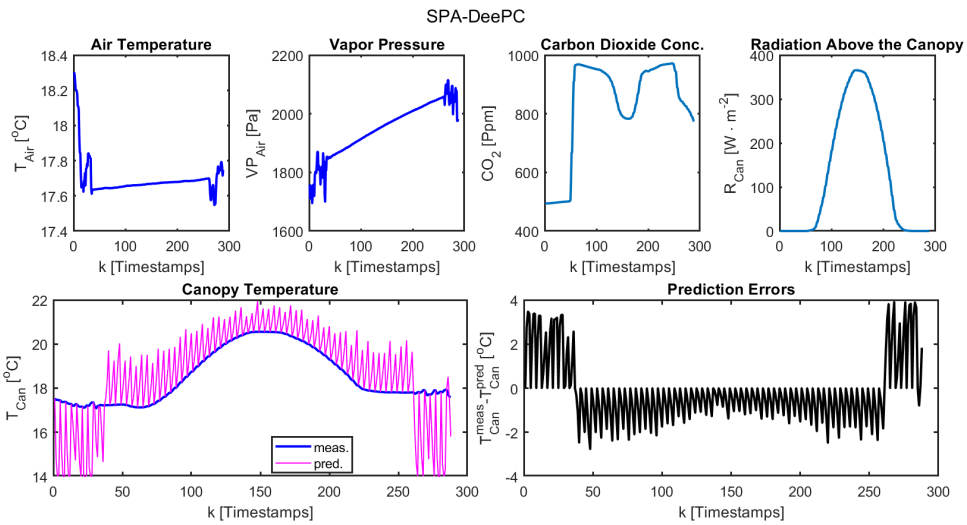

Figure 5.10: Simulation results of the SPA-DeePC. The inputs and exogenous inputs are displayed in the first row of the plots. The lower left plot demonstrates the system's output where the solid blue line represents the measures output, and the magenta line shows the DeePC predictions. The lower right plot demonstrates the prediction errors that occurred during the simulation. The experiment runs for one day.

The missing behavioral characteristics could be added as constraints in the optimization problem. For instance, the term $C_5 \cdot ||T_{Air} - T_{Can}||_2^2$ can be introduced in the objective function of the Problem (5.17). In this way, the temperature difference between the air and the canopy is penalized and a decrease in night-time prediction errors is expected. Figure 5.11 presents the simulation results after the addition of the penalty term with $C_5 = 10^5$ for $D = 0$ and $C_5 = 10^3$ for $D = 1$. The figure maintains the same structure and color code regarding the plots. As expected, the night-time prediction errors are decreased as the solver knows that $T_{Air} - T_{Can}$ should be limited. High prediction errors are observed during the daytime, the reason is that compared to T_{Air}, R_{Can} has a greater influence on T_{Can}. Nevertheless, such an approach has two major disadvantages. At first, there is no way to ensure that the introduced constraints are exhaustive and as a result, no guarantees can be given about the predictor's accuracy. Hence, the data-driven representation cannot be considered validated. Moreover, these constraints express the system's behavior that should be indicated by the system representation. Semantically they should not be used as constraints in the optimization problem.

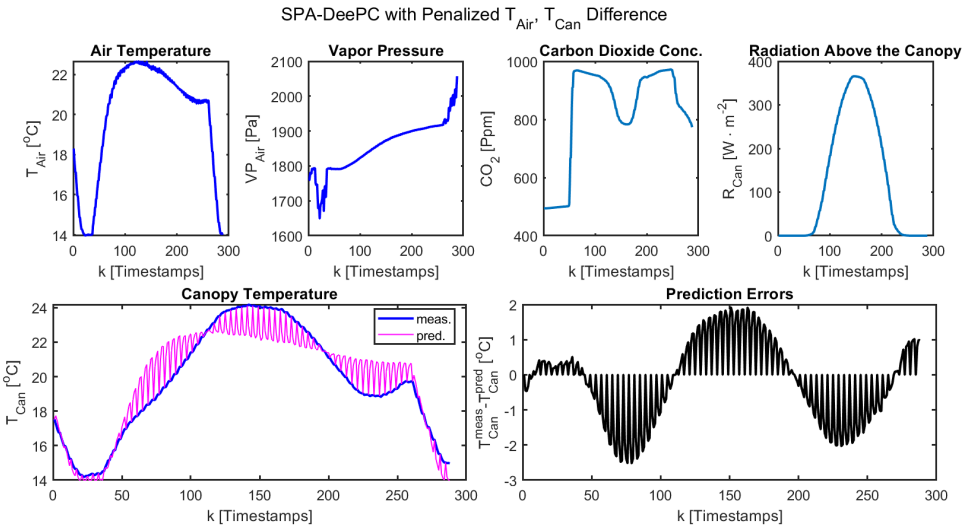

Figure 5.11: Simulation results of the SPA-DeePC with penalized T_{Air}, T_{Can} difference. The inputs and exogenous inputs are displayed in the first row of the plots. The lower left plot demonstrates the system's output where the solid blue line represents the measures output, and the magenta line shows the DeePC predictions. The lower right plot demonstrates the prediction errors that occurred during the simulation. The experiment runs for one day.

5.5.1. DISCUSSION

The data-driven representation is a linearized version of the ground truth system for the state space included in the training dataset. Thus, despite the violation of the PE condition, the data-driven predictor can be used to approximate the system's behavior in a predictive controller when the controller tries to steer the system's output on a, *a*

priori known, reachable trajectory. However, as shown by the SPA-DeePC attempt, the violation of the PE condition let the solver consider, physically unreachable trajectories as feasible solutions. Therefore, the data-driven predictor is not capable to be used in a nominal predictive controller. Error compensation techniques have to be applied as the data do not efficiently express the system's behavior. However, this data-based representation approach can be considered promising for the partial representation of the system's behavior, especially when the testing dataset contains inputs with similar patterns to the training dataset.

6

NONLINEAR MODEL PREDICTIVE CLIMATE CONTROL

6.1. INTRODUCTION - CASE STUDY

This chapter performs a case study of three different control regimes. All cases were tested on the GreenLight climate-crop model for the same exogenous inputs. The first controller is based on the SPA predictive control approach presented in Chapter 4. In this case, a Non-Linear Model Predictive Controller (NMPC) founding its decision-making process on SPA, developed and tested on the ground truth model. The second control regime is rule-based and it is implemented using a combination of sigmoid P controllers. Rule-based controllers may be considered out of date but they are still in use as they do not require any system representation and they can achieve decent results under proper tuning. The aforementioned controller acts as a benchmark representing conventional climate control. In the context of a fair comparison, a controller able to represent the optimal ratio between the final yield and the energy consumption is required. For this reason, it was decided to design and test an NMPC algorithm including a yield maximization economic objective function. The yield maximization NMPC controller is the third climate control approach being a part of the comparison process and it is considered a benchmark. The objective of the numerical simulations is the comparison of the novel SPA-based control algorithm with the benchmark control regimes. The comparison takes place in terms of energy consumption, yield production, carbohydrate balance, and execution time.

The following paragraphs outline the weather profile used in the simulations and the inputs controlled by the climate controllers. Section 6.2 demonstrates the SPA ENMPC algorithm and the design decisions regarding the objective's function weights and cost matrices. Section 6.3 provides the overview of the benchmark control algorithms. The comparison between the various simulated results is presented in Section 6.5.

WEATHER CONDITIONS

The weather conditions are considered as the exogenous inputs of the system and they follow the formulation introduced in (3.5) respecting the mentioned units of measurement. The data used describe the weather in Amsterdam, the Netherlands, for the period between March and May of the year 1995.

CONTROL INPUTS

The control inputs are a subset of the inputs presented in (3.6). The tested climate controllers do not apply control on artificial lighting, thus all corresponding input signals can be excluded from the climate control algorithms. Then, the control inputs vector becomes:

$$u = \begin{bmatrix} Boil \\ BoilGro \\ ThScr \\ Roof \\ ExtCo_2 \end{bmatrix} \tag{6.1}$$

Where $Boil$ and $BoilGro$ are the boiler values used to heat the heating pipes and the grow pipe system, accordingly. $ThScr$ the deployment of the thermal screen, $Roof$ the roof's window aperture, and $ExtCo_2$ the actuator corresponding to the CO_2 enrichment system. As shown in Equation (6.2) all control inputs take values in the range $[u_l,\ u_u]$ where $u_l = 0$ and $u_u = 1$.

$$u_l \leq u_i \leq u_u \quad \forall i \in \{1, \ldots, 5\} \tag{6.2}$$

6.2. SPEAKING PLANT APPROACH NMPC

The proposed SPA-NMPC climate control algorithm can be considered as the extension of the climate controller introduced in Section 4.6.1 for the whole climate-crop system. The optimal future input trajectories are selected for a prediction horizon N. In other words, SPA-NMPC is responsible to calculate the future greenhouse actuator inputs able to minimize the greenhouse energy consumption while ensuring the maximum production and distribution of carbohydrates in the crop and the minimum carbohydrates losses. At this point should be reminded that carbohydrate production is monitored by stomatal resistance and carbohydrate distribution by the reference tracking of the mean canopy temperature in combination with the preservation of canopy temperature in its optimal range. Finally, carbohydrates losses are a result of maintenance respiration. The optimization problem formulating the SPA-NMPC approach is given below:

$$
\begin{aligned}
\min_{u} \quad & \Big(\sum_{k=0}^{N-1} ||u_k||_{R_{\Delta_k}} + ||u_k - u_{k-1}||_{R_1} + C_0 \cdot ||p_1 \cdot e^{\frac{p_3 \cdot T_{Can}}{T_{Can}+p_2}} - VP_{Air}||_2^2 \\
& + C_1 \cdot ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 + C_2 \cdot ||T_{Can_{24}}||_2^2 + C_3 \cdot ||CO_2^* - CO_2||_2^2) \cdot D + P + \\
& \Big(\sum_{k=0}^{N-1} ||u_k||_R + ||u_k - u_{k-1}||_{R_1} + C_4 \cdot ||T^*_{Can_{24}} - T_{Can_{24}}||_2^2 + C_5 \cdot ||T_{Can_{24}}||_2^2) \cdot (1 - D) \\
\text{s.t.} \quad & \\
& x_{k+1} = f(x_k, u_k, d_k, k), \ \forall k \in \{0, \ldots, N-1\} \\
& x(0) = \hat{x}(k) \\
& u_l \leq u \leq u_u, \ \forall k \in \{0, \ldots, N-1\}
\end{aligned}
\tag{6.3}
$$

Where $R_{\Delta_k} \succcurlyeq 0$ and $R \succcurlyeq 0$ are the quadratic positive semi-definite control input cost matrices. $R_1 \succcurlyeq 0$ is the quadratic positive semi-definite cost matrix for the rate of change of the control actions. u_l and u_u the lower and upper actuator bounds respectively and $x_{k+1} = f(x_k, u_k, d_k, k)$ the ground truth climate-crop model needed for the prediction of the future trajectories, where $x0$ the initial state conditions. P denotes the penalty function incorporating the canopy temperature and relative humidity inequality constraints, as displayed in Equation (6.4).

$$
\begin{aligned}
P = \sigma_0 \cdot \max(0, \sum_{k=0}^{N-1} (14 - T_{Can_k}))^2 + \sigma_1 \cdot max(0, \sum_{k=0}^{N-1} (T_{Can_k} - 27))^2 + \\
\sigma_2 \cdot max(0, \sum_{k=0}^{N-1} (RH_{in_k} - 87))^2
\end{aligned}
\tag{6.4}
$$

Finally, D defines the state of the day where $D = 1$ if a non-negative value of the global radiation I_{Glob} exists in the prediction horizon and $D = 0$ when $I_{Glob_i} = 0$, $\forall i \in \{1, \ldots, N\}$ as displayed in Equation (6.5).

$$
D = \begin{cases} 1, & \text{if } \exists i \in \{1, \ldots, N\} \text{ for which } I_{Glob_i} > 0 \\ 0 & \text{if } \forall i \in \{1, \ldots, N\} \ I_{Glob_i} = 0 \end{cases}
\tag{6.5}
$$

6.2.1. WEIGHTS AND COST MATRICES

The weights C_i for $i \in \{0, \ldots, 5\}$ and σ_i for $i \in \{0,\ 1,\ 2\}$ are necessary to properly scale the effect of each term in the objective function. The demonstrated controller uses $C_0 = 1 \cdot 10^{-10}$, $C_1 = 5 \cdot 10^2$, $C_2 = 1 \cdot 10^{-3}$, $C_3 = 2 \cdot 10^{-3}$, $C_4 = 1$, $C_5 = 3 \cdot 10^{-2}$ and $\sigma_i = 3 \cdot 10^4$ $\forall i \in \{0,\ 1,\ 2\}$. The extremely small C_0 value has to do with the fact that the stomatal resistance term included both T_{Can} and VP_{Air}. Vapor pressure control becomes very hard when no humidifying and dehumidifying elements are used as greenhouse actuators. The ground truth greenhouse model does not include any of these devices. Therefore, VP_{Air} control requires quick changes in actuator values and the use of counter-intuitive recipes including heating and cooling at the same time. However, energy consumption

has a peculiar role in climate control, thus it was decided to scale the carbohydrates' production term to the level where it can be efficiently affected by the canopy temperature.

Another remark that should be added is the selection of C_1, C_2, C_4, and C_5. Carbohydrate production is achieved only during the daytime through photosynthesis. Consequently, the monitoring of carbohydrates distribution has a high priority during the day to prevent buffer overflow. For this reason, C_2 is two orders of magnitude greater than C_4. On the other hand, during the nighttime limiting carbohydrates losses is prioritized, hence C_5 is one order of magnitude greater than C_3. Last, the penalty function P is enabled only in the extreme cases where canopy temperature hits the limits of the optimal range or relative humidity becomes extremely high. Such scenarios can induce destructive consequences on the tomato plant and have to be avoided. Accordingly, the high σ_i values authorize the controller to act aggressively and return the system to a safe operational point.

With respect to the cost matrices, it should be noted that they are $n_u \times n_u$ diagonal square matrices. Their diagonal entries are presented in (6.6). As can be seen, the R_{Δ_k} cost matrix is variable, and its values are dependent on global radiation. In particular, the control input costs increase as incident radiation increases. In this way, the higher the incident radiation, the less aggressive the actuator signals and the higher the exploitation of the sun's energy.

$$\begin{aligned} R_{\Delta_k} &= diag([1\ 0.2\ 2.25\ 1\ 0.6]) \cdot I_{Glob_k} \\ R &= diag([1\ 0.2\ 0.001\ 1\ 1]) \\ R_1 &= diag([1\ 1\ 2\ 2\ 0.5]) \end{aligned} \tag{6.6}$$

The cause behind the selection of input costs relates to energy consumption. On days with intense global radiation, the thermal screen should not be deployed to avoid unnecessary shading, lower photosynthetic rate due to limited PAR above the canopy, and reduced thermal energy storage capacity in greenhouse air. Therefore, the thermal screen's cost entries have increased penalties in comparison with the rest actuators in both R_{Δ_k} and R_1. The climate controller intents to use the greenhouse air as a battery of thermal energy charging from the incident radiation. Therefore, the air exchange between the indoor and outdoor environment has to be regulated. The fourth entry of R_1 penalizes the window's flickering, preventing quick changes in the indoor-outdoor airflow. Thermal screen deployment and window aperture impact relative humidity similarly. Thus, when RH_{in} approaches its upper limit (during the daytime) the climate controller should act either by deploying the thermal screen or by opening the window. However, thermal screen deployment will shade the crop and negatively affect the final yield. It is preferable to lose some heat energy rather than reduce carbohydrate production. Because of that, it was decided to use a lower penalty value on window aperture in R_{Δ_k} compared to the penalty of screen deployment. By this approach, the climate controller prefers opening the window when the screen deployment is not obligatory.

Regarding nighttime, the costs R_1 applied on the slew rate of control inputs remain the

same as the fluctuations in actuators' signals should be regulated as in daytime. On the other hand, the control input matrix acting during nighttime is different. The main difference is applied to the cost of the thermal screen. Thermal screening has a great impact on energy saving, especially during cold nights, as it reduces energy losses and increases the efficiency of the heating system. Therefore, the corresponding cost value is 0.001.

Finally, it has to be mentioned that the introduced weights and cost matrices were selected for a 4ha Venlo-type greenhouse located in the Netherlands. Patterns existing in the climate conditions of a geographical region have a crucial role in the expectation that we can have from the climate controller. Therefore, the tuning process should be repeated for greenhouses located in regions with different climate characteristics.

6.3. BENCHMARK ALGORITHMS

6.3.1. RULE-BASED CLIMATE CONTROL

A rule-based climate controller is introduced in order to achieve a valid comparison between the novel SPA-based NMPC control algorithm and the conventional way of climate control. This controller consists of several sigmoid P controllers, where each proportional controller follows the below mathematical formulation:

$$Action = \frac{1}{1 + e^{\frac{-2ln100}{pBand} \cdot (x - setPoint - 0.5 \cdot pBand)}} \; [0 - 1] \tag{6.7}$$

Where x is the controlled variable, $setPoint$ is the desired value for the controlled variable, $pBand$ is the band defining the width of the proportional control, and $Action$ is the control action. $Action$ takes values in range [0, 1] where $Action = 1$ represents the full actuator capacity and $Action = 0$ corresponds to zero actuator's effort. As depicted in Figure 6.1, the controller is close to full action when $x = setPoint + pBand$ and close to no action at $x = setPoint$. The sigmoid curve of Figure 6.1 is flipped horizontally for negative $pBand$ values.

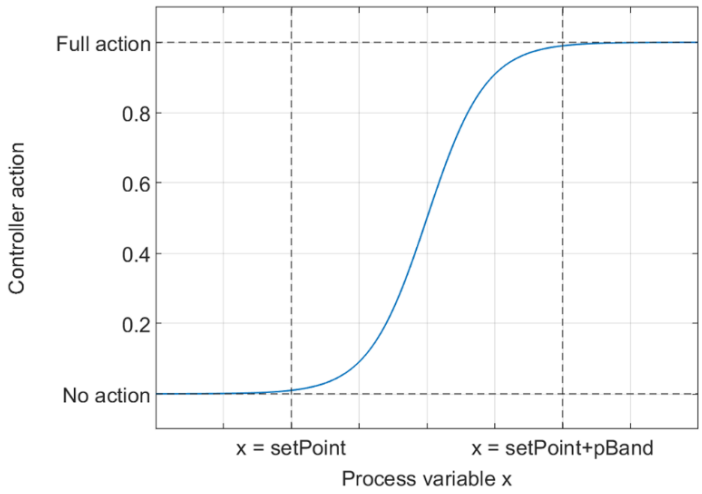

Figure 6.1: Response of the smoothed proportional controller to a process variable x according to a sigmoid function. For negative $pBand$ values, the curve is flipped horizontally [41]

The presented rule-based controller was originally demonstrated in [41] and it was inspired by the work done in [42].

For climate control, the ground truth climate-crop system requires the proportional controllers presented in Table 6.1 to ensure the control of air temperature, CO_2 concentration and prevent the violation of an upper relative humidity level. Each controller in Table 6.1 is characterized by its $setPoint$ and $pBand$ values. Generally, the air temperature has to be maintained on a greater level during the day and on a lower level during the night, CO_2 concentration should be as high as possible to enhance photosynthesis in the daytime but no CO_2 is consumed through dark hours. According to these facts, the rule-based controller should be able to switch modes, thus different setpoints are used for the day and nighttime. The various setpoints used are displayed in Table 6.1.

In reality, climate control requires the excitation of several actuators at the same time. A typical example corresponds to relative humidity control where, in some cases, ventilation and thermal screen deployment needs to be triggered concurrently. Another case is when temperature control requires the cooperation of the boilers with the thermal screens to protect the crop from destructive cold conditions. Consequently, in the rule-based approach, multiple proportional controllers act simultaneously. Figure 6.2 provides an intuitive description of the rules acting in this control method by showing when and how the actuators have to be mobilized. It should be noted that Figure 6.2 is the modified version of Figure 3 presented in [41].

Table 6.1: Necessary P Sigmoid Controllers for the climate control of the ground truth greenhouse model where P_{Heat} the heating controller, P_{CO_2} the CO_2 concentration controller, P^{ExHeat}_{ThScr} the controller for the thermal screen opening due to excess heat, P^{Cold}_{ThScr} the controller for thermal screen closure due to cold outdoor temperatures, P^{ExRH}_{ThScr} the controller for thermal screen opening due to excessive relative humidity, P^{ExHeat}_{Vent} the controller for ventilation due to excess heat, P^{Cold}_{Vent} the controller for ventilation closure due to low indoor temperature, P^{ExRH}_{Vent} the controller for ventilation opening due to excess humidity and D the variable declaring the day and nighttime (See Equation (6.5)).

Necessary P Sigmoid Controllers		
Controller	$setPoint$	$pBand$
P_{Heat}	$T_{sp} = \begin{cases} 18.5\,^oC, & \text{D=0} \\ 19.5\,^oC, & \text{D=1} \end{cases}$	$T_{pBand} = -0.5^oC$
P_{CO_2}	$CO_{2\,sp} = \begin{cases} 0, & \text{D=0} \\ 1000ppm, & \text{D=1} \end{cases}$	$CO_{2\,pBand} = -100ppm$
P^{ExHeat}_{ThScr}	$ThScr^{ExHeat}_{sp} = T_{sp} + 0.5\,^oC$	$ThScr^{ExHeat}_{pBand} = 1$
P^{Cold}_{ThScr}	$ThScr^{Cold}_{sp} = \begin{cases} 10\,^oC, & \text{D=0} \\ 5\,^oC, & \text{D=1} \end{cases}$	$ThScr^{Cold}_{pBand} = -1^oC$
P^{ExRH}_{ThScr}	$ThScr^{ExRH}_{sp} = 85\%$	$ThScr^{ExRH}_{pBand} = 10\%$
P^{ExHeat}_{Vent}	$Vent^{ExHeat}_{sp} = T_{sp} + 0.5\,^oC$	$Vent^{ExHeat}_{pBand} = 0.75^oC$
P^{Cold}_{Vent}	$Vent^{Cold}_{sp} = T_{sp} - 1\,^oC$	$Vent^{Cold}_{pBand} = -1^oC$
P^{ExRH}_{Vent}	$Vent^{ExRH}_{sp} = 87\%$	$Vent^{ExRH}_{pBand} = 50\%$

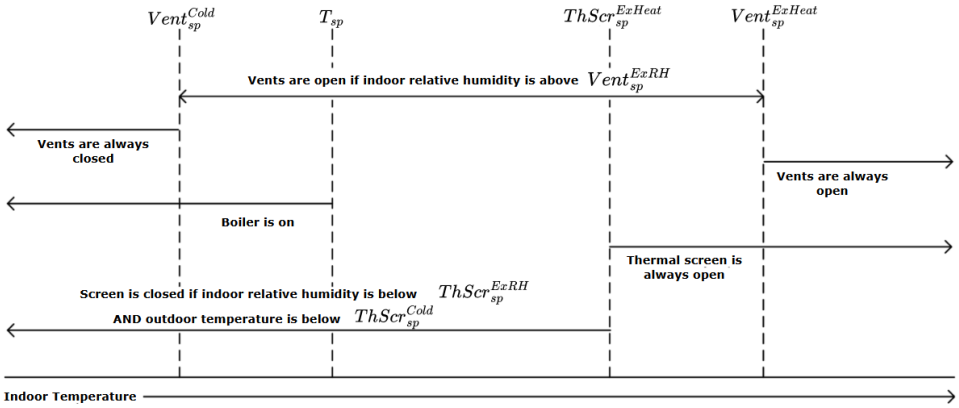

Figure 6.2: Modified version of the Figure 3 originally presented in [41] displaying the behavior of the rule-based climate controller. The figure shows the control of heating, ventilation, and thermal screen in the reference setting where $Vent_{sp}^{Cold}$ the setpoint of the ventilation controller adjusting window opening due to cold temperatures, T_{sp} the setpoint of the heating controller, $ThScr_{sp}^{ExHeat}$ the setpoint of the thermal screen controller due to excess heat, $Vent_{sp}^{ExHeat}$ the setpoint of the controller adjusting window aperture due to excess heat, $Vent_{sp}^{ExRH}$ the setpoint of the controller for the ventilation due to excessive relative humidity, $ThScr_{sp}^{ExRH}$ the setpoint of the controller for the thermal screen opening due to excessive relative humidity and $ThScr_{sp}^{Cold}$ the setpoint of the controller for the thermal screen deployment due to low outdoor temperatures. In respect to carbon dioxide control, CO_2 is supplied whenever the CO_2 concentration is below $CO_{2\,sp}$.

6

6.3.2. YIELD MAXIMIZATION NMPC

The idea behind the design and implementation of the yield maximization NMPC is the generation of a benchmark algorithm able to conclude the optimal trade-off between yield and energy consumption. The results of this algorithm are going to be compared with the results of the SPA-based NMPC approach. To achieve a fair comparison, both algorithms should use similar state spaces, thus they have to incorporate similar constraints regarding the system's states. For this reason, both controllers embed the penalty function P in their objective functions. The potential of the final crop yield can be approximated by the number of carbohydrates in the fruits per square meter. The energy consumption can be regulated by introducing the quadratic costs of the actuator's effort and the actuator's fluctuations as explained in Section 4.7. Photosynthesis is positively correlated with CO_2 concentration, thus in the context of balanced comparison, the yield maximization climate controller should operate under a similar CO_2 concentration with the other two climate controllers. To satisfy this requirement, a reference tracking term steering the CO_2 concentration to $CO_2^* = 1000\ ppm$ is required. The analytical formulation of the yield maximization NMPC method is displayed in Equation (6.8).

$$\min_{u} \quad (\sum_{k=0}^{N-1} ||u_k||_{R_{\Delta_k}} + ||u_k - u_{k-1}||_{R_1} - cFruit(N) + C_0 \cdot ||CO_2^* - CO_2||_2^2) \cdot D + P + (\sum_{k=0}^{N-1} ||u_k||_R + ||u_k - u_{k-1}||_{R_1} - cFruit(N)) \cdot (1 - D) \tag{6.8}$$

s.t.

$$x_{k+1} = f(x_k, u_k, d_k, x_0), \ \forall k \in \{0, \dots, N-1\}$$
$$u_l \leq u \leq u_u, \ \forall k \in \{0, \dots, N-1\}$$

where $cFruit(N)$ is the system's state representing the number of carbohydrates in the fruits at the last time interval of the prediction horizon. The switching function D is defined in Equation (6.5) and the penalty function P in Equation (6.4). $C_0 = 2 \cdot 10^{-2}$ the weight of the CO_2 reference tracking term. The quadratic positive semi-definite cost matrices R_Δ, R_1, and R are defined as shown in Equation (6.6).

6.4. TECHNICAL REMARKS ON THE SIMULATED EXPERIMENTS

RULE-BASED CLIMATE CONTROLLER

The publicly available `MATLAB` code of the Greenlight model was used for the simulation of the rule-based climate controller. Minor changes applied to the original `MATLAB` code in order to introduce Amsterdam's weather dataset, the initial state of the crop, and the control parameters (setpoints and $pBands$) presented in Table 6.1. As mentioned, the simulation starts with an already mature crop. The reason is that the provided photosynthesis model shows abnormal behavior in the simulation of young crops. Specifically, for immature crops, the photosynthesis model concludes on optimal canopy temperatures below the chilling point (12^oC) which is unrealistic.

NMPC CLIMATE CONTROLLERS

Both NMPC algorithms execute the same steps for the simulation of the experiments. A high-level presentation of the founding steps is demonstrated in Algorithm 4.

Algorithm 4 High-level representation of the steps ruling the NMPC Algorithms

1: **Define** N and N_c
2: **while** True **do**
3: **Define the initial conditions** $x_0 = [x_{g_0}^T \ x_{c_0}^T]^T$
4: **Define the future** N **exogenous inputs** d_f
5: **Consider the previous** N **control inputs as the initial guess** $u_0 = u_{past}^*$
6: **Solve the optimization problem** (6.8)**/**(6.3) **and conclude to the optimal future control inputs** $u^* = \{u_0, \cdots, u_{N-1}\}$
7: **Feed** x_0, d_f, **and** $u^* = \{u_0, \cdots, u_{N_c-1}\}$ **in the Greenlight model and receive the system's state trajectories** x
8: **end while**

The NMPC algorithms require a system representation for the prediction of the system's future behavior. From the programming perspective, the predictor is called several times during the solution of the optimization problem. As a result, the predictor's execution time plays a major role in the Algorithm's 4 total execution time. However, the original Greenlight model requires approximately 4 seconds in order to simulate 12 time intervals with a sampling time of 300 seconds. The main reason behind the increased execution time is the use of the `DyMoMa` framework. `DyMoMa` is an object-oriented framework for defining and running linear and non-linear dynamic models in `MATLAB` [43]. Its object-oriented nature makes it a user-friendly tool for the development and analysis of a dynamic model. Each input, state, and auxiliary state is an object containing the label, the symbolic function, and the numerical values of each variable. Despite the provided convenience, the original Greenlight model is considered slow to be used in the context of a predictive controller. For this reason, it was decided to implement an accelerated, non-object-oriented version and incorporate it as a predictor in the NMPC algorithms. The accelerated climate-crop model does not use the `DyMoMa` framework, and provides as outputs only the system's states (See Equation (3.2)). As in the original Greenlight code, the solver `ode15s` is exploited to solve the needed differential equations.

Table 6.2: Comparison of the measured execution times between the original Greenlight model and its non-object-oriented, accelerated version

Measured Execution Times		
	Original Greenlight Model (`DyMoMa` Framework)	Accelerated Version (No `DyMoMa` Framework)
Execution Time for 12 timestamps	3.89 sec	0.36 sec
Execution Time for 1000 timestamps	15.07 sec	12.57 sec

The comparison between the two climate-crop model implementations, in terms of execution time, is displayed in Table 6.2. It can be seen that the accelerated climate-crop model is significantly faster for the relatively small simulation horizon, as it does not have to initialize and define an object for each variable. On the other hand, the execution times are comparable for longer simulation horizons. In this case, the majority of the execution time is consumed by the `ode15s` solver. Figure 6.3 displays the canopy temperature and vapor pressure in air outputs of the accelerated and the original Greenlight codes. The mismatches are considered negligible and the normalized mean square error between the results are $9.99 \cdot 10^{-4}$ and 0.0032 for T_{Can} and VP_{Air} respectively.

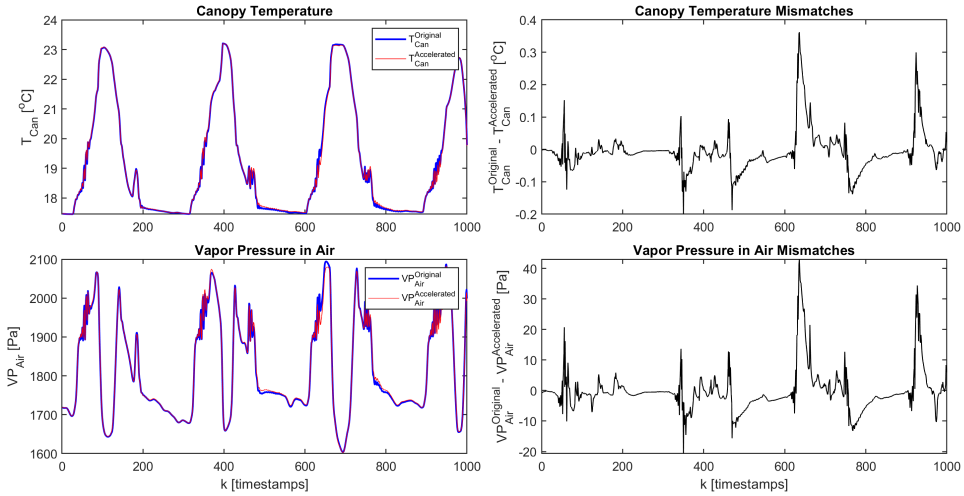

Figure 6.3: Results from the comparison between the original and the accelerated Greenlight codes. The diagram contain the outputs of both codes and the measured mismatches. The results of the original and the accelerated Greenlight code are displayed with blue and red lines, respectively.

6.5. COMPARISON OF SIMULATED RESULTS

The simulation of the experiments was implemented in `MATLAB` using the publicly available Greenlight model as ground truth (plant). The solver `ode15s` was used for the solution of the differential equations describing the model's state trajectories. Regarding the NMPC algorithms, an accelerated, non-object-oriented version of the Greenlight code used as the predictive model. It should be reminded that the original and the accelerated Greenlight models are ruled by the same differential equations, their differences are located in their implementation. The prediction horizon was set to $N = 6$ and the control horizon to $N_c = 3$ timestamps. The sampling time of the model was set to 300 seconds, hence NMPCs predict the future 30 minutes and apply the control inputs for the future 15 minutes. The main reason behind the selection of the horizons was the execution time of the NMPC-based algorithms. The `TOMLAB snopt` solver was used for the calculation of the optimization problems. Finally, the experiments simulate 35 days based on the same weather dataset for a mature crop.

The comparison of the simulated results is conducted on three main pillars, the input energy, the yield, and the algorithm's execution time. The input energy consists of energy originating from the sun and the heating energy from the boilers.

6.5.1. PAR LIGHT FROM THE SUN REACHING ABOVE THE CANOPY

The incident radiation above the canopy is calculated by Equation (6.9) originally introduced in [22],

$$R_{PARGhSun} = (1 - p_{etaGlobAir}) \cdot \tau_{CovPAR} \cdot p_{etaGlobPAR} \cdot I_{Glob} \tag{6.9}$$

where $p_{etaGlobAir}$ the ratio of global radiation absorbed by the greenhouse construction, τ_{CovPAR} the PAR transmission coefficient of the cover, $p_{etaGlobPAR}$ the ratio of PAR in global radiation, and I_{Glob} the incident global radiation in $W \cdot m^{-2}$.

The total PAR light from the sun reaching above the canopy is calculated in $mol \cdot m^{-2}$ and in $MJ \cdot m^{-2}$ as

$$PAR_{Sun} = 10^{-6} \cdot \int (p_{PAR\_Sun}^{J to \mu mol} \cdot R_{PARGhSun}) dt \quad [mol \cdot m^{-2}] \tag{6.10}$$

$$PAR_{Sun} = 10^{-6} \cdot \int R_{PARGhSun}\, dt \quad [MJ \cdot m^{-2}] \tag{6.11}$$

where $p_{PAR\_Sun}^{J to \mu mol}$ expressing the conversion factor of sun's PAR from J to $\mu mol_{photons}$ in $\mu mol_{photons} \cdot J^{-1}$.

The measured PAR_{Sun} are demonstrated in Table 6.3. From the table can be extracted that the SPA-NMPC achieved 0.65% less PAR above the canopy than the rule-based benchmark and 4.68% more PAR than the fruit maximization NMPC.

6

Table 6.3: Demonstration of the measured total PAR light from the sun reaching above the canopy in $mol \cdot m^{-2}$

Measured PAR_{Sun}			
	SPA-NMPC	Rule-Based	Yield Max. NMPC
Measured PAR_{Sun} $[mol \cdot m^{-2}]$	565.2644	568.9227	538.7984
Measured PAR_{Sun} $[MJ \cdot m^{-2}]$	122.8836	123.6789	117.1301

All three experiments run on the same weather dataset entailing that all algorithms faced identical global radiation. Figure 6.4 demonstrates the incident PAR radiation above the canopy for a sunny and cloudy day. As can be seen in Figure 6.4, every crop experienced different PAR above the canopy. The differences in the PAR values above the canopy emerge because of the thermal's screen behavior.

In Figure 6.5, the thermal's screen action is depicted for the two days presented in the previous figure. The advantage introduced by the NMPC algorithms is that the thermal screen opens earlier during the first morning hours (See Figures 6.5 and 6.4 for $k \in [4035, 4065] \cup [4324, 4355]$). This behavior is achieved through the thermal's screen penalty in R_{Δ_k}. On the other hand, the NMPC approaches conclude on lower PAR levels above the canopy during the sun's peak hours on sunny days. This pattern appears on

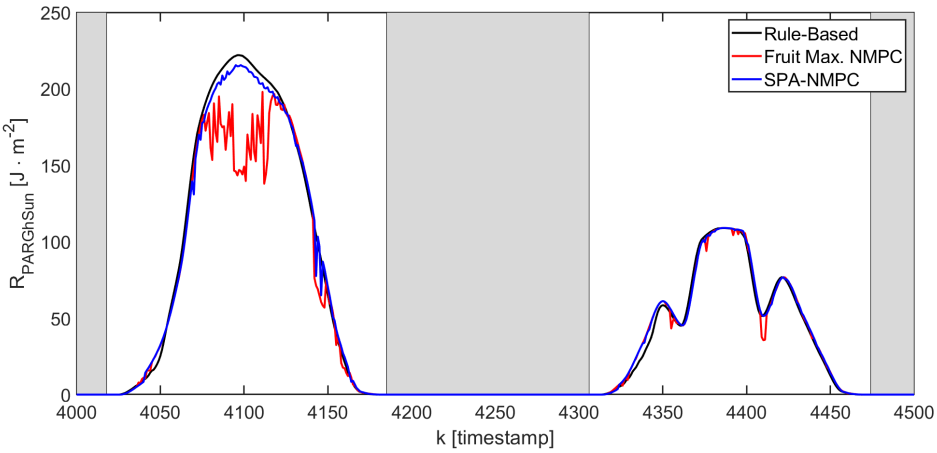

Figure 6.4: The incident PAR radiation above the canopy for two consecutive days. The first day corresponds to a sunny and the second on a cloudy one. The grey (x, y) plane background represents the nighttime and the white corresponds to the daytime.

every sunny day with different intensities, concluding on the lower PAR_{Sun} presented in Table 6.3 for the NMPCs. This pattern occurs due to the penalty function P living in the objective function of the NMPCs. When the canopy temperature or the relative humidity approaches their upper limits, P is activated, permitting the aggressive behavior of the actuators. As a result, thermal's screen fluctuations occur shading the crop (See Figure 6.5 for $k \in [4050,\ 4200] \cup [4330,\ 4450]$).

6.5.2. HEATING ENERGY INPUT

The simulated heating system consists of two boilers, the first boiler is responsible for the control of the heating rails, and the second for the control of the grow heating pipes running on the canopy level. The heat from the boiler to the rails is expressed in Equations (6.12) and (6.13) in $W \cdot m^{-2}$.

$$H_{BoilPipe} = u_{Boil} \cdot \frac{p_{pBoil}}{p_{aFlr}} \tag{6.12}$$

$$H_{BoilGroPipe} = u_{BoilGro} \cdot \frac{p_{pBoilGro}}{p_{aFlr}} \tag{6.13}$$

where u_{Boil} and $u_{BoilGro}$ the corresponding control inputs, p_{pBoil} and $p_{pBoilGro}$ the capacity of the heating systems in W, and p_{aFlr} the floor area of the greenhouse in m^{-2}.

The total energy consumption of the heating system is calculated in $MJ \cdot m^{-2}$ as:

$$Boil_{In} = 10^{-6} \cdot \int (H_{BoilPipe} + H_{BoilGroPipe}) dt \tag{6.14}$$

Table 6.4 contains the total energy consumption of the boiler for each control algorithm. The rule-based climate controller results in the greatest energy consumption. This fact

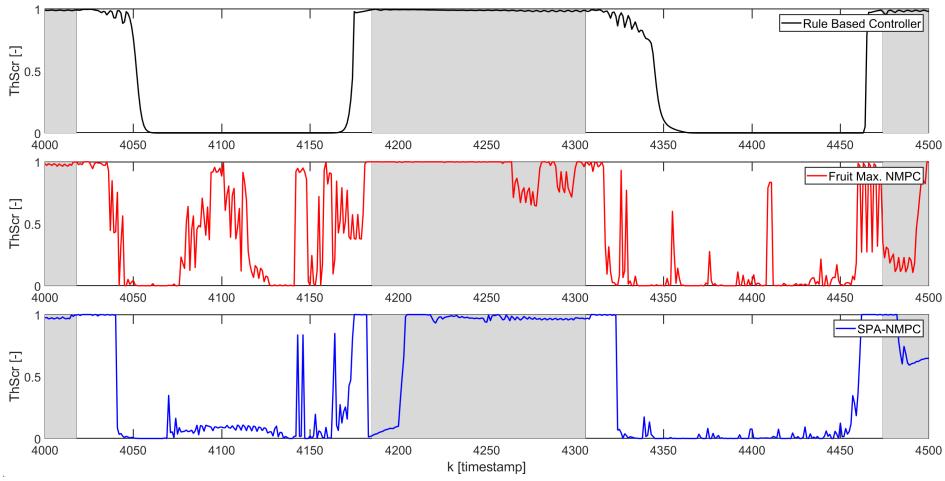

Figure 6.5: The thermal's screen action for the two consecutive days displayed in Figure 6.4. The grey (x, y) plane background represents the nighttime and the white corresponds to the daytime.

is justified by two main reasons. First, the rule-based algorithm consists of several P-type controllers which do not take into consideration either the actuator's effort or the actuator's slew rate. Second, the nighttime air temperature setpoint is set to $18.5^{o}C$ which is higher than the nighttime air temperatures indicated by the NMPCs. The SPA-NMPC algorithm steers the air temperature to a low level during the night intending to minimize the maintenance respiration and maintain non-zero carbohydrates flow from the buffer to the organs. The fruit maximization NMPC follows a similar pattern to increase energy efficiency. Considering the SPA-NMPC energy consumption as the comparison's reference point, it can be noticed that the SPA-NMPC consumed 6.31% more energy than the fruit maximization NMPC and 4.48% less energy than the rule-based controller.

Table 6.4: Demonstration of the measured total energy consumption from the boilers in $MJ \cdot m^{-2}$

Measured Heating Energy Consumption			
	SPA-NMPC	Rule-Based	Yield Max. NMPC
Measured $Boil_{In}$ $[MJ \cdot m^{-2}]$	178.5488	186.5508	167.2745

Figure 6.6 displays the air temperature trajectories for the same two, simulated, days. The figure is introduced to experimentally validate the aforementioned findings regarding the boilers' energy consumption. A noteworthy remark displayed in Figure 6.6 has to do with the reference tracking capability of the rule-based controller. As can be seen, the rule-based climate controller cannot achieve zero tracking error as it consists of pro-

portional controllers without integral terms. Another remark regards the SPA-NMPC approach and specifically the increased air temperature values appearing in the morning and late evening hours. This repetitive pattern makes the SPA-NMPC the second energy-consuming approach. The global radiation values are low during the morning and late evening resulting in the reduced cost of the actuator's effort (R_{Δ_k}) which let the boiler run at full capacity.

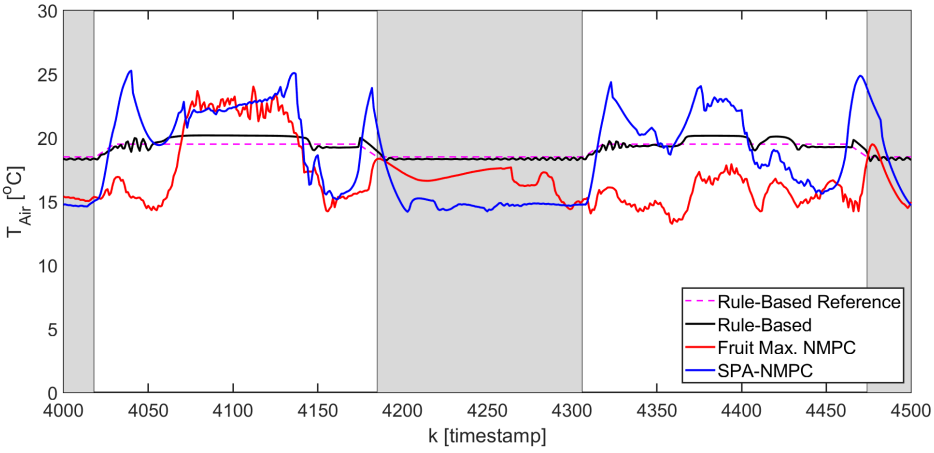

Figure 6.6: The implemented air temperature trajectories (T_{Air}) for the two consecutive days displayed in Figure 6.4. The grey (x, y) plane background represents the nighttime and the white corresponds to the daytime.

6.5.3. HARVESTED YIELD

Concerning harvesting, it is assumed that fruits are harvested when the stored carbohydrates in fruits (C_{Fruit}) have reached a certain amount (C_{Fruit}^{Max}). Hence, the harvesting rate of dry fruits is described as:

$$MC_{FruitHar} = \max\{0,\ C_{Fruit} - C_{Fruit}^{Max}\} \tag{6.15}$$

Where $MC_{FruitHar}$ in $mg_{CH_2O} \cdot m^{-2} \cdot s^{-1}$. The total harvested yield is calculated by Equation (6.16) through the integration of the harvesting rate and the introduction of the dry matter content (dmc). The dmc constant is used to convert the dry to wet harvested matter and its value was defined to 6% according to [44].

$$Yield_{Total} = 10^{-6} \cdot \frac{1}{dmc} \cdot \int MC_{FruitHar}\, dt \tag{6.16}$$

Table 6.5 presents the calculated wet yield in $kg \cdot m^{-2}$ as also the energy efficiency in $MJ \cdot kg^{-1}$. From Table 6.5 is calculated that the SPA-NMPC produced 7.02% and 4.67% more wet yield than the fruit maximization NMPC and the rule-based climate controller accordingly. Moreover, the SPA-NMPC outperformed the fruit maximization NMPC algorithm in terms of energy efficiency by 0.76% and 9.60% respectively. For reasons of completeness, it is noticed that the SPA-NMPC efficiency was regarded as the reference point of the aforementioned comparison of percentages.

Table 6.5: Demonstration of the harvested wet yield in $kg \cdot m^{-2}$ and the heating energy input needed per tomato yield in $MJ \cdot kg^{-1}$

Wet Yield and Heating Energy Efficiency			
	SPA-NMPC	Rule-Based	Yield Max. NMPC
Measured $Yield_{Total}$ $[kg \cdot m^{-2}]$	3.3928	3.2341	3.1546
Energy input needed per tomato yield $[MJ \cdot kg^{-1}]$	52.6258	57.6817	53.0257

CARBOHYDRATE BALANCE

The idea behind the proposal of the SPA-NMPC is to monitor the crop's carbohydrates production, distribution, and consumption. Therefore, it would be interesting to clarify the reason behind the increased yield production. In other words, to examine if the increased yield is just a result of the increased photosynthetic activity or if the SPA-NMPC managed to control the biochemical processes governing carbohydrates balance more beneficially. Better carbohydrate balance could mean reduced maintenance respiration, reduced growth respiration, or carbohydrate distribution focused on fruit development instead of the rest organs. Consequently, the following paragraphs are devoted to the presentation of the carbohydrate balance analysis.

According to the crop's physiology, carbohydrates are produced through photosynthesis, some of them are consumed by growth respiration and another portion is consumed by maintenance respiration. Parts of the produced carbohydrates stay in the organs (fruits, leaves, stems) and another part is harvested in the form of fruits or pruned in the form of leaves and stems. The amount of carbohydrates consumed by respiration processes is calculated by the integration of the Equations (4.13) and (4.17). Carbohydrates stored in the organs are expressed as the crop's states (See Equations (3.25), (3.26), and (3.27)). The dry harvested matter is calculated by integrating Equation (6.15). The pruning rate is defined similarly to the harvesting rate where leaves are pruned when the simulated LAI exceeds a maximum threshold value (Equation (6.17)). The total, dry, pruned carbohydrates can be found by integrating Equation (6.17). Finally, the integral of Equation 10 displayed in [20] was used for the computation of the produced carbohydrates. The presentation of the photosynthesis model is out of the scope of this study and the reader is referred to the Electronic Appendix of [12].

$$MC_{LeafHar} = max\{0,\ C_{Leaf} - C_{Leaf}^{MaX}\} \tag{6.17}$$

Table 6.6 presents the amount of produced carbohydrates per control algorithm in $[kg_{CH_2O} \cdot$

m^{-2}]. As can be seen, SPA-NMPC produced the most carbohydrates and the rule-based approach showed the second-best results.

Table 6.6: Demonstration of the total amount of carbohydrates produced by photosynthesis [$kg_{CH_2O} \cdot m^{-2}$]

Produced Carbohydrates			
	SPA-NMPC	Rule-Based	Yield Max. NMPC
Measured MC_{AirBuf} [$kg_{CH_2O} \cdot m^{-2}$]	0.9424	0.9244	0.9051

The increased photosynthetic activity of the SPA-NMPC algorithm is also depicted in Figure 6.7 which displays the photosynthesis rates of all climate controllers for the reference days. The increased photosynthetic rate of the SPA-NMPC is justified by two main factors. First, as mentioned in Section 6.5.1, the crop experiences increased PAR levels during the early morning hours entailing increased photosynthetic activity (Figure 6.7 for $k \in [4035, 4065] \cup [4324, 4355]$). Second, an increased photosynthetic rate is observed during the sun's peak hours. The reason behind this pattern is the increased canopy temperature achieved by the SPA-NMPC. In general, photosynthesis has a positive correlation with canopy temperature when $T_{Can} \in [12^oC, 30^oC]$ and $R_{ParGhSun} \geq 100$ [$W \cdot m^{-2}$].

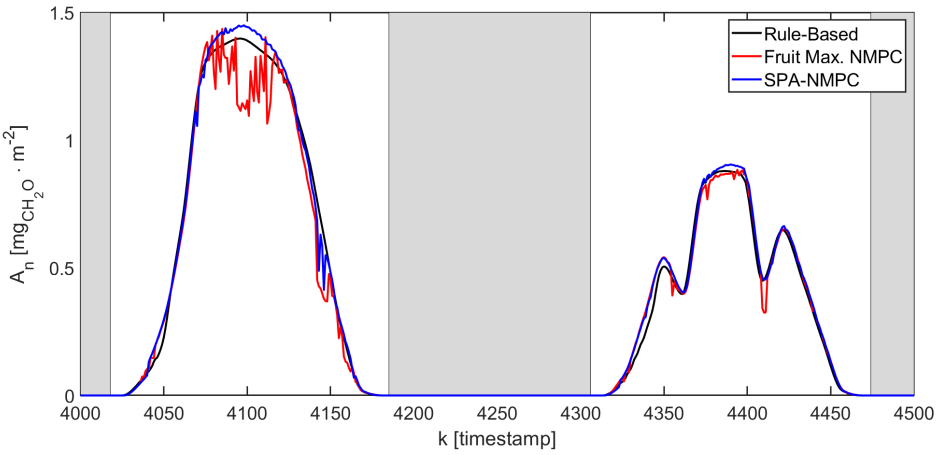

Figure 6.7: The simulated photosynthetic activity for the two consecutive days displayed in Figure 6.4. The grey (x, y) plane background represents the nighttime and the white corresponds to the daytime.

Regarding carbohydrates balance, Figure 6.8 demonstrates the distribution of the produced carbohydrates using pie charts. It can be noticed that the distribution of carbohydrates is similar in all three cases. The χ^2 test was applied to the measured results in or-

der to conclude if the differences between the control regimes are significant. The χ^2 calculation was conducted using the `crosstab` MATLAB function which returned $\chi^2 = 366$ and $p = 0.3114$ indicating that the distribution differences are not significant. Concluding, the increased yield of the SPA-NMPC occurred due to increased photosynthetic activity as the algorithm did not achieve any improved carbohydrate management.

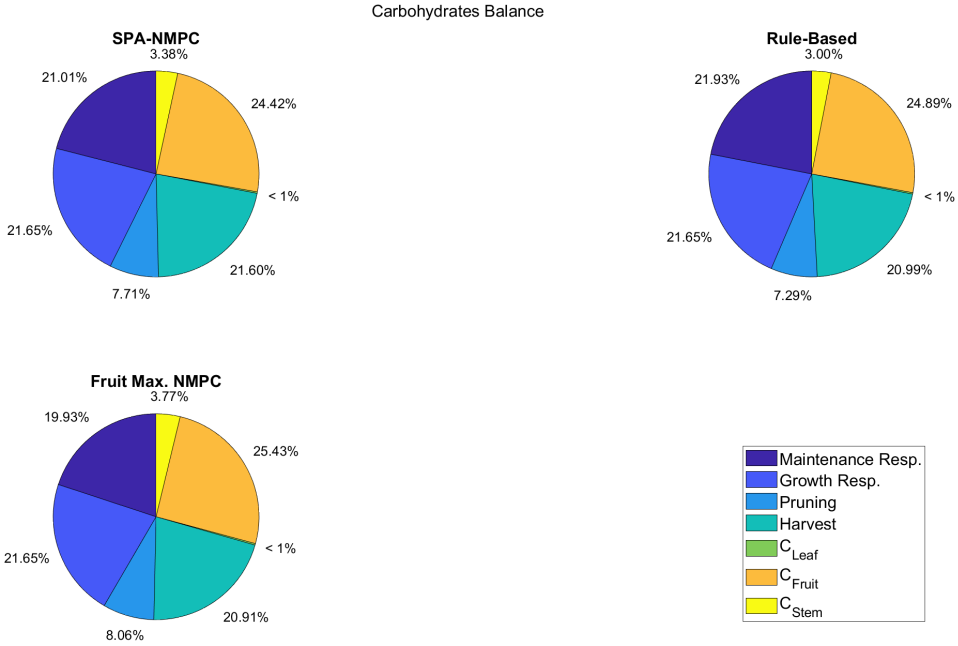

Figure 6.8: Pie charts representing each control's algorithm distribution of the produced carbohydrates. The total amount of produced carbohydrates is distributed to fulfill the respiration needs, to be stored in the leaves, fruits and stems (C_{Leaf}, C_{Fruit}, and C_{Stem}) and a part of it has been removed through pruning and harvesting.

6.5.4. EXECUTION TIME

The measured mean, minimum, and maximum execution times of the climate control algorithms are presented in Table 6.7. The rule-based climate controller showed the fastest performance as it does not require the solution of an optimization problem in order to conclude on the future control inputs. Moreover, the rule-based controller has to decide only for the next timestamp, as no control horizon is applied. The increased execution times of the NMPC algorithms are a byproduct of the computationally expensive predictive model.

A noteworthy observation is that the SPA-NMPC controller is 40.76% faster than the fruit maximization NMPC algorithm in terms of mean execution time. However, the maximum execution times measured for both NMPC algorithms declare that sometimes the algorithms do not manage to provide the control inputs on time. The control horizon

Table 6.7: Measured execution times for the SPA-NMPC, the rule-based climate controller, and the fruit maximization NMPC in *seconds*.

Measured Execution Times of the Control Regimes			
	SPA-NMPC	Rule-Based	Yield Max. NMPC
Mean Execution Time [*sec*]	461.7942	0.0079	779.6588
Minimum Execution Time [*sec*]	3.2045	0.0062	432.7669
Maximum Execution Time [*sec*]	3335.0	0.0106	1247.5

of the climate controllers are set to $N_c = 3$ (timestamps) = 900 (sec) and the maximum execution times for the SPA and the fruit maximization NMPC, accordingly, are 3335 and 1247.5 seconds. During the simulation process, computational delays have not been taken into consideration.

6.5.5. CONCLUSION & REMARKS

The SPA-NMPC concluded on the greatest tomato yield with the lowest heating energy consumption per produced unit tomato mass. Despite the fact that SPA-NMPC came second in terms of PAR above the canopy, it managed to achieve the greatest total photosynthetic activity. This control approach also managed to implement meaningful control patterns. Such patterns are the low night temperatures for the regulation of maintenance respiration, and the increased morning and evening temperatures aiming to increase $T_{Can_{24}}$ when the actuator's cost is low. It should also be noticed that SPA-NMPC managed to satisfy the soft state constraints. Based on all these facts, the canopy temperature control idea governing SPA-NMPC can be considered promising. However, there is still room for improvement in several aspects. At first, the solution to the online optimization problem can become a great issue as there are cases where the execution time was greater than the control horizon. The execution's time acceleration could be achieved by replacing the used predictive model with a simpler one and then compensating for the prediction errors. Another option would be the execution of the control algorithm in application-oriented embedded hardware. Moreover, the acceleration of the execution time would permit the use of a longer prediction horizon. A longer prediction horizon is expected to decrease the actuator's aggressive behavior when P is activated. Potential for growth also exists in the thermal's screen tuning. Smoother thermal screen effort, during the sun's peak hours, could result in increased PAR values above the canopy enhancing the crop's photosynthetic activity. In summary, the SPA-NMPC outperformed the benchmark algorithms in terms of yield and heating energy efficiency but there is

still development potential regarding execution time and PAR above the canopy.

7

CONCLUSION

7.1. SUMMARY

This thesis project is engaged with the development and testing of a SPA-based cost function able to be used as the optimization criterion of a greenhouse climate predictive controller. Moreover, it was tested if the data-driven predictor introduced in [11] was capable to be used as a system representation for predictive climate control. The examination of these subjects hopes to assist the transition from conventional to autonomous greenhouses, the insurance of food security, and the minimization of fossil fuel consumption.

Before proceeding with the analysis of the aforementioned topics, an introduction to the fundamental components and the variants of predictive control was carried out. This introduction took place to help the reader understand the meaning of the subjects in question as the current study does not necessarily refer to control engineers. Afterward, the presentation of a coupled climate-crop system satisfying the thesis' project requirements was conducted. This model and its components were considered ground truth. The climate and the crop model were used for the generation of training and testing data for the data-driven predictor. Additionally, the climate-crop model acted as the plant and the predictive model for the testing of the climate controllers.

Subsequently, it was examined how the crop's carbohydrates production, distribution, and loss can be affected through the control of crop and climate variables concluding with the formulation of the SPA-based objective function. Through this process, the necessary state constraints were also defined. Then, the SPA objective function proved its potential in an experiment introduced as a proof of concept. A SPA predictive canopy temperature controller was compared with a fruit maximization MPC concluding with similar results. Finally, the novel objective function was transformed in order to properly fit in a greenhouse climate controller.

Chapter 5 showed that despite the violation of the PE condition, the data-driven predictor can describe the greenhouse system when the testing input signals are similar to the training input trajectories. Experimentally was shown that the constantly changin dynamics of the ground truth model entail the evolution of the prediction error over time. Moreover, the use of dynamic Hankel matrices was tested concluding that their frequent update diminishes the over-time evolution of prediction errors. Multiple DeePC algorithms were implemented to examine how the PE violation affects the plant-predictor mismatches. DeePC managed to track a reachable reference trajectory. However, when it is not known that the desired trajectory can be described by the predictor, the controller considers physically unreachable trajectories as feasible solutions.

Afterward, a case study was performed to test the SPA cost function on a greenhouse climate control problem. As the SPA-DeePC was considered failed, the experiments were conducted in a SPA-NMPC formulation. For the sake of comparison, a conventional rule-based climate control algorithm and a fruit maximization NMPC illustrating the optimal yield-cost ratio were introduced as benchmarks. The comparison indexes were the input energy, yield production, and execution time. The SPA-NMPC came first in terms of yield production and heating energy efficiency despite that it was not always capable to calculate the control inputs on time. However, it was shown that the SPA-NMPC can produce meaningful control patterns and outperform the benchmark algorithms even if there is still room for improvements regarding the PAR above the canopy and the execution time.

7.2. ANSWERS TO RESEARCH QUESTIONS

This section intends to gather the answers to the research questions and sub-questions noted in Chapter 1. The answers can be found below:

- *Is the linear data-driven predictor, presented in [11], able to describe the behavior of the complex climate-crop system presented in [12] without persistently exciting training data?*

  - The data-driven predictor, presented in [11], is capable to describe the complex climate-crop system presented in [12] when the testing input signals are similar to the training input trajectories, despite the violation of the PE condition. However, the predictor fails to describe the system's behavior for inputs that do not conform with the training behavior. It was shown that the introduction of state constraints, in the prediction optimization problem, are capable to reduce the prediction errors. Nevertheless, there is no mathematical confirmation that these constraints are exhaustive and semantically they should not be placed in the control problem. Rephrasing, the data-driven predictor is able to describe the climate-crop system behavior under specific conditions.

The sub-question should be answered before proceeding with the answer to the second research question.

- *What greenhouse, climate, and crop system models should be selected for the purposes of ground truth, disturbance description, model-based, and/or data-based predictive controller design?*

  – The pursued model should be able to describe the climate's thermo-dynamics, vapor pressure, carbon dioxide concentration, and artificial lighting in order to represent a modern greenhouse. Moreover, it should properly describe the effect of the crop stage on the climate dynamics. The model should efficiently describe the effects of high and low temperatures and not only a narrow operating range. In this way, experimenting with novel climate control approaches is allowed. As the model will be used as ground truth for the simulation of SPA-based experiments, it has to be mechanistic for the proper interpretation of the results. It should also be validated permitting the support of the results. This set of requirements is satisfied by an exploratory, validated, process-based greenhouse model named `Greenlight`, presented in [12].

- *Which of the available actuators are going to be controlled and why?*

  – The boilers for the control of the pipe-rail heating system, the thermal screen as it can control energy leakages and provide shading, the roof aperture for ventilation and humidity control, and the carbon dioxide enrichment system to ensure increased photosynthetic activity. The proposed climate control algorithm should be tested on different lighting conditions but artificial lighting creates repetitive lighting patterns. For this reason, the artificial lighting actuators were excluded from the control design.

- *Which crop processes should be incorporated in the SPA-based objective function?*

  – The SPA-based objective function should monitor carbohydrate production, ensure a high degree of carbohydrate distribution from the buffer to the organs, and eliminate carbohydrate losses. To do so, stomatal conductance should be incorporated for carbohydrate production. Assimilates partitioning for the control of the distribution flow and maintenance respiration for the elimination of carbohydrate losses. Growth respiration should not be included in the loss elimination as it will negatively affect the carbohydrate flow.

- *What modifications does the SPA-based objective function need in order to fit in the NMPC greenhouse control scheme?*

  – As the SPA-based objective function is unable to generate meaningful CO_2 setpoint values, a reference CO_2 term has to be added. The SPA-NMPC has to control the greenhouse actuators and monitor their energy consumption,

thus two more terms have to be added. A quadratic term penalizing the control inputs and another quadratic cost penalizing the actuator's slew rate. Finally, a penalty function for the introduction of soft state constraints in the objective function is considered necessary.

- *Which are the optimal performance metrics to become benchmarks for the intended comparisons?*
  - The optimal performance metric for the intended comparison is the heating energy efficiency per produced mass unit as the introduction of advanced control techniques aims for the optimal exploitation of the energy resources. The produced wet yield and the heating energy consumption are also considered benchmarks as they permit us to sanity-check the efficiency index. Finally, the photosynthetic active radiation reaching the canopy is another performance index that should be exploited. The proper control of PAR above the canopy can have a significant impact on photosynthesis and as a result on the final yield. Finally, the algorithm's execution time is another performance benchmark that should be introduced to examine if the proposed approach is able to produce results on time.

- *How does the designed SPA-based climate control algorithm perform compared to a conventional rule-based climate controller and a yield maximization NMPC algorithm?*
  - The SPA-NMPC outperformed the benchmark control algorithms in terms of yield production and heating energy efficiency. Particularly, it produced 4.67% and 7.02% more wet yield than the rule-based climate controller and the fruit maximization NMPC accordingly. In terms of energy efficiency, SPA-NMPC outperformed the benchmarks by 0.76% and 9.60% respectively. Regarding heating energy consumption, the SPA-NMPC consumed 4.48% less energy than the rule-based controller and 6.31% more energy than the fruit maximization NMPC. The proposed SPA algorithm came second on the total PAR above the canopy with 0.65% less total PAR than the rule-based benchmark and 4.68% more PAR than the fruit maximization NMPC. Concerning the execution time, no fair comparison can be conducted between the NMPCs and the proportional rule-based climate controller. It should be underlined, though, that the mean execution time of the SPA-NMPC was 40.76% lower than the corresponding time of the fruit maximization NMPC. However, there were cases where both NMPC algorithms failed to provide future control inputs on time.

During Chapter 1 a side-question was formulated to examine the applicability of the SPA-NMPC in an industrial, real-world application. The question and its answer are given below:

- *Under which assumptions could the proposed approach be considered an applicable real-world application?*

– The SPA-NMPC algorithm requires the capability of conducting canopy temperature measurements in a fast, accurate, and automated manner. However, in reality such sensors are not included in the standard greenhouse equipment. In the context of this study, it was assumed that the greenhouse was equipped with thermal cameras. In addition, it was assumed that the greenhouse is a "stirred tank" and the temporal climate variable values are uniformly distributed. Finally, the crop model performs under the "big leaf" approach with perceives the canopy as a single layer. In order to use the SPA-NMPC algorithm in a real-life scenario, it should be assumed that the greenhouse is properly equipped, the predictive model incorporates the differences in spatial climate dynamics, and the objective function is properly modified to handle multiple canopy layers.

7.3. RECOMMENDATIONS FOR FUTURE WORK

- **Greenhouse-Crop Modelling:** The used ground truth model was developed under the "stirred tank" and the "big leaf" assumptions. These assumptions simplify the modeling of the crop and climate processes but they do not fit in a real-world scenario. Therefore, it is proposed to reconstruct the SPA-NMPC approach so it can handle multiple canopy layers and spatial differences in the distribution of the climate variables. Such a modification would test the applicability of the SPA-NMPC in a more realistic scenario.

- **Data-Driven System Representation:** It was shown that the data-driven predictor was capable to describe the greenhouse behavior when the given inputs conform with the training input trajectories. However, a training dataset containing the system's behavior on the whole reachable state space would require an enormous amount of data. Such a dataset would also conclude in great computation time for the predictor. A solution to this problem could be found in the Linear-Parameter Varying DeePC theory where the system representation requires the inputs and a set of parameters defining the system's operating region in order to predict the future output trajectories. This approach could also be combined with the Online DeePC algorithm to eliminate the evolution of prediction errors over time.

- **SPA-NMPC Parameter Tuning:** The SPA-NMPC controller contains several parameters for the scaling of the individual terms living in the objective function and the penalization of the control inputs and their slew rate. According to the simulated results presented in Chapter 6, the SPA-NMPC could have achieved greater PAR levels above the canopy enhancing the production of carbohydrates by the smoother control of the thermal screen. Thus, it is recommended to retune the control algorithm aiming to achieve higher photosynthetic activity during the sun's peak hours.

- **SPA-NMPC Computation Time:** The simulated experiments showed that the proposed climate control algorithm requires increased computation time which sometimes may be greater that the control horizon. The execution time can be accelerated by replacing the currently used climate-crop model with a simpler one and

adjusting the algorithm to compensate for the prediction errors. Moreover, the acceleration of the execution time would permit the evaluation of the algorithm under the use of a longer prediction horizon. Therefore, it is recommended to test the SPA-NMPC using a canopy temperature model as a predictive model combined with prediction error compensation techniques.

- **SPA-NMPC Simulated Experiments:** Because of the lack of time, the experiments presented in Chapter 6 did not run for a whole crop season. It is recommended to be repeated for a whole year.
- **Uncertainty on Exogenous Inputs:** The development and testing of the SPA-NMPC algorithm was conducted under the assumption that future exogenous inputs were identical to the implemented ones. In reality, this assumption does not always hold and mismatches occur between the predicted and the real weather. Thus, testing the algorithm in a more realistic scenario would require the introduction of uncertainty in exogenous inputs.
- **Thermal Cameras and Rule-Based Climate Control:** The system identification process of a modern greenhouse remains an open problem. As a result, rule-based climate controllers will remain the standard in industrial climate control for some time. Thus, it is recommended to explore how rule-based control could take advantage of crop signals such as stomatal conductance, canopy temperature, and mean canopy temperature. For example, rules could be applied to ensure that the canopy temperature is in its optimal range, or that the canopy temperature is greater than the dew point.

BIBLIOGRAPHY

[1] J. Cheeseman, "Food security in the face of salinity, drought, climate change, and population growth," in Dec. 2016, pp. 111–123, ISBN: 9780128018545. DOI: 10.1016/B978-0-12-801854-5.00007-8.

[2] J. Sengupta, *Declining number of farmers: The alarming trend must be reversed*, https://www.orfonline.org/research/declining-number-of-farmers-the-alarming-trend-must-be-reversed/, [Online; accessed 28-August-2022], 2013.

[3] 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, 2019, ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2019.09.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1537511019308323.

[4] Y. Hashimoto, "Recent strategies of optimal growth refulation by the speaking plant concept," *International Society for Horticultural Science (ISHS)*, 1989. DOI: https://doi.org/10.17660/ActaHortic.1989.260.5.

[5] R. Vadivambal and D. Jayas, "Applications of thermal imaging in agriculture and food industry—a review," *Food and Bioprocess Technology*, vol. 4, pp. 186–199, Feb. 2011. DOI: 10.1007/s11947-010-0333-5.

[6] I. LEINONEN, O. M. GRANT, C. P. P. TAGLIAVIA, M. M. CHAVES, and H. G. JONES, "Estimating stomatal conductance with thermal imagery," *Plant, Cell & Environment*, vol. 29, no. 8, pp. 1508–1518, 2006. DOI: https://doi.org/10.1111/j.1365-3040.2006.01528.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3040.2006.01528.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3040.2006.01528.x.

[7] T. Lawson and S. Vialet-Chabrand, "Speedy stomata, photosynthesis and plant water use efficiency," *New Phytologist*, vol. 221, Jul. 2018. DOI: 10.1111/nph.15330.

[8] S. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, no. 7, pp. 733–764, 2003, ISSN: 0967-0661. DOI: https://doi.org/10.1016/S0967-0661(02)00186-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967066102001867.

[9] C. Verhoek, H. S. Abbas, R. Tóth, and S. Haesaert, "Data-driven predictive control for linear parameter-varying systems," *IFAC-PapersOnLine*, vol. 54, no. 8, pp. 101–108, 2021. DOI: 10.1016/j.ifacol.2021.08.588. [Online]. Available: https://doi.org/10.1016/j.ifacol.2021.08.588.

[10] S. Baros, C.-Y. Chang, G. E. Colón-Reyes, and A. Bernstein, "Online data-enabled predictive control," *Automatica*, vol. 138, 2022. DOI: 10.1016/j.automatica.2021.109926. [Online]. Available: https://doi.org/10.1016/j.automatica.2021.109926..

[11] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the deepc," pp. 307–312, Jun. 2019. DOI: 10.23919/ECC.2019.8795639.

[12] D. Katzin, S. van Mourik, F. Kempkes, and E. J. van Henten, "Greenlight – an open source model for greenhouses with supplemental lighting: Evaluation of heat requirements under led and hps lamps," *Biosystems Engineering*, vol. 194, pp. 61–81, 2020, ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2020.03.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1537511020300738.

[13] E. F. Camacho and C. Bordons, *Model Predictive Control*. Springer London, ISBN: 978-0-85729-398-5. DOI: https://doi.org/10.1007/978-0-85729-398-5.

[14] P. Biswas, A. East, E. Hewett, and J. Heyes, "Chilling injury in tomato fruit," in Jan. 2017, pp. 229–278, ISBN: 9781119281245. DOI: 10.1002/9781119281269.ch5.

[15] L. Grüne, "Nominal Model Predictive Control," in *Encyclopedia of Systems and Control*, J. Baillieul and T. Samad, Eds., 5 p., to appear in Encyclopedia of Systems and Control, Tariq Samad, John Baillieul (eds.), Springer, 2014, 10 p. DOI: 10.1007/978-1-4471-5102-9_1-1. [Online]. Available: https://hal.inria.fr/hal-00944284.

[16] A. Bemporad, "Model predictive control design: New trends and tools," in *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006, pp. 6678–6683. DOI: 10.1109/CDC.2006.377490.

[17] R. Findeisen and F. Allgöwer, "An introduction to nonlinear model predictive control," Jan. 2002.

[18] M. Y. E. Ghoumari, D. Megías, J. I. Montero, and J. Serrano, "Model predictive control of greenhouse climatic processes using on-line linearisation," in *2001 European Control Conference (ECC)*, 2001, pp. 3452–3457. DOI: 10.23919/ECC.2001.7076468.

[19] M. Ellis, H. Durand, and P. D. Christofides, "A tutorial review of economic model predictive control methods," *Journal of Process Control*, vol. 24, no. 8, pp. 1156–1178, 2014, Economic nonlinear model predictive control, ISSN: 0959-1524. DOI: https://doi.org/10.1016/j.jprocont.2014.03.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959152414000900.

[20] B. Vanthoor, P. de Visser, C. Stanghellini, and E. van Henten, "A methodology for model-based greenhouse design: Part 2, description and validation of a tomato yield model," *Biosystems Engineering*, vol. 110, no. 4, pp. 378–395, 2011, ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2011.08.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1537511011001413.

[21] H. de Zwart, English, WU thesis 2071 Proefschrift Wageningen, Ph.D. dissertation, 1996, ISBN: 9789054855330.

[22] B. Vanthoor, C. Stanghellini, E. van Henten, and P. de 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, no. 4, pp. 363–377, 2011, ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2011.06.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1537511011000948.

[23] G. van Straten and E. J. van Henten, "Optimal greenhouse cultivation control: Survey and perspectives," *IFAC Proceedings Volumes*, vol. 43, no. 26, pp. 18–33, 2010, 3rd IFAC Conference in Modelling and Control in Agriculture, Horticulture and Post-Harvest Processing - Agricontrol, ISSN: 1474-6670. DOI: https://doi.org/10.3182/20101206-3-JP-3009.00004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474667015310326.

[24] S. Adams, K. Cockshull, and C. Cave, "Effect of temperature on the growth and development of tomato fruits," *Annals of Botany*, vol. 88, no. 5, pp. 869–877, 2001, ISSN: 0305-7364. DOI: https://doi.org/10.1006/anbo.2001.1524. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0305736401915240.

[25] N. Baytorun, S. Topcu, K. Abak, and H. Dasgan, "Growth and production of tomatoes in greenhouses at different temperature levels," *Gartenbauwissenschaft*, vol. 64, pp. 33–39, Jan. 1999.

[26] J. Bowyer and R. Leegood, "2 - photosynthesis," in *Plant Biochemistry*, P. Dey and J. Harborne, Eds., London: Academic Press, 1997, 49–p4, ISBN: 978-0-12-214674-9. DOI: https://doi.org/10.1016/B978-012214674-9/50003-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780122146749500035.

[27] T. T. Kozlowski and S. G. Pallardy, "Chapter 7 - carbohydrates," in *Physiology of Woody Plants (Second Edition)*, T. T. Kozlowski and S. G. Pallardy, Eds., Second Edition, San Diego: Academic Press, 1997, pp. 159–173, ISBN: 978-0-12-424162-6. DOI: https://doi.org/10.1016/B978-012424162-6/50024-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780124241626500247.

[28] R. Fischer, D. Rees, K. Sayre, Z.-M. Lu, A. Condon, and A. Saavedra, "Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies," *Crop Science - CROP SCI*, vol. 38, Nov. 1998. DOI: 10.2135/cropsci1998.0011183X003800060011x.

[29] G. Farquhar, S. von Caemmerer, and J. Berry, "A biochemical model of photosynthetic co," *Planta*, vol. 149, pp. 78–90, Jun. 1980. DOI: 10.1007/BF00386231.

[30] G. D. Farquhar and S. von Caemmerer, "Modelling of photosynthetic response to environmental conditions," in *Physiological Plant Ecology II: Water Relations and Carbon Assimilation*, O. L. Lange, P. S. Nobel, C. B. Osmond, and H. Ziegler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, pp. 549–587, ISBN: 978-3-642-68150-9. DOI: 10.1007/978-3-642-68150-9_17. [Online]. Available: https://doi.org/10.1007/978-3-642-68150-9_17.

[31] S. Caemmerer, *Biochemical Models of Leaf Photosynthesis*. Jan. 2000, ISBN: 9780643103405. DOI: 10.1071/9780643103405.

[32] C. Bernacchi, J. Bagley, S. Serbin, U. Ruiz Vera, D. Rosenthal, and A. VanLoocke, "Modeling c3 photosynthesis from the chloroplast to the ecosystem," *Plant Cell and Environment*, Apr. 2013. DOI: 10.1111/pce.12118.

[33] Q. Liao, R. Ding, T. Du, S. Kang, L. Tong, and S. Li, "Stomatal conductance drives variations of yield and water use of maize under water and nitrogen stress," *Agricultural Water Management*, vol. 268, p. 107 651, 2022, ISSN: 0378-3774. DOI: https://doi.org/10.1016/j.agwat.2022.107651. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378377422001986.

[34] Y. P. Wang and P. G. Jarvis, "Influence of shoot structure on the photosynthesis of sitka spruce (picea sitchensis)," *Functional Ecology*, vol. 7, no. 4, pp. 433–451, 1993, ISSN: 02698463, 13652435. [Online]. Available: http://www.jstor.org/stable/2390031 (visited on 08/07/2022).

[35] T. N. Buckley, "Modeling Stomatal Conductance," *Plant Physiology*, vol. 174, no. 2, pp. 572–582, Jan. 2017, ISSN: 0032-0889. DOI: 10.1104/pp.16.01772. eprint: https://academic.oup.com/plphys/article-pdf/174/2/572/38273530/plphys_v174_2_572.pdf. [Online]. Available: https://doi.org/10.1104/pp.16.01772.

[36] C. Stanghellini, "Transpiration of greenhouse crops : An aid to climate management," English, WU thesis 1152 Proefschrift Wageningen, Ph.D. dissertation, 1987.

[37] A. de Koning, "Development and dry matter distribution in glasshouse tomato : A quantitative approach," English, Ph.D. dissertation, 1994, ISBN: 9789054853329.

[38] E. HEUVELINK, "Tomato growth and yield: Quantitative analysis and synthesis," *Wageningen: Wageningen University*, 1996.

[39] I. Markovsky, J. Willems, S. Van Huffel, and B. De Moor, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*, English, Number 11 in Monographs on Mathematical Modeling and Computation. SIAM, Mar. 2006, ISBN: 978-1-4471-2227-2.

[40] J. Willems, P. Rapisarda, Markovsky, and B. De Moor, "A note on persistency of excitation," English, *Systems Control Letters*, vol. 54, no. 4, pp. 325–329, Apr. 2005, ISSN: 0167-6911. DOI: 10.1016/j.sysconle.2004.09.003.

[41] D. Katzin, L. F. Marcelis, and S. van Mourik, "Energy savings in greenhouses by transition from high-pressure sodium to led lighting," *Applied Energy*, vol. 281, p. 116 019, 2021, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2020.116019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261920314628.

[42] B. Vanthoor, E. van Henten, C. Stanghellini, and P. de Visser, "A methodology for model-based greenhouse design: Part 3, sensitivity analysis of a combined greenhouse climate-crop yield model," *Biosystems Engineering*, vol. 110, no. 4, pp. 396–412, 2011, ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2011.08.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1537511011001425.

[43] D. Katzin, *Dymoma - dynamic modelling for matlab*, English, Mar. 2020. DOI: 10.5281/zenodo.3697908.

[44] D. Katzin, "Energy saving by led lighting in greenhouses: A process-based modelling approach," English, WU thesis 7846 Includes bibliographical references. - With summaries in English, Dutch, and Hebrew, Ph.D. dissertation, Wageningen University, 2021, ISBN: 9789463957649. DOI: 10.18174/544434.

[45] A. Baghban, M. Bahadori, J. Rozyn, *et al.*, "Estimation of air dew point temperature using computational intelligence schemes," *Applied Thermal Engineering*, vol. 93, pp. 1043–1052, 2016, ISSN: 1359-4311. DOI: https://doi.org/10.1016/j.applthermaleng.2015.10.056. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1359431115011035.

[46] M. G. Lawrence, "The relationship between relative humidity and the dewpoint temperature in moist air - a simple conversion and applications," *Bulletin of the American Meteorological Society*, vol. 86, pp. 225–233, 2005.

[47] O. A. Alduchov and R. E. Eskridge, "Improved magnus form approximation of saturation vapor pressure," *Journal of Applied Meteorology*, vol. 35, pp. 601–609, 1996.

A

ASSUMPTIONS

Assumptions are an indivisible part of research as they provide a basis for developing research questions and interpreting results. Particularly, assumptions are used to fill the knowledge gaps, guide the research design and limit the scope of the study. The scope of this appendix is the presentation of all assumptions made in the context of this project.

1. *The greenhouse air is considered to be a "perfectly stirred tank", assuming no spatial differences in temperature, CO2, and vapor concentration.*

   This assumption allows all fluxes to be expressed in units of energy or mass per m^{-2} of the greenhouse floor. Furthermore, it simplifies climate-crop modeling and climate control, in any other case, modeling, and control should incorporate fluid dynamics.

2. *The greenhouse is divided into two main compartments. The top compartment contains the air volume above the level of the screens and the main compartment consists of the volume below the screens.*

   According to greenhouse climate datasets, the variation of air temperature, VPD, and CO_2 concentration on the top compartments differs from the variation in the main compartment. The introduction of this assumption limits the scope of climate control to the variables living in the greenhouse's main compartment.

3. *Crop modeling has been implemented based on the big-leaf approach.*

   The big-leaf approach is a modeling framework used in ecosystem ecology to estimate the fluxes of energy, water, and carbon dioxide between the land surface and the atmosphere. The advantage introduced is that the ecosystem-level fluxes are estimated by averaging the properties of the leaves over the entire canopy. In other words, the crop model works at the scale of the whole plant, rather than considering individual leaves separately. In this way, the climate controller considers a single canopy layer for the measurement of the canopy temperature.

4. *The simulated greenhouse system contains thermal cameras as temperature sensors.*

   This assumption has great importance as this thesis project explores ways to exploit the advantages of thermal cameras in climate control. The introduction of thermal cameras lets us consider that all surface temperatures are measurable. Surface temperatures (floor, soil, cover, screens, etc) are states of the ground truth climate-crop model. As a result, the climate controller can measure the canopy temperature and the need for a state observer is eliminated.

5. *The CO_2, VPD, and carbohydrate related states are considered measurable.*

   The scope of this project is the design of a novel SPA-based objective function for predictive climate control. Another objective is to examine if the ground truth climate-crop model can be effectively represented by data. Because of a lack of time, the development of a state observer for the greenhouse system was not possible. Thus, it should be assumed that the simulated greenhouse is equipped with all necessary sensors in order to consider The CO_2, VPD, and carbohydrate states measurable.

B

Dew Point

The dew point of saturated air corresponds to the amount of water vapor in the air [45]. In other words, the dew point is the temperature where for constant air pressure and water content the air would become saturated with water vapor. When the air temperature is below the dew point, the water vapor condenses and forms liquid water.

The dew point can be calculated using the Magnus-Tetens formula ([46]) as shown below:

$$T_s = \frac{b \cdot (ln(\frac{RH}{100}) + \frac{a \cdot T}{b+T})}{a - ln(\frac{RH}{100}) + \frac{a \cdot T}{b+T}} \tag{B.1}$$

Where a and b the Magnus coefficients, RH the relative humidity, T the temperature in ${}^oC$ and T_s the dew point in ${}^oC$. For the dew point calculation, the coefficients proposed by Alduchov and Eskridge are used where $a = 17.625$ and $b = 243.04$ $[{}^oC]$ [47].

In the context of this study, is has to be ensured that the canopy temperature achieved by the SPA-NMPC algorithm is always greater than the dew point. In this way, it is guaranteed that stomata can operate properly as no liquid water is formed on the leaf surface. Based on the post-simulation analysis, it occurred that the canopy temperature was always greater than the dew points calculated by Equation (B.1). Figure B.1 is introduced to visualize the aforementioned outcome.

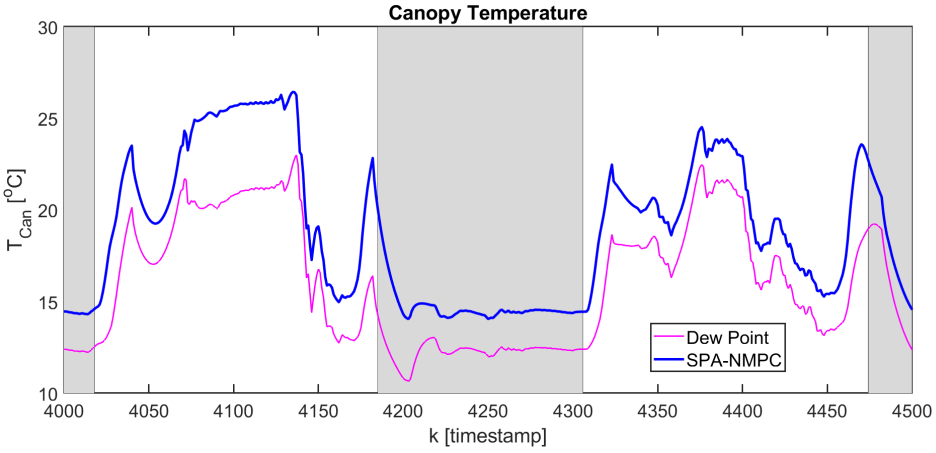

Figure B.1: The canopy temperature and the dew point trajectories for the two consecutive days presented in Figure 6.4. The first day corresponds to a sunny and the second on a cloudy one. The grey (x, y) plane background represents the nighttime and the white corresponds to the daytime.

C

SPA-NMPC Points to Be Improved

This appendix is introduced in order to depict the simulated greenhouse actuator inputs and the climate-crop variables for two consecutive days. The first day represents a sunny day with relatively high incident radiation and the second day is a cloudy day. Figure C.1 intends to demonstrate the points on which the SPA-NMPC controller can be improved.

In $k = 4040$ and $k = 4322$ the air temperature and as a result, the canopy temperature starts decreasing. This is an effect of the concurrent thermal screen closure and the heating's boiler deactivation. As the incident radiation is still on a low level, limited energy input is provided by the sun, and the greenhouse loses thermal energy (for $k \in [4040, 4055]$ and $k \in [4322, 4360]$) till the sun radiation increases or the boiler is activated again. This event could be prevented by the smoother closure of the thermal screen. A similar event takes place for $k = 4137$ where the greenhouse loses energy for the next two hours ($k \in [4137, 4163]$). In this case, the energy is lost because of the ventilation. However, this energy loss is triggered by the fact that the relative humidity reaches its upper limit. As a result, the P function is activated and the controller permits more aggressive behavior. In timestamps $k = 4180$ and $k = 4480$ the greenhouse starts losing thermal energy again. This time the reason has to do with the fact that the SPA-NMPC decides to implement low canopy temperatures during the nighttime. The main problem, in this case, is that the controller was heating the greenhouse during the last hours of the day and finally, this energy was wasted. A solution to this problem could be the use of a longer prediction horizon or the introduction of a different mechanism for the classification between day and night time.

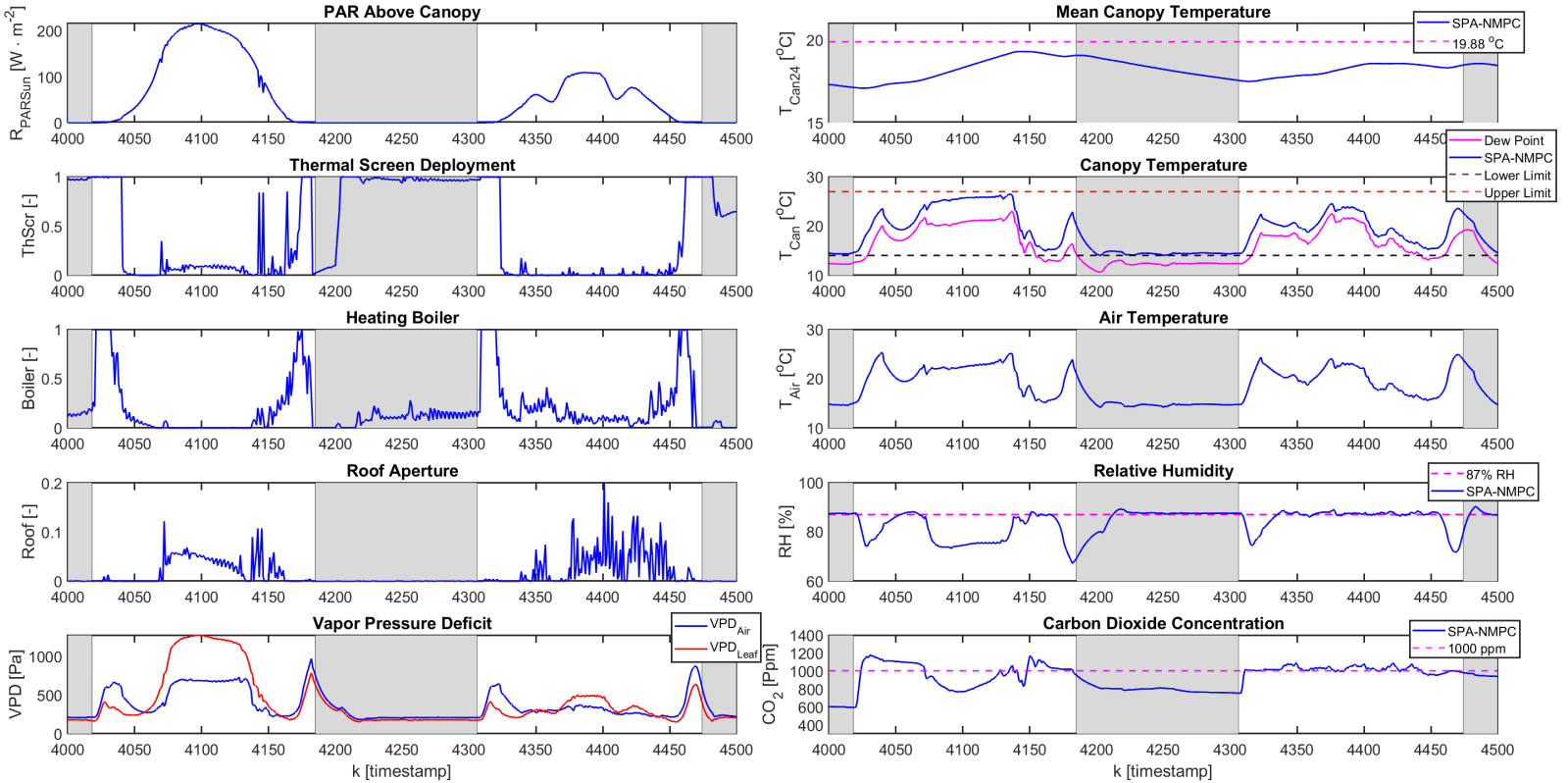

Figure C.1: Simulated greenhouse actuator inputs and climate/crop variable for two consecutive days. The first day corresponds to a sunny and the second on a cloudy one. The grey (x, y) plane background represents the nighttime and the white corresponds to the daytime.