



Identification of Plant Stress in Greenhouses by Remote Sensing of Dynamic Plant Fluorescence Response

A thesis within signal processing and system identification of biological systems

Master's thesis in Biomedical Engineering

Martin Granström
Philip Jansson

MASTER'S THESIS 2017:EX030

Identification of Plant Stress in Greenhouses by Remote Sensing of Dynamic Plant Fluorescence Response

A thesis within signal processing and system identification of
biological systems

Martin Granström
Philip Jansson



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Research and Development, Heliospectra AB
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2017

Identification of Plant Stress in Greenhouses by Remote Sensing of Dynamic Plant
Fluorescence Response

A thesis within signal processing and system identification of biological systems

Martin Granström

Philip Jansson

© Martin Granström, Philip Jansson, 2017.

Supervisor: Johan Lindqvist, Lukas Wikander, Heliospectra AB

Examiner: Torsten Wik, Department of Electrical Engineering

Master's Thesis June, EX030/2017

Department of Electrical Engineering

Research and Development, Heliospectra AB

MISTRA Innovation research program

Chalmers University of Technology

SE-412 96 Gothenburg

Telephone +46 31 772 1000

Cover: Plant and LED-bar

Typeset in L^AT_EX

Gothenburg, Sweden 2017

Abstract

As a result of a continuing increase in population the UN predicts that a 70% increase in worldwide food production, by the year 2050, is needed. This puts pressure on the horticulture industry to increase its yield to cope with the increasing demands. A potential way to increase the yield of crops grown in greenhouses is to implement a biological feedback control system that utilises LED-lights that can alter its spectral properties, and use this to diagnose the plants and adjust the lamp light depending on the status of the plant. The plant emit a light called chlorophyll fluorescence, where the slow chlorophyll fluorescence dynamics, induced by light intensity changes in the LED light, have been shown to provide information about its current stress level. This plant light response can be measured and the ultimate goal is to use it as a feedback signal in a controller concept. The problem is that the fluorescent signal is weak and sometimes drowns in reflected sunlight, and there is currently no robust method to extract the fluorescence signal when the sunlight is present. The goal of the work presented is to, by the help of various signal processing methods, extract the correct chlorophyll fluorescence signal from a plant in presence of sunlight. Two main methods have been studied; linear curve fitting and parametric modelling. We found that the curve fitting method is too sensitive to noise, rendering it useless in most situations where the sunlight intensity it high. With parametric modelling, two ways of performing the numeric search for the system parameters were used; Prefiltering and modified damped Gauss-Newton algorithms, where the latter seems to be able to extract the correct signal in a lot of cases, sometimes even at high sunlight intensities. This means that it is possible to extract the chlorophyll fluorescence from plants in a cheap and effective way even when sunlight is present.

Keywords: Plant stress, Chlorophyll fluorescence, Sensor fusion, Signal processing, System identification, MATLAB, Simulink

Sammanfattning

Världsbefolkningens fortsatta ökning har lett till att FN har förutspått att en ökning på 70% av matproduktion måste ske fram till år 2050 för att kunna hålla jämna steg. Detta sätter stor press på hortikultur-industrin att öka sin skörd för att kunna matcha det ökade behovet. Implementering av ett biologiskt återkopplat reglersystem, som utnyttjar LED-ljus speciella egenskaper, kan användas till att diagnostisera planter genom att styra ljusnivåer och ljusspektrum. Planter ger ifrån sig ljus som kallas klorofyllfluorescens, till följd av ljusexponering. Hos en planta har klorofyllfluorescensen visat sig kunna tillhandage information om dess hälsa. Denna ljussignal som plantorna avger kan mätas och det finns en ambition att använda denna som återkopplingssignal i ett reglersystem. Problemet är dock att fluorescens-signalen ibland drunknar i störningar, främst reflekterat solljus, när mätningar görs. För närvarande finns inget billigt eller tillräckligt robust system som kan extrahera fluorescens-signalen när solljus är närvarande. Målet med detta projekt är därför att, med hjälp av olika signalbehandlingsmetoder, extrahera en korrekt klorofyll fluorescens-signal från en växt. Två huvudmetoder har använts; linjär kurvanpassning och parametrisk modellering. Kurvpassningsmetoden är för känslig för brus vilket gör den oanvändbar i de flesta situationer där det är hög solljusintensitet. När det kommer till parametrisk modellering så har två olika sätt för att beräkna de varierande systemparametrarna använts; Förfiltrering och en dämpad Gauss-Newton algoritmen, där den senare metoden har visat sig kapabel till att få ut en korrekt signal. ibland även vid starkt solljus. Det betyder att det är möjligt att extrahera klorofyllfluorescens-signalen från växter på ett billigt och effektivt sätt, till och med när solljus är närvarande.

Keywords: Plantstress, Klorofyllfluorescens, Sensorfusion, Signalbehandling, Systemidentifiering, MATLAB, Simulink

Preface

This work is for a Master of science degree in Electrical engineering and has been conducted at Heliospectra AB in a collaboration with Chalmers University of Technology. The research took place during the period January 2017 - June 2017 and was supervised by Torsten Wik, Professor in Automatic control at Chalmers, Johan Lindqvist, R&D Engineer at Heliospectra AB and Lukas Wikander, R&D Engineer at Heliospectra AB.

Acknowledgements

We would like to send out our gratitude to both Johan Lindqvist and Lukas Wikander for, in a pedagogical way, explaining theories and methods that were crucial for this thesis. Also for always being available to answer questions and brainstorming ideas. We would also like to thank Ida Fällström and Grazyna Bochenek for always being available to answer questions regarding biology in a clear and understandable way, even though our biological knowledge prior to this work was very limited. We would also like to thank Torsten Wik and Anna-Maria Carstensen for giving meaningful insights and ideas at crucial points of the project. We would also like to send appreciation to Karel Keesman and Lennart Ljung for, in spite of no initial involvement giving expert advise in a friendly and helpful manner. Special thanks to Daniel Bänkestad for proofreading this report as well as lending us his desk at Heliospectra. Finally, we would like to thank Heliospectra and all its employees, not only for giving us this opportunity but also for making us feel at home in their office.

Martin Granström, Philip Jansson, Gothenburg, 12/6/17

Nomenclature

OE	Output Error
HPS	High Pressure Sodium
ARX	Auto Regressive with eXogenous input
ARMAX	Auto Regressive Moving Average with eXogenous input
PEM	Prediction Error Method
LSE	Least Squares Error
IV	Instrumental Variable
dGN	damped Gauss-Netwon
ChFl	Chlorophyll Fluorescence
G tube	Gershun tube
IT	Integration Time
LED	Light Emitting Diode
PAR	Photosynthetically Active Radiation
PFD	Photon Flux Density
DFRA	Dynamic Fluorescence Response Analysis
FS	Filtered Spectrum
FLS	Filtered Line Spectrum

The use of hat over variables, e.g. $\hat{\theta}$, indicate that it is an estimate.

Note that, in the pdf text, the expansion of the acronyms can be seen by hovering over said acronym.

Contents

List of Figures	xii
List of Tables	xvi
1 Introduction	1
1.1 Background	2
1.2 Aim	3
1.3 Objectives	3
2 Theory and related work	5
2.1 Electromagnetic spectrum	5
2.2 Photosynthesis and fluorescence	6
2.3 System identification	6
2.4 Black box modelling	7
2.5 Least squares	8
2.6 Output error model	8
2.7 Prediction error methods	9
2.7.1 Prefiltering algorithm for OE systems	9
2.7.2 Damped Gauss-Newton	9
2.8 Linear curve fitting	11
2.9 Bode plot	11
2.10 Fit percentage and performance index	12
2.11 Related work	12
2.11.1 Dynamic Fluorescence Response Analysis	12
2.11.2 Methods for estimating chlorophyll fluorescence response	13
2.11.3 Measurement model	14
3 Method	16
3.1 Equipment	16
3.2 Data evaluation	17
3.2.1 PSMT evaluation	17
3.3 Filtering	18
3.4 Linear curve fitting implementation	19
3.5 Parametric modelling of the plant system	20
3.6 Assumptions and analysis	22
3.7 Data collection - simulations	28

3.7.1	31st of January 2017	28
3.7.2	8th of March 2017	29
3.7.3	4th of April 2017	30
3.8	Simulation setup	30
3.8.1	Simulation results visualised	32
3.9	Data collection - realistic test environment	33
3.9.1	12th of April 2017 - Greenhouse experimental setup	35
3.9.2	26th of April 2017 - Greenhouse experimental setup	36
4	Results	38
4.1	Curve fitting assumption analysis	38
4.2	Parametric modelling assumption analysis	42
4.3	Simulation results	45
4.3.1	Simulation experiment 31/1/2017	45
4.3.2	Simulation experiment 8/3/2017	47
4.3.3	Simulation experiment 4/4/2017	51
4.4	Greenhouse results	53
4.4.1	Greenhouse Experiment 1 and 2 - 12/4/2017	53
4.4.2	Greenhouse experiment 3 - 26/4/2017	62
4.5	PSMT evaluation results	69
4.5.1	Curve fitting	69
4.5.2	Prefiltering algorithm	70
4.5.3	dGN algorithm	71
5	Discussion	72
5.1	Simulation data analysis	72
5.2	Experimental data analysis	73
5.3	Greenhouse setup - sources of error	73
5.4	Quota	73
5.5	Different methods	74
5.6	Differences between simulation and experiment	75
5.7	Difference between sunlight intensities	76
5.8	Future work	76
6	Conclusion	78
	Bibliography	79
A	Appendix	I
B	Appendix	II

List of Figures

1.1	A concept of a future system that uses the Chlorophyll Fluorescence (ChlF) from the plants to feed back information in order to adjust lamp light intensity. <i>Source: Torsten Wik</i>	3
2.1	A visual representation of the spectrum of the sunlight radiation at a certain point in time. The unit <i>Counts</i> , represents the number of incoming photons during a certain amount of time called the Integration Time (Integration Time (IT)).	5
3.1	An example of the fluorescence response over time with different stress levels.	18
3.2	The bode plot for the Butterworth filter used.	19
3.3	The spectrum around the ChlF band. The fluorescence is considered to be negligible above 800 nm in the estimation of the sunlight.	20
3.4	The system, showing the canopy, inputs and output. Where $l[n]$ is the lamp light, $d[n]$ is the sunlight between 735-755 nm, $d_{PAR}[n]$ is the sunlight within Photosynthetically Active Radiation (PAR), $u[n]$ is the input to the plant, consisting of $l[n]$ and $d_{PAR}[n]$, $y_f[n]$ is the fluorescence, ρ is the reflectance, $g[u[n], n]$ is a model of the canopy and $y[n]$ is the measured output.	21
3.5	Example data visualising the difference between the static ρ in red and the dynamic $\rho(t)$ in blue.	23
3.6	Normalised $q(t)$ between dynamic $\rho(t)$ and static ρ from Figure (3.5).	24
3.7	$q(t)$ for all wavelengths during one sample. Blue line shows $q(t)$, red mark shows where the fluorescence band starts, yellow mark shows where the fluorescence band ends and where the substitute band starts and the purple mark shows where the substitute band ends.	25
3.8	A comparison of $q(t)$ between band 735-755 nm and band 800-820 nm.	26
3.9	Further estimations of $q(t)$ for the 735-755 nm band.	27
3.10	Experimental setup with dill plants and spectrometers.	28
3.11	The experimental setup used for data collection on the 8th of March 2017.	29
3.12	Lamp intensity and corresponding fluorescence response, with no sunlight present, was used as the verification data.	30

3.13	The fluorescence response from both the lamp light and sunlight put together to simulate a real fluorescence response. Note the Light Emitting Diode (LED)-light step excitation at 20 seconds, induced by a change in the lamp light.	31
3.14	The raw sensor input from which the fluorescence response from Figure (3.12) need to be extracted.	32
3.15	A visual representation of the sunlight and background light produced by the LED-lamp. Both spectra produced the same Photon Flux Density (PFD) in PAR	34
3.16	The stall used in the greenhouse experiments, note that the tarpaulin is not included in this Figure.	36
4.1	Ratio of 735-755 nm to 800-820 nm of the reflected sunlight in the experiment in Section. 3.7.1.	39
4.2	Incoming sunlight from 10:16 to 11:16, see data collection from Section 3.7.1	39
4.3	Ratio of reflected 735-755 nm and 800-820 nm from the second data collection, see section 3.7.2	40
4.4	Incoming sunlight from 11:18 to 12:18, see data collection from Section 3.7.2	41
4.5	The ratio is clearly depending on the sunlight intensity. It decreases when the sunlight gets strong and increases when the sunlight decreases.	41
4.6	Incoming sunlight from 14:06 to 15:06, see data collection 3.7.3	42
4.7	ρ for the fluorescent wavelength band during the one hour of measurements made in the end of January.	43
4.8	ρ for the non-fluorescent wavelength band during the one hour of measurements made in the end of January.	44
4.9	ρ for the fluorescent wavelength band during the one hour of measurements made in March (no plants). Note the peak between 100 and 300 seconds, which is probably the result of a shadow passing over sensor M1.	44
4.10	Incoming and reflected light integrated over the 735-755 nm band with incoming light in PAR at an average of $40 \mu\text{mol m}^{-2} \text{s}^{-1}$	45
4.11	The samples are sorted, in rising order, with respect to its fit percentage. The error shows little correlation to the fit. Mean fit: 74,45% Median fit: 81,90%	46
4.12	A gathering of all the fits and corresponding errors from the experiment in Section 3.7.1 when using the prefiltering algorithm, using Filtered Spectrum (FS). The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 87,64% Median fit: 96,10%	46
4.13	A gathering of all the fits and corresponding errors from section 3.7.1 when using Filtered Line Spectrum (FLS). The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 88,61% Median fit: 95,94%	47
4.14	Measured light intensities from experiment in Section 3.7.2	48

4.15	Results of performance of curve fitting on the 3.7.2 data. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 2,08% Median fit: 0%	49
4.16	A gathering of all the fits and corresponding errors from the experiments in Section 3.7.2, using FLS. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 85,74% Median fit: 93,18%	49
4.17	A gathering of all the fits and corresponding errors from the experiments in Section 3.7.2, using FS. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 93,72% Median fit: 96,23%	50
4.18	Measured light intensities from experiment in Section 3.7.3	51
4.19	A gathering of all the fits and corresponding errors from 3.7.3. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 0,63% Median fit: 0%	51
4.20	A gathering of all the fits and corresponding errors from the experiments in Section 3.7.3, using FLS. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 13,72% Median fit: 0%	52
4.21	A gathering of all the fits and corresponding errors from 3.7.3, using $q(t)$ for band 800-820 nm (without FLS or FS). The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 47,56% Median fit: 52,54%.	52
4.22	Collection of all four test measurements in Experiment 1. The fluorescence response is hard to distinguish from the raw data. There were five minutes between each test, starting with (a)	53
4.23	Collection of all four reference measurements in Experiment 1. The fluorescence response is relatively easy to distinguish from the raw data. There were five minutes between each data collection start, starting with (a)	54
4.24	The four step responses and corresponding OE430 response from the reference measurements. In order from first to last.	55
4.25	Results from the Curve fitting estimations of the Experiment 1 data. The results are shown in both time domain and frequency domain.	56
4.26	Experiment 1 results when using the Prefiltering algorithm.	56
4.27	Results from the dGN algorithm estimation of the Experiment 1 data.	57
4.28	Collection of all four test measurements in Experiment 2. The fluorescence response is much easier to distinguish from the raw data in Experiment 2 because of the low sunlight intensity. There were five minutes between each data collection, starting with (a)	58
4.29	Collection of all four reference measurements in Experiment 2. The fluorescence response is clearly visible in the raw data. There were five minutes between the start of each data collection, starting with (a)	59

4.30	The four step responses and the corresponding OE540 responses for the reference measurements in Experiment 2. In order from first to last.	60
4.31	Results from the Curve fitting estimations for Experiment 2 data. . .	61
4.32	Results from the Prefiltering algorithm estimations for Experiment 2 data.	61
4.33	Results from the dGN algorithm estimations for Experiment 2 data .	62
4.34	Collection of the six test measurements in Experiment 3. The fluorescence response is hardly distinguishable in the raw data for any of the collected data sets. There were five minutes between the start of each data collection, starting with (a)	63
4.35	The four reference measurements in Experiment 3 taken directly after the tarpaulin had been pulled down with five minutes between each measurement starting with (a)	64
4.36	The final four reference measurements where the plant had been acclimated to total darkness. These final measurements were made approximately an hour after the sunlight tests. The data collections are made directly after one another, i.e. 80s between each start. . . .	65
4.37	The first four step responses and the corresponding OE430 responses from the reference measurements. In order from first to last.	66
4.38	The last four step responses and their corresponding OE430 responses from the reference measurements. In order from first to last.	67
4.39	Results from the Curve fitting estimations to the data from Experiment 3.	68
4.40	Results from the Prefiltering algorithm estimations to the data from Experiment 3.	68
4.41	Results from the damped Gauss-Newton (dGN) algorithm estimations for the data in Experiment 3.	69
4.42	The PSMT evaluation for the Curve fitting method. For the data with lower sunlight intensity there was a much higher chance of observing a PSMT shape of the estimation.	70
4.43	The PSMT evaluation for the Prefiltering algorithm. The result for test 3 where the background sunlight intensity was strong caused bad results with the Prefiltering algorithm	70
4.44	The PSMT evaluation for the dGN algorithm. This algorithm performed by far the best results.	71
B.1	The ratio is clearly depending on the sunlight intensity. It decreases when the sunlight gets strong and increases when the sunlight decreases. II	II

List of Tables

3.1	Spectrometer specifications	16
3.2	Spectrometer setup for the experiment performed on the 31st of January.	28
3.3	Spectrometer setup for the experiment performed on the 8th of March.	29
3.4	Spectrometer setup for the experiment performed on the 4th of April.	30
3.5	Spectrometer setup for the experiments performed on the 12th of April.	35
3.6	Spectrometer setup of the experiment performed on the 26th of April.	36
4.1	Temperatures in Experiment 1.	54
4.2	Temperatures for Experiment 2.	60
4.3	Temperatures for Experiment 3 with background light.	65

1

Introduction

TO COPE with the increase in population, the United Nations have predicted that a 70% increase in worldwide food production needs to be reached by the year 2050¹. This will put pressure on the crop growing greenhouses. Traditional greenhouse lighting comes from High Pressure Sodium (HPS) lamps which are used to grow crops on a large scale. The crops can be sensitive to the wrong amount of light causing stress, which can lead to worsened plant health. Too high exposure causes photoinhibition, which affects the photosynthetic capacity of plants in a negative way².

Heliospectra AB is a company focused on developing LED-lights and specialises in producing different spectral light for growing plants. The main difference between LED- and HPS-lights is that the latter have a fixed spectrum that is not customizable, while LED lights can have diodes that can alter a spectrum's magnitude, for given wavelengths. During the last decade the efficiency and usefulness of LED based lights has been vastly improved. As more food is needed, there is also an increased desire to decrease power consumption during production. The power consumption can be up to 50% less with LED lights when comparing to HPS-lights while still outperforming in terms of spectral customisation³.

Previous work has concluded that plant health can be determined by evaluating the Chlorophyll Fluorescence (ChlF) response to changes in emitted light^{4,5,6}. After a plant is exposed to light, three things can happen; the light is transmitted through the leaves, the light is reflected off the leaves or the plant absorbs the light, where most of it is used in photosynthesis and some is re-emitted as ChlF. ChlF occurs when the absorbed light starts a process where the chlorophyll molecules return from higher energy levels, and therefore emit light. ChlF contains information about the plant physiology and can thereby be used to determine the health of plants⁷.

Studies have shown that the spectral distribution of light affects the biomass of the plants⁸. The LED lights developed at Heliospectra AB are able to not only utilise a more diverse light spectrum than the standard HPS lamps, which can improve plant health and biomass, but also consumes less power. This thesis work is a continuation on work that revolves around trying to use the plant health status to adapt the lamp power to yield a better and healthier plant, as well as lower energy costs, and therefore provide a more environmentally friendly way of growing crops. Other benefits would be the ability to automatically control harvest dates and growth trajectories.

1.1 Background

Today's technology used for determining a plants health operates either on-leaf or remotely. The on-leaf technology is very time and resource consuming. A LED that excites a pulse of light is put on top of the leaf, initiating a ChlF response where its development over time is measured and evaluated. Some measurement techniques also require the plant in question to be in a dark-adapted state, like the Fv/Fm method⁹. There are ways to measure a ChlF response remotely, and one of those is the τ -LIDAR. LIDAR uses a laser to induce a fluorescence response from a remote location^{10,11}. The major drawback is its high price¹², which makes it non-viable in a commercial greenhouse. Also, this is not a sufficient technique to use in larger plantations since it can only get a handful of results from different plants within a limited time. It would be of great value to be able to probe larger sections of plants at once during any time of the day to determine their status.

Work is being made at Heliospectra AB with the goal of producing a control system that will be able to control the light intensity of the LED lights with respect to plant health. Figure (1.1) depicts the concept around the inner workings of such a control system. In order to monitor and adjust light intensity in the LED light, there must be a good way of measuring the ChlF emitted from the plants with a sensor near the lamp to determine whether the plants need any change in the light. The problem with measuring the fluorescence signal lies in the many disturbances present, mainly sunlight. The direct and reflected sunlight is picked up by the sensor and makes it hard to distinguish what is disturbances and what is fluorescence. The intensity of the reflected sunlight alone can easily be thousandfold of the intensity of the fluorescence signal.

This report will revolve around research regarding extraction of the fluorescence signal in the presence of varying amount of light with the intent of improving one or multiple methods to do this in an efficient manner. More specifically, the research question is:

- Is it possible to extract a Chlorophyll fluorescence response, induced by a light intensity change of a LED-lamp, using signal processing methods when sunlight is present?

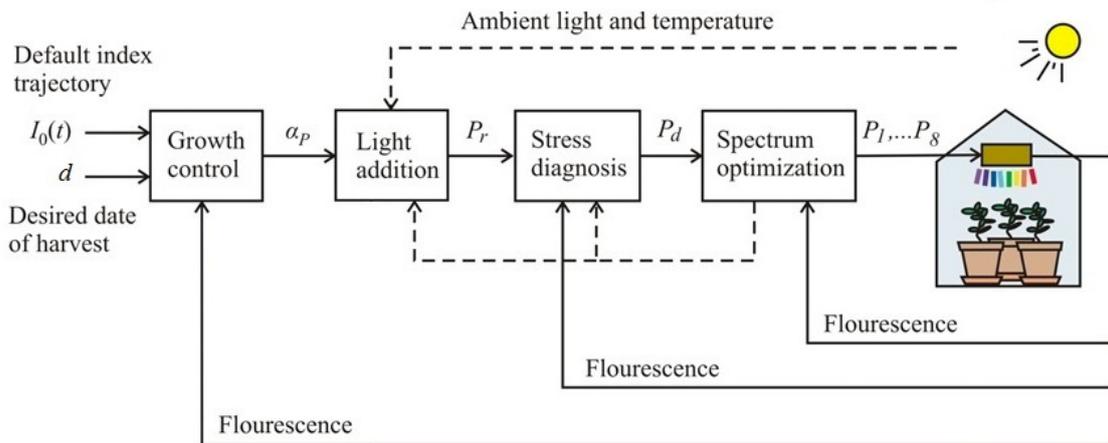


Figure 1.1: A concept of a future system that uses the ChlF from the plants to feed back information in order to adjust lamp light intensity. *Source: Torsten Wik*

1.2 Aim

The main aim of this project is to obtain a ChlF step response using sensors, LED lights and signal processing methods, that is free of sunlight disturbance. The goal is to be able to do this under any level of background sunlight intensity.

1.3 Objectives

The objective of the presented work was to continue on earlier work carried out on modelling and simulation of the LED lights and plant systems by Carstensen et.al.⁴, Lindqvist⁵ and Wikander⁶, as well as utilise sensors in junction with the LED lights to construct and gather data from simulation and realistic environments. Different signal processing methods are tested and evaluated. In the latest work, Lukas Wikander concluded that model parameter estimation is the key to be able to proceed with the most promising methods for extracting a ChlF response⁶.

The objectives were:

- Go further in depth with the signal processing methods for identification of ChlF response. The two types of methods examined are; Linear curve fitting and Parametric modelling.
- Setting up a simulation environment of the plant and light systems with the help of collected data for sunlight and fluorescence responses, in MATLAB and Simulink.
- Remove sunlight and simulate a ChlF step response with the help of the various signal processing methods in the simulated environment.

- Perform simulations where measured reflected and direct sunlight are used to produce perspicuous results of the overall performance of the specific method at certain sunlight intensities.
- Plan and execute experiments under realistic circumstances in a greenhouse environment.
- Analyse and evaluate the results obtained both in the simulation environment, by comparing the simulation results to actual fluorescence responses, as well as in the realistic greenhouse experiments by visual inspection.

A demarcation is that there will be no study regarding classification of plant stress. The focus is on extracting the correct chlorophyll fluorescence signal rather than determining plant health.

2

Theory and related work

IN THE following chapter, the underlying theory of this thesis work will be explained. It should give the necessary knowledge for the reader to understand the methods used and results obtained.

2.1 Electromagnetic spectrum

Electromagnetic radiation comes in different wavelengths. The collection of all these wavelengths is called the electromagnetic spectrum. An example of the spectrum of incoming sunlight can be seen in Figure (2.1).

The materials, which the light hits, have different properties and absorption rates for each wavelength. This makes the relation between wavelength bands for incoming and reflected light spectrum different from one another.

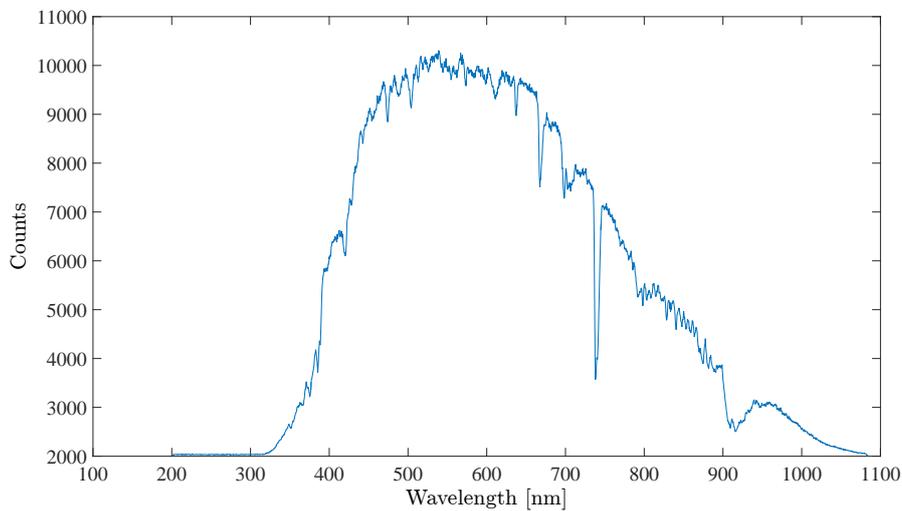


Figure 2.1: A visual representation of the spectrum of the sunlight radiation at a certain point in time. The unit *Counts*, represents the number of incoming photons during a certain amount of time called the Integration Time (IT).

2.2 Photosynthesis and fluorescence

Photosynthesis is a process where bacteria, some protists and plants use water, carbon dioxide and light between 400-700 nm, called Photosynthetically Active Radiation (PAR), to create chemically bounded energy and oxygen. The chemical energy can be used to fuel the plant's activity and is stored as sugar. Chlorophyll is a green pigment associated with this energy conversion, and absorbs most visible light. In photosynthesis an electron gains energy and is excited to a higher energy state. Increasing photosynthesis is always preferable since more chemical energy can then be stored in the plants¹³.

The photosynthetic efficiency from conversion of light to chemical energy is 3-6%. The rest of the absorbed light is dissipated as heat and fluorescence, of which only a few percent is re-emitted as fluorescence^{14,15}. In short, the fluorescence is the excess energy associated with an electron excitation to a higher state. ChlF is a form of luminescence within 670-800 nm, with two peaks in its spectrum; one at 685nm and one around 735-755 nm.

A fluorometer or spectrometer can be used to measure the intensity of the ChlF, which usually measured as Photon flux [$\mu mol s^{-1}$]. The Photon flux is made independent of collection area by dividing it with m^2 , which converts it to Photon Flux Density (PFD) [$\mu mol m^{-2} s^{-1}$], which is the standard unit for light intensity used throughout the thesis. The basic steps are to take in light through a fiber optic cable, collected over a chosen IT. Light is then reflected by a concave mirror into a grating, where the grating disperse the light with varying angles into another concave mirror. This mirror focuses the light into a detector, where the photons are converted into electrons. A computer software then bases the spectral distribution on the detected electrons where the output is one scalar of PFD for each wavelength. This information can then be integrated over a chosen wavelength interval where the PFD for the actual IT can be determined. This makes it possible to see the difference in fluorescence over time for any given interval¹⁶.

2.3 System identification

System identification is a way of constructing mathematical models of dynamic systems from measured input and output data. Prior information about the dynamic system itself can be included and vary from partly known (gray box) to completely unknown (black box). The procedure to follow when performing system identification is to first measure the input and output data. Then selecting a proper model structure and an estimation method for determining the adjustable parameters of the structure, and finally evaluating the results of the estimated model¹⁷.

2.4 Black box modelling

Standard black-box modelling uses the sampled input (u) as well as the output (y) of a system to determine the system parameters. Linear systems can be described using transfer functions ($G(q)$), where q is the discrete time shift operator:

$$y[n] = G(q) u[n] \quad (2.1)$$

$$\text{where} \quad (2.2)$$

$$G(q) = \frac{B(q)}{A(q)}, \quad (2.3)$$

where $B(q)$ and $A(q)$ are polynomials having the following standard notation

$$\begin{cases} B(q) = b_0 + b_1 q^{-1} + b_2 q^{-2} + \dots + b_{n_b} q^{-n_b} \\ A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + \dots + a_{n_a} q^{-n_a}. \end{cases} \quad (2.4)$$

Accordingly, n_b is the order of $B(q)$ and n_a is the order of $A(q)$. The orders also represent the number of zeros and poles of the system, i.e. the roots of B and A, respectively. Once input and output data as well as the order of the system is chosen, the following formula is used to describe the system.

$$y[n] = \frac{B(q)}{A(q)} u[n] = \frac{b_0 + b_1 q^{-1} + b_2 q^{-2} + \dots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + a_2 q^{-2} + \dots + a_{n_a} q^{-n_a}} u[n], \quad (2.5)$$

which gives a corresponding difference equation

$$\begin{aligned} y[n] + a_1 y[n-1] + a_2 y[n-2] + \dots + a_{n_a} y[n-n_a] &= \\ = b_0 u[n] + b_1 u[n-1] + b_2 u[n-2] + \dots + b_{n_b} u[n-n_b] & \end{aligned} \quad (2.6)$$

for the sampled data. Separating $y[n]$ to one side of the equation yields

$$y[n] = -a_1 y[n-1] - a_2 y[n-2] \dots - a_{n_a} y[n-n_a] + b_0 u[n] \quad (2.7)$$

$$\begin{aligned} &+ b_1 u[n-1] + b_2 u[n-2] + \dots + b_{n_b} u[n-n_b] \\ &= \theta^T \varphi[n] \end{aligned} \quad (2.8)$$

where

$$\theta = [-a_1 \quad -a_2 \quad \dots \quad -a_{n_a} \quad b_0 \quad b_1 \quad b_2 \quad \dots \quad b_{n_b}]^T \quad (2.9)$$

and

$$\varphi[n] = [y[n-1] \quad y[n-2] \quad \dots \quad y[n-n_a] \quad u[n] \quad u[n-1] \quad u[n-2] \quad \dots \quad u[n-n_b]]^T. \quad (2.10)$$

A dataset of measured outputs $y[n]$ and inputs $u[n]$ is then used to find the parameter values for a model that best approximates the outputs for the given inputs. The determination of these parameters differs depending on the linearity of the regression. There are standardised models for many system dynamics, such as Auto Regressive with eXogenous input (ARX) which is a model with a linear regression and Output Error (OE) which results in a nonlinear parameter identification. Both can be seen in Equation (2.11) below.

$$\begin{cases} ARX : & y[n] = \frac{B[q]}{A[q]}u[n] + \frac{1}{A[q]}\epsilon[n] \\ OE : & y[n] = \frac{B[q]}{A[q]}u[n] + \epsilon[n] \end{cases} \quad (2.11)$$

where $\epsilon[n]$ is assumed to be a white noise term representing the error between measured values and ideally predicted values, i.e. when $\epsilon[n] = 0 \forall n$ the model should represent the system perfectly. Estimation techniques for the two models will be explained in Sections 2.5 and 2.7.1-2.7.2.

2.5 Least squares

In order to determine the approximating linear models, the least squares formula is used to minimise the error $\epsilon[n, \hat{\theta}^{(i)}]$. Given a set of outputs $y[n]$ the unknown parameter vector θ can be estimated using the following formula¹⁸:

$$\hat{\theta} = [\Phi[n]^T \Phi[n]]^{-1} \Phi[n]^T y[n] \quad (2.12)$$

where $\Phi(n)$ is the regressor matrix:

$$\Phi(n) = [\varphi[1]^T \quad \varphi[2]^T \quad \dots \quad \varphi[n]^T]^T \quad (2.13)$$

2.6 Output error model

From previous work^{5,6} it has been shown that when attempting to model ChlF responses to changes in incoming light, an OE model with 3 poles, 3 zeros and no delay (OE430) is the highest order needed to catch the dynamics.

2.7 Prediction error methods

There are several different methods for iterative calculation of the parameter vector θ for an OE model. One of these is the Prediction Error Method (PEM)^{18,19}. PEM is a way to determine the global minimum of the least squared error by using the error between the output $y[n]$ and the predicted output $\hat{y}[n]$ ^{18,19}:

$$\epsilon[n, \theta] = y[n] - \hat{y}[n] \quad (2.14)$$

where

$$\hat{y}[n] = \frac{B[q, \theta]}{A[q, \theta]} u[n] \quad (2.15)$$

2.7.1 Prefiltering algorithm for OE systems

If the OE model structure is rewritten as:

$$A(q, \theta)y[n] = B[q, \theta]u[n] + A[q, \theta]e[n, \theta] \quad (2.16)$$

a filter consisting of $A[q, \hat{\theta}^{(i-1)}]^{-1}$ can be applied preemptively to the input and output data ($u[n]$ $y[n]$). With the prefiltering algorithm this is done iteratively via the the prediction error calculation

$$\begin{aligned} e[n, \hat{\theta}^{(i)}] &= A[q, \hat{\theta}^{(i)}][A[q, \hat{\theta}^{(i-1)}]^{-1}y[n] - B[q, \hat{\theta}^{(i)}][A[q, \hat{\theta}^{(i-1)}]^{-1}u[n] \\ &\Rightarrow \tilde{y}[n] = \frac{B[q, \hat{\theta}^{(i)}]}{A[q, \hat{\theta}^{(i)}]} \tilde{u}[n] + \frac{1}{A[q, \hat{\theta}^{(i)}]} e[n, \hat{\theta}^{(i)}] \end{aligned} \quad (2.17)$$

where $\hat{\theta}^0$ is the biased Least Squares (LS) parameter estimate from Equation (2.12).

From Equation (2.17) an ARX structure is formed with the new filtered input and output with a prediction error that is close to white where $\tilde{y}[n] = [A[q, \hat{\theta}^{(i-1)}]^{-1}y[n]$ and $\tilde{u}[n] = [A[q, \hat{\theta}^{(i-1)}]^{-1}u[n]$. By the use of ordinary LS a non-biased estimation of the parameter vector $\hat{\theta}$ can then be acquired from Equation (2.17) using $\tilde{y}[n]$ ²⁰.

2.7.2 Damped Gauss-Newton

Given the measured output ($y[n]$) and the measured input ($u[n]$), the damped Gauss-Newton (dGN) will make a prediction of $y[n]$, which follows from Equation (2.14), i.e. when the error $e[n, \hat{\theta}^{(i)}]$ is 0, the prediction represent the system exactly. The Least Squares Error (LSE) is minimised w.r.t. $\hat{\theta}^{(i)}$ using the following formula, where N is the number of samples

$$V_N(\hat{\theta}^{(i)}) = \frac{1}{N} \sum_{t=1}^N e^2[n, \hat{\theta}^{(i)}] \quad (2.18)$$

$$\hat{\theta}_N^{(i)} = \arg \min_{\theta} V_N(\hat{\theta}^{(i)}) \quad (2.19)$$

The dGN minimisation technique is given by

$$\hat{\theta}^{(i)} = \hat{\theta}^{(i-1)} - \mu [V''(\hat{\theta}^{(i-1)})]^{-1} V'(\hat{\theta}^{(i-1)}) \quad (2.20)$$

$$\text{where } \mu \text{ is step length and} \quad (2.21)$$

$$V'(\hat{\theta}) = -\frac{1}{N} \sum_{t=1}^N (y[n] - \hat{y}[n] \frac{\delta}{\delta \theta} \hat{y}[n]) \quad (2.22)$$

$$V''(\hat{\theta}) \approx \frac{1}{N} \sum_{t=1}^N \left(\frac{\delta}{\delta \theta} \hat{y}[n] \right) \left(\frac{\delta}{\delta \theta} \hat{y}[n] \right)^T \quad (2.23)$$

The initial value of $\hat{\theta}^0$ is calculated by linear least squares from Section 2.5.

$$\theta^{(i)} = [\Phi^T[n]^{(i-1)} \Phi[n]^{(i-1)}]^{-1} \Phi^T[n]^{(i-1)} y[n] \quad (2.24)$$

In order to find the global minimum a prefilter, as in Equation (2.17) is used on the inputs and outputs. This is followed by the derivation of a new $\Phi[n]$ in the same way as in Equation (2.9), but with filtered measurements $\tilde{y}[n]$ and $\tilde{u}[n]$. The regressor components then looks as follows:

$$\varphi[n] = [\tilde{y}[n-1] \tilde{y}[n-2] \dots \tilde{y}[n-n_a] \tilde{u}[n] \tilde{u}[n-1] \tilde{u}[n-2] \dots \tilde{u}[n-n_b]]^T \quad (2.25)$$

The Jacobian, $\frac{\delta}{\delta \theta} \hat{y}[n]$, will be non-conventional, since the standard one does not incorporate $e[n, \hat{\theta}^{(i)}]$. A new error, $\hat{e}[n, \theta^{(i)}]$, will also be calculated. This is done by applying orthogonal-triangular decomposition on a matrix A , consisting of $\Phi[n]$ and $e[n, \hat{\theta}^{(i)}]$.

$$A = [\Phi[n] \quad e[n, \hat{\theta}]] \quad (2.26)$$

If the size of $\Phi[n]$ is $(n_a + n_b + 1) \times (N - n_a)$ and the size of $e[n, \hat{\theta}^{(i)}]$ is $1 \times (N - n_a)$, the size of A is $(n_a + n_b + 2) \times (N - n_a)$.

The orthogonal-triangular decomposition decomposes the matrix A into a product, QR , where R is the upper triangular part and Q is an orthogonal matrix satisfying the following criterion:

$$A = QR \quad (2.27)$$

where

$$Q^T Q = Q Q^T = I \quad (2.28)$$

giving the relation that Q is orthogonal if: (2.29)

$$Q^{-1} = Q^T \quad (2.30)$$

The upper triangular part of Q , of size $(n_a + n_b + 2) \times (n_a + n_b + 2)$, will then be extracted. Note that R will not be used as the upper triangular matrix, it will be extracted from the Q matrix²¹.

Extract all data, except for the last column, to get a new Jacobian, $J(\theta^{(i)}, e)$, with size $(n_a + n_b + 2) \times (n_a + n_b + 1)$, where the error is included. $\hat{e}[n, \theta^{(i)}]$ is simply the last column of the upper triangular part matrix^{22,23}. The dGN formula will then be as follows:

$$\hat{\theta}^{(i)} = \hat{\theta}^{(i-1)} + \mu(J(\theta^{(i-1)}, e) J(\theta^{(i-1)}, e)^T)^{-1}(\hat{e}[n, \theta^{(i-1)}] J(\theta^{(i-1)}, e)) \quad (2.31)$$

2.8 Linear curve fitting

Curve fitting is used to determine the linear equation used to fit two curves. Given two vectors consisting of data and slack variables called $C(t)$, and another vector of different data but with same length, $f(t)$. The following equation is then used, where $x = \begin{bmatrix} k & m \end{bmatrix}^T$ is a 2×1 vector representing the multiplication factor (k) and offset (m) of the linearity.

$$C(t) = \begin{bmatrix} z(1) & z(2) & \dots & z(N) \\ 1 & 1 & \dots & 1 \end{bmatrix}^T \quad (2.32)$$

$$f(t) = \begin{bmatrix} y(1), & y(2) & \dots, & y(N) \end{bmatrix}^T \quad (2.33)$$

$$\min_x \|C(t)x - f(t)\|^2 \quad (2.34)$$

where z and y represents the points in the two equally sized curves.

2.9 Bode plot

A bode plot utilises the frequency domain to visualise properties of a linear system, where the phase (ϕ) and magnitude (k_{mag}) of different frequencies can be evaluated. Using a transfer function $H(j\omega)$, the magnitude is calculated through

$$k_{mag} = |H(j\omega)| \quad (2.35)$$

where ω is the input frequency. The phase is calculated through

$$\phi = \arg(H(j\omega))^{24} \quad (2.36)$$

2.10 Fit percentage and performance index

In order to get an assessment ground to stand on when comparing the response of a system and an estimation of it, the normalised root mean square error (NRMSE) can be used according to²³.

$$fit(t) = 100 \left(1 - \frac{\|y(t) - \hat{y}(t)\|_2}{\|y(t) - \text{mean}(y(t))\|_2} \right) \quad (2.37)$$

Where $y(t)$ is a system output and $\hat{y}(t)$ is the estimated system output. The fit is a percentage of how well the two outputs match. This is especially useful when comparing simulation results and references. A high fit percentage is not always equivalent to a good result, due to the occurrence of missing crucial dynamics. It is, however a very effective tool to assess model quality during large simulations which is why this is the main performance index used throughout this thesis.

2.11 Related work

As mentioned earlier, this thesis aims to take up previous work conducted on plant stress and remote fluorescence sensing. The following Sections will explain some related work within Dynamic Fluorescence Response Analysis (DFRA), made by Carstensen et.al.²⁵ and Lindqvist⁵, and signal processing methods for DFRA in a real environment by Wikander⁶.

2.11.1 Dynamic Fluorescence Response Analysis

DFRA is a method to remotely detect ChlF in order to evaluate stress levels, where Lindqvist⁵ defines plant stress as the plant reaction when subjected to sub-optimal growth. The DFRA studies have been carried out by Carstensen et. al.²⁵ and Lindqvist⁵, at Chalmers and at Heliospectra AB.

From a height of 1-2 m, the ChlF response from a step of incoming lamp light in a controlled environment is evaluated. Using an identified OE model with 4 zeros and 3 poles, the frequency domain properties are examined⁵. The bode plot is used for this and the phase properties are studied. The conclusion is that light intensity shift the plant dynamics in the frequency domain. The plant dynamics are also examined in the time domain where the responses complexity is reduced with increasing stress, (see 3.2.1 for more details). At higher stress levels only 3 zeros and 2 poles are needed and a plant in extreme stress only need, 2 zeros and 1 pole. The lateral position of some dynamics also changed with the stress level^{25, 5}.

Instead of only using a step excitation to trigger a ChlF response, Carstensen et. al. also used a sinusoidal varying excitation light, and compared the resulting frequency properties with those found in the bode analysis. The sinus with a frequency

of 0.1 rad/s yielded similar results for phase shifts as using a step²⁵. Due to quantitative differences for data with low PAR using sinusoidal excitation, the step excitation seems to be more consistent as a whole, where it can also be studied in time domain²⁵. Lindqvist did further research using only the step excitation, making its underlying research more stable⁵. Steps are therefore used in this study instead of sinusoidal excitation.

The ChlF responds to sudden changes in light in a non-linear way. Carstensen et. al. suggests that this is due to a buffer system of the plant where it uses internal feedback and the buffers are likely to be metabolite pools²⁵. It is also found that photosynthesis only takes place after the plant is subjected to a high enough magnitude of PFD within PAR, working as a dead zone for a certain threshold.

2.11.2 Methods for estimating chlorophyll fluorescence response

Wikander attempted to estimate the ChlF response in presence of disturbing sunlight using multiple approaches, utilising signal processing methods but also spectral data treatment to gain access to a wider range of options. The spectral data treatments use only spectral data for one sample at a time but take better advantage of the data within one spectrum, while the signal processing methods use multiple time samples. Fraunhofer line discrimination (FLD) and spectral quotient are both methods using the spectral data and are thus not affected by sensor placement in the same way as the signal processing methods⁶. They are, however, more sensitive to noise⁶ and comes with their own set of problems, which suggested that the signal processing methods were to be further pursued in this follow-up study, rather than the spectral data treatment methods.

Curve fitting has already been evaluated by Wikander in a greenhouse environment with sunlight present⁶. A similar investigation in the performance of curve fitting is made in this thesis, although some things differ in terms of fitting time and slack variable value, which might play a role in its ability to properly extract a ChlF response.

Wikander derived a model of the system, where the reflection properties of the canopy also was included. As such, the model also included sunlight. The same model will be used in this thesis for parametric modelling and is described further in Section 2.11.3 and 3.5. The author concluded that by correctly estimating the system and sunlight feed forward parameters, a ChlF response could be determined.

For the iterative parameter search in the model, Wikander used the Instrumental Variable (IV) method⁶. However, this method did not handle noise well⁶, finding the wrong minimum and frequently leading to a faulty estimation of the parameters. More details on the IV method can be found in appendix A. Despite this the parametric modelling method for estimation of the ChlF was still considered to have the best prospects for solving the DFRA problem in the presence of sunlight⁶. To progress and enhance the parametric model method, other iterative solutions need to be investigated hopefully handling the noise of the input and output data better. Thereof the usage of prefiltering and dGN, which handles the error in a more efficient way by filtering the input and output data for every iteration.

2.11.3 Measurement model

The measurement model used for the signal processing methods has been defined by Wikander and structures the available data and light behaviour in a realistic way⁶. The fluorescent output y_f from the plant is a function of the system $g[u[\lambda, n], \lambda, n]$ and the input $u[\lambda, n]$ in PAR, i.e.

$$y_f[\lambda, n] = g[u[\lambda, n], \lambda, n], \quad (2.38)$$

where λ is the wavelength and n is the sample number. The input, in turn, is a function of the incoming sun $d[\lambda, n]$ and the lamp light $l[\lambda, n]$:

$$u[\lambda, n] = l[\lambda, n] + d[\lambda, n] \quad (2.39)$$

The sunlight acts as a disturbance in the ChlF wavelength band (670-800 nm) both directly via the reflected light and indirectly through its induction of ChlF. The system output $y[\lambda, n]$ is thus:

$$y[\lambda, n] = \begin{cases} g[u[\lambda, n], \lambda, n] + \rho[\lambda] \cdot d[\lambda, n], & \text{if } \lambda \in (670, 800) \\ \rho[\lambda] \cdot u[\lambda, n], & \text{if } \lambda \notin (670, 800) \end{cases} \quad (2.40)$$

where ρ is a wavelength dependent reflectance.

Each sample of the signal $x[n]$ is measured in *Counts per wavelength*. If the response is to be shown as a function of time, it is often preferable to integrate the data over specific wavelength ranges⁶.

$$x[n] = \sum_{i=\lambda_1}^{\lambda_2} x[i, n] \quad (2.41)$$

This enables the signal to be converted to *Counts* in a frequency range of interest, e.g. PAR (400-700 nm), peak of ChlF (735-755 nm) and a fluorescence free wavelength band, for instance (800-820 nm). This interval is then converted to PFD ($\mu\text{mol m}^{-2} \text{s}^{-1}$) which, for multiple sample data can be displayed over time.

3

Method

IN this section the approach and the equipment are described more in detail. The evaluation methods will also be presented as well as all the different experiments and their corresponding setups.

3.1 Equipment

The specifications for the sensors used to collect data are given in Table (3.1). Two sensors were used throughout the project with one collecting the incoming sunlight and lamp light, and the other collecting the reflected light and chlorophyll fluorescence.

Table 3.1: Spectrometer specifications

Reference name	<i>M1</i>	<i>M2</i>
Measurement task (relative plant)	<i>Incoming light</i>	<i>Reflected light and chlorophyll fluorescence</i>
Ocean Optics model	<i>Maya2000 Pro</i>	<i>Maya2000 Pro</i>
Wavelength resolution (FWHM) [nm]	<i>1.85</i>	<i>1.85</i>
Wavelength range [nm]	<i>199.1-1116.7</i>	<i>198.2-1085.6</i>
Fibre diameter [μm]	<i>600 or 50</i>	<i>600</i>
Calibration date	<i>2016-01-29</i>	<i>2016-01-29</i>

The spectrometers also had attachments with the purpose of altering the incoming light into the fibre. These were a Gershun tube (G-tube) and a cosine corrector. The G-tube was used to reduce the field of view to focus solely on the canopy. The cosine corrector increased the field of view to make the sensor less sensitive to the angle of the incoming light. The sampling frequency used throughout all measurements was 1 Hz. The light used in the greenhouse experiments was provided by two Heliospectra LX601G LED lights. These lights had three different channels: Blue (450 nm), Red (660 nm) and White (5700K). The red and white LEDs were used for constant background light and the blue for the light excitation. The reason for this is that the excitation light should not interfere with the ChlF band between 735-755 nm, and it should be as strong as possible. Both the red and white LED-light contained some energy in the ChlF band while the blue could excite light without any interference on the ChlF.

3.2 Data evaluation

The fluorescence signal obtained from the plants was evaluated in time and frequency domains. To confirm model accuracy, the shape of the resulting fluorescence response after a lamp light intensity step change was studied in the time domain. The accuracy of the model gave an instantaneous value of the model goodness in terms of fit percentage when comparing the fluorescence response from the model with that of the sensor data (see Section 2.10).

With the help of this type of information, bad models were discarded, aiding in the process of finding suitable parameters for either curve fitting or parametric modelling. For a better overview of the health of a plant, the frequency domain was examined as well, during step excitation. The phase shift trumped the time domain observations in terms of stress indication in one important aspect; it was more or less independent of the signal amplitude. On a practical level, this means that it is robust to changes in leaf area, morphology and distance between canopy and measurement device⁴. It was also shown by Nedbal that at the frequency $\omega \approx 0.1 \text{ rad/s}$ there are dynamical features in the fluorescence signal, which can be of interest when determining plant health²⁶. The frequency span used to study the bode plot was therefore chosen to be between 0.01 and 1 rad/s⁵.

3.2.1 PSMT evaluation

One way of classifying the dynamics exhibited in a fluorescence step response is to analyse the PSMT pattern. The letters define the step (P), the initial peak (S), the second peak (M) and the settling time (T). The shape and position of these have to do with the current dynamics of the plant and was therefore expected to be a valid evaluation method when it comes to plant diagnostics. Lindqvist et.al. observed that the time between the S and the M-peak shifted depending on the health of the plant²⁷. The position and maximum of the M-peak were therefore used, in junction with the fit, to validate model estimations. Figure (3.1) shows the difference in the M-peak for different stress levels.

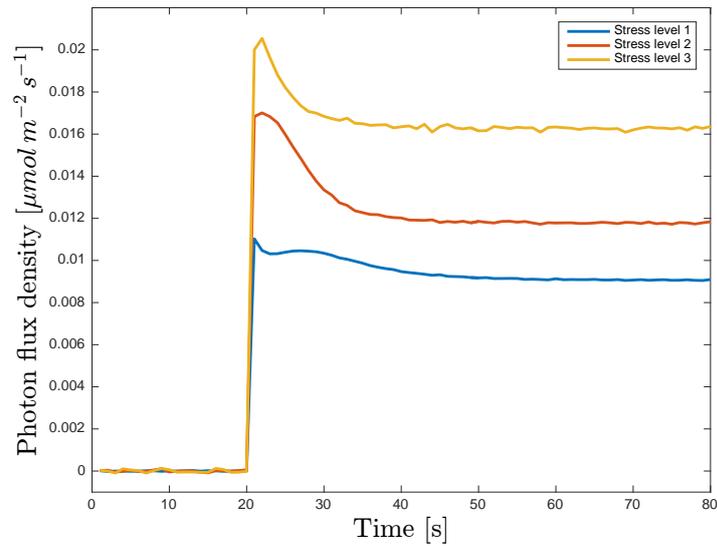


Figure 3.1: An example of the fluorescence response over time with different stress levels.

3.3 Filtering

In order to filter out noise, a low pass filter of order 5 and cutoff frequency of 1 rad/s was used on both the input $u(t)$ and the output $y(t)$. The filter used was of Butterworth type given by

$$H(j\omega) = \frac{1}{\sqrt{1 + \left(\frac{j\omega}{j\omega_0}\right)^{2n}}} \quad (3.1)$$

where ω_0 is the cutoff frequency, ω the current frequency and n the filter order. The normalized phase and magnitude plot of the filter is seen in Figure (3.2).

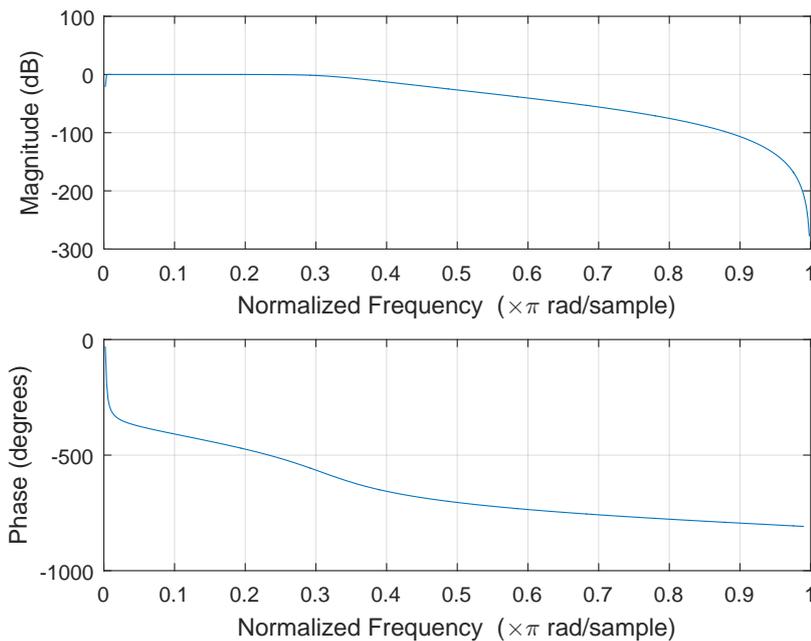


Figure 3.2: The bode plot for the Butterworth filter used.

3.4 Linear curve fitting implementation

Using the theory from Section 2.8, the $C(t)$ matrix was the reflected sunlight in the 800-820 nm band ($\rho d_{800-820}(t)$) and the $f(t)$ vector was the 735-755 nm band ($\rho d(t)$). The experiments will have a step after 20 s in $\rho d(t)$, therefore only the first 20 s was used for the fitting, where $\rho d_{800-820}(t)$ was warped ($\rho \hat{d}(t)$) with respect to the parameters $x = [k \ m]$. By subtracting $\rho \hat{d}(t)$ signal from $\rho d(t)$, the output should be a fluorescence response induced by lamp excitation with some noise. The band of choice was 800-820 nm because of the absence of fluorescence (see Figure (3.3)).

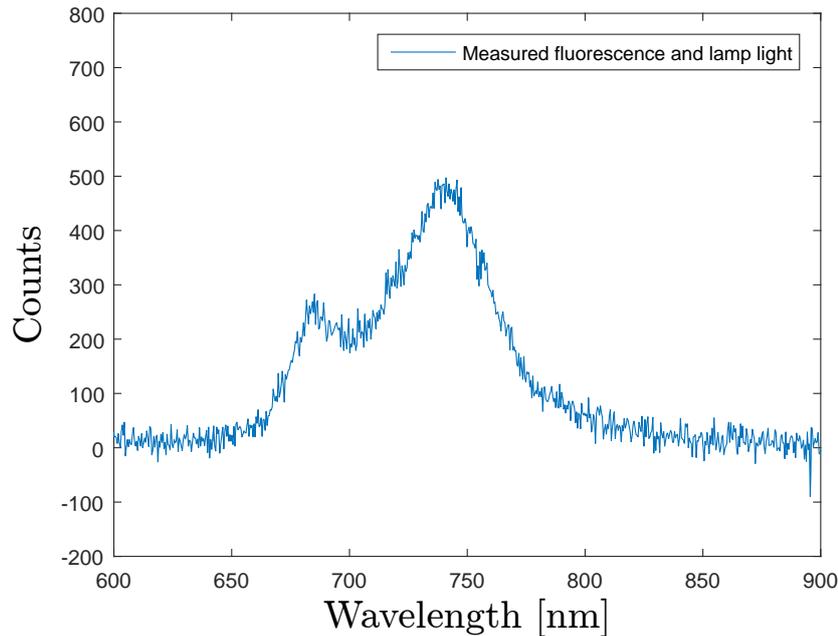


Figure 3.3: The spectrum around the ChlF band. The fluorescence is considered to be negligible above 800 nm in the estimation of the sunlight.

3.5 Parametric modelling of the plant system

To be able to interpret plant health from input-output data, a plant model needed to be acquired. The system that was being modelled consisted of the plant dynamics $g[u[n], n]$ and the reflection variable ρ , based on the measurement model described in Section 2.11.3. Figure (3.4) shows an overview of the system. The input, $u[n]$, to the plant comprised of two signals, i.e. the lamp light $l[n]$ and the sunlight within PAR $d_{PAR}[n]$. The output from the system, $y[n]$, therefore consisted of both fluorescence, $y_f[n]$, and reflected sunlight, $\rho d[n]$. $y_f[n]$ denotes the fluorescence that the plants emit when excited by light within the PAR region. This was the signal that was ultimately desired.

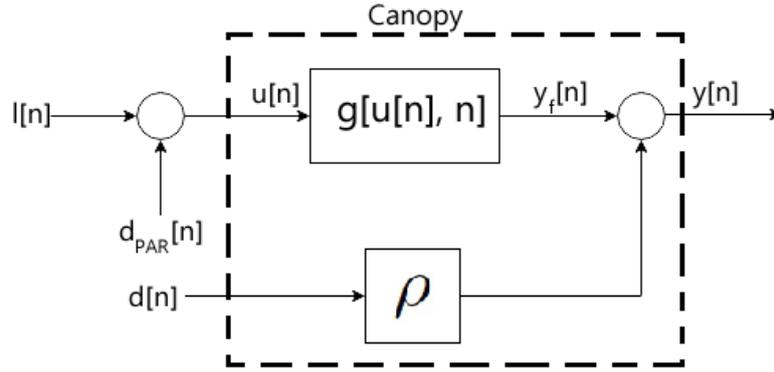


Figure 3.4: The system, showing the canopy, inputs and output. Where $l[n]$ is the lamp light, $d[n]$ is the sunlight between 735-755 nm, $d_{PAR}[n]$ is the sunlight within PAR, $u[n]$ is the input to the plant, consisting of $l[n]$ and $d_{PAR}[n]$, $y_f[n]$ is the fluorescence, ρ is the reflectance, $g[u[n], n]$ is a model of the canopy and $y[n]$ is the measured output.

By accurate parametric modelling, the phase response in frequency domain as well as the step response in time domain of the model could be used to determine plant health⁵. Plants that are more stressed generally exhibits a less complex fluorescent response²⁷, although for the purpose of this project a OE430 model was sufficiently complex to describe the necessary dynamics of the systems. From the system in Figure (3.4) a modified OE model for this project was obtained, where reflected sunlight was added as a feed-forward disturbance⁶:

$$y[n] = g[u[n], n] + \rho d[n] = \frac{B}{A}u[n] + \rho d[n] \Rightarrow Ay[n] = Bu[n] + A\rho d[n] \quad (3.2)$$

An OE430 then gives the following difference equation:

$$y[n] = -a_1y[n-1] - a_2y[n-2] - a_3y[n-3] + \rho d[n] + \rho a_1d[n-1] + \rho a_2d[n-2] + \rho a_3d[n-3] + b_0u[n] + b_1u[n-1] + b_2u[n-2] + b_3u[n-3] \quad (3.3)$$

Written in terms of parameter and regressor vectors

$$y[n] = \theta^T \varphi[n] \quad (3.4)$$

Where the parameter vector θ was

$$\theta = [a_1 \ a_2 \ a_3 \ \rho \ \rho a_1 \ \rho a_2 \ \rho a_3 \ b_0 \ b_1 \ b_2 \ b_3]^T \quad (3.5)$$

and the regressor vector φ was

$$\varphi[n] = \begin{bmatrix} -y[n-1] & -y[n-2] & -y[n-3] & d[n] & d[n-1] & d[n-2] & d[n-3] \\ u[n] & u[n-1] & u[n-2] & u[n-3] \end{bmatrix} \quad (3.6)$$

When a set of data $l(t)$, $d(t)$ and $y(t)$, had been collected it was put in separate aggregated matrices:

$$\varphi(t) = [\varphi(n_a + 1) \quad \cdots \quad \varphi(N)] \quad (3.7)$$

$$\mathbf{y}(t) = [y(n_a + 1) \quad \cdots \quad y(N)]^T \quad (3.8)$$

The parameter vector estimate $\hat{\theta}$ was then calculated using the least squared error¹⁸ from Section 2.5, which minimised the squared error of the aggregated matrices above. For the OE model the ordinary least squares will produce a biased least squared error¹⁸. Iterative methods were therefore used to successively calculate the best estimate for θ for this parameter nonlinear system. The Prefiltering (2.7.1) and dGN (2.7.2) methods were used for the numerical search.

3.6 Assumptions and analysis

All methods used in this thesis lean on certain assumptions about the dynamics of the system. In order for the methods to work, first in theory and then practically, these assumptions must be valid, at least to a certain extent. The fundamental reason for these assumptions originate from the fact that nonlinear systems, such as the plant system considered here, are hard to handle. Therefore, linearization often takes place to make better use of optimal methods to minimise the mean square error.

Concerning the reflectance the following assumptions are used:

- The ratio between the reflected light of the wavelengths 735-755 nm and 800-820 nm is constant throughout all sunlight intensities during the time it takes to collect the necessary data. Curve fitting (see Section 3.4) rests upon the assumption that, for the time it takes to fit the different wavelength bands and estimate the fluorescence, this ratio is constant. This is required in order to get a good, usable, approximation of the plant fluorescence. The reason 800-820 nm is used as counterpart is because there is no fluorescence present in that wavelength band (see Figure (3.3)).
- The reflection ρ is constant during the amount of time it takes to stimulate a fluorescence response and letting it settle. For example; if an instance of 80 samples, collected with a sample time of 1 s, is needed the reflection ρ needs only be constant during those 80 seconds.

When the above assumption turns out to be false, a solution was to multiply the incoming sunlight data ($d(t)$) with a correction vector quota ($q(t)$) to compensate for ρ not being scalar, that was estimated by the parametric model. This correction vector $q(t)$ was obtained by calculating the reflectance ($\rho(t)$) as in Equation (3.9)

and divide by ρ . This was illustrated in Figure (3.5), where the red line shows the scalar ρ as a line and the blue line shows $\rho(t)$. Now normalising the static ρ with the dynamic $\rho(t)$ as in Equation (3.9) gave $q(t)$, as illustrated in Figure (3.6). Note that $d_\rho(t)$ is the measured reflected light, $d(t)$ and $d_\rho(t)$ both have to contain the same wavelength intervals and ρ will always be estimated from the 735-755 nm interval. This is due to the wanted fluorescence signal being in this band.

$$\rho(t) = q(t) \cdot \rho = d_\rho(t) d(t)^{-1} \Rightarrow q(t) = d_\rho(t) d(t)^{-1} \cdot \rho^{-1} \quad (3.9)$$

where

$$\rho = \rho_{LS} = (d_\rho(t)^T d_\rho(t))^{-1} \cdot d_\rho(t)^T d(t) \quad (3.10)$$

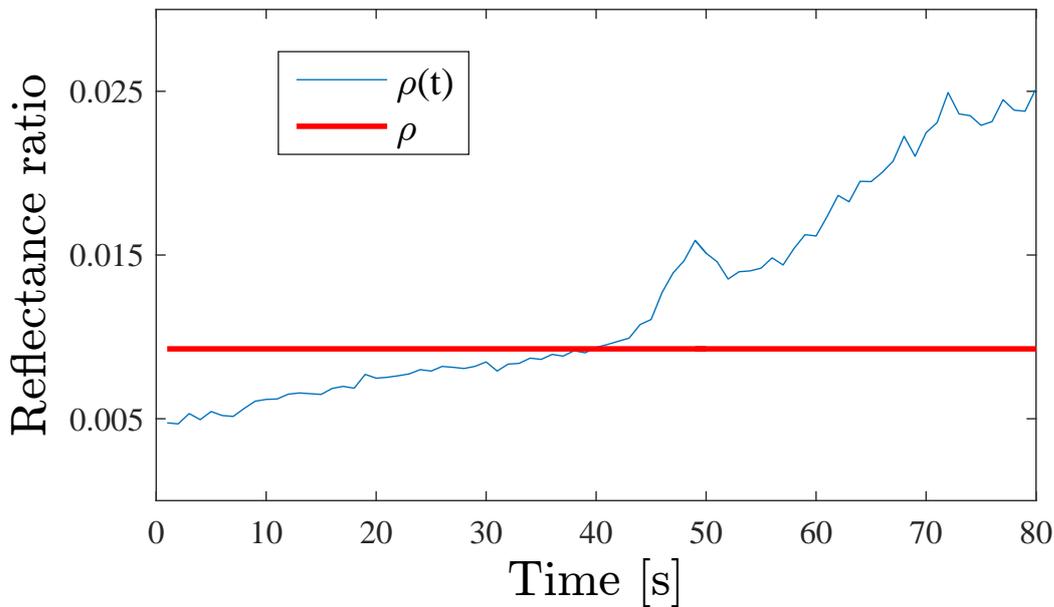


Figure 3.5: Example data visualising the difference between the static ρ in red and the dynamic $\rho(t)$ in blue.

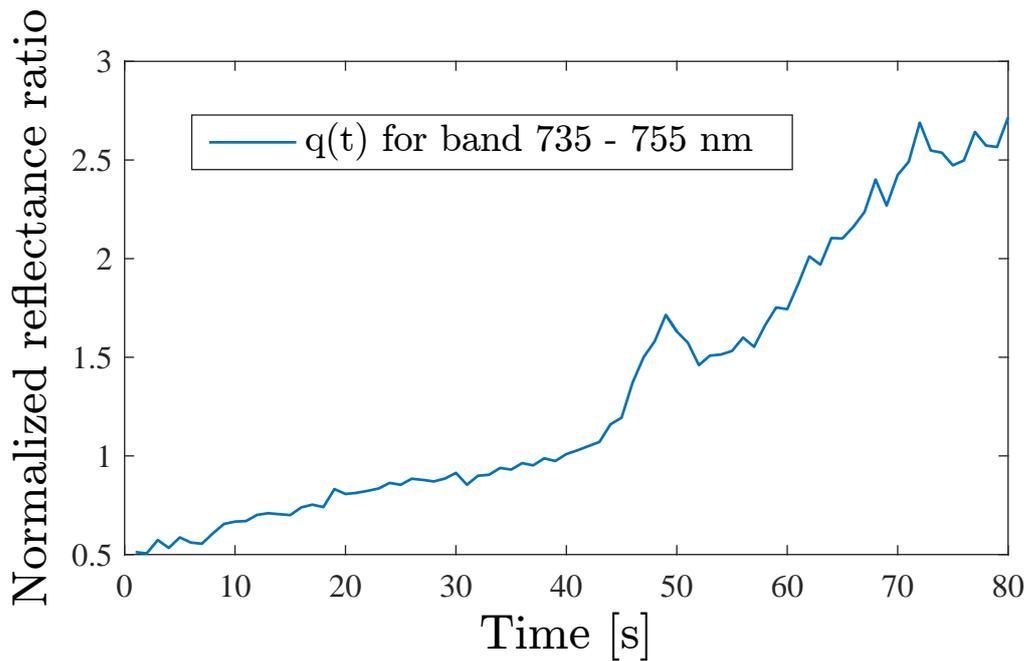


Figure 3.6: Normalised $q(t)$ between dynamic $\rho(t)$ and static ρ from Figure (3.5).

Due to the presence of fluorescence, there was no way of directly obtaining $q(t)$ from the 735-755 nm band. Another spectral band must therefore be used to estimate $q(t)$. $q(t)$ of this spectral band had to be very similar to that of 735-755 nm in order to work as a substitute $q(t)$. Figure (3.7) shows $q(t)$ for all wavelengths during one sample. The black part of $q(t)$ ranges from 670 -800 nm and was deemed unfit to use as substitute since fluorescence may be present within that wavelength band. The band 800-820 nm was, in most samples, the wavelength interval that was most alike the 735-755 nm band while also being free from fluorescence. It was therefore used as a substitute for $q(t)$. Comparing the 735-755 nm $q(t)$ to 800-820 nm for one sample gave the plot in Figure (3.8).

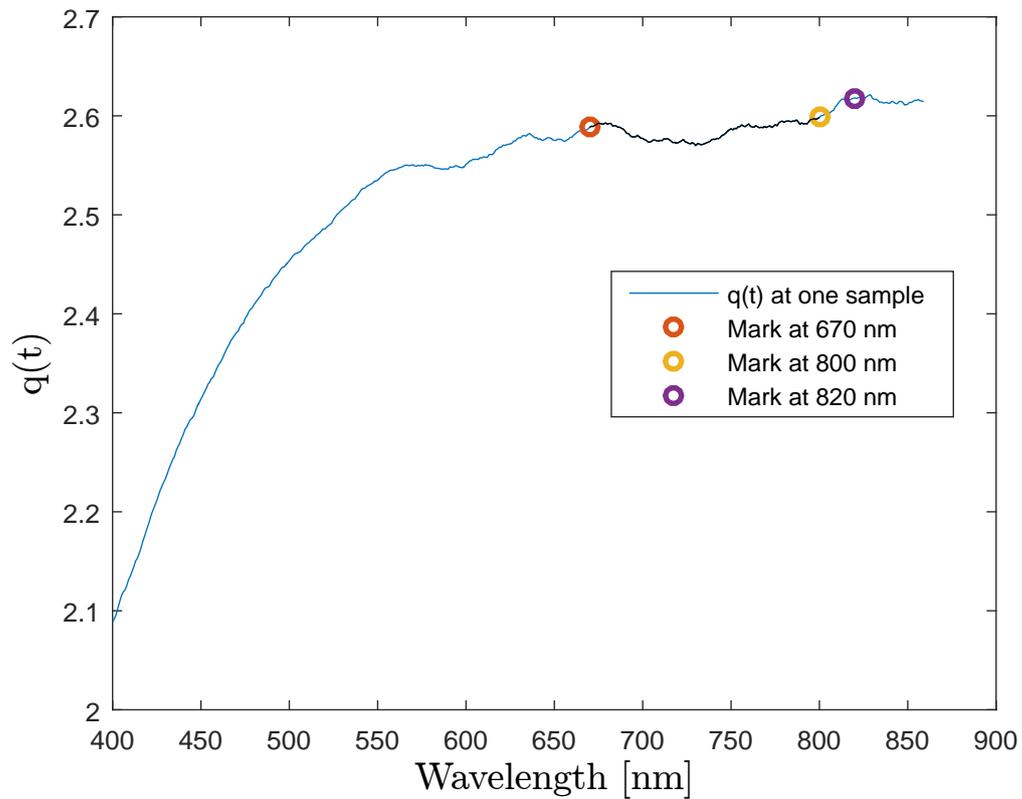


Figure 3.7: $q(t)$ for all wavelengths during one sample. Blue line shows $q(t)$, red mark shows where the fluorescence band starts, yellow mark shows where the fluorescence band ends and where the substitute band starts and the purple mark shows where the substitute band ends.

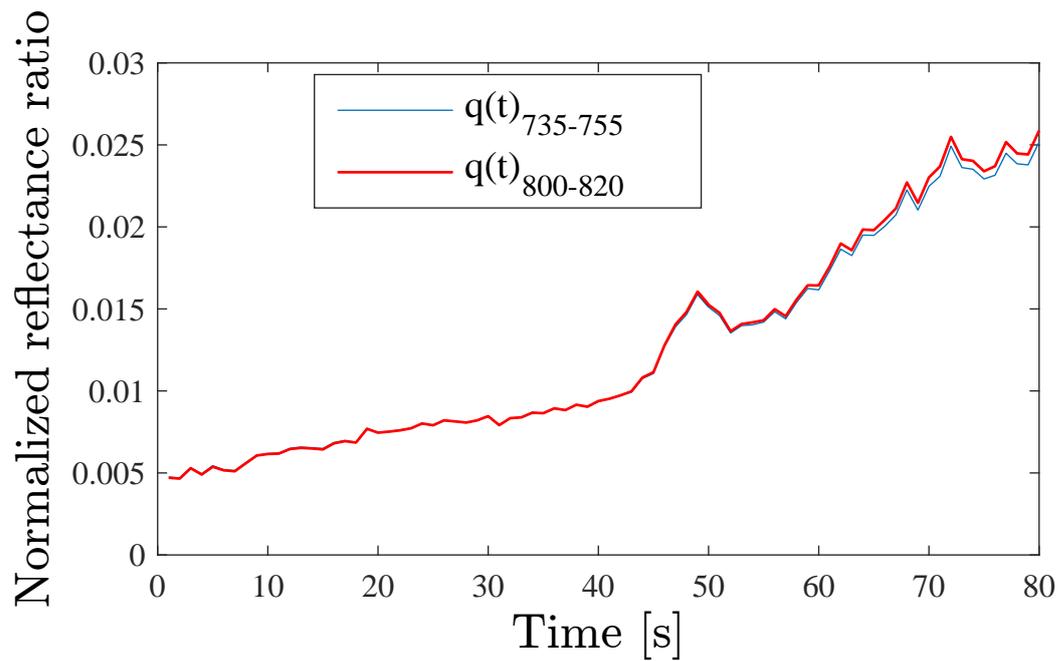


Figure 3.8: A comparison of $q(t)$ between band 735-755 nm and band 800-820 nm.

To further estimate a substitute $q(t)$, the spectrum signal from Figure (3.7) was filtered using a mean filter, called Filtered Spectrum (FS). By interpolating between the start and the end of the fluorescence band, followed by extracting the value at 735-755 nm from this interpolation, an improved $q(t)$ called Filtered Line Spectrum (FLS) could be used. Both the true $q(t)$ and the substitute can be seen in Figure (3.9).

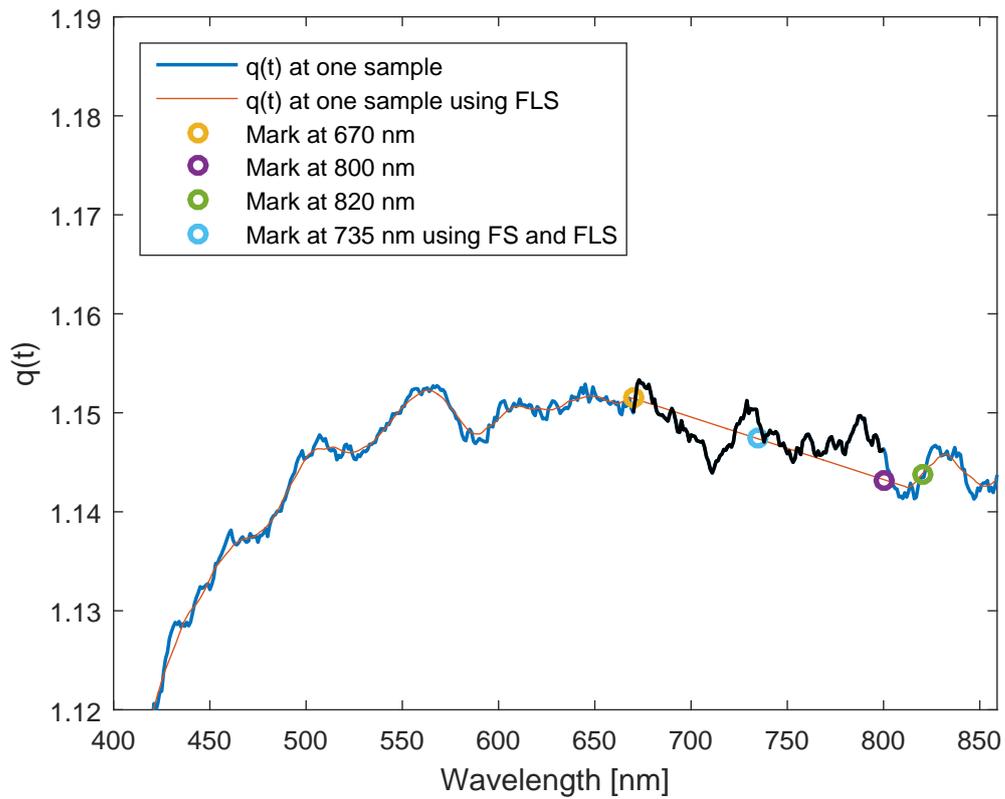


Figure 3.9: Further estimations of $q(t)$ for the 735-755 nm band.

The assumptions are analysed in Section 4.

3.7 Data collection - simulations

The following subsections describe the different data collections used for the simulation experiments (SE).

3.7.1 31st of January 2017

Table 3.2: Spectrometer setup for the experiment performed on the 31st of January.

Reference name	<i>M1</i>	<i>M2</i>
IT [<i>ms</i>]	300	300
Field of view	180°	20°
Fiber width [μm]	600	600
Distance to canopy [<i>cm</i>]	0	46

An experiment containing six dill plants standing on a green plastic carpet and two spectrometers was setup according to Figure (3.10) in a greenhouse environment in order to collect data. The goal was to collect data of the direct sunlight and the corresponding reflected light to be used in simulations. The reason for using dill plants was that different surfaces have different light absorption properties. Thus, to use plants in the data collection would potentially mean a more realistic dataset for the simulations.

The test started 10:16 in the morning and ended 11:16 the same day. It was constantly cloudy throughout the experiment but an increasing sunlight intensity was expected since the sun was still rising. The spectrometer setup can be seen in Table (3.2)

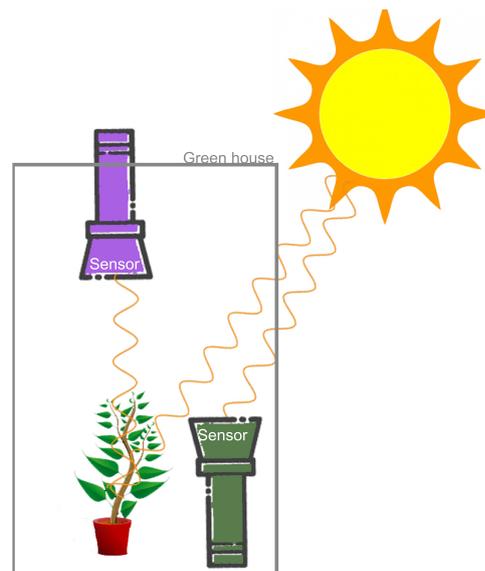


Figure 3.10: Experimental setup with dill plants and spectrometers.

3.7.2 8th of March 2017

Table 3.3: Spectrometer setup for the experiment performed on the 8th of March.

Reference name	<i>M1</i>	<i>M2</i>
IT [<i>ms</i>]	13	35
Field of view	180°	20°
Fiber width [μm]	600	600
Distance to canopy [<i>cm</i>]	0	46

An experiment similar to the one in Section 3.7.1 was made but on a day with much stronger sunlight. Another difference was also that there were no plants present, but only the green plastic carpet used in the previous experiment. This experiment was made to acquire data for stronger disturbances to test the different methods on, as well as to better be able to analyse the reflective properties of the different wavelengths in the spectrum of the sunlight.

The experiment was conducted between 11:18 and 12:18, during which the weather was partly cloudy. The spectrometer setup can be seen in Table (3.3)

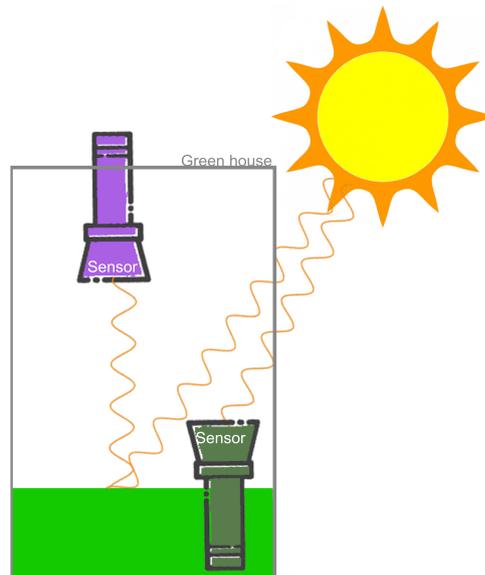


Figure 3.11: The experimental setup used for data collection on the 8th of March 2017.

3.7.3 4th of April 2017

Table 3.4: Spectrometer setup for the experiment performed on the 4th of April.

Reference name	<i>M1</i>	<i>M2</i>
IT [ms]	100	15
Field of view	180°	20°
Fiber width [μm]	50	600
Distance to canopy [cm]	0	46

This experiment imitated the one in Section 3.7.1, where the experiment was setup as in Figure (3.10). Basil plants were used together with a spectrometer at canopy level facing up (M1) and another spectrometer a certain distance above the canopy facing the plants (M2). The experiment was conducted under clear sky between 14:06 and 15:06. The spectrometer setup can be seen in Table (3.4).

3.8 Simulation setup

To be able to test and verify the different signal processing methods, a known reference fluorescence response needed to be used in order to see if the methods worked. The fluorescence response from a basil plant illuminated by a LED light at $100 \mu\text{mol m}^{-2} \text{s}^{-1}$ within PAR in a controlled environment was used as verification data; the data was measured by Lindqvist⁵. The parameters of a system with the transfer function $\left(\frac{B(q)}{A(q)}\right)$ was identified using a OE430 model with lamp light as input and fluorescence response as output, seen in Figure (3.12).

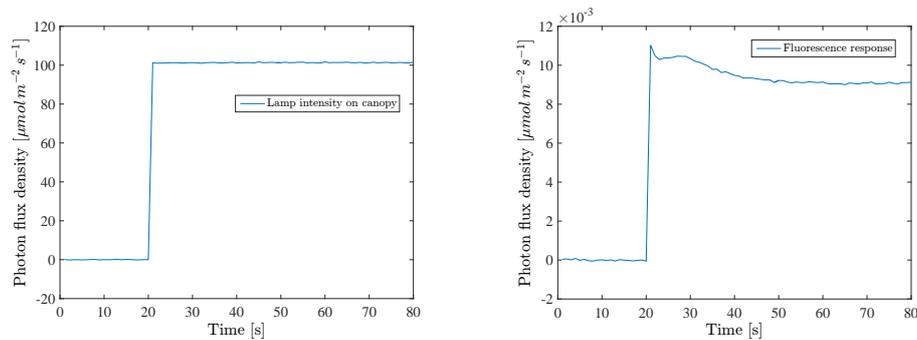


Figure 3.12: Lamp intensity and corresponding fluorescence response, with no sunlight present, was used as the verification data.

The simulation environment was set up in stages. Reflected and incoming data from 3.7.1 and 3.7.2 was used together with the verification data to create simulated signals similar to what is measured in a real environment. The lamp light $l(t)$ is on the left and the verification data $y_{f,lamp}(t)$ is seen on the right in Figure (3.12).

The model used in Section 3.5 have a fluorescence response, $y_f(t)$, that depends on sunlight as well as lamp light. The given reference was induced from only lamp

light, therefore the reference fluorescence response also needs to be dependent of sunlight. The sun induced fluorescence ($y_{f,sun}(t)$) was then added to $y_{f,lamp}(t)$ to create $y_f(t) = y_{f,sun}(t) + y_{f,lamp}(t)$. $y_{f,sun}(t)$ was calculated in Simulink by filtering the sunlight within PAR, $d_{PAR}(t)$, with the same system that was obtained through the verification data $\frac{B(t)}{A(t)}$, producing the signal shown in Figure (3.13).

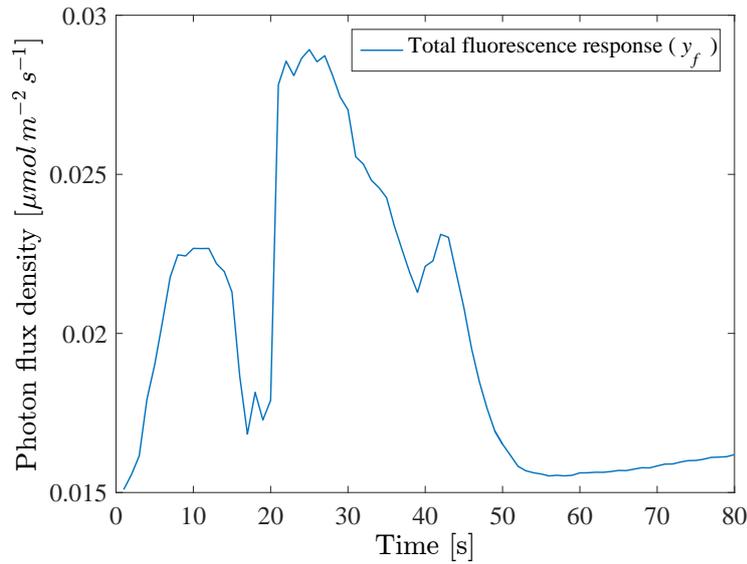


Figure 3.13: The fluorescence response from both the lamp light and sunlight put together to simulate a real fluorescence response. Note the LED-light step excitation at 20 seconds, induced by a change in the lamp light.

$y_f(t)$ was then added to the reflected light $\rho d(t)$ producing the total system output $y(t)$ that can be seen in Figure (3.14).

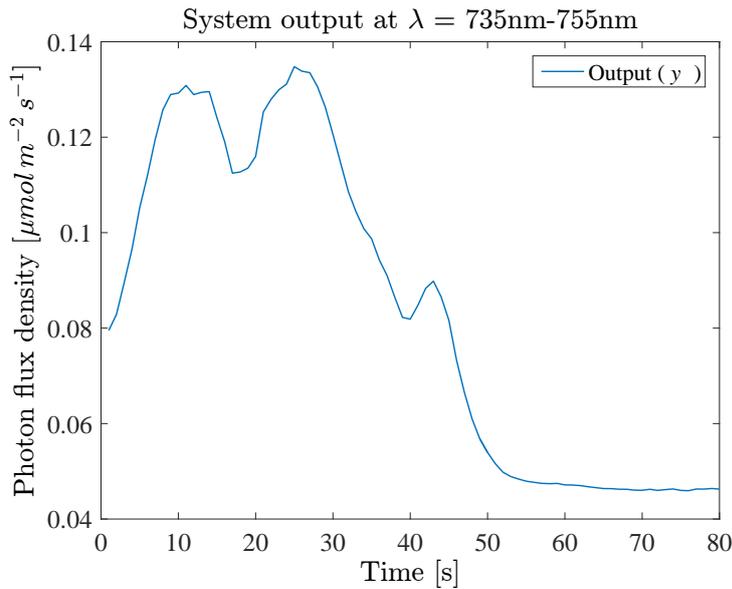


Figure 3.14: The raw sensor input from which the fluorescence response from Figure (3.12) need to be extracted.

The plant system model is thereby written as

$$y(t) = y_f(t) + \rho d(t) + e(t) \Rightarrow \quad (3.11)$$

$$y(t) = \frac{B(q)}{A(q)} \underbrace{(l(t) + d_{PAR}(t))}_{u(t)} + \rho d(t) + e(t) \quad (3.12)$$

Where $d(t)$ was the incoming sunlight within the 735-755 nm band, this in order to extract the ChlF response. Even though $d(t)$ was the incoming sunlight, $\rho d(t)$ was measured as reflected sunlight using sensor M2 in Section 3.7.1. $e(t)$ was the error between estimated signal and true true signal.

Finally the input $u(t)$ and output $y(t)$ was filtered with the same low-pass Butterworth filter to remove noise, see section 3.3 for details on the filter.

A data set length of 80 s was used throughout the project as it is a sufficient amount of time to observe the full ChlF response while still having 20 s prior to the step.

3.8.1 Simulation results visualised

The dataset needed in order to perform one of these simulations needed to match the length of the ChlF response, i.e. the same 80 second length. The data was, however, collected during a full hour with a 1 second sampling time, which means 3520 different points where the $y_{f,lamp}(t)$ and LED-light step $l(t)$ could be added when doing the setup. Since all collected data were 3600 samples, but only 80 samples were used for calculations, the collected data was divided into 80 samples long simulation sets. The first simulation set was acquired by using sample 1 to 80, from

the 1 hour dataset, adding $y_{f,lamp}(t)$ and do estimations where the fit (2.10) and error was saved. The next dataset started from sample 2 to 81 and so on. This meant 3520 separate tests for each method, each with its own fit and error, which is described below. When evaluating curve fitting an OE430 was fitted to the final result to be able to calculate a fit, and when using parametric modelling the estimated parameters made out the system that was to be compared to the known reference step response. All of these fits were saved and sorted in a rising order to visualise how well the methods performed on every dataset.

There were two ways of calculating the error of the results, one for curve fitting and one for the parametric modelling methods. The error of curve fitting was calculated as the difference between $\rho d(t)$ and $\rho \hat{d}(t)$, which was supposed to be zero if a perfect fit was obtained. The sum of squared errors was then calculated from that difference. The same sorting index that was used to sort the fit was used to sort all 3520 error values.

As for the parametric modelling, the difference between $q(t)$ for the 735-755 nm band and the $q(t)$ substitute was squared and summed. This substitute band is explained further in Section 3.6. The error was then sorted with the same indices as the fit. This was done to investigate if there was a correlation between the fit and $q(t)$ error.

3.9 Data collection - realistic test environment

The final tests were performed in the Heliospectra greenhouse where five basil plants were used as canopy along with two lamps (LX601G). Three thermometers were also used to measure the average temperature near the canopy. The lamps excited the canopy with blue light in the 450 nm spectrum in order to get a ChlF step response but the red and white LEDs were also needed to act as background light. Comparing the lamps background light spectrum to that of the sun, both with the same intensity in PAR, gave data like the one in Figure (3.15).

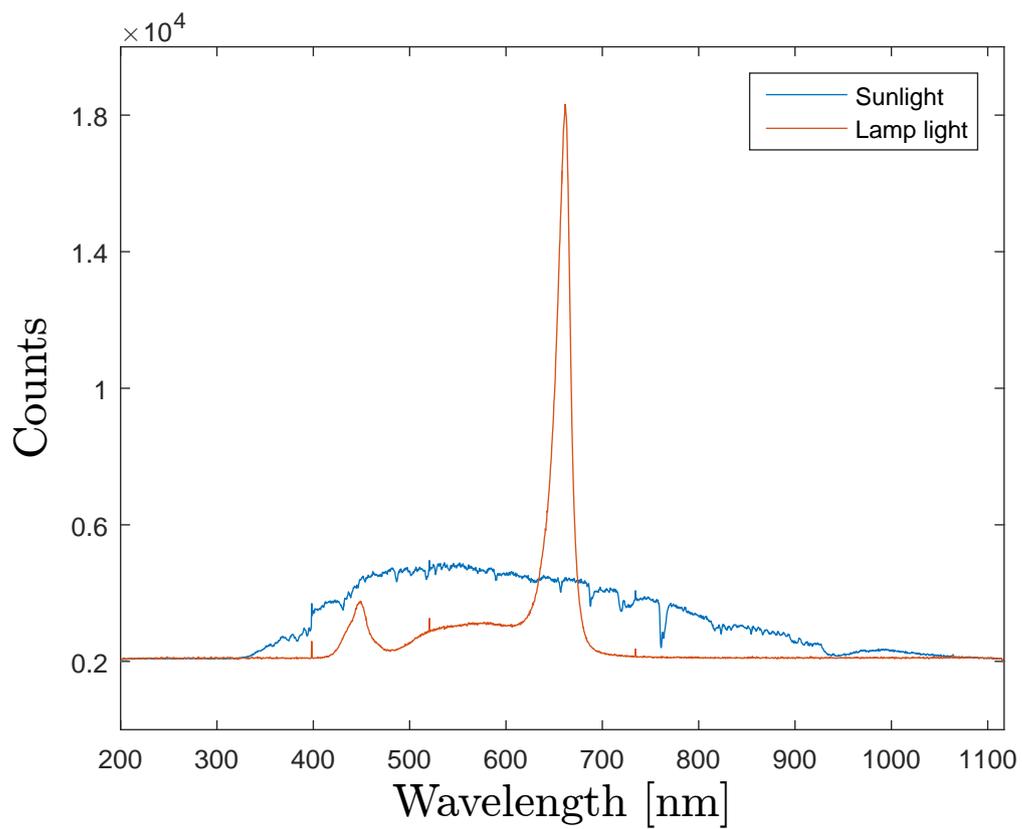


Figure 3.15: A visual representation of the sunlight and background light produced by the LED-lamp. Both spectra produced the same PFD in PAR

3.9.1 12th of April 2017 - Greenhouse experimental setup

Table 3.5: Spectrometer setup for the experiments performed on the 12th of April.

Reference name	<i>M1</i>	<i>M2</i>
IT [<i>ms</i>]	100	15 / 100*
Field of view	180°	20°
Fiber width [μm]	50	600
Distance to canopy [<i>cm</i>]	0	40

Figure (3.16)) with a tarpaulin covering the stall and blocking the sunlight, which could easily be removed. This allowed for swift blackout of the canopy after making measurements in sunlight. The lamps were then used to mimic the sunlight intensity by illuminating the canopy with similar intensity as the sun outside, using red and white LEDs. One difference between the light sources was that the lamp did not have a significant output in the ChlF band. Therefore, all of the reflected data collected in the 735-755 nm range would only contain fluorescence and no reflected light. The reason the lamps needed to have approximately the same background light as the sun was because the fluorescence response of a plant changes depending on what light the plant was adapted to⁴. Using the same background light intensity in both cases will get a reference that should be as similar as possible to the true ChlF step response and can therefore be used as an indication of the quality of the results.

The experiments were first conducted in sunlight, where the lamps excited the canopy with blue light at $100 \mu mol m^{-2} s^{-1}$ in PAR after 20 seconds. When the measurement was completed the tarpaulin was pulled down and the red and white LEDs were turned on with intensity that matched the sun's. The blue LEDs then turned on, like before, and excited the canopy to obtain a reference response.

The measurements were performed in intervals of five minutes. First of four measurements were made under sunlight to collect different data for analysis. When the initial phase was completed, the tarpaulin was pulled over the stall and four reference measurements were made. A measurement was made with five minute intervals, giving the plant time to acclimate to the background light of the lamp during the reference data collection. This routine was done for two different light intensity levels (around 50 and $100 \mu mol m^{-2} s^{-1}$ in PAR), where 8 references and 8 test measurements were made in total.

* 15 ms was used during the first data collection and 100 ms during the second.

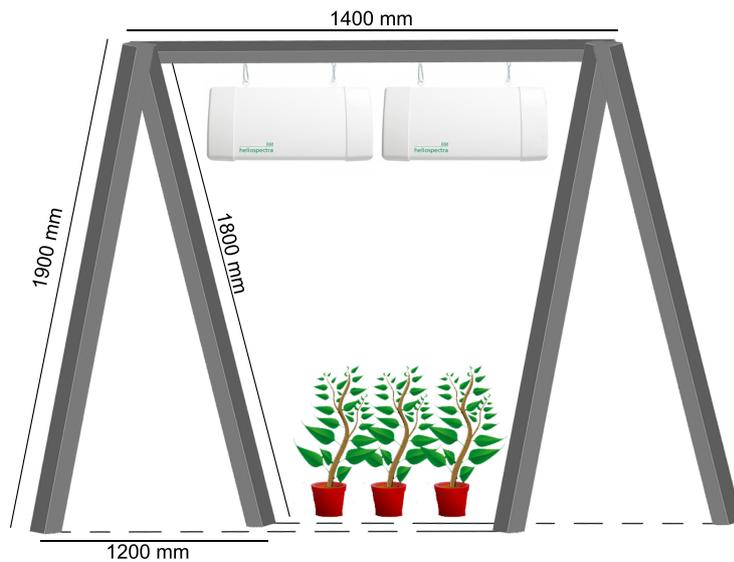


Figure 3.16: The stall used in the greenhouse experiments, note that the tarpaulin is not included in this Figure.

3.9.2 26th of April 2017 - Greenhouse experimental setup

Table 3.6: Spectrometer setup of the experiment performed on the 26th of April.

Reference name	<i>M1</i>	<i>M2</i>
IT [<i>ms</i>]	150 / 100*	13
Field of view	180°	20°
Fiber width [μm]	50	600
Distance to canopy [<i>cm</i>]	0	40

This test was also done in the Heliospectra greenhouse, and the same setup was used as in Section 3.9.1. The sunlight intensity for this measurement was higher compared to previous measurements, which was also intended in order to evaluate the methods under different light conditions.

Since the plants have an adaptation time of roughly 15 minutes²⁸, the reference data was this time collected 20 minutes after the tarpaulin was pulled down. The following 4 measurements were made every 2 minutes. A shorter time between the measurements was used to minimise the effect of the increasing temperature on the plants in the stall. Note that the sunlight intensity for the tests were higher than the lamps could imitate, the lamps could only go as high as $500 \mu mol m^{-2} s^{-1}$ in PAR.

Another reference measurement was made where no background lights was used, where a response of only an excitation of the blue light was made. This was done since the noise level of the signal increases with increasing lamp intensity, which was a problem in the first 4 reference measurements. The idea is that since the first 4 reference measurements all contain noise but a shape can clearly be seen, then if the

* 150 ms for the test with sunlight and 100 ms for the reference measurements.

final 4 reference measurements has the same dynamics but without the noise, they can be used to get a noiseless reference signal.

4

Results

THE RESULT SECTION contains the results that have been obtained through the experiments. It also encloses the results of the assumption analysis to answer the questions put forward in Section 3.6. The simulations are shown before the greenhouse testings, and all tests are presented in chronological order of the different methods used. To show the simulation results in a way that is easy to overview the resulting fit has been sorted in increasing order with its corresponding error.

4.1 Curve fitting assumption analysis

Curve fitting builds upon the assumption that the ratio between the spectrum 735-755 nm and 800-820 nm is constant throughout different reflected sunlight intensities. To test this assumption the data from the sunlight collection experiments was used to see if the ratio between the two wavelength bands was, in fact, constant. Each data sample from the 735-755 nm band was divided with the corresponding sample from the 800-820 nm band.

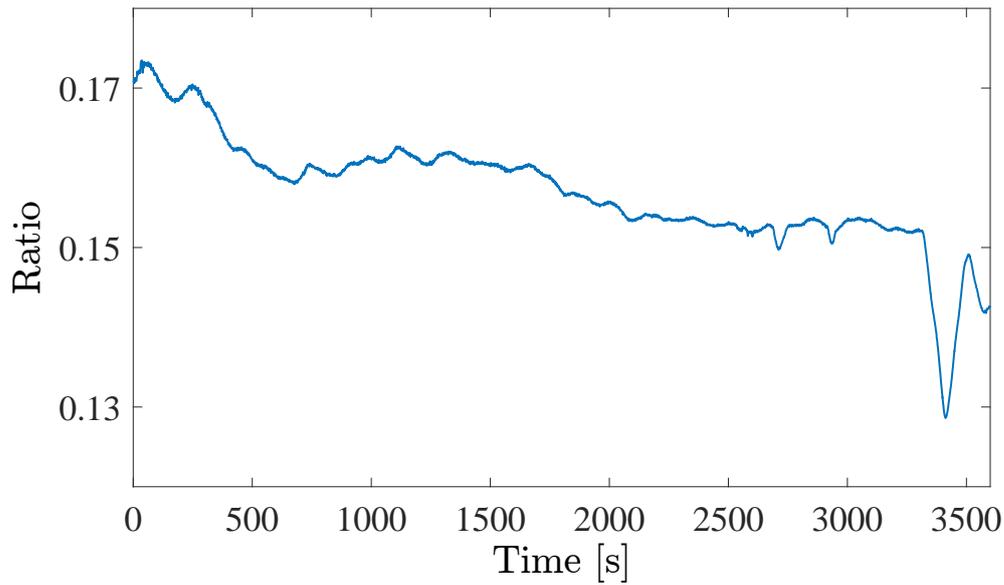


Figure 4.1: Ratio of 735-755 nm to 800-820 nm of the reflected sunlight in the experiment in Section. 3.7.1.

From Figure (4.1) it can be concluded that the ratio is not perfectly constant. It is also not correct to state constant stress levels during all periods of 80s. However, this particular data set contains some fluorescence from the dill plants which can be the reason for the varying behaviour. Figure (4.2) shows the same dataset but incoming light intensity.

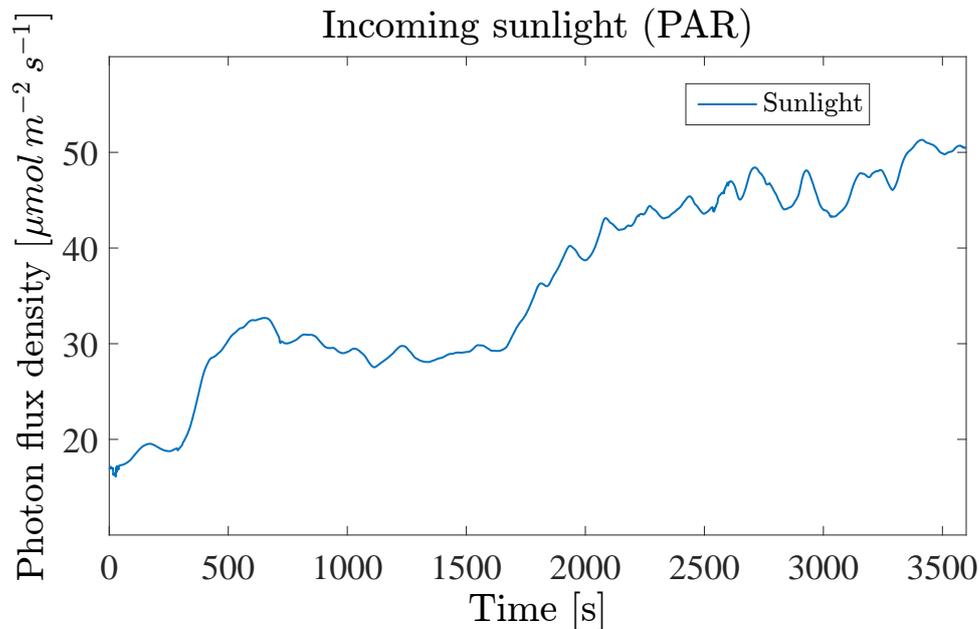


Figure 4.2: Incoming sunlight from 10:16 to 11:16, see data collection from Section 3.7.1

During the data collection from Section 3.7.2, the weather was more varied. This was found to create a much more nonlinear ratio between the wavelength bands.

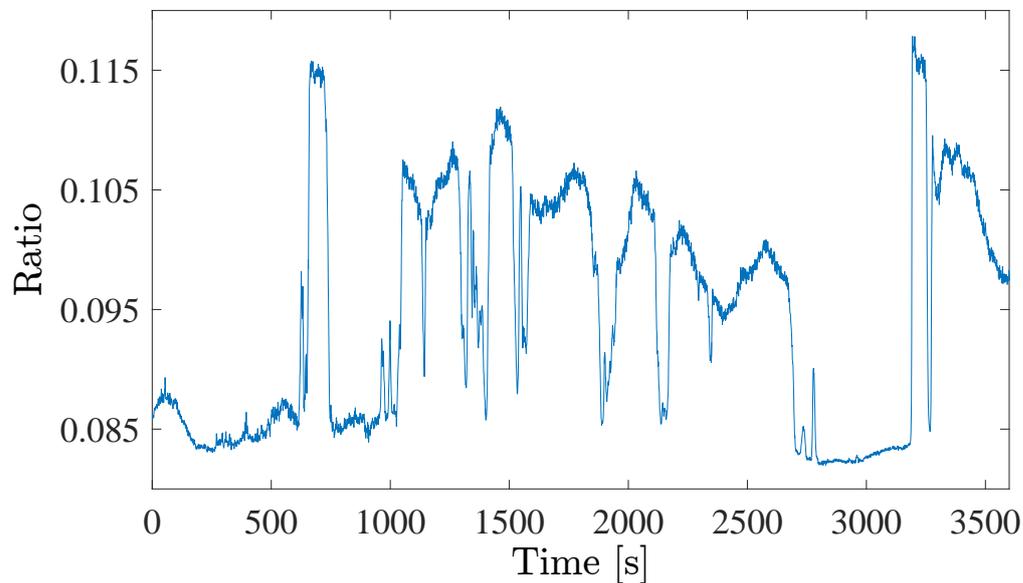


Figure 4.3: Ratio of reflected 735-755 nm and 800-820 nm from the second data collection, see section 3.7.2

From Figure (4.3) it can be observed that it would be hard, even for a period of 80 s, to capture constant behaviour of the ratio which is, again, needed for curve fitting to work properly. A connection can be drawn to the incoming sunlight which varied intensely. Note that comparing it to the incoming light intensity from Figure (4.4), the ratio is correlated with the light intensity.

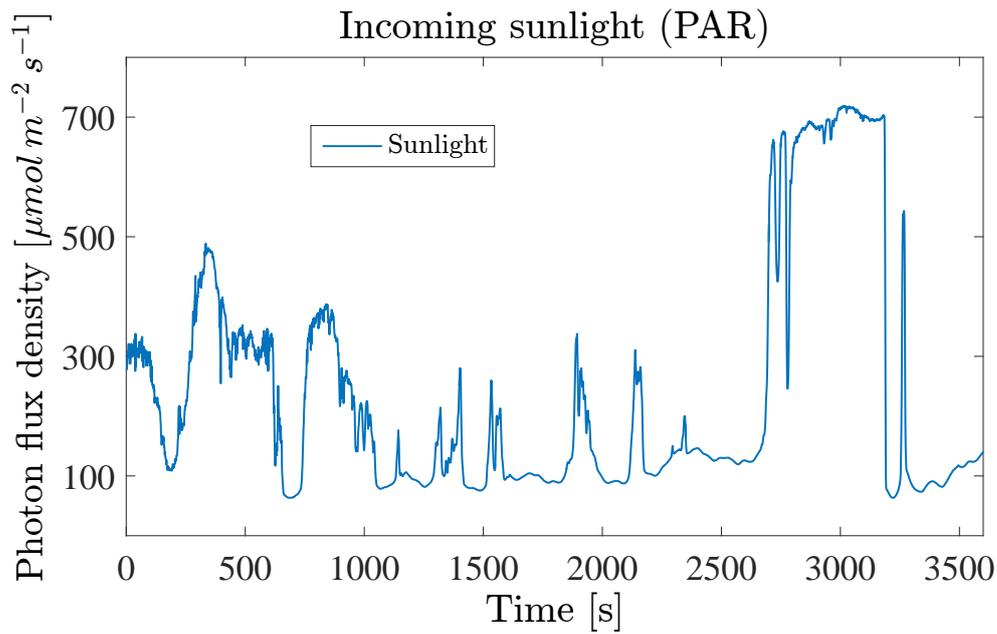


Figure 4.4: Incoming sunlight from 11:18 to 12:18, see data collection from Section 3.7.2

In the final data collection, the sunlight was very strong, creating a nonlinear ratio between the wavelength bands where Figure (4.5) also shows a varying reflectance. Figure (4.6) shows the dataset for the corresponding incoming light intensity.

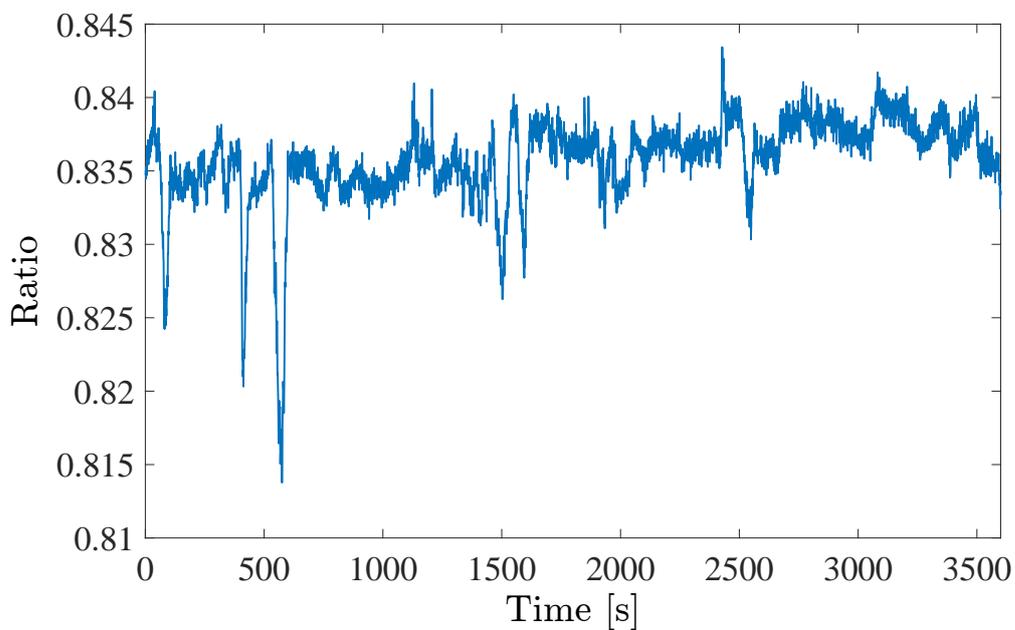


Figure 4.5: The ratio is clearly depending on the sunlight intensity. It decreases when the sunlight gets strong and increases when the sunlight decreases.

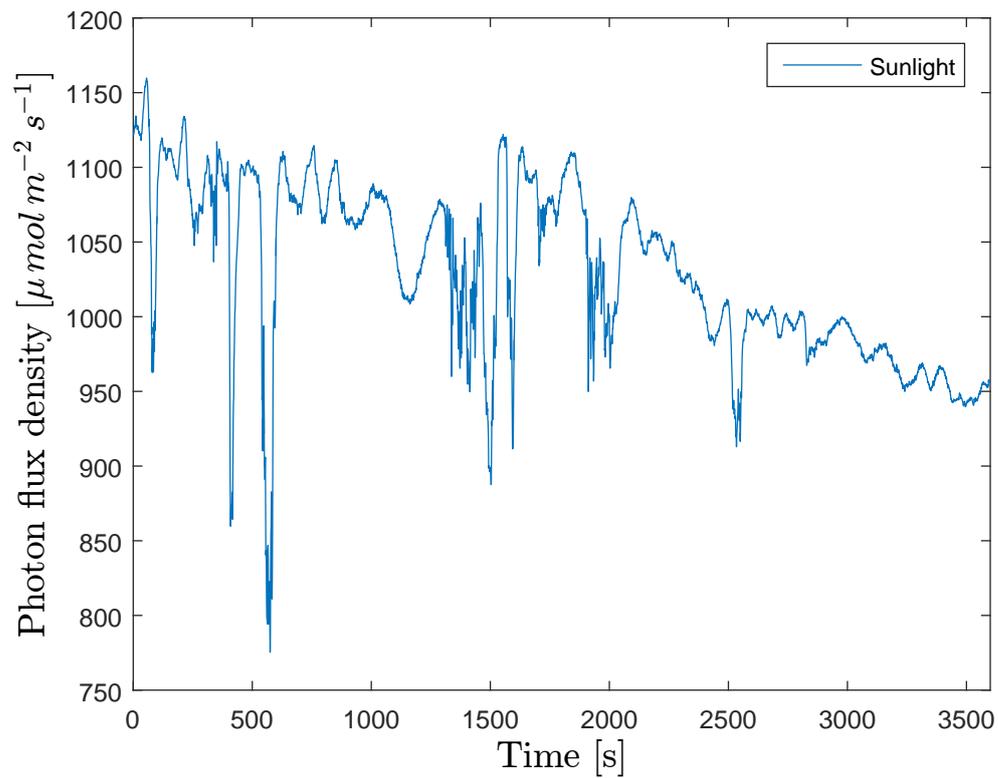


Figure 4.6: Incoming sunlight from 14:06 to 15:06, see data collection 3.7.3

4.2 Parametric modelling assumption analysis

The idea that the reflection ρ would be constant during an arbitrary time period was something that was generally believed to be true. Also, the assumption that the reflection ratio would occur for all wavelengths that is not conflicting with PAR was made.

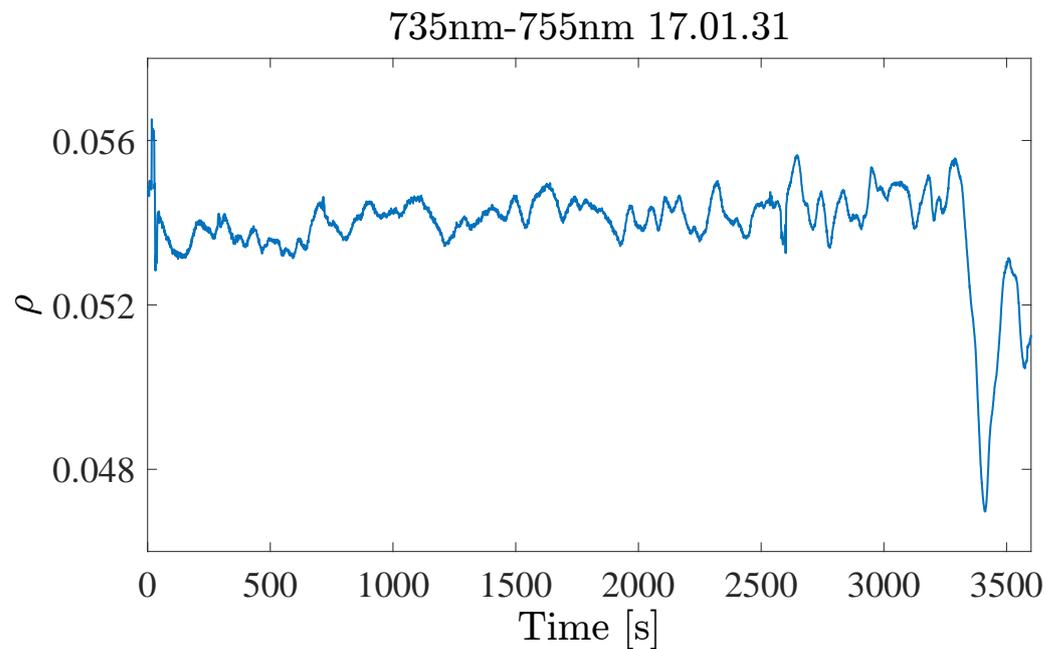


Figure 4.7: ρ for the fluorescent wavelength band during the one hour of measurements made in the end of January.

Figure (4.7) show that ρ appears constant within relatively small boundaries except for the dips in the end of the measurement. The reason for the dip is most likely due to one sensor being partly shadowed. The same dip can be seen in Figure (4.1) since this also is from the first data collection, (3.7.1).

The assumption that ρ would be the same for different wavelengths is tested by doing the same calculations for the 800-820 nm band. From Figure (4.8), it was found that ρ behaves in another manner, and is thereby not the same for all wavelengths.

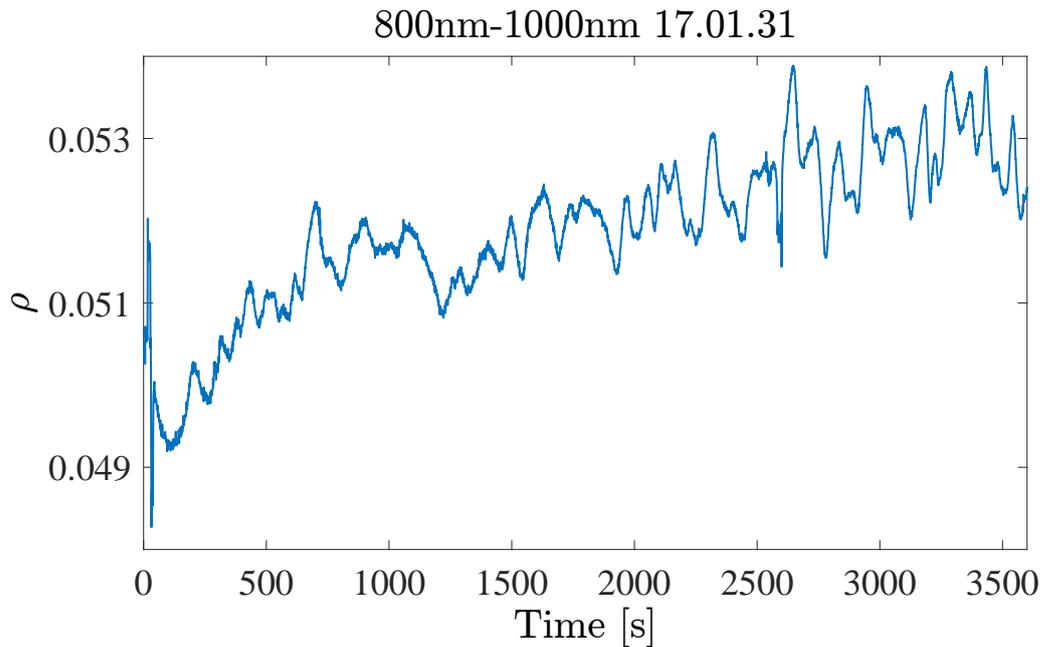


Figure 4.8: ρ for the non-fluorescent wavelength band during the one hour of measurements made in the end of January.

When testing the "constant ρ assumption" on the March data, the assumption was found to be wrong. The large change in sunlight intensity caused the reflection to vary a lot (see Figure (4.9)).

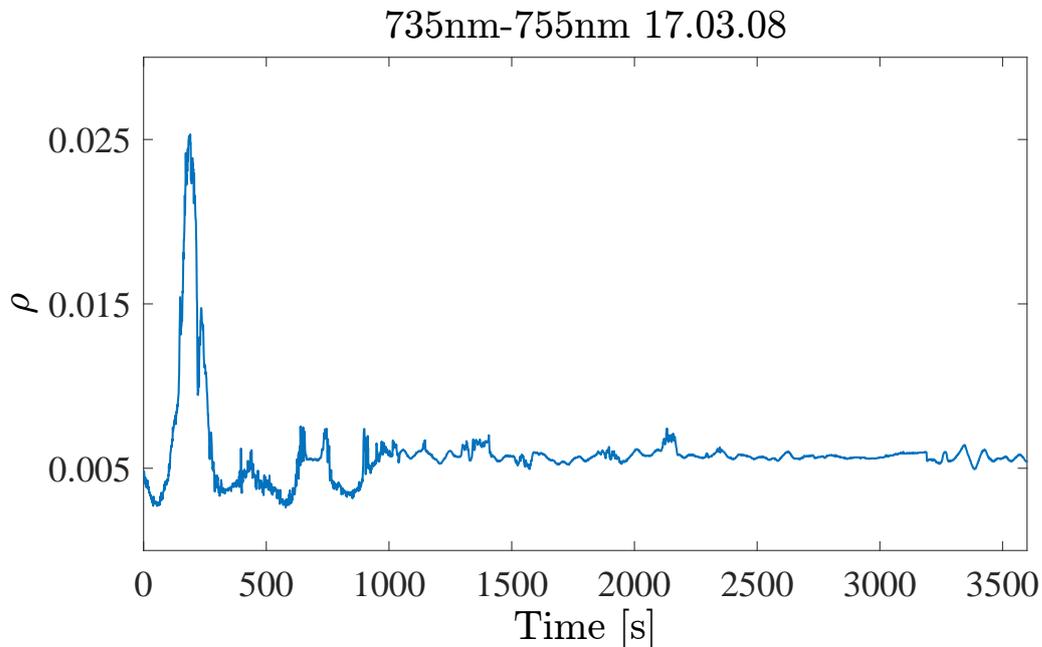


Figure 4.9: ρ for the fluorescent wavelength band during the one hour of measurements made in March (no plants). Note the peak between 100 and 300 seconds, which is probably the result of a shadow passing over sensor M1.

4.3 Simulation results

The following section shows the results from all three data collection simulations. The visualisation of the simulation results are explained more thoroughly in Section 3.8.1.

4.3.1 Simulation experiment 31/1/2017

The collected sunlight data from Section 3.7.1 can be seen in Figure (4.10).

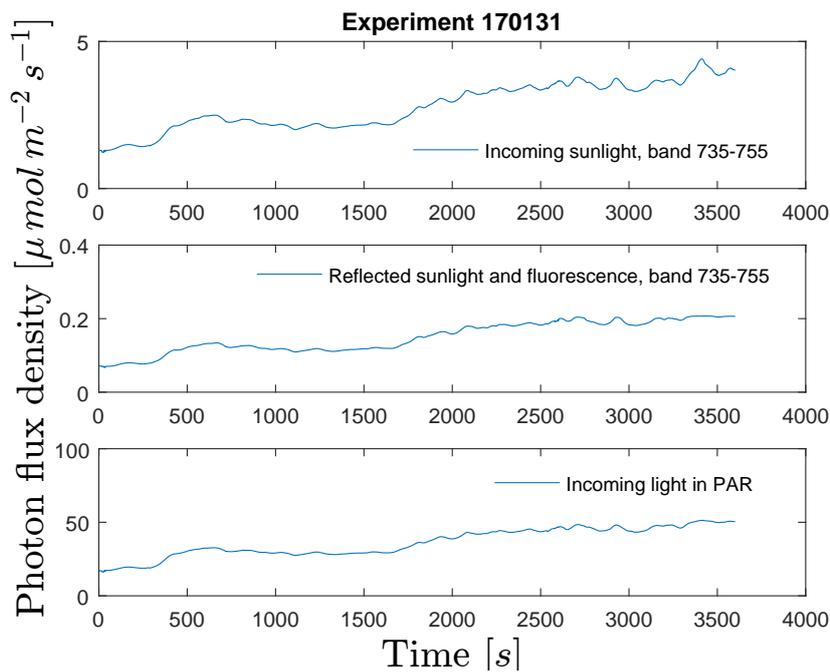


Figure 4.10: Incoming and reflected light integrated over the 735-755 nm band with incoming light in PAR at an average of $40 \mu\text{mol m}^{-2} \text{s}^{-1}$.

Curve fitting

Figure (4.11) shows the fit, from Equation (2.37), and error for all collected data simulated at each sample throughout the full hour of measurements. The error is the difference between $\hat{d}(t)$ and the ground truth $d(t)$, from Section 3.4.

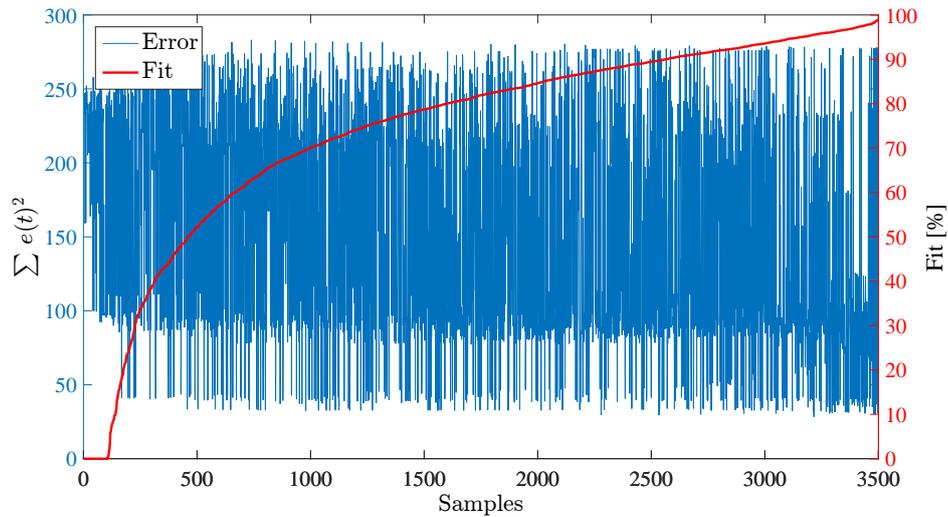


Figure 4.11: The samples are sorted, in rising order, with respect to its fit percentage. The error shows little correlation to the fit. Mean fit: 74,45% Median fit: 81,90%

Prefiltering algorithm

The results for the parametric models are evaluated according to the error between the real and substitute $q(t)$, explained in Section 3.6, and the fit between true output and estimated output. The fit has been sorted in rising order with its corresponding error. A clear connection between the two can be observed in Figure (4.12).

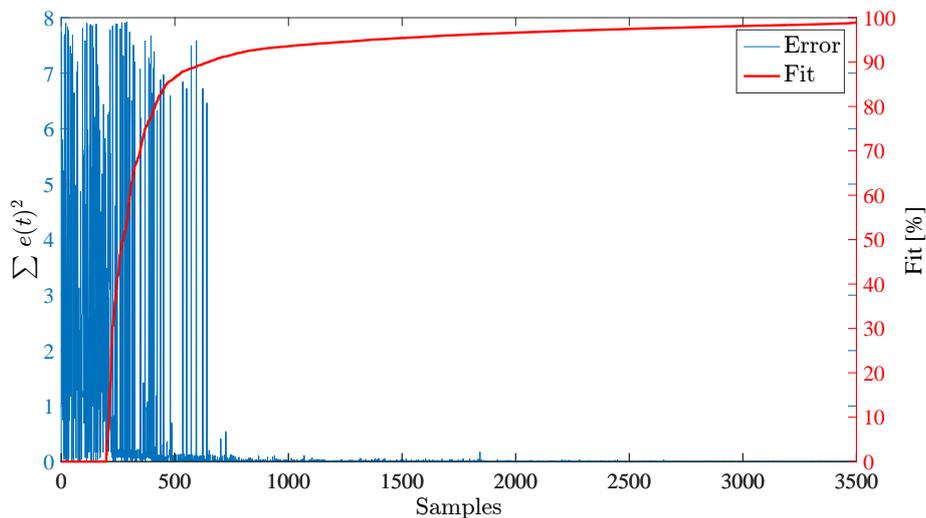


Figure 4.12: A gathering of all the fits and corresponding errors from the experiment in Section 3.7.1 when using the prefiltering algorithm, using FS. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 87,64% Median fit: 96,10%

dGN algorithm

Similar results are seen when using the dGN algorithm (see Figure (4.13)).

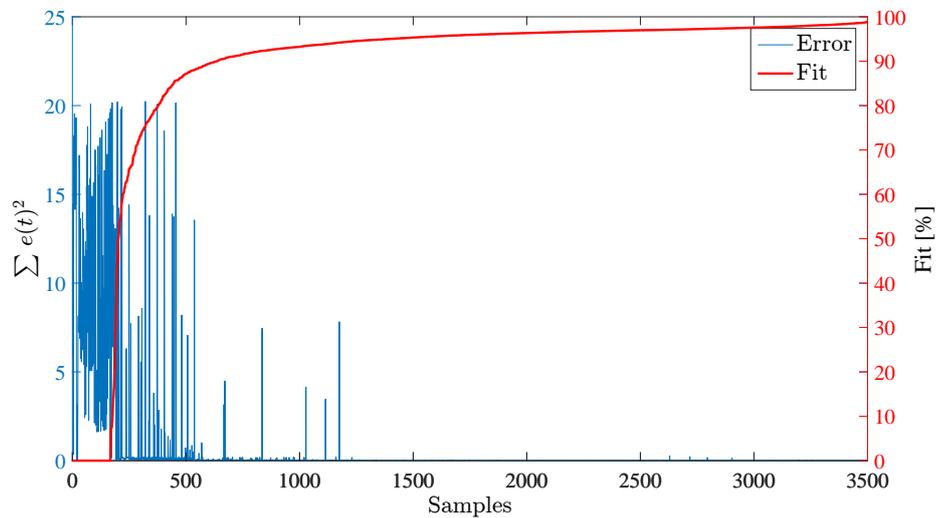


Figure 4.13: A gathering of all the fits and corresponding errors from section 3.7.1 when using FLS. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 88,61% Median fit: 95,94%

Although the $q(t)$ error is varying between the simulations, because of FLS and FS, the trend is still clearly visible. Less error means a better fit and result.

4.3.2 Simulation experiment 8/3/2017

Figure (4.14) show the measured light of the data collected in the experiments in Section 3.7.2.

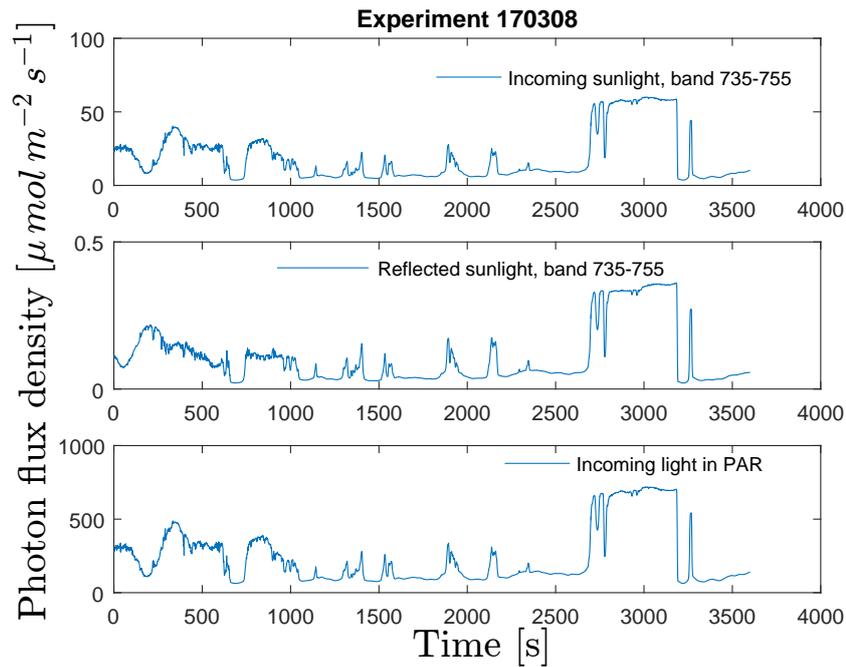


Figure 4.14: Measured light intensities from experiment in Section 3.7.2

The sunlight intensity is shifting a lot and is significantly stronger than the data collected in Section 3.7.1. Figure (4.15) shows the performance when simulating and evaluating the final fit and the error between the treated incoming sunlight and the reflected sunlight.

Curve fitting

The simulation results for curve fitting can be observed in Figure (4.15).

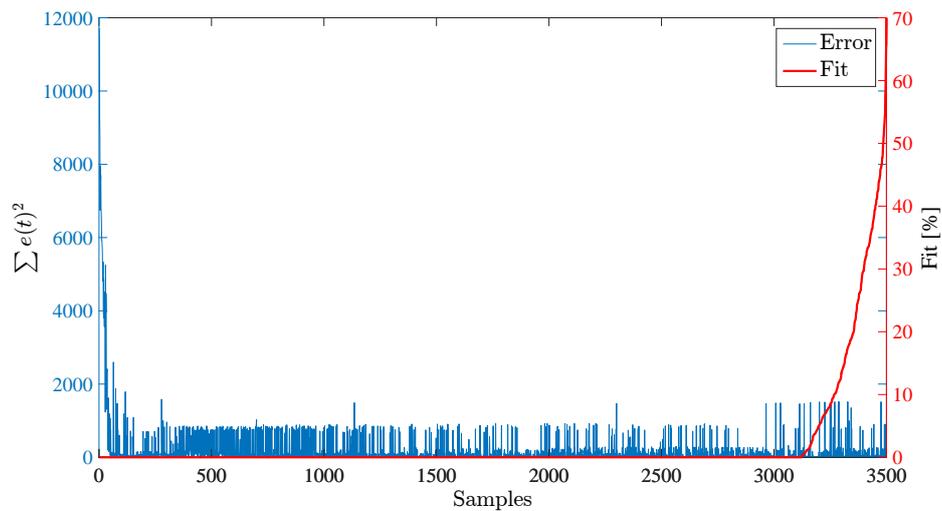


Figure 4.15: Results of performance of curve fitting on the 3.7.2 data. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 2,08% Median fit: 0%.

Figure (4.15) demonstrates the lack of performance of the curve fitting method when stronger sunlight is present. Only about a tenth of the samples produced a fit over 0%.

Prefiltering algorithm

The simulation results for the Prefiltering algorithm can be observed in Figure (4.16).

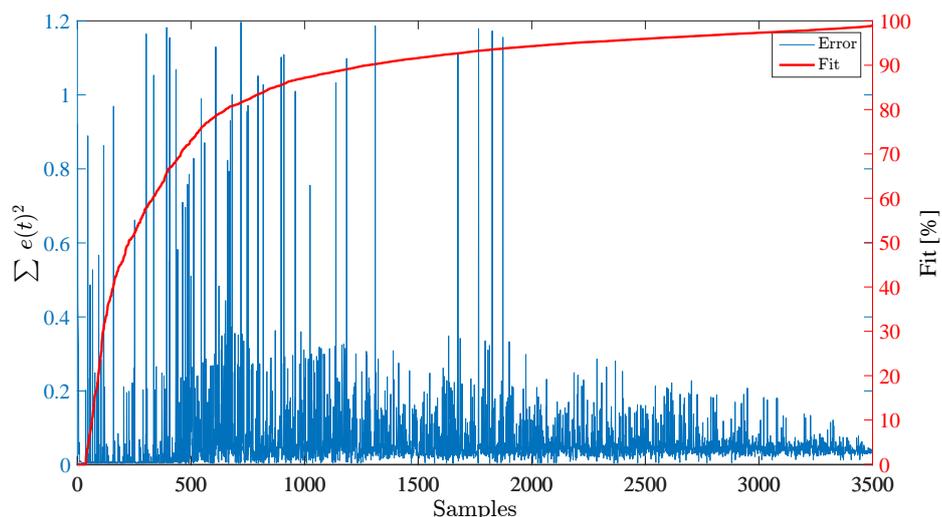


Figure 4.16: A gathering of all the fits and corresponding errors from the experiments in Section 3.7.2, using FLS. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 85,74% Median fit: 93,18%

The results are significantly better with the prefiltering algorithm when comparing

to curve fitting for the same data.

dGN algorithm

The simulation results for the dGN algorithm can be observed in Figure (4.17).

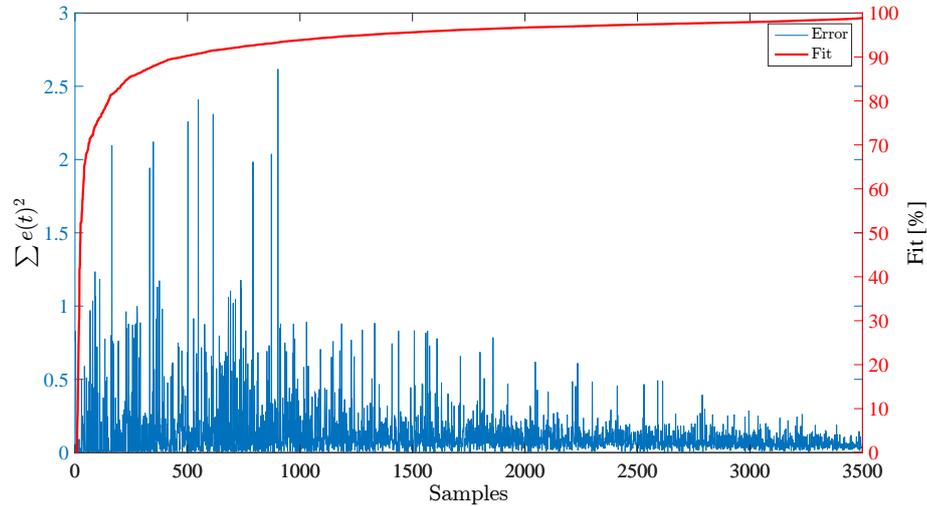


Figure 4.17: A gathering of all the fits and corresponding errors from the experiments in Section 3.7.2, using FS. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 93,72% Median fit: 96,23%

The results show that parametric modelling drastically outperforms curve fitting for higher sunlight intensities.

4.3.3 Simulation experiment 4/4/2017

Figure (4.18) shows the results of the data collection from the experiment in Section 3.7.3.

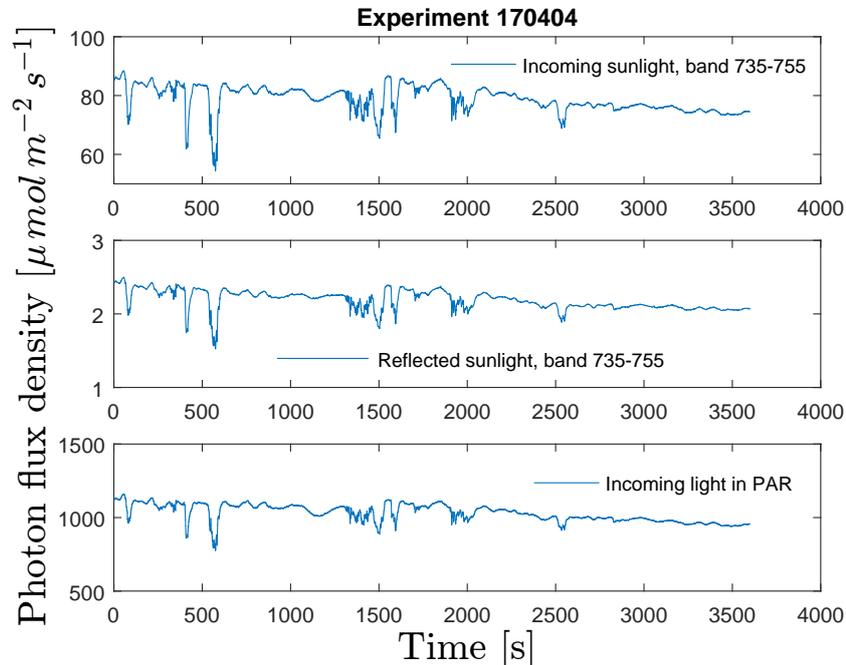


Figure 4.18: Measured light intensities from experiment in Section 3.7.3

Curve fitting

The simulation results for curve fitting can be observed in Figure (4.19).

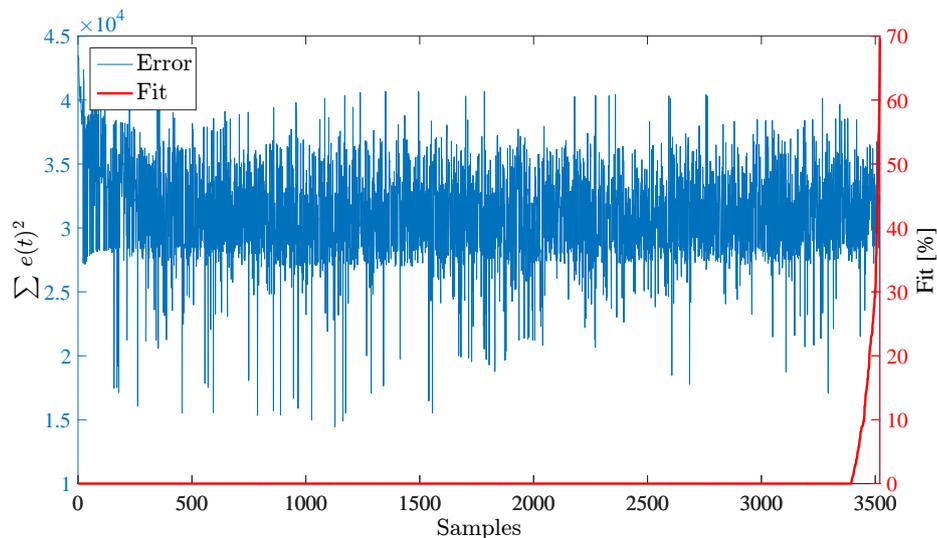


Figure 4.19: A gathering of all the fits and corresponding errors from 3.7.3. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 0,63% Median fit: 0%.

Prefiltering algorithm

The simulation results for the Prefiltering algorithm can be observed in Figure (4.20).

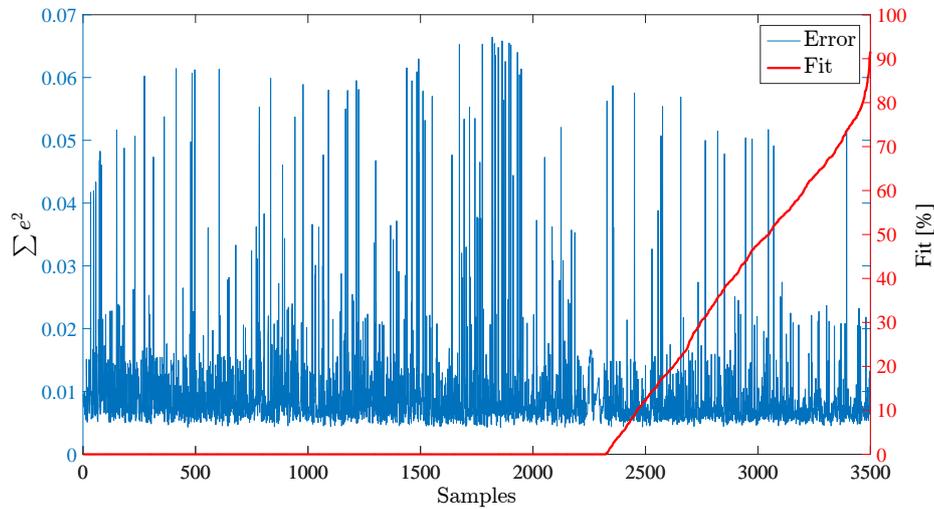


Figure 4.20: A gathering of all the fits and corresponding errors from the experiments in Section 3.7.3, using FLS. The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 13,72% Median fit: 0%.

dGN algorithm

The simulation results for the dGN algorithm can be observed in Figure (4.21).

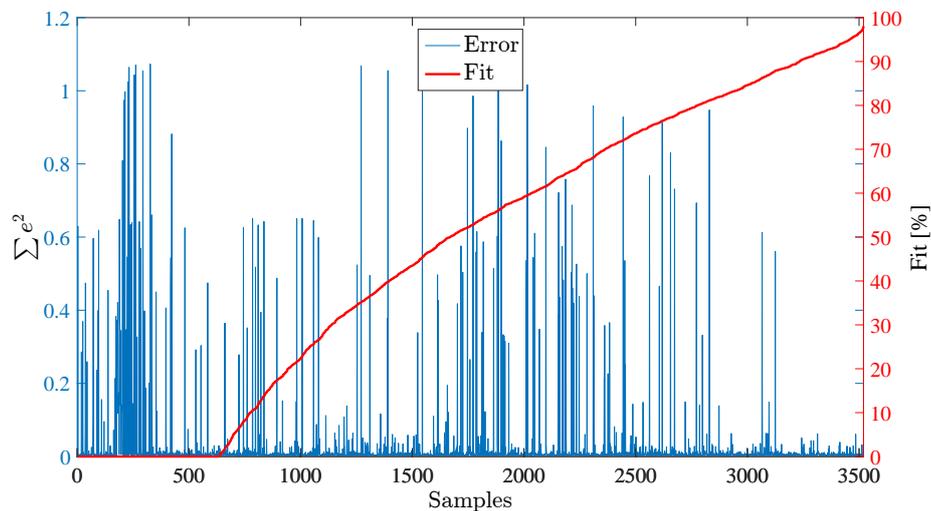


Figure 4.21: A gathering of all the fits and corresponding errors from 3.7.3, using $q(t)$ for band 800-820 nm (without FLS or FS). The samples are sorted, in rising order, with respect to its fit percentage. Mean fit: 47,56% Median fit: 52,54%.

4.4 Greenhouse results

In the following the results are shown for the tests made in a realistic environment in the Heliospectra greenhouse. These results are made as a practical test for the methods tested in previous simulations.

4.4.1 Greenhouse Experiment 1 and 2 - 12/4/2017

The greenhouse experiments are divided into the order of which they were made, starting with the first test.

1)

Figure (4.22) shows the data collected from the experiment in Section 3.9.1 and is called Experiment 1. Figure (4.23) contains the reference data of the experiment. The sunlight intensity in PAR was around 100-130 [$\mu\text{mol m}^{-2} \text{s}^{-1}$] during these 4 tests. Table (4.1) shows the temperature for all reference measurements. The mean temperature during the sunlight tests were 20, 0°.

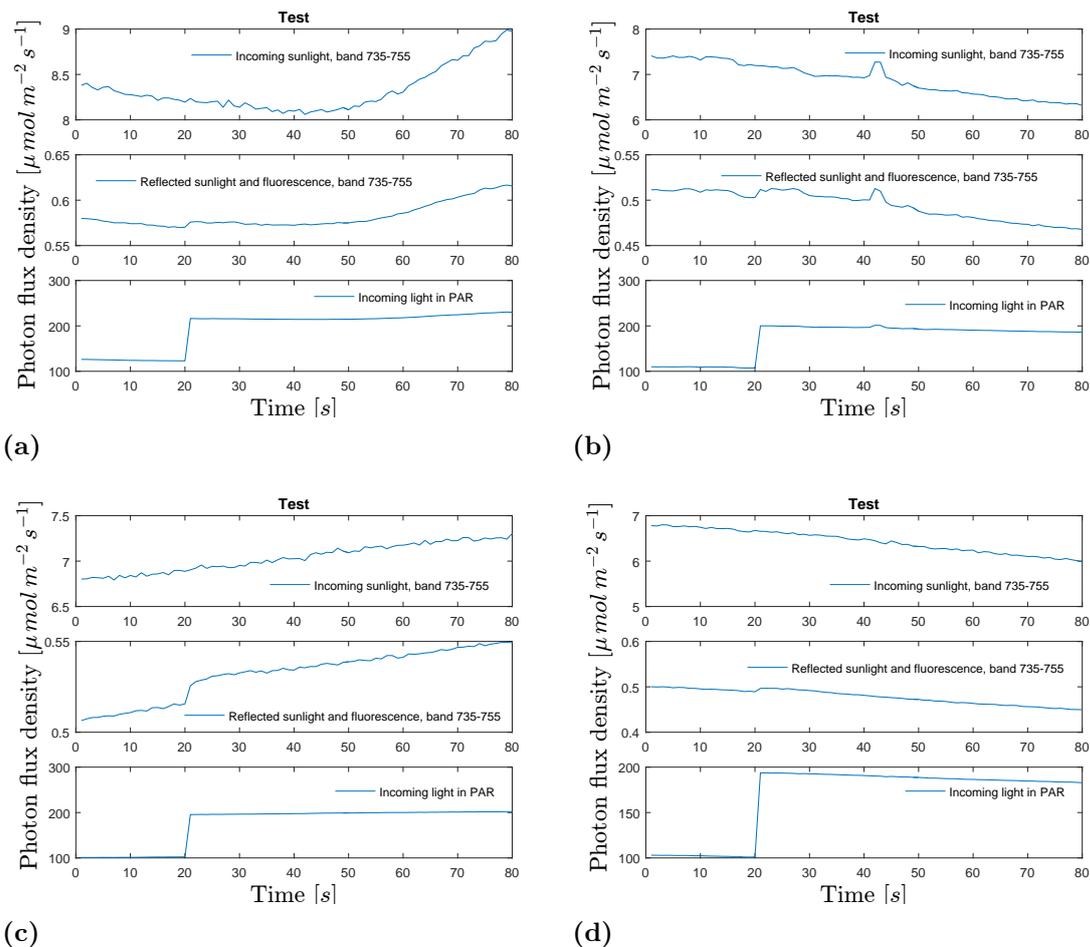


Figure 4.22: Collection of all four test measurements in Experiment 1. The fluorescence response is hard to distinguish from the raw data. There were five minutes between each test, starting with (a).

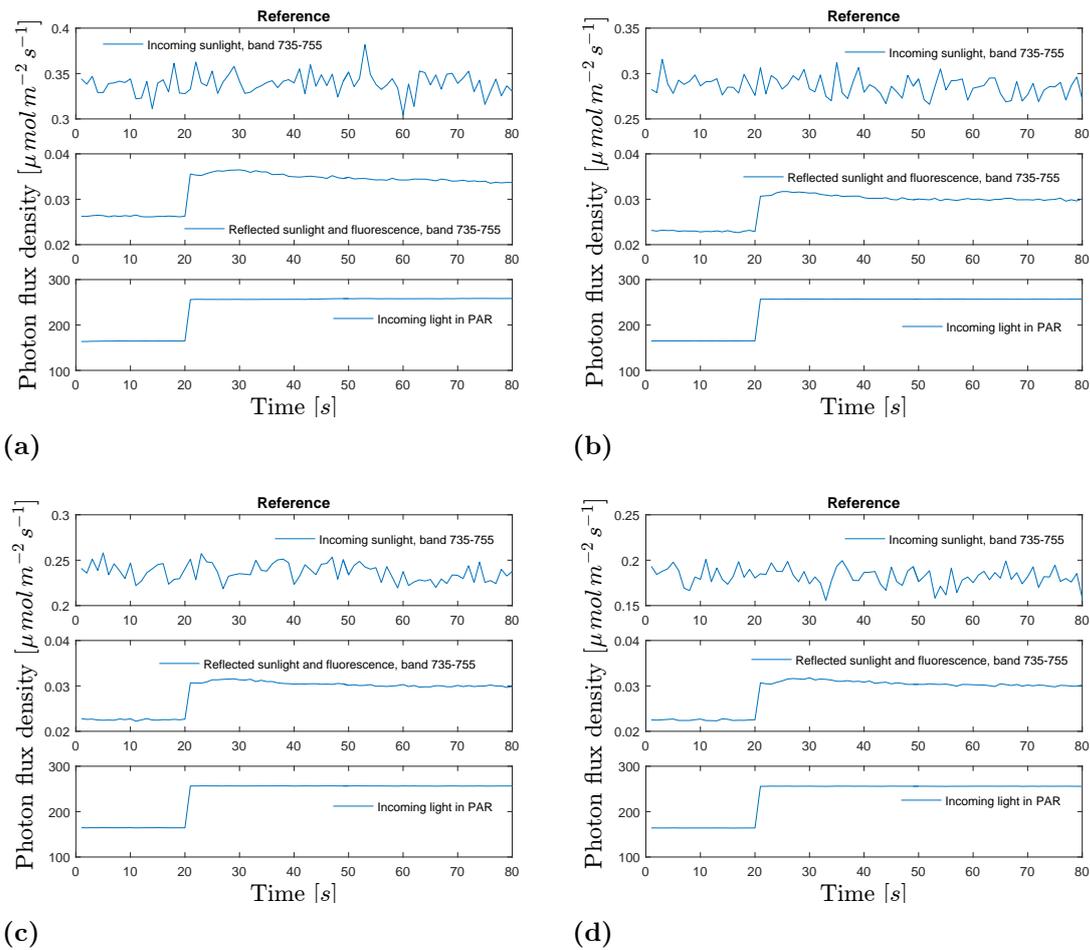


Figure 4.23: Collection of all four reference measurements in Experiment 1. The fluorescence response is relatively easy to distinguish from the raw data. There were five minutes between each data collection start, starting with (a).

Table 4.1: Temperatures in Experiment 1.

<i>Reference</i>	<i>Temperature</i>
(a)	25,1°
(b)	25,6°
(c)	25,8°
(d)	25,5°

In order to be able to use the reference measurements and compare them to the sunlight tests, the system needs to be estimated with a parametric model from the raw data. Using an OE430 model, results like the ones in Figure (4.24) were achieved for the identifications from the reference measurements.

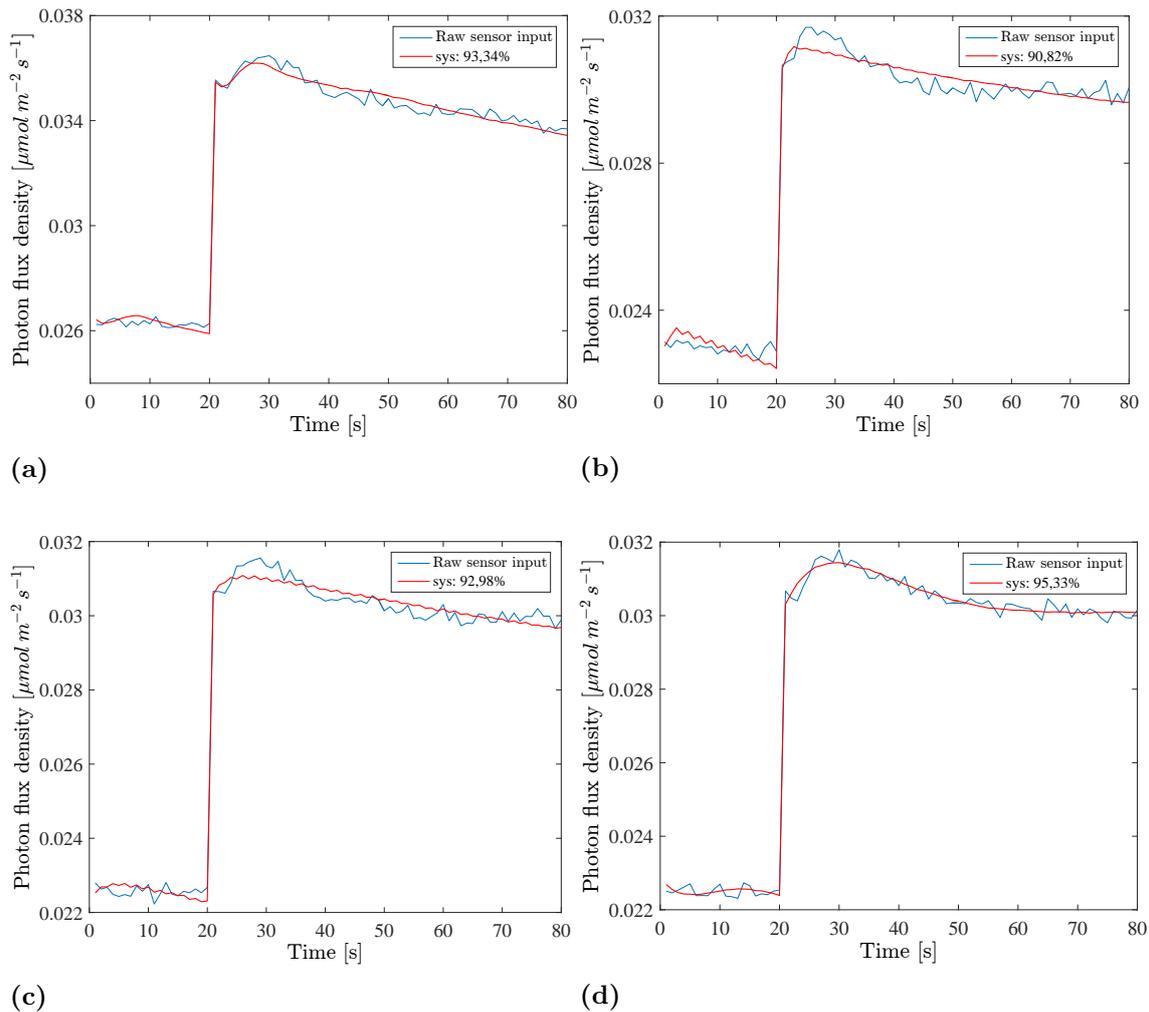


Figure 4.24: The four step responses and corresponding OE430 response from the reference measurements. In order from first to last.

In this case, system (a) is used for validation since it showed plausible dynamics and was the reference closest in time to the sunlight tests. However, a downward trend can be spotted in the first three measurements. This is believed to occur due to the change from sunlight to lamp light. Even though the intensity is roughly the same the change in spectrum may cause the plant to behave slightly different. The trend would therefore be a result of the plant adapting to the new light spectrum and thus decreasing its fluorescence response during the first 10 to 15 minutes. Note that reference raw data with obvious trends will be de-trended in the upcoming validation tests to make a more fair comparison.

Curve fitting

Each curve fitted result was estimated, and thus smoothed, with an OE430 model to better be able to compare the results. These results are seen in Figure (4.25).

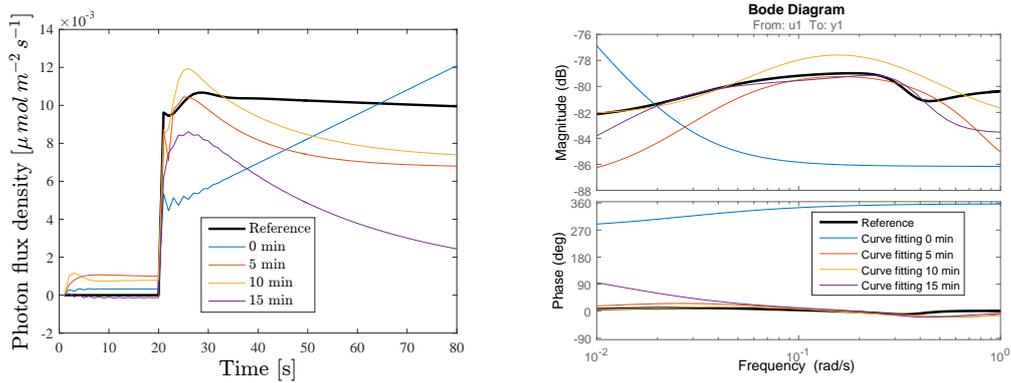


Figure 4.25: Results from the Curve fitting estimations of the Experiment 1 data. The results are shown in both time domain and frequency domain.

Prefiltering algorithm

Figure (4.26) shows the estimated results for both time and frequency domain using the Prefiltering algorithm.

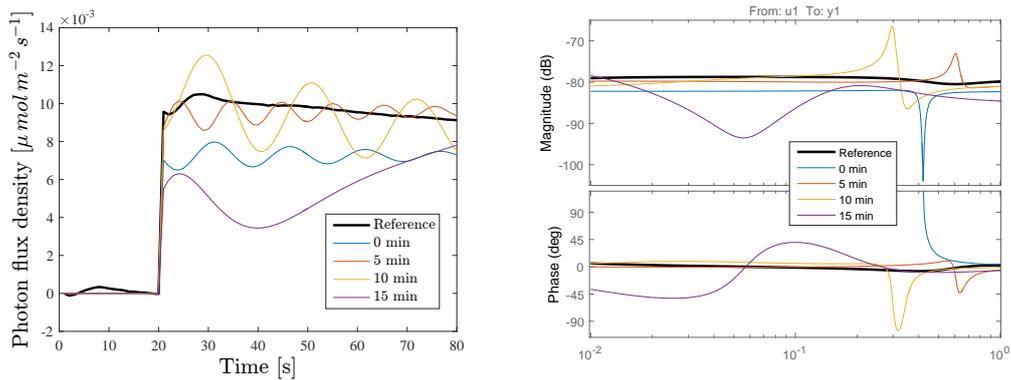


Figure 4.26: Experiment 1 results when using the Prefiltering algorithm.

dGN algorithm

Figure (4.27) shows the estimated results for both time and frequency domain using dGN.

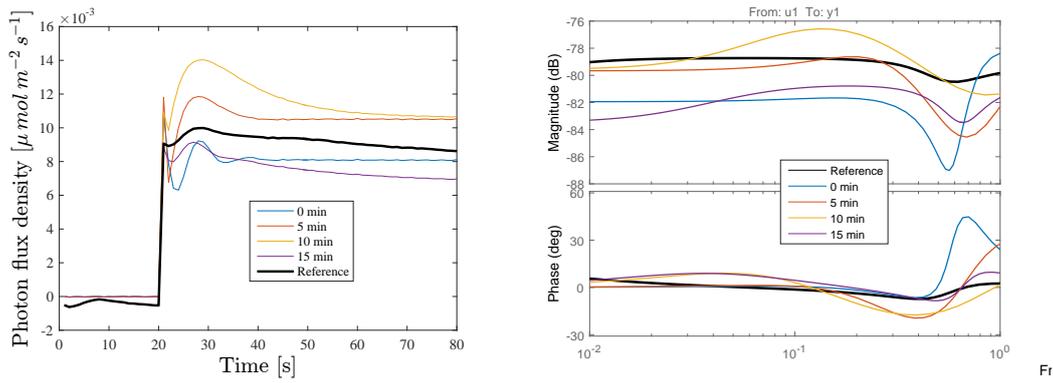


Figure 4.27: Results from the dGN algorithm estimation of the Experiment 1 data.

2)

Figure (4.28) shows the data collected from the Experiment in Section 3.9.1. The PFD in PAR was around 60-80 $\mu\text{mol m}^{-2} \text{s}^{-1}$ during these 4 tests. Figure (4.29) shows the corresponding reference measurements. Table (4.2) shows the temperature for all reference measurements. The mean temperature during sunlight tests were 16, 0°.

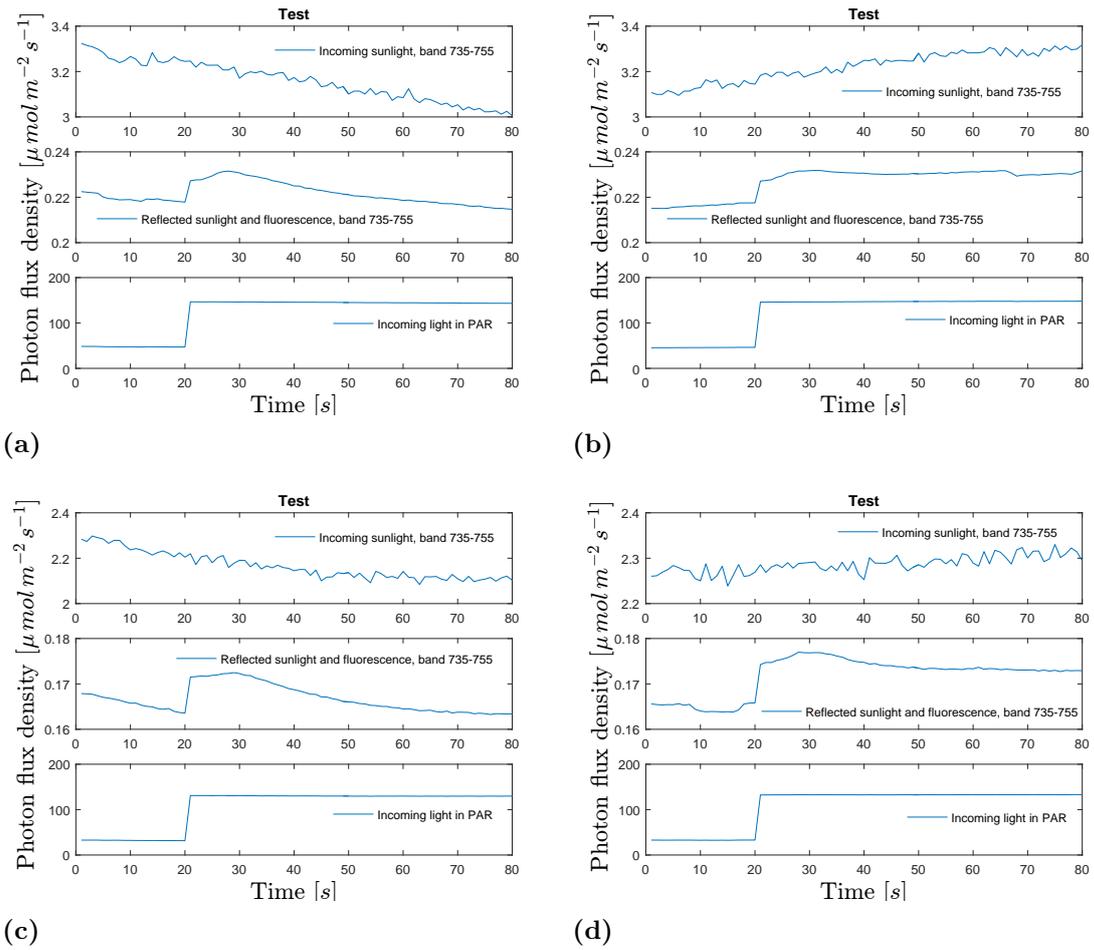


Figure 4.28: Collection of all four test measurements in Experiment 2. The fluorescence response is much easier to distinguish from the raw data in Experiment 2 because of the low sunlight intensity. There were five minutes between each data collection, starting with (a).

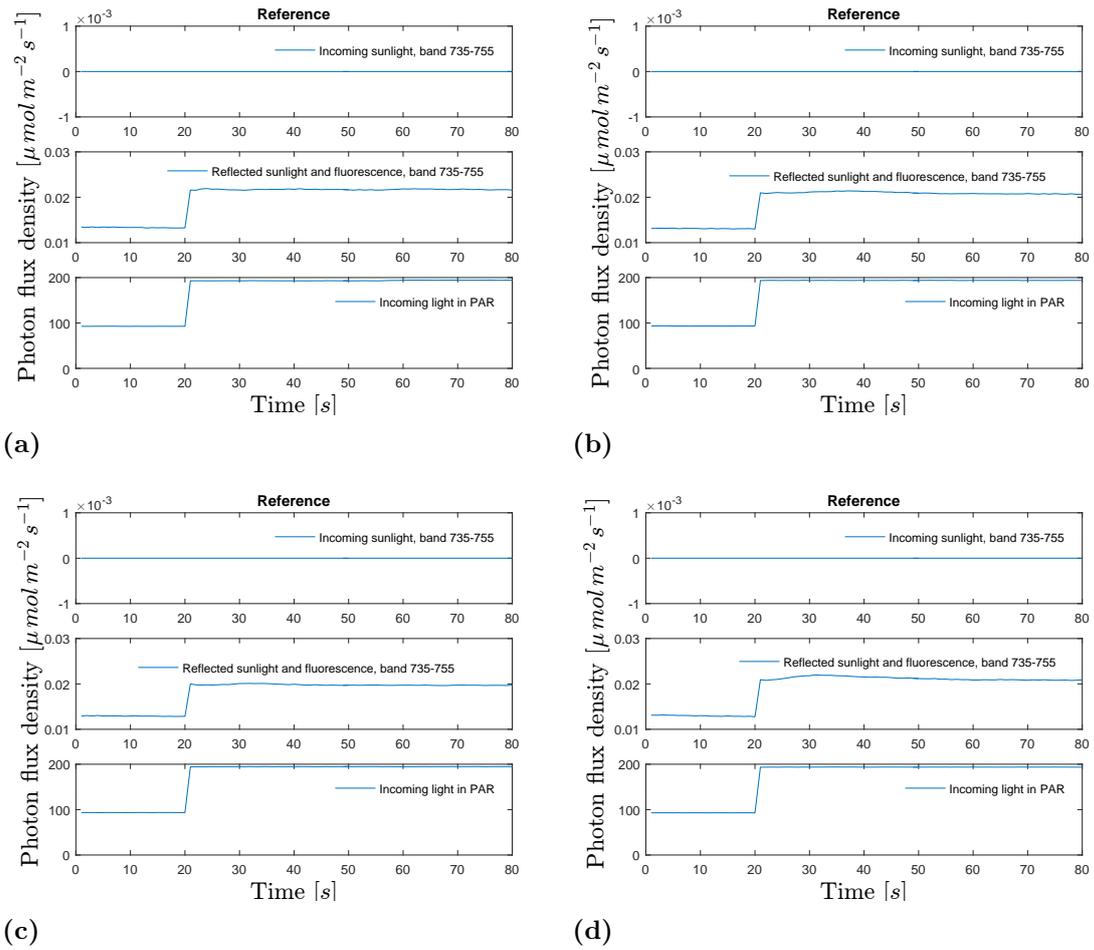


Figure 4.29: Collection of all four reference measurements in Experiment 2. The fluorescence response is clearly visible in the raw data. There were five minutes between the start of each data collection, starting with (a).

In order to get a good estimation this dataset, an OE540 model had to be used, the estimations are illustrated in Figure (4.30). System (d) is used in the upcoming validation since the model estimation contains all necessary dynamics seen in the raw data.

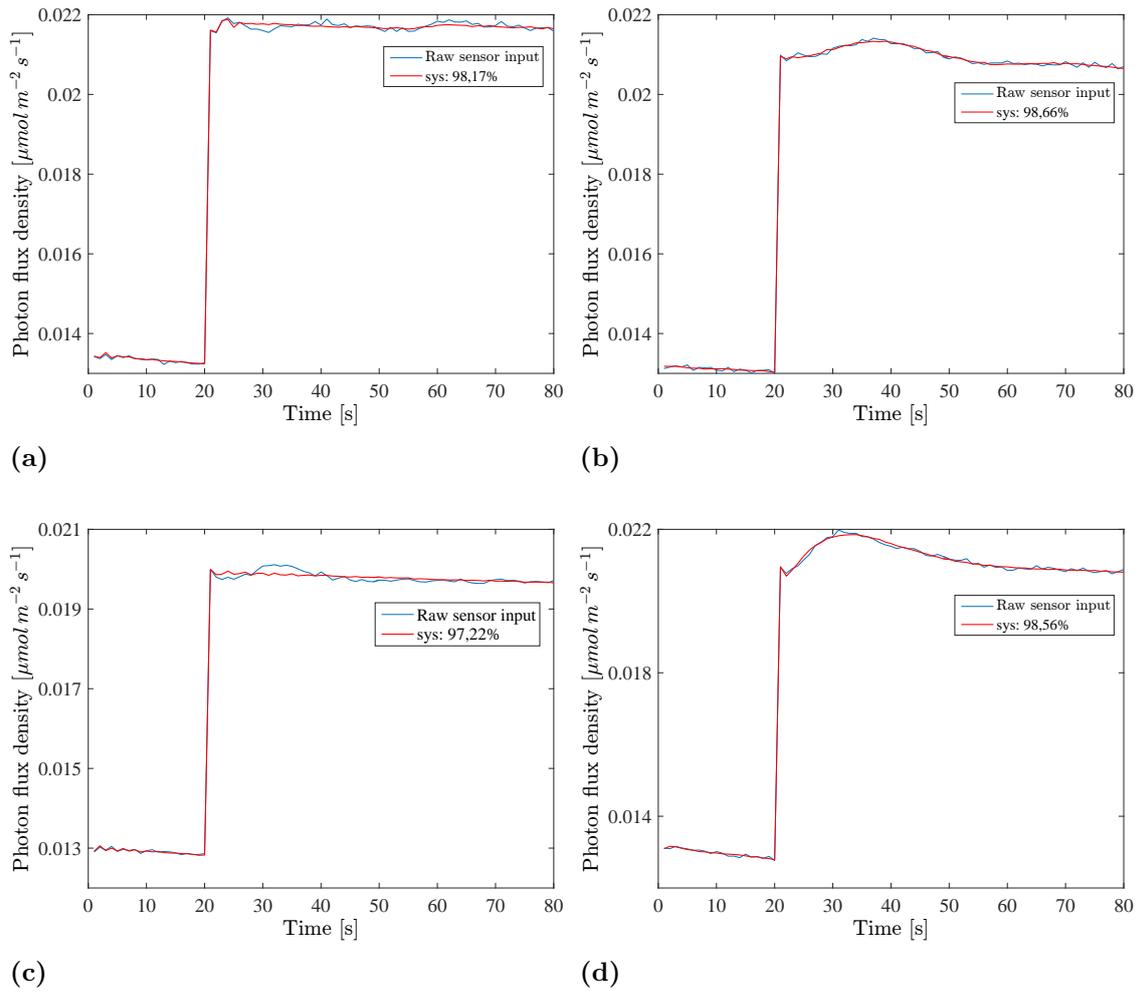


Figure 4.30: The four step responses and the corresponding OE540 responses for the reference measurements in Experiment 2. In order from first to last.

Table 4.2: Temperatures for Experiment 2.

<i>Reference</i>	<i>Temperature</i>
(a)	18,3°
(b)	18,5°
(c)	22,4°
(d)	22,9°

Curve fitting

Figure (4.31) shows the estimated results for both time and frequency domain using curve fitting.

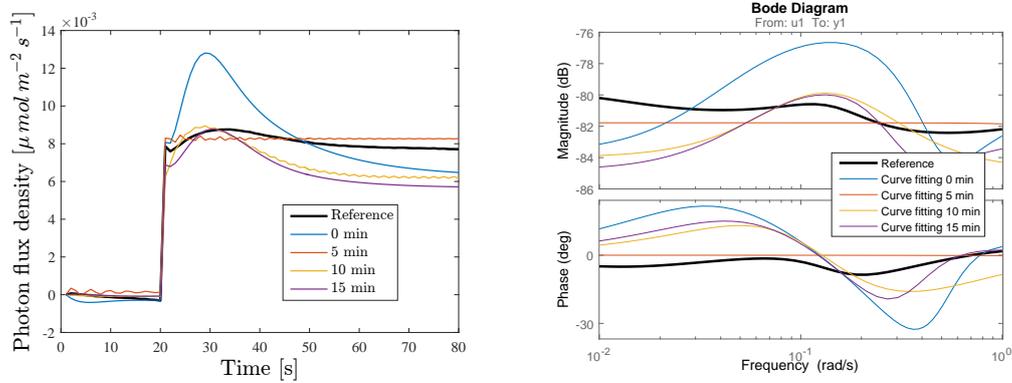


Figure 4.31: Results from the Curve fitting estimations for Experiment 2 data.

Prefiltering algorithm

Figure (4.32) shows the estimated results for both time and frequency domain using the Prefiltering algorithm.

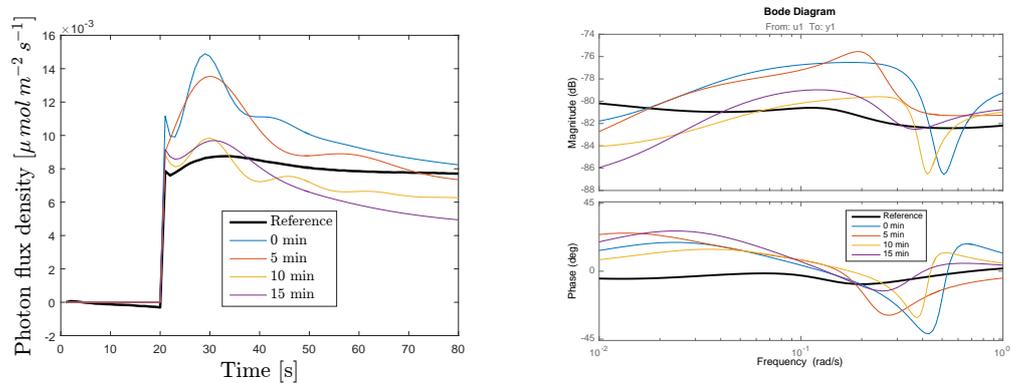


Figure 4.32: Results from the Prefiltering algorithm estimations for Experiment 2 data.

dGN algorithm

Figure (4.33) shows the estimated results for both time and frequency domain using dGN.

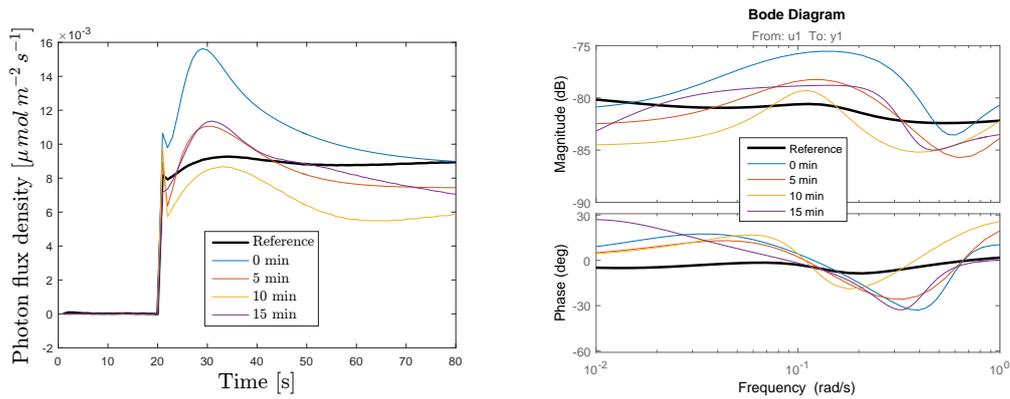


Figure 4.33: Results from the dGN algorithm estimations for Experiment 2 data

4.4.2 Greenhouse experiment 3 - 26/4/2017

Figure (4.34) shows the test data collected from Section 3.9.2, which is called Experiment 3. The sunlight intensity in PAR was between $400\text{--}800 \mu\text{molm}^{-2} \text{s}^{-1}$. Six tests were made instead of four because of the increased estimation difficulty that comes with higher background light. Four reference measurements were taken where the lamp intensities were set at their maximum (see Figure (4.35)) but since the noise level increases with increasing lamplight intensity, four references were also measured without any background light (Figure (4.36)). These four final measurements were taken with 80 second intervals instead of 5 minutes. Table (4.3) shows the temperature for all reference measurements. The mean temperature during the sunlight tests were $23, 4^\circ$.

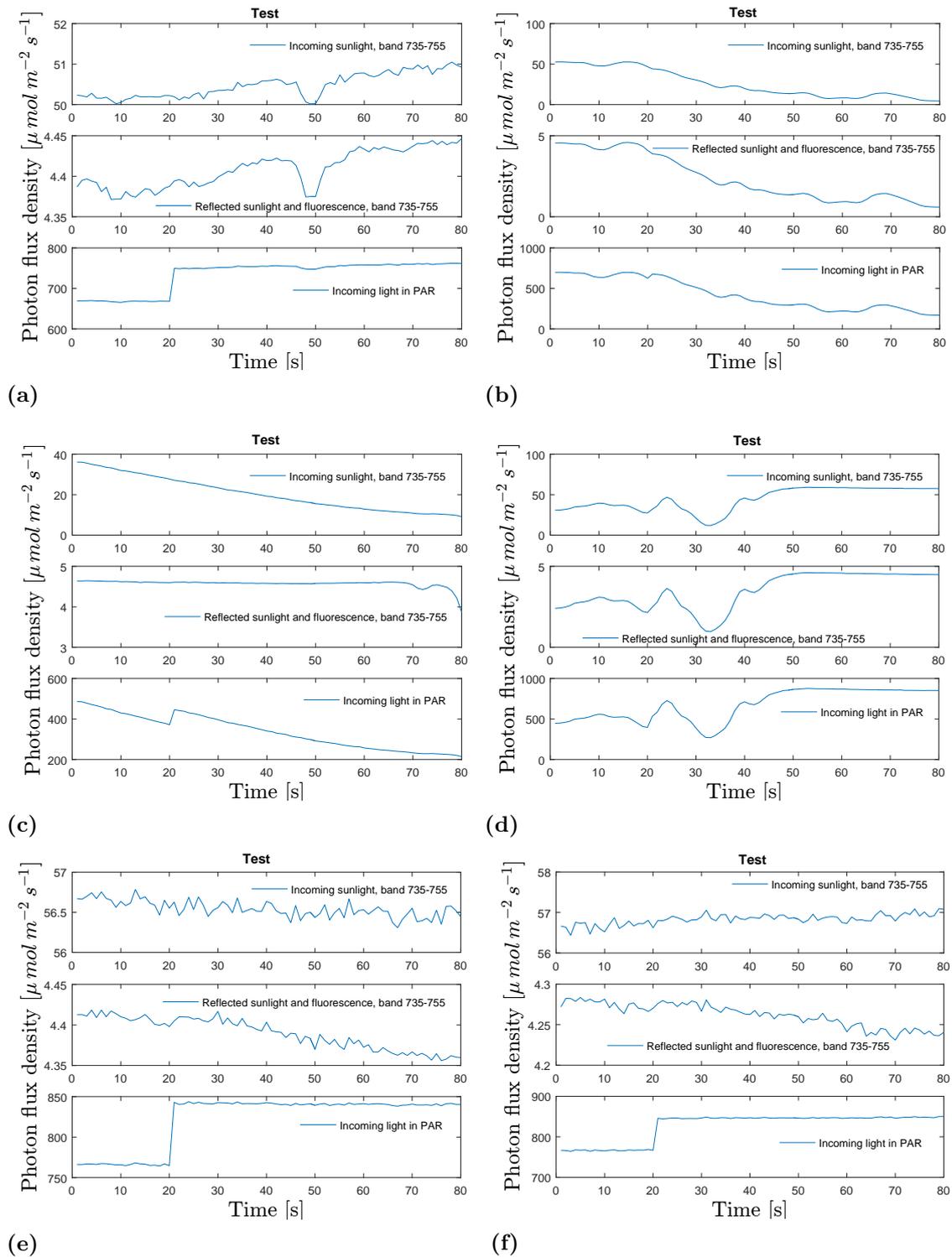


Figure 4.34: Collection of the six test measurements in Experiment 3. The fluorescence response is hardly distinguishable in the raw data for any of the collected data sets. There were five minutes between the start of each data collection, starting with (a).

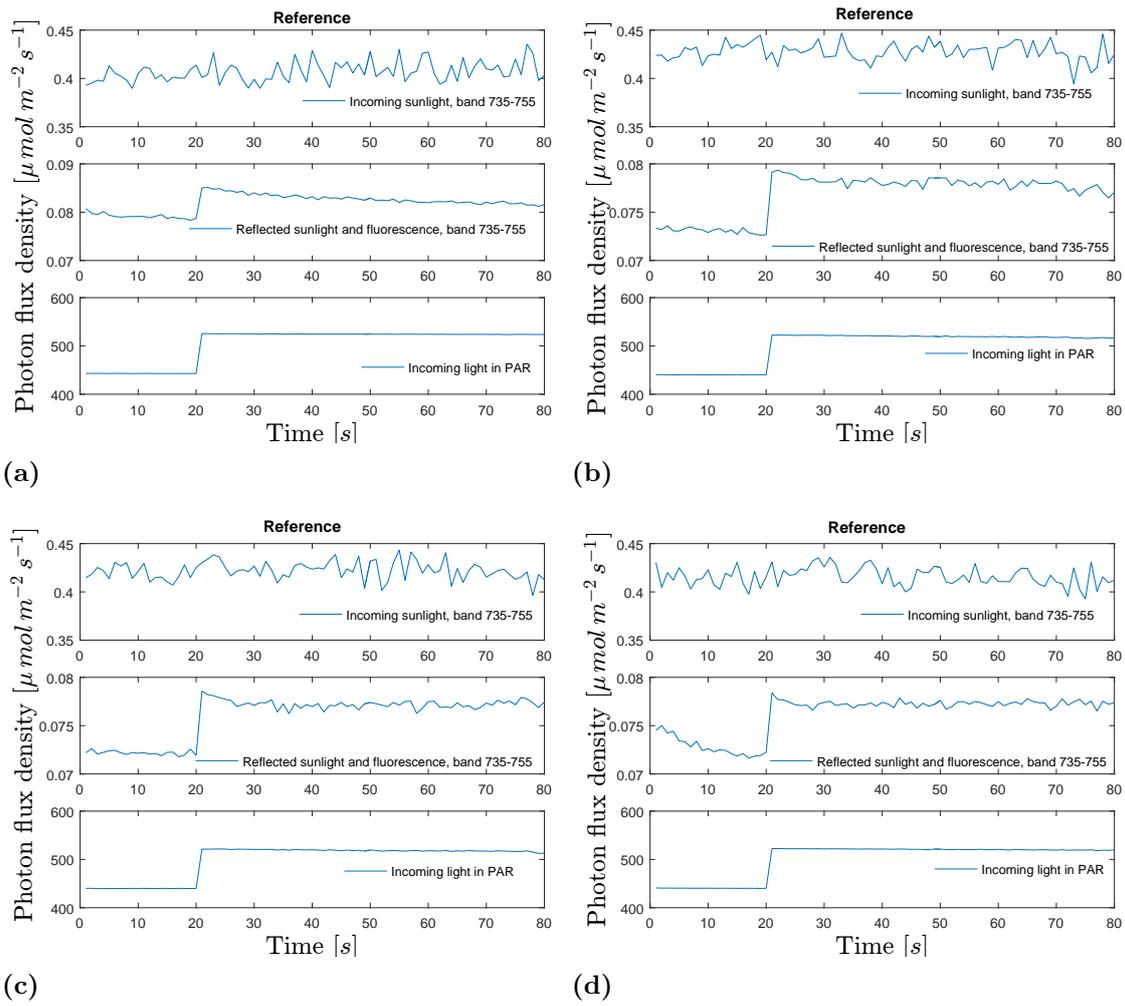


Figure 4.35: The four reference measurements in Experiment 3 taken directly after the tarpaulin had been pulled down with five minutes between each measurement starting with (a).

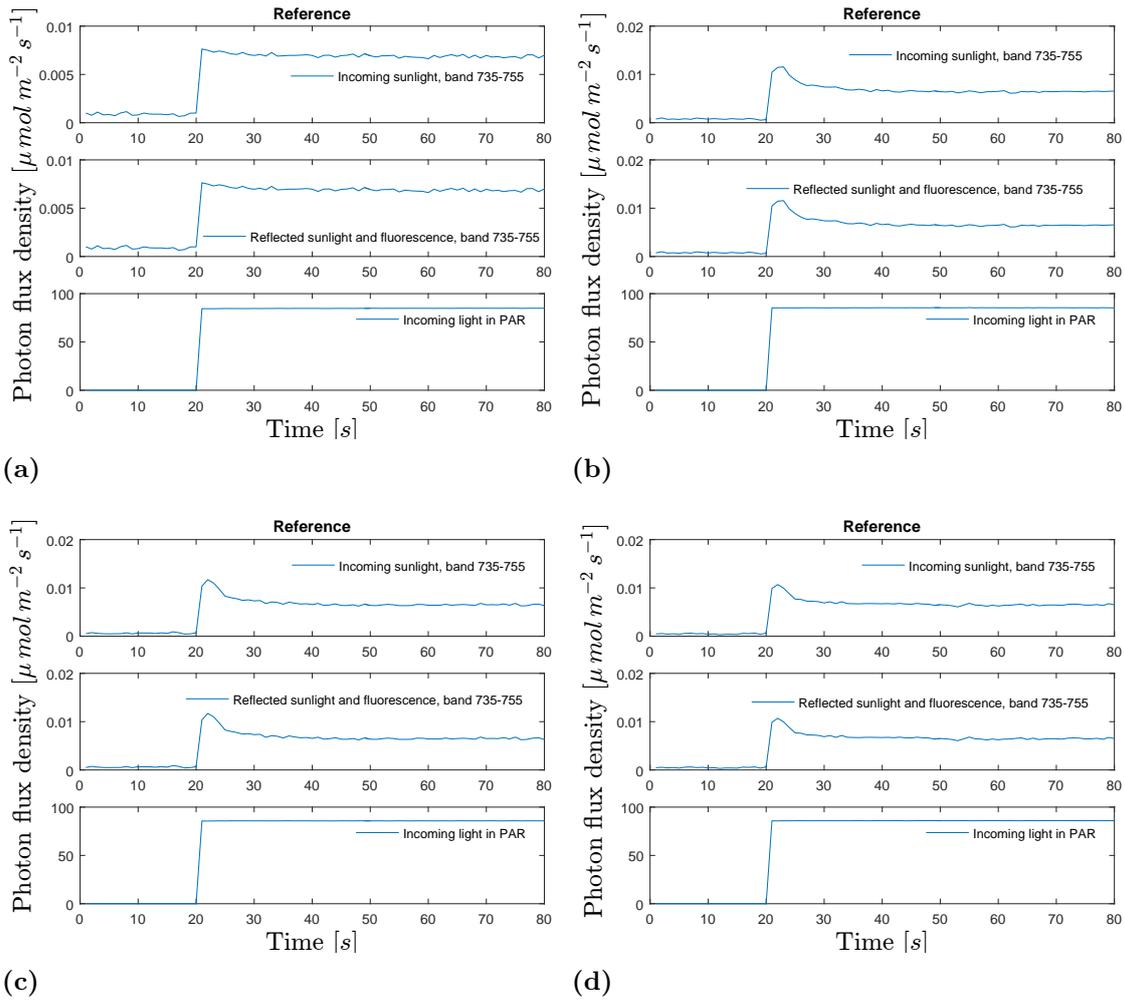


Figure 4.36: The final four reference measurements where the plant had been acclimated to total darkness. These final measurements were made approximately an hour after the sunlight tests. The data collections are made directly after one another, i.e. 80s between each start.

Table 4.3: Temperatures for Experiment 3 with background light.

<i>Reference</i>	<i>Temperature</i>
(a)	24,7°
(b)	29,9°
(c)	33,3°
(d)	39,3°

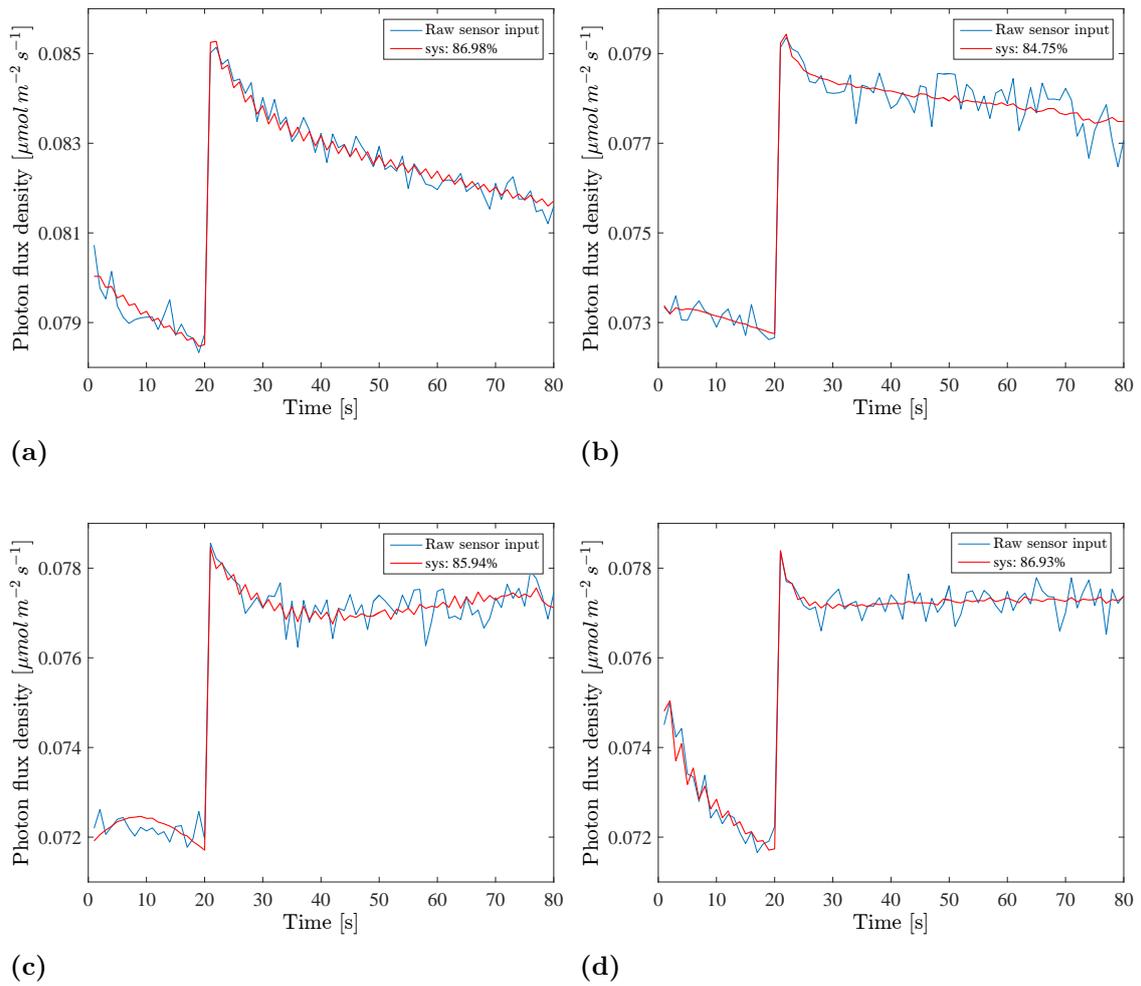


Figure 4.37: The first four step responses and the corresponding OE430 responses from the reference measurements. In order from first to last.

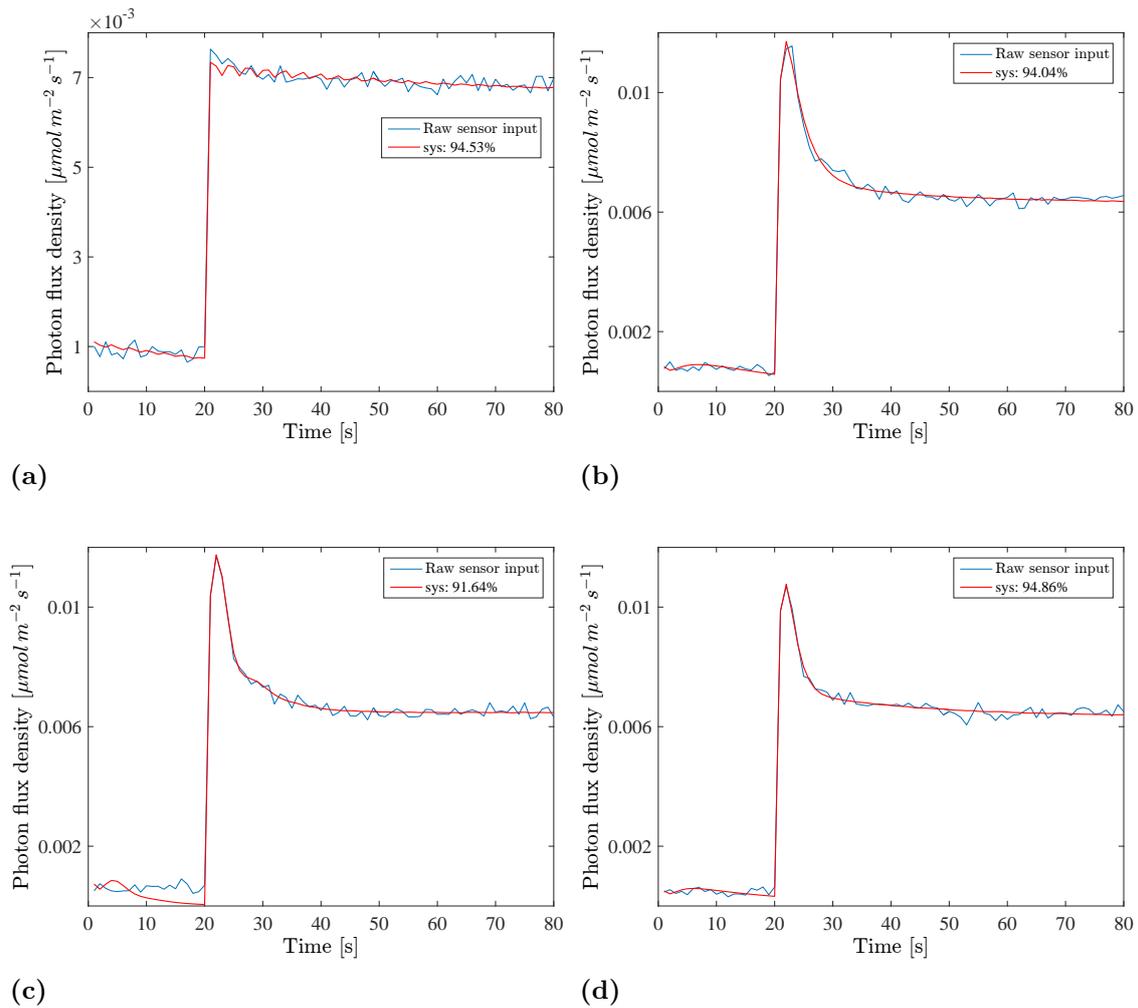


Figure 4.38: The last four step responses and their corresponding OE430 responses from the reference measurements. In order from first to last.

OE430 models were fitted to all eight fluorescence reference datasets and the results can be seen in Figure (4.37) and (4.38). The first plots are for the data with background light and the latter ones without. System (d) in Figure (4.38) will be used for upcoming validation. Note that the second set of measurements was done without background lights. Although this is the case, the S and M peak have merged together in the first set as well as in the second, making their dynamics relatively alike. This yields the second set of reference measurements a decent substitute for the one with background lights. Temperature data is missing for the Experiment 3 references without background light. However the temperature is most likely above 40° on average.

Curve fitting

Figure (4.39) shows the estimated results for both time and frequency domain using curve fitting.

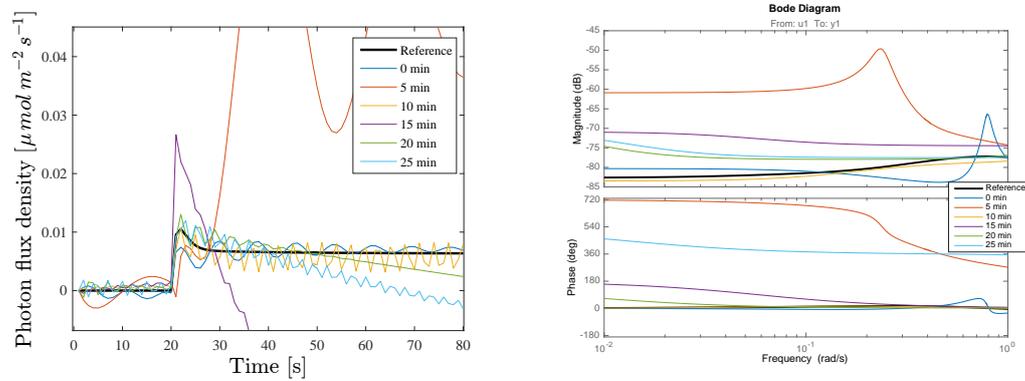


Figure 4.39: Results from the Curve fitting estimations to the data from Experiment 3.

Prefiltering algorithm

Figure (4.40) shows the estimated results for both time and frequency domain using the Prefiltering algorithm.

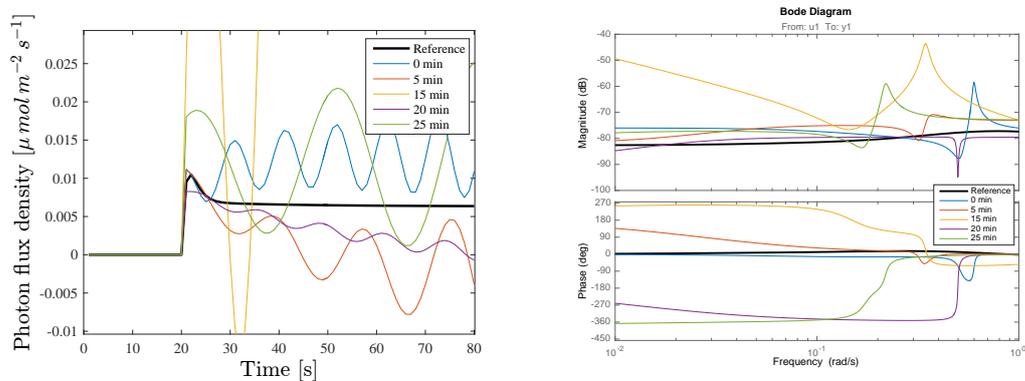


Figure 4.40: Results from the Prefiltering algorithm estimations to the data from Experiment 3.

dGN algorithm

Figure (4.41) shows the estimated results for both time and frequency domain using dGN.

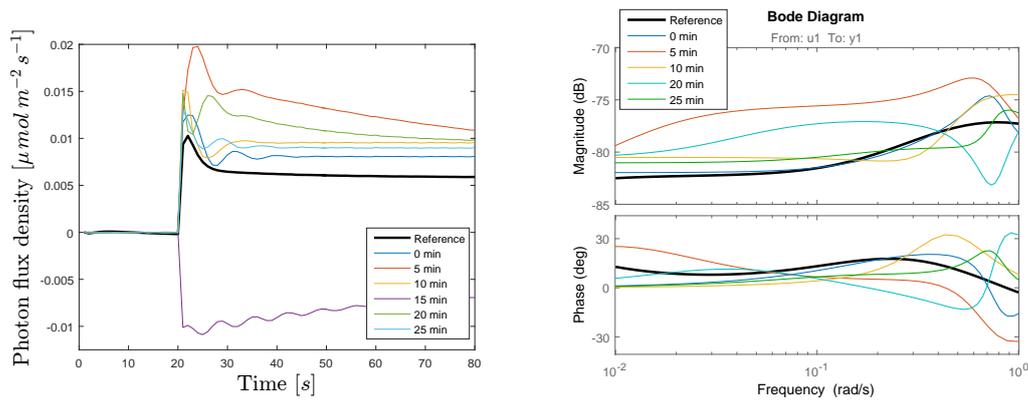


Figure 4.41: Results from the dGN algorithm estimations for the data in Experiment 3.

4.5 PSMT evaluation results

In the following results, the time until the M-peak can be seen for each experiment and method used. Note that the plant stress level is assumed to be roughly the same for each marking of the same colour. As an example, the plants stress level in Experiment 3 is roughly the same despite large variations in the PFD in Figures (4.43) and (4.44). The number of data for each experiment is 4 test data for greenhouse Experiment 1 and 2 (described in Section 3.9.1), and 6 test data for greenhouse Experiment 3 (described in Section 3.9.2). Note that the absence of data points is due to the inability of the method to find a system from which an M-peak could be determined.

4.5.1 Curve fitting

For curve fitting, there are only 3 datasets for Experiment 1 and 2, and 0 datasets for Experiment 3 (see Figure (4.42)). The reasons are the curve fitting results in Sections 4.4.1 and 4.4.2, where an M-peak could only be established in a few estimation cases.

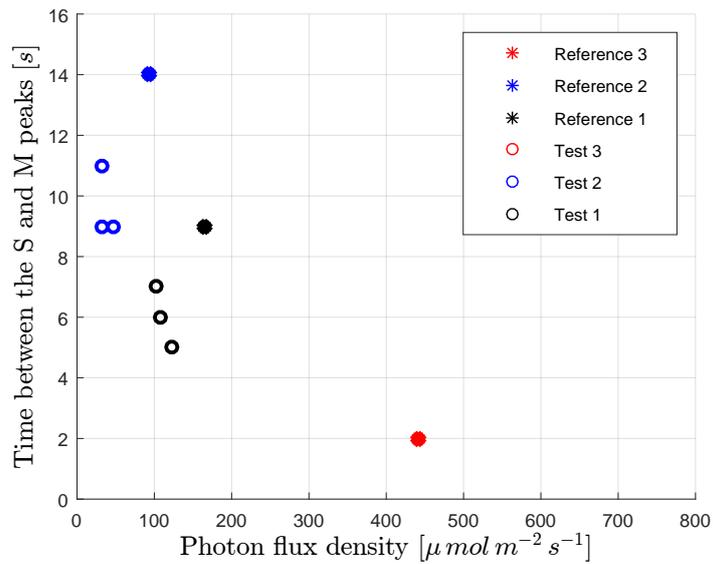


Figure 4.42: The PSMT evaluation for the Curve fitting method. For the data with lower sunlight intensity there was a much higher chance of observing a PSMT shape of the estimation.

4.5.2 Prefiltering algorithm

The Prefiltering algorithm could find the M-peak for 4 datasets for Experiment 1 and 2, and 5 datasets for Experiment 3 (see Figure (4.43)).

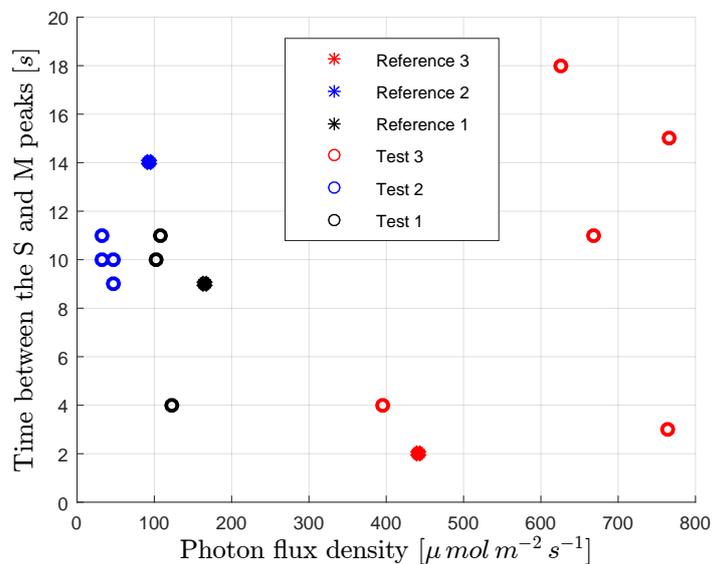


Figure 4.43: The PSMT evaluation for the Prefiltering algorithm. The result for test 3 where the background sunlight intensity was strong caused bad results with the Prefiltering algorithm

4.5.3 dGN algorithm

The dGN algorithm could distinguish an M-peak in 4 datasets for Experiment 1 and 2, and 5 datasets for Experiment 3 (see Figure (4.44)).

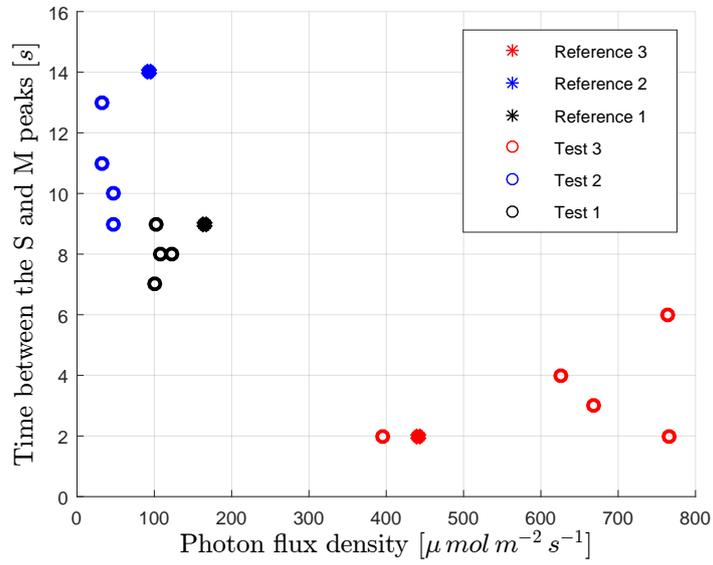


Figure 4.44: The PSMT evaluation for the dGN algorithm. This algorithm performed by far the best results.

5

Discussion

BECAUSE OF the project's many research elements, the results are sometimes hard to interpret. In this chapter we will try to explain the results and some phenomena observed to the best of our abilities.

5.1 Simulation data analysis

Some of the data that gave poor estimation results went through a major shift somewhere along the dataset. It could for example be when the data went from being varying to constant. One reason for this could be cloud interference, where the cloud could contribute to sudden change in the dynamics of the data. This makes ρ more varying but it also seems like the $q(t)$ solution can not handle this large difference in ρ during a dataset. What is strange is that this seems to mostly occur when the data goes from dynamic to static, not the other way around.

Another problem is when a shadow is cast on the area where the sensors are facing, or on the sensor directly. For instance if, for a few samples, one sensor is in the sun while the other is covered completely, or partly, by shadows, the estimation of parameters could potentially yield corrupt data.

The error used in linear curve fitting is the error that represent the methods flaw the most, though the error does not seem to correlate with the fit in any way. This is probably due to the methods sensitivity to pure noise where the measurement noise is overwhelming. The $q(t)$ error used for parametric modelling estimation seem to be more correlated, since using $q(t)$ for the band 735-755 nm yields a close to perfect estimation, making the substitute $q(t)$ closer to the actual was expected to give better estimation.

An interesting relationship between the wavelength ratios and the incoming sunlight was observed. The ratio seems to be dependent on the sunlight intensity. This connection could potentially prove useful during, for instance Curve fitting, since it provides us with an extra piece of information. See Appendix B for a plot from the experimental data collected on 08/03 - 2017 that illustrates this connection.

5.2 Experimental data analysis

There is a disturbance in the middle left plot in Figure (4.34), where only one of the sensors caught a shadow, although this corrupted the data, the response could be estimated.

The same phenomenon as in the simulation data occurs in the middle right plot in Figure (4.34), where the data went from being dynamic into being static. This resulted in a very poor estimation for all methods.

5.3 Greenhouse setup - sources of error

Problems with the experimental setup are factors that differentiate the test signal from the reference signal. One big problem is the matching of light intensity, where it is hard to reproduce the same light intensity in the reference as in the realistic test. For example, in Experiment 2, where the sunlight intensity was as low as $30 \mu\text{mol m}^{-2} \text{s}^{-1}$ in PAR, the lamps could not reach that low light intensity without going into PWM-mode, affecting the plants and making the reference signal differ from the test. On the other end of the spectrum, in Experiment 3 the sunlight intensity was too strong for the lamps to match. The plants will in these cases adapt to the changes in light, where a bigger difference leads to longer adaptation time. Even if a matching intensity between reference and test is obtained, the difference in electromagnetic spectrum distribution could potentially also add to the adaptation time making the reference and test signals vary (see Figure (3.15)).

The fluorescence response is only known from a constant background light with a step in the blue light. With dynamic sunlight interference, the receptivity of the plants fluorescence could vary - making the magnitude of the responses in time domain differ. This may explain the differences between test and reference in Figures (4.27), (4.33) and (4.41).

Since the reference measurements were taken inside a black tarpaulin, the temperature around the plants during the reference measurements was higher compared to the tests. This changes the stress level of the plant and makes the reference conditions different from the test signal. It could also prolong the adaptation time for the plants.

5.4 Quota

It has been confirmed that the reflectance ρ is not constant over time. Things like cloud interference and O_2 -absorption rates makes the signals themselves vary but do not necessarily cause a varying ρ . What was found is that the reflection ρ most likely deviates because of more palpable factors. Wind drafts cause movements of

the canopy and thus the areas and angles at which the sensor is facing. The greenhouse, however, is windproof but the fans from the lamps caused slight perturbation of the canopy. Although this could be the case for the experiments, the simulations never utilised a lamp. Another cause could be varying transmission rates when light transmits through the plants as well as varying absorption rates, which would directly affect the reflectance. These rates depend mostly on incoming light intensity but also on the stress of the plant, which should not change too much over a period of shorter time. What can be concluded by comparing Figure (4.7) and (4.2) is that ρ does not vary in a linear relation to sunlight intensity.

Another reason for the difference in reflectance over time could be that the spectrometers have different calibrations. Each spectrometer should be calibrated regularly and the calibration files used at the time of the thesis were quite old. The different spectrometers also use different calibration files, making them different from one another regarding offset and gain, possibly yielding a varying reflectance.

If the reflectance would be specular, ρ would possibly be constant. Since the used surfaces are all rough though, diffuse reflection probably occurs where the light reflects in a more stochastic way - possibly triggering the varying reflectance.

The sunlight intensity could vary in between the sensors iteration times. For example, if one sensor is using 15 ms integration time and the other is using 100 ms, they are both open for 15 ms while only one of them is open 85 ms longer. If the sunlight intensity is not constant during these 85 ms, the reflectance should vary because of this. In Experiment 2, from Section 3.9.1, the same integration time for both sensors was used. Even though this still yielded a varying reflectance the problem might still lie in the sensors ability to provide simultaneous data.

5.5 Different methods

Out of all three algorithms used; Curve fitting, parametric modelling with Prefiltering and with dGN, the latter generally gives the best estimations in all datasets. For low as well as middle light intensities, the Prefiltering algorithm produces good estimation and in some cases so did Curve fitting, but they are still outperformed by dGN. The dGN method is the default method used in the MATLAB System Identification Toolbox²⁹. It is preferable because of its efficiency, especially near the minimum, although it has a drawback in that it is sensitive to bad initial estimates. This is, however, compensated for somewhat with the damping of the step size. Many different ways of obtaining a better initial guess than the one obtained using linear least squares were tried, but no general method could be suggested.

As for the simulations, the dGN algorithm appeared to find good estimations up to around $800 \mu mol m^{-2} s^{-1}$ in PAR, after which the intensity got too high, compare Figure (4.17) and (4.21). The Prefiltering algorithm started to produce poor estimates earlier at around $400-600 \mu mol m^{-2} s^{-1}$ in PAR but sometimes even as low as $200 \mu mol m^{-2} s^{-1}$ in PAR, compare Figure (4.17) and (4.21). Curve fitting

estimated poorly throughout most middle to high sunlight intensities but also sometimes as low as $30 \mu mol m^{-2} s^{-1}$ in PAR, compare Figure (4.13) and (4.17).

As for the experiments, the dGN algorithm produced decent estimations up to around $800 \mu mol m^{-2} s^{-1}$ PAR (see Figure (4.44)) - where all the red dots are clearly separated from the others and are fairly close to the reference. The prefiltering algorithm started to give poor estimations around $80-150 \mu mol m^{-2} s^{-1}$ PAR (see Figure (4.43)). Curve fitting started to give poor estimations somewhere between 150 and $500 \mu mol m^{-2} s^{-1}$ PAR (see Figure (4.42)).

From the greenhouse experiment data, some estimations had some ripples after the M-peak. This could be an effect of the Gibbs phenomenon³⁰, although it should not occur if both input and output have been filtered the same way. Therefore, the ripples seem to be an effect of the estimation itself, where the optimal minimum was not found. The ripples seem to have the same frequency as the M-peak, making the bode plots increase in phase but not shift in frequency.

The Curve fitting itself could be studied more, for example by manipulating the spectrum seen in Figure (2.1) to better match the 735-755 nm band. Although improvements could be made, the parametric modelling felt more promising to develop further. The difference between both parametric modelling methods were the number of tuning parameters as well as the incorporation of the estimation error vector. Prefiltering only had one tuning parameter, the number of iterations, while dGN also included stepsize and direction. Prefiltering did not have any direct feedback of the previously estimated error, where it was incorporated both in the Jacobian as well as the numerator in Equation (2.21) for dGN.

5.6 Differences between simulation and experiment

The simulations were set up in a way to reflect a real experiment as accurately as possible, where the only difference from the real test is the emitted fluorescence signal. The fluorescence signal used in simulation had a two step approach, where the first step was to estimate a model from lamp light and its corresponding fluorescence signal. The next step was to simulate the fluorescence induced by the sun. This was made by using the previously estimated model and the raw sunlight data as input to that system, then the resulting output is the simulated fluorescence signal. In the simulation, it is assumed that the model of the plant that is estimated from only lamp light and lamp induced fluorescence would be the same model as if being estimated from the sun and lamp. There can, however, be problems with how the fluorescence signal behaves after a response from a more dynamical light source, such as the sun. In the simulation case, the model was estimated with a linear input signal, where the derivative was roughly around zero except for the step itself. The input for the model in the greenhouse experiment had a more varying derivative that was both negative and positive. Hysteresis in the plant system probably plays

a role in the creation of strange results where different signs of the derivative causes the system to behave differently. Something that was not calculated for in the simulations.

Another difference between simulation and experiment is the wavelengths that can be used for calculating $q(t)$, which should be free from fluorescence. In the simulated data, the 800-820 nm span could be used without any problems since there were no ChlF present. In the experiment, a span of 800-820 was also used and it was assumed to not contain any fluorescence. What is not considered is that with increased light intensity, the fluorescence band gets a small "tail" into the adjacent wavelengths. This "tail" is always there but too small at low light to have any significant impact. Therefore fluorescence could exist in the 800-820 band for higher light intensities resulting in a poor estimation where fluorescence is cut out. Since the estimation with $q(t)$ gets better the closer it is to 735-755 nm (see Figure (3.7)), a more adaptive wavelength span could be used that take incoming light intensity into account.

5.7 Difference between sunlight intensities

From Figure (4.13), 4.17 and 4.21, it can be seen that the error in $q(t)$ decreases with increasing sunlight intensity. This is probably due to the sensor noise being uncorrelated with the sunlight intensity, yielding a high Signal to Noise Ratio (SNR).

Although the error between $q(t)$ from the 735-755nm band and the 800-820 nm band is reduced with increased sunlight intensity, the estimation gets worse because the fluorescence signals magnitude decreases relative to the amount of reflected light.

5.8 Future work

An experiment to compare the step induced fluorescence response with static and dynamic background light would be a way to learn how the ChlF response behaves during the different excitations. A dark room, where the lamp excites a step as in Figure (3.12) as well as a sinusoidal step could be used for this. Parametric models based on the input and output would then be estimated and compared between the two cases. Note that for this to work, the canopy would have to stay in roughly the same stress level for the period of time it takes for the test to be done.

The next step could be to do the greenhouse experiments with a better reference method. Instead of using one canopy and one tarpaulin to black it out, two set of canopies could be used where one is constantly used as a reference and one is used as test. The reference would have to be in a dark environment with matched background light, temperatures and plant stress with the test.

We also believe that if the fluorescence-less $q(t)$ for 735-755 nm band could be found, the dGN will find a perfect estimation of the system. Therefore, more work could be put in to find the correct $q(t)$. A start would be to collect incoming and reflected light from a flat surface, where more specular reflection occurs, and see if ρ still varies. Trying this in a controlled environment with a static lamp light as well as a dynamic lamp light would be most beneficial. If it does not yield a more constant ρ , it could be the sensors that is causing the problem. As mentioned earlier, the importance of simultaneous sensor measurements might play a role in acquiring more correlated data in terms of reflectance. To make the sensors more synchronised with the ability to have a higher sample rate could be of great use.

A higher sampling rate would also give more data points for the parametric modelling estimation, making the global minimum easier to find. It could also give more opportunities to find better filters, where the current cutoff-frequency of 1 *rad/s* is dangerously close to the sampling rate being used.

The current PSMT method seems like a decent classification method, although a lot of research has to go into creating reference data for a large variety of plants. With the advancements in deep learning a neural network that has been trained on a lot of ChlF data could potentially be used instead as a classification method, in the future. This method requires, however, a vast data pool to make the network reliable in its training.

To improve upon the dGN a combination can be made with so called subspace methods. This could help in eliminating some of the local minima that might be caught when using only the dGN³¹. One could also try different model structures. For instance, the Nonlinear Auto Regressive Moving Average with eXogenous input (NARMAX) might have some unforeseen properties that the OE model structure lacks.

A possible way around a lot of the problems regarding sensor noise and uneven calibrations is to only use one sensor. This would mean that the incoming light would have to be estimated from the sole sensor measuring the reflected light and ChlF-response from the plant. Issues with this technique is that the estimation of the incoming light would have to be fairly accurate to make up for the difference between M1 and M2. An issue which, after this project, has been deemed very hard to compensate for. Since light is behaving so volatile and inconsistently it is plausible to believe that the benefits of using only one sensor does not outweigh the problems it creates.

6

Conclusion

THE FUTURE holds many challenges. As the earth's population is increasing, it puts a large strain on the food industry to keep equal pace while still maintaining high quality of the products they sell. As more crops need to be grown, more surface area is needed and the growth cycle have to be as short as possible. Artificial light using LED-lamps and vertical farming is an efficient way of both decreasing the space needed and to keep the growth of plants independent of good weather or hour of the day. To be able to control the intensity of light in order to maximise growth, minimise spill and to minimise the effects on the environment is the next step in the evolution of greenhouse farming. By using techniques to model the plant system and thus estimating their ChlF response the results of this project show promise.

Visual inspection of the results both in time- and frequency domain indicate the capabilities of parametric modelling and dGN to solve the minimising argument (Equation (2.19)). The results reflects the difficulties of dealing with highly non-linear biological systems as well as volatile and noisy light signals, as evident by most results of the less extensive prefiltering method and the linear curve fitting. Despite this, it is clear that with proper identification methods a ChlF response to a lamp-induced light intensity variation can be extracted from plants, even with sunlight present. This further strengthens the notion that remote sensing of plant health can be reliable to the extent that it can be used as feedback to control artificial light for growth in greenhouses, in a not so distant future. A way to further enhance the chance of localising the correct minimum in the numerical search for the parameters is by finding ways of better estimating the initial guess or to complete the dGN with methods to make the search itself better. The answer to the research question posed in the start of the paper is simply, yes. It is clear that the fluorescence can be estimated by the use of parametric modelling if only the correct algorithm is used and the collected data stays within reasonable standards.

As living standards are increasing rapidly in a lot of developing countries. the demand on leafy greens, like herbs, will rise. Therefore, the need for energy efficient solutions is on top of the wish list for many companies and producers, not only because of the potential positive economical aspect but also because the awareness of the need to preserve the environment.

Bibliography

- [1] UN-report. World must sustainably produce 70 per cent more food by mid-century. *United Nations News Centre*, December, 2013.
- [2] N Murata, S Takahashi, Y Nishiyama, and S I. Allakhverdiev. Photoinhibition of photosystem ii under environmental stress. *Biochimica et Biophysica Acta (BBA) - Bioenergetics Volume 1767, Issue 6, June 2007, Pages 414–421*, November, 6, 2006.
- [3] Heliospectra. Heliospectra homepage. <https://www.heliospectra.com>, Visited in January 2017.
- [4] Anna-Maria Carstensen, Torsten Wik, and Tessa Pocock. Remote light stress detection for greenhouse led lighting control. *Proceedings 19th Nordic Process Control Workshop*, pp. 20:1-6, March, 2015.
- [5] Johan Lindqvist. Remote detection of plant stress by analysis of the dynamic behaviour of chlorophyll a fluorescence response, EX030. Master's thesis, Chalmers University of Technology, Gothenburg, 2015.
- [6] Lukas Wikander. Remote Plant Stress Detection in Greenhouse Environments, EX030. Master's thesis, Chalmers University of Technology, Gothenburg, 2016.
- [7] Joseph Lakowicz. Principles of fluorescence spectroscopy. *Kluwer Academics*, 1999.
- [8] Ida Fällström and Grazyna Bochenek. The effect of diurnal light intensity distribution on plant productivity in a controlled environment. *Heliospectra AB*, February, 2016.
- [9] K Maxwell and G N Johnson. Chlorophyll fluorescence – a practical guide. *Journal of Experimental Botany Vol. 51, No. 345, pp. 659-668*, April, 2000.
- [10] Ismael Moya and Zoran Cerovic. Remote sensing of chlorophyll fluorescence: Instrumentation and analysis. *Advances in Photosynthesis and Respiration, volume 19, pp 429-445*, 2004.
- [11] A Ounis, J Bach, A Mahjoub, F Daumard, I Moya, and Y Goulas. Combined use of lidar and hyperspectral measurements for remote sensing of fluorescence and vertical profile of canopies. *REVISTA DE TELEDETECCIÓN 45, Special Issue, 87-94*, 2016.

-
- [12] Tom Simonite. Self-driving cars' spinning-laser problem. <https://www.technologyreview.com/s/603885/autonomous-cars-lidar-sensors/>, Visited in June , March 20, 2017.
- [13] RJ Heyden NA Campbell, B Williamson. Biology exploring life. *Prentice Hall*, 2006.
- [14] K Miyamoto. Renewable biological systems for alternative sustainable energy production. *Food and Agriculture Organization of the United Nations*, 2009.
- [15] K Maxwell and G N Johnson. Chlorophyll fluorescence – a practical guide. *Journal of Experimental Botany Vol. 51, No. 345, pp. 659-668*, April 2000.
- [16] John Brand. Lines of light: The sources of dispersive spectroscopy. *Gordon and Breach publishers*, 1995.
- [17] MathWorks. System identification overview. <https://se.mathworks.com/help/ident/gs/about-system-identification.html>, Visited in May 2017.
- [18] Karel J. Keesman. *System Identification: An Introduction*. Springer, Wageningen, Netherlands, 2011.
- [19] Lennart Ljung. Prediction error estimation methods. *Report no.: LiTH-ISY-R-2365, Department of Electrical Engineering, Linköping University*, October 3, 2001.
- [20] Karel J. Keesman. *System Identification: An Introduction, Algorithm 6.5, pp. 131*. Springer, Wageningen, Netherlands, 2011.
- [21] Lennart Ljung and Torkel Glad. System identification: Theory for the user, second edition. *Prentice Hall*, 1999.
- [22] Lennart Ljung and Torkel Glad. System identification: Theory for the user, second edition, methods, chapter 7 and 11. *Prentice Hall*, 1999.
- [23] Lennart Ljung and Rajiv Singh. Compare model output and measured output. <https://se.mathworks.com/help/ident/ref/compare.html>, Visited in November 1993.
- [24] Sigurd Skogestad and Ian Postlewaite. Multivariable feedback control. *John Wiley Sons*, 2005.
- [25] Anna-Maria Carstensen, Tessa Pocock, Daniel Bänkestad, and Torsten Wik. Remote detection of light tolerance in basil through frequency and transient analysis of light induced fluorescence. *Computers and Electronics in Agriculture, Volume 127, Pages 289-301*, September, 2016.
- [26] Ladislav Nedbal. Negative feedback regulation is responsible for the non-linear modulation of photosynthetic activity in plants and cyanobacteria exposed to a dynamic light environment. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, October, 2003.

- [27] Johan Lindqvist, Daniel Bånkestad, Anna-Maria Carstensen, Björn Lundin, and Torsten Wik. Complexity of chlorophyll fluorescence dynamic response as an indicator of excessive light intensity. *Elsevier, Volume 49, Issue 16, 2016, Pages 392-397*, 2016.
- [28] HM Kalaji, G Schansker, and et al. RJ Ladle. Frequently asked questions about in vivo chlorophyll fluorescence: practical issues. *Photosynthesis Research, Vol. 122. pp. 121-158*, 2014.
- [29] MathWorks. System identification. <https://se.mathworks.com/help/ident/ref/systemidentification-app.html>, Visited in May 2017.
- [30] G Arfken. Gibbs phenomenon. §14.5 in *Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press, pp. 783-787*, 1985.
- [31] Bo Wahlberg, Magnus Jansson, Ted Matsko, and Mats A. Molander. Experiences from subspace system identification - comments from process industry users and researchers. *LNCIS, volume 364, pp 315-327*.

A

Appendix

Instrumental variables (IV) method

The φ used in many methods looks as follows

$$\varphi = \begin{bmatrix} y(n_f - 1) & \dots & y(0) & u(n_f - 1) & \dots & u(n_f - n_b) \\ y(n_f) & \dots & y(1) & u(n_f) & \dots & u(n_f - n_b + 1) \\ y(n_f + 1) & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ y(N - 1) & \dots & y(N - n_f) & u(N - 1) & \dots & u(N - n_b) \end{bmatrix} \quad (\text{A.1})$$

The IV method takes this one step further and uses a matrix Z which is a lot like φ but uses the estimated output instead of measured, and looks as follows

$$Z^{-1} \begin{bmatrix} \hat{y}(n_f - 1, \theta^{i-1}) & \dots & \hat{y}(0, \theta^{i-1}) & u(n_f - 1) & \dots & u(n_f - n_b) \\ \hat{y}(n_f, \theta^{i-1}) & \dots & \hat{y}(1, \theta^{i-1}) & u(n_f) & \dots & u(n_f - n_b + 1) \\ \hat{y}(n_f + 1, \theta^{i-1}) & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \hat{y}(N - 1, \theta^{i-1}) & \dots & \hat{y}(N - n_f, \theta^{i-1}) & u(N - 1) & \dots & u(N - n_b) \end{bmatrix} \quad (\text{A.2})$$

The initial parameter guess $\theta^{(0)}$ is approximated using linear least square. The updated θ is then calculated using the following least squares formula

$$\theta^{(i)} = [Z^{(i-1)T} \varphi]^{-1} Z^{(i-1)T} y \quad (\text{A.3})$$

For each iteration, a new θ is calculated and used in Z^{21} .

B

Appendix

The ratio gets lower with increasing sunlight intensity (see Figure (B.1)).

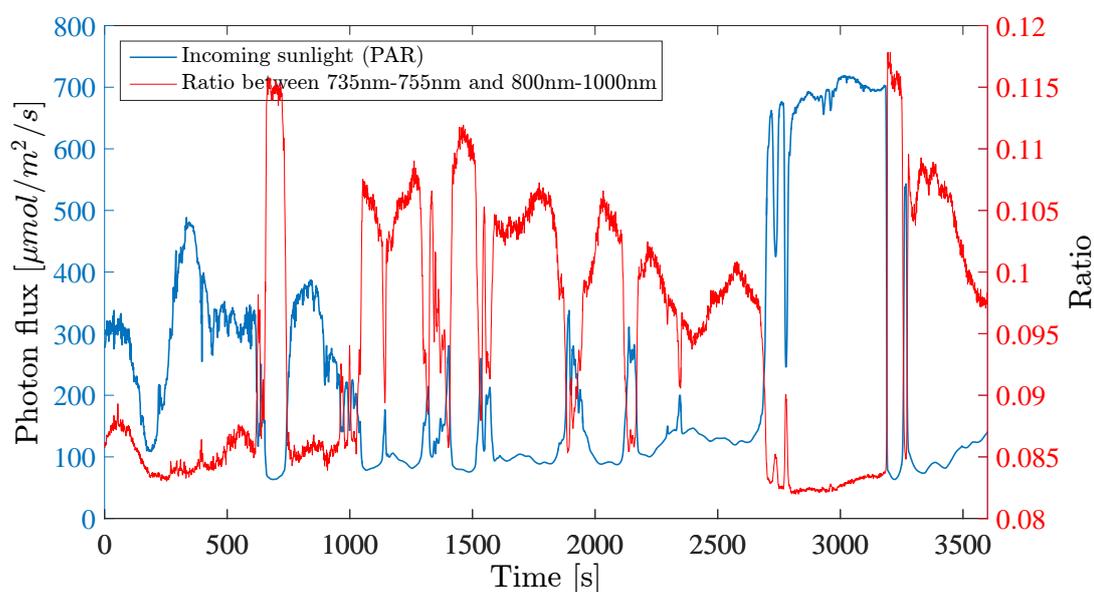


Figure B.1: The ratio is clearly depending on the sunlight intensity. It decreases when the sunlight gets strong and increases when the sunlight decreases.

This gives the impression that a connection between the two can be found which can help balancing the nonlinearity and thus making the ratio more constant.