



Norwegian University
of Life Sciences

Master's Thesis 2017 30 ECTS
Department of Mathematical Sciences and Technology

Methods for detection of powdery mildew in agricultural plants with hyperspectral imaging

Gudbrand O. Askvig
Mechanical Engineering and Product Development

“The master’s eye is the best fertilizer”

- Pliny the Elder -

PREFACE

Here at NMBU, previously Norwegian College of Agriculture, we pride ourselves in keeping sustainability close at heart at heart at all times. Norway may perhaps be one of the wealthiest countries in the world (per capita), giving the option to rely on import to supply the population with food. However, in my opinion, with great wealth comes a responsibility to develop knowledge and technology to ensure a sustainable future for the rest of the world, as well as for ourselves.

As I grew up on a farm, I have heard my father talk repeatedly of how it is important to know when to sow, when to fertilize, which pesticides to use and when to harvest. To know this, one must know the soil in the field, keep an eye on the weather and continuously inspect the field for pests of any kind. To me, this is the very essence of precision agriculture – increasing the ability to apply the right amount of effort at the right time.

I was excited when I found a subject for my thesis that allowed me to pursue my passion for agriculture and sustainability as well as to learn more about remote sensing and precision agriculture. Although I have had only the bare minimum of courses in programming and statistics, I thought that this was an excellent opportunity to learn more on these subjects.

I would like to give a great thanks to my supervisor, associate professor Ingunn Burud, for providing me with the fundamental codes for image processing, guiding me through the process of data analysis, and for being understanding and patient with my lack of previous knowledge on the subject. Also, thanks to my parents, who both helped nurture my fascination for science from a young age, and whose wisdom remind me to always try to see the bigger picture. Thanks to my bro Jan Vegard for helping me get into the concept of coding, to Runa, for helping me get up in the morning, and to Erling, for figuratively giving me a kick in the behind when he figured I was slacking off.

Finally, thanks to all the people I had the pleasure of meeting during my years at NMBU. You all made my time as a student so much more interesting, fun and and rewarding and I feel I have grown a lot as a person.

Sign: _____

Gudbrand Askvig

Ås, IMT/NMBU 01.12.16

ABSTRACT

With the human population still on the rise, it is important to continue developing technology for agriculture that will allow us produce enough food for the entire population. An important part of this is to become able to perform early detection of pathogens, as this will both reduce crop losses and reduce necessary amount of pesticide.

This study was conducted to investigate methods involving hyperspectral image analysis for early detection and discrimination of powdery mildew on three greenhouse plants; cucumber, strawberry and tomato. Two separate experiments were conducted, one with a VNIR hyperspectral camera (400-1000 nm) and one with a SWIR hyperspectral camera (1000-2500 nm). For each plant type, multiple healthy and infected samples were scanned and analysed.

The study had difficulty achieving a conclusive detection of powdery mildew with previously developed spectral disease indices. Results might improve if more appropriate methods for pre-processing and data filtering are applied.

Principal component analysis (PCA) was performed in an attempt to identify informative spectral bands, and a PCA including a 1st derivative 2nd order polynomial Savitzky-Golay algorithm showed that wavelengths around 700-730 nm contributed the bulk of loadings, and thus >90% of the variation in an image with healthy and mildew-infected areas. This

Parallel to the analysis performed in MATLAB, the graphical hyperspectral software Scyven was used as a supporting tool, and some comparisons were made to evaluate its usefulness as a tool for high precision scientific experiments. Though this proved useful for a beginner in the field of hyperspectral image analysis, its limitations makes it unsuitable for analysing the large image files used in this study.

SAMMENDRAG

Verdens befolkning vokser stadig, og det er viktig at vi forsetter å utvikle ny landbruksteknologi for å kunne føre befolkningen i årene framover. Å ha evnen til å oppdage sykdommer og skadedyr på tidlig stadium vil være en viktig del av dette, da tidlig deteksjon både vil hindre tap av avling og redusere nødvendig mengde plantevernmidler.

Denne studien har hatt som formål å undersøke hvorvidt bruk av hyperspektral bildeanalyse kan brukes for tidlig deteksjon av meldugg på tre ulike drivhusvekster; agurk, jordbær og tomat. To ulike forsøk ble gjennomført, hvert forsøk med hver sin type hyperspektrale kamera. Det første forsøket tok i bruk et VNIR-kamera som tar bilder i 400-1000 nm, og det andre brukte et SWIR-kamera som tar bilder i 1000-2500 nm. For begge forsøk ble det tatt bilder av flere syke og friske planter for hver sort, som senere ble analysert i programvare.

Det viste seg vanskelig å oppnå konkluderende resultater ved bruk av tidligere utviklede indekser for deteksjon av plantesykdommer. Dette kunne muligens blitt forbedret ved bruk av bedre tilpasset dataprosessering.

Prinsipalkomponentanalyse ble utført i et forsøk på å identifisere informative bølgelengder. Analyse med en førstederivert Savitzky-Golay-algoritme av andre orden viste at bølgelengder rundt overgangen mellom rødt og infrarødt lys bidro til hoveddelen av vektningen til første prinsipalkomponent. Denne prinsipalkomponenten forklarte igjen over 90% av variansen i den nye datamatriksen.

Parallellt med analysen som ble foretatt i MATLAB, ble programmet Scyven brukt som støttende hjelpemiddel. Samtidig ble det gjort noen sammenligninger for å finne ut hvorvidt dette programmet kunne brukes til denne typen eksperiment. Det viste seg at selv om programmet fungerte som en amatørvennlig introduksjon til hyperspektral bildeanalyse, hadde det enkelte begrensninger som gjorde det lite egnet til denne studien.

Table of Contents

PREFACE	II
ABSTRACT	III
SAMMENDRAG	IV
List of figures	VII
List of tables	VIII
List of Abbreviations and acronyms	IX
1 INTRODUCTION	1
1.1 Background for the study	1
1.1.1 Motivation	1
1.1.2 The evolution of agriculture	1
1.1.3 Chemical free pest control	2
1.1.4 Precision Agriculture	2
1.2 Spectral imaging	3
1.2.1 Fields of application	3
1.2.2 Spectral imaging in agriculture	3
1.3 Research objectives	3
2 THEORY AND FORMULAS	4
2.1 Multi- and hyperspectral imaging	4
2.1.1 Methods of spectral imaging	5
2.1.2 Spectral data	6
2.2 Measuring plant health from spectral data	7
2.2.1 Plant physiology and effects of powdery mildew infections	7
2.2.2 Spectral vegetation indices	9
2.3 Statistics – Dimensionality reduction and PCA	10
2.4 Results from previous similar studies	12
3 EXPERIMENTS	14
3.1 Equipment	14
3.1.1 Plant Samples	14
3.1.2 Optics, lighting and rig	15
3.1.3 Software: MATLAB and PLS/MIA-toolbox	17
3.1.4 Software: Scyllarus and Scyven	17

3.2	Acquiring data	17
3.2.1	Calibrations	17
3.2.2	Taking the pictures	19
3.3	Data analysis	19
3.3.1	Early inspection with Scyven	19
3.3.2	Importing and extracting data into MATLAB	22
3.3.3	Analysis with SVI and SDIs in MATLAB	22
3.3.4	Identifying informative bands with PCA in MATLAB	22
4	Results and discussion	23
4.1	Early data from Scyven	23
4.1.1	Illuminant recovery	23
4.1.2	Illuminant recovery – section discussion	23
4.1.3	PCA and analysis of spectrum shapes	25
4.1.4	PCA and spectrum shapes – section discussion	26
4.1.5	Classification with Linear Spectral Unmixing and SVM	27
4.1.6	Classification – Section discussion	30
4.2	Analysis in MATLAB	30
4.2.1	Spectral index results	30
4.2.2	Spectral indexes – section discussion	32
4.2.3	PCA results	32
4.3	Closing reflections	33
5	Conclusion and further research	34
6	References	35
7	Appendixes	38
A -	HySpec_Analysis-program	I
B -	readenvihdr-function	II
C -	Scyven illuminant spectra	IV
D -	VNIR Samples Areas	IX
E -	PCA scores and loadings	XII

List of figures

Figure 2-1: Illustration of a Bayer filter	4
Figure 2-2: Optical scheme for an imaging spectrometer.....	5
Figure 2-3: Push broom-scanning	6
Figure 2-4: Hyperspectral datacube.....	6
Figure 2-5 a and b: Mildew infected grape leaves	7
Figure 2-6: Cellular composition of a leaf.	8
Figure 2-7 a and b: Light interaction and colonization strategies	8
Figure 2-8: Hyperspectral PCA.....	11
Figure 2-9: Dimension reduction.....	12
Figure 2-10: Spectral reflectance signatures for barley leaves	13
Figure 3-1: Early, intermediate and late stages of mildew infection	14
Figure 3-2: Camera rig	15
Figure 3-3: 3D-printed plastic mount	15
Figure 3-4: Experiment setup	16
Figure 3-5: Corrected and uncorrected spectralon irradiance.....	18
Figure 3-6: Cropped areas for analysis	20
Figure 3-7: Selected polygons for material classification.	21
Figure 4-1: PC1 over an area of the spectralon	24
Figure 4-2: Irradiance spectrum for an area of the spectralon.	24
Figure 4-3: Error message.....	24
Figure 4-4: Colour representation of scores in PC1.....	25
Figure 4-5: Colour representation of scores in PC2.....	25
Figure 4-6: Reflectance spectra for points of interest.....	26
Figure 4-7: Classification performed on an infected tomato leaf	27
Figure 4-8: Classification performed on an infected tomato leaf 2	27
Figure 4-9: Classification performed on an infected cucumber leaf.....	28
Figure 4-10: Classification performed on an infected cucumber leaf 2.....	28
Figure 4-11: Classification performed on an infected strawberry leaf.....	29
Figure 4-12: Classification performed on an infected strawberry leaf 2.....	29
Figure 4-13: MCARI2 and MRENDVI values for sample points on plant JI2.....	30

Figure 4-14: Image visualisation of MCARI2 and MRENDVI	31
Figure 4-15: Image visualisation of PMI and the difference between PMI and HI	32
Figure C-1: Illuminance spectra, NICTA-method, unfiltered.	IV
Figure C-2: Illuminance spectra, NICTA-method, CMA-filter.....	IV
Figure C-3: Illuminance spectra, NICTA-method, SavGol-filter.	V
Figure C-4: Illuminance spectra, Gray Edge-method, SavGol-filter.....	V
Figure C-5: Illuminance spectra, NICTA-method, SavGol-filter.	VI
Figure C-6: Illuminance spectra, Gray Edge-method, SavGol-filter.....	VI
Figure C-7: Illuminance spectra, Gray World-method, SavGol-filter.	VII
Figure C-8: Illuminance spectra, White Patch-method, SavGol-filter.....	VII
Figure C-9: Illuminance spectra, NICTA-method, SavGol(7)-filter.	VIII
Figure C-10: Illuminance spectra, Grey Edge-method, SavGol-filter.	VIII
Figure D-1: Locations of sample areas on cucumber leaves.	IX
Figure D-2: Locations of sample areas on strawberry leaves.....	X
Figure D-3: Locations of sample areas on tomato leaves.	XI
Figure E-1: Loading plot and scores for cucumber, healthy #1	XII
Figure E-2: Loading plot and scores for cucumber, infected #2	XIII
Figure E-3: Loading plot and scores for strawberry, healthy #1.....	XIV
Figure E-4: Loading plot and scores for strawberry, infected #2	XV
Figure E-5: Loading plot and scores for tomato, healthy #1	XVI
Figure E-6: Loading plot and scores for tomato, infected #2.....	XVII

List of tables

Table 2-1: Examples of spectral vegetation indices and their applications.	9
Table 3-1: Camera settings for the two experiments.....	18
Table 4-1: Healthy and infected areas and their scores on HI and PMI.....	31

List of Abbreviations and acronyms

Abbreviation	Full meaning
ANN	Artificial Neural Network
CMA	Central Moving Average
GMO	Genetically Modified Organism
HR	Hypersensitive Response
MCARI2	Modified Chlorophyll Absorption Ratio Index Improved
MRENDVI	Modified Red Edge Normalized Difference Vegetation Index
NIR	Near Infrared
NMBU	Norwegian University of Life Sciences (Norwegian: Norges Miljø- og Biovitenskapelige Universitet)
nm	Nanometers
PC#	Principle Component nr # (replace # with appropriate number)
PCA	Principle Component Analysis
RCV	Remotely Controlled Vehicle
RGB	Red/Green/Blue
SavGol	Savitzky-Golay
SDI	Spectral Disease Index
Sensel	Sensing Element
SKP	Centre for Plant Research in Controlled Climate (Norwegian: Senter for Klimaregulert Planteforskning)
SVI	Spectral Vegetation Index
SVM	Support Vector Machine
SWIR	Short Wave Infrared (radiation)
UAV	Unmanned Aerial Vehicle
UV	Ultra Violet (radiation)
VNIR	Visual and Near Infrared (radiation)

1 INTRODUCTION

1.1 Background for the study

This section will introduce the topics covered in this work. This will include a brief introduction to the evolution of agriculture in general and some of its challenges, as well as precision agriculture and the use of drones and remote sensing.

1.1.1 Motivation

“The greatest shortcoming of the human race is our inability to understand the exponential function.” This is a quite famous quote by A. A. Bartlett; professor emeritus of physics at the University of Colorado at Boulder and known for his talks on sustainable growth [1]. Already in the late 1700s, Thomas Malthus noticed that an increase in food production only led to an increase in population, meaning that there would always be a shortage of food for the lower classes of society [2]. In essence, both these statements claim that the growth of the human population must eventually come to a halt, because there will simply not be enough arable land available to grow crops.

In theory, one could increase the amount of arable land, but this often comes at the cost of removing forest areas. This is often considered undesirable, as forests play a vital part in the binding of CO₂. Furthermore, Norway is a cold, weathered rock with only 2.2% arable land area [3], making it even more important to cultivate the available soil in the best way possible.

1.1.2 The evolution of agriculture

The proverb “necessity is the mother of invention” opposes the somewhat pessimistic views of Bartlett and Malthus, mentioned in the previous section. Agriculture has seen a couple of major technological revolutions, including (but not limited to) invention of the plough, implementation of mechanized tools and the ability to create genetically modified organisms (GMOs). Humankind has developed ways to increase crop yield by breeding crops with higher yield per acre, developing bigger and more time efficient farming machines and inventing artificial fertilizer and pesticides. However, many of these come with a cost. Bigger machines mean more soil compaction. Intensive cultivation leads to soil depletion and excessive use of chemicals have a wide variety of consequences.

1.1.3 Chemical free pest control

Pest control has always been a major part of growing crops. Some early recordings include descriptions of the Sumerians' use of sulphur compounds to control insects and mites, as well as the Roman ritually where they burned straw, fish guts, dung or animal horns to battle mildew and blight [4]. Unfortunately, many pest have a nasty habit of developing immunities to certain types of pesticide, and many chemicals used for pest control can be harmful to humans or the environment surrounding the field. Therefore, a chemical free form of pest control would have many benefits.

Of course, chemical free pest control is just as ancient as agriculture itself, simple examples being cats to counter mice, and manually weed out plants that undesired in the field. However, as the fields have grown in size, many such methods have been impractical for some time. Recently, advances in machine vision and automated robots could once again make mechanical weeding a valid alternative for some types of crops. For example, the RoboCrop InRow Weeder from Garford Farm Machinery Ltd [5] uses video image analysis to locate both crop and weed in real-time, and a robot arm uproots the weed while leaving the crops unharmed.

For other pests, such as fungi, mechanical methods are impractical, and researchers are searching for new solutions. A patent by A. Michaloski [6], as well as several newer studies, shows that exposure to UV-radiation has a suppressing effect on different kinds of powdery mildew [7] [8] [9]. To suppress the fungi at an early stage, and thereby prevent sporulation and spread, one could either continuously expose the crops to UV-radiation, or apply a detector system to determine where and when to apply UV.

1.1.4 Precision Agriculture

The goal of precision agriculture is to optimize returns while preserving resources. As this implies reduced use of i.e. fertilizer and pesticide, it would also mean a reduction of the negative environmental impact from industrialized agriculture. GPS-assisted machinery can improve driving patterns, reducing both fuel consumption and soil compression. Automated lightweight robots using machine vision can perform weeding or harvesting, and tractor manufacturers like Case are developing autonomous full size tractors that could handle heavier tasks, like ploughing [10]. As UAVs (commonly referred to as "drones"), sensors and computing power becomes cheaper and more available, it enables the farmer to gather more information about his crops. He will then have a better chance of making good decisions with respect to where and when to apply agricultural input for maximum effect [11].

1.2 Spectral imaging

Spectral imaging as well as multi- and hyperspectral sensors will be explained in more detail in chapter 2, and this section will mainly focus on presenting existing applications and potential future uses, as well as how it will be applied in this study.

1.2.1 Fields of application

Ever since Joseph von Fraunhofer discovered black absorption lines in the solar spectrum, physicists and chemists has used spectroscopy to both classify stars and identify different elements or substances. Spectral imaging combines spectroscopy and photography and grants the ability to analyse the spectrum of each pixel within a single image. Each pixel can then be classified based on the shape and/or intensity of its spectrum. Being a powerful tool for classification, spectral imaging is already widely used in a number of fields, including astronomy, geology, military applications, medicine, food quality control and agriculture [12].

1.2.2 Spectral imaging in agriculture

An example is the Yara N-Sensor, which measures the chlorophyll levels in crops in real-time using a multispectral sensor. By automatically and strategically varying the amount of fertilizer spread, the use of N-sensor has reportedly increased the yield in cereal crops by 3%, while reducing the amount of applied fertilizer by 13% [13]. The WeedSeeker from AgriOptics New Zealand LTD use a multispectral sensor to detect weeds and apply herbicide only where needed. AgriOptics proclaim that the WeedSeeker can reduce the costs of weed-control chemicals by an impressive 90% [14].

A key element in reducing the necessary amount of pesticide is early detection and classification of pathogens in the crops. Several studies has found that hyperspectral imaging shows great promise in the detection of diseases like mildew in a variety of crops, examples being maize [15], cucumber [16], grapes [17] and date fruit [18].

1.3 Research objectives

The overarching goal of this thesis is to investigate the possibility of using spectral data to early detect and identify powdery mildew on cucumber-, strawberry- and tomato plants. Relevant previously developed methods found will be tested, and PCA will be used to identify informative wavelengths in the images taken. Primary tools will be MATLAB, PLS_Toolbox and Scyven. The study will not go into subjects like multivariate image regression, artificial neural nets, support vector machines or other forms of machine learning.

2 THEORY AND FORMULAS

This chapter will address the theory and technologies necessary to perform this study. The first section will explain the concepts of spectral imaging, technologies and methods for acquiring these images and the nature of spectral image data. Section 2.2 will give a brief introduction to how powdery mildew affects the infected plants, how plants defend themselves from such attacks, and how this could affect the optical properties of the plant. Section 2.3 will introduce the statistical tools used to process said data, and section 2.4 will cover some results from similar experiments.

2.1 Multi- and hyperspectral imaging

As light of different wavelengths hits an object, it is absorbed, reflected, transmitted and scattered in certain ways determined by the chemical composition of the object. The rate of reflection is what determines what colour we perceive. A red apple reflects mostly red light, while absorbing light of the other visible wavelengths. The human eye has three types of sensory cells called rods that are sensitive red, green and blue light. As each cell is stimulated, it sends a signal to our brain, which process the information into an image, aka, what we “see”. Quite similarly, consumer grade 3-channel RGB-cameras regularly uses a Bayer filter, which uses an array of red-, green- and blue filtered sensing elements, or sensels. The light intensity at each sensel is combined with the intensities from surrounding sensels to create a vector with three dimensions. Each dimension represents one type of filter (or channel), namely red, green and blue for RGB-cameras. Hence, for each pixel in an image is a 3D-vector that tells our computer the RGB colour values in that pixel.

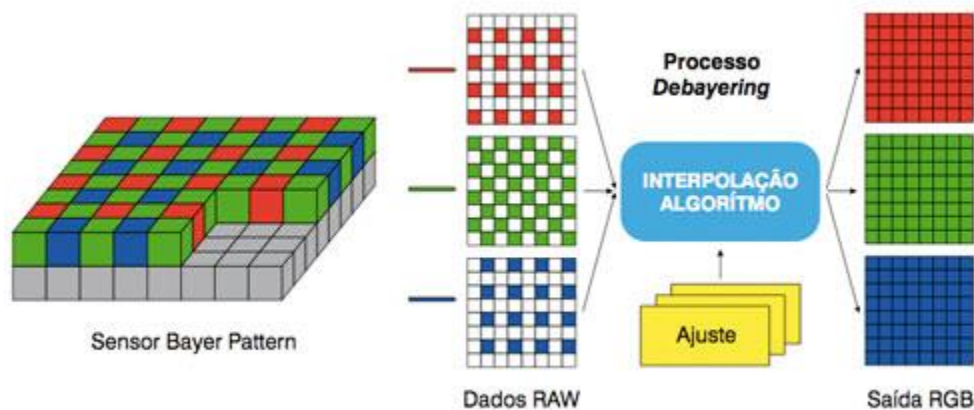


Figure 2-1: Illustration of a Bayer filter and the interpolation process to acquire RGB color values for every pixel. Found at http://www.colorizemedia.com/detalhe_tecnologia.php?pag=53

With multi- and hyperspectral imaging, it is possible to detect variations in light reflectance outside the visible spectrum. As mentioned in chapter 1, this can be a powerful tool for determining the chemical composition of objects in the image.

2.1.1 Methods of spectral imaging

One way to create an image with wavelengths different from those observable to the human eye is to replace the RGB-filters with filters of other wavelengths. A disadvantage of this method is that for each type of filter added to the filter array, the spatial resolution of the image is reduced. This lack of precision implies that this method is suboptimal for research applications that require high precision.

There are several other methods for creating multispectral images, including multi-layered sensors, filter wheels, spectral scanners and interferometers. Although only the spectral scanner will be covered in this study, the data generated from each of these sensor types would be quite similar. This will be explained in section 2.1.2.

Figure 2-2 illustrates the optical scheme for a spectral scanner. The sensor captures a single line of pixels, and a spectrum is recorded for each pixel in the line. To get a full image, the sensor records continuously while either the camera or the object to be recorded moves is moved along an axis perpendicular to the scan line. This is often referred to as “push broom scanning” (illustrated in Figure 2-3), and this is the method that will be used in this study.

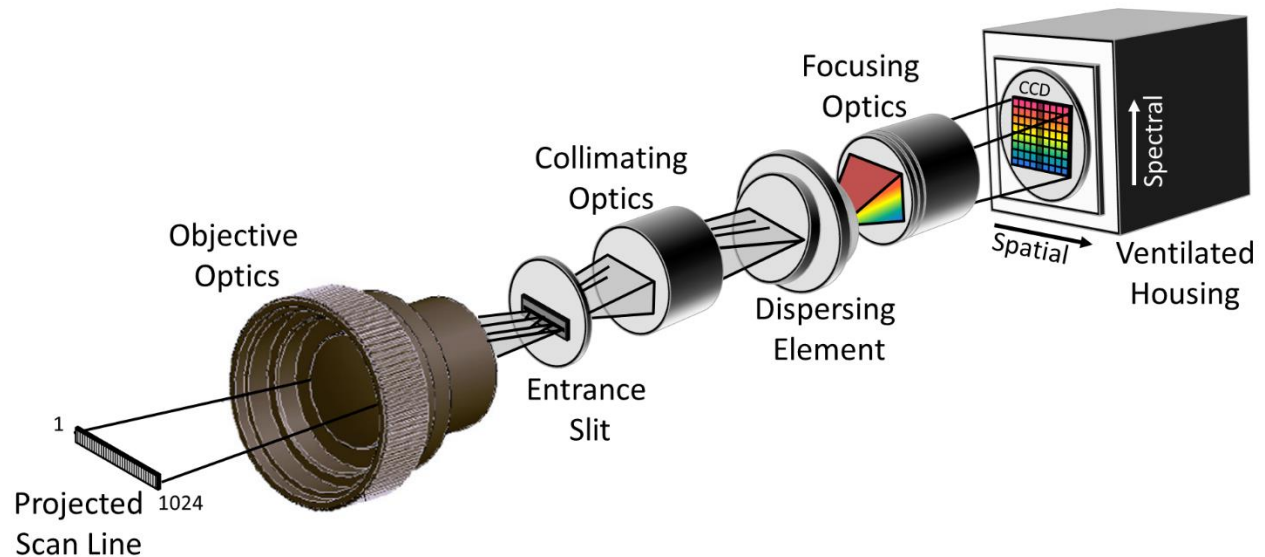


Figure 2-2: Optical scheme for an imaging spectrometer.
Found at: http://home.uni-leipzig.de/strahlen/web/general/de_index.php?goto=measurements

2.1.2 Spectral data

While an RGB image has three colour values per pixel, multi- and hyperspectral images can have many more. Similar to RGB cameras, for or each pixel there is an n-dimensional vector, where n is the number of spectral bands available to the sensor. Although there is no clear defining border between multi- and hyperspectral, a multispectral sensor typically have 3-10 relatively broad spectral bands, while a hyperspectral sensor can have hundreds of narrower bands [19] and thus higher spectral resolution. Figure 2-4 illustrates a hyperspectral dataset as a datacube, with two spatial and one spectral dimension. Each value in this 3D data matrix represents the reflection of a given wavelength in a given pixel.

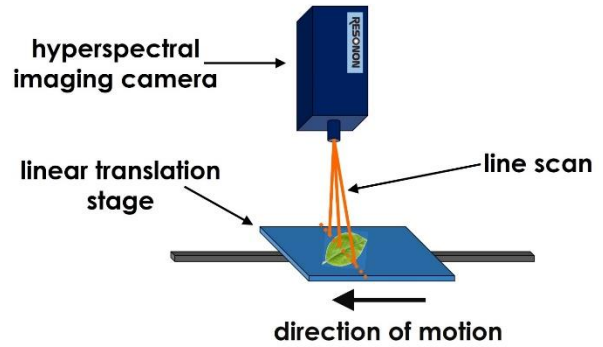


Figure 2-3: Hyperspectral imaging with the push broom principle. Found at: https://www.resonon.com/knowledge_base.html

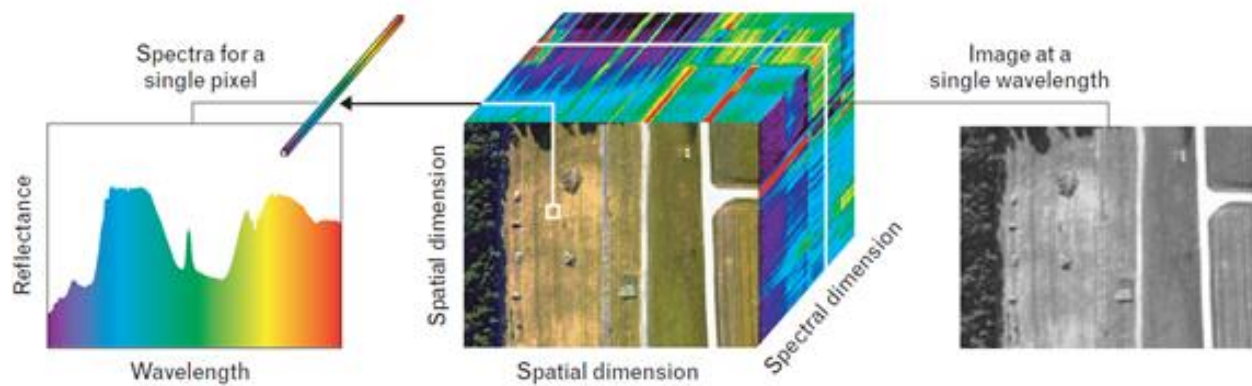


Figure 2-4: Example of a hyperspectral datacube with two spatial and one spectral dimension. Found at: http://www.tankonyvtar.hu/en/tartalom/tamop425/0032_terinformatika/ch04s04.html

Although the data is indeed recorded in a 3D matrix, the *order* of a hyperspectral image is not truly three dimensional, as pointed out by Burger in [20]. The reason for this is that the data has two spatial dimensions, meaning each data point in the matrix is given by location vs location vs wavelength. This means that a hyperspectral image can be unfolded to create a two dimensional dataset, with only one spatial dimension and the spectral dimension, which allows for second order analysis procedures [20]. This will be further explained in section 2.3.

2.2 Measuring plant health from spectral data

Various diseases have different symptoms, in plants as well as in animals. A basic understanding of plant physiology could therefore be beneficial in order to have some idea of what to expect from the reflectance spectra of sick plants vs healthy ones. Moreover,

2.2.1 Plant physiology and effects of powdery mildew infections

Powdery mildew is a collective term for plant diseases caused by a variety of fungi from the order Erysiphales [21]. Symptoms include white, powdery fungal growth on the leaf surfaces and stems, and sometimes the fruit. Figure 2-5a show a heavily infected grape leaf, while Figure 2-5b show the fungus on a microscopic level with round spores attached to long conidiophores.

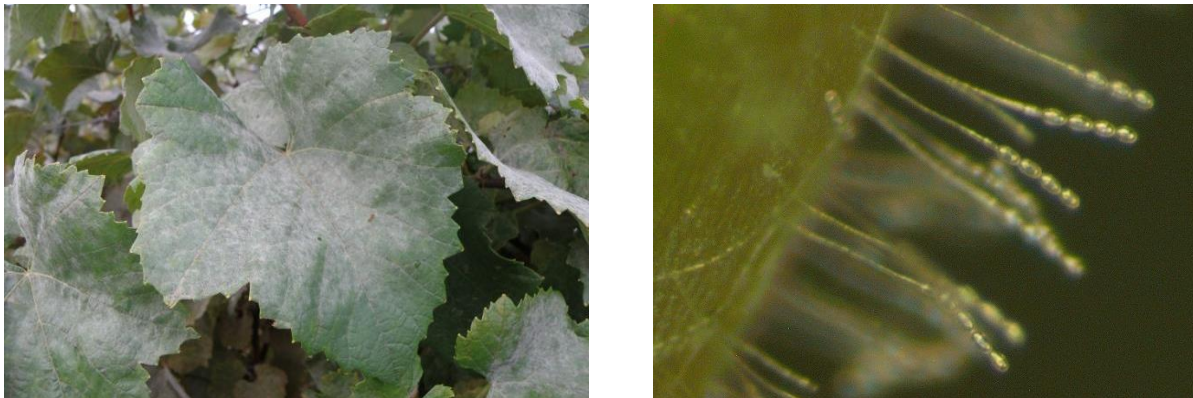


Figure 2-5 a and b: Mildew infected grape leaves, makro (left) and microscope view (right). Found at <https://www.agric.wa.gov.au/table-grapes/powdery-mildew-grapevines-western-australia>

Being a parasitic organism, the fungus grows on the surface of the leaves. Here, they penetrate the epidermal cells (the outermost cell layer of the leaf, see Figure 2-6) and extract water and nutrients from the host, which usually results in a slowed rate of photosynthesis in the host leaf [22] [23]. Put in simple terms, the plant is more concerned about defending itself than performing its day-to-day tasks of performing photosynthesis and gathering energy. In addition, the very presence of the fungi blocks out a portion of sunlight, and the “attack” also affects the ability of the stomata to perform gas exchange, both of which weakens the host leaf.

Most plants have several defence mechanisms to protect itself from pathogens. The first line of defence is usually the cell wall itself, forming a physical barrier as well as containing a variety of chemical defence mechanisms [24]. However, as some pathogens have evolved to either tear down or bypass these mechanisms, the plants can apply the scorched earth-tactics, or “hypersensitive response” (HR). As cells in the infected area commit suicide, the plant seeks to limit or deny pathogen access to water and nutrients, thereby limiting the pathogens ability to survive and reproduce. This local water deficiency could be a way to tell if a plant is sick, as water has two characteristic absorption bands around 1400 and 1900 nm [25], and the absorption rate for a sick plant could deviate from that of a healthy plant. This is especially true for necrotrophic pathogens, which outright kills the host cells to access nutrients.

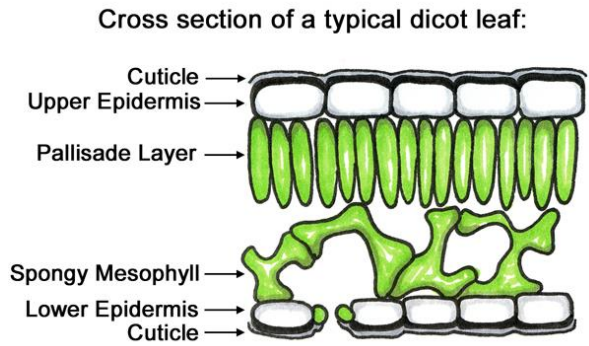


Figure 2-6: Illustration of the cellular composition of a leaf.

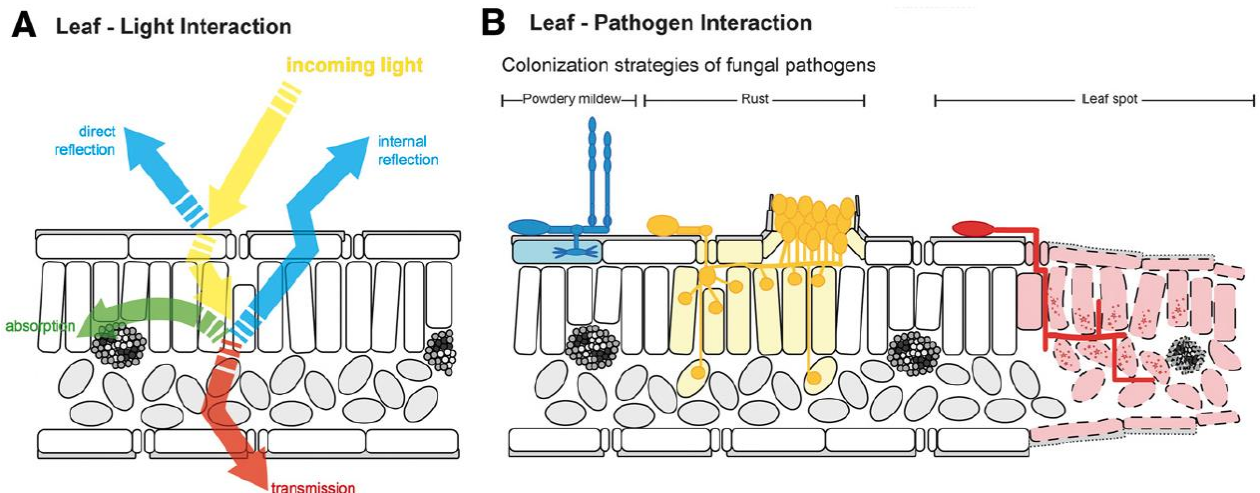


Figure 2-7 a and b: Light interaction with leaf tissue (a) and colonization strategies for fungal pathogens (b). Modified from (Mahlein 2016)

Mahlein (2016) [26] points out that powdery mildew infections has a “relatively low impact on tissue structure and chlorophyll composition during early infection”, as powdery mildew is a biotrophic fungus and does not trigger HR in the same degree as some other pathogens. However, it does produce fungal structures on the leaf surface, which can influence the optical properties in the affected areas.

2.2.2 Spectral vegetation indices

As mentioned in chapter 1, several studies has shown that spectral data has high potential for detecting diseases in crops. However, spectral data in its raw form is difficult to interpret. Spectral vegetation indices (SVIs) uses the relationships between the reflectance of a few selected wavelengths to express specific physiological parameters. This is usually presented as a value ranging from -1 to +1 or 0-1. A collection of SVIs can be found on [27], and Table 2-1 from [28] presents a table over which index is suitable for different types of applications.

Table 2-1: Examples of spectral vegetation indices and their applications.

Index	Equation	Application	References
Carter & Miller stress (CMS); Carter stress (CSc and CSd)	$\frac{R_{694}}{R_{760}}, \frac{R_{605}}{R_{760}}, \frac{R_{710}}{R_{760}}$	Chl content	Carter & Miller (1994); Carter (1994)
Datts index (DattsCI)	$\frac{R_{850}-R_{710}}{R_{850}-R_{680}}$	Chl content	Datt (1999)
Difference vegetation index (DVI)	$R_{810} - [(R_{610} + R_{661})/2]$	Canopy Chl density	Broge & Leblanc (2001)
Enhanced vegetation index (EVI) ¹	$\frac{G(R_{NIR}-R_{red})}{RNIR+C_1(R_{red})-C_2(R_{Blue})+L}$	Photosynthetic activity	Huete <i>et al.</i> (2002)
Gitelson & Merzlyak chlorophyll index (GM)	$\frac{R_{750}}{R_{700}}$	Chl content	Gitelson & Merzlyak (1994)
Green normalized difference vegetation index (GNDVI)	$\frac{R_{810}-[(R_{510}+R_{561})/2]}{R_{810}+[(R_{510}+R_{561})/2]}$	Chl content	Gitelson & Merzlyak (1997)
Modified red edge ratio (mSR)	$\frac{R_{750}-R_{445}}{R_{705}+R_{445}}$	Chl content	Sims & Gamon (2002)
Normalized difference vegetation index (NDVI)	$\frac{R_{800}-R_{680}}{R_{800}+R_{680}}$	Chl content and energy absorption	Gamon <i>et al.</i> (1997); Gamon & Surfus (1999)
Normalized difference water index (NDWI)	$\frac{R_{860}-R_{1240}}{R_{860}+R_{1240}}$	Liquid water content of vegetation canopies	Gao (1996)
Photochemical reflectance index (PRI)	$\frac{R_{531}-R_{570}}{R_{531}+R_{570}}$	Carotenoids; xanthophyll cycle	Gamon <i>et al.</i> (1992), (1997)
Pigment-specific normalized difference (PSND)	$\frac{R_{800}-R_{675}}{R_{800}+R_{675}}, \frac{R_{800}-R_{650}}{R_{800}+R_{650}}$	Chla and Chlb content	Blackburn (1998)
Pigment-specific simple ratio (PSSRa, PSSRb and PSSRc)	$\frac{R_{800}}{R_{675}}, \frac{R_{800}}{R_{650}}, \frac{R_{800}}{R_{470}}$	Chla and Chlb content; carotenoids	Blackburn (1998), (1999); Sims & Gamon (2002)
Plant senescence reflectance index (PSRI)	$\frac{R_{680}-R_{500}}{R_{750}}$	Carotenoids; Chl content	Merzlyak <i>et al.</i> (1999); Sims & Gamon (2002)
Ratio analysis of reflectance spectra (RARS _a , RARS _b and RARS _c)	$\frac{R_{675}}{R_{700}}, \frac{R_{675}}{R_{650}}, \frac{R_{760}}{R_{700}}, \frac{R_{760}}{R_{500}}$	Chla and Chlb content; carotenoids	Chappelle <i>et al.</i> (1992); Blackburn (1999)
Ratio vegetation index (RVI)	$\frac{R_{800}}{R_{680}}$	Chl content	Pearson & Miller (1972); Broge & Leblanc (2001); Sims & Gamon (2002)
Red edge NDVI (rNDVI)	$\frac{R_{750}-R_{705}}{R_{750}+R_{705}}$	Chl content	Gitelson & Merzlyak, (1994), Sims & Gamon (2002)
Red/Green Index (RGI)	$\frac{\sum_{i=600}^{700} R_i}{\sum_{i=500}^{600} R_i}$	Anthocyanin	Gamon & Surfus (1999); Sims & Gamon (2002)
Second soil-adjusted vegetation index (SAVI2) ²	$\frac{RNIR}{R_{red+(a/b)}}$	Canopy Chl density	Broge & Leblanc (2001)
Structure-insensitive pigment index (SIPI)	$\frac{R_{800}-R_{445}}{R_{800}-R_{680}}$	Carotenoids; Chl	Peñuelas <i>et al.</i> (1995); Sims & Gamon (2002)
Water index	$\frac{R_{900}}{R_{970}}$	Canopy water content	Peñuelas <i>et al.</i> (1997)

R, reflectance.

¹EVI variables: G, gain factor; L, canopy background adjustment factor; C_{1,2}, aerosol influence coefficients.

²a and b in SAVI2 are soil line coefficients (e.g. a = 1 and b = 0 for bare soil).

Wavelength numbers are in nanometers.

Two such indices, listed below, were used early in this study, namely the Modified Chlorophyll Absorption Ratio Index Improved (MCARI2) from [29] and the Modified Red Edge Normalized Difference Vegetation Index (MRENDVI) from [30] (featured as “modified Normalized Difference index, or mND_{index}”).

$$MCARI2 = \frac{1.5[2.5(R_{800} - R_{670}) - 1.3(R_{800} - R_{550})]}{\sqrt{(2 \cdot R_{800} + 1)^2 - (6 \cdot R_{800} - 5 \cdot \sqrt{R_{670}}) - 0.5}} \quad (2-1)$$

$$MRENDVI = \frac{R_{750} - R_{705}}{R_{750} + R_{705} - 2 \cdot R_{445}} \quad (2-2)$$

As in Table 2-1, R indicates relative reflection at the wavelength indicated by the subscripts. It should be mentioned that the primary function of both these indices is to estimate chlorophyll content. The previously section stated that early stage mildew infection has a low impact on chlorophyll levels, and therefore these two indices should be of little practical value to this study. However, that literature was not obtained before MCARI2- and MRENDVI- analysis was already complete, so they will still be mentioned in the results section.

2.3 Statistics – Dimensionality reduction and PCA

As mentioned, hyperspectral data can be quite overwhelming if not processed or reduced in some way. The aforementioned spectral indices provide one way of simplifying spectral data, but their use imply that it is known which spectral bands contain useful information. This section will cover the statistical tools used to obtain this information, as well as other methods to classify data.

For discriminating between two materials, the “useful information” is the portion of the data that has the greatest deviation between the two classes. In the case of this study, the useful information will be the spectral bands that has the greatest deviation in reflectance when comparing healthy and infected plants. In contrast, the wavelengths that has the same reflection for both healthy and infected plants are of no interest, as they do not provide us with information to separate the two plants. There idea of removing the non-informative variables in data is often referred to as dimensionality reduction.

PCA was developed as a tool to find the greatest variance in datasets by studying the eigenvectors of the covariance in the data matrix. For readers not familiar with the method, introductions are provided by multiple sources, e.g. [31] and [32]. In short, the data is rotated onto a new dimensional space where the principal components are the new dimensions. A set of data has equally many principal components as it has variables, and the first principal component (PC1) will be the vector that spans the greatest variance in the data. This is illustrated in Figure 2-9, where a group of individuals is classified based on three different genes. The same samples are just as effectively classified using only the first two principal

components. The third component is of little informational value, and can be ignored, thus reducing the dimensionality of the data. As hyperspectral images can have several hundreds of variables, in the form of spectral bands, and PCA can potentially reduce this to a handful of more useful principal components.

Burger explains the process in [20], but a short summary is presented here. As mentioned in section 2.1.2, when performing PCA on hyperspectral images, the image is first unfolded into a 2D-matrix X as illustrated in Figure 2-8. The following rotation of the data matrix onto the new dimensional space results in a bi-linear decomposition, also illustrated in Figure 2-8. The decomposed vectors, namely the loadings (column vectors P) and scores (row vectors T) are found in decreasing order of importance, so that the first loading vector, p_1 , is the loading vector that accounts for the most variance in the data structure, p_2 second most, etc.

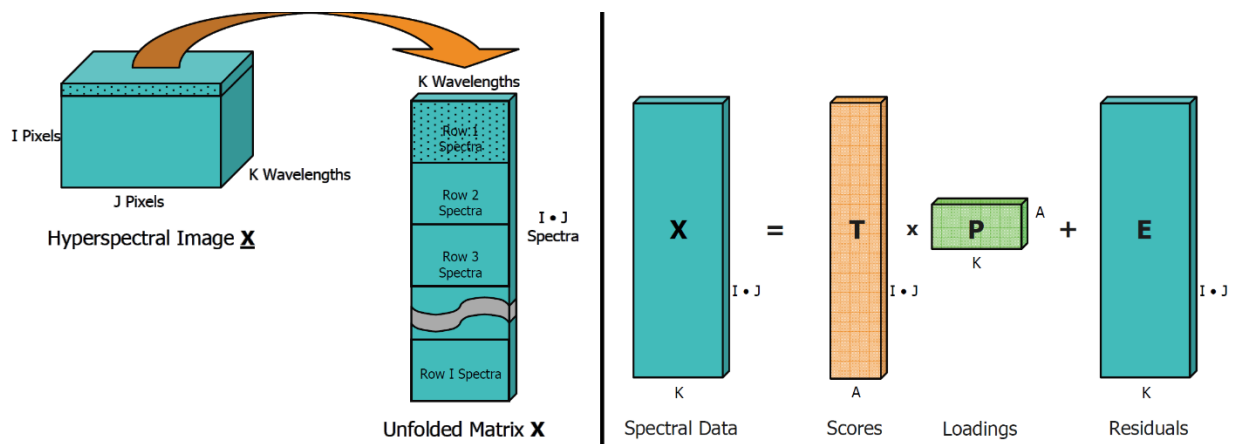


Figure 2-8: A hyperspectral image unfolded (left) and decomposed into scores, loadings and residuals (right). The loadings are the projections of the PCA factors onto the original variables, and the scores are projections of samples onto the PCA factors.

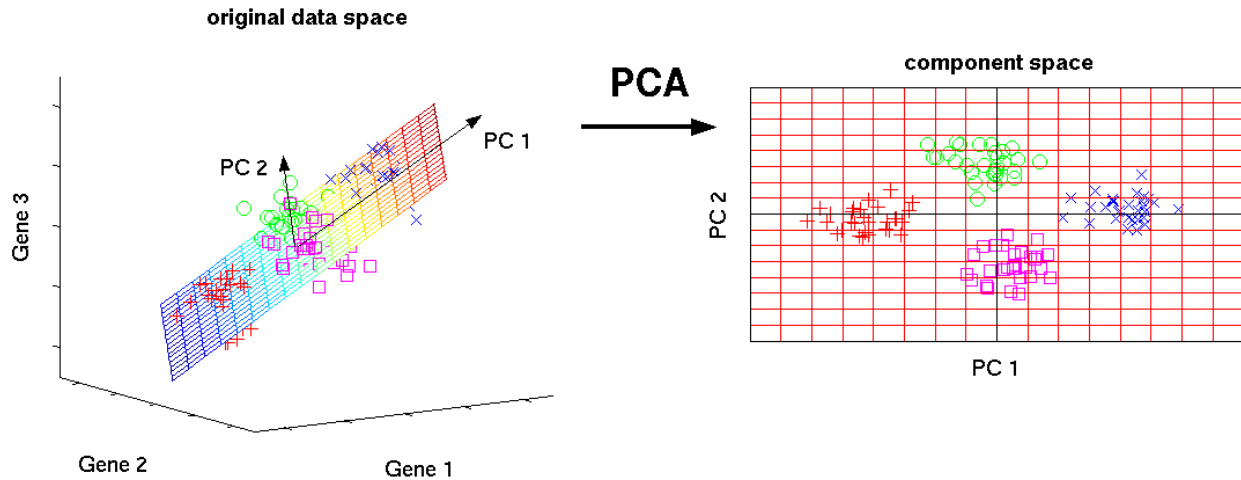


Figure 2-9: A three dimensional dataset can be reduced down to two principal components without severe loss of information. http://www.nlpc.org/pca_principal_component_analysis.html

2.4 Results from previous similar studies

In 2010, Rumpf used a collection of SVIs as parameters to not only determine if a plant is healthy or ill, but also discriminate between different diseases, which has not yet been proven possible using a single SVI [33]. Mahlein and Rumpf later developed specific spectral disease indices (SDIs) to detect *Cerospora* leaf spot, sugar beet rust and powdery mildew in sugar beets [34], shown in equations (2-3) to (2-6).

$$\text{Healthy - index(HI)}: \frac{R_{534} - R_{698}}{R_{534} + R_{698}} - \frac{1}{2} \cdot R_{704} \quad (2-3)$$

$$\text{Cerospora leaf spot - index(CLS)}: \frac{R_{698} - R_{570}}{R_{698} + R_{570}} - R_{734} \quad (2-4)$$

$$\text{Sugar beet rust - index(SBRI)}: \frac{R_{570} - R_{513}}{R_{570} + R_{513}} + \frac{1}{2} \cdot R_{704} \quad (2-5)$$

$$\text{Powdery mildew - index(PMI)}: \frac{R_{520} - R_{584}}{R_{520} + R_{584}} + R_{724} \quad (2-6)$$

Although disease detection at early stages of infection proved difficult due to only minor change in leaf reflectance, Mahlein and Rumpf claims that “SDIs will improve and simplify plant disease detection based on hyperspectral data.” Figure 2-10, taken from [26], illustrates variations in the reflectance spectra of barley leaves when infected with net blotch, rust and powdery mildew. The spectra demonstrate obvious differences, making it quite intuitive that it should be possible to do some discriminating analysis on it.

The HI and PMI indices will be used in this study to investigate if these indexes could also be applied to other plants, despite being developed by studying sugar beet leaves.

Further supporting the potential for discriminating diseases are studies done by Eady et.al. to classify different serotypes of *Salmonella enterica* [35] [36]. Here, informative spectral bands were identified using PCA, and these were used for classification by the use of SVM.

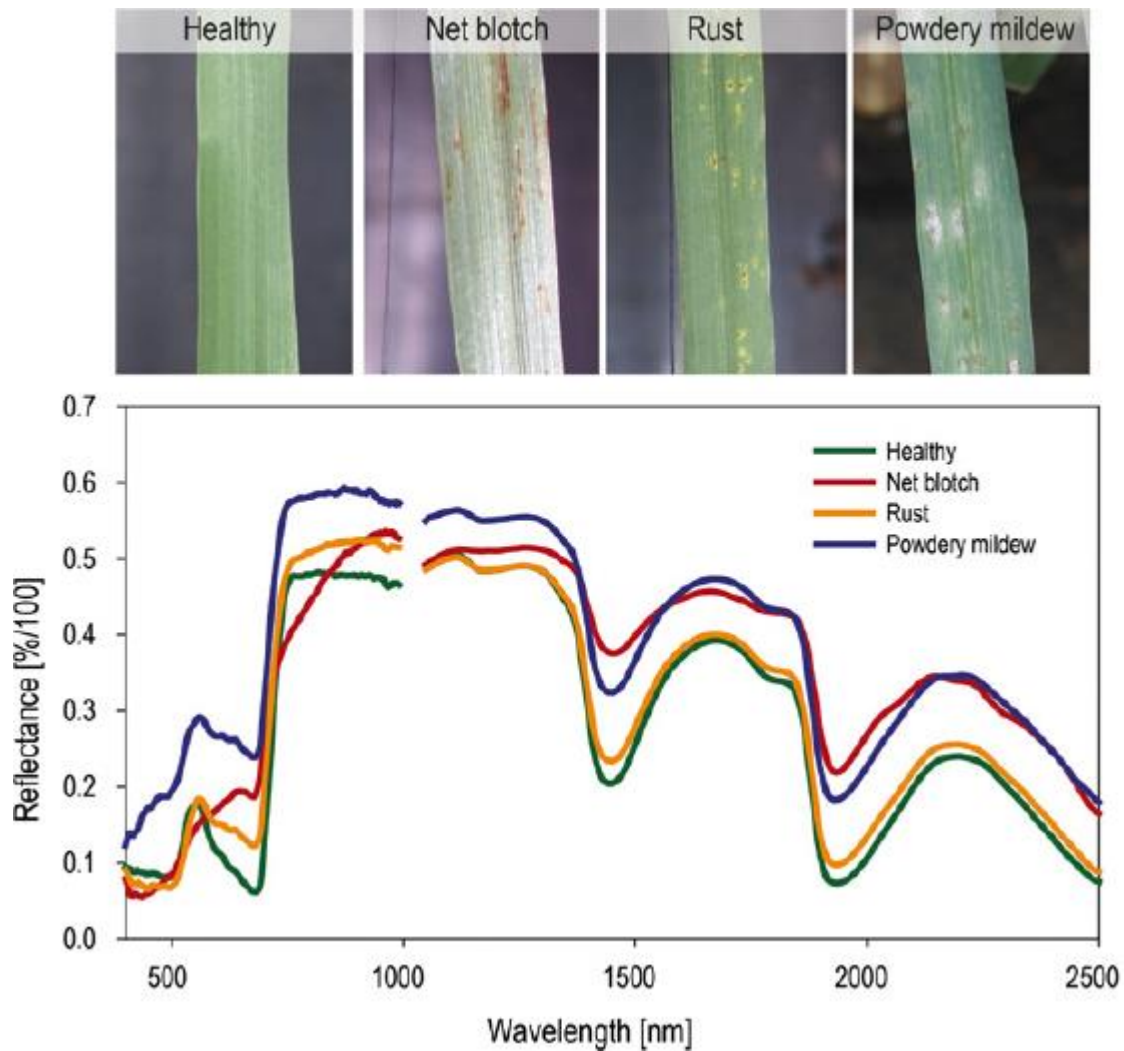


Figure 2-10: Spectral reflectance signatures for barley leaves diseased with net blotch, rust and powdery mildew. From Mahlein (2016).

3 EXPERIMENTS

3.1 Equipment

3.1.1 Plant Samples

The samples for both experiments were acquired from NMBU's own Center for Plant Research in Controlled Climate (SKP). Healthy leaves of cucumber, tomato and strawberry were collected from their respective greenhouses, and mildew infected leaves were collected from isolated containment chambers. Samples were divided into the four categories listed below, and examples from the different “infected”-categories are shown in Figure 3-1:

- **Healthy:** Green, happy plant. Used for control.
- **Infected, early stage:** mildew barely, if at all detectable by the human eye
- **Infected, intermediate stage:** mildew easily visible, but not completely dominant.
- **Infected, late stage:** majority of leaf covered in spores

For the VNIR-experiment, four samples were taken for each plant type; two healthy control samples, one with early stage infection and one with intermediate infection.

For the SWIR-experiment, nine samples were taken for each plant; three healthy control samples, three intermediately infected, one with late stage infection and two samples where mildew was placed directly in agar in an attempt to observe the fungi without the leaf.

A sample would be placed in a closed petri dish with an appropriate agar, administered by staff engineers at SKP. The petri dishes themselves were of the disposable plastic type, as we did not have glass dishes available. Special care was always taken to not accidentally infect the healthy control samples with mildew.



Figure 3-1: Early, intermediate and late stages of mildew infection on strawberry leaves (left to right, respectively). The spores are barely detectable in the early stage, and are marked by two red circles.

3.1.2 Optics, lighting and rig

Cameras

Both of the two experiments used hyperspectral cameras from Specim, Spectral Imaging Ltd.

The first experiment used the VNIR-model. This camera use a silicon sensor, and has a spectral range of 390-1020 nm. The scan line is 1312 pixels wide, and each pixel has 900 spectral bands.

The second experiments used a SWIR camera with a spectral range of 930-2530 nm. The scan line is 320 pixels wide, and each pixel has 256 spectral bands.

In both experiments, the camera in question was mounted in a fixed position, lens facing downward, as seen in Figure 3-2.

Lighting

Three lamps were used for the experiments; two wall mounted lamps and a single standing lamp

The SWIR-experiment included the two wall mounted lamps, as shown in Figure 3-4.

The VNIR-experiment utilized all three light sources, with the standing lamp positioned approximately at the position of the camera in Figure 3-4.

Spectralon rod

Spectralon is a white polymer material with reflectance values generally above 99% over a range from 400 nm to 1500 nm [37]. The intensity of light emitted by the artificial light sources will vary with wavelength, and the spectralon rod his will be used to obtain a white reference for reflection (~100% reflection for all wavelengths).

Lift rig and petri dish mount

To ensure that a given plant sample would be placed in the exact same position from one day to another, a plastic cross was glued to the bottom of every samples petri dish. A plastic



Figure 3-2: The SWIR-camera is shown in the top of the picture, looking down at the sample and spectralon.



Figure 3-3: 3D-printed plastic mount for the petri dishes.

mount, shown in Figure 3-3, was designed and 3D-printed so that the crosses on the petri dishes could only fit in the centre of the mount, with one possible orientation.

The mount was fastened to an adjustable height platform, which in turn was mounted on a linear actuator.

Linear actuator

The cameras both capture images using the push broom-scanning principle presented in chapter 2, and therefore requires that either the camera or the sample is moved along a line perpendicular to the scan line in order to record the second spatial dimension. This was achieved by the use of a ball screw actuator, and the entire system is shown in Figure 3-4.

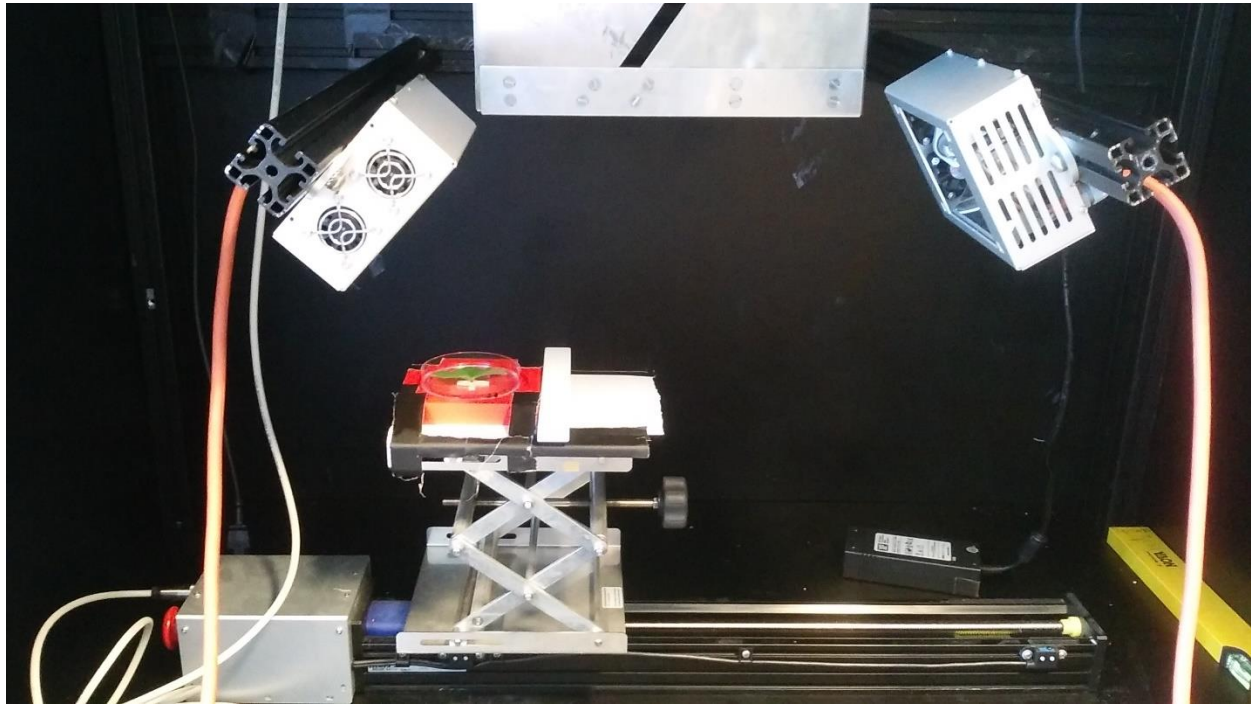


Figure 3-4: The setup with linear actuator, lift, petri dish mount, spectralon rod and two lamps. Camera not visible in top of picture.

Camera software

Both cameras and the linear actuator were connected to a computer with software provided by the camera manufacturer, namely Specim DAQ Solution Software v.3.62. This software was used to adjust settings for spectral and spatial binning, framerate, exposure time and scan speed of the actuator. Real time visualisation of the scanned image as well as sensor saturation was also provided.

For each scan, the software would output two files to disk; a big .raw-file containing the recorded values, and a smaller .hdr-file (header, more specifically ENVI-format), containing metadata. This metadata include the used settings and a list of the wavelengths corresponding to each spectral band (i.e. band 1 = 392 nm for the VNIR-camera), i.a..

3.1.3 Software: MATLAB and PLS/MIA-toolbox

MATLAB by MathWorks Inc. is a matrix-based programming platform optimized for engineering and scientific problems [38]. A library of prebuilt toolboxes contain tools for many standardized applications, an example being the PLS_Toolbox from Eigenvector Research Inc. [39], which will be used to solve many of the statistical tasks in this study.

The software versions used are listed below;

- MATLAB R2016b, v9.1.0
- PLS_toolbox v8.2.1

3.1.4 Software: Scyllarus and Scyven

Scyllarus™ is a set of tools for processing hyperspectral data, developed by Australia based Data 61, formerly NICTA (National Information and Communications Technology Ltd) and CSIRO (Commonwealth Scientific and Industrial Research Organisation) [40]. The visual analysis program Scyven (Scyllarus Visual Environment) will be used in early analysis to get a general idea of how the spectral signatures in the image looks like. Scyven also has its own functions for performing material classification, PCA i.a. However, being a simplified graphical interface, it has some limitations, primarily with respect to using the spectralon as a 100% reflectance area. This software will therefore only be used to perform simple, early inspection of the images, as well as experimenting with different methods for illuminance estimation, pre-processing etc. The software version used was 1.2.0.

3.2 Acquiring data

3.2.1 Calibrations

The camera settings were found primarily by trial and error, and is listed in Table 3-1. In order to achieve a high spatial resolution, the sample would ideally be placed as close to the sensor as possible (while still fitting the entire sample within the image), by raising the adjustable platform. However, this would present a problem with illumination, as the two XXX-lamps would have to be angled in such a way that they would shed light directly onto

the sensor and severely oversaturate it. Thus, a compromise was made with regard to spatial resolution, and the positioning of the lamps and exposure times were adjusted in parallel to achieve maximum illumination without over-saturating the sensors. Spectral binning of the VNIR camera was set to 4, thus reducing the spectral resolution from 900 to 200 bands, implying a wavelengths delta of ~ 3 nm per band. The reason for this was to avoid impractically big data files, and 3 nm per band was assumed to be precise enough. The spectral resolution of the SWIR camera was already at ~ 6 nm per band, and spectral binning was not deemed necessary. Using a print of a chessboard pattern, scan speeds were calibrated to avoid that the image was stretched in either direction.

Table 3-1: Camera settings for the two experiments.

	VNIR camera settings	SWIR camera settings
Spectral Bin	4	1
Spatial Bin	1	1
Framerate	20 ms	20 ms
Exposure	25 ms	0.8 ms
Scan speed	3.18 mm/s	9.20 mm/s
Scan length	180 mm	180 mm

A black reference was achieved by scanning for a couple of seconds with the shutter closed at the end of each picture, and this area would be used to remove background noise and correct pixel-to-pixel variations in the sensor itself. The impact of this correction can be seen in Figure 3-5.

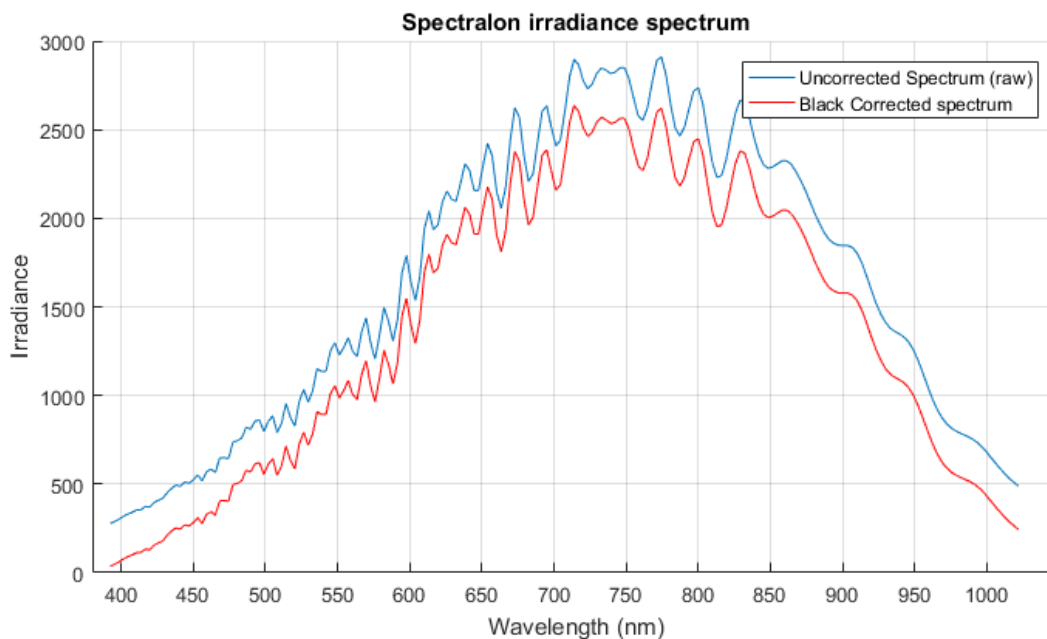


Figure 3-5: Corrected and uncorrected spectralon irradiance spectra in the VNIR-area.

3.2.2 Taking the pictures

VNIR-samples

One picture was taken for each sample nearly every day for 14 days. Days 5, 11 and 12 was not included because of lab accessibility during weekends. With the given settings, each image was saved as an 1194x1312x200 data cube.

SWIR-samples

A slightly different approach was used for the SWIR-samples, as only one picture was taken for each sample instead of taking pictures over several days. With given settings, each image was saved as a 416x320x256 data cube.

In both cases, files were given unique, recognizable names, i.e. TH2_Day01 would imply “Tomato, Healthy sample #2 at Day 1”. Samples were stored openly (though in their respective petri dishes) in a dark room at room temperature.

3.3 Data analysis

3.3.1 Early inspection with Scyven

To get a simple visualisation of the data, a few selected images were loaded into Scyven and analysed using some of its standard functions. The built-in tools were used run images through a series of processes, listed below. To map the accuracy of the software, this would later be compared with the corresponding results from the analysis done in MATLAB.

- Pre-processing: reduce noise by spatial or spectral filters.
- Illuminant estimation: an estimation of the spectral density of external light sources.
- Reflectance recovery: calculates relative spectral reflection in the image.
- Material discovery: materials are found and clustered according to reflectance spectra.
- Principal Component Analysis: performs PCA on the reflectance cube.

First off, different regions in an image were observed to investigate how different settings and circumstances affected the illuminant estimation, material classification and PCA. One picture was chosen for each spectral category (VNIR and SWIR), and three different regions were selected from each picture. The regions are illustrated in Figure 3-6, and [x, y] ranges indicating the spatial pixel coordinates used to crop them from the original images are listed:

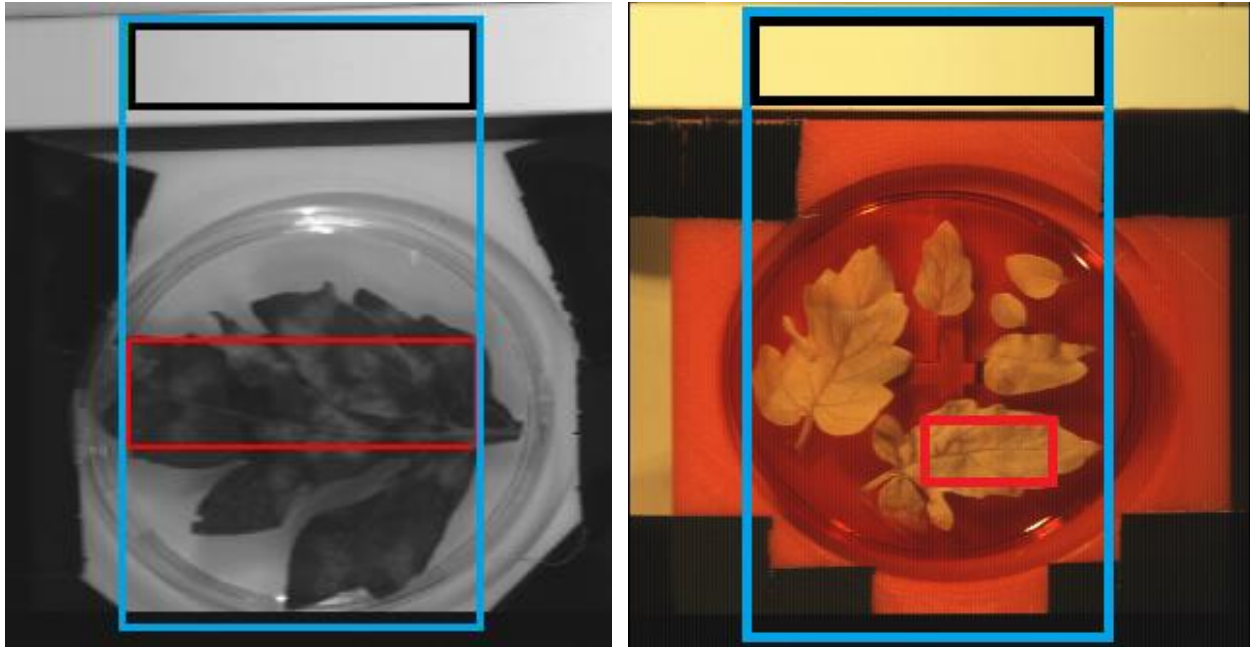


Figure 3-6: Illustration of the cropped areas from VNIR (left) and SWIR (right) tomato leaf samples. The white spectralon can be seen in the top of the picture, as well as the black reference line in the bottom. Blue, black and red rectangles respectively mark "complete sample", "spectralon" and "leaf" areas.

- Complete sample and spectralon ([301:1150, 401:900] and SWIR: [46:360, 61:245])
- Spectralon area (VNIR: [301:420, 451:890] and SWIR: [46:165, 61:245])
- Infected leaf w/ white mildew spots ([756:915, 451:890] and SWIR: ([256:288, 151:215])

For pre-processing, three of the four available alternatives were tested. Primarily, the default filter size of 5 was used, but sizes 7 and 9 were also tested in some cases.

- No filter
- Central Moving Average (CMA) (spectral filter)
- Savitzky-Golay (SavGol) (spectral filter)

Similarly, six different built-in methods for illuminant estimation were tested:

- NICTA
- Finlayson and Schaefer
- Grey Edge
- Grey World
- Shade of Grey
- White Patch

After a sensible illuminant was recovered, strategical points were inspected to identify differences in the reflective spectra of healthy and infected areas. The built-in PCA-tool was also applied to aid in the selection of such points.

A spectral library was created by selecting different areas and labelling them with their respective material, namely spectralon, healthy leaf, mildew (both lightly and heavily infected), red plastic and black tape as shown in Figure 3-7. This library would then be saved and used for classification in different images.

NB!

By visually inspecting irradiance at each band in Seyven, it was not possible to observe any mildew at all in the SWIR images. They will therefore not be given much attention further in the study.



Figure 3-7: Selected polygons for material classification.

3.3.2 Importing and extracting data into MATLAB

Images were imported into the MATLAB workspace with the `readenvihdr`-function (see Appendix B) written at NMBU. Data was extracted with the `HySpec_Analysis` algorithm (Appendix A), which is a modified version of a program also written at NMBU. In short, the program would load raw data into a 3D-matrix, and use the black reference line and areas from the spectralon to, respectively, subtract sensor errors and normalize a “white” value in order to represent data as relative reflectance instead of measured irradiance.

Furthermore, sample areas would be selected for further analysis by simply cropping the data matrix, as the full images contained much irrelevant data, i.e. areas showing the red plastic mount or black duct tape. The program also contains code to unfold the three-dimensional data matrix into 2D, in order to be processed as 2nd order data by PLS-toolbox.

Finally, the program contained code for calculating four of the spectral vegetation- and spectral disease indices presented in chapter 2, namely the MCARI2 and MRENDVI indices from section 2.2.2, and the “healthy-index” and “powdery mildew-index” from section 2.4.

3.3.3 Analysis with SVI and SDIs in MATLAB

By applying the spectral index-equations on each pixel in a sample area, the 200 spectral bands was reduced down to a single index. Thus, new images were created where the value of each pixel was the calculated spectral index. Three areas were analysed; a small area that seemed green and healthy, a small area heavily infected, and a bigger area with both healthy and infected areas. The intention of this was to establish a basis for what index values to expect in healthy and infected areas, and then use this information to classify such areas in the larger sample image.

In addition, both healthy and infected leaves were studied over time to investigate how the MCARI2 and MRENDVI indices changed over time.

It should be noted that analysis with SVI and SDI are only done for the VNIR-images, as the wavelengths used to calculate the indices are all 800 nm or below.

3.3.4 Identifying informative bands with PCA in MATLAB

A sample area would be unfolded into a spectral-by-spatial 2D-matrix and loaded into PLS_Toolbox. Pre-processing was done with a first derivative, 2nd order Savitzky-Golay algorithm.

4 Results and discussion

4.1 Early data from Scyven

4.1.1 Illuminant recovery

Illuminant spectra were recovered using the NICTA-method and three different pre-processing methods; no filter, CMA and SavGol. The spectra recovered from the spectralon area are shown in Figure C-1 to Figure C-3, and the curves demonstrate slight differences in estimated illuminant, CMA smoothing the curve notably more than SavGol, the latter only smoothing slightly compared to the unfiltered spectrum. For the spectralon area, most of the different methods for illuminant recovery demonstrated only small differences in estimated illuminant spectrum. For Gray World- and Shades of Gray- methods, the spectra were identical. The Gray Edge-method stood out, however, as seen in Figure C-4.

When studying the more complex, full sample image, the choice of recovery method had a greater impact on the resulting illuminant spectrum. The spectra resulting from using the NICTA- or Finlayson & Schaefer- methods were quite similar to those found when studying the spectralon, while Gray Edge-, Gray World- and Shades of Gray-methods differentiated from spectralon results. The spectrum from the White Patch-method appear to be cut off by a threshold at power value 0.09. These results are shown in Figure C-5 to Figure C-8.

Attempts to recover an illuminant spectrum from the leaf area proved even more difficult. A curious discovery was the fact that the NICTA-method was unable to estimate an illuminant spectrum using SavGol pre-processing unless the filter size was set to 7. Even more curiously, it was found that the opposite was true when studying the full sample image; filter size 5 and 9 worked, but size 7 produced the same flat power spectrum as 5 and 9 did for the leaf area. In any case, the Gray Edge-method again stood out from the rest. NICTA-, Gray World/Shade of Gray- and White Patch-methods gave similar results, although very different from those found in the two previous images.

4.1.2 Illuminant recovery – section discussion

The factor that had the biggest impact on the illuminance curve, was the chosen area, and if that area contained a part of the spectralon. From the quite large variation in the spectra, is clear that the presence of a reference area have an impact on the way Scyven's estimates the illuminant. It was attempted to save the illuminant recovered while only observing the spectralon, but the exported CSV-file containing the spectrum would not be re-loaded into the software correctly. The error message is shown in Figure 4-3. This meant it was necessary to choose between two evils. To study a big image that included the spectralon would result

in an estimated illuminant that seemed ok, but also a lot of irrelevant data that would greatly affect PCA. The other option was to study an area with only leaf or mildew pixels, that would contain little irrelevant data, but the illuminant would then be estimated incorrectly, and the reflectance likewise.

Another observation was that illumination was not uniform along the vertical axis of the image, shown by PC1 in Figure 4-2 and the variation in the spectrum shown in Figure 4-1. This seems reasonable, since the standing lamp illuminated the sample from the right side of the image. This would mean there will be errors related to position in x-direction.

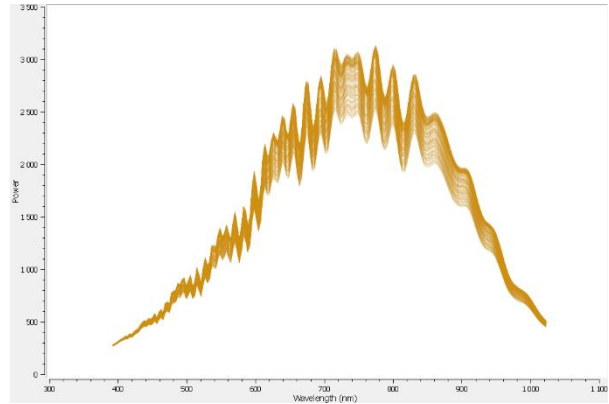


Figure 4-1: Irradiance spectrum for an area of the spectralon.

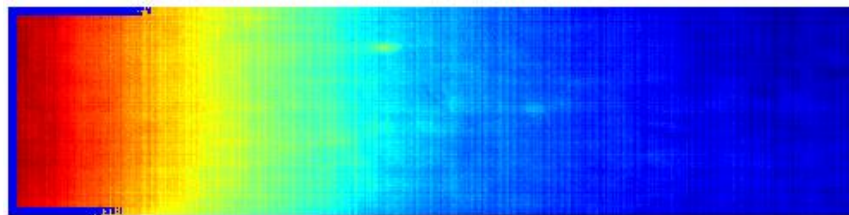


Figure 4-2: PC1 over an area of the spectralon show that reflectance vary with vertical position.

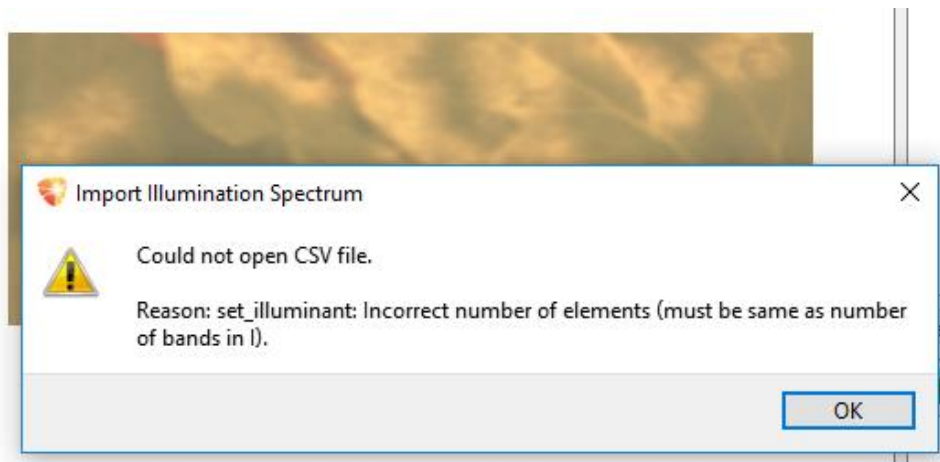


Figure 4-3: Error message received when attempting to import illuminant spectrum.

4.1.3 PCA and analysis of spectrum shapes

Figure 4-4 and Figure 4-5 respectively shows colour representations of scores for PC1 and PC2 from the selected leaf area. Three points of interest, shown in Figure 4-5, were selected in order to see if there was any difference in reflectance when studying the entire sample picture or just the selected area. These reflectance spectra are shown in Figure 4-6, and there is a clearly observable difference between the spectra depending on if the only sample area or the full image was used for calculation illuminance. In either cases, infected areas seem to have a less steep slope in the red edge area, and both healthy and infected areas show a reflection that is greater than that of the spectralon for wavelengths above ~ 780 nm.

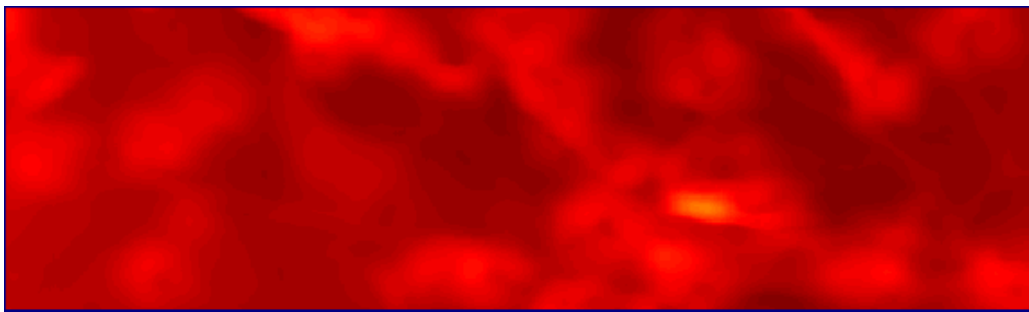


Figure 4-4: Colour representation of scores in PC1 from the selected area.

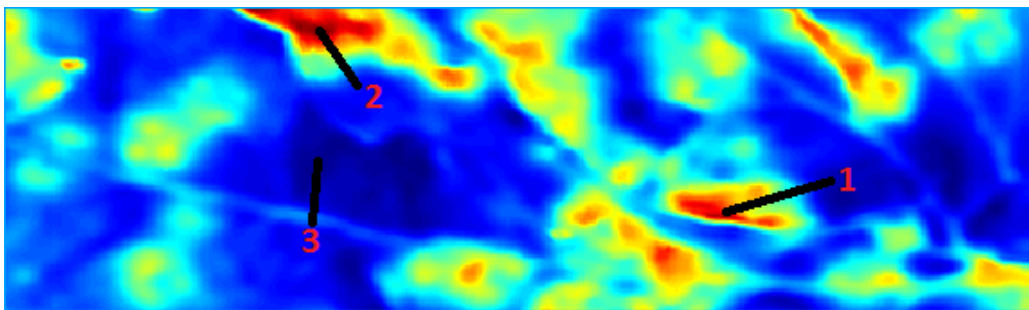


Figure 4-5: Colour representation of scores in PC2 from the selected area, with areas of interest marked with numbers 1-3.

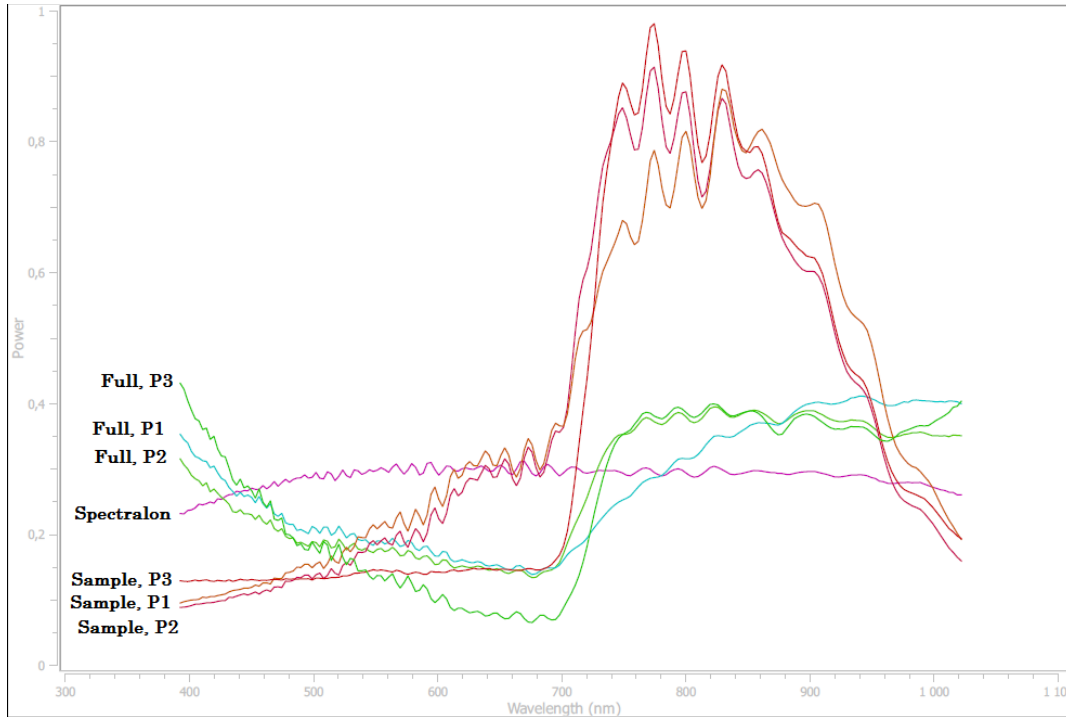


Figure 4-6: Reflectance spectra for selected points of interest while looking at either sample area or full picture.

4.1.4 PCA and spectrum shapes – section discussion

It should be noted that, by studying a small area with only leaf and mildew, the PCA was performed using an incorrectly estimated illuminant. However, the areas with high scores in PC2 are indeed covered with mildew, so it can be assumed that this principal component contain some useful information.

The fact that reflection is greater than spectralon is highly unlikely, due to the spectralon's reflective properties. However, it is possible that this is linked to the non-uniformity of illumination, pointed out in section 4.1.2, and that an extreme value on the more strongly illuminated right side could “outshine” a mean spectralon spectrum on the darker left side.

However, one thing that does agree with results from previous studies, is that mildew infected areas have a higher reflection ratio than healthy areas in the visible part of the spectrum, but lower reflection above 700-720 nm [41].

4.1.5 Classification with Linear Spectral Unmixing and SVM

Figure 4-7 show areas classified as mildew infected after material classification was performed using the generated spectral library. In Figure 4-8, a class for “lightly infected” areas was included to investigate how well the software could tell detect degrees of infection. Figure 4-9 to Figure 4-12 show results the same analysis on cucumber and strawberry leaves.

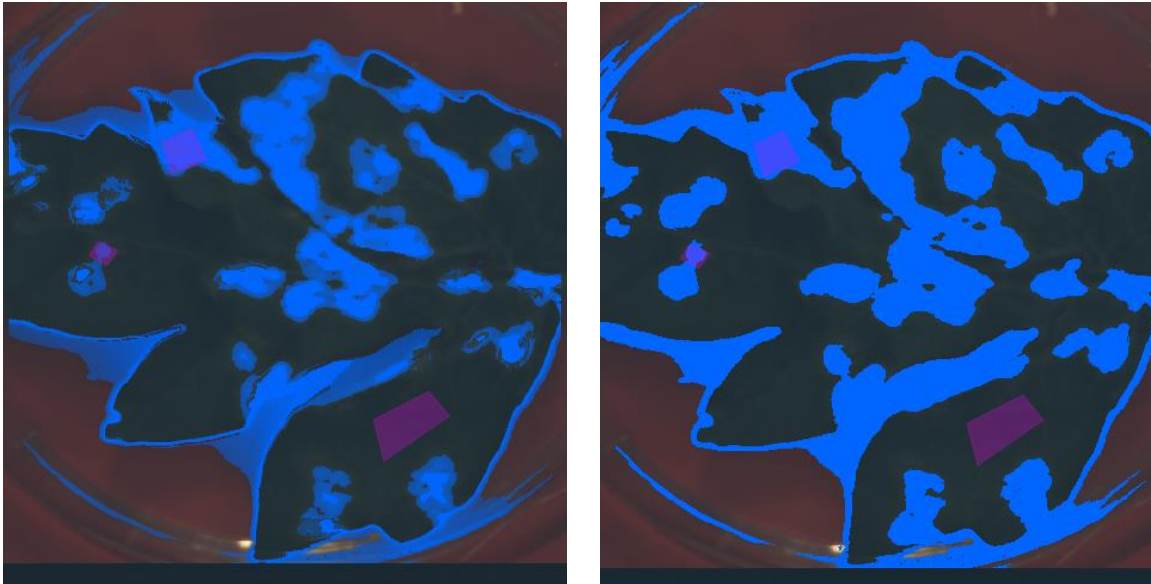


Figure 4-7: Classification performed on an infected tomato leaf with the integrated Spectral Unmixing (left) and SVM-functions (right) in Scyven. Blue colouring indicate that an area is classified as mildew infected.

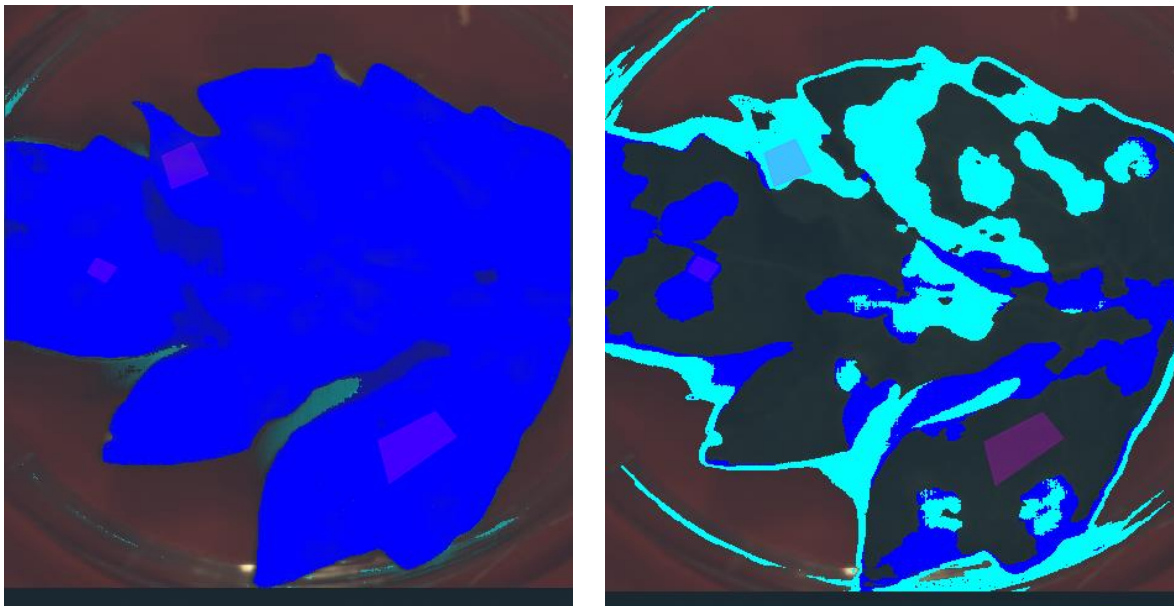


Figure 4-8: Classification performed on an infected tomato leaf with the integrated Spectral Unmixing (left) and SVM-functions (right) in Scyven. Cyan colouring indicate that an area is classified as mildew infected, while blue colouring indicate lightly infected areas.

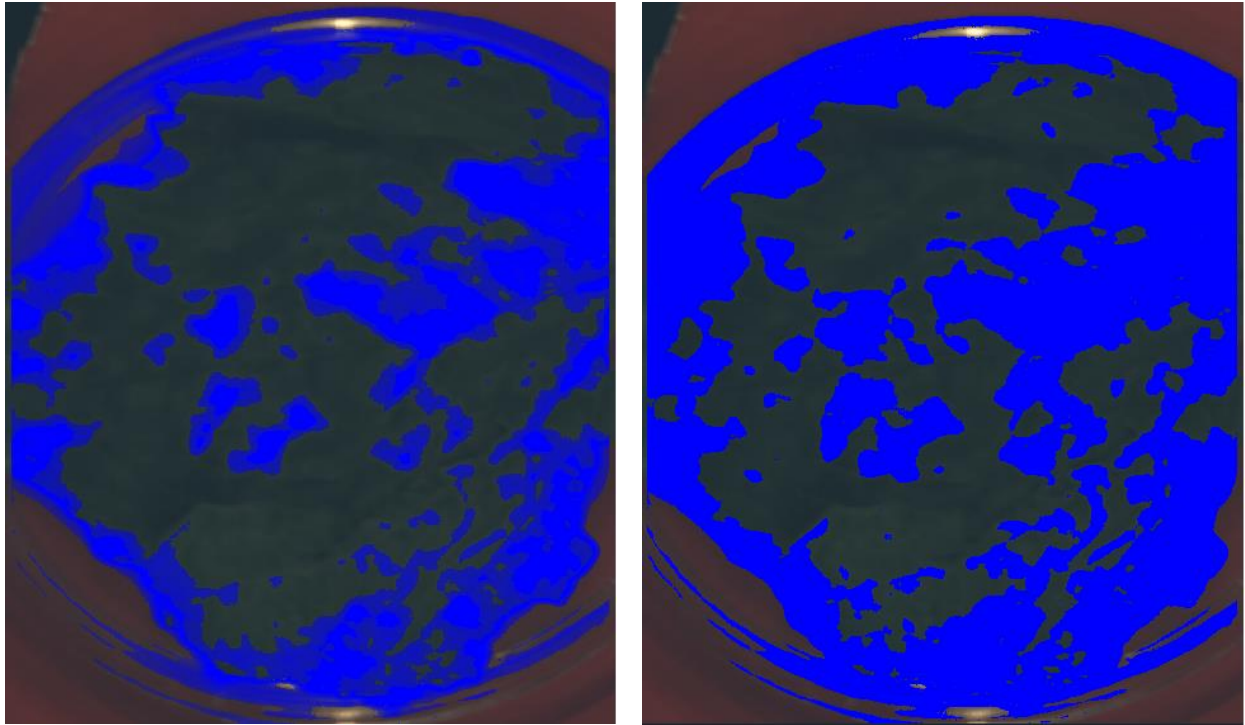


Figure 4-9: Classification performed on an infected cucumber leaf with the integrated Spectral Unmixing (left) and SVM-functions (right) in Scyven. Blue colouring indicate that an area is classified as mildew infected.

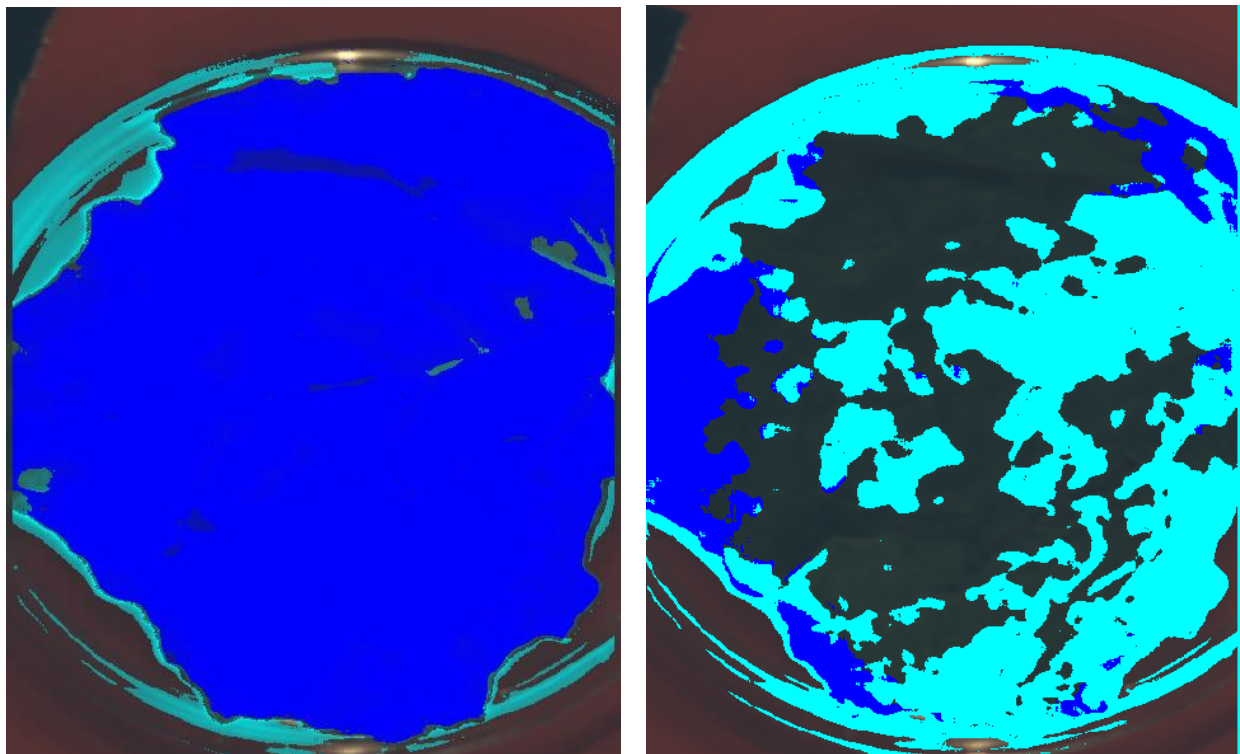


Figure 4-10: Classification performed on an infected cucumber leaf with the integrated Spectral Unmixing (left) and SVM-functions (right) in Scyven. Cyan colouring indicate that an area is classified as mildew infected, while blue colouring indicate lightly infected areas.

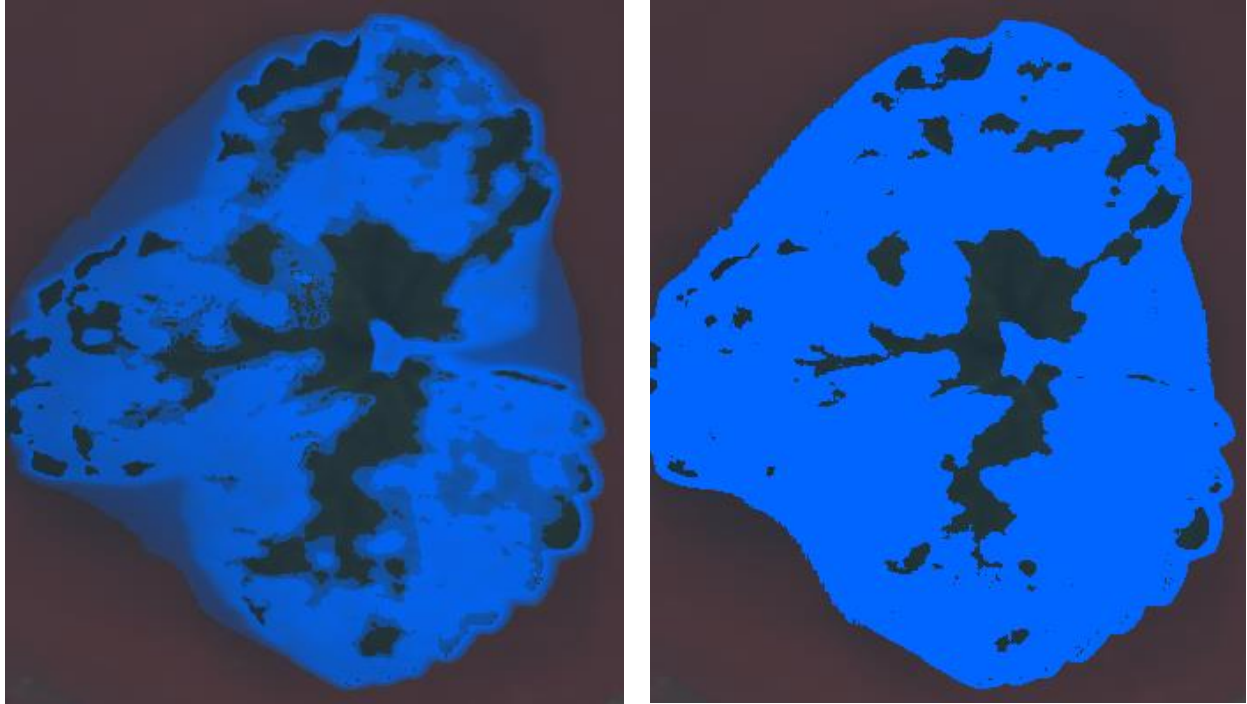


Figure 4-11: Classification performed on an infected strawberry leaf with the integrated Spectral Unmixing (left) and SVM-functions (right) in Scyven. Blue colouring indicate that an area is classified as mildew infected.

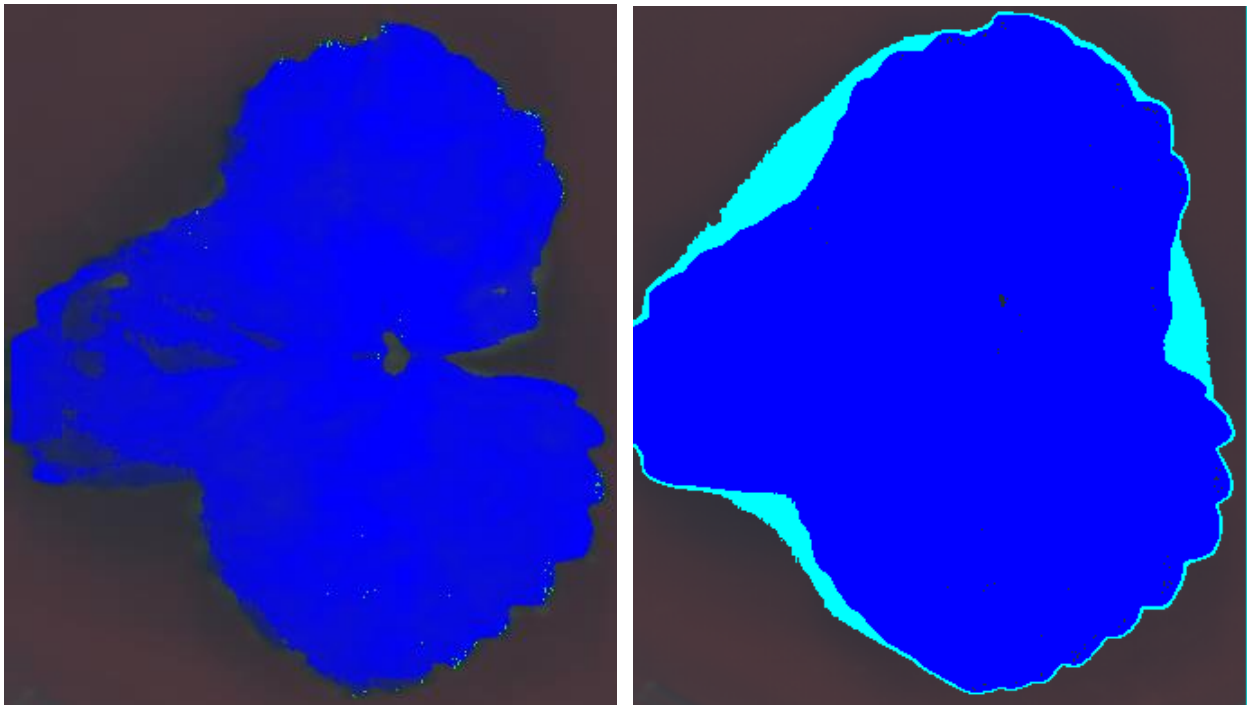


Figure 4-12: Classification performed on an infected strawberry leaf with the integrated Spectral Unmixing (left) and SVM-functions (right) in Scyven. Cyan colouring indicate that an area is classified as mildew infected, while blue colouring indicate lightly infected areas.

4.1.6 Classification – Section discussion

By comparing the classifications with the physical samples, it is apparent that the methods within Scyven does indeed work to some extent. A reoccurring situation is that it mistakes either shadowy areas or parts of the petri dish as mildew. An attempt was made to define more classes, such as shadowy leaf and shadowy plastic, in order to get more precise classification, but result was rather that more shadowy areas would be mistaken for mildew, as well as the other way around.

4.2 Analysis in MATLAB

4.2.1 Spectral index results

For each of the VNIR-images, twelve points were selected, and MCARI2 MRENDVI indices were calculated. The result is visualised on Figure 4-13. MCARI2 show a tendency to decrease slightly over time.

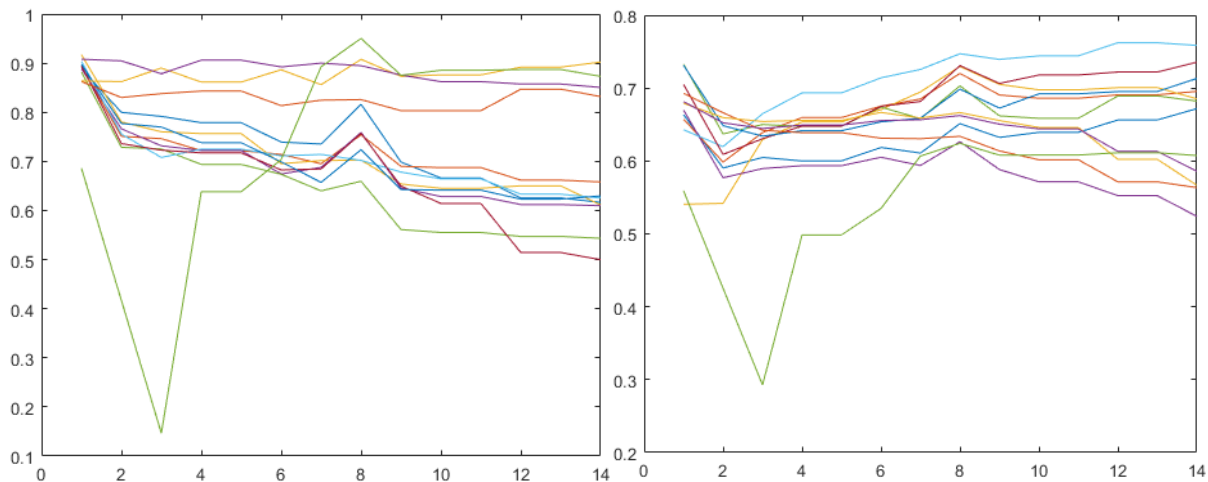
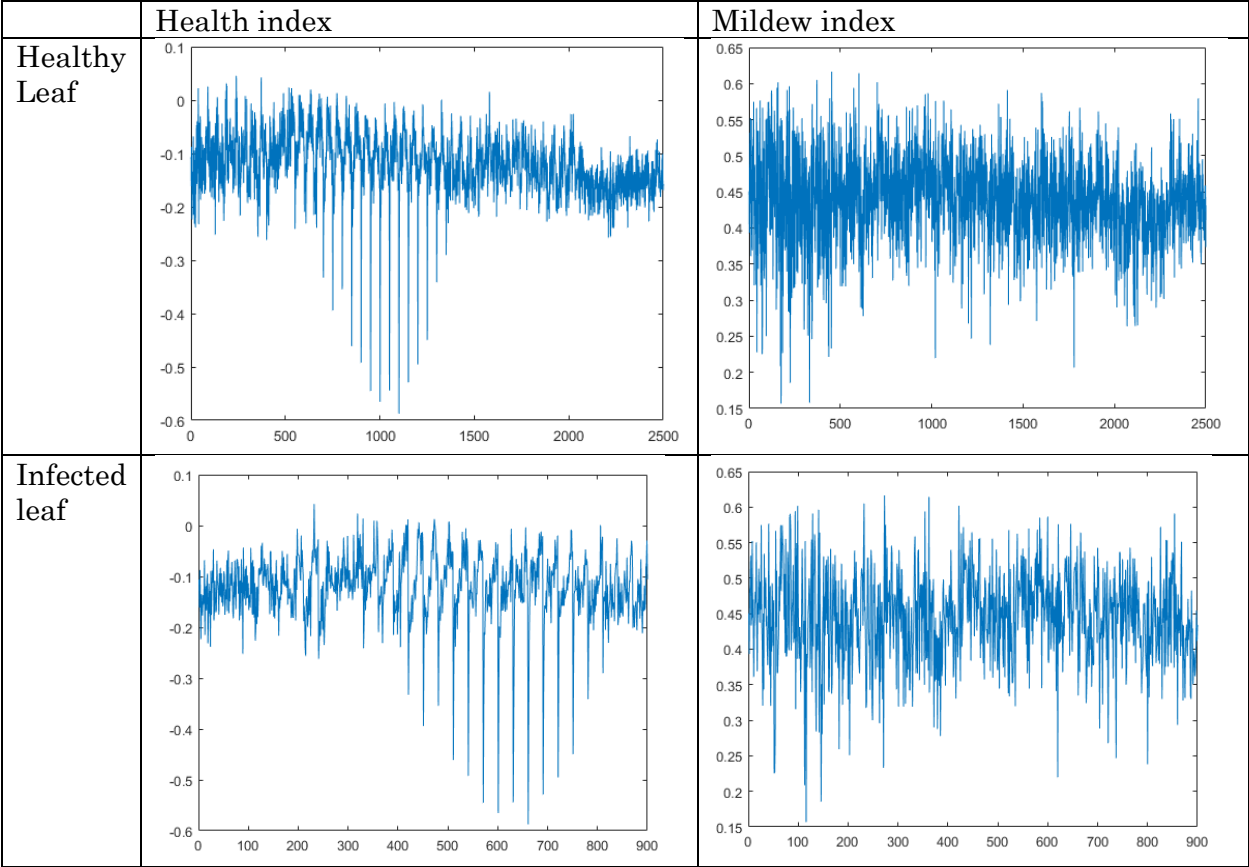


Figure 4-13: MCARI2 (left) and MRENDVI (right) values for the twelve sample points on plant JI2.

Table 4-1 shows scores on HI and PMI for small selected areas; one heavily infected of 30x30 pixels and one without visible signs of infection of 50x50 pixels. Both were unfolded in order to more easily observe the index at each pixel.



Table 4-1: Healthy and infected areas and their scores on HI and PMI



The HI lies between 0 and -0.2, with a notable dip down to -0.6, regardless if the sample area is healthy or infected. Similarly, PMI lies between 0.55 and 0.35, regardless the state of the sample area. Figure 4-14 shows MCARI2 and MRENDVI values represented as pixels in the larger sample image from which they were calculated. Higher index value equals brighter pixel. An obvious difference is apparent between the two.

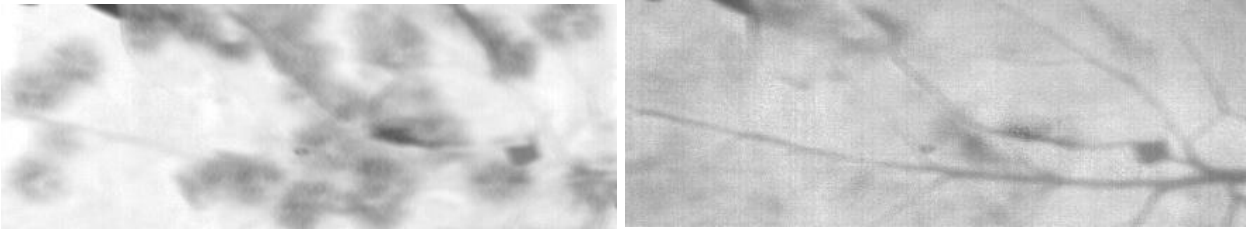
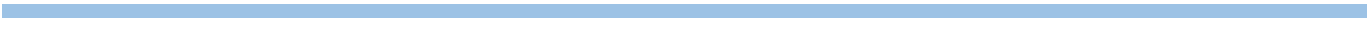


Figure 4-14: Image visualisation of MCARI2 (left) and MRENDVI (right).



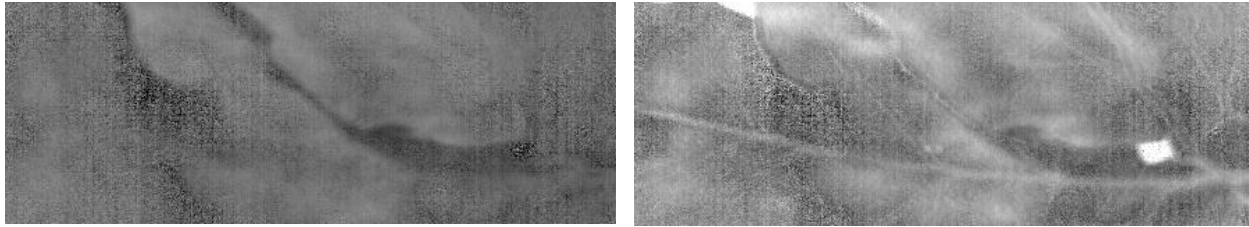


Figure 4-15: Image visualisation of PMI (left) and the difference between PMI and HI (right).

Figure 4-15 show PMI, and also PMI when the HI was subtracted from each respective pixel. The most notable result from this was that the small hole down to the red plastic became visible, and the “veins” in the leaf came out more clearly. The areas covered with mildew also became slightly brighter and more visible.

4.2.2 Spectral indexes – section discussion

The result from this section was somewhat disappointing, as none of them seem to produce any conclusive results with respect to identify the presence of mildew. A surprising result was that the MCARI2-index seemed to detect mildew relatively well, while the PMI worked rather poorly, even when paired with the HI.

One possible reason for lack of conclusive results from Mahlein and Rumpf’s indices could be that the wavelengths used in the SDIs were selected based on test done on sugar beet leaves. Still, errors in the present experiment seem just as (if not more) likely. Author’s limited experience with the equipment, as well as lack of knowledge on statistical tools for extracting valuable data prior to this study are also major factors that might have reduced the precision of the performed experiments.

4.2.3 PCA results

The resulting loadings and score plots can be found in Appendix E.

Each VNIR image from day 1 was loaded into MATLAB. A sample area was selected that covered as much leaf area as possible. For all samples, PC1 explained >90% of the variance in the image, and the loading plot for PC1 had a notable peak just above band #100. Closer inspections revealed that the band which contributed maximum loading varied slightly from sample to sample, between 103 (708 nm) for healthy cucumber, and 108 (723 nm) for diseased tomato.

Looking at the score plots, it can be seen that infected plants show a notably greater span on the scores in PC1.

4.3 Closing reflections

The previous discussion sections have focussed primarily on their respective topics. Some sources for error that have not been mentioned these previous section shall therefore be covered here. As an experiment that rely on high precision, a small error could disturb the results quite a lot. Such a simple thing as getting the right focus on the camera proved to be a small challenge, as the focus wheel needed to be turned manually to adjust focus, while looking at the computer screen to see when the focus was perfect. With the given position of camera and computer, this proved difficult for a single individual. Other hardware-related errors revolved around illumination and the spectralon. Adjusting distance and angle of the lamps

A practical note on Scyven is that the software requires a lot of processing power, and it was even prone to crash while loading big image files (700 MB). It also took far more time to load an image into Scyven (~10 min) than when loading into MATLAB (~1 min). Studying the Windows Task Manager while loading an image showed that memory use by Scyven would slowly increase (>1MB/s) until the entire image was loaded onto memory, before dropping after loading was complete and Scyven again became responsive.

However, as a tool for very simple, early analysis like visual inspection of different reflection rates, it works quite nicely. It also serves well as an introductory software for someone starting with hyperspectral analysis.

5 Conclusion and further research

The study had difficulty achieving a conclusive detection of powdery mildew with previously developed spectral disease indices. Results might improve if more appropriate methods for pre-processing and data filtering are applied.

When a Savitzky-Golay transformation algorithm (1st derivative, 2nd order polynomial) was applied, PCA showed that wavelengths around 700-730 nm contributed the bulk of loadings, and thus >90% of the variation in the (now transformed) image with healthy and mildew-infected areas. This points to the fact that there should be some useful information in this region of the spectrum.

To develop a complete method for discrimination and classification of plant diseases, more tools likely needs to be implemented. Such tools could include some form of artificial intelligence (or machine learning) like support vector machines or artificial neural networks.

The graphical spectral analysis software Scyven proved useful for simple, preliminary studies, but issues regarding illuminant recovery and a high demand for memory makes it less suitable for studies that require high precision and/or analysis of large images.

Even though no conclusive results were found in this study, the field of spectral show great promise, as demonstrated by many previous studies.

6 References

Reference library generated by EndNote X8:

- [1] University of Colorado at Boulder. (2005, 16.04). *Arithmetic, Population and Energy - a talk* by Al Bartlett. Available: http://www.albartlett.org/presentations/arithmetric_population_energy.html
- [2] T. R. Malthus, *An essay on the principle of population, as it affects the future improvement of society*. London, J. Johnson, 1798.
- [3] F. a. A. O. o. t. U. Nations. (17.04). *FAOSTAT*. Available: <http://www.fao.org/faostat/en/#data/RL>
- [4] J. Unsworth. (2010, 21.04). *History of Pesticide Use*. Available: http://agrochemicals.iupac.org/index.php?option=com_sobi2&sobi2Task=sobi2Details&catid=3&sobi2Id=31
- [5] G. F. M. Ltd. (24.04). *RoboCrop InRow Weeder*. Available: http://www.garford.com/products_robocropinrow.html
- [6] A. J. Michaloski, "Method and apparatus for ultraviolet treatment of plants," 1991. Available: <http://www.google.ch/patents/US5040329>.
- [7] W. J. Janisiewicz and et.al., "Use of Low Dose UV-C Irradiation to Control Powdery Mildew Caused by *Podosphaera aphanis* on Strawberry Plants," *Canadian Journal of Plant Pathology*, vol. 38, no. 4, pp. 430-439, 2016.
- [8] A. Suthaparan and et.al., "Suppression of Powdery Mildews by UV-B: Application Frequency and Timing, Dose, Reflectance and Automation," *Plant Disease*, 2016.
- [9] A. Suthaparan *et al.*, "Determination of UV action spectra affecting the infection process of *Oidium neolycopersici*, the cause of tomato powdery mildew," (in eng), *J Photochem Photobiol B*, vol. 156, pp. 41-9, Mar 2016.
- [10] Agriculture.com. (22.04). *Case Unveils Autonomous Concept Tractor*. Available: <http://www.agriculture.com/news/machinery/case-ih-unveils-autonomous-concept-tractor>
- [11] C. Anderson. (17.04). *Agricultural Drones - Relatively cheap drones with advanced sensors and imaging capabilities are giving farmers new ways to increase yields and reduce crop damage*. Available: <https://www.technologyreview.com/s/526491/agricultural-drones/>
- [12] G. P. Thomas. (2013, 25.04). *Hyperspectral Imaging - Methods, Benefits and Applications*. Available: <http://www.azom.com/article.aspx?ArticleID=8495#5>
- [13] Yara. (17.04). *Utvikling og prestasjoner for Yara N-Sensor®*. Available: <http://www.yara.no/gjodsel/Tools-and-Services/n-sensor/Utvikling.aspx>
- [14] AgriOptics. (17.04). *WeedSeeker*. Available: <http://www.agrioptics.co.nz/portfolio/weedseeker/>
- [15] X. Zhao and et.al., "Early Detection of *Aspergillus parasiticus* Infection in Maize Kernels Using Near-Infrared Hyperspectral Imaging and Multivariate Data Analysis," *Applied Sciences*, vol. 7, no. 90, 2017.
- [16] Y. Tian and L. Zhang, "Study on the Methods of Detectiong Cucumber Downy Mildew Using Hyperspectral Imaging Technology," *Physics Procedia*, vol. 33, pp. 743-750, 2012.
- [17] A. Picon and et.al., "Early detection of downy mildew in vineyards based on hyperspectral imaging."

- [18] M. A. Teena and et.al., "Near infrared (NIR) hyperspectral imaging to classify fungal infected date fruits," *Journal of Stored Products Research*, vol. 59, pp. 306-313, 2014.
- [19] GISGeography. <http://gisgeography.com/multispectral-vs-hyperspectral-imagery-explained/>. Available: <http://gisgeography.com/multispectral-vs-hyperspectral-imagery-explained/>
- [20] J. E. Burger, "Hyperspectral NIR Image Analysis," Doctoral Thesis, Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, **Umeå, Sweden**, 2006.
- [21] B. Kendrick, *The Fifth Kingdom*, 3rd ed. Focus Pub., 2000.
- [22] R. Diez. (2014). *Powdery Mildew - Effects on Plants*. Available: <http://dyna-gro-blog.com/powdery-mildew-effects-on-plants/>
- [23] T. R. Gordon and J. M. Duniway, "Effects of Powdery Mildew Infection on the Efficiency of CO₂ Fixation and Light Utilization by Sugar Beet Leaves," *Plant Physiology*, vol. 69, 1982.
- [24] B. C. Freeman and G. A. Beattie, "An Overview of Plant Defenses against Pathogens and Herbivores," *The Plant Health Instructor*, 2008.
- [25] R. B. Smith, *Introduction to Hyperspectral Imaging*. MicroImages Inc., 2012.
- [26] A. K. Mahlein, "Plant Disease Detection by Imaging Sensors - Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping," *Plant Disease*, vol. 100, no. 2, pp. 241-251, 2016.
- [27] Harris Geospatial Solutions. *Narrowband Greenness*. Available: <https://www.harrisgeospatial.com/docs/NarrowbandGreenness.html>
- [28] S. V. Ollinger, "Sources of variability in canopy reflectance and the convergent properties of plants," *New Phytologist*, vol. 189, no. 2, pp. 375-394, 2011.
- [29] D. Haboudane *et al.*, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote sensing of environment*, vol. 90, no. 3, pp. 337-352, 2004.
- [30] D. A. Sims and J. A. Gamon, "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages," *Remote sensing of environment*, vol. 81, no. 2, pp. 337-354, 2002.
- [31] L. I. Smith, "A Tutorial on Principal Components Analysis," 2002.
- [32] J. Shlens, "A Tutorial on Principal Component Analysis - Derivation, Discussion and Singular Value Decomposition," 2003.
- [33] T. Rumpf *et al.*, "Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance," *Computers and Electronics in Agriculture*, vol. 74, no. 1, pp. 91-99, 10// 2010.
- [34] A. K. Mahlein *et al.*, "Development of spectral indices for detecting and identifying plant diseases," *Remote Sensing of Environment*, vol. 128, pp. 21-30, 1/21/ 2013.
- [35] M. Eady *et al.*, "Rapid and early detection of Salmonella serotypes with hyperspectral microscopy and multivariate data analysis," *J Food Prot*, vol. 78, no. 4, pp. 668-74, Apr 2015.
- [36] M. P. Eady, B., "Classification of Salmonella enterica serotypes with selective bands using visible/NIR hyperspectral microscope images," *J Microsc*, vol. 263, no. 1, pp. 10-9, Jul 2016.

-
- [37] G. T. Georgiev and J. J. Butler, "Long-term calibration monitoring of Spectralon diffusers BRDF in the air-ultraviolet," *Applied Optics*, vol. 46, no. 32, pp. 7892-7899, 2007/11/10 2007.
- [38] MathWorks. *MATLAB Overview - The Language of Technical Computing*. Available: <https://se.mathworks.com/products/matlab.html>
- [39] Eigenvector Research Inc. *Software*. Available: <http://www.eigenvector.com/software/index.htm>
- [40] Data 61. *Scyllarus - Software*. Available: <https://scyllarus.data61.csiro.au/software/>
- [41] L. Yuan and et.al., "Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects," *Field Crops Research*, no. 156, 2014.

7 Appendixes

Appendix A.....	I
Appendix B.....	II
Appendix C.....	IV
Appendix D	IX
Appendix E	XII

A - HySpec_Analysis-program

```
FileName='Day01_T_S2';
%Day = 1;

%----- Importing raw data -----
thdr = readenvihdr([FileName '.hdr']);
Data_RAW = multibandread([FileName '.raw'], [str2num(thdr.lines), str2num(thdr.samples), str2num(thdr.bands)], 'uint16', 0, thdr.interleave, 'ieee-le');
NumberOfLines = str2num(thdr.lines);
NumberOfSamples = str2num(thdr.samples);
NumberOfBands = str2num(thdr.bands);
Wavelengths = str2num(thdr.Wavelength)';

%imtool(Data_RAW(:,:,80),[]) %Shows whole image, radiance at band 80

%----- Black/white references -----

% Creating black reference
Data_CorrB = zeros(str2num(thdr.lines), str2num(thdr.samples), str2num(thdr.bands));
ref_black = squeeze(mean(Data_RAW(NumberOfLines-9:NumberOfLines, :, :), 1));
for i = 1:NumberOfLines
    Data_CorrB(i, :, :) = squeeze(Data_RAW(i, :, :)) - ref_black; %Correcting for black
end

% Creating white reference
Data_CorrBW = zeros(str2num(thdr.lines), str2num(thdr.samples), str2num(thdr.bands));
ref_white = squeeze(mean(Data_CorrB(316:415, :, :), 1))';
for i=1:NumberOfLines
    for j=1:NumberOfBands
        Data_CorrBW(i, :, j) = Data_CorrB(i, :, j) ./ ref_white(j, :); % Correcting for white
    end
end
```

B - readenvihdr-function

```
function [hdr]= readenvihdr(filename)
%Read in header.
[fid,msg] = fopen(filename,'rt');

if fid<0
    error(msg)
end

frewind(fid);
hcell = {' '};

while ~feof(fid)
    hcell = [hcell {fgetl(fid)}];
end
fclose(fid);

%Remove first empty cell.
hcell = hcell(2:end);
i = 1;

flines = strfind(hcell,'='); %Lines with field names.
hinfo = [];

while i <= length(hcell)
    if ~isempty(flines{i})
        %Field name.
        fname = strtrim(hcell{i}(1:flines{i}-1));
        fname = strrep(fname,' ','_');
        fname = genvarname(fname);
        if i==length(hcell) || ~isempty(flines{i+1})
            %Value is contained on single line.
            myval = strtrim(hcell{i}(flines{i}+1:end));
            i = i+1;
        elseif ~isempty([flines{i+2:end}])
            %Multiline value but not the end. Find the next field name line and
            %create value from lines in between.
            j = i+2;
            while j<length(hcell)
                if ~isempty(flines{j})
                    %Found next field line.
                    loc = j;
                    break
                else
                    j = j+1;
                end
            end
            myval = [hcell{i:j-1}];
            myval = myval(strfind(myval,'{')+1:strfind(myval,'}')-1);
            i = j;
        else
            %Value is rest of cell array.
```

```
    myval = [hcell{i:end}];
    myval = myval(strfind(myval, '{')+1:strfind(myval, '}')-1);
    i = length(hcell);
end
hinfo.(genvarname(fname)) = myval;
else
    i = i+1;
end
end
hdr=hinfo;
```

C - Scyven illuminant spectra

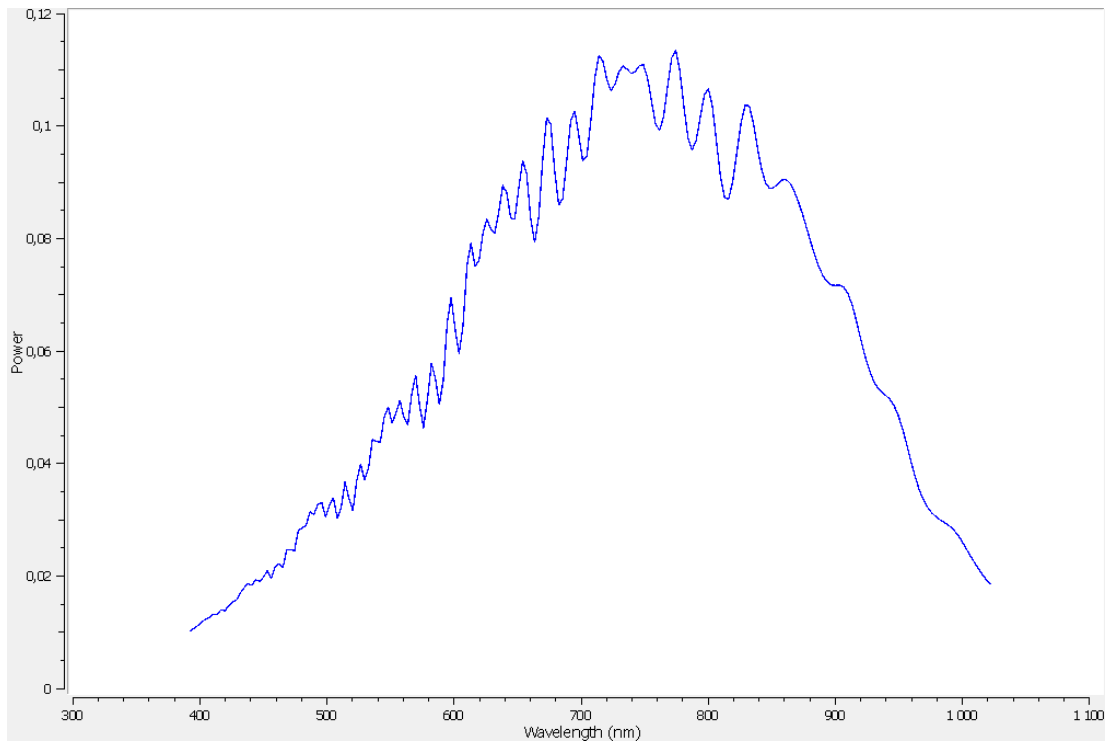


Figure C-1: Illuminance spectra recovered from spectralon area, NICTA-method, unfiltered.

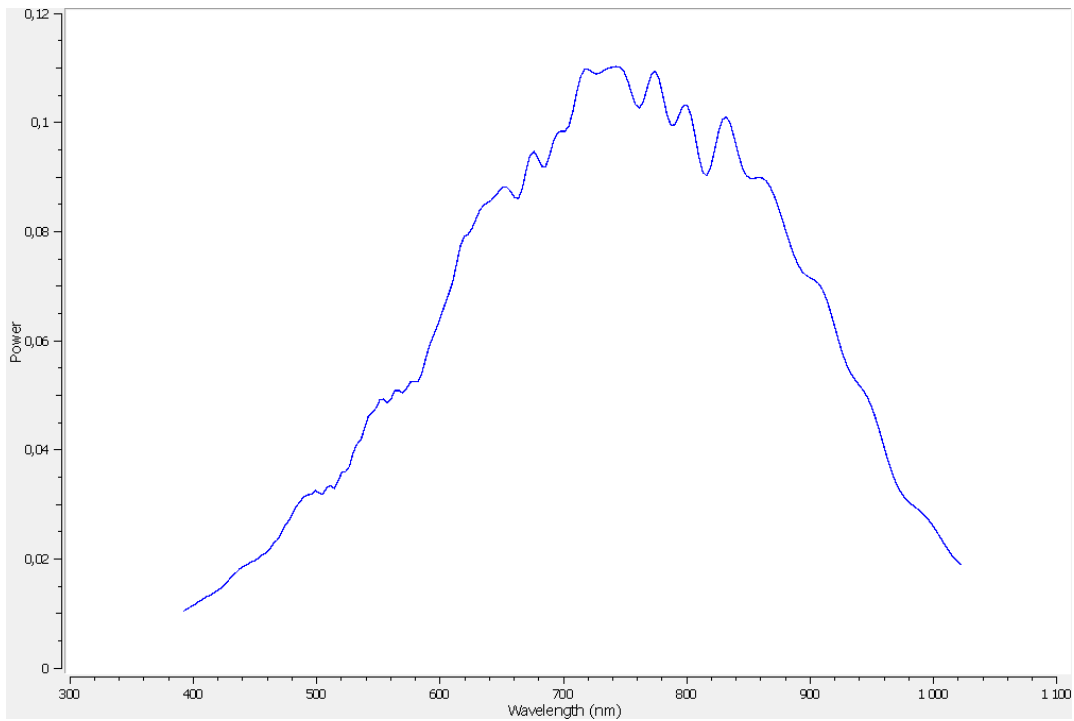


Figure C-2: Illuminance spectra recovered from spectralon area, NICTA-method, CMA-filter

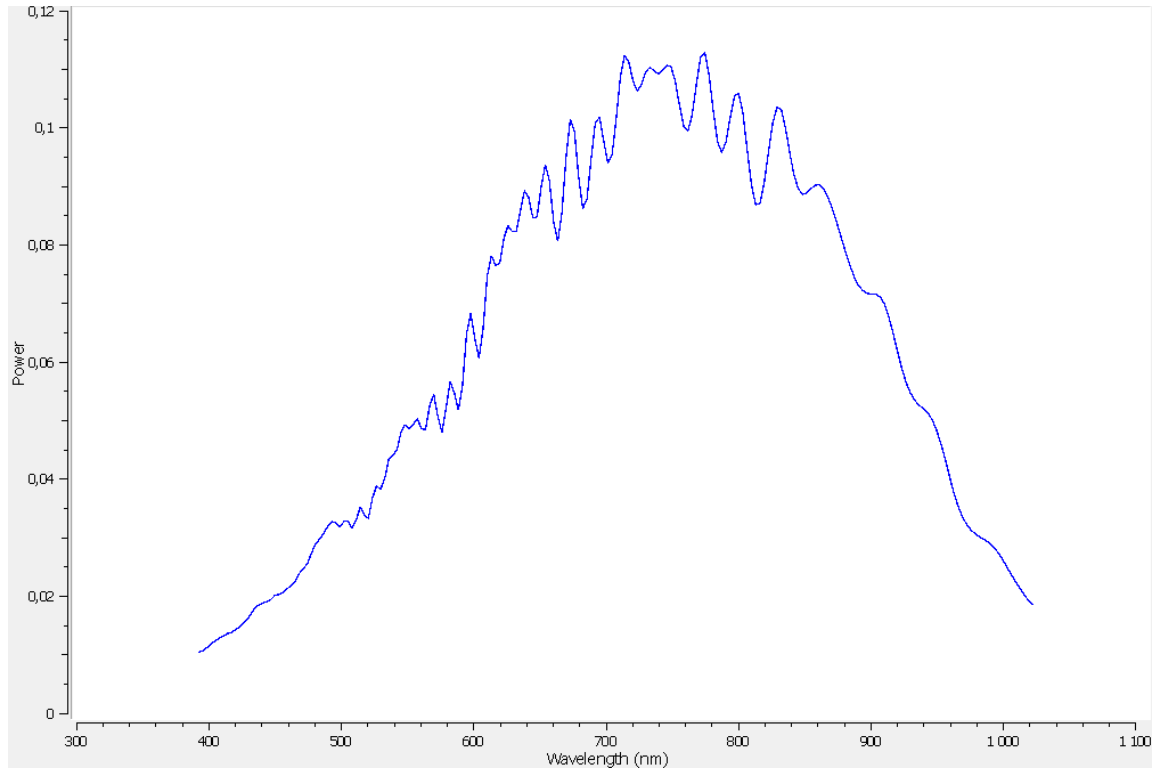


Figure C-3: Illuminance spectra recovered from spectralon area, NICTA-method, SavGol-filter.

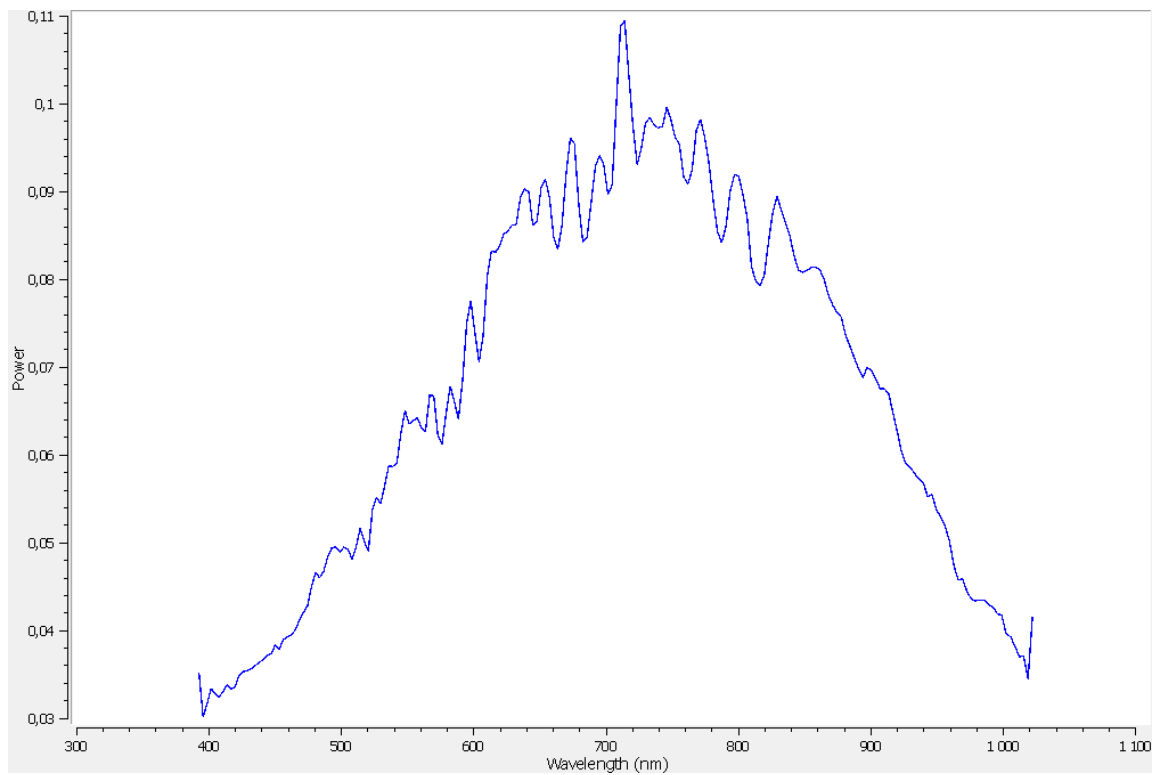


Figure C-4: Illuminance spectra recovered from spectralon area, Gray Edge-method, SavGol-filter.

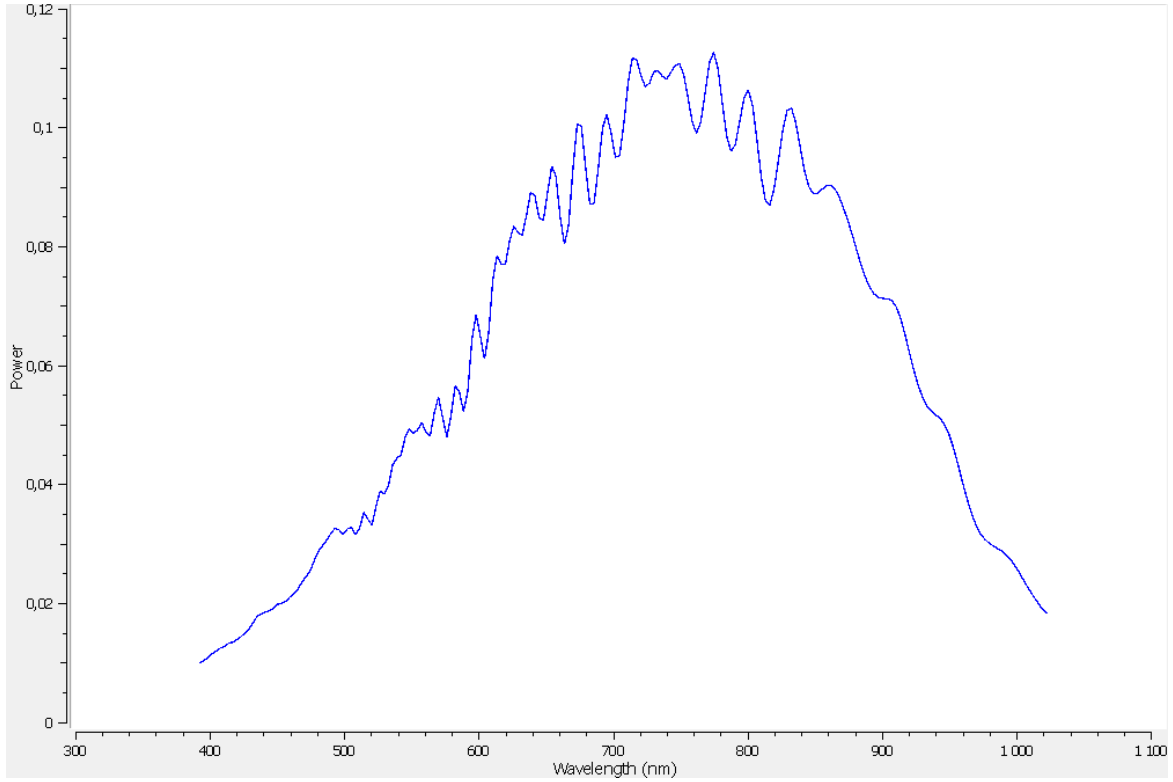


Figure C-5: Illuminance spectra recovered from full sample area, NICTA-method, SavGol-filter.

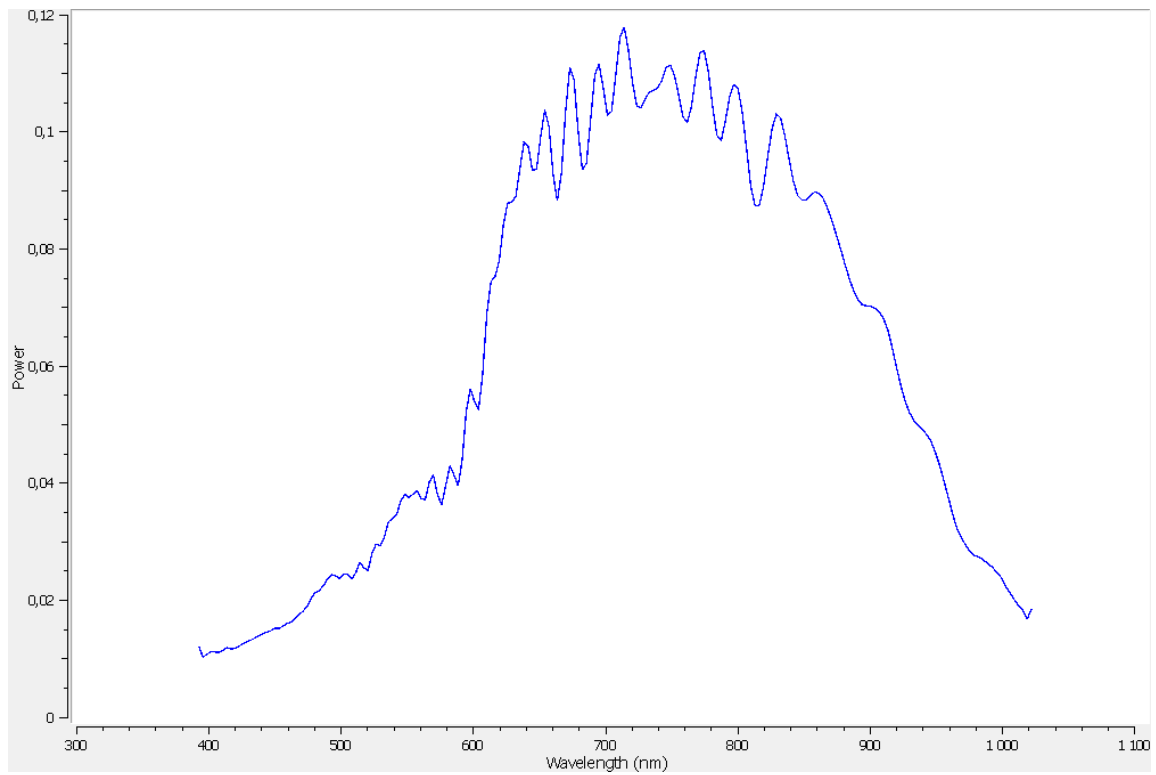


Figure C-6: Illuminance spectra recovered from full sample area, Gray Edge-method, SavGol-filter.

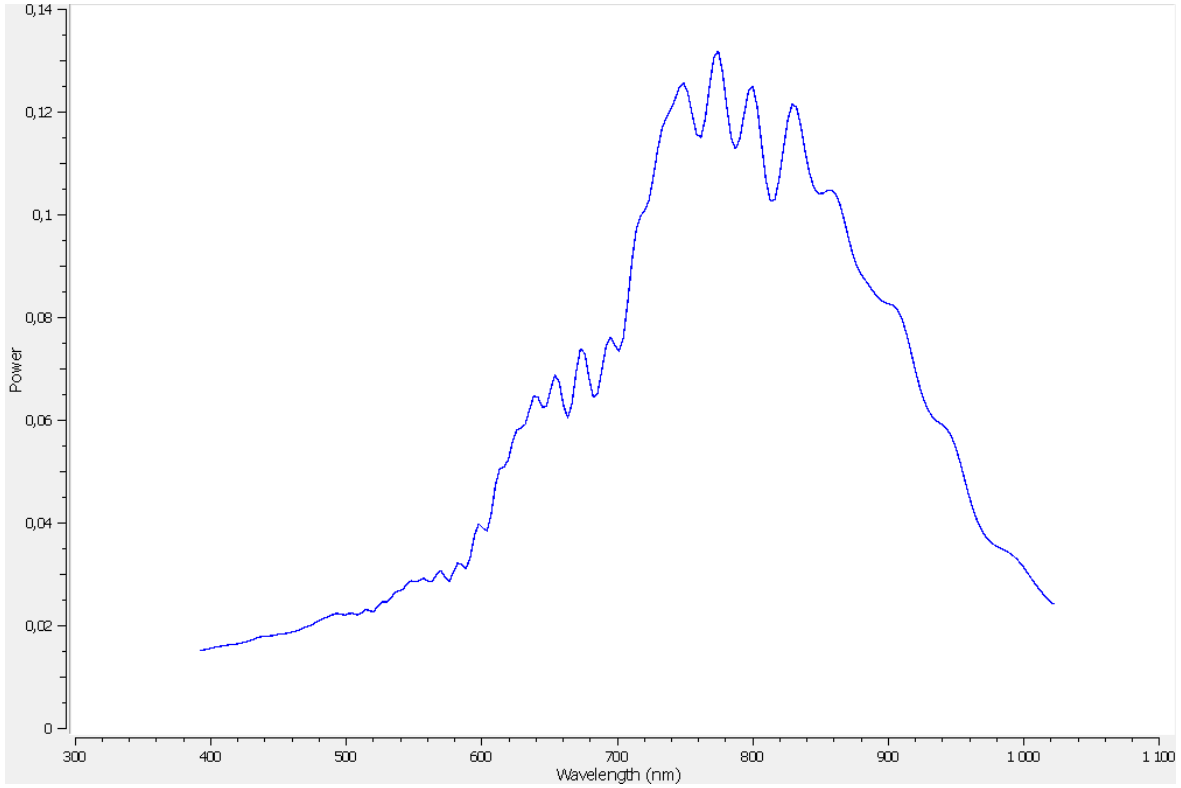


Figure C-7: Illuminance spectra recovered from full sample area, Gray World-method, SavGol-filter.

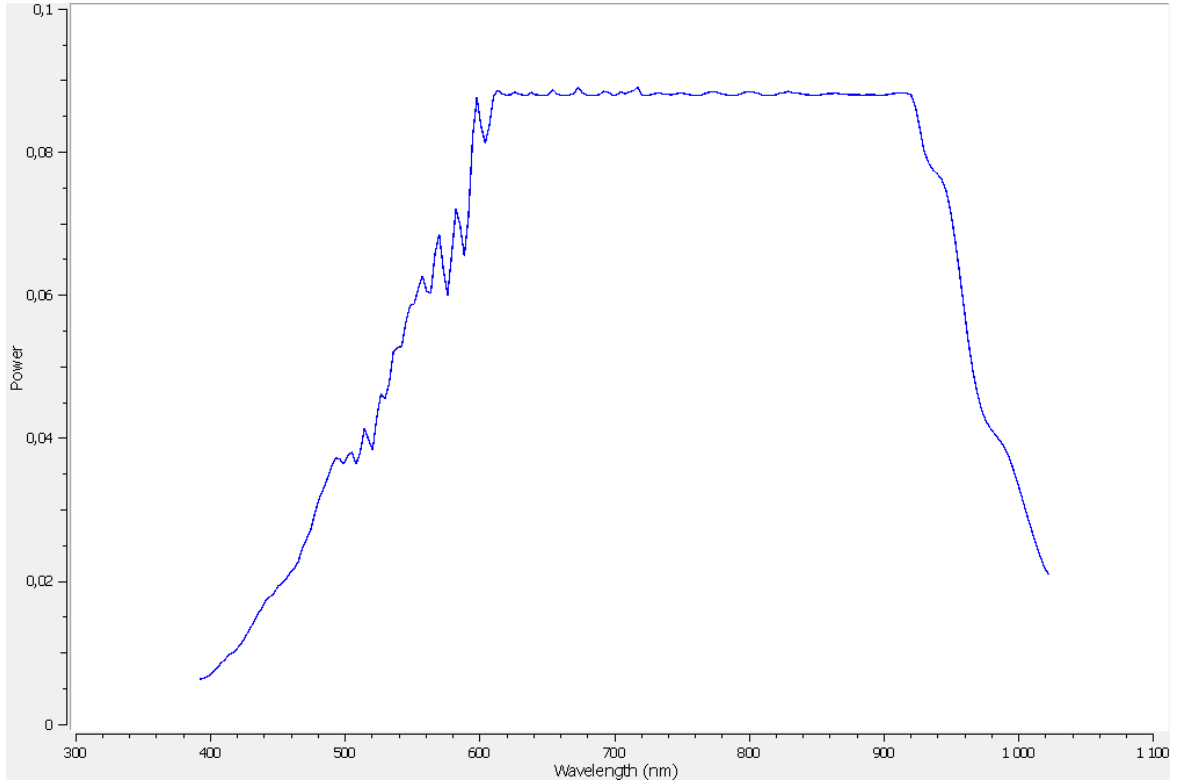


Figure C-8: Illuminance spectra recovered from full sample area, White Patch-method, SavGol-filter.

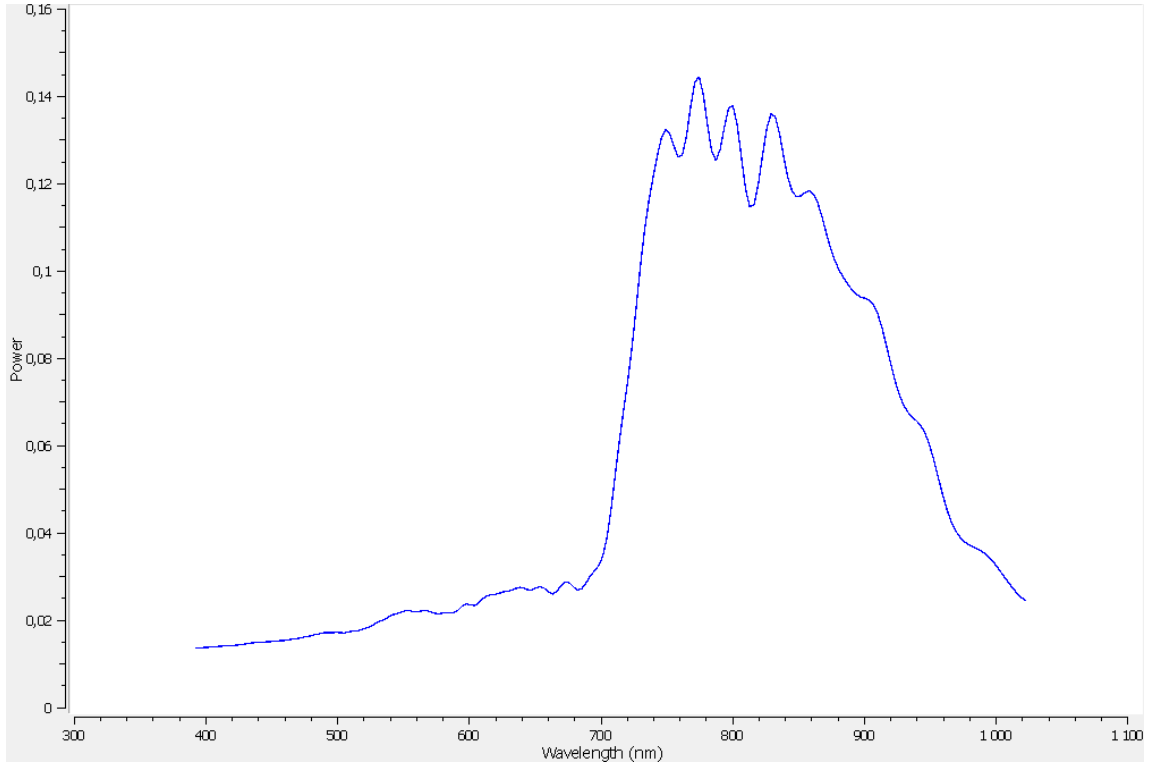


Figure C-9: Illuminance spectra recovered from leaf area, NICTA-method, SavGol(7)-filter.

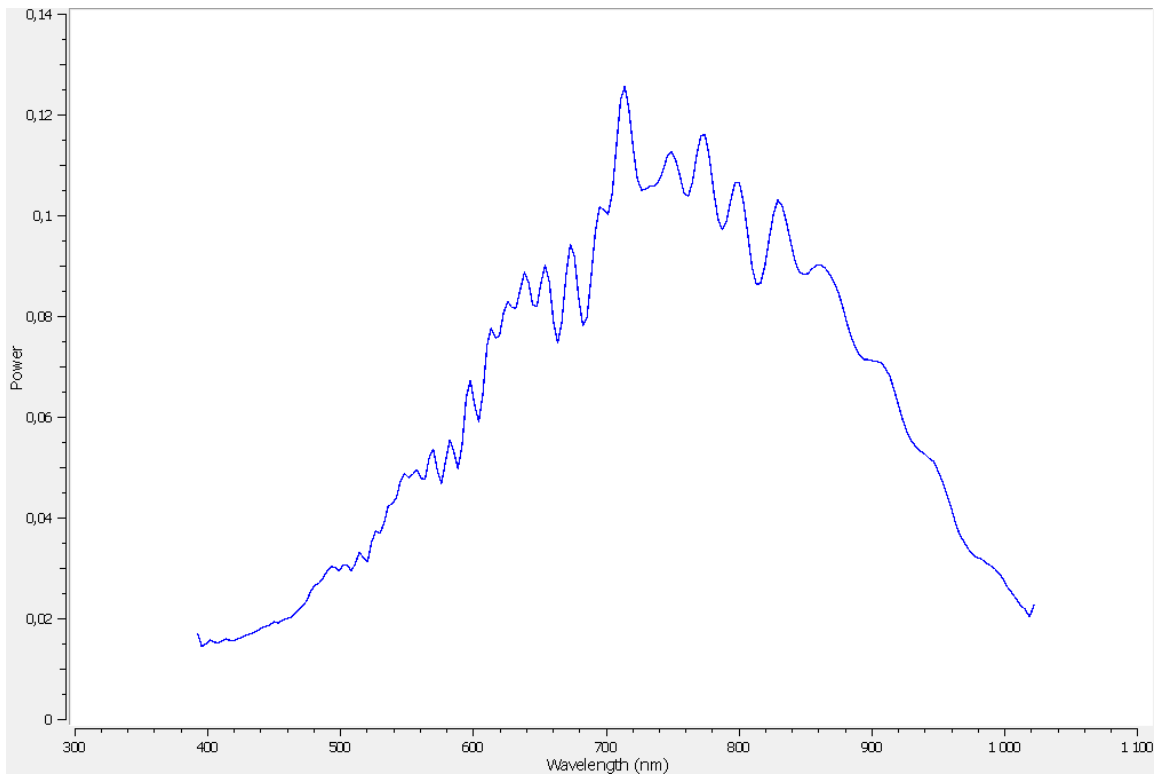


Figure C-10: Illuminance spectra recovered from leaf area, Grey Edge-method, SavGol-filter.

D - VNIR Samples Areas

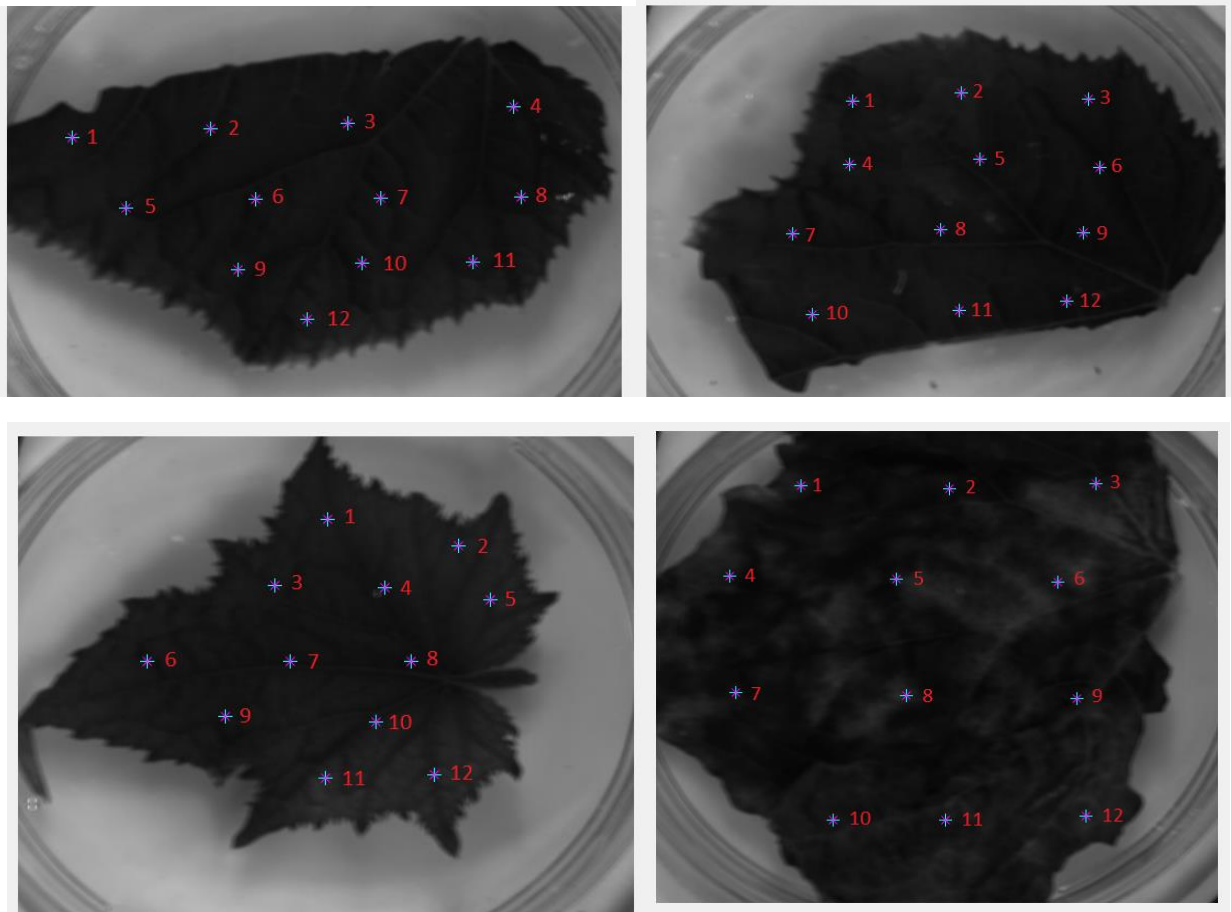


Figure D-1: Locations of sample areas on cucumber leaves.

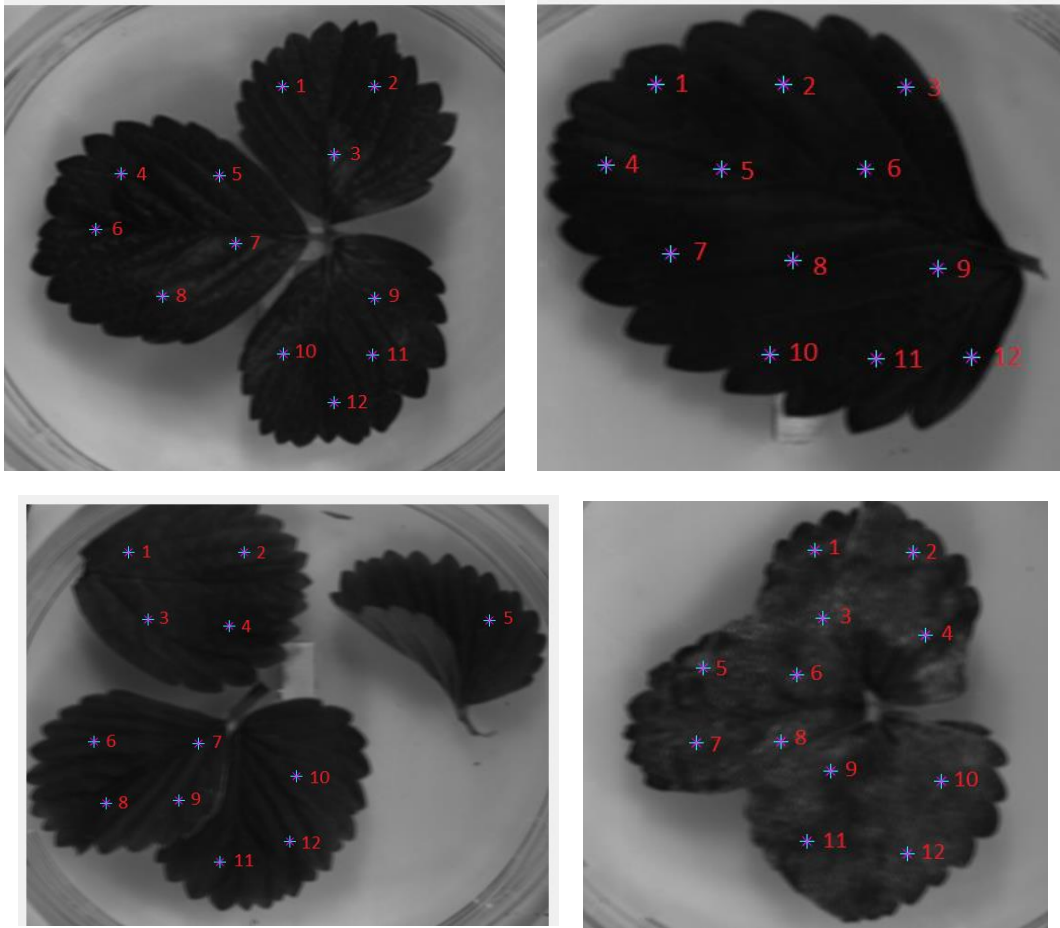


Figure D-2: Locations of sample areas on strawberry leaves.

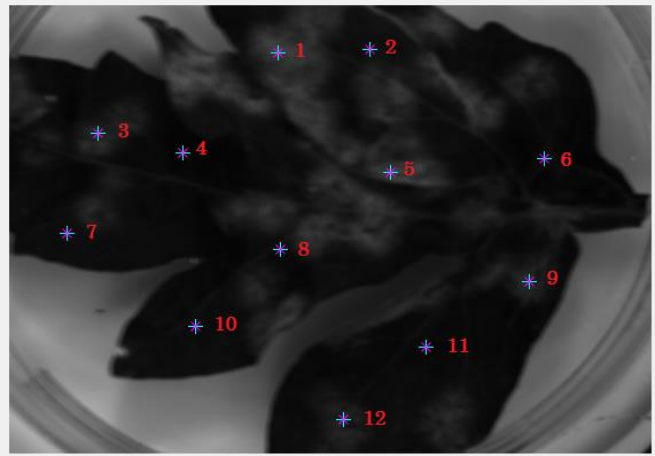
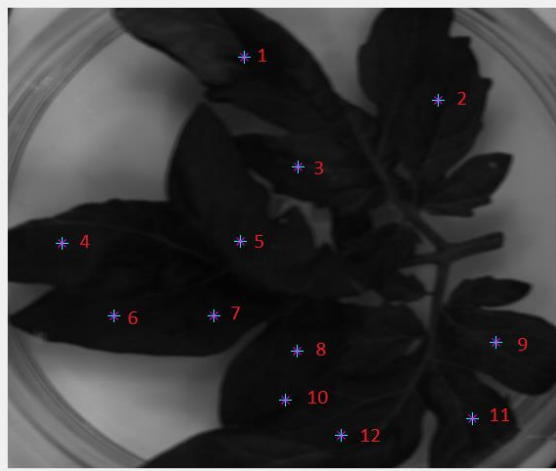
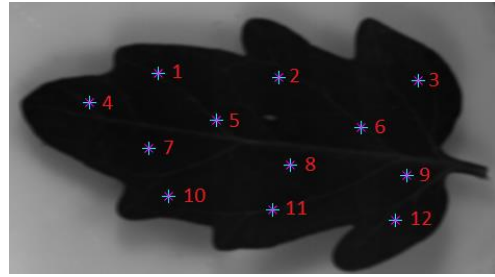
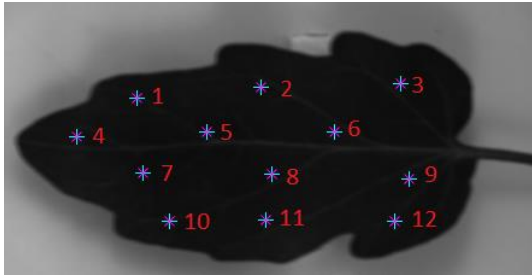


Figure D-3: Locations of sample areas on tomato leaves.

E - PCA scores and loadings

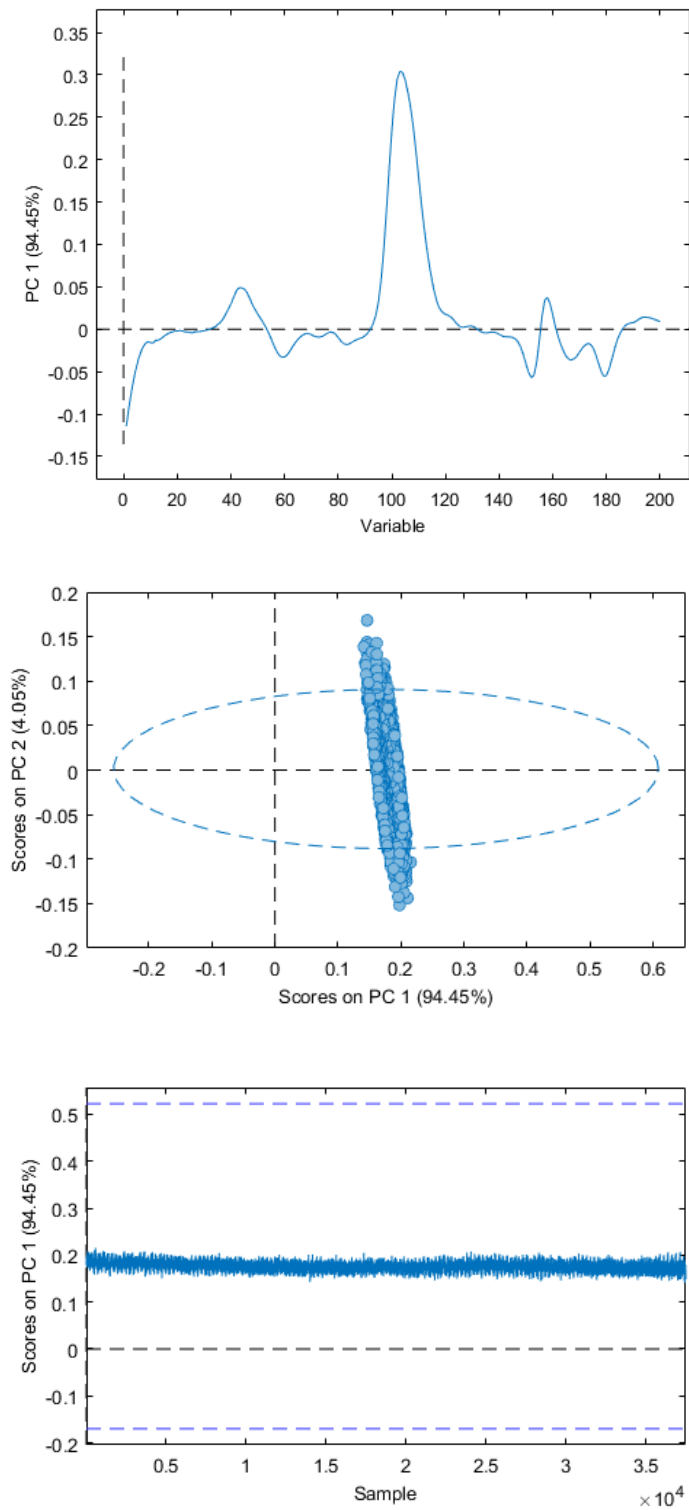


Figure E-1: Loading plot and scores for cucumber, healthy #1

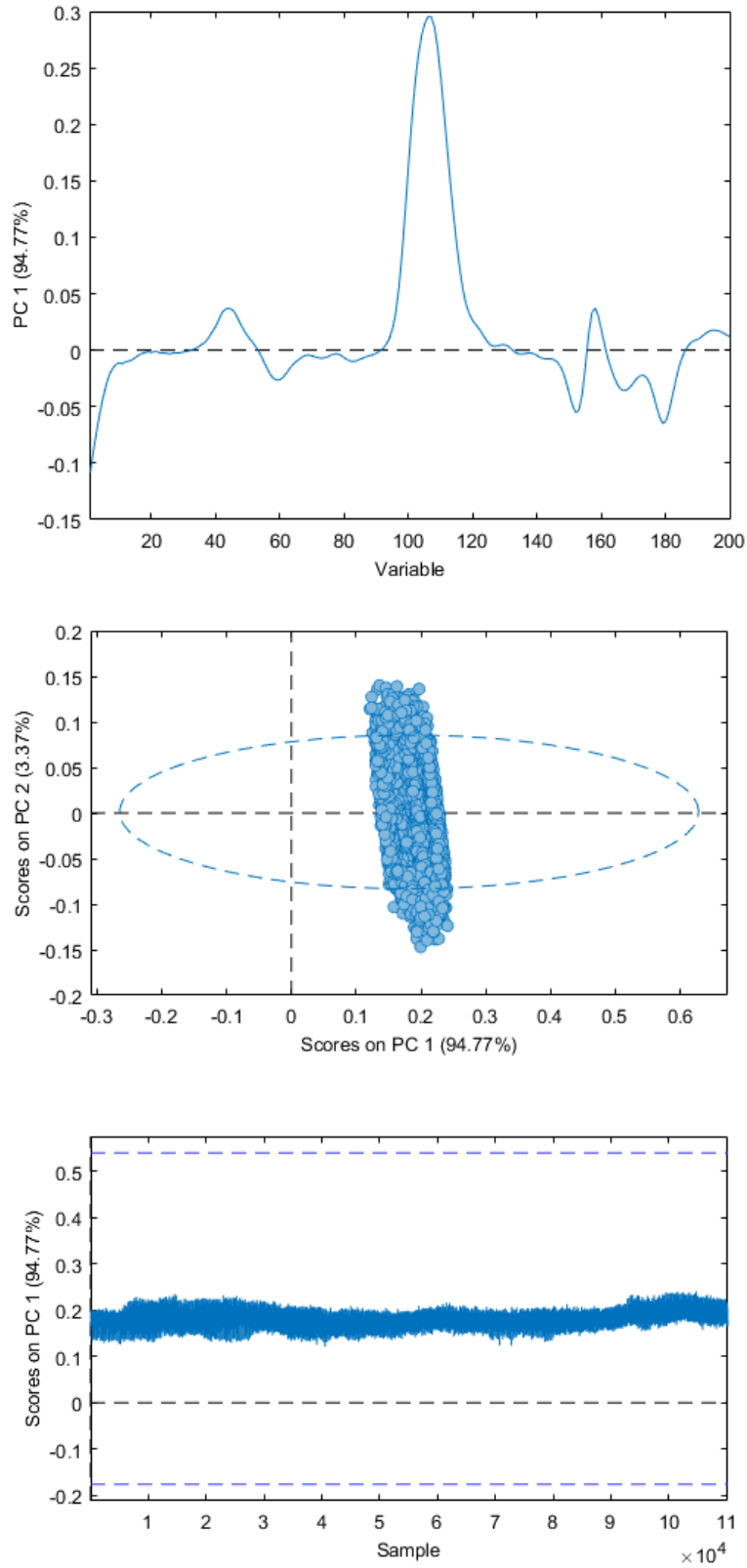


Figure E-2: Loading plot and scores for cucumber, infected #2

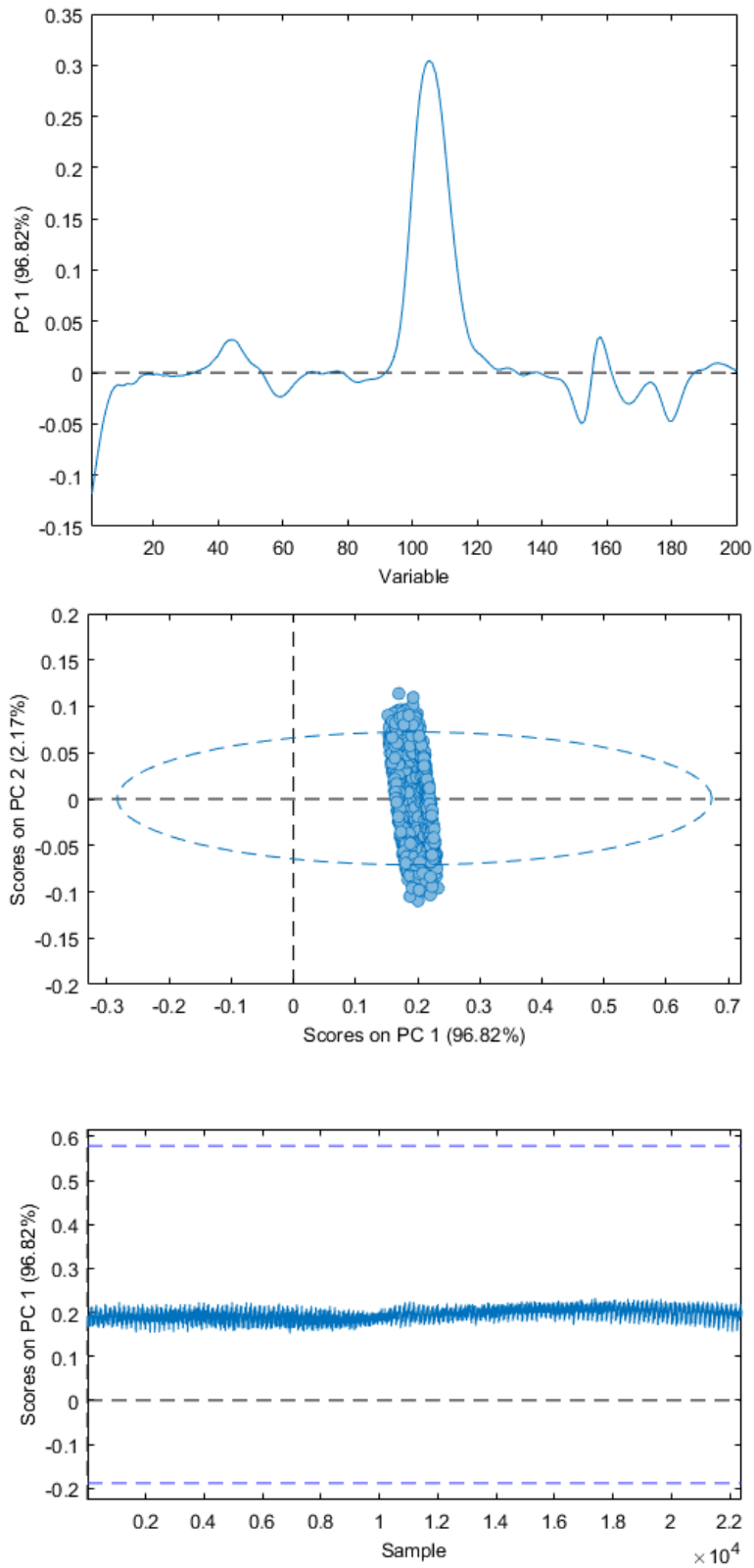


Figure E-3: Loading plot and scores for strawberry, healthy #1

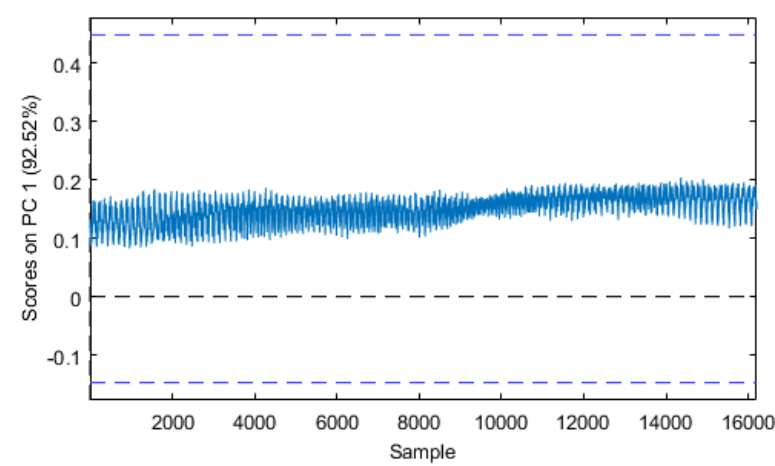
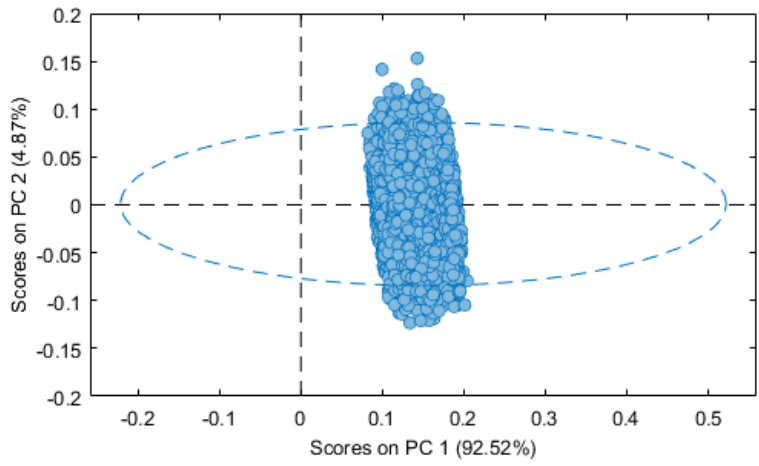
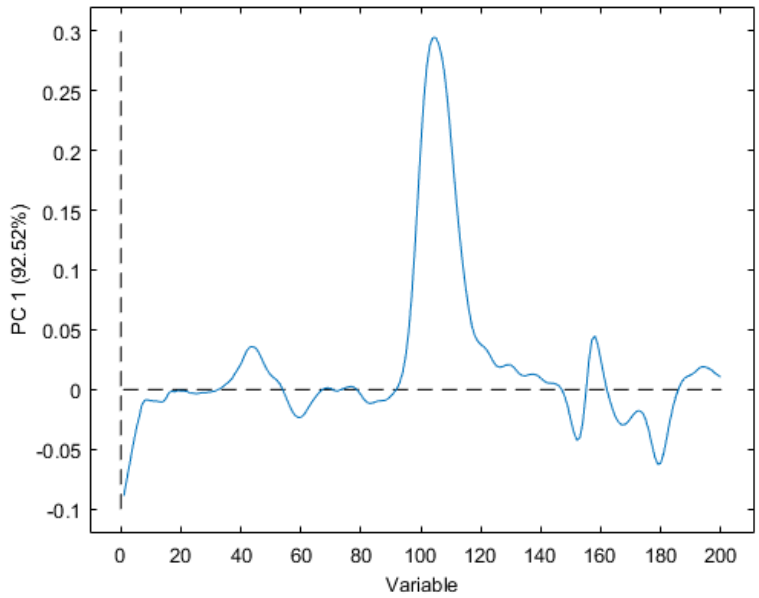


Figure E-4: Loading plot and scores for strawberry, infected #2

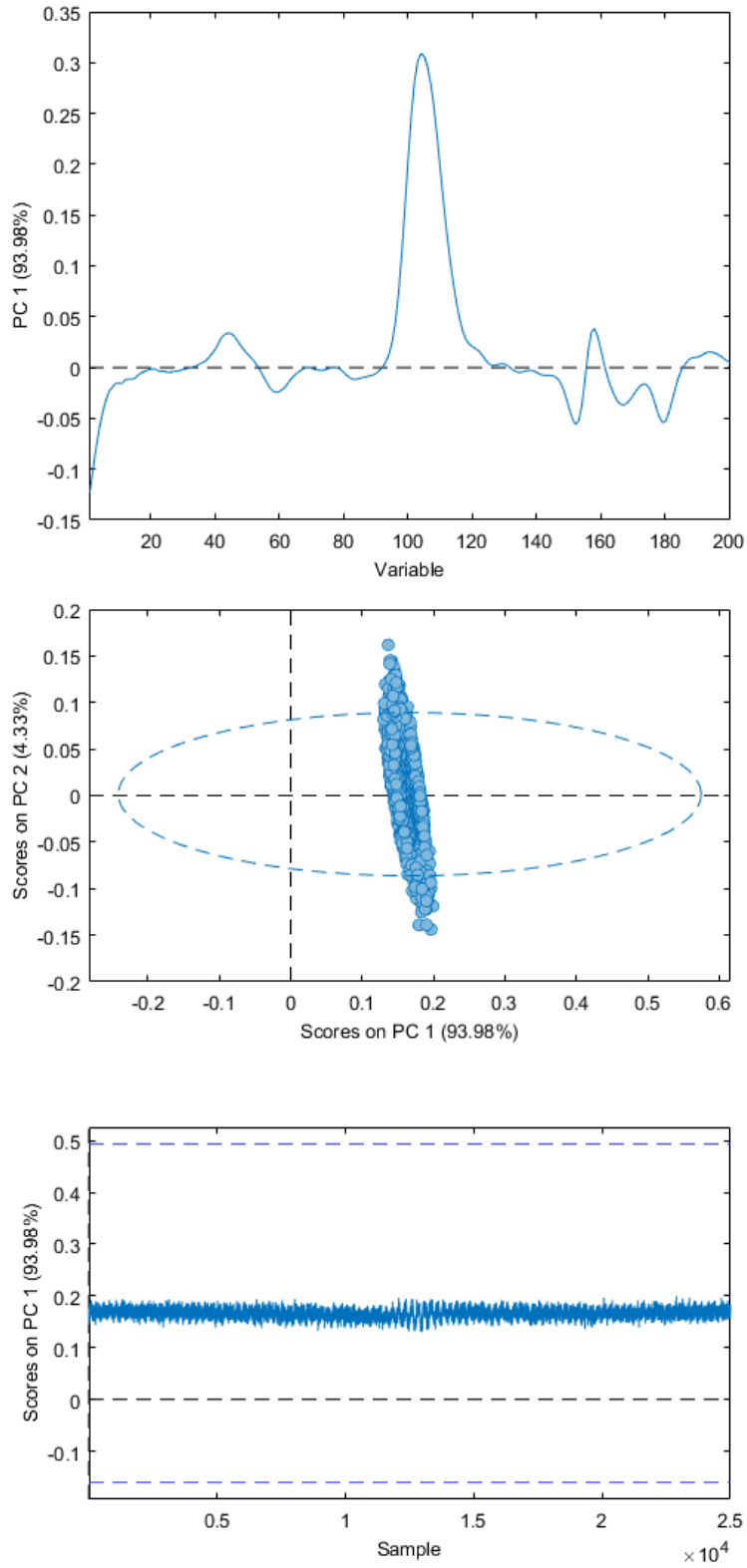


Figure E-5: Loading plot and scores for tomato, healthy #1

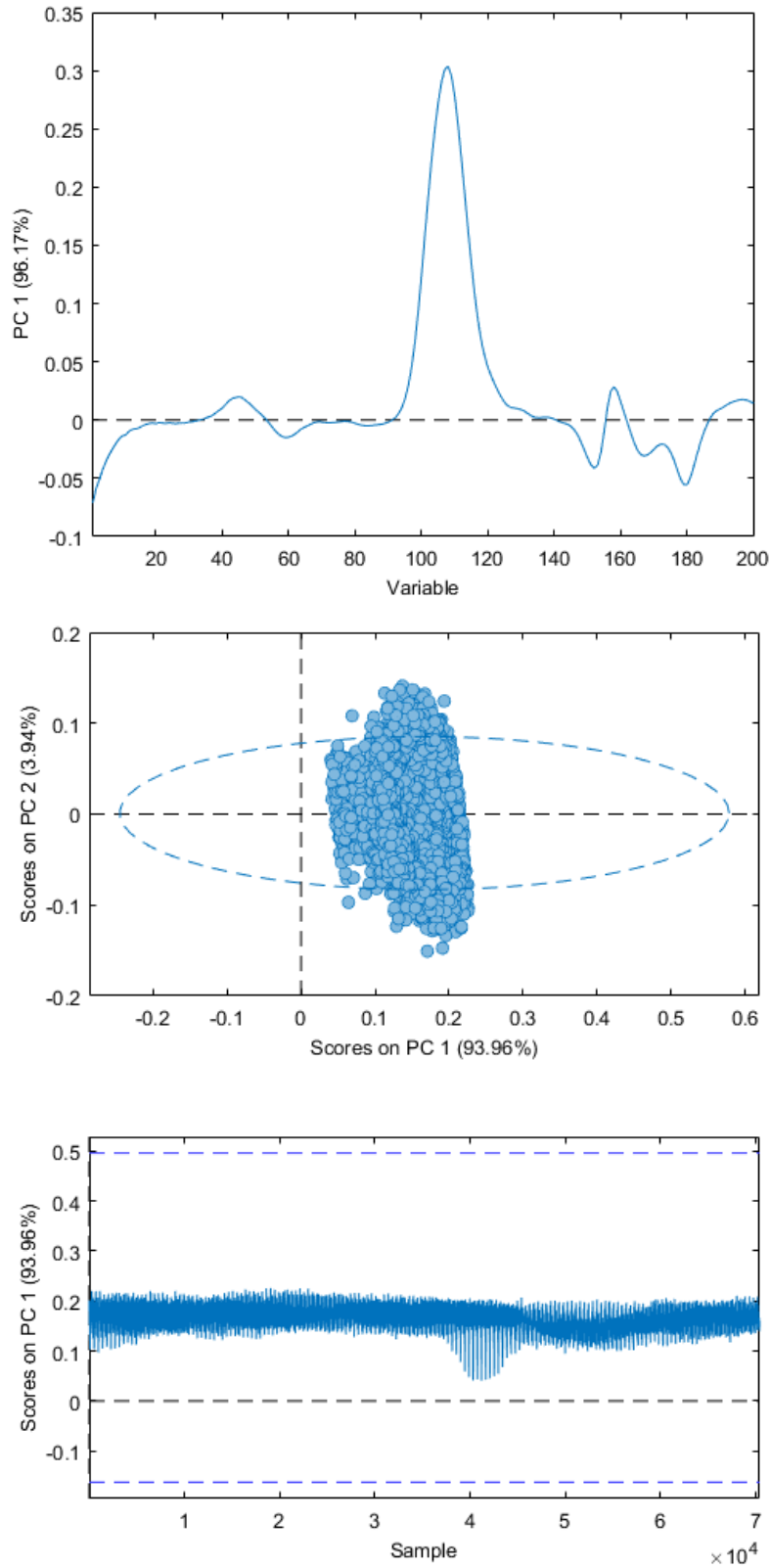


Figure E-6: Loading plot and scores for tomato, infected #2



Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway