

Ella Mahlamäki

**HYPERSPECTRAL IMAGING TO ESTIMATE
INTRINSIC VISCOSITY OF CELLULOSE-BASED
FABRICS**

Master of Science Thesis

Master of Science Thesis

Faculty of Engineering and Natural Sciences

Examiners: Res. Prof. Mikko Mäkelä

Prof. Juha Toivonen

January 2022

ABSTRACT

Ella Mahlamäki: Hyperspectral imaging to estimate intrinsic viscosity of cellulose-based fabrics
Master of Science Thesis
Tampere University
Science and engineering
January 2022

Production of cellulose and cotton fabrics from virgin materials requires large amounts of natural resources. Recycling and reusing waste textiles make the textile production more sustainable. Efficient recycling requires methods for waste textile sorting. Previously it has been shown that natural and synthetic fibers, and different cellulose fibers can be classified using near-infrared hyperspectral imaging and chemometrics. The object of this thesis was to classify cellulose-based textile samples based on intrinsic viscosity. The intrinsic viscosity value was known for each sample. The methods used in this thesis were kept simple and the results easily interpretable. Basic theory behind light and material interactions and molecule absorption process are presented to provide deeper understanding of the information acquired from hyperspectral images. In addition, principles of hyperspectral imaging and some exploratory and multivariate analysis methods are reviewed. Different methods and their effect on the classification results are discussed.

Hyperspectral camera used in this thesis operated in line-mode in near-infrared wavelength region. In the near-infrared region (1000-2500 nm) cellulose-based fabrics have characteristic interactions with light. Each material has unique chemical composition, and they can be identified from their unique absorption spectra. Absorption bands seen in the spectra are generated by energy level transitions in the molecules. In the infrared region absorption bands seen in the spectra are mainly caused by overtones and combination bands. Absorption in the infrared region happens if the molecular vibration changes the dipole moment of the molecule.

Differences between the sample average absorbance spectra were found with preprocessing and chemometric methods. Standard normal variate transformation with mean centering were shown to find the differences between intrinsic viscosities of the samples. The classification model was determined using average absorbance spectrum of all sample pixels of each sample. Treating each pixel separately created too much noise for accurate classification. Linear discriminant analysis with canonical variables were shown to produce promising classification results. Classification model created for this thesis correctly classified 100 % of the training samples (n=74) and 94.7 % of the test samples (n=38). Wavelength dimensions were reduced using principal component analysis to be able to conduct linear discriminant analysis. The number of used principal components were shown to have an important effect on the classification results.

The textile samples consisted of two different sets of textiles. These sets had different chemical compositions and despite the chemical differences they could be classified based on the intrinsic viscosities. Being able to sort cellulose textiles based on intrinsic viscosity helps to control the viscoelastic properties of dissolved fibers during chemical recycling.

Keywords: Hyperspectral imaging, Near-infrared, Textiles, Cellulose, Intrinsic viscosity, Classification

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Ella Mahlamäki: Hyperspektrikuvantamisen soveltaminen selluloosapohjaisten kankaiden sisäisen viskositeetin määrittämiseen

Diplomityö

Tampereen yliopisto

Teknis-luonnontieteellinen

Tammikuu 2022

Selluloosa- ja puuvillapohjaisten kankaiden valmistaminen alkuperäisistä raaka-aineista vaatii paljon luonnonvaroja. Tekstiilien tuotannon ympäristövaikutuksia saadaan pienennettyä kierrättämällä ja valmistamalla uusiomateriaaleja tekstiilijätteestä. Tekstiilien tehokas kierrättäminen vaatii tehokkaita lajittelumenetelmiä. Optisten kuvantamismenetelmien hyödyntäminen on osoittanut lupaavia tuloksia tekstiilien tunnistamisessa. Aikaisemmin on osoitettu, että luonnonkuidut pystytään erottamaan synteettisistä kuiduista ja eri selluloosapohjaiset kuidut toisistaan hyperspektrikuvantamisen ja kemometrinen menetelmien avulla. Tämän työn tavoitteena oli luokitella selluloosapohjaiset tekstiilinäytteet niiden sisäisten viskositeettiarvojen perusteella. Työssä hyödynnetyt menetelmät olivat yksinkertaisia, jolloin tulokset pysyivät helpommin tulkittavissa. Teoriaosiossa käsiteltiin valon ja materiaalien välisten vuorovaikutusten ja absorptio- taustalla olevaa teoriaa sekä hyperspektrikuvantamisen ja työssä hyödynnettyjen data-analyysimenetelmien periaatteita. Työssä käsiteltiin myös eri menetelmien vaikutuksia luokittelutuloksiin.

Valon ja materiaalien väliset vuorovaikutukset ovat monimutkaisia ja kullekin materiaalille ominaisia. Materiaalit voidaan tunnistaa niiden absorptiospektrien avulla. Tässä työssä tekstiilinäytteitä kuvattiin hyperspektrikameralla lähi-infrapuna-alueella (1000–2500 nm). Selluloosa-pohjaisilla kuiduilla on tällä aallonpituusalueella tunnistettavia absorptio-ominaisuuksia, jotka johtuvat pääosin vibraation overtone ja combination -transitioista.

Näytteiden keskiarvospektrien välillä havaittiin eroja esikäsittely- ja kemometrinen menetelmien avulla. Esikäsittelymenetelmät löysivät eroja keskiarvospektrien sisäisten viskositeettiarvojen väliltä. Työssä kehitetty luokittelumenetelmä luokitteli näytteet kolmeen eri sisäisen viskositeetin luokkaan hyödyntäen lineaarista diskriminanttianalyysiä ja kanonisia muuttujia. Luokittelumenetelmä luokitteli harjoitusnäytteistä oikein 100 % (n=74) ja testinäytteistä 94,7 % (n=38). Jokaiselle näytteelle laskettiin keskiarvospektri hyödyntäen jokaista näytteen pikseliä. Spektrien luokittelu pikselitasolla ei osoittautunut mahdolliseksi liiallisen kohinan takia. Aallonpituuksien dimensioiden lukumäärää vähennettiin pääkomponenttianalyysin avulla, jotta lineaarinen diskriminanttianalyysi voitiin toteuttaa. Käytettyjen pääkomponenttien lukumäärällä osoitettiin olevan huomattava vaikutus luokittelutuloksiin.

Tekstiilinäytteet sisälsivät kaksi näytejoukkoa, jotka erosivat toisistaan kemiallisesti. Siitä huolimatta näytteet pystyttiin luokittelemaan niiden sisäisen viskositeetin perusteella. Tekstiilikuitujen käyttäytymistä kemiallisen kierrätyksen aikana voidaan paremmin kontrolloida, kun tekstiilijäte pystytään lajittelemaan sisäisen viskositeetin perusteella.

Avainsanat: Hyperspektrikuvantaminen, Lähi-infrapuna, Tekstiilit, Selluloosa, Sisäinen viskositeetti, Luokittelu

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

PREFACE

I would like to thank all who have contributed to the completion of this thesis. This thesis has been written in Functional Cellulose Research Group at VTT. Firstly, I want to thank our research team leader Hannes Orelma and my supervisor Mikko Mäkelä for giving me the opportunity to write a meaningful thesis and opening the doors to the research world. Special thank you to Mikko for guidance, patience, support and for being inspiring supervisor during this project. Also, special thank you to my other supervisor Juha Toivonen for supervising this thesis. I would also like to thank Marja Rissanen and Inge Schlapp-Hackl from Aalto University Biorefineries Research Group at Department of Bioproducts and Biosystems for providing the textile samples and reference values.

In my mind master's thesis has been some big project to get over with in order to graduate. It ended up teaching me more than I expected. Getting the opportunity to work in a meaningful and interesting project, and being part of a team full of talented researchers, this project has given me more than just a finished thesis. Thanks to all coworkers, lunch and coffee company for making every day at work something to look forward to.

I must thank all the people and communities without whom I wouldn't be graduating. To Luukas, Matias and other squad members, thank you for carrying me through some of our courses and patiently helping me with exercises. To Sporrts team NMKSV, thank you for all the good times, peer support and help no matter what the problem. To all people I have got to know during my studies, thank you for making the years at Tampere University fun and memorable. Finally, to my parents and brother, thank you for all the support and always being there for me.

In Espoo, 31st January 2022

Ella Mahlamäki

CONTENTS

1.	Introduction	1
1.1	Cellulose fibers	1
1.2	Textile imaging applications.	2
2.	Hyperspectral imaging	4
2.1	Basic principles	5
2.2	Hyperspectral image acquisition	6
3.	Physical background	9
3.1	Electromagnetic spectrum	9
3.2	Material surface reflectance	10
3.3	Vibrational motion in molecules	11
3.4	Quantum mechanical harmonic and anharmonic oscillator	12
3.5	Infrared activity and absorption	16
4.	Multivariate image analysis.	18
4.1	Calibration and correction	18
4.2	Chemometrics	20
4.3	Preprocessing	20
4.3.1	Standard normal variate transformation and mean centering	21
4.3.2	Derivative methods	22
4.4	Exploratory data analysis and classification	23
4.4.1	Principal Component Analysis	23
4.4.2	Linear Discriminant Analysis	26
4.4.3	Wavelength selection and cross-validation	28
5.	Experimental methods	30
5.1	Textile samples	30
5.2	Image acquisition.	31
5.3	Image processing	32
5.4	Differences between sample spectra	33

5.5	Sample classification	33
6.	Results	36
6.1	Image calibration and preprocessing	36
6.2	Differences between sample sets	38
6.3	Sample classification performance	39
7.	Discussion	43
7.1	Textile samples	43
7.2	Preprocessing methods	45
7.3	Pixels vs object	46
7.4	The number of used principal components	48
7.5	Wavelength selection	49
8.	Conclusion	52
	References	55

LIST OF FIGURES

1.1	Chemical structure of cellulose and repeating cellobiose units. [7]	2
2.1	Hyperspectral data cube.	5
2.2	Hyperspectral imaging techniques whiskbroom, pushbroom and staring. . .	6
2.3	Structure of line-scan hyperspectral camera. Modified from [20]	7
3.1	A) Mirror, diffuse and glossy reflectance models and B) geometry of bidirectional reflectance distribution function coordinates.	11
3.2	Directions of molecular vibrations for A) non-linear H_2O molecule and B) linear CO_2 molecule.	12
3.3	Vibrational motion of diatomic molecule. Modified from [22]	13
3.4	Potential energy curve and energy levels for diatomic anharmonic and harmonic oscillator. Modified from [22]	15
4.1	Directions of standard normal variate transformation and mean centering for data matrix X	21
4.2	The effect of adding 1st and 2nd order derivation on sinusoidal signal. . . .	22
4.3	PCA model for a $I \times J$ data matrix X	23
4.4	Principal component analysis visualization on random set of multivariate normal distributed points.	24
4.5	Classification of an unknown sample based on a linear class boundary. . .	27
4.6	N-fold cross-validation visualization.	29
5.1	Specim SWIR 3 hyperspectral camera (left), sample imaging set up (center) and samples imaged to set the moving table speed and the field of view (right).	31
5.2	Hyperspectral image processing workflow.	32
5.3	Sample pixel separation from the background and the calculated respective average absorbance spectrum.	33

5.4	Classification model creation and validation workflow.	34
6.1	A) Hyperspectral image of one textile sample and B) illustration of separating sample pixels from the background.	36
6.2	Raw average signals for one imaged sample.	37
6.3	Reflectance transformed average sample spectra and -log10 transformed average absorbance spectra of the samples in the first sample set.	37
6.4	SNV transformed and mean centered average absorbance spectra of the first sample set. Spectra colored according to measured intrinsic viscosity values.	38
6.5	Score plots of first two principal components showing the differences of the two sample sets.	39
6.6	Principal component loading plots of the first two principal components calculated for the two sample sets.	39
6.7	Classification results of the training sample set.	40
6.8	Classification results of the test sample set.	41
7.1	Effect of applying 1st and 2nd order Savitsky-Golay derivation on the SNV transformed and mean centered average absorbance spectra on the first set of samples.	45
7.2	Calculation of average absorbance spectrum of chosen sample ROI.	46
7.3	Effect of changing the ROI size on the sample average spectra of the first sample set.	47
7.4	Effect of changing the number of used principal components on the classification results.	49
7.5	Wavelength selection results.	50

LIST OF TABLES

3.1	Absorption band intensities in mid-IR and NIR regions. [32]	17
3.2	Examples of absorption band assignments of cellulose and water. [32] [9]	17
4.1	Confusion matrix example.	27
5.1	Sample sets used in the analysis and their intrinsic viscosity regions.	30
5.2	Intrinsic viscosities of the classification classes.	34
6.1	Number of samples in each class in training set, test set and overall.	40
6.2	Confusion matrix of the training set classification results. Correctly classified rate 100 %.	41
6.3	Confusion matrix of the test sample set classification results. Correctly classified rate 94.7 %.	42
6.4	Classification results for each class, training and test sets as correctly classified percentage %CC.	42
7.1	Effect of different preprocessing methods on the classification results of the training sample set. Mean centering used as last method in all combinations. Results shown as correctly classified percentage %CC.	46
7.2	Effect of changing the ROI size on the training sample set and training set 3-fold cross-validation classification results. Results shown as correctly classified percentage CC%.	48
7.3	Effect of changing the number of used principal components in the classification model. Classification results shown as correctly classified percentage CC%.	49
7.4	Effect of wavelength selection on the classification results of the training sample set. Results are shown as correctly classified percentage %CC.	51

ABBREVIATIONS

%CC	Correctly Classified percentage
BRDF	Bidirectional Reflectance Distribution Function
CCR	Correctly Classified rate
CV	Canonical Variable
CVA	Canonical Variate Analysis
IR	Infrared
LDA	Linear Discriminant Analysis
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares
QDA	Quadratic Discriminant Analysis
ROI	Region of Interest
SNV	Standard Normal Variate
SVD	Singular Value Decomposition

1. INTRODUCTION

The demand for new textiles is constantly growing and sustainable textile fiber production is currently not able to meet the needs. The production of all textile fibers has doubled to 111 million tons within ten years in 2000-2019. [1] 60 % of global textile consumption is due to clothing and only 12 % of those textiles is recycled. Rest of the waste textiles are currently incinerated or landfilled. [2] In 2019 polyester, cotton and cellulose were globally the most produced textile fibers. In total 57.7, 25.7 and 7.1 million tons, respectively. [1] Production of cotton and cellulose -based textiles has large ecological footprint. Recycling and reusing textiles could reduce the environmental impact. [3] The European Commission has set a directive, which states that all EU Member States are obligated to arrange collection of textile waste from the beginning of 2025. This directive will increase the number of textiles to be recycled. [4]

Efficient textile waste recycling process and production of reused textiles requires methods for textile identification. Optical methods have shown promising results that different textile fibers can be identified based on the differences in their chemical structures. Identified cellulose-based materials can be chemically recycled and re-spun into new fibers [5]. The recycling process would be easier to control if the textiles are sorted according to their specific properties.

1.1 Cellulose fibers

Cellulose is the most abundant organic polymeric material in the world and most of it is produced in plants. Other origins for cellulose are, for example, marine animals, fungi and bacteria. The advantages of cellulose as a clothing material has been discovered already thousands of years ago. [6]

Cellulose consists of several hundreds to thousands of β -D-glucopyranose units linked to each other with 1,4-glycosidic linkages. Cellulose polymer chain is linear, and each glucan ring is folded to arm-chair configuration. Each hydroxyl group (O-H) is in equatorial position and methine groups (C-H) in axial positions. Each adjacent glucopyranose unit rotation to each other is 180° . The repeating unit in cellulose polymer chain is called cellobiose. Structure of the repeating cellobiose unit is shown in Figure 1.1.

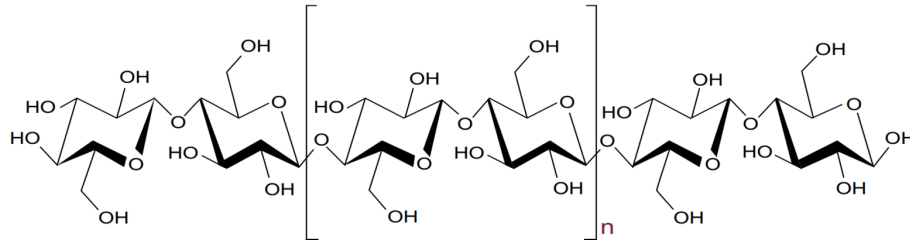


Figure 1.1. Chemical structure of cellulose and repeating cellobiose units. [7]

The degree of polymerization of cellulose polymer depends on how many repeating cellobiose units there are in each fiber. Regenerated cellulose fibers are produced from cellulose solution. The degree of polymerization and the molecular weight influence the viscoelastic properties of the cellulose fiber solution and the viscosity of the solution affects the spinning behavior in the regeneration process. Chemical recyclers are interested in sorting the intrinsic viscosity region of 350-550 ml/g out of other textile samples. Controlling the spinning behavior of the cellulose solution enables also the control of the strength-properties of the regenerated cellulose fiber. [8]

Cellulose is an organic material containing only C, O and H atoms. When light interacts with cellulose, these atoms provide most characteristic information in near-infrared (NIR) wavelength region. [9] Near-infrared range covers the wavelength region 1000-2500 nm.

1.2 Textile imaging applications

Many applications have been developed for textile material identification in near-infrared region. Applications on different wavelength regions provide different information of the sample composition. In the NIR region the molecular absorptivity is lower and not all the light is absorbed at the very surface molecules of the materials. Organic materials have characteristic interactions with light in the NIR region. That is why most of the textile identification applications use the wavelength region of around 1000-2500 nm. Also

visible near infrared region 450-950 nm shows promising results in pure and blend textile identification. [10]

Previous studies have shown promising results in fiber identification [2] [10], especially in discriminating natural and synthetic fibers, and in textile surface finishing property determination [11]. Many of the studies aimed to identify different textile blends. [2] [12] [13] It has been shown that synthetic and natural textile fibers can be identified with many different methods. Going into smaller textile properties, Mäkelä et. al classified different cellulose fibers using hyperspectral imaging. The results seemed promising and for example cotton, viscose, lyocell and modal were classified correctly. [14] Hyperspectral imaging in NIR wavelength region and usage of chemometric methods have shown potential in textile discrimination. [15]

Near-infrared spectroscopy is not the only method to analyse the composition of textiles. NIR methods are limited to measure the surface material properties of the samples. Like NIR spectroscopy, nuclear magnetic resonance (NMR) spectroscopy does not have to deal with differences in textile colors, changing particle sizes or sample surface structures. Predicting material compositions, for example, of multilayer textiles, NMR provides more accurate results. [12]

The interactions between light and materials are complex and hard to interpret. The desired information needs to be extracted from the spectroscopic data to estimate the chemical properties of the samples. Chemometric methods are widely used in analysing multivariate spectroscopic data. Principal component analysis and partial least squares regression are the most used methods in textile identification found in literature. Preprocessing methods vary in different studies and applications, but standard normal variate transformation and mean centering are found in almost all. [2] [11] [13]

The objective of this thesis was to present the process of taking hyperspectral images, analysing and processing the images to create simple classification model from theory to practice. The theory of light and material interactions are also discussed. The aim was to determine the potential of classifying cellulose-based textiles based on the measured intrinsic viscosities.

2. HYPERSPECTRAL IMAGING

Human eye can see electromagnetic radiation only in the visible light wavelength region. When electromagnetic radiation interacts with materials, many interactions happen in wavelength regions outside of human vision. Studying the interactions between electromagnetic radiation and matter, materials can be identified, different objects detected or even some processes recognized. This field of science is called spectroscopy. Spectroscopy measures and studies the spectrum of materials. Spectrum includes the information of how much radiation the material has emitted, reflected or transmitted based on the principle known as Beer-Lambert law [16]. Spectroscopy has many applications from macro to atomic scale structure and compositional analysis. [17] More of the physical background of spectroscopic interactions is reviewed in Chapter 3.

Many different methods and devices can be used to gather the spectroscopic information. Wide range of applications provides also wide range of spectroscopic devices. Hyperspectral imaging is a useful method to measure spectroscopic information and hyperspectral cameras can operate in different wavelength regions. Hyperspectral image consists of spatial and spectral information as spectrum is imaged for each pixel. This provides large amount of information and creates a wide range of analysis possibilities. Handling hyperspectral images in the molecular level makes it possible to recognize minor changes in the sample composition. [18] Hyperspectral imaging can also be called chemical imaging as the aim is to gain information of the chemical composition of the samples. [17]

Hyperspectral imaging methods can be in-line, on-line, at-line or off-line. These methods vary in where the camera is placed and where the images are taken during the process. In in-line methods the camera is part of the process flow, and the samples are imaged so that the flow is not disturbed. On-line methods also have the camera as a part of the process flow, but the samples are guided one at a time through the imaging device. With

at-line methods, the samples are removed from the process flow and images are taken near the other process devices. Off-line methods remove the samples from the flow and takes the samples to other environment where the images are taken. [17]

2.1 Basic principles

Traditional digital camera gathers data of three wavelength channels, red, green and blue, and combines them to form digital color images. In hyperspectral image, every pixel has information of the continuous spectrum in the imaged wavelength region. The spectra of each pixel are stacked to form hyperspectral data cube. Each hyperspectral data cube contains spatial coordinates and spectrum of each pixel. In other words, hyperspectral imaging combines digital imaging and spectroscopy. [9] Formation of the data cube is visualized in the Figure 2.1. The spatial dimensions of the image pixels are in x and y axes and the spectral dimension in the λ axis.

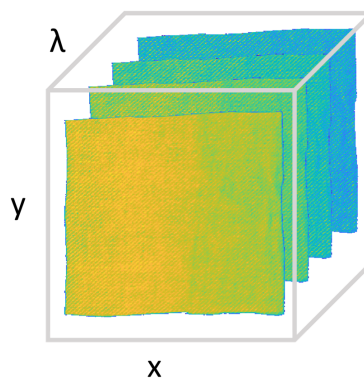


Figure 2.1. Hyperspectral data cube. Spatial information located in x and y axes, and spectral information in λ axis.

The measured spectrum can be used to identify materials based on the spectral fingerprints. This means that each material interacts differently with electromagnetic radiation and each material has unique spectrum. Measured spectra can be used to identify different areas or, for example, to identify material compounds in the sample. Hyperspectral imaging methods are increasingly used in various areas such as remote sensing, agriculture, medical diagnostics and textiles. [15]

Near-infrared hyperspectral imaging is a useful tool to analyse chemical compositions of organic materials. In NIR region light and organic compounds interacts characteristically. Imaging in NIR region is costly, because imaging higher NIR wavelengths (1100-2500

nm) sets some requirements for the detectors. Silicon-based detectors are suitable option for applications measuring wavelengths of 1100 nm or lower. More costly options such as indium gallium arsenide (InGaAs) or mercury cadmium telluride (HgCdTe) based detectors are required when the wavelength range rises above 1100 nm. [9]

2.2 Hyperspectral image acquisition

There are different ways to measure hyperspectral images. Most used methods are pushbroom, whiskbroom and staring imaging. Whiskbroom imager measures each pixel separately, pushbroom imager one line of pixels simultaneously and staring imager all the pixels at one wavelength at a time. Figure 2.2 illustrates the principles of these imaging techniques.

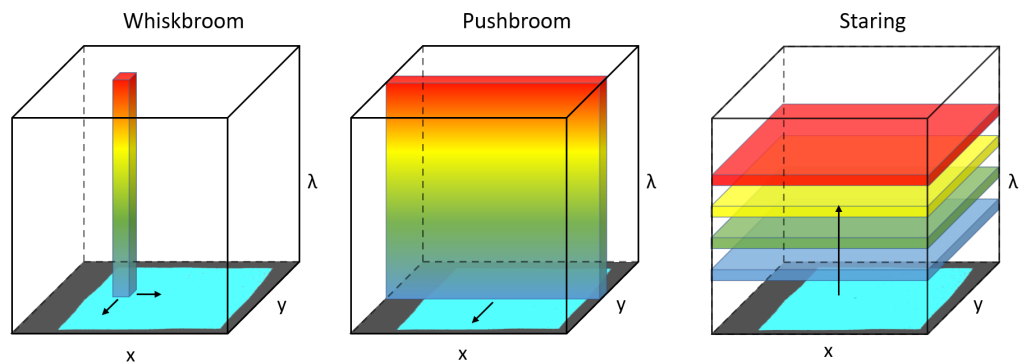


Figure 2.2. Hyperspectral imaging techniques whiskbroom, pushbroom and staring.

Device using whiskbroom imager scans the sample point-by-point with usually one single detector. One pixel at a time is imaged in x-y-plane. Main benefits for whiskbroom technology are high spectral resolution and high discriminating power between similar objects in samples. [17] The variation between sample sizes and spectral ranges that can be imaged with whiskbroom imagers is wide. However, because of the point-by-point scanning, such technology is time-consuming, and the images have low spatial resolution [9]. Usually, this technology is found from off-line applications. [17]

With pushbroom imaging systems all the spectral information of each line is acquired at the same time. Pushbroom systems scan the whole line of pixels at a time while usually the sample moves below the detectors. [9] This technique is also called line-scanning. [18] Taking images this way takes only few seconds and the compromise between the spectral and spatial resolutions is optimal. [9] Line-scan technology is the most used

hyperspectral imaging technique, because the structure of line-scan system allows imaging wide range of wavelength ranges and sample sizes. The applications for line-scan technology are on-line and in-line. [17] Samples moving below the detector in an in-line system makes real-time analysis possible. [9]

Staring imager takes the whole image of one wavelength at a time. [9] To get all the spectral data, wavelength selection is required. The most used method to do the selection is to use a rotating wheel in front of the detector with bandpass filters. [17] Filter wheel in the optical path determines the number of imaged wavelengths. The wavelengths chosen to the wheel must be determined with prior knowledge of the spectral behavior of the sample. [18] This provides high spatial resolution, but low spectral resolution. [9] This technique is also called tunable filter scanning, wavelength scanning or focal plane scanning. [18] The number of applications for staring imagers is narrow, because the sample needs to be stationary during the imaging. [17] That is why there are only at-line or off-line applications.

Hyperspectral camera commonly consists of an objective, a spectrograph and a grayscale camera. These components need to be optimized for each technique and used wavelength range. In line-scan cameras, spectrograph consists of slit, collimating and dispersive optics, and focusing lens. [19] The structure of a line-scan hyperspectral camera is illustrated in Figure 2.3.

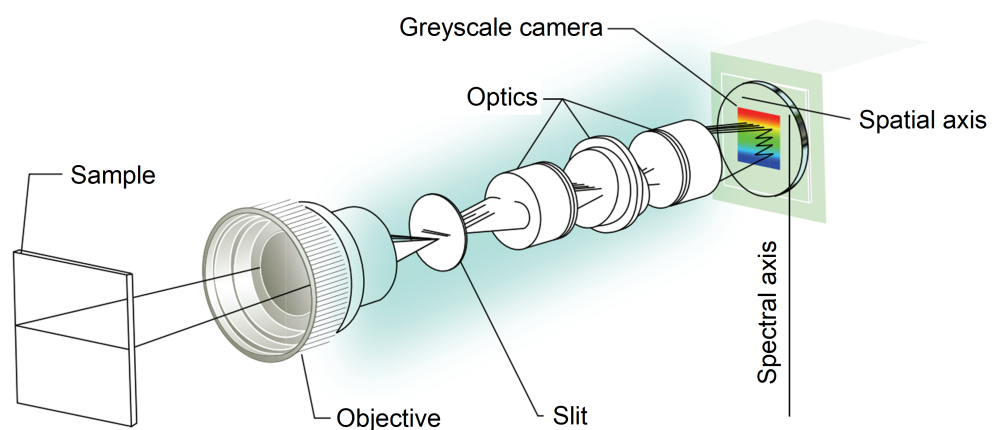


Figure 2.3. Structure of line-scan hyperspectral camera. Modified from [20]

The slit collects the incoming information one line at a time, and the width of the entrance slit determines the accuracy of the spectra. Narrower slit provides more accurate spectra. After the slit, light goes through collimating optics, which directs the light to the dispersive optics where light is dispersed into spectra. Dispersed spectral information is directed to lens, which focuses the image to grayscale camera. Grayscale camera measures the intensities of each wavelength in the coming spectra and gathers the information of each spectrum in each pixel in the line. Each line is imaged similarly as sample moves below the camera. [19]

Hyperspectral imaging set ups also includes a source of illumination. Correct illumination sources and angles of detection have crucial influence on the image quality. Some photons detected in the spectrum are specularly reflected from the sample surface and, depending on the application, do not carry relevant chemical information. Specularly reflected light may overlap and cover the spectral information of the chemical compounds of interest. Photon diffusion might cause spatial artefacts and shift the spatial information to another spatial region. Diffuse illumination, detection and correct image preprocessing methods can be used to avoid these spectral and optical artefacts. Also, the angles of the illumination source and detector needs to be set optimally to reduce these effects and to obtain reliable imaging results. [17] Hyperspectral imaging has great potential in textile imaging applications. The development of cost-effective and faster hyperspectral cameras has made the technology already available for industrial use. [10]

3. PHYSICAL BACKGROUND

When light hits material it interacts with it either via absorption, reflection, scattering or transmission. The electromagnetic spectrum of light consists of different frequencies, wavelengths and photon energies of electromagnetic radiation. Interactions with material changes these properties in the spectrum and the changes can be analysed. By studying the behavior of material interactions with light, we can get information of the chemical and physical composition of the material. Vibrational spectroscopy explains the reasons for the changes in the absorption spectrum at infrared wavelengths. Infrared (IR) and Raman are linear vibrational spectroscopic methods widely used for analysing for example cellulose and other organic materials. [21]

3.1 Electromagnetic spectrum

Light, visible and non-visible, is electromagnetic radiation. Important unit, when studying light, is photon. Photon is the quantum of light and one of the elementary particles. Photon has energy that depends on the wavelength of the light according to equation

$$E = hf = h\frac{c}{\lambda}, \quad (3.1)$$

where h is Planck's constant ($h = 6.6 * 10^{-34} Js$), f is the frequency, c is the speed of light and λ is the wavelength of light. [22]

Molecule can absorb a photon and move to higher energy level, if the photon energy matches the energy level difference in the molecule. The absorbed photons and energy level changes can be seen in the absorption spectrum. [21] Process where energy level is changed can also go to other direction where photon is emitted, and molecule moves to lower energy level. [23]

The absorption of photons also depends on the chemical composition of the absorbing molecule. Each molecule has unique structure and consequently unique absorption spectrum. The intensity or the "strength" of the absorption seen in the spectrum depends on the concentration of the absorbing component in the sample through the Beer-Lambert law. The equation for Beer-Lambert law is

$$A = \epsilon lc = \log_{10}\left(\frac{I_0}{I}\right), \quad (3.2)$$

where A is the absorbance, ϵ the absorptivity, l optical path length in cm, c the concentration of the absorbing component, I_0 is the incident intensity of light and I the transmitted intensity. [21] If optical path length and absorptivity are held constant, absorbance is directly related to the concentration of the absorbing component.

Estimating sample concentrations using Beer-Lambert law is easier if different components show their own characteristic bands in the spectrum without interferences of other components. In real life different components behave similarly and they are hard to identify from the spectrum. [21] Spectroscopic applications usually measures the transmittance or reflectance spectrum to get the information of the absorption properties of the samples.

3.2 Material surface reflectance

When light interacts with material, the reflected light can be approximated to be the incident light subtracted with the absorbed and transmitted light. Term reflectance refers to the fraction of the incident light that is reflected. Imaging applications for object recognition usually consists of light source, object that reflects the incident light and camera that measures the reflected light. Light source is directed to the object in desired angle and the reflected light, and the reflection angle are measured.

There are different possibilities for light to reflect from a material surface. In perfectly diffuse reflection the reflected radiance is constant in all directions. If the incident light is only reflected to direction with same incident and reflection angle, it is called perfectly specular reflectance or mirror reflectance. In real life specular reflectors reflects the light also in other directions. This is called glossy reflectance. [24] The different reflectance

models are illustrated in Figure 3.1A. The angle and the intensity of the reflected light depends on the composition of the illuminated material, and the wavelength and the angle of the incident light. [25] Reflected light is usually detected in predefined angles to get desired information of the material. [24]

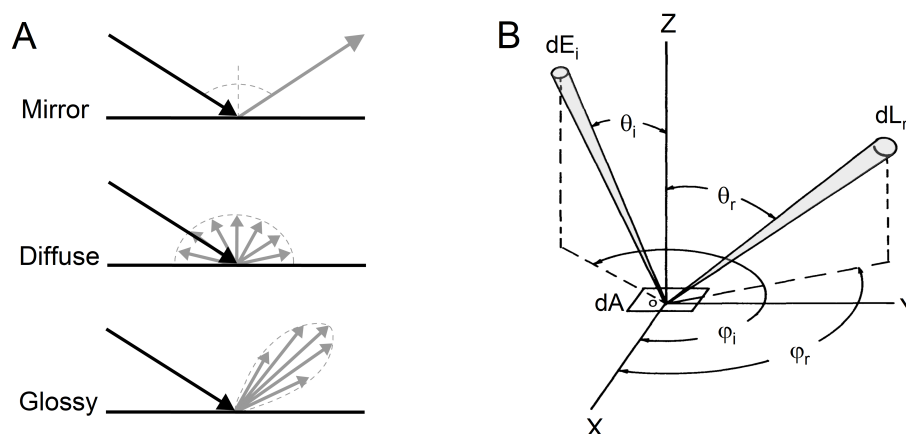


Figure 3.1. A) Mirror, diffuse and glossy reflectance models and B) geometry of bidirectional reflectance distribution function coordinates. B modified from [24].

Bidirectional reflectance distribution function (BRDF) can be used to describe how light is reflected from material surface. [26] The incident angle of the light coming from the light source towards the illuminated material is denoted as (θ_i, ϕ_i) and the reflection angle (θ_r, ϕ_r) . [24] Spectral BRDF can be calculated using equation

$$f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{dL_r(\theta_r, \phi_r)}{dE_i(\theta_i, \phi_i)}, \quad (3.3)$$

where $dL_r(\theta_r, \phi_r)$ is the reflected radiance and $dE_i(\theta_i, \phi_i)$ the incident irradiance. [27] The directions and angles of the reflected radiance and incident irradiance in BRDF is illustrated in Figure 3.1B. In the figure, dA is the illuminated area of the material. The reflectance spectra are affected by the material composition and absorption properties. Absorption of the light causes the molecule to move to higher energy level, which excites the vibrational motion in the molecule.

3.3 Vibrational motion in molecules

Molecules in matter are in constant motion and the motion is due to the characteristic vibrations of molecules. [22] The amplitude of the movement correlates with temperature,

but does not reach absolute zero, even at absolute zero temperature. The motion is not possible to observe in real time and if it was it would seem random. [23] There are possible places to "find" the atoms at a time though and these places can be described as "normal modes of vibration". [22]

Molecules can be described as systems, and they have a specific number of directions where they can vibrate or move freely. These are called the degrees of freedom. Degrees of freedom of the molecule depends on the number of atoms N and the linearity of the molecule. Molecules with N atoms have $3N$ degrees of freedom of motion to x , y and z directions. In addition to motional freedom there are vibrational degrees of freedom that change the distance and angles between the atoms. For non-linear molecules the degrees of freedom are $3N - 6$ and for linear $3N - 5$. [22] For linear molecule $3N - 5$ is because of two moments of inertia. Rotation along longitudinal axis is considered as zero moment of inertia. [23] Molecular vibration directions are shown in Figure 3.2 for non-linear H_2O molecule and linear CO_2 molecule. These characteristic vibrations are responsible for the characteristic bands in the absorption spectra. [22]

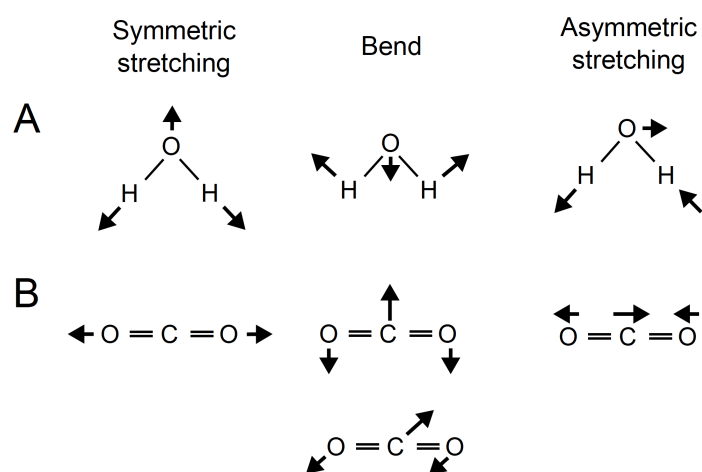


Figure 3.2. Directions of molecular vibrations for A) non-linear H_2O molecule and B) linear CO_2 molecule.

3.4 Quantum mechanical harmonic and anharmonic oscillator

To understand these molecular vibrations better, let's consider classical mechanics model of harmonic oscillator. Atoms in molecules can be described as masses connected with

a spring that follows Hook's law

$$F = -kx, \quad (3.4)$$

where F is the restoring force of the spring, k is the spring constant, or the bond stiffness and x is the length of the spring. [21] [23] Diatomic system with two masses m_1 and m_2 is illustrated in Figure 3.3. Vectors x_1 and x_2 are the harmonic oscillation displacements from the equilibrium state.

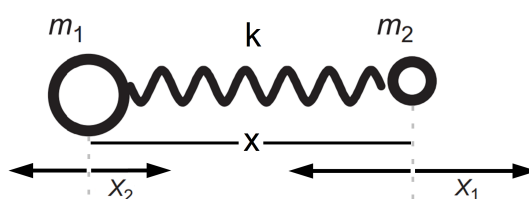


Figure 3.3. Vibrational motion of diatomic molecule. Modified from [22]

The mass of diatomic system can also be expressed as reduced mass

$$m = \frac{m_1 m_2}{m_1 + m_2}. \quad (3.5)$$

When this system vibrates the atoms are displaced from equilibrium state with x_1 and x_2 . The two atoms have same frequency, and they pass their equilibrium states at the same time. For diatomic molecule the classical vibrational frequency can be calculated from equation [22]

$$\nu = \frac{1}{2\pi} \sqrt{k \left(\frac{1}{m_1} + \frac{1}{m_2} \right)} = \frac{1}{2\pi} \sqrt{\frac{k}{m}}. \quad (3.6)$$

This is the frequency where the IR absorption peak can be observed. [21] As seen from the Equation 3.6, the vibrational frequency of diatomic harmonic oscillator depends on the force constant k and the masses of the atoms in the molecule. The force constant is affected by the bond energy between the two atoms. [22] For larger molecules the vibrations become more complex, and these calculations are no longer valid. Each molecule has unique structure and vibrations, so the absorption of photons is different in each molecule. [23]

Classical mechanics explains the normal modes of molecular vibrations and their frequency, but it does not explain the transitions between the states that are the observations in IR and Raman spectroscopy. [23] Quantum mechanical approach is used to explain the transitions. Quantum mechanics states that molecules have only quantized energy states, which leads to the conclusion that vibrational energy has discrete values. [22]

Spectroscopy studies the transitions between energy states and the harmonic transitions can happen only under certain conditions ($\Delta v = \pm 1$, where v is the quantum number). The quantum mechanical harmonic oscillator energy levels have potential energies

$$E_i = (v_i + \frac{1}{2})h\nu \quad v_i = 0, 1, 2, \dots, \quad (3.7)$$

where ν is the vibrational frequency of the oscillator and v is the quantum number. [22] According to classical mechanics and Equation 3.7, the position and the momentum of particle in motion can be observed simultaneously. [23] Quantum mechanical Heisenberg's uncertainty principle states that we cannot be certain to say where the particle is at a time, so we need to use probabilities of finding a particle. [22]

Electrical and mechanical behavior is not harmonic in real atoms. In real life molecules behave more like anharmonic oscillators. Anharmonic oscillation happens when something disturbs the harmonicity. Harmonic oscillator allows transitions only between fundamental levels where ($\Delta v = \pm 1$) and for the anharmonic oscillator also overtones and combination bands ($\Delta v = \pm 2, \pm 3, \dots$) are possible. The potential energy function needs anharmonic correction

$$E_v = h\nu_e(v_i + \frac{1}{2}) - h\chi_e\nu_e(v_i + \frac{1}{2})^2, \quad (3.8)$$

where the term $h\chi_e$ defines the amount of anharmonicity in the oscillator. [22]

Electrical anharmonicity happens when the change in the dipole moment and polarizability are not linearly proportional to the nuclear displacement coordinate. Mechanical behavior of real molecules follows Hooke's law only when internuclear distance r is not very different from the equilibrium state r_e . The molecule dissociates when r is large and the atoms in the molecule do not influence each other anymore. That is why potential en-

ergy curve goes to zero at dissociation energy D_e . The repulsion of the two nuclei makes the potential curve slope steeper at small r compared to harmonic oscillator curve. [28] Both types of anharmonicity usually have an effect on vibrational overtone intensities. Figure 3.4 shows the potential energy curve for diatomic molecule that behaves as an anharmonic oscillator. Anharmonicity makes the energy level differences smaller. Dashed line in the figure shows the difference to harmonic oscillator.

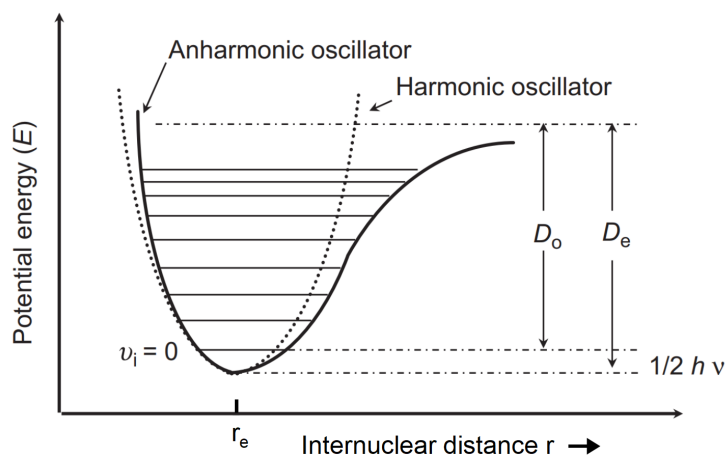


Figure 3.4. Potential energy curve and energy levels for diatomic anharmonic and harmonic oscillator. Modified from [22]

Overtone band occurs when the transition happens from the ground state to the second or to a higher excited state. The transition from $v = 0$ to $v = 2$ is called the first overtone and $v = 0$ to $v = 3$ the second overtone, and so on. Combination band occurs if more than one fundamental vibration is excited at the same time. The intensities of overtones and combination bands are weaker than of fundamental vibrations. [23] In addition to overtones and combination bands, one phenomenon that can happen for vibrational levels is called Fermi resonance. Fermi resonance happens when a fundamental band interacts with overtone or combination band and produces two quite strong bands. [22] Differences between the IR and the NIR region are that in IR region the vibrational transitions are mainly due to fundamental vibrations and in NIR due overtones and combination modes. [29]

3.5 Infrared activity and absorption

In the infrared region absorption of IR photon happens only if the molecular vibration changes the dipole moment of the molecule. [22] Therefore molecules that have changing dipole moments due to chemical bond vibrations are IR active. [30] Dipole moment is caused by the charge differences in the molecule. [21] NIR absorption process also needs the photon energy to match the frequency of the overtones and combination modes.

Homonuclear diatomic molecules such as H_2 and O_2 do not have electronegativity differences and dipole moments, which leads to the fact that they are not IR active. Heteronuclear diatomic molecules such as HCl and CO have atoms with different electronegativities and therefore also have dipole moment and are IR active. [22]

Absorption of photon by IR active molecule leads to vibrational energy level change. [22] Absorption intensity (the "strength" of the IR activity) depends on the change in the dipole moment during the absorption. [31] The intensity of the IR absorption also depends on the concentration of the absorbing component in the sample according to the Beer-Lambert law (Equation 3.2). [21] Collected information in the form of IR spectrum consists of IR band intensities, which are proportional to the square of the change in the dipole moment. [22] The wavelength where the absorption band can be detected depends on the masses of the atoms in the chemical bond that produces the overtones and combination modes. The wavelength of the absorption band increases with the increase in the masses of the atoms and decreases when the chemical bond gets stronger. Absorption is much weaker in NIR region, compared to, for example, to mid-IR region, because of the transitions in the combination and overtone modes (in mid IR the fundamental transitions produce stronger absorption intensities). [31] The absorption band intensities in the mid-IR and NIR regions are shown in Table 3.1 to illustrate how the strength of the bands gets weaker in the overtones.

Most modes that are active in NIR region consists of vibrational modes of molecular functional groups of XH , where $X = C, O, N$ or S . [31] Materials containing hydrogen bonds have a lot of information of their chemical composition in their NIR spectra. This makes the analysis of organic materials in the NIR region useful. [9] Also stronger chemical bonds like $C=O$ can be observed with NIR spectroscopy methods. [31] In the NIR region

Table 3.1. Absorption band intensities in mid-IR and NIR regions. [32]

Band	Wavelength range	Average intensity (cm^2mol^{-1})
Fundamental	Mid-IR	10 000
1st overtone	Mid-IR or NIR	100
2nd overtone	NIR	10
3rd overtone	NIR	1
4th overtone	NIR	0,05

with wavelengths 800-1200 nm bands created by electronic transitions, overtones and combination modes can be observed. When the wavelength range rises to 1200-1800 nm the main band source is from the first overtones of stretching vibrations or combination modes of XH, where X = C, O, N. The bands in the wavelength range 1800-2500 nm are from combination modes. [29] Absorption bands seen in the spectrum needs to be assigned to make conclusions of the chemical composition of samples. [32] Some typical absorption band assignments of cellulose and water are listed in Table 3.2.

Table 3.2. Examples of absorption band assignments of cellulose and water. [32] [9]

Band location (nm)	Component	Bond vibration
1440-1470	Water	1st overtone O-H stretch
1471-1592	Cellulose	1st overtone O-H stretch
1920-1940	Water	O-H asymmetric stretch
2110	Cellulose	O-H deformation vibration + O-H stretch
2335	Cellulose and hemicellulose	C-H stretch + C-H deformation vibration

In the NIR region the absorption bands are difficult to interpret. The analysis of the NIR spectra have been made easier due to the bright radiation sources and sensitive detectors that are available nowadays. [31] Overlapping peaks and broad absorption bands makes recognizing the important chemical components and relevant information from the spectra complicated. [9] With the right data processing methods the information can be extracted from the spectral data.

4. MULTIVARIATE IMAGE ANALYSIS

Hyperspectral images contain a lot of information, relevant and irrelevant. The important and relevant information needs to be extracted from the images to make accurate analysis of the target samples. Multivariate data analysis methods are a useful tool to do this for hyperspectral applications. The image analysis workflow consists of image acquisition, calibration, preprocessing, data selection and model construction and validation.

4.1 Calibration and correction

The untreated raw hyperspectral images include both spectral and spatial errors. The errors can be noise, drift or errors due to wavelength dependencies, nonuniform lighting or changes in, for example, the temperature of the measurement environment. Calibration and correction steps need to be performed before applying analysis methods to the image. [33]

The imaged raw signal data is first acquired as detector signal intensity counts. Those need to be transformed into reflectance values to get the data into more useful units. This can be done using maximum reflectance and dark current measurements. Maximum reflectance spectra are measured with standard reference materials for which the amount of reflectance is known. Dark current is obtained by measuring the spectra after blocking the camera lens so that light does not enter the detector. [33] Common way to measure the maximum reflectance is to use Spectralon white reference standard materials. [15] Spectralon is a material that has the highest reflectance value of known materials in ultraviolet and infrared regions. Transformation of the detector signal intensity counts to reflectance values can be calculated using equation

$$R = \frac{H - D}{W - D}, \quad (4.1)$$

where H is the sample spectra, D the dark current and W the maximum reflectance reference value. This is called a two-point linear calibration model. Also, more complex calibration models can be used. [33] Measuring spectra of more than one standard reference material improves the accuracy of the calibration but makes the calculation more complicated and time-consuming. The aim of adding the calibration models is to standardize the spectra to get accurate spectral values for further analysis. [34]

There are two ways to correct the spatial differences in sample illumination and camera sensor response sensitivities. These two wavelength-dependent calibration models are global and pixelwise calibration. Global calibration model does not consider the spatial dependencies in the hypercube. This approach uses wavelength dependencies and takes into account each wavelength channel and assumes that sensitivities in the channels are distributed uniformly in the spatial locations. Median spectrum of each calibration standard is used to build the calibration model at each wavelength. Pixelwise calibration models are built to calculate transforms separately for each pixel in the sample. The calibration models are calculated for every pixel in the sample and for the corresponding reference standard sample pixel. This corrects the spatial changes in illumination and detector sensitivities across the sample. [33]

After conducting desired calibrations, reflectance values can be changed into absorbance units using equation

$$A = -\log_{10}(R), \quad (4.2)$$

where R contains the unitless reflectance values. [15] Absorbance transformed spectra shows the absorbance band locations, which can be assigned to different substances found in literature.

The images contain data of the sample and usually also of the background. To get accurate information only of the sample for further analysis, the sample pixels can be separated from the background. Conducting reflectance transforms and background separations mentioned above removes the imaging device and measurement related errors and prepares the images for further preprocessing.

4.2 Chemometrics

Hyperspectral images contain large amounts of information. The spectra need to be simplified and the interesting information needs to be extracted to understand it better. Chemometrics is a field of science trying to create data-driven models to estimate the desired chemical properties of the samples [35]. In other words, chemometrics focuses on using mathematical and statistical methods to analyse the structure and properties of a chemical system. [30] Chemometric methods are widely used in spectroscopic applications and in interpretation of other multivariate data. [36]

The early development of chemometrics started in the late 1960s in several analytical chemistry research groups. When scientific computing and computer-aided methods became widely accessible, it made more complex analysis possible. [37] Also the introduction of multivariate analysis instruments enabled the development of chemometrics. [36]

Most of chemometric methods deals with two-dimensional data matrixes and extracts the useful information from them. [36] There are many ways to analyse the data matrix, such as component or factor analysis, regression analysis and classification. [38] Principal Component Analysis (PCA) and Partial Least Squares (PLS) are examples of basic chemometric methods. [35] PCA will be reviewed in Section 4.4.1. Chemometric methods produce better results if some preprocessing methods are used on the data set before applying them.

4.3 Preprocessing

Spectral preprocessing methods are commonly used in spectroscopic image analysis. [15] Multivariate analysis provides better results if the preprocessing methods have been used to transform the spectra, highlight the underlying features, remove errors and noise, and to emphasize the relevant information from the image spectra. Some unwanted effects in the spectra can be scattering, detector artefacts or differences in the sample surfaces. [18] There are various spectral preprocessing methods and the ones to use need to be chosen for each application case-by-case. [39]

4.3.1 Standard normal variate transformation and mean centering

Scattering, different particle sizes and multicollinearity are some effects that make absorbance spectra complex and hard to interpret. Standard Normal Variate (SNV) transformation is a mathematical row-wise centering and scaling method used to remove unwanted effects from spectra. [15] SNV removes the effects of scatter and changing particle sizes. SNV can be calculated with equation

$$SNV = (x - \bar{x}) / \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}, \quad (4.3)$$

where x is the absorbance spectra, \bar{x} is the mean of the absorbance spectra and n is the size of the number of samples. [40] SNV transformed spectrum has mean value 0 and standard deviation 1. The shape of SNV transformed spectra is the same as in the absorbance spectra before applying the transformation, only the unwanted effects has been removed.

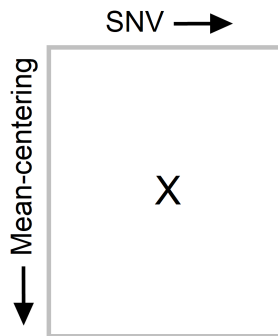


Figure 4.1. Directions of standard normal variate transformation and mean centering for data matrix X .

SNV handles the row-wise processing of the spectra and mean centering is used to do the column-wise centering. Mean centering is conducted by subtracting the mean values of the column variables from the spectra. Figure 4.1 illustrates the directions of SNV and mean centering methods for data matrix X .

4.3.2 Derivative methods

Spectral derivative methods improve the resolution of IR spectra. The aim for derivative methods is to remove the overlapping spectral signals. [18] One of the most used derivation and smoothing method in chemometrics is Savitsky-Golay method. Savitsky-Golay smoothing removes the high frequency noise and derivation removes the low frequency signals. [41] Savitsky-Golay smoothing is a row-wise method which is based on moving-window polynomial fitting. Window with a chosen range of points is used to fit the polynomial of given order and to estimate new value for given wavelength. [18] Polynomial fitting smooths the spectra, and the derivative is calculated from the fitted values. [32] Savitsky-Golay method can be used for smoothing, derivation and noise reduction. [18]

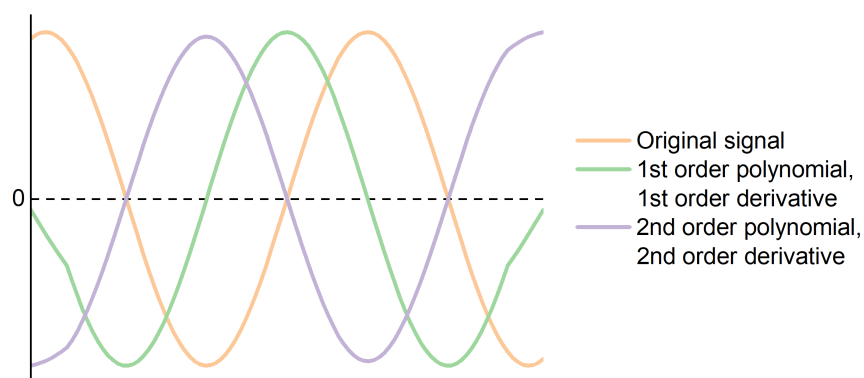


Figure 4.2. The effect of adding 1st and 2nd order derivation on sinusoidal signal.

Adding derivative methods on spectra makes the interpretation more difficult. [40] The locations of the peaks are moved as illustrated in Figure 4.2 for sinusoidal signal. The second order derivation leads to absorption peaks pointing to the opposite directions. Also possible band shifts need to be considered when analysing the second derivative of the spectrum. For example, the shape of the NIR spectrum and the sizes of the sample particles affects the band shifts. [32] As can be seen in Figure 4.2, derivation creates aberrations around the edges of the signal. Aberrations can be eliminated by, for example, selecting the wavelengths of interest after applying the derivative methods.

4.4 Exploratory data analysis and classification

After calibrating and preprocessing the hyperspectral images, exploratory data analysis methods can be applied to find correlations and patterns in the sample set. Studying the image spectra with exploratory analysis methods helps to determine the classification methods which produce accurate results.

4.4.1 Principal Component Analysis

Feature extraction methods can be used to reduce the dimensions of the data and to find the important information from it. Principal component analysis is a multivariate method that is commonly used to do this. [35] PCA is used to find correlations between variables and to reduce the dimensions of the data set into variables that contain the relevant information. These variables are linear combinations of the original variables, and they explain most of the variation in the data set. The variables are called principal components (PC). [42] [43]

The principal components are calculated so that the first PC explains most of the variation in the data and the second PC the second most and so on. [43] PCA model consists of the data, the scores, the loadings and the residuals as shown in Figure 4.3. Vectors T and P are called score and loading vectors. Matrix E contains the residual information that is not included in T and P. [44] All of the information in the original set of data is saved in the vectors T, P and E. Scores can be explained as the projections of the observations to the principal components and loadings as the weights required to transform the original variables, or geometrically as the rotation of the original axes in the sample space. [42]

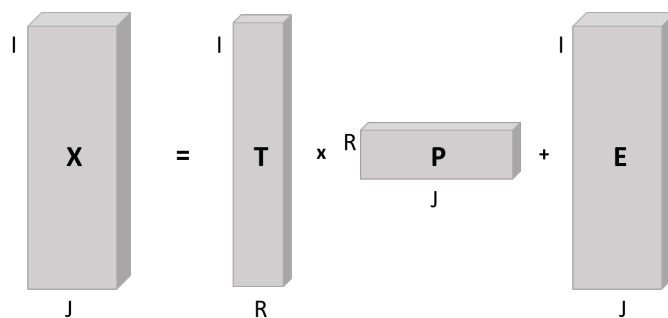


Figure 4.3. PCA model for a $I \times J$ data matrix X .

Calculation of the principal components starts with mean centering the data. First principal component is calculated by finding the direction of the largest variance in the data set. The first principal component therefore explains most of the variation in the data. The second principal component is orthogonal to the first one. [42] This is illustrated in Figure 4.4 for a two-dimensional set of random points from a multivariate normal distribution. Other principal components are calculated similarly. Score values can be plotted and visualized for example in score plots. Score plots show the possible groupings and helps to detect patterns in the data set. [43] Part D in Figure 4.4 shows an example of a score plot.

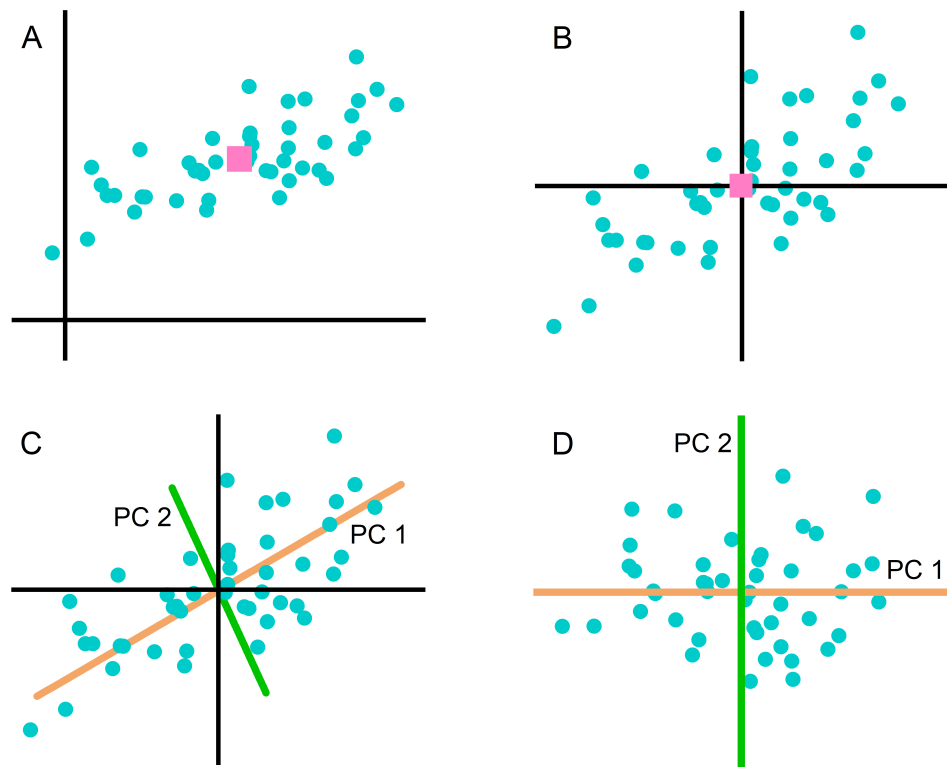


Figure 4.4. Principal component analysis visualization. A) Random set of multivariate normal distributed points. Pink square indicating to the mean center of the points. B) Mean centered set of points, C) directions of the first two principal components, and D) score plot of the first two principal components.

Common method to implement PCA and to calculate these principal components is to use singular value decomposition (SVD). To implement SVD, hyperspectral data cube needs to be unfolded to 2D data matrix, where rows contain the samples and columns

the wavelength variables of the sample spectra. SVD can be calculated using equation

$$X = U\Sigma V^T, \quad (4.4)$$

where X is the preprocessed and mean centered data matrix, U and V are orthogonal matrixes containing eigenvectors and Σ diagonal matrix containing eigenvalues. Calculating this for hyperspectral data can be time-consuming. To reduce the computation time, SVD can be also calculated for matrix X using its cross-product $X^T X$ with equation

$$X^T X = V\Sigma V^T, \quad (4.5)$$

where V is the matrix with same eigenvectors as in Equation 4.4. [44] Vectors in matrix V are also called loading vectors and they can be used to calculate the score vectors:

$$T = XV. \quad (4.6)$$

The first components of T and V contain the information of the first PC, the second component the information of the second PC and so on. The quality on how well the combination of T and P summarizes the data in matrix X can be evaluated with calculating the percentage of explained variation. The number of used principal components needs to be chosen first to calculate the E . When the percentage of explained variation is good, the combination of T and P creates good summary of the original data. [43] This can be calculated using equation

$$R^2 = \frac{\|X\|^2 - \|E\|^2}{\|X\|^2} 100\%, \quad (4.7)$$

where statistic R^2 can take values between 0% and 100%. Values closer to 100% means better summarization.

4.4.2 Linear Discriminant Analysis

Principal components acquired from PCA might not explain the desired variations. [45] There are alternative methods to discriminate the samples in the data set. Discriminant analysis is widely used supervised classification method. [18] While PCA tries to find the directions of the largest variance in the data set, discriminant analysis methods try to find the maximum variance between predefined classes and the minimum variance between the samples within a class.

Most used discriminant analysis methods are Quadratic Discriminant Analysis (QDA) and Linear Discriminant Analysis (LDA). QDA does the classification using quadratic boundaries separating the classes. LDA classifies the samples into classes using linear projections as boundaries. [18]

Linear Discriminant Analysis can be conducted using Canonical Variate Analysis (CVA). Canonical variables (CV) are calculated finding the minimum variance within a class and maximum variance between classes. Within-class scatter matrix is the average of covariance matrices of each class weighted by the class sizes. It can be calculated using equation

$$S_w = \frac{1}{n - c} \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T, \quad (4.8)$$

where n is the number of samples, n_i the number of samples in the i th class, c is the number of classes, x_{ij} the j th sample in i th group and \bar{x}_i the mean vector of the i th class. Between class scatter matrix is the variance matrix of the class means weighted by the weight of the class according to equation

$$S_b = \frac{1}{c - 1} \sum_{i=1}^c n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T, \quad (4.9)$$

where \bar{x} is the overall mean vector and other variables same as in Equation 4.8. Canonical Variates Analysis tries to find the direction w which maximizes the equation

$$J(w) = \frac{w^T S_b w}{w^T S_w w}, \quad (4.10)$$

where $J()$ is the criterion function. [46] CVA cannot be applied on a data set where the number of variables exceeds the number of samples. To reduce the number of variables, PCA can be used first and then CVA calculated using principal component scores. [45] CVA works well if the variance between the classes is large compared to the variance within the classes. [47]

Calculated scores on the canonical variables can be used to calculate the linear class boundaries using linear discriminant analysis. An unknown sample is considered to belong to a specific class based on which side of the boundary it falls. This is illustrated in Figure 4.5. Unknown sample X is classified to class 2.

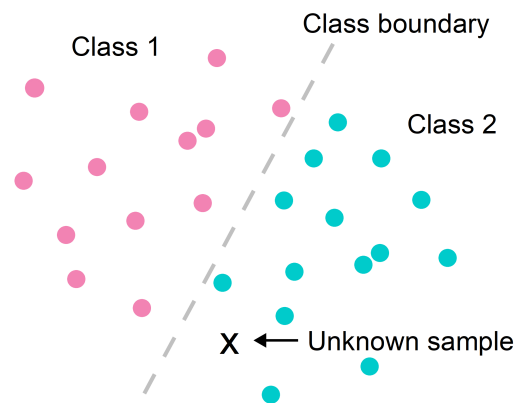


Figure 4.5. Classification of an unknown sample based on a linear class boundary.

The classification accuracies can be calculated as correctly classified rates (CCR) or in correctly classified percentages (%CC). %CC can be calculated dividing amount of correctly classified samples with all the samples and multiplying with 100. %CC can also be calculated for each class separately. These values are used to evaluate the quality of the classification method and how well the samples can be classified. [48]

Table 4.1. Confusion matrix example.

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Results can also be shown in confusion matrixes where the number of correctly and incorrectly classified samples can be seen clearly. Table 4.1 shows an example of confusion matrix. Content of confusion matrix can be explained, for example, by assuming two

classes, positive and negative as shown in the table. Classification result of samples can be divided into four categories: true positives (TP) are samples that are known to belong to positive class and predicted to belong there, true negatives (TN) are samples that are correctly predicted to belong to negative class, false positives (FP) are samples that are incorrectly predicted to belong to positive class and false negatives (FN) samples that are incorrectly predicted to belong to negative class. [49] The CCR of the positive class is also called sensitivity and for the negative class selectivity. [48]

4.4.3 Wavelength selection and cross-validation

One way to improve the accuracy of the classification results is to choose and use only wavelengths that contain most of the wanted information. Wavelength selection before a classification model improves classification accuracy if a subset of wavelengths can be found that better separates the sample classes. [31] Other wavelength regions that do not contain useful information may decrease the model accuracy.

There are many ways to conduct wavelength selection. The main idea is to find sets of wavelengths that improve the classification accuracy. Wavelength selection should be made for each application case-by-case depending on the chemical composition of the samples. Samples imaged in NIR region have mostly information of hydrogen functional groups. It would seem useful to choose the wavelengths where the absorption bands of these groups are seen. [31]

One way to perform wavelength selection is by dividing the wavelengths into intervals and calculating the classification accuracy for all possible interval combinations. The wavelength combination that provides the best classification results can be chosen for further analysis. Classification accuracies for different wavelength interval combinations can be calculated using cross-validation, which is a model validation method. The sample set is divided into folds which are divided to training and test sets. N-fold cross-validation is visualized in Figure 4.6.

The classification model is calculated using the training set, and the test set is used to calculate the accuracy of the model. This is done so that each fold of samples is used as a test set once. The classification accuracy is calculated each round and as a result the overall accuracy for the classification model is obtained.

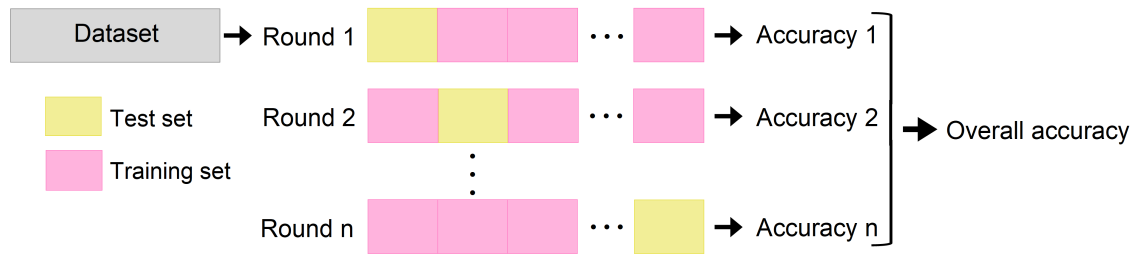


Figure 4.6. *N-fold cross-validation visualization.*

5. EXPERIMENTAL METHODS

The objective of this thesis was to find simple hyperspectral image processing and classification methods to classify cellulose-based textile samples based on their intrinsic viscosity values. In this chapter the samples and used methods are presented.

5.1 Textile samples

The textile samples were provided by Aalto University, and they consisted of two different sets of samples. The first set had 18 different textiles and three to four samples of each, in total 55 textile samples. The second set had samples of 19 different textiles and in total 57 samples. The samples were approximately squares with 10 cm side length.

The intrinsic viscosity values for each sample were measured at Aalto University using the SCAN-CM 15:88 standard. Intrinsic viscosity of a sample was measured by dissolving pulp of the sample to cupri-ethylenediamine (CED) solution. A specified amount of sample solution and solvent were put through two capillary viscometers. The viscosities were calculated using the efflux times. [50] Intrinsic viscosity values of the samples were around 300-1660 ml/g. Table 5.1 shows the number of samples in each sample set and the measured intrinsic viscosity region.

Table 5.1. *Sample sets used in the analysis and their intrinsic viscosity regions.*

Sample set	Number of samples	Intrinsic viscosities (ml/g)
1	55	300-1270
2	57	590-1660

The optimal intrinsic viscosity region for the classification, defined by the chemical recyclers at Aalto University, was 350-550 ml/g. In this region there were 12 samples and all of them were from sample set 1.

5.2 Image acquisition

Hyperspectral images were measured using Specim SWIR 3 (Specim, Spectral Imaging, Ltd.) hyperspectral camera. [51] The camera is shown in Figure 5.1. The operating principle of the camera was similar as reviewed in Section 2.2. The images were scanned in line-mode with 384 spatial pixels in one line and 288 spectral wavelength bands for each pixel. Spectral bands covered the wavelength region 967-2560 nm with a spectral sampling width of 5.6 nm and spectral resolution of 12 nm. Samples were illuminated with polychromatic light produced by quartz halogen lamps. Figure 5.1 shows the imaging set up. Each sample was imaged separately and reference standard material with 99 % reflectance was imaged with each sample.

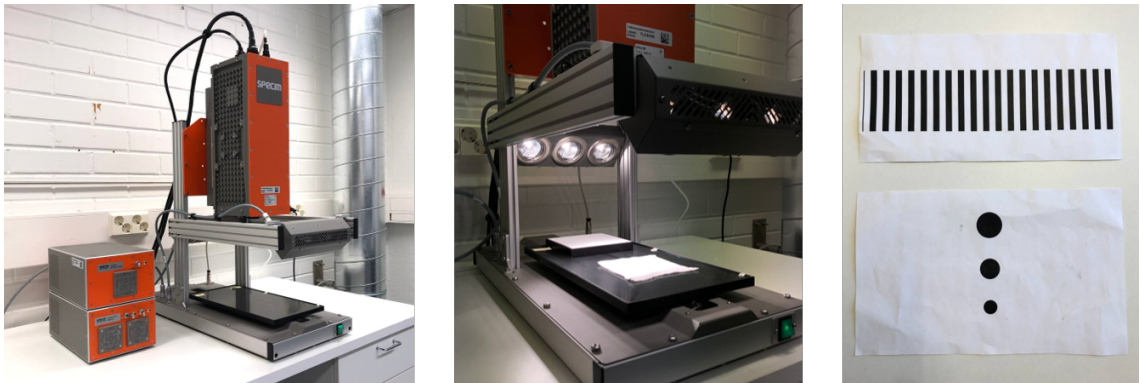


Figure 5.1. Specim SWIR 3 hyperspectral camera (left), sample imaging set up (center) and samples imaged to set the moving table speed and the field of view (right).

The field of view was set to cover the textile samples, approximately to 12 cm. It was determined by imaging a sample with lines. During imaging the samples moved below the camera, approximately 27 cm distance away from the objective. The speed of the moving table was set so that round objects appeared approximately round. This was examined by imaging sample with round objects. This and the sample imaged to set the field of view are shown in Figure 5.1. The integration time was adjusted so that the maximum reflectance signal was approximately 90 % of the maximum signal output. Both sets of samples were imaged with the same camera settings in a dark room to eliminate the effect of room lighting. Imaging all the samples produced 112 large hyperspectral data cubes.

5.3 Image processing

Hyperspectral images were loaded to Matlab® using an in-house Matlab function developed at VTT. Images were processed in Matlab and most of the code was written specifically for this thesis. Commercial Matlab functions *svd* [52], *canoncorr* [53], *classify* [54] and *PLS – toolbox* software [55] were incorporated inside the written code. Figure 5.2 shows the image processing work flow.

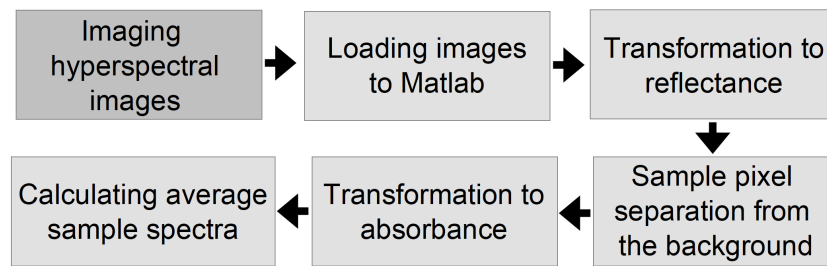


Figure 5.2. Hyperspectral image processing workflow.

Raw hyperspectral images were measured as detector signal intensity counts. Reference material and dark current measurements were used to transform the raw signal values into reflectance using a simple two-point linear reflectance transform shown in Equation 4.1.

The reflectance transformed images were still large hyperspectral data cubes containing also information of the background. The sample pixels were separated from the background pixels using principal component analysis. The images were reshaped to two dimensional matrixes and PCA was conducted after mean centering the data using singular value decomposition. SVD was calculated using the Matlab function *svd*, which returned the eigenvectors of the matrix. These eigenvectors were used to calculate the principal components. PCA found the direction of the maximum variance between the sample and the background pixels, which was seen in the first principal component. Limit was set to score values to separate these pixels from each other. Only the sample pixels were chosen for further analysis.

All the sample pixels were used to calculate the average spectrum for each sample. The average reflectance spectra were converted into absorbance units using Equation 4.2. An example of separating sample pixels from the background and the calculated respective average absorbance spectrum is illustrated in Figure 5.3.

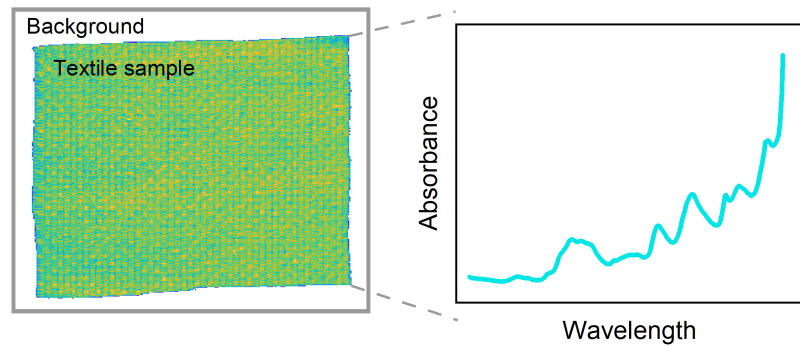


Figure 5.3. Sample pixel separation from the background and the calculated respective average absorbance spectrum.

Images were calibrated and corrected and ready for further processing and analysis. From 112 hyperspectral image cubes, the information of each sample was now in one average absorbance spectrum.

5.4 Differences between sample spectra

The whole sample set consisted of two sets of textile samples. The sample sets were compared to each other using unsupervised exploratory analysis to find possible differences. PCA was used to find the possible variance between the samples, and principal components were calculated using Matlab function *svd*. Principal component scores were visualized in score plots and different sets of textiles and intrinsic viscosity values were compared. The confidence ellipses for score plots were calculated using PLS-toolbox. Loading plots of the first two principal components were used to explain the differences between the two sample sets.

5.5 Sample classification

The two sample sets combined together were allocated to three classification classes according to their intrinsic viscosity values. Intrinsic viscosity values were known for each sample and the optimal intrinsic viscosity region was 350-550 ml/g. The intrinsic viscosities of the classes are shown in Table 5.2. Colors for the classes are the same class colors used for the classification results in Chapter 6. The samples were then divided into training sample set and test set for classification. The training set was used to develop

the classification model and the test set to validate the final model. From all the samples every third sample was chosen to the test set. The training set consisted of 74 (66 %) samples and the test set of the remaining 38 (34 %) samples. Both training and test set included samples of each classification class. Division of the samples to training and test set, and the workflow of developing and validating the classification model is shown in Figure 5.4.

Table 5.2. *Intrinsic viscosities of the classification classes.*

Class	Intrinsic viscosity (ml/g)
Below optimal viscosity	< 350
Optimal viscosity	350-550
Above optimal viscosity	> 550

Training set samples were used to train and adjust the model. Different preprocessing methods were tested to evaluate their effect on the classification performance. The average absorbance sample spectra were preprocessed using SNV and mean centering. SNV was used to remove the effects of scatter row-wise, and mean centering column-wise to make the differences in the spectra more visible. SNV transformation was done using Equation 4.3 and mean centering by subtracting the mean values from the wavelength variables. These calculations were conducted using Matlab code written for this thesis. Wavelengths outside the range of 1000-2500 nm were excluded to analyse only the wavelength region of interest.

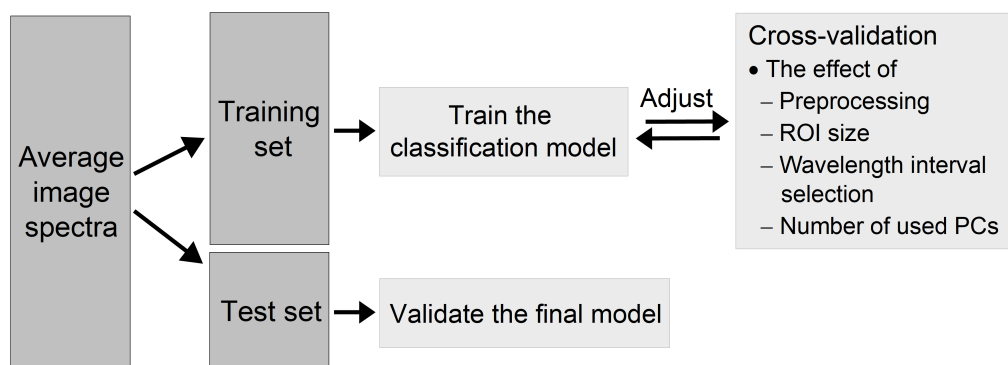


Figure 5.4. *Classification model creation and validation workflow.*

The classification model was trained using the samples in the training set. Linear discriminant analysis using canonical variables was used as classification method. Covariance matrices for calculating the canonical variables cannot be calculated if the number of

wavelength variables exceeds the number of samples. So PCA was performed first to reduce the dimensions of the wavelength variables. The mean of the spectra was calculated for mean centering and the principal components were calculated using the *svd* function. The optimal number of used principal components were calculated using 3-fold cross-validation illustrated in Figure 4.6. Cross-validation was calculated for all possible numbers of principal components. The correctly classified rate was determined for each cross-validation round, and the lowest number of principal components that produced 100 % correctly classified rate was chosen.

Canonical variate analysis was used to find the directions which maximized the variance between classes and minimized the variance within a class. Canonical variate analysis used the allocated classification classes of the samples to calculate the canonical variables. Canonical variables were calculated using the principal component scores and Matlab function *canoncorr*. The function returned the canonical coefficients and the scores.

Test set sample spectra were preprocessed with the same methods as the samples in the training set. Spectra were then mean centered with the mean of the training samples. The dimensions of the wavelength variables in the spectra were reduced using the principal component loadings calculated for the training samples. Then the test sample spectra were projected into the canonical variable space determined for training set using the canonical coefficients.

Matlab function *classify* was used to train the discriminant analysis classifier and to calculate linear class boundaries for the three classes. Classify-function takes the test samples, training samples and classes for the training samples. The linear class boundaries are then calculated based on the training samples and their classes. Function assigns each test sample into one of the classification classes based on which side of the class boundaries the sample fall as illustrated in Figure 4.5.

The classification results were calculated as correctly classified percentages. %CC was calculated by dividing the number of correctly classified samples with the number of all samples. %CC was calculated for each class, training set and test set separately. These values were used to evaluate the accuracy of the classification method and how well the samples could be classified.

6. RESULTS

The results of hyperspectral image processing and textile sample classification using methods described in Chapter 5 are presented in this chapter.

6.1 Image calibration and preprocessing

The textile sample sets were imaged and treated as reviewed in Chapter 5. Imaging all the samples produced 112 hyperspectral images. In Figure 6.1A is illustrated hyperspectral image of one textile sample. Each image consisted of 384 pixels per imaged line, 425 lines and 288 wavelength bands. The sample pixels were separated from the background using PCA, which is illustrated in Figure 6.1B.

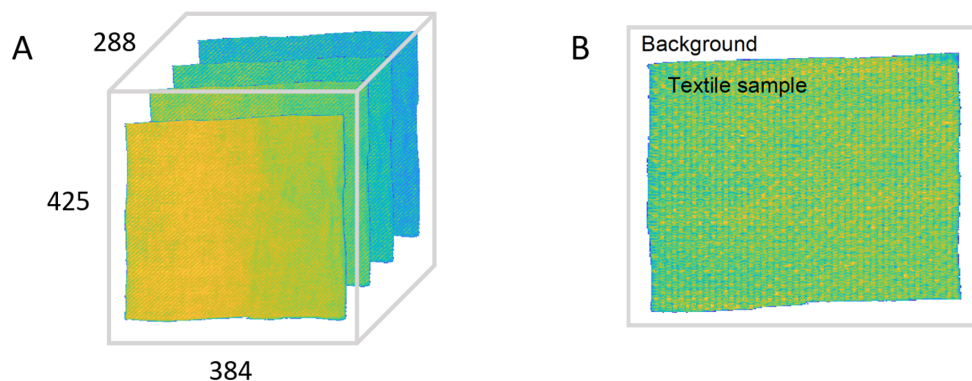


Figure 6.1. A) Hyperspectral image of one textile sample and B) illustration of separating sample pixels from the background.

Raw sample images, white reference material and dark current were imaged as detector signal intensity counts. Figure 6.2 shows raw average signals for one imaged sample. White reference signal W refers to imaged white reference standard with 99 % reflectance and dark reference signal D to dark current. Signal H is the average spectrum of the sample.

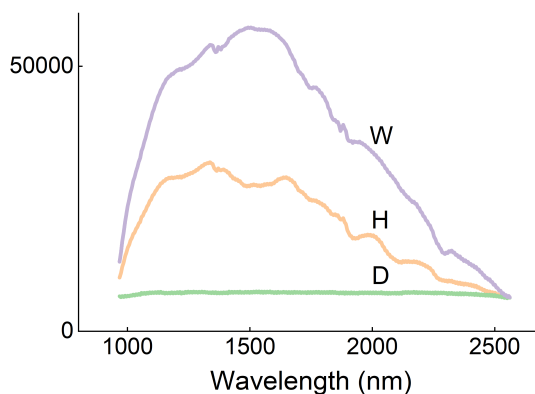


Figure 6.2. Raw average signals for one imaged sample. W referring to white reference signal, D to dark reference signal and H to the sample spectrum.

Simple two-point linear reflectance transform was used to convert the raw image values into reflectance units. Then images were converted into absorbance units for easier interpretation of the spectra. All the sample pixels separated from the background were used to calculate the average spectra for each sample. Figure 6.3 shows reflectance transformed average spectra and the $-\log_{10}$ transformed average absorbance spectra of the samples in the first sample set. Average reflectance spectra do not show the differences between the samples that can be observed in the $-\log_{10}$ transformed average absorbance spectra.

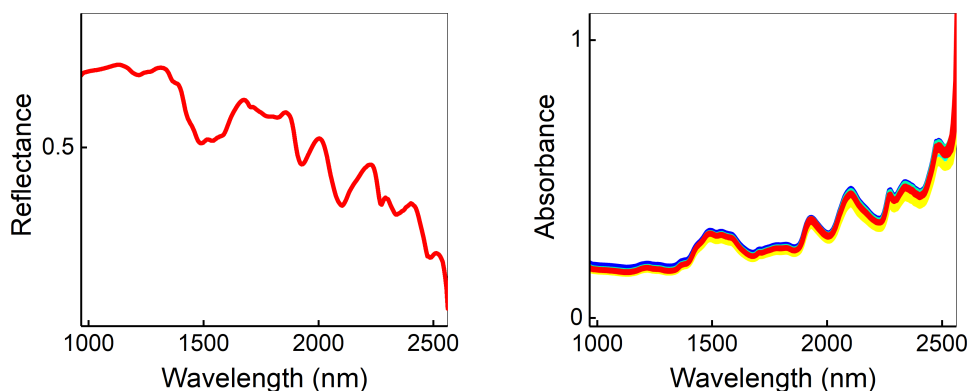


Figure 6.3. Reflectance transformed average sample spectra and $-\log_{10}$ transformed average absorbance spectra of the samples in the first sample set.

Standard normal variate transformation was used to remove the effect of scatter and decrease the effects of uneven lighting. Figure 6.4 shows the effects of SNV transformation and mean centering on the average absorbance spectra of the first sample set. Spectra are colored according to the measured intrinsic viscosity values. Lower viscosity values are colored in blue and higher in red.

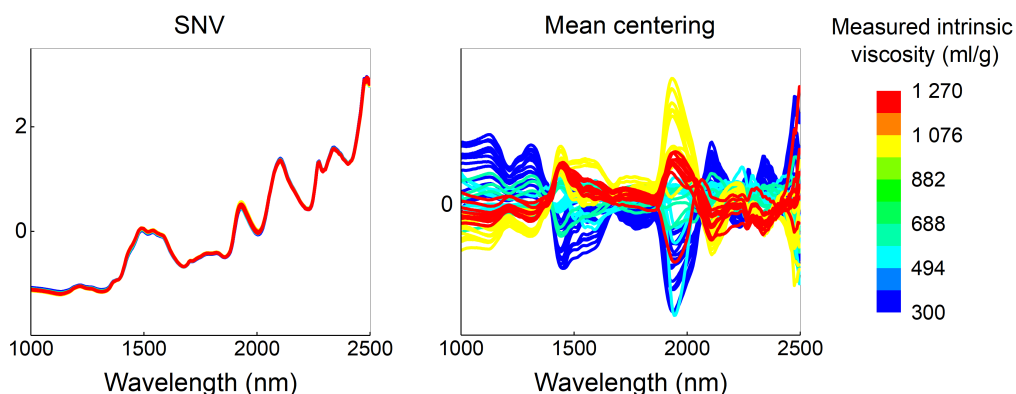


Figure 6.4. SNV transformed and mean centered average absorbance spectra of the first sample set. Spectra colored according to measured intrinsic viscosity values.

After SNV transformation, absorbance spectra showed less variation between the samples as can be seen from the Figures 6.3 and 6.4. Mean centering the SNV transformed spectra started to show differences between the intrinsic viscosities of the samples. This can be visually seen from the mean centered spectra in the Figure 6.4. SNV transformed and mean centered absorption spectra shows the largest variations between absorption bands at wavelengths around 1440 nm and 1930 nm. These bands can be assigned to water according to Table 3.2. Another large absorption band variation can be observed at 2108 nm, which is close to cellulose band assignment reported in Table 3.2.

6.2 Differences between sample sets

The textile samples consisted of two different sample sets. Examining the two sets with PCA showed that the two sets of samples could be separated from each other. This is shown in score plot of the first two principal components in Figure 6.5, where the samples are colored based on the sample sets. The first two principal components account for 79 % of the variance between the sample spectra. In the figure the scores are also colored based on the measured intrinsic viscosity values. Differences and discrimination also between different intrinsic viscosity values can be seen. Principal component loadings show which wavelength regions explain most of the variation in the spectra. Loading plots of the first two principal components calculated for the two sample sets are illustrated in Figure 6.6. The highest peaks in the first principal component loading plot can be assigned to water and the loading peaks in the second principal component to cellulose

according to Table 3.2. The differences between the two sample sets are due to different chemical compositions.

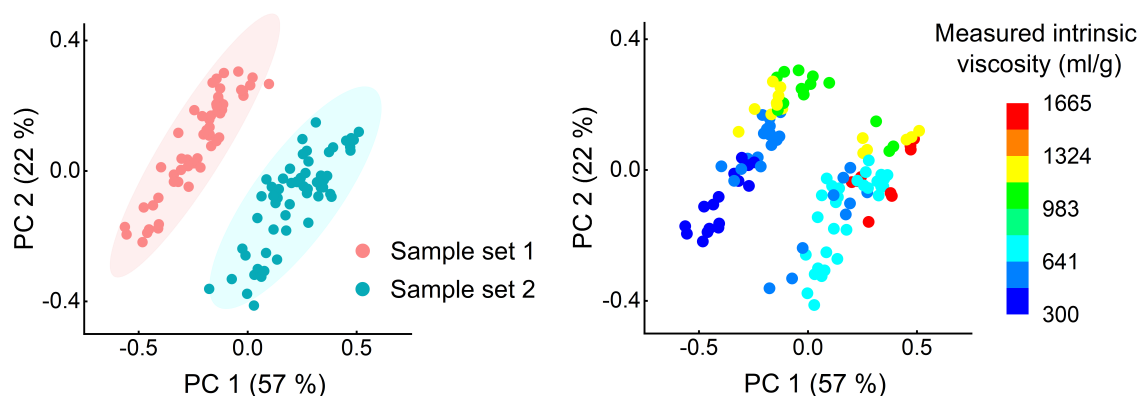


Figure 6.5. Score plots of first two principal components showing the differences of the two sample sets. Colored based on the sample sets (left) and measured intrinsic viscosities (right).

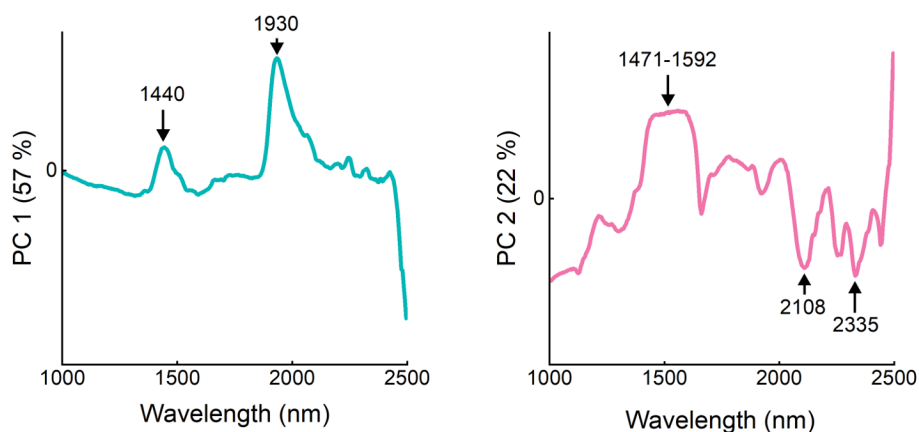


Figure 6.6. Principal component loading plots of the first two principal components calculated for the two sample sets.

6.3 Sample classification performance

All the samples were divided into training and test sets for classification. Training set consisted of 74 samples and the test set of 38 samples. The classification model was trained and adjusted using on the training set samples. Methods used for determining the classification model were PCA to reduce the number of wavelength variables, CVA to find the separation between the classes and LDA to calculate the class boundaries. Classification model was trained to classify the samples into three classes according to their intrinsic viscosity values, which is shown in Table 5.2. Table 6.1 shows how the

samples were divided into classification classes in training and test sets. As seen from the table, the available textile samples were not distributed evenly to each class.

Table 6.1. Number of samples in each class in training set, test set and overall.

Class	Training set	Test set	Overall
< 350	8	5	13
350-550	8	4	12
> 550	58	29	87

The optimal number of used principal components were estimated using cross-validation on the training set. Cross-validation was conducted using 3 folds, which ensured that there were samples of each class in each fold. The smallest number of principal components that produced 100 % correctly classified rate for cross-validation was 34. First 34 principal components were chosen to be used in the classification.

LDA classification model was created using methods reviewed in Section 5.5. The results of the training set classification are shown in Figure 6.7. The class regions are visualized in the figure with colors. Class with viscosity values <350 ml/g is colored in pink, 350-550 ml/g in green and >550 ml/g in blue. The samples, dots in the figure, are colored based on the real class they belong to. A sample is counted as misclassified if it is not located in the region of its own color in the figure. Samples in each class forms clusters and can be separated from each other. Figure 6.7 shows that there were no misclassified samples during the model training. Table 6.2 shows the confusion matrix of the training set classification results.

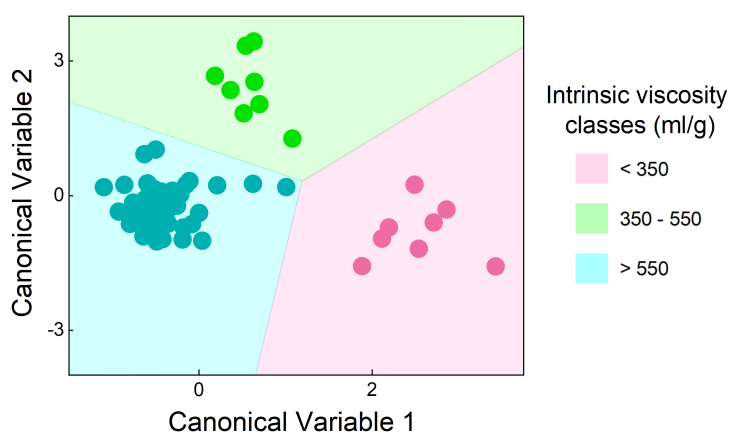


Figure 6.7. Classification results of the training sample set.

Table 6.2. Confusion matrix of the training set classification results. Correctly classified rate 100 %.

		Measured		
		< 350	350-550	> 550
Predicted	< 350	8		
	350-550		8	
	> 550			58

The test set was used to validate the determined classification model. Test set samples were treated with the same preprocessing methods which were used to train the classification model. Test samples were classified as reviewed in Section 5.5. The test set classification results are shown in Figure 6.8. The class regions are colored similarly as in Figure 6.7.

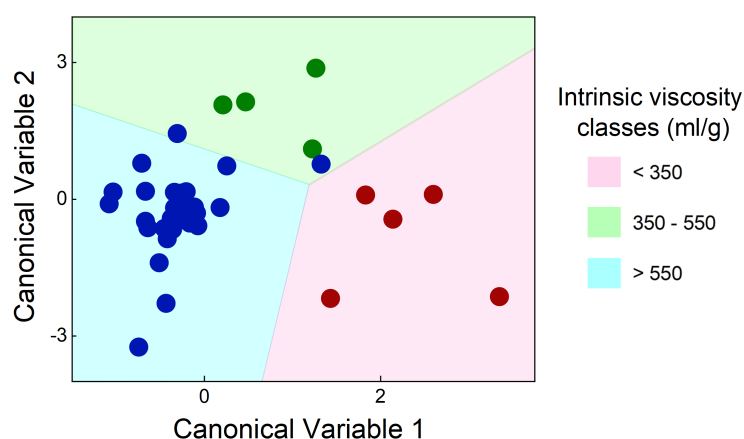


Figure 6.8. Classification results of the test sample set.

As can be seen from the Figure 6.8, two samples were misclassified. These misclassified samples belong to the class with viscosity values >550 ml/g and they were misclassified to the optimal viscosity class 350-550 ml/g. The two misclassified samples were close to the class boundaries as seen from the figure. Table 6.3 shows the confusion matrix for test sample set classification results. The correctly classified rate for the test sample set was 94.7 %.

Table 6.4 shows the correctly classified percentages for training and test sets in each classification class. Overall, there were only two misclassified samples. Classification results showed the possibility of classifying textile samples according to their intrinsic viscosity values. Image processing and classification methods used in this thesis produced

promising classification results.

Table 6.3. Confusion matrix of the test sample set classification results. Correctly classified rate 94.7 %.

		Measured		
		< 350	350-550	> 550
Predicted	< 350	5		
	350-550		4	2
	> 550			27

Table 6.4. Classification results for each class, training and test sets as correctly classified percentage %CC.

Class	Training set	Test set
< 350	100	100
350-550	100	100
> 550	100	93.1
Overall	100	94.7

7. DISCUSSION

Analysing hyperspectral images showed promising results that cellulose-based fabrics could be classified based on intrinsic viscosity. The samples were classified to predefined classes with an overall test set accuracy of 94.7 %. The methods used in the analysis were chosen to be as simple as possible. Using simple methods makes the interpretation of the results easier. This chapter discusses the classification results, the effects of using different image processing methods and other observations made during the research process.

7.1 Textile samples

The textile samples consisted of two sets of different textiles. Comparing the two sample sets using PCA showed that the two sets could be separated from each other. This was seen in Figure 6.5. Principal component loadings showed that the separation in PCA could be explained by differences in the chemical composition of the sample sets. The highest peaks in the first loading vector could be assigned to water and the peaks in the second loading vector to cellulose. This was seen in loading plots of the first two principal components in Figure 6.6. Even though the sample sets differ chemically from each other, it was shown that they can be classified based on the intrinsic viscosities.

The intrinsic viscosities of the samples were measured three times by hand based on the SCAN-CM 15:88 standard [50]. The average value of the three measurements was calculated for each sample. These intrinsic viscosity measurements create standard deviation for the viscosity values. Multivariate classification methods used in this thesis takes the given viscosity values and does not consider the deviation. The results need to be analysed with that in mind that the measured viscosity values used to train the model might not be as accurate as what the hyperspectral camera measures. The viscosities were

measured with same method and standard for both sample sets.

The classification model was trained and tested using predefined classes based on measured intrinsic viscosity values. Linear discriminant analysis and canonical variables were shown to provide promising classification results for our samples. Correctly classified percentages of 100 % for training set and 94.7 % for test set showed that chemically slightly different textile samples could be classified in object level according to their intrinsic viscosity values. Two samples in the test set were misclassified. The measured intrinsic viscosities of these samples were 634 ml/g and 592 ml/g, and the standard deviations ± 18 ml/g and ± 6.2 ml/g respectively. The standard deviations do not explain the reason for misclassification, but the measured sample viscosities are quite close to the optimal viscosity class. All the samples in the optimal intrinsic viscosity region (350-550 ml/g) were classified correctly.

The number of available samples also affected the classification results. There was imbalance in the number of samples in the classification classes as could be seen from Table 6.1. 78 % of the samples belonged to the class with viscosities >550 ml/g. Rest of the samples were distributed to the optimal 350-550 ml/g and <350 ml/g viscosity classes. Dividing the samples to training and test sets made the number of samples in the training set even smaller. The classification model was trained with only eight samples in the 350-550 ml/g and <350 ml/g viscosity classes.

The method for classification model quality evaluation was accuracy, which was calculated as correctly classified percentages. The class imbalance can make the accuracy misleading, because it may vary between classes. Using this metric also when training the classification model may give too much weight for the classes with more samples and inadequate weight for others. Classification model may not find the characteristics of the samples in classes with less samples, and these samples are more likely to be misclassified. [56] This effect was not seen in our classification results as the only misclassified samples belonged to viscosity class >550 ml/g. Classes could be rebalanced by collecting more samples to viscosity classes 350-550 ml/g and <350 ml/g. Other ways for rebalancing could be removing samples from class >550 ml/g to make the class sizes equal or oversampling the classes 350-550 ml/g and <350 ml/g to increase the number of samples in them.

7.2 Preprocessing methods

Different preprocessing methods were applied on the training set sample spectra to find the differences between intrinsic viscosity values. Standard normal variate transformation and mean centering were found to be the simplest methods to provide accurate classification results. The absorption bands seen in SNV transformed and mean centered average spectra in Figure 6.4 show band locations which can be assigned to water and cellulose. These substances are found in cellulose-based fabrics.

Savitsky-Golay derivation was studied as one of the preprocessing methods. Derivative methods change the positions of the absorption peaks, which makes the interpretation of the spectra harder. This effect was illustrated in Figure 4.2. The effect of adding Savitsky-Golay 1st and 2nd order derivation with frame length of 13 on the average absorbance spectra of the first sample set is shown in Figure 7.1. Derivation was applied before SNV and mean centering. The differences between intrinsic viscosities can be seen, but the locations of the absorption peaks are moved, and absorption bands cannot be so clearly assigned after adding the derivative method.

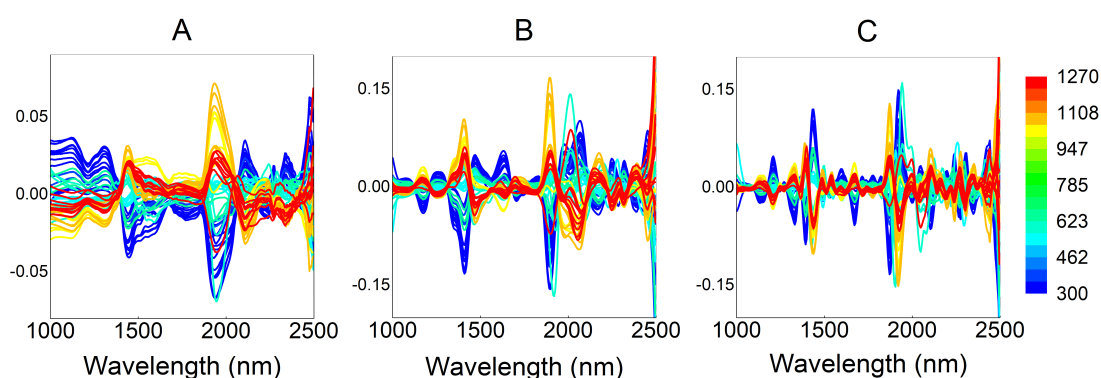


Figure 7.1. Effect of applying B) 1st and C) 2nd order Savitsky-Golay derivation on the A) SNV transformed and mean centered average absorbance spectra of the first set of samples. Sample spectra colored based on the measured intrinsic viscosity values (ml/g).

Table 7.1 shows the training set and training set 3-fold cross-validation classification results with different preprocessing methods. Results are calculated using all sample pixels and the number of used principal components is calculated individually for each used method combination. Mean centering was used as the last preprocessing method in all the combinations. Using Savitsky-Golay derivation did not clearly improve the classifica-

tion results.

Table 7.1. Effect of different preprocessing methods on the classification results of the training sample set. Mean centering used as last method in all combinations. Results shown as correctly classified percentage %CC.

Preprocessing methods	Training set	Cross validation
SNV	100	86.5
1st order derivation	98.7	85.2
2nd order derivation	98.7	87.8
1st order derivation, SNV	97.3	82.4
2nd order derivation, SNV	98.7	87.8

7.3 Pixels vs object

Taking images with hyperspectral camera provides spatial and spectral information of each separate pixel. It would seem useful to classify each pixel separately. Using average spectra of the samples loses the spatial information and treats the samples in object level. Figure 6.4 showed that preprocessing the sample average absorbance spectra with SNV and mean centering finds differences between intrinsic viscosities. Choosing to use all the sample pixels and calculating the average spectra provided the highest classification accuracy.

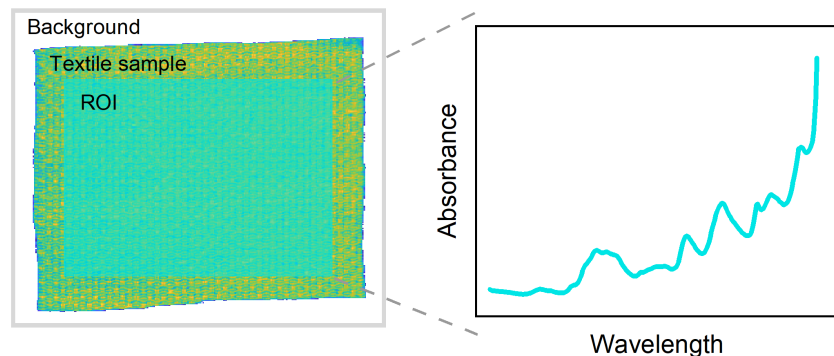


Figure 7.2. Calculation of average absorbance spectrum of chosen sample ROI.

Instead of using all the sample pixels to calculate the average spectra, some region of interest (ROI) can be chosen. Figure 7.2 shows an example of a sample and the 200x200 pixel ROI that is used to calculate the average absorbance spectrum. Using same sized ROI of all samples instead of all the sample pixels to calculate the average spectra makes

the further analysis of all the samples more comparable. Classifying each pixel separately was shown to create too much noise to be able to classify the samples. The effect of decreasing the ROI size on the average spectra of the first sample set is shown in Figure 7.3.

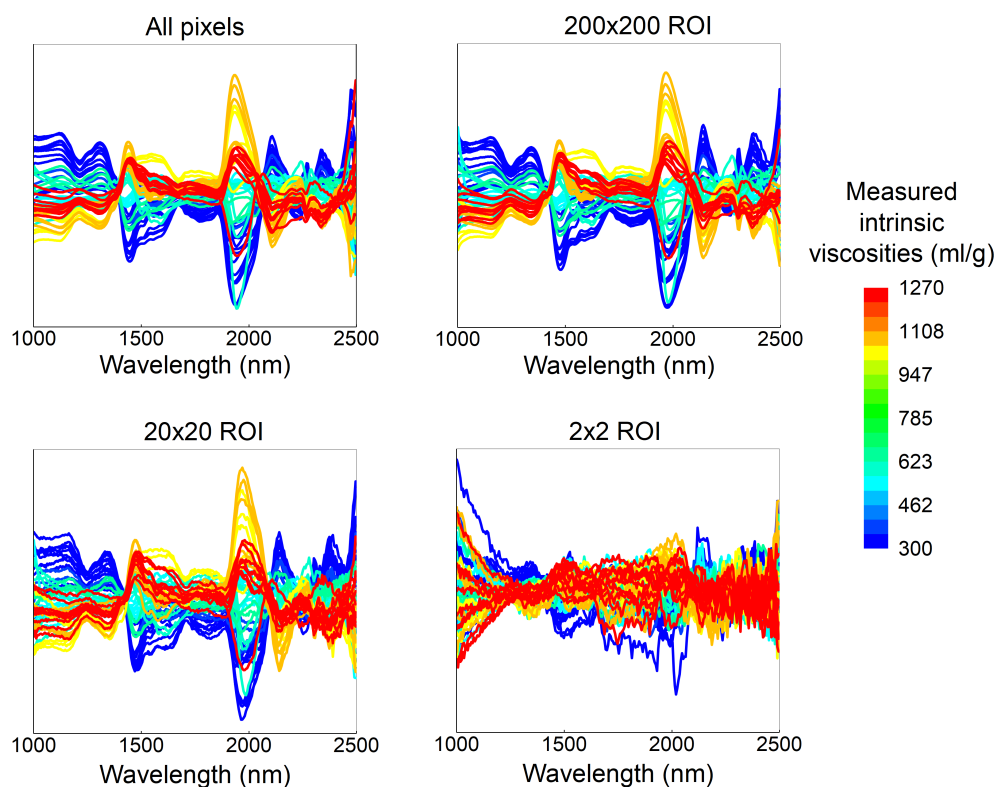


Figure 7.3. Effect of changing the ROI size on the sample average spectra of the first sample set.

Quite small ROI sizes still creates discrimination between different viscosities. Classification results for training set and training set 3-fold cross-validation are shown in Table 7.2. Results in the table are calculated with one average spectrum per each sample, and spectra are preprocessed using SNV and mean centering. 200x200 ROI shows the highest cross-validation accuracy, but lower training set accuracy compared to results using all pixels. Choosing multiple small ROIs from one sample could make it possible to classify small areas in the samples.

The intrinsic viscosities of the samples are correlated to the cellulose polymer fiber lengths, which are molecular scale properties. Pixel level calculations are possible when studying larger scale properties than intrinsic viscosity of similar textile samples. Mäkelä et. al showed that polyester content of waste textiles can be predicted in pixel level with promis-

ing accuracy [2]. Different textile fibers have different chemical compositions, which can be more clearly detected from hyperspectral image pixel spectra.

Table 7.2. Effect of changing the ROI size on the training sample set and training set 3-fold cross-validation classification results. Results shown as correctly classified percentage CC%.

ROI size	Training set	Cross-validation
All pixels	100	86.5
200x200	98.7	90.6
20x20	98.7	86.5
2x2	97.3	82.4

7.4 The number of used principal components

To be able to conduct LDA, dimension of the average sample spectra wavelength variables was reduced using PCA and the optimal number of used principal components was determined using cross-validation. It was shown that increasing the number of used principal components started to form denser clusters of the training samples in each class. At the same time after certain point the accuracy of test sample classification decreased. Choosing too few or too many principal components caused under- and overfitting of the classification model.

Figure 7.4 shows the classification results of training and test sets with 5, 40 and 70 used principal components. The class regions and samples are colored similarly as in Figures 6.7 and 6.8. Figure 7.4 shows the formation of denser clusters of the training samples when increasing the number of principal components and the simultaneous decrease in the test set classification accuracy. Using too few principal components did not find the underlying features in the average sample spectra for efficient classification.

Table 7.3 shows the classification results of training and test sets with different number of used principal components as correctly classified percentages %CC. Promising classification results were achieved with the lowest number of principal components that provided 100 % correctly classified rate for the training set cross-validation, which was 34.

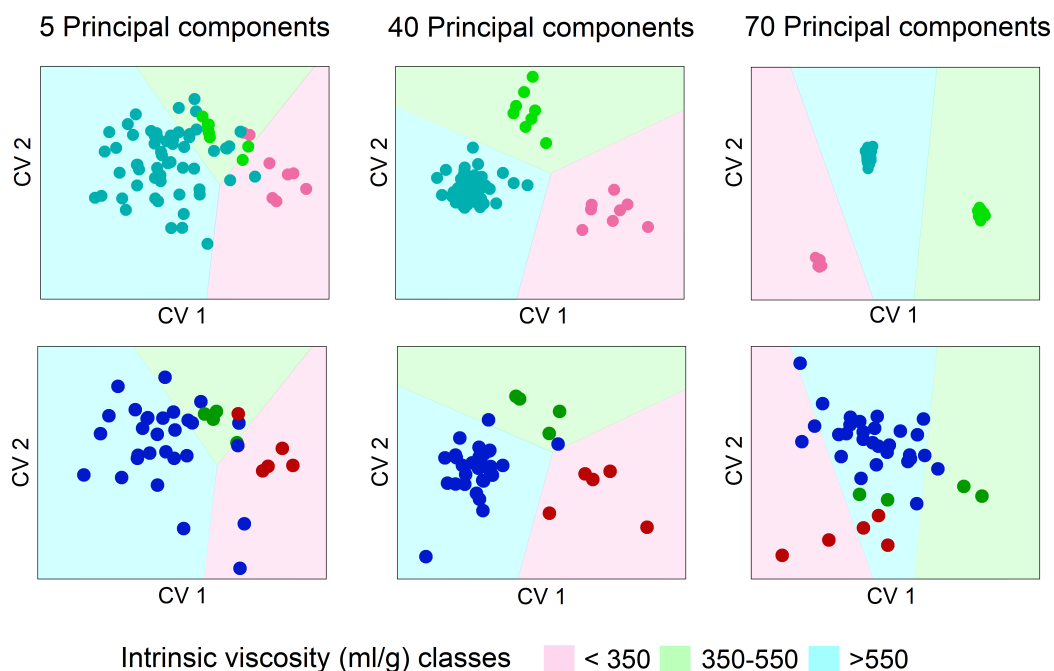


Figure 7.4. Effect of changing the number of used principal components on the classification results. Top row images show the training set classification results and bottom row the test set results.

Table 7.3. Effect of changing the number of used principal components in the classification model. Classification results shown as correctly classified percentage CC%.

Number of used PC's	Training set %CC	Test set %CC
5	78.4	73.7
34	100	94.7
40	100	94.7
70	100	79.0

7.5 Wavelength selection

The interactions between light and materials depend on the wavelength of the light and the composition of the materials. Choosing the wavelengths to the analysis that are the most relevant for the specific sample analysis can improve the results. Wavelength selection was studied in this thesis.

Wavelength region was divided into 10 intervals, each including 27 wavelength bands. Classification accuracies were calculated for every possible wavelength interval combination to find the intervals that contribute most on the classification results. The number of

used wavelength intervals was changed in each iteration. 20 % of the wavelength combinations that produced the best correctly classified rate in each iteration were chosen and the relative occurrence of each interval was calculated. Wavelength selection showed that there were intervals which contributed more on the results than others. Results of the wavelength selection are shown in Figure 7.5. The figure is colored based on the relative occurrence of each interval in each iteration. The figure also shows the division of the wavelength region to the intervals.

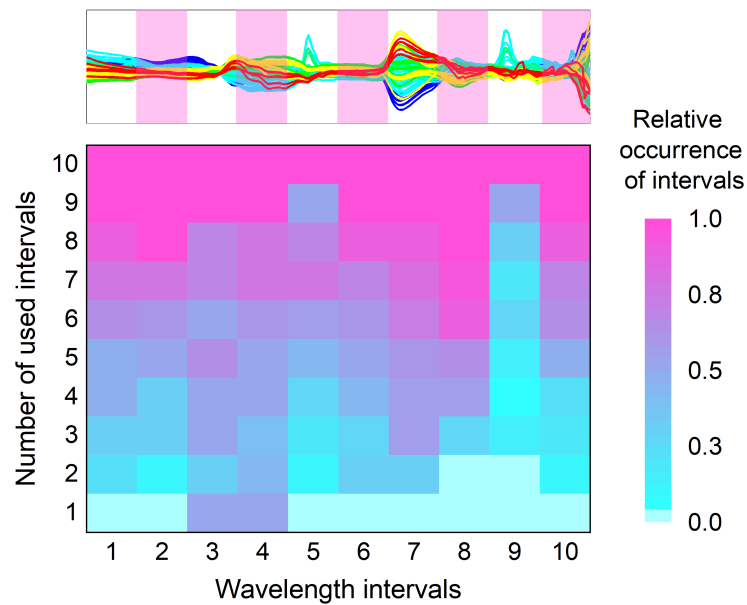


Figure 7.5. Wavelength selection results. Figure colored based on the relative occurrence of intervals that produced the best classification accuracies during the wavelength selection iterations.

Choosing wavelength intervals to train the classification model did not clearly improve the classification results. The classification results of the training set and training set 3-fold cross-validation with some combinations of used wavelength intervals are shown in Table 7.4. Wavelength combinations used in the table are chosen based on the wavelength selection results seen in Figure 7.5.

The cross-validation results were slightly improved when choosing for example wavelength intervals 7 and 8 instead of using the whole wavelength region. Training set classification accuracies were decreased. Wavelength selection was also studied on spectra with derivative preprocessing methods. The classification results were not improved using derivative method and wavelength selection.

Table 7.4. Effect of wavelength selection on the classification results of the training sample set. Results are shown as correctly classified percentage %CC.

Used wavelength intervals	Training set	Cross-validation
7 8	97.3	90.6
3 4 7 8	97.3	90.5
3 4 6 7 8	97.3	91.9
2 3 4 6 7 8	97.3	87.8

8. CONCLUSION

In this thesis cellulose-based textile samples were classified based on their intrinsic viscosity values using hyperspectral imaging and chemometric methods. Aim was to find simple methods to produce accurate classification results. Using simple methods made the interpretation of the results easier. It was shown that the samples could be classified with promising accuracy. Hyperspectral camera used in this thesis operated in the near infrared wavelength region. The interactions between light and samples in NIR region are complex and hard to interpret. This makes the evaluation and the analysis of the absorbance spectra difficult. Chemical recyclers are interested in distinguishing textile samples with intrinsic viscosities between 350 ml/g and 550 ml/g out of other samples. Intrinsic viscosities of the textile samples were measured using SCAN-CM 15:88 standard. Being able to sort cellulose-based textiles based on the intrinsic viscosities helps to control the viscoelastic properties of dissolved fibers during the chemical recycling process.

The samples consisted of two different sets of cellulose-based textiles. PCA showed differences between the two sample sets. According to principal component loadings the largest variance in the spectra could be assigned to water and cellulose. Despite the differences in the chemical composition of the two sample sets, they could be classified based on their intrinsic viscosity values.

Raw hyperspectral images consisted of a lot of information and the relevant information needed to be extracted. After reflectance and absorbance transformations and background separation, standard normal variate transformation with mean centering were shown to generate differences between intrinsic viscosities of the sample average spectra and to produce promising classification accuracies. Also, Savitsky-Golay derivative methods were applied but they did not clearly improve the classification accuracy and made the interpretation of the spectra more difficult.

Classification model created for this thesis was determined using linear discriminant analysis, which could not be conducted when the number of wavelength variables exceeded the number of samples. Dimensions of the wavelength variables were reduced using principal component analysis. The number of used principal components was shown to have an important effect on the classification accuracy. Using too many principal components caused overfitting of the model and using too few caused underfitting. Increasing the number of used principal components made the training samples of each class to form denser clusters. At the same time the accuracy of the test set classification decreased. The number of used principal components was calculated using 3-fold cross-validation for the training sample set and the smallest number of principal components that produced 100 % correctly classified rate was chosen.

Linear class boundaries were calculated using linear discriminant analysis and canonical variables. Samples were classified to three viscosity classes. The samples were not distributed evenly to each classification class. Class with intrinsic viscosities >550 ml/g included 78 % of the samples, and rest of the samples divided to 350-550 ml/g and <350 ml/g viscosity classes. Canonical variables found the maximum variance between classes and minimum variance within a class. The final classification model correctly classified 100 % of the training samples and 94.7 % of the test samples.

The samples were classified using average absorbance spectrum of all image pixels of each sample. The classification accuracy decreased with the decrease in ROI size. It was shown that pixels alone were too noisy to be classified correctly. Small ROI sizes still produced promising classification results, which showed that small areas from textile samples could be possible to classify. Previously it has been shown that different cellulose fibers could be classified on a pixel level with promising accuracy [14]. Bigger differences in the chemical structure makes also the differences in the absorbance spectra bigger, which enables the pixel level classification.

As far as we know, this was the first time it was shown that textiles could be classified based on their intrinsic viscosity values. Combining the classification model classifying textiles based on their intrinsic viscosity values with classification models classifying samples based on other properties would make the waste textile sorting more efficient. In practice, textile sorting in-line applications could consist of object detection and classi-

fication models. Textiles could be detected as objects, classified by fiber types and then different detected textiles classified further, for example, based on the intrinsic viscosity. Classification model determined for this thesis should be trained and validated with bigger sample set to see the effects on the classification results and to rebalance the class sizes.

Being able to sort cellulose-based textiles based on the intrinsic viscosities helps to control the chemical recycling process of waste cellulose textiles. Increasing the reuse of waste textiles makes the production of new textiles more sustainable and saves natural resources. Recycling waste textiles is a sustainable way to meet the growing demand of new textiles.

REFERENCES

- [1] *Kuitutilastot*. Tech. rep. Accessed: 2021-12-22. 2021. URL: <https://stjm.s3.eu-west-1.amazonaws.com/uploads/20210429144931/Kuitutilastot-2021-pdf.pdf>.
- [2] Mäkelä, M., Rissanen, M. and Sixta, H. Machine vision estimates the polyester content in recyclable waste textiles. *Resources, conservation and recycling* 161 (Oct. 2020), p. 105007. URL: <https://dx.doi.org/10.1016/j.resconrec.2020.105007>.
- [3] Dahlbo, H., Aalto, K., Eskelinen, H. and Salmenperä, H. Increasing textile circulation—Consequences and requirements. *Sustainable Production and Consumption* 9 (Jan. 2017), pp. 44–57. URL: <https://dx.doi.org/10.1016/j.spc.2016.06.005>.
- [4] *Guidance for textile collection*. Tech. rep. Accessed: 2022-01-10. 2019. URL: <http://www.ecap.eu.com/wp-content/uploads/2019/07/Guidance-for-Textiles-Collections.pdf>.
- [5] Haslinger, S., Wang, Y., Rissanen, M., Lossa, M. B., Tantt, M., Ilen, E., Määttä, M., Harlin, A., Hummel, M. and Sixta, H. Recycling of vat and reactive dyed textile waste to new colored man-made cellulose fibers. *Green chemistry : an international journal and green chemistry resource : GC* 21.20 (Oct. 2019), pp. 5598–5610. URL: https://explore.openaire.eu/search/publication?articleId=dedup_wf_001::52c108a2dfa14987503a598ec6806abc.
- [6] Jiang, F. Cellulose Fibers. *Handbook of Fibrous Materials*. John Wiley & Sons, Ltd, 2020. Chap. 4, pp. 95–124. ISBN: 9783527342587. DOI: <https://doi.org/10.1002/9783527342587.ch4>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527342587.ch4>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527342587.ch4>.

- [7] George, J. and Sabapathi, S. N. Cellulose nanocrystals: synthesis, functional properties, and applications. *Nanotechnology, science and applications* 8 (2015), pp. 45–54. URL: <https://www.ncbi.nlm.nih.gov/pubmed/26604715>.
- [8] Sixta, H., Michud, A., Hauru, L., Asaadi, S., Ma, Y., King, A. W. T., Kilpeläinen, I. and Hummel, M. Ioncell-F: A High-strength regenerated cellulose fibre. *Nordic pulp & paper research* 30.1 (Jan. 2015), pp. 43–57. URL: <http://www.degruyter.com/doi/10.3183/npprj-2015-30-01-p043-057>.
- [9] Manley, M. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. English. *Chemical Society reviews* 43.24 (Nov. 2014), pp. 82–8214. URL: <https://www.ncbi.nlm.nih.gov/pubmed/25156745>.
- [10] Blanch-Perez-del-Notario, C., Saeys, W. and Lambrechts, A. Hyperspectral imaging for textile sorting in the visible–near infrared range. *Journal of spectral imaging* 8.1 (Oct. 2019), a17. URL: <https://explore.openaire.eu/search/publication?articleId=doajarticles::ab0ad10a6603c57ded85819769ba5468>.
- [11] Mirschel, G., Daikos, O., Scherzer, T. and Steckert, C. Near-infrared chemical imaging used for in-line analysis of functional finishes on textiles. *Talanta (Oxford)* 188 (Oct. 2018), pp. 91–98. URL: <https://dx.doi.org/10.1016/j.talanta.2018.05.050>.
- [12] Haslinger, S., Hietala, S., Hummel, M., Maunu, S. L. and Sixta, H. Solid-state NMR method for the quantification of cellulose and polyester in textile blends. *Carbohydrate polymers* 207 (Mar. 2019), pp. 11–16. URL: <https://dx.doi.org/10.1016/j.carbpol.2018.11.052>.
- [13] Huang, J. and Yu, C. Fiber content determination of linen/viscose blends using NIR spectroscopy. *Bioresources* 15.2 (Mar. 2020), pp. 3006–3016. URL: <https://search.proquest.com/docview/2519839053>.
- [14] Mäkelä, M., Rissanen, M. and Sixta, H. Identification of cellulose textile fibers. *Analyst (London)* 146.24 (Dec. 2021), pp. 7503–7509. URL: <https://www.ncbi.nlm.nih.gov/pubmed/34766958>.
- [15] Mäkelä, M., Geladi, P., Rissanen, M., Rautkari, L. and Dahl, O. Hyperspectral near infrared image calibration and regression. *Analytica chimica acta* 1105 (Apr. 2020), pp. 56–63. URL: <https://dx.doi.org/10.1016/j.aca.2020.01.019>.

- [16] Burger, J. *Hyperspectral NIR image analysis*. Vol. 2006:60. Acta Universitatis agriculturae Sueciae. Umea: Unit of Biomass Technology and Chemistry Swedish University of Agricultural Sciences Umea, 2006. ISBN: 9789157671097. URL: <http://libris.kb.se/resource/bib/10217716>.
- [17] Boldrini, B., Kessler, W., Rebner, K. and Kessler, R. W. Hyperspectral Imaging: A Review of Best Practice, Performance and Pitfalls for in-line and on-line Applications. English. *Journal of near infrared spectroscopy (United Kingdom)* 20.5 (Oct. 2012), pp. 483–508. DOI: 10.1255/jnirs.1003. URL: <https://journals.sagepub.com/doi/full/10.1255/jnirs.1003>.
- [18] Ravikanth, L., Jayas, D. S., White, N. D. G., Fields, P. G. and Sun, D.-W. Extraction of Spectral Information from Hyperspectral Data and Application of Hyperspectral Imaging for Food and Agricultural Products. *Food and bioprocess technology* 10.1 (Nov. 2016), pp. 1–33. URL: <https://link.springer.com/article/10.1007/s11947-016-1817-8>.
- [19] SpecimSpectral. *How Specim line scan hyperspectral cameras work*. Youtube. 2019. URL: <https://www.youtube.com/watch?v=4JvKrBaxFgY>.
- [20] Paviotti, A., Ratti, F., Poletto, L. and Cortelazzo, G. M. Multispectral Acquisition of Large-Sized Pictorial Surfaces. *EURASIP journal on image and video processing* 2009.1 (Nov. 2009), pp. 1–17. URL: <https://link.springer.com/article/10.1155/2009/793756>.
- [21] Makarem, M., Lee, C. M., Kafle, K., Huang, S., Chae, I., Yang, H., Kubicki, J. D. and Kim, S. H. Probing cellulose structures with vibrational spectroscopy. *Cellulose (London)* 26.1 (Jan. 2019), pp. 35–79. URL: <https://search.proquest.com/docview/2181120060>.
- [22] Larkin, P. *Infrared and raman spectroscopy principles and spectral interpretation*. eng. 2011008524. 2011.
- [23] Diem, M. *Modern Vibrational Spectroscopy and Micro-Spectroscopy*. Chicester: Wiley, 2015. ISBN: 9781118824863. URL: [https://ebookcentral.proquest.com/lib/\[SITE_ID\]/detail.action?docID=1895629](https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=1895629).
- [24] Nicodemus, F. E. *Geometrical considerations and nomenclature for reflectance*. Vol. 160. Washington, DC: US Dep. of Commerce, Nat. Bureau of Standards, 1977.

- [25] Ferrero, A., Rabal, A. M., Campos, J., Pons, A. and Hernanz, M. L. Spectral and geometrical variation of the bidirectional reflectance distribution function of diffuse reflectance standards. *Applied optics. Optical technology and biomedical optics* 51.36 (Dec. 2012), pp. 8535–8540. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23262591>.
- [26] Kim, D. B., Jang, I. Y., Choi, H. K. and Lee, K. H. Recovery and representation of spectral bidirectional reflectance distribution function from an image-based measurement system. *Color research and application* 41.4 (Aug. 2016), pp. 358–371. URL: <https://api.istex.fr/ark:/67375/WNG-JVB0914C-W/fulltext.pdf>.
- [27] Zhu, J., Wang, K., Liu, H., Du, B. and Guo, F. Modified model of polarized bidirectional reflectance distribution function for metallic surfaces. *Optics and laser technology* 99 (Feb. 2018), pp. 160–166. URL: <https://dx.doi.org/10.1016/j.optlastec.2017.08.029>.
- [28] Hollas, J. M. *Modern spectroscopy*. 4. ed., reprinted. Chichester [u.a.]: Wiley, 2008. ISBN: 9780470844168.
- [29] Ozaki, Y. Near-Infrared Spectroscopy—Its Versatility in Analytical Chemistry. *Analytical sciences* 28.6 (May 2012), pp. 545–563. URL: https://www.jstage.jst.go.jp/article/analsci/28/6/28_545/_article/-char/en.
- [30] Morais, C. L. M., Lima, K. M. G., Singh, M. and Martin, F. L. Tutorial: multivariate classification for vibrational spectroscopy in biological samples. English. *Nature protocols* 15.7 (July 2020), pp. 2143–2162. URL: <https://www.ncbi.nlm.nih.gov/pubmed/32555465>.
- [31] Pasquini, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica chimica acta* 1026 (Oct. 2018), pp. 8–36. URL: <https://dx.doi.org/10.1016/j.aca.2018.04.004>.
- [32] Schwanninger, M., Rodrigues, J. C. and Fackler, K. A Review of Band Assignments in near Infrared Spectra of Wood and Wood Components. *Journal of near infrared spectroscopy (United Kingdom)* 19.5 (Oct. 2011), pp. 287–308. URL: <https://journals.sagepub.com/doi/full/10.1255/jnirs.955>.
- [33] Burger, J. and Geladi, P. Hyperspectral NIR image regression part I: calibration and correction. English. *Journal of chemometrics* 19.5-7 (May 2005), pp. 355–363.

- DOI: 10.1002/cem.938. URL: <https://api.istex.fr/ark:/67375/WNG-H15MFCPF-0/fulltext.pdf>.
- [34] Burger, J. and Geladi, P. Hyperspectral NIR image regression part II: dataset pre-processing diagnostics. English. *Journal of chemometrics* 20.3-4 (Mar. 2006), pp. 106–119. URL: <https://api.istex.fr/ark:/67375/WNG-0V9NHK50-F/fulltext.pdf>.
- [35] Wise, B. M. and Gallagher, N. B. The process chemometrics approach to process monitoring and fault detection. *Journal of process control* 6.6 (1996), pp. 329–348. URL: [https://dx.doi.org/10.1016/0959-1524\(96\)00009-1](https://dx.doi.org/10.1016/0959-1524(96)00009-1).
- [36] Geladi, P. *Chemometrics in spectroscopy. Part 1. Classical chemometrics*. 2003. URL: [https://dx.doi.org/10.1016/S0584-8547\(03\)00037-5](https://dx.doi.org/10.1016/S0584-8547(03)00037-5).
- [37] Brereton, R. G. A short history of chemometrics: a personal view. *Journal of chemometrics* 28.10 (Oct. 2014), pp. 749–760. URL: <https://api.istex.fr/ark:/67375/WNG-16P53442-R/fulltext.pdf>.
- [38] GELADI, P., SETHSON, B., NYSTROM, J., LILLHONGA, T., LESTANDER, T. and BURGER, J. Chemometrics in spectroscopy Part 2. Examples. *Spectrochimica acta. Part B: Atomic spectroscopy* 59.9 (Sept. 2004), pp. 1347–1357.
- [39] Dotto, A. C., Dalmolin, R. S. D., Caten, A. ten and Grunwald, S. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. *Geoderma* 314 (Mar. 2018), pp. 262–274. URL: <https://dx.doi.org/10.1016/j.geoderma.2017.11.006>.
- [40] Barnes, R. J., Dhanoa, M. S. and Lister, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. English. *Applied spectroscopy* 43.5 (July 1989), pp. 772–777. DOI: 10.1366/0003702894202201. URL: <http://www.ingentaconnect.com/content/sas/sas/1989/00000043/00000005/art00007>.
- [41] Savitzky, A. and Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical chemistry (Washington)* 36.8 (July 1964), pp. 1627–1639. URL: <http://dx.doi.org/10.1021/ac60214a047>.

- [42] Abdi, H. and Williams, L. J. Principal component analysis. *Wiley interdisciplinary reviews. Computational statistics* 2.4 (July 2010), pp. 433–459. URL: <https://api.istex.fr/ark:/67375/WNG-38VPNXS4-W/fulltext.pdf>.
- [43] Bro, R. and Smilde, A. K. Principal component analysis. *Analytical methods* 6.9 (2014), pp. 2812–2831. URL: <https://pubs.rsc.org/en/content/articlehtml/2014/ay/c3ay41907j>.
- [44] Burger, J. and Gowen, A. Data handling in hyperspectral image analysis. *Chemometrics and intelligent laboratory systems* 108.1 (2011), pp. 13–22. URL: <https://dx.doi.org/10.1016/j.chemolab.2011.04.001>.
- [45] Nørgaard, L., Bro, R., Westad, F. and Engelsen, S. B. A modification of canonical variates analysis to handle highly collinear multivariate data. *Journal of chemometrics* 20.8-10 (Aug. 2006), pp. 425–435. URL: <https://api.istex.fr/ark:/67375/WNG-JVCJM89G-C/fulltext.pdf>.
- [46] Duda, R. O., Hart, P. E. and Stork, D. G. *Pattern classification*. 2nd edition. New York [u.a.]: Wiley, 2000, pp. 121–124. ISBN: 9781118586006.
- [47] Vandeginste, B. G., Massart, D. L., Buydens, L. M., De Jong, S., Lewi, P. J. and Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part B*. Vol. Part A. Data handling in science and technology. Oxford: Elsevier Science & Technology, 1998, pp. 1–713.
- [48] Brereton, R. G. Contingency tables, confusion matrices, classifiers and quality of prediction. *Journal of chemometrics* 35.11 (Nov. 2021), n/a. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.3331>.
- [49] Brereton, R. G. One-class classifiers. *Journal of chemometrics* 25.5 (May 2011), pp. 225–246. URL: <https://api.istex.fr/ark:/67375/WNG-K3X309KD-6/fulltext.pdf>.
- [50] *SCAN-15:88: 1988: Pulps. Viscosity in cupri-ethylenediamine solution*. 1988.
- [51] *Specim, Spectral Imaging Ltd*. <https://www.specim.fi/hyperspectral-cameras-us/> Tiedosto. [Accessed October 19th 2021].
- [52] *svd (Singular value decomposition)*. <https://se.mathworks.com/help/matlab/ref/double.svd.html>. Accessed: 2021-12-10.
- [53] *canoncorr (Canonical correlation)*. <https://se.mathworks.com/help/stats/canoncorr.html>. Accessed: 2021-12-10.

- [54] *classify (Classify observations using discriminant analysis)*. <https://se.mathworks.com/help/stats/classify.html>. Accessed: 2021-12-21.
- [55] *PLS-toolbox, Advanced Chemometrics Software for use with MATLAB®*. <https://eigenvector.com/software/pls-toolbox/>. Accessed: 2021-12-13.
- [56] Megahed, F. M., Chen, Y.-J., Megahed, A., Ong, Y., Altman, N. and Krzywinski, M. The class imbalance problem. *Nature methods* 18.11 (Nov. 2021), pp. 1270–1272. URL: <https://www.ncbi.nlm.nih.gov/pubmed/34654918>.