UNIVERSITY OF VAASA

FACULTY OF TECHNOLOGY

AUTOMATION

Tommi Rintala

NEAR-INFRARED SPECTROSCOPY IN FOOD QUALITY CONTROL

Master's thesis for the degree of Master of Science in Technology submitted for inspection, Vaasa, 18 December 2012.

Supervisor	Jarmo Alander
Instructors	Vladimir Bochko, Birgitta Martinkauppi

VAASAN YLIOPISTO
Teknillinen tiedekunta
Tekijä: Tommi Rintala
Diplomityön nimi: Lähi-infrapuna-alueen spektroskopia ruuan laadun valvonnassa
Valvojan nimi: Jarmo Alander
Ohjaajan nimi: Vladimir Bochko, Birgitta Marginkauppi
Tutkinto: Diplomi-insinööri
Koulutusohjelma: Sähkö- ja energiatekniikan koulutusohjelma
Suunta: Automaatiotekniikka
Opintojen aloitusvuosi: 2005
Diplomityön valmistumisvuosi: 2012 **Sivumäärä: 113**

TIIVISTELMÄ:
Ruuan laadunvalvonta on tärkeä aihe koko tuotantoketjussa pellolta kuluttajalle. Olemme tarkastelleet lähi-infrapuna-alueen spektroskopian käyttämistä suomalaisten elintarvikkeiden laadun analysoinnissa. Tavoitteena oli tutkia tomaattien luokittelua laadun perusteella sekä mustikkoiden antosyaanipitoisuuden määritystä. Tutkimme myös kemikaalipitoisuuksien mittaamista hyödyntäen lähi-infrapunaaluespektroskopiaa natrium bimetasulfaatilla ja morfiinilla.

Osoitimme että lineaariset regressiomallit, kuten PCA ja PLS, muodostavat hyvän perustan spektroskooppisen tiedon analysoinnille. Vertasimme perinteisiä regressiomalleja lineaarisiin monimuuttujamalleihin ja havaitsimme jälkimmäisten olevan helpompia käyttää, koska tällöin ei tarvitse etsiä ennen analyysiä mielenkiintoisia absorptiopiikkejä vaan ne saadaan selville yhdessä mallin tulosten kanssa. Osoitimme myös kuinka PLS malli ennustaa mustikoiden antosyaanipitoisuuksia.

AVAINSANAT: Lähi-infrapuna spektroskopia, analyysi, laadunvalvonta, elintarvikkeet

UNIVERSITY OF VAASA
Faculty of Technology

Author:	Tommi Rintala
Topic of the Thesis:	Near-infrared Spectroscopy in Food Quality Control
Supervisor:	Jarmo Alander
Instructor:	Vladimir Bochko, Birgitta Martinkauppi
Degree:	Master of Science in Technology
Degree Programme:	Degree Programme in Electrical and Energy Engineering
Major of Subject:	Automation
Year of Entering the University	2005
Year of Completing the Thesis:	2012 **Pages:** 113

ABSTRACT:

Food quality control is an important issue in wholo production chain, from soil to consumer. We have examined the usage of near-infrared spectroscopy in analysis of Finnish agriculture products, such as tomatoes, where we aimed to classify them based on the quality and anthocyanin content on blueberries. Our measurements of aqueous solutions, such as sodium bimetasulfide and morphine were used in study of applying near-infrared spectroscopy in concentration measurement of chemical.

We have shown that with linear multivariate methods, specially PCA and PLS, give good basis of the spectroscopy data analysis. On comparison of traditional regression models the linear multivariate methods were easier to use. There is no need to find important absorption peaks in spectral data, they are by-product of multivariate model. And we have illustrated how the PLS model predicts blueberry anthocyanin concentrations.

KEYWORDS: Near-infrared spectroscopy, analysis, quality control, food quality

TABLE OF CONTENTS Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

SYMBOLS AND ACRONYMS

A	Absorption intensity (vector)
c	Concentration (e.g. of sample liquid)
c_0	Speed of light in vacuum
c_m	Speed of light in matter
ξ	Molar extinction coefficient
ε	Error term, e.g. noise in measurement
f	Frequency
l	Path-length (of light ray)
λ	Wavelength (in nm)
h	Planck's constant $6.62606896 \times 10^{-34}\mathrm{Js}$
$\hbar$	Planck's constant $\frac{h}{2\pi} = 1.05457266 \times 10^{-34}\mathrm{Js}$
I_0	Intensity (vector)
I_B	Black reference intensity (vector)
I_W	White reference intensity (vector)
k_B	Boltzmann's Constant, $1.380658 \times 10^{-23}\mathrm{J/K}$
n, n_i	Index of refraction
ν	Vibrational frequency
R	Reflectance
R_n	Reflectance in wavelength n
R_W	Reflection coefficient (intensity)
S	Scattering
T	Transmission intensity (vector)
θ_i	Angle between light ray and perpendicular [to sample surface]

x_i	Measured reflectance or absorbbance on wavelength λ_i
$\bar{x}$	Mean vector x
v	Speed (of light)

CARI	Chlorophyll Absorption Ratio Index
DVI	Difference Vegetation Index
EVI	Enhanced Vegetation Index
IR	Infrared
FT-NIR	Fourier Transform Near-infrared Spectroscopy
LAI	Green Leaf Area Index
LV	Linear Variable
MAP	Modified-Atmosphere Package
MCARI	Modified Chlorophyll Absorption Ratio Index
MSAVI	Modified Soil-Adjusted Vegetation Index
MSR	Modified Simple Ratio
NAI	Normalised Anthocyanin Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIRS	Near-infrared Reflectance Spectroscopy
NIR	Near-infrared
NMI	Nuclear Magnetic Imaging
NMR	Nuclear Magnetic Resonance

PCA	Primary Component Analysis
PLS	Partial Least Squares
PRESS	Predicted Residual Estimated Sum of Squares
PVI	Perpendicular Vegetation Index
RDVI	Re-normalized Difference Vegetation Index
RMSEP	Root Mean Square Error of Prediction
SEP	Standard Error or Prediction (or Perfomance)
RII	Reflectance Integral Index
RVI	Ratio of Vegetation Index
SAVI	Soil-Adjusted Vegetation Index
SNR	Signal to Noise Ratio
SR	Simple Ratio
TA	Titratable Acidity
TMS	Tomato Maturity State
TVI	Transformed Vegetation Index
UV	Ultraviolet (light)
VI	Vegetation Index
VIS	Visual Spectra
VIS/NIR	Visible/Near-infrared
WI	Water Band Index
YI	Yellowness Index

PREFACE

I thank TEKES for financial support for this work via TarhaNIRSo -project carried out at University of Vaasa between 2010 and 2012. I wish to thank for my supervisor professor Jarmo Alander, and my instructors Prof. Vladimir Bochko and Prof. Birgitta Martinkauppi for guidance during writing of this thesis. Many thanks to prof. Paul Geladi for introducing me to the Chemometrics.

I would also like to thank my dear wife for all the support while I was occupying myself with something else when I should have been writing the thesis.

While working to finish this thesis, I learned a lot. Not everything is reflected from the thesis, and I strongly believe for the 80/20 rule; 80 percent of the work executed is shown only in 20 percent of the pages. But the lessons learned which are shown in the pages of this thesis are still lessons learned.

I believe that creativity and thought require some degree of indolence;

So, ye three Ghosts, adieu! Ye cannot raise

My head cool-bedded in the flowery grass;
For I would not dieted with praise,

A pet-lamb in a sentimental farse!
Fade softly from my eyes, and be once more

In masque-like figures on the dreamy urn;

Farewell! I yet have visions for the night,
And for the day faint visions there is store;

Vanish, ye Phantoms! from my idle spright,
Into the clouds, and never more return!

John Keats: Ode on Indolence, VI

1. INTRODUCTION

Analytical consumers make their food purchase decisions based on price and perceived quality. The perceived quality is based on previous experiences of the product, testimonial from other consumers and the product brand among other subjective measures.

Quality of food is extremely important for our health and well-being, at least if we believe the advertisements in media. The marketing industry is claiming that the quality of food products will give us long and enjoyable life full of happiness and enjoyment.

Producers wish to produce higher quality products because the customers, both consumers and food industry, are willing to choose the higher quality product if a choice is present. Inexpensive and effective methods allow better quality control.

Non-destructive quality control and defect detection methods are important application area for food industry. Applying destructive methods in sorting is not viable since if the food items are damaged or destroyed during the sorting they cannot be used in production. If, for example, chemical content analysis is performed using traditional mass spectrometry there is no food item to be used in food manufacturing process.

Since organic products, such as tomatoes, are unique individuals the [destructive] quality analysis explains only the individual tomatoes which have been investigated—there might still be faulty ones in the rest of the production batch. This would require the batch to be removed from production and therefore it would not produce any profit for manufacturer, which would also affect to the income of the farmer, eventually.

There are many different non-destructive analysis methods which all have one thing in common; they utilise non-hazardous method(s) to investigate and analyse the content, composition, or structure of product. Some features such as surface defects are easy to find visually and remove the faulty ones. Other defects such as nearly all internal defects require other methods, such as measuring the floating capabilities of potatoes and remove those ones, which float. But there are still other defects which are much harder to find and require more complex methods. These methods include applying X-rays, ultrasound, nuclear magnetic resonance, nuclear magnetic imaging, computer vision systems, acoustics and photospectrography. This thesis focuses on optical methods such

as Visible/Near-infrared (VIS/NIR) spectroscopy.

In the first Chapter, a short introduction to the relevant terminology and basic concepts is presented. We define the criteria which food quality control method should fulfil to be acceptable. The definition of quality is given with brief history of Near-infrared Reflectance Spectroscopy (NIRS). And in the end of the first Chapter the background of a anthocyanin measurement on blueberries is reviewed.

The second Chapter begins with physics concepts and background of photon radiation, which is the base of NIRS. The absorption, reflectance transmittance, and scattering of light are explained in terms of energy transfer. Next the mathematical tools and multivariate analysis are explained briefly together with chlorophyll indices. The Chapter ends with theory of spectrophotometry and a short walk-through of the devices used.

In third Chapter the measurements made for this thesis are explained. The measurement classes are defined, and the measurement conditions are described.

The fourth Chapter we present the analysis of the measurement data. Simple indice analysis and multivariate analysis methods, such as PCA and PLS are used. In the fifth Chapter the conclusions are drawn from the measurements and the analysis of the results.

My main contribution is to utilise near-infrared spectroscopy, as non-destructive quality control method, in quality control measurements for Finnish food producers and confirm that it can be applied in testing.

During the work I have learned a great deal more of multispectral measurement and analysis, signal processing and analysis and refreshed my knowledge of molecular vibration.

1.1. Related Work

It is easy to understand that food quality is an important topic and a huge amount of scientific research work has been carried out to analyse and improve the methods which are used to assure the quality of food. According to Statistics Finland in the year 2007 retail market revenue of fruits and vegetable sales was 1133.2 million euro in Finland alone. This means that market for food industry has big economical impact. The same applies to both human consumable food and feed for domesticated animals.

Using Google Scholar we can find out that there is quite active research work going on around NIR and food. In Fig. 1 we have a plot showing some most frequent food items (fish and milk) compared to those which we have worked with in this work (salad, tomato, blueberries, potato).

Figure 1. Google Scholar article count for search terms "+NIR +food" 2002–2011 on April 9th 2011.

Why there are so many articles on fish and milk and so little articles on blueberries? This is most likely due to pure volume; people nourish themselves more with fish and milk than with blueberries. Fish is a fresh product which does not stay safe to eat long after killed unless prepared, e.g. freezing. However, berries and vegetables are a good source of nutrition and interesting because of their vitamin and nutritional values.

Milk is one of FAO's favourite products in their work of raising the nutrition level of rural populations of the world. It is a volume product and the consumption of milk has grown in countries previously thought as poor, in Asian countries such as China, who can afford now to consume milk with their huge population. This is interesting, since frequency of lactose intolerance is higher in Asian (90 %) than in European countries (5–75 %) (T 1994). Blueberries and tomatoes are globally quite marginal products, therefore small number of research is understandable.

Applying the Google Scholar and for a simple query "food quality control" we will find

over two million articles on food quality control. This is such a huge amount that even if would be estimated that a 20 percent of the result would consist of overlapping or review articles, one would still be unable to analyse them all. Therefore some relevant research work may have been neglected.

According to study conducted by Kummu et al. (2012), one quarter of the globally produced food is lost within the food supply chain. They estimate that one billion people could be fed, if the amount of food loss could be halved. This is good motivation to think of ways to improve supply chain and methods for determining the food quality.

1.2. Methods for determining food quality

In this work we will demonstrate usage of NIR methods for food quality control. Reason for this is that we try to find methods which can be used in food industry in Finnish food production. To make sure that the methods found are possible to use by food industry they should comply with the following criteria:

1. measurement is possible [feasible] to do with resources available,
2. method produces objective, as comparable and repeatable, data where results can be extracted,
3. measurement is not destructive, ie. it does not destroy sample(s), and
4. measurement method is scalable, ie. it can be used from small scale up to mass production, depending on the production scale.

If measurement method requires expensive hardware or is otherwise difficult to conduct, it is obvious that it is not very feasible to apply early in production chain.

Repeatable and predictable results allow method to be used for quality control. If results achieved with a method are ambiguous or do not correspond with reality the method is not a good one.

Destructive method can only be used once for each sample. Therefore it is not suitable for continuous monitoring, which would be desirable in food production.

Measurement method which is feasible only on large number of samples is not good for small producer, and simple measurement which is not applicable for large number of samples is not good for big manufacturer.

1.3. Quality

What is the food quality all about? As Shewfelt (1999) points out, the term quality has been defined in many ways and in multiple contexts and therefore there is little agreement what it is, how it can be measured and what is the relationship between quality and consumer acceptability, but several opinions.

Defect in food is usually visually detectable; bruise in the surface, dis-continuance in texture, colourisation error, but it can also be internal. Examples of internal defects include e.g. error in taste or smell, caused by extraneous material. Also diseases, such as fungi, can cause defects to food. Wrong structure, e.g. too much fat in pork, can also be seen as a defect.

Quite often the quality is defined as the absence of defects (Shewfelt 1999). This may seem an easy measurement of quality but it arises nasty question; what are the defects whose absence we are measuring. To apply this kind of method we need to be able to decide whether defect exists or not. This sounds easy but requires that also the amount of defect is taken into consideration. If we look into harvested fruit with enough magnification we can quite often find something which can be categorised as a defect or contamination.

Another method of defining the quality is to list out all the relevant attributes which can be observed or measured and rank the quality as excellence in these variables. As Auerswald et al. (1999) found out, consumers are willing to pay for better flavor, or if the product is produced harmless to environment, there are less chemical residue, product has higher vitamin content or less nitrate, has been tested by government, has been produced using quality assurance, contains more fibre etc. Some of these attributes are non-existing, from a producers point-of-view, since they can be satisfied by other means than quality control carried out by producer.

As Arvanitoyannis et al. (1999) stated, proofing the wine authenticity is a complex problem, but each wine has it's own characteristics based on phenolic compounds and minerals. Each geographical location has its unique combination of soil and environmental characteristics and therefore the certification of geographical origin of wine is quite straightforward. The wine farm either is located within a certificated area and it has "correct chemical fingerprint", or is not.

When listing the attributes one must not forget, however, that food should also contain enough carbohydrates, proteins, fibre, vitamins and minerals. The main purpose of eating food should be filling the energy and nutrient needs of body.

The quality parameters for different kind of food can share the same parameters or they can be very different. Fresh-cut fruits and vegetables are highly perishable for handling defects (Watada & Qi 1999). Beef and poultry may have much longer freshness period with them—if stored correctly(Lin et al. 2004).

People, as consumers, use all their senses to evaluate quality: sight, smell, sense, taste, touch and even hearing (Abbott 1999; Mattheis & Fellman 1999). Sound can be even used to analyse ripeness of apples as showed by Belie et al. (2000b). But, as is commonly known human perception is not perfect and even there exists some unique individuals who have knowledge and experience to rate food quality that is not the solution for food industry.

Preharvesting affects on flavor of fresh fruits (Ferguson et al. 1999; Mattheis & Fellman 1999). As Nerd & Mizrahi (1999) demonstrated with yellow pitaya, the ripening in the plant produced more solubles sugars and acidity than the fruits which were harvested early and ripened with controlled temperature.

Polder et al. (2004) showed that the surface distribution of carotenes and cholorophyll correlates with ripening of the tomatoes. Later the emission ratio of carotenoids and chlorophyll (R_{525}/R_{672}) was demonstrated by (Lai et al. 2007) to indicate tomato ripeness. According to Lai et al. (2007), carotenoids increase ten times and chlorophyll decrease during ripening.

As Kays (1999) stated the appearance is a primary criterion for customer when they make

purchase decision. The appearance in mainly composed of size, shape, form, colour, condition and absence of defects. He also classified preharvest factors into (1) biological factors [pathological, entomological, animal], (2) physiological factors [physiological disorders, nutritional imbalances, maturity], (3) environmental/cultural factors [climate, weather, soils, water relations, light intensity, pruning, thinning, bagging, herbicide injury], (4) mechanical damage [weather damages; chilling, burning, freezing, frost, high temperature stress, wind damage, hail damage, flooding soils, pollution], (5) extraneous matter [growing medium, vegetable matter, chemical residues]; and (6) genetic variation and aberrations.

Sams (1999) described postharvest factors which may affect on texture of food and categorised them as to: (1) abiotic [soil moisture, temperature, relative humidity and nutrient availability] factors, (2) genetic influence [cultivates within a species vary significantly in textural attributes]; and (3) preharvest environmental, cultural, physiological and genetic factors. First category includes items which are attended by farmer and other primary producers by changing attributes of environment, second category can be altered with selection of species to use in production and third category requires also other actors to co-operate.

Post-harvest factors such as temperature and humidity (Paull 1999) have big an effect on food quality. Artificial storage atmospheres (Artés et al. 1999), such as ethylene (Saltveit 1999) oxygen, and carbon-dioxide (Beaurdy 1999; Siomos et al. 2001) can help to maintain the quality. Post-processing handling and packaging (Barth & Zhuang 1996; Zagory 1999) either maintains or destroys quality already achieved previously.

The importance of bacterial pathogen control was raised by Brackett (1999). The work starts before planting and continues to the end of the delivery chain of the final product. Some products can be packaged e.g. in Modified-Atmosphere Package (MAP) or applied other protective measures to minimise the bacterial effect.

Quality can also be seen as a time-dependent variable. Naturally the storage time affects on fruit texture (Johnston et al. 2001), but also other factors, such as harvesting time, are time dependent. Some food items, such as vegetables, do not survive well long storage. For some food items, the optimal harvesting season is short and harvesting outside of the

optimal season degrades quality.

From this we can conclude that attributes for food quality control methods should be divided into three categories; (1) process, (2) product and (3) consumer oriented attributes.

The process oriented attributes of quality control include methods which are non-destructive, scalable, allow usage of statistical methods, can be applied for production management, have fast response, can be used to loop control back to production system on-line and allow real-time [quality] control.

Product oriented attributes of quality control method attributes include safety, nutrient level, colour, handling resistance, composition, and ripeness.

Customer oriented attributes of quality control method attributes include usage friendliness, taste, appeal, odour, structure, texture, feel, aroma, appearance, flavor, chewing sounds, hand-feel, and mouth-feel. Most of them are inter-related so that one cannot exists without another.

The quality control method attributes contribute to the food quality attributes. A summary of food quality attributes is shown in Table 1. The table includes weather the attribute is immediately interesting to the actor in question. Direct interest, (x) in table, means that the quality attribute is meaningful to the actor, such as for example size is important to the cucumber farmer because of the rules define that if cucumbers (*Cucumis sativus* L.) do not fulfil measurements for size and shape it will not be classed as *Extra*, but more likely to *Class I* or *Class II* (EEC 1677/88 1988).

A summary of the previous quality factors has been shown in Table 2. The summary lists those factors which affect to the food quality and which can be changed in each state of food item life span. First state “pre harvest” refers to agricultural decisions

Table 1. Food quality attributes affected to interest groups.

	Consumer	Market	Farmer	Manufacturer
absence of defects	x	x	x	
appearance	x	x		
colour	x	x		
condition	x			
firmness	x			
form	x			
shape	x		x	
size	x		x	x
taste	x	x		
texture	x			x

Table 2. Summary of the effect of quality factors in the food production chain. (x = direct causality, + = can be improved, †= negative effect, * = may affect, o = no effect)

	pre harvest	post harvest	process	delivery chain	usage
animals	x	x	*	*	*
entomological	x	x	o	o	
nutritional	+	+	+	o	o/†
physiological	x	x	x	x	+/o
pruning	x	o	o	o	o
thinning	x	o	o	o	o
pollution	x	x	*	*	*
temperature	x	x	+	+	+
time	x	x	x	+	+
maturity	x	x	x	x	x
condition	x	x	x	x	x
humidity	x	x	x	*	o
storage atmosphere	o	*	*	*	o
condition	x	x	x	x	o
bagging/packaging	o	*	+/*	+	o
pathological factors	x	x	o	o	o
physical disorders	x	x	*	o	o
light intensity	x	x	*/†	*/†	†/o
biological	x	x	*	*	o
environmental	x	x	*	*	*
extraneous matter	x	x	x	x	x
genetic variance	x	x	o	o	o
mechanical damage	x	x	x	x	x

1.4. Visible and Near-infrared Spectrum Methods in Food Quality

The first published NIRS application in 1949 was a measurement for using visible to short-wavelength NIR radiation to grade eggs (Workman & Weyer 2008: 108). During

the 20th century applying NIRS in food quality control has evolved from experiments in laboratory environment to actual applications with increasing accuracy (Cen & He 2007). The NIRS is applied in quality control of pears (Liu et al. 2008), peaches (Fu et al. 2008), maize (Volkers et al. 2003), beef, (Prieto et al. 2009), and figs (Burks et al. 2000).

Visual Spectra (VIS) is used in colour measurement of food in general (McCaig 2002), potato chips (Pedreschi et al. 2006), cherries Crisosto et al. (2002), dough proofing (Sinelli et al. 2008a), sweet potatoes' starch physiochemical quality and pasting properties (Lu et al. 2006) and many more.

There are several articles on applying VIS/NIR in identification such as Casale et al. (2010) on identification of table olive cultivates, milk powder protein content such as Chang et al. (2007), and applying hyper-spectral reflectance imaging in poultry, such as Cho et al. (2007) on detection of organic residues.

Peirs (2000) also applied NIRS for predicting the optimal picking date of apple cultivates. Reflectance spectroscopy is also applied in defect detection, such as brown heart detection in pears by Zerbini et al. (2002). Assessment of the quality parameters in grapes using VIS/NIR spectroscopy Kemps et al. (2010). VIS/NIR estimation quality indices for Royal Gala -apple McGlone et al. (2002).

Usually the target is in the food industry, producing food for human consumption, but we must not forget that NIRS is also applied in the raising of domesticated animals (Givens et al. 1997) and livestock. Often the main focus is in the development of new methods and tools for the quality control (Grevesmuehl et al. 1991).

Apparently the target is to improve the productivity of food industry. Author wishes to state that this is merely an observation, not a criticism against research carried out. The producers point-of-view to food quality is probably not exactly the same.

As there has been long tradition of research of consumers' food choice and quality perception, Brunsø et al. (2002) introduced the Total Food Quality Model, as seen in Fig. 2. In the model, the quality is divided into charasteristics before and after purchase. According to the Total Food Quality Model the quality itself is not the aim, but it is desired, because it helps via satisfying purchase motives (or values).

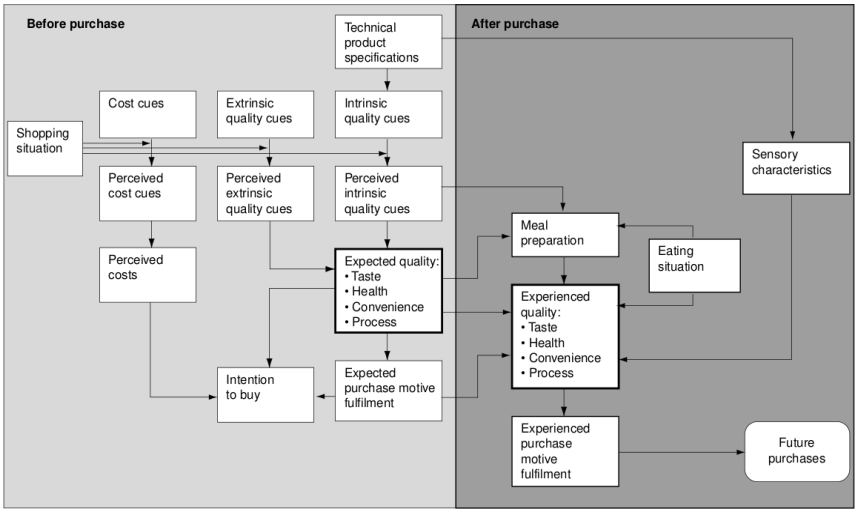


Figure 2. Total Food Quality Model (Brunsø et al. 2002).

The reason for NIR usage in food quality control methods is that it has quite good and predictable penetration properties to biological material (Lammertyn et al. 2000) and can be used to measure sugar and acidity content (Schmilovitch et al. 2000). Scalability of the method [estimated by author of this thesis], was one the criterias used in this thesis. Scalability with NIRS is assumed to be better than applying acoustical methods (Belie et al. 2000a) or electrical impedance spectroscopy (Bauchot et al. 2000), since acoustical methods are prone to environmental noise and electrical impedance measurement requires contact to object measured.

According to DeEll et al. (1995); Prange et al. (2002) chlorophyll fluorescence has been shown to be an effective indicator of low O_2 and/or high CO_2 stress in apples. Schofield et al. (2005) applied it to determine storage potential of iceberg lettuce, DeEll et al. (1996) showed earlier that chlorophyll fluorescence of apples can be used to detect scald development. Some care seems to be needed when chlorophyll concentration is used in analysis.

Toivonen & DeEll (2001) looked into chlorophyll fluorescence of broccoli and found out that there exists a strong correlation between presence of CO_2. The same correlation existed even when broccoli was packaged in MAP. DeEll et al. (1996) showed that this measurement could be taken even through the MAP bag.

Ren & Arnold (2007) showed that quantifying glucose, lactate, and urea dose independently is possible in mixed solutions and that near-infrared gives more accurate results than Raman spectroscopy. The latter resulting from the greater analytical sensitivity of

NIR spectroscopy.

As Alander et al. (2012) reviewed over 200 papers on optical non-destructive near-infrared multispectral methods published during the last twenty years, they confirmed that a lot of research and development has been done on food safety and quality, but still more work needs to be done to find economic ways of monitoring food safety.

1.5. Quality in this research

In this research work the quality is seen and used as a goal which each actor is actively pursuing for. Consumer recognises good product quality and is willing to pay more for it. Manufacturers work for improving the quality of their products because it will give edge in competition of consumer market. This naive model is used because it allows us to simplify the terminology; the quality consists of parameters, which can be objectively measured and therefore we can compare the level of quality.

Therefore the quality is defined differently for each actor. Of course the parameters are somewhat overlapping between them but importance of these parameters differ between actors.

1.6. Blueberries

Research work on blueberries has been targeted on anthocyanin and phenolic content of the berries. The content of sugars, anthocyanins and phenolic in blueberries vary among different cultivates and *Vaccinum* species.

The anthocyanin pigments degrade during processing and storage of foodstuff (Wrolstad et al. 2005). As anthocyanins have main responsibility of the blue colour of blueberry, which can be noticed also visually from the colour changes of blueberries.

Wrolstad et al. (2005) noted that anthocyanins change colour with pH change. This means that if we measure the absorbancy spectra of blueberries, it might also be required to measure the pH to predict the anthocyanin content.

In Fig. 3 and 4 the generalised structure of anthocyanin is shown. The molecules R_1 to R_7 represent substitution locations where different molecules can connect to. With different

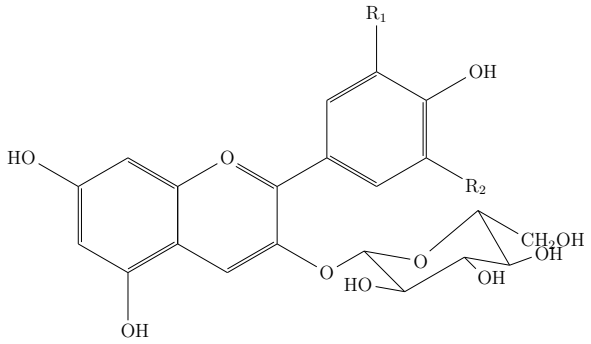


Figure 3. Generalised anthocyanin structure with glycoside (Wrolstad et al. 2005).

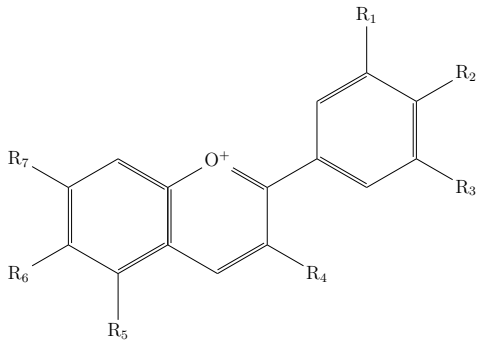


Figure 4. Generalised anthocyanin structure with all substitution locations marked.

substitutes the anthocyanin may have different colour. Possible substitutes are H, OH, and $\mathrm{OCH_3}$.

The difference of Fig. 4 compared to 3 is that $\mathrm{R_4}$ is glucoside substitute. If $\mathrm{R_4}$ will contain galactoside the structure will be anthocyanide, instead of anthocyanin.

Lohachoompol et al. (2004) tracked the total anthocyanin content and antioxidant activity of blueberries (*Vaccinium corymbosum* L). They found out that significant loss of anthocyanin content occurs if the blueberries were dried compared to frozen blueberries during three months of storage. However, drying of blueberries didn't affect significantly to the antioxidant activity level.

Drying of blueberries is not always required. As possible option Perkins-Veazie et al. (2008) showed, that UV treatment might even rise the antioxidant level, decrease decay caused by rise rot and have no effect on anthocyanin level. As ripe rot is also caused by fungus (*C. acutatum*) that might explain the UV response.

Ordidge et al. (2010) showed that the phenolic levels of cultivated blueberry crops are unresponsive to the UV transparency of plastic film which were used to cover crops. This

can also be seen as a positive result if one wishes to apply UV treatment in an industrial blueberry handling process.

Solomakhin & Blanke (2010) measured also that by using coloured hailnets over apple cultivates increased the chlorophyll content (analysed with NDVIndex as shown in Eq. (62)) and lowered the anthocyanin content (NAIndex shown in Eq. (73)). As foliage is a kind of natural hailnet, one could still ask that is it possible to cultivate "super blueberries" with higher anthocyanin content if foliage is removed. This could also be interpreted so, that perhaps the blueberries cultivated on open areas have higher content of anthocyanin compared to those in thicker forest. Or perhaps this could be interpreted that cultivated blueberries will produce more anthocyanin if hailnets, with similar optical properties to foliage, is used.

Rein (2005) in her research used wavelengths of 280 nm for catechins, 320 nm for phenolic acids, 365 nm for flavonols, and 520 nm for anthocyanins in content measurements. Therefore we also use 520 nm for blueberry anthocyanin amount calculation.

2. METHODS

In this Chapter the basic physics concepts of photon energy transmittance is explained. The theory is aimed to give enough background information to understand the physics where NIRS is based. An overview of the measurement devices commonly used in NIRS are shown. The relevant mathematical tools and analysis methods are given, which will be utilised in analysis of the data.

Spectroscopy is a science where the absorption, reflection and transmission of light in specimen is observed and analysed. It is divided into several areas, named according to the wavelengths used. This thesis focuses mainly on near-infrared light spectroscopy, which means wavelength range from 800 to 2500 nm respectively.

Over the centuries there have been evolution of scientific theory which have been used to explain the effects of light with matter.

As nowadays is commonly stated, quantum physics gives us the most accurate model of light. As accurate model it is also a bit too complex for this work and as Hecht (2002: ch 1, p. 65) states we can safely apply more simplified version of light and matter interference without significant errors. In this simplified model we handle light just as a combination of electromagnetic waves and particle (photon) radiation.

2.1. Energy in Light

Light can be seen as combination of particles and wave energy. The photons are the particles and their movement is wave like. Therefore we can use wave models when we are talking about light. A frequency of a light wave is determined by the period of oscillations of the light wave. The speed and wavelength of light are related as shown in equation (1), where λ refers to the wavelength and f is the frequency.

$$E = hf = \frac{hc_0}{\lambda} \rightarrow h = \frac{\lambda}{c_0} \tag{1}$$

In this simplification we also can ignore the momentum of particle. The speed of light in

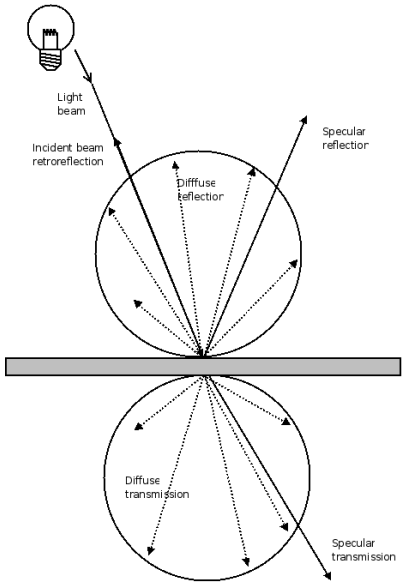


Figure 5. Light beam strikes material (redrawn from Räty et al. (2004), some additions by author).

medium is typically described by the index of refraction n, which is related to the speed of light in vacuum by Eq. (2):

$$n = c_0/c_m \tag{2}$$

The importance of the speed of light in sample is explained by the fact that light also includes energy, since as Planck [or Planck-Einstein] relation Eq. (1) shows, a photon with speed contains energy. In the equation c_m stands for the speed of light, h Planck's constant, and λ the wavelength of light. The speed of light refers to speed in medium, not in vacuum, which is referred in this thesis with symbol c_0.

When light (or other radiation) strikes a sample (object of interest) reflection, absorption, and transmission occur (Räty et al. 2004). Illustration of this can be seen in Fig. 5 where the grey block represents the sample, light ray enters from the top left corner (illustrated by light-bulb). Arrow heads show primary directions of light rays.

2.1.1. Molecular Vibration

According to Workman & Weyer (2008: 1) the energy level of a molecule is described as a sum of atomic and molecular motions due to translational, rotational, vibrational, and

electrical energies.

Translational energy can be thought to be similar to the potential energy of an object in traditional mechanics; external energy is required to move a stationary object. Similarly external translational energy is required to move an atom further away from other atoms in a molecule. An atom does not start rotating without external energy. Analogy to classical mechanics is also very obvious. External energy is also required for an atom to vibrate and energy is required for exitation of electrons to upper energy levels.

A molecule should be thought as a three dimensional object rather than a two dimensional representation, though in this thesis illustrations are given in two dimensions. In Fig. 6 a two dimensional illustration of methyl molecule is given. Each atom in a molecule has some freedom of movement relative to other atoms in a molecule before the atom breaks free. Workman & Weyer (2008: 1) states that this distance is approximately 10-15 % of the average distance between atoms in a molecule. This movement is called stretching. Stretching back and forth is analogously therefore called stretching vibration.

The variation and change of angle of the bond is called bending. It is just another type of translational change which requires energy. Bending is either symmetric or asymmetric.

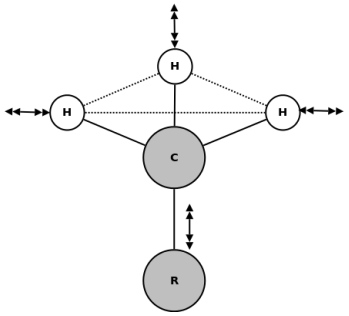


Figure 6. Methyl molecule with arrows illustrating possible stretching directions. (Workman & Weyer 2008: 3)

In Fig. 7 the double headed arrows illustrate directions for atoms to stretch. The stretching can be symmetric, for example if left and right hydrogen atoms stretch both simultaneously. However, if only the left and top hydrogen atoms stretch, as is seen in Fig. 8, is the stretching asymmetric. Of course symmetric stretching cannot exist without even number of atoms in stretching.

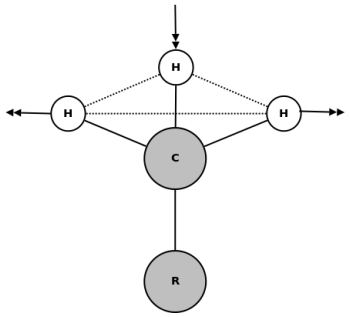


Figure 7. Symmetric molecule stretching of methylene. Arrows illustrate possible stretching directions. (Workman & Weyer 2008: 3–4)

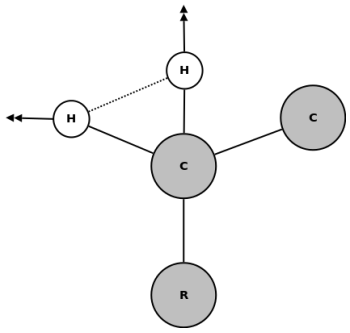


Figure 8. Asymmetric molecule stretching of methylene. Arrows illustrating possible stretching directions. (Workman & Weyer 2008: 3–4)

When photon (light) enters into sample two things can happen: the energy of the photon does or does not match to the vibrational energy of the molecule. When there is a match between the photon energy and the natural vibrational frequency of a molecule the molecule absorbs this energy. Absorbed energy increases the vibration amplitude of the molecule. The frequency of the absorbing vibration remains constant (Workman & Weyer 2008: 13). Therefore we can say that molecular vibration is quantified, and only certain energy levels can exist.

Using the normal mode theory we can calculate the frequency of molecules absorbing vibration. This is achieved by using Eq. (3), where K is the force constant value, which has different values for different bonds, m_1 is the mass of the atom, m_2 is the mass of second atom. This model is called *Ideal Harmonic Oscillator.* (Workman & Weyer 2008: 12–14)

$$\nu = \frac{1}{2\pi}\sqrt{\frac{K}{\left(\frac{1}{m_1}+\frac{1}{m_2}\right)}} \tag{3}$$

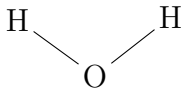


Figure 9. Simplified presentation of water molecule structure. Pure water is composed from two hydrogen and a single oxygen atoms.

The discrete vibrational energy levels for any molecule E_{VIB} are given by Eq. (4), where h is Planck's constant, ν is vibrational frequency of the bond and v is the vibrational quantum number ($v = [0, 1, 2, 3, ...]$) (Workman & Weyer 2008: 13). As can be seen from the equation and Ozaki, McClure & Christy (2007: 15) point out, the lowest possible vibrational energy ($v = 0$) becomes to $\frac{1}{2}h\nu$, in contrast to classic mechanics, and therefore molecule has always some vibrational energy.

$$E_{\mathrm{VIB}} = h\nu\left(v + \frac{1}{2}\right) \tag{4}$$

The liquid water (see Fig. 9) has quite complex absorption spectrum, even the molecule is pretty simple. Angle between hydrogen covalent bonds is 104.45 degrees. This complexity is partly caused by the fact that water is not pure $\mathrm{H_2O}$ but it is mixture of molecules, such as $\mathrm{O^{2-}}$, $\mathrm{OH^-}$ HDO, and $\mathrm{D_2O^+}$.

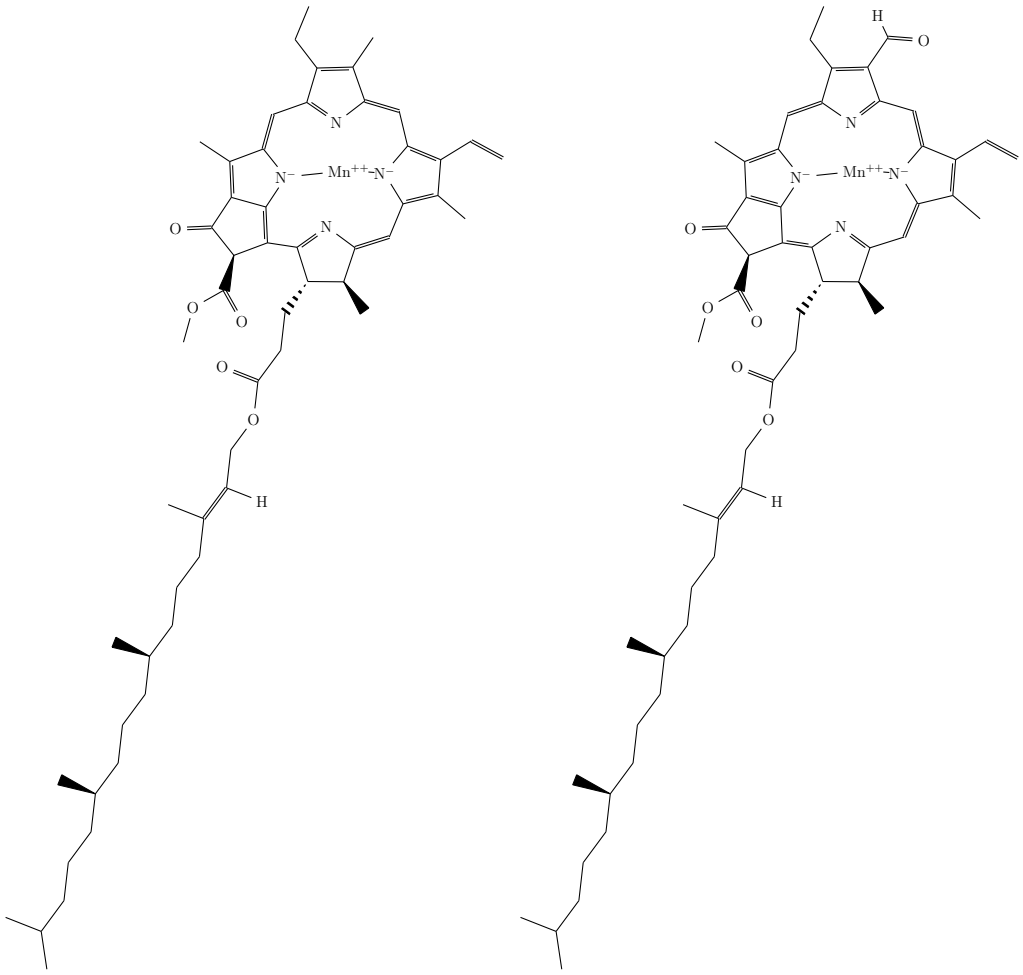
As can be seen the chlorophyll a- and b-molecule (see Figs. 10(a) and 10(b)) and anthocyanin (see Fig. 4) are complex compared to water molecule. As the calculation of absorption peaks would be extremely complex, we shall not calculate it in this thesis. We are, however, interested in the absorbance peaks of the chlorophyll and in the amount of absorbance which correlates with the amount of the chlorophyll in our measured samples.

2.1.2. Rotational energy

The rotational energy is inversely proportional to the moment of inertia, as can be seen in Eq. (5). If we calculate rotational energy for $\mathrm{O_2}$ -molecule, we found it to be: $E_{0r} = \frac{(1.05\times10^{-34}\mathrm{Js})^2}{(16\mathrm{u})(10^{-10}\mathrm{m})^2}\frac{1\mathrm{u}}{1.66\times10^{-27}\mathrm{kg}} = 2.59\times10^{-4}\mathrm{eV}$.

$$E_{0r} = \frac{\hbar^2}{2I} = \frac{\hbar^2}{mr_0^2} \tag{5}$$

As Tipler (1999: p. 1215) states, the typical measured frequency of a transition between



(a) chlorophyll-a (b) chlorophyll-b

Figure 10. Molecular structure of chlorophyll a- and b. More complex structure than with pure water causes manual calculation of the absorbance peaks to be unmanageable.

vibrational states is 5×10^{13}Hz, which gives the magnitude of vibrational energy for O_2 -molecule to be $E = h\nu = (4.14 \times 10^{-15}\text{eVs})(5 \times 10^{13}s^{-1}) = 0.2\text{eV}$. From this we can conclude that vibrational levels cannot be exited by molecular collisions, in ordinary temperatures since there is difference of three decades (Tipler 1999: 1215-1216).

2.1.3. Temperature

All objects radiate, which was first discovered by Max Karl Ernst Ludwig Planck at 1905, and the idea was formulated later (by Planck) into Planck Radiation Law (see Eq. (6).

$$I_\lambda = \frac{2\pi hc^2}{\lambda^5}\left[\frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1}\right] \tag{6}$$

If we just focus in the effect of temperature change in black radiator object, we assume that only the temperature changes and therefore we write Eq. (6) in form:

$$\frac{dI_\lambda}{dT} = a\frac{1}{1/e^{bT}-1} \hat{=} ae^{bT}, \tag{7}$$

where $a = \frac{2\pi hc^2}{\lambda^5}$, $a \ll 1$, $b = \frac{hc}{\lambda k_B}$, and k_B is Boltzmann Constant.

Therefore we can deduce that small change in temperature causes small change in the spectral irradiance of a black radiating object. The irradiance from a NIRS measurement object does not change radically if the measurement temperature has small change. However, changes may occur because the chemical reactivity or properties of measured object may change due to the change of temperature.

The molecular interpretation of temperature is that the molecules of a substance have kinetic energy of:

$$(\frac{1}{2}mv_x^2)_{av} = \frac{1}{2}kT \tag{8}$$

where T is the temperature of the substance and v_x is the motion along the x axis. Since this applies to also the y and z axes, we see that in average:

$$(v_x^2)_{av} = (v_y^2)_{av} = (v_z^2)_{av} \tag{9}$$

and hence the total kinetic energy becomes:

$$K_{av} = (\frac{1}{2}mv^2)_{av} = \frac{3}{2}kT \rightarrow K = N(\frac{1}{2}mv^2)_{av} = \frac{3}{2}NkT = \frac{3}{2}nRT \tag{10}$$

Naturally molecules may have also rotational or vibrational kinetic energy. (Tipler 1999: 550-551)

In order to get reliable measurements, and without a need for temperature correction in calculations, the environmental temperature should be kept constant during repeated measurements.

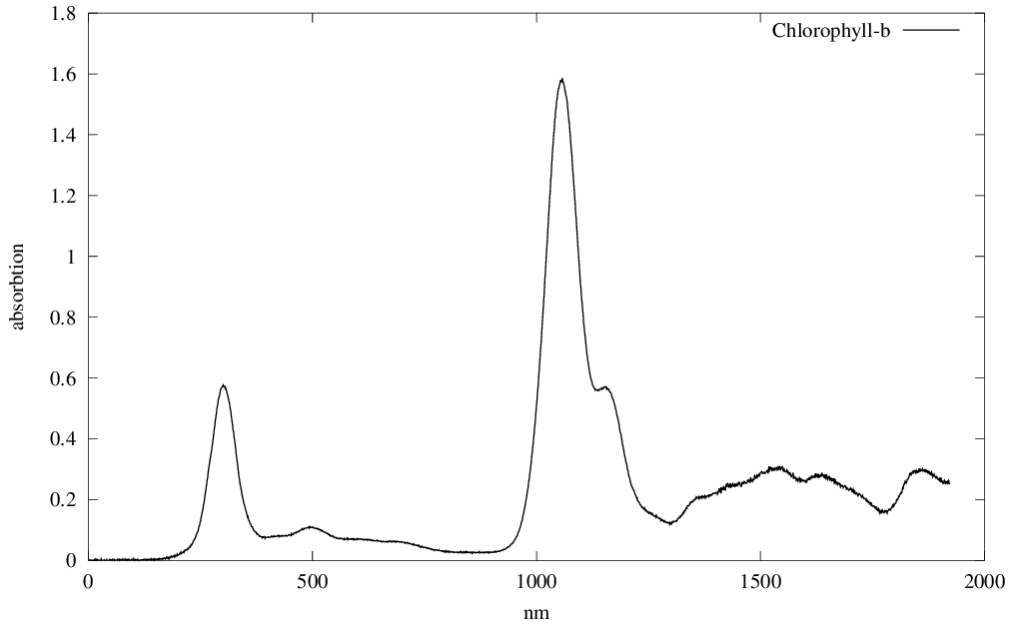


Figure 11. Chlorophyll-B diethyl solvent absorbance spectra (image created from data of PhotoChemCAD -software).

2.2. Absorption

If a photon collides into a molecule and the energy of the photon equals to an energy which is required for molecule's excitation an absorption may occur. The photon releases it's kinetic energy to an electron, which is exitated, which is illustrated in Fig. 12(c). In this process the photon itself is destroyed.

The relaxation of the exitation occurs quite often very quickly but can sometimes be stable for longer time—of course in the perspective of atom-scale time. According to Tipler (1999: 1030) this usually occurs within 10^{-8}s. In relaxation the electron returns to it's previous energy level and the released energy can emit a new photon.

As the photons energy is dependent on the wavelength, which can be seen in Eq. (1), a given molecule can absorb only photons of certain wavelengths. This means that each molecule has it's own "absorption footprint", which can be used to identify it. This applies to both single atoms and composite molecules. In Fig. 11 an absorption spectra of chlorophyll-b is shown. The absorption has higher and lower areas in the spectra but the high peak is located at 435.8 nm.

The concentration of molecule has a relationship with absorbance, which is usually writ-

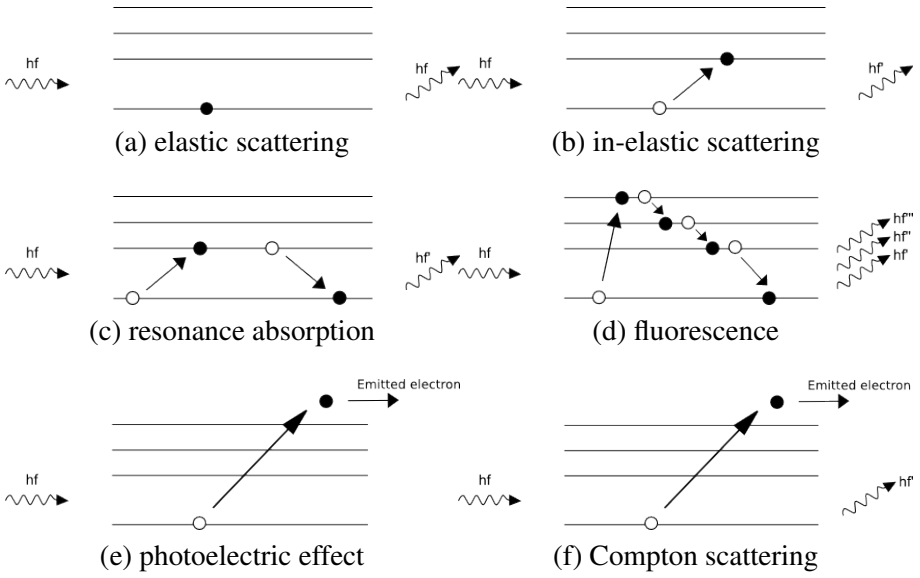
ten as in Eq. (11), where A is the absorbance, ξ is the product of absorptivity, c is the concentration of the molecule and l is the path length of the light ray in the sample.

$$A = \xi cl \implies \xi = \frac{A}{cl} \tag{11}$$

From this we can draw the following conclusions; (a) the longer the path light travels within the sample, the more absorbance occurs; (b) the higher the concentration of the substance in the sample, the more absorbance occurs. Therefore, if we keep the path length constant, we can determine the concentration of the substance in the sample from the absorbance, since the absorptivity does not change, if temperature and other variables are constant.

2.3. Reflectance

Reflection of light is divided into specular reflection and diffuse reflection. Specular reflection describes reflection on simple smooth surfaces, such as a mirror, which reflect light in a simple predictable way. Diffuse reflection describes opaque reflection on a flat surface such as paper, or fabric. Also transmission of light can be divided into diffuse transmission and specular transmission. (Räty et al. 2004)



(a) elastic scattering (b) in-elastic scattering

(c) resonance absorption (d) fluorescence

(e) photoelectric effect (f) Compton scattering

Figure 12. Photon-atom interactions according Tipler (1999: 1031).

If a light ray enters to a sample it will refract. The Snell's law, seen in Eq. (12), gives us the relationship between refraction index and angle which refracted light ray will have. Outer substance, usually air, has refraction index of n_1 and sample substance n_2. If light, having intensity I_0, enters a sample from angle θ_1 it will refract and continue it's way with angle θ_2.

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{12}$$

However, The Snell's law does not tell us how much light will be refracted or reflected. That is explained by Fresnel equation, in Eq. (13) and (14). In specular reflection the amount and direction of reflection can be approximated by Lambert's cosine law, see Eq. (15), which describes surfaces with equal luminance, i.e. smooth and flat theoretical surfaces.

$$T \equiv \quad \frac{I_t \cos\theta_t}{I_i \cos\theta_i} \tag{13}$$

$$R \equiv \quad \frac{I_r A \cos\theta_r}{I_i A \cos\theta_i} \tag{14}$$

The Fresnel equations, Eq. (13) and (14), gives the amount of transmittance (T) and reflectance (R) of light, where the area of the illumination (A) and angle (θ) are known.

$$I_0 = \frac{I \,\mathrm{d}\,\Omega \,\mathrm{d}\,A}{\mathrm{d}\,\Omega_0 \,\mathrm{d}\,A_0} \frac{\text{photons}}{\text{s} \cdot \text{cm}^2 \cdot \text{sr}} \tag{15}$$

In a diffuse reflection in a smooth and a flat surface the intensity can be approximated by using the Lambert's reflectance law (Eq. (16)). If the surface is not smooth then the amount of reflected light can be approximated with statistical methods.

$$I_D = L \cdot NCI_L \tag{16}$$

In Eq. (16) the I_D is the intensity of the diffusely reflected light, N is normalised light vector, C is colour and I_L is the intensity of incoming light. Since $L \cdot N = |N||L| \cos\alpha =$

$\cos\alpha$, where α is the angle between the incoming and reflected light, the highest intensity will be in the direction of incoming light. (Hecht 2002: 98)

The relation between reflectivity and absorbance can therefore be written as in Eq. (17), where R is the reflectance, ξ is the molecular absorbance, c is the concentration, and l is the path length of the light ray.

$$R = \frac{I}{I_0} = 10^{\xi cl} \implies A = -\log_{10}\left(\frac{I}{I_0}\right) = -\log_{10} R \tag{17}$$

2.4. Transmittance

When we are measuring liquid or other samples, we can also measure the absorbance of light ray travelling through our sample. Then we can write our relationship between absorbance and transmittance as in Eq. (18), where T is the transmittance, ξ is the molecular absorbance, c is the concentration, and l is the path length of the light ray. Usually on gas measurements the natural logarithm is used instead.

$$T = \frac{I}{I_0} = 10^{\xi cl} \implies A = -\log_{10}\left(\frac{I}{I_0}\right) = -\log_{10} T \tag{18}$$

Again we can see that the absorbance is depending on the molecular concentration, if the path length is kept constant.

2.5. Scattering

If a photon interacts with an atom (or molecule) and it has different energy than is required for absorption, scattering occurs. When the size of a molecule is larger than the photon wavelength, only forward scattering occurs. This means that only small change in the direction of the photon occurs. If the molecule is smaller than the photons wavelength, scattering occurs uniformly to every direction, which is called Rayleigh Scattering.

In Fig. 12 different phenomenons are shown which can occur when photons interact with an atom. In the elastic scattering Fig. 12(a) the energy of a photon is too low for exciting the atom to an excited state, atom remains in the ground state and the photon is

scattered. The energy of a photon remains the same as before scattering, hence naming elastic scattering (Tipler 1999: 1031).

When the energy of an incident photon has enough energy to excite the atom to an excited state, the energy of the scattered photon (hf') decreased by ΔE, which is the energy required for the excitation of the atom. Inelastic scattering (Fig. 12(b)) is often referred as Raman scattering, named after it's first observer C.V. Raman.

On fluorescence, see Fig. 12(d), the photon excites an atom to a higher energy level. The atom later looses the energy and spontaneously emits a photon while transition to a lower energy level happens. If the luminance of photons is visible (in UV or VIS area) this is called autofluorescence.

In photoelectric effect, see Fig. 12(e), the absorption of the photon energy ionises the atom and an emission of electron occurs. The possible extra energy is transformed into kinetic energy of the electron. The emission absorbs photon energy totally and not even a low-energy level photon is emitted. However, if the energy of the photon is much greater than is required for electron emission, the extra energy can emit a new photon. This is referred to as Compton scattering (Fig. 12(f)). In NIRS, the Compton scattering is not usually present.

In real life, some materials reflect and transmit light simultaneously. This phenomena is called transflection. A transflection reflector includes a body for simultaneously reflecting and transmitting light rays. The body is formed by a plurality of layers of a polymeric reflective material and has reflected rays at each interface between adjacent ones of the multiple layers. The body is both selectively configured and positioned relative to a light source for a predetermined distribution of reflected and transmitted rays.

2.6. Analysis tools

The spectrometer instrument measures intensity of light (I_o) for several wavelengths, as described earlier, (λ_i) simultaneously. Each measurement for wavelength λ_i is then referred as x_i. We write each measurement j, i.e. spectrum, as a vector Eq. (19):

$$X_j = \begin{pmatrix} x_1 & x_2 & \cdots & x_i \end{pmatrix} \tag{19}$$

and if we have more than one measurement, then all the measurements will form a matrix, see Eq. (20), with dimensions $[i, j]$. From this point on, and if not otherwise stated, all measurement variables are in fact matrices, where each measured intensity is presented as a vector element.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,i} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,i} \\ \vdots & \vdots & \ddots & \vdots \\ x_{j,1} & x_{j,2} & \cdots & x_{j,i} \end{pmatrix} \tag{20}$$

2.6.1. Data Autoscaling

In NIRS we are interested most of the time in the intensity of the absorbency and reflectance of the light at the object(s) at certain wavelength, $I_i(\lambda_i)$ and $R_i(\lambda_i)$ respectively. We therefore use the matrix form, see Eq. (20).

When handling the spectral data it is a custom to use mean centering of data, see Eq. (21). This is done because mean centering does not change relative distance between data points and therefore does not change standard deviation but allows us to move the center-of-gravity (mean) of data points to zero, which simplifies calculations.

$$\bar{X_j} = X_j - \frac{1}{n}\sum_{n=1}^{i} X_j \tag{21}$$

If, however, the data has different scaling(s) or units, then the mean centering does not give correct results. In these cases some other methods, such as autoscaling needs to be applied.

For analysis purposes, the data should sometimes be scaled before analysis is carried out. This can be written as matrix operation $\mathbf{X}_s = \mathbf{XS}$ where the scaling vector $\mathbf{S}$ is a diagonal matrix. One most often used scaling method is autoscaling. In autoscaling the mean

centered data X_m is column-wise divided with the standard deviation of that column. After autoscaling the data has mean of zero and standard deviation of one.

2.6.2. PCA

The most often applied statistical tool in chemometrics seems to be PCA. It is used in making predictive models based on spectral data. When building a predictive model, we try to find the principal components of a set of data, i.e. the variables which describe major trends of the data. If, for example, we have a set of measurements from known items for NIR reflectance, we can use the set as an input to the PCA model. The resulting primary components of training set can be used to differentiate, or classify, similar type item measurements with projecting the measurement data into primary component space.

The basic idea of PCA is simple; the measurement matrix **X**, with j measurements and i reflectance wavelengths, is written as a sum of r products of score vector (t_n) and loading's (p_n), as shown in Eq. (22) for a vector element j of **X**. The $r \leq \min i, j$ is the rank of the matrix **X**.

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + \cdots + t_r p_r^T \tag{22}$$

Since there are always noise in measurements, it's common to truncate the equation after k components and move the truncation to residual matrix **E**, as in Eq. (23). If **X** would include absorbancy measurement of liquid (following Beer's law), t_n could be replaced with concentrations and p_n with pure component spectra. Now we can write X as a sum of score weighted concentrations and the residual matrix E:

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E \tag{23}$$

PCA can be solved as an eigenvector decomposition of the covariance or correlation matrix of the process variables, therefore the covariance matrix of data matrix **X** can be written as in Eq. (24), provided that columns have been mean centered.

$$\mathrm{cov}(\mathbf{X}) = \frac{\mathbf{X}^T \mathbf{X}}{j - 1} \tag{24}$$

As p_i vectors are the eigenvectors of the covariance matrix we can state that for each p_i in Eq. (25).

$$\mathrm{cov}(\mathbf{X})p_i = e_i p_i \tag{25}$$

where e_i is the eigenvalue associated with the eigenvector p_i. Score vectors t_i form an orthogonal set ($t_i^T t_j = 0$ for $i \neq j$) and p_i are orthonormal ($p_i^T p_j = 0$ for $i \neq j$, $p_i^T p_j = 1$ for $i = j$. Therefore we can write the score vector t_i as the linear combination of the original $\mathbf{X}$ multiplied by p_i:

$$\mathbf{X}p_i = t_i \tag{26}$$

2.6.3. PLS

Partial Least Squares Regression is based on, like PCA, a number of orthogonal score vectors. The main differences between PCA and PLS are that PCA is used to reduce dimensionality, and it is unsupervised learning technique, whereas PLS finds the relation between two different data sets, and can be used for dimensionality reduction, regression (value prediction) and pattern recognition. PLS is also label aware technique. The model in PLS is written as in Eq. 27, where X is the measurement data matrix, y is the known concentration, or predicted variable, and b includes the weights solved from the model.

$$Xb = y \tag{27}$$

There exists several algorithms, the most common of which are NIPALS and SIMPLS that can be used to solve the equation. One key aspect of PLS is that after the model has been built it is used to predict output based on new measurements.

In this work the model variables are called Linear Variable (LV), which is common convention, to distinguish them from PCA model variables (PC).

2.6.4. Noise Reduction

According to (Ozaki et al. 2007: 53), the most generally used method for improving the signal to noise ratio in spectra is accumulation-average processing. This basically means that the measurement signal is calculated as an average of several measurements. This reduces the amount of high-frequency noise.

When the accumulation average does not remove noise from the signal enough, filtering or smoothing is generally applied. The most simple smoothing method is the moving-average method, in Eq. (28). The x_i is the measured data, w_k is the weight parameter. The average is calculated from the value x_i and the nearest neighbours, from x_{i-n} to x_{i+n}.

$$x_i = \sum_{k=-n}^{n} w_k x_{i+k} \tag{28}$$

The Savitzky-Golay method uses the same equation, but the weights are determined with fitting the spectrum data with a low-degree polynomial. The values for convolution weights can be found in numeral tables. For example, if $n = 5$, the weights $w_k = [-3/35, 12/35, 17/35, 12/35, -3/35]$ and $k = [-2, -1, 0, 1, 2]$.

2.6.5. Derivation of the Signal

In computer applications we work with discrete data. The measured signal is equispaced. The distance between nodes is written as $h = x_{i+1} - x_i = \Delta x$. We could, as shown by Kreyszig (1998); Sun et al. (2010), to apply Taylor series expansion for a discrete function $f(x)$:

$$f(x) = f(x_i) + (x - x_i)f^{(1)}(x_i) + \frac{(x - x_i)^2}{2!}f^{(2)}(x_i) + \frac{(x - x_i)^3}{3!}f^{(3)} + \cdots \tag{29}$$

For this analysis we will look only the first four term of Taylor series expansion. In real applications we might use more. We shall write the error as E, which is the difference of our Taylor Series expansion and actual value:

$$f(x) = f(x_i) + (x - x_i)f^{(1)}(x_i) + \frac{(x - x_i)^2}{2!}f^{(2)}(x_i) + \frac{(x - x_i)^3}{3!}f^{(3)} + E \qquad (30)$$

From this we can see that:

$$\begin{aligned}
E &= \frac{(x - x_i)^4}{4!}f^{(4)}(x_i) + \frac{(x - x_i)^5}{5!}f^{(5)} + \cdots \Rightarrow & (31)\\
E &= \frac{(x - x_i)^4}{4!}f^{(4)}(\xi) \qquad x_i \leq \xi \leq x \Rightarrow & (32)\\
E &\cong \frac{(x - x_i)^4}{4!}f^{(4)}(x_i) \Rightarrow & (33)\\
E &\cong O(x - x_i)^4 & (34)
\end{aligned}$$

Our Taylor series is convergent, and therefore each higher term in the error series should be decreasing. Therefore we write the error term as $O(h)^n$, where n is the dimension, in this examples $n = 4$.

Now we evaluate $f(x_{i+1})$, and $f(x_{i-1})$ as follows:

$$\begin{aligned}
f(x_{i+1}) &= f(x_i) + (x_{x+1} - x_i)f^{(1)}(x_i) + \frac{(x_{i+1} - x_i)^2}{2!}f^{(2)}(x_i) & (35)\\
&\quad + \frac{(x_{i+1} - x_i)^3}{3!}f^{(3)} + O(x_{i+1} - x_i)^4 \Rightarrow & (36)\\
f_{i+1} &= f_i + hf_i^{(1)} + \frac{h^2}{2}f_i^{(2)} + \frac{h^3}{6}f_i^{(3)} + O(h)^4 & (37)\\
f(x_{i-1}) &= f(x_i) + (x_{x-1} - x_i)f^{(1)}(x_i) + \frac{(x_{i-1} - x_i)^2}{2!}f^{(2)}(x_i) & (38)\\
&\quad + \frac{(x_{i-1} - x_i)^3}{3!}f^{(3)} + O(x_{i+1} - x_i)^4 \Rightarrow & (39)\\
f_{i-1} &= f_i - hf_i^{(1)} + \frac{h^2}{2}f_i^{(2)} - \frac{h^3}{6}f_i^{(3)} + O(h)^4 & (40)
\end{aligned}$$

By using these evaluation results we will write the discrete first derivative. First we calculate the difference of $f_{i+1} - f_{i-1}$ to be:

$$f_{i+1} - f_{i-1} = \left(f_i + h f_i^{(1)} + \frac{h^2}{2} f_i^{(2)} + \frac{h^3}{6} f_i^{(3)} + O(h)^4 \right) \tag{41}$$

$$- \left(f_i - h f_i^{(1)} + \frac{h^2}{2} f_i^{(2)} - \frac{h^3}{6} f_i^{(3)} + O(h)^4 \right) \Rightarrow \tag{42}$$

$$= 2h_i^{(1)} + \frac{h^3}{3} f_i^{(3)} + O(h)^4 \tag{43}$$

So, the approximation to the first derivative is:

$$f_i^{(1)} = \frac{f_{i+1} - f_{i-1}}{2h} + E \tag{44}$$

And the error term can be written as:

$$E = \frac{h^2}{6} f_i^{(3)} \cong O(h)^2 \tag{45}$$

By using calculus we can also show that the second derivative will be:

$$f_i^{(2)} = \frac{f_{i+1} - 2f_i + f_{i-1}}{h^2} + E, \qquad E \cong -\frac{1}{12} h^2 f_i^{(4)} \tag{46}$$

As Sun et al. (2010: 358) explains that the truncated Taylor series expansion can also be used to calculate the second derivative as is shown by Eq. (47):

$$f^2(x) = \frac{f(x + \Delta x) - 2 \times f(x) + f(x - \Delta x)}{\Delta x^2} \tag{47}$$

which can be applied for calculus as:

$$\mathrm{d}^2 A_i = A_{i+k} - 2 \times A_i + A_{i-k} \tag{48}$$

In the analysis section of measurements, we use signal derivation to measure the change rate of the signal slope. Sudden change in change rate of signal can be used to locate interesting frequency peaks.

However, it should be noted that each time signal is derivated, the signal to noise ratio deteriorates (Ozaki et al. 2007: 54). Therefore noise reduction, and in some cases polynomial filtering, should be applied prior to derivation of the signal.

2.6.6. Data Size Reduction

The amount of data of the spectral measurement can be a problem in practical NIR applications. For reliable results, a lot of measurements is required, with as much spectral channels as possible. This data is usually handled as matrices where we have a column for each spectral channel used in measurements and a row for each measurement. If we use full measurement sets in data analysis, the matrices quickly become too big for efficient calculation.

In NIR problems, reduction of the measurement data size before the analysis is often necessary. If we can reduce the number of the measured wavelengths it will reduce the computation time and memory requirements considerable. This is due to the fact that e.g. the PCA solving algorithms, as seen previously, for j measurements has computational memory requirements of $O(n^2)$.

Another problem can be that the measured spectra wavelengths are redundant especially in NIR applications. This is mainly caused by overlapping wavelengths in absorption.

In this work we have applied the polynomial function fitting. In the polynomial function fitting the solution is to find the polynomial function $p(x)$ of degree n, such that maximum value of the difference between actual data and fitted polynomial function, also called as residual, is minimal, also called as Chebyshev approximation criterion (Giordano et al. 2003: 108). Therefore, polynomial function fitting is a minimisation problem, where the target of the minimisation can also be written as in Eq. (49), where the y_i is the actual data value at the index i, and the $p(x_i)$ is the function value at the same index. The most frequently used criteria for the coefficient selection is the least-squares-error (Giordano et al. 2003: 110).

$$\sum_{i=n}^{m} |y_i - f(x_i)|^2 \tag{49}$$

Least-squares-error -method is not perfect, one must be vary that there exists an implicit assumption that the error in the independent variable is zero or at least so close to zero that it can be neglected, otherwise there may be stability problems.

When the suitable polynomial function has been found, it is used to calculate smaller set of data, which can then be used in analysis in place of the original data thus saving resources. Polynomial fitting should be used with care, since polynomial function may oscillate.

In this work we have applied the saturation of polynomial function error for criteria to select appropriate degree of the polynomial function. The saturation is calculated as a sum of absolute error between the measurement data and values calculated with polynomial function in measurement data space, as is shown in Eq. (50). In other words, the saturation is the total error of polynomial fitting for the measurement data. If we increase the degree of polynomial fitting function after the saturation point, the saturation value does not decrease meaningfully.

$$\mathrm{Sat} = \sum_{i=n}^{m} |y_i - f(x_i)| \tag{50}$$

In the process of data size reduction, the discrete data is converted to continuous polynomial function. This conversion is, although convenient, artificial. The data should still be investigated properly for possible discontinuous points. The reason for this is two-folded; firstly the error of the describing polynomial function may grow considerable close to discontinuous parts, and secondly, the probabilistic assumption that error distribution is uniform, may not apply. One possible solution to the problem of the discontinuous parts is to convert the continuous polynomial function into splines and therefore avoid problems close to the data discontinuous areas, since spline is piecewise defined.

2.6.7. Noise in Measurements and Data

When a measurement is done, the result is composed of actual data and some error. When we repeat the measurement we can see that the measurement results are not exactly identical. We illustrate this in Fig. 13 where some measurements are shown (circles) and the actual (real) value is shown as a bullet. In the figure, the spread means that for the same

data point (black dot) we get several results, which are quite close the actual value but not identical. They are spread around the actual value. We might also receive another cluster of measurement results further from the actual value. The further from the actual value the result cluster is located, the bigger is the bias value of that cluster.

During the model build-up and in error analysis, we are interested to minimise both spread and bias of error. Spread can be thought as a precision of measurements. The smaller the spread, the more reliable model can be built from the data. This can be achieved by taking larger samples and therefore locating the actual data value by minimising the measurement error value.

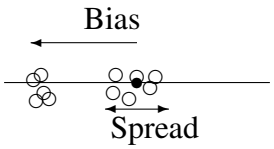


Figure 13. Illustration of measurement spread and bias of error (○ = measurement, ● = real value).

When we apply NIRS we try to collect large sample sets because this is the best way to increase the precision. The does not improve the accuracy automatically, which can be achieved by eliminating possible systematic measurement errors. This is due to the assumption that the distribution of data is following normal distribution.

The normal distribution assumption allows us to assume that if our mean value is zero (μ), most of our measurements are close to the zero, if σ^2 is small. The difference to μ shows the measurement error.

Variance of the standard distribution describes the spread of measurements. When spread (variance) grows, measurements don't describe the true value very well.

Root Mean Square Error of Prediction (RMSEP), also known as Standard Error or Prediction (or Perfomance) (SEP), is commonly used in NIR spectroscopy to describe the models prediction ability (Bellon-Maurel et al. 2010). The SEP is a measure of the difference between predicted and reference values. It can be computed (see Eq. (51)) as the sum of squares of the differences between the predicted $\hat{y}_i$ and the true values of $\forall y_i \in i : \{1..m\}$ are defined by

$$\mathrm{RMSEP}^2 = \mathrm{SEP}^2 = \sum_{i=1}^{m} \frac{(\hat{y}_i - y_i)^2}{m} \tag{51}$$

where m is the number of measurements. As Bellon-Maurel et al. (2010) point out, it is generally compared to expected accuracy, or expected error, when deciding if a method provides acceptable results.

We decompose the SEP as a sum of bias squared and SEP with bias corrected ($\mathrm{SEP}^2\mathrm{c}$, as (Bellon-Maurel et al. 2010) did:

$$\mathrm{SEP}^2 = \mathrm{Bias}^2 + \mathrm{SEP}^2\mathrm{c} \tag{52}$$

where the bias of data is

$$\mathrm{Bias} = \sum_{i=1}^{m} \frac{\hat{y}}{m} - \sum_{i=1}^{m} \frac{y_i}{m} = \bar{\hat{y}} - \bar{y} \tag{53}$$

and the bias corrected SEP is

$$\mathrm{SEP}^2\mathrm{c} = \sum_{i=1}^{m} \frac{(\hat{y}_i - \mathrm{Bias} - y_i)^2}{m}. \tag{54}$$

With Partial Least Squares (PLS) models, the Predicted Residual Estimated Sum of Squares (PRESS) and Correlation Coefficient (R^2) are often used. They are calculated for PLS model as:

$$\mathrm{PRESS} = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \tag{55}$$

$$\mathrm{R}^2 = 1 - \frac{\mathrm{PRESS}}{y^2} \tag{56}$$

where y contains the real values.

Therefore, the aim in the measurement should be to reduce the spread and to control the bias of measurements, since measurement is composed of the actual value and noise. The measurement $\overline{\lambda}_j$ can be divided into corresponding parts, if it is written as in Eq. (57), where ε_a is relative to the wavelength of λ_j, as the measurement value in wavelength i, and ε_b stands for the base noise in measurement. ε_b can be extrapolated with random function, since it includes also the background noise.

$$\overline{\lambda}_j = \lambda_j + \varepsilon_a(j) + \varepsilon_b \tag{57}$$

Basic assumption is, however, that the value of ε_b is small. If we state that in our measurement (λ_i) we always have some error component (ε_i) and we are comparing two measurements we could write our comparison ratio as:

$$\text{ratio} = \frac{(\lambda_1 + \varepsilon_1) - (\lambda_2 + \varepsilon_2)}{(\lambda_1 + \varepsilon_1) + (\lambda_2 + \varepsilon_2)} = \frac{(\lambda_1 - \lambda_2) + (\varepsilon_1 - \varepsilon_2)}{(\lambda_1 + \lambda_2) + (\varepsilon_1 + \varepsilon_2)}. \tag{58}$$

If the noise values are small,

$$\text{SNR}(\lambda_1) \gg \varepsilon_1; \text{SNR}(\lambda_2) \gg \varepsilon_2; \varepsilon_1 \approx \varepsilon_2 \approx 0 \tag{59}$$

we can ignore them. But if Signal to Noise Ratio (SNR) is low the weight of the noise terms in the ratio grows and the measurement data describes only the noise.

2.6.8. Vegetation Indices

These indices are used mostly in remote sensing. Reader should be aware that most of them include weights for atmospheric disturbance correction. Since the measurement distance applied in NIRS is very short compared to satellite imaginary the atmospheric effect can be ignored.

However, remote sensing and NIR are still not that different; both apply wavelength analysis of light in order to find out if the analysed object contains interesting wavelengths. This is based on the basic idea that each molecule compound has it's own absorption curve of photon energy.

As described by Haboudanea et al. (2004), Green Leaf Area Index (LAI) is a key variable which is used by crop physiologist and modellers when they estimate the foliage cover, crop growth and yield. Different models exist which give LAI estimates and in the following they, such as Ratio of Vegetation Index (RVI) and Difference Vegetation Index (DVI), are all described under the term Vegetation Indices.

Research work around vegetation indices has been quite active. In *Remote Sensing of Environment* -journal alone there exists (January 25th 2011) in total 2494 articles which handle vegetation indices. Though most of them are applying vegetation indices, not so much of developing them further.

The one of the earliest (Schowengerdt 2007) vegetation indices was RVI shown in Eq. (60). In literature it is also quite often called as Simple Ratio (SR) or Vegetation Index (VI). In Eq. (60), the nir is the spectral range of near-infrared light and red is the spectral range of red light.

$$\mathrm{RVI} = \frac{R_{\mathrm{NIR}}}{R_{\mathrm{red}}} \tag{60}$$

The simple DVI, which is calculated as difference of intensity of light in NIR and red region, is given in Eq. (61). DVI is usually written as

$$\mathrm{DVI} = \mathrm{nir} - \mathrm{red} \tag{61}$$

The modulation ratio for NIR and reflectance is called the Normalized Difference Vegetation Index (NDVI) (see Eq. (62)) (Schowengerdt 2007). It is based on the contrast between the maximum absorption in the chlorophyll and maximum reflection leaf cellular structure. In other words, if there exist deep layer of biomass (forest), the index will have greater value than thin layer. NDVI is defined as:

$$\mathrm{NDVI} = \frac{R_{800} - R_{670}}{R_{800} + R_{670}} = \frac{\mathrm{RVI} - 1}{\mathrm{RVI} + 1} \tag{62}$$

NDVI is widely used but improved versions have been implemented such as Renormalized Difference Vegetation Index (RDVI) in Eq. (63) by (Roujean & Breon 1995).

It was meant to combine the advantages of DVI and NDVI. RDVI is defined as:

$$\text{RDVI} = \frac{R_{800} - R_{670}}{\sqrt{(R_{800} + R_{670})}} \tag{63}$$

In Eq. (64) Modified Simple Ratio (MSR) is shown which has been suggested as improvement of RDVI (Haboudanea et al. 2004). It is stated (Roujean & Breon 1995) to be less sensitive to geometrical and optical properties of plant canopies. MSR is defined as:

$$\text{MSR} = \frac{\left(\frac{R_{800}}{R_{670}} - 1\right)}{\sqrt{\left(\frac{R_{800}}{R_{670}} + 1\right)}} \tag{64}$$

In Eq. (65) is shown the Soil-Adjusted Vegetation Index (SAVI) (Schowengerdt 2007). L is an empirically determined constant, which tries to minimise the indices sensitivity to the reflectance variation of soil background. The value of L is typically around 0.5 according to Schowengerdt (2007: 189).

$$\text{SAVI} = (1 + L)(R_{800} - R_{670})/(R_{800} + R_{670} + L) \tag{65}$$

Qi et al. (1994) has developed an improved version of SAVI, which is called Modified Soil-Adjusted Vegetation Index (MSAVI) and is shown in Eq. (66). The advantage, compared to SAVI, is that the weight variable L is self-adjusted and does not appear in the formulation.

$$\text{MSAVI} = \frac{1}{2}\left[2R_{800} + 1 - \sqrt{(2R_{800} + 1)^2 - 8(R_{800} - R_{670})}\right] \tag{66}$$

The Enhanced Vegetation Index (EVI) is shown in Eq. (67) (Schowengerdt 2007), where C_1 and C_2 are aerosol resistance term coefficients and G is the gain factor. In MODIS-EVI algorithm these usually are set as $L = 1$, $C_1 = 6$, $C_2 = 7.5$ and $G = 2.5$ (Schowengerdt 2007).

The EVI is an “optimised” index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a

Table 3. Summary of vegetation indices.

RVI	$\frac{R_{\text{nir}}}{R_{\text{red}}}$
DVI	nir $-$ red
NDVI	$\frac{R_{800}-R_{670}}{R_{800}+R_{670}} = \frac{\text{RVI}-1}{\text{RVI}+1}$
RDVI	$\frac{R_{800}-R_{670}}{\sqrt{(R_{800}+R_{670})}}$
MSR	$\frac{\left(\frac{R_{800}}{R_{670}}-1\right)}{\sqrt{\left(\frac{R_{800}}{R_{670}}+1\right)}}$
SAVI	$(1+L)(R_{800}-R_{670})/(R_{800}+R_{670}+L)$
MSAVI	$\frac{1}{2}\left[2R_{800}+1-\sqrt{(2R_{800}+1)^2-8(R_{800}-R_{670})}\right]$
EVI	$G\left(\frac{R_{\text{nir}}-R_{\text{red}}}{L+R_{\text{nir}}+C_1R_{\text{red}}-C_2R_{\text{blue}}}\right)$
TVI	$0.5\left[120(R_{750}-R_{550})-200(R_{670}-R_{550})\right]$

decoupling-coupling of the canopy background signal and a reduction in atmosphere influences. The R_{nir}, R_{red} and R_{blue} are atmospherically corrected or partially atmosphere corrected surface reflectance of near-infrared, red and blue colour, accordingly. EVI is more responsive to canopy structural variations than NDVI (Schowengerdt 2007). EVI is defined:

$$\text{EVI} = G\left(\frac{R_{\text{nir}}-R_{\text{red}}}{L+R_{\text{nir}}+C_1R_{\text{red}}-C_2R_{\text{blue}}}\right) \tag{67}$$

Transformed Vegetation Index (TVI) is given in Eq. (68) (Schowengerdt 2007). It's general idea is to determine the area between green peak, near-infrared and the minimum reflectance in the red region. The area will increase if chlorophyll absorption increases, which means decrease of red reflectance and increase of NIR reflectance due to leaf tissue abundance. TVI is:

$$\text{TVI} = 0.5\left[120(R_{750}-R_{550})-200(R_{670}-R_{550})\right] \tag{68}$$

The indices are shown in Table 3. From the table it can be seen that there exist several models which can be used to find estimates for LAI. What is interesting, is that plant

Table 4. Summary of chlorophyll indices.

MCARI	$[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})] \frac{R_{700}}{R_{670}}$
RII	$\int_{705}^{750}(R_\lambda/R_{705} - 1)d\lambda$
YI	$-10(R_{\lambda_{-1}} - 2R_{\lambda_0} + R_{\lambda_{+1}})/\Delta\lambda^2$
Chl NDI	$\frac{R_{750}-R_{705}}{R_{750}+R_{705}}$

foliage has quite clear reflectance and absorption behaviour, as compared to soil. As in green foliage there is a lot chlorophyll, we will next look for more detailed view of that.

2.6.9. Chlorophyll and Anthocyanin Indices

Kim et al. (1994) developed the Chlorophyll Absorption Ratio Index (CARI) which measures the chlorophyll absorption at 670 nm relative to green 550 nm and reflectance at 700 nm. It was then simplified by Daughtry et al. (2000) to Modified Chlorophyll Absorption Ratio Index (MCARI), which is shown in Eq. (69).

$$\text{MCARI} = [(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})] \frac{R_{700}}{R_{670}} \tag{69}$$

The Reflectance Integral Index (RII) was proposed by Gitelson et al. (1999). It is calculated using a discrete summation approximation of the integral:

$$\text{RII} = \int_{705}^{750}(R_\lambda/R_{705} - 1)d\lambda \tag{70}$$

The Yellowness Index (YI) is used to indicate chlorosis in stressed leaves by using the shape of of reflectance spectra around 600 nm. Adams et al. (1999). It can be seen in Eq. (71) where λ_{-1} is 580 nm, λ_0 is 624 nm, λ_{+1} is 668 nm and $\Delta\lambda$ = 44 nm.

$$\text{YI} = -10(R_{\lambda_{-1}} - 2R_{\lambda_0} + R_{\lambda_{+1}})/\Delta\lambda^2 \tag{71}$$

Chlorophyll-a has an absorption peak at 680 nm and chlorophyll-b has a peak at 645-650 nm. According to Blackburn (1998), the 635 nm range can be used for testing of chlorophyll-b concentration, since it correlates better with chlorophyll-b content.

Also the wavelengths R_{730}, R_{706} and R_{680} have successfully been used in chlorophyll concentration analysis.

Richardson et al. (2002) stated that in their study the best reflectance indice for measuring chlorophyll content was Chl NDI, which can be seen in Eq. (72). However, it was not among those of commonly used in the literature.

$$\text{Chl NDI} = \frac{R_{750} - R_{705}}{R_{750} + R_{705}} \tag{72}$$

Solomakhin & Blanke (2010) presented Normalised Anthocyanin Index (NAI), see Eq. (73), on their apple analysis. The structure of index is easily explained by the fact that anthocyanin is a breakdown product of chlorophyll, therefore the less chlorophyll is in apples, the more anthocyanin can exist.

$$\text{NAI} = \frac{I_{780} - I_{570}}{I_{780} + I_{570}} \tag{73}$$

2.6.10. Water in Measurements

Water is a strong absorber in the IR region of spectrum. In measurement of high water content food samples, where water content is over 80 %, the spectra is strongly dominated by water absorption.

Many high-watery food items, such as milk, fresh potatoes and most fruits, show strong absorption near 1400–1410 nm region. However, the 1900–1950 nm region has been applied in quantitative analysis of water content in foods. (Büning-Pfaue 2003)

Water absorption indices are commonly used in remote sensing applications. They are used directly, in determining the land versus water mass and indirectly determining the properties of object [liquid water content]. However, frozen water, liquid water and vapour have all different absorption wavelengths.

Penuelas et al. (1993) defined the so called Water Band Index (WI), shown in Eq. (74), which can be used in determining the liquid water content in plants by applying reflectance in the 950–970 nm region. WI is defined as:

$$\mathrm{WI} = \frac{R_{900}}{R_{970}} \tag{74}$$

Gao (1996) defined and presented Normalized Difference Water Index (NDWI) (as shown in Eq. (75), which uses 860 nm and 1240 nm, which are both liquid waters absorption peaks. According to Gao (1996) it is less sensitive to atmospheric distortions than NDVI, when used as vegetation index. However, the main usage function is for determining the water content of plants [from space].

$$\mathrm{NDWI} = \frac{R_{860} - R_{1240}}{R_{860} + R_{1240}} \tag{75}$$

According to Bobelyn et al. (2009) the liquid water absorption peaks are at 970, 1170 and 1450 nm. However, as Blakey et al. (2009) has shown, reliable NIR model for predicting water content in Kiwi -fruit can be achieved by using only the 1100–2000 nm range.

Due to the fact that liquid water is not only $\mathrm{H_2O}$ but includes also OH -groups, four maximum are usually used in water content analysis. These peaks are located at about 970, 1190, 1450 and 1940 nm. Water also associates strongly with ions, organic monomers and polymers by hydrogen bonds, which causes that water absorption spectra is influenced by effects of these solutes (Büning-Pfaue 2003). Quite often this is used to measure these solutes from the water peaks.

All this means that applying NIR spectra in measurement of water content is not as simple as it might seem.

2.7. Spectrophotometry

The Bouguer, Lambert and Beer relationship assumes that there exists a change in spectrometer response, which is usually expressed in form (Workman & Springsteen 1998):

$$T = \frac{I_t}{I_0} = 10^{-\xi cl} \tag{76}$$

where T is the transmittance, I_t is the intensity of the transmitted light, I_0 is the intensity of actual light, ξ is the molar extinction coefficient, c is concentration and, l is the path length. Similarly if we look into reflectance and use I_r for reflected intensity, then we can write the reflectance R as in Eq. (77):

$$R = \frac{I_r}{I_0} \tag{77}$$

To analyse the amount of absorption, we use equations (17) and (18). We can write transmittance T in form, shown in Eq. (78), where I is the intensity, R reflectance, A absorbency, S scattering, F fluorescence, ϵ measurement error, and ε stands for noise.

$$T = I - R - A - S - F - \epsilon - \varepsilon \tag{78}$$

During measurement setup the white- and black references need to be measured. The calibration black reference includes the background noise of the environment and the measurement system. The white reference should explain the largest possible value of saturation of the system. White reference value should be selected so that the maximum measurement value of the sample would not be more than 80 % of the white reference value. In the equations in this thesis the black reference is referenced as I_B and white reference as I_W.

Because of the non-linearity of light sources, the measured signal has to be corrected. This is done (see Eq. (79)) by removing background noise from the signal. R_W in the equations is the reflection coefficient, which describes how much light intensity reflects back (in average) from standard white reference. We are using in our measurements standard white with manufacturer reported $R_W = 0.98$. Therefore the actual light intensity can be written as:

$$I_0 = \frac{I - I_B}{I_W - I_B} \cdot R_W \tag{79}$$

And the SNR is:

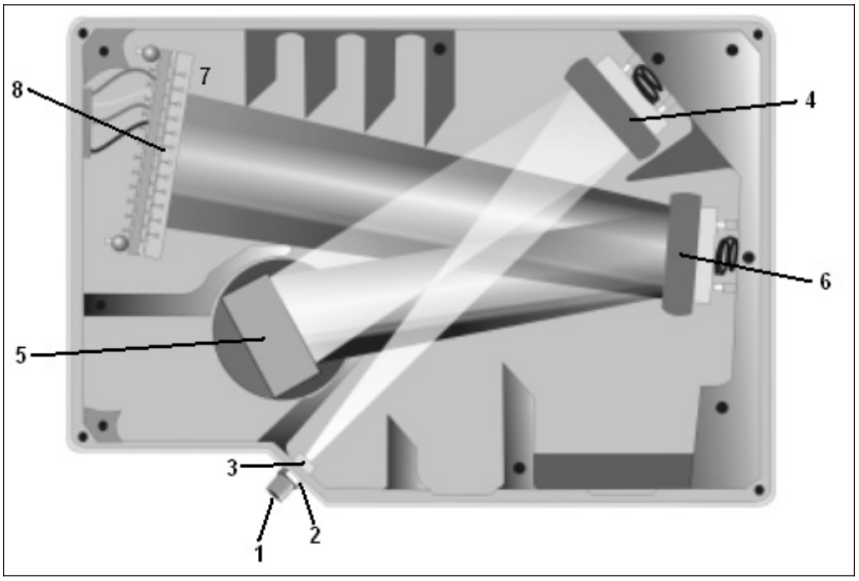


Figure 14. Typical components and light path in spectrometer device. 1) connector; fiber connector used to connect probe to spectrometer, 2) slit; it is used to control the amount of light entering into spectrometer, 3) filter; pass only those wavelengths which are needed, 4) collimating mirror; focuses light towards grating, 5) grating; diffracts the light from collimating mirror, 6) focusing mirror; focuses light into L2 detector collection lens, 7) L2 detector collection lens; focuses the light into CC detector elements, 8) CCD detector; CCD detector collects the light received from the focusing mirror and L2 detector collection lens and converts the measurement to digital signal. (Ocean Optics 2008)

$$\mathrm{SNR} = 10 \log \left(\frac{\mathrm{mean}(\lambda)}{\mathrm{std}(\lambda)} \right), \tag{80}$$

where $\mathrm{mean}(\lambda)$ is the mean, and $\mathrm{std}(\lambda)$ is the standard deviation of measured values in wavelength λ.

2.7.1. Devices

Fig. 14 illustrates a typical spectrometer. Different spectrometers may have different components, but the basic principle is similar. Applying mirrors and filters can be used to change the wavelengths recorded and sensitivity of the spectrometer. Some devices may be equipped with interchangeable filters.

Usually filtering is applied in probes, if it is needed for measurement purposes. Applying mathematical filtering in a post-measurement analysis is much easier, and gives us more possibilities if we have all measurement data available. Best dynamics from the measurements is achieved when both these filtering methods are applied correctly.

2.7.2. White and Black Reference

In real life, a measurement signal consists of background noise and signal. In order to minimise measurement error in spectrophotometrical measurement we apply two known measurements; white and black reference. In Fig. 15 we have a simulated signal (shown as thick line), white reference (shown as thin line) and background noise (abs(λ) in this example, shown as line with bullet).

Background noise consists of environmental electrical noise, irradiance, and other sources such as the noise in CCD. We can measure this noise by blocking all light entering into the spectrophotometer. The reference signal is any known and repeatable light entering into the spectrophotometer probe. In reflection measurement, a standard reflector is used, which is a surface whose reflectance and absorption properties are known. In absorption measurement direct light from the light source can be used for full saturation of CCD.

The measured signal in this simulation is calculated by adding background noise and actual signal, as can be seen in Fig. 15. We then use Eq. (79) to separate the actual measurement signal from the noise by subtracting the noise (black reference) and normalising the measurement by constant (white reference $-$ black reference). The corrected signal is shown in Fig. 16.

In Fig. 17 we have calculated the SNR for the simulated signal. It shows, also in Fig. 16, that SNR value is higher at the center area. This is due to the fact that there is no big difference between signal value and background noise in the near ends of the measurement range. This is quite often similar in spectroscopic measurements.

2.7.3. Integrating Sphere

Integrating sphere is one of the most usable probes in diffuse reflection measurements. Workman & Springsteen (1998) describes the integrating sphere to be a hollow ball coated (inside) with a highly reflective and diffuse material. The important thing is that the material used for coating is optically stable, wavelength independent, and gives known reflectance in wavelengths used in measurement. This results in repeatable measurements.

Fig. 18 gives an schematic illustration of an integrating sphere. The angle between the sample normal and the light ray entering into sphere is usually between 8 to 10 degrees.

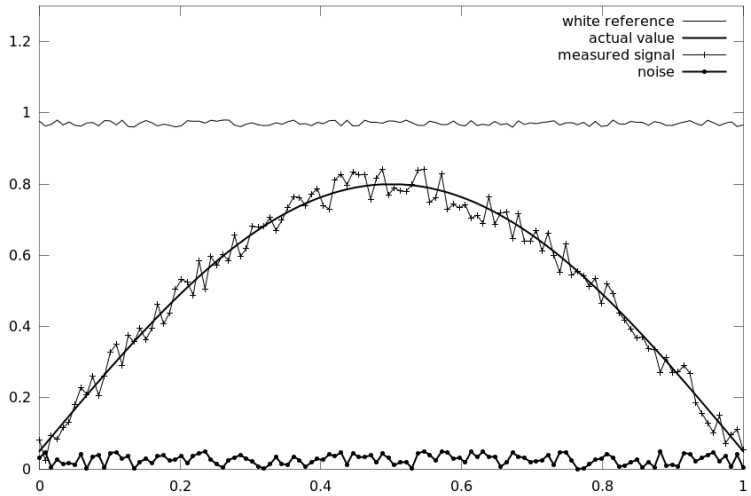


Figure 15. Simulated white reference (thin line), actual value (thick line), measured signal (line with $+$) and background noise (line with $\bullet$).

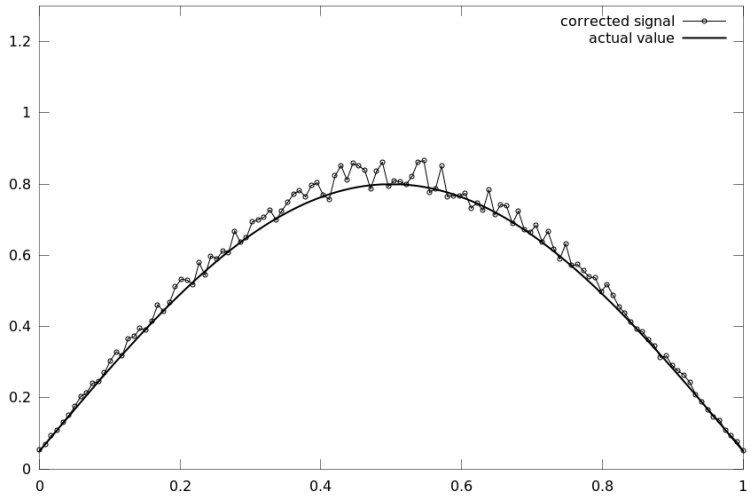


Figure 16. Comparison of simulated corrected signal (single line) and the actual value (line with $\circ$). The significance of the noise is greater where the signal is weak, in this case in both ends of the spectra.

The thick line under the detector is a baffle to block the direct irradiation from the sample entering into the detector. The sample port is a hole where light ray exits from the sphere and can collide with the sample. Some light absorbs, and some reflects from the sample and enters back to the sphere.

A part of the light will be absorbed and another part will be reflected from the sample. Reflections occur at several angles, as shown in Fig. 5. Rays may reflect several times

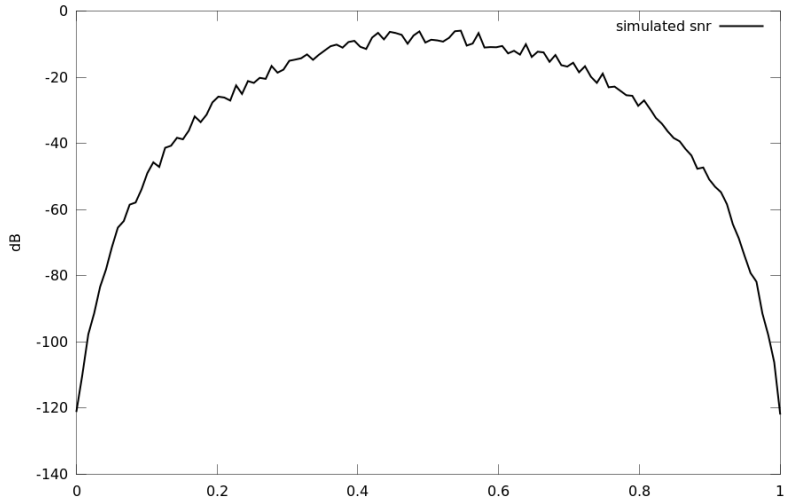


Figure 17. Simulated signal (Fig. 15) SNR in dB. From this illustration can be seen that the noise is greater in the middle of measurement area.

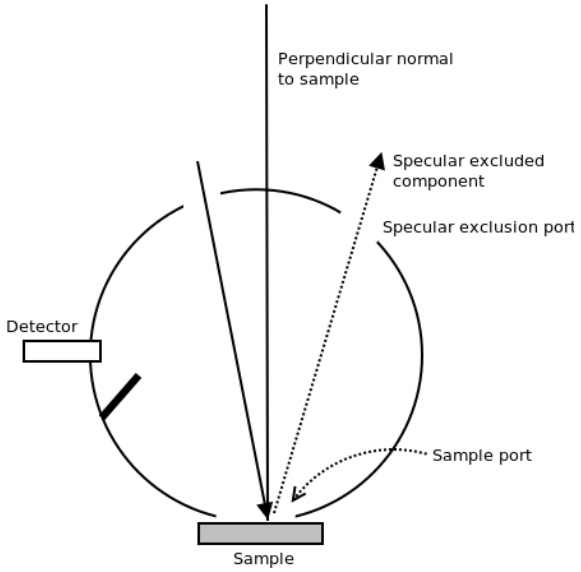


Figure 18. Cross section of single-beam integrating sphere. Adapted from (Schanda 2007: 52).

before entering into the detector. This causes the intensity of light entering to the detector to be a sum of several rays. Hence the name integrating sphere. The specular exclusion trap allows the specular reflection exclusion, if required.

According to Schanda (2007) if the reflectance of sphere coating $p(\lambda)$, is high (> 97 %) the integrating sphere magnifies spectral differences by $\frac{p(\lambda)}{1-p(\lambda)}$. This, together with the fact that sphere is angle independent due to the integration, explains why the integrating sphere is such a popular probe in chemometrics measurements.

3. MEASUREMENTS

In this Chapter the NIR measurements made are described and the measurement classes are explained. Also the measurement results are brifly commented.

When measuring liquids we most commonly measure the transmission through the sample. When light passes through the somewhat transparent sample, some light is absorbed. In this case Eq. (18) applies.

If we have the measurement setup illustrated in Fig. 19, we can write the equation of the amount of light entering into the test head as in Eq. (81), where a_1 is the absorption ratio of the liquid when light first time travels through it. Parameter a_2 is the absorption ratio when light travel second time through the liquid—after reflection from mirror below sample whose reflectance constant is r_m Parameter s_1 is used for scattering modelling of light when light ray enters through surface first time and s_2 for scattering at the bottom of the liquid.

$$I = (I_0 - R)(a_1 r_m a_2 s_1 s_2) \tag{81}$$

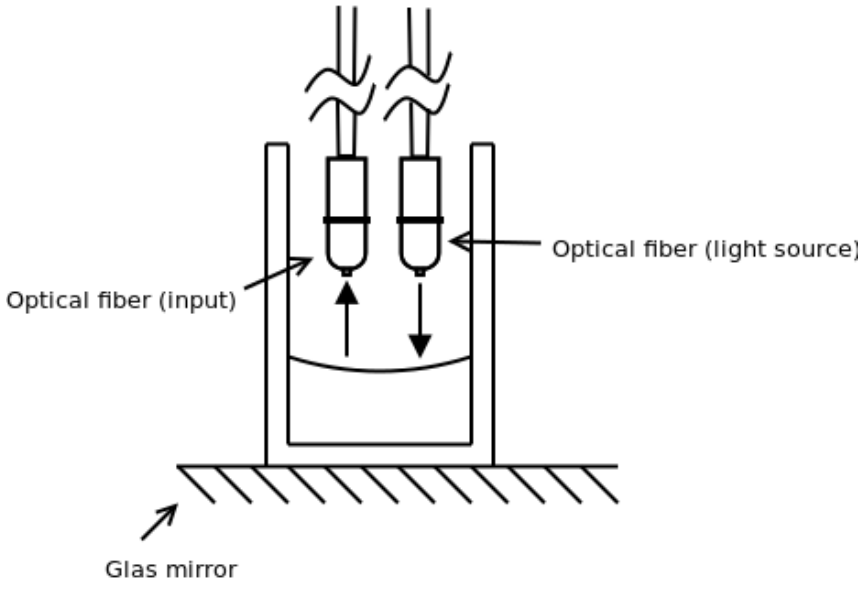


Figure 19. Testing liquid in tube.

3.1. Tomatoes

Tomato (*Solanum lycopersicum* L) belongs to eggplant family and is originated from Peru and Chile area. The edible part of tomato is a berry.

The target of the measurements was to find if NIR spectral measurement can be used in tomato classifying. For this purpose we received 5 classes of different tomatoes [denoted by 'A', 'B', 'C', 'D', 'E'] as can be seen in Table 5. Each class was measured using Ocean Optics HR4000 NIR spectrometer. The data measured was analysed using PCA and PLS methods. The condition and ripeness were based on visual observation without actual knowledge of their conditions.

Reflectance measurements were done from six directions to five individual tomatoes from each of the five classes. Four measurements were made from four side faces of the tomato, one from top and one from bottom of the tomato, see Fig. 20. In the measurements 30 ms integration time were applied, with boxcar width 5 and averaging 5 measurements. Boxcar is a smoothing method, where measurements are averaged with moving average on measurements near each other.

First measurements were made 1st of November and second measurements 3rd of November 2010. Between the measurements the tomatoes were stored in paper bags, each class members in their own semi-closed bag in a dark room having room temperature 23^{o}C. It seems that this preservation method allowed ripening of tomatoes since the reflectance intensity raised between measurements.

As can be seen in Fig. 20 the reflection spectra measured from different sides of tomato differs mostly in the navel of tomato, the top-left curve in the image. The measurement

Table 5. Tomato measurement classes. Table column ripeness was determined by visual inspection from the tomato samples. The value 1 stands for raw whereas 5 was ripe class. Different number of measurements in the first measurement series were caused by human error during measurement.

Class	Condition	Ripeness	Measurements
A	Good	2	30 + 30
B	Good	2	29 + 30
C	Good	2	29 + 30
D	Good	3	30 + 30
E	Good	3	30 + 28

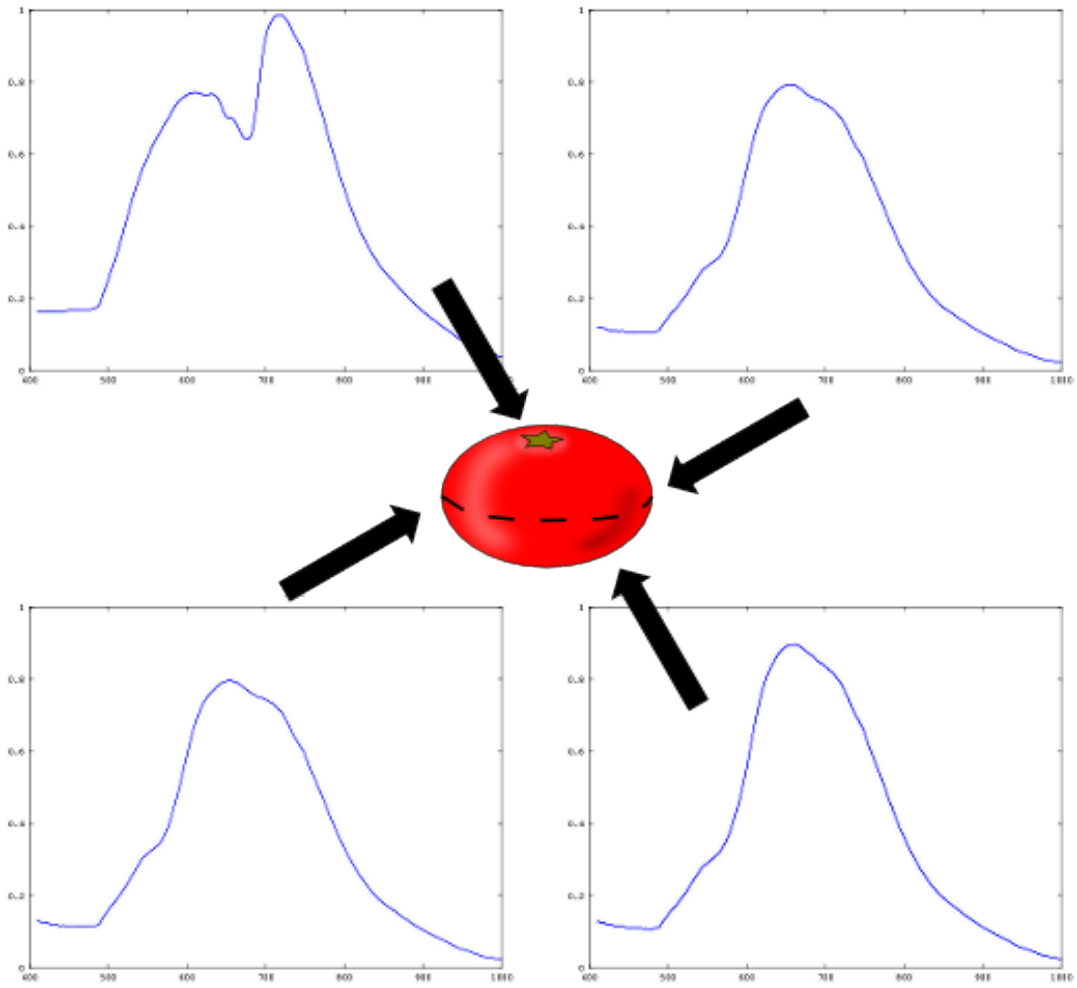


Figure 20. NIR measurements of NIR reflectance of tomato from different sides. Navel (top-left figure), side on equator (top-right figure), on equator other side (bottom-left figure) and below (bottom-right figure) measurements. The navel area produces clearly different reflectance spectra than the other.

from the sides of tomato are very similar to the bottom measurement, which differs only slightly from the equatorial measurements.

In Fig. 21 SNR of tomato measurements can be seen. The SNR could have been better and this estimates that the reliability of the classification based on tomato NIR spectra is not very strong in this case.

In the post-processing phase measurement data below 400 nm and above 1000 nm were discarded, since the measurement data outside of range 400–1000 nm consisted mostly from noise generated by the measurement equipment itself, since the light source didn't generate radiation in UV range. The range from 100 nm to 380 nm belongs to UVA (315–400 nm), UVB (280–315 nm) and UVC (100–280 nm).

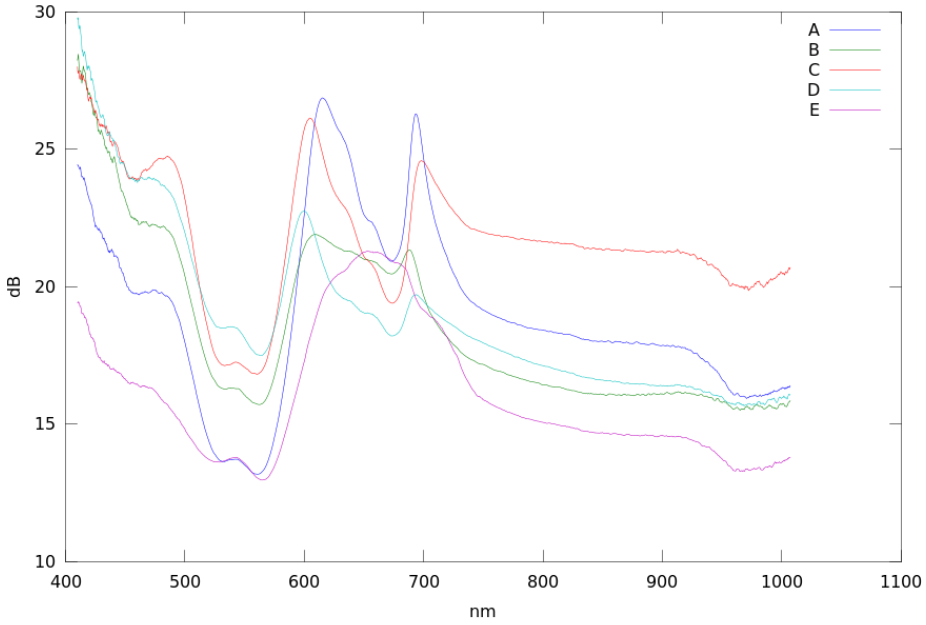


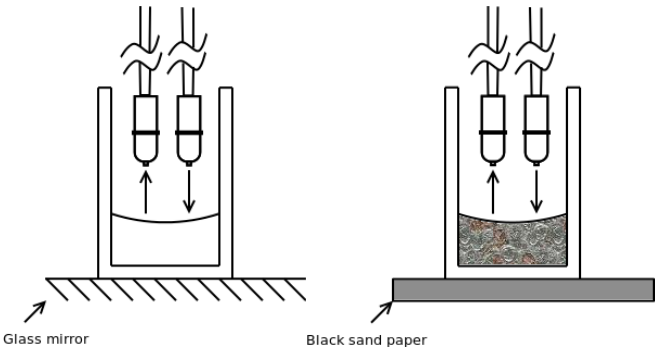
Figure 21. SNR for first tomato measurement in dB/nm. Each class represents tomato class mean. Higher values values denote higher signal values versus noise.

We assumed that our tomato samples were all healthy, since no information was given to suppose otherwise.

3.2. Medicine solution

The purpose of the measurements done was to find out if NIR spectrometry could be used to determine the concentration of medicine aqueous solution (*morphine hydrochlorium*). The medicine is water soluble and for the measurements it was diluted into distilled water. The chemical composition, according to WHO Pharmacopoeia Library (2006), is $C_{17}H_{19}NO_3, HCl, 3H_2O$. Aqueous solutions were prepared by laboratory personnel before they were given to us.

All measurements were done in a dark room (in the basement of the building) using black sandpaper or glass mirror as a background with Ocean Optics HR4000 NIR spectrometer. The sandpaper background was used because the scattering of black sandpaper is quite close to uniform scattering. Ordinary glass mirror allows good reflectance and is quite durable against aqueous contact. The measured liquids included both aqueous solutions which were filtered during the normal production process and solutions which were non-



(a) Test tube containing the liquid standing over glass mirror

(b) Test tube containing medicine powder over sandpaper

Figure 22. Measurement setup of aqueous medicine.

filtered.

The measurement setup of the liquids can be seen in Fig. 22. One optical fiber was used to transmit light from the halogen source to the test tube and another optical fiber was used as the measurement probe. The measurement probes of the spectrophotometer where inside of the test tube but not touching the liquid or the wall of the test tube.

One major measurement setup error was done during measurements. The amount of liquid solutions was not controlled between different measurement series. As the level of liquid in test tube varies also the intensity measured varies. This can most clearly be seen between the first and second liquid measurement. In the third measurement the amount of liquid was nearly the same as in the second. Due to the difference in liquid level in the test tubes a direct comparison between the whole series cannot be done.

Another mistake happened with integration time. In the first liquid measurement 5000 μs was used, but in the second and third measurement 4000 μs was used. This was due to the fact that reflectance intensity changed dramatically—due to the uncontrolled distance between probe and the liquid level in the tube.

3.3. Sodium bimetasulfide solution

In Finland, sodium [pyro]sulfide and sodium bimetasulfide (see Fig. 23 for chemical structure) is used to inhibit darkening of peeled potatoes. The darkening occurs due to oxidation of the starch. Inhibition of the darkening is achieved utilising easily soluble sodium

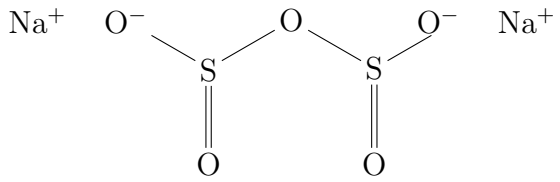


Figure 23. Generalised chemical structure of $Na_2S_2O_5$.

ions in aqueous solution, which prevents the oxidation process. Usually $\frac{20}{10000}$ (v/v) is applied.

The peeling process involves diluting the peeled potatoes in aqueous solution. When potatoes are diluted, they engulf some sulfide. If the concentration of solution is too low, potatoes need to be kept longer in the solution for equal amount of sulfide. On the other hand, if concentration is too high, potatoes will have too high amount of sulfide. Therefore if one wishes to control the perceived quality of peeled potatoes, it is required to also be able to control the amount of sodium bimetasulfide in the solution.

The following series of aqueous $Na_2S_2O_5$-solutions (v/v) were created: $\frac{5}{10000}$, $\frac{10}{1000}$, $\frac{20}{10000}$, $\frac{40}{10000}$, $\frac{80}{10000}$ and $\frac{160}{10000}$. Method for series creation was to weight amount needed for the strongest solution with accuracy scale and then dilute each sample to larger amount of ion stabilised water. The room temperature was 23^oC.

Liquids were put into test tubes and the absorption was measured using measurement head in the test tube, so that measurement head was located equally distance from the tube walls. Ocean Optics HR4000 measurements were done in dark room and LT Industries ParaFuel measurements with Technobothnia environment laboratory.

The lower region of NIR (400–1000 nm) was measured with Ocean Optics HR4000 and higher NIR region (1200–2400 nm) with LT Industries ParaFuel. As can be seen in Fig. 24 and 25, the differences were really small and in some wavelengths the absorption seems not to be linear with concentration.

The higher NIR region, however, seemed more interesting. The differences in Fig. 24 were too small to be noticed in the graph. The small differences were expected because water absorbs strongly NIR light.

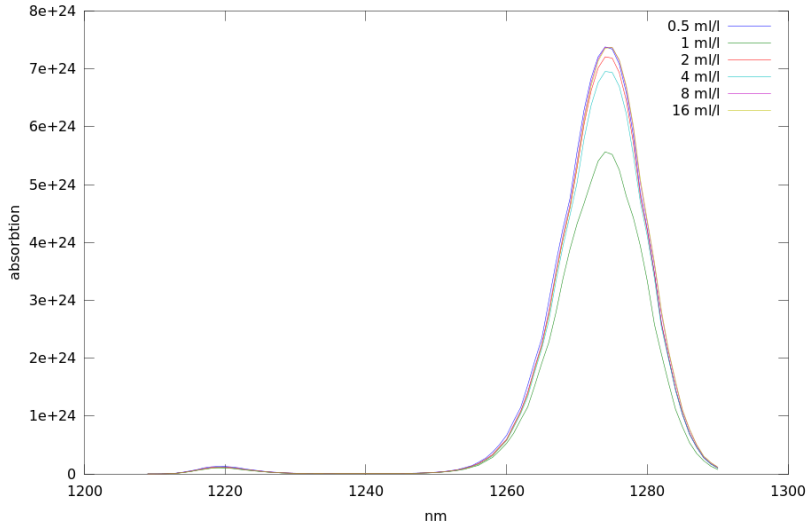


Figure 24. $Na_2S_2O_5$ absorption spectra measured using LT Industries ParaFuel. In the order of increasing concentration: 0.5, 16, 8, 2, 4, 1 ml/l.

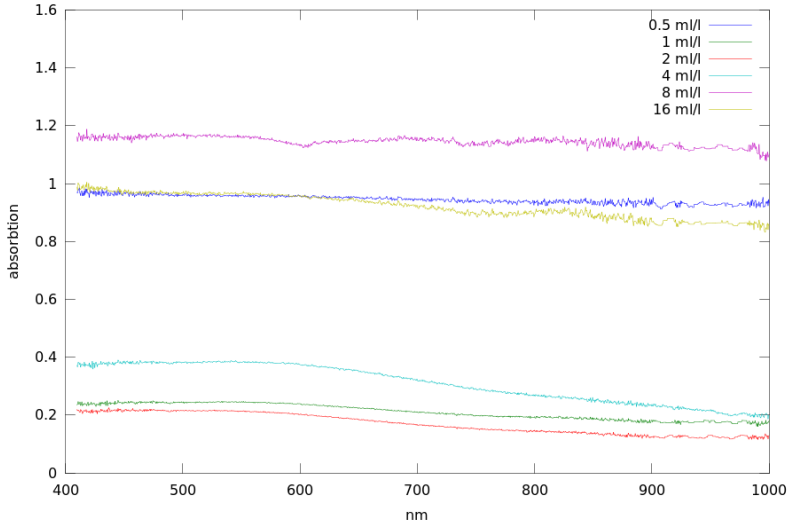


Figure 25. $Na_2S_2O_5$ absorption spectra measured using Ocean Optics HR4000. Differences in absorption level can be seen, but they do not follow exactly the order of concentrations of the dilutes: 8, 0.5, 16, 4, 1, 2 ml/l.

3.4. Blueberries

Blueberry (*Vaccinium myrtillus* L) grows wild in Finnish forests, from south to north. As the name states, the ripe berry is blue coloured. Chemically blueberries are composed of anthocyanins, polyphenols, sugars (sucrose, glucose, and fructose), water, fibre, ascorbic acid and other elements (Sinelli et al. 2008b).

From the consumers point of view the blueberry is a rich source of natural sugars, anthocyanins, and vitamins. Motivation for the blueberry measurements is to classify blueberry batches based on the amount of anthocyanin using NIR reflectance. This is interesting from the food manufacturer and distributor point of view.

Sugiyama et al. (2010) applied NIR to discriminant analysis for detecting foreign materials, mostly leaves and stems which are commonly picked among blueberries since blueberries are usually handpicked. They found that optimal wavelengths for foreign material determination were 1268 nm and 1317 nm. The foreign materials were clearly distinguished from blueberries, since they have low intensity and therefore also appeared black in images.

Giangiacomo (2006) studied interactions between water and sugar [glucose, fructose and sucrose]. A strong correlation between concentration of sugars and water was found in absorbance at 1928 nm wavelength area. Absorption increased when sugar concentration was higher. The study included samples where sugar concentration varied from 5 to 65%.

Rodriguez-Saona et al. (2001) applied Fourier Transform Near-infrared Spectroscopy (FT-NIR) for determining the sugar content in fruit juices. They applied aqueous solutions of sugar mixtures from 0–8% weight/volume for calibration and used transflection reflectance and integrating sphere.

Sinelli et al. (2008b) looked into blueberry (*Vaccinium corymbosum* L - from Northern Italy) and analysed their quality and nutraceutical content using near and mid-infrared spectroscopy. The results were promising.

Absorption of smashed frozen blueberries can be seen in Fig. 26. From this absorption curve we can see that absorption is strongest in the region between 2000 nm and 2300 nm. The high peaks found are listed in Table 6.

From this we can deduce that the sugar and water ratio, amount of anthocyanins and acids of blueberry batch can be predicted from NIR spectra so that comparison between batches of blueberries can be done.

For the blueberry measurements seven blueberry samples were created, as can be seen in Table 7. The samples 1, 2 and, 4 were randomly selected blueberries from given samples.

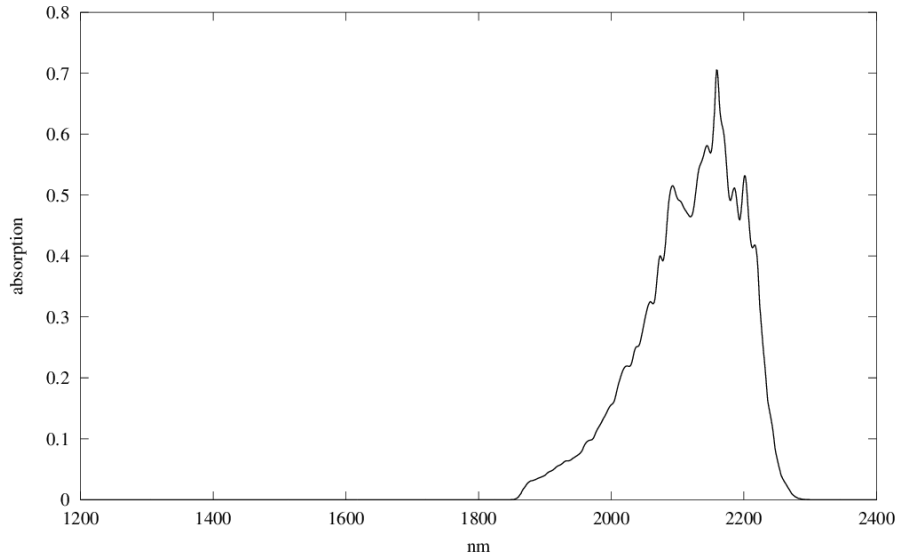


Figure 26. Frozen smashed blueberry absorption measured with LT Industries ParaFuel. High absorption can be seen in the 2000–2300 nm region.

Table 6. High peak wavelengths from blueberry absorption.

nm	relative absorbancy
2093	0.515
2145	0.581
2159	0.705
2186	0.512
2201	0.532
2216	0.418

The samples 3 and 3b were sorted so that 3 contained all sized blueberries while 3b contained only small sized blueberries.

Table 7. Blueberry sample sets.

Sample set	Origin
A	Riitan herkku frozen sample 1
B	Riitan herkku frozen sample 2
C	Riitan herkku frozen sample 3
D	Riitan herkku frozen sample 3b
E	Riitan herkku frozen sample 4
F	Supermarket frozen blueberries, origin Suonenjoki, Finland
G	Supermarket frozen blueberries, origin Poland

The blueberry samples selected when still frozen and possible leaves and other visual debris were removed before melting and smashing the samples. The reflectance measurement was done for the blueberries first when they were still frozen using Ocean Op-

tics HR4000-spectrometer. Then the blueberries were melted and smashed and trans-reflectance was measured using LT Industries ParaFuel-spectrometer. We then measured the pH using Mettler Toledo SevenEasy meter and measured the weights of samples. This data can be seen in Table 8. The weight itself does not have importance. It is included just to show that measurement samples were small and of similar size. Next the trans-reflectance was measured using again Ocean Optics HR4000-spectrometer. After this the blueberry samples were shipped to Kajaani University Consortium, CEMIS-Oulu/Analytical chemistry and bioanalytics laboratory for quantitative chemical analysis of anthocyanin and sugar content.

Table 8. Blueberry sample pH and weight. Weight was measured so that sample sizes were not too different.

Sample	pH	Weight g
A	3.02	72.40
B	2.93	80.60
C	3.20	64.75
D	3.15	63.90
E	3.15	71.50
F	3.09	63.55
G	3.07	61.95

Table 9 shows the anthocyanin contents from Kajaani University Consortium, CEMIS-Oulu/Analytical chemistry and bioanalytics laboratory HPLC-DAD analysis results for our measurement samples. The amounts are measured as mg/100g and calculated as equivalents of cyanidin-3-glucoside. From these figures we can see that the measured anthocyanin content varies between samples.

Table 9. Laboratory HPLC-DAD analysis of anthocyanin content in blueberries.

Anthocyanin	A	B	C	D	E	F	G
Delphinidin-3-galactoside	69	87	65	69	65	47	35
Delphinidin-3-glucoside	60	80	63	66	58	42	38
Cyanidin-3-galactoside	43	54	42	44	42	41	38
Delphinidin-3-arabinoside	57	75	59	61	52	39	26
Cyanidin-3-glucoside	43	59	47	51	45	43	52
Petunidin-3-galactoside	18	20	16	17	18	14	11
Cyanidin-3-arabinoside	36	52	42	42	36	34	31
Petunidin-3-glucoside	39	50	41	43	39	30	30
Peonidin-3-galactoside	4	5	4	3	3	4	5
Petunidin-3-arabinoside	13	16	13	14	12	10	8
Meonidin-3-glucoside	17	20	17	18	17	18	26
Malvidin-3-galactoside	17	20	16	15	16	12	14
Peonidin-3-arabinoside	2	2	2	1	1	2	2
Malvidin-3-glucoside	40	46	38	43	38	30	33
Malvidin-3-arabinoside	9	11	8	9	9	7	6
Total anthocyanin	468	597	474	497	451	372	357

4. ANALYSIS

In this chapter we explain the analysis of the measurements from the previous chapter and discuss of the conclusions which have been drawn based on the results.

4.1. Tomato Analysis

Data below 400 nm was discarded, since the light source used does not give efficiently light below that so that the measurement data consists mostly of background noise (light).

In Fig. 27 the spectral emission ratio change of carotenoids/chlorophyll (in 525 nm/ 672 nm) is shown from the first and second measurement. As the second measurement was done next day, some change in ratio was assumed. Since these tomato measurements are absorbtion measurements, decreased ratio means more ripe tomato.

The most important antioxidants in tomatoes are carotenoids (Clinton 1998) and phenolic compounds (Hertog et al. 1992). Most of the carotenoids in tomatoes consists of lycopene (about 93 %), and of β-carotene (about 7 %). Lycopene content varies with ripening and is mainly responsible for the red colour (Sun et al. 2010: p. 369). Lycopene content has been measured from 400-1000 nm range by Clément et al. (2008b) and founded that the result was almost as accurate as in 400-1500 nm range.

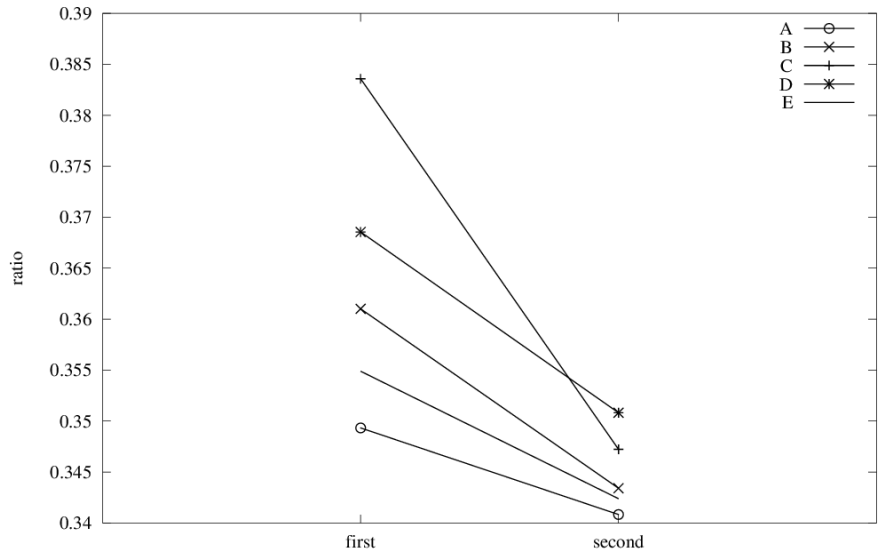


Figure 27. Tomato carotenoids/chlorophyll ratio as ripeness indicator.

The ripening of tomatoes consists of several processes, including the breakdown of chlorophyll and build-up of carotene (Sun et al. 2010: p. 370). Clément et al. (2008a) found that there are four factors which will explain up to 81 % of the total variance in tomato ripening:

1. Tomato Maturity State (TMS) is related to colour, lycopene content, firmness,
2. Titratable Acidity (TA), pH, and soluble solids contents,
3. gustatory index is related to electrical conductivity, soluble solids, TA and pH; and
4. soluble solids alone, and variety.

So we will determine the ripeness of the tomato based on estimated chlorophyll-carotene ratio. As previously was shown the main absorption wavelength for chlorophyll-b is at 435.8 nm.

Sugar content of tomatoes was analysed by applying the Eqs. (48) and (82) for wavelengths 902 nm and 874 nm. In Fig. 28 the measurements of tomatoes and the calculated sugar content with Brix equation (Sun et al. 2010: p. 363) is given by:

$$\text{Brix} = a\mathrm{d}^2 A_{874} + b\mathrm{d}^2 A_{902} + c \tag{82}$$

In function (82) the weights $a = 70.32$, $b = -438.84$ and $c = 19.01$ were used. These weights were obtained by Junichi and Mizuki (Sun et al. 2010) from melon sugar content measurements.

From Fig. 28 we can see that even when using the same weights for Brix equation that Junichi and Mizuki used for melon analysis, clear differences between tomato classes can be observed. According to results classes 'A' and 'B' had low sugar content, when compared to 'C', 'D' and 'E' classes, which would indicate that the latter were less mature.

We have plotted each measurement of tomato in the Brix graph to illustrate the fact that the Brix equation is also very vulnerable to navel of the tomato. All peaks in Fig. 28 are the measurement where light has reflected from the greenish-brownish navel.

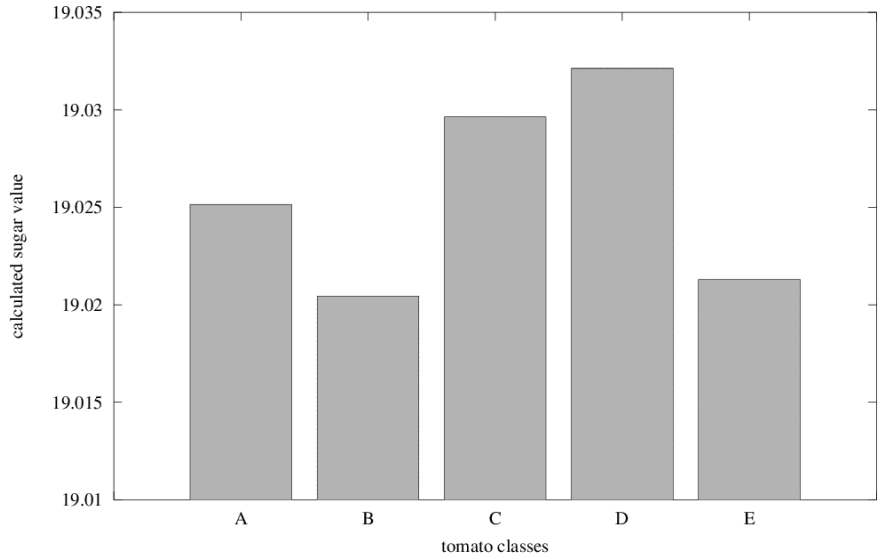


Figure 28. Tomato sugar content for each measurement, applying Brix equation.

As explained previously, the ripening of tomato includes chlorophyll breaking, so one way to describe the ripeness state of tomato is to measure the chlorophyll content. Several methods for this have been shown in the previous chapter.

The measurement data for tomatos had 3648 wavelengths. The wavelenths outside of 400 and 1000 nm was discarded because the light source has low intensity for these wavelengths. This left us with 2374 wavelengths. For the tomato PCA model, the data was reduced using polynomial fitting with 23 degree polynomial function and then 100 wavelengths were reconstructed with polynomial function. Missing data was replaced with arrays of zeroes.

In the Fig. 29 can be seen that the error of polynom fitting does not decrease in tomato measurement data much after the saturation point, approximately at degree of 23. Even with much higher, such as 100 degree, polynomial function, the error value does not decrease remarkably.

In the PCA models of the measurements, the first primary component has a clear dominance, which can be seen in Fig. 30. This means that most of the data can be described with the two first components and the difference between the measurement classes should be explained by these components.

In Fig. 31 the data from the first measurement is projected to the PC1 and PC2. Classi-

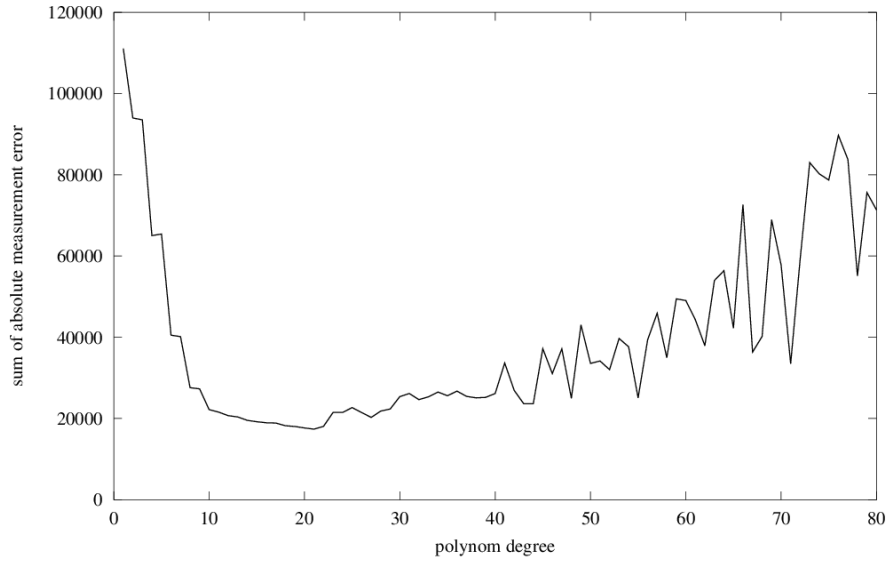
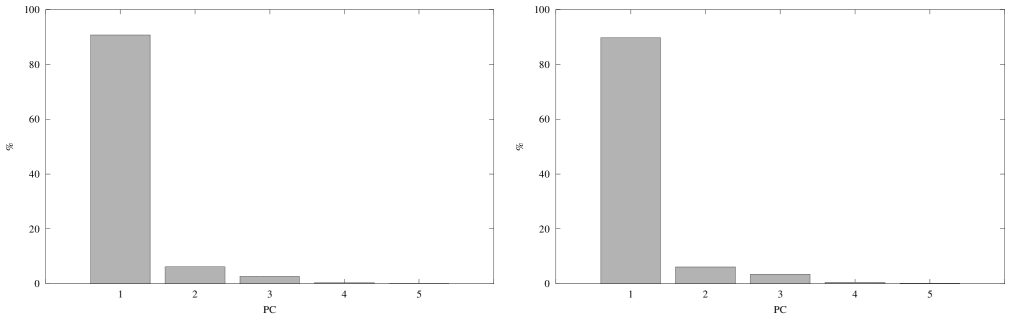


Figure 29. Sum of absolute error in polynom fitting. The actual sum of absolute error in polynom fitting reaches the saturation point at degree 23 and starts to increase after that.



(a) first measurement (b) second measurement

Figure 30. Tomato PCA model loadings. The first primary component dominates in the model and explains most of the measurement data ($> 97\ \%$).

fication of the tomato classes is not clear. Although the class 'A' tomatos are clustered quite nicely, which can be interpreted to indicate that the tomatoes in class 'A' are more similar than in other classes.

The third primary component does not help us in class separation, which can be seen in Fig. 32 and 33, where the data is projected to the PC2 and PC3, and to PC1 and PC3.

In Fig. 34 the data from the second measurement is projected to the PC1 and PC2. Classification of classes is even harder. After ripening 'A' and 'E' classes is still quite clear, but the 'B', 'C' and 'D' class are now almost identical. If we compare to the first PCA model, we can see that the spread of data on both primary components increases, which

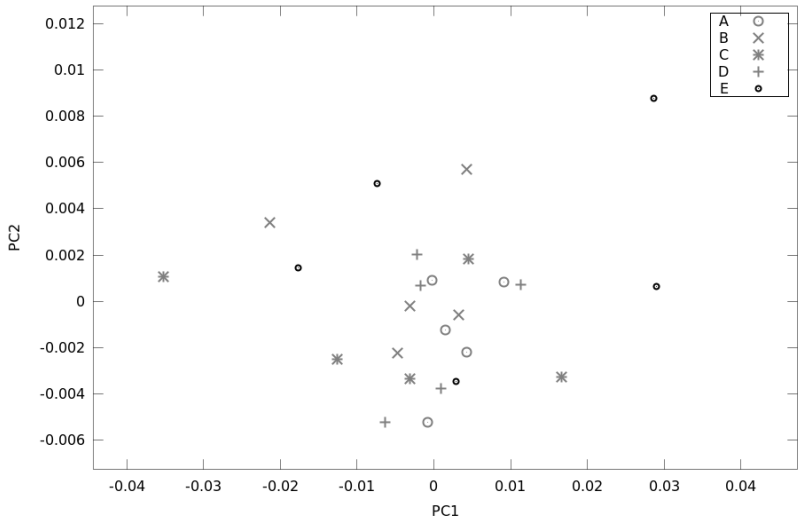


Figure 31. Tomato PCA model from the the first measurement day series. Data projected on PC1 and PC2. Each mark represents a measurement on individual tomato from classes A, B, C, D, and E. No clear classification can be seen.

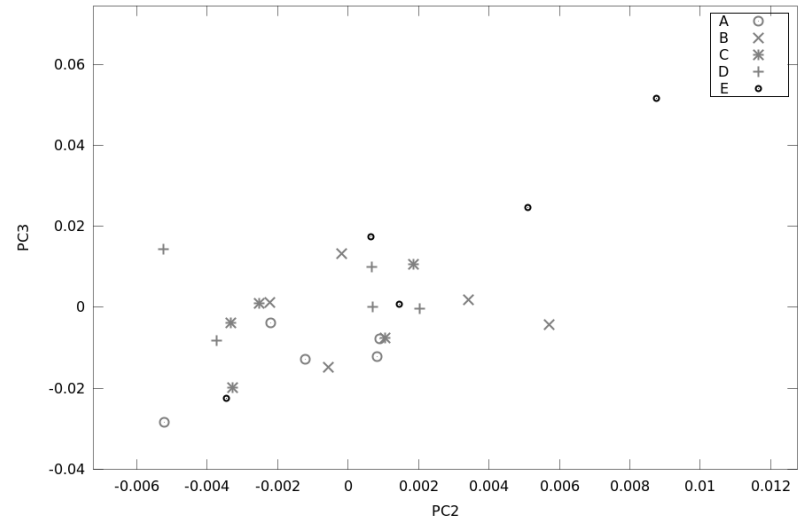


Figure 32. Tomato PCA model from the first measurement day series. Data projected on PC2 and PC3. Each mark represents a measurement on individual tomato from classes A, B, C, D, and E. No clear classification can be seen.

is mostly due to the ripening of the tomatoes.

As was shown in the model from first the measurement day series, the third primary component does not help us to classify tomatoes after ripening period either. Perhaps a look into the measurement data is required. The data consists of measurements from six sides of a tomatoes.

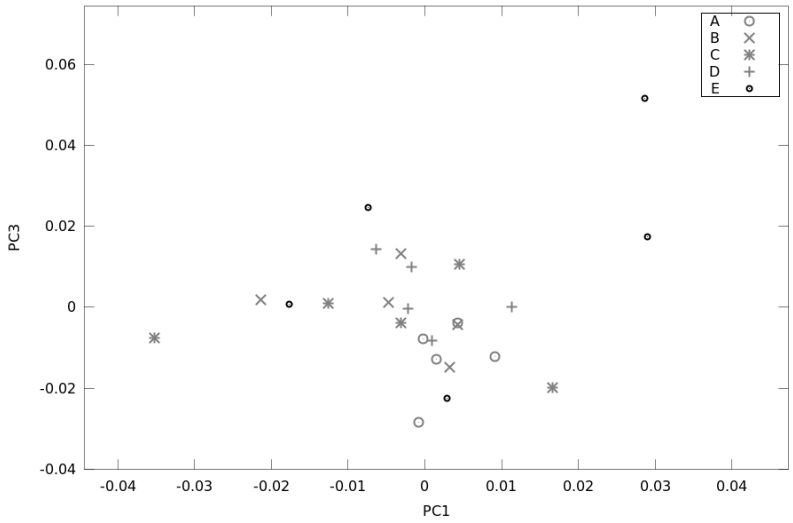


Figure 33. Tomato PCA model from the first measurement day series. Data projected on PC1 and PC3. Each mark represents measurements of individual tomato from classes 'A', 'B', 'C', 'D', and 'E'. No clear classification can be seen.

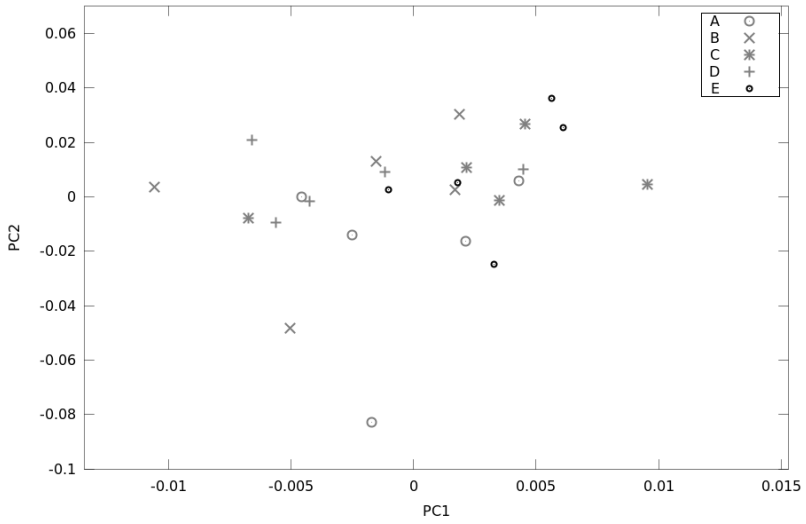


Figure 34. Tomato PCA model from second measurement. Data projected on two first primary components. Each mark represents measurements of individual tomato from classes 'A', 'B', 'C', 'D', and 'E'. No clear classification can be seen

Our next step is to discard the first measurement of the navel, the stem side of tomato, since the navel is mostly green or light brown and therefore PCA model tends to describe the size and colorization of the navel more than anything else. In our improved PCA model, see Fig. 36 and 37, we have removed navel measurements. The first primary component describes now 99.7% in both measurement data models and second component

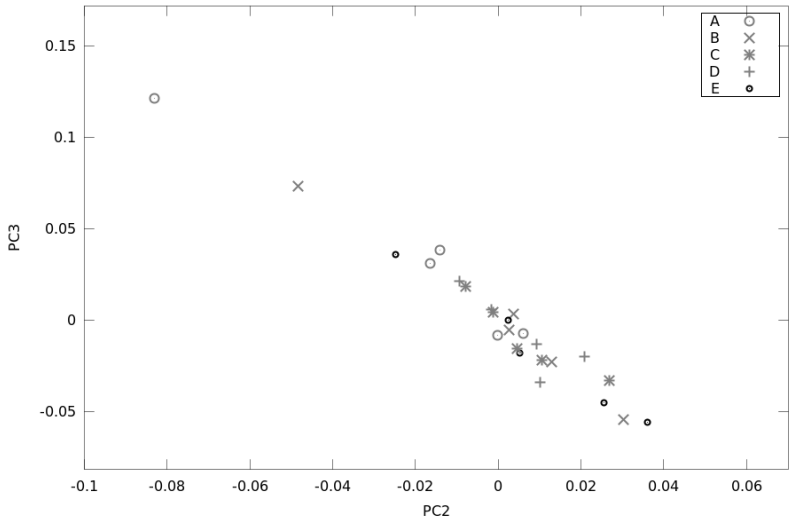


Figure 35. Tomato PCA model from second measurement. Data projected on PC2 and PC3. Each mark represents measurements of individual tomato from classes 'A', 'B', 'C', 'D', and 'E'. No clear classification can be seen.

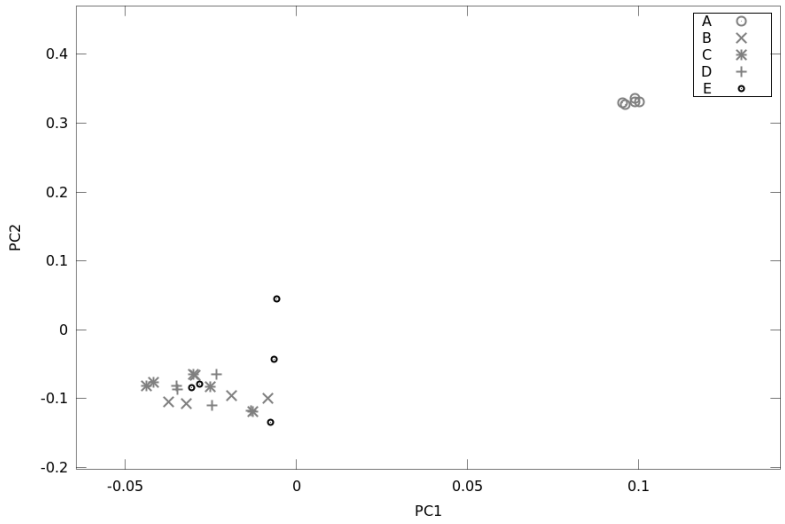


Figure 36. Improved tomato PCA model from first measurement, navel side measurement data removed. Data projected on two first primary components. Each mark represents measurements of individual tomato from classes 'A', 'B', 'C', 'D', and 'E'. Now the class 'A' is clearly separated from other classes.

2.6 % in the first, and 2.8 % in the second measurement model.

The class 'A' is now clearly separated from the other classes in both measurements. The removal of the navel measurement data improved the model, and the similarity of class 'A' tomatoes is stronger than previously. However, in this PCA model, we are not able

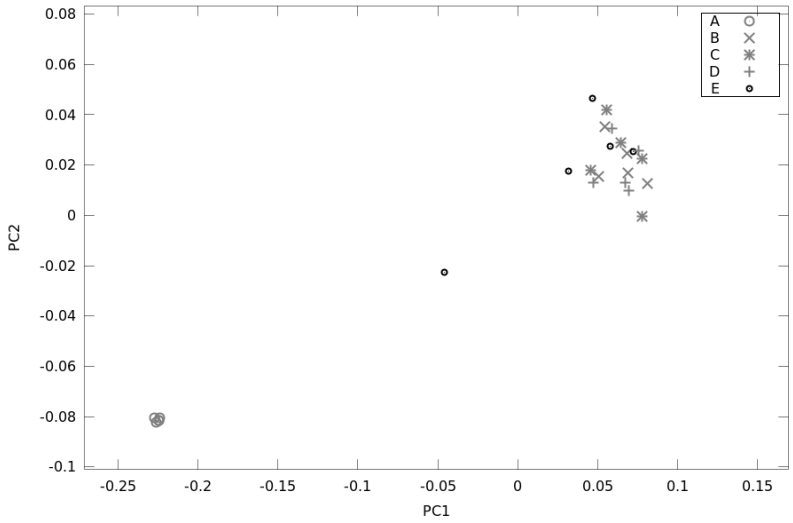


Figure 37. Improved tomato PCA model from second measurement, navel side measurement data removed. Data projected on two first primary components. Each mark represents measurements of individual tomato. Now the class 'A' separates celarly.

to separate 'B', 'C', 'D', and 'E' classes from each other. In 'E' class, there is more differences within class.

In other words, we are not able to confirm the Brix result that 'A' and 'B' classes would have the highest sugar content with our PCA model. Our PCA model does indicate that class 'A' is different from the other classes, but since we don't have sugar content measurements available and our only classification is visual inspection, we are not able to draw further conclusions. Due to the small number of samples, we didn't taste them, thus saving the samples for further measurements.

4.2. Blueberry Analysis

In Fig. 38 we have compared the different vegetation indices to blueberry anthocyanin index. It would be interesting if picking of blueberries can be timed, e.g. applying aerial photography, so that berries contain maximum amount of anthocyanin. The NAI seems to correlate best to anthocyanin content. As a preliminary study, we here answer to the question whether the anthocyanin content has influence to vegetative indices.

In Fig. 39 we have applied the NAIndex (see Eq. (73)) to our first blueberry sample measurements and compared it to the total anthocyanin content given by the HPLC-DAD

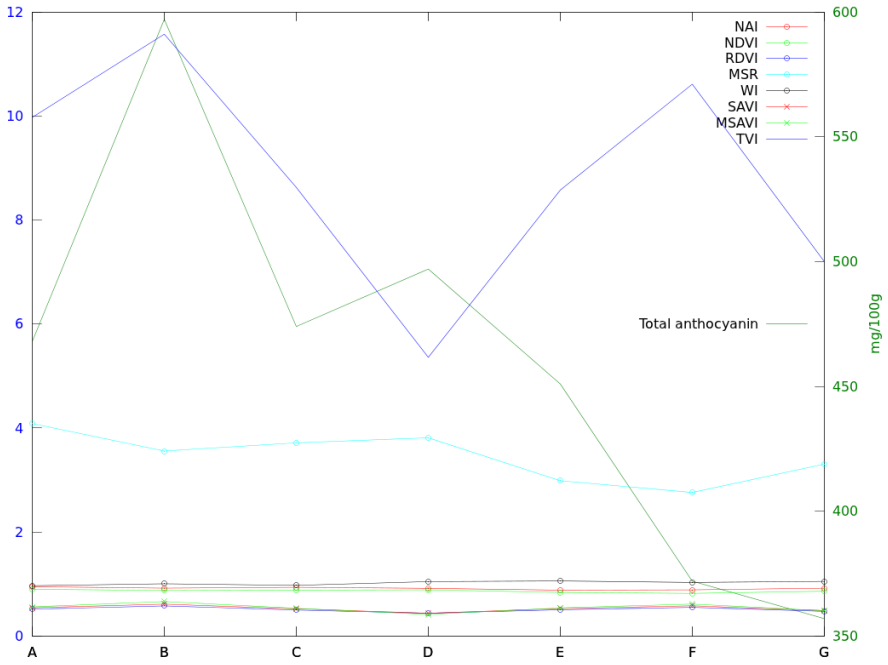


Figure 38. Blueberry vegetative indices and anthocyanin content. TVI and total anthocyanin correlate on classes 'A', 'B', 'C', and 'D' more than for classes 'E', 'F', and 'G'. Other indices do not correlate with anthocyanin amount.

analysis (Table 9). As we had three measurements per sample, the median of measurements is used in the graph.

In Fig. 40 we can see correlation plot between NAIndex results and total anthocyanins. If linear correlation existed, the graph would consists of straight horizontal line. Therefore we can safely say, that NAIndex is not a good approximation for the total anthocyanin content.

The NAIndex does not seem to be very good approximation for analysed anthocyanin components either. This can be seen from Fig. 41 since the correlation(s) are not even close to one for all tested series. At least based on our measurements, the NAIndex does not give confident results of the anthocyanin contents for blueberries.

Although our blueberry measurement were made using plastic containers, we assumed that the plastic container was not major error source. This safe decicion was made, based on our measurements of absorption of plastic container, and since Ozaki et al. (2007) lists that the plastics we used (PET), have OH stretching and deformation in range 1879–1908

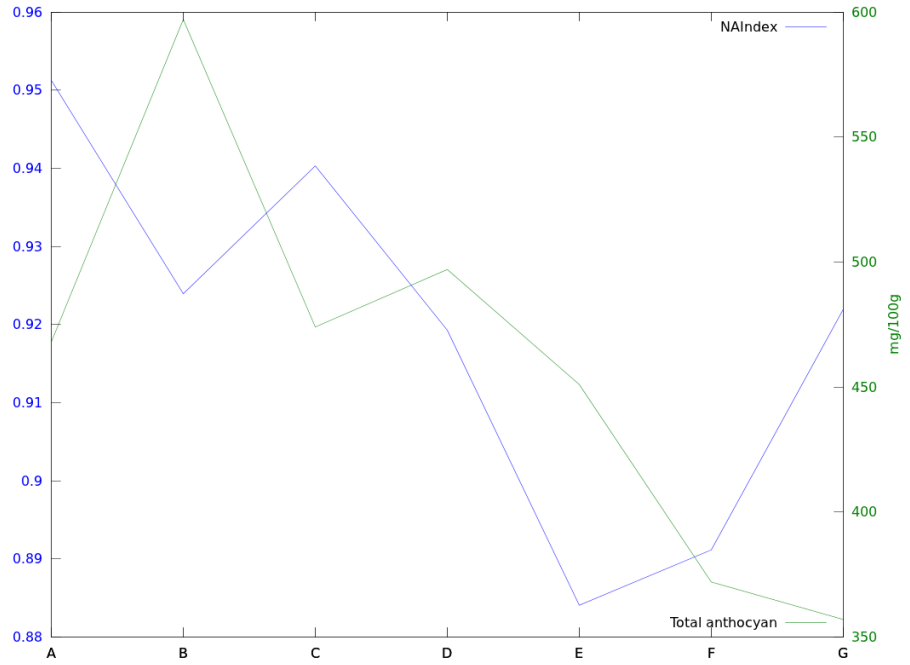


Figure 39. NAIndex and total anthocyanin amount of blueberry sample series. For some sample sets the correlation between NAIndex and total anthocyanin content seems to be quite good.

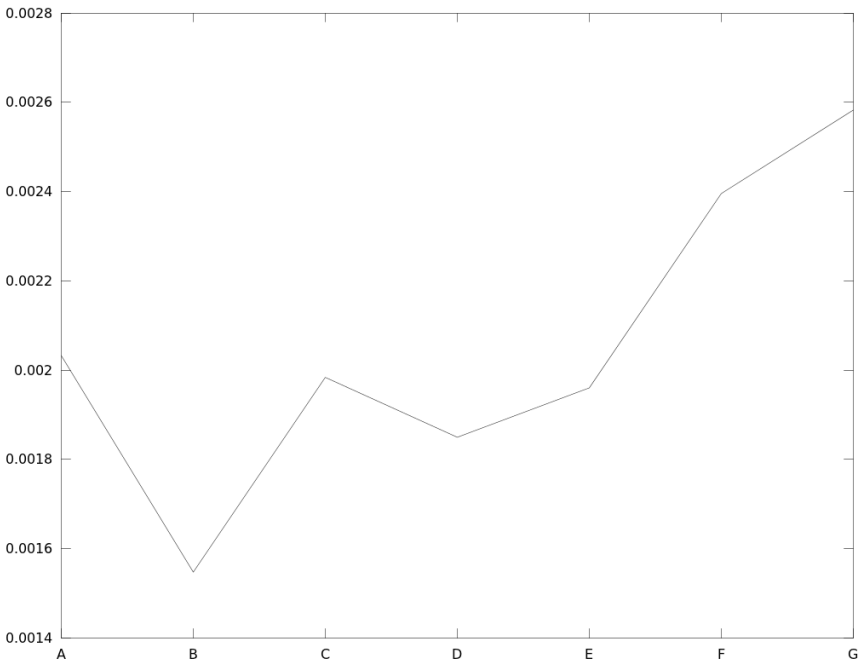


Figure 40. Correlation of blueberry NAIndex and total anthocyanin analysis results.

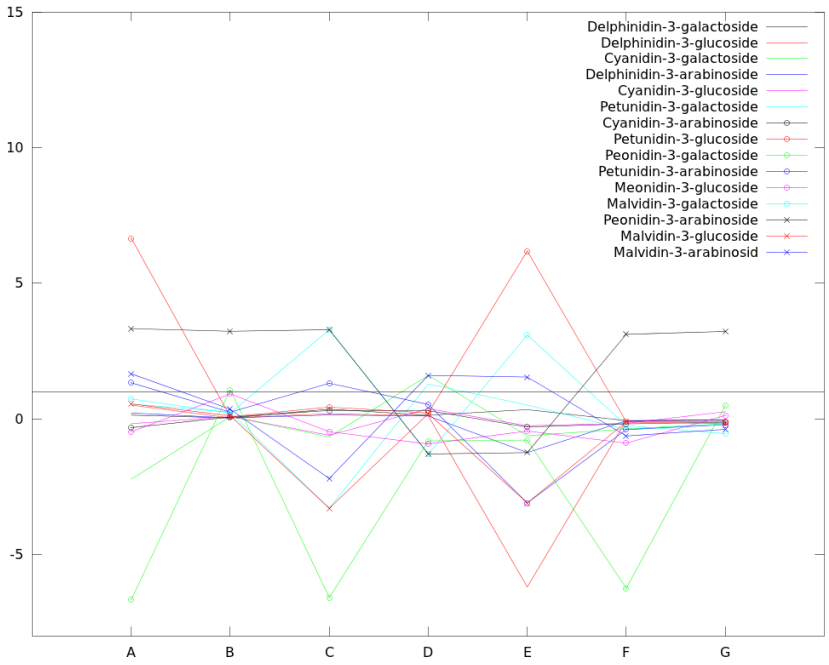


Figure 41. Blueberry NAIndex correlation to anthocyanin components. Correlation results are shown for each anthocyanin component obtained from th chemical analysis of blueberries.

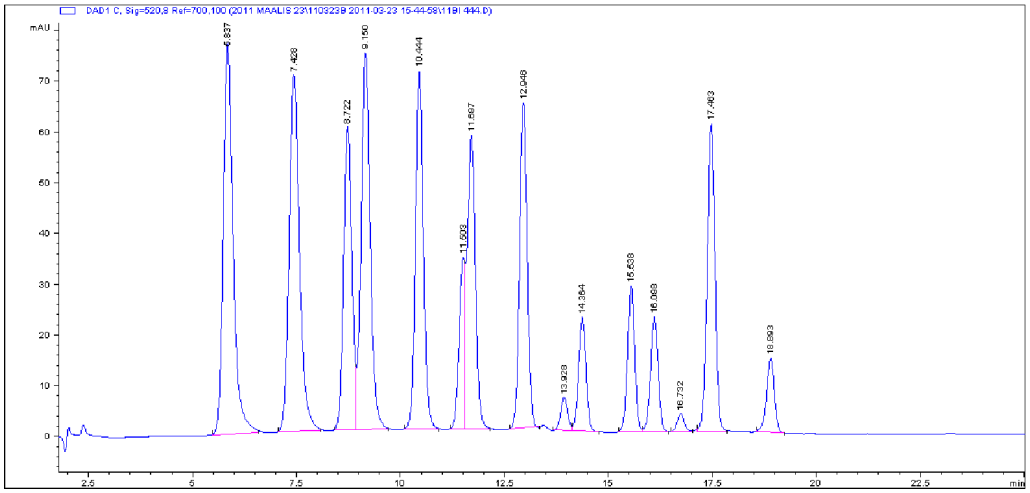


Figure 42. Cromatogram of blueberry samples. (Measured by Kajaani University Consortium, CEMIS-Oulu/Analytical chemistry and bioanalytics). Peaks in cromatogram identify and show the amount of unique anthocyanin components.

nm, which is outside of the range of our Ocean Optics HR4000 -spectrometer. In addition it seemed that there was no visible distortation in LT Industries ParaFuel -measurements either.

Fig. 43 shows the correlation between blueberry pH and the total anthocyanin content.

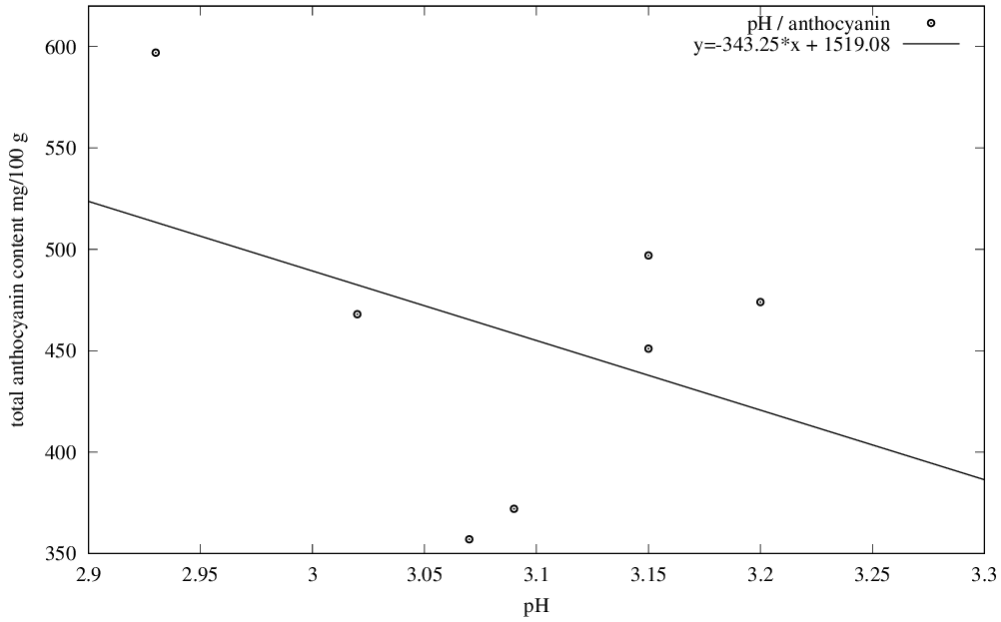


Figure 43. Blueberry pH and total anthocyanin correlation. When pH rises, the total anthocyanin decreases.

The higher the pH of blueberry sample was, the lower the total anthocyanin content was. There is, however, quite high spread in correlation graph. Propably pH is also connected to sugar or water content of blueberries. To eliminate this, sugar and water content of blueberries should have been measured too.

Based on previous analysis we can state that anthocyanin content does not reflect to these vegetative indices. This was expected, since vegetative indices focus mainly on relationship between red and green (or blue) colour. Since our blueberry samples were closely ripe they are not that different in red colour region. Therefore aerial photography, applying the studied vegetative indices, is not a good tool. Other solutions, such as NIR based indice, is required.

From the LDA analysis in Fig. 44 we can see that our measurements do not have a good grouping. If the first two primary components would explain the total anthocyanin content, it would be easy to see.

For PCA model building we apply the polynomial fitting to reduce measurement data size. The saturation of absolute polynomial fitting error is shown in Fig. 45. From this we can see that using more complex than 17 degree polynomial function for fitting does not result much improvement.

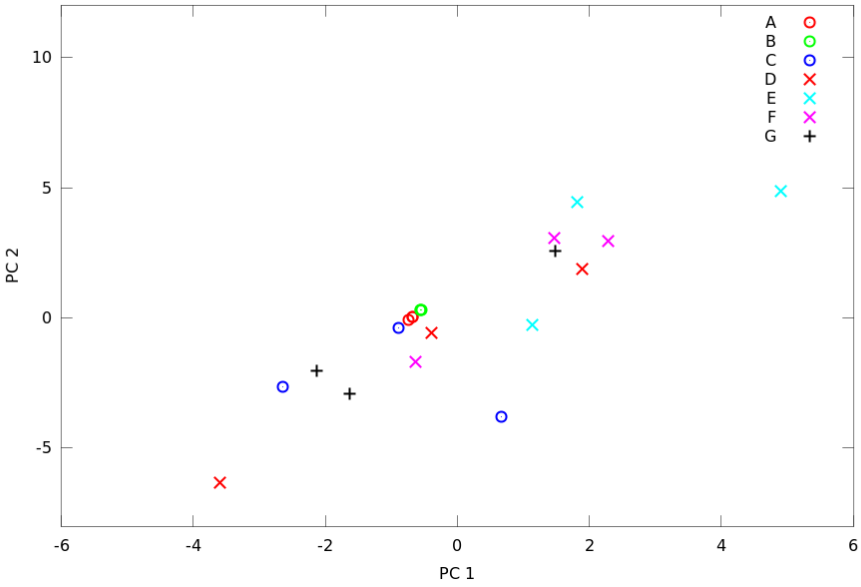


Figure 44. Blueberry LDA model. First two primary components have been shown.

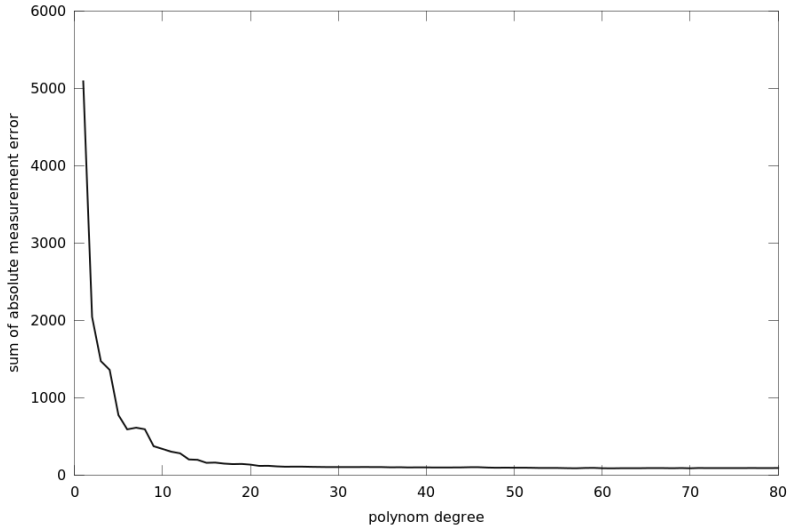


Figure 45. Blueberry polynomial fitting error saturation. Fitting error does not decrease much after the 17th degree of the polynom function.

The PCA model loadings of the first measurement series are given in Fig. 47. First primary component describes 72.7 % and the second 25.8 % of the data. Data is spread and no grouping can be seen in the PCA model result, which can be seen in Fig. 47.

The PCA model loadings for the second measurement series are in Fig. 49. First primary component describes 99.7 % and the second 0.24 % of the data. Again, the PCA model does not show grouping in Fig. 49.

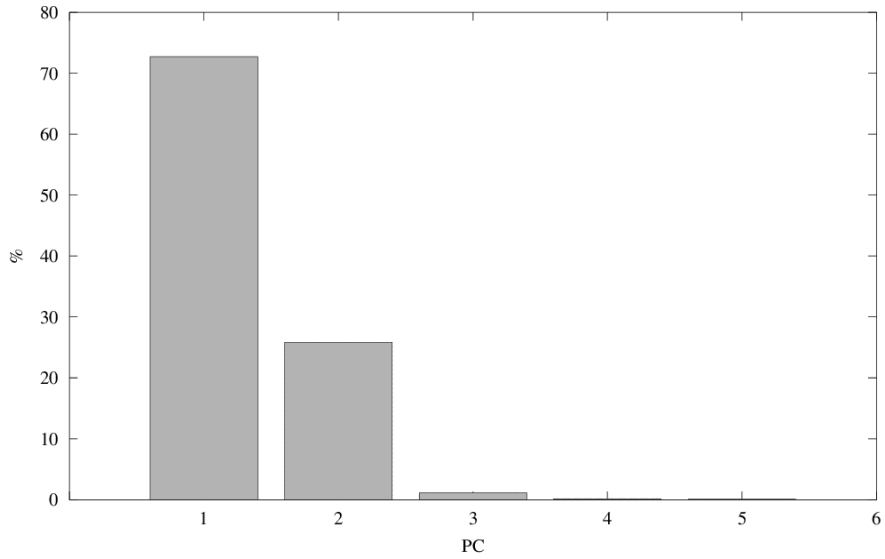


Figure 46. Blueberry PCA model loadings for the first measurement series.

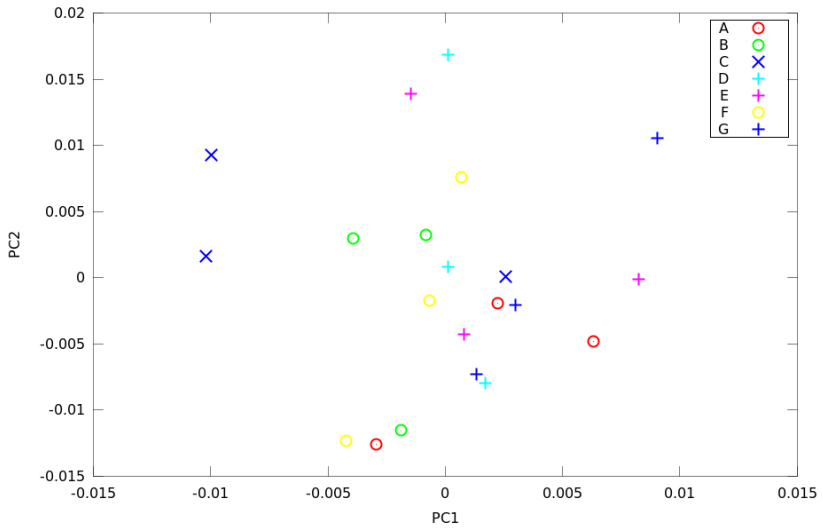


Figure 47. Blueberry PCA model PC1 and PC1 from the first measurement series.

For some reason, the PCA and LDA models do not show good groupings. Based on other research, there should be some grouping found. In order to test the validity of our measurements, we apply PLS model to the measurement data. Since in PLS we can apply the known anthocyanin content we should be able to find components which describe the data.

The measurement data from the LT Industries ParaFuel was used to create a PLS model. The prediction error for all the anthocyanins and for each measurement class are shown in Fig. 50. If the mark is on the diagonal, then the measured and model prediction equals.

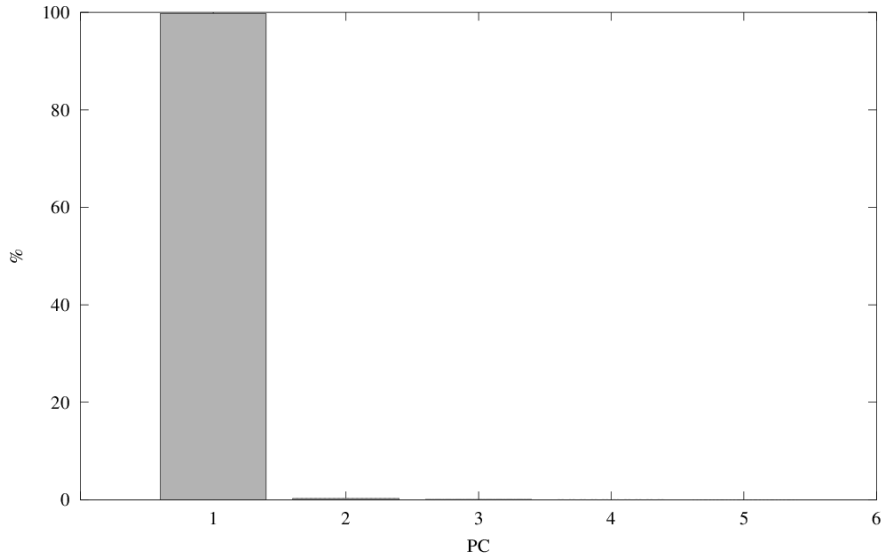


Figure 48. Blueberry PCA model loadings for the second measurement series. First component describes 99.7 % of the data.

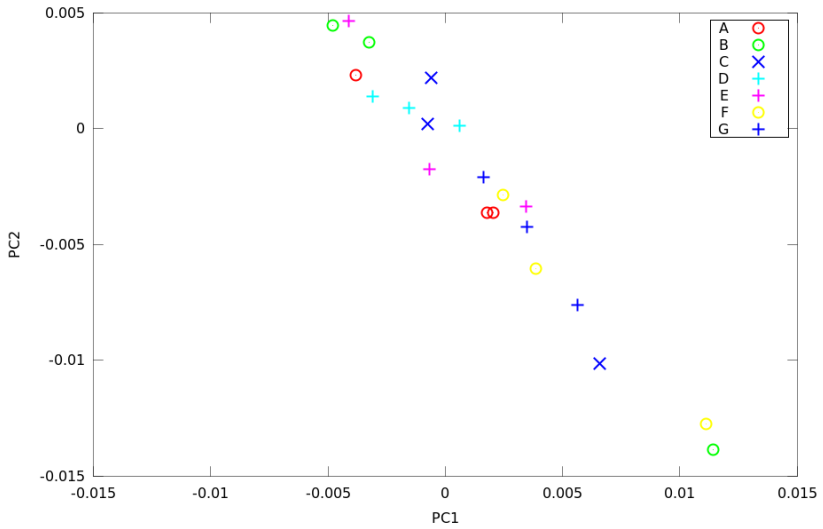
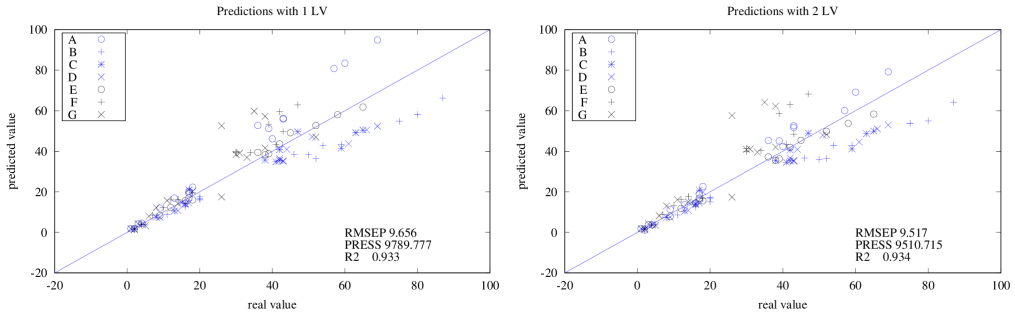


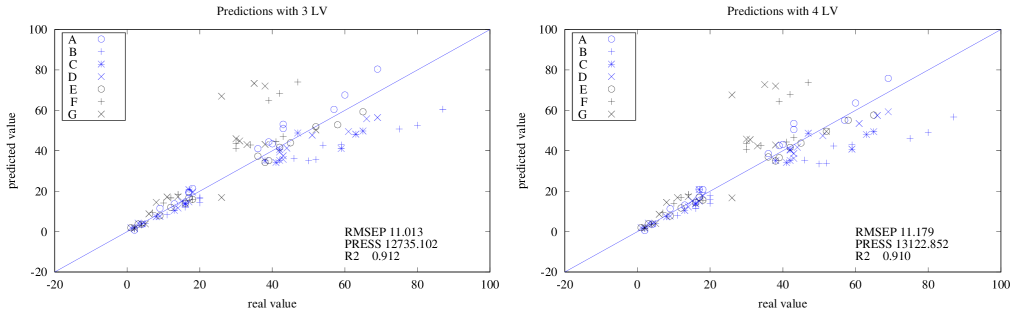
Figure 49. Blueberry PCA model PC1 and PC2 for second measurement series.

From this it is easy to see that the best option is to use two LV model, where we have RPMSEP = 9.517, PRESS = 9510.715, and R^2 = 0.934. The PRESS and RMSEP values actually grow when more LVs are used, therefore figures with five and six LVs is omitted. Because we have so few measurements, using more than six LV is not recommended, as in PLS the number of LVs should not exceed number of measurements.

When the measurement data from Ocean Optics HR4000 was used to build PLS model the best model is with three LVs, where we have RMSEP = 5.630, PRESS = 16643.513 and



(a) one LV (b) two LVs



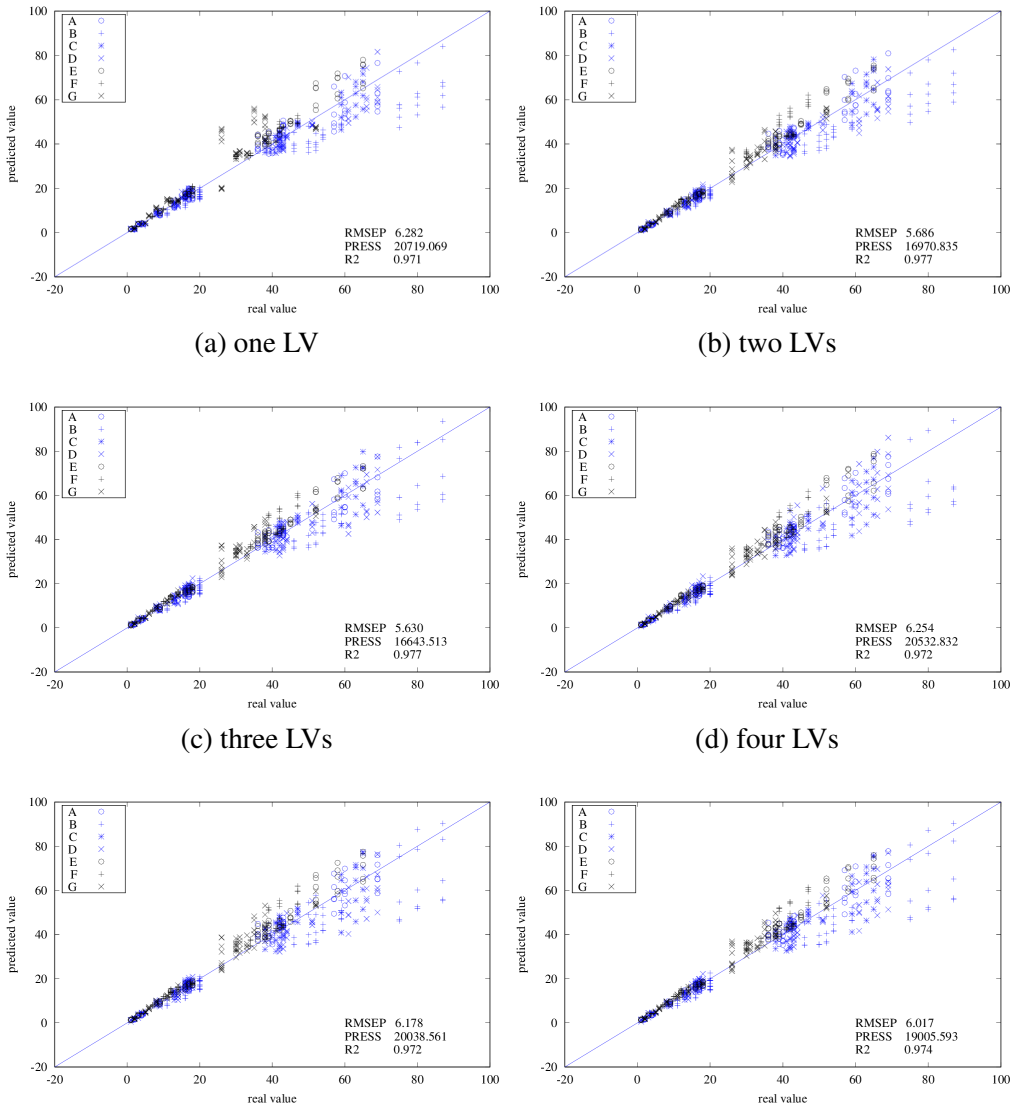
(c) three LVs (d) four LVs

Figure 50. Blueberry PLS model predictions for measurement data done with LT Industries ParaFuel. Each measurement is plotted against the predicted value. If values are far from the diagonal line, then the predicted value is not close to real value.

R^2 = 0.977. This can be seen from Fig. 50, where we have plotted comparison between the true value and model predicted value. Marks close to the diagonal mean equality. Again, adding more LVs after three to model does not help significantly. Both PLS model evaluations were made using the leave-one-out -method, where each measurement row is in turn used as testing record.

The two measurement datasets were used separately in PLS model building, because this gave better results than when the two wavelength ranges were combined. This is mostly due to the fact that there is a non-continous point between two appended measurements. This problem could perhaps be removed if proper scaling and centering of data would be applied.

Based on the PLS models applying NIR measurements to measuring anthocyanin content, we can say that it is atleast promising. As can be see in Fig. 52, some anthocyanins were better predicted with this model than others. Quite good results were obtained, which most likely will improve with larger sample sets. In order to improve measurements, also



(a) one LV (b) two LVs

(c) three LVs (d) four LVs

(e) five LVs (f) six LVs

Figure 51. Blueberry PLS model predictions for Ocean Optics HR4000 measurement data. Each measurement is plotted against the predicted value. If values are far from the diagonal line, then the predicted value is not close to the true value.

the sugar content, at least for the major sugars, should be analysed.

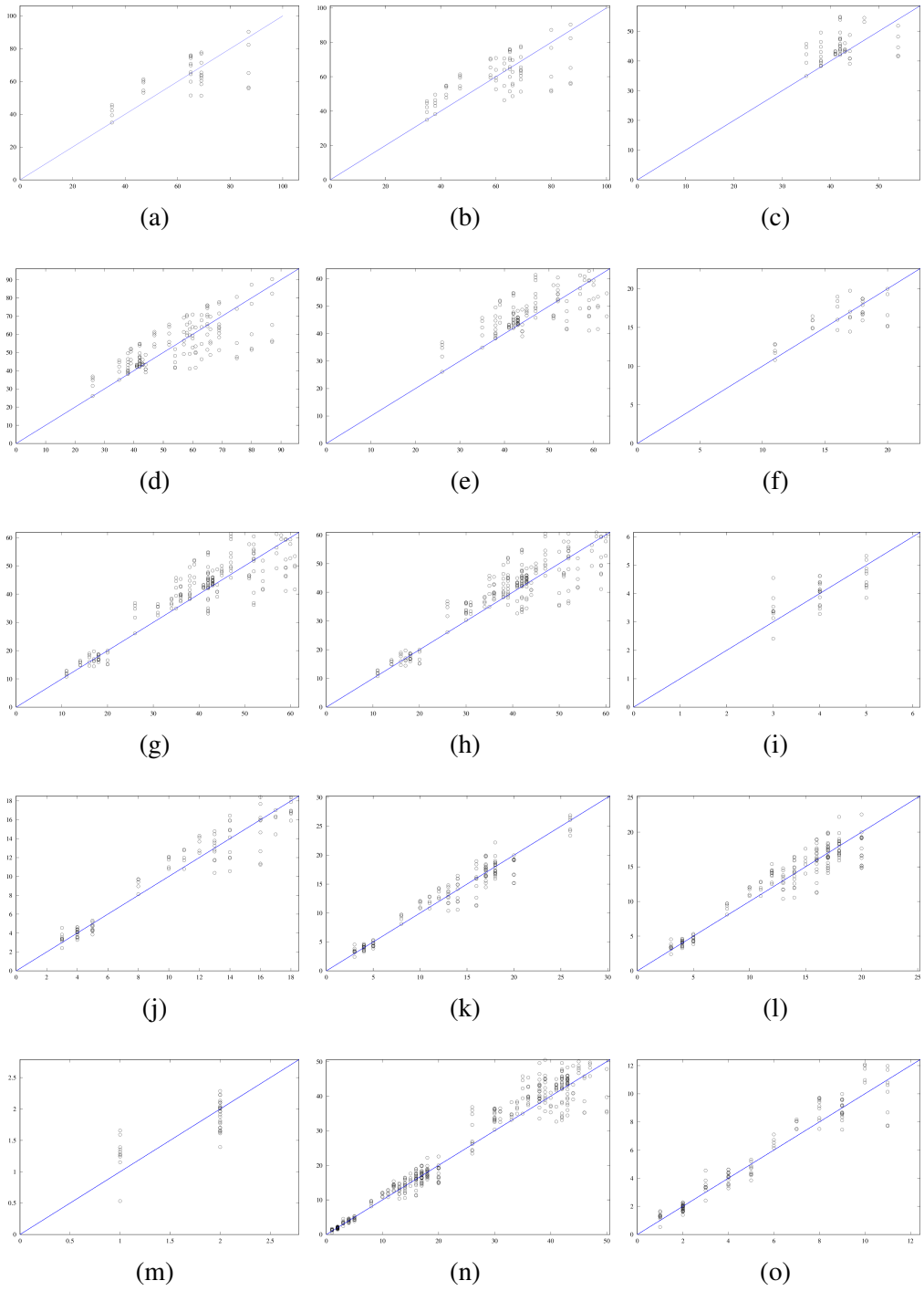
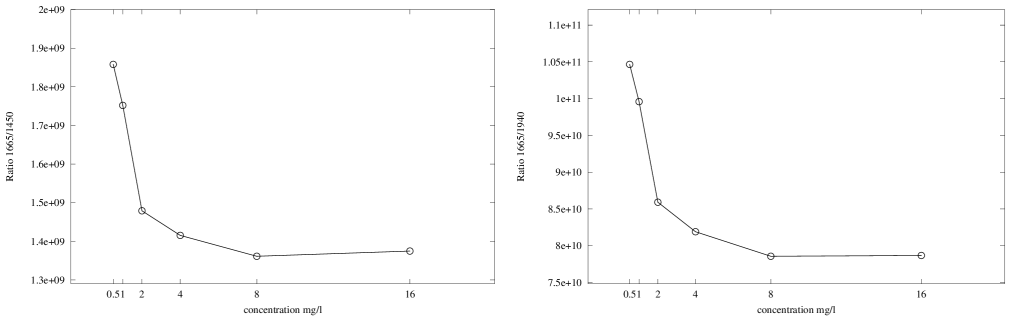


Figure 52. Anthocyanin predictions for PLS model from measurements done with Ocean Optics HR4000. Some anthocynanin predictions were better than others. a) Delphinidin-3-galactoside b) Delphinidin-3-glucoside c) Cyanidin-3-galactoside d) Delphinidin-3-arabinoside e) Cyanidin-3-glucoside f) Petunidin-3-galactoside g) Cyanidin-3-arabinoside h) Petunidin-3-glucoside i) Peonidin-3-galactoside j) Petunidin-3-arabinoside k) Meonidin-3-glucoside l) Malvidin-3-galactoside m) Peonidin-3-arabinoside n) Malvidin-3-glucoside o) Malvidin-3-arabinoside

4.3. Sodium bimetasulfide solution analysis

In Fig. 53 the $Na_2S_2O_5$ content correlation is viewed. The probe was under the solution surface. However, it is possible that some diffuse scattering occurs in water, therefore the ratio 1665 to 1450 nm (and 1940) were used to calibrate possible errors caused by water level differences. As shown previously the 1450 and 1940 nm are both used in NIR spectra analysis to determine water content in samples. The value 1665 nm was chosen based on the standard deviation divided with mean calculus, which is shown in Fig. 54

The results from determining the $Na_2S_2O_5$ content applying NIR seem promising. However, more measurements, with larger variety of aqueous solutions is needed. In order to



(a) Ratio 1665/1450 nm (b) Ratio 1665/1940 nm

Figure 53. The correlation of sodium bimetasulfide ($Na_2S_2O_5$) content and NIR absorbency. There seems to be logarithmic correlation in the absorption at these ratios.

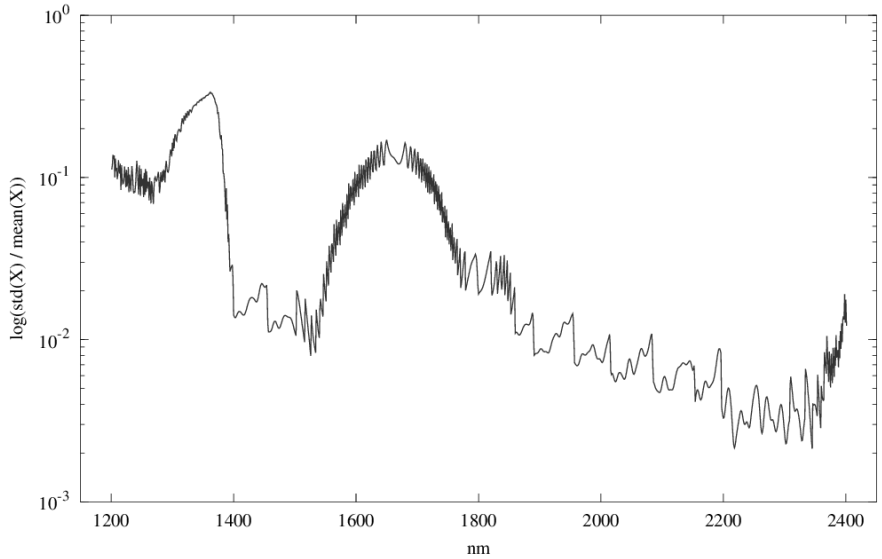
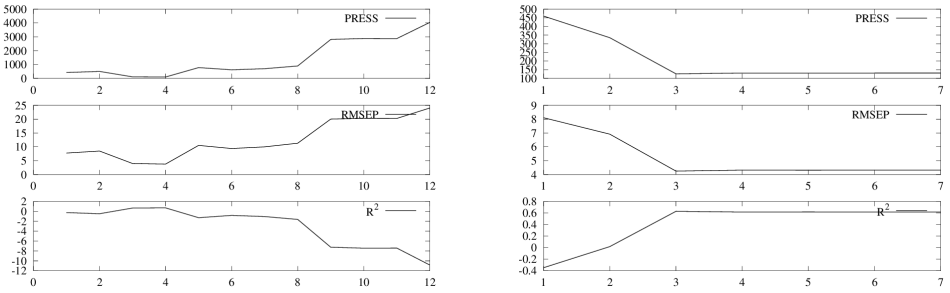


Figure 54. Standard deviation of $Na_2S_2O_5$ measurement divided by measurement means. High peaks are the absorption peaks.



(a) (b)

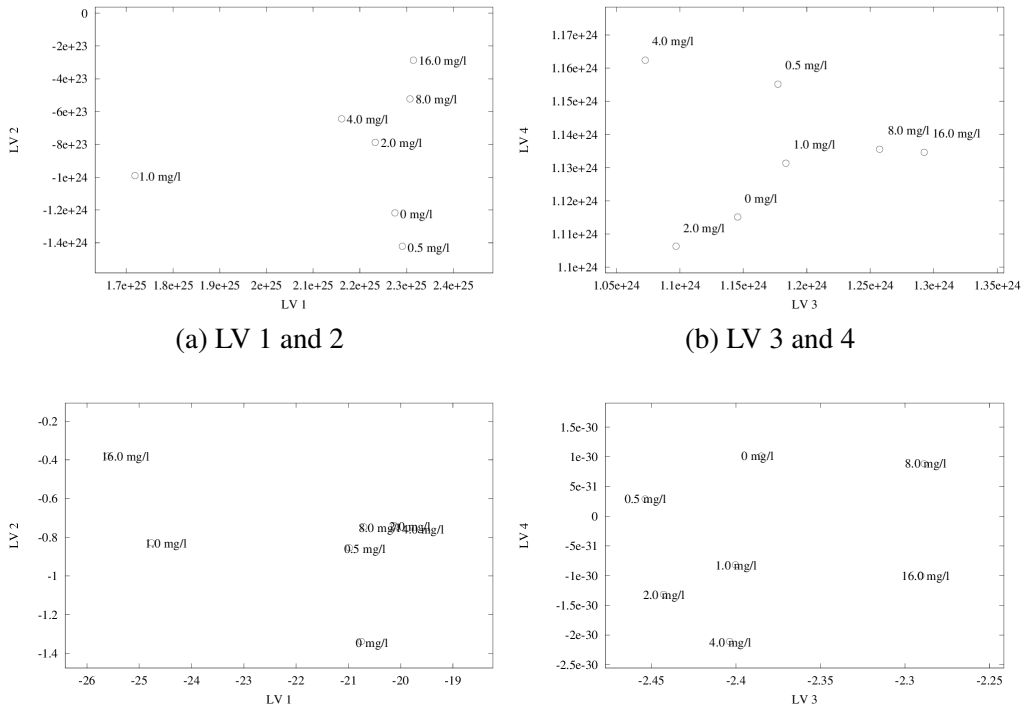
Figure 55. $Na_2S_2O_5$ PLS model LV selection. From this we can see that the model error has local minimum when four (a) and three (b) LV is used. Only LVs from one to seven were calculated. The PLS models were created from the measurement data of a) LT Industries ParaFuel, and b) Ocean Optics HR4000.

get closer to real situation, also some potato mash and dirt should be added to solutions to verify that the wavelengths found here do not get overruled by some other part.

As the PLS is adapting to measurements, the water absorption is not a problem, because each measurement should have equal absorption in the wavelengths where water absorption peaks are. However, the selection of the number of LVs can be a problem, when we have such small data set. To ease up the selection, the error estimates were calculated using LVs from one to ten. Results can be seen in Fig. 55, and from this we can safely select four LVs to be the best model for LT Industries ParaFuel data and three for Ocean Optics HR4000 data. The local minimum of RMSEP and PRESS is most likely dependent on small number of measurements.

After building the PLS model for LT Industries ParaFuel $Na_2S_2O_5$ measurement data, we can see that the second LV describes the concentration of $Na_2S_2O_5$ almost perfectly. It looks as if the measurements of pure water and 0.5 mg/l measurements would have been replaced. The same phenomena was shown, but not mentioned earlier (in Fig. 53) when the ratio at 1665/1450 nm was presented.

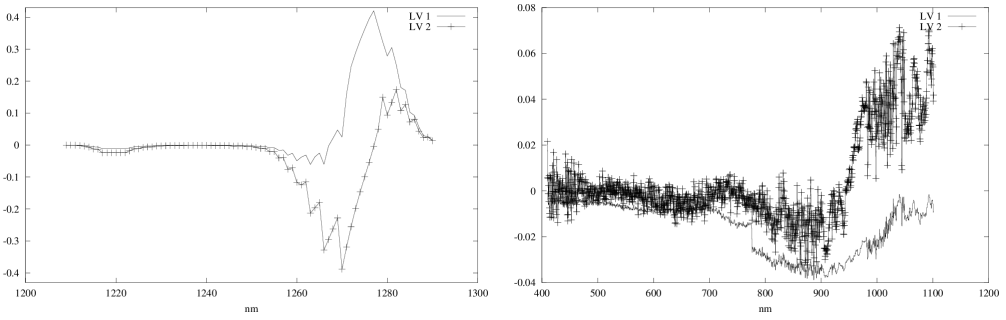
In Fig. 57 the weight of LV is plotted against the measurement wavelength. It can be seen that in the first LV the absorption at 1270 nm has big influence, shown as negative peak in the Figure. This supports the previous finding, shown in Fig. 54, that this absorption peak correlates with $Na_2S_2O_5$ content. However, interesting is that in PLS model the peak wavelengths 1665 and 1940 nm do not have any significance. Only the range from 1210



(a) LV 1 and 2 (b) LV 3 and 4

(c) LV 1 and 2 (d) LV 3 and 4

Figure 56. PLS model LVs of $Na_2S_2O_5$ measurements. The second LV describes almost correctly the concentration of aqueos solution. a,b) LT Industries ParaFuel, c,d) Ocean Optics HR4000.

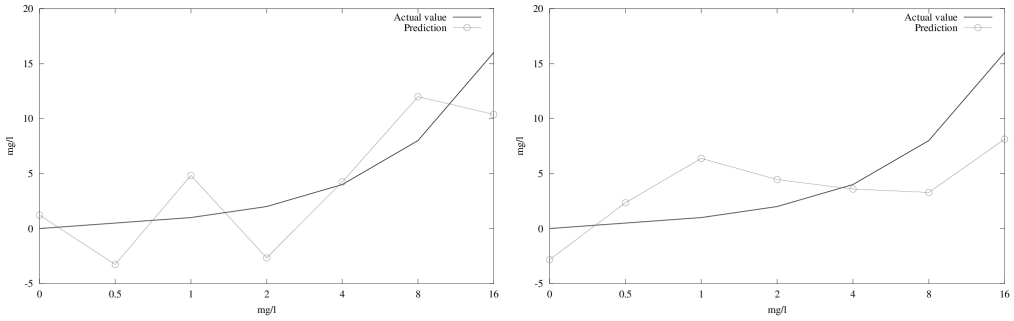


(a) (b)

Figure 57. $Na_2S_2O_5$ PLS model LVs 1 and 2 weights on wavelength from a) LT Industries ParaFuel, b) Ocean Optics HR4000 data.

to 1290 nm is shown, since outside that it is approximately constant.

Because of the low number of measurements, the PLS model was verified by using leave-one-out -method. It is meant to be used in situtations when only a low number of measurements is available. Fig. 58 shows the predicted values from PLS model. The model error estimates are in LT Industries ParaFuel model PRESS = 99.642, RMSEP = 3.7729 and



(a) (b)

Figure 58. $Na_2S_2O_5$ PLS model of LT Industries ParaFuel data predicted values compared to the measurements. The two curves should overlap for perfect prediction. a) LT Industries ParaFuel b) Ocean Optics HR4000

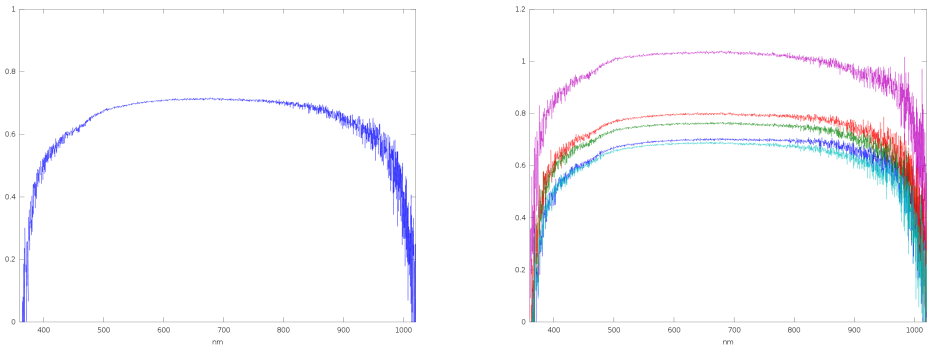
R^2 = 0.70801 and in Ocean Optics HR4000 model PRESS = 126.28, RMSEP = 4.2474, and R^2 = 0.62994.

Based on the error estimates, we can say that the LT Industries ParaFuel is the better choice, when the concentration of aqueous $Na_2S_2O_5$ solution is measured. The wavelength range from 1200 to 1300 nm seems to contain enough information for PLS model to predict the concentration. This is illustrated in Fig. 58, where the true and predicted values are compared. With a larger sample size, the model would most likely to be more accurate.

4.4. Analysis of the Aqueous Medicine Solution

NIR absorption of medicine in powder form can be seen in Fig. 59(a). As can be observed, it is quite flat in the range from 400 nm to 1000 nm. The shape of spectra at the extreme ends of the spectrum is not fully reliable because of the light source used. In Fig. 59(b) are given the absorption spectra of powder from different positions—this means that the powder sample was moved in x- and y- directions. The change of intensities is explained by the fact that the surface of powder was not totally flat. As it was placed into test tube it was shaked for leveling. The important thing is that the shape of spectra, as can be seen, is similar in different locations. Therefore the reflectance of the powder is close to uniform.

In the measurement of distilled water and aqueos medicine absorption can be seen, in Fig. 60, that the aqueos medicine absorbs light less than water in the range from 390 nm



(a) single measurement

(b) several measurements from different positions

Figure 59. Absorption spectra of morphine in powder form. In (b) the effect of changing the measurement position can be seen. As can be seen, the shape of powder surface affects on the reflectance. For good measurement results control of surface shape is required.

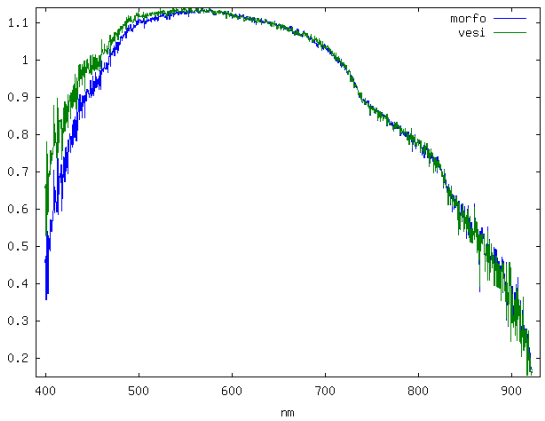


Figure 60. Comparison of distilled water and aqueos medicine

to 550 nm. Because the difference was very small, which can be seen from curve shapes Fig. 60, between water and aqueos mecicine, a suspicion was raised that perhaps the filtering process is too efficient—it may filter also some medicine away. Therefore to the next measurement an unfiltered aqueos medicine was prepared, with equal concentration, was added. Drop of absorption, as compared to powder form morphine, is visible from 600 nm in both liquids.

The measurement results of the second measurement series are shown in Fig. 61. Here we have added also the unfiltered liquid. The reader should not pay attention to the fact that some of the measurement results are negative, the relevant information is in the intensity changes. The negative numbers are due to difficulties in setting white reference—aqueos

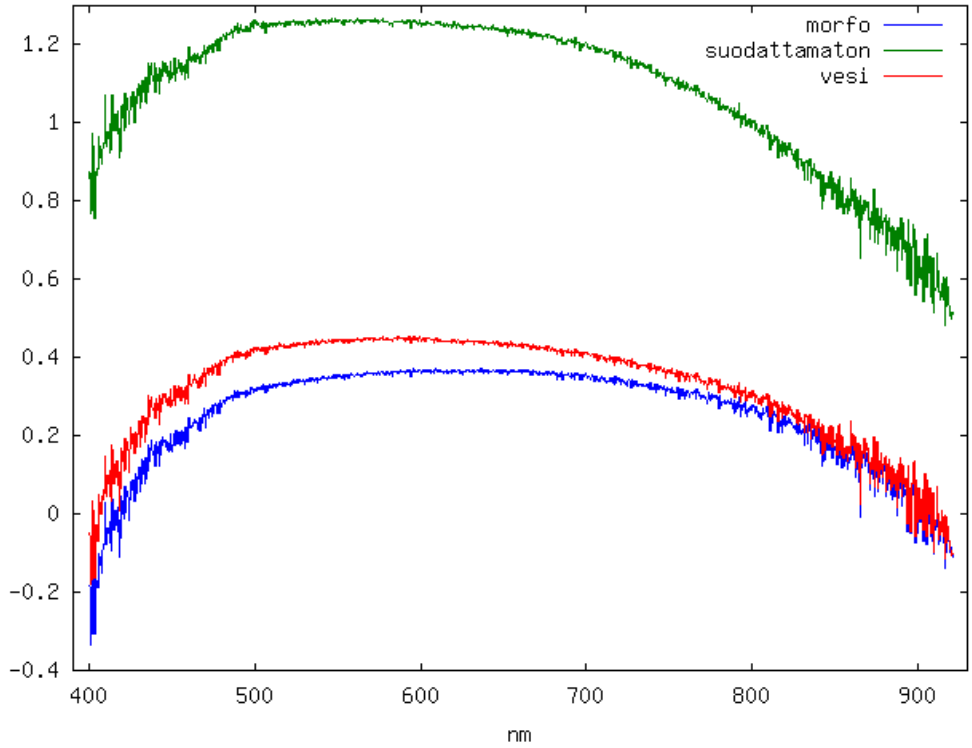


Figure 61. Comparison of water, filtered aqueos and unfiltered aqueos medicine.

medicine has very high reflectance and it has been measured in a test tube over a glass mirror.

In Fig. 61 the absorption levels of filtered and non-filtered solution are different. This is mostly caused by the fact that in the unfiltered liquid there are propably a lot of medicine particles, clusters of molecules, which are not dissolved into water but remain in crystal form. The scattering of the unfiltered liquid and the powder are quite close to each other.

In the third measurement series a new liquid solution was added. To the filtered aqueos medicine was added powder (and mixed properly) but no post-filtering was done. The measurement of these (four solutions) can be seen in Fig. 62. The shapes of the curves are in Fig. 62. Again the unfiltered solution had the highest transmission. The filtered liquid with added medicine was clearly raising above normal filtered liquid and it's shape was closer to water's spectral shape than to the shape of aqueous medicine spectra.

In Fig. 62(a) a comparison between series one to three is shown. The graphs are corrected by subtracting the water measurement from each aqueous medicine measurement. As waters transmission changes only due to the amount of the liquid the transmission of

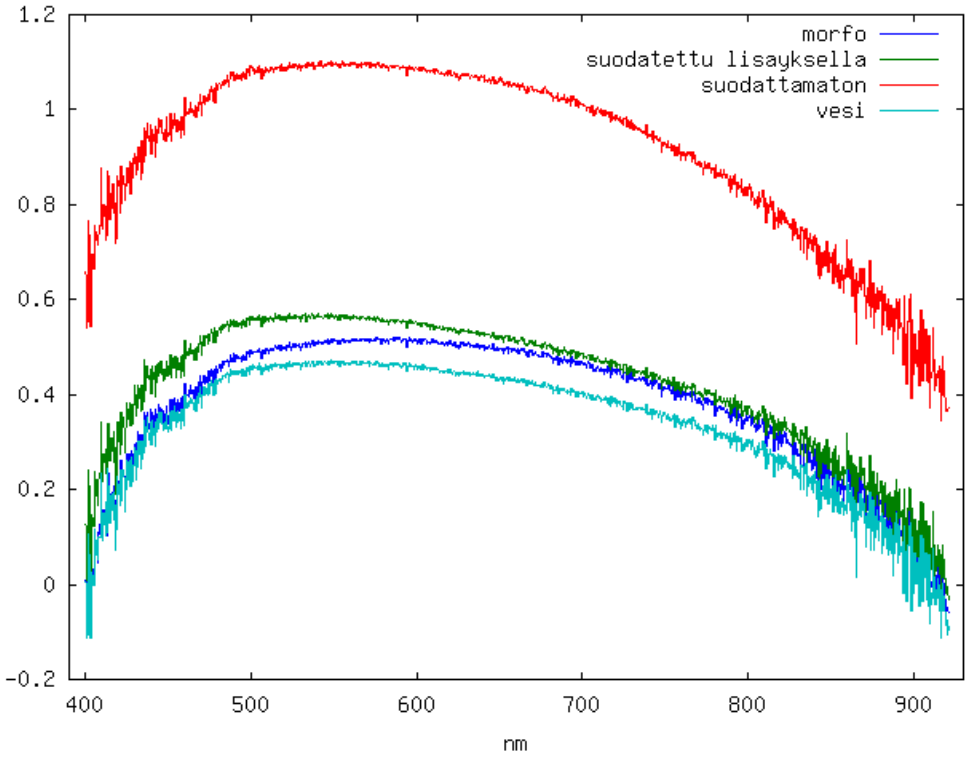
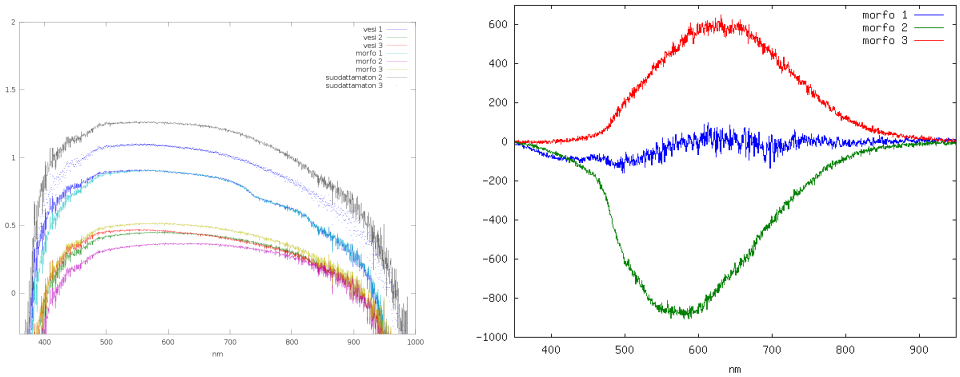


Figure 62. Results from medicine measurement series number three. Again we can see the large change in the absorption level, most likely due by medicine crystals.



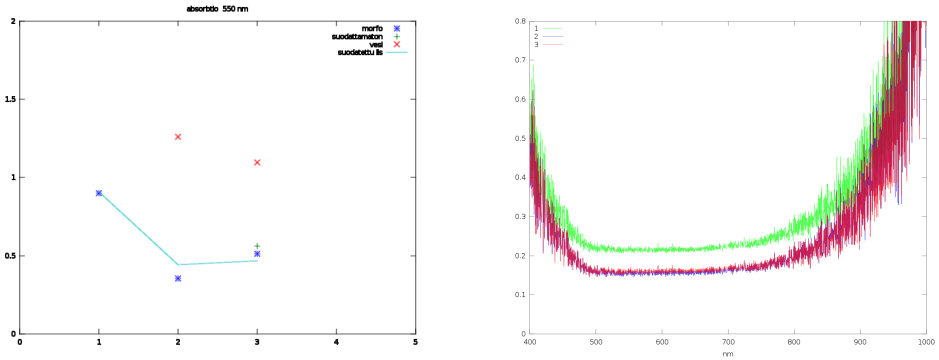
(a) aqueous medicine series 1-3 and unfiltered aqueous medicine with added medicine.

(b) different aqueous medicines with water correction.

Figure 63. Comparison of all medicine measurements. In (a) can be seen the similar shape for all spectra. The effect of water correction can be seen in (b).

aqueous medicine change also because of the change in concentration.

As can be seen in Fig. 63(b) the intensities of measurements are not linear. During the process of mixing the medicine, the concentration should drop from measurement one towards measurement three. Measurement one does not fully comply to this.



(a) absorption in 550 nm

(b) absorption of second and third aqueous medicine in range 400-1000 nm

Figure 64. Aqueos medicine absorption at 550 nm. The absorption seems to be somewhat linearly correlated to concentration.

In Fig. 64(a) we have compared the transmission at 550 nm for all liquids. As can be seen, the first measurement results are clearly too high. Even though in this figure the integration time has been compensated, the amount is too high. This is caused by the fact that there were much more liquid in the first samples than in the second and third.

On the other hand, the transmission results for the second and third measurement, in Fig. 64(a), are logical and suggest that there exists a linear dependency between the intensity and concentration of aqueos medicine at 550 nm.

In Fig. 64(b) is given a comparison between the second and third measurements of unfiltered aqueous medicine. The shape and intensity changes are almost identical.

As a visual observation we found out that the surface tension of liquid is highly affected by the concentration of medicine. This can quite easily be seen in Fig. 65. It seems that the higher the concentration of the aqueous medicine, the higher the surface tension.

The PCA model does not help us much because the number of comparable measurements was too low. In Fig. 66 it can be seen the model loadings that the first component dominates and explains 96.53 % of data while the second component only 3.23 %. The primary components of the PCA model can be seen in Fig. 67. The measurements ’m1’ and ’m2’ should be very close to each other, since they are measurements from the same solution. Similarly ’m3’ and ’m4’ should be close to each other and ’m5’, ’m6’, and ’m7’ should be close to each other.

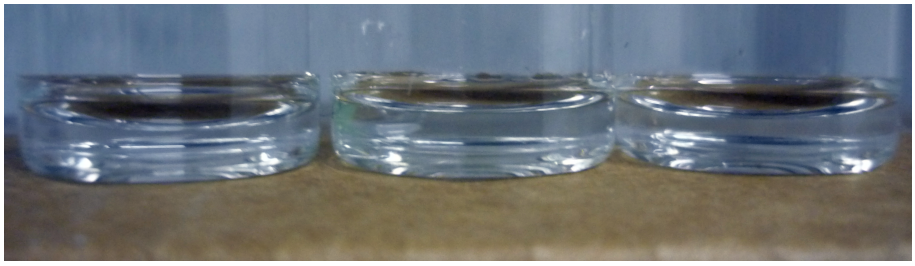


Figure 65. Surface tension of aqueous medicine has relatively big vertical effect in the test tube, due to the small diameter of the tube. From left to right, water, filtered aqueous medicine and unfiltered aqueous medicine. The concentration of aqueous medicine was equal before filtering process.

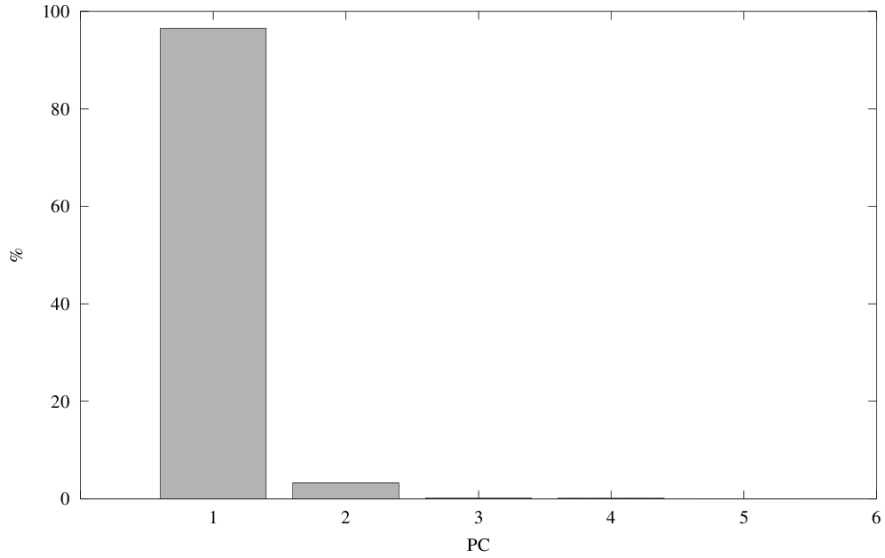
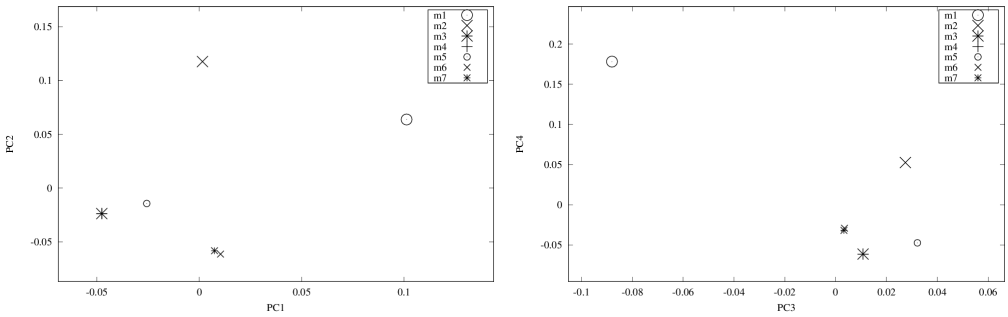


Figure 66. Aqueous medicine solution PCA model loadings.

Because of the small number of comparable measurements it makes no sense to do error estimation. The measurement of the aqueous solutions is difficult, and the poor PCA model shows this clearly.

Based on the experiments we can say that the filtering process is efficient. It will remove most of the undissolved medicine from the aqueous solutions. Even though this was not the goal of filtering.

NIR measurement is clearly sensitive to the crystallisation of the medicine within solution and it seems that in some cases we were measuring the amount of crystals in the solution. However, if the filtering process is controlled, then it seems to be possible to apply NIR for concentration measurement. Also the amount of the liquids needs to be controlled more carefully e.g. by a suitable probe design.



(a) PC1 and PC2 (b) PC3 and PC4

Figure 67. Aqueous medicine solution PCA model primary components.

The results clearly indicate that more exact measurements are needed, if regression model or linear multivariate model for concentration measuring is desired. These measurements were not controlled enough to allow building of a reliable model.

5. CONCLUSIONS

In Table 1 we listed some food quality attributes from previous research. Based to literature we have shown that with NIRS we can detect some defects, address colouring problems with imaging system, monitor sugar and water content and acidity which are partly responsible for ripeness and taste. Condition and firmness can be monitored by measuring the water content and the reflectance.

We have verified results from previous research, that with a PCA model of NIRS measurements can be used for tomato classifying. The colour, water content and sugar are detectable by NIR, as shown e.g. by Giangiacomo (2006); Lai et al. (2007); McCaig (2002); Polder et al. (2004). Data size reduction helps analysis of the data, and if done properly, does not too much interfere with actual data.

Intensity of the reflectance was not used in the analysis of the tomatoes, but our results indicate that changes in the reflectance intensity is one of the key variables describing the ripening of a tomato. Intensity change occurs on wide spectral range, also on NIR spectra.

Our results showed also that carotenoids chlorophyll ratio (525/672) is an other easy detectable key variable, describing the changes happening during ripening process of tomatoes. Our results also support the fact that with NIR, it is possible to detect also internal changes in tomatoes, such as change in sugar content, as side product from ripening. Full sugar content and chemical analysis of tomatoes should be carried out so that the PCA model can fully be utilised.

We have also shown that PLS models built from NIR measurement data can be used in predicting anthocyanin content in blueberries and the results are quite promising. With larger testing samples where we have even leaves and other foreign material we can build a quite good model, which can be used to predict the anthocyanin contents.

We found out that applying NIRS to measuring the concentration of aqueous natrium bimetasulfide solution might be possible. The best results were obtained from the wavelength range of 1200 to 1300 nm, where our LT Industries ParaFuel spectrometer works. No reliable results were obtained from the range 400 to 1100 nm. However, more measurements are needed to verify that the good results also apply, when the solution is not

“laboratory grade”, in this case the solution was created from ion balanced water and pure powder, but also contains dirt, potato starch and other foreign substances diluted from potatoes.

The measurements of aqueous medicine didn’t succeed. Due to time limitations and lack of better measurement probes, at the time when the measurements were executed, the results were not reliable. Perhaps better solution would be to use a special probe which is designed to liquid measurements or the measurements should be done through the walls of the container. That would raise new problems, like high reflectivity of glass used, but it would be more controllable.

If we are willing to further develop the NIRS measurement setup, our measurements also fulfil the four criterias, which we set in Chapter 1: measurements can be feasible (economically), they can produce repeatable and comparable results, actual measurements are not destructive and they can be developed to be scalable.

More research work is required before the methods explained in this thesis can fully be applied as tools in food quality control work. We need to do more measurements with larger sample size so that the results are really statistically reliable. Measurement setup requires more thought in order to eliminate error sources in measurements; ie. measurements need to be simpler to execute and easier to calibrate, therefore more automation is needed.

One of our contributions is now the knowledge that some methods seem to work while others don’t and we have detected certain problem areas, which should be focused when planning of the measurements. And because we know that PCA and PLS work in many of these problems, we can plan the measurement setup so, that we can fully utilise it.

BIBLIOGRAPHY

Abbott, Judith A. (1999). Quality measurement of fruits and vegetables. *Postharvest Biology and Technology* 15, 207–225.

Adams, ML, WD Philpot & WA Norvell (1999). Yellowness index: An application of spectral second derivatives to estimate chlorosis of leaves in stressed vegetation. *International Journal of Remote Sensing* 20, 3663–3675.

Alander, Jarmo T., Vladimir Bochko, Birgitta Marginkauppi, Sirinnapa Saranwong & Timo Mantere (2012). *A review of optical non-destructive visual and near-infrared methods for food quality and safety*. Submitted to Journal of Spectroscopy.

Artés, F., M.A. Conesa, S. Hernández & M.I. Gil (1999). Keeping quality of fresh-cut tomato. *Postharvest Biology and Technology* 17, 153–162.

Arvanitoyannis, I.S., M.N. Katsota, E.P. Psarra, E.H. Soufleros & S. Kallithraka (1999). Application of quality control methods for accessing wine authenticity: Use of multivariate analysis (chemometrics). *Trends in Food Science & Technology* 10, 321–336.

Auerswald, Helga, Peter Peters, Bernhard Brückner & Angelika Krumbein (1999). Sensory analysis and instrumental measurements of short-term stored tomatoes (*Lycopersicon esculentum* Mill.). *Postharvest Biology and Technology* 15, 323–334.

Barth, M. Margaret & Hong Zhuang (1996). Packaging design affects antioxidant vitamin retention and quality of broccoli florets during postharvest storage. *Postharvest Biology and Technolgy* 9, 141–150.

Bauchot, Anne D., F. Roger Harker & W. Michael Arnold (2000). The use of electrical impedance spectroscopy to assess the physiological condition of kiwifruit. *Postharvest Biology and Technology* 18, 9–18.

Beaurdy, R.M. (1999). Effect of O_2 and CO_2 partial pressure on selected phenomena affecting fruit and vegetable quality. *Postharvest Biology and Technology* 15, 293–303.

Belie, N. De, S. Schotte, P. Coucke & J. De Baerdemaeker (2000a). Development of an automated monitoring device to quantify changes in firmness of apples during storage. *Postharvest Biology and Technology* 18, 1–8.

Belie, N. De, V. De Smedt & J. De Baerdemaeker (2000b). Principal component analysis of chewing sounds to detect differences in apple crispness. *Postharvest Biology and Technology* 18, 109–119.

Bellon-Maurel, Véronique, Elvira Fernandez-Ahumada, Bernard Palagos, Jean-Michel Roger & Alex McBratney (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry* doi:10.1016/j.trac.2010.05.006.

Blackburn, GA (1998). Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *International Journal of Remote Sensing* 19, 657–675.

Blakey, Robert J., John P. Bower & Isa Bertling (2009). Influence of water and ABA supply on the ripening pattern of avocado (*Persea americana* Mill.) fruit and the prediction of water content using near infrared spectroscopy. *Postharvest Biology and Technology* 53:1-2, 72 – 76. ISSN 0925-5214. doi:10.1016/j.postharvbio.2009.03.004.

Bobelyn, Els, Anca-Sabina Serban, Mihai Nicu, Jeroen Lammertyn, Bart Nicolai & Wouter Saeys (2009). Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance. *Postharvest Biology and Technology* doi:10.1016/j.postharvbio.2009.09.006.

Brackett, Robert E. (1999). Incidence, conributing factors, and control of bacterial pathogens in produce. *Postharvest Biology and Technology* 15, 305–311.

Brunsø, Karen, Thomas Ahle Fjord & Klaus G. Grunert (2002). *Consumers' food choice and quality perception*. MAPP Working Papers 77, University of Aarhus, Aarhus School of Business, The MAPP Centre. <URL:http://ideas.repec.org/p/hhb/aarmap/0077.html>.

Burks, C. S., F. E. Dowell & F. Xie (2000). Measuring fig quality using near-infrared spectroscopy. *Journal of Stored Products Research* 36:3, 289 – 296. doi:10.1016/S0022-474X(99)00050-8.

Büning-Pfaue, Hans (2003). Analysis of water in food by near infrared spectroscopy. *Food Chemistry* 82:1, 107 – 115, 2nd International Workshop on Water in Foods. ISSN 0308-8146. doi:10.1016/S0308-8146(02)00583-6.

Casale, Monica, Paola Zunin, Maria E. Cosulich, Erika Pistarino, Patrizia Perego & Silvia Lanteri (2010). Characterisation of table olive cultivar by NIR spectroscopy. *Food Chemistry* doi:10.1016/j.foodchem.2010.03.080.

Cen, Haiyan & Yong He (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology* 18:2, 72 – 83. doi:10.1016/j.tifs.2006.09.003.

Chang, Min, Peng jiao Chu & Ke xin Xu (2007). Study on noninvasive detection using NIR diffuse reflectance spectrum for monitoring protein content in milk powder. *Guang Pu Xue Yu Guang Pu Fen Xi* 27:1, 43–45.

Cho, Byoung-Kwan, Yud-Ren Chen & Moon S. Kim (2007). Multispectral detection of organic residues on poultry processing plant equipment based on hyperspectral reflectance imaging technique. *Computers and Electronics in Agriculture* 57:2, 177–189. doi:10.1016/j.compag.2007.03.008.

Clinton, S. K. (1998). Lycopene: chemistry, biology, and implications for human health and disease. *Nutrition Reviews* 56:2, 35–51.

Clément, Alain, Martine Dorais & Marcia Vernon (2008a). Multivariate approach to the measurement of tomato maturity and gustatory attributes and their rapid assessment by Vis-NIR spectroscopy. *Journal of Agriculture and Food Chemistry* 56:5, 1538–1544. doi:10.1021/jf072182n.

Clément, Alain, Martine Dorais & Marcia Vernon (2008b). Nondestructive measurement of fresh tomato lycopene content and other physicochemical characteristics

using visible-NIR spectroscopy. *Journal of Agricultural and Food Chemistry* 56:21, 9813–9818. doi:10.1021/jf801299r.

Crisosto, Carlos H., Gayle M. Crisosto & Mark A. Ritenour (2002). Testing the reliability of skin color as an indicator of quality for early season 'Brooks' (*Prunus avium* L.) cherry. *Postharvest Biology and Technology* 24, 147–154.

Daughtry, C. S. T., C. L. Walthall, M. S. Kim, E. Brown de Colstoun & J. E. McMurtrey (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment* 74:2, 229 – 239. doi: 10.1016/S0034-4257(00)00113-9.

DeEll, Jennifer R., Robert K. Prange & Dennis P. Murr (1996). Chlorphyll fluorescence of Delicious apples at harvest as a potential predictor of superficial scald development during storage. *Postharvest Biology and Technology* 9, 1–6.

DeEll, J.R., R.K. Prange & D.P. Murr (1995). Chlorophyll fluorescence as a potential indicator of controlled-atmosphere disorders in 'Marshall' McIntosh apples. *HortScience* 30:5, 1084–1085.

EEC 1677/88 (1988). *Commission Regulation (EEC) No 1677/88 of 15 June 1988 laying down quality standards for cucumbers*. Location: `<URL:http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=EN&numdoc=31988R1677&model=guichett>` [cited 9th April 2011].

Ferguson, Ian, Richard Volz & Allan Woolf (1999). Preharvest factors affecting physiological disorders of fruit. *Postharvest Biology and Technology* 15, 255–262.

Fu, Xia-Ping, Yi-Bin Ying, Ying Zhou, Li-Juan Xie & Hui-Rong Xu (2008). Application of NIR spectroscopy for firmness evaluation of peaches. *J Zhejiang Univ Sci B* 9:7, 552–557. doi:10.1631/jzus.B0720018.

Gao, Bo-cai (1996). NDWI–A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58:3, 257 – 266. doi:10.1016/S0034-4257(96)00067-3.

Giangiacomo, R. (2006). Study of water-sugar interactions at increasing sugar concentration by NIR spectroscopy. *Food Chemistry* 96:3, 371 – 379, 3rd International Workshop on Water in Foods. ISSN 0308-8146. doi:10.1016/j.foodchem.2005.02.051.

Giordano, Frank R., Maurice D. Weir & William P. Fox (2003). *A First Course in Mathematical Modeling*. 3rd edition. 511 Forest Lodge Road, Pacific Grove, CA 93950, USA: Thompson Brooks/Cole.

Gitelson, AA, C Buschmann & HK Lichtenthaler (1999). The chlorophyll fluorescence ratio F_{735}/F_{700} as an accurate measure of the chlorophyll content in plants. *Remote Sensing of Environment* 69, 296–302.

Givens, D. I., J. L. De Boever & E. R. Deaville (1997). The principles, practices and some future applications of near infrared spectroscopy for predicting the nutritive value of foods for animals and humans. *Nutrition Research Reviews* 10:1, 83–114. doi:10.1079/NRR19970006.

Grevesmuehl, B., C. Kradjel & H. Kellner (1991). Designing optimized industrial process analysers for closed loop control. *Journal of Automatic Chemistry* 13:4, 139–142. doi:10.1155/S1463924691000263.

Haboudanea, Driss, John R. Millera, Elizabeth Patteyc, Pablo J. Zarco-Tejadad & Ian B. Strachan (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* 90, 337–352.

Hecht, Eugene (2002). *Optics*. Fourth international edition edition. Addison-Wesley.

Hertog, M. G. L., P. C. H. Hollman & M. B. Katan (1992). Content of potentially anticarcinogenic flavonoids of 28 vegetables and 9 fruits commonly consumed in the Netherlands. *Journal of Agricultural and Food Chemistry* 40:12, 2379–2383.

Johnston, Jason W., Errol W. Hewett, Nigel H. Banks, F. Roger Harker & Maarten L.A.T.M. Hertog (2001). Physical change in apple texture with fruit temperature: effects of cultivar and time in storage. *Postharvest Biology and Technology* pages 13–21.

Kays, Stanley J. (1999). Preharvest factors affecting appearance. *Postharvest Biology and Technology* 15, 233–247.

Kemps, Bart, Lorenzo Leon, Stanley Best, Josse De Baerdemaeker & Bart De Ketelaere (2010). Assessment of the quality parameters in grapes using VIS/NIR spectroscopy. *Biosystems Engineering* doi:10.1016/j.biosystemseng.2010.02.002.

Kim, M.S, C. S. T. Daughtry, E. W. Chappelle, J. E. McMurtrey III & C. L. Walthall (1994). *Proceedings of the 6th Symp. on Physical Measurements and Signatures in Remote Sensing*.

Kreyszig, Erwin (1998). *Advanced Engineering Mathematics*. sixth edition. John Wiley & Sons.

Kummu, M., H. de Moel, M. Porkka, S. Siebert, O. Varis & P.J. Ward (2012). Lost food, wasted resources: Global food supply chain losses and their impacts on freshwater, cropland, and fertiliser use. *Science of the Total Environment* 438, 477–489.

Lai, A., E. Santangelo, G.P. Soressi & R. Fantoni (2007). Analysis of the main secondary metabolites produced in tomato (*Lycopersicon esculentum*, Mill.) epicarp tissue during fruit ripening using fluorescence techniques. *Postharvest Biology and Technology* 43, 335–342.

Lammertyn, Jeroen, Ann Peirs, Josse De Bardemaeker & Bart Nicolaï (2000). Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biology and Technology* 18, 121–132.

Lin, M., M. Al-Holy, M. Mousavi-Hesary, H. Al-Qadiri, A. G. Cavinato & B. A. Rasco (2004). Rapid and quantitative detection of the microbial spoilage in chicken meat by diffuse reflectance spectroscopy (600-1100 nm). *Letters in Applied Microbiology* 39:2, 148–155. doi:10.1111/j.1472-765X.2004.01546.x.

Liu, Yande, Xingmiao Chen & Aiguo Ouyang (2008). Nondestructive determination of pear internal quality indices by visible and near-infrared spectrometry. *LWT - Food Science and Technology* 41:9, 1720 – 1725. doi:10.1016/j.lwt.2007.10.017.

Lohachoompol, Virachnee, George Srzednicki & John Craske (2004). The change of total anthocyanins in blueberries and their antioxidant effect after drying and freezing. *Journal of Biomedicine and Biotechnology* 5, 248–252.

Lu, Guoquan, Huahong Huang & Dapeng Zhang (2006). Prediction of sweetpotato starch physiochemical quality and pasting properties using near-infrared reflectance spectroscopy. *Food Chemistry* 94:4, 632 – 639. doi:10.1016/j.foodchem.2005.02.006.

Mattheis, James P. & John K. Fellman (1999). Preharvest factors influencing flavor of fresh fruit and vegetables. *Postharvest Biology and Techonology* 15, 227–232.

McCaig, T. N. (2002). Extending the use of visible/near-infrared reflectance spectrophotometers to measure colour of food and agricultural products. *Food Research International* 35:8, 731 – 736. doi:10.1016/S0963-9969(02)00068-6.

McGlone, V. Andrew, Robert B. Jordan & Paul J. Martinsen (2002). Vis/NIR estimation at harvest of pre- and post-storage quality indices for 'Royal Gala' apple. *Postharvest Biology and Technology* 25, 135–144.

Nerd, Avinoam & Yosef Mizrahi (1999). The effect of ripening stage on fruit quality after storage of yellow pitaya. *Postharvest Biology and Technology* 15, 99–105.

Ocean Optics (2008). *HR4000 and HR4000CG-UV-NIR Series High-Resolution Fiber Optic Spectrometers Installation and Operation Manual*. Ocean Optics, Inc, 830 Douglas Avenue, Dunedin, FL, USA 34698.

Ordidge, Matthew, Paulina Garcia-Macas, Nicholas H. Battey, Michael H. Gordon, Paul Hadley, Philip John, Julie A. Lovegrove, Eleni Vysini & Alexandra Wagstaffe (2010). Phenolic contents of lettuce, strawberry, raspberry, and blueberry crops cultivated under plastic films varying in ultraviolet transparency. *Food Chemistry* 119:3, 1224 – 1227. ISSN 0308-8146. doi:10.1016/j.foodchem.2009.08.039.

Ozaki, Yukihiro, W. Fred McClure & Alfred A. Christy, editors (2007). *Near-Infrared Spectroscopy in Food Science and Technology*. John Wiley & Sons, Inc., Hoboken, New Jersey: John Wiley & Sons, Inc., Hoboken.

Paull, Robert E. (1999). Effect of temperature and relative humidity on fresh commodity quality. *Postharvest Biology and Technology* 15, 263–277.

Pedreschi, Fracon, Jorge León, Domingo Mery & Pedro Moyano (2006). Development of a computer vision system to measure the color of potato chips. *Food Research International* 39, 1092–1098.

Peirs, Ann (2000). Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR-spectroscopy. *Postharvest Biology and Technology* 21, 189–199.

Penuelas, J, I. Filella, C. Biel, L. Serrano & R. Save (1993). The reflectance at the 950-970 nm region as an indicator of plant water status. *International Journal of Remote Sensing* 14:10, 1887–1905.

Perkins-Veazie, Penelope, Julie K. Collins & Luke Howard (2008). Blueberry fruit response to postharvest application of ultraviolet radiation. *Postharvest Biology and Technology* 47, 280–285.

Polder, G., G.W.A.M van der Heijden, H. van der Voet & I.T. Young (2004). Measuring surface distribution of carotenes and chlorophyll in ripening tomatoes using imaging spectrometry. *Postharvest Biology and Technology* 34, 117–129.

Prange, Robert K., John M. DeLong, Jerry C. Leyte & Peter A. Harrison (2002). Oxygen concentration affects chlorophyll fluorescence in chlorophyll-containing fruit. *Postharvest Biology and Technology* :24, 201–205.

Prieto, N., D.W. Ross, E.A. Navajas, G.R. Nute, R.I. Richardson, J.J. Hyslop, G. Simm & R. Roehe (2009). On-line application of visible and near infrared reflectance spectroscopy to predict chemical-physical and sensory characteristics of beef quality. *Meat Science* 83:1, 96 – 103. doi:10.1016/j.meatsci.2009.04.005.

Qi, J., A. Chehbouni, A.R. Huete, Y.H. Kerr & S. Sorooshian (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment* 48:2, 119 – 126. doi: 10.1016/0034-4257(94)90134-1.

Rein, Maarit (2005). *Copigmentation reactions and color stability of berry anthocyanins*. PhD thesis, University of Helsinki, Faculty of Agriculture and Forestry.

Ren, Min & Mark A. Arnold (2007). Comparison of multivariate calibration models for glucose, urea, and lactate from near-infrared and Raman spectra. *Analytical and Bioanalytical Chemistry* 387:3, 879–888.

Richardson, Andrew D., Shane P. Duigan & Graeme P. Berlyn (2002). An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New Phytologist* 153.

Rodriguez-Saona, L. E., F. S. Fry, M. A. McLaughlin & E. M. Calvey (2001). Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy. *Carbohydrate Research* 336:1, 63–74.

Roujean, Jean-Louis & François-Marie Breon (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment* 51:3, 375 – 384. doi:10.1016/0034-4257(94)00114-3.

Räty, J., K. E. Peiponen & T. Asakura (2004). *UV-Visible Reflection Spectroscopy of Liquids*. Optical Sciences. Springer.

Saltveit, Mikael E. (1999). Effect of ethylene on quality of fresh fruits and vegetables. *Postharvest Biology and Technology* 15.

Sams, Carl E. (1999). Preharvest factors affecting postharvest texture. *Postharvest Biology and Technology* 15, 249–254.

Schanda, János, editor (2007). *Colorimetry*. John Wiley & Sons Inc., Hoboken, New Jersey.

Schmilovitch, Ze'ev, Amos Mizrach, Aharon Hoffman, Haim Egozi & Yoram Fuchs (2000). Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biology and Technology* 19, 245–252.

Schofield, R. Andrew, Jennifer R. DeEll, Dennis P. Murr & Sylvie Jenni (2005). Determining the storage potential of iceberg lettuce with chlorophyll fluorescence. *Postharvest Biology and Technology* 38, 43–56.

Schowengerdt, Robert A. (2007). *Remote Sensing, Modelds and Methods for Image Processing*. Academic Press.

Shewfelt, R.L. (1999). What is quality? *Postharvest Biology and Technology* 15, 197–200.

Sinelli, Nicoletta, Ernestina Casiraghi & Gerard Downey (2008a). Studies on proofing of yeasted bread dough using near- and mid-infrared spectroscopy. *Journal of Agricultural and Food Chemistry* 56:3, 922–931. doi:10.1021/jf0727138.

Sinelli, Nicoletta, Anna Spinardi, Valentina Di Egidio, Ilaria Mignani & Ernestina Casiraghi (2008b). Evaluation of quality and nutraceutical content of blueberries (*Vaccinium corymbosum* L.) by near and mid-infrared spectroscopy. *Postharvest Biology and Technology* 50:1, 31 – 36. doi:10.1016/j.postharvbio.2008.03.013.

Siomos, Anastasios S., Constatinos C. Dogras & Evangelos M. Sfakiotakis (2001). Color development in harvested white asparagus spears in relation to carbon dioxide and oxygen concentration. *Postharvest Biology and Technology* 23, 209–214.

Solomakhin, Alexey & Michael M. Blanke (2010). Can coloured hailnets improve taste (sugar, sugar: acid ratio), consumer appeal (colouration) and nutritional value (anthocyanin, vitamin C) of apple fruit? *LWT - Food Science and Technology* 43:8, 1277 – 1284. doi:10.1016/j.lwt.2010.02.020.

Sugiyama, Takehiro, Junichi Sugiyama, Mizuki Tsuta, Kaori Fujita, Mario Shibata, Mito Kokawa, Tetsuya Araki, Hiroshi Nabetani & Yasuyuki Sagara (2010). NIR spectral imaging with discriminant analysis for detecting foreign materials among blueberries. *Journal of Food Engineering* 101:3, 244 – 252. doi:10.1016/j.jfoodeng.2010.06.026.

Sun, Xiaochun, Felix Marza, Hongxiang Ma, Brett F Carver & Guihua Bai (2010). Mapping quantitative trait loci for quality factors in an inter-class cross of US and Chinese wheat. *Theor Appl Genet* 120:5, 1041–1051. doi:10.1007/s00122-009-1232-x.

T, Sahi (1994). Hypolactasia and lactase persistence. Historical review and the terminology. *Scand J Gastroenterol Suppl* :202, 1–6.

Tipler, Paul A. (1999). *Physics for Scientists and Engineers*. fourth edition. 41 Madison Avenue, New York, NY 10010: W.H. Freeman and Company.

Toivonen, Peter M.A. & Jennifer R. DeEll (2001). Chlorophyll fluorescence, fermentation product accumulation, and quality of stored broccoli in modified atmosphere packages and subequent air storage. *Postharvest Biology and Technology* 23, 61–69.

Volkers, K. C., M. Wachendorf, R. Loges, N. J. Jovanovic & F. Taube (2003). Prediction of the quality of forage maize by near-infrared reflectance spectroscopy. *Animal Feed Science and Technology* 109:1-4, 183 – 194. doi:10.1016/S0377-8401(03)00173-1.

Watada, Alley E. & Ling Qi (1999). Quality of fresh-cut produce. *Postharvest Biology and Technology* 15, 201–205.

WHO Pharmacopoeia Library (2006). *Monographs: Pharmaceutical substances: Morphini hydrochloridum - Morphine hydrochloride*. <URL:http://apps.who.int/phint/en/d/Jb/6.1.273.html> [cited 20th April 2011].

Workman, Jerry Jr. & Art Springsteen, editors (1998). *Applied Spectroscopy, a Compact Reference for Practitioners*. Academic Press.

Workman, Jerry Jr. & Weyer Weyer (2008). *Practical Guide to Interpretive Near-Infrared Spectroscopy*. CRC Press.

Wrolstad, Ronald E., Robert W. Durst & Jungmin Lee (2005). Tracking color and pigment changes in anthocyanin products. *Trends in Food Science & Technology* 16, 423–428.

Zagory, Devon (1999). Effect of post-processing handling and packaging on microbial populations. *Postharvest Biology and Technology* 15, 313–321.

Zerbini, Paola Eccher, Maurizio Grassi, Rinaldo Cubeddu, Antonio Pifferi & Alessandro Torricelli (2002). Nondestructive detection of brown heart in pears by time-resolved reflectance spectroscopy. *Postharvest Biology and Technology* 25, 87–97.

A. NIR SPECTROPHOTOMETERS

The following information is taken from the user and installation manuals of the corresponding products.

Table 10. NIR spectrophotometer comparisation. (*) Parafuel operates with single acquisition (sample).

Property	Ocean Optics HR4000	LT Industries ParaFuel	Control Development
Range (nm)	198.5 – 1118.5	1201–2401	905 – 1682
Resolution (nm)	0.03 – 1.0	1	
Channels	1	1	1
Integration time	3.8 ms – 10 s	30 ms	-
Focal length (input)	F/4 101mm	n/a	-
Focal length (output)	68 mm	n/a	-
Input Fiber connector	SMA 905	LT	SMA 905
Signal-to-Noise	300:1(*)	20 000:1	-