Master's Thesis

Petra Kaukonen, 2021

LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY LUT

LUT School of Engineering Science

Chemical Engineering

Petra Kaukonen

ONLINE NIR ANALYSIS FOR HEMI-HYDROLYZED WOOD PULP

Examiners   Prof. Satu-Pia Reinikainen
                    Assoc. Prof. Ritva Tuunila
Supervisors M.Sc. (tech.) Sami Turunen
                    M.Sc. (chem.) Tuomas Mäkinen

Puupohjaisen lignoselluloosabiomassan hyödyntämistä prosesseissa biokemikaaleiksi ja uusiksi tuotteiksi on tutkittu useita vuosia. Biomassan muuttaminen tuotteiksi sisältää useita vaiheita, joista erityisesti tehokas esikäsittelyprosessi on tärkeä. Lignoselluloosabiomassa koostuu pääasiassa selluloosasta, ligniinistä ja hemiselluloosasta. Uusien biojalostamoiden esikäsittelyprosessit edellyttävät edistynyttä prosessiohjausta, jossa online-analytiikalla on merkittävä rooli kemiallisten muutosten mittaamisessa nopeasti reaaliajassa, jotta prosessia voitaisiin ohjata parhaalla mahdollisella tavalla. Tavanomaiset kemialliset menetelmät, joita käytetään lignoselluloosabiomassan karakterisointiin, ovat aikaa vieviä ja vahvasti laboratoriointensiivisiä.

Spektroskooppiset menetelmät tarjoavat vaihtoehdon lignoselluloosaprosessin reaaliaikaiselle seuraamiselle mitattujen spektrien perusteella. Lähi-infrapunaspektroskopian (NIR) sopivuutta testattiin online-mittauksiin esikäsitellylle biomassalle ligniinin ja kappaluvun määrittämiseksi. Ensimmäiset mittaukset ja kalibroinnit suoritettiin laboratorio-olosuhteissa kiinteälle ja nestemäisille näytefraktioille. Laboratoriomittaukset osoittivat hyvää korrelaatiota ligniinin ja kappaluvun laboratorioreferenssiarvioihin, jonka jälkeen NIR-menetelmää testattiin inline-mittauksina prosessin pilotointiolosuhteissa lietefraktiolle hyödyntämällä monimuuttujaisia tilastollisia prosessinohjausmenetelmiä. Online-mittaukset eivät olleet täysin suoraviivaisia, mutta lupaavia. NIR on uusi menetelmä tällaisen esikäsitellyn biomassan mittaamisessa ja aiheesta on tällä hetkellä saatavilla niukasti tutkimustietoa, erityisesti online-mittauksista. Tämä tutkimus osoitti, että NIR-mittauksilla on potentiaalia biomassan online-analyyseihin, mutta tulevaisuudessa kalibrointeja tulisi parantaa. Tämä voidaan toteuttaa käyttämällä ulkopuolisia korjauskertoimia, kuten kosteuspitoisuutta ja lämpötilaa, jotka sisältyvät kalibrointimalliin.

ABSTRACT

Lignocellulosic wood biomass utilization and process development into value added biochemicals have been under research for several years. Biomass conversion for final products include several steps where especially efficient pretreatment is important. Lignocellulosic biomass is mainly composed of cellulose, lignin, and hemicelluloses. New advanced biorefinery pretreatment processes require advanced process control where online analysis is in significant role for measuring chemical substance changes and process control for the best possible ways. Conventional laboratory based chemical methods that are used for lignocellulosic biomass characterization are time consuming and strongly laboratory intensive.

Spectroscopical methods provide option for process real time analysis for lignocellulosic biomass based on spectral data collected from the process. Near-infrared spectroscopy (NIR) feasibility as online analysis was tested for pretreated lignocellulosic biomass to measure lignin and kappa number. NIR calibrations created first laboratory measurements for pretreated biomass solid and liquid fractions. Laboratory measurement showed good correlation against lignin and kappa number laboratory references. Therefore, NIR was tested as inline measurements in real process piloting conditions for slurry applying multivariate process control charts. Results from online measurements were not straightforward but presented promising option. NIR is new method to measure this kind of pretreated lignocellulosic material and there is available only some information now of this topic especially online analysis. This study proved NIR potential to be utilized for online analysis, but calibrations should be improved in the future. This can be implemented with using outside corrections such as moisture content and temperature control that are included calibration.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

NIRS	Near-infrared spectroscopy
FTIR	Fourier-transform infrared spectroscopy
IEA	International Energy Agency
EU	European Union
BioMEG	monoethylene glycol
BioMPG	monopropylene glycol
RFF	functional fillers
PET	polyethylene terephthalate
PAT	process analytical technology
APC	advanced process control
IR	infrared radiation
v_0	natural frequency
k	force constant of the chemical bonding
μ	atoms reduced mass
E	vibration level
X_e	an anharmonic constant
PCA	principal component analysis
PLS	partial least squares
CCD	charged coupled device
S/N	signal-to-noise ratio
FT	Fourier-transform
MSCP	multivariate statistical process chart
T	object score matrix
X	data matrix
P	variable loadings
E	noise and errors
A	number of PCA factors
m	number of vars
n	number of samples
$SPE_x(Q)$	Squared predicted error chart
MEWMA	Multivariate Exponentially Weighted Moving Average

SPEC	squared predicted error
MCUSUM	Multivariate Cumulative Sum
UV-VIS	ultraviolet/visible spectroscopy
SNV	standard normal variate
AACC	American Association for Clinical Chemistry
NREL	National Renewable Laboratory
RMESV	root mean squared error of validation
R^2	goodness of fit
HMF	5-hydroxymethyl-2-furfural
HPLC	high-performance liquid chromatography
O	original biomass pulp
W	washed solids
L	liquid/filtrate
S	slurry
OL	online loop tests
MSC	multiplicative scatter correction
A_λ	wavelength-dependent absorbance
T	light transmittance
ε_λ	wavelength-dependent molar absorptivity
I	path length of light through the sample
c	concentration
R	detected reflectance
x_i'	denotes 1st derivative
x_i''	2nd derivative at wavelength points i
RMSEC	Root mean square error of calibration
PRESS	Predicted error sum of squares
$w - \%$	weight percent

1 INTRODUCTION

Biomass utilization and process development into biofuels and value added biochemicals have been under research for several years. Factors such as sustainability and green based preferences have driven corporations use renewable feedstock sources to replace fossil-based materials where biomass is one of the most potential and promising solution to respond this need. (Özdençi, 2017) (Shinde et al., 2018) Especially lignocellulosic based biorefinery processes are growing trend and market field of biochemicals from lignocellulosic materials is predicted to increase from 2 % to 22 % by 2025. (Bello et al., 2015) European Commission has set a bioeconomy strategy 2012 which is valid until 2030. In that strategy circular economy and effective utilization process of lignocellulosic biomass materials into valuable products are at centre. (Hassan et al., 2019) In upcoming years an economic aspect for efficient biomass utilization processes and development are even more in remarkable role. (Menon & Rao, 2012)

Lignocellulosic biomass represents an inexpensive and abundant material that has huge opportunity to be used in fuels, biochemicals, energy and biobased materials. Lignocellulosic feedstock is mainly composed of hemicellulose, cellulose and lignin that form major structure of woody biomass. Biomass conversion in biorefinery concept into valuable products include several steps where diligent and efficient pretreatment is in key role. Pretreatment has straight impact further biorefinery process and product applications which affect its high process costs. (Bello et al., 2018) Lignocellulosic wood material pretreatment process is challenging into individual products due to material's complicated chemical structure. Wood fermentable sugars hard releasing into individual units to be utilized further for different end products require strong and effective pretreatment conditions to break the structure. (Hassan et al., 2019) (Kucharska et al., 2018) (Ewanick et al., 2014) During the last two decades pretreatment processes have been developed to understand lignocellulosic chemical structure changes, behavior, and natural recalcitrance including interactions between lignin units and carbohydrate polymers. (Sills & Gossett, 2011) (Shinde et al., 2018) There are available several options for pretreatment prospective process but roughly they can be categorized physical, chemical, biological, and physico-chemical operations. (Bello et al., 2018)

Lignocellulosic processes and especially pretreatment step need careful real time process control and monitoring. Nowadays used conventional analyses for lignocellulosic wood

material characterization are time consuming and expensive but also very labor-intensive. Conventional wet chemical methods have been proven ability to work well for wood-based material analysis in laboratory scale, but large number of samples and industrial applications need instant improvements and fast online analysis. Analysis of lignocellulosic biomass with two-staged sulfuric acid hydrolysis has been used a century as chemical method for total lignin, carbohydrates, and sugars in laboratory scale. Wet chemical methods weaknesses are expensive costs and laboratory necessity. In addition, these methods require pretreatment where some extractives are removed, and sample particle size designated to achieve reliable results. Disadvantage of these methods is that that they are not able to separate certain types of hemicellulose mixtures like arabinoxylan and xyloglucan. (Xu et al., 2013) (Jiang et al., 2013)

Spectroscopic methods have increased attraction online, inline, and at-line applications in process scale for several years to be developed and specified more. (Sun et al., 1997) These methods differ from conventional process measurements such as pH, temperature, and flow rate meters whereas information is based on more physical changes. Spectroscopic online analysis provides information of process chemical changes based on spectral information. Near-infrared spectroscopy (NIRS) provides non-invasive, environmentally friendly, and rapid method to characterize lignocellulosic woody mass component chemical features, process changes and conditions in online measurements simultaneously. (Xu et al., 2013) Therefore, industrial process applications especially in developed lignocellulosic biorefinery concepts needs real-time process control to measure quality and substance properties in pretreatment step. NIRS has proven potential to be used for lignocellulosic material characterization and component analysis providing information of process chemical changes. (Xue et al., 2015) Accordingly, Raman and Fourier-transform infrared spectroscopy (FTIR) have also been used and proven feasible for measurements of biomass features also in process scale environment. (Ewanick et al., 2014) (Xu et al., 2013)

1.1 Objectives of Master Thesis

The objective of this Master Thesis work is to investigate NIR capability and feasibility for online measurements of solid and liquid fractions from pretreated lignocellulosic woody biomass. NIR technique is new method to measure this kind of pretreated woody biomass. There are available only some research publications related to this topic of online analysis and piloting. NIR as online measurement has huge potential to be investigated more in this field of study to create fast chemical process measurements form spectral data changes. Literature part

represents background for this work where biorefinery concept, lignocellulosic material, and spectroscopic methods for online process measurements are described. Experimental part is focusing first on NIR laboratory measurements and after that online test for evaluating significant process variable changes affecting online analysis at pilot scale.

LITERATURE PART I

2 BIOREFINERY CONCEPT PRINCIPLE

The biorefinery concept can be comprised process to convert or fractionate biomass further into energy, bio-based materials or value-added biochemicals using nature resources respectfully. The main target in biorefinery processes is to maximize the utilization of biomass feedstock while decreasing emissions and minimizing waste and residue streams. (Alén, 2011) In the long run, meaning of cost-effective and economical aspects show more ways to develop biorefineries to build up sustainable lifecycle in the upcoming years. Thereby development and implementation of biorefineries need to be considered as sustainable aspects and supply chain point of views that meets the vision of using feedstock materials in the future effectively. (Menon & Rao, 2012)

Biorefineries are classified into different types based on used raw materials and process routes as first, second and third generation biorefineries. (Pandey et al., 2015) First order biorefinery target is to produce one focused main product and use one feedstock material such as producing for instance biodiesel. Examples of first order biorefineries are also today's paper and pulp mills as well as corn grain to ethanol plants. Second type biorefineries produce one feedstock but differ from that part because they are able produce various end products and energy. Third order biorefineries produce chemical products and energy from various feedstock materials that represent the most developed and advanced biorefinery of these three types. (Clark & Deswarte, 2008) Biorefineries process steps are feedstock handling and collection, pretreatment, conversion process, product separation and products classification. (Özdençi et al., 2017) The International Energy Agency (IEA) biorefinery classification system is used for classification biomass utilization processes that are thermochemical, biological, chemical, and mechanical conversion routes. (OECD, 2017)

2.1 Lignocellulosic biorefinery

Lignocellulosic biorefinery uses lignocellulosic feedstock derived from wood, corn stover, straw, sugarcane bagasse or rice for producing energy, bio-based materials and biochemicals or just one of these end-products. (De Bhowmick et al., 2018) Lignocellulosic biomass provides most abundant resource which leads into the upcoming trend which consist bioconversion of lignocellulosic feedstock up to value-added green based products. (Özdençi et al., 2017) (Menon & Rao, 2012) Lignocellulosic feedstock is commonly treated with using alkaline agents or acids for lignin, cellulose, and hemicellulose to create form that could be easily used in hydrolysis step which further produces C5 and C6 sugars. These sugar products can be used straight as feedstock material in fermentation process for producing biofuels such as ethanol, butanol, and hydrogen but also nowadays for value added biochemicals. Before lignin fraction has been sold or used as energy source but now it could be converted into future application products to replace fossil-based materials that are for instance phenolic components or composites. (Pandey et al., 2015) Lignocellulosic biorefinery concept route to valuable products is presented in Figure 1.
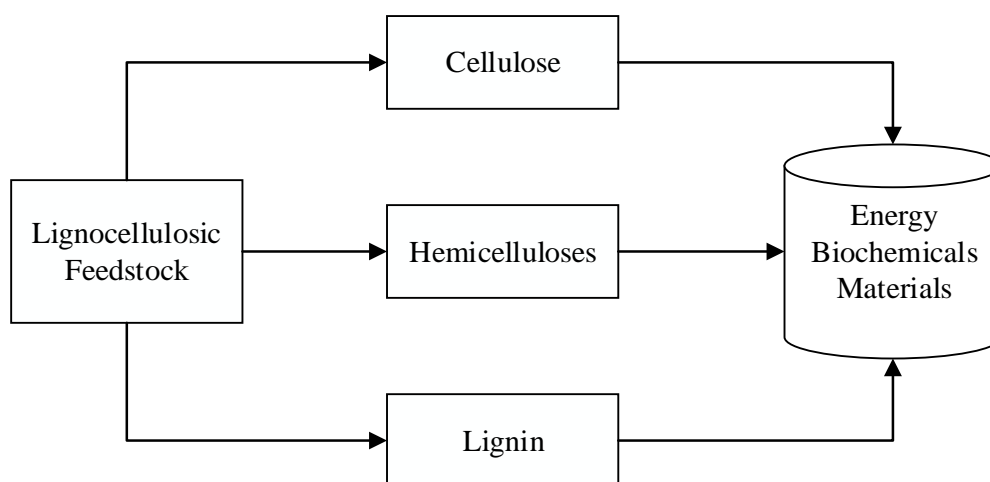
Figure 1.   Lignocellulosic biorefinery route to valuable products (modified Clark & Deswarte, 2008)

2.2 Lignocellulosic biorefineries current state and prospects

Biorefineries that use lignocellulosic feedstock material are representing one of the most important feedstock materials in European bioeconomy development. Lignocellulosic biorefineries product markets are increasing which means that by the year 2030 approximately 30 % of oil-based chemicals are replaced with biochemicals which form important

infrastructure to the European bioeconomy development. Even if lignocellulosic biorefineries will have challenges such as raw material logistics, conventional process technology problems and market field in the future, EU will support this industry development for achieving environmental targets and bio-based material replacements. (Hassan et al., 2019) Lignocellulosic biorefineries have become as research hotspots lately and upcoming years the development study will continue and lignocellulosic biomass process to valuable products will increase and form new industries. (Mussatto, 2016) (Hongzhang, 2015) It is estimated that market field is 40 million EURO by 2020 and reach up to 50 million EURO until 2030 with annual growth rate of 4 %. European Commission released the bioeconomy strategy in 2012 achieving the sustainable biorefinery and cost-effective lignocellulosic feedstock utilization. Totally, 67 biorefineries in lignocellulosic field are operating now worldwide whereas one-third at commercial scale. Lignocellulosic biorefineries have attracted interesting among scientists while approximately 130-150 patents are submitted yearly especially related on pre-treatment which is the biggest current bottleneck and object of development in the upcoming years. (Hassan et al., 2019)

2.3 Pretreatment process of lignocellulosic feedstock

Lignocellulosic material has its own characteristics to be further utilized. Fermentable sugars are locked in recalcitrant structure which is major limitation and challenge. (Fatma et al., 2018) Lignocellulosic material pretreatment is focusing on enhancing hemicellulose and cellulose affordance for further process steps such as saccharification to be more easily handled and converted. (Kucharska et al., 2018) In the Figure 2 is presented pretreatment effects for lignocellulosic wood structure. This kind of pretreatment step is efficient when the formation of sugars is effective in following process steps such as enzymatic hydrolysis where carbohydrate components are formed, and inhibitor compounds formation is prevented to minimum. From another point of view, the pretreatment step of lignocellulosic material could be route to fractionate components valuable products and potential biochemicals. This pretreatment approach does not only focus on enhancing the production of biofuels than separating these valuable chemicals from lignin and hemicellulose its own fractions. However, even if biomass fractionation to individual fractions is still bottleneck. This provides challenges and huge opportunities to be solved that in the future lignocellulosic feedstock could be utilized the best possible way into individual building blocks from critical pretreatment step. (Kucharska et al., 2018)
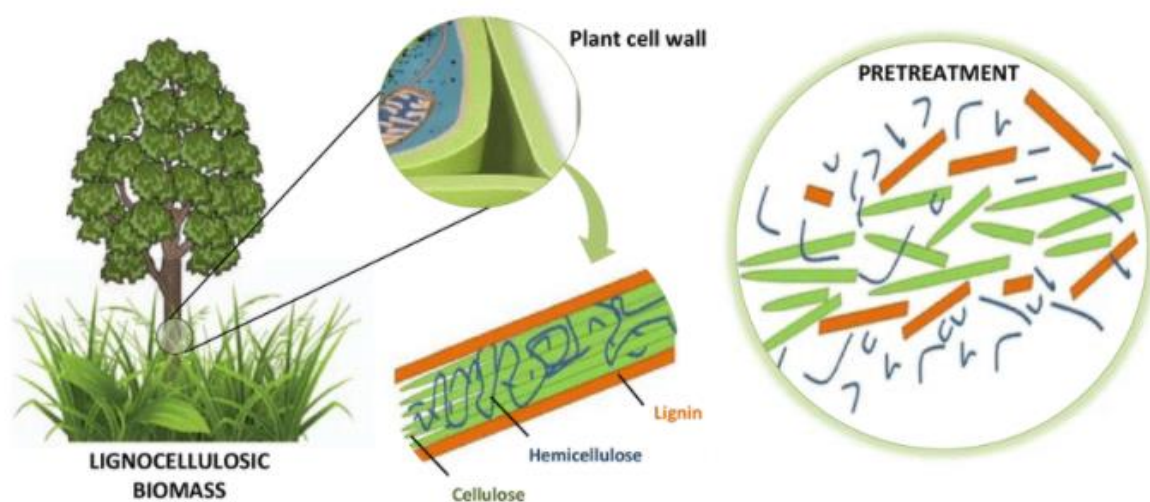
Figure 2.   Pretreatment effects to the lignocellulosic wood structure (Mussatto, 2016)

Comparing conventional process development and first-generation biofuels biorefineries scaling up was achieved relatively fluent due to easy extraction of fermentable sugars and oils from biomass material that was easily straight scaled up into bigger extraction process. Lignocellulosic material characteristic feature remains challenging to be utilized but in last ten years development has gone forward and lignocellulosic biorefineries have attracted therefore notice in the EU. (Hassan et al., 2019) Pretreatment of lignocellulosic feedstock material biofuels and valuable products is the most critical, sensitive, and decisive step. (Hassan et al., 2019) For achieving effective pretreatment process several factors should be considered such as reduction of particle size of biomass, low catalyst costs, energy minimization and preserving hemicellulose fractions as well as minimizing inhibitor formation. (Menon & Rao, 2012) (De Bhowmick et al., 2018)

2.4 Process description

Wood as material provides environmentally friendly, climate neutral and renewable feedstock. Woody based lignocellulosic material does not compete with food production industry. The process route represents new second generation industrial lignocellulosic biorefinery concept which end products are renewable glycols and lignin for replacing fossil-based materials. Primary products for this biorefinery are sugars and lignin but also residue streams which are converted into biofuels and bioenergy. Process sugars are further processed monoethylene glycol (BioMEG) and monopropylene glycol (BioMPG). Lignin is further converted into renewable functional fillers (RFF) that can be used replacing silica and carbon black as well as raw material to different kind of rubber applications. BioMEG can be applied in PET bottles,

textiles, packaging, and deicing fluids. Accordingly, BioMPG is utilized in composite, pharma, detergents, and cosmetic applications. The used feedstock material for process is biomass made of beech which include sawdust and responsible forest thinning. All wood that is used is based on certified and regionally sourced feedstock from sustainably managed forests. The considered process and focused process area marked as dashed line is presented in the Figure 3 which performs as a background for this study. (UPM Biochemicals, 2020)
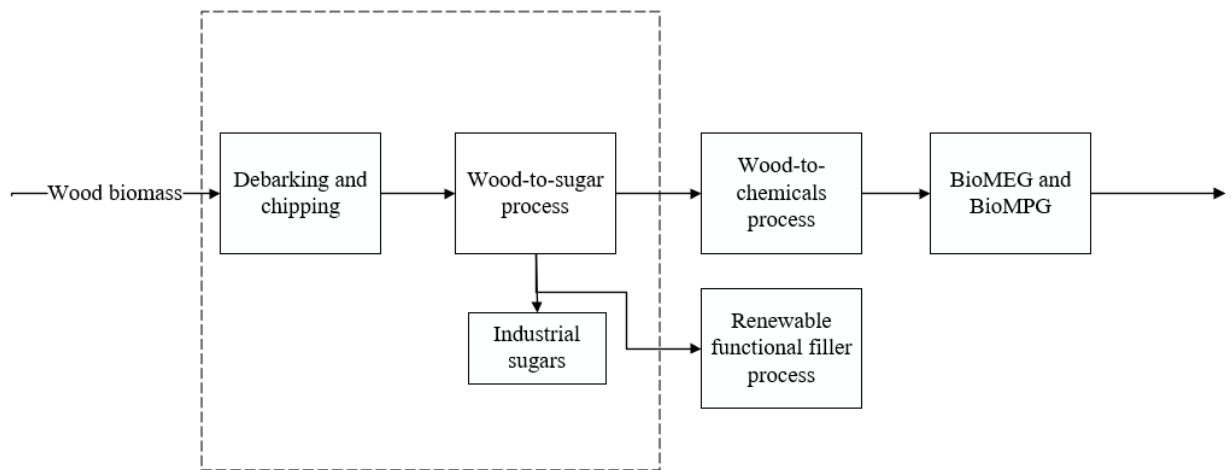
Figure 3.   Wood to sugars process route (modified UPM Biochemicals, 2020)

3 LIGNOCELLULOSIC WOOD BIOMASS

Growing environmental reasons and more sustainable society aspect fossil fuel-based materials and process routes want to be replaced with renewable ones. (Matsakas et al., 2019) Biomass have huge potential to be utilized as one of the best available resources on Earth for producing more greener options to produce energy and bio-based products. Biomass is term to explain all kind of biologically produced feedstock matter such as wood, energy crops and some agricultural and forests residues that form the main renewable sources to be used environmentally friendly solutions which have attracted much interest globally. (Brys et al., 2016) Especially, wood based lignocellulosic biomass including cellulose, hemicellulose and lignin have increased interest lately and forms important role in forest-based process development nowadays as renewable source of feedstock material in biorefinery development purposes. (Matsakas et al., 2019)

3.1 Composition of lignocellulosic wood biomass

Generally, the wood biomass can be categorized as trees, wood waste, bark, sawdust, and timber. Woody biomass structure understanding as the polymeric chemical consistency of wood is referred because wood modification processes affect strongly on biomass structure and applications. The main components for wood structural polymeric components are hemicellulose, lignin, and cellulose but also some extractives are included to the structure too. (Hill, 2006) Additional components beside cell walls such as water and extractives are appeared in commonly in microfibers and dissolved in the matrices. However, add to cell walls these compounds can be found in lumens, filling, or flowing elements or in coating parts. These wall structure compounds content varies a lot in different wood species in the total mass of tree that notably affects straight to wood properties. (Kettunen, 2006) The main components of woody biomass are presented in the Figure 4. (Basu, 2013) Now considered biomass components are cellulose, hemicellulose, and lignin because studied wood based lignocellulosic biomass at this purpose does not include extractives or ash or the content of these are relatively very low.
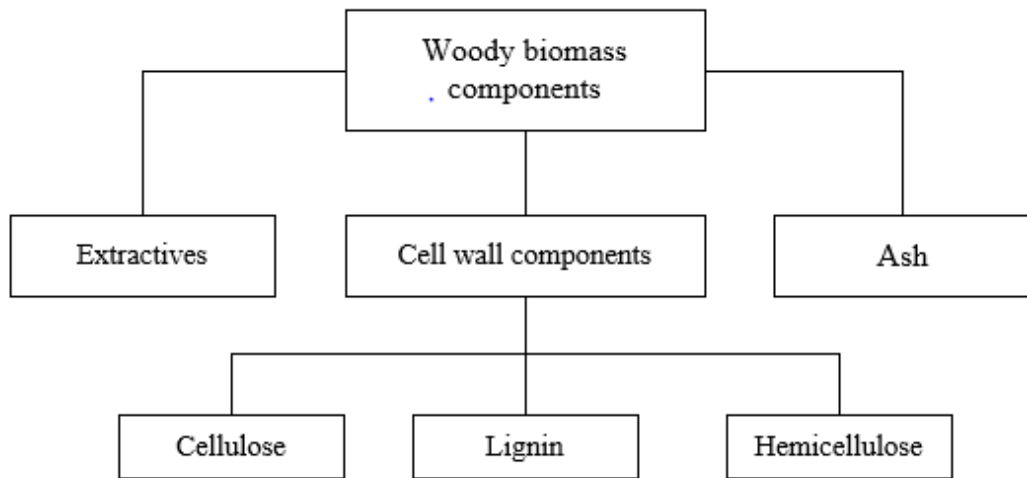
Figure 4.   Woody biomass main components (modified from Basu, 2013)

3.2 Cellulose

Cellulose is linear long chain polymer with crystal structure of many glucose units. This makes its chemical structure very strong and provides base structure to biomass that content in wood is 40 to 50 %. (Basu, 2013) Therefore cellulose forms the most abundant resource of natural polymer characteristics in worldwide. (Henze et al., 2018) Cellulose chemical structure is

almost crystallized natural thermoplastic polymer found in microfibers cell walls. Cellulose is formed during wood growth with polymerization reaction of glucose monomers linked into molecule chain as β-1,4-polyglucoside where the 1,4-bonds make cellulose chemistry strong, rigid, and straight forming D-anhydroglucopyranose units. (Kettunen, 2006) (Mussatto, 2016) Cellulose is composed of 5000-10 000 glucose units. Cellulose crystalline structure is harder to be hydrolysed than amorphous units and prevent chemical or enzymatic degradation and make it insoluble into most solvents. (Mussatto, 2016) This stable and hard crystalline structure of cellulose is caused by intra and inter molecular hydrogen bonds that are formed between anhydroglucan chains. (Mäki-Arvela et al., 2011) The polymer structure of cellulose is presented in Figure 5.
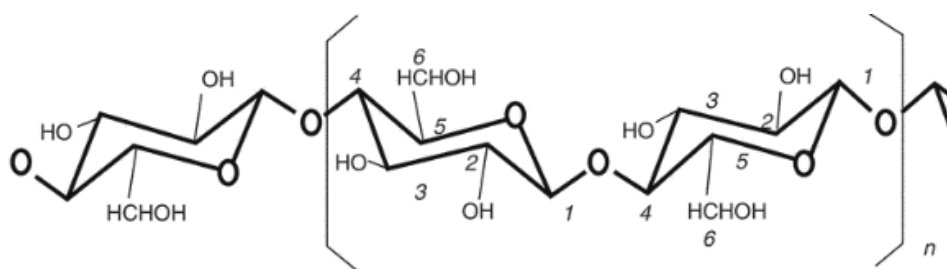
Figure 5.   Linear cellulose polymer structure with β-1,4-polyglucoside linkage (Woodings, 2001)

3.3 Hemicellulose

Hemicellulose forms 20-35 % of total lignocellulosic biomass and it can be characterized as the second most common polysaccharide occurring in the nature. (Saha, 2003) Hemicellulose is part of the plant cell walls where it creates a strong crystallized structure. (Basu, 2013) Hemicelluloses are consisted of heterogenous polymers (arabinose, xylose), hexoses (mannose, glucose, galactose) and some sugar acids where xylose present largest quantity which is about 90 % of hemicellulose consistency. (Clark & Deswarte, 2008) (Mussatto, 2016) Comparing hemicellulose structure to cellulose, the composition of hemicellulose is not homogenous because it is composed of these different sugar units. (Hill, 2006) (Saha, 2003) Hemicelluloses form amorphous structures where short chains are more easily degraded with chemical treatment or enzymatic hydrolysis than cellulose. (Mussatto, 2016) The principle of hemicellulose in wood structure is build up a linkage between cellulose and lignin compounds. (Fatma et al., 2018) The main chemical composition of hemicellulose structure is presented in Figure 6. (Mohan et al., 2006)
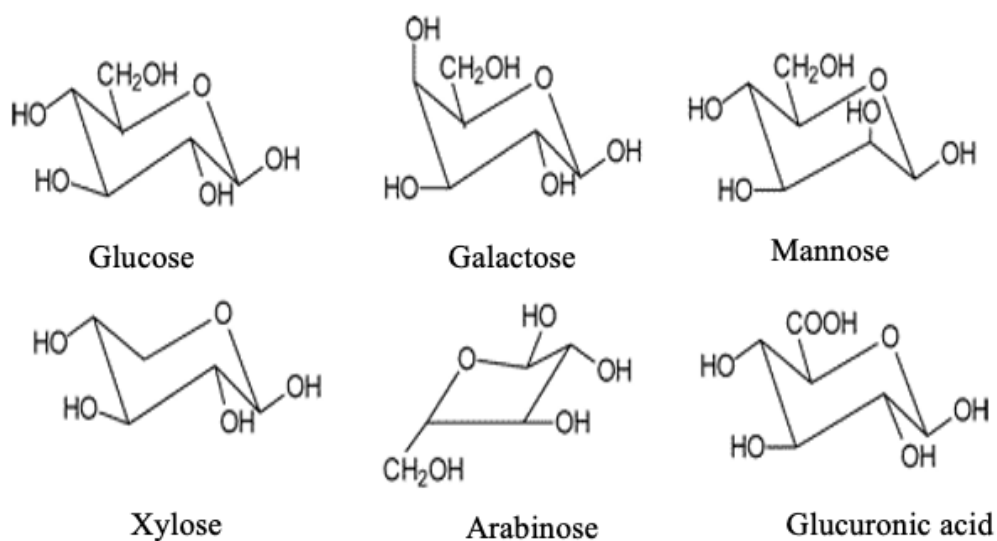
Figure 6.   Hemicellulose main component structures (Mohan et al., 2006)

Among different wood species there is clear variation between hemicellulose content and chemical composition where hardwood and softwood species content can be comparable, but behaviour of them differs that has impact biomass features as well. Distribution of different hemicellulose content distributions in weight-% for hardwood and softwood is presented in Table I.

Table I   Different hemicelluloses in wood structure in hardwood and softwood as w-% (Mäki-Arvela et al., 2011)

Hemicellulose, w-%	Hardwood	Softwood
arabinomethylglucuronoxylans	0,1-1	15-30
methylglucuronoxylans	80-90	5-15
glucomannans	1-5	1-5
galactoglucomannans	0,1-1	60-70
arabinogalactans	0,1-1	1-15
other galactans	0,1-1	0,1-1
pectins	1-5	1-5

Softwood hemicellulose is formed galactoglucomannans and arabino-glucuronoxylan as xylan units and in hardwood species. (Basu, 2013) The most significant hemicelluloses in softwood species are galactoclugomannans and arabinoglucurunoxylans. Softwood also contains xyloglucans, arabinogalactan, and some other glucans. Examples of softwood such as spruce

and pine are composed of 20 % O-acetyl-galactoglucomannan and 5-10 % of arabino-4-O-methylglucuronoxylan. Accordingly, in hardwood species major component is xylan and particularly O-acetyl-4-O-methylglucuronoxylan which amount is 80-90 % of wood's hemicellulose structure. Acetylglucuronoxylan amount in hardwood is approximately between 13-50 % and glucomannan 1-5 %. Arabinogalactans are more common in softwood species than in hardwoods. (Mäki-Arvela et al., 2011)

3.4 Lignin

Lignin is natural polymer that is one of the major components of lignocellulosic structure after cellulose and hemicellulose as lignin content is approximately 15-30 %. (Lu et al., 2017) Lignin's composition and chemical structure are complicated and more understanding needs to be achieved to create more efficient and environmentally friendly processes to utilize lignin features even better in new value-added applications. Lignin based modifications can be utilized widely in industrial applications such as fuels, functional polymers, carbon fibres as well as in aromatic chemicals. Apart from this lignin is still quite hard and slow component to be used because its complex chemistry. Lately there have been research related to separating it from cellulosic biomass to industrial process purposes. (Akpan, 2019)

Lignin chemical structure is strongly dependent on biomass type, plant, age, location of tissue and on growing conditions. (Lu et al., 2020) Lignin physicochemical structure can be modelled as amorphous and three dimensional and its main function in the plan cell is structural support between cellulose and hemicellulose structures, protecting from biological and chemical hazardous components but also transporting water and some essential nutrients. It consists of methoxylated phenylpropane structures and lignin chemical functional groups are carboxyl, methoxyl, carbonyl and hydroxyl linked into aromatic or aliphatic components with various properties and amounts which lead into different compositions and structures of lignin modules. (Akpan, 2019) (Gillet et al., 2017) Lignin is consisted with three primary monolignols units which are sinapyl (3.5-dimethoxy-4-hydroxycinnamyl), p-coumaryl alcohols (4-hydroxycinnamyl) and coniferyl (3-methoxy-4-hydroxycinnamyl) that are linked to ether or carbon bonds. These monolignols are known as monomer units are p-hydroxyphenyl (H), syringyl (S) and guaiacyl (G) which are presented in Figure 7. (Akpan, 2019)
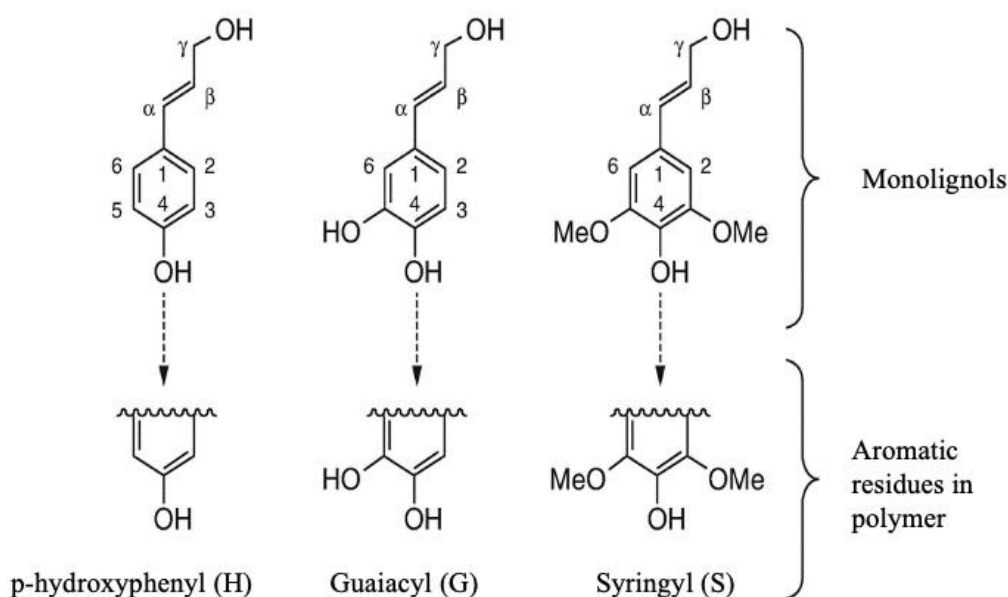
Figure 7.   Three basic lignin monomeric units (H), (G) and (S) (Akpan, 2019)

Lignin can be divided into softwood, hardwood, and non-wood lignin. Lignin amount in softwoods is in rage of 25-35 % and in hardwoods 18-25 %. Lignin amount varies a lot with different sources and wood species. (Akpan, 2019) Grass lignin content is only 10-15 % of the considered total plant mass. (Gillet et al., 2017) In softwood species like spruce more than 90 % of lignin are G monolignol units. Accordingly, in hardwood species such as beech, aspen and birch the lignin content in S-units are between 45-75 % that is thus relatively bigger than G-units. Grasses have higher H-unit content than softwood or hardwood species. (Lu et al., 2020) Chemical structure of lignin between grass and softwood does not differ a lot but hardwood lignin's have special characteristics and variation between different wood species like amounts of specific dimers, functional groups, and degree of condensation. Add to that hardwood lignin has less condensed structure when comparing to the softwood lignin structure. Softwood species lignin structure forms compact compound structures that have very strong resistance features to degrading attacks. H/G/S-units in wood structure have straight impact biomass features such as utilization, pretreatment and digestibility. Especially the S/G-units ratio impacts the pretreatment performance on biomass processes and lignin degradability reactions as higher S-unit content causes that lignin is easier to remove. Effects to enzymatic

hydrolysis are strongly related to the S/G-unit ratio content based on biomass hydrolysable operations as well. (Lu et al., 2020)

Lignin can be typically classified into technical and native lignin. Native lignin is original lignin structure in the lignocellulose without any modifications and it does not occur pure in nature rather part of lignocellulose. Accordingly, the main sources of technical lignin are available from industries that are kraft lignin and its several lignin types as for example hydrolysis lignin, pyrolytic lignin and organosolv lignin. Chemical composition of these lignins varies due to different sources and extraction methods. Technical lignin as by-product from industry field can be converted straight as raw material for chemical production. Due to aliphatic and aromatic hydroxyl groups which are major components in technical lignin, it can be utilized widely replacing fossil-based feedstock materials. (Chio et al., 2019) Especially lignin has very important role in biomass utilization because it provides highly strong feedstock material and add to that it has good rigidity features and good defence against microbiological infections. (Kameshwar & Qin, 2017)

4 LIGNOCELLULOSE INTO VALUE ADDED CHEMICALS AND SUGARS

Hemicellulose can be converted into valuable products with biological and chemical process routes where it is utilized further to value-added sugars and chemicals. Generally, hemicellulose is converted using enzymatic hydrolysis or chemical process treatment where the first step is the critical pretreatment process. (Saha, 2003) Process options for hemicellulose pretreatment are acid, water (steam or liquid), alkaline agents or organic solvents. After hemihydrolysis process step, the important sugars in acid hydrolysis are formed that are mannose, xylose, galactose, arabinose and rhamnose. It has been noted that acid hydrolysis method has more advantages than dilute acid pretreatment. (Khanchanalai et al., 2016) (Mäki-Arvela et al., 2011) Organic solvents and alkaline treatment are not good solutions when considering lignin valorisation. In these processes, lignin is removed which affects straight for the fermentation process or in bioconversion of sugars from hemicellulose material. Biomass containing both pentoses and hexoses can derive in hydrolysis process although it is more usual with using strong acids or enzymatic catalysts. By-products such as hydroxymethylfurfural and furfural that are both very strong inhibitors. From these two, formic and levulinic acids can be formed

as degradation products. Lignocellulosic material possible reaction pathways in hydrolysis are presented in the Figure 8.
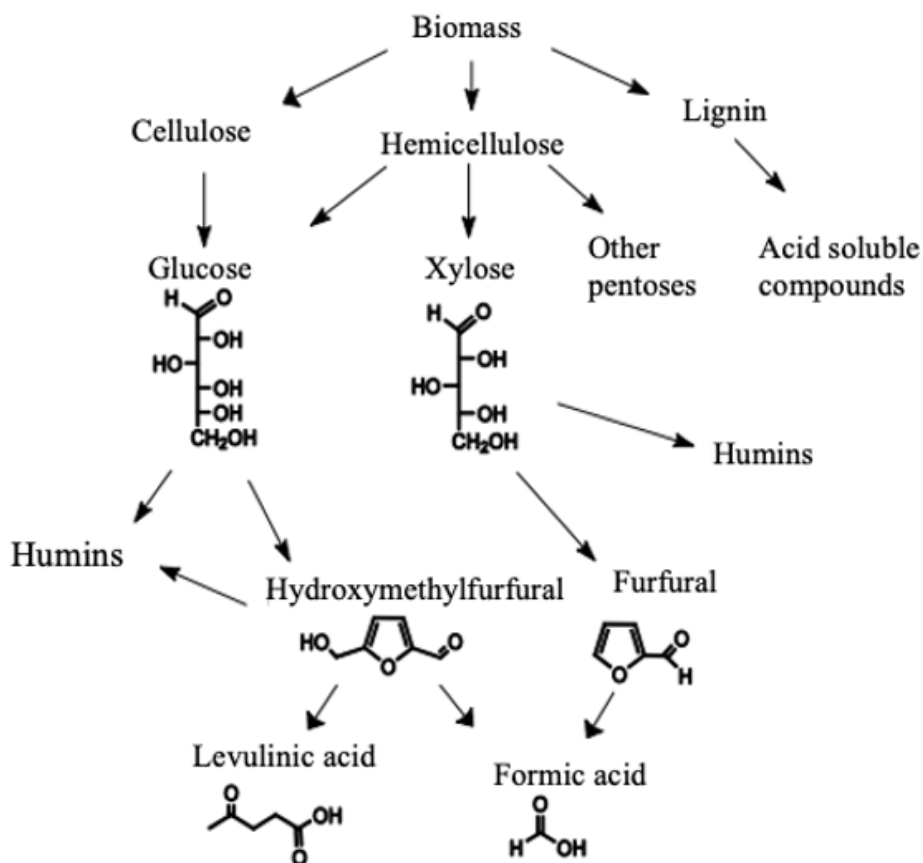
Figure 8.   Reaction routes for biomass in acid hydrolysis pretreatment process (Kanchanalai et al., 2016)

Most used pretreatment are acid, water, and steam process routes. In these processes lignocellulosic biomass is provided with route to produce good selectivity for hemicellulose and solutions that are partly or totally hydrolysed monomeric or oligomeric sugars as well as cellulose rich solids for further applications and bio-based processes. Pretreatment methods examples and conventional process routes are presented in the Table II.

Table II   Pretreatment methods and example process routes for lignocellulosic biomass (Saha, 2003)

Method	Example process route
Thermo-mechanical	Milling, shearing, grinding, extruder
Acid treatment	Dilute acid (HCl, H_2SO_4), concentrated acid (H_2SO_4, HCl)
Organic solvents	Ethanol, butanol, phenol, methanol
Alkali treatment	Ammonia, sodium hydroxide, alkaline hydrogen peroxide
Autohydrolysis	Steam pressure, steam explosion, supercritical carbon explosion

5 ONLINE PROCESS MEASUREMENTS

Process analytical and advanced measurements importance have increased as part of chemical industry nowadays. Development of online process sensors for creating real time information from process conditions during production is important. (Rønnest et al., 2011) Process analytical technologies (PAT) guidance is published framework by the USA FDA for industries especially for pharmaceutical industry in 2004. However, PAT gives guidance to other industrial fields to be utilized as well in the future together with high-throughput equipment achieve product quality aspects. (Chen et al., 2011) Process analytical technologies provide process monitor and control of critical features to respond the end-product quality requirements. Technologies like these form a system which is build up based on the most critical process parameters. Information is collected such as chemical and physical point of views that affect most process working principle. PAT provides also better process understanding where technologies can be divided into process control tools, process analyzers, multivariate design for data analysis and some continuous tools for improving real time process knowledge and management. (Challa & Potumarthi, 2013)

Advanced process measurements can be applied as online, inline, and at-line scale. Online tools are typically measurements where collected sample from process line is conducted to the analyzer and later returned to the process. Accordingly, in-line measurements working principle is based on measurements where sample is analyzed straight from the process stream which can be invasive or noninvasive. Third process measurement in real time is at-line option where sample is removed totally from the main process stream and analyzed nearby to the process line. (Challa & Potumarthi, 2013) Nowadays processes need automatic systems that are able

gather information in real time and reliable optimize process conditions but as properly working online scale process needs to be carefully modelled and methods for this could be used in- or closed-loop together with these measurement options. (Stanke & Hitzmann, 2012) (Folfgang & Macdonald, 2005) Generally this framework is consisted of well-known process where the most critical sources are explained and specified, final products quality can be predicted reliably fast and process variability is strictly controlled and changes for the process are possible to carry out if needed. (Marison et al., 2012)

The reason why online process measurements are needed in several processes is that conventional methods in process monitoring applications require large sample amounts, they have batch delays and methods are labor-intensive. These tools possibilities are rapid testing and sample collecting as well as fast straightforward results to control process more reliable and effective than traditional measurements. (Challa & Potumarthi, 2013) Conventional process analyzers such as pH, temperature and flow meters are not able to provide information of chemical changes occurring the process. Lately several methods for biomass monitoring for spectroscopic and imagining field have been developed working as online, inline, and at-line scale. On-line biomass measurement technologies have advantages and opportunities to be discovered and developed in the upcoming years. Quality control of the process and products whereas time consuming methods can be replaced with spectroscopic methods for measuring chemical and physical changes. Advanced online techniques play critical role in future biorefineries especially investigating pretreatment process changes. (Ferrer et al., 2016) These advance process control (APC) tool function is to control and keep process conditions in constant or limited levels that have been set for the process input system where real time optimization is gathered with multivariate and linear systems. Therefore, APC systems and online measurements in biorefining, petrochemical, pulp and paper industry are used to process stability and material quality control. (Ewanick et al., 2014)

6 SPECTROSCOPIC METHODS AS ONLINE MEASUREMENTS

Spectroscopic methods for lignocellulosic material characterization have been utilized rapidly in the past twenty-first century because they provide benefits that by conventional methods are not able to achieve especially in critical online scale. This development has showed way to effective utilizing fiber optics and valuable biomass process analyses. (Lupoi et al., 2014)

Spectroscopic methods such as near-infrared reflectance spectroscopy (NIRS) and Raman spectroscopy have increased attention in online scale measurement development in recent years. (Ewanick et al., 2014) (Tsuchikawa & Kobori, 2015) NIR has shown potential to be used for lignocellulosic material online measurements as well as Raman spectroscopy could be utilized for continuous monitoring of biorefinery processes. (Ewanick et al., 2014) (Xue et al., 2015) These both technologies provide simply, fast, and non-invasive measurement for lignocellulosic biomass where composition and chemical characterization could be analyzed. (Ewanick et al., 2014)

Suitability of NIR for analyzing biomass was already expressed last decade and after that Raman investigation has achieved footprint. (Sun et al., 1997) NIR and Raman have been studied to be used for example rapid lignin measurements from lignocellulosic biomass when normally measurement is carried out in laboratory scale as kappa number titration method. (TAPPI, 1999) Sun et al., 1997) (Xue et al., 2015) Spectroscopic methods such as UV-VIS and FTIR provide accurate solutions to biomass analysis in lignin field and investigating cellulose and hemicellulose as main components of lignocellulosic biomass (Xu et al., 2013) (Chai & Zhu, 2014) (Ferrer et al., 2016) Accordingly NIR is commonly used for biomass characterization purposes, and it is taking more place monitoring changes occurring in the pretreatment processes. (Ferrer et al., 2016) Figure 9 presents basic principle of spectroscopic measurement system where light radiation is passed through the sample to the optical system and last for the data processing unit. (Xu et al., 2013)

| Radiation/Light source | → | Optical system | → | Sample holder | → | Detector | → | Data analysis |
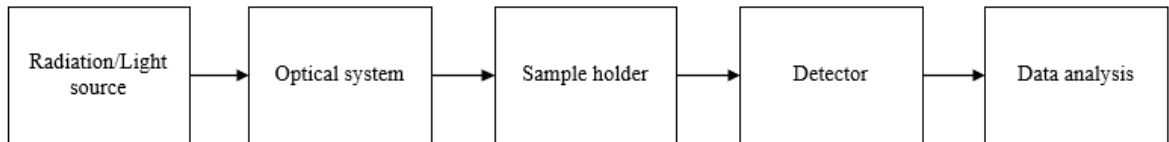
Figure 9.　　Principle of spectroscopic online measurement operation (modified Xu et al., 2013)

6.1 IR spectroscopy

Infrared radiation (IR) causes chemical bonds vibration with stretching or bending and vibrational transition changes are result from changes in dipole when IR-region is active. Infrared spectroscopy is based on molecular vibrations that are active when IR-region occurs for collecting information from compounds. (Kuda-Malwathumullage, 2013) The chemical vibrations are explained with harmonic oscillator model which means that transitions between temporal energy levels in molecules form possible changes in dipole moment. This model is not able to explain actual molecules behaviour entirely like quantum mechanical model that strongly reminds a harmonic oscillator. (Cozzolino, 2015) Fundamental or natural frequency v_0 can be calculated with equation (1) where it is depending on the chemical bonding force constant and the atoms reduced mass. (Kuda-Malwathumullage, 2013)

$$v_0 = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad (1)$$

where v_0 is natural frequency

             k force constant of the chemical bonding

             μ atoms reduced mass

Harmonic oscillator model can be applied to energy by transitions with equation in vibration level with equation (2). (Kuda-Malwathumullage, 2013)

$$E = \left(v + \frac{1}{2}\right)\frac{h}{2\pi}\sqrt{\frac{k}{\mu}} = \left(v + \frac{1}{2}\right)hv_0 \qquad (2)$$

where v is vibrational quantum number

             h Planck's constant

             v_0 vibration natural frequency

In fact, these occurring vibrations in chemical bonds restrict harmonic oscillator model. These chemical bonds break while reaching the dissociation energy as a harmonic oscillator model which can be seen in equation (3). Equation (3) determines the real occurring vibrations in molecules. (Kuda-Malwathumullage, 2013)

$$E = \left(v + \frac{1}{2}\right)hv_0 + \left(v + \frac{1}{2}\right)^2 v_0 X_e + higher\ terms \tag{3}$$

where E is given vibration level

             v quantum number

             Xe an anharmonicity constant

NIR causes two major effects to compound that are molecular overtones and combinations. (Ozaki & Morisawa, 2020) Bending and stretching are common vibrations models and can be further classified into symmetric and antisymmetric stretching modes. Accordingly, bending includes four different movements that are scissoring, rocking, wagging, and twisting. In scissoring mode atoms move towards or away from each other where in rocking mode atoms are moving clockwise or anticlockwise. Scissoring mode atoms move away or towards and during wagging mode atoms move as a V form. While twisting vibration, an atom moves forward and others backward. Temperature and physical changes have straight impact on chemical bond behave and how NIR measure them. (El-Azazy, 2018) NIR effects to chemical compounds behaviour can be modelled as different model vibrations that are presented in the Figure 10.
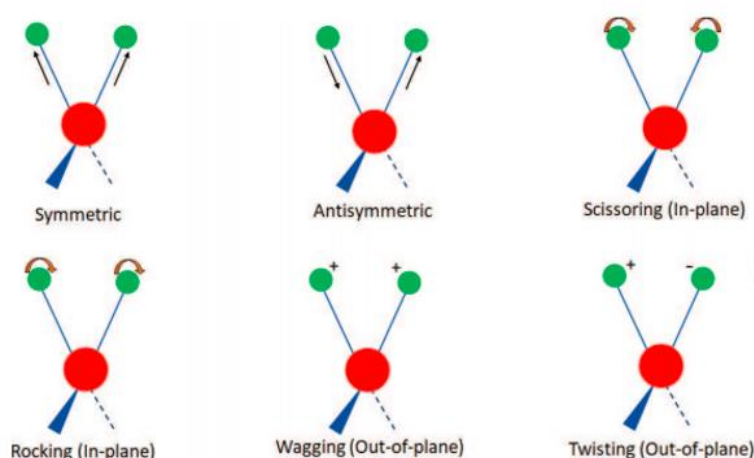

Figure 10.   Molecular vibrations occurring with NIR spectroscopy (El-Azazy, 2018)

6.1.1 Near-infrared spectroscopy

Near-infrared spectroscopy (NIRS) is method which is based on high-energy vibrations in near IR-wavelength areas of 4,000-12,800 cm^{-1} which is called near-IR region. (Pasquini, 2018) (Türker-Kaya & Huck, 2017) (Kuda-Malwathumullage, 2013) This near-infrared region is powerful and very energetic field that is close to visible light spectrum. (Agelet & Hurburgh, 2010) NIR can measure chemical band and fundamental vibrations of N-H, O-H and C-H in the mid-infrared area. (Cheng et al., 2010) Working principle of NIR is based on electromagnetic radiation adsorption where spectra are formed from overtones and combination bands where energy is absorbed with specified chemical bond and single photon is exciting two vibrations. (Kuda-Malwathumullage, 2013) (Agelet & Hurburgh, 2010) (Cozzolino, 2014) NIR spectra bands are broad, overlapped and have lower baseline resolution than MID as mid-infrared bands. However, low absorption makes possible to adjust thickness and analysis depth of the samples. This is huge analytical benefit because straight measurements from scattering samples, high turbid solids or liquids and high adsorbed samples are possible measure with reflectance or transmittance mode without requiring sample pretreatment. (Cozzolino, 2015) NIR regions with specified wavenumber areas and possible applications are presented in the Table III. NIR region was less used than the mid-IR region due to overlapping spectra and weak signal earlier but during 20th century instrumentation and data analysis have developed and gone further that NIR region has achieved more attention in various applications and studies.

Table III   NIR regions with specified wavenumber areas and possible applications (Kuda-Malwathumullage, 2013)

NIR regions	Wavenumber area, cm^{-1}	Applications
Near-IR	12,800-4,000	• Solid • Liquid • Gas
Mid-IR	4,000-200	• Complex solids • Liquids • Gaseous mixtures
Far-IR	200-10	• Inorganic material • Organometallic species

Considering factors of NIR measurements are wavelength range and spectral resolution. NIR chemical vibrations and characterized absorption bands in each wavelength area are presented in the Figure 11 where first, second and overtone regions are showed. There are available

various instruments that are covering all these regions or part of them but for NIR it is typical to use in mid-IR absorptions whereas harmonic changes and overtones represent same information along spectrum. (Streamer, 2013) Overtones of NIR are electron excitations and the all spectra include information structures of methyl C-H, methoxy C-H, methylene C-H, primary and secondary amides N-H, aromatic C-H, O-H alcohols and water, S-H, and some C=O groups. (Agelet & Hurburgh, 2010)
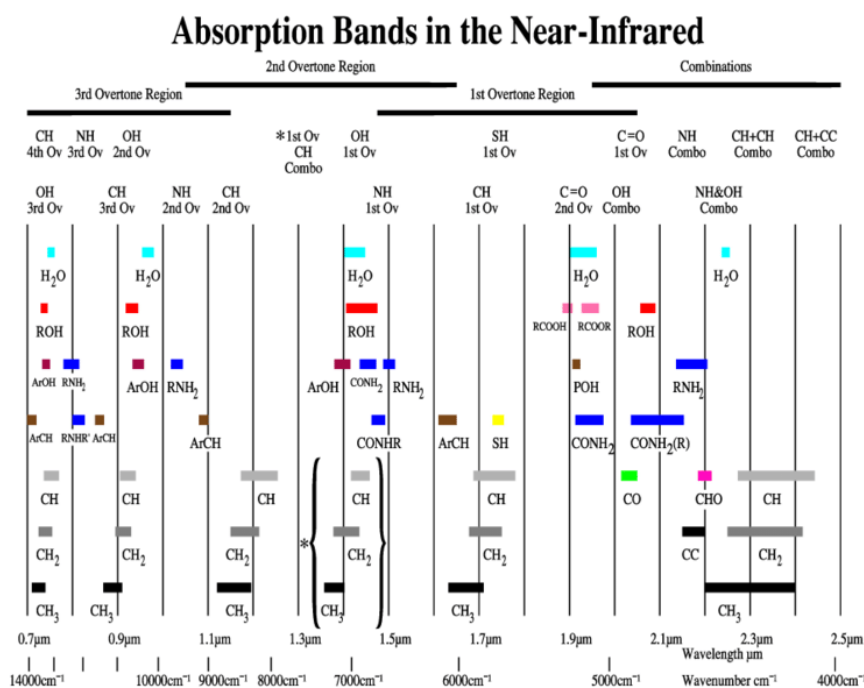
Figure 11. NIR absorption bands (Streamer, 2013)

NIR has been used in different industrial fields and it has over 30 years history as characterization tool for pharmaceutical, agricultural, forestry and chemical applications. (Ferrer et al., 2016) (Cozzolino, 2015) It is method that provides non-destructive and non-invasive way to measure physical and chemical properties for qualitative and quantitative characteristics which are able further build up into online process measurements. (Patel, 2007) Whether NIR has advantages to be used method still requires careful model calibration and multivariate modelling methods to work properly and understand completely the spectral information. Statistical and mathematical tools are usually needed to handle and modify data for example principal component (PCA), partial least squares analysis (PLS) or using numerical methods such as first and second derivates. (Agelet et al., 2010) (Ferrer et al., 2016) In addition

to this NIR requires large reference data and calibration modelling with multivariate methods. In the NIR system, the spectra correlate with sample concentration or feature which is wanted to be measured and there needs to be reference method where dependent variables rely on, and model can be calibrated. Calibration sample set should be wide because the main purpose is to test the created model and its effectiveness. (Agelet et al., 2010)

6.1.2 Near-infrared spectroscopy instrumentation

As an equipment NIR is quite flexible technique that can be modified based on sample features and analysis requirements. NIR main components are a light source which is usually a tungsten halogen lamp, a monochromator, sample holder and presentation interface and a detector for reflectance and transmittance measurement purposes. NIR multi-channel detectors various elements are in rows as diode arrays or planes as charged coupled devices (CCD) to gather several wavelengths at once for analysing. (Cozzolino, 2015) NIR sources that are typically used are Nerst glower (ZnO_2 & Y_2O_3), tungsten-halogen lamps, nichrome wires or Globar (SiC). Usually, detectors are photon detectors made from semiconducting materials that are indium antimonide (InSb), lead sulphide (PbS), indium arsenide (InAs) and indium gallium arsenide (InGaAs). Detectors' working principle is based on interaction of semiconducting materials by NIR radiation. (Kuda-Malwathumullage, 2013) Commonly data is gathered and classified with using transmittance, reflectance or transflectance measurement modes. Reflection is used for granular, solid, and powdered samples whereas transmission is for clear samples. In transflectance mode is able analyse slurries, emulsions, and liquids. The used mode for NIRS is consequently strongly dependent on sample characteristics and optical features as well. (Bart et al, 2013) (Cozzolino, 2015) In transmittance mode the emitted light radiation crosses the sample and emitted light pass through the detector. Accordingly, in reflectance measurements emitted light source is partly passed through the sample or probe and after this it is reflected to the detector. Transflectance mode the light source reflects to the mirror at first and then back to the sample and last to the detector. (Cozzolino, 2015) Two common method principles, transmittance and reflectance are presented in the Figure 12.
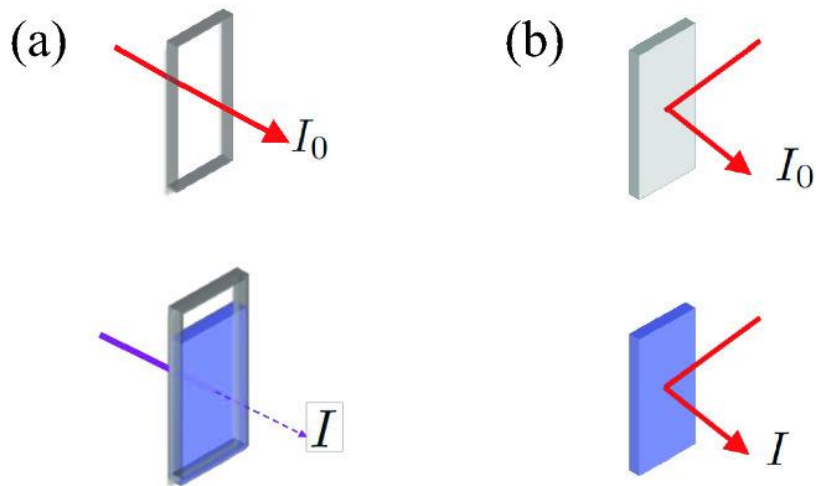
Figure 12.   Schematic of NIR (a) transmittance and (b) reflectance measurements modes (Ikehata, 2020)

For the NIR equipment can be used Fourier transform (FT) instrument for fast data data acquisition and achieving better signal-to-noise ratio (S/N) in measurements. (Kuda-Malwathumullage, 2013) FT-NIR has benefits including instrumental simplicity, high resolution, and measurement accuracy. (Agelet et al., 2010) In FT-NIR the used technology is similar compared to used FT-IR spectrometer systems. There are multiple accessories available to FT-NIR spectrometers for both laboratory and online measurements such as fibre optic equipment. Add to this it is commonly possible to transfer measured data and calibration models from instruments including same sampling unit system which allows easier process implementation. (Streamer, 2013)

6.1.3 Sampling and measurement methods

NIR spectroscopy is suitable for various materials and sample types including suspensions, liquids, solids, pastes, powders, and fibres which leads to that different kind of sampling methods and techniques are applied with NIR. Sampling and measurement methods are strongly dependent on sample type and light transparency. Sample pretreatment and measurement condition stabilizing may be necessary. Pretreatment of samples can be milling, crushing, homogenizing, or cutting and for some type of samples moisture control is very important. However, when considering inline or online analysis, sample pretreatment should be fast and easy to carry out and possibility to get representative sample from the process. Samples are divided into calibration and validation samples that include variation of total sample matrix which leads into that the more complicated process samples are, the more samples are required for creating representative measurements, calibration, and sample set. (Ikehata, 2010)

Transmittance is used for liquid and clear solution samples. The used measurement method is then transmittance liquid sample cell where selection of window material and path length of cells are important factors. This measurement mode needs to be considered if solvents are needed. Cell wall is commonly made of quartz whether limitation of material like this is that it cannot be used in harsh alkaline process conditions. Several options of cell wall transmittance samplers are available at markets nowadays. Second option for cell wall sampler is suitable transflection cell where the transmitted light is reflected from a mirror. Transflection cell material plate can be gold, ceramic, aluminium, or stainless steel. Transflection cell is practical option due to it is easy to clean and repair. Measuring liquid or solutions temperature should be considered during measurement whereas cell can be thermostatted with a holder. For suspensions, turbidity is key factor to be considered during NIR measurements because high turbidity sample that scatter the incoming light and small part is transmitted and therefore diffuse reflection method is good because light goes inside the material. Opposite for low turbidity samples transmittance and transflectance are better options achieving good absorption to spectra. While analysing suspensions with NIR it is good to mix the sample effectively and analyse homogeneous part of samples. (Ikehata, 2010)

Otherwise, while analysing different solid type of samples, diffuse reflectance method is used but for such as thin polymer films also transmittance measurement can be applied. During solid material measurement, avoiding reflection from the surface due to its high intensity compared to its low information from the sample, intrreractance is used during NIR measurements of bulk and solid type of samples. Interactance probe avoid cyclic reflection from the surface of the sample. Add to solid samples, diffuse reflectance method is used for powders, particles, and grains with cells. Before measurement in cells, this type of samples may need homogenizing for ensuring the sample smooth particle size because it has strong impact on NIR measurement due to scattering variation, shape of particles and surface changes. Therefore, stabilization of measurement conditions ensures the quantitative and qualitative analysis as well as reproducibility. (Ikehata, 2010)

6.1.4 NIR probes

NIR probes are commonly divided into online flow through and immersion probes which are installed directly for inline or online measurements with bypass system. Both two categories provide different options and opportunities for process measurements. When using transmission probe liquid is basically illuminated straight, and absorption of sample is measured and for this case, the light passes the medium and then reflects from a mirror. Reflection probes then are backscattering type where light is scattered or detected from sample particles and objects. Accordingly, special type probe is the attenuated total reflectance probe where coming light is reflected totally from the surface of the probe and fleeting light that is interacting with measured solution is finally the signal which is measured. Depth of light source incursion is about a few hundred nanometres to few microns to the sample after total reflectance. Depth of light is dependent on used wavelength and refractive indices. Different type of typical NIR probes is presented in the Figure 13. (Kessler & Kessler, 2020)
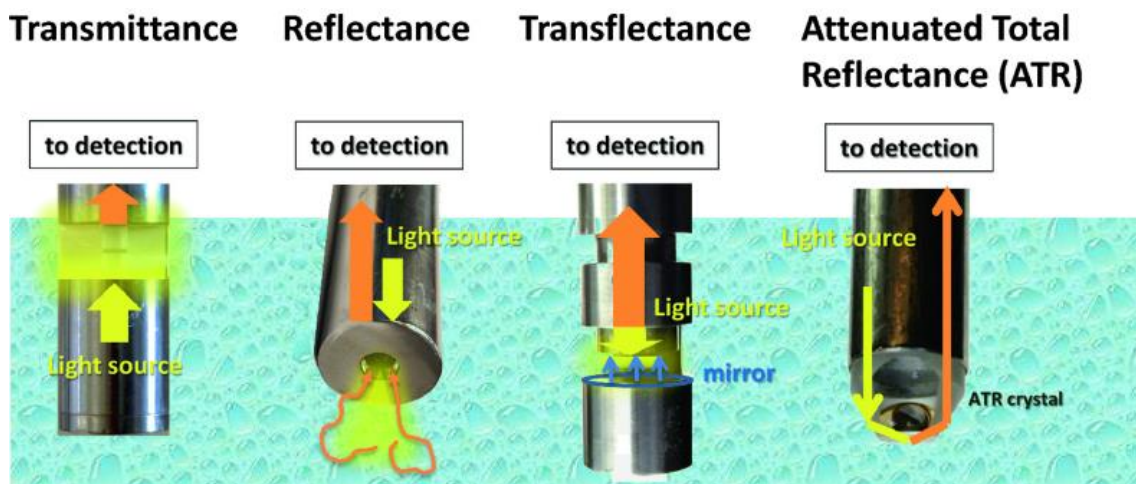
Figure 13.   Different type of typical NIR probe types including transmittance, reflectance, transflectance and attenuated total reflectance probes (Kessler & Kessler, 2020)

6.2 Raman spectroscopy

Raman spectroscopy presents vibrational technique that has been an important technique characterizing lattice and electronic structure for determining the quantitative and qualitative features (Wu et al., 2018) (Ewanick et al., 2013) Raman spectroscopy was found in 1928 and it is suitable measurement for gas, solid and liquid samples where water is invisible unlike in NIR. Raman spectral data contains narrow peaks in large area which enables complex compound analysis where the bands are categorized by frequency, shape, and intensity. Because

vibrational behaviour of each component is unique, Raman provides a fingerprint of them which is strongly dependent on geometric, atomic masses and chemical bonds. (Larkin, 2011) Raman spectra forms when laser is focused to the sample and photons scattering from vibrations inside the molecules can be measured and molecule move into higher energy level (Ewanick et al., 2013) (Lupoi et al., 2014) It is fast and non-destructive measurement, and it is possible to scale up from laboratory to industrial applications with minimum sample preparation. (Ewanick et al., 2014) Raman equipment contains monochromatic light source, optics, spectral analyser which is monochromator or interferometer and detector. Light source is laser which can be UV, visible or near-IR. Multivariate modelling methods need to be used also for Raman spectral data to build up reliable and feasible measurement calibrations. (Larkin, 2011) (Challa & Potumarthi, 2013)

6.3 Fourier-transform infrared spectroscopy

In Fourier-transform infrared spectroscopy (FTIRS) working principle is based on interferometry and for that case it differs from traditional IR spectroscopy. Basically, IR spectroscopy mainly rely on molecular vibrations and collect spectra in the mid-infrared region. (Rees, 2010) (Challa & Potumarthi, 2013) It provides versatile and non-destructive analysis technique for investigating samples qualitative and quantitative organic and inorganic compounds from gas, liquid, and solid phase. (Tucureanu et al., 2016) Each molecular vibration has own characteristics frequency of certain energy level. (Moore, 2017) (Tucureanu et al., 2016) Basic function of FTIR is similar than NIRS but data for the spectra is collected from large area. Sample is directed to the light source with different frequencies compared to conventional monochromatic light whereas absorbed light from the sample is measured. (Challa & Potumarthi, 2013) FTIR needs chemometrics and mathematical analysis tools such as PCA and PLS that are used for treating spectral data and it is fairly new method whereas first IR microscopes are from 1980's. (Ferrer et al., 2016) (Kuda-Malwathumullage, 2013)

6.4 Comparison of spectroscopic methods

Analytical spectroscopic methods such as NIR, FTIR and Raman provide opportunities to be used for lignocellulosic material analysis and these all have advantages but also some drawbacks. The most significant feature in all these techniques is their non-invasiveness and ability for development from laboratory to online process measurement applications. These all methods require multivariate modelling methods for creating representative results that should be considered. (Lupoi et al., 2014) Advanced process control (APC) is possible with these spectroscopic methods for process monitoring and control purposes from spectral data that conventional methods are not able to provide. These high-throughput spectroscopic

technologies have been considered the best option to replace conventional chemical methods whereas NIR and FTIR are better known. NIR, FTIR and Raman have similar features and benefits but differ from measurements principles from each other. NIR can measure samples that include solids whereas FTIR prefers liquid samples. The biggest difference between these methods are spectrum measurement range and accuracy to separate peaks. NIR and FTIR are more water sensitive while Raman is not sensitive for water peaks. Measurements can be modified based on used equipment and considering required method specifications. (Ewanick et al., 2014) (Lupoi et al., 2013) (Cahalla & Potumarthi, 2013) Comparison between spectroscopic methods is presented in the Table IV where advantages, disadvantages, working principle and commonly used chemometric method are mentioned.

Table IV   Spectroscopic methods comparison where shortly presented each method advantages, disadvantages, working principle and commonly used chemometrics method

	Advantages	Disadvantages	Working principle	Chemometric method	Reference
NIR	-Flexible sample preparation -Non-invasive, non-destructive -Fast < 1 min -Online applications -No chemicals or organic solvents needed -Water compatible -Several features from single spectrum -Can be set in dangerous and harsh process environments	-Calibration -Data analysis and modelling -Sample temperature control strict -Large number of reference samples needed -Probe fouling -Changes of operating conditions may affect data and performance	-Reflectance -Transmission	-PLS -PCA	(Bart, 2006) (Patel, 2007) (Bart et al., 2013) (Cozzolino, 2015) (Ozaki & Morosawa, 2020) (Cahalla & Potumarthi, 2013)
FTIR	-No sample preparation -Rapid analysis (3-5 min) -No chemicals needed -Non-destructive	-Weak analytic sensitivity -Probe fouling -Gas bubbles and oscillation may cause errors -Not good for solid samples	-Absorption -Reflectance	-PLS -PCA	(Bart, 2006) (Bart et al., 2013) (Lupoi, 2014) (Sonnleitner, 2012) (Cahalla & Potumarthi, 2013)
Raman	-Small sample size -Non-invasive, non-destructive -Accurate peaks -Fast identification of samples -Online application suitable -Suitable for solid, liquid and gas samples -High resolution -Not sensitive for water	-Fluorescence effect -Validation -Weak signal -Limited area of sampling -Overlapping peaks	-Raman effect	-PLS -PCA	(Bart, 2006) (Lupoi, 2014) (Cahalla & Potumarthi, 2013) (Ewanick et al., 2014)

7 DATA ANALYSIS AS PART OF SPECTROSCOPIC MEASUREMENTS

Nowadays chemical processes provide huge amount of collected data. This requires multivariate modelling process methods to utilize increased data effectively and fast. Routes can be roughly divided into two sections where first is predict quality of the product and second the real time process control. (Kohonen et al., 2007) Chemometrics including multivariate modelling methods is way to use mathematic, computer science and statistics for obtaining information from material system where multivariate regression models are used to treat large amount of chemical data extracting relevant information from collected process data sets. Chemometrics include steps like monitoring, modelling, calibration and controlling. Multivariate modelling methods state of art is to identify and derive information from the process, thereby support the use of process sensors, equipment, and analyzers. These modelling tools provide benefits that safe process operations and real-time information, monitoring and control can carry out for maintaining or obtaining process conditions. (Challa & Potumarthi, 2013) Commonly when applying these methods, predictor variables from the process are collected into one data matrix called X which further is analyzed using multivariate methods. Multi-block methods are option especially when treating NIR data where data gives more information than one matrix X and then data is divided into individual blocks. (Kohonen et al., 2007)

NIR development and history started by Karl Morris applied with multivariate data analysis in the late of 1950's. (Nkansah et al., 2010) Well-working calibration model in multivariate modelling is typical example that is required in NIR applications. Commonly used data modelling tools are correlation analysis, partial least squares (PLS) and principal component analysis (PCA). PCA and PLS are widely used tools in spectroscopy spectral data calibrations and modelling online processes. (Challa & Potumarthi, 2013) (Kuda-Malwathumullage, 2013) These multivariate modelling methods are important and in significant role because large number of data and process variables are collected to online process computers. In addition to these multivariate methods, process control can be carried out by using multivariate statistical control charts (MSPC). (Kourti & MacGregor, 1995) Multivariate data analysis tools can solve in NIR applications overlapping as well as broad spectral peaks, sensitiveness to samples physical variations and large amount of data to create working calibrations. (Agelet &

Hurburgh, 2010) This work is concentrated on using correlation analysis, PCA and PLS for analyzing the NIR data.

7.1 Correlation analysis

The principle of basic correlation analysis is based on linear relationship or regression where certain wavelength area has correlation with measured reference value. Correlation and linear regression represent the most common technique between considered two variables. The aim of correlation coefficient analysis is to reveal positive or negative correlation between laboratory measured reference values and spectral wavelengths. Value close to zero represent that model compounds do not have linear relationship. High absolute values of correlation coefficients indicate high linear correlation. It should be noted when data includes outliers, the correlation decreases, and this could lead to incorrect conclusions. Outliers should be removed, and it should be noted that there can be multiple spectral ranges where correlation occurs. This does not represent an analytical calibration in this work for only one compound rather mixture. Correlation analysis requires independent samples and then narrower wavelength area can be applied to create correlation coefficient in that range. (Bewick et al., 2003) (Challa & Potumarthi, 2013)

7.2 Principal component analysis

Principal component analysis (PCA) is multivariate modelling method which enables interpretation of correlated variable set, compressing data dimensionality and noise filtering. It represents one of the most used linear projection methods. (Challa & Potumarthi, 2013) In PCA modelling, method forms a matrix that is called score matrix (T) where scores include the spectral changes in suitable matrix form to create calibration model where information compression and condensation in matrix X is in critical role to understand model function. (Olivieri, 2018) Principal component analysis shows clusters and outliers as well as other data structures. Linear PCA is applied with principal components that are orthogonal vectors. The more PCA components are used, the more variation is included to the model. In addition to component number, loadings are important factor that generate values which is distance from the origin and those suppose describe how meaningful number of components are in the model. Every component has loading with each wavelength that is modelled with direction cosine between the response variable and principal component. Typical samples are placed near to origin. Increased loadings lead into bigger regression coefficient. PCA method change data to new coordinate system where highest variance is in first axis and accordingly variance gradually declines in successive axis. PCA can separate collected samples to clusters based on spectra variance which corresponds. Especially NIR spectra include many dimensions, PCA

reduce these dimensions into two or three most valuable components which describe the spectra data variation and data is then more understandable and each of these variables can be analyzed separately. Before applying PCA, data is often pretreated with scaling or mean centering methods or other noise filter to pack data smaller component sets to be modelled easily. Otherwise, it is harder to separate dimensions based on several wavelengths. (Challa & Potumarthi, 2013) (Eriksson et al., 1999) PCA working principle and data information matrix (X), score matrix (T), variable loadings (P) and noise errors (E) are presented in the Figure 14.

Figure 14.   PCA working principle where X collected data matrix, T object scores, P variable loadings, E noise and errors, A is number of PCA factors, m number of vars and n number of samples

7.3 Multivariate statistical process charts

Multivariate statistical process control charts (MSPC) are used to model and visualize process condition changes which are caused by simultaneously occurred changing variables. Therefore, these statistical process charts are needed for process control and monitoring purposes in real time analysis. The main objective of these charts is helping to identify possible problems occurring during online process rapidly and give signal that process operators react to condition changes. In practice this means, if there is unrecognized change in the process, used control charts should fast show negative control signal. There is presented many processes controls charts option in the literature that have assumption where processes assume certain probability and distribution limits. Process control charts are not accurate and reliable in all situations, some nonparametric charts have been developed to correct problems. (Koutras & Trianrafyllou, 2020) Commonly used MSPC methods are Hostelling's T^2 chart, squared residuals predicted error chart (SPEC), Multivariate Exponentially Weighted Moving Average (MEWMA) chart, Multivariate Cumulative Sum (MCUSUM) chart, control ellipsoid or χ^2 control chart or PCA

based chart. Part of this work online analysis tests, Hostelling's T^2 and squared prediction error $(SPE_x(Q))$ charts were tested for NIR data. Hostelling's T^2 chart describes the systematic changes among samples that occur in the process. SPEx chart working principle is based on showing changes that are caused by outside factors such as temperature and raw material changes. (Santos-Fernández, 2012) (Sofikitou & Koutras, 2020) PCA is applied when using process charts and route for it is presented in the Figure 15a. Benefit for using these multivariate statistical process charts is that reference values are not needed, only measured spectral data. Visualization examples of used process charts later in this work experimental part is presented in the Figure 15b for Hostelling's T^2 and 15c for SPEx(Q) chart of online NIR measured data. These process charts are used simultaneously, and both are followed during online measurements.
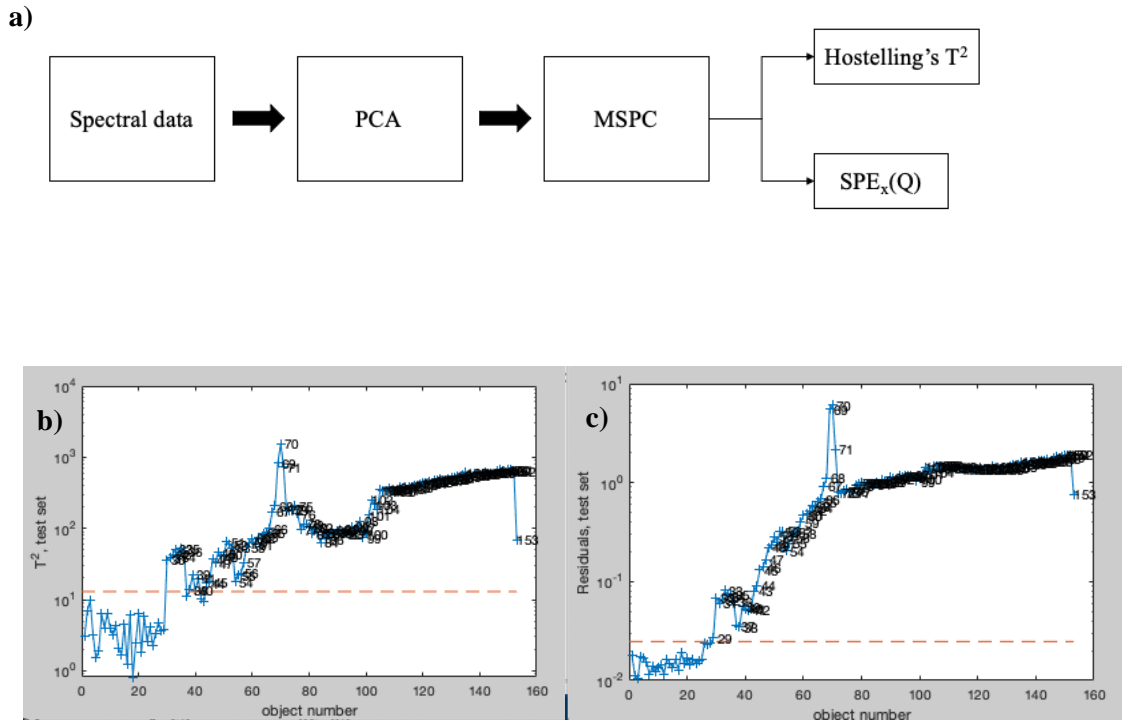
Figure 15.   (a) MSPC route when using PCA method
(b) Hostelling's T^2 control chart example of NIR online measured data of systematic variation
(c) Squared predicted residual error SPEx(Q) process chart example of NIR online data

7.4 Partial least squares

Partial least square (PLS) analysis is one of the most widely known parametric and linear method in chemometrics regression to create powerful predicting models. PLS working principle is based on two matrices that include information, X and Y which are connected. In PLS modelling, X is variable which is measured NIR spectra and Y represents response variable

that is reference values. The solution is rotated to creating maximized covariance between these matrices where covariance represents linear relationship. For example, now when measured NIR spectra is placed in matrix X, lignin and kappa number as predicted values are placed in the matrix Y which is modelled with PLS method. Model loadings are used for calculating the regression coefficients similar way like in normal regression modelling which indicates how well PLS model reference values. To work properly PLS requires enough variable samples and components included to the model. PLS method has been used in analytical chemistry where it is used for multivariate calibration to predict concentrations of new sample's spectral data. (Challa & Potumarthi, 2013) (Eriksson et al., 1999)

8 ONLINE ANALYSES OF LIGNOCELLULOSIC BIOMASS

Lignocellulosic biomass online measurement applications with spectroscopic methods have been investigated in recent years providing rapid and relatively simple analysis whether lignocellulosic material has complex chemical structure. (Lupoi et al., 2014) These applications and biomass sensors generate accurate information of critical process parameters also in hard and challenging process conditions from spectral data associated to chemical changes occurring in process. (Petersen Rønnest, 2011) Process development of NIR, Raman and FTIR applications for monitoring biomass features represent tools to be explored even more in upcoming years and especially effects of major challenges such as temperature, turbulence flow and other variables that are related to instrumentation or external issues are investigated. (Chen et al., 2011) In the last 20 years, online and in-situ measurements have gone further which make possible for non-measurable process variables to be monitored with software sensors and estimators. Online applications are dependent on mathematical and hybrid models which include expert knowledge and the real time data which are used for process monitoring and control. (Schügerl, 2001)

8.1 NIR process application studies

NIR application for lignocellulosic material analysis was discovered last decade where technique development for different kind of applications have started to control process conditions. Especially in paper and wood industry field, NIR has been widely used for density, physical and chemical analyses such as moisture, char, and ash. NIR has been applied for wood structure measurements from chips and bark where Klason lignin, extractives and size

distribution were analysed. Lignin investigation with NIR technique has been popular and results have showed correlation with coefficient values more than 0,80.

Jiang et al. (2010) investigated coniferous biomass lignin-carbohydrates from southern pine samples with NIR. Spectral data was collected from 21 solid wood samples. Samples were ground into different particle levels 3 mm, 20-mesh, 40-mesh, and 80-mesh with Wiley mill to original, air- and oven-dried and delignified samples in ambient conditions. FT-NIR measurement was tested for determining Klason lignin, extractives, cellulose, holocellulose and hemicellulose contents where holocellulose represents sum of the monosaccharides. Reference values behind the calibration model were based on basic NREL standard measurements with HPLC to measure monosaccharides and wet chemical method to measure lignin. Spectrum range was between 10 000 – 4000 cm^{-1} with resolution of 4 cm^{-1} and average spectrum measured scans were 32. Collected NIR spectra for this Jiang et al. (2010) study was based on diffused reflection and pretreatment method for spectral data was first derivate which was discovered to model and decrease the spectra noise well. For NIR model 16 samples were used as calibration and 5 as prediction set for creating model for each compound with PLS chemometric method at selected spectrum range. This showed best suitability for modelling wood samples. Result and findings of this study were that particle size has effect for NIR measurements. Calibration was more precise when particle size decreases especially when samples particle size of 80-mesh was used, NIR calibration results improved significantly. Models for holocellulose and lignin fitted best to the created model and accordingly, cellulose and extractives fitted good.

NIR has been used for biomass structure analysis and investigating structural changes in lignocellulosic biomass in enzymatic digestibility. (Xu et al., 2013) Xu et al. (2013) investigated both FTIR and NIR techniques in published review study where PLS showed the best ability. The composition analysis of lignocellulosic biomass with NIR included lignin, xylan and glucan. Biomass residues were studied by Fitoussi et al. (2011) where NIR capability was studied for analysing biomass features such as cellulose, lignin, moisture, ash, and extractives from pretreated sugarcane bagasse with ammonia. In the Table V is presented NIR applications for lignocellulosic biomass where each compound wavelength and correlation with cross validation results are summarized. From the cross-validation results can be noticed that good

and high correlation values achieved from biomass analysis with NIR from different kind of compounds. (Ferrer et al., 2016)

Table V   NIR wavelength and cross validation examples for biomass analysis (Ferrer et al., 2016)

Compound	Wavelength, nm	Cross validation
Lignin	500-2400	0,77
Klason lignin	1333-2239	0,89
Cellulose	1299-2439	0,93
Hemicellulose	1299-2439	0,91
Xylan	1000-2500	0,88
Galactan	1000-2500	0,88
Glucan	1000-2500	0,94
Lignin S/G	1100-2500	0,74
Glucose	500-2400	0,88
Xylose	500-2400	0,76

NIR as a spectroscopic measurement has been one of the biggest developments recently because this kind of analytics provides rapid biomass feedstock screening and promise lower process costs in the long run. (Ferrer et al., 2016) Add to wood-based industry NIR has been applied in pharmaceutical, cosmetic, food industry and petrochemical fields. Especially in petrochemical industry NIR is promising technology to replace conventional methods to measure fuels, hydrocarbons, different petroleum fractions and derivates such as polymers. From wood industry sector NIR has been used widely for different measurements for quality and consistency of structure especially in milk product processes whether NIR is still quite new method from solid wood analysis or prediction of chemical composition.

Zhou et al. (2015) predicted mixed carbohydrate content and mixed hardwood lignin with ATR-FTIR and FT-NIR. Sample set contained different wood species that were cotton wood, eucalyptus, aspen, and poplar with total amount of 37 samples. Analysed wood samples were set in ambient conditions at 23-26 ℃ with moisture content less than 50 % and thickness of 3 mm for three weeks and the last they were ground with Wiley mill into 80-mesh and dried for spectral measurements to homogenize samples. The used NIR equipment was PerkinElmer 400 FT-IR/FT-NIR spectrometer in the range 4,000-10,000 cm^{-1} with 32 scans and resolution of 4 cm^{-1}. Spectrum Quant software was used for creating PLS and PCA multivariate models. Results by Zhou et al. (2015) showed that ATR-FTIR is good for screening and FT-NIR can be

used as quality monitoring purposes. The best predictions from model were built for lignin, extractives, and xylose.

Research and studies have also showed that with NIR is possible to measure wood degradation process, density and wood stiffness that are important especially in solid timber applications. He & Hu (2013) investigated with Bruker FT-NIR that is possible to predict chemical composition of 116 different wood species with resolution of 8 cm^{-1} and spectrum range between 12,800–3600 cm^{-1} using NIR spinning cup module sampling during measurements. Different wood samples were chipped and ground with Whiley mill with integrated 1 mm sieve further into 40-60 mesh. Analyzed samples collected as retaining on the 60-mesh sieve for chemical and spectral analyses. Based on their study cellulose and hot-water extractives in wood samples achieved the highest R^2 values that were 0,962 and 0,963. Pentosan results were good with value of high R^2 0,98. Outcome of this study was good model for predicting these values in different applications and excellent data fit. Tsuchikawa & Kobori (2015) stated that density of wood material can be predicted based on wood components such as cellulose, lignin, and hemicellulose with NIR. Other factor for measuring wood density is that it can be predicted from material moisture content. NIR absorptions 7000, 7160 and 7320 cm^{-1} are in significant role when measuring wood material density.

Wolfrum & Sluiter (2009) studied correlations and multivariate calibration models between NIR data to corn stover feedstock and dilute acid pretreated corn stover. NREL methods was used to measure lignin and carbohydrate content of calibration and validation samples. Pretreated samples were first washed, dried, and finally milled smaller particle size down to 1 mm with bench-top mill. Non-structural material was removed from feedstock samples with water and ethanol before two-stage acid hydrolysis with sulfuric acid. Biomass samples fractionated into solid and hydrolysate liquor, which was consisted of sugar degradation products, acid-soluble lignin, organic acids, protein, monosaccharides, and some part of ash. Solid phase of the sample was consisted of acid-insoluble lignin (Klason), ash and some protein. HPLC was used to analyse hydrolysate samples and UV-VIS for acid-soluble lignin. Acid-insoluble lignin measured gravimetrically.

NIR spectra was collected with Foss NIR Systems model 6500 Forage Analyzer together with transport reflectance module set. All samples were air-dried to 10 % moisture content and measured 32 spectra in range 400-2,500 nm which were averaged. PLS method was used for building calibration models with WinISI v1.50 (Inrasoft International) and Unscrambler for PLS-1 and PLS-2 models. In PLS-1 model reacted spectral data variables against the independent variables whereas PLS-2 model principle was to react directly for predicting more than one constituent. Calibration models and results were created for lignin, glucan, xylan, arabinan, mannan and galactan. In calibration model, 28 samples were used as validation samples to test model performance and standard normal variate (SNV) and first derivates were used to pretreat the model data. Based on Wolfrum & Sluiter (2009) this study results showed that when combining these two different type samples for single calibration was not successful and therefore acid pretreated and original corn stover samples needed own calibration model to predict reference values. Results showed that dilute acid pretreated and feedstock corn stover NIR measurements provide fast biomass composition characterization where different mathematical pretreatments of spectra did not show statistically significant different to results. However, calibrations need to be developed where larger number of samples are included to original calibration set. This study proved that NIR calibrations are correctly used inexpensive method applied for various materials. (Wolfrum & Sluiter, 2009)

Ye et al. (2008) studied fast classification and compositional analysis of corn stover fractions with FT-NIR techniques where objective was to determine different botanical fractions composition. Analyzed samples were collected six months beforehand and storage to less than 50 % moisture content and separated into leaves, nodes, internodal piths and rinds, husks, and sheath. Sample's storage in ambient conditions in 20-30 ℃ and before FT-NIR analysis samples were dried at 105 ℃ and cut smaller 0,5-1,0 cm pieces with Wiley mini-mill and passed through the 40-mesh screen and finally cooled to room temperature. Reference values were measured based on NREL standard to lignin, ash and carbohydrates which were analyzed with HPLC. Acid-soluble lignin was measured with UV-VIS with spectra range 230 nm and other fractionate, insoluble lignin with gravimetric analysis. Correction values for C5 (xylose and arabinose) was used value of 0,88 and for C6 sugars (mannose, glucose, galactose) 0,90 to adjust the concentration for these sugars from dilute acid hydrolysis reference method. Spectra of samples were collected in range of 4,000-10,000 cm^{-1} with 8 cm^{-1} resolution and 64 scans using diffuse reflectance mode. PLS regression method was used for calibration model and validation sample testing were noted that cross-validation is the best suitable option when used

data set has variability. PCA was applied to separate variables into different fractions with using two components towards increasing sugar content. Highest fractions were sheath and husk.

Ye et al. (2008) study showed that by NIR is able analyze different corn stover fraction effectively whether chemical composition strongly varies in different part of plant fractions. PLS model showed good quality results compared to AACC standard for xylan, glucan, lignin, mannan and ash. Benefits of NIR application can been seen in the Figure 16 where wet chemistry and FT-NIR are compared. Wet chemistry methods require more workload than NIR measurements of ash, sugars, and lignin but however wet chemistry methods are required for creating the calibration model first. In the future focus could be focusing on improving these wet chemical methods for calibration and spectral pretreatment. Future studies should consider reducing physical interference and wider range in used calibration samples to balance both model robustness and required time for model development but also costs. NIR provide instant feedstock information of lignin, sugars and ash content which is benefit for example biomass feedstock suppliers and bioethanol producers. (Ye et al., 2008)
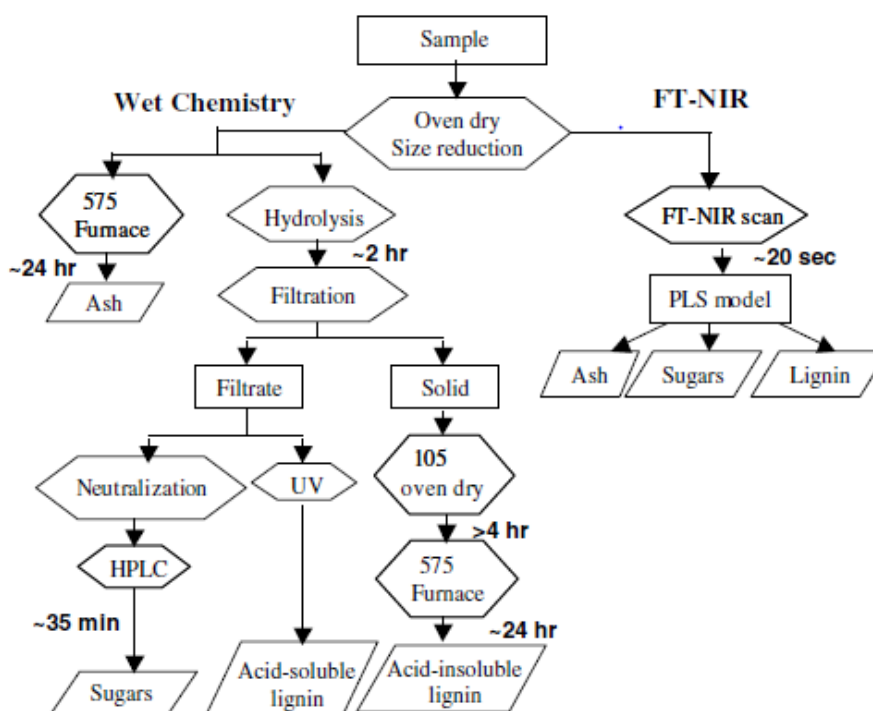

Figure 16.   Process steps for original wet chemical analysis and FT-NIR analysis route (Ye et al., 2008)

Liu et al. (2010) studied from the same aspects corn stover and switchgrass chemical composition variations with FT-NIR technique with chemometric models. The used FT-NIR spectrometer was Excalibur 3100 including diffuse reflectance probe. Analyzed spectral range was 4,000-10,000 cm^{-1} with 64 scans. Studied reference values were glucan, lignin, xylan, arabinan, galactan, mannan and ash that were measured by wet chemical analyses as two-stage hydrolysis and HPLC. Calibration results for corn stover, galactan, mannan and ash achieved AACC requirements and industrial levels. On the other hand, corn stover's glucan, xylan, arabinan and lignin models were good quality control meters. Accordingly, chemometrics model for individual switchgrass can predict both lignin and mannan for screening and other variables as quality control. Based on that study presented models are suitable both materials with individual models and FT-NIR is promising for biomass chemical composition characterization to be developed multiple biomass species. In the future, it would be good that model would include more biomass feedstock species due to model robustness and wider industrial applications.

Several herbaceous and wood-based lignocellulose feedstock materials were studied by near NIRS in laboratory scale creating calibration models to 121 samples that included sugarcane, switchgrass, corn stover, hybrid poplar, black locust, and American sycamore. Collected wood samples storage outdoors for 26 weeks as large pieces and corn stover and switchgrass for 3, 6, 13 or 26 weeks. Before NIR measurement with Pacific Scientific 6250 monochromator, samples were cut into 2 mm with shear mill and further with cyclone mill to 1 mm to achieve homogenous. Spectral data was collected in range 1,100-2,500 nm with using 64 scans in rotating spinning cup module. Samples that used were analyzed based on NREL reference method for uronic acids, arabinose, xylose, lignin, ash, mannose, galactose, carbon, nitrogen, hydrogen, and oxygen contents. All samples were not used for the creating calibration thus 20 of them were used as validation samples to test calibration model. Calibration model was created with using PLS method and 20 of the samples were randomly chosen to be used as validation set to test model capacity and robustness and some outliers from data set were removed. This study proved that NIRS is capable to analyze various biomass feedstock material with good results. Calibrations concentrating on narrower models including only feedstock materials such as wood and herbaceous, might enhance precision. However, it seems that calibration including this kind of diverse feedstock materials can be used to predict composition and that way minimize calibration costs. (Sanderson et al., 1996)

In biorefinery the pretreatment step is the most critical point of the process and often slurries are formed that contain liquid phase with solubilized components and solid phase with insoluble compounds. Due to these two separated phases, wet chemical methods require more time and phase separation into individual two phases. Sluiter & Wolfrum (2013) studied NIRS and created calibration model without separation to predict the solids in situ using the whole spectra of analyzed corn stover slurry samples. They have investigated that NIRS is capable to predict solid phase chemical composition after separation, washing and drying steps. It is noted while removing the separation step, time savings are achieved which could help the process scale analysis of biorefinery pretreatment step. Liquid phase of the sample contains xylan derived from hemicellulose, sugars such as glucose, arabinose and mannose but also solubilized fraction of lignin and other sugar degradation compounds. The solid phase was composed of cellulose and parts of lignin and hemicellulose that are not dissolved into liquid phase. Constituents such as protein and ash are included in both phases.

The objective of Sluiter & Wolfrum was to present a calibration model for pretreated, washed, and dried slurry solids from corn stover pretreatment. Spectral data was collected using Foss NIR Systems XDS analyzer with reflectance probe with 32 scans and the whole slurry spectra was measured. Reference data results for solids were based on NREL method where two-stage hydrolysis used to determine lignin, carbohydrates, and acetate. Ash and moisture content measured gravimetrically. Temperature effect noted while measuring the NIR spectral changes. Temperature for the slurries was stabilized at 30 ℃ in water bath and homogenized with hand mixing. Spectral data was collected after this using a Foss transflectance Cam-lock cup. Based on this Sluiter & Wolfrum (2013) study the slurry calibration turned out quite good apart from large association part of the spectrum with water parts. When removing these water absorption sections from the calibration model, higher RMSEV values and lower correlations achieved. It was also noted that pH might affect slurries and biomass spectra. There could be some chemical variation among samples when solids are separated and further dried to wet chemical analysis but slurries including total solids chemical composition might vary. This showed that one pitfall might be when analyzing highly acidic slurries because it might affect how transflectance cell works. (Sluiter & Wolfrum, 2013)

Final slurry model was consisted of acid-pretreated, steam pretreated, acid-impregnated and base-deacetylated samples. Model included 262 calibration samples and 19 validation samples

which were removed from the calibration model. Reference sample results varied especially for lignin from 10,6 to 32,5 and for glucan concentration was between 29,7-80,2 and some variables constituent level started from zero whether this variation did not have specific negative impact on results. Sluiter & Wolfrum (2013) study showed that in situ measurement for solids with NIR is promising whether model was not like the same for washed solids and dried solids from uncertainty and correlation aspects. These solids fractions need to be measured with using separate calibration models.

Multivariate data analysis and NIR were applied when Nkansah et al. (2010) studied yellow poplar and red oak chemical and physical properties. In their study long term goal was to investigate that NIR could be in the future applied in biomass feedstock analysis and biorefinery processes to ensure quality aspects and material characterization to monitor the process conditions based on that and reduce costs. Reference values that were studied with NIR were total lignin, ash, extractives, acid-insoluble lignin, holocellulose and bulk density that were measured using traditional wet chemical standard methods. The used NIR equipment was Bruker Matrix-F FT-NIR spectrometer including a fiber optic probe for liquid and solids samples. Probe was set perpendicular to the sample surface at a 5 mm distance. Spectral data was analyzed with Unscrambler MVDA software. Models were built covering all the NIR spectral range but for selected spectral areas where full spectral model provided better results because R^2 improved, and RMSEP/SEP values were smaller. Outcome of this study showed that acid-insoluble lignin did not model effectively ($R^2 < 0,40$) compared to total lignin and other variables that were fitted with good results. Using first derivative pretreatment for spectral data enhanced the model and predictiveness. Final models showed quite high correlation and created PLS models with separated spectral range in 1,300-1,800 nm and including all spectra in range of 800-2,400 nm results were quite similar. Reduced spectra region results for all chemical and physical variables with value of $R^2 < 0,77$ and except total lignin $R^2 < 0,66$.

Spectroscopy has potential to be used as valid and analytical technique for physical and chemical material composition analysis by the standard limits. This leads into that method could be utilized as online applications due to lack of conventional methods problems to respond rapid online analysis need. (Mancini et al., 2019) Mancini et al. (2019) studied 94 wood chip and pellet samples to predict moisture, ash, and calorific content with multivariate modeling methods. These samples were dried and ground into 1 mm to ensure sample being homogenous

before spectral analysis. Spectra collecting was carried out with Buchi online NIR equipment. Measurement range was 400-1700 nm with 32 scans with resolution of 7 cm^{-1}. Used instrument in this study based on diode-array and it was placed like rotating module to simulate real pellet process. Results of this study showed that online application could be possible for predicting these values, but calibration model needs to be increased validation set before online implementation. Moisture content and calorimetric values of wood chips were predicted correlation value of 0,97 and 0,92 whether compared to ash content that did not give successful results. This study showed that implementation directly from the laboratory is not easy rather it requires modelling, but laboratory measurements give base to create online applications.

All in all, NIR potential upcoming years in forest-based industrial applications is huge. These industries have grown in recent years and production capacities are bigger which leads into that feedstock material characterization rapidly and reliably is more important to guarantee quality and production. Further development should be done investigating NIR working in hard process conditions and because of this technique has its development path to follow and more improvements should do on noise reduction to reduce tune-up timing and increase method robustness. Development like this will lead into measurements which handle unstable industrial environment and deal with raw material fluctuations to evolution of continuous process measurements. Sample preparation is one development stage which will be in key point especially in online measurements. Now presented studies are mainly based on laboratory scale and publications related to pretreated lignocellulosic wood biomass are limited. Sample preparation which was used in these studies was strongly based on homogenizing the sample, stabilizing moisture content, and eliminating particle size variations. Common for these all are that conventional reference method is based on NREL guidelines and total spectrum range is used for model calibration.

8.2 FTIR and Raman applications

FTIR and Raman are techniques that could be considered characterization of lignocellulosic biomass. FTIR has history also in the food industry, but it has shown potential to be applied in biorefinery applications for online measurements and in-situ monitoring. FTIR and Raman are both new especially in online process measurement scale compared to NIR but have opportunities to be discovered more because both methods provide chemical information that conventional methods are not able to measure in online scale. (Xu et al., 2013) (Ferrer et al., 2016) FTIR has been used for example in fermentation processes, investigating enzymatic

hydrolysis, cellulose fractions, and crystallinity. FTIR advantage is its in-situ measurement and it has been used for example fatty acid component measurement which normally is very difficult. For different lignocellulosic material pretreatment processes FTIR provides chemical compounds analysis where for example lignin structure can be easily followed while lignin and hemicellulose both are at the target as intermediate or final products in biorefineries nowadays. Raman is more unknown technique for biomass analysis and narrowly research studies are published related to its utilization for biomass samples laboratory or online scale. (Ferrer et al., 2016)

Pretreatment process of biomass has been investigated with FTIR for characterizing pretreated slurries and dissolved components in aqueous solutions. Limitation to use FTIR has been the aqueous samples water peaks strong absorbance effect. However, attenuated total reflection (ATR) has provided way to control water absorbance disturbance that measured spectra will not be totally covered by water peaks. An example of FTIR usage was studied to measure sugars in dilute acid pretreated biomass liquors whereas xylose, mannose, glucose, galactose, lactid acid and 5-hydroxymethyl-2-furfural (HMF) were measured in soft- and hardwood slurry samples. PLS method was used to compose HPLC wet chemical reference data to predict values with TQ Analyst Software. (Tucker et al., 2000) (Lupoi et al., 2014) Used regions for PLS model were 1500 -830 cm^{-1} to sugars, 1683-84 cm^{-1} to acetic acid and for HMF 1562-1503 cm^{-1} including 23 samples of liquors. Same aspects were used for 18 washed samples for pretreated solids. Samples that were analysed in this study consisted of chipped trees, limbs, needles, and bark.

An another FTIR measurements for process monitoring were carried out for pretreated yellow poplar slurry. In this study was used an ASI Applied Systems ReactIR 1000 equipment and six-reflection diamond-composite Hastelloy C-276 insertion BioProbe with 64 scans and resolution of 8 cm^{-1} in neutralization when lime is added into reactor to adjust the process pH level. Reactor slurry included 16 % insoluble and 19 % total solids. Spectra from liquor samples was analysed with a six-reflection diamond-composite ASI DurasampIIR cell included to Nicolet Avatar 360 spectrometer. Measurements included 512 scans at resolution of 2 cm^{-1}. Washed solids were analysed using Nicolet Impact spectrometer with 512 averaged scans and resolution of 8 cm^{-1}. This study showed that FTIR applied with PLS regression can predict these variables in biomass from solids, liquors, and slurries including both dissolved and insoluble solids.

Based on results in this study was noted that when predicting washed solids with FTIR needs accurate lignin and glucan reference values. Hastelloy probe can predict process changes in reactor where monitoring of the process is fast and probe material is suitable for harsh process conditions. In the future models for individual samples should be developed for FTIR. (Tucker et al., 2000)

Sills & Gossett (2010) investigated enzymatic hydrolysis saccharification changes and alkaline pretreatment of six different lignocellulosic biomass samples with FTIR and PLS modelling where Bruker Optics Vertex 80 FTIR spectrometer was used for measurements with 64 average scans in range of 4,000 – 400 cm^{-1} with 2 cm^{-1} resolutions. Compositional analysis was made for glucan, xylan and Klason lignin in solid samples based on reference values measured followed by NREL standard method and high-performance liquid chromatography (HPLC) before and after pretreatment step. Result based on their study was promising whether models were constructed only from six FTIR spectra. Study main outcome was that PLS with FTIR measurements raw biomass samples from pretreatment and enzymatic hydrolysis is possible and able to carry out successfully. Add to FTIR Raman spectroscopy is possible option for lignocellulosic biomass continuous monitoring. (Ewanick et al., 2014) Lupoi & Smith (2012) investigated by Raman at 1064 nm lignin content in various woody lignocellulosic and herbaceous biomasses and lignin monomers G and S. The used multivariate modelling methods were PCA and principal component regression. This study showed Raman as alternative technique to analyze biomasses and provided information with high resolution and rapidly. (Lupoi & Smith, 2012)

SUMMARY OF THE LITERATURE PART

Based on literature and research studies related to NIR technology, published studies are made in laboratory scale in ambient conditions. Samples are homogenized and particle size effects minimized as well as moisture content and sample temperature effects. Biomass reference method that is based on NREL guidelines are in the center. Online applications or studies for biomass analysis are still quite limited available, however NIR has been tested for several materials to replace old conventional methods to reach accurate results faster.

EXPERIMENTAL PART II

NIR feasibility and online measurements possibility for pretreated lignocellulosic beech wood biomass samples from pretreatment process was tested in thesis experimental part. Experimental tests were carried out first at laboratory scale as pretests for online loop measurements to test online process control possibilities of NIR. Online loop tests composed of different trials where process conditions were changes and spectral data collected from online setup continuously. Sample sets that were used in experimental part are presented in the Figure 17. First sampling set is original biomass (O). Adding water to sample O, new samples such as washed solids (W) and slurry (S) can be created. Washed solids do not include insoluble fraction which is removed by two-staged re-slurring filtration. All these four fractions were tested with NIR and created individual calibrations. These fractions are possible in real process environment, but online testing objective was to plan measurement that would be fast, reliable, and easy to carry out in process conditions. Therefore, slurry was chosen as sample type for online loops (OL) tests. Final tests were done by online loops where slurry was measured as inline NIR measurement to test process scale measurements with different modifications.
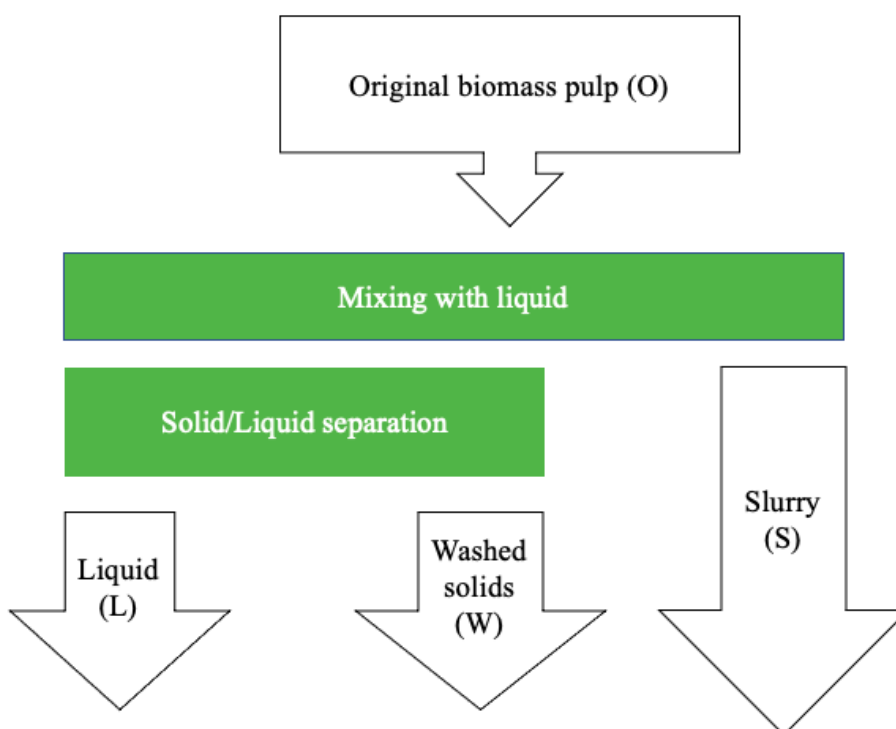

Figure 17. Sampling sets for pretreated lignocellulosic biomass in experimental tests including original biomass pulp (O), liquid (L), washed solids (W) and slurry (S)

Laboratory tests as pretests included these previous presented O, W, L and S sampling sets. Sensitivity for different process changes such as moisture and temperature were tested for original pulp and slurry samples. These results are discussed in more details in the chapter 13. OL tests included six different inline slurry tests and five of them are presented in this work. Sample sets were collected to capture different variations from time span of three years. In the Table VI is collected samplings sets and amount of reference parameters for laboratory and online scale tests. Reference parameters that investigated with NIR were lignin and kappa number.

Table VI   Experimental sampling set setup, parameters and sample amount for laboratory and online scale tests

Sample set	Parameter number	Sample amount
Laboratory scale		
Original biomass (O)	2	28
Washed solids (W)	2	28
Liquid (L)	2	28
Slurry (S)	2	28
Online scale		
Online Loop tests (OL)	- based on laboratory measurements - taken samples - changes in measured online spectrum	- 6 loops total - 4 loops results presented

Based on literature research and latest publications presented in the literature part considering NIR, successful measurements and calibration models are dependent on process conditions, sample type and its chemistry. This means that always the spectral data correlation is not linear to reference value. For this biomass type samples investigated this part, affecting process indicators are raw material type, moisture, soluble content, particle size and temperature. Special feature of investigated biomass is that it includes content of 30 % insoluble and 70 % soluble components. Changes are occurring in both phases which create some challenges to gather reliable information, so neither cannot be eliminated. The main objective was to find out NIR suitability to monitor pretreatment process variations. NIR feasibility tested from aspect if it can control occurring changes but also process indicators and this biomass complexity features such as insoluble content changes. Indicators like these do not affect process concentrations itself but these have straight impact on measured NIR spectra and its quality. Especially temperature and moisture content variations had big effects for calibrations and validation samples. Therefore, in experimental part was tested how impact of insoluble fraction

changes and moisture variation of samples could be minimized to achieve reliable NIR measurements when taking these into account during calibration process. Process indicators affecting on NIR spectra and sampling sets are presented in the Figure 18 which formed hypothesis base to be investigated in experimental part.
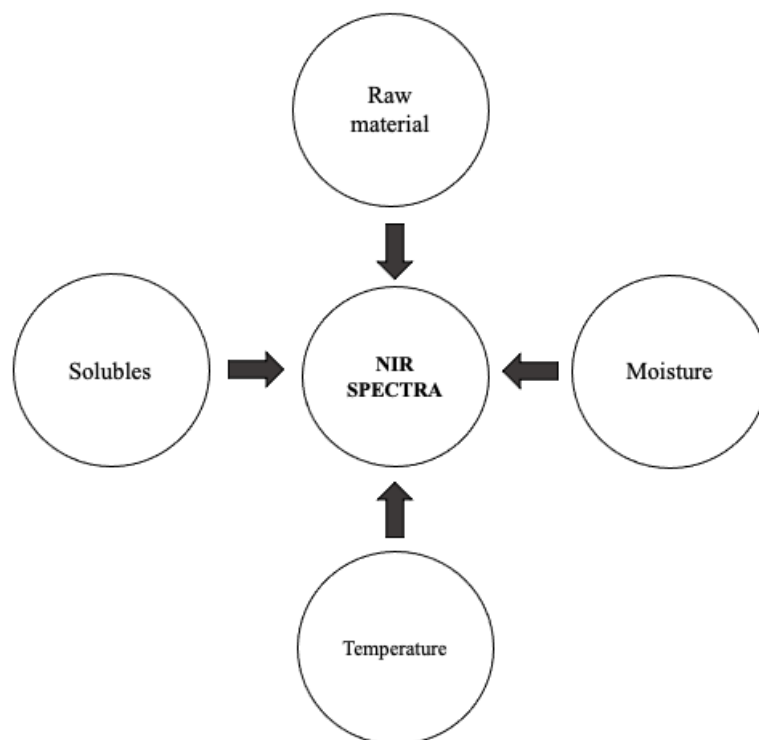

Figure 18.   Process indicators investigated in this study affecting on measured NIR spectra

9 REFERENCE METHOD AND SAMPLE SELECTION

NIR applications and new method development requires representative calibration sample set. There are not available straight instructions the number of samples that good calibration include based on literature. It is better to include several samples into calibration with great capture, but it can be built for 20 to 30 samples. This is sample amount that would be good to include calibration model at first. After first calibration, model can be updated including more samples over time. An ideal calibration model set will include the chemical as well as physical and spectral information of samples and avoid coming extrapolations while predicting new samples. It should be noted that NIR calibration is totally based on reference values and measurements and therefore suitable laboratory measurement should be chosen carefully. Also, calibration set might include multiple outliers and removing of them is not always easy because model should

include after removing certain samples still similar information of samples that are removed and this way is possible to ensure model goodness and representativeness. (Agelet & Hurburgh, 2010) Therefore parameter range needs to be adequate which is included to calibration set and model captures this.

The used laboratory reference method for measure total lignin was based on NREL standard guidelines for monomeric sugars and total lignin for dried and grinded biomass samples from sample set O. Biomass samples kappa number was measured based on modified ISO 302:2004 method. NREL is international and widely used measurement guideline for structural carbohydrates and lignin, but experimental study results only for lignin and kappa number are discussed from this part. Method is based on two-step acid hydrolysis with sulfuric acid to fractionate biomass samples that are not containing extractives. During this measurement, biomass samples containing lignin forms two different phases that are acid insoluble and acid soluble lignin. The latter is measured with UV-VIS spectroscopy. Acid insoluble lignin is measured gravimetrically. In acid hydrolysis polymeric carbohydrates are hydrolysed further monomeric and after this these are soluble in the liquid fraction. This part of biomass is analysed with HPLC. (Sluiter et al., 2011) Accordingly, biomass kappa number measurement is based on titration with potassium permanganate solution consumed by one gram of dried pulp. Kappa number was measured from washed pulp samples (W) where biomass samples containing insoluble fraction is washed before analysis that insolubles do not disturb measurement accuracy. (TAPPI, 1999) Variation range for lignin parameters were between 34-37 % and for kappa number 38-70.

10 NIR EQUIPMENT AND MEASUREMENT SETUP

The used equipment at first in the laboratory scale measurements was Thermo Scientific Antaris II FT-NIR-analyzer where used light source was HeNe-laser and InGaAs detector. This equipment included separated rotating sampling cuvette and liquid sampling transmission module, but in this work only fiber optic probe was used. After laboratory measurements, four-channel Thermo Scientific Antaris MX FT-NIR Process Analyzer was used for OL tests. Online NIR equipment was similar compared laboratory model, including the same detector but it was more robust and suitable for online testing with different inline probe. Used light source in online model was halogen. Biggest difference between these NIR equipment was that online

model included internal background correction. NIR equipment used during this experimental part is presented in the Figure 19 where (a) is laboratory NIR and (b) online-NIR.

Figure 19.   (a) Thermo Scientific Antaris II FT-NIR-analyzer for laboratory measurements
(b) Four-channel Thermo Scientific Antaris MX FT-NIR Process Analyzer for online loop tests

10.1 Probes

Selected fiber optic probe for measurements in laboratory scale was SabIr diffuse-reflectance probe which can analyze samples directly or indirectly through materials based on reflection. Reflection probes include two fiber optic cables where first is for illumination and other for detection. Transmission probes usually have only one optic cable. This probe covers spectral range of 12 000- 4 000 cm^{-1} (833-2500 nm). Length of probe was 15,8 cm and diameter 1,6 cm. Window material was high-quality sapphire which is resistant for strongly chemicals and material of probe stainless steel (316). SabIR probe is designed to measure solids and powder samples. Therefore, it is not straight suitable for liquid and transparent sample analysis, however there is available adapter that can be attached to the probe for this purpose. This probe was used in laboratory measurements for sample fractions O, W, L and S. SabIR diffuse-reflectance probe is presented in the Figure 20.
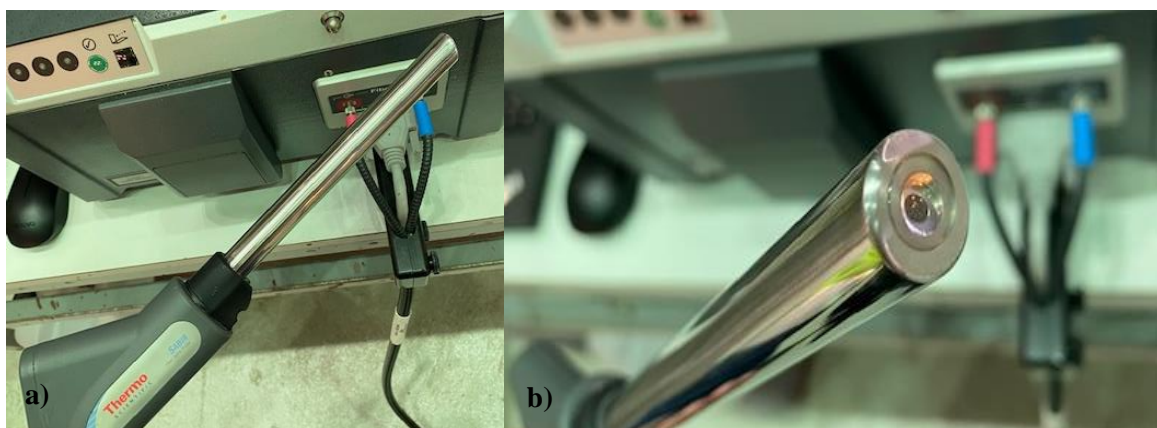
Figure 20.　(a) Thermo Scinetific SabIR diffuse-reflectance stainless steel (316) probe
　　　　　　(b) Tip side of the probe

In OL tests was used fiber optic diffuse-reflectance Series 400 probe also made of stainless steel (316) together with separated Swagelok coupling system for the inline setup. This probe is designed for in-process liquid, solid and slurry analysis. Series 400 probe working principle is based on first illuminating sample and then detecting light once it has passed through the analyzed sample. Operating conditions for this probe are up to 300℃ and 5000 PSI. Tip of the probe is smooth which enables cleaning and easy sampling. This probe has high-sensitivity features to measure samples with minimum spectral reflection or background noise. For this kind of short online test evaluation stainless steel probe was suitable but in the long run online or inline application requires probe that is made of more harsh material such as Hastelloy. Probe for used for online test is presented with Swagelok inline coupling in the Figure 21.
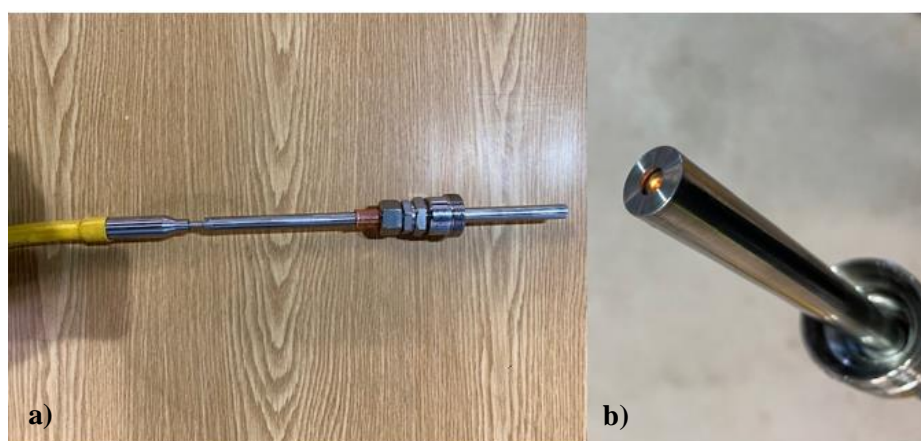

Figure 21.　(a) Thermo Fisher Series A300 online diffuse-reflectance probe made of stainless steel (316) with Swagelok inline coupling
　　　　　　(b) Online probe tip

10.2 Measurement setup

Laboratory NIR measurement spectral range was 12,000 – 3,800 cm^{-1} where now used range was 10,000 – 4,300 cm^{-1} for collecting total spectral range variations. Resolution of measurements was 4 cm^{-1} and 32 averaged scans. The optimal measurement setup for resolution was tested with filtrate samples testing increasing resolution and GAIN value. Number of averaged scans were constant, 32 scans. Filtrate spectra treated with Savitzky-Golay smoothing filter which impact to spectra is eliminate and reduce noise while processing data. Filtrate samples included more noise than other sampling sets. Smoothing of spectra basically change all sharp peaks to smoother also sample peaks in analyzed region and therefore it should be careful not to remove important spectral information. (Wu et al., 2015) GAIN changes have impact detector signal and light intensity to the sample. GAIN increasing helps achieve better signal when it is weak, especially then while using separate sampling accessories. Typically used GAIN setting is 2 or 4 in diffuse-reflectance measurements. However, while increasing both GAIN and resolution, impact was negative to the calibration model and therefore used setup was 32 scans and resolution of 8 cm^{-1} for other fractions, solids, and slurry samples at laboratory scale. In online NIR equipment spectral range covers 12,000 – 4,800 cm^{-1} but used range in OL tests was same as laboratory pretests but number of scans were increased to 64 scans for creating representative online data.

In the Figure 22(a-c) is presented filtrate sample example spectra when measurement conditions are changing. It can be seen in the Figure 22c that spectral data is noisy compared to ones where is used increased GAIN and resolution, but changes of measurement conditions did not give significant changes to results and spectral information was noisy. (Thermo Fisher Scientific manual, 1996-2011)
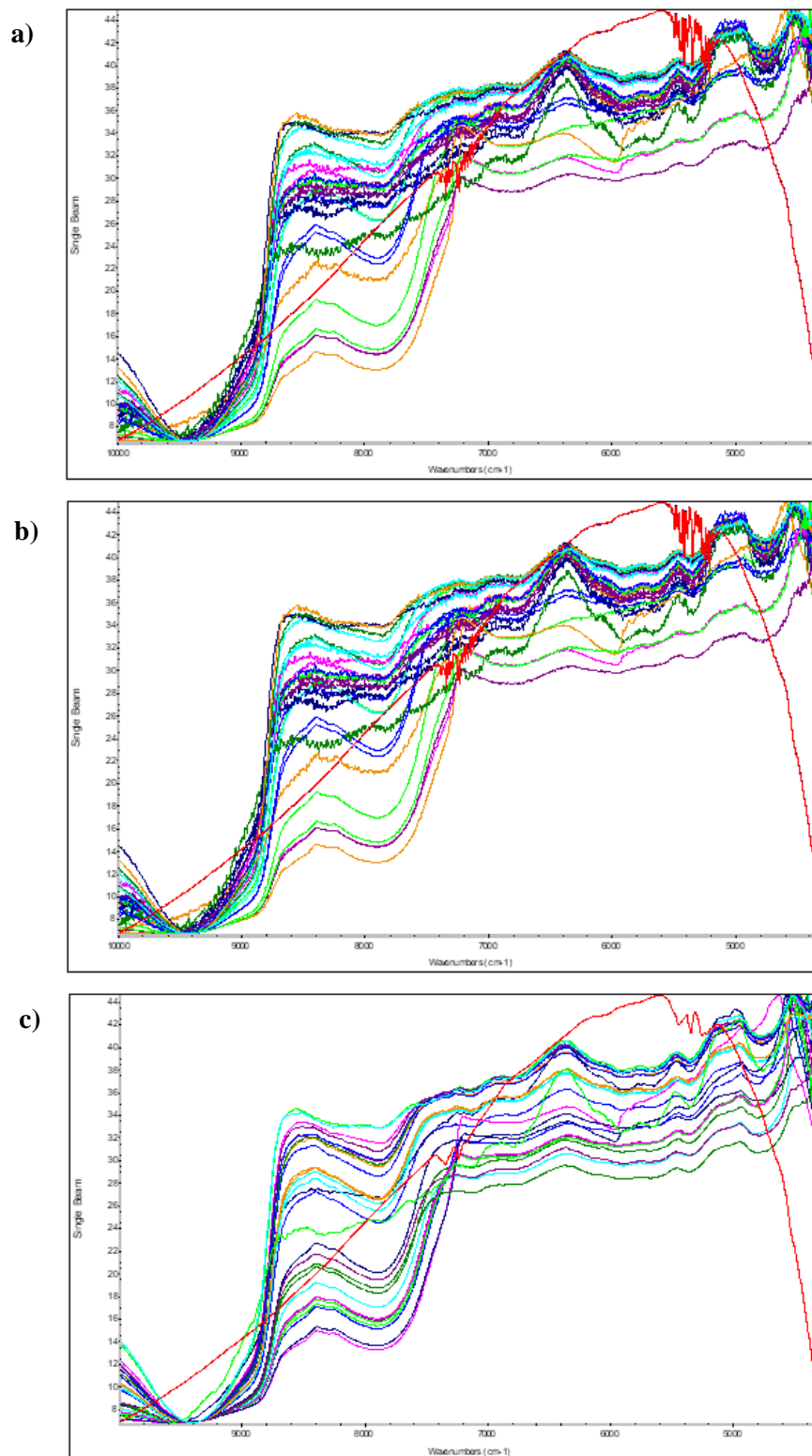
Figure 22.   (a) Filtrate samples with SabIR diffuse-reflectance probe at resolution setup of 32 cm⁻¹, 32 scans and GAIN of 6x. Savitzky-Golay smoothing used to treat spectra
(b) Filtrate samples with SabIR diffuse-reflectance probe at resolution of 16 cm⁻¹, 32 scans and GAIN of 4x with Savitzky-Golay smoothing
(c) Filtrate samples measured with SabIR diffuse-reflectance probe at resolution 8 cm⁻¹, 32 scans and GAIN of 2x. Red spectrum is collected background. This measurement setup is used in experimental part measurements at laboratory scale also for other fractions. Spectra do not include Savitzky-Golay smoothing

10.3 NIR software

Thermo Scientific TQ Analyst Method Developer Software was used for calibration model development from measured spectral data. TQ Analyst is suitable for multiple spectroscopic applications such as mid-infrared, far-infrared, near-infrared and for Raman. This software includes versatile package which can be moved from laboratory measurements to online analysis and equipment. Add to TQ Analyst, software package includes OMNIC program which purpose is to handle and collect spectral data from measurements and move data to TQ Analyst for calibration model which is done manually. In the Figure 23 is presented TQ Analyst (a) user interface and below (b) OMNIC program overview for collecting spectral data at laboratory scale.

Figure 23.   (a) TQ Analyst Software user interface
(b) OMNIC overview for spectral data modifications and measurements
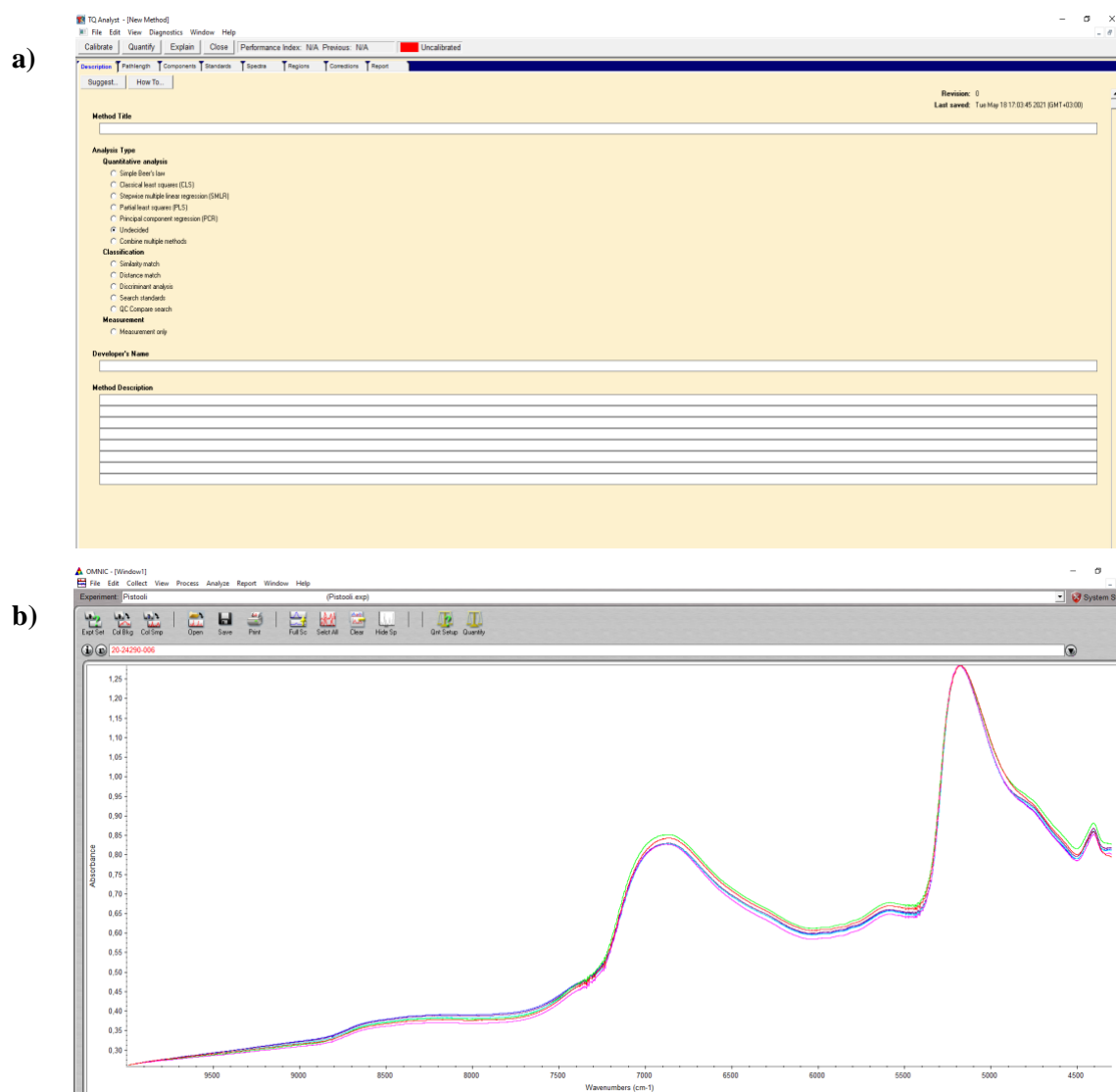
NIR measurement process route is presented in the Figure 24 with Thermo Scientific TQ Analyst chemometric package. Measurement process is based on these two programs, OMNIC and TQ-Analyst Software at laboratory scale. OMNIC is software where measurements and possible spectral pretreatment are done. By OMNIC program spectral data can be saved and stored outside but in addition outside data can be brought to the program and create new model. OMNIC specific feature was that program normalize automatically spectral data and visualize it following the way where spectrum is forced to go from minimum to maximum absorbance without no variation of individual samples, but it could be removed in OMNIC program. After collecting spectral data and saving information from OMNIC, TQ-Analyst is used for classification or quantitative analysis of spectral data. In quantitative analysis software calculates concentrations of one or more samples or mixtures. For this work only qualitative analysis is used where spectral data intensity is correlating to absorbance changes occurring samples. More details of calibration process are discussed in chapter 12.1.
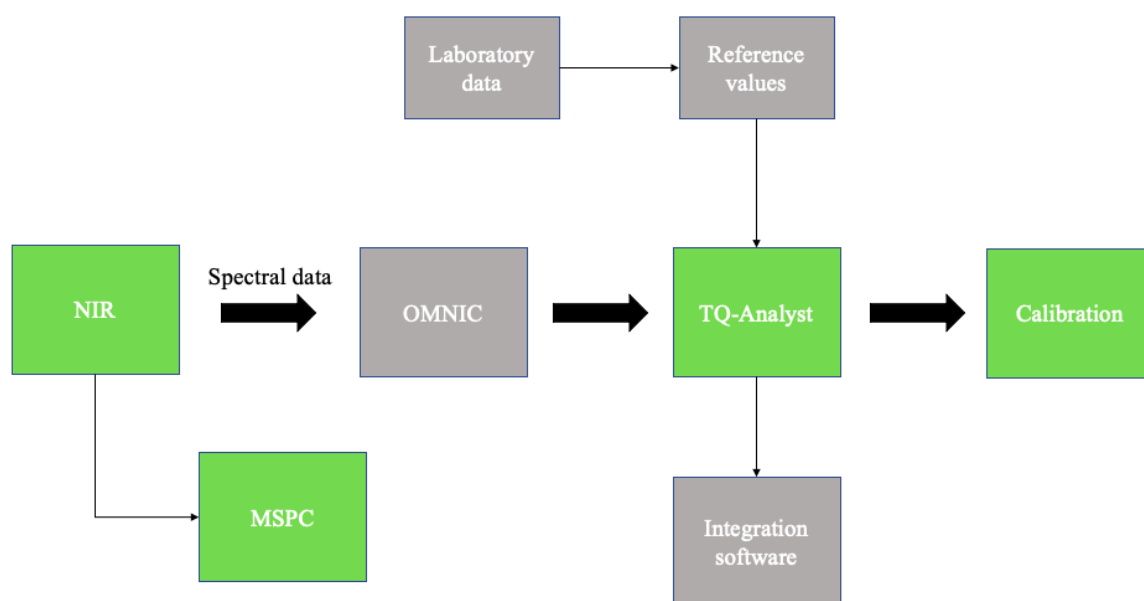

Figure 24. NIR measurement process route using Thermo Fisher software package and MSPC utilization from spectral data

Online measurement software is similar and based on these two programs, TQ-Analyst and OMNIC but for online measurements created software requires separated Integration and RESULT Operation programs to create automation background for continuous measurements. Integration program includes automation workflow where setups for online loop tests were set

such as spectrum range, resolution, number of scans, spectral data delays and data saving settings. Calibration for online loops were created by TQ-Analyst which moved into Integration workflow program. Only Result Operation needed in online loop tests. These programs work quite smoothly together and moving calibration settings to online equipment was quite easy but included several steps. Measurements for online calibration were made by using online probe and used online NIR equipment.

11 PRETREATMENTS OF SPECTRAL DATA

Processing NIR spectral data is important part of creating representative model because it is very important understand spectral information and its quantitative features before applying bilinear modelling such as PLS and PCA. NIR spectra is consisted of various nonlinear contributions which commonly are light scattering occurred by particles, intermolecular interactions, phase, and surface transitions. (Sørensen et al., 2020) Especially in NIR reflectance and transmittance applications this light scattering occurs. Analyzing solid samples by NIR reflectance method, systematic variations is caused by light scattering and variation to path length as well while diffusively reflected and mirror-like reflections are measured. Undesired systematic variations in solid samples forms a major part of sample set. These spectral reflections can be minimized by using instruments or sampling geometry. Diffusively reflected light contains the chemical information of the sample which is source of information where in addition to absorption changes is micro-structure changes like scattering. Scattering forms are Lorenz-Mie and Rayleigh where is not information of energy transitions. Lorenz-Mie and Rayleigh scattering are both electromagnetic radiation which can be caused by bubbles, particles, roughness of sample surface, fibers, and density fluctuations. Lorenz-Mie scattering is not that wavelength dependent in turn Rayleigh when particles have smaller diameter than wavelength of electromagnetic radiation. (Rinnan et al., 2009)

These effects to NIR spectra can be removed well by pre-processing methods. In pre-processing target is enhance spectral data linear relationship between absorbance and the sample concentrations. This means that wanted result is to remove nonlinearity before chemometrics is applied and used spectra contains additional information of chemical changes. There are multiple methods that are proposed to use pretreat NIR data which handle physical and chemical variations in spectra. (Sørensen et al., 2020) Pre-processing methods can be divided

into spectral derivatives and scattering correction which are used in both transmittance and reflectance NIR data. Scattering correction methods include multiplicative scatter correction (MSC), inverse MSC, standard normal variation (SNV), normalization, extended MSC and extended inverse MSC. Spectral derivative pre-processing methods include 1st and 2nd derivatives, Savitzky-Golay polynomial and Norris-Williams derivates. These methods are for spectral data smoothing and decreasing signal-to-noise ratio to minimum. In addition, pre-processing of NIR data is forced obeying Lambert-Beer's Law which is presented in equation (4) which describes linear relation between spectral absorbance and concentration of samples of interest. (Rinnan et al., 2009)

$$A_\lambda = -\log_{10}(T) = \varepsilon_\lambda \cdot 1 \cdot c \qquad (4)$$

where A_λ wavelength-dependent absorbance

T light transmittance

ε_λ wavelength-dependent molar absorptivity

I path length of light through the sample

c concentration of samples of interest

Equation (5) is valid strictly only in pure transmittance measurements and therefore in reflectance is redefined by equation (4).

$$A_\lambda = -\log_{10}(R) \cong \varepsilon_\lambda \cdot 1 \cdot c \qquad (5)$$

where R detected reflectance

Selecting optimal method for pretreatment of NIR spectral data is considered by further used modelling step. It is in some situations challenging because too severe pre-processing should be avoided because that may lead to lose some valuable spectral information. It is hard to

evaluate preprocessing successful before model validation process which is discussed more in chapter 12. Overall, while pre-processing spectral data it is advisable not to use several methods rather that data complexity is decreased without losing information. In this work, 1st derivative is used as pre-processing NIR data with TQ Analyst and MATLAB software. TQ Analyst software provides in addition 2nd derivative and smoothing methods such as Norris-Williams and Savitzky-Golay filters.

Objective of derivative is to remove multiplicative and additive effects from spectral data. Working principle of 1st derivative is based on removing only the spectrum baseline while 2nd derivative removes both, baseline, and linear trend. Pretreatment of spectral data with derivatives correct spectrum overlapped peaks and baseline slopes. Derivate based calibration models are usually more robust requiring smaller amounts of variables. Overall 1st and 2nd derivates are the mostly used as well as Savitzky-Golay smoothing filter. However, the optimal pre-processing is strongly dependent on transmittance or reflectance, samples and NIR equipment itself. (Agelet & Hurburgh, 2010) Derivatives in both options smooth spectral data while reducing not too much signal-to-noise ratio. Derivative method is easy, in 1st derivative it is difference between spectral measured points and accordingly in 2nd derivative, calculated differences between 1st derivative spectra points as follows with equations (6) and (7). (Rinnan et al., 2009)

$$x' = x_m - x_{m-1} \tag{6}$$

$$x_i'' = x_i' - x_{i-1}' = x_m - 2x_m + x_{m-2} \tag{7}$$

where x_i' denotes 1st derivative

             x_i'' 2nd derivative at wavelength points i

The used variables are X and Y where X is NIR spectra and Y response variables as concentrations in this work. In the Figure 25a is presented original measured sample without pretreatment and 25b 1st derivative of measured spectra without other smoothing tool filters.
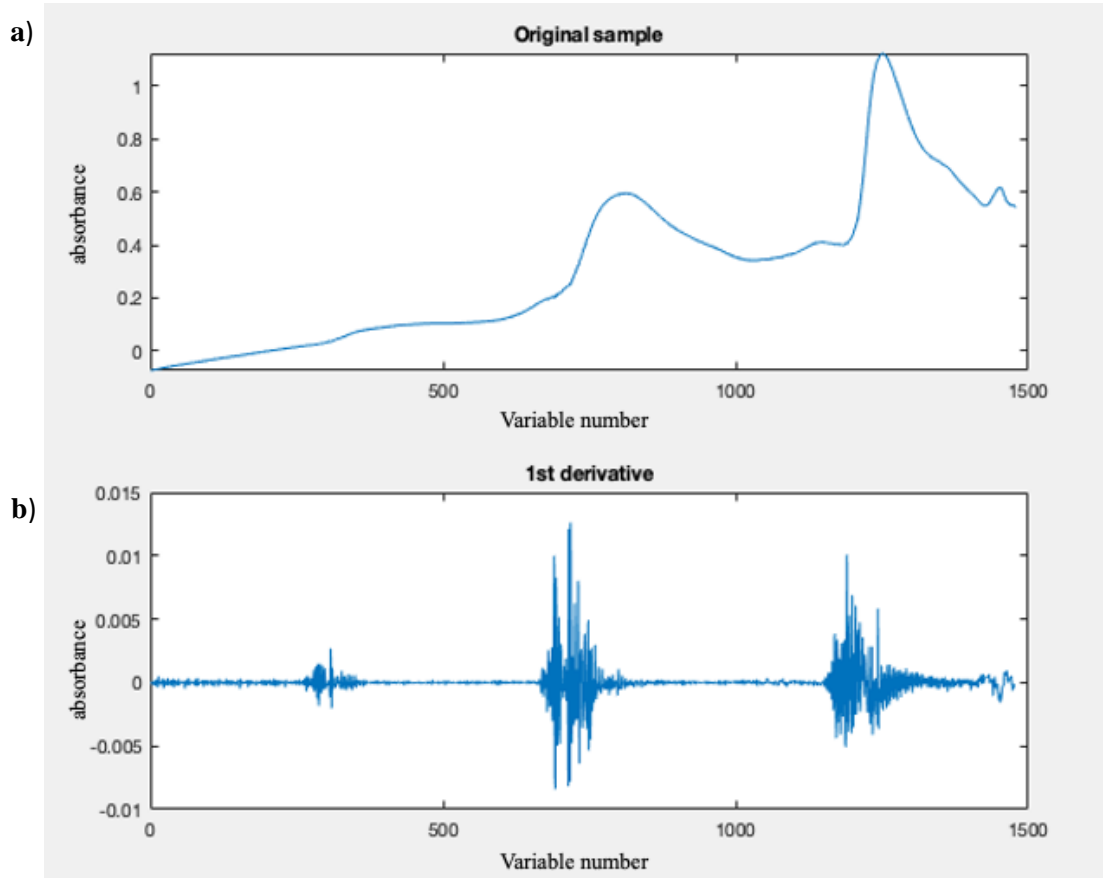

Figure 25.　　　　(a) Original sample NIR spectra, x-axis variable number and y-axis absorbance
(b) 1st derivate of original spectra presented in the figure (a)

Examples of 1st derivative effects to measured NIR spectra is presented in the Figure 26 when dry matter content varies. One original sample is measured in different moisture content levels and temperatures. In the figure 26a temperature is at laboratory temperature but dry matter content level varies from normal. Normal dry matter content for original pulp is approximately between 52-54 %. Blue spectra dry matter content is 28 %, yellow 42 % and red 38 %. There is variation in spectral data shapes when dry matter content varies as well as temperature change. These are discussed in more details in chapter 13.3 which contains information of different measurements variations. In the Figure 27 is presented one sample in different temperatures; 30, 40 and 50 °C. Figure 27b presents first derivative of samples.
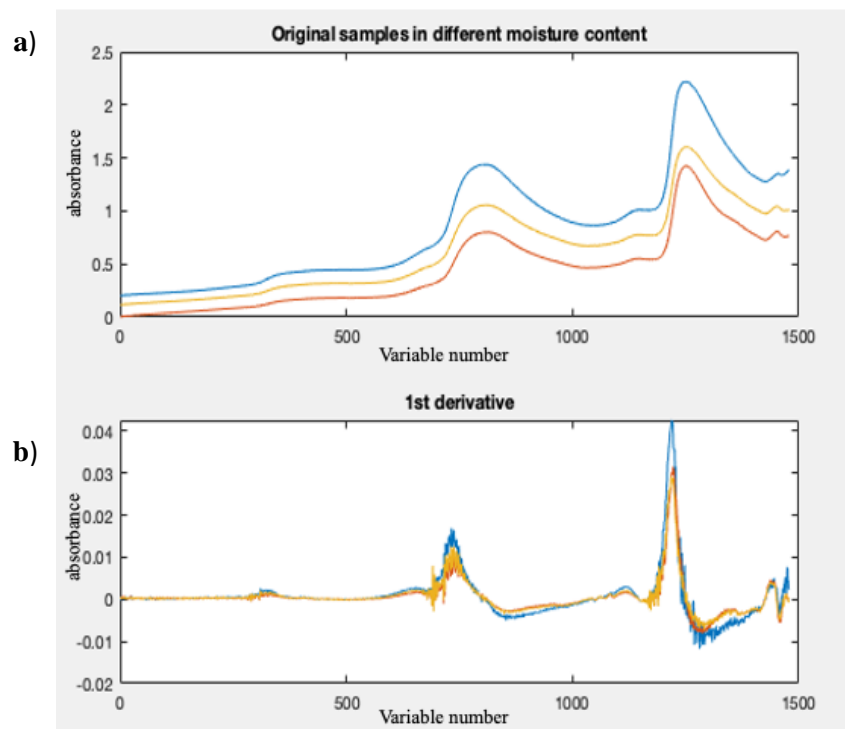
Figure 26. (a) Original sample in different dry matter contents. Blue spectra 28 %, yellow 42 % and red 38 % where x-axis variable number and y-axis absorbance
(b) First derivate of moisture variations spectra where x-axis variable number and y-axis absorbance
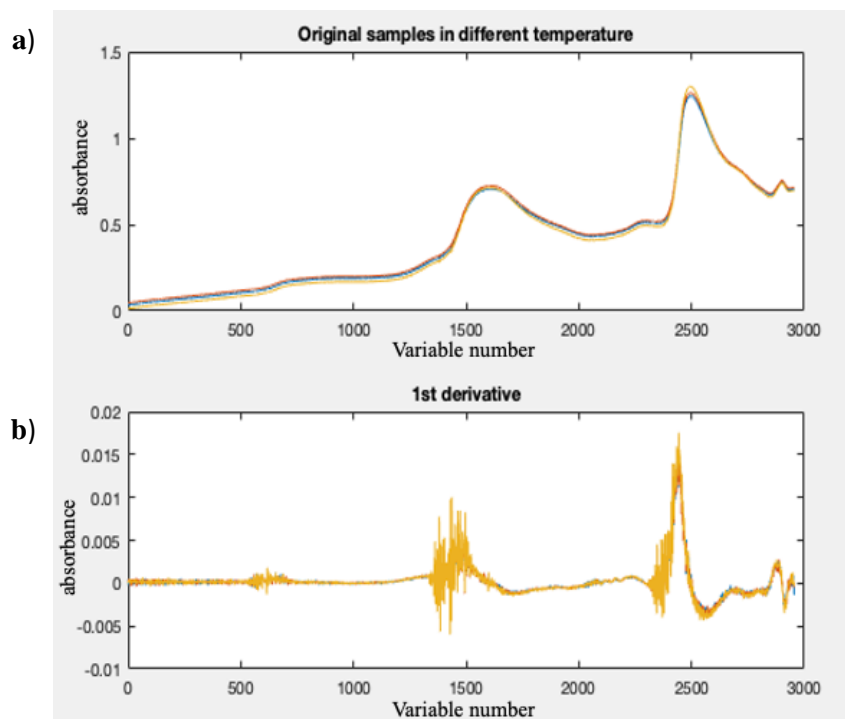

Figure 27. (a) Original sample in different temperatures without moisture variation at 30, 40 and 50 ℃ where x-axis variable number and y-axis absorbance
(b) First derivative of temperature variations where x-axis variable number and y-axis absorbance

12 CALIBRATION PROCEDURE AND VALIDATION

Calibration is a process where mathematical model is formed between the instrumental response and the reliable concertation of reference data in certain range. In case of spectroscopic calibration modelling, relationship between spectral intensity and concentration are in interest.

12.1 Calibration

A representative sampling set from original biomass samples (O) with required capture at different concentrations was selected in the starting point of the calibration process. Created calibrations were based on 28 different samples with reference values in one model at same spectrum range. Software suggests based on spectral data areas that could be used in calibration model, but now total spectrum range was used in first calibrations to including also smaller changes and information to the model. Already at this point noted that selected spectrum range impact strongly to calibration model. After certain samples were selected and measured by NIR, they were included to new calibration sheet as calibration samples in TQ Analyst software together with reference values. Option to ignore and validate listed samples was not used in calibration stage. Table VII summarizes calibration settings and model performance results to original biomass set (O) where pre-processing method, number of calibration samples included to model, spectrum range, number of factors, and root mean square error of calibration (RMSEC) are presented. PLS method was used as multivariate calibration method.

Table VII   Calibration settings for and original biomass set (O) to lignin and kappa number

Parameter	Pre-processing method	Calibration samples	Chemometrics method	Spectrum range, cm^{-1}	Factor number	R^2	RMSEC
Lignin	1st derivative, no smoothing	28	PLS	4300-10 000	5	0,9998	0,0251
Kappa number	1st derivative, no smoothing	28	PLS	4300-10 000	5	0,9996	0,224

Performance tools for estimating the multivariate calibration model provided by TQ Analyst Software were R^2, RMSEC values and PRESS plot. R^2 value called goodness of fit describes how well regression prediction points will approximate data. Value of 1,0 estimates that model

is perfectly fitted but directly this does not give answer if model would work properly and estimate future samples correctly. RMSEC as root mean square error of calibration is dimension how well calibration works and how it could estimate future samples. RMSEC is result of obtained independent sample set and performance of calibration result. Model PRESS is value which describes predicted error sum of squares and in PLS model it is sum of residual concentration error to standards that are removed by cross-validation iteration.

Cross-validation is mostly used technique for model performance testing and validation which is based on calibration of samples by resampling when all samples are once removed from the sample set and modelled. It is also method for limit predictive model overfitting. (Kucheryavskiy et al., 2020) Cross-validation and PRESS were used in TQ Analyst to optimize used PLS components in model as internal validation process. TQ Analyst propose spectrum area and number of PLS components to be used in calibration model but for calibration model number of components were increased based on PRESS and residuals plots and how model estimate individual samples outside the calibration plot. Program suggested factor numbers were low which affected that calibration results were worse and therefor needed smaller variation was not included to the model. However, number of used PLS factor in model need to be set minimum but TQ Analyst provided 1-2 was way too low.

PLS calibration model in normal conditions to lignin is presented in the Figure 28a including total calibration sample set, 28 samples and five PLS factors for original samples (O). The used spectrum range for calibration is total measured area, 4300-10 000 cm^{-1}. Residual's plot is presented in the Figure 28b which shows that samples are not that much spread. In the residual plot, on the vertical axis are residuals and horizontal the independent variables. Factor plot to recommended PLS components is showed in the Figure 28c which shows that line flattens towards the end and RMSECV values are not changing anymore. PRESS and RMSEC results are summarized into table in the Figure 28d.
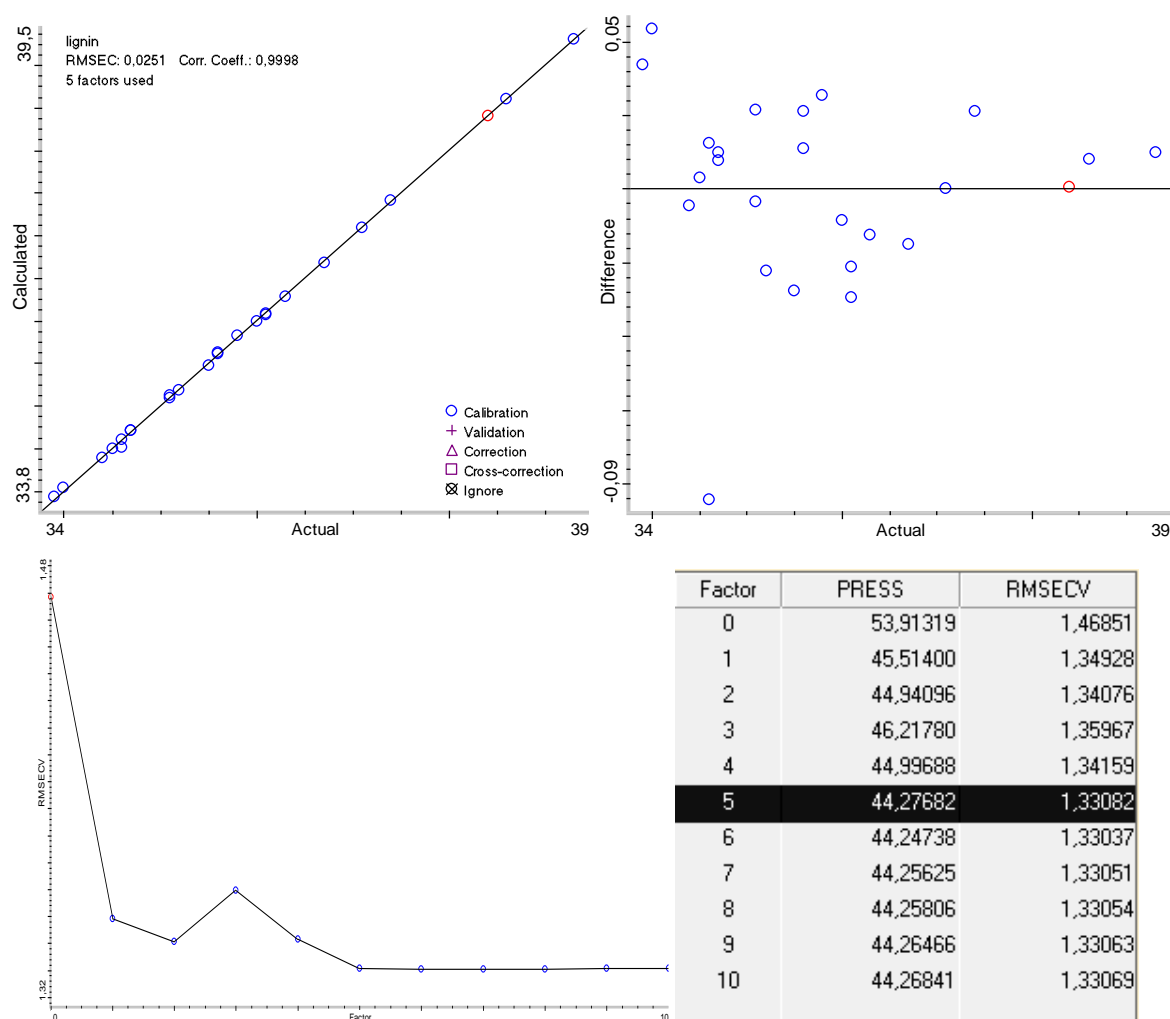
Figure 28.   (a) Calibration for lignin covering all spectrum area, 4300-10 000 cm^{-1} without validation samples for original sample set (O)
(b) Residuals plot for lignin calibration
(c) Factor plot for lignin
(d) Calibration PRESS and RMSECV table for lignin calibration results

Kappa number calibration results of original sample set (O) are presented in the Figure 29a with five PLS factors and residual plot 29b. From the residual plot can be seen that samples are now more spread compared to lignin results. Despite calibration results were as good as compared to lignin. Factor plot is presented in the Figure 29c and 5 PLS components is used in this calibration. These both calibrations represent ideal examples how calibration process work and response with references is obtained from the measured NIR data. When these models are being tested, results are not straightforward anymore and this is discussed more in the next validation chapter.
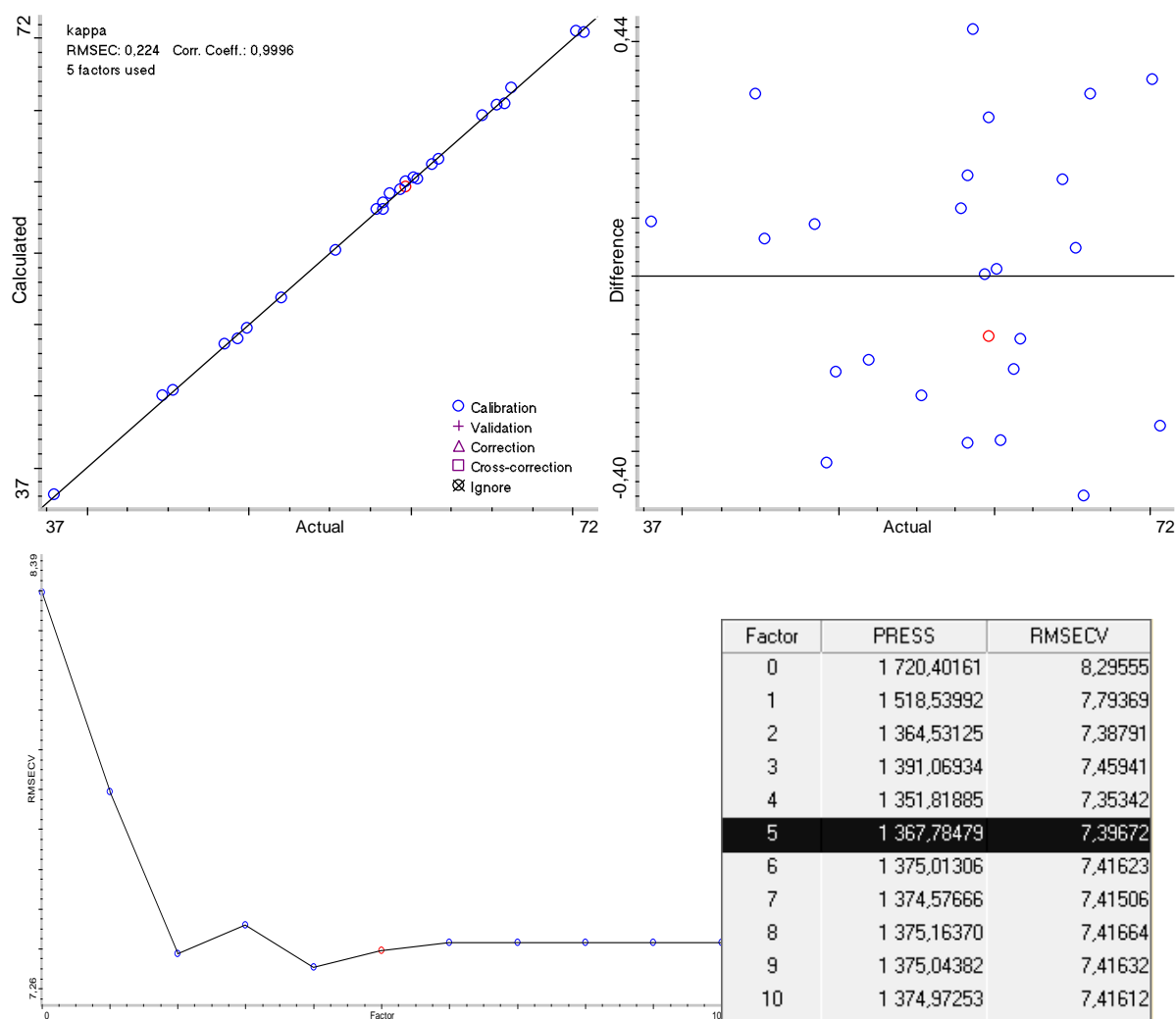
Factor	PRESS	RMSECV
0	1 720,40161	8,29555
1	1 518,53992	7,79369
2	1 364,53125	7,38791
3	1 391,06934	7,45941
4	1 351,81885	7,35342
5	1 367,78479	7,39672
6	1 375,01306	7,41623
7	1 374,57666	7,41506
8	1 375,16370	7,41664
9	1 375,04382	7,41632
10	1 374,97253	7,41612

Figure 29. (a) Calibration for kappa number covering all spectrum area, 4300-10 000 cm^{-1} without validation samples for original sample set (O)
(b) Residuals plot for kappa number calibration
(c)   Factor plot for kappa number
(d) PRESS and RMSECV table for kappa number calibration results

12.2 Model validation

Calibration model validation plays crucial part for performance testing in multivariate modelling analysis. Main target in validation is to show how well model can predict new samples and give reliable results. Ideally validation is done by using samples that are not included to the calibration sample set and this was used in this experimental part. Before showed examples of lignin and kappa number calibrations were overly well fitted models that did not included sample variations or when other influencing factors such as the effects of moisture and temperature are not considered. This leads into those models do not represent and predict reality without that these changes are noticed while creating calibration model these biomass samples. Presented calibrations as model validation examples are made to cover total spectral data (4300 -10 000 cm^{-1}) using 1st derivative as pre-processing and PLS multivariate

72

method. The used spectral data for original samples as calibration set is presented in the Figure 30a at 20 °C and 1ˢᵗ derivate of sample set in 30b. Beside in the Figure 31a is presented used validation sample spectra at 30 °C and 1ˢᵗ derivates of it in 31b. From these figures can be noted that spectrum shapes are changing due to increased temperature.
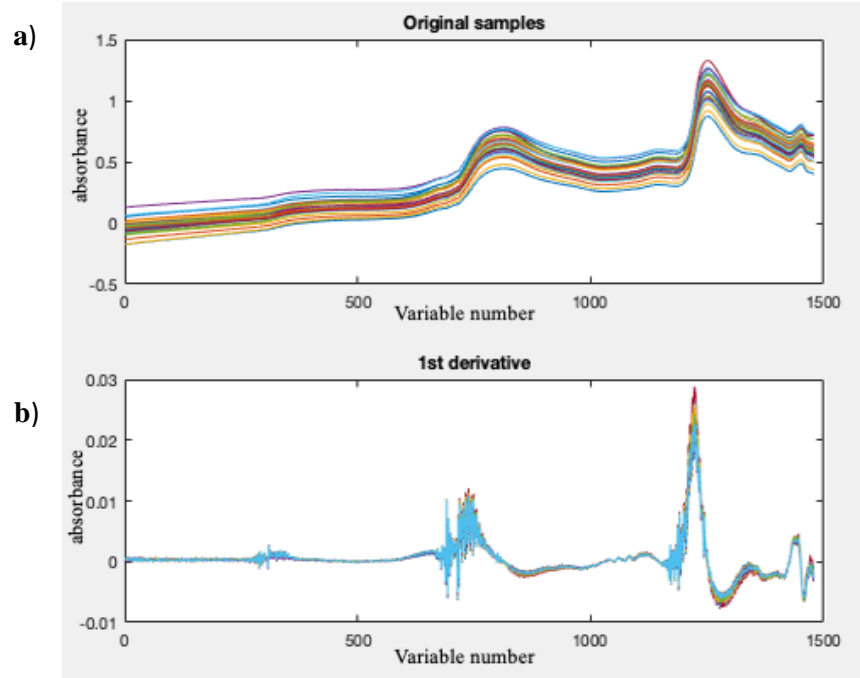

Figure 30.   (a) Original sample set (O) as calibration material at 20 °C where x-axis variable number and y-axis absorbance
(b) Original sample set 1ˢᵗ derivative where x-axis variable number and y-axis absorbance
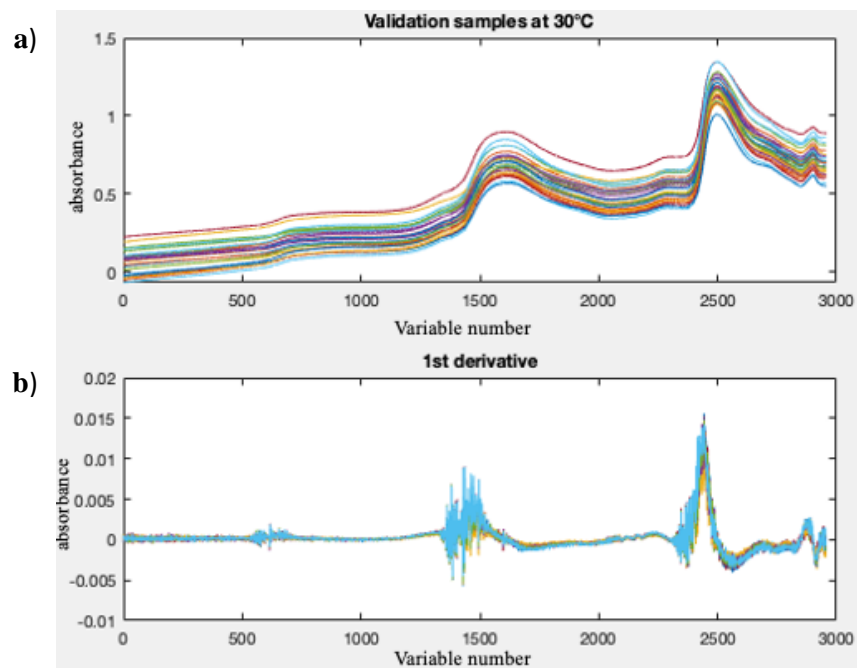
Figure 31.   (a) Validation sample set at 30 °C where x-axis variable number and y-axis absorbance
(b) Validation sample set 1ˢᵗ derivative where x-axis variable number and y-axis absorbance

In the Figure 32a is presented validation set to lignin calibration of original sample set (O) when temperature changes from 20 ℃ calibration to 30 ℃ for validation samples. From the validation set can be seen that those 10 degrees temperatures effect is big, and samples do not estimate well with using this calibration model. Model included both temperatures in calibration predicts better validation samples which is presented in the Figure 32b for lignin, but model still does not represent required efficiency perfectly.
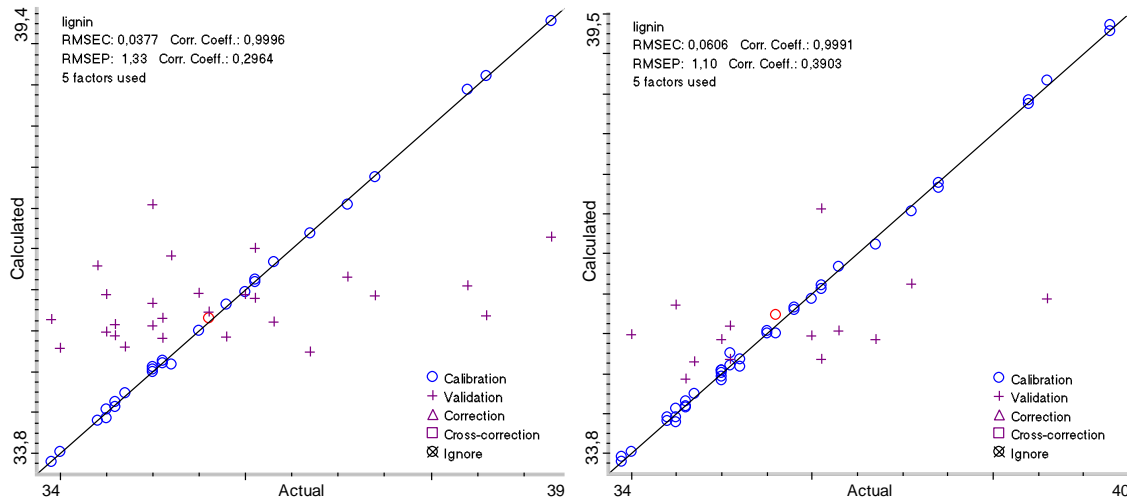
Figure 32.   (a) Calibration for lignin at 20 ℃ to original sample set (O) and validation samples at 30 ℃. Calibration includes total spectral area 4300-10 000 cm^{-1}
(b) Calibration including samples at 20 °C and 30 °C with validation samples

In the Figure 33a is presented calibration for kappa number of original sample set (O) with validation samples at 30 ℃. Figure 33b presents combined temperature calibration with validation samples for kappa number.
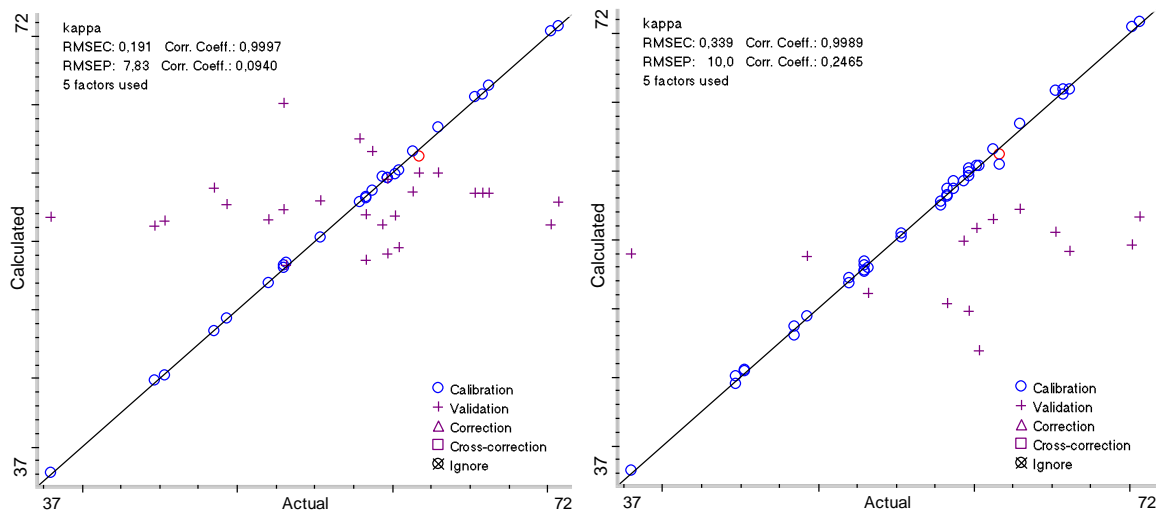

Figure 33.   (a) Calibration for kappa number at 20 ℃ to original sample set (O) and validation samples at 30 ℃. Calibration includes total spectral area 4300-10 000 cm^{-1}
(b) Calibration including samples at 20 °C and 30 °C with validation samples

Based on presented calibration and validation results for lignin and kappa number in Figure 32 and 33, calibration model can enhance and minimize temperature effect with lower temperature range that is included to the model. Temperature effect calibration noted as well as moisture variations. Water regions occur dominant 5200 and 6890 cm^{-1} in NIR and this can be seen from 1st derivate plots where peaks are smaller and sharp. (Padalkar et al., 2013) For this reason, calibration spectrum area cannot be cut to remove water areas without losing the information. The other option is to use external water moisture factor included to the calibration. In the Figure 34a is presented original sample set (O) calibration spectral data and in the Figure 34b its 1st derivates for creating calibration model to describe moisture content changes in measured samples.
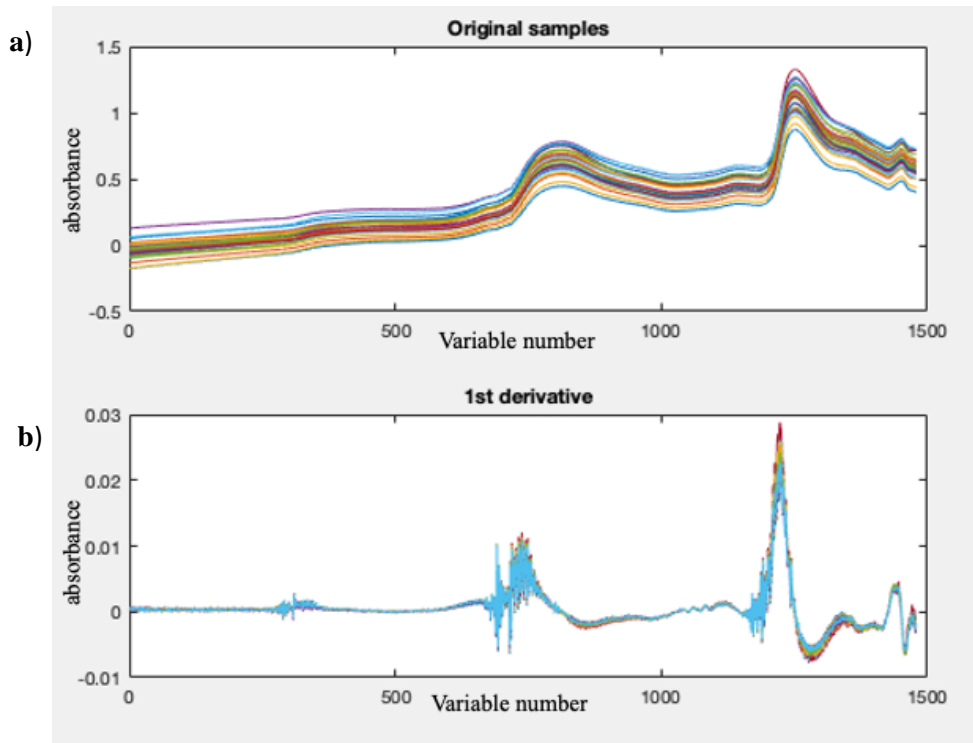

Figure 34.   (a) Original sample (O) calibration set where x-axis variable number and y-axis absorbance
(b) 1st derivative of original sample set, variable number in x-axis and absorbance in y-axis

Original samples in different moisture levels that are used as validation samples are presented in the Figure 35a. Dry matter for samples are 28 %, 31 %, 34 %, 38 % and 48 %. First derivative of validation samples is presented in the Figure 35b. In the Figure 35a can be seen that moisture changes impact more on spectrum shapes than temperature presented in the Figure 31 a.
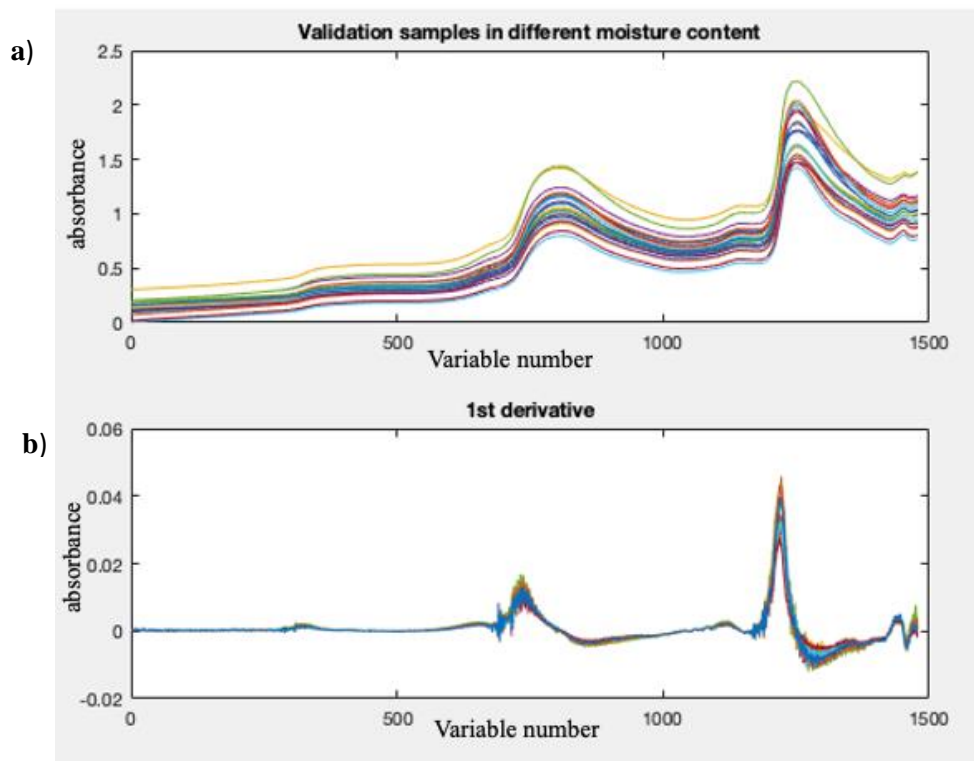
Figure 35. (a) Selected samples from set O in different total solids; 28%, 31%, 34%, 38% and 48% as validation samples, x-axis variable number and y-axis absorbance

(b) Validation sample set 1st derivatives where x-axis variable number and y-axis absorbance

Lignin calibration set for original mass samples is presented in the Figure 35a with moisture validation samples. Figure 35b presents calibration model including moisture variation samples and after this validation samples for different dry matter for lignin that are 28 %, 31 %, 34 %, 38 % and 48 %. Based on calibration and validation plots, dry matter content effect model efficiency.
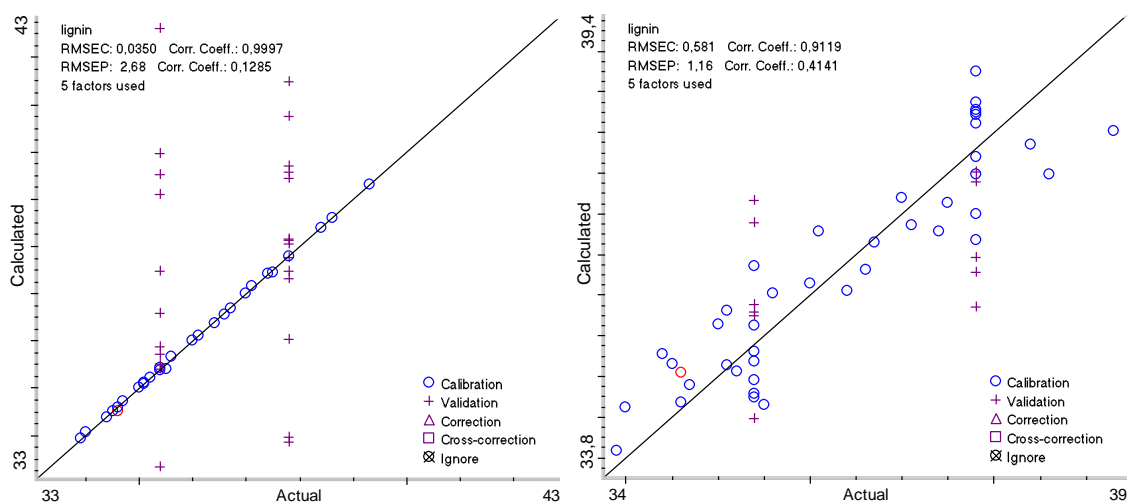

Figure 35. (a) Lignin calibration set with validation samples in dry matter levels; 28%, 31%, 34%, 38% and 48%. Calibration includes total spectral area

(b) Calibration capturing different dry matter levels with validation samples for lignin

Kappa number calibration with validation samples in different dry matter levels is presented in the Figure 36. In the Figure 36b is presented calibration model that includes samples also within different dry matter levels and validation samples.
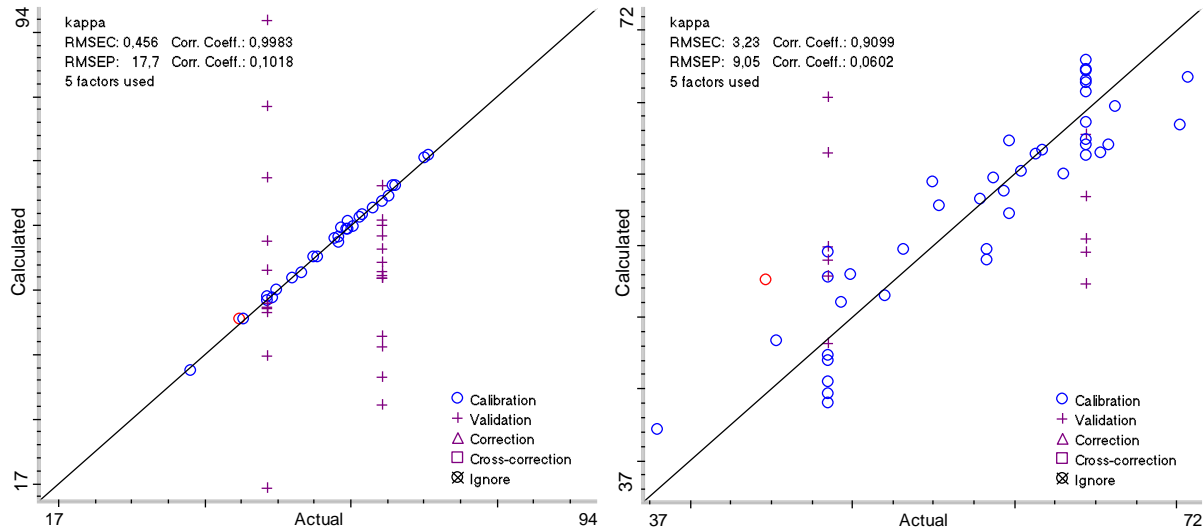

Figure 36.   (a) Kappa number calibration set with validation samples in different dry matter levels; 28%, 31%, 34%, 38% and 48%. Calibration includes total spectral area
(b) Calibration capturing different dry matter with validation samples for kappa number

Based on these calibration model validations that were not successful, calibration model testing and improving were done. These both, lignin and kappa number have same trend in validation testing, both parameters do not predict properly with these tested models. Temperature and moisture effect is too severe. This led into that model needs improvements and more careful spectral area limitations due to water peaks which is more detailed discussed in NIR laboratory measurement chapter 13. Elements of successful NIR measurements require calibration model that can predict samples and model is updated with new calibration samples over time because if this is not done regularly, impacts to measurements are negatively. Used reference parameter uncertainties that are straight associated with wet chemical methods, are directly translated to NIR model prediction. NIR calibration model must capture process variations and model is not able to predict samples which have temperature and moisture changes even if this variation included to the calibration model.

13 NIR IN LABORATORY

NIR laboratory calibrations were first tests towards final online analysis tests and creating suitable calibration model to those final experiments. In addition, NIR laboratory measurement's biggest objective was to test how sensitive NIR is measuring the pretreated woody biomass. Previously presented sample set was tested under laboratory with different modifications. Focus on measurements were set to original biomass (O) and slurry (S) sampling sets due to these would be easy and reasonable to implement in real process environment. On the other hand, washed solids (W) and liquid (L) as filtrate were not first fractions in priority order but despite these sampling sets were evaluated to see how NIR is able to measure them. Measurement campaigns for NIR laboratory measurements, modifications, main structure of fractions and usage of validation samples is presented in the Table VII.

Table VII   NIR laboratory measurements sample sets, tested modifications and whether measurements are used as validation samples

Measurement campaign	Modifications	Composition distribution	Validation samples
Original biomass (O)	• temperature changes • moisture changes	- 70 % solubles - 30 % insolubles	Yes
Washed solids (W)	• no modifications	- insolubles	No
Liquid (L)	• sensitivity testing for the probe • resolution	- solubles	No
Slurry (S)	• TS-% changes • insoluble changes • temperature	- solubles - insolubles	Yes

13.1 Measurement campaigns

Laboratory measurements started from basic point for this type of pretreated biomass and based on other NIR related studies to measure other materials. All measurement campaign sample matrix sets were measured by NIR at ambient laboratory conditions with SabIR probe and then created calibration model based on reference data as pretests towards online tests. In the Table X are presented each measurement campaign individual calibration model goodness of fit and number of used PLS factors to lignin and kappa number. Conclusion was based on these laboratory measurements that all sample campaigns were possible to be measured with NIR. Laboratory NIR probe was designed to measure samples that include solids in samples, despite

it was able to measure liquid fraction but that calibration model needs more modification to achieve reasonable results. Below presented Table IX is an indication of NIR suitability to measure these sampling sets locally, and calibration models have high R^2 value. The number of used PLS factor during calibration process is within approved limits. These first pretest formed a base towards further online tests and creating its calibration model and calibration modifications testing to create more global model. Same calibration model was used to predict both lignin and kappa number.

Table IX   Individual calibration results for each measurement campaign without validation samples, goodness of fit and number of used PLS factors during calibration process

Measurement campaign	Lignin		Kappa number	
	R^2 of the calibration	Number of PLS factors	R^2 of the calibration	Number of PLS factors
Original biomass (O)	0,9998	5	0,9996	5
Washed solids (W)	0,9992	5	0,9996	5
Liquid (L)	0,9930	4	0,9939	4
Slurry (S)	0,9969	5	0,9978	5

13.2 Selection of NIR laboratory measurement calibration area

After all measurement campaign fractions were measured and individual calibration models were created to predict reference data, next step was test model globality and sensitivity to process changes. Now samples were not predicted as desired which was the key challenge point during this study. Calibrations were more local than global because kappa number and lignin values were not able to predict. The focus was set then on these two models, O and S fractions and their improvements to more global calibration models. NIR is strongly temperature and moisture dependent and sensitive to changes of these process conditions. In addition to extreme process condition changes and sensitiveness, the effect of spectral area delimiting and cutting was applied to the models.

Spectral area selecting is important part and strongly influencing factor of calibration models in how the best spectral data information is obtained from the samples. While delimiting the spectral area in the calibration models it is possible to reduce e.g., temperature and moisture

effect, but it is also possible through spectrum preprocessing techniques. In first calibration models total spectral area was applied to the models but eventually spectral area review was needed to find out if there are certain wavenumber areas that are highly correlating with lignin and kappa number which is noted in literature studies too. Correlation to certain wavenumber area was applied using basic regression model in Excel to find the highest correlation value between reference data and measured NIR absorbance from 1st derivate spectral data. However, use of the entire spectral data area proved to be a successful choice with calibration models including the spectral information variations that are important to model. In the Figure 37 is presented lignin correlation plot to original mass spectral NIR absorbance against reference data where x-axis wavelength number (cm^{-1}) and y-axis correlation coefficient. From the total lignin correlation plot can be seen that there are not certain distinctive areas which would have extremely high correlation and highest correlation value reaches 0,68. Overall, lignin spectral response is spread quite evenly over the spectrum range that is hard to choose certain areas for the model because even little changes are occurring throughout the total measurement section. In the end of the measurement range at 8000-10 000 cm^{-1} spectra is noisy and lignin correlation is clearer in the beginning at 4300-6000 cm^{-1}.
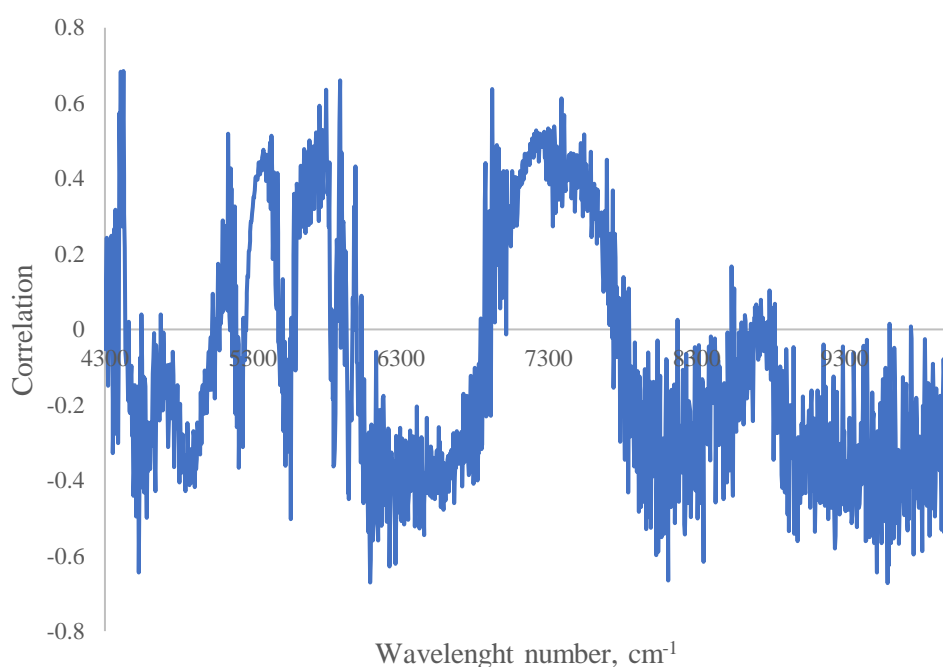

Figure 37. Total lignin correlation plot for original mass samples where measured NIR absorbance against reference data. 1st derivate used for preprocessing the spectral data before correlation analysis. Samples measured normal laboratory conditions without modifications in spectral range of 4300-10 000 cm^{-1}

Kappa number correlation plot to original biomass samples is presented in the Figure 38 from the 1st derivatives of spectral data against references. Kappa number correlation plot differs from the lignin plot, due to it is more even in the entire measuring spectral range is occurring correlation variation. Same goodness of fit is repeated in this situation and only certain wavelength areas are hard to choose for the calibration based on these results. There are peaks that show quite good correlation but R^2 values are not higher than 0,65 at this point. Both positive and negative correlation both are occurring evenly. To the correlation plot results it does not matter whether the correlation occurs in the negative side if it is clear and within acceptable limits that the calibration model is able to build representative.
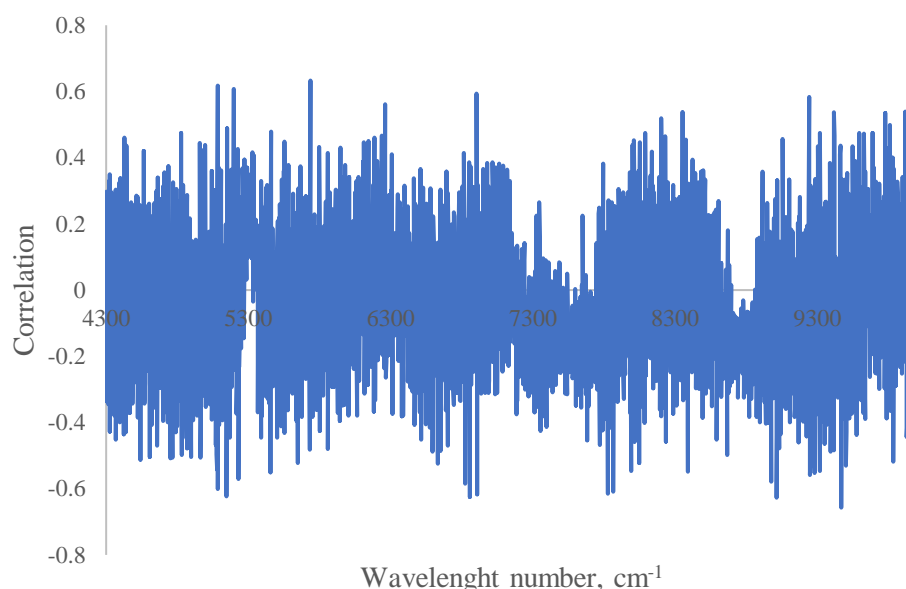

Figure 38.   Kappa number correlation plot for original mass samples where measured NIR absorbance against reference data. 1st derivate used for preprocessing the spectral data before correlation analysis. Samples measured normal laboratory conditions without modifications in spectral range of 4300-10 000 cm^{-1}

Same correlation data analysis was applied to the slurry samples. In the Figure 39 is presented lignin correlation against reference data to TS 12 % slurry. NIR data is collected at laboratory conditions without extreme sample modifications and data preprocessed with using 1st derivative before correlation analysis. For the slurry samples results are the same and certain correlation peaks are not found in the spectrum range.
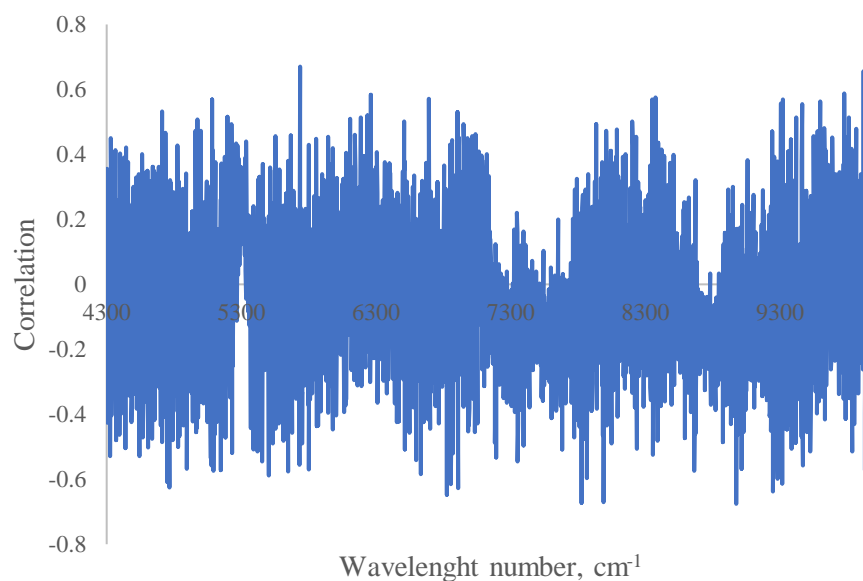
Figure 39. Lignin correlation plot for TS 12 % slurry samples. NIR measured data pretreated with 1st derivative before correlation analysis and samples measured at laboratory conditions in spectral range of 4300-10 000 cm^{-1}

Correlation plot analysis for kappa number is presented in the Figure 40 for slurry samples without sample modifications such as total solids or insoluble content changes. Correlation plot for kappa number shows that there is not certain area that highly correlates and stands out from the measurement range.
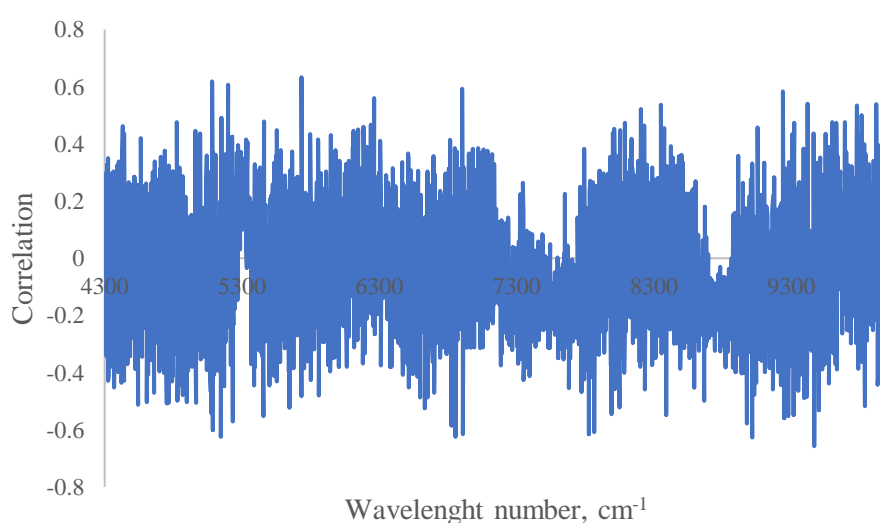

Figure 40. Kappa number correlation plot for TS 12 % slurry samples. NIR measured data pretreated with 1st derivative before correlation analysis and samples measured at laboratory conditions in spectral range of 4300-10 000 cm^{-1}

In addition to correlation analysis, temperature and moisture effect can be minimize with using preprocessing methods. In the Figure 41a is presented an example of original mass sample measured spectra at 18, 23, 30, 40 and 50 °C temperature. Overall, lignin and kappa number both have similar correlation to each other. Now an example case where is presented effect of temperature on spectra of one sample changes to make it easier to visualize. In the Figure 41b is presented original sample spectra in different temperatures after pretreatment with using 1st derivative. The purpose for this is justify that 1st derivative is also able to react to the spectrum change rates as well as it removes the baseline effect and change absorption sensitivity better. There is occurring always systematic change when temperature varies but with 1st derivate it no longer occurs so strongly which improve calibration model.
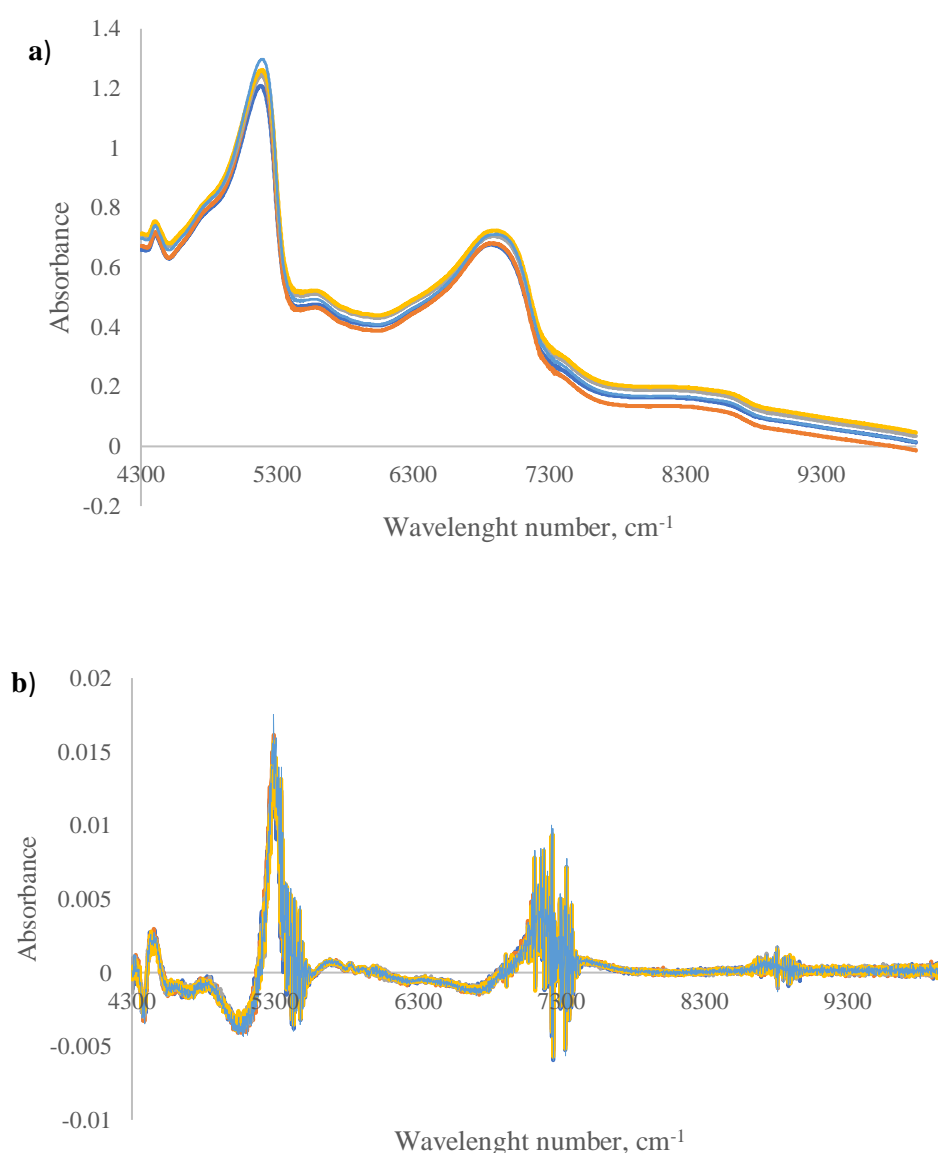

Figure 41.   (a) Original sample spectrum at 18, 20, 30, 40 and 50 °C
(b) 1st derivate of spectra (a) measured in different temperatures

In the figure 42 is presented 1st derivate correlation plot against original measured samples which model preprocessing effects more clearly. Blue line is 1st derivative correlation and orange is original sample data measured at 18, 20, 30, 40 and 50 °C temperatures. This indicates that temperature effect has been eliminated with preprocessing spectra by 1st derivative based on one sample by achieving higher correlation to reference data. Presented spectral range is delimited in the Figure x from 7000-8000 cm^{-1} to visualize it more clearly. However, it should be noted that this is an extreme example and in the actual process and the used calibration in it, the temperature would not vary so extensively. The end conclusion based on these correlation plots was that only certain wavelength area cannot be chosen based on good R^2 values to improve calibration model effectiveness.
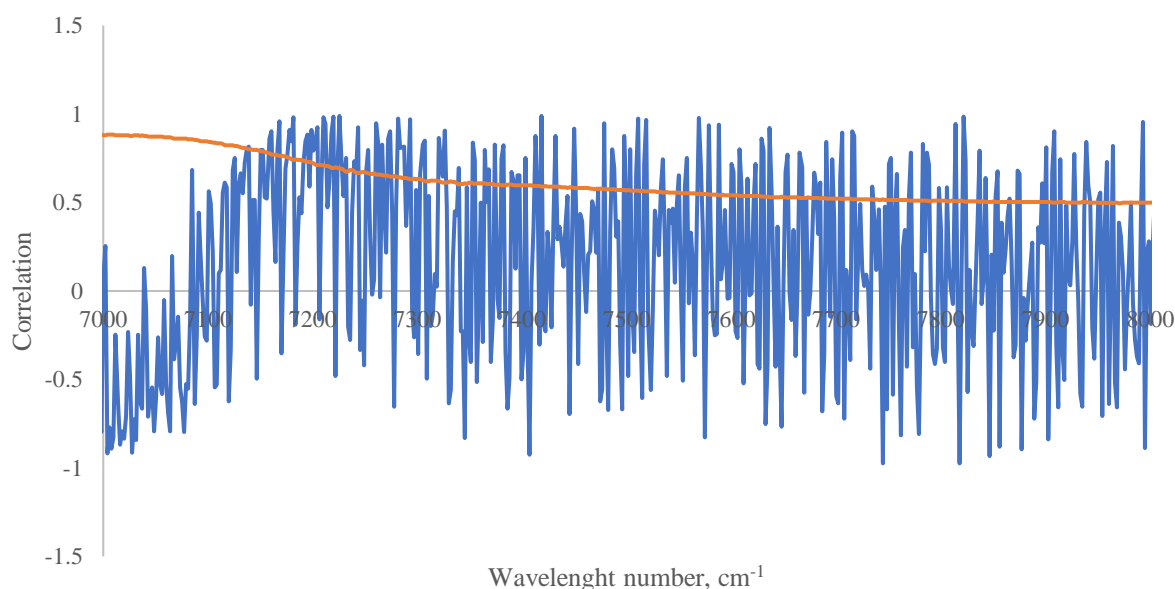

Figure 42. An example sample correlation effect to temperature changes where blue is first derivative correlation and orange sample measured spectral data at 18, 20, 30 and 40 °C

13.3 Calibration model sensitivity

After correlation plot analysis to normal samples at laboratory scale it was applied to the extreme examples when moisture content and temperature changes and how calibration model could be modified to predict these extreme samples and possible process variations better. This comparison to normal condition samples helped to evaluate which of these changes, temperature or moisture affect the most to original pulp sample measurements. Original samples dry matter content is approximately within level of 50-55% where water peaks effects should be less intensive compared to analyzing the slurry samples. Six selected samples with good capture were chosen of total 28 calibration set and used to test moisture content effects.

84

Dry matter content variations were for two analyzed samples 28 %, 31 %, 34 %, 38 %, 42 % and 48 %. In the Figure 43a is presented dry matter content changes as variation samples to original mass calibration model and in the Figure 43b is presented residuals plot for kappa number. As before stated that total spectrum area includes important information and it is hard to split smaller pieces to certain component, now due to increased water effect calibration model spectrum range was modified to achieve better fitting for validation samples based on coefficients. From the calibration figure can be seen that new validation samples (cross) are spread around the calibration sample line but still try to model along the curve. However, results presented in the Figure 43 were improved when comparing the starting point in previous presented in the Figure 35.
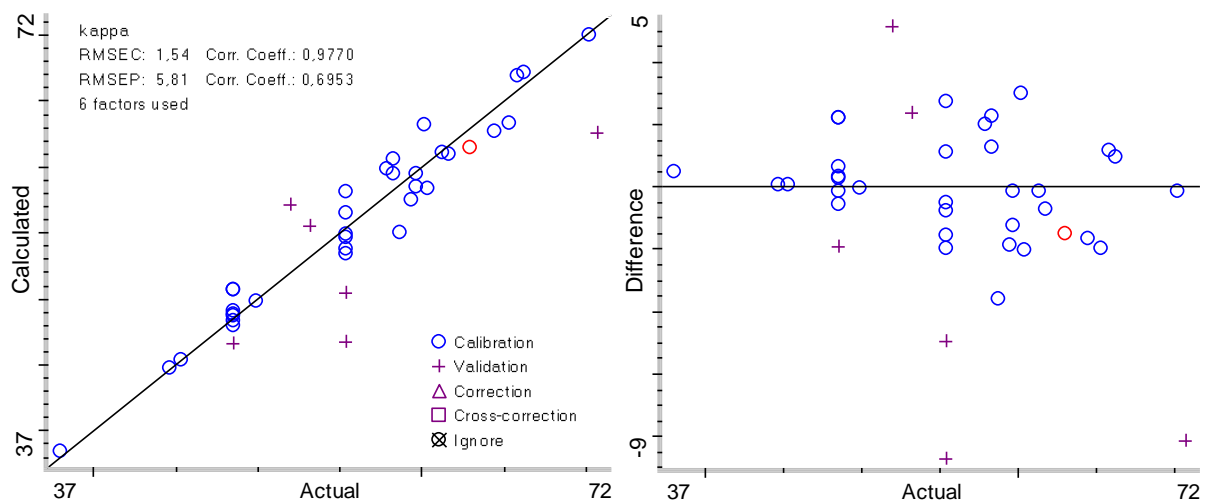

Figure 43.   (a) Calibration for original biomass samples in normal process moisture content with extreme moisture content variation validation samples marked as cross
(b) Residuals plot for calibration

Moisture content variation to kappa number and lignin from created calibration model results are presented in the Figure 44. From the Figure 44 can be seen that kappa number linearization is little higher than lignin which indicates that for the lignin dry matter content variations effect more during validation process. Even if neither R^2 are not that high but behave still linear and apart from the moisture variation, samples are able to be predicted with NIR. Of course, the final model in real process would not include this scale moisture variations and that would be considered and set certain moisture content correction including inside the calibration model.
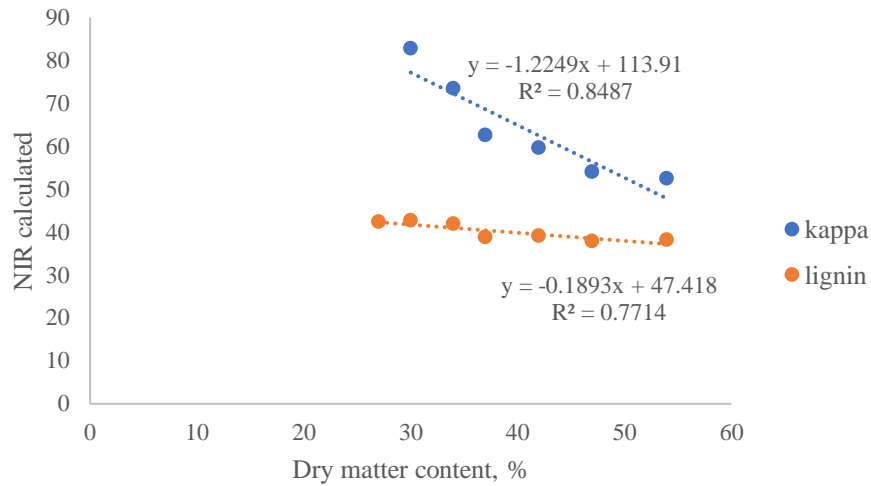
Figure 44. NIR calibration calculated values against dry matter content -% changes for predicted samples

Temperature models were created at 18, 20, 30, 40 and 50 °C for original biomass sample sensitivity including total spectrum range area. In the Figure 45a is presented calibration for lignin that is created to all temperature variations and residuals plot in the Figure 45b. Temperature model also shows that all samples are fitted quite well along the calibration curve against the reference data and R^2 value is high too, 0,9861. Naturally there is some variation due to temperature variation, but overall calibration model is stable without validation samples including only calibration ones.
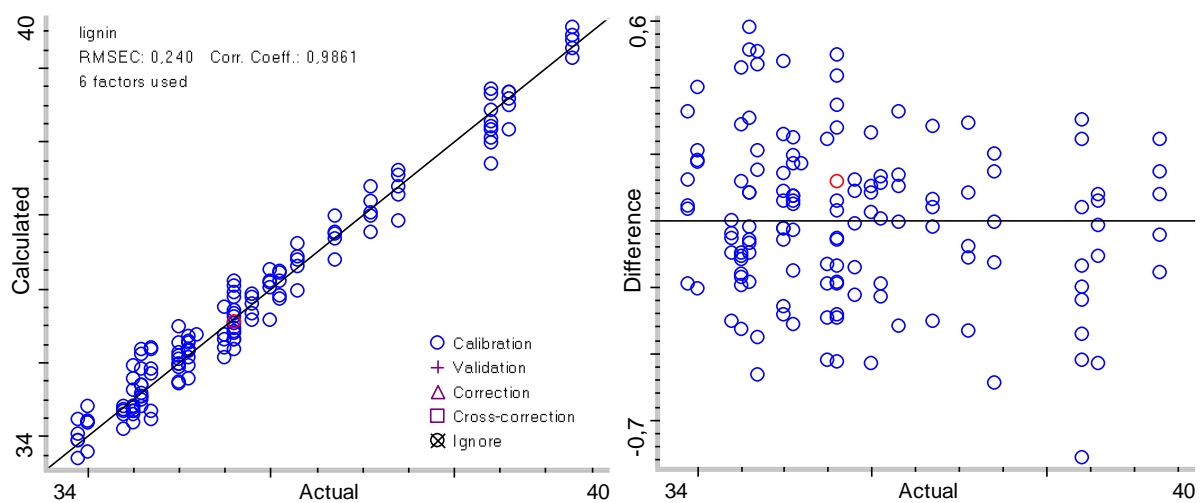

Figure 45. (a) Calibration for lignin original biomass samples including all samples with temperature variations at 18, 20, 30, 40 and 50 °C
(b) Residuals plot for calibration lignin samples

In the Figure 46a is presented kappa number calibration and 46b residuals plot to original biomass samples with temperature variations. Kappa number calibration plot is like lignin and models these samples quite well within high R^2 value: 0,9867.
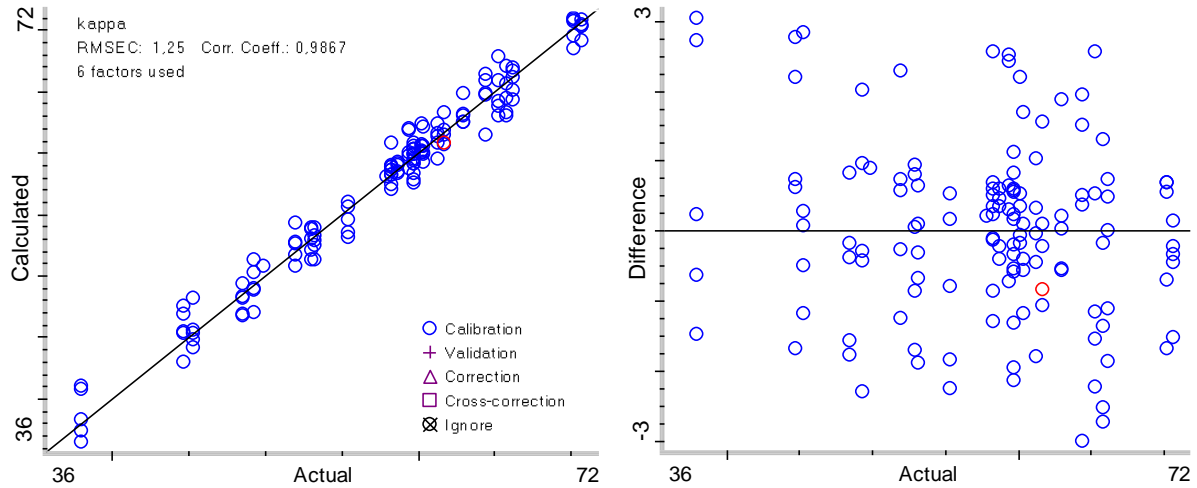

Figure 46. (a) Calibration for kappa number original biomass samples including all samples with temperature variations at 18, 20, 30, 40 and 50 °C
(b) Residuals plot for calibration kappa number samples

After this temperature effect and model was validated with each temperature and model working principle was not that workable and straightforward. This model was not able to predict new samples without including temperature variation also in calibration. This led into that common model including samples in several reasonable temperatures and variations could be included to the calibration model which further utilized when creating online calibration to final online loop experiments. This also stated that calibration could be created in certain temperature level to predict changes the best possible way but extreme conditions such as 18 °C and 50 °C have too high variation in these biomass sample measurements. Calibration model including samples at 18, 20 and 30 °C for kappa number is presented in the Figure 47a and residuals plot for samples in the Figure 47b with relatively high R^2 value.
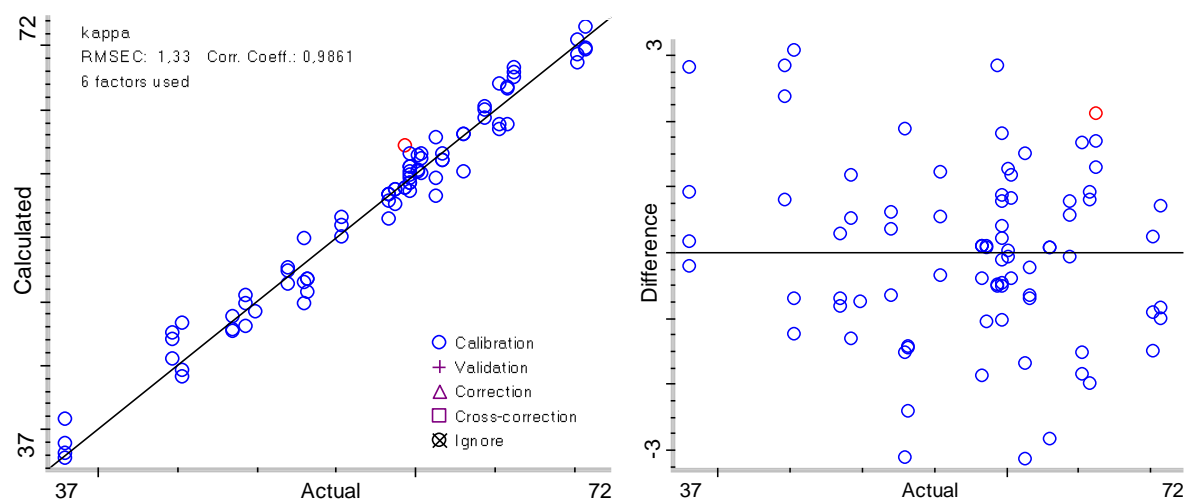
Figure 47. (a) Calibration for kappa number at 18, 20 and 30 °C to original biomass samples
(b) Residuals plot for kappa number calibration

In the Figure 48a is presented lignin common calibration model at 18,20 and 30 °C and residuals plot in the Figure 48b. Results are like kappa number and R^2 for the model is high apart from the temperature variations.
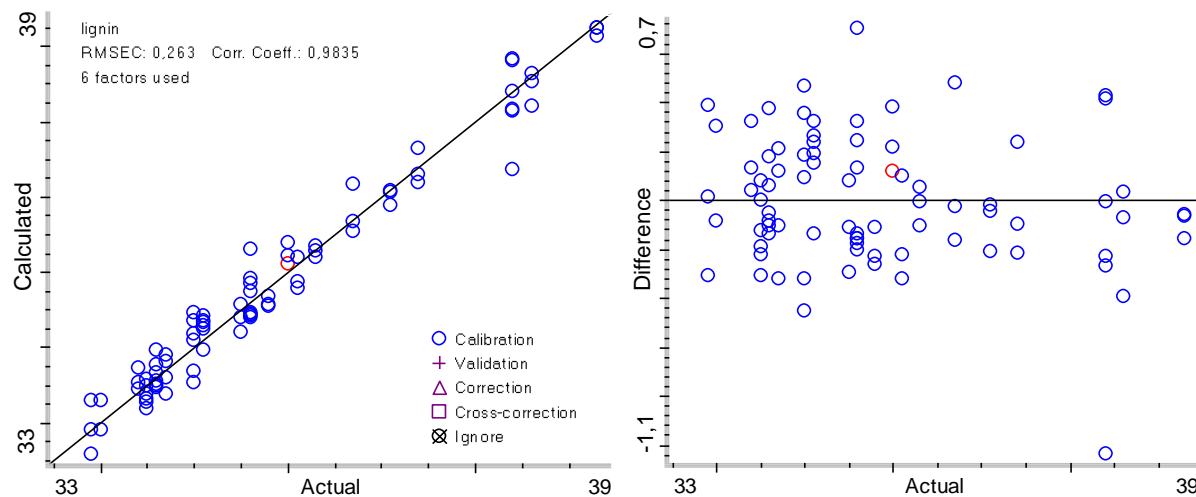

Figure 48. (a) Calibration for lignin at 18,20 and 30 °C to original biomass samples
(b) Residuals plot for lignin calibration

In addition to temperature and moisture content testing for original biomass fractions, slurry sample calibration sensitivity testing was applied too. Slurry sensitivity tests focused on solid fraction total solids TSS-% changes and liquid fraction insoluble changes. Selected samples for slurry test were taken from the original calibration sample set where six samples with good capture of reference values of lignin and kappa number were used for this purpose. In the Figure

49a is presented starting point calibration for kappa number TSS 10% slurry and residuals plot in the Figure 49b. Calibration samples fit well and have high R^2 value: 0,9978.

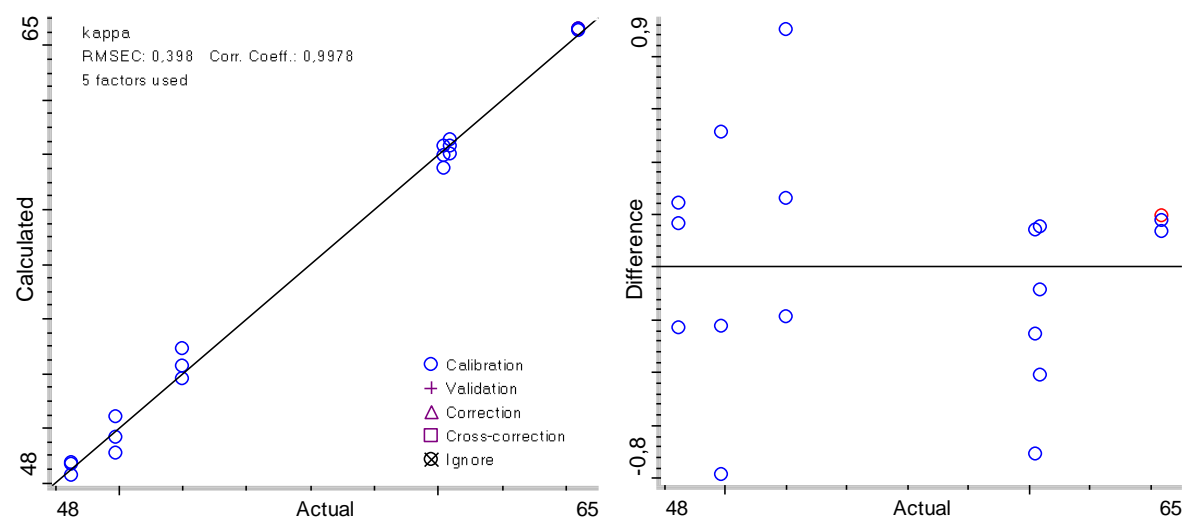**Figure 49.**   (a) Calibration for TSS 10 % slurry for kappa number
(b) Residuals plot for slurry calibration samples

In the Figure 50a is presented lignin calibration for starting point TSS 10 % slurry and residuals in the Figure 50b with relatively high R^2 value: 0,9969.
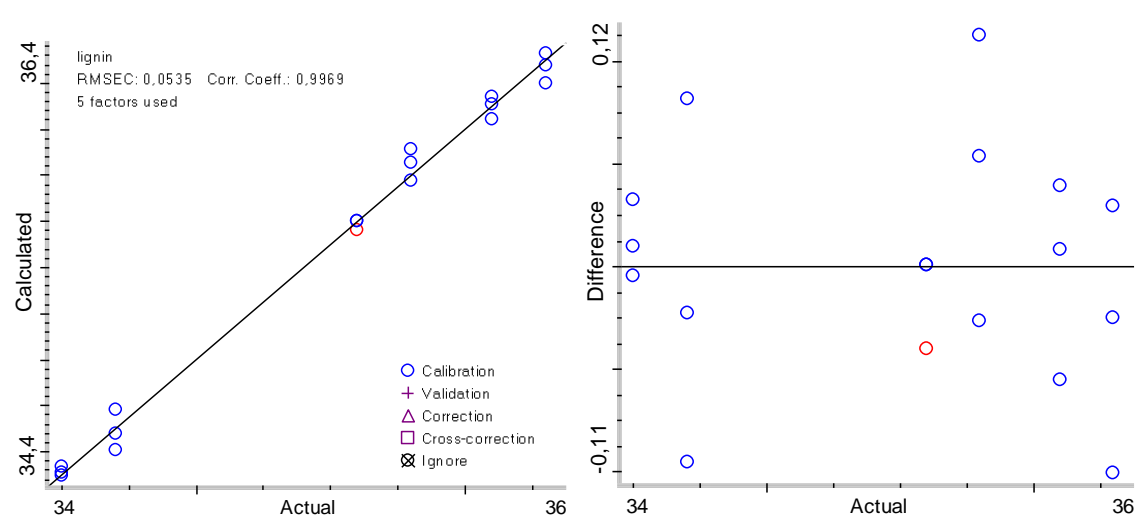

Figure 50.   (a) Calibration for TSS 10 % slurry for lignin
(b) Residuals plot for slurry calibration samples

Presented model above was tested when validation samples between TSS 12-17 % were used. Brix in slurry calibration samples was set constant, at level 10 %. In the Figure 51a is presented calibration with validation samples and residuals plot in the Figure 51b. Even calibration model R^2 is high, model is not able straight predict this variation occurred in slurry total solid content change. In the Figure 51b is presented model that includes in calibration set with TSS% variation samples and validation sample set. Now model goodness of fit is little worse but new samples model better because calibration captures the variation in total soluble solids. However, validation includes still samples which are not correctly predicted which can be seen in the Figure 51b.

Figure 51. (a) Calibration for TSS 10 % slurry and validation samples including TSS % variation between 12-17 %
(b) Calibration model with TSS% variations for kappa number and validation slurry samples

Lignin validation results are presented in the Figure 52a where calibration model is not including all TSS% variations. In the Figure 52b is presented calibration including TSS% variation samples for lignin and new validation set. From the 52b plot can be seen that now some of the lignin samples are now estimated better than at first in the Figure 52a. Lignin samples are estimated similar like kappa number when calibration also includes slurry samples of TSS 12-17% and validation sample set.
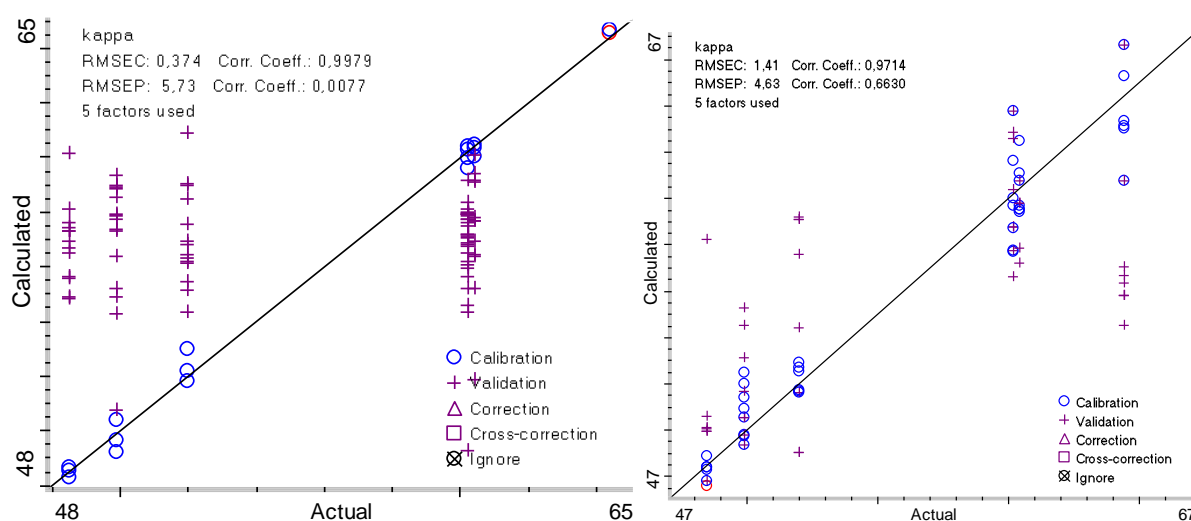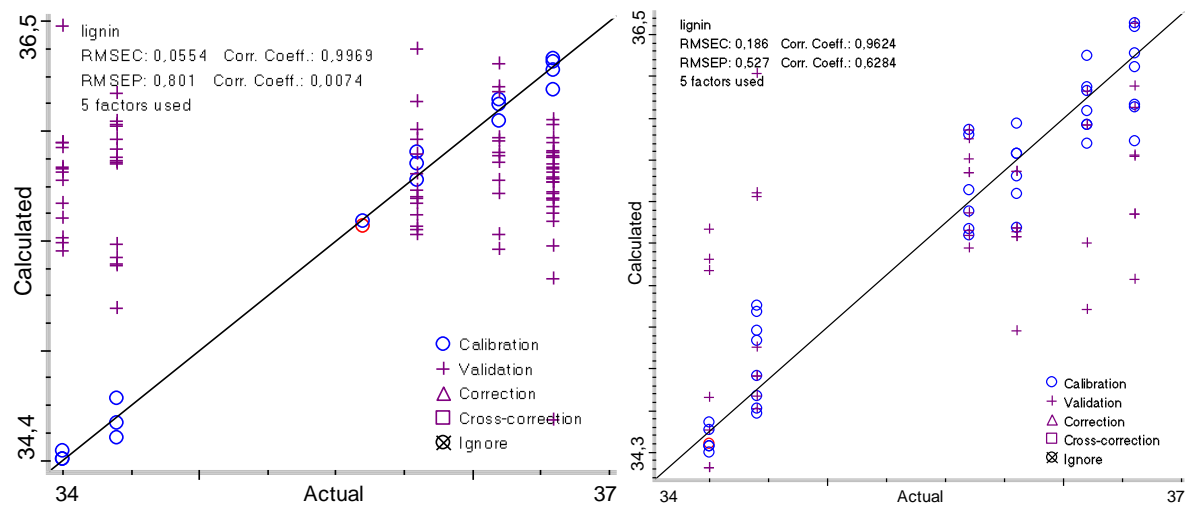
Figure 52. (a) Calibration for lignin TSS 10 % slurry and validation samples including TSS % variation between 12-17 %
(b) Calibration model including different TSS% slurries and validation slurry samples for lignin

After this same model tested when brix-% in insoluble phase is changing. In the Figure 53a is presented calibration model for kappa number with validations samples including brix-% changes of 10 %, 12 %, 15 % and 17 % while total solids amount was constant. In the Figure 53b is presented new calibration including insoluble changes and validation samples.

Figure 53. (a) Calibration for kappa number and validation samples with insoluble variation of 10 %, 12 %, 15 % and 17 % and validation samples
(b) Calibration model including different slurries and validation slurry samples for kappa number

91

Lignin calibration result with validation samples is presented in the Figure 54a. Calibration including also different samples with insoluble changes is presented with validation samples in the Figure 54b.
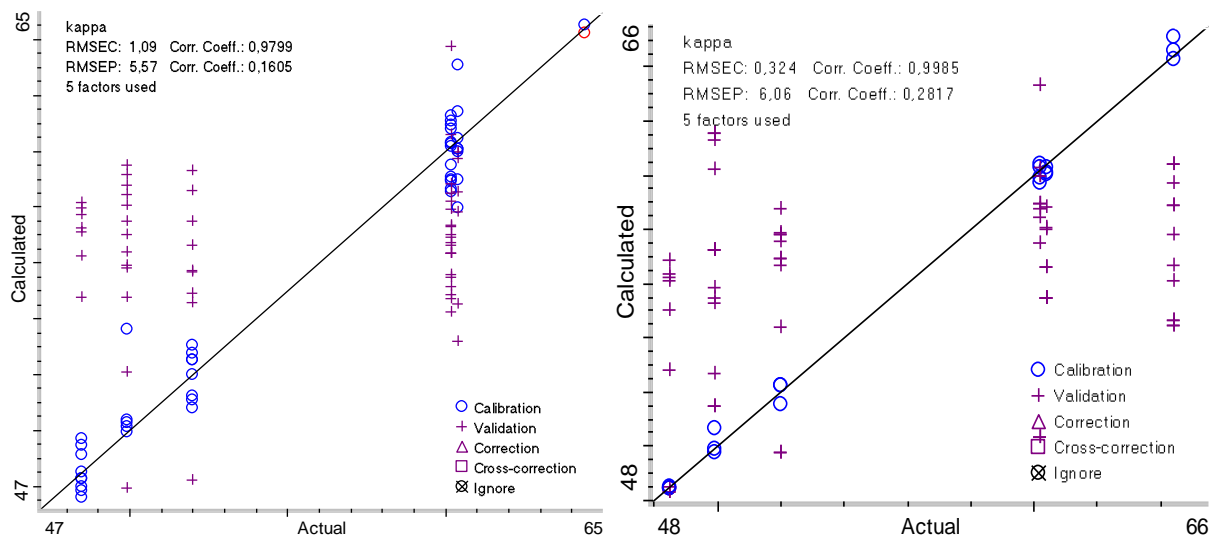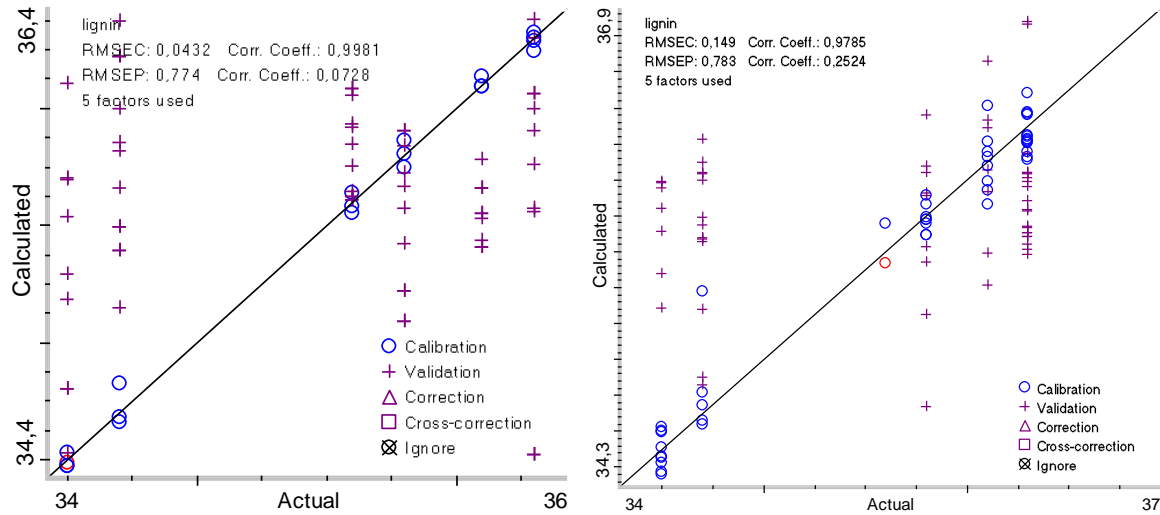

Figure 54.   (a) Calibration for lignin and validation samples with insoluble variation of 10 %, 12 %, 15 % and 17 %
(b) Calibration model including different slurries and validation slurry samples for lignin

Based on these slurry sample sensitivity test conclusions were that both have impact on validation process, total soluble solids as well as liquid fraction changes that diffuse-reflectance probe can notice. Overall previous presented slurry validation models did not predict the optimal solution, but the main target was to investigate the NIR sensitivity to these changes and it turned out to be that it is possible to see these changes in spectral data as well as in calibration models. In real process temperature does not vary this much. Temperature impact on measurements is smaller than moisture content. Even if these extreme validation samples did not fit perfectly, NIR showed potential its sensitiveness to these changes in slurry samples and calibration models can be further improved.

13.4 Calibration model improvements

Created calibration models to each sample set showed potential especially for original biomass and slurry set were important fractions when thinking the goal which is online analysis in real process environment. Even if everything did not estimate perfectly, considering process conditions such as temperature and moisture variation, calibration models were improved from the beginning. Wider models including more variation and samples with good capture can

improve models further. Multivariate statistical process charts can be utilized also creating predicative calibration models in laboratory scale to see how spectral information is changed. An important factor developing NIR calibration model is used calibration sample references and those values. Now used samples were collected within long time which could have affected to the results a bit but apart from this model were able to predict new samples. NIR huge potential lies on that it can see the bigger and deeper picture of the sample variation and interactions, not only the surface of the sample or one chemical bond change. This is good feature especially while analyzing this kind of biomass that has complicated chemical structure and not only one feature is changing.

There is possibility to create model that has correction for temperature, moisture content or insoluble change with brix-% included to the calibration model to minimize these effects. As a whole calibration is strictly limited to work in certain process conditions that occur in that statement which smooth external changes is calibration model. Spectral pretreatment as 1st derivative is already doing a lot what noted earlier and created calibration models if it is not possible to use in strict model process conditions. Process charts provided warnings can also be utilized in calibrations models when calibration area is exceeding over process charts limits. Wider calibration model was tested during online tests including three different temperatures and these laboratory test modifications gave important base towards online test experiments. Process charts working principles in practice is discussed more in the next chapter.

14 ONLINE LOOP TESTS

Online loop tests were final experimental setup to evaluate NIR feasibility and sensitivity to real process changes that could occur during operation after biomass pretreatment step in lignocellulosic biorefinery. Slurry samples were chosen to be utilized in these experiments due to easy inline setup of NIR. Slurry was made straight to the mixing tank including constant 10 % C5 sugar solution as insoluble content, water, and pretreated biomass. Biomass dry matter content was between 52-54 % depending on used biomass sample. In the Table X is presented the process variable changes, temperature, pressure, and operation time for each loop tests trials. Process variables that were investigated during these continuous online measurements were biomass raw material change to another wood specie, temperature operation changes, total

solids content changes, adding different type of pretreated biomass and final continuous overnight loop without process changes to see if fouling occurs.

Table X   Online Loop test process variables and operating setups for each trial

Loop test	Process variable	Temperature, °C	Pressure, bar	Operation Time
1	• different raw material doses from beech to oak	60	2	2,5 h
2	• temperature changes • flow rate changes	40–65	2	6,0 h
3	• total solid content changes	60	2	5,0 h
4	• adding different type pretreated biomass	60	2	2,0 h
5	• continuous overnight loop without process changes, fouling	60	2	1,5 d

Experimental setup was consisted of 60 liters mixing tank, membrane pump, computer and online NIR with probe that had Swagelok coupling. Mixing tank bath was heated with hot water and loop tests temperature was measured straight from the bath of the tank. System was not vacuumed and therefore, the top of the tank was covered with hard aluminum foil to prevent biomass evaporation during heating and operation. Samples were taken from the pump valve, which created some pressure and foam to the slurry samples. However, pressure of all loop test was low which was target because of small pipelines and this scale operation conditions. The used natural pressure formed by the membrane pump and mixing ensured that slurry is running properly along the pipelines and back to the tank. Online loop test setup at pilot scale is presented in the Figure 55a and cross section of inside the mixing tank and slurry in the Figure 55b.
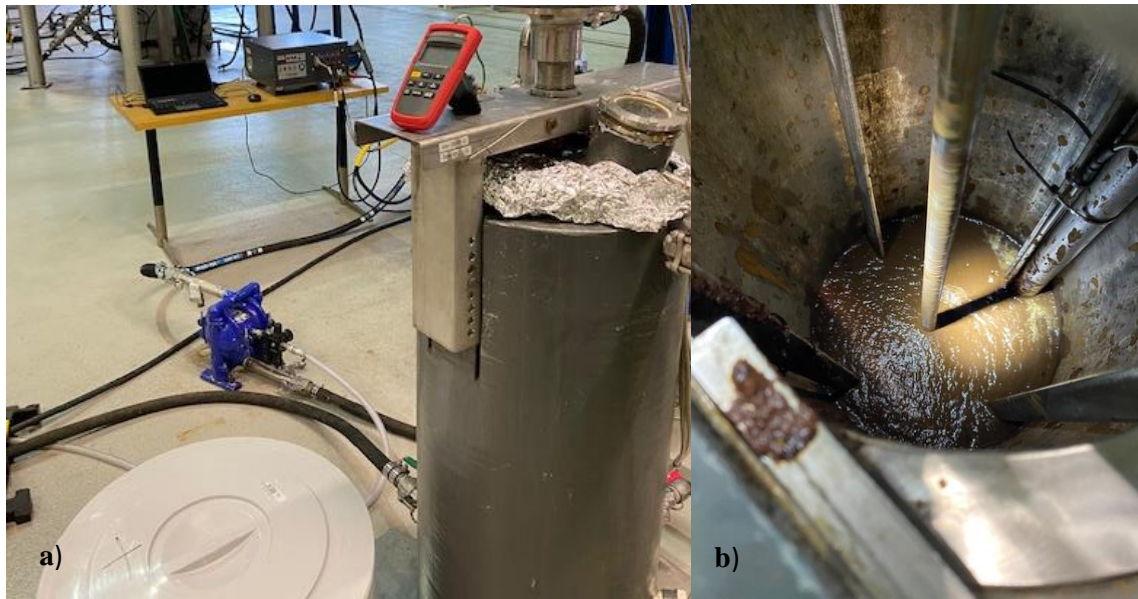
Figure 55. (a) Online loop setup where heated mixing tank, membrane pump, computer and online NIR
(b) Cross section inside of the tank, slurry sample and electronic mixer

The connection to the inline setup for the probe is presented in the Figure 56a and above in the Figure 56b. Probe was installed at 90 degrees angle to the process line to ensure the best possible contact between the probe tip and moving slurry sample. Approximately 1,0 cm of free space wanted to be left from the pipeline wall to the probe tip that slurry moves and does not accumulate to the measurement point. Nonetheless, in the online loop tests probe was accidentally installed too close to the pipeline that was originally planned. After all measurements probe was not fouled but due to probe installation too near to the pipeline wall where bigger particles over 1,0 cm may not have flow past the probe tip. This could have affected to significantly the spectral data intensity collected from online loop tests and probe contact.

Figure 56. (a) Online probe installation with Swagelok at 90 degrees to the loop
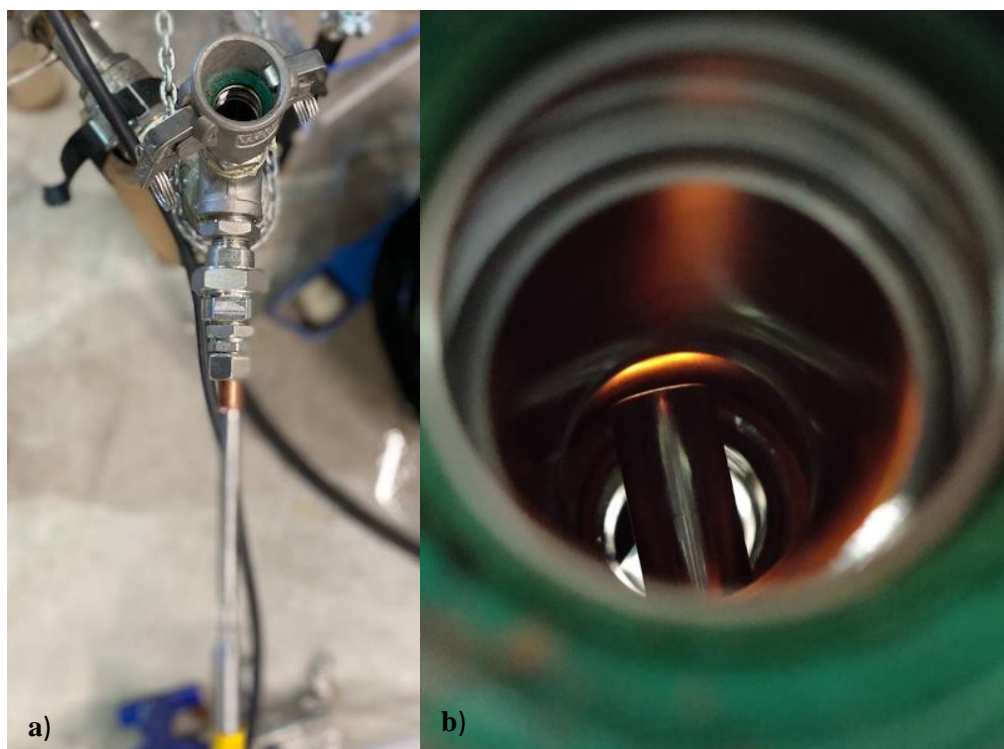(b) Connection to the pipeline from above where free space left from the pipeline wall
for slurry to move forward and connect with the tip of the probe

Slurry fraction was chosen to the online measurements due to its close application to the real process conditions and future needed measurements from this biomass. Original solids and washed solids online measurements were not possible to create and demonstrate real process conditions such as temperature and measurement place. Calibration model to slurry was created in three different temperatures at 58, 60 and 62 ℃ with online probe and moved after this to the online NIR and its automation. Three different calibration temperatures were used because of real process application environment and its changes. At laboratory scale measurements calibration sample set was 28 samples, now at online calibrations only 6 samples which were measured four times to increase model. These six samples were selected from the sample set of 28, to predict best process changes and these samples had the biggest variation of concentrations as well. The online calibration results for lignin are presented in the Figure 57(a-d).

Figure 57. (a) Calibration for lignin at 58, 60 and 62 ℃ without validation samples including total spectrum range, 4300-10 000 cm^{-1}
(b) Residuals plot for lignin samples
(c) Factor plot for lignin
(d) Calibration PRESS and RMECV table for lignin online calibration results to optimal used factor number

Kappa number online calibration is presented in the Figure 58(a-d). There can be seen that both kappa number and lignin behave quite similar, and model has variation between sample fitting whether temperature difference is not big, only four degrees. Also, number of used factors to create calibration model based on PRESS results was quite big but acceptable.

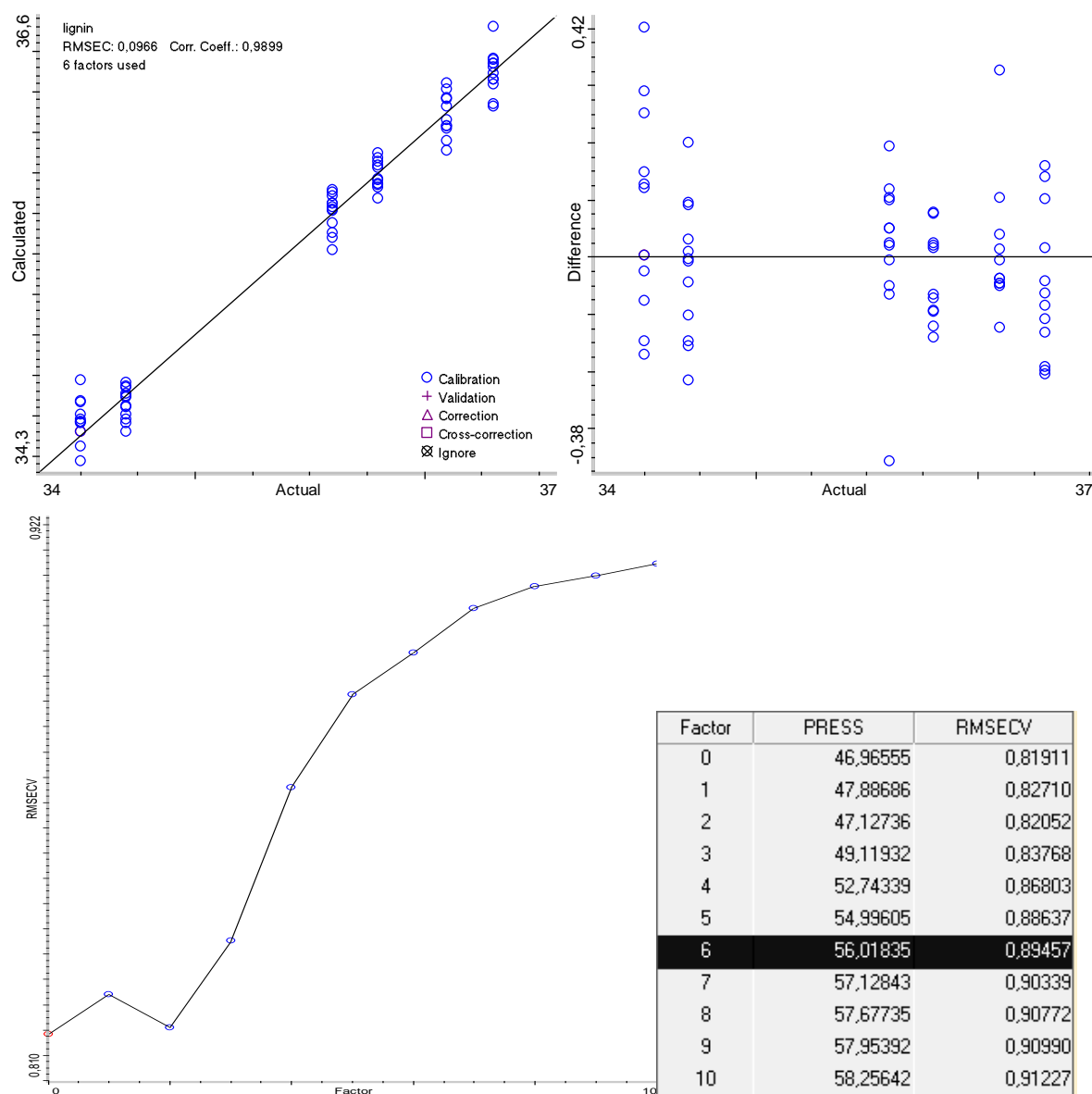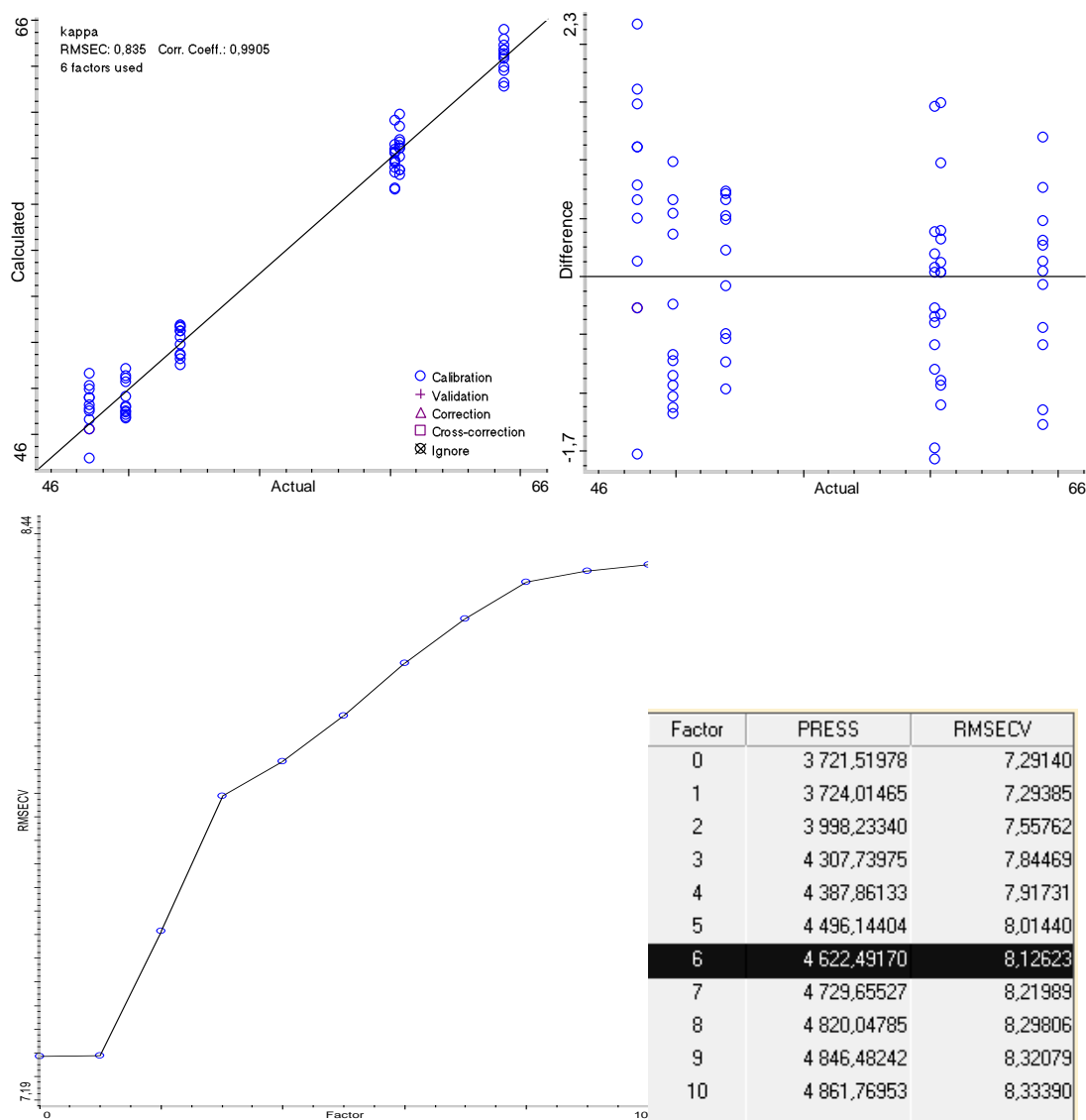**Figure 58.** (a) Calibration for kappa number at 58, 60 and 62 °C without validation samples including total spectrum range, 4300-10 000 cm^{-1}
(b) Residuals plot for kappa number
(c) Factor plot for kappa number
(d) Calibration PRESS and RMSECV table for kappa number online calibration to optimal used factor number

14.1 Online loop (OL) tests

First online testing loop test was based on changing the beech biomass to totally another type of wood specie slurry at 60 ℃ that was made of oak. First normal level slurry containing 11 % of total solids was heated up to operation temperature. Disturbing biomass at same concentration was added to the mixing tank slowly by little doses between 5-30 % of volume of original slurry sample and after every addition of different slurry, sample was taken from the membrane pump. Before taking the sample, a moment was waited that slurry composition and temperature are stabilized. From these taken samples, quick dry matter content and brix value were measured straightaway from pressed slurry filtrate and later reference measurements in the laboratory based on NREL. Brix for this purpose is to describe average of the soluble solids content. Overall, the brix content for slurry in the loops was set to be constant, 10 %. After this experiment, mixing tank was emptied after experiment because the biomass could not at this point be reused like other in these experiments were utilized and after every loop test tank was not emptied if possible.

Second online loop test was related to changing the operation temperature below and above the optimum processing temperature considering real life process that would be for the slurry sample 60 ℃. In addition, this second loop included small changes to flow rate which was made by the membrane pump to increase or decrease material flow. Now used slurry contained TS 12 % which was quite easily pumped and close to the real process environment. However, difference between loop 1 slurry containing TS 11 % could be seen while mixing this slurry portion. Temperature changes were made at 20 minutes timelines and for this water was used to heat or cool down the mixing tank reactor bath. Samples were taken from the start and after experiment, to see if these changes have had effects to starting point or to the end. Brix and quick dry matter content were measured straight.

After this in the 3rd loop test, TS changes were tested and utilized 2nd loop slurry. The used TS-% levels for this test were 12 %, 11 % and 10 %. Samples were taken from the starting point and after every dose where brix and quick dry matter content also measured straight. Final slurry which contained TS 10 % left to spin overnight to continuous loop to see whether possible fouling occurring and if it is affecting to the NIR results within this still relatively short testing time. The four online loop, was quite like the second one, but now added slurry had

different pretreatment time and process history. For this loop tests adding of another slurry were made at same level, from 5 % to 30 % and samples taken from all these steps.

Target of these all loops were to see how measured NIR spectral data and online NIR equipment given trend change when process conditions at not stabilized. An example of online NIR given trend is presented in the Figure 59a where kappa number and lignin measurements are presented. Example result online spectra is presented in the Figure 59b where green is measured background spectrum automatically every hour. Add to these, there is available to get result summary of each measured parameters and spectral data summary. Online loop tests and conclusions are presented in the next chapter.
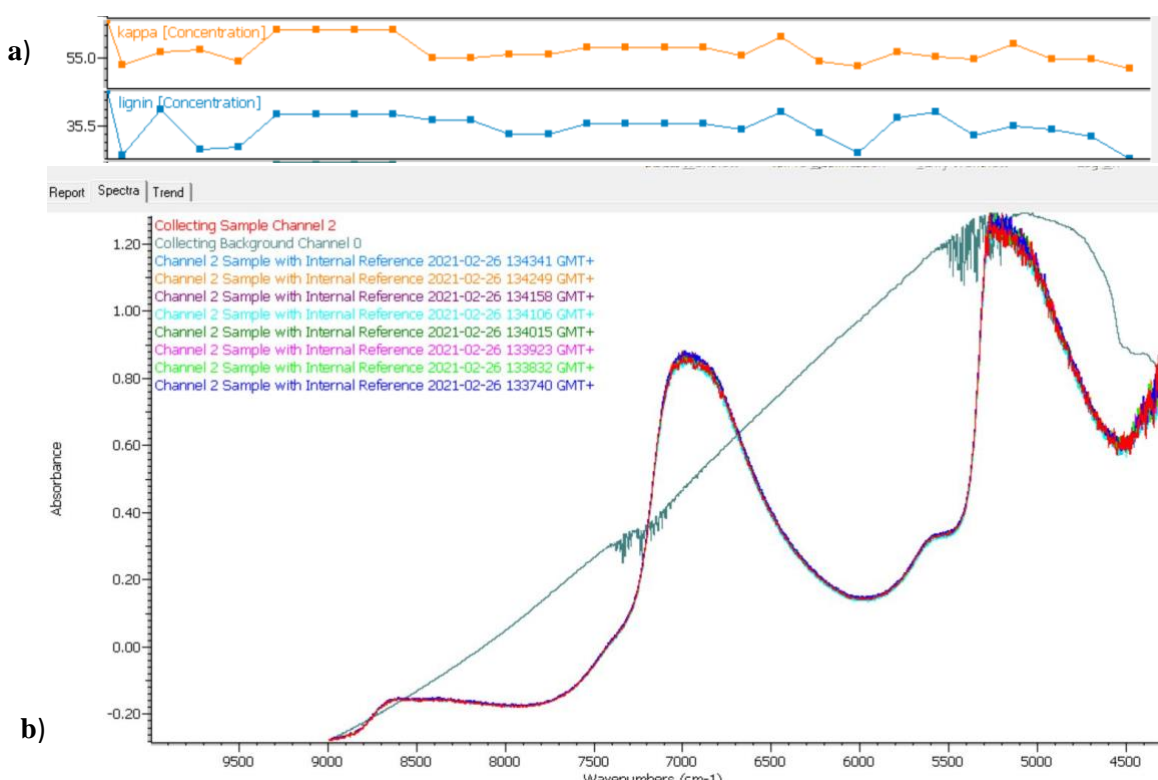

Figure 59.   (a) An example trend plot from online NIR to kappa number and lignin concentrations during online loop testing
(b) Example spectra from the online NIR were automatically measured background spectrum marked green colour

14.2 Process charts and results

Multivariate statistical process control charts were used to final test to NIR at pilot process conditions to measure occurring changes continuous and online. The biggest target was to see how these in real processes used control charts can see changes in measured online spectra. Control charting is widely used to monitor and follow processes, and these charts give statistical control for the process. Process control charts were applied by using MATLAB and principal component analysis toolbox whereas Hostelling's T^2 and $SPE_x(Q)$ charts were used for online NIR loop data sets as individual cases. These process charts provide way to view spectral data without requiring the parameter values and only occurring changes in measured online spectrum is sufficient and that make data modelling faster. First measured online data was moved into MATLAB and scaled into one Excel from individual measured spectra from online NIR. Spectral data was pretreated with using 1st derivative and then PCA-menu toolbox to create first calibration model to data. After calibration model, test set from the online loop changes could be created and then use process charts.

In the Figure 60(a-e) is presented online loop test 1 results where 60a shows original measured online spectral data. Contributions for spectrum changes are presented for the T^2 in the figure 60b and Q in the Figure 60d where x-axis is spectrum wavenumber (cm^{-1}) and y-axis time as sample number. These contributions present systematic change in measured spectrum as color change in the figures at certain wavenumber areas. From both these contribution plots can be seen that changes are occurring at wavenumbers areas starting at beginning 4300 to 5500 cm^{-1} and in the middle area at 6500-7000 cm^{-1} especially in T^2 plot. In the T^2 contribution color changes are clearer than in Q plot but similar in both figures is that the spectral at 8000-9000 cm^{-1} changes are smaller. Multivariate process control charts results are presented in the Figure 60c where Hostelling's T^2 results and in the Figure 60e where results for the residuals as $SPE_x(Q)$ chart is presented. Logarithmic axis is used to see easier the changes in the control chart figures and x-axis presents time as sample number. Calibration area is separated in both control charts as red dashed line. Red vertical lines describe time point when process conditions are changed from the calibration set. Contribution T^2 and Q plots can be compared to the original measured spectra and wavenumber areas.
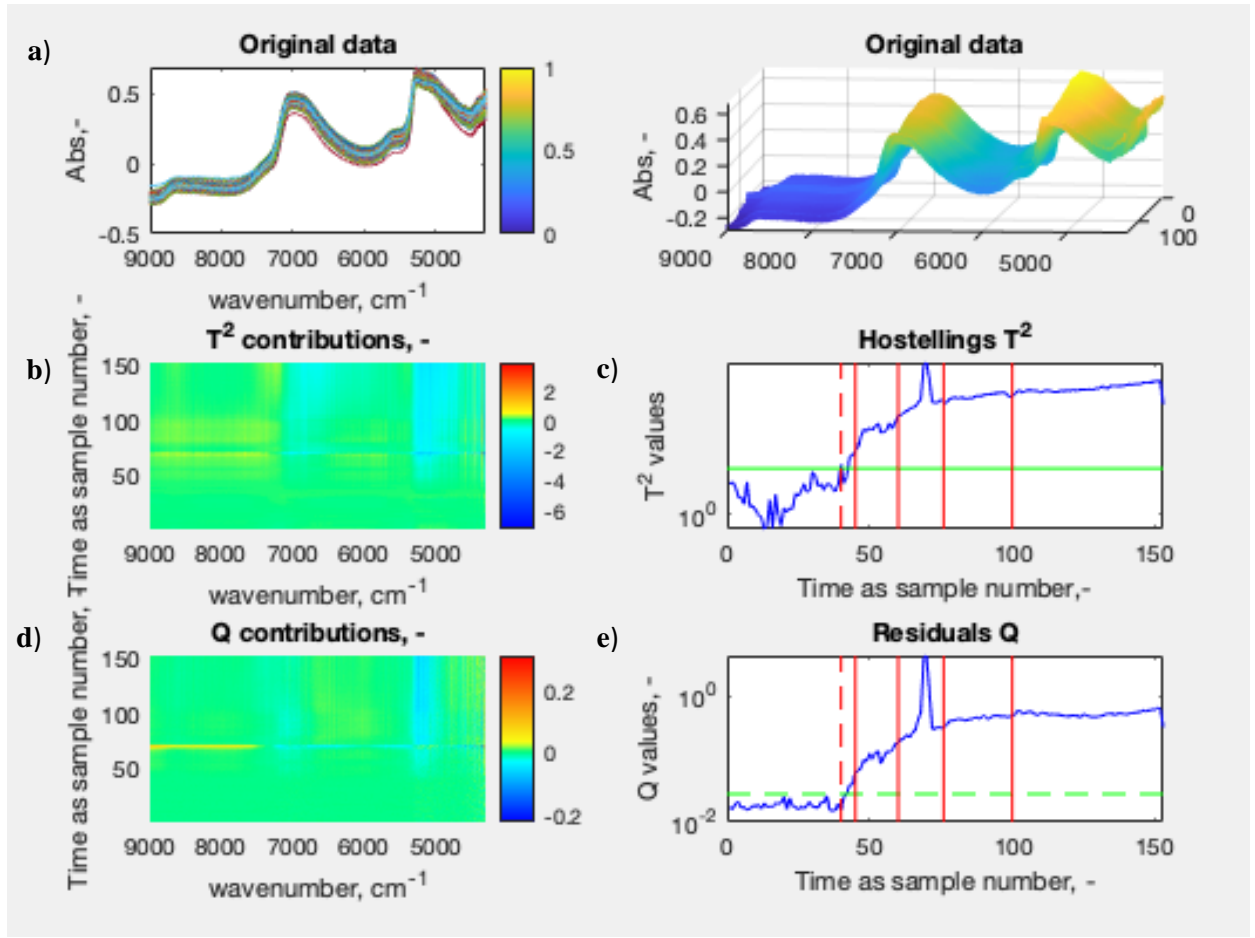
Figure 60. (a) Original online measured spectral data for loop 1
(b) Contribution plot for T^2 where x-axis wavenumber and y-axis time as sample number
(c) Hostelling's T^2 chart results for loop 1 where used calibration set is separated as red dashed line from the test set. Horizontal line describes when model run away from the calibration set. Red lines describe oak dosages to the process
(d) Contribution plot for Q where x-axis wavenumber and y-axis time as sample number
(e) Residuals $SPE_x(Q)$ chart for loop 1 where used calibration set is separated as dashed line from the test set and oak dosages marked by red lines. Horizontal line describes when model run away from the calibration set

Hostelling's T^2 chart shows these cases summary of principal component changes and residuals plot, $SPE_x(Q)$ that part of the model which is left outside the calibration. Multivariate process control charts working principal example can be seen in more details in the Figure 61(a-b) for the online loop 1 in bigger scale than in the Figure 60. This way all online loop collected data is modelled with MATLAB. Calibration set is marked by red vertical dashed line and samples that run over the model as number plot which is separated by horizontal dashed line. Calibration set for the loop 1 included samples 1-40 and after sample 40, changes occurred when from 5 % up to 30 % oak slurry was added to the mixing tank which is presented with vertical red lines. Results showed that process control charts were able to notice sudden process changes within a short delay and after oak doses, process runs out of balance. However, there is one big peak

for samples 69 and 70 and after these samples process is stabilized to the end, after the last dosage of 30 % oak is added to the mixing tank. During these measurements, online NIR react to another solid material and its spectral changes to create response.
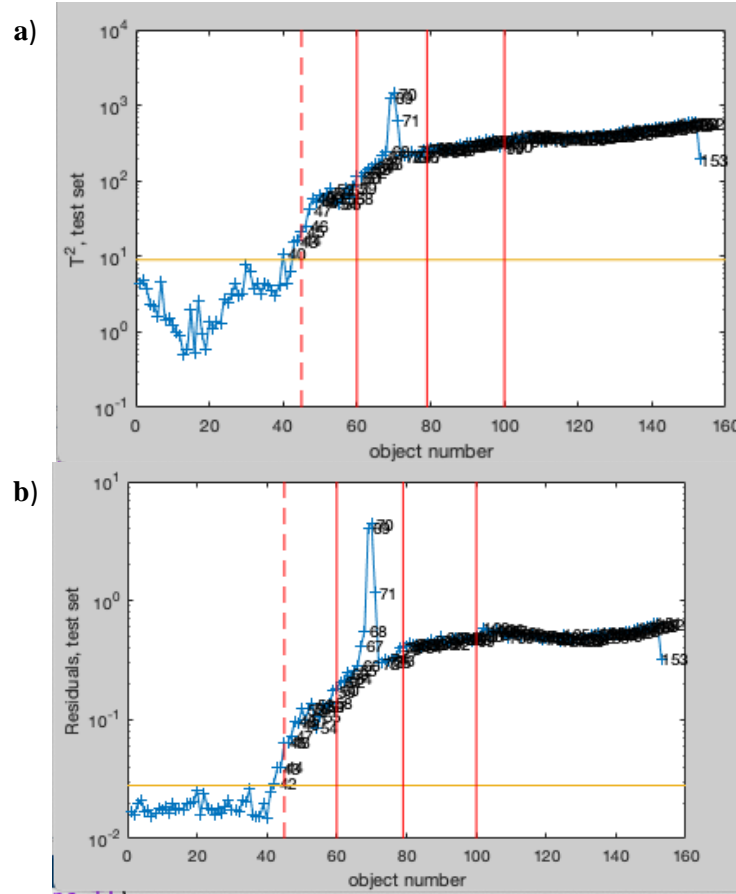

Figure 61.   (a) Hostelling's T^2 chart test set and sample plot including calibration set until sample number 40 marked with red dashed line and oak dosages marked red vertical lines
(b) Residuals as $SPE_x(Q)$ test set and sample plot including calibration set marked red dashed line. Oak dosages marked red vertical lines

Online loop 2 results are showed in the Figure 62(a-e) where temperature and flow rate were changed process variables, slurry itself stayed constant. Calibration sample set during this loop was quite low, including 25 samples at 60 ℃ normal process temperature and normal flow rate. Test set contained samples where flow rate and temperature changes occurred. Temperature steps were 60-65 ℃, 65-55 ℃, 55-45 ℃, 45-40 ℃ and 40-60 ℃ in this order. Flow rate meter was not available to use in this experiment and changes were made with using the membrane pump. Hostelling's T^2 chart is presented in the Figure 61c and residuals, $SPE_x(Q)$ chart in the Figure 61e. There is some variation between process charts because Hostelling's T^2 note the changes occurred in spectral intensity and absorbance due to temperature changes. Accordingly, $SPE_x(Q)$ chart show physical changes when temperature changes. Temperature changes during

this loop were not huge but sufficiently exceeded the calibration temperature and both process charts recognized these temperature variations. This shows that even smaller temperature variations are enough no change spectral data which is reasonable due to vibration changes in molecules affected by temperature. Therefore, NIR measures at this point new possible vibrations and bonds that may have break up during heating the biomass up and down from the normal process condition temperature which is 60 °C. Temperature changes also cause those molecules are twisting and rotating in different directions which is straight correlated to measured online spectral changes. Contributions plot for T^2 in the Figure 62b shows changes in the entire spectral range. Contribution of Q in the Figure 62d did not change significant but from the beginning at 4300 to 5200 cm^{-1}, there is some color changes.
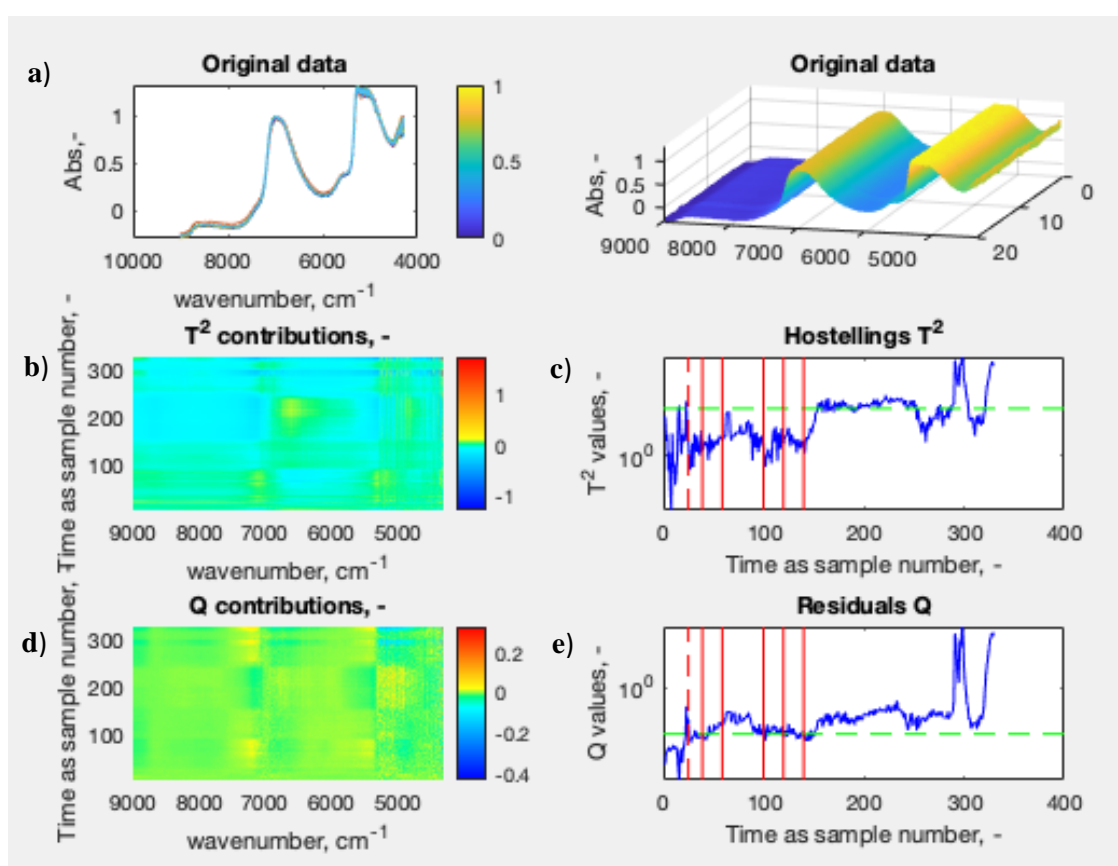
Figure 62.   (a) Original online measured spectral data for loop 2
(b) Contribution plot for T^2 where x-axis wavenumber and y-axis time as sample number
(c) Hostelling's T^2 chart results for loop 2 where calibration set is separated as dashed line from the test set. Horizontal line describes when the model run away from the calibration and vertical red lines when temperature change occurs
(d) Contribution plot for Q where x-axis wavenumber and y-axis time as sample number
(e) Residuals SPE$_x$(Q) chart for loop 2 where calibration set separated as dashed line from the test set. Horizontal line describes when the model run away from the calibration and vertical red lines describe temperature change

Loop 3 results including TS-% content changes from 12 % to 10 % is presented in the Figure 63(a-e). Calibration set included first 50 samples with TS content of 12 % slurry and rest of the samples were used as test set when diluting the slurry and changing total solids content. Liquid fraction soluble content stayed constant at value of brix 10 %. In the Figure 63a where original measured online spectra presented, can be seen flattened peak at 5167 cm^{-1} which was removed and replaced to zero in data set because its affected error to data modelling. Hostelling's T^2 chart is presented in the Figure 63c and residuals, $SPE_x(Q)$ chart in the Figure 63e. From these charts can be seen that at this situation when slurry is diluted and total solids amount is changed in the slurry, residuals chart reacts this more easily and faster than Hostelling's T^2 chart.

Contributions plot for loop 3 are presented for T^2 in the Figure 63b and Q in the Figure 63d. In contribution T^2 plot spectral changes are occurred quite evenly over the measured spectrum range but some yellow color peaks are occurred in the end point areas at 7000-9000 cm^{-1}. Variation occurs also at 5500-6800 cm^{-1} in T^2 contribution plot. Contribution for Q plot at this point is not clear and does not show changes as good as T^2 plot in certain wavenumber areas. However, in the Q contribution some changing can be seen at 5000-5200 cm^{-1} areas and the end of the measurement range is smooth without clear changes. Based on this loop experiment can be noted that changing the solid fraction, diffuse-reflectance probe react this kind of process change even if the scale was not that big but amount of water in slurry increases. During this loop, slurry is diluted when concentration is decreased. Increased water amount cause that chemical bonds in the slurry are looser and structure is changed.
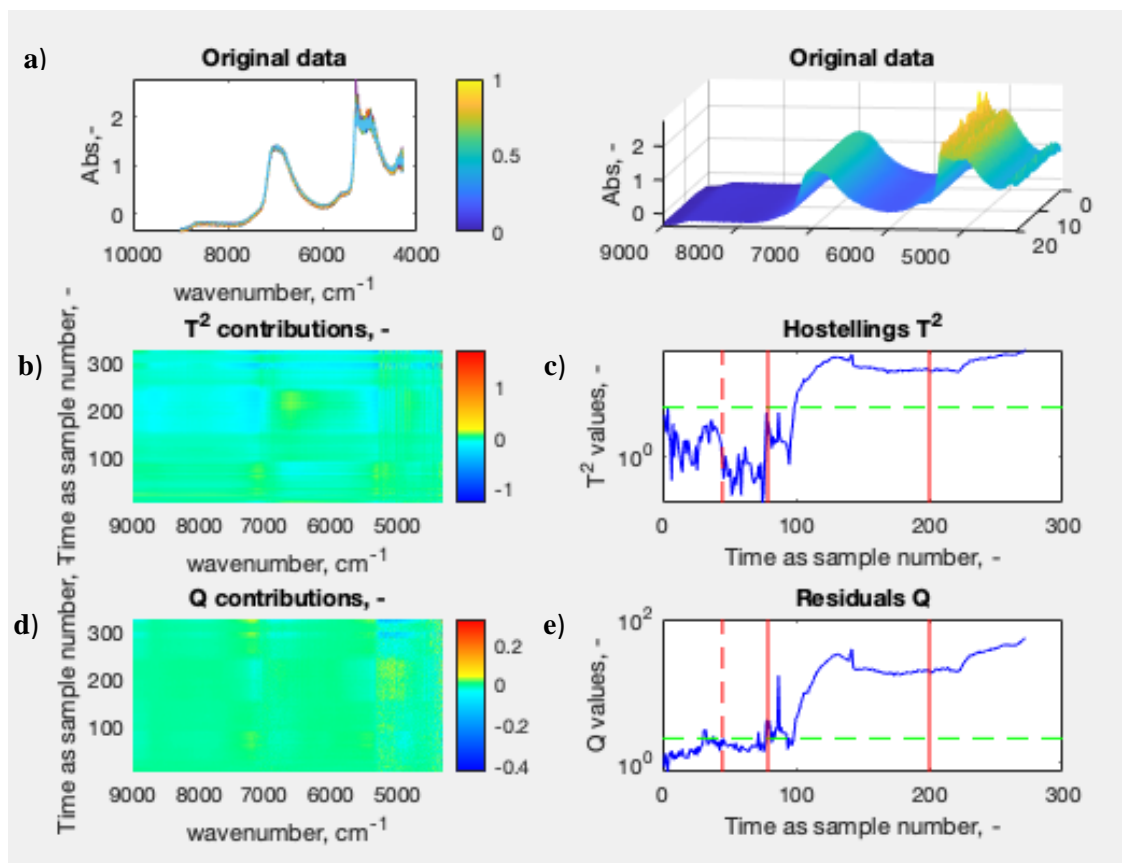
Figure 63. (a) Original online measured spectral data for loop 3 including removed peak at 5167 cm^{-1}
(b) Contribution plot for T^2 where x-axis wavenumber and y-axis time as sample number
(c) Hostelling's T^2 chart results for loop 3. Calibration set is separated as dashed red line from the test set. Horizontal line describes when the model run away from the calibration and vertical red lines solid content changes in slurry
(d) Contribution plot for Q where x-axis wavenumber and y-axis time as sample number
(e) Residuals SPE$_x$(Q) chart for loop 3 where calibration set separated as dashed line from the test set. Horizontal line describes when the model run away from the calibration and solid content changes in slurry

In the Figure 64(a-d) is presented online loop 4 results where two different type of TS 11 % pretreated slurries were mixed for adding another biomass using individual doses from 5 % to 30 % to the mixing tank. Soluble fraction was constant during these measurements and set as brix value of 10 %. Calibration set included 25 samples of first biomass before additions. Additions are marked as red vertical lines in the process control charts. Original measured spectral online data is seen in the Figure 64a. Hostelling's T^2 chart is presented in the Figure 64c and SPE$_x$(Q) chart in the Figure 64e. From multivariate process charts in this loop SPE$_x$(Q) react directly to process change unlike Hostelling's T^2 chart. Contribution plot for T^2 is shown in the Figure 63b and Q contribution in the Figure 63d. Both contribution colour plots are similar, but T^2 includes variation all over the spectrum range. This could be caused that spectral data is flattened when bigger peaks are lowered, and baseline increased. Overall, process

control charts results are quite moderate even if these two pretreated slurries differ in their properties and concentrations.
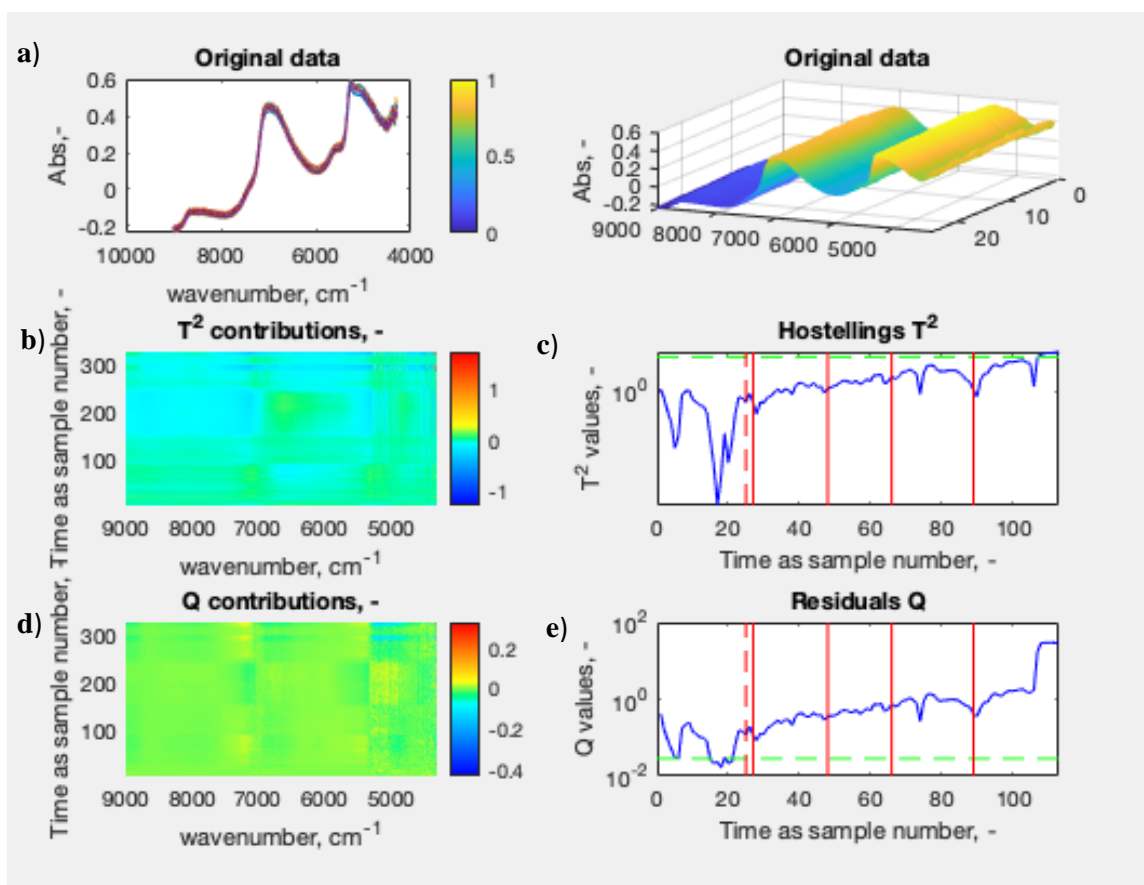

Figure 64.   (a) Original online measured data for loop 4
(b) Contribution plot for T^2 where x-axis wavenumber and y-axis time as sample number
(c) Hostelling's T^2 chart results for loop 4 where calibration set is separated dashed red line from the test set. Horizontal line describes when the model run away from the calibration and red vertical lines when another pretreated material is added
(d) Contribution plot for Q where x-axis wavenumber and y-axis time as sample number
(e) Residuals $SPE_x(Q)$ chart for loop 4. Calibration set is separated dashed red line from the test set samples and vertical red lines describe when another pretreated material is added

During these measurements online, probe was not removed from the inline setup rather water/NaOH mixture wash used to clean pipelines and probe then when material was changed between different loops. Loop 5 target was to evaluate how easily probe will be fouling inline setup which was one question when considering real life process measurements of this slurry after pretreatment step and would it be in the future application to use separate cleaning module inline or as at-line solution. In addition to fouling, other analytical problems that could occur and change spectrum information from NIR during online process, was also studied at this point. In the Figure 65a is presented online probe without cleaning and removed straight from

the inline setup. After this probe was able to be cleaned easily with using pure water. However, Swagelok coupling gained some slurry in it but itself the probe was quite clean afterwards online loop testing. In the Figure 65b is presented slurry sample taken from the membrane pump and some foaming is formed but no antifoam was used during the online loops because in the mixing tank was not foam. It is noted that foam has effect to the spectrum sensitivity and results, therefore inline application it should be considered if foam will be some reason formed during the process. During the online loop measurements, used pipeline and loop was quite short and normal membrane pump was used which could cause foaming. Foaming decreases spectral intensity and sample contact to the probe for achieving accurate spectral data. Pressure changes was prevented with taking care of good flow and mixing the slurry in the tank.
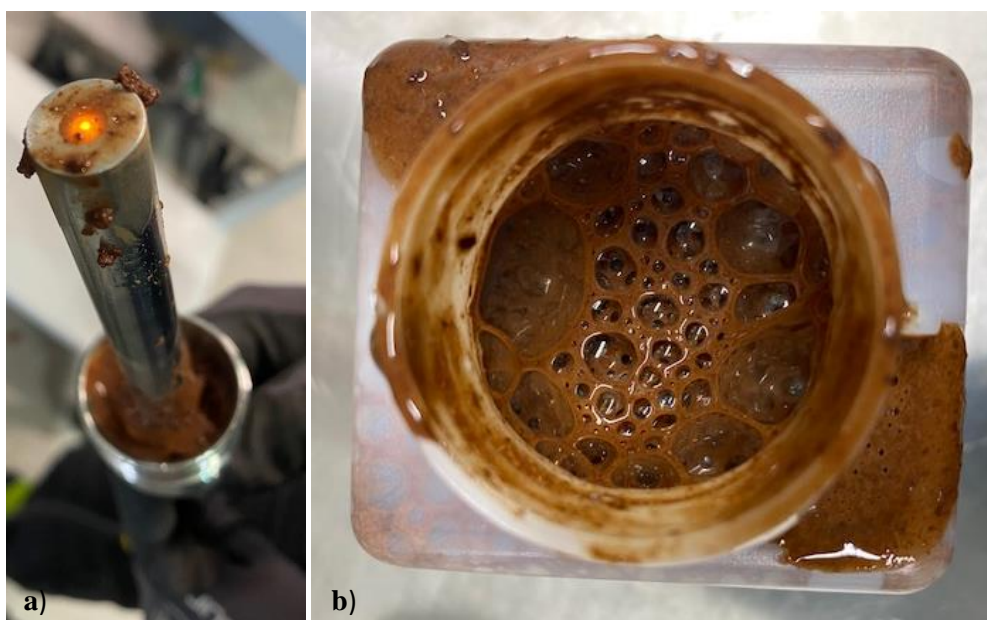
Figure 65.   (a) Probe after continuous online loop 5 and fouling test
(b) Slurry sample and an example of foaming and bubbles after sampling from membrane pump

Online loop tests within this timeframe in addition to laboratory tests showed NIR potential to be used as online analysis tool for pretreated biomass considering that these tests were first time evaluation by NIR as online scale. Nonetheless, these tests showed some drawbacks as well and future improvements. The biggest drawback was that used calibration model did not work properly during online measurements and online NIR results in monitor did not showed reasonable results at that point. Online calibration model calculated values stayed both side of the reference values averages for both kappa number and lignin. Therefore, online loop data was modelled by MATLAB and utilized measured spectral changes. Reasons why created

calibration model did not give successful results could be the model capture that was not enough at this point and was not able to predict new samples. Online calibration testing and validation before online loop tests should have been done and spectral data quality tested compared to laboratory model more carefully. Now from laboratory measurements moved straight into online analysis testing with online NIR equipment.

Tested multivariate control process charts noticed these systematic process changes in spectra whether automation itself in NIR equipment with this created calibration model was not completely straightforward. In addition to analytical problems such as possible fouling and probe contact angle could also have affected to the results. This point the most critical point was the calibration used during the online measurements which turned out more local than global model. Despite of local calibration model, online loop results gave valuable information how to improve these measurements and drawbacks that during NIR measurements might occur towards possible process implementation. Next step would be creating new wider calibration model to the automation or increase the old model with more samples and validate online calibration before online measurements. Parameter variation of now used calibration was at its extreme especially in lignin values but it seems that NIR requires more samples in calibration and even little variation is acceptable.

SUMMARY OF THE EXPERIMENTAL PART

NIR measurements were done for analyzing pretreated biomass samples started from the basics. Laboratory measurements in controlled conditions provided good correlation between measured spectral data and reference values of kappa number and lignin for all sampling sets. Validation of models showed that calibrations turned out to be more local than global. Temperature and moisture had direct impact on calibration results and prediction of new samples. Correlation analysis did not give certain area that could have been used for kappa number and lignin in calibrations. Therefore, total spectrum range was used in all created models. Similar aspect compared to the literature was reference method that was based on NREL standards. (Jiang et al., 2010) (Wolfrum & Sluiter, 2009) (Ye et al., 2008) Recent studies of using NIR in biomass measurements have focused on laboratory scale from homogenized, ground, and dried samples. Biomass samples were not dried or ground into certain particle size in this study. In addition, samples included both insoluble and soluble fraction that neither could not be eliminated.

During online loop tests used multivariate process control charts noticed spectral changes from process changes. Online loop calibration model needs to be developed with new samples and validated in the future. There are advantages to modify NIR calibrations, it can analyze samples' chemical changes, rotations, and vibrations deeper. Method is strongly dependent on reference data and used values in calibration and multivariate modelling methods are needed to create representative NIR models. Compared to conventional wet chemical methods based on NREL guidelines, NIR is new area that is advisable to be developed and models further improved to save costs and time. Next step would be online calibration improvements by increasing used data in calibration and its validation into certain process conditions or applying outside correction methods such as temperature, brix-% and moisture content to the calibration model. Analytical problems that may have occurred due to probe installation at 90° angle should be considered future experiments. Upcoming online tests probe could be installed e.g., at 45° angle with sufficient distance of the pipe wall to ensure bigger particles moving past the probe. In addition, calibration and model update needs to be developed from the laboratory model to online analysis as a routine.

15 CONCLUSIONS

Based on calibration models created at laboratory scale and online analysis tests NIR has potential to be utilized as pretreated biomass measurements. NIR provides faster analysis than old conventional laboratory measurements for kappa number and lignin. NIR as online analysis tool provides reliable information directly of the pretreated lignocellulosic woody biomass from the spectral data. Target of this thesis represents totally new field of study to use NIR predicting this kind of pretreated biomass material. There were only few publications for such as biomass samples and online scale measurements. These publications are discussed in the literature part. Thus, study of the investigation of suitability of NIR online measurement in this work was started from the basics. In addition to this the properties of the pretreated biomass gave their own challenge to the measurement and creating calibration models when both soluble and insoluble fractions are changing at the same time. Despite this feature measurements in laboratory and online scale achieved potential and positive results forgetting some challenges in the beginning when creating calibration model validations. These challenges to calibration models such as temperature effect, moisture content and insoluble fraction changes in liquid fraction were suggestions for further improvements to eliminate these in future calibration

models. Outside correction method included to calibration for brix-% to describe average of the soluble solids content, moisture content and temperature could be used if preprocessing method, 1st derivative does not eliminate these. Based on now created calibration models, with using 1st derivative at least temperature effect can be smoothed and minimized. Also replacing old wet chemical methods in laboratory for measuring these critical process factors with NIR, would create huge improvements in process monitoring as well as economical point of view in the long run.

Chemometrics and spectral data analysis are the key point to create representative NIR measurements. NIR calibration models are based on strongly on reference data, and it is important to choose samples that capture the most process variables in real environment where measurements are planned. Calibration model can be updated within new samples and validated later in the future. Laboratory scale measurement indicated that with diffuse-reflectance probe is easier to measure changes in solid properties, however slurries were quite well established but those samples need more preprocessing. Transflectance probe would be ideal to filtrate samples, but now used diffuse-reflectance probe managed nevertheless to model them effectively. While creating calibration model to sample campaigns, there were not certain areas that could be used in calibration models to model lignin and kappa number based on correlation analysis. Accordingly, total spectrum area is good to include calibration models while changes are occurring along the spectral determination region.

Even if material characteristics was challenging and measurements were not at first straightforward, there are more advantages to use and develop NIR implementation towards pretreated biomass process step online measurements. NIR would save process costs and time, due to old conventional wet chemistry method based on NREL guidelines. These measurements are time consuming compared to NIR which can measure sample within a half-minute. Of course, NIR requires working calibration model that is adjusted to the process conditions, but it is possible. Online loop tests in this work proved that NIR recognize spectral changes and with multivariate process control charts it is possible to monitor the process based on only spectral data. Hostelling's T^2 and $SPE_x(Q)$ do not require laboratory references. Slurry was chosen the easiest option to inline NIR measurements, but calibration model needs to be improved and increase its capture with more samples to cover wider variation area. In addition,

probe installation should be improved, and probe setup change that slurry moves evenly past the probe. This way sufficient contact with sample can be ensured, but also prevent fouling and any plugs in the tip of the probe. Spectral data quality also may increase when contact to the sample is enhanced and multivariate statistical process charts contribution figures will show more clearly spectral area changes when process conditions are changed. It is significant to avoid spectral data flattening where baseline is increased, and bigger peaks lowered. In this manner online spectral model data is accurate and provide better calibration results.

MATLAB software turned out to be very useful and handy tool to model and preprocess the spectral data easily. In addition, the most important is to create multivariate process charts to monitor the process measurements changes in online loop tests with PCA. TQ Analyst was at some point a bit unwieldy program but able to create basic calculations and calibration models. Next step could be models processing and if possible, these outside correction methods is good to test in practice with multivariate process charts to create more global model to lignin and kappa number.

REFERENCES

Alén, R., 2011. Biorefining of Forest Resources. Papermaking Science and Technology. ISBN 978-952-5216-39-4.

Akman, E.I., 2019. Chemistry and Structure of Lignin. In: Sustainable Lignin for Carbon Fibers: Principles, Techniques, and Applications. Springer. ISBN:978-3-030-18791-0. pp.1-12.

Agelet, L.E., Hurburgh, C.R., 2010. A tutorial on near Infrared Spectroscopy and Its Calibration. Critical Reviews in Analytical Chemistry. Vol.40. pp. 246-260.

Bart, J.C.J., 2006. Plastics Additives: Advanced Industrial Analysis.

Bart, J.C.J., Gucciardi, E., Cavallaro, S., 2013. Quality assurance of biolubricants. Science and Technology.

Basu, P., 2013. Biomass Gasification, Pyrolysis and Torrefaction. Practical Design and Theory. pp.47-86.

Bello, S., Rǐos, C., Feijoo, G., Moreira, M.T., 2018. Comparative evaluation of lignocellulosic biorefinery scenarios under a life cycle assessment approach. Biofuels, Bioproducts and Biorefining. Vol.12. pp.1-37.

Brys, A., Brys, J., Ostrowska-Ligeza, Kaleta, A., Gornicki, K., Glowacki, S., Koczon, P., 2016. Wood biomass characterization by DSC or FT-IR spectroscopy. Journal of Thermal Analysis and Calorimetry. Vol.126. pp.27-35.

Chio, C., Sain, M., Qin, W., 2019. Lignin utilization: A review of lignin depolymerization from various aspects. Renewable and Sustainable Energy Reviews. Vol.107. pp.232-249.

Clark, J., Deswarte, F., 2008. Introduction to Chemicals from Biomass. WILEY. ISBN 978-0-470-05805-3. pp.6-8, 35.

Challa, S., Potumarthi, R., 2013. Chemometrics-Based Process Analytical Technology (PAT) Tools: Applications and Adaptation in Pharmaceutical and Biopharmaceutical Industries. Applied Biochemistry & Biotechnology. Vol.169. pp-66-76.

Chai, X.S., Zhu, J.Y., 2014. Rapid pulp kappa number determination using spectroscopy. Institute of Paper Science and Technology.

Chen, Z., Lovett, D., Morris, J., 2011. Process analytical technologies and real time process control a review of some spectroscopic issues and challenges. Journal of Process Control. Vol.21. pp. 1467-1482.

Cheng, H., Yang, J., Liu, Q., Zhang, J., Frost, R.L., 2010. A spectroscopic comparison of selected Chinese kaolinite, coal bearing kaolinite and halloysite – A mid-infrared and near-infrared study. Spectrohimica Acta. Vol.77. pp.856-861.

Cossolino, D., 2014. Infrared spectroscopy: theory, developments and applications. ISBN. 1-62948-523-3.

De Bhowmick, G., Sarmah, A.K., Sen, R., 2018. Lignocellulosic biorefinery as a model for suitable development of biofuels and value added products. Biosource Technology. Vol.247. pp.1144-1154.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 1999. Introduction to Multi- and Megavariate Data Anlaysis using Projection Methods (PCA & PLS). Umetrics. pp.

Ewanick, S.M., Thompson, W.J., Marquardt, B.J., Bura, R., 2013. Real-time understanding of lignocellulosic bioethanol fermentation by Raman spectroscopy. Biotechnology for Biofuels. Vol.6. pp. 1-8.

Ewanick, S., Schmitt, E., Gustafson, R., Bura, R., 2014. Use of Raman spectroscopy for continuous monitoring and control of lignocellulosic biorefinery processes. Vol.86(5). pp.867-879.

El-Azazy, 2018. Infrared Spectroscopy – A Synopsis of the Fundamentals and Applications. Infrared Spectroscopy – Principles, Advances and Applications.

Fatma, S., Hameed, A., Noman, M., Ahmed, T., Shahid, M., Tariq, M., Sohail, I., Tabassum, R., 2018. Lignocellulosic Biomass: A Sustainable Bioenergy Source for the Future. Protein & peptide Letters. Vol.25. pp.1-16.

Ferrer, A., Alciaturi, C., Faneite, A., Rios, J., 2016. Analyses of Biomass Fibers by XRD, FT-IR, and NIR. Analytical Techniques and Methods for Biomass. Springer. pp.45-83.

Folfgang, A., Macdonald, D., 2005. Practical process control for engineers and technicians. ISBN 1-281-00945-8.

Fu, T., Elie, N., Brunelle, A., 2018. Radial distribution of wood extractives in European larch *Larix decidua* by TOF-SIMS imagning. Photochemistry. Vol.150. pp.31-39.

Fitoussi, C., Chiesa, S., Villegas, J., Gnansounou, E., Alciaturi, C., Ferrer, A., 2011. Compositional analysis of biomass feedstocks via near infrared spectroscopy for second-generation bioethanol production.

Gillet, S., Aguedo, M., Petitjean, L., Morais, A.R.C., da Costa Lopes, A.M., Lukasik, R.M., Anastas, P.T., 2017. Lignin transformations for high value applications: towards targeted modifications using green chemistry. Royal Society of Green Chemistry. Vol.19. pp.1-34.

Hassan, S.S., Williams, G.A., Jaiswal, A.K., 2019. Lignocellulosic Biorefineries in Europe: Current State and Prospects. CellPress Reviews. Trends in Biotechnology. Vol.37. pp. 231-234.

Hassan, S.S., Williams, G.A., Jaiswal, A.K., 2019. Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges and opportunities. Renewable and Sustainable Energy Reviews. Vol.101. pp. 590-599.

Henze, T., El Seoud, O.A., Koschella, A., 2018. Production and Characteristics of Cellulose from Different Sources. Springer. pp. 1-38.

Hill, C., 2006. Wood Modification: Chemical, Thermal and Other Processes. WILEY. ISBN 10 0-470-02172-1. pp.25.

He, W., Hu, H., 2013. Prediction of hot-water-soluble extractive, pentosan and cellulose content of various wood species using FT-NIR spectroscopy. Biosource Technology. Vol.140. pp.299-305.

Hongzhang, C., 2015. Lignocellulose biorefinery engineering: principles and applications. Elsevier. ISBN 0-08-100145-2.

Ikehata, A., 2020. NIR Optics and Measurements Methods. Near-Infrared Specroscopy. Springer. pp.211-233.

Jiang, W., Han, G., Brian, K.V., Tu, M., Liu, W., Fasina, O., 2013. rapid assessment of coniferous lignin-carbohydrates with near-infrared spectroscopy. Wood Science Technology. Vol. 48. pp. 109-122.

Jun, X-J., Huang, H., Nie, Z-K., Qu, L., Xu, Q., Tsao, G.T., 2012. Fuels and Chemicals from Hemicellulose Sugars.

Kettunen, P.O., 2006. WOOD Structure and Properties. Material Science Foundation 29-30. ISB. 0-87-849-487-1. pp.44-45, 54-55.

Kessler, R.W., Kessler, W., 2020. Inline and Online Process Analytical Technology with an Outlook for the Petrochemical Industry. Near-Infrared Spectroscopy. Springer. pp.553-593.

Kameswhar, A.K.S., Qin, W., 2017. Qualitative and Quantitative Method for Isolation and Characterization of Lignin-Modifying Enzymes Secreted by Microorganisms. Bioenergy Research. Vol.10. pp. 248-266.

Kanchanalai, P., Temani, G., Kawajiri, Y., Realff, M.J., 2016. Reaction kinetics of concentrated-acid hydrolysis for cellulose and hemicellulose and effect of crystallinity. BioResources. Vol. 11. pp.1672-1689.

Kiviharju, K., Salonen, K., Moilanen, U., Meskanen, E., Leisola, M., Eerikäinen, T., 2007. On-line biomass measurements in bioreactor cultivations: comparison study of two on-line probes. Journal of Industrial Microbiology and Biotechnology. Vol.43. pp.561-566.

Kucharska, K., Rybarczyk, P., Hołowacz, I., Łukajtis, M.G., Kamiński, M., 2018. Pretreatment of Lignocellulosic Materials as Substrates for Fermentation Processes. MDPI. pp.1-32.

Kucheryavskiy, S., Zhilin, S., Rodionova, O., Pomerantsev, A., 2020. Procrustes Cross-Validation – A Bridge between Cross-Validation and Independent Validation Sets. Analytical Chemistry. Vol.92. pp. 11842-11850.

Kuda Malwathumullage, C.P., 2013. PhD thesis: "Applications of near-infrared spectroscopy in temperature modelling of aqueous-based samples and polymer characterization. Iowa Research Online.

Kourti, T., MacGregor, J.F., 1995. Process analysis, monitoring and diagnosis, using multivariate projection methods. Chemometrics and intelligent laboratory systems. Vol.28. pp.3-21.

Koutras, M.V., Triantafyllou, I.S., 2020. Recent Advances on Unvariate Distribution-Free Stewhart-Type Control Charts. Distribution-Free Methods for Statistical Process Monitoring and Control. pp.1-56.

Kohonen, J., Reinikainen, Satu-Pia, Aaljoki, K., Perkiö, A., Väänänen, T., Höskuldsson, A., 2007. Multi-block methods in multivariate process control. Journal of Chemometrics. Vol.22. pp.281-287.

Liu, R., Li, L., Yin, W., Xu, D., Zang, H., 2017. Near-infrared spectroscopy monitoring and control of the fluidized bed granulation and coating processes – A review. International Journal of Pharmaceutics. Vol.530. pp. 308-315.

Liu, L., Ye, P., Womac, A.R., Sokhansanj, S., 2010. Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. Carbohydrate Polymers. Vol.81. pp.820-829.

Lu, Y., Lu, Y-C., Hu, H-Q., Xie, F-J., Wei, X-Y., Fan, X., 2017. Structural Characterization of Lignin and Its Degradation Products with Spectroscopic Methods.

Lu,Y-C., Lu, Y., Fan, X., 2020. Structure and Characteristics of Lignin. Springer. pp.17-75.
Mohan, D., Pittman, C.U., Steele, P.H., 2006. Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review. Energy & Fuels. Vol.20. pp.848-889.

Larkin, P.J., 2012. Infrared and raman spectroscopy principles and spectral interpretation. ISBN 1-4933-0127-6.

Lupoi, J.S., Singh, S., Simmons, B.A., Henry, R.J., 2014. Assessment of Lignocellulosic Biomass Using Analytical Spectroscopy: an Evolution to High-Throughput Techniques. BioEnergy Research. Vol.7. pp.1-23.

Lupoi, J.S., Smith, E.A., 2012. Characterization of Woody and Herbaceous Biomasses Lignin Composition with 1064 nm Dispersive Multichannel Raman Spectroscopy. Applied Spectroscopy. Vol. 66. pp. 903-910.

Mancini, M., Duca, D., Toscano, G., 2019. Laboratory customized online measurements for the prediction of the key-parameters of biomass quality control. Journal of Near Infrared Spectroscopy. Vol.27. pp.15-25.

Matsakas, L., Raghavendran, V., Yakimenko, O., Persson, G., Olsson, E., Rova, U., Olsson, L., Christakopoulos, P., 2019. Lignin-first biomass fractionation using a hybrid organosolv – Steam explosion pretreatment technology improves the saccharification and fermentability of spruce biomass. Biosource Technology. pp.521-528.

Marison, I., Hennessy, S., Foley, R., Schuler, M., Sivaprakasam, S., Freeland, B., 2012. The Choice of Suitable Online Analytical Techniques and Data Processing for Monitoring of Bioprocesses. Measurement, Monitoring, Modelling and Control of Bioprocesses. pp.249-280.

Menon, V., Rao, M., 2012. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. Progress in Energy and Combustion Science. Vol 38. pp. 522-550.

Mussatto, S., 2016. Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery. Elsevier. ISBN 0-12-802561-1.

Moore, E., 2017. Fourier transform infrared spectroscopy (FTIR): methods, analysis and research insights. Chemical engineering methods and technology. ISBN 1-5361-0401-9.

Mäki-Arvela, P., Salmi, T., Holmbom, B., Willfor, S., Murzin, D.Y., 2011. Synthesis of Sugars by Hydrolysis of Hemicelluloses – A Review. Chemical Reviews. Vol.111. pp.5638-5666.

Nkansah, K., Danwson-Andoh, B., Slahor, J., 2010. Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part 1 yellow-poplar (*Liriodendron tulipifera L.*). Biosource Technology. Vol. 2010. pp.4570-4576.

Rees, O.J., 2010. Fourier transform infrared spectroscopy developments, techniques and applications. Chemical engineering methods and technology. ISBN 1-61324-383-9.

Olivieri, A.C., 2018. Principal Component Analysis. Introduction to Multivariate Calibration. pp.57-71.

Sanderson, M.A., Agblevor, F., Collins, M., Johnson, D.K., 1996. Compositional analysis of biomass feedstocks by near-infrared reflectance spectroscopy. Biomass and Bioenergy. Vol.11. pp.365-370.

Santos-Fernándes, E., 2012. Multivariate Control Charts. Multivariate Statistical Quality Control Using R. pp.17-61.

Santos-Fernándes, E., 2012. Study Cases. Multivariate Statistical Quality Control Using R. pp.107-119.

Sonnleitner, B., 2012. Automated Measurement and Monitoring of Bioprocesses: Key Elements of the M^{3}C Strategy. Measurement, Monitoring, Modelling and Control of Bioprocesses. pp.1-33.

Sofikitou, E.M., Koutras, M.V., 2020. Multivariate Nonparametric Control Charts Based on Ordered Samples, Signs and Ranks. Distribution-Free Methods for Statistical Process Monitoring and Control. pp.57-105.

Sun, Z., Ibrahim, A., Oldham, P.B., 1997. Rapid Lignin Measurement in Hardwood Pulp Samples by Near-Infrared Fourier Transform Spectroscopy. Journal of Agricultural & Food Chemistry. Vol.45. pp.3088-3091.

Streamer, R., 2013. What NIR Instruments to Use in: Workshop on commercial application of IR spectroscopies to solid wood. Wood Technology Research Centre.

Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter. J., Templeton, D., Crocker, D., 2011. Determination of Structural Carbohydrates and Lignin in biomass. Laboratory Analytical Procedure (LAP). National Renewable Energy Laboratory. NREL.

Sluiter, A., Wolfrum, E., 2013. Near infrared calibration models for pretreated corn stover slurry solids, isolated and *in situ*. Journal of Near Infrared Spectroscopy. Vol.21. pp.249-257.

Schügerl, K., 2001. Progress in monitoring, modelling and control of bioprocesses during the last 20 years. Journal of Biotechnology. Vol.2. pp. 149-173.

Sørensen, K.M., van den Berg, F., Engelsen, S.B., 2020. NIR Data Exploration and Regression by Chemometrics – A Primer. Near-Infrared Spectroscopy. Springer. pp.127-189.

OECD, 2017. Working Party on Biotechnology, Nanotechnology and Converging Technologies. Biorefineries models and Policy. [Accessed 30.11.2020]. Available at: https://www.innovationpolicyplatform.org/www.innovationpolicyplatform.org/oa_folders/too lbox/download/20815/11579/Biorefineries%20Models%20and%20Policies%20%5BDSTI_S TP_BNCT(2016)16_FINAL%5D/index.pdf

Ozaki, Y., Morisawa, Y., 2020. Principles and Characteristics of NIR Spectroscopy. Near-Infrared Spectroscopy. Springer. pp.11-35.

Patel, H., 2007. Near Infrared Spectroscopy: Basic principles and use in tablet evaluation. International Journal of Chemical and Life Sciences. pp.2006-2015.

Pandey, A., Höfer, R., Taherzadeh, M., Nampoothiri, M., 2015. Industrial Biorefineries and White Biotechnology.

Pasquini, C., Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. Analytica Chemical Acta. Vol.1026. pp.8-36.

Petersen Rønnest, N., Stocks. S.M., Lantz, A.E., Gernaey, K.V., 2011. Introducing process analytical technology (PAT) in filamentous cultivation process development: comparison of advanced online sensors for biomass measurement. Journal of Industrial Microbiology and Biotechnology. Vol.38. pp.1679-1690.

Rinnan, Å., van der Berg, Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry. Vol.28. pp.1201-1222. Saha, B.C., 2003. Hemicellulose bioconversion. Journal of Industrial & Biotechnology. Vol.30. pp.279-291.

Shinde, S.D., Meng, X., Kumar, R., Ragauskas, A.J., 2018. Recent advances in understanding the pseudo-lignin formation in a lignocellulosic biorefinery. Green Chemistry. Vol.20. pp. 2161-2400.

Stanke, M., Hitzmann, B., 2012. Automatic Control of Bioprocesses. Measurement, Monitoring, Modelling and Control of Bioprocesses. pp.35-63.

Stenius, P., 2000. Forest Products Chemistry. Papermaking Science and Technology. ISBN 952-5216-03-9. pp.44.

Sun, Z., Ibrahim, A., Oldham, P.B., 1997. Rapid Lignin Measurement in Hardwood Pulp Samples by Near-Infrared Fourier Transform Raman Spectroscopy. Journal of Agricultural and Food Chemistry. Vol.45. pp.3088-3091.

TAPPI. 1999. Kappa number of pulp. [ONLINE]. [Accessed 15.12.2020]. Available at: https://research.cnr.ncsu.edu/wpsanalytical/documents/T236.PDF

Tucureanu, V., Matei, A., Avram, A.M., 2016. FTIR Spectroscopy for Carbon Family Study. Critical Reviews in Analytical Chemistry. Vol.46. pp. 506-520.

Tucker, M.P., Mitri, R.K., Eddy, F.P., Nguyen, Q.A., Gedvilas, L.M.,Webb, J.D., 2000. Fourier Transform Infrared Quantification of Sugars in Pretreated Biomass Liquors. Applied Biochemistry and Biotechnology. Vol.84. pp.84-86.

Tsuchikawa, S., Kobori, H., 2015. A review of recent application of near infrared spectroscopy to wood science and technology. Journal of Wood Science. Vol.61 pp.213-220.

Thermo Fischer Scientific manual, 1996-2011. TQ Analyst Software User Guide.

Türker-Kaya, S., Huck, C.W., 2017. A Review of Mid-Infrared and Near-Infrared Imaging: Principles, Concepts and Applications in Plant Tissue Analysis. Molecules. Vol.22. pp.1-20.

UPM Biochemicals. 2020. [ONLINE]. [Accessed 8.12.2020]. Available at: https://www.upmbiochemicals.com/biorefinery/

Woodings, C., 2001. Regenerated Cellulose Fibres. Woodhead Publishing. pp.2.

Wu, J-B., Lin, M-L., Tan, P-H., 2018. Raman Spectroscopy of Monolayer and Multilayer Graphenes. In Raman Spectroscopy of Two-Dimensional Materials. Springer. vol.276. pp. 1-27.

Wu, H., Sun, J., Zheng, L., 2015. The analysis of the different filtering algorithm effects on the fluorescence spectrum data processing. ICMII. pp.46-51.

Wolfrum, E.J., Sluiter, A.D., 2009. Improved multivariate calibration models for corn stover feedstock and dilute-acid pretreated stover. Cellulose. Vol.16. pp.567-576.

Xue, J., Yang, Z., Han, L., Liu, Y., Zhou, C., 2015. On-line measurement of proximates and lignocellulose components of corn stover using NIRS. Applied Energy. Vol.137. pp.18-25.

Xu, F., Yu, J., Tesso, T., Dowell, F., Wang, D., 2013. Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. Applied Energy. Vol.104. pp.801-809.

Ye, X.P., Liu, L., Hayes, D., Womac, A., Hong, K., Sokhansanj, S., 2008. Fast classification and compositional analysis of cornstover fractions using Fourier transform near-infrared techniques. Biosource Technology. Vol.99. pp. 7323-7332.

Zhou, C., Jiang, W., Via, B.K., Fasina, O., Han, G., 2015. Prediction of mixed hardwood lignin and carbohydrate content using ATR-FTIR and FT-NIR. Carbohydrate Polymers. Vol.121. pp.336-341.

Özdençi, K., De Blasio, C., Muddassar, H.R., Melin, K., Oinas, P., Koskinen, J., Sarwar, G., Järvinen, M., 2017. A novel biorefinery integration concept for lignocellulosic biomass. Vol.149. pp.974-987.