Master’s Thesis

Elina Heinonen 2020

LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY LUT

School of Engineering Science

Degree programme in chemical engineering

Elina Heinonen

ONLINE ANALYSIS OF WOOD EXTRACTIVES

Examiners: Professor Satu-Pia Reinikainen

PhD Risto Kotilainen

Tiivistelmä

Online analysis of wood extractives

Hakusanat: Lähi-infrapuna spektroskopia, NIRS, mäntyöljy, uusiutuva diesel, online-analytiikka, PCA, PLS.

130 sivua, 51 kuvaa, 22 taulukkoa, 2 liitettä

Elina Heinonen

Diplomityö 2020

Lappeenrannan-Lahden teknillinen yliopisto LUT

School of Engineering Science

Kemiantekniikan koulutusohjelma

Työn tarkastajat: Professori Satu-Pia Reinikainen, LUT

Filosofian tohtori Risto Kotilainen, UPM

Uusiutuvaa dieseliä voidaan jalostaa sellunkeitossa syntyvästä jäännöstuotteesta, mäntyöljystä. Tämän diplomityön kirjallisessa osassa tehtiin kirjallisuus tutkimus erilaisista online-analysaattoreita, jotka voitaisiin sovittaa mäntyöljyn esikäsittelyprosessiin. Tarkoituksena on analysoida mäntyöljykomponenttien ja epäpuhtauksien määrää. Tutkimuksen perusteella analysaattoriksi valittiin lähi-infrapuna spektroskooppi (NIR). Kokeellisessa osassa erilaisia mäntyöljynäytteitä analysoitiin laboratoriossa lähi-infrapuna spektroskoopilla.

Mäntyöljykomponenteille luotiin NIR kalibrointimallit laboratorion referenssimittausten perusteella. Validointimittausten perusteella mallit eivät kuitenkaan olleet täysin paikkansapitäviä, mutta vesipitoisuuden, hartsihappojen, rasvahappojen ja happoluvun kalibrointi näyttää olevan mahdollista. Ongelmallisimpia kalibrointeja olivat neutraaliaine- ja saippuanumerokalibroinnit. Kaikkien mallien ongelmat johtuvat todennäköisesti liian pienestä kalibrointialueesta ja mahdollisesta mallien ylimallinnuksesta. Lisäksi työssä tutkittiin näytteen lämpötilan vaikutusta mittaustulokseen sekä heijastavuus- ja transmittanssi-menetelmiä verrattiin keskenään.

Abstract

Online analysis of wood extractives

Keywords: Near infrared spectroscopy, NIR, online-analysis, crude tall oil, renewable diesel, PCA, PLS.

130 pages, 51 figures, 22 tables and 2 appendices

Elina Heinonen

Master's thesis 2020

Lappeenranta-Lahti University of Technology LUT

School of Engineering Science

Degree programme in chemical engineering

Thesis supervisors: Professor Satu-Pia Reinikainen, LUT

PhD Risto Kotilainen, UPM

Renewable diesel can be refined from the residual product of Kraft pulping, tall oil. In the theoretical part of this master's thesis, a literature review was conducted on various online analyzers that could be adapted to the tall oil preprocessing. The purpose to analyze the amount of tall oil components and impurities. Based on the study, a near-infrared spectroscope was selected as the analyzer. In the experimental part, different tall oil samples were analyzed in the laboratory by near-infrared spectroscopy.

NIR calibration models were created for tall oil components based on reference measurements. Based on the validation measurements, the models were not completely accurate, but calibration seems possible for water content, resin acid, fatty acid and acid number. Most problems occurred with neutral substances and residual soap number. The problems of the models are probably due to the too small calibration range and possible overfitting of the models. In addition, the effect of sample temperature to the measurement results was studied, and reflectance and transmittance methods were compared.

Acknowledgements

The last six months have been challenging, but I am happy to have taken on the challenge and coming out stronger from the other side. It is an understatement to say I learned a lot, about the work and about myself. I am also grateful to have so many people in my life who play a part in how I got here, writing this text as I look over my finished thesis.

Thank you to everyone at UPM and LUT who helped with the thesis. To my supervisor, Kati Vilonen, who gave me the opportunity to do the work, and who helped, taught, and guided me throughout all of it. To professor Satu-Pia Reinikainen, who shared her Matlab expertise, without which I would not have been able to do it. To Hosmed representative Petra Honkavirta, who taught me how to use the near-infrared analyzer and always happily helped me with the measurements and creating the calibration models. Thanks also to the laboratory staff and others who helped and shared their professional knowledge with me.

I also want to thank my friends and family, especially my parents, who always supported me and believed in me and my abilities, even when I wasn't so sure myself. Special thanks to my friends Anna, Tara and Marianne, who went through this journey with me from the first day of university to the last, and all that goes with it. Lastly, thank you to my boyfriend Tatu, who always made me happy and laugh even through all my troubles and stress.

I would not be here without all your help, love, and support. Thank you.

Elina Heinonen

Lappeenranta 13.7.2020

Table of contents

List of symbols and abbreviations

^{13}C	carbon-13 isotope
^{15}N	nitrogen-15 isotope
^{1}H	hydrogen-1 isotope
^{31}P	phosphorous-31 isotope
AAS	atomic absorption spectroscopy
Am-241	americium-241 isotope
AOTF	acousto-optic tunable filter
bPLS	correlation coefficient
C	carbon
℃	Celsius
CCD	charged coupled detector
Co-57	cobalt-57 isotope
CSS	crude sulphate soap
CTO	crude tall oil
E	residual noise of X
EDXRF	energy dispersive x-ray fluorescence
F	residual noise
FPXRF	field portable x-ray fluorescence
H	hydrogen
HDXRF	high definition x-ray fluorescence
HPLC	high performance liquid chromatography
ICP-OS	inductively coupled plasma-optical spectrometry
InGaAs	indium-gallium-arsenide (detector)
IR	infrared

KNN	K-nearest neighbors
LCTF	liquid crystal tunable filter
LED	light emitting diode
LFXRF	low-field x-ray fluorescence
mbar	millibar
MHz	megahertz
N	nitrogen
N	number of samples
Ni	nickel
NIRS	near infrared spectroscopy
O	oxygen
ODt	oven dry tonne
P	loadings of X
PbS	lead-sulfur (detector)
PC	principal component
PCR	principal component regression
PDA	photodiode array
PLS	partial least square
PLS-PM	partial least squares path model
PLS-SEM	partial least squares structural equation model
ppm	parts per million
PRESS	predicted residual sum of squares
PTO	purified tall oil
Q	loadings of Y
R&D	research and development
R2	goodness of fit

RMSECV	root mean square of cross-validation
S	sulfur
Si	silicon
SLD	superluminescent light-emitting diode
SPEx	squared prediction error of X
T	scores of X
T	tesla
t	time
T2	Hotelling's T2 statistic
T_2	spin-spin relaxation time
TFE	thin film evaporator
TOFA	tall oil fatty acids
TOR	tall oil resin
U	scores of Y
UV/VIS	ultraviolet/visible spectrometry
W	variable weight matrix
w-%	weight percent
WDXRF	wavelength dispersive x-ray fluorescence
XRD	x-ray diffraction
XRF	x-ray fluorescence
$\bar{y}$	mean value of response values
$\hat{y}$	predicted response values
y	response values
Zn	Zinck

1. Introduction

With the decline of fossil fuel resources, the use of renewable biofuels has increased. Biofuels offer a more sustainable future and they produce lower greenhouse gas emissions than fossil fuels, that contributes to climate change. According to the EU directive RED II, in its member states 14% of the energy consumed by road or rail transport, must be renewable energy by the year 2030. Of those 14%, 3.5% should be advanced biofuels, Finland plans that by the year 2030, 30% of fuels used in road and rail transport in Finland would be biofuels, from where 10% should be accounted by advanced biofuels. Finnish forest industry company UPM produces renewable diesel and renewable naphtha in Lapppeenranta biorefinery, that is made from crude tall oil (CTO), that is an advanced biofuel resource. [UPM] [European commission] [Voegele, 2019]

CTO is a residue from Kraft pulp production. It is a dark brow liquid, that is viscous and foul smelling before refining. Typical chemical composition of CTO is 38-53 w-% of fatty acids, 38-53 w-% of rosin acids and 6,5-20 w-% of neutral unsaponifiable compounds and some impurities, such as water, soap and metals. As CTO is not a main product, but a residue, it doesn't have any quality guidelines and its composition varies. The CTO composition changes based on different wood species, growing regions, season of the year and other extraneous aspects like wood storage and pulping. In addition to these aspects, CTO quality can also be changed when it is transported and stored. [Aro&Fatehi,2017]

The changes in CTO produce a challenge for the refining process. The changes in the quality of the CTO can be fast and changes elsewhere in the process need to be made in according to the composition parameters. At the moment, the component parameters and impurities are measured in the laboratory a few times a day. However, as the changes in the CTO composition can occur in hours or minutes, this is not enough. Online analysis could be the answer to produce measurements in real time in order to improve process automatization and control, adjust product quality and improve yield.

In this thesis online analysis equipment were researched and compared, in effort to choose the right analyzer for the process in question. For the experimental part near

infrared spectroscopy was chosen, and samples were analyzed in the laboratory to construct calibration models for crude tall oil and purified tall oil. Samples were measured with two methods, reflectance and transmittance methods. With transmittance method samples were measured in three different temperatures to see the effect of temperature in the sample spectra and its effect on the models. Lastly purified tall oil samples were collected daily for 45 days to see the trend of the process.

1.1 Production process

In this section the production process from Kraft pulping to renewable diesel is generally described, just to serve as a background. More detail will be put in the pre-treatment process part. More detailed description of other process parts has been done previously by Leivo (2018) and Halme (2019) in their master's thesis on related subjects.

CTO production process starts at the Kraft pulp mill. The flowsheet of the production process is presented in Figure 1. In pulp production wood chips are treated with cooking liquor containing sodium hydroxide and sodium sulfide to dissolve the lignin. Wood extractives such as resin acid, fatty acid, neutral and oxidized substances react with the cooking liquor, and after the cooking stage the residual cooking liquor that is now black liquor is separated from the pulp during washing. The weak black liquor is concentrated by evaporation. When the solid content rises during the evaporation in the now strong black liquor, a layer of soap called crude sulphate soap (CSS) forms to the surface of the liquor. CSS is removed and turned into CTO by acidulation process with sulfuric acid. [Peters&Stojcheva,2017] [Hase et all. 2003] In Figure 2. CTO's part of the pulping process is shown.

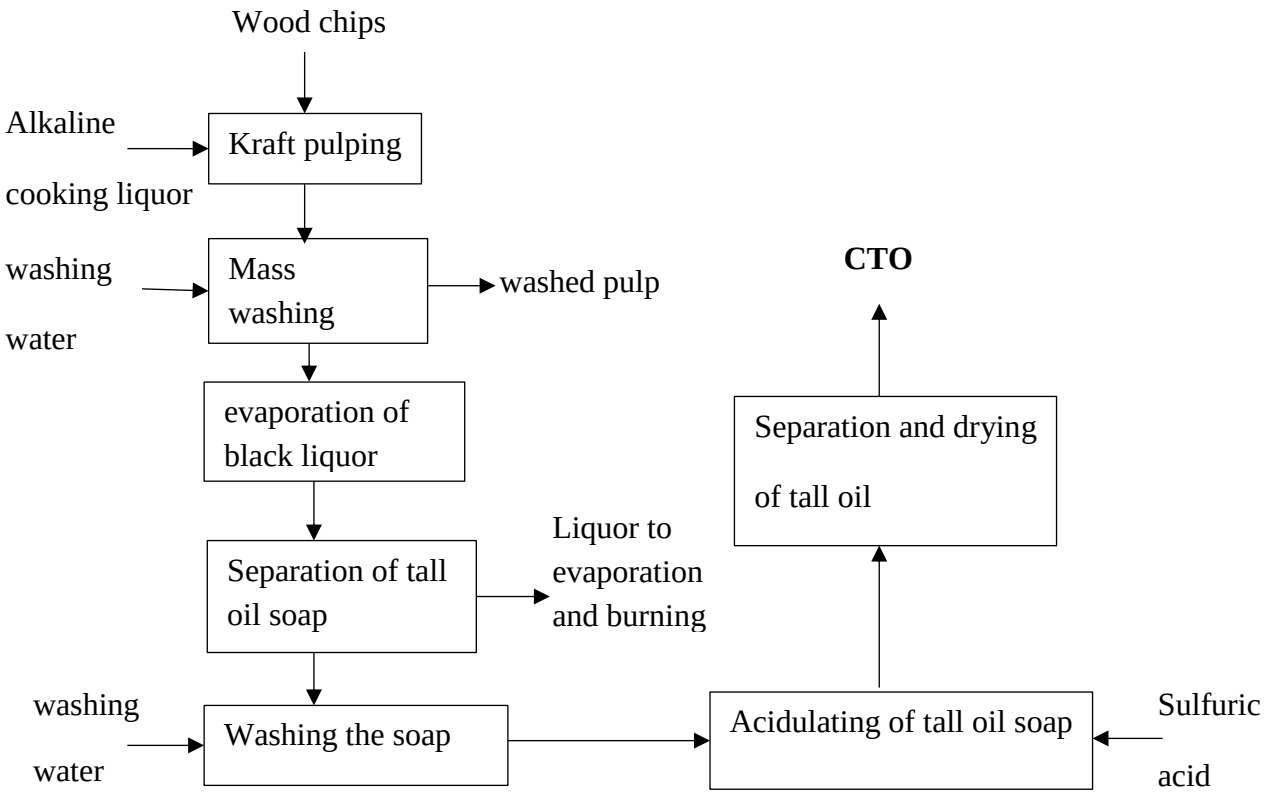


Figure 1. Flow sheet of CTO production from wood to CTO. Modified from Prokkola et all.

Figure 2. Amount of CTO from wood. ©UPM. [Mannonen, 2014]

CTO is then stored to tanks before transportation from the pulp mill and before feeding it to the biorefinery. As the CTO is acidic, the storage tanks and transportation tanks

used must be lined with acid resistant material, as the CTO may be corrosive to the tank walls otherwise. For example, in some cases the train carriages that are used to transfer the CTO, are not acid resistant, and excess iron dissolves into the CTO due to the corrosion of the carriage walls. Iron can cause problems later in the hydrotreatment if not removed [Arora et all., 2018].

Around 48 truckloads of CTO are collected into one storage tank from multiple suppliers and from different pulping processes. During the time in the tank, some components like water and water-soluble impurities and some solid particles settle at the bottom of the tank. This decreases the amount of water in the CTO that is fed into the pretreatment, as the tank is not emptied from the bottom. However, the composition of CTO and its quality varies between separate tanks and sometimes even inside one tank if mixing is not proper. [Vilonen, K.,2020]

Pretreatment of CTO consists evaporation steps. In the pretreatment steps impurities such as water, solid particles, metals and salts are removed. The pretreatment steps produce purified feed stock, the light fraction that consists of gasses, turpentine, light hydrocarbons and water and the residue fraction that consists of heavy hydrocarbons that have 30 or more carbon atoms, pitch and most impurities like metals are removed here. The purified feed stock contains mainly fatty acids, resin acids and neutral components such as sterols and stanols. The flowsheet of different feeds is shown in Figure 3. [Mannonen, 2014] [FI 126029B] [Kotoneva, J., 2020]

In the evaporation system two or more components are separated from each other, by utilizing the differences of vapor pressure of the components. The evaporation is done in vacuum, in order to improve evaporation at lower temperatures. By creating multi step evaporation, it is possible to accomplish the evaporation in a controlled and effective manner that leaves small amount of residue. In the pretreatment multi step evaporation, the evaporators can be either the same or different type evaporators. Evaporators using thin fill evaporation technique are suitable for the pretreatment of CTO. Such evaporators include thin film evaporators (TFE), short path evaporators, falling film evaporators and plate molecular stills. The evaporators operate in conditions between range 0,1-100 mbar and 50-450°C. [FI 126029B]

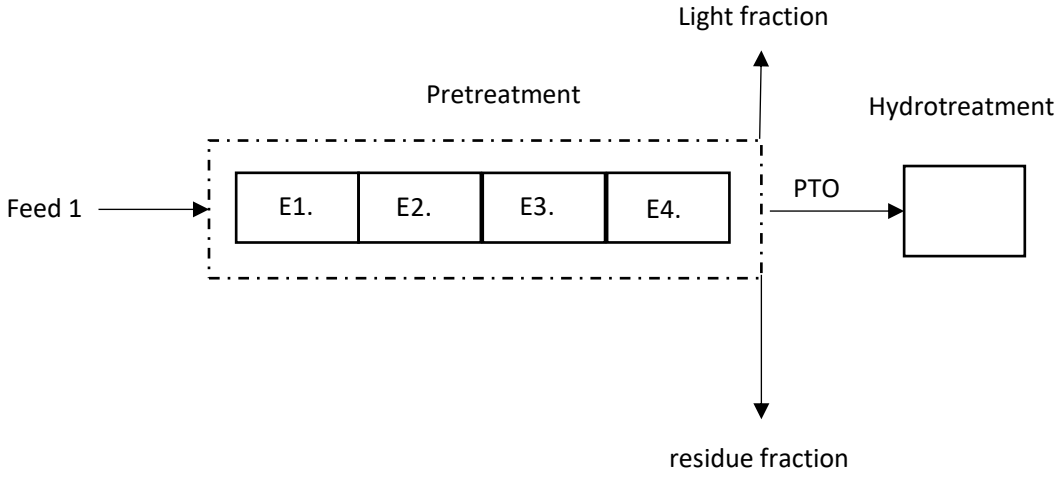


Figure 3. Flowsheet of different feeds and residues in pretreatment. E1-E4 are different evaporation steps. Feed 1 is untreated CTO. PTO is purified tall oil, from where the light fraction and residue fraction have been removed from. The light fraction contains gases, light hydrocarbons and water. The residue fraction contains hydrocarbons with more than 30 carbons in the chemical structure. Modified from [FI 126029B].

In Figure 4 below, the flow sheet of the process is represented from unpurified CTO to renewable diesel. After the pre-treatment, purified tall oil (PTO) is fed to hydrotreatment, where the chemical structure is modified with the addition of hydrogen in the presence of a catalyst to produce hydrocarbons. In fractioning, co-products (water, hydrogen sulfide and other uncondensable gasses), from the hydrotreatment are removed. The remaining liquid can be distilled into renewable diesel. [Mannonen, 2014] [Bezergianni et all., 2013] [Perego et all.,2012] [FI 126029B]

Figure 4. Simplistic flowsheet of CTO to renewable diesel. [Mannonen, 2014] ©UPM

1.2 components and impurities of CTO

The amount of wood extractives and its three main components vary widely, depending on the wood species, growth site (climate), soil quality and season of the year. The composition of the CTO also changes due to wood storage and transport of wood and CTO. In Figure 5, variation of monthly soap recovery average is shown. [Laxen et all.2008] In Table I, the differences in component constituents between different wood species are shown. [Norlin, 2012] CTO composes from fatty acids (TOFA), resin acids (TOR) and from neutral unsaponifiable compounds and some impurities. [Aro&Fatehi,2017] In Table II, the composition of CTO depending on the growth place is shown. [Gullichsen et all, 1999]

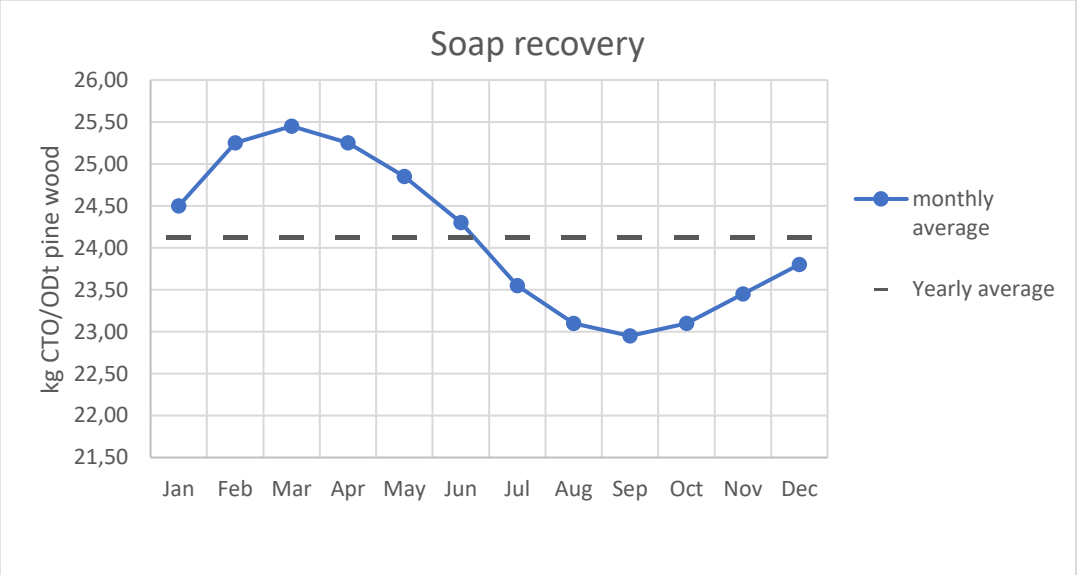


Figure 5. Seasonal variation of soap recovery. Modified from [Laxen et all.2008]

Table I. Primary components of CTO in different wood species. [Norlin, 2012]

Wood species	extractives %	Fatty acids, kg/t wood	Rosin acids, kg/t wood	unsaponifiables, kg/t wood
Pine	≈ 3	15-20	5-10	2-5
Spruce	≈ 1	4-8	1-2	1-3
Birch	≈ 2	12-18		5-10

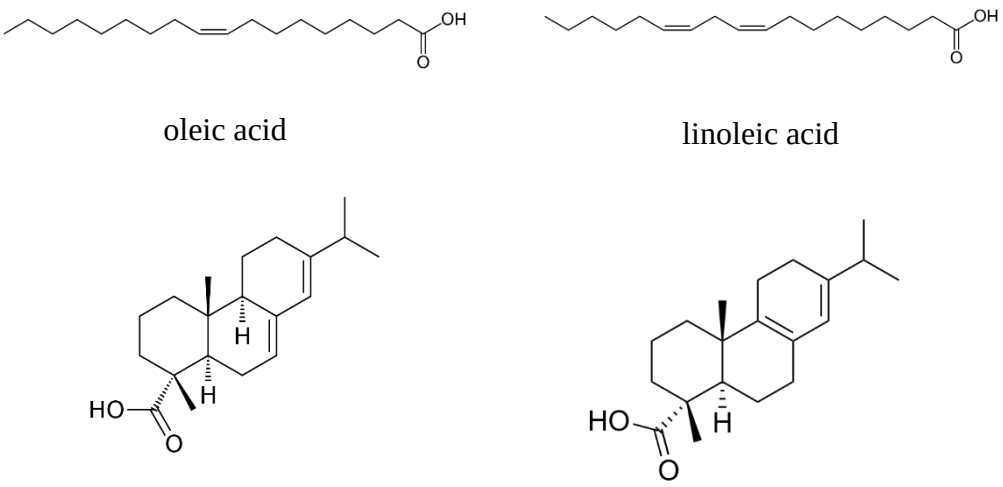
Table II. Composition of CTO depending on the growth site. The wood used in the USA and Canada columns are pine when Scandinavian example is from the use of wood mix of pine and spruce. [Gullichsen et all, 1999]

	Southeastern USA	Northern USA and Canada	Scandinavia
Acid number	165	135	132
saponification number	172	166	142
Resin acids, %	40	30	23
Fatty acids, %	52	55	57

Unsaponifiables, %	8	15	20

The fatty acids in tall oil are usually mostly oleic, linoleic and linolenic acids, with smaller amount of pinolenic, palmitic and other fatty acids that are contained in wood. In Figure 6, chemical structures of oleic and linoleic fatty acids contained in CTO are presented. [Hase et all., 2003]

The most important resin acids in CTO are abietic, dehydroabietic, palustric, pimaric, levopimaric and neoabietic acid, but CTO includes smaller amounts of other resin acids as well. In Figure 6, the chemical structure of abietic acid and palustric acid in CTO are presented. [Hase et all., 2003]



oleic acid linoleic acid

abietic acid palustric acid

Figure 6. The chemical structures of fatty acids and resin acids in CTO. Fatty acids are presented on the upper row, and resin acids on the lower row.

Neutral components in CTO are also called unsaponifiables, because they do not turn into soap in alkaline hydrolysis during the Kraft pulping. This fraction consists mostly of sterols, different alcohols, aldehydes, hydrocarbons, and turpentenes, but it can also have small amounts of other components. [Wang et all. 2001][Vikström et all. 2005] Some neutral components, like the main phytosterol in tall oil, β-sitosterol, could be further processed and used in medicine, cosmetic or food applications. β-sitosterol is

an antioxidant, that has anti-inflammatory and antibacterial actions, it has also been found to lower cholesterol and inhibit the growth of tumors as well as renew and moisturize skin. [Wang et all., 2005] Chemical structures of some neutral components in CTO are presented in Figure 7.

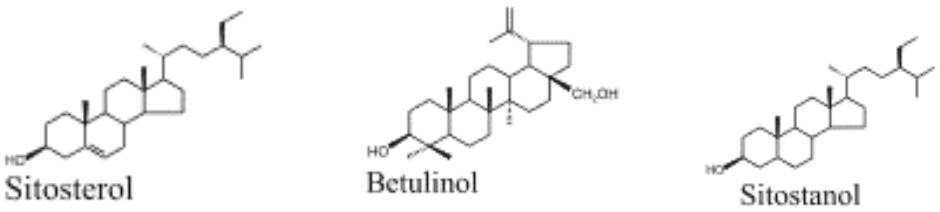


Figure 7. The chemical structures of neutral components in CTO. [Vikström et all. 2005]

CTO also contains some impurities, that need to be removed in the pretreatment. Such impurities are water, soap residue, residual metals, salts, sulphates, solid particles, sodium, phosphorous and silicon. [FI 126029B]

Water percentage in the CTO varies from 0% to 10%, but the desirable amount is less than 2%. Some of the water content in the CTO settles in the storage tanks, so it doesn't reach the pretreatment step. Water that is left is removed in the first pretreatment step in the light fraction with turpentine. [Vilonen, K., 2020] [FI 126029B]

Some impurities are left from the Kraft pulping process, where black liquor cooking and tall oil soap acidulation process. These are for example sodium, sulphate and calcium compounds. Only a small amount of silicon is in the wood material itself, but the majority of silicon in the CTO comes from the pulp mill where it is used as an antifoaming agent. Some soap can also be present in the CTO as a residue from the acidulating process. The amount of soap residue can be minimized by adding more acid, but this can lead to excess of sulfuric acid that increases corrosion risk, which can be problematic for example in transportation. [Vilonen, K., 2020] [Wansbrough et all. 2008]

As some metal and mineral impurities come from the production process, some come from the wood and others, like iron dissolves into the CTO due to corrosion in the

storage and transportation carriages. Minerals like phosphorous appears in the wood itself as it is vital nutrient for plants like nitrate. Metals like sodium and potassium are also present in the wood itself. [Arora et all., 2018] [Kotoneva, J., 2020]

Solid particles are usually precipitants from the process. Lignin is a component present in the wood. In the alkaline Kraft pulping lignin produces sodium lignite in the tall oil soap, and after the acidulation process, the lignin precipitates. Cellulose or hemicellulose can be present in CTO as fibers. Compounds can also precipitate as salts, for example when calcium ions, that are present in the soap form calcium sulfate, when introduced to sulfate ions from the acidulation process. [Wansbrough et all. 2008]

2. Online analysis

Online analyzers are analyzing equipment that are connected to a process, where they conduct automatic sampling for part of the flow, analyze said sample and return it to the flow. Online analyzers report the collected data to the operator in real time. [AAVOS International] In this chapter different analysis equipment is studied, in an effort to find suitable online analysis for the process. The analyzers are studied keeping the process in question in mind, thus not all related theory behind the method is explained, mostly basic working principles, and applications. The equipment would be installed on multiple points in the process. Places for online analyzers are shown in Figure 8 below. The places are before and after the storage tanks, possibly between the evaporators and after the pretreatment step.

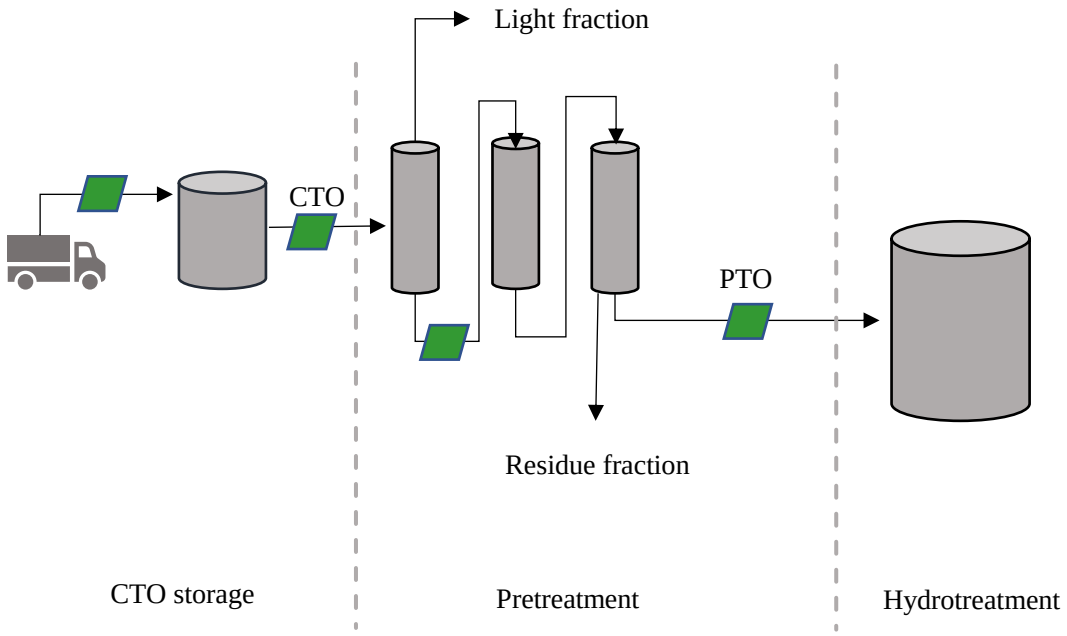


Figure 8. Green blocks represent the possible points of measurement for the online analyzers in the process.

Because the analyzers will be in different places, they will have some differences in the operating conditions, for example changes in temperatures, which can be problematic for the analyzers. The wanted measurements from the process by the online analyzer are at least water content, acid value and the concentrations of impurities. Preferably also the detection of lignin and fiber structures.

Chemical compounds and functional groups present in CTO that could be used to identify compounds or determine the amount of said compounds in the samples are: COOH, C-H, C=C, O-H, -O-CH3, aromatics, N-H, S-H and complex metals.

2.1 NIRS

Near infra-red spectroscopy (NIRS) is based on the absorption of electromagnetic radiation in the near infrared wavelength region from 780 to 2500 nm (13300 to 4000 cm^{-1}). NIRS uses molecular overtones and combination vibrations in the said region to identify compounds, as energy absorbance is specific to chemical bonds. Functional groups and bonds like C-H, O-H, N-H, S-H C=O, =C-H, COOH and aromatic C-H are strong NIR absorbers, and thus easy to detect. [Bart, 2006] Metals can also be detected,

if they are in complex form with organic molecules that contain C-H, N-H or O-H bonds, as pure metals do not absorb in the NIR spectra region. [Nomngongo et all., 2016]

The equipment principle is similar to other spectroscopy methods, where a light from controlled source is being directed at the sample where it is reflected at diffraction grading and forward to detectors. Figure 9 shows a simplified principle of NIRS, note that this is only one way of constructing NIRS equipment, as there are many ways to do this. Depending on the need, the detector can be programmed to use different measurement modes, depending whether the energy is reflected, transmitted or absorbed when the NIR radiation interacts with the sample. Most common measurement modes are transmittance, interactance, transflectance, diffuse transmittance and diffuse reflectance. These modes are presented in Figure 10. The choice of the measurement method is mostly affected by the chemical characteristics of the sample, for example, the phase (solid, liquid or gas), translucency and size of particles in the sample. NIRS is a widely used technology in different fields, for example in pharmacy, medical applications, dairy, agriculture, astrology, wood and forestry and petroleum industry applications. [Balabin et all. 2011][Bart et all. 2013] [Prieto et all. 2017][Roberts et all. 2018]

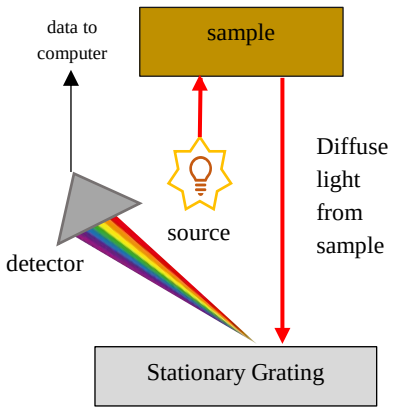


Figure 9. Simplified working principle of NIRS. NIR radiation is directed at the sample, the reflected radiation is then collected to the diffraction grating, which splits the radiation into constituent wavelengths. These wavelengths are then detected and transferred into a computer. Modified from [Buchi]

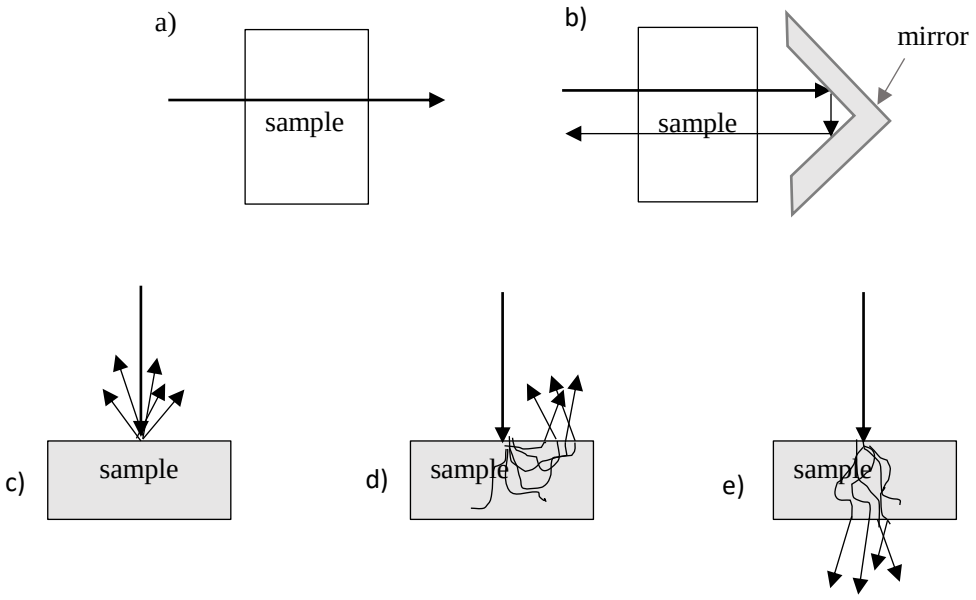


Figure 10. Most used measurement modes in NIR spectroscopy. a) transmittance, b) transflectance, c) diffuse reflectance, d) interactance and e) diffuse transmittance. Modified from [Pasquini, 2003]

NIRS can be used for quantitative measurements (component concentrations) as well as qualitative measurements (the identification of components) and it can determine chemical and physical product properties. It is a rapid analysis method, where the measurements can be done within one minute or less per sample. It doesn't need sample preparation and is does not destruct the sample and there is no chemical waste or additional chemical costs. There is no fixed sample size and multiple results can be obtained from the same measurement data. [Bart et all., 2013] As NIRS can be used to measure a wide array of properties, equipment manufactures provide custom designed applications, to suit different needs. Some physical and chemical product properties reportedly measured are: density, viscosity, flash point, cloud point, color, transparency, total acid number, molecular weight, identification of raw material, composition determination, concentrations of organic and inorganic components, content of volatile components and water content. [Bart et all.,2013]

Even if NIRS has a wide range of advantages, it does have some disadvantages as well that need to be taken into account. NIRS is a secondary method, which means it needs calibration against reference methods, and all in all it needs robust calibration, as each

component needs its own calibration for each type of sample analyzed. The samples should also be in controlled temperature. [Bart et all., 2013] However, even if NIRS can't give specific about the sample components without thorough calibration, changes in the spectra can be used as indicative information for example with water content levels.

2.1.1 NIRS instrumentation

When assembling an NIR spectrometer, it is important to keep in mind the end application, as different instrumentation in NIRS can affect the end result. The price of the equipment also depends on the instruments used. [Pasquini, 2013] There are five basic sections in any commercial NIR spectrometer: 1. sample compartment, 2. light source, 3. Light wave selection system, 4. detector, and 5. signal processor. [Agelet et all. 2010] Example instrumentation from these sections are presented below.

In laboratory, if working with reflectance methods no sample compartment is needed, and if working with transmittance, open silica and quartz sample compartments can be used, as they are transparent to NIR light. In online analysis sample compartments are not necessarily needed, depending on the type of the sample. Free flowing liquid samples can be analyzed with flow through cells, but extremely viscous liquids should be analyzed with the diffuse reflectance method. NIR light can travel through glass without losing much of its signal integrity. Fiber optics can also be used, if the sample is far from the source, or if the sampling is done by immersion in liquid. [Agelet et all. 2010] [Bart, 2006]

Most commonly used light sources in NIR are the tungsten halogen lamp, LED and SLD. The most popular is the tungsten halogen lamp, which wavelength emission ranges from 320 to 2500 nm. LED (light emitting diode) has low price and power consumption, it is small in size but has a long lifetime. LEDs are usually used in small NIR equipment that are used the laboratory environment. LED can produce radiation with a narrow band width of 30-50 nm, but it can be centered in any wavelength of the spectral region. One instrument can also utilize multiple LEDs set in different wavelengths or they can be used to produce a polychromatic source, whose radiation

is dispersed by using monochromator devices such as gratings or filter optics. SLDs (superluminiscent light-emitting diodes, acronym SLED can be used, but it is also used for surface-emitting LEDs.) are semiconductor tunable diode lasers. It is powerful and bright like a laser diode, but smaller in size and cheaper, and it can be adjusted to narrow or wide band width within the NIR wavelength range. [Agelet et all., 2010] [Pasquini, 2013]

Wavelength selectors collect the energy from the sample and feed the wanted wavelengths to the detector. Wavelengths can be selected by filters. The simplest filters work by absorbing all wavelengths, but the one that is of interest. These filters can be arranged in a wheel, that allows selection of multiple wavelengths. Some filters, like acousto-optic tunable filter (AOTF) and liquid crystal tunable filter (LCTF) can create several wavelengths, thus don't need mechanical devices to switch the filter used. Dispersive type instruments use a prism or grating to collects the energy from the sample. presented in Figure 9. Grating usually is the most low-cost option when compared with more modern technologies. Like [Agelet et all., 2010] [Pasquini, 2013]

Detectors transform the light energy from the wavelength selectors into electric signal, which can be transformed into digital data that can be processed by the computer. Detectors are selected based on the wavelength region to be covered. Most used detectors are silicon (Si) detectors, lead sulfide (PbS) detectors, indium gallium arsenide (InGaAs) detectors or charged coupled devices (CCD). Silicon is the most common detector material in the 400 to 1100 nm range. Si detectors are stable, fast, sensitive to low light intensity and low cost. PbS and InGaAs are used in higher wavelength regions and they can be used together in the same instrument. Multiple InGaAs and CCDs detectors can be set in photodiode array (PDA) spectrographs in groups, to achieve faster measurements. CCDs have higher signal sensitivity and resolution than InGaAs detectors, while InGaAs detectors in PDA have higher signal precision and higher signal-to-noise ratio than CCDs in PDA. [Agelet et all., 2010]

NIRS can also use Fourier transform (FT) mathematical technique, that makes the NIRS faster, more sensitive and precise with lower background noise. The FT-NIR's working principle differs from the traditional NIR slightly. FT-NIR uses polychromatic light instead of monochromatic light, which is then directed to a beam

splitter that produces an interferogram, that shows the intensity of light as a function of time. The interferogram is then transferred to actual spectra by Fourier transformation. [Chmielarz et all., 2019] The FT-NIR can be seen as the updated version of the traditional NIRS.

NIRS offers a rapid and cost-effective online analysis, but to get analytical data from the spectra, multivariate data analysis methods should be applied. NIR spectra contain vast amounts of information about the physical and chemical properties of molecules, but the weak NIR bands can be wide and overlapping. [Bart, 2006] Thus chemometrics should be applied. This means using methods like for example, partial least squares (PLS), K-nearest neighbors (KNN) or/and principal components regression (PCR) to analyze the digitized spectra data. [Balabin et all. 2011]

2.2 Online HPLC

High performance liquid chromatography (HPLC) is an analytical technique that can be used to separate, identify, and quantify components in organic and inorganic mixtures. In HPLC, a pump feeds the solvent through a column in high pressures up to 400 atm. This makes the HPLC a much faster method than other column chromatography methods. The high pressure also allows the column to use much smaller particle size packing material, that are usually silica or polymers. The smaller particle size in packing material means higher surface area for interactions between the stationary phase and mobile liquid phase, which results for example in better separation of components in mixture. The principle of liquid chromatography is that different components have different degrees of interaction with the stationary absorption phase, thus they have different elution rates and can be separated from each other as they flow out of the column. Basic flowsheet of HPLC is presented in Figure 11. [Koester, 2016]

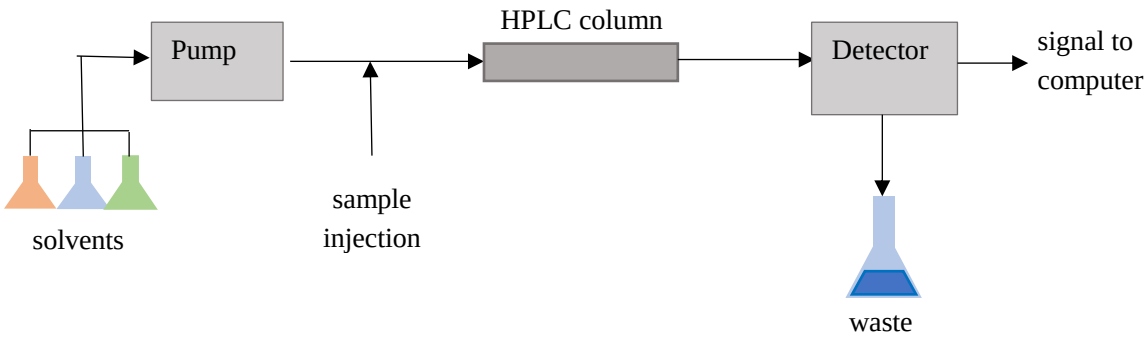


Figure 11. Basic flowsheet of HPLC equipment. Modified from [Koester, 2016]

There are a few limitations to the sample for the use of HPLC. The sample must dissolve completely to the solvent chosen. The sample can't react with the solvent and a suitable detector must be found. [Jaarinen et all., 2005]

If there are solid particles in the mixture, that do not dilute to the solvent, the sample can be filtered before feeding the eluent to the column. Pre-columns are also recommended, as they capture the non-dissolved particles, that could otherwise end up in the analytical column. This prolongs the life of the analytical column and the pre-column can be changed when necessary. The condition of the analytical column should be measured during use by measuring the plate number with a suitable testing liquid. The plate number is first tested for a clean column and at regular intervals. Fouling of the column can be seen as the decrease of the plate number. [Jaarinen et all., 2005]

Depending on the sample, the packing material, type of chromatography and detector should be chosen. Packing material is chosen depending on the pH of the eluent (limited between 2-8 with most packing materials), eluent and sample (no precipitation, or other unwanted reactions can occur) and the type of the chromatography. Chromatography type can be for example normal-phase chromatography, reverse-phase chromatography or ion exchange chromatography. Normal-phase chromatography is the first kind of HPLC that was developed, thus it is named *normal*-phase, but reverse-phase chromatography is most commonly used HPLC type. There are also numerous detector types, UV-VIS-, fluorescence-, electrochemical conductivity-, IR-, and differential refractive index detector to name

a few. The detectors are selected based on the needed selectivity and sensitivity. [Jaarinen et all., 2005] [Higson, 2004]

Eluents are usually a mixture of two or more solutions whose ratios need to be optimized right in order to get the components to separate from one another. The type of the chromatography usually determines the type of eluent used. For example, in reverse-phase chromatography, the stationary phase is nonpolar, thus the eluent needs to be polar. Here an example of eluent is a combination of water and methanol. In reverse-phase chromatography, the most polar components elute first. In normal-phase chromatography, the situation is reverse, as the stationary phase is polar, thus the eluent must be nonpolar, here hexane and its mixtures can be used as eluents. The viscosity of the eluent effects the retention time in the column, the smaller the viscosity, the smaller is the backpressure caused by the stationary phase is. [Jaarinen et all., 2005]

When the composition of the sample stays constant, and only one type of solvent can be used, it is called isocratic elution. Sometimes, the polarity of the sample components changes so drastically, that the eluent that is used for some components, cannot separate the others out of the column in a sensible time. In these situations, the eluent composition can be changed in intervals to get all of the sample components to elute and separated. This is called gradient elution. [Jaarinen et all., 2005] [ChromAcademy]

According to literature, HPLC can identify and quantify metal ions, [Yang et all., 2003] the fatty acids and resin acids of tall oil [Murray et all., 1981], water content [Stevens et all., 1987] and lignin [Ungureanu et all., 2009] but to determine all of these at one go can produce problems for selecting a suitable detector, eluent and stationary phase. It should also be noted that the literature examples mentioned most have studied pure substances, and CTO is very complex substance. HPLC provides fast results, but in online applications, when sampling straight from the product stream, it must be made sure that sample temperature and pressure are not damaging to the sampler or analyzer instruments. [Dionex, 2002] Some disadvantages of HPLC are the cost, that accumulates from the equipment itself, solvents used, hazardous waste management and replacement of columns from time to time. It is also a complex method and coelution of similar components can happen. [Burdick, 2018]

2.3 Low field NMR

Nuclear magnetic resolution (NMR) spectroscopy is used to identify and quantify components in a mixture, and it is the only method that can also identify the three-dimensional chemical structure of components. For NMR to work, the studied nuclei need to have a spin. Nucleus has a spin when it has an uneven number of nuclear particles. The most common types of NMR spectroscopy are the proton (^{1}H-NMR) and carbon 13-isotope (^{13}C-NMR), but it can be done with all nuclei that possess spin, for example ^{15}N and ^{31}P. [Myllyviita, 2019]

NMR is based on the interactions with the magnetic moment of the atom nucleus and an outer magnetic field. The magnetic moment of the nucleus is determined by the nuclear spin and magnetogyric ratio. [Reichenbächer et all., 2012] When the sample is placed in external magnetic field, its nuclei can align either with or opposed to the external magnetic field. When radio frequency in emitted to the nuclei, their orientation is reversed. This is known as resonance condition. When the spin returns to its base level, the energy is released at the same frequency as it was absorbed by the nuclei. This energy is then detected, and the signal can be processed into NMR spectra. [chemquide] Figure 12 shows the basic principle of NMR.

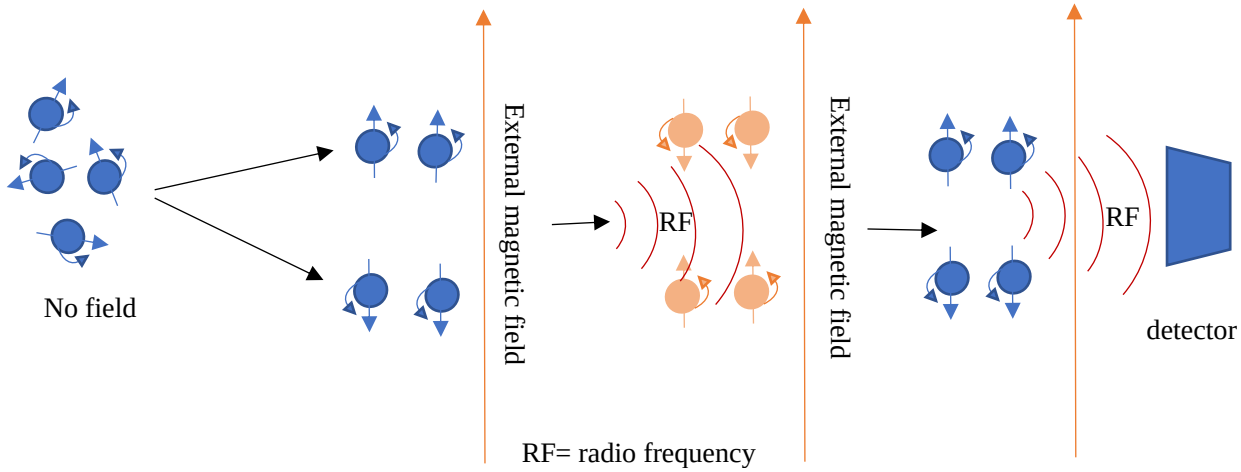


Figure 12. simplified basic principle of NMR.

NMR spectrum shows the chemical sift of the atoms, signal intensity and signal distribution. Chemical sift indicates the nuclei resonance frequency compared with a standard sample in a magnetic field. The signal intensity shows the number of the nuclei of the same kind in the sample, and signal distribution gives information of other nuclei nearby and the chemical structure of the molecule. [Solunetti] [Myllyviita, 2019] For smaller molecules, the spectra are usually very simple to interpret, but big and complex molecules can cause problems like spectral overlapping. In these situations, measurements in higher magnetic field can reduce overlapping or 2D-measurements may help to get more information about the bonds between nuclei. Sometimes chemometric methods are needed to gain information about the spectra. [Martonen, 2019]

For proper measurement and to interpret the spectra properly, it is good to understand the theory behind the method. For the purpose of this thesis, I shall not go into full detail, but there is a lot of literature explaining it all. For example, Edwards (2009), Rule et all., (2006) and Günther (2013) to mention a few.

As NMR equipment is usually very large and costly, low field NMR can offer smaller and less expensive equipment, that is often used online for process control. Low field NMR uses permanent magnets, with the typical strength of 0,5-1 T (10-60 MHz) do not require much maintenance. [Blümich,2005] [Nordon et all., 2002] Low-field NMR has lower sensitivity and resolution than high-field NMR. In low magnetic fields, peak overlapping is common and multivariate regression models, like PLS, can be used to get information from the spectra. In online NMR applications, the process temperature and pressure can be handled by choosing suitable flow NMR probes and sampling equipment. [Zientek et all., 2016] Low field NMR can detect acid concentrations based on the chemical shift of the acid/water signal, which comes from the exchangeable proton of acid and water. [Nordon et all., 2002] Water can also be detected in different physical and chemical environments. [Thybring et all., 2017] LF-NMR can be used in various stages of the process, to show the changes in sample composition. [Nordon et all., 2002]

2.4. XRF

There are many x-ray analyzer methods like x-ray fluorescence (XRF), x-ray diffraction (XRD) and x-ray transmittance (XRT), that can be used as an online analysis method. These methods can be used in quantitative and qualitative measurements, identification, detection, quality and process control. XRF provides the elemental analysis of a sample when XRD gives the compound analysis of minerals. XRD and XRF can be used together. [911 metallurgist] [Baensch, 2014] XRT detects mostly metals, thus can hardly detect organic materials. [Huang et all., 2019] As XRF is the recommended method for biofuel products, it is the focus of this chapter.

XRF is a non-destructive analysis method, that can analyze liquid, solid and gas samples and does not produce any waste. For sample preparation, the XRF method only acquires the homogenization of the sample, as the analysis happens mostly on the sample surface. XRF can analyze concentrations from 0.0001 % to 100%. [Laine-Ylijoki et all., 2003a]

In XRF, x-ray radiation is directed to an atom, where it can remove an electron from the innermost electron shell. In the pursuit of the minimum energy state, the atom fills the inner shell with an electron from the outer shells. This process emits an energy in the x-ray region, that is specific to each element. This energy can then be detected, and the atom identified, the intensity of the signal determines the concentration of the same element in a sample. [Laine-Ylijoki et all., 2003a] This mechanism is presented in Figure 13.

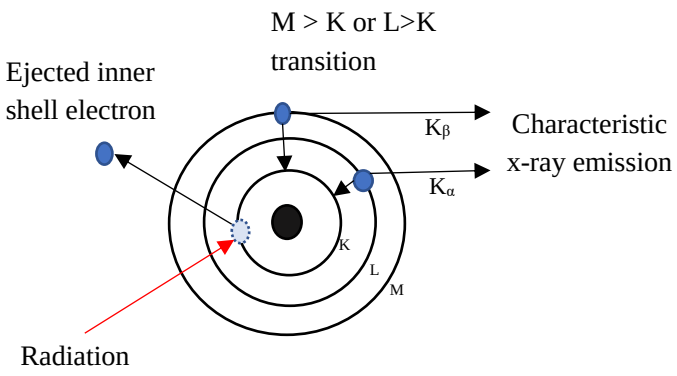


Figure 13. XRF mechanisms modified from [Laine-Ylijoki et all., 2003a]

XRF are usually classified to wavelength dispersion XRF (WDXRF) or energy dispersion XRF (EDXRF) based on the energy detection method. WDXRF physically separates the emitted x-rays according to their wavelengths, when EDXRF directly measures the different x-ray energies emitted from the sample. WDXRF is mostly used in laboratory settings, as EDXRF is usually used in field portable XRF (FPXRF) equipment. [Kalnicky et all., 2001] Compared with WDXRF, EDXRF is cheaper and smaller in size, but their resolution is weaker. The typical detectable atomic weight range for EDXRF is from sodium to uranium. [Laine-Ylijoki et all., 2003b]

FTPXRF are small in size, user friendly and don't require specific expertise to be used. FPXRF can use either isotope sources or x-ray tubes for radiation sources, the x-ray tubes are becoming more popular, as they increase radiation power, decrease measurement time (5-15s) and are safer to use than isotope sources, as isotopes can release toxic radiation into the environment, if not properly capsuled. When using x-ray radiation, there is a lot of safety regulation to be followed. With isotope sources, the measurement time is 30-100s. [Laine-Ylijoki et all., 2003b] [Laine-Ylijoki et all., 2003a] Some disadvantages of the isotope sources are that they emit radiation at certain energy levels, thus they only excite atoms at certain atomic number range. No isotope alone is enough to excite the whole atomic number range. Also, with radiation isotopes, the half-lives determine the chancing of the isotope source as their intensity decreases. The half-lifes of isotopes can vary from 270 days (Co-57) to 470 years (Am-241). [Kalnicky et all., 2001].

There are a few different detectors used in XRF equipment, that vary in resolution, speed and analyte sensitivity. The most common types of detectors are the gas flow proportional detector, scintillation detector and the solid-state semiconductor detector, from which the latter is most commonly used in FPXRF because of the best resolution. Some detectors may require liquid nitrogen as coolant or other electronic cooling. [Kalnicky et all. 2001]

For the quantitative analysis of oil and liquid samples, robust calibration is required in the beginning, some things to consider during calibration are detector resolution, the sample matrix, accuracy of calibration standards and sample homogeneity. The detectors have varying resolution and some element combinations can be impossible

to calibrate due to detector limitations. It is important to choose the right detector for the process in question. The sample matrix can interfere with the measurement of the target element. The percentages of substances in a liquid sample that cannot be measured, for example N, C and O, should not affect the accuracy of XRF. However, in the case that other elements do interfere, it is usually possible to correct this in calibration. [Kalnicky et all., 2001] [Gazulla et all., 2013] The online equipment is customizable for different types of process conditions (the temperature, pressure, viscosity etc.) In the online equipment, the sample flows to a sample cell, where it is analyzed by XRF, and released back to the process. The x-ray passes through a thin window before hitting the sample. The window material is selected based on the elements measured from the sample. [Rigagu] [Hombre] [Malvern Panalytical]

3. Analyzer comparison

In this chapter, the online analyzers studied before are compared with one another on cost, possible problems in process, maintenance needs and analysis abilities to find a suitable equipment. Few process examples for each technique are collected from literature.

3.1 Process examples

As previously mentioned, NIRS works in a variety of fields. One example is in dairy industry, where Nûñez-Sanchez et all. (2016) determined the fatty acid profile in goat milk. Analysis was done with liquid milk in transflectance mode and with dried milk in reflectance mode. Transflectance gave similar or better results than reflectance. This was a favorable outcome, as analyzing liquid milk with accurate results keeps the analysis fast and the sample doesn’t need any pretreatment.

There are multiple reports about the use of NIR and multivariate regression methods in order to predict the quality parameters of biodiesel. Some parameters analyzed were the oxidative stability index, acid number, water content, iodine number and the w% of different oils from different feedstocks (palm, rapeseed and soybean oil) in a mixture. Multiple multivariate methods were used for the quality analysis of NIR

spectra and to create calibration models between analytical reference data and NIR spectra. [Baptista et all., 2008] [de Lira et all., 2010]

Murray et all., (1981) studied the qualitative and quantitative analysis of tall oil fatty and resin acids with HPLC. The experiments were done in the laboratory conditions with distilled and filtered samples. For the qualitative analysis it was found that resin acids need different solvent composition than fatty acids to separate in the column. Different ratios of tetrahydrofuran, acetonitrile and water were used as solvents. For the quantitative analysis of certain acid components (abietic and oleic) calibration curves were constructed to measure the amount of said components in a mixture. In quantitative analysis, the peak area is related to the percentage of the components in a mixture, but for some components, for example linoleic and palmitic acids in the fatty acid mixture, the peaks can overlap, leading to the need for chemometric methods.

Ungureanu et all. (2009) used HPLC to characterize wheat straw lignin, and to analyze difference between the unmodified lignin and hydroxymethylated lignin. Lignin is a chemically active macromolecular compound, due to phenolic and aliphatic hydroxyl functional groups, found in plants and wood structures. The hydroxymethylation contributed to changes in the lignin molecular structure, allowing lignin to be used in more effective practical applications. This change in the molecular structure by hydroxymethylation was proven with multiple analysis methods, including HPLC. The HPLC analysis occurred in laboratory conditions, with both samples the solvent was a mixture of water, methanol and acetic acid. It could be seen in the resulting spectra that the unmodified lignin gave only one peak, as the modified lignin gave two. This proved the changes in lignin polymolecularity as a result of different functionalization of the lignin fractions or of the occurrence of some condensation reactions.

Low field ^{1}H-NMR has been used to analyze light, medium and heavy petroleum fractions, that include substances like phenols, carboxylic acid, alcohols and others. Determined analyses were the refractive index, viscosity, density and total acid number. Results were seen to have correlations with the standard reference methods, so LF-NMR was recommended for the quality control of petroleum fractions, as it is non-destructive and a fast analysis method. [Barbosa et all., 2013]

Silva et all. (2012) studied low field ^{1}H-NMR on crude oil-water biphasic mixtures. Two series of samples with different water content were made. First series was with oil and pure water and in the second series 360 ppm aqueous solution of Mn^{2+} was used. Hydrogen is present in both, water and oil, but there is a difference in the two substances ^{1}H- nuclei relaxation behavior, which is the distinguishing factor. With the first series, there was a big difference in the 1H-nuclei relaxation behavior, so the quantities of different phases were easy to detect straight from the spectra. With the second series it was found that the concentration of paramagnetic ions accelerated the relaxation of ^{1}H-nuclei in the water rich phase, so that the corresponding T_2 (spin-spin relaxation) values were in the same range as the values of the oil. In this case PLS regression was used to help calculate the water content from the spectra data.

XRF can be used to determine metals and minerals in oil and petrochemicals. Zajac et all., (2015) analyzed heavy metals in used oil. Simultaneous multicomponent analysis in the hydrocarbon matrix was performed with high definition XRF (HDXRF). HDXRF was able to detect high (1389 mg/kg of Zn) and low (0,14 mg/kg of Ni) concentrations of chosen metals. XRF was comparable to other metal detection methods like atomic absorption spectrometry (AAS) and inductively coupled plasma-optical spectrometry (ICP-OS) and had the advantage of being nondestructive and didn’t require any sample preparation.

3.2 Analyzer comparison

Analyzers are compared in tables III and IV. In table III analyzers are compared based on their ability to carry out the measurements wanted for CTO. The fiber is thought to consist of cellulose and hemicellulose or other similar substances that could be left from the pulping. For the determination of lignin and fiber with LF-NMR, there are multiple reports to determine these substances in high field NMR, however, none could be found for low-field applications after extensive search, so it could be possible, but could also be that these are too large molecules for low field applications.

In Table IV the analyzers are compared based on disadvantages and problems that could occur in the process. It should be noted that only problems that distinguish and

are specific for the said analyzer are mentioned, not basic or annual maintenance that most equipment has by default and is usually provided by the manufacturer. The cost is not specifically defined for these analyzers, but any chemical addition, waste treatment or frequent equipment part change adds up the cost. Based on the literature study and tables III and IV, NIRS is chosen as an online analysis method to be further tested.

Table III. Comparison of analysis methods based on wanted measurements. [Nomngongo et all., 2016] [Bart et all.,2013] [Li et all., 2015] [Yang et all., 2003] [Murray et all., 1981], [Stevens et all., 1987] [Ungureanu et all., 2009] [Laine-Ylijoki et all., 2003b] [Thybring et all., 2017] [Nordon et all., 2002]

	NIRS	HPLC	LF-NMR	XRF
Acid number/ acid content	x	x	x	
Water content	x	x	x	
Metal/mineral impurities	x	x	x	x
Lignin	x	x		
fiber	x	x		

Table IV. Disadvantages and possible problems of the analysis equipment. [Jaarinen et all., 2005] [Burdick, 2018] [Zientek et all., 2016] [Kalnicky et all., 2001] [Romañach, 2010] [Laine-Ylijoki et all., 2003b]

	NIRS	HPLC	LF-NMR	XRF
dangers/ waste/ toxins	-	liquid waste	-	possible radiation if isotopes are used
chemical addition	-	solvent chemicals	-	-

maintenance/ calibration issues	Secondary method, samples and components need to be individually calibrated and checked with reference methods.	Filters, column or/and precolumn must be changed from time to time.	-	Renewal of radiation source depending on the type used.
other problems	Relies on chemometrics. May have problems of analyzing trace substances less than 1% in a sample.	If any undissolved particles are present, without filtering, risk of column fouling. As CTO is so complex, eluting and detecting all components at the same time can be problematic.	Complex molecules may be problematic to detect in low magnetic fields.	Detectable atomic weight ranges from sodium to uranium.

4. Mathematical methods and methodology in modeling online NIR

Multivariate methods for analysis, monitoring and diagnosis of process operating performance are becoming more important, because of the large amount of data

collected in an online process. With multivariate methods, the aim is to maximize the amount of information, and at the same time reduce the information in to a more easily interpreted form. Some main objectives of multivariate models are data reduction, grouping and classification of observations and modeling of relationships between variables. For NIR data, multivariate methods are accepted as standard methods due to their ability to handle cross correlated variables, such as wavenumbers in spectra. [Cordella, 2012] [Kourti et all., 1994] Methods applied in this thesis are principal component analysis (PCA) and partial least squares (PLS) regression.

In different parts of this chapter, some examples are made to visualize pretreatment of the data, PCA and PLS, using NIR-spectral data from a set of 21 samples of CTO. This dataset is referred to as CTO21 in the text. Matlab software is being used to treat the data.

4.1 Pretreatment of spectra

4.1.1 Scaling for multivariate models

Before using the models, the raw data usually needs to be pretreated into a suitable form. Most common pretreatment methods are autoscaling and centering. In centering, the variables are moved around the origin. Since all latent variables (PCA and PLS components) are always fitted via origin in direction of dominating variance, the first latent variable would be drawn from the origin to means of values, unless variables are centered. This would be an unnecessary component. Centering is done by calculating the mean value of the data samples and subtracting that value from the data points. This is pictured in Figure 14. [Teppola et all., 1997] [Eriksson et all., 2001]

In autoscaling, the variables are scaled into unit variance, where the value gotten from centering, is divided with the standard deviation of the data. Autoscaling to unit variance is done with data that has variables that have widely different numerical values and/or are in different units, like height (m) and weight (kg), or in spectral data, baseline variation and peaks with different heights and variation range. Autoscaling removes the effect of the absolute values, this ensures that each variable has an equal opportunity to affect the model. This is visualized in Figure 15. Variance of the original

variable equals to its weight or importance in all least squares fitting procedures. The importance of each variable can be adjusted via its variance. When variables have equal variances, the method handles them equally, and extracts from variable matrix the systematic variation unbiased by the variances. When handling spectra, the small peaks often hold the most interesting phenomena although the high peaks tend to dominate the variation. Furthermore, the narrow high peaks are often the most unstable ones. Autoscaling might also increase noise due to scaling baseline, but that is seldom reported as a problem. Thus, autoscaling is often applied in NIR based model. [Teppola et all., 1997] [Eriksson et all.,2001]

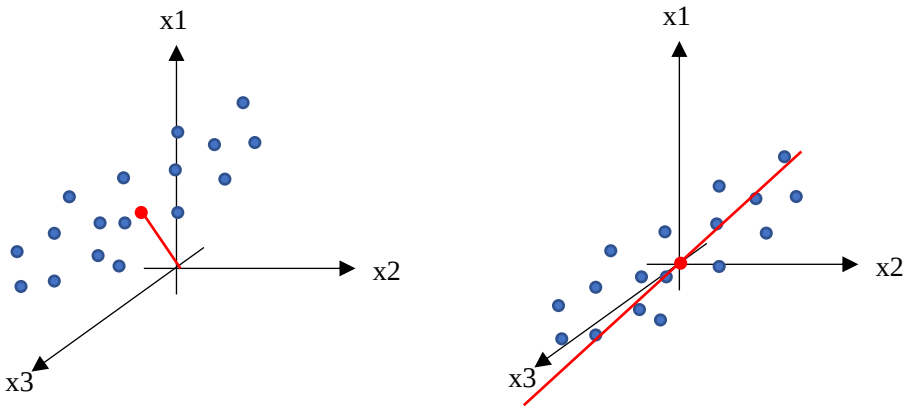


Figure 14. Centering. Line fitted via origin.The variable averages are calculated from the data and then subtracted from the data, this moves the data in a way that the average of the data is moved to the origin. Modified from [Eriksson et all, 2001]

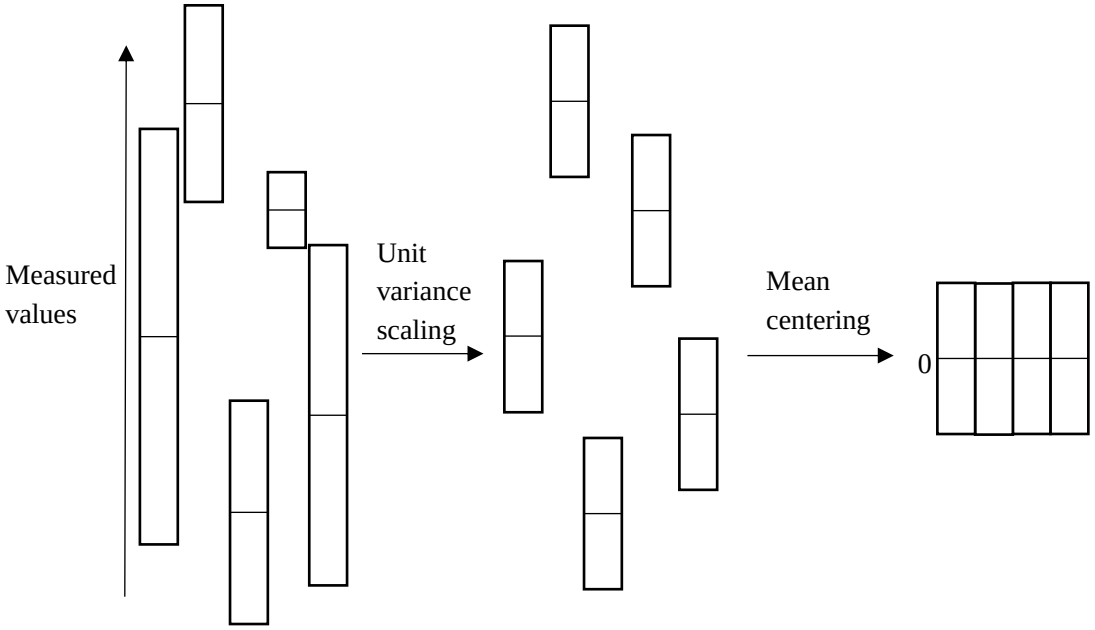


Figure 15. Autoscaling. The columns represent variables with different measured data and the horizontal line in each column represents the mean value of that data. Prior to pretreatment, the variables have different lengths and different mean values. After unit variance scaling, the variables all have identical lengths, but the mean values are still different. After mean centering, the variables all have equal length and their mean value is zero. Modified from [Eriksson et all., 2001]

Normalization of data is similar to scaling and standardization. In normalization variables are scaled to have values between 0 and 1. In standardization, a mean value is subtracted from data, and the data is divided by standard deviation. This creates variables with mean of 0 and standard deviation of 1. Standardization is usually used when the algorithm used makes assumptions about the distribution of data, like linear regression. Normalization on the other hand, is used when the algorithm used does not make assumptions of the distribution of data, for example KNN. [Lakshmanan, 2019]

When using spectral data, baseline correction is an important pretreatment step. In spectrum signal, can cause uneven amplitude sifts across the wavenumber range and lead to bad results. These sifts in amplitude can be caused by phenomenona like fluorescence, phosphorescence and black body radiation, and they should be compensated before any further analysis. The effect of baseline correction is shown in Figure 16. There are multiple ways to do this step, and the algorithm used should be

chosen based on the spectral data. Manual baseline correction relies mostly on user expertise, noise level and the characteristic of the baseline, thus it has some room for error, and is not always completely accurate. In process application, the analysis program usually has automatic baseline correction algorithm in use. In figure 16, the baseline correction has been done manually, using 3rd degree polynomial fitting. [Qian et all., 2017]

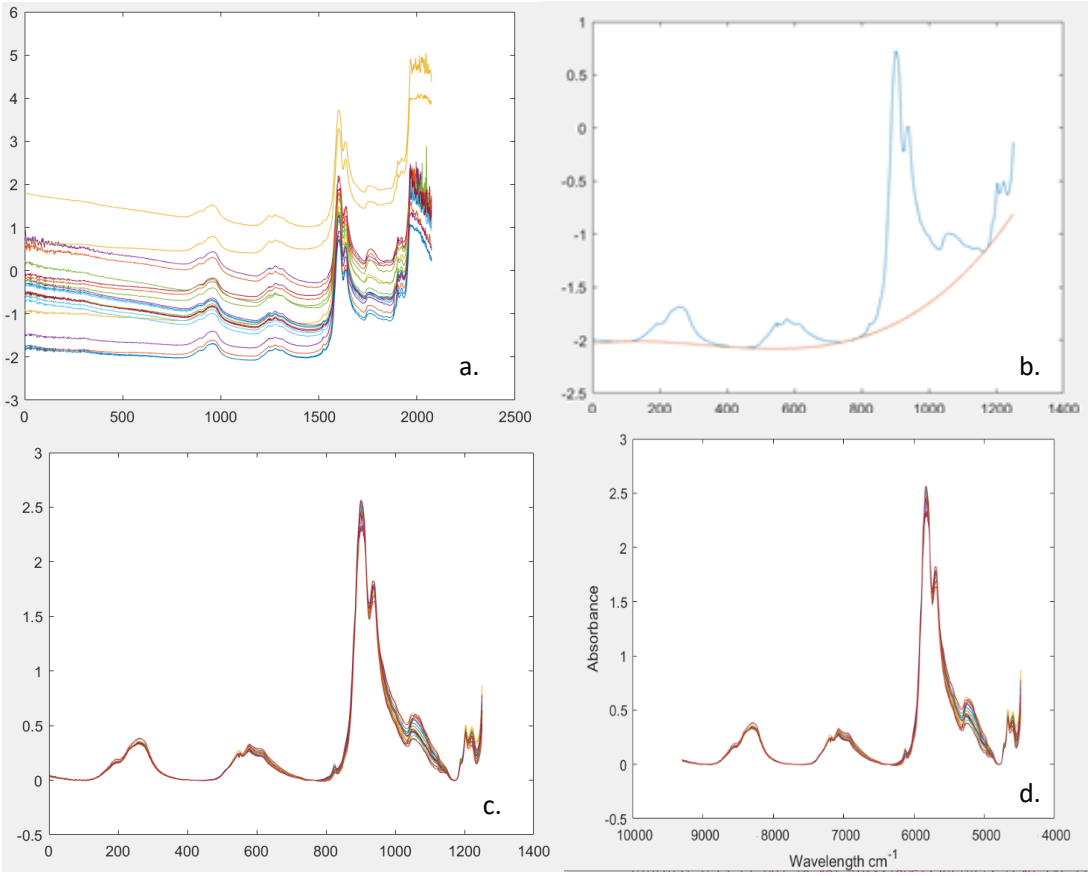


Figure 16. a: spectral data of CTO21 samples that has not been baseline corrected. b: one sample spectrum that is fitted 3rd degree polynomial line. Where the polynomial line hits the spectrum, are the spots that are set to go via zero when the baseline is removed. c: the same data as in figure a, but with the baseline removed. d: spectral data with real wavelength values in x-axis.

In Figure 16a, the spectrum is shown as whole, in figure b, noisy spectra from the beginning and end has been cut, As can be seen from Figures 16a, b and c, the real wavenumber values are not shown as default. The range goes from 0 to 1251 they are

the numbers of measurement points. These values represent the wavenumber values from 9299 to 4478 cm^{-1}. Real wavenumber values are shown in figure 16d.

4.2 Monitoring process and quality of spectra

4.2.1 PCA theory

A large dataset consists of multiple samples and possibly of hundreds of variables, which can make it hard to interpret. Each variable ads a new dimension to the analysis, and if there is more than 3 dimensions in a dataset, the visualization becomes impossible. Principal component analysis (PCA) is an analysis tool that is used to reduce the number of dimensions, without the loss of information and variance, by creating new variables, principal components. If there was a dataset that had a hundred variables, PCA tries to compress that information down to fewer principal components. The first principal component should explain most of the data, by finding correlated variables and the second should explain the residual variance as good as possible. The different principal components are perpendicular to each other, and thus not correlated with each other. This is visualized in Figure 17. After choosing the principal components that explain most of the data, the rest can be discarded without losing too much information. Choosing to use too many principal components usually just weaken the model's prediction abilities, as the last principal components typically represent only noise. [Jolliffe et all., 2016] [Zhang et all., 2017] [Mujunen et all., 1997] [Teppola et all., 1997]

Optimal number of principal components can be chosen with many different methods, for example with cross-validation, proportion of variance accounted for and eigenvalue-one criterion. In cross-validation, some of the samples are left out from the calibration sequence and for these samples PRESS (predicted residual sum of squares) value is calculated as a function of the number of principal components. The optimal number of principal components is when this value is at minimum. Proportion of variance accounted for -method, retains a component, if it accounts for specified percentage of variance in the data set. For example, it can be decided to retain all components, that account for at least 5% or 10% of the total variance. This percentage

can be calculated by dividing the eigenvalue of a component by total eigenvalues of the correlation matrix. Eigenvalues represent the amount of variance that is accounted for by a given component. In the eigenvalue-one criteria, any component that has eigenvalue larger than one, is taken into the model. If the eigenvalue of a component is larger than one, the component is accounting for greater amount of variance than had been contributed by one variable. Thus, if the eigenvalue is less than one, the component accounts for less variance than one variable. [Mujunen et all., 1997] [Teppola et all., 1997] [O'Rourke et all., 2013]

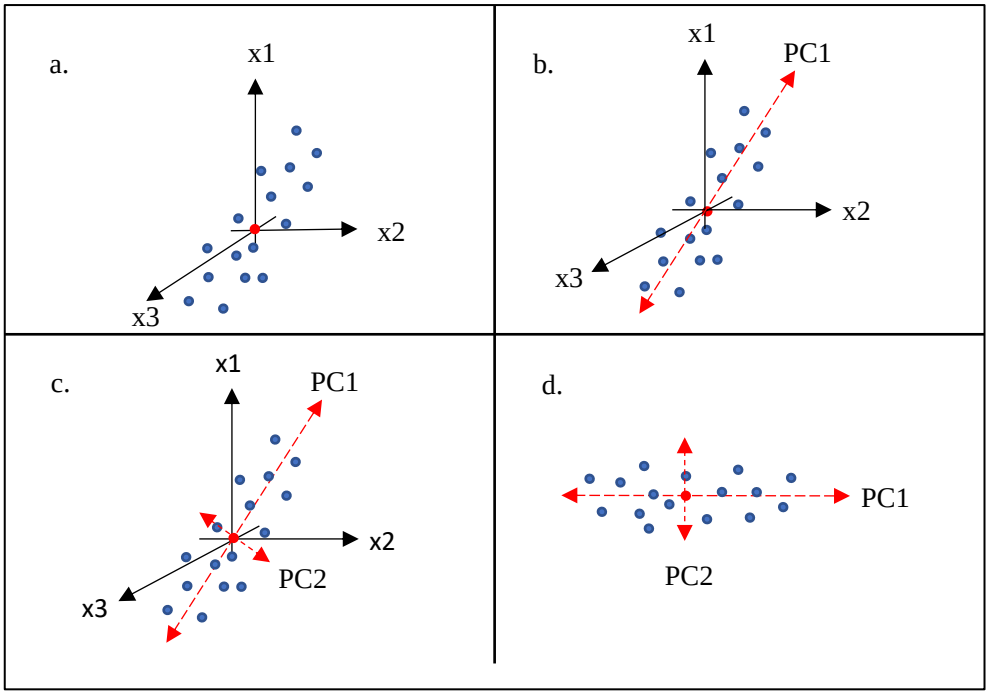


Figure 17. Creation of principal components. a) pretreated data in multidimension b) First principal component explains as much of the variation as possible c) second principal component explains the rest d) the principal components create new axis for the data.

Analysis data is first collected in a matrix X, that has *n* lines and *p* columns. Each line corresponds with an observation, what in this case would be an individual NIR spectrum of one sample. Each column corresponds to a variable. Typically, this data matrix is huge and multidimensional. In PCA this matrix is represented with equation

(1), that has much smaller matrices T and P to represent the original matrix. [Cordella, 2012]

$$X = TP' + E \quad (1)$$

Where X matrix

T scores

P’ loadings

E residual noise

4.2.2 PCA in process monitoring

As can be seen from Figure 16, it is nearly impossible to gain specific information or asses the similarities of the samples from the raw spectra without any additional analysis. For this reason, different monitoring charts can be used in to identify abnormalities and follow the process trend, to ease the routine monitoring of the process. With warning limits, spectral areas that are abnormal to the process can be identified, and thus the reason for the warning can be indirectly identified.

4.2.3 Score and loading plots

The scores correspond to observations, aka the samples, and the loadings correspond to variables. Both of these are calculated for one principal component at a time. Scores are new coordinates of the original observations projected on the principal component axis. Loadings tell the orientation of the principal component axis in the original variable space. The larger the loading, the more significant the variable is for the formation of the principal component. The score and loading plots can be used to visualize the data, to see for example the nature of the variation, variable correlations for the principal components and the grouping of the samples. [Mujunen et all.,1997]

In Figure 18, the 1st and 2nd principal component scores of CTO21 are presented. If these samples were taken in continuous process, the trend of the process could be seen here. Plotting the scores one PC at a time doesn't tell much else from the process though. In Figure 19 the scores of PCs 1 and 2 plotted against each other are presented with 95% and 99% confidence lines. This figure tells us much more. As can be seen, there are no outliers in the data that would go outside of the confidence ellipses. In the right-side figure groups are outlined. Same groups can be seen in Table V. It can be seen that samples in these groups have similar qualities. Samples in Table V are arranged by increasing resin acid value.

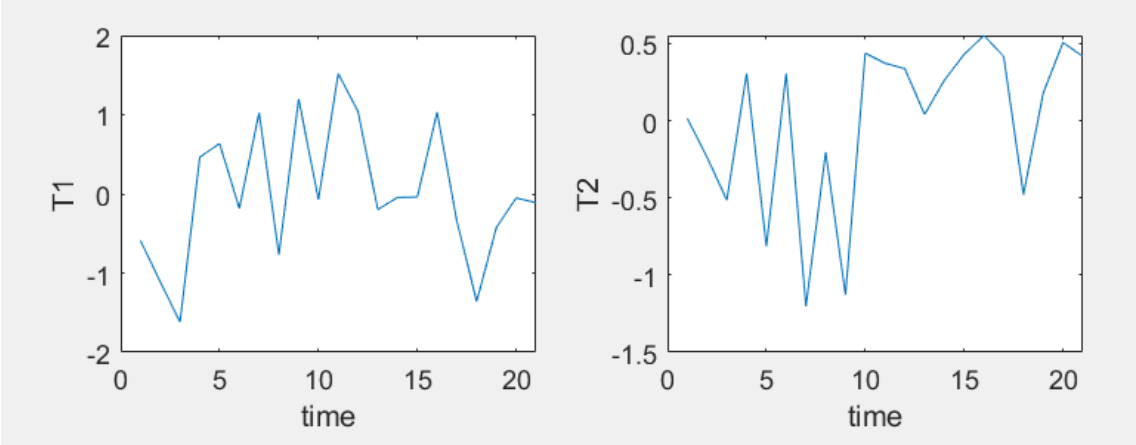


Figure 18. Score plots of PC1 and PC2 of CTO21 samples.

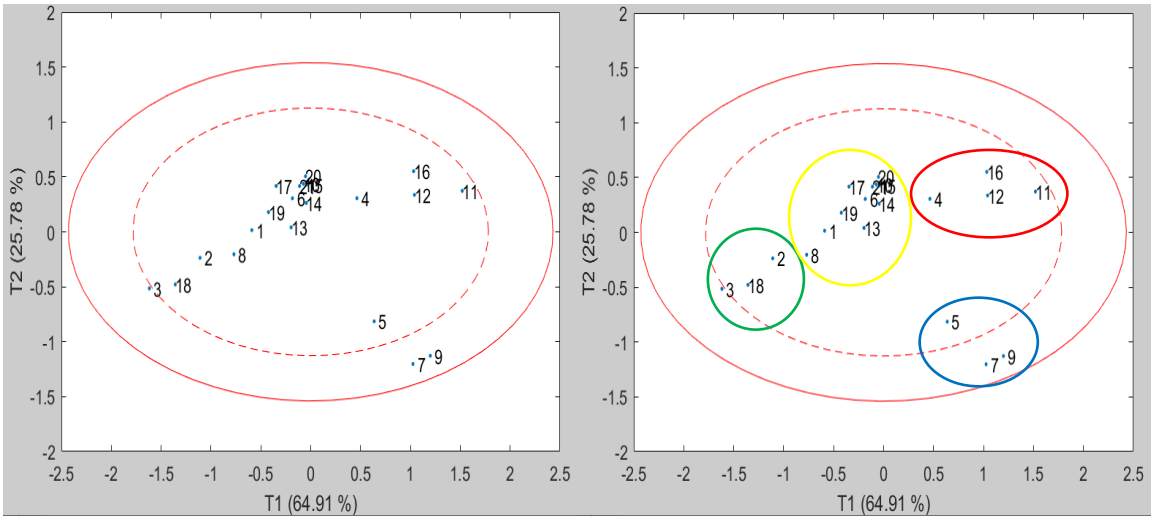


Figure 19. Scores of PC1 and PC2 plotted against each other. Sample groups presented in the right side.

Table V. Samples in groups. Samples are arranged by increasing resin acid number

Sample number/group		11	16	12	9	7	5	4	1	10	14	17	8	19	21	20	13	15	6	2	18	3
Sample ID		19-21178-001	19-23138-001	19-22147-001	19-20948-001	19-20525-001	19-19828-001	19-19818-001	19-18692-001	19-21175-001	19-22777-001	19-23143-001	19-20947-001	20-00589-001	20-01666-001	20-00590-001	19-22148-001	19-23135-001	19-20524-001	19-19456-001	19-23144-001	19-19811-001
Resin acids	Wt-%	10.5	12.5	13.0	13.6	14.7	16.6	17.2	21.0	21.9	22.4	22.8	24.4	24.5	24.9	25.2	25.3	25.9	26.7	29.8	35.3	38.3
Acid number	mg/KOH/g	124.2	130.8	127.0	111.5	116.1	114.7	126.1	127.2	140.2	140.3	140.1	141.3	141.5	138.5	140.3	136.0	140.2	138.0	146.1	155.5	161.7
Neutral substances	Wt-%	36.9	33.4	35.3	43.1	40.7	41.3	35.5	34.7	28.1	28.0	28.1	27.4	27.2	28.7	27.8	30.0	27.8	28.8	24.6	19.5	16.2
Fatty acids	Wt-%	52.6	54.1	51.7	43.3	44.6	42.1	47.3	44.3	50.0	49.6	49.1	48.2	48.3	46.4	47.0	44.7	46.3	44.5	45.6	45.2	45.5
residual soap number		0.13	0.11	0.13	0.06	0.06	0.09	0.02	0.07	0.38	0.21	0.17	0.01	0.26	0.10	0.17	0.01	0.04	0.02	0.00	0.06	0.08
Water	Wt-%	0.23	0.09	0.45	0.71	1.17	0.90	0.44	0.53	0.24	0.49	0.39	0.52	0.27	0.36	0.20	0.55	0.10	0.43	0.71	0.31	0.51
metals	mg/kg	56	40	51	38	38	43	38	22	250	226	446	388	268	163	212	102	194	75	576	59	43

In this dataset, loadings represent the wavenumbers where measurements were taken. In Figure 20, a loading plot is presented. In figure 21, two biplots of loadings and scores are presented. On the left side, all loadings are presented, and on the right side for clarity only some loadings are presented.

In the loadings plot, the farther away from the plot origin the variable is, the stronger the impact it has on the model. In figure 21, we can see that the three variables that are furthest away from the Origo, represent the biggest peak in the spectra. Also, variables that have similar information are close to each other, like variables 1050 and 1100. The variable 900 is further away from these two, so these two peaks do not present the same information, yet they are side by side in the spectra. When variables are on opposite sides of the plot, they are inversely correlated, meaning when other increases the other decreases. [Eriksson et all., 2001]

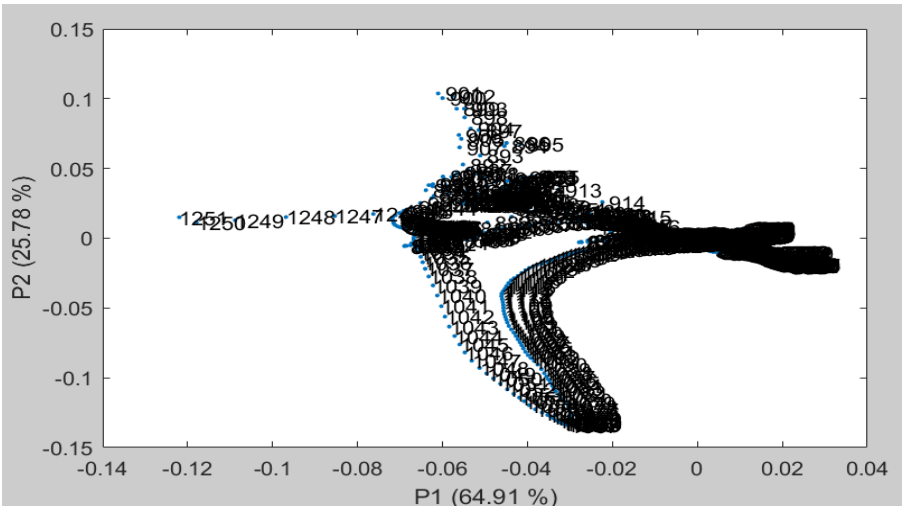


Figure 20. Loadings.

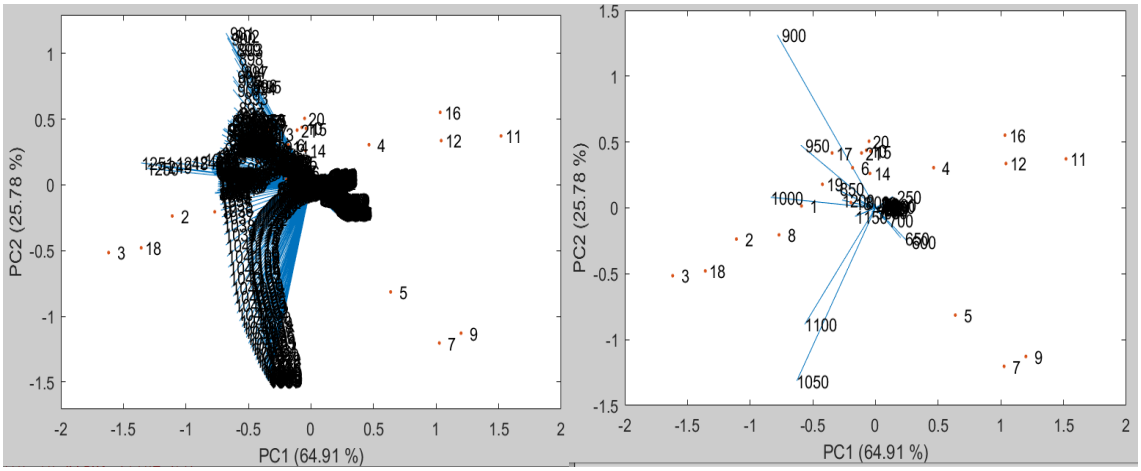


Figure 21. Biplot of scores and loadings. Scores correspond to the samples and loadings correspond to wavenumbers. On the left, all 1251 loadings are presented, and on the right, only some are presented for clarity.

When number of PCs is relatively low (2-3), visualization of the process can be easily done with PCA scores and loadings plots to identify sample groups and show variation. But examples presented here only visualize the principal components 1 and 2. With this dataset, there are five principal components in total, that each give different plots, thus controlling the process this way becomes impossible in routine use. In routine process control, multivariate statistical process control (MSPC) charts are needed.

4.2.4 Multivariate statistical process control charts

Product quality often depends on several variables, that occur simultaneously. Thus, measuring just one or two variables does not give a comprehensive image of the product. Multivariate statistical process control (MSPC) charts can be used in process control, to visualize the process in real time with ease, and to indicate any malfunctions in the process. Any number of variables can be controlled two charts, Hotelling's T^2 and squared prediction error of X SPEx (sometimes called Q statistic). Note, that these charts need to be used simultaneously.

The T^2-value can be used to visualize the systematic variation of each sample within the model. It tells how far from the center of the model the object is. If the T^2 plot is

made for a stable and good process situation, when the T^2-value for an object is small, the process works as planned. High SPEx values indicate of (new) variation unfamiliar to model – in process modeling this could be for example due to change in raw material, process units, malfunctioning of instrument, etc.

T^2 and SPEx plots are mostly used to detect disturbances and malfunctions in the process. After the malfunctions in the process are detected, a more precise investigation can be conducted to detect the reason for the malfunction. [Mujunen et all., 1997] [Kohonen, 2009] [Kourti et all., 1994]

In Figure 22, T^2 and SPEx MSPC charts can be seen for the CTO21 data. In Figure 22, we can see that according to the T2 chart the process is under control, but the SPEx chart indicates abnormalities in sample 16. From the contributions chart in Figure 23, the variables, in this case the wavenumbers, that affect the abnormalities can be seen in the contributions chart. If it is known, what quality parameter the wavenumber area represents, the reason for the abnormality can be detected.

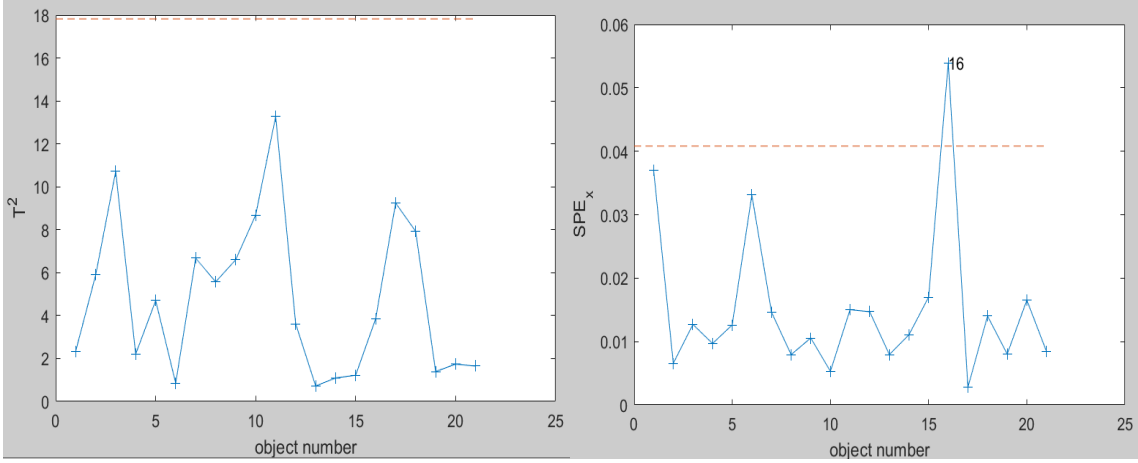


Figure 22. T^2 and SPEx MSPC charts of the data. Here the dashed line represents the 95% confidence limit, the 95% SPEx limit is 0.0408063 and the T^2 limit is 17.8209. The confidence limit values could be changed manually to better fit the process in hand, or they could be statistically calculated as they are here.

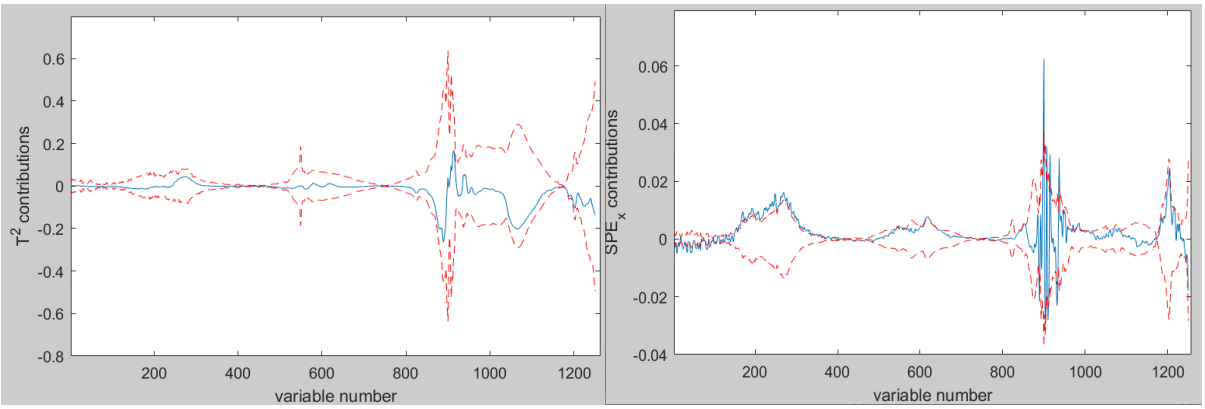


Figure 23. T^2 and SPEx contribution charts for sample 16. The red dotted line represents the confidence limits, and the blue line the contributions for each variable. In the SPEx contribution chart the sample 16 goes over the confidence limits in multiple spots, which causes the sample to cross the MSPC SPEx chart limit.

4.3 Predictive models with PLS

4.3.1 PLS theory

Partial least squares regression is an extension of the PCA method, but as PCA is an analysis for one matrix X, PLS models collinearity of two matrices X and Y. The X matrix includes the observable variables, e.g. spectra and the Y matrix consists of response variables, in this case the concentrations to be modeled. In PLS, X and Y are first modeled with PCA type model, and then the solution in the X block is rotated so that covariance between X and Y is maximized. Covariance indicates the direction of the linear relationship between the variables. In the case that data is autoscaled, the correlation between X and Y is maximized. Correlation values are standardized and indicate the strength and direction of the linear relationship between the variables. The matrix equation for X is the same matrix as in PCA equation (1) and the matrix for Y can be written as equation (2).

$$Y = UQ' + F \quad (2)$$

Where U scores

Q' loadings

F residuals

In PLS the components are rotated so, that the correlation between scores T and U is maximized. The geometrical model of this phenomena is presented in Figure 24. [Mujunen et all., 1997] [Mujunen et all., 1996] [Saha, 2018]

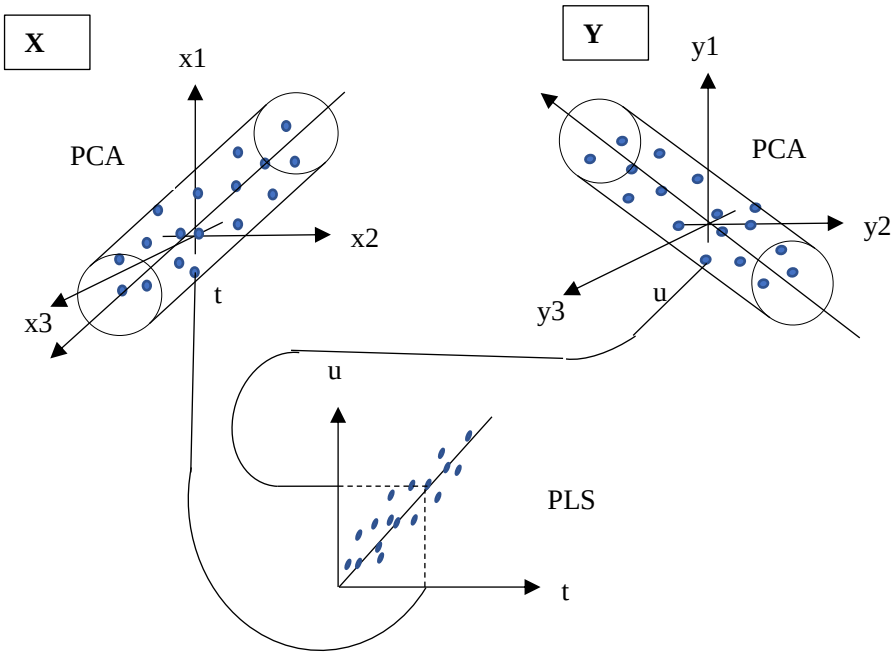


Figure 24. The geometrical interpretation of PLS. Modified from [Mujunen et all., 1997]

In rotation, a new matrix W (variable weight matrix) is formed. It describes how the original X variables affect each PLS component. It can also be applied in calculating the regression coefficients of the model. PLS is a regression model and regression coefficients can be applied in a similar way to any other regression models. Regression coefficient can be calculated based on equation (3). [Kohonen, 2009]

$$b_{PLS} = W(P'W)^{-1}Q \quad (3)$$

Where	b_{PLS}	correlation coefficient
	W	weight matrix
	P	loadings of X
	Q	loadings of Y

To demonstrate PLS with a simplified example, we can study a set of different wines and analyze them in the lab. The lab results can include sugar content, alcohol content and acidity. These values and variables are placed in the predictor matrix X. Matrix Y can consist of depending variables of how well the wine goes with meat or dessert or what the price for the wine is. With PLS, we can use the matrix X to predict the matrix Y and detect that e.g. wines with a larger sugar content are collinear with dessert wines. Or that the price of the wine is correlated on the alcohol content. [Abdi, 2007] In the case of CTO, we can use water content as an example. Here the NIR-spectra would consist of the X matrix, while Y matrix would consist of water content in CTO that would be predicted based on the NIR-spectra.

4.3.2 PLS in process modeling

The goal of PLS process modeling is usually to produce precise estimates for the present time (t (0)), or predictions for near future (t (+1, +5)). PLS method can be applied as a modeling tool for calibration of laboratory models in research and development (R&D) or as a daily routine tool. In this thesis the most common applications are in focus. However, there are various advanced PLS based modeling strategies. Some of these are shortly presented in Table VI.

Table VI. Advanced PLS based modeling strategies.

Method	Short description	Reference
Non-linear PLS	In some cases, the data can have nonlinear behavior, and linear approaches to model the data does not work. Mild nonlinearity can be attempted to be modeled linearly by adding a few extra latent variables. Or nonlinear modeling methods can be applied. There are multiple PLS based models for nonlinearity, here only a few are mentioned. Mild nonlinearities can be modeled by quadratic polynomial PLS, where polynomial is fitted between the score vectors of X and Y. Moderate nonlinearity requires higher-order polynomials and can be modeled with methods like neural networks PLS and spline-PLS. Severe nonlinearity can make the data look discontinuous and clustered and can't be modeled by a single continuous model.	[Eriksson et all., 2001] [Kohonen, 2009]
PLS path modeling	In a production process with multiple units and process variables, the importance and effect of these units and variables to the outcome product, its quality and production cost can be estimated with PLS path model (PLS-PM). PLS-PM gives an estimate of the end product or cost, based on the raw material, that changes and becomes more accurate when the product moves through each unit in the production process. PLS-PM highlights the connections between production units, variables and latent variables, to show their effect to the product,	[van Kollenburg et all., 2020]

	and indicate what parts of the process should be controlled firmly, are there important variables that could have otherwise been deemed uninformative and what action to take when something goes wrong.	
multiblock PLS	Related to PLS regression, but instead of one concentrated data matrix X, one obtains model parameters for multiple individual data blocks X_1, X_2...X_i that are all used to predict Y. The blocks can be divided by the type of the variable or according to the part of the process. It is meaningful to combine the data blocks to obtain a reliable prediction model. This method also provides information on how much each block X_i contributes to prediction of Y.	[Baum et all., 2019] [Kohonen et all., 2008] [Kohonen, 2009]
Priority PLS	Like in multiblock PLS, the data is in different blocks Xi, but instead of modeling all the data simultaneously, in priority PLS the blocks are treated in series. Each variable is given a priority number of 1,2,3 etc. First the variables with priority 1 are modeled, and all other variables have their weight set to zero. When modeling is done to priority 1 variables to a satisfying end, the second set of variables with priority 2 are introduced, and all other variables are once again weighed to zero. This is done until all the variables have been included. Priority PLS can help assessing the importance of variables. It can be done when modeling is being studied in the light of groups of variables.	[Kohonen et all., 2008] [Kohonen, 2009]

N-way PLS	A multidimensional PLS model with 3 or more dimensions used. For example: object × variable × time. In the case of online NIR, spectra from multiple instruments before and after process unit could be placed in the same matrix. Decomposing the spectra with an N-way model gives additional information of the chemical system and reactions.	[Stordrange et all.,2004]

4.3.3 Stationary or dynamic model

Creating a model usually starts in a laboratory with a stationary model, where calibration samples with certain variation range are chosen, and the product qualities are modeled based on the reference values. This gives information on what can be modeled e.g. with NIR spectra. A model based on calibration data is not reliable outside the variation range of the calibration samples. In industrial processes, the quality of the product measured usually tends to change a lot along time, whether due to product change, raw material change, cleaning or fouling of measurements instruments etc. Due to these changes it is not recommended to use a single stationary model for a long period of time. This is why its dynamic models are often used in continuous industrial processes and a strategy for routine process monitoring and validation should be created, to ensure the accuracy of the measurements and estimates. [Kohonen, 2009]

There are multiple ways to construct a dynamic model, and it is highly dependent on the case, which type of model should be used. Dynamic model approach also holds some risks, that should be considered. The models created can be local or global. Being local means that the model is more concentrated at the present time and forgets its history at some point. Global model on the other hand saves its history for longer. If the model has a short history it gives a good estimate on the near future, this method is also recommended if the process is constantly changing, as it would require a huge effort to remember all the possible situations, and with long history the model could

become slow. If on the other hand the process only has a few different product types, it would be useful for the model to be more global and remember the previous products and learn new products as they come along. It is possible to create a model that updates automatically, or a model can also have a constant stationary component to it and an updating part. In all cases, it is highly important to validate the reference values and spectra used for the calibration update, to ensure the data is valid before it is used. This can be done for example with MSPC charts. [Kohonen, 2009]

Visualization of changes in dynamic PLS model (and process) can be via regression coefficients. In Figure 25 regression coefficients from a stationary model and a dynamic model are shown for comparison and to indicate the importance of a dynamic model in a continuous process. In figure 25a spectra for NIR diesel measurements used to model viscosity index from a continuous production process is presented. In Figure 25b regression coefficients from the data for a stationary model is shown. Regression coefficients indicate the effect of each variable for the model. Regression coefficients can be positive or negative. Here if the value is positive, increase on those parts means higher viscosity, and if it is negative, further decrease here means lower viscosity.

When updating regression coefficients are drawn, it gives and in-sight to process dynamics in timewise. Figure 25d illustrates regression coefficients drawn as a pseudo color plot. Time axis is the vertical axis and variables, or wavenumbers are in the horizontal axis. For the dynamic model, only a short range of the variables is selected to show here, this selection is from variable 170 to 250, marked by two vertical lines in Figure 25a. Figure 25c, illustrates the same regression coefficients as Figure 25d, but they are visualized as a colormap on spectra. It can be seen from the color plots that the effect of the variables changes over time. There are time periods when this spectral range has had no effect on viscosity. Coefficients near zero are marked with light blue as shown with the color bar (Fig 25d). There are short time periods when certain wavelengths or small peaks have affected positively or negatively to viscosity of the product. These are marked with yellowish or dark blue colors.

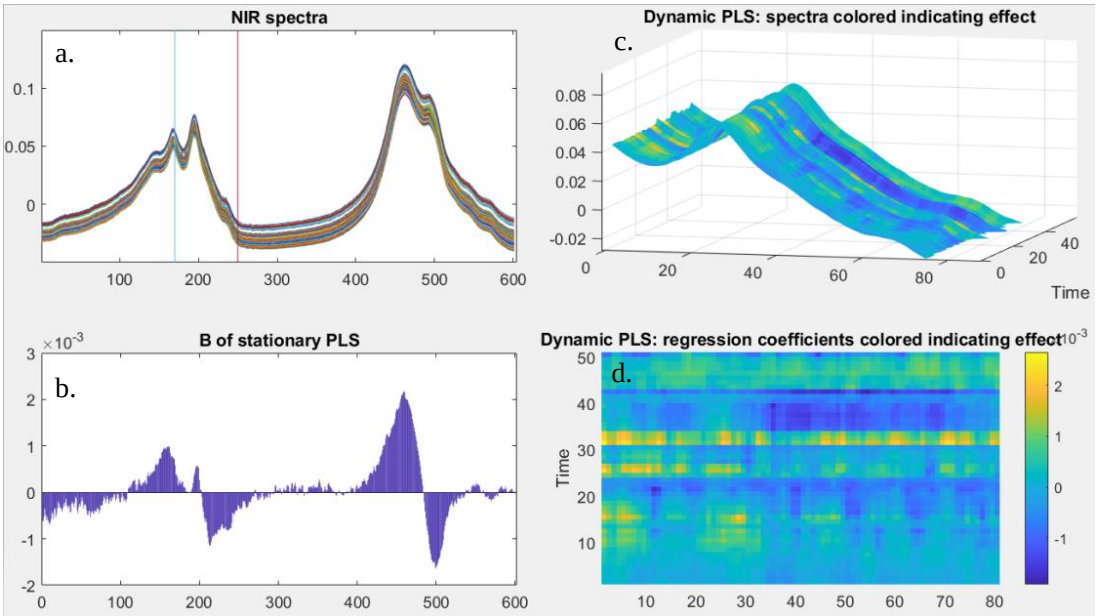


Figure 25a) Diesel spectra used to model viscosity index. b) Regression coefficients for stationary model. c) Dynamic model regression coefficients changing over time illustrated as a colormap on spectra of variables170-250. d) Same information as in figure c, but illustrated as pseudo color plot.

In Figure 26a, a stationary PLS model is presented for the same diesel data that in Figure 25, here the. In the Figure 26 the blue lines represent the real process and red lines represent the predictions, from here it is clear that a single stationary model does not predict the process properly to a satisfactory end. In Figure 26b, the same stationary model is used, with one exception. Here more history is added to the model as the last measured Y reference value (at time t-1) is added to the X variables and more weight is added to it. This makes the models predictions better, but they are still lagging.

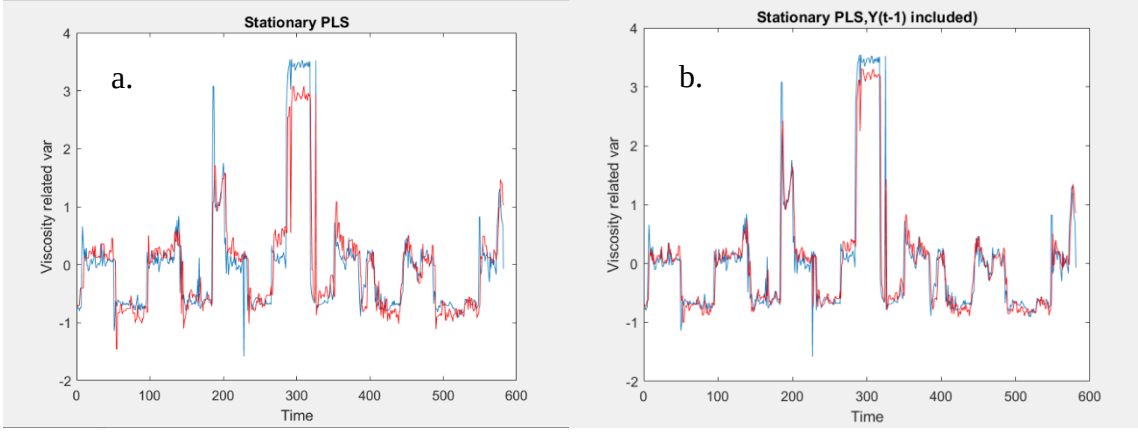


Figure 26. a) Single stationary PLS model used for modeling viscosity index for continuous diesel process. b) Same stationary model, with Y(t-1) included in X. Blue line represents the real process, red line represents the predictions.

In Figure 27 a dynamic model is presented. Like in Figure 26b, here also the last weighted measured reference values Y(t-1) is added to X variables. The dynamic updates the PLS model after predetermined number of steps. In Figure 27 multiple images are presented with different number of steps. The smaller the number of steps, more accurate the model is.

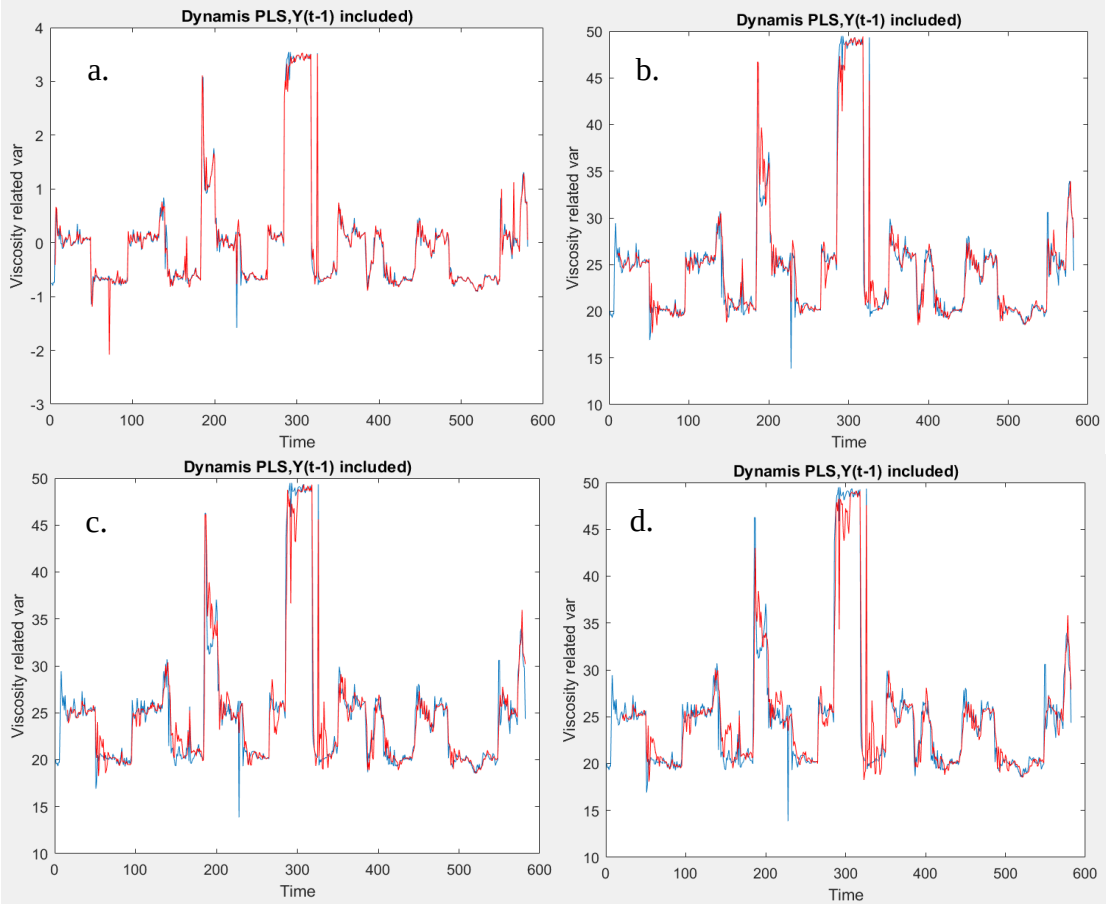


Figure 27. Dynamic model for continuous diesel production process, with different number of steps between PLS models. a) 5 steps b)10 steps c)15 steps and d) 20 steps. Smaller the number of steps in each PLS, the better the prediction.

4.4 Complexity of models

It is always possible to get a 100% fit with a PLS model, but to make sure the model is not overfitted the model needs to be validated. Good regression model must have a high predictive capability achieved with the minimum number of latent variables. This assures the model is not overfitting. The complexity of the model equals to the amount of PLS components. There are many methods for interpreting the model fit and considering the optimal number of latent variables used in the model. Some methods are mentioned here. However, any method does not compare to the usage of external data set for validation of the model. [Kohonen, 2009]

Inner validation with cross-validation is a standard procedure in PLS model validation. It is a leave-one-out type validation, where part of the calibration samples is left out from modeling and estimated based on the others. This is repeated until all samples are left out and predicted once and only once. Prediction errors of the samples left out are combined as a sum of squares (PRESS). The cross-validation is then repeated after introducing a new latent variable at a time. At each step, the PRESS value is calculated, and the optimal number of latent variables can be found from minimal PRESS. PRESS can be calculated with equation 4. Other ways to interpret the fit of the model is to calculate the root mean square error of cross-validation RMSECV or goodness of fit in cross validation Q^2 and fir of model R^2-value. In RMSECV the optimal amount of PLS components is also at the minimum value. RMSECV is calculated with equation 5. The R^2- value indicates good model when the value is close to 100%. The fit of model is sometimes deemed un-informational, as it can be greatly affected by a simple sample. [Kohonen, 2009]

$$PRESS = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \quad (4)$$

Where N number of samples

$\hat{y}$ predicted response values

y response values

$$RMSECV = \sqrt{\frac{PRESS}{N}} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \quad (6)$$

Where $\bar{y}$ mean value of response values

The fit of model needs to also be assessed visually. Kohonen (2009) gives a visual example of different types of results of modeling with R^2 values in his dissertation. The image is presented in Figure 28.

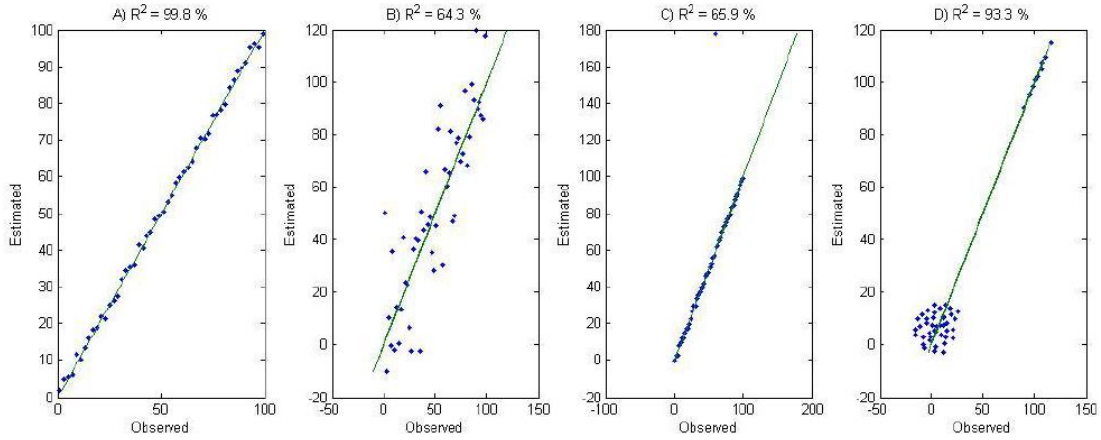


Figure 28. Different types of results from modeling showing R^2 values. Figures A and B represent the most common situations. A) near perfect fit B) model with lots of dispersion. C) R^2 worsened by a single sample. D) unsuccessful model with good R^2 value. [Kohonen, 2009]

EXPERIMENTAL PART

In this part of the thesis, some experiments are made in the laboratory with Thermo Fisher Antaris II NIR- analyzer. Two different measurement methods, reflectance, and transmittance were used for method comparison. For the transmittance method, samples in three different temperatures were measured. Calibration was made with a variety of CTO and PTO samples, and the calibrations were validated with an external validation set.

5. Analysis of CTO and PTO with NIR

5.1.1 Variable and sample selection

In modeling, there is always a change that variation is modeled instead of the real phenomena. Unnecessary data can lead to erroneous results or weaken modeling power. Uncorrelated variables or faulty samples should be removed. With spectral data, together with improvement in fit, initial variable selection can be done with *a priori* knowledge since there are listed tables of spectral regions and corresponding response available. [Kohonen, 2009] One such table is presented in Figure 29. These kinds of figures can usually work as guidelines, not exact rules, as peaks can shift depending on substance and its temperature. [Bono, 2014]

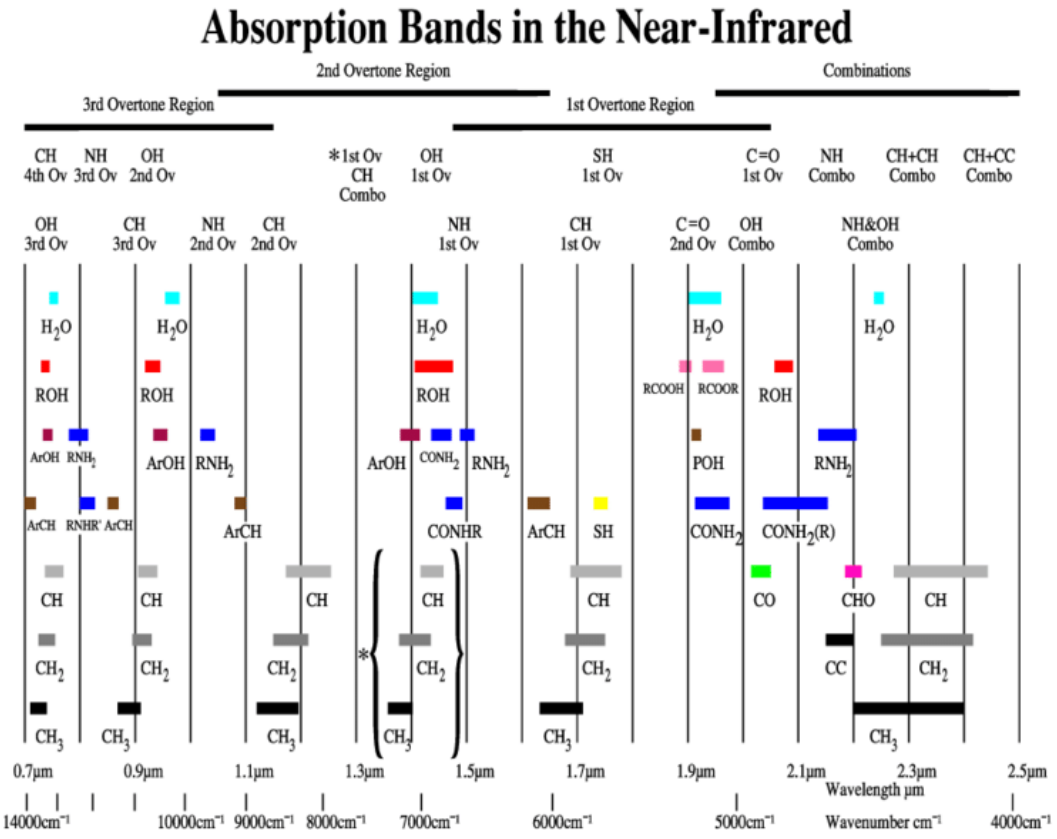


Figure 29. Absorption bands in the near-infrared [Bono, 2014]

In the calibration, spectral ranges were chosen for each component variable according to steps mentioned above. Calibration samples were selected to best represent the normal variability of the concentration values of the variables.

5.1.2 Spectra derivatives and smoothing

In spectroscopy, 1st, or higher order derivates of absorbance with respect to wavelength can be used to improve accuracy of quantification in the presence of interference caused by a broad absorbing component, matrix or scattering. In this thesis zero, first and second order derivatives are used, spectra for these derivatives are illustrated in Figure 30 by Owen (1995) for a Gaussian band. The 1st order derivative starts and finishes at zero, and it also passes through zero point at the maximum point of the zero-order absorbance band. The 2nd order derivative is a negative band with a minimum point at the maximum point of the original zero-order band. It also has two positive satellite bands on both sides of the main band. The first order derivative removes the spectral baseline, and the second

order derivative also removes the linear trend. An undesired effect of derivatization of spectra is that the signal-to noise ratio decreases when higher orders of derivates are used. This can be reduced by using smoothing filters, such as Savitzky- Golay filter or Norris derivate filter. [Owen, 1995] [Rinnan et all., 2009]

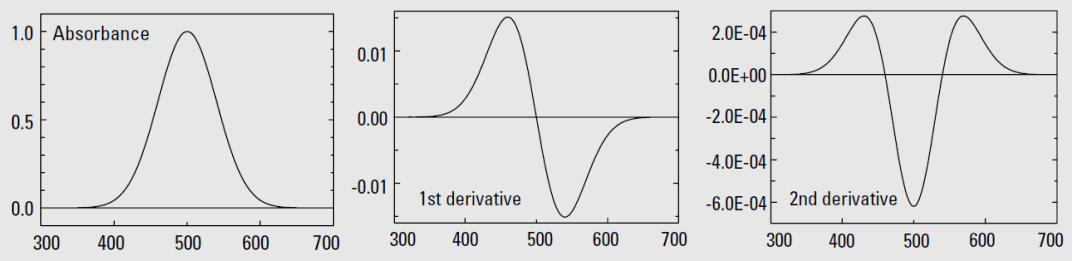


Figure 30. Zero order derivative, 1st derivative and 2nd derivative spectra [Owen, 1995]

In the Savitzky-Golay filter a polynomial is fitted in a symmetric window on the raw data. After calculating the parameters for this polynomial, a derivative estimate for the center point can be found analytically for derivative of wanted order (Figure 31a). This operation is then applied to all spectra points sequentially. The window size (number of points used to calculate the polynomial) and degree of the polynomial need to be decided. The order of the derivative can't be higher than the degree of the polynomial. For example, third degree polynomial can be used up to third-order derivative. Too high degree of smoothing can also distort the derivative spectrum. [Rinnan et all., 2009] [Owen,1995]

Norris derivative filter is developed to reduce the decreasing the signal-to-noise ratio. In smoothing an average of given number of points (smoothing window centered around the measurement point) is calculated. For the 1st order derivation, the difference between two smoothed values with a given gap size between them is taken, and for 2nd order derivative, the smoothed value at the current measurement point is taken twice and a smoothed value at gap distance on either side is taken. An illustration of the Norris derivative filter is given for the 1st order derivative in Figure 31b. [Rinnan et all., 2009]

.

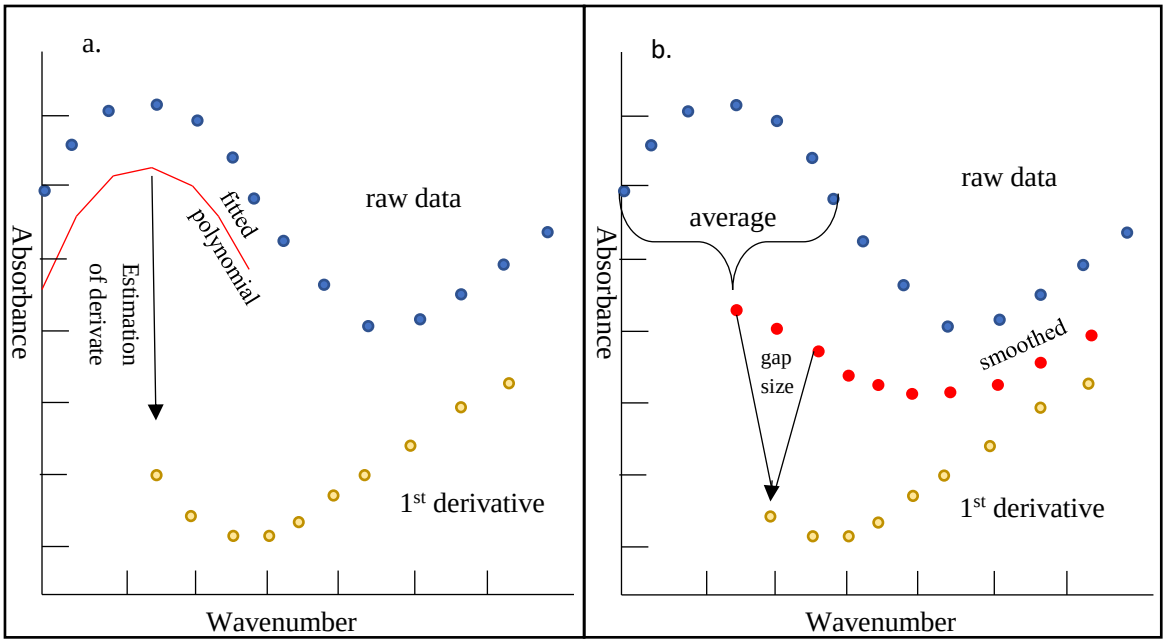


Figure 31. a.) Principle of Savitzky-Golay filer with a seven-point smoothing window and a second-degree polynomial. b.) Principle of Norris derivative filter with a 7-point smoothing window and a gap size of three. Modified from [Rinnan et all., 2009]

5.2 Calibration of CTO and PTO with reflectance method

Different calibration models were made for CTO and PTO, as CTOs composition changes as it transitions to PTO and there are noticeable differences in the spectra. This is shown in Figure 32. There is a lot of noise in the original spectra, especially in the PTO spectra, this was fixed by using more smoothing.

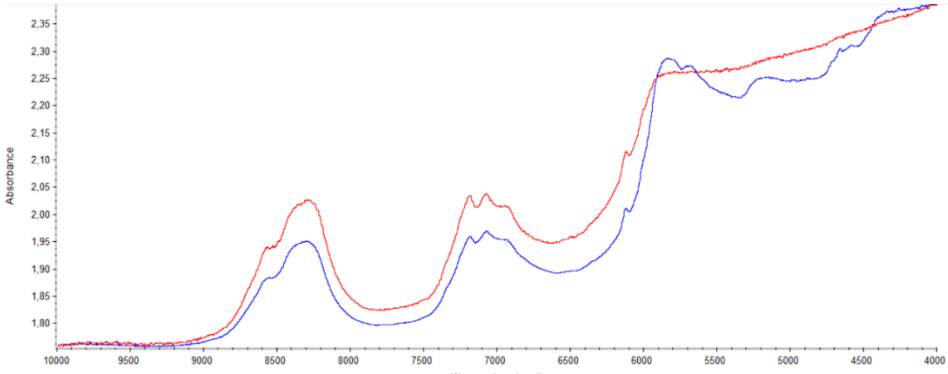


Figure 32. Spectra of PTO (red) and CTO (blue)

The calibration models were checked with PRESS chart and RMSEC value and validated with an external validation set. The optimal RMSEC value is below 1 and the optimal shape of the PRESS curve is descending curve that levels out in the end. Measurement errors for the laboratory measurements are shown in Table VII, these values need to be taken into consideration when assessing the validation results. The calibration range for PTO and CTO measurements are also presented in Table VII. Validation results are only valid within the validation range. For neutral substances and fatty acids, the measurement error is assessed to be the similar to resin acids. As acid number and resin acids are analyzed in the laboratory and fatty acids and neutral substances are calculated from their subtraction. The measurement error is especially significant in the residual soap and water content measurements.

Table VII. Measurement error for the reference methods and calibration ranges for CTO and PTO calibrations for the NIR reflectance method.

component	CTO calibration range Min-Max	PTO calibration range Min-Max	Measurement error of the reference method with 95% confidence
Resin acid	9.50–37.80	24.20–29.30	±1.8 %
Acid number	110.90–158.70	145.80–154.90	±1.03 %
Neutral substances	26.00–44.00	19.90–25.10	~±1.8 %
Fatty acid	42.80–56.00	49.50–53.20	~±1.8 %
Residual soap	0.00–0.62	-	± 21.1 %
Water content	0.25–5.07	-	± 30.52 %

5.2.1 CTO calibration

Calibration was done for all components individually using same set of calibration samples, but different settings based on the component. In Table VIII. settings and wavenumber range values used to calibrate each component are showed.

Table VIII. settings used in CTO calibration for different components

component	**Wavenumber range, cm^{-1}**	**derivative**	**smoothing**
Acid number	8916-6023	2nd	No smoothing
Resin acid	8852-6055	1st	Savitzky-Golay Data points: 7 Polynomial order 3
Neutral substances	8916-6600	2nd	Savitzky-Golay Data points: 7 Polynomial order 3
Fatty acid	7793-4995	1st	No smoothing
Soap number	8897-4995	2nd	Savitzky-Golay Data points: 7 Polynomial order 3
water	5455-4754 7689-6298	2nd	Norris derivative filter, Segment length 5, gap between segments 1

The calibration was checked to be good based on the RMSEC value, the difference chart, and the PRESS chart. Outliers were detected based on the difference chart and removed. In Figure 33. on the left side the calibration charts for acid number is shown before any outliers were removed. Total number of samples in the calibration was 86, these same samples was used for all components. On the right-side same calibration is shown, but with 7 outliers removed. The ideal PRESS curve is descending curve that levels down in the end. The number of PCs can be determined based on the PRESS value, either when the value is at its minimum, or the decrease in the value is no longer statistically significant. In some cases, it is better to choose the amount of PC's in the model based on the latter rule, to avoid possible overfitting of the model. The optimal difference value was told to be equal or less than 10% of the value range of the samples. In the case of acid number, the values ranged from 110,90 mg/KOH/g to 158,7 mg/KOH/g. There is a

difference of 47,8 mg/KOH/g between these values, and 10% from this value is 4,78. The difference can be either negative or positive. All calibration models were made with same principal. Charts for all components are shown in Appendix I.

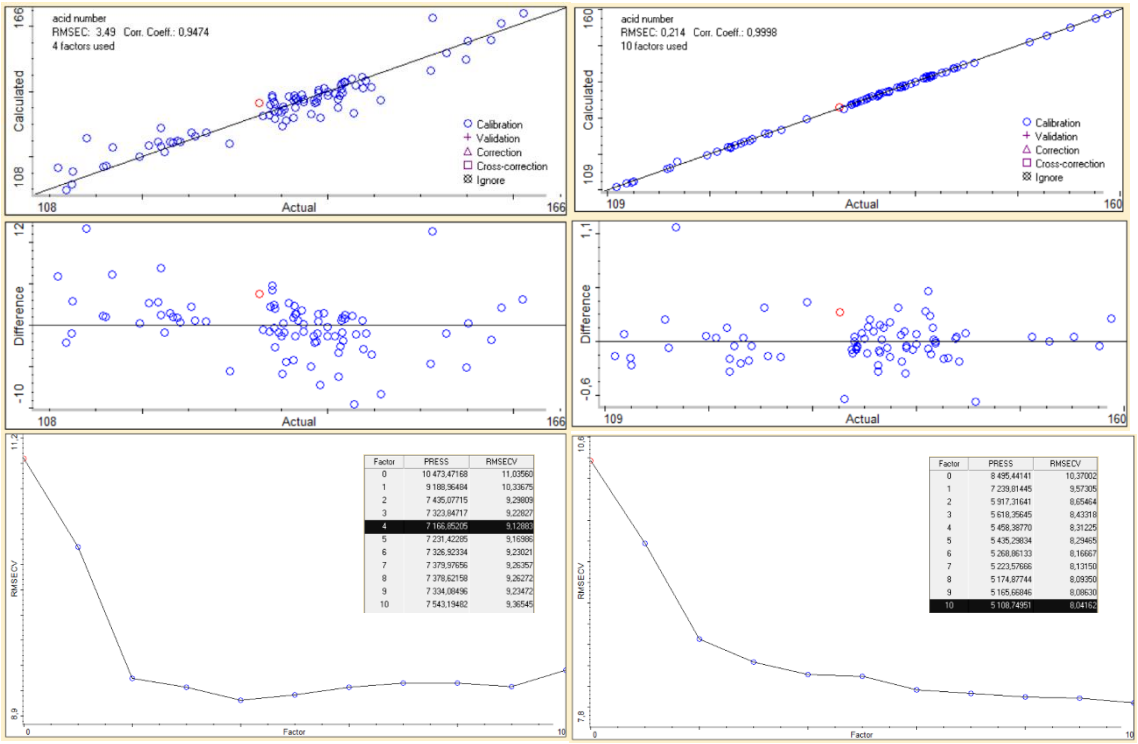


Figure 33. Calibration charts for reflectance method CTO acid number before removing outliers on the left side, and after removing outliers on the right side.

5.2.2 Calibration of metal and mineral salts

Calibration of impurities such as metal and mineral compounds was attempted but it proved to be unsuccessful. There is little certain knowledge of the bonds and substances these minerals are connected to. They are mostly expected to be connected to a carbonyl group -COOX, but the concentrations are believed to be too small to show in the NIR spectra.

For most of the minerals it was possible to create calibration models with good correlation coefficient, RMSEC and PRESS values, but the validation showed the calibration was overfitted and false. In Figure 34. calibration charts for silicon and potassium are presented. There are no reference method measurement error values available for metals. In Table IX a few validation results are shown. Regardless of efforts, the models could

not be bettered. After the modeling failed for reflectance method CTO samples, calibration of metals and mineral was not tried on PTO samples, as they have even lower concentrations of metals. Metals were not also done with the transmittance method.

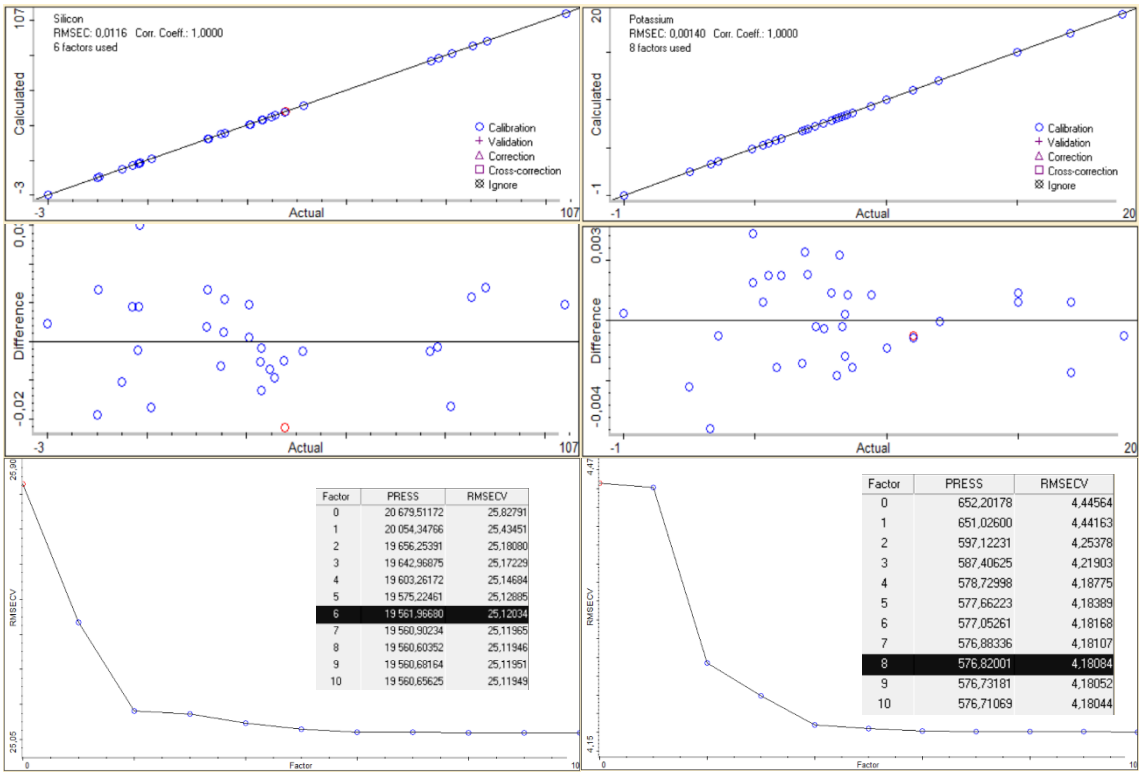


Figure 34. Calibration charts for silicon on the left side, and for potassium on the right side.

Table IX. Validation results for reflectance method CTO silicon and potassium samples

Reflectance CTO validation silicon, mg/kg			
Sample ID	Reference value	NIR measurement	Difference %
20-11114	16.6	45.58	-174.58
20-11540	4.76	23.41	-391.81
20-11545	42.3	24.15	42.91
20-11803	17.8	32.2	-80.90
20-11816	33.4	24.57	26.44
Reflectance CTO validation potassium, mg/kg			
Sample ID	Reference value	NIR measurement	Difference %
20-11114	6.4	11.15	-74.22
20-11540	13	7.68	40.92
20-11545	13	9.31	28.39
20-11803	3.4	12.02	-253.53

20-11816	20	9.97	50.15

The difference values are similar to residual soap number results. For the same reasons as the metals, the calibration of residual soap was also mostly unsuccessful. The chemical structure of soap is combined of a long carbon chain and a carbonyl group that has a metal (usually natrium) attached to it instead of a hydrogen. As with metals, the amount of residual soap is too small to be detected behind other resin and fatty acids that make up most of the CTO. The residual soap calibration was still done for transmittance and reflectance method.

5.2.3 PTO calibration

PTO calibrations were made with same principals as CTO calibrations. Charts for all calibrations are in Appendix I. The settings used in PTO calibrations are shown in Table X. The number of samples used for the calibration before removing any outliers was 16. The minimum number of samples used in calibration for one component is 11.

Table X. values used in PTO calibration for different components

Component	**Wavenumber range, cm^{-1}**	**Derivative**	**Smoothing**
Resin acid	8763-6055	2^{nd}	Savitzky-Golay Datapoints: 15 polynomial order: 3
Acid number	8916-6023	2^{nd}	Savitzky-Golay Datapoints: 15 polynomial order: 3
Fatty acid	7463-5465	2^{nd}	Savitzky-Golay Datapoints: 15 polynomial order: 3
Neutral substances	8802-6753	2^{nd}	Savitzky-Golay Datapoints: 15 polynomial order: 3

5.3 CTO and PTO validation for reflection method

All validations are done with an external validation set. The samples chosen are similar to the ones used in calibration and inside the calibration range. The validation samples are also analyzed with the laboratory reference methods, and the NIR results are compared to the laboratory results to see whether the calibration was successful. For reliable results, the number of validation samples should be and is at least ¼ of the number of calibration samples.

5.3.1 CTO validation

Validation for the CTO and PTO calibration was done with external validation set. The results of CTO validation for each component are presented in Figure 35. Here the blue line represents the line where the samples should place based on the laboratory measurement results. The yellow blocks represent the validation samples measured with NIR. From the charts it is clear that some validations have a lot of errors. Some samples are outliers with many components, and some are outliers in one and fine in others, no sample was under the measurement error limit with all components.

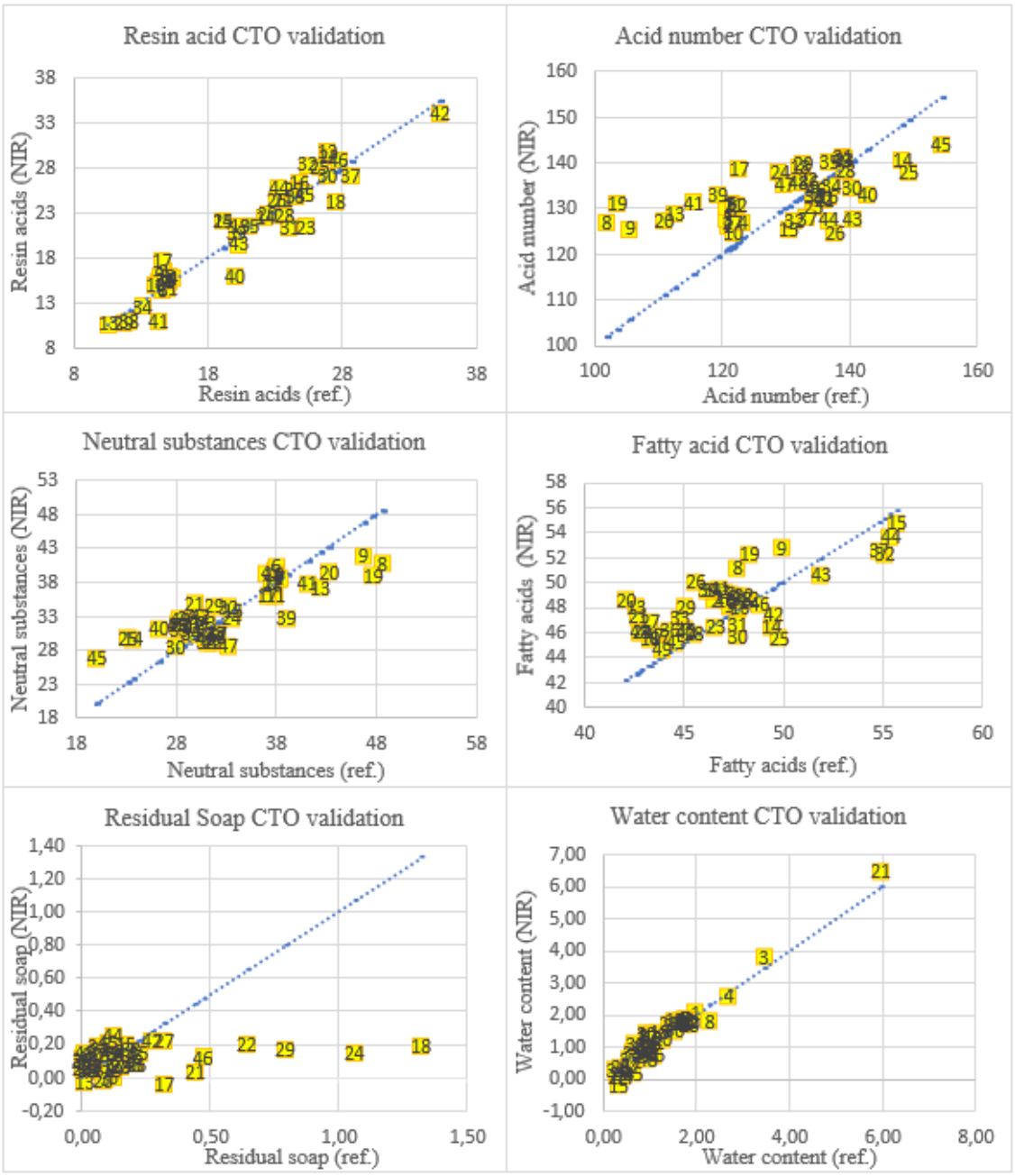


Figure 35. CTO validations for different components. Blue line represents the line where the samples should place based on laboratory results, yellow blocks represent the samples measured with NIR.

In Table XI some of the validation result values are shown, to better showcase the results in Figure 35. Results from all samples are in Appendix II. Positive note was made with resin acids, as the NIR measurement values were, despite some difference, in similar range to the reference value, as when the reference value was small the NIR value was

small and so on. In resin acids the calibration range was the largest and could include such variation.

Table XI. CTO validation results. Results from all samples are in Appendix II.

CTO validation resin acid, W-%				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
8	20-03335	3.7	4.92	-32.97
18	20-03931	24.8	26.41	-6.49
23	20-04213	22.4	22.78	-1.70
45	20-11814	35.30	34.03	3.60
CTO validation acid number, mg/KOH/g				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
8	20-03335	101.9	126.97	-24.60
18	20-03931	132.2	139.14	-5.25
23	20-04213	134.3	130.28	2.99
45	20-11814	154.30	144.19	6.55
CTO validation neutral substances, W-%				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
8	20-03335	48.6	40.81	16.03
18	20-03931	31.9	30.55	4.23
23	20-04213	31.0	31.91	-2.94
45	20-11814	20.10	26.84	-33.53
CTO validation fatty acids, W-%				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
8	20-03335	47.7	51.20	-7.34
18	20-03931	43.3	45.45	-4.97
23	20-04213	46.6	46.43	0.37
45	20-11814	44.60	45.16	-1.26
CTO validation residual soap				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
8	20-03335	0.14	0.16	-14.29
18	20-03931	1.32	0.19	85.61
23	20-04213	0.10	0.14	-40.00
45	20-11814	0.10	0.21	-110.00
CTO validation water content, W-%				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
8	20-03335	2.29	1.77	22.71
18	20-03931	1.21	1.45	-19.84

23	20-04213	1.41	1.71	-21.28
45	20-11814	0.65	0.16	75.39

In Table XII. percentages of reflectance CTO samples with different difference percentages. The samples were checked with a Matlab PCA model to find any abnormal samples that could explain some of the results. The T^2 and SPEx charts for the baseline corrected spectra are shown in Figure 36. Based on these charts, sample 21 should be automatically removed. Sample 15 and 33 should be checked. These samples do not solve all the problems in the validation results as there are many more samples with problematic results. This could mean the calibration model is faulty and can be overfitted.

Table XII. Percentages of reflectance CTO samples with different difference percentage. Reference method measurement error for resin acid, neutral substances and fatty acids is 1,8%. For acid number the value is 1,03%, for residual soap 21,1% and for water content the value is 30,52%.

Difference %	**resin acid, %**	**Acid number, %**	**Neutral substances, %**	**Fatty acids, %**	**Difference %**	**residual soap, %**	**Difference %**	**Water, %**
≤ Reference method measurement error	18.37	4.08	12.24	22.45	**≤ 21,1%**	17.39	**≤ 30,52%**	71.43
≤ 5 %	48.98	48.98	38.78	61.22	**≤ 25 %**	17.39	**≤ 35 %**	75.51
≤ 10 %	71.43	83.67	69.39	93.88	**≤ 30 %**	17.39	**≤ 40 %**	83.67

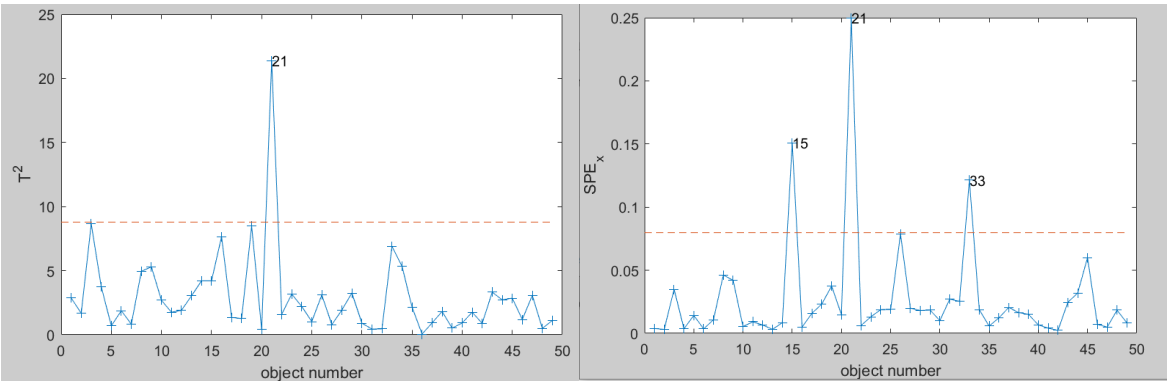


Figure 36. T2 and SPEx charts for CTO validation samples. Two PCs was used. The 95% SPEx confidence limit is 0.0797207 and the 95 T^2 confidence limit is 8.79029.

5.3.2 PTO validation

Like the CTO validation results the validation results for PTO reflectance method did have some errors in them, which is probably mostly due to overfitting. With reflectance method, the spectra of PTO were noisy (Figure 37) and they required a lot of smoothing, this may be the reason for overfitting. The results of validation are presented in Table XIII and Figure 38. In table XIII only some of the validation results are shown. All results are shown in appendix II.

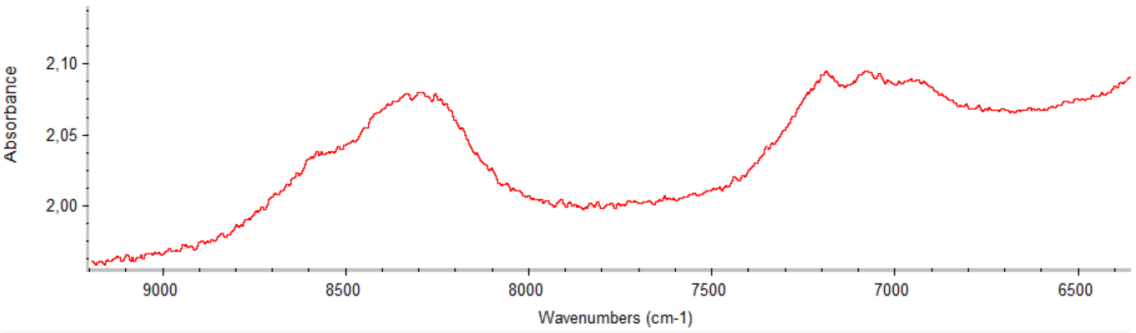


Figure 37. Noisy PTO spectra. Note that the image does not display the whole spectra.

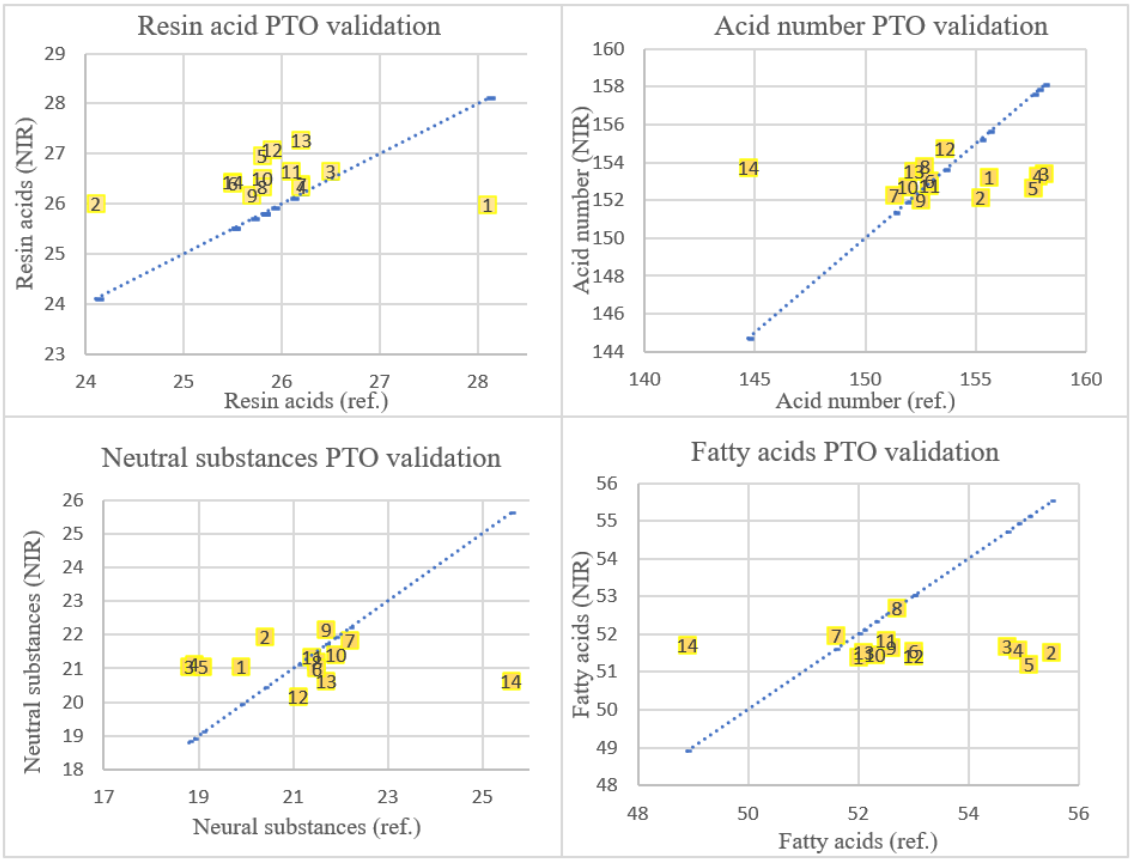


Figure 38. Reflectance method validation results for PTO samples for different components. Blue dotted line represents the real laboratory values and yellow blocks represent the validation results of same samples.

Table XIII. Values for PTO validation samples for reflectance method. Unlike in CTO validation, the NIR measurement gives similar values to all samples, even though there is more variation in the reference values. Values for all 14 PTO validation samples are presented in appendix II.

PTO validation resin acid, W-%				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
1	20-10954	28.1	25.97	7.58
2	20-12099	24.1	26.01	-7.93
3	20-12068	26.5	26.66	-0.60
4	20-11981	26.2	26.33	-0.50
14	20-11165	25.5	26.44	-3.69
PTO validation acid number, mg/KOH/g				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
1	20-10954	155.6	153.24	1.52

2	20-12099	155.2	152.17	1.95
3	20-12068	158.1	153.43	2.95
4	20-11981	157.8	153.30	2.85
14	20-11165	144.7	153.75	-6.25
PTO validation neutral substances, W-%				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
1	20-10954	19.9	21.06	-5.83
2	20-12099	20.4	21.95	-7.60
3	20-12068	18.8	21.07	-12.07
4	20-11981	18.9	21.12	-11.75
14	20-11165	25.6	20.64	19.38
PTO validation fatty acid, W-%				
Sample number	Sample ID	Reference value	NIR measurement	Difference %
1	20-10954	52	51.41	1.14
2	20-12099	55.5	51.52	7.17
3	20-12068	54.7	51.68	5.52
4	20-11981	54.9	51.6	6.01
14	20-11165	48.9	51.7	-5.73

As it can be seen from Figure 38 and table XIII there are some samples have very large difference between the reference value measured in the laboratory and the value measured with the NIR. In Table XIV, the sample percentages at different difference percentages is shown. Measurement errors for reference methods were presented in Table VII Validation spectra were checked with Matlab to find out if any samples were pointed out as outliers by MSPC charts. T2 and SPEx charts for reflection PTO validation samples are presented in Figure 39. If both charts show same samples outside the limits, the results of PLS model should not be believed to be correct. If only one chart shows samples outside its limits, the sample should be checked but not automatically deleted. In process singular outlier samples are not alarming, but if several consecutive samples go over the 95% confidence limits, there is a problem in the process that needs to be fixed or the model needs to be updated. Outliers found in the Figure 39 MSPC charts do not explain most of the samples that have larger difference percentage, so there is reason to believe that the calibration models are overfitted.

Table XV. Percentage of samples with different difference percentages. The reference method measurement error for resin acid, neutral substances and fatty acids are 1,8% and for acid number the value is 1,03%.

Difference %	Resin acid, %	Acid number, %	Neutral substances, %	Fatty acids, %
≤ Reference method measurement error	21.4	57.1	14.3	42.9
≤ 5 %	85.7	92.9	57.1	64.3
≤ 10 %	100	100	71.4	100

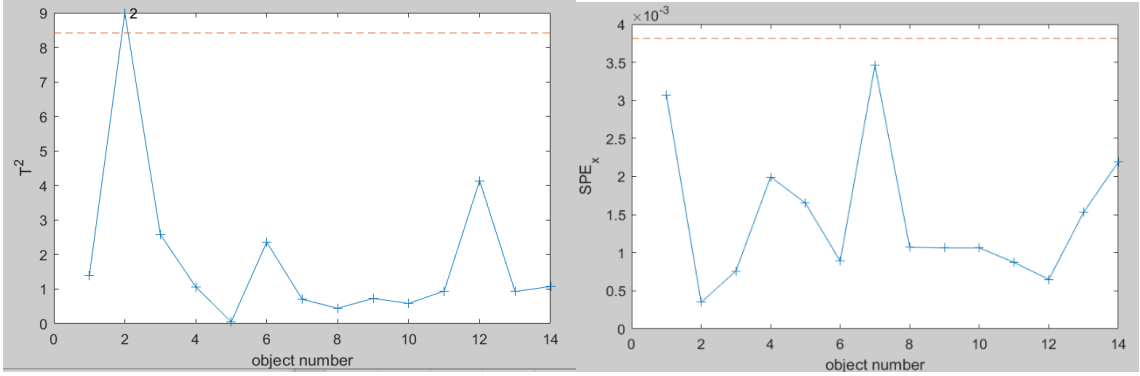


Figure 39. MSPC charts with 2 PC's used for reflectance PTO validation samples.

6. Analysis of CTO and PTO with NIR transmittance method

With the transmittance method, the samples were measured with 1ml cuvette and in three different temperatures, to see whether the temperature affects the results. The reference measurements are done in room temperature in the laboratory and in the process the analysis would be done with hot sample with the online analyzer. Individual calibration models were done to all three temperatures (28℃, 40℃ and 60℃). With transmittance method, daily PTO and CTO samples were used in analysis. The composition variation within these samples is much smaller compared to samples used with the reflectance method. These samples were selected, and fewer samples were analyzed than in reflectance method because of the limited number of single use cuvettes available. In this chapter the effect of temperature is examined, and transmittance and reflectance method results are compared to each other.

6.1 Calibration of CTO and PTO with transmittance method

In Figure 40. The spectra of CTO and PTO is presented. With transmittance, the CTO and PTO spectra are similar to each other and neither are especially noisy. There are large peaks in the wavenumber range 6000-4000. Peaks with high absorbance over 3 were not used in calibration. Before removing outliers, 15 samples were used in CTO calibration and 17 samples were used in PTO calibration. Other than that, the calibration was done similarly to the reflectance calibrations.

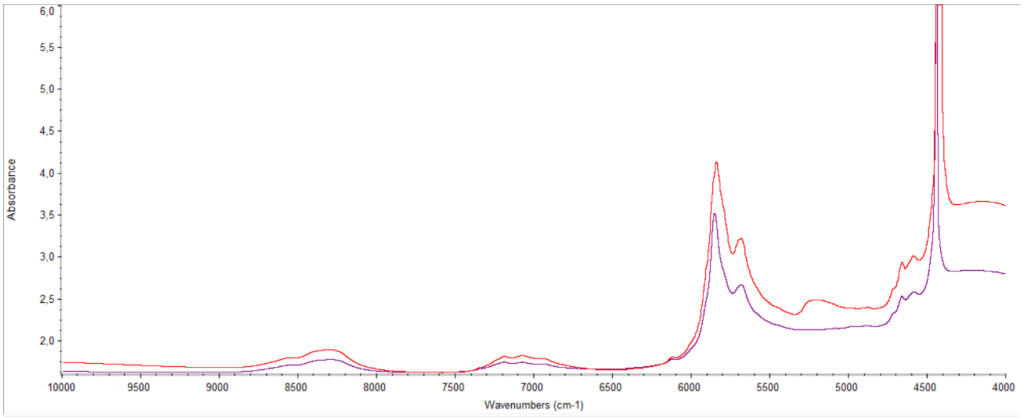


Figure 40. CTO and PTO spectra measured with transmittance method in 28℃. Red spectra is CTO sample; blue is the PTO sample.

In Table XVI calibration range and other calibration settings are shown for CTO. Settings for PTO are shown in Table XVII. From the Tables it can be seen that the settings were different for different temperatures when calibrating one component. This is not ideal. If the models are trustworthy as they are now, this could mean the models for the actual process need to be tested and made in the temperature they are going to be tested in the process, in the case the models are premade in the laboratory before implementing them to the online analyzer.

Table XVI. Transmittance PTO calibration settings.

Component	Temperature	Calibration range CTO	Wavenumber range CTO	derivative	smoothing
Acid number	28℃	137,8-142,9	8704-6037	2nd	Norris Derivative filter Segment length: 5 Gap between segments: 1
	40℃	137,8-142,9	8852-5997	2nd	Savitzky-Golay Datapoints: 7 Polynomial order: 3

	60℃	137,8-142,9	8852-5997	2nd	Savitzky-Golay Datapoints: 7 Polynomial order: 3
Resin acid	28℃	23,1-28,6	8780-6607	2nd	Norris Derivative filter Segment length: 5 Gap between segments: 1
	40℃	23,1-26,6	8780-6011	1st	Savitzky-Golay Datapoints: 7 Polynomial order: 2
	60℃	23,1-26,8	8780-6011	1st	Savitzky-Golay Datapoints: 7 Polynomial order: 3
Neutral substances	28℃	26,4-29,7	8812-5999 5754-4516	2nd	Savitzky-Golay Datapoints: 7 Polynomial order: 2
	40℃	26,4-29,1	8812-5999 5754-4516	2nd	Savitzky-Golay Datapoints: 7 Polynomial order: 2
	60℃	26,4-29,0	8812-5999 5754-4516	1st	Savitzky-Golay Datapoints: 7 Polynomial order: 3
Fatty acids	28℃	46,1-48,2	8812-6081	2nd	Savitzky-Golay Datapoints: 9 Polynomial order: 2
	40℃	45,9-48,1	8812-6594	1st	Savitzky-Golay Datapoints: 7 Polynomial order: 2
	60℃	46,1-48,1	8812-6024	1st	Savitzky-Golay Datapoints: 7 Polynomial order: 3
Residual soap	28℃	0,07-0,16	8806-6011	2nd	Savitzky-Golay Datapoints: 7 Polynomial order: 2
	40℃	0,07-0,16	8806-6011	2nd	Savitzky-Golay Datapoints: 7 Polynomial order: 2
	60℃	0,07-0,16	8806-6011	2nd	Savitzky-Golay Datapoints: 7 Polynomial order: 2
Water content	28℃	0,66-0,86	7507-6582 5442-4789	2nd	Savitzky-Golay Datapoints: 7 Polynomial order: 2
	40℃	0,53-0,88	7507-6582 5442-4789	1st	Savitzky-Golay Datapoints: 5 Polynomial order: 2
	60℃	0,66-0,88	7507-6582 5442-4789	1st	Savitzky-Golay Datapoints: 5 Polynomial order: 3

Table XVII. Transmittance PTO calibration settings.

Component	Temperature	Calibration range PTO	Wavenumber range PTO	derivative	smoothing
Acid number	28℃	151,3-156,0	8704-6037	1st	Savitzky-Golay Datapoints: 5 Polynomial order: 2

	40	149,9-156,0	8704-6037	1st	Savitzky-Golay Datapoints: 7 Polynomial order: 3
	60	149,9-156,0	8704-6037	1st	Savitzky-Golay Datapoints: 11 Polynomial order: 2
Resin acid	28	25,2-29,3	8780-6607	2nd	Norris Derivative filter Segment length: 5 Gap between segments: 1
	40℃	25,2-29,3	8780-6607	2nd	Norris Derivative filter Segment length: 5 Gap between segments: 1
	60	25,2-29,3	8780-6607	2nd	Norris Derivative filter Segment length: 5 Gap between segments: 1
Neutral substances	28	19,7-22,2	8812-5999 5754-1516	1st	Savitzky-Golay Datapoints: 11 Polynomial order: 2
	40	19,7-23,0	8812-5999 5754-1516	2nd	Savitzky-Golay Datapoints: 9 Polynomial order: 3
	60	19,7-21-7	8812-5999 5754-1516	1st	Savitzky-Golay Datapoints: 9 Polynomial order: 2
Fatty acids	28	49,5-53,8	8812-6081	2nd	Norris Derivative filter Segment length: 5 Gap between segments: 1
	40	49,5-53,1	8812-6081	1st	Norris Derivative filter Segment length: 5 Gap between segments: 1
	60	49,5-53,1	8812-5999	2nd	Norris Derivative filter Segment length: 5 Gap between segments: 2

6.2 Validation of CTO and PTO with transmittance method

Validation charts for CTO are presented in Figure 41. and for PTO in Figure 42. All temperatures are presented in same chart for all components. Specific validation result values are in Appendix II. In the CTO acid number validation, there was a calibration model or analyzer error, that resulted only four validation samples to be analyzed. The samples could be measured for all other components. More samples could not be analyzed to replace these samples, as there was limited amount of single use cuvettes available.

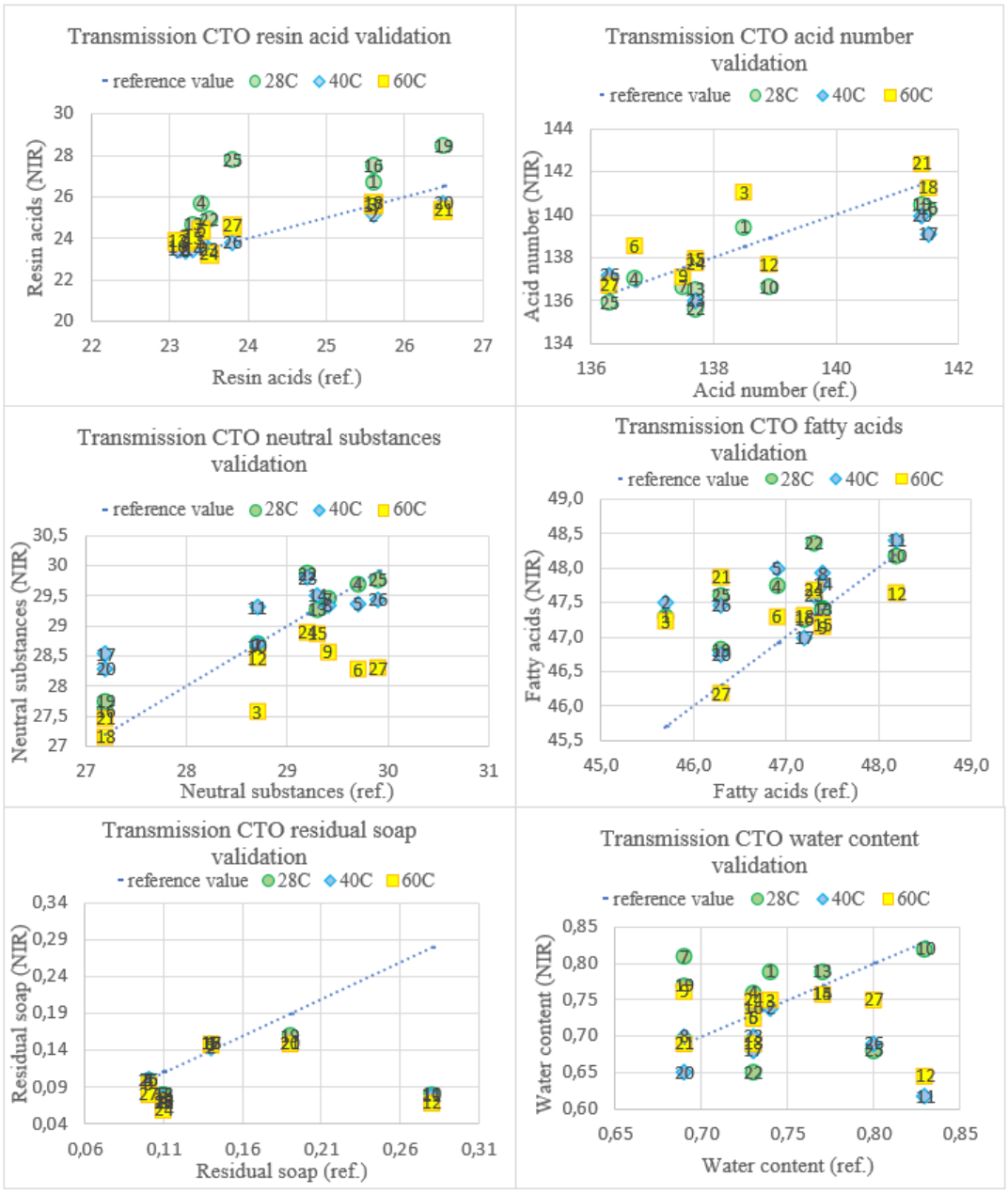


Figure 41. Transmittance CTO validation charts

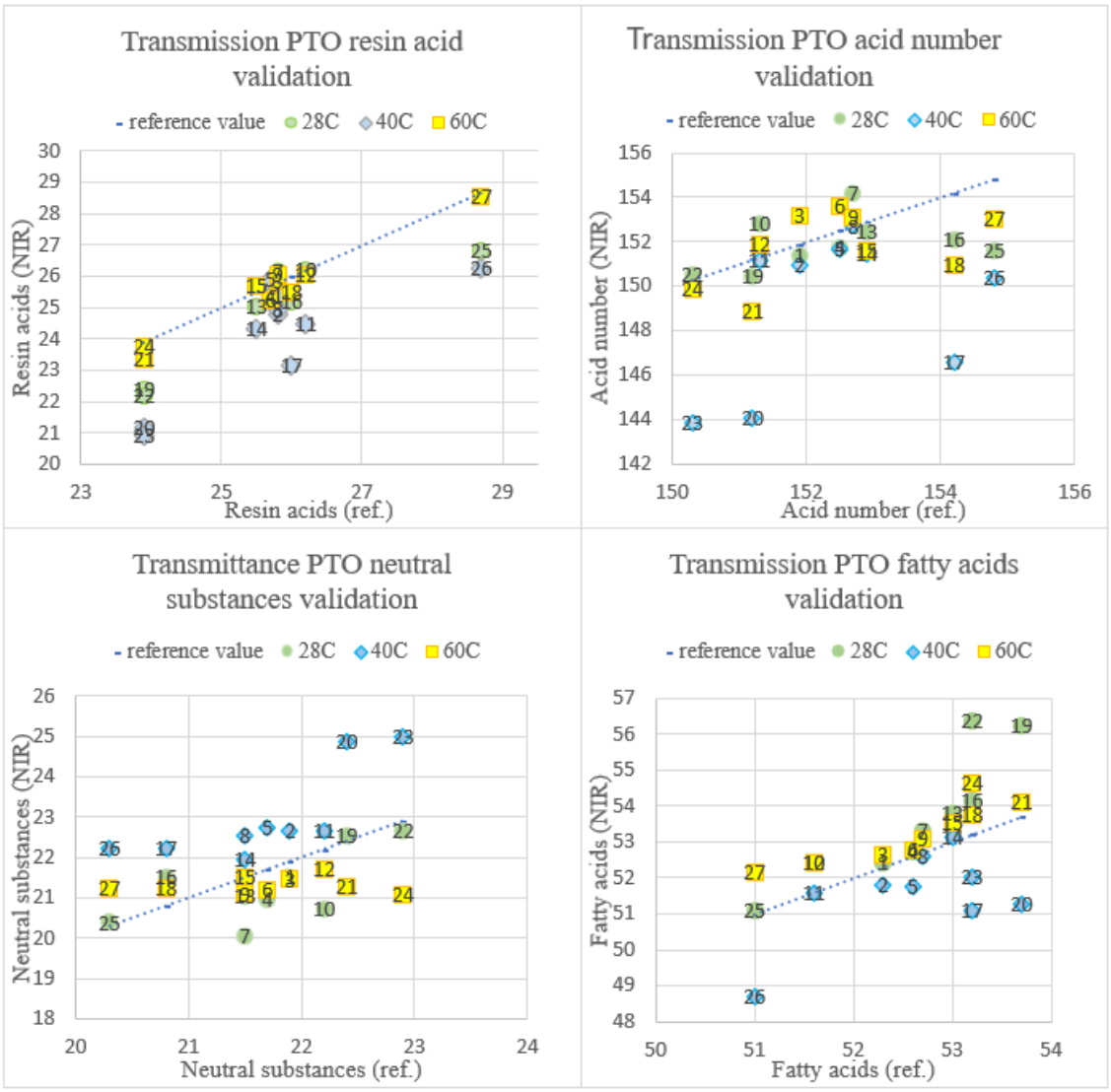


Figure 42. Transmittance PTO validation charts. In resin acid validation at 40C only four calibration samples were analyzed due to analysis errors.

Percentages of CTO and PTO samples at different temperatures with different difference percentages are presented in Table XVIII for CTO and Table XIX for PTO. Based on these tables there seems to not be any significant difference between the temperatures. Most likely the differences here from the calibration model errors, and not due to temperature.

Table XVIII. Percentages of transmittance CTO validation samples with different difference percentage. With acid number at 40C, due to analyzer or model error only 4 validation samples could be analyzed.

Difference %	**Resin acids, %**			**Acid number, %**			**Neutral substances, %**			**Fatty acids,%**		
temperature	**28**	**40**	**60**	**28**	**40**	**60**	**28**	**40**	**60**	**28**	**40**	**60**
≤ Reference method measurement error	0.0	66.7	33.3	77.8	50.0	77.8	77.8	55.6	55.6	66.7	66.7	77.8
≤ 5 %	44.4	100	100	100	100	100	100	100	100	100	100	100
≤ 10 %	88.9	100	100	100	100	100	100	100	100	100	100	100
Difference %	**Residual soap**			**Difference %**			**Water content, %**					
temperature	**28**	**40**	**60**	**temperature**			**28**	**40**	**60**			
≤ Reference method measurement error	44.4	55.6	55.6	**≤ Reference method measurement error**			100	100	100			
≤ 25 %	44.4	55.6	55.6	**≤ 35 %**			100	100	100			
≤ 30 %	77.8	55.6	55.6	**≤ 40 %**			100	100	100			

Table XIX. Percentage of transmittance PTO validation samples with different difference percentage.

Difference %	**Resin acids, %**			**Acid number, %**			**Neutral substances, %**			**Fatty acids, %**		
temperature	**28**	**40**	**60**	**28**	**40**	**60**	**28**	**40**	**60**	**28**	**40**	**60**
≤ Reference method measurement error	55.5	11.1	77.8	77.8	55.6	66.7	44.4	0	11.1	66.7	55.6	77.8
≤ 5 %	66.7	44.4	100	100	100	100	77.8	88.9	88.9	88.9	100	100
≤ 10 %	100	66.7	100	100	100	100	100	88.9	100	100	100	100

The validation samples were checked with PCA model in Matlab to find abnormal samples. In Figure 43. MSPC charts for CTO validation are presented. In the top row all calibration and validation CTO samples are presented. From the samples that go over the limit only sample 54 is validation sample. The bottom row are only the validation samples are drawn.

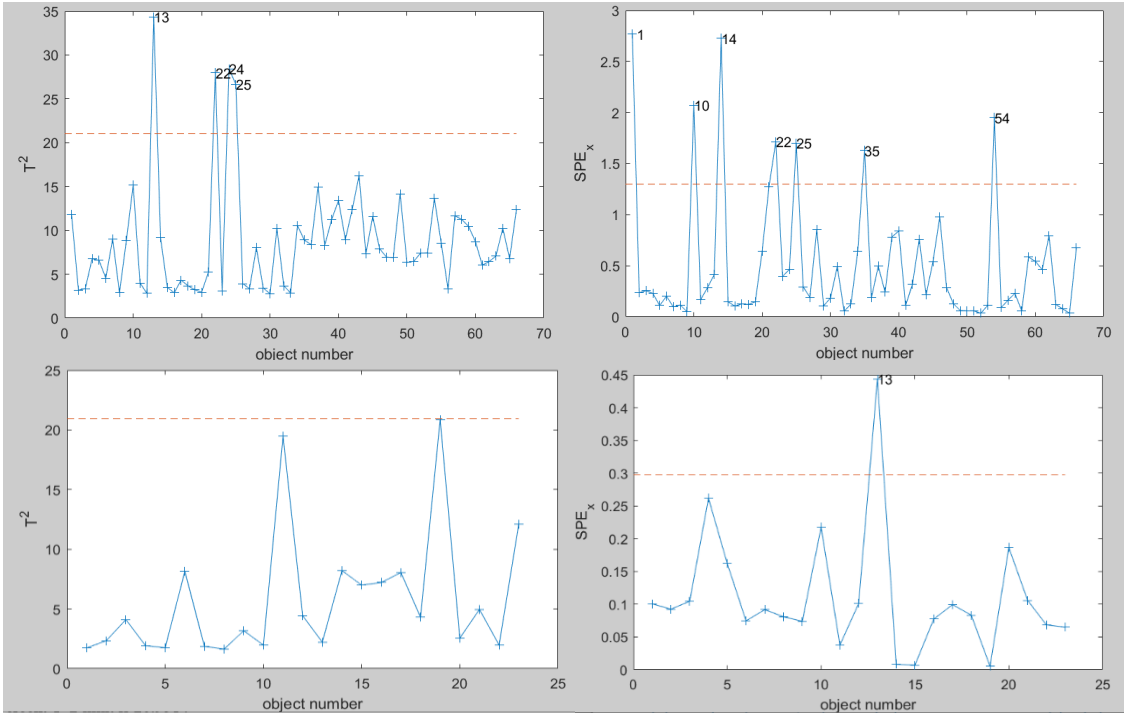


Figure 43. MSPC charts for transmittance CTO validation. In top row are all calibration and validation samples and in bottom row only the validation samples are checked. For the top row the 95% SPEx confidence limit is 1.29851 and the 95% T^2 confidence limit is 21.0226. For the bottom row the values are 0.297754 and 20.9514 respectively.

In Figure 44. MSPC charts for PTO validation samples are presented. In top row all calibration and validation PTO samples are presented. Here samples after 48 are validation samples, so only samples 55,58 and 67 are abnormal validation samples. Then all validation samples at different temperatures were checked and are presented in the bottom row. Here sample 7 is the same as sample 55 and sample 19 is the same as sample 67.

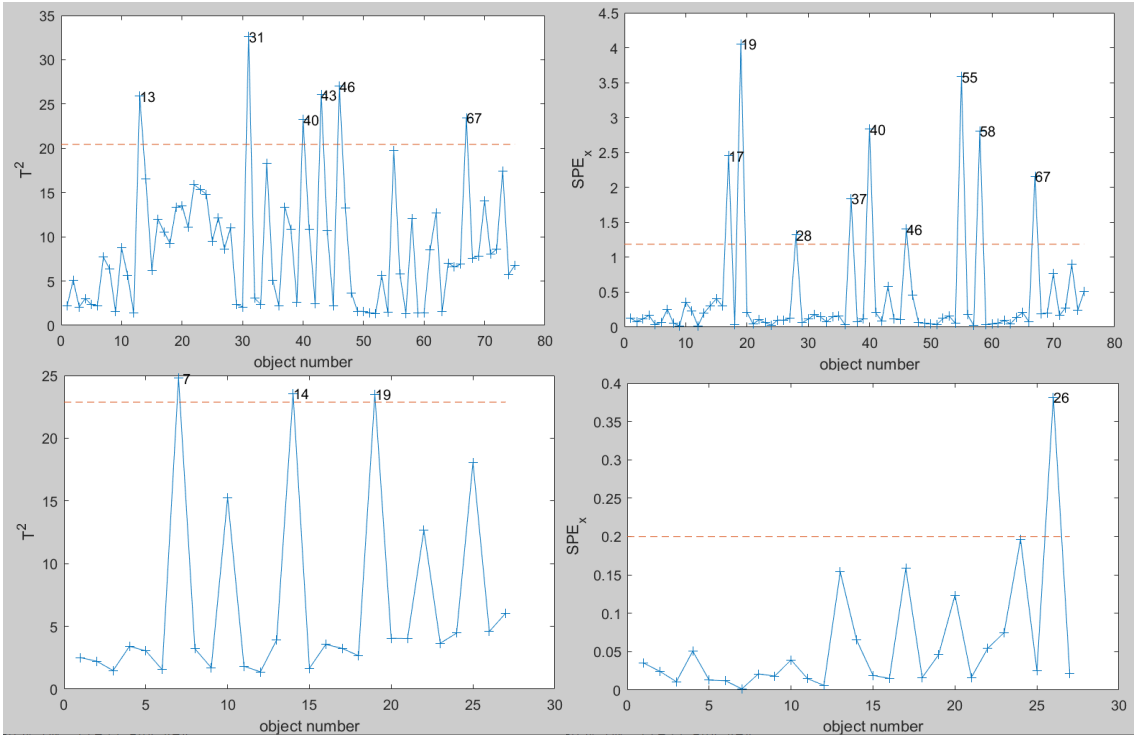


Figure 44. MSPC charts for transmittance PTO validation. In top row are all calibration and validation samples and in bottom row only the validation samples are checked. For the top row the 95% SPEx confidence limit is 1.29851 and the 95% T^2 confidence limit is 21.0226. For the bottom row the values are 0.199882 and 22.8738 respectively.

6.3 Effect of temperature

As could be seen from the Tables XVIII and XIX, there is not a big difference between the temperatures on the validation results. The difference in the results is most likely due to the calibration models and not the temperatures. This indicates that the reference values are reliable to use even if the NIR sample is analyzed hot in the process. In Figure 45. a few samples in different temperatures are presented to show that there is not a huge difference in the sample spectra regardless of the temperature. It should also be kept in mind that peaks with absorbance over 3 was not used in calibration.

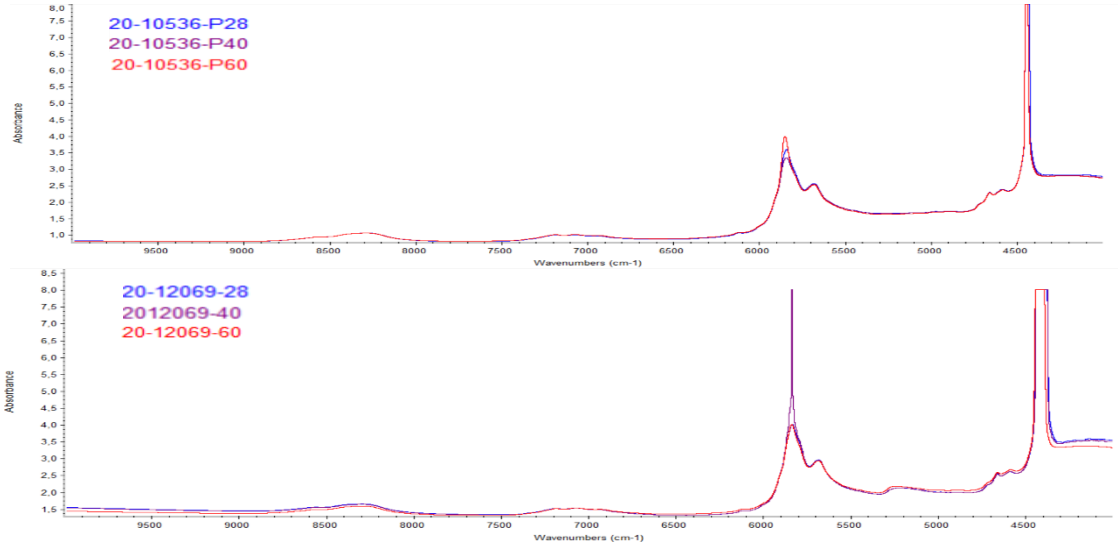


Figure 45. Top: PTO sample spectra at different temperatures. Bottom: CTO sample spectra at different temperatures. Note that large spikes with absorbance value over 3 are not used in calibration.

6.4 Comparison to reflectance method

Apart with the obvious shape difference in the spectra with the two methods, with reflectance method, the PTO spectra was noisy and with transmittance CTO and PTO spectra were more similar to each other.

When comparing the two methods, the samples from transmittance validation are used. In Table XX and XXI the percentages of these validation samples measured with reflection method with different difference percentage are presented. When comparing these tables to Tables XII and XV, it seems transmittance method is equal or better at all temperatures, except for PTO neutral substances. where temperatures 40C and 60C get worse results. At the moment this would mean the transmittance method was overall more accurate. It would however be recommendable to do more measurements and even redo the models, so they were always reliable.

Table XX. Percentages of CTO validation samples with different difference percentages. Same samples as in transmittance method are used to compare the two methods to each other.

Difference %	**resin acid, %**	**Acid number, %**	**Neutral substances, %**	**Fatty acids, %**	**Difference %**	**residual soap, %**	**Difference %**	**Water, %**
≤ Reference method measurement error	0	12.5	12.5	50	**≤ 21,1%**	25	**≤ 30,52%**	87.5
≤ 5 %	37.5	75	37.5	87.5	**≤ 25 %**	25	**≤ 35 %**	100
≤ 10 %	62.5	100	87.5	100	**≤ 30 %**	37.5	**≤ 40 %**	100

Table XXI. Percentages of PTO validation samples with different difference percentages. Same samples as in transmittance method are used to compare the two methods to each other.

Difference %	**Resin acid, %**	**Acid number, %**	**Neutral substances, %**	**Fatty acids, %**
≤ Reference method measurement error	22.2	77.8	44.4	33.3
≤ 5 %	75	100	88.9	100
≤ 10 %	100	100	100	100

7. tracking the daily PTO process

PTO samples were collected almost daily, that show the development of the process as a function of time. The samples were measured with the reflectance method. The data has been normalized between 0 and 1 with 3 principal components used in the PCA modeling. The differences in the samples can be detected by drawing the PCA scores one PC at a time (Figure 46) or following the T^2 MSPC chart.

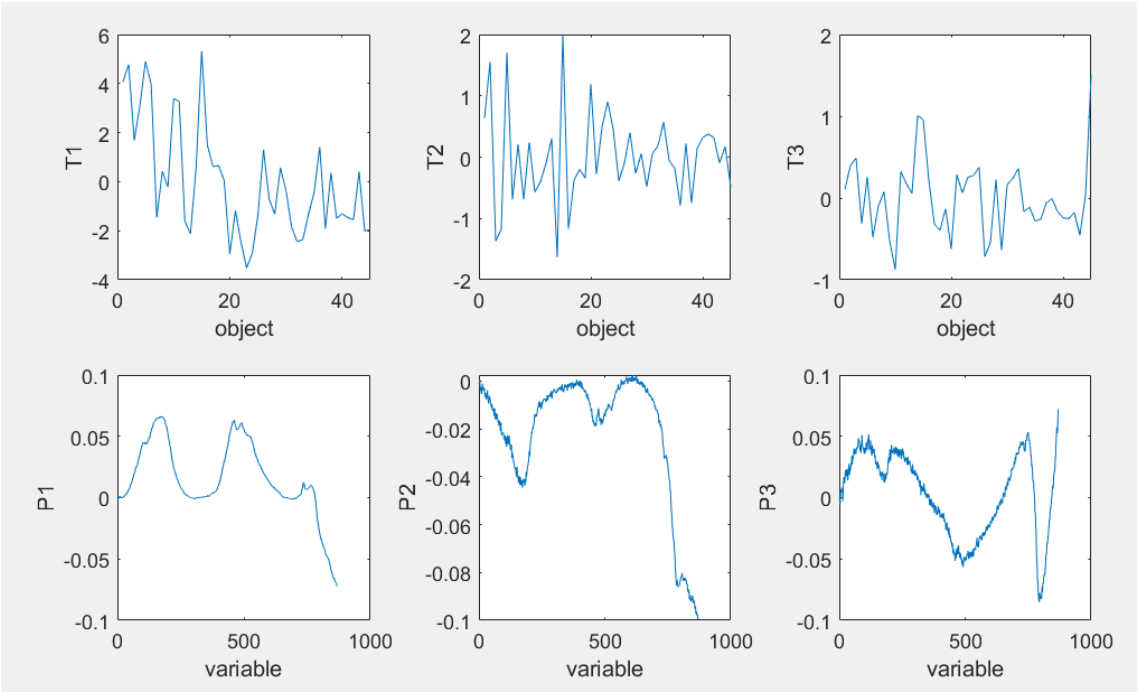


Figure 46. Scores (top row) for all 3 principal components one at a time and loadings (bottom row) for each PC. Note the change in axis values. It is hard to interpret this plot to detect abnormalities in the process.

Using the T2 chart is easier, as it shows all the PC's at a time (Figure 47). The chart shows the process leveling out after the 15th sample, and spike back again in the 45th sample. It could be assumed that during that gap the process works as it should There are a variety of possible reasons for some samples to go over the limit. The abnormalities may be in the sample quality, process, or analyzer. Reference values and spectra for samples selected based on Figure 47. are shown in Table XXII and Figure 50, to try to find the reason for these abnormalities.

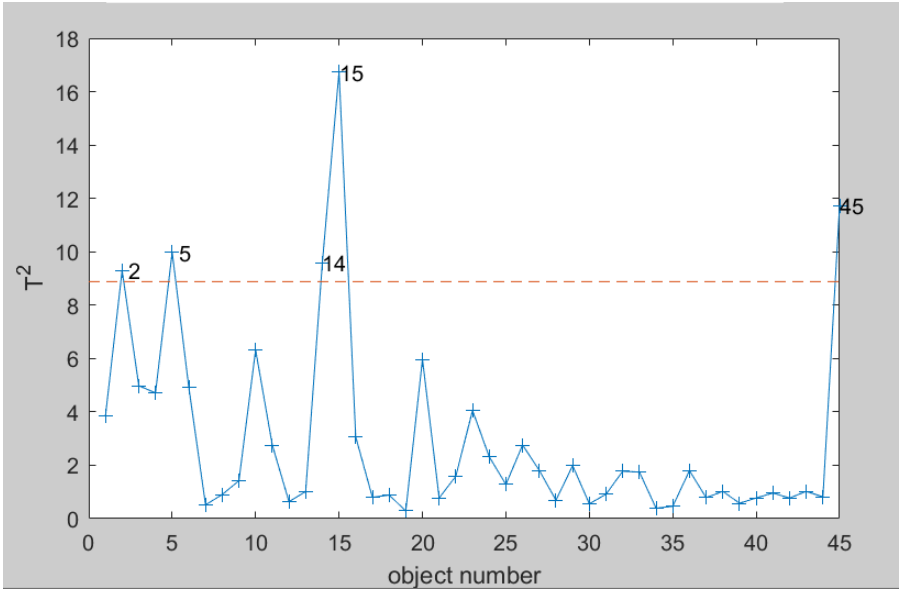


Figure 47. T2 chart for PTO process. After the 15th sample the process levels out and that can be interpret as the normal process situation when both the analyzer and process work properly.

In Figure 48. contributions for sample 15 are drawn. The red dotted line represents the confidence limits. This is shown as an example that the outlier sample contributions go over the limit at almost all variables.

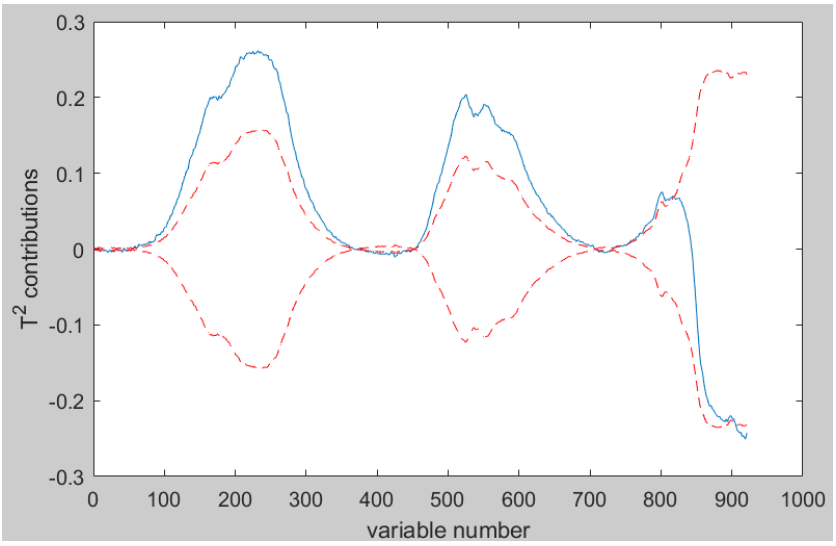


Figure 48. T^2 contributions of sample 15. The blue line represents the sample and the red dotted line represent the minimum and maximum confidence limits. It can be seen the sample crosses the limits in almost every variable.

In Figure 49a, the spectra of PTO samples are drawn with baseline correction, and in Figure 49b, is the same data normalized between 0 and 1. In these images a lot of variation can be seen between the spectra. In Figure 49c the samples from 28 to 44 are drawn and can be seen to be a lot more similar to each other. In Figure 49d, the samples from 28 to 44 are drawn and normalized between 0 and 1.

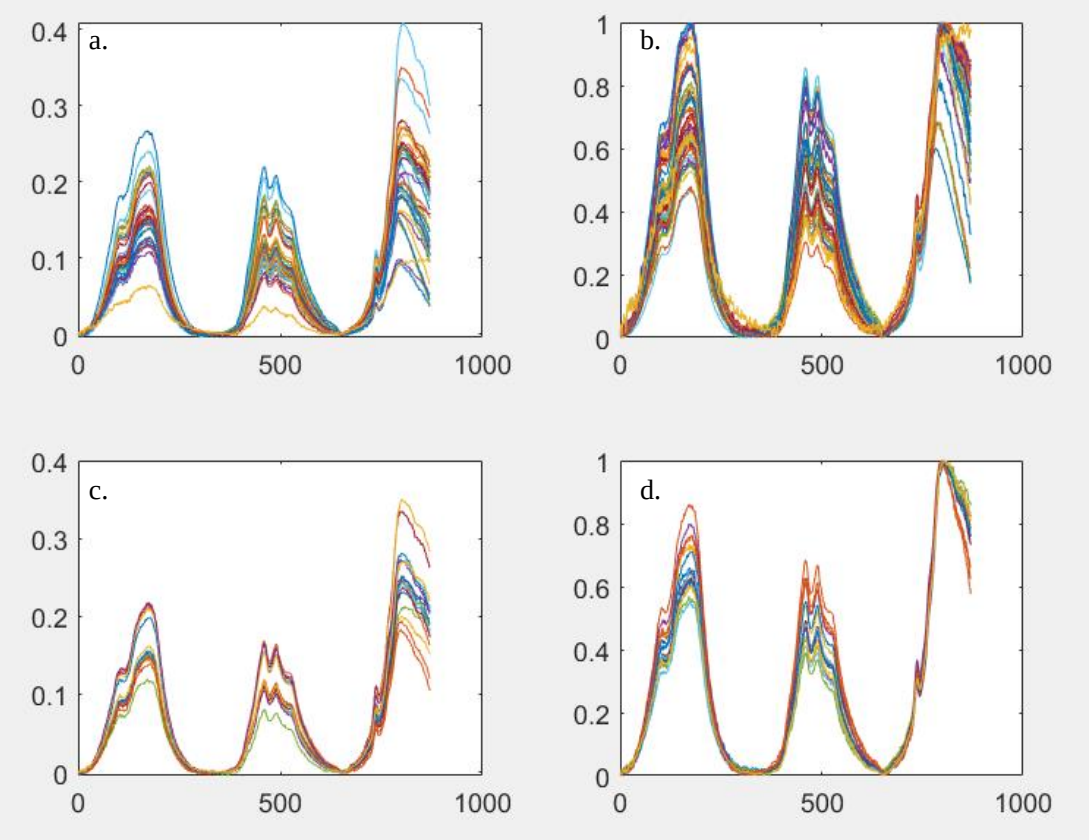


Figure 49a. Spectra for all 45 PTO samples with background correction. b. Spectra for all 45 PTO samples with background correction and normalization between 0 and 1. c. Spectra for samples 28-44 with background correction. d. Spectra for samples 28-44 with background correction and normalization between 0 and 1. Samples 28 to 44 represent a leveled part of the process, and the spectra can be a lot more similar with each other.

Some samples and their reference values are shown in Table XXII. Half of the samples go over the limit (marked with red in Table) and some stay under the limit (marked with green in Table). Based on these values it is impossible to pinpoint one clear reason why some samples go over the limit and why some stay under it. In Figure x. spectra of samples are shown to try to find differences that could explain the abnormalities.

Table XXII. Samples that go over the 95% limit marked with red (2,5,14,15,45). Samples that stay under the limit marked with green (10,26,34,35,42). From the values in the table we can't see any clear reason why some samples go over the limit and some don't.

sample no.	sample ID	Resin acids	Acid number	Neutral substances	Fatty acids
2	20-09731	25	149.5	23.2	51.8
5	20-09846	24.3	150.9	22.5	53.2
14	20-10536	28.7	154.8	20.3	51
15	20-10626	28.7	159	18.2	53.1
45	20-13450	24	149	23.5	52.5
10	20-10267	27.5	151.1	22.2	50.3
26	20-11588	25.7	152.5	21.7	52.6
34	20-12099	24.1	155.2	20.4	55.5
35	20-12807	18.9	145.9	25.4	55.7
42	20-13313	24.2	151.3	22.3	53.5

In Figure 50a, all ten sample spectra from Table XXII are drawn. Spectra does have some variance, but other than sample 45 (red), it is hard to tell whether the samples are under or over the limit. The sample 45 is also very noisy, compared to all other sample spectra. In Figure 50b. only the bad sample spectra that goes over the limit are drawn. Comparing this image to Figure 50c, where only samples that stay under the limit are drawn, it can be seen that the bad spectra has a lot more variance within themselves, and the good spectra are lot more similar to each other. In Figure 50d, two bad samples and two good samples are drawn. In addition to the difference in the absorbance values, there is a difference in the spectra in the 5500-4500 cm-1 wavenumber range. Based on this and the fact that the good samples were randomly selected, it could be assumed that the reason for the bad samples was the sample components.

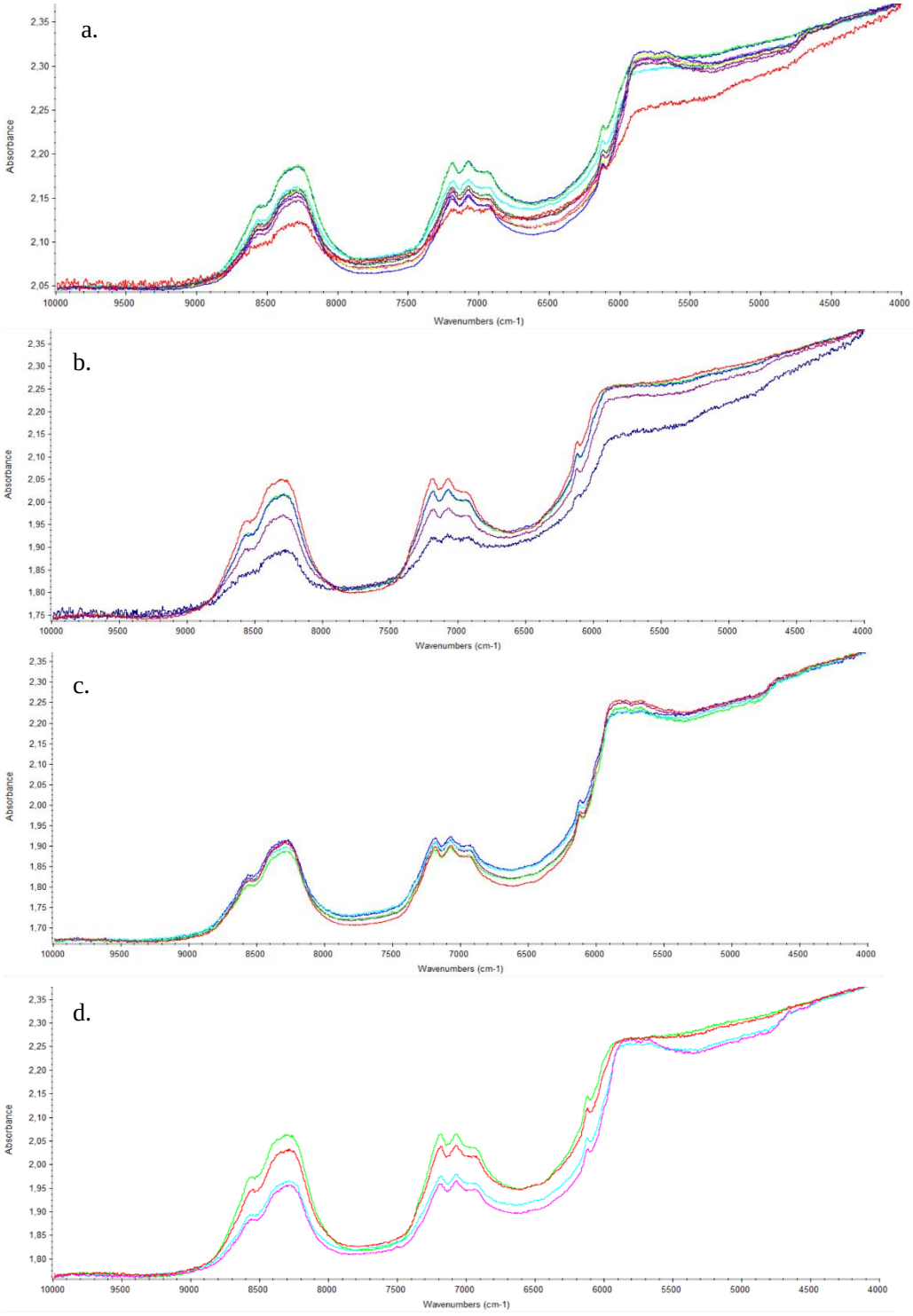


Figure 50a. all 10 samples from Table XXII. b. samples that go over the limit. c. samples that stay under the limit. d. two bad samples and two good samples.

One PCA model was made where the ‘normal’ spectra, aka samples from 28 to 44 were used to form the PCA model, where all samples are below the 95% confidence limit. The model was then tested with the original sample set of 45 samples. In Figure 51, T^2 and SPEx charts are shown for the test set. As can be excepted, most of the test samples are outside the 95% confidence limit, but the ‘normal’ samples from 28 to 44 that the model was based on, stay under the confidence limits. Note that the SPEx chart has a logarithmic scale, so that the confidence limit is easier to see. When comparing this model to the original model (Figure 47.) it is clear that the calibration data and the pretreatment of the spectra can largely affect the outcome of the model.

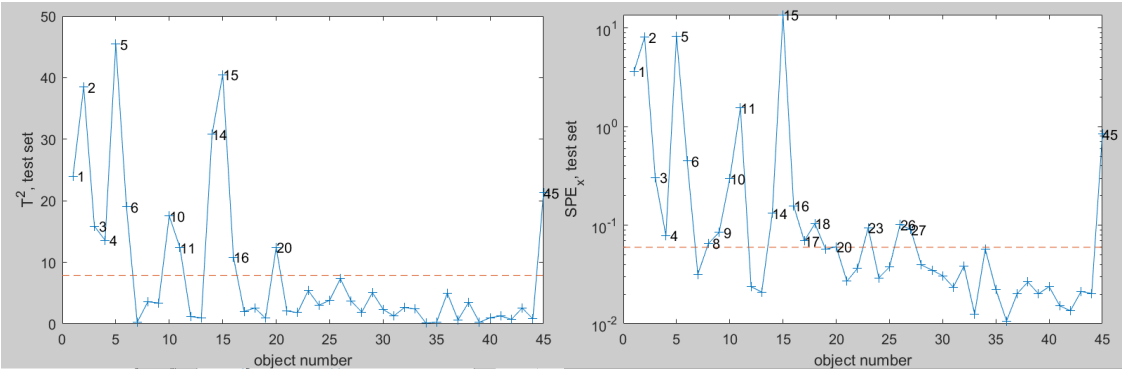


Figure 51. T^2 and SPEx charts for the test set. Model was calibrated with ‘normal’ samples from 28 to 44 and original data was used to test the model. In the SPEx chart logarithmic scale is used for the Y-axis, so that the confidence limit is easier to see. Two PCs were used. The 95% confidence limit for T^2 is 7.8573 and for SPEx 0,0596924.

8. Conclusions

Out of all the equipment researched and compared NIR seemed to have most advantages and less disadvantages. As NIR is a secondary method, the results rely largely on the accuracy of the reference methods and the mathematics of the model. Based on the validation results, the models were most likely little overfitted and could not be trusted as is. During the calibration and validation, it became clear that settings used in calibration, spectra preprocessing, variable and sample selection had a big impact on the outcome of the calibration. In future and before implementing an actual online instrument to the process, the models need more processing.

The models were made with very small calibration value range, which can make the models very local. The largest variation in values was in reflectance method resin acid with CTO samples. Even if the NIR and reference results had a large difference with some samples, the results were still in the same value range. In the validation charts the NIR measurement result differences from the reference values look more severe than they are, as the value range is so small. Based on all measurements, it seems possible to calibrate all components, while keeping in mind the reference method measurement errors.

With the reflectance method the PTO models seemed to be more accurate, but it is important to note that the PTO samples were lot more consistent to each other, and CTO samples were collected from different sources with a lot more variation in the composition. Less PTO samples were also included in the calibration, as there was less PTO samples available at the time. In the future, more samples should be included in the calibration. With transmittance method, both PTO and CTO calibrations were mostly equally accurate. Reflectance and Transmittance methods were compared to each other using the same set of validation samples. In the case the models are assumed reliable, transmittance method would be more accurate method. This would, however, need more analysis done with the transmittance method, as there was limited amount of single use cuvettes available. In the case transmittance method would be used in the online analyzer, the correct light path length should be established, as it could bring the large peaks present in the transmittance method spectra down.

With the transmittance method the samples were analyzed in three different temperatures. When checking the spectra with Matlab MSPC charts, most samples were below the 95% confidence limit, despite the temperature they were analyzed. The calibration models for the components at different temperatures could not however be made with same settings for some reason. But when the settings were optimized for temperatures, all components could be calibrated with the same reference values.

In the future, the models would need more processing, and if possible, more unison settings for all components would be good. The calibration samples would need to have a wider calibration range to include all the variation of the process. A dynamic model could also be constructed and tested for the process.

References

911 metallurgist, web page, online, obtainable: https://www.911metallurgist.com/blog/difference-between-xrf-and-xrd, sited: 10.2.2020

AAVOS international, Web page, Online, inline, atline and bypass analysers, obtainable: https://aavos.eu/glossary/online-inline-atline-bypass-analysers/ read 23.1.2020

Abdi, H., 2007, Partial least square regression (PLS regression), *in book* Encyclopedia of measurement and statistics. Sage.

Agelet, L.E., Hurburgh, C.R., 2010, A tutorial on near infrared spectroscopy and its calibration, Critical reviewa in analytical chemistry, vol 40, issue 4, pp. 246-260

Aro, T., Fatehi, P., 2017, Tall oil production from black liquor: challenges and opportunities, Separation and purification technology, vol 175, pp.469-480.

Arora, P., Ojagh, H., Woo, J., Grennfelt, E., Olsson, L., Greaser, D., 2018, Investigating the effect of Fe as a poison for catalytic HDO over sulfide NiMo alumina catalysts, Applied catalysis B: Environmental, vol 227, pp. 240-251.

Baensch, A.,2014, New generation portable XRF &XRD-mars rover technologies landing on earth, Unlocking SA's mineral wealth technical forum, conference paper, pp. 44-46

Balabin, R., Safieva, R., 2011, Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data, Analytica chimica acta, vol 689, issue 2, pp. 190-197.

Baptista, P., Felizardo, P., Menezes, J.C., Correia, M.J.N., 2008, Monitoring the quality of oils for biodiesel production using multivariate near infrared spectroscopy models, Journal of near infrared spectroscopy, vol. 16 issue 5, pp.445-454

Barbosa, L., Knock, F., Silva, R., Freitas, J., Lacerda, V., Castro, E., 2013, Application of low-field NMR for the determination of physical properties of petroleum fractions, Energy & fuels, vol.27, issue 2, pp. 73-679

Bart, J., 2006, Plastic additives: advanced industrial analysis, IOS press, pp. 34-52

Bart, J., Gucciardi, E., Cavallaro, S., 2013, Quality assurance of biolubricants, Chapter 8 in book biolubricants; Science and technology, pp.396-450.

Baum, A., Vermue, L., 2019, Multiblock PLS: Block dependent prediction modeling for python, The journal of open source software, Vol 4. pp. 1190

Bezergianni, S., Dimitriadis, A., 2013, Comparison between different types of renewable diesel, Renewable and sustainable energy reviews, vol 21, pp.110-116.

Blümich, B., 2005, Essential NMR; for scientist and engineers, Springer, pp. 215-238

Bono, V., 2014, Characterization of fibrous fractions from wine industry by-products and their use in baked goods, doctoral thesis, University of Milan, pp. 38-55

Buchi, NIR-Online process analyzer, webpage, obtainable: https://www.buchi.com/en/products/nirsolutions/nir-online-process-analyzer sited: 28.1.2020

Burdick, R., 2018, What are the disadvantages of HPLC? Sciencing, online, obtainable: https://sciencing.com/list-7237518-disadvantages-hplc-.html sited: 4.2.2020

Chemguide, what is nuclear magnetic resonance (NMR)?, online, obtainable: https://www.chemguide.co.uk/analysis/nmr/background.html sited: 5.2.2020

Chmielarz, M., Sampels, S., Blomgvist, J., Brandenburg, J., Wende, F., Sandgren, M., Passoth, V., 2019, FT-NIR: a tool for rapid intracellular lipid quantification in oleaginous yeasts, Biotechnology for biofuels, vol. 12, article number: 169

ChromAcademy, The theory of HPLC; gradient HPLC, online, obtainable: https://www.chromacademy.com/lms/sco8/Theory_Of_HPLC_Gradient_HPLC.pdf sited: 4.2.2020

Cordella, C., 2012, PCA: the basic building block of chemometrics, Chapter 1 *in book* analytical chemistry, Intechopen

de Lira, L., de Albuquerque, M, Pacheco, J., Fonseca, T., Calvanti, E., Stragevitch, L., Pimentel, M., 2010 Infrared spectroscopy and multivariate calibration to monitor stability quality parameters of biodiesel, Microchemical Journal, vol 96, issue 1, pp. 126-131

Dionex, 2002, Chromatography with on-line HPLC and ion chromatography for process control, *written by* Doyle, M., Newton, B., online, obtainable: http://dionex.su/en-us/webdocs/6647-DX-800_CAST_Article.pdf sited 4.2.2020

Edwards, J.C., 2009, Principles of NMR, obtainable: https://www.researchgate.net/publication/268026544_Principles_of_NMR sited: 17.2.2020

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001, Multi- and megavariate data analysis, principles and applications, Umerics academy, Sweden, pp.43-111,215-269

European commission, Renewable energy-recast to 2030(RED II), EU science hub, online: https://ec.europa.eu/jrc/en/jec/renewable-energy-recast-2030-red-ii sited: 20.1.2020

FI 126029B, UPM-Kymmene corporation, 2015, Process for producing hydrocarbons, (Nousiainen J., Kukkonen, P., Kotoneva, J., Lindberg, T., Äijälä, T.), Patentti ja rekisterihallitus

Gazulla, M., Orduña, M., Vicente, S., Rodrigo, M., 2013, Development of a WD-XRF analysis method of minor and trace elements in liquid petroleum products, Fuel, vol 108, pp. 247-253.

Gorai, A., Tuluri, F., Tchounwou, P., 2015, Development of PLS-path model for understanding the role of precursors on ground level ozone concentration in Gulfport, Mississippi, USA, Atmospheric pollution research, vol 6, issue 3, pp. 389-397

Gullichsen, J., Lindberg, H., 1999, Byproducts of chemical pulping, *in book* Papermaking science and technology; Chemical pulping 6B, Fabet Oy, pp. 378-387

Günther, H., 2013, NMR spectroscopy: basic principles, concepts and applications in chemistry, Wiley, 3rd edition.

Halme, N., 2019, Optimization of crude tall oil evaporation, Master's thesis, Aalto university.

Hase, A., Koppinen, S., Riistama, K., Vuori, M., 2003, Suomen kemianteollisuus, Chemas Oy, pp. 128-138

Higson, S., 2004, Separatory methods and chromatography *in book* analytical chemistry, Oxford university press, pp. 230-241

Hombre analyzer solutions, C-quand: on-line XRF elemental analysis in liquids, web page, obtainable: https://hobre.com/products/cquand-elemental-analysis-in-liquids/ sited: 11.2.2020

Huang, S., Wang, X., Chen, Y., Xu, J., Tang, T., Mu, B., 2019, Modeling and quantitative analysis of x-ray transmittance and backscatter imaging aimed at security inspection, optics express, vol 27, issue 2.

Jaarinen, S., Niiranen, J., 2005, laboratorion analyysitekniikka, Edita Helsinki, 5th edition pp. 153-177

Jolliffe, I., Cadima, J., 2016, Principal component analysis: a review and recent developments, The Royal Society publishing, Philosophical transactions of the royal society a mathematical, physical and engineering sciences.

Kalnicky, D., Singhvi, R., Field portable XRF analysis of environmental samples, Journal of hazardous materials, vol. 83, pp. 93-122

Koester, V., 2016, What is HPLC? Chemistry views magazine, online, obtainable: https://www.chemistryviews.org/details/education/9464911/What_is_HPLC.html sited: 3.2.2020

Kohonen, J., Reinikainen, S-P., Aaljoki, K., Perkiö, A., Väänänen, T., Höskuldsson, A., 2008, Multi-block methods in multivariate process control, Journal of chemometrics, vol. 22, pp. 281-287.

Kohonen, J., 2009, Advanced chemometric methods: Applicability on industrial data, Doctoral dissertation, Lappeenranta university of technology.

Kotoneva, J., 2020, Laboratory manager, UPM biofuels, personal interview.

van Kollenburg, G, van Es, J., Gerretzen, J., Lanters, H., Bouman, R., Koelewiin, W., Davies, A., Buyndens, M., van Manen, H-J., Jansen, J., 2020, Understanding chemical production processes by using PLS path model parameters as soft sensors, Computers and chemical engineering, vol 139. article no.106841

Kourti, T., MacGregor, J., 1994, Process analysis, monitoring and diagnosis, using multivariate projection methods, Chemometrics and intelligent laboratory systems, vol 28, pp. 3-19.

Laine-Ylijoki, J., Rustad, I., Syrjä, J-J., Wahlström M., 2003b, Suitability of XRF-methods to on-site testing of waste materials, Nordtest report, Obtainable: http://www.nordtest.info/index.php/technical-reports/item/suitability-of-xrf-methods-on-on-site-testing-of-waste-materials-nt-tr-545.html sited: 10.2.2020

Laine-Ylijoki, j., Syrjä, J-J., Wahlström, M., 2003a, Röntgenflueresenssimenetelmät kerrätyspolttoaineiden pikalaadunvalvonnassa, VTT prosessit, Otamedia Oy, obtainable: https://www.vtt.fi/inf/pdf/tiedotteet/2003/T2215.pdf sited: 10.2.2020

Lakshmanan, S., 2019, How, when and why should you normalize/ standardize/ rescale your data?, online, obtainable: https://medium.com/@swethalakshmanan14/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff sited 17.4.2020

Laxen, T., Tikka, P., 2008, Chemical pulping part 2: recovery of chemicals and energy, *in book* papermaking science and technology, Paper engineer's association, Second edition, pp.360-380

Leivo, J., 2018, Optimizing supply chain costs by minimizing the variance in raw material composition, Master's thesis, Aalto university, school of chemical engineering.

Li, X., Sun, C., Zhou, B., He, Y., 2015, Determination of hemicellulose, cellulose, and lignin in Moso bamboo by near infrared spectroscopy, Scientific reports, vol.5 article number 17210.

Malvern panalytical, Epsilon Xflow, web page, obtainable: https://www.malvernpanalytical.com/en/products/product-range/epsilon-range/epsilon-xflow sited: 11.2.2020

Mannonen, S., 2014, UPM biofuels; biofuels from wood based raw materials, conference material, Bryssels 15.10.2014.

Martonen, H., 2019, Johdatus kemiallisten siirtymien kvanttimekaaniseen laskemiseen, Master's thesis, University of Jyväskylä.

McWirth, A., Weindorf, D., Zhu, Y., 2012, Rapid analysis of elemental concentrations in compost via portable x-ray fluorescence spectrometry, Compost science& utilization, vol. 20, issue 3, pp. 185-193

Mujunen, S-P, Minkkinen, P, Wirkkala, R-S., 1996, Metsäteollisuuden aktiivilietelaitosten puhdistustehon ja häiriötilanteiden ennustaminen monimuuttujamenetelmien avulla: Osa 2, puhdistetun veden typpi, fosfori ja COD, Lappeenrannan teknillinen korkeakoulu, kemiantekniikan osasto, pp. 8-10

Mujunen, S-P., Teppola, P., Minkkinen, P., 1997, Metsäteollisuuden aktiivilietelaitosten toiminnan monimuuttujainen seuranta ja mallintaminen, Suomen ympäristö, Kaakkois-suomen ympäristökeskus, pp. 17-21

Murray, T.P., Lehr, J.R., 1981, Compositional analysis of tall oils used as phosphate flotation reagent, National fertilizer development center, Tennessee valley authority.

Myllyviita, A., 2019, NMR-spektroskopia, Peda.net, online, obtainable: https://peda.net/p/myllyviita/OrbitaaliMarvinSketch/opettajalle/spektroskopia/63n sited: 5.2.2020

Nomngongo, P., Munonde, T., Mpupa, A., Biata, N., 2016, Near-infrared spectroscopy combined with multivariate tools for analysis of trace metals in environmental matrices, *in book* Deveploments in near-infrared spectroscopy, InTechOpen

Nordon, A., McGill, C., Littlejohn D., 2002, Evaluation of low-field nuclear magnetic resonance spectrometry for at-line-process analysis, Applied spectrometry, vol. 56, issue 1, pp.75-82

Norlin, L-H.,2012, Tall oil, in Ullmann's encyclopedia of industrial chemistry, vol 35, pp. 583-396.

Núñez-Sánchez, N., Martinez-Marin, A., Polvillo, O., Fernández-Cabanás, V., Carrizosa, J., Urrutia, B., Serradilla, J., 2016, Near infrared spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats, Food chemistry, vol. 190, pp.244-252

O'Rourke, N., Hatcher, L., 2013, A step-by-step approach to using SAS for factor analysis and structural equation modeling, Second edition, USA, SAS institute Inc. pp.16-31

Owen, J., 1995, Uses of derivative spectroscopy, Application Note, Agilent technologies, Germany.

Pasquini, C., 2003, Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications, Journal of the Brazilian chemical society, vol. 14 nro.2

Perego, C., Ricci, M., 2012, Diesel fuel from biomass, Catalysis and science technology, vol. 2/2012, pp. 1776-1786

Peters, D., Stojcheva, V.,2017, Crude tall oil low ILUC risk assessment; comparing global supply and demand, Ecofys, obtained: https://www.upmbiofuels.com/traffic-fuels/upm-bioverno-diesel-for-fuels/ sited 14.1.2020

Prieto, N., Pawluczyk, O., Dugan, M.E.R., Aalhus, J.L., 2017, A Review of the principles and applications of near-infrared spectroscopy to characterize meat, fat and meat products, Sage Journals, Applied spectroscopy, vol. 7, issue 7, pp.1403-1426

Prokkola, H., Kuokkanen, T., Vähäoja, P., Kangas, T., Karhu, M., Rämö, J., Lassi, U., 2014, Characterization and biodegradation rates of tall soaps in different water and soil environments, Water, air & soil pollution, vol 225, article number 2070

Qian, F., Wu, Y., Hao, P, 2017, A fully automated algorithm of baseline correction based on wavelet feature points and segment interpolation, Optics and laser technology, vol. 96, pp. 202-207

Reichenbäcker,M., Popp, J., Chapter 4; nuclear magnetic resonance spectroscopy, *in* book Challenges in molecular structure determination, Springer, pp. 215-218

Rigagu, NEX OL process elemental analyzer, web page, obtainable: https://rigakuedxrf.com/nex-ol.php?gclid=EAIaIQobChMIo7nFsJjJ5wIVSOaaCh1nYwJlEAAYAiAAEgL2pPD_BwE sited: 11.2.2020

Rinnan, Å., van der Berg, F., Engelsen, S., 2009, Review of the most common pre-processing technigues for near-infrared spectra, Trends in analytical chemistry, vol 28, issue 10, pp. 1201-1222

Roberts, J., Power, A., Chapman, J., Chandra, S., Cozzolino, D., 2018, Vibrational spectroscopy methods for agro-food product analysis, Chapter Three *in book* Comprehensive Analytical Chemistry, vol 80, pp.51-68.

Romañach, R.J., 2010, Vibrational Spectroscopy for Pharmaceutical Analysis, Part IX. Introduction and Origin of NIR Bands, teaching material, Engineering research center for structured organic particulate systems, online, obtainable: https://www.slideserve.com/libitha/vibrational-spectroscopy-for-pharmaceutical-analysis sited: 26.2.2020

Rule, G.S., Hitchens, T.K., 2006 Fundamentals of protein NMR spectroscopy, Springer, Netherlands

Saha, S., 2018, Baffled by covariance and correlation??? Get the math and the application in analytics for both the terms, Towards data science, Online, obtainable: https://towardsdatascience.com/let-us-understand-the-correlation-matrix-and-covariance-matrix-d42e6b643c22 sited 28.4.2020

Silva, R., Carneiro, G., Barbosa. L., Lacerda, V., Freitas, J., Castro, E., 2012, Studies on crude oil-water biphasic mixtures by low-field NMR, Magnetic resonance in chemistry, vol. 50, issue 2, pp. 85-88

Solunetti, NMR (nuclear magnetic resonance) eli ydinmagneettinen resonanssi, online, obtainable: http://www.solunetti.fi/fi/solubiologia/nmr/ sited: 5.2.2020

Stevens, T., Chritz, K., Small, H., 1987, Determination of water by liquid chromatography using conductometric detection, Analytical chemistry, vol 59, pp.1716-1720.

Stordrange, L., Rajalahti, T., Libnau, F., 2004, Multiway methods to explore and model NIR data from batch process, Chemometrics and intelligent laboratory systems, vol. 70, pp.137-145

Teppola, P., Mujunen, S-P., Minkkinen, P., Wirkkala, R-S., 1997, Monimuuttujamenetelmien käyttö sellu- ja paperitehtaiden aktiivilietelaitosten toiminnan seurantaan; Pääkomponenttianalyysipohjaiset menetelmät sekä sumea luokittelu, Lappeenrannan teknillinen korkeakoulu, Kemiantekniikan osasto, pp. 12-27

Thybring, E., Kymäläinen, M., Rautkari, L., 2017, Experimental techniques for characterizing water in wood covering the range from dry to fully water-saturated, Wood science and technology, vol 52, pp.297-329

Ungureanu, E., Ungureanu, O., Căpraru, A-M., Popa,V., 2009, Chemical modification and characterization of straw lignin, Cellulose chemistry and technology, vol. 43, pp. 263-269.

UPM, UPM biofuels web page, obtainable: https://www.upmbiofuels.com/traffic-fuels/upm-bioverno-diesel-for-fuels/ sited: 14.1.2020

Vikström, F., Holmbom, B., Hamunen, A., 2005 Sterols and triterpenyl alcohols in common pulpwoods and black liquor soaps, Holz als Roh- und Werkstoff, vol 63, pp. 303-308.

Vilonen, K., 2020, Spesialist, UPM, Personal interview 22.1.2020

Voegele, E., 2019, Finland to require 30% biofuel, 10% advanced biofuel by 2030, Biomass magazine, online: http://biomassmagazine.com/articles/15930/finland-to-require-30-biofuel-10-advanced-biofuel-by-2030, sited: 20.1.2020

Wang, S-F., Furuno, T., Cheng, Z., 2005, Study of extraction of phytosterol from masson pine raw tall oil, The Japan wood research society, Vol 48. pp- 505-511

Wang, S-F., Furuno, T., Cheng, Z., Katoh., S., 2001, Composition of neutral fractions in Chinese raw tall oil, The Japan wood research society, Vol. 47 pp. 400-405

Wansbrough, H., Rough, M., Cooney, S., 2008, tall oil production and processing, obtainable: https://nzic.org.nz/app/uploads/2017/10/4G.pdf sited: 23.1.202

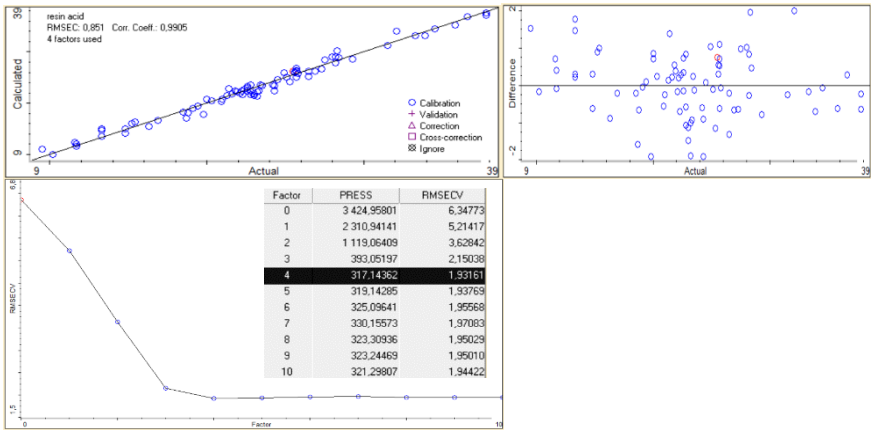
Yang, G., Zhang, C., Hu, Q., Yin, J., 2003, Simultaneous determination of four heavy metal ions in tobacco and tobacco additive by online enrichment followed by RP-HPLC and microwave digestion, Journal of chromatographic science, vol 41, pp. 195-199.

Zajac, G., Szyszlak-barglowicz, J., Slowik, t., Kuranc, A., Kaminska, A., 2015, Designation of chosen heavy metals in used engine oils using the XRF method, Polish journal of environmental studies, vol 24, issue 5, pp. 2277-2283

Zhang, Z., Castelló, A., 2017, Principal components analysis in clinical studies, Annals of translational medicine, vol 5(17), article number 351.

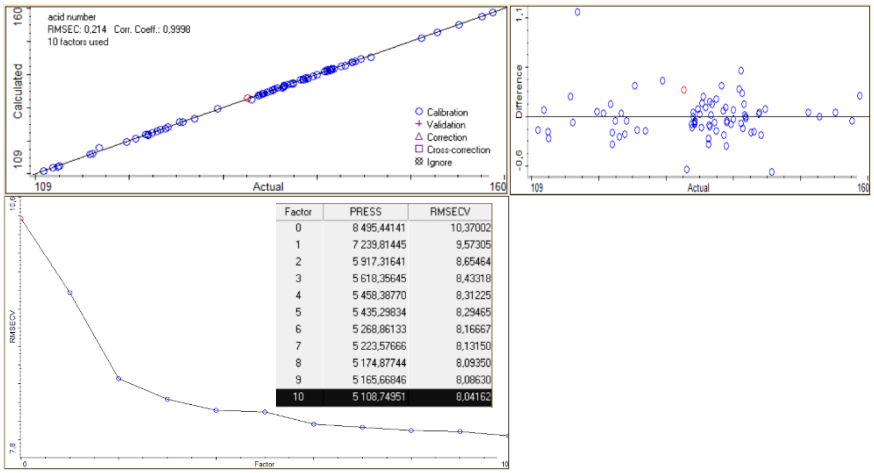
Zientek, N., Meyer, K., Kern, S., Maiwald, M., 2016, Quantitative online NMR spectroscopy in a nutshell, Chemie Ingenieur Technik, vol 88, issue 6, pp. 698-709

APPENDIX I. Calibration charts



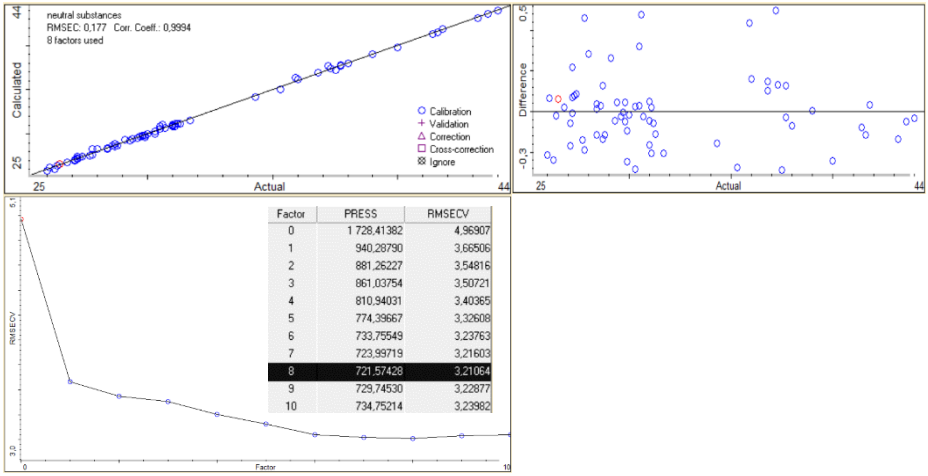
Factor	PRESS	RMSECV
0	3 424,95801	6,34773
1	2 310,94141	5,21417
2	1 119,06409	3,62842
3	393,05197	2,15038
4	317,14362	1,93161
5	319,14285	1,93769
6	325,09641	1,95568
7	330,15573	1,97083
8	323,30936	1,95029
9	323,24469	1,95010
10	321,29807	1,94422

Figure 1. Reflectance CTO resin acid calibration charts



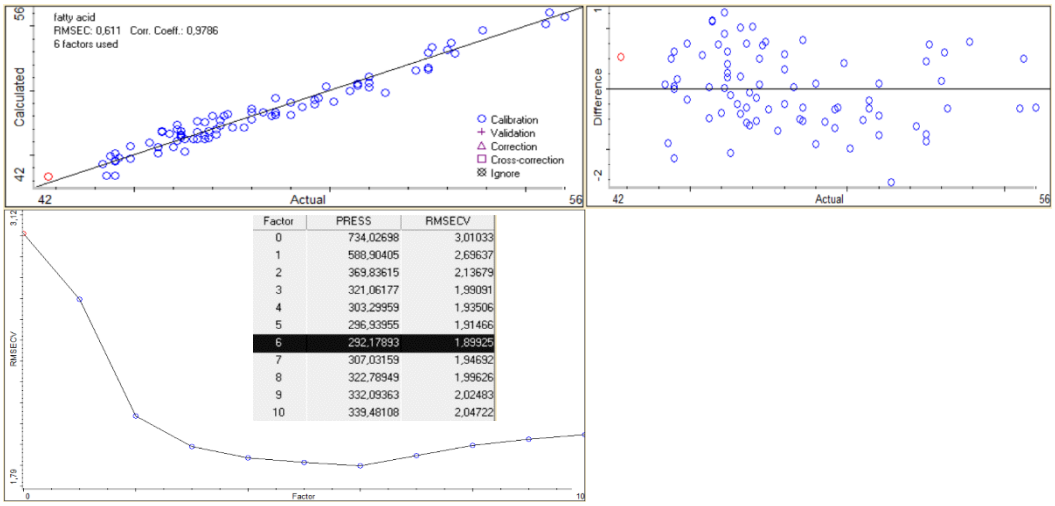
Factor	PRESS	RMSECV
0	8 495,44141	10,37002
1	7 239,81445	9,57305
2	5 917,31641	8,65464
3	5 618,35645	8,43318
4	5 458,38770	8,31225
5	5 435,29834	8,29465
6	5 268,86133	8,16667
7	5 223,57666	8,13150
8	5 174,87744	8,09350
9	5 165,66846	8,08630
10	5 108,74951	8,04162

Figure 2. Reflectance CTO acid number calibration charts



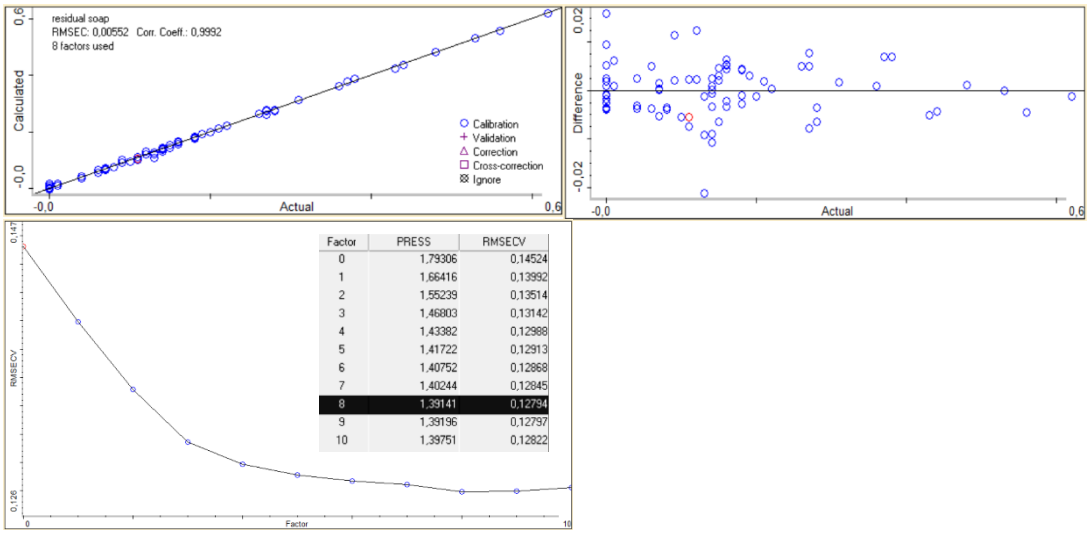
Factor	PRESS	RMSECV
0	1 728,41382	4,96907
1	940,28790	3,66506
2	881,26227	3,54816
3	861,03754	3,50721
4	810,94031	3,40365
5	774,39667	3,32608
6	733,75549	3,23763
7	723,99719	3,21603
8	721,57428	3,21064
9	729,74530	3,22877
10	734,75214	3,23982

Figure 3. Reflectance CTO neutral substances calibration chart



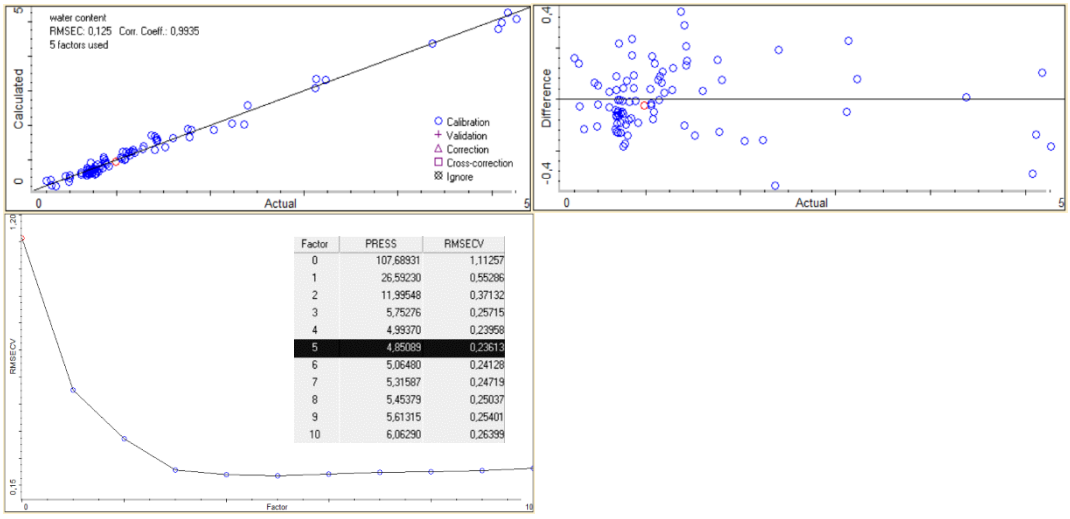
Factor	PRESS	RMSECV
0	734,02698	3,01033
1	588,90405	2,69637
2	369,83615	2,13679
3	321,06177	1,99091
4	303,29959	1,93506
5	296,93955	1,91466
6	292,17893	1,89925
7	307,03159	1,94692
8	322,78949	1,99626
9	332,09363	2,02483
10	339,48108	2,04722

Figure 4. Reflectance CTO fatty acid calibration charts



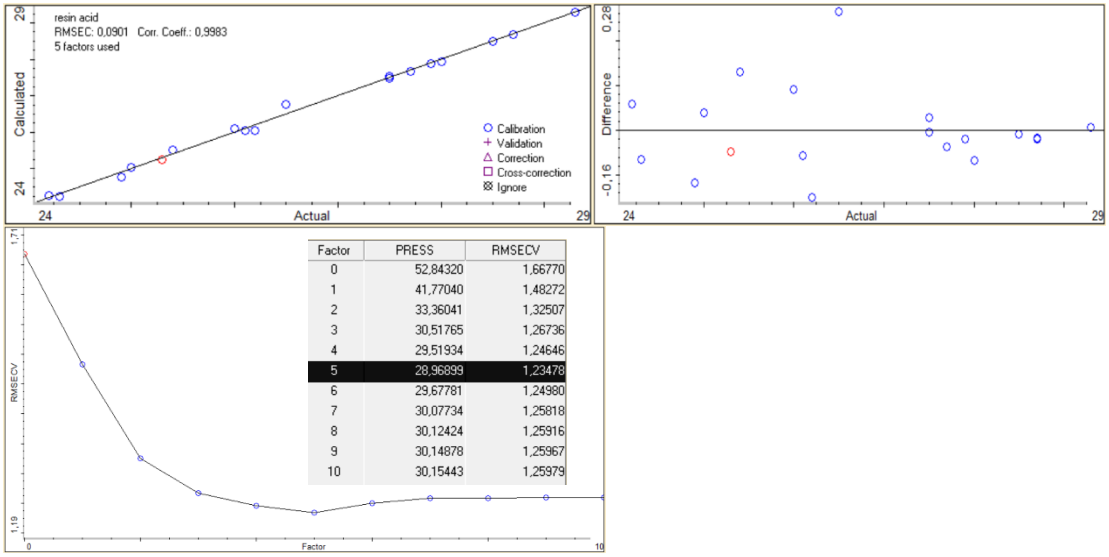
Factor	PRESS	RMSECV
0	1,79306	0,14524
1	1,66416	0,13992
2	1,55239	0,13514
3	1,46803	0,13142
4	1,43382	0,12988
5	1,41722	0,12913
6	1,40752	0,12868
7	1,40244	0,12845
8	1,39141	0,12794
9	1,39196	0,12797
10	1,39751	0,12822

Figure 5. Reflectance CTO residual soap calibration charts



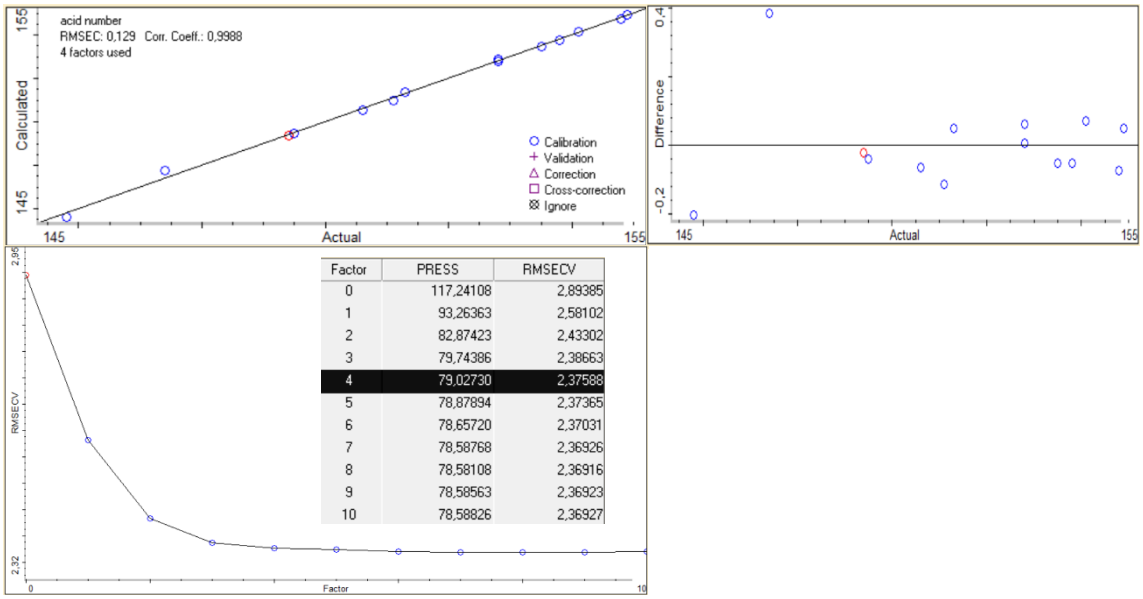
Factor	PRESS	RMSECV
0	107,58931	1,11257
1	26,59230	0,55286
2	11,99548	0,37132
3	5,75276	0,25715
4	4,99370	0,23958
5	4,85089	0,23613
6	5,06480	0,24128
7	5,31587	0,24719
8	5,45379	0,25037
9	5,61315	0,25401
10	6,06290	0,26399

Figure 6. Reflectance CTO water content calibration charts



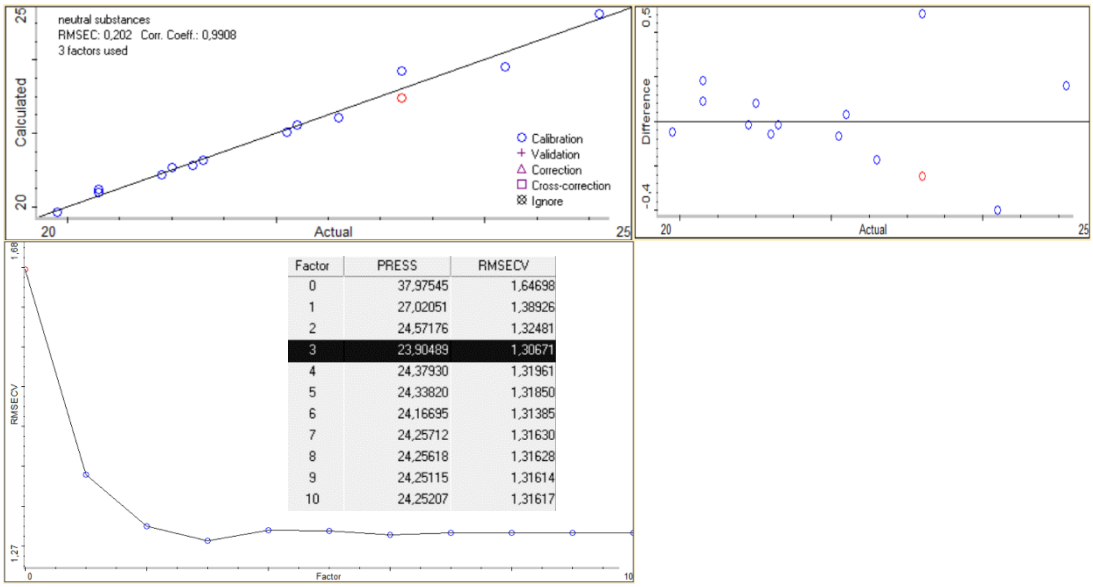
Factor	PRESS	RMSECV
0	52,84320	1,66770
1	41,77040	1,48272
2	33,36041	1,32507
3	30,51765	1,26736
4	29,51934	1,24646
5	28,96899	1,23478
6	29,67781	1,24980
7	30,07734	1,25818
8	30,12424	1,25916
9	30,14878	1,25967
10	30,15443	1,25979

Figure 7. Reflectance PTO resin acid calibration charts



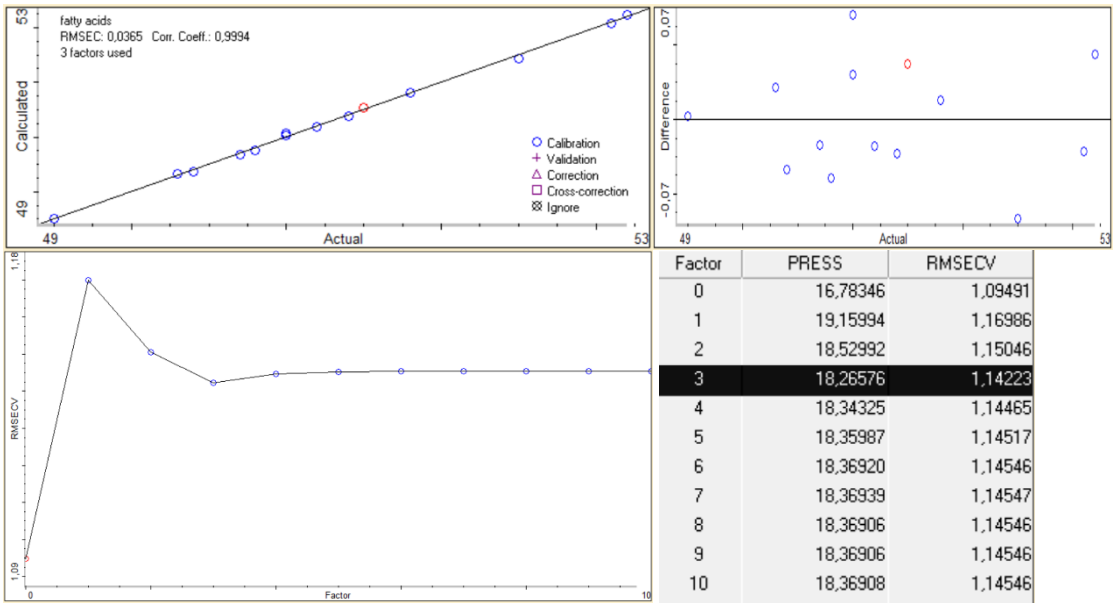
Factor	PRESS	RMSECV
0	117,24108	2,89385
1	93,26363	2,58102
2	82,87423	2,43302
3	79,74386	2,38663
4	79,02730	2,37588
5	78,87894	2,37365
6	78,65720	2,37031
7	78,58768	2,36926
8	78,58108	2,36916
9	78,58563	2,36923
10	78,58826	2,36927

Figure 8. Reflectance PTO acid number calibration charts



Factor	PRESS	RMSECV
0	37,97545	1,64698
1	27,02051	1,38926
2	24,57176	1,32481
3	23,90489	1,30671
4	24,37930	1,31961
5	24,33820	1,31850
6	24,16695	1,31385
7	24,25712	1,31630
8	24,25618	1,31628
9	24,25115	1,31614
10	24,25207	1,31617

Figure 9. Reflectance PTO neutral substances calibration charts



Factor	PRESS	RMSECV
0	16,78346	1,09491
1	19,15994	1,16986
2	18,52992	1,15046
3	18,26576	1,14223
4	18,34325	1,14465
5	18,35987	1,14517
6	18,36920	1,14546
7	18,36939	1,14547
8	18,36906	1,14546
9	18,36906	1,14546
10	18,36908	1,14546

Figure 10. Reflectance PTO fatty acids calibration charts.

CTO transmittance calibration charts

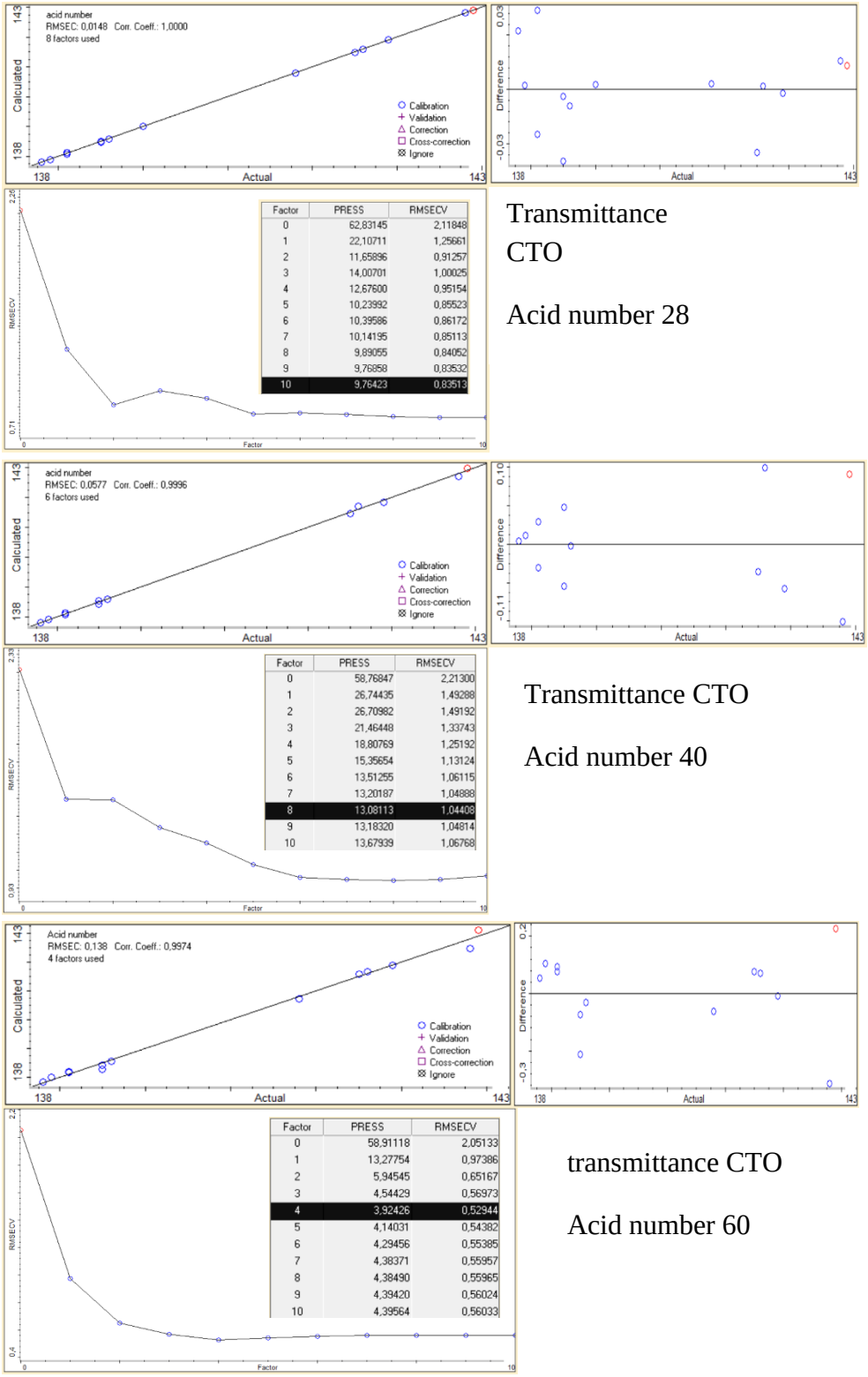


Figure 11. Transmittance CTO Acid number calibration charts

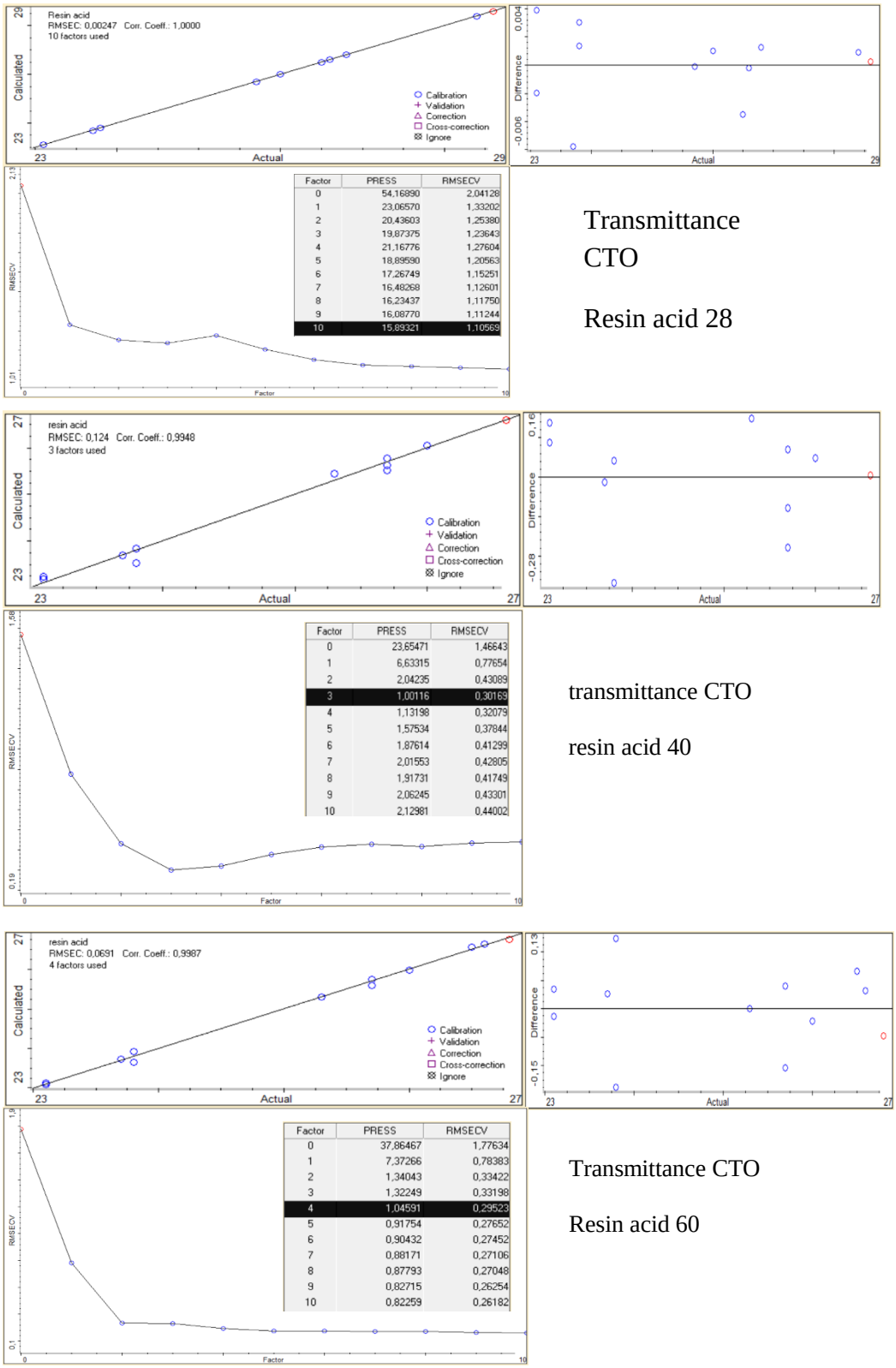
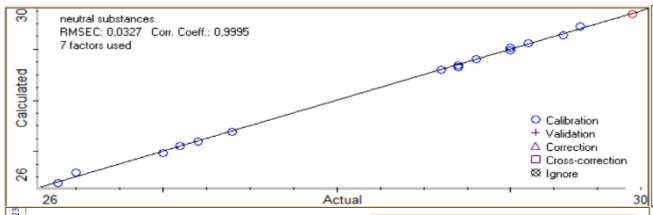
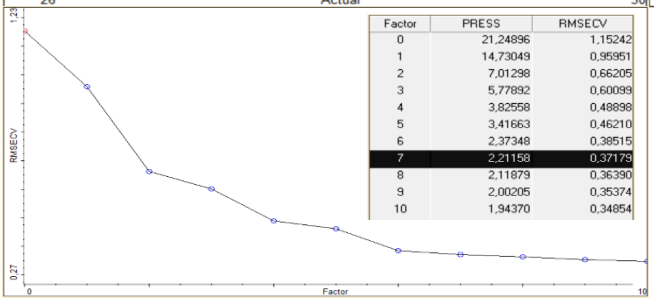


Figure 12. Transmittance CTO resin acid calibration charts.

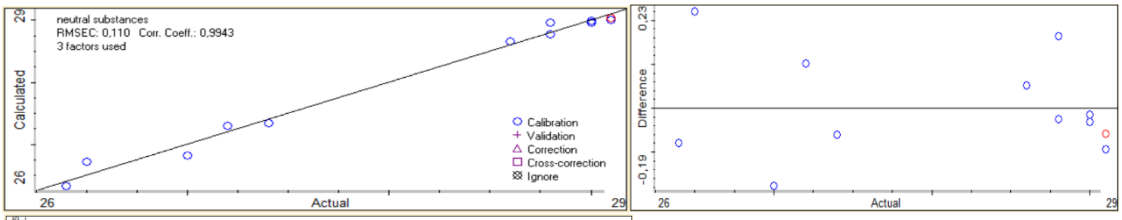


Factor	PRESS	RMSECV
0	21,24896	1,15242
1	14,73049	0,95951
2	7,01298	0,66205
3	5,77892	0,60099
4	3,82558	0,48898
5	3,41663	0,46210
6	2,37348	0,38515
7	2,21158	0,37179
8	2,11879	0,36390
9	2,00205	0,35374
10	1,94370	0,34854

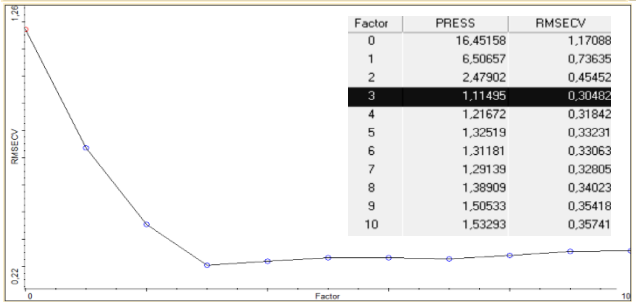


Transmittance CTO

neutral substances 28

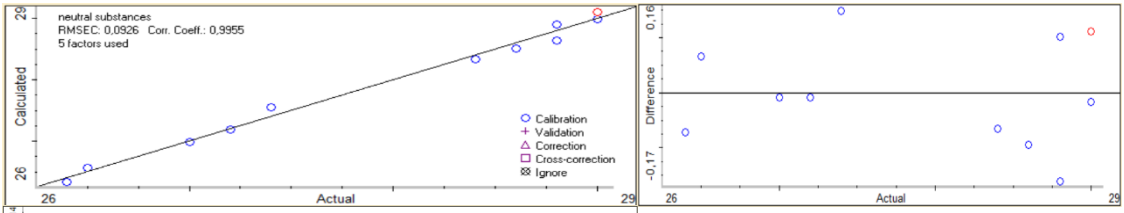


Factor	PRESS	RMSECV
0	16,45158	1,17088
1	6,50657	0,73635
2	2,47902	0,45452
3	1,11495	0,30482
4	1,21672	0,31842
5	1,32519	0,33231
6	1,31181	0,33063
7	1,29139	0,32805
8	1,38909	0,34023
9	1,50533	0,35418
10	1,53293	0,35741

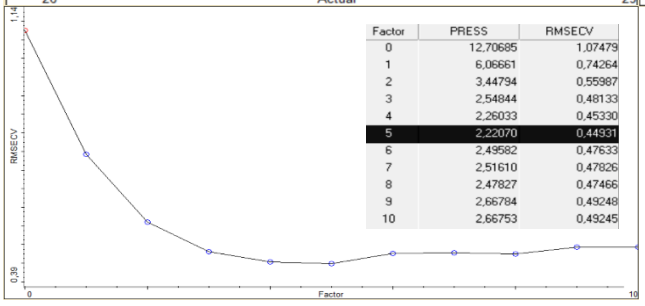


transmittance CTO

Neutral substances 40



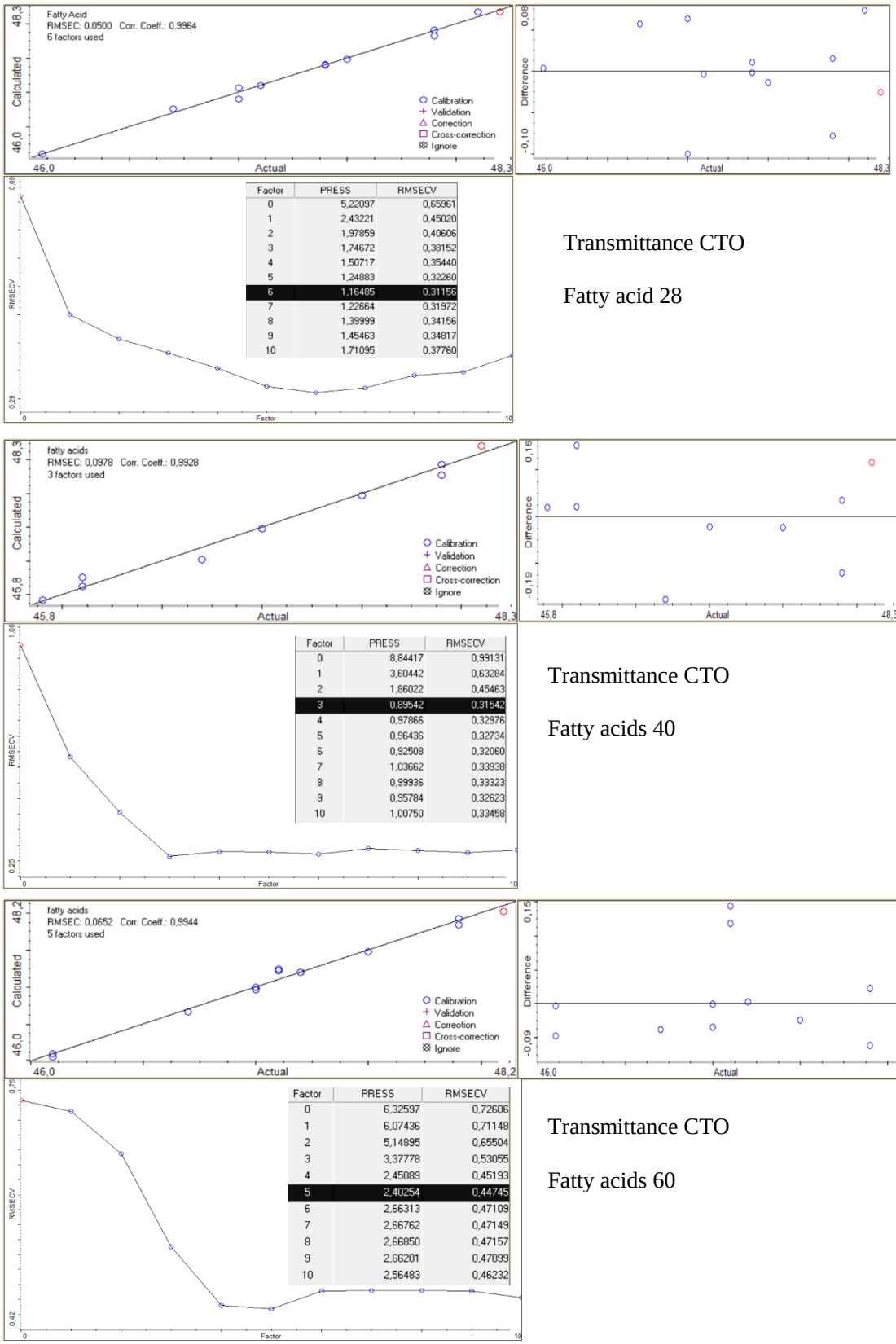
Factor	PRESS	RMSECV
0	12,70685	1,07479
1	6,06661	0,74264
2	3,44794	0,55987
3	2,54844	0,48133
4	2,26033	0,45330
5	2,22070	0,44931
6	2,49582	0,47633
7	2,51610	0,47826
8	2,47827	0,47466
9	2,66784	0,49248
10	2,66753	0,49245



transmittance CTO

Neutral substances 60

Figure 13. Transmittance CTO neutral substances calibration charts



Transmittance CTO

Fatty acid 28

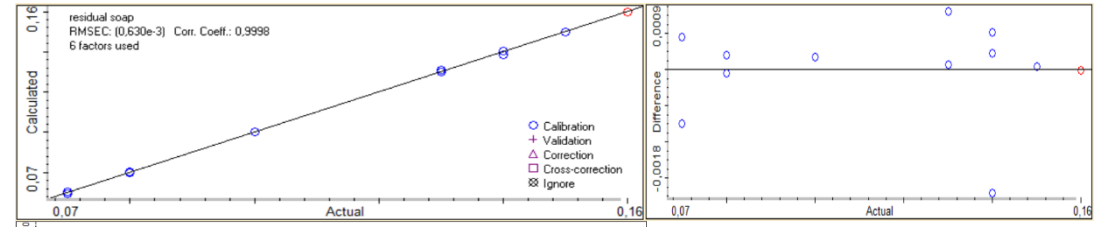
Transmittance CTO

Fatty acids 40

Transmittance CTO

Fatty acids 60

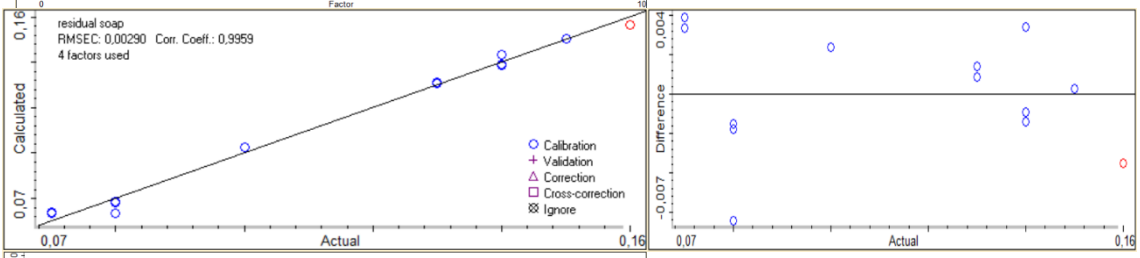
Figure 14. Transmittance CTO fatty acids calibration charts



Factor	PRESS	RMSECV
0	0,01745	0,03664
1	0,00280	0,01467
2	0,00180	0,01178
3	0,00135	0,01020
4	0,00090	0,00830
5	0,00077	0,00769
6	0,00058	0,00666
7	0,00061	0,00684
8	0,00061	0,00687
9	0,00060	0,00680
10	0,00053	0,00640

Transmittance CTO

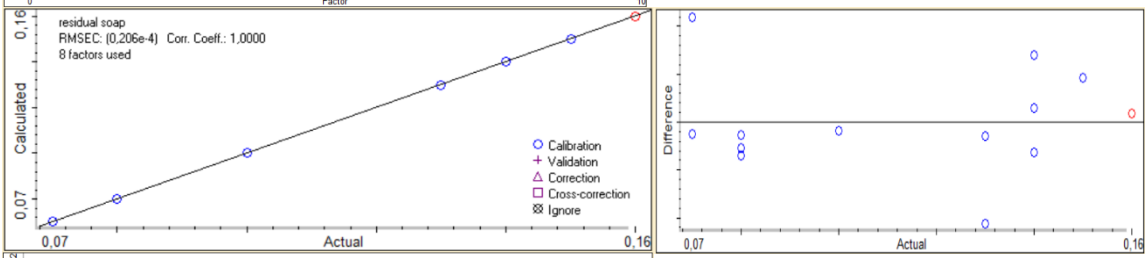
Residual soap 28



Factor	PRESS	RMSECV
0	0,01959	0,03741
1	0,00295	0,01451
2	0,00238	0,01303
3	0,00144	0,01014
4	0,00109	0,00883
5	0,00109	0,00881
6	0,00117	0,00914
7	0,00133	0,00976
8	0,00216	0,01241
9	0,00233	0,01289
10	0,00235	0,01297

Transmittance CTO

Residual soap 40

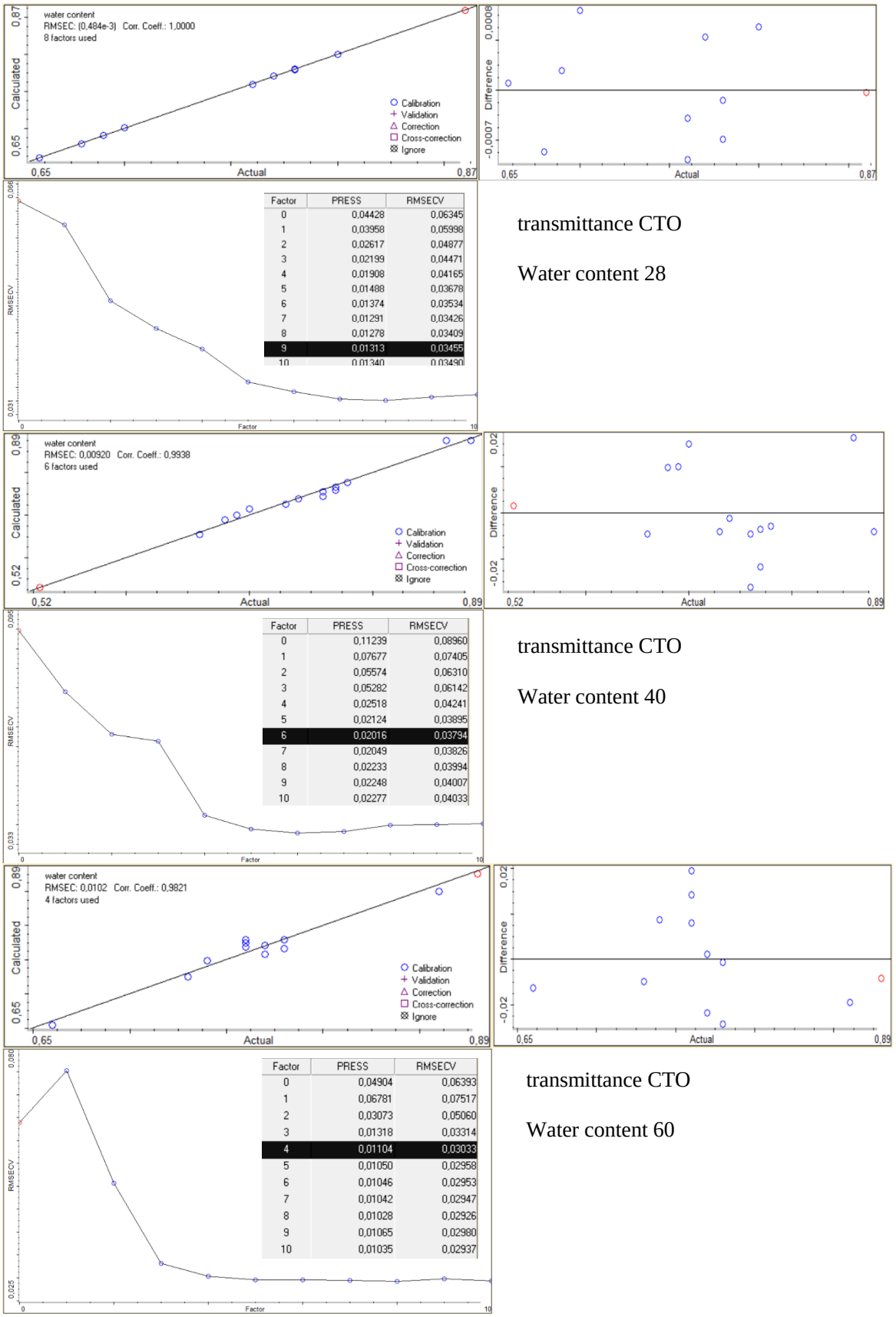


Factor	PRESS	RMSECV
0	0,01959	0,03882
1	0,00276	0,01456
2	0,00141	0,01040
3	0,00080	0,00784
4	0,00072	0,00743
5	0,00067	0,00716
6	0,00065	0,00709
7	0,00065	0,00707
8	0,00064	0,00704
9	0,00064	0,00704
10	0,00064	0,00700

Transmittance CTO

Residual soap 60

Figure 15. Transmittance CTO residual soap calibration charts



transmittance CTO

Water content 28

transmittance CTO

Water content 40

transmittance CTO

Water content 60

Figure 16. Transmittance CTO water content calibration charts.

Calibration charts for transmittance PTO

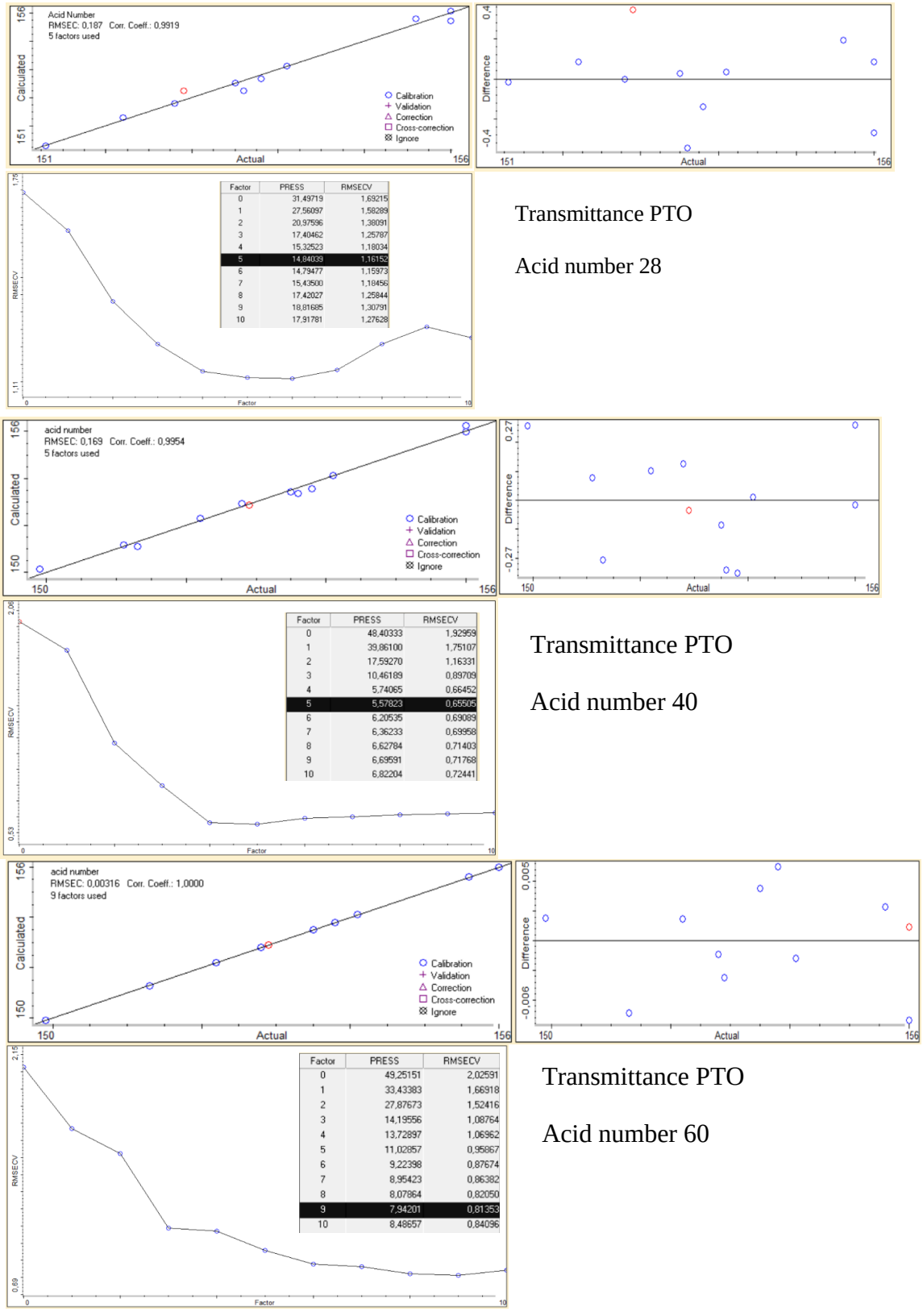
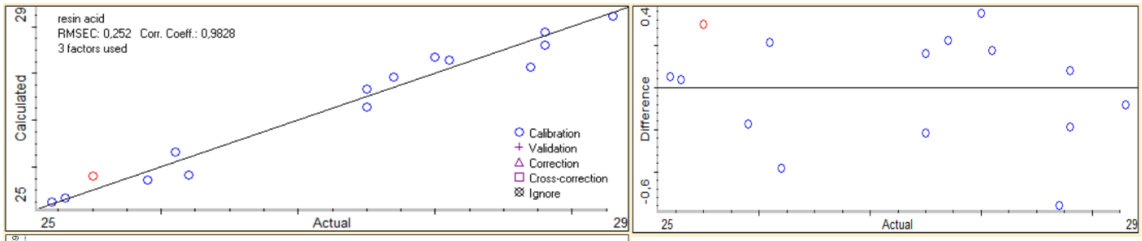
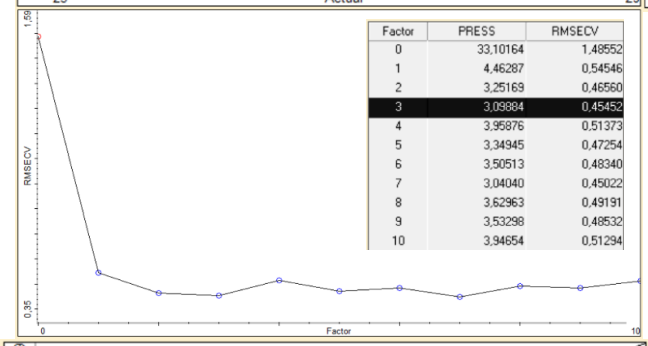


Figure 17. Transmittance PTO acid number calibration charts

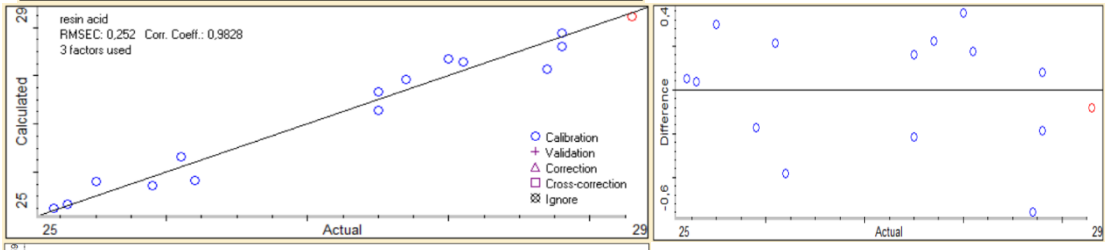


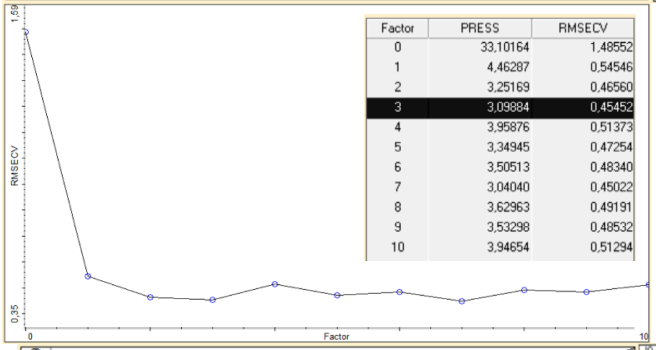


Factor	PRESS	RMSECV
0	33,10164	1,48552
1	4,46287	0,54546
2	3,25169	0,46560
3	3,09884	0,45452
4	3,95876	0,51373
5	3,34945	0,47254
6	3,50513	0,48340
7	3,04040	0,45022
8	3,62963	0,49191
9	3,53298	0,48532
10	3,94654	0,51294

transmittance PTO

Resin acid 28

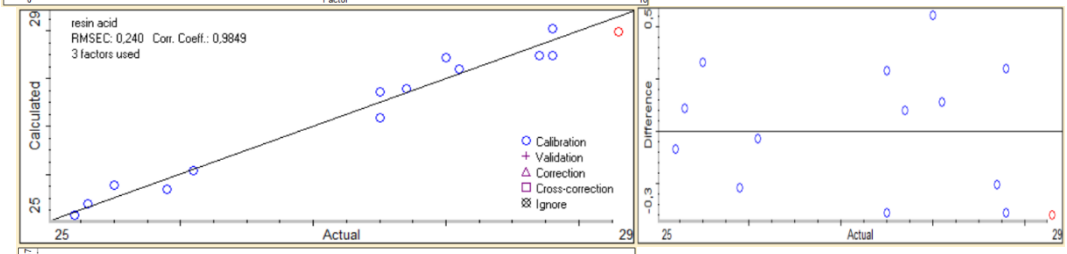


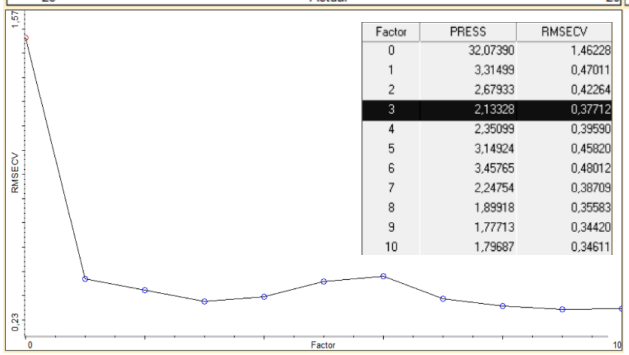


Factor	PRESS	RMSECV
0	33,10164	1,48552
1	4,46287	0,54546
2	3,25169	0,46560
3	3,09884	0,45452
4	3,95876	0,51373
5	3,34945	0,47254
6	3,50513	0,48340
7	3,04040	0,45022
8	3,62963	0,49191
9	3,53298	0,48532
10	3,94654	0,51294

Transmittance PTO

Resin acid 40



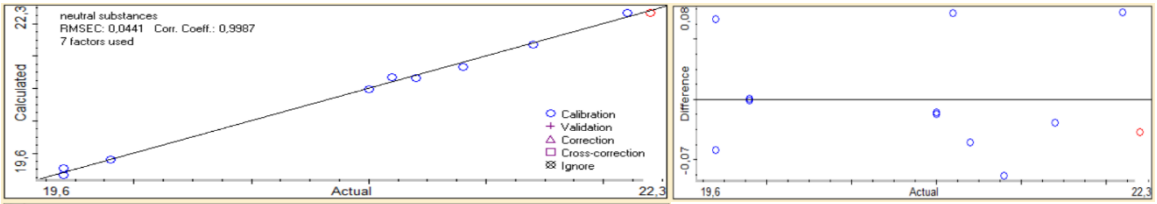


Factor	PRESS	RMSECV
0	32,07390	1,46228
1	3,31499	0,47011
2	2,67933	0,42264
3	2,13328	0,37712
4	2,35099	0,39590
5	3,14924	0,45820
6	3,45765	0,48012
7	2,24754	0,38709
8	1,89918	0,35583
9	1,77713	0,34420
10	1,79687	0,34611

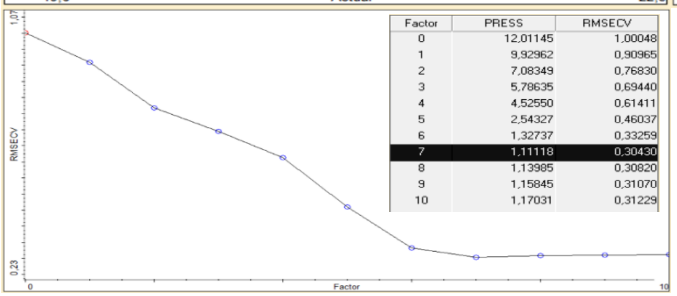
Transmittance PTO

Resin acid 60

Figure 18. Transmittance PTO resin acid calibration charts

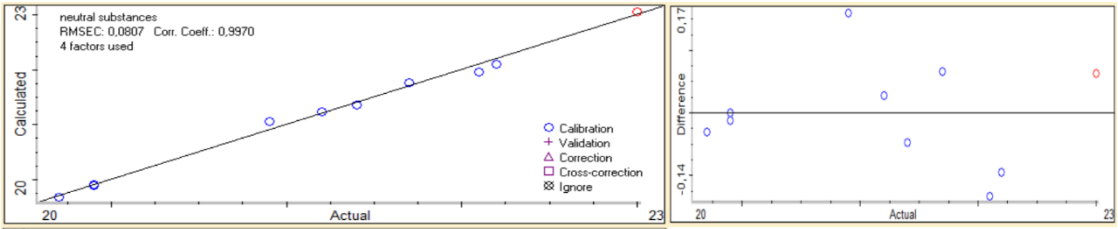


Factor	PRESS	RMSECV
0	12,01145	1,00048
1	9,92962	0,90965
2	7,08349	0,76830
3	5,78635	0,69440
4	4,52550	0,61411
5	2,54327	0,46037
6	1,32737	0,33259
7	1,11118	0,30430
8	1,13985	0,30820
9	1,15845	0,31070
10	1,17031	0,31229

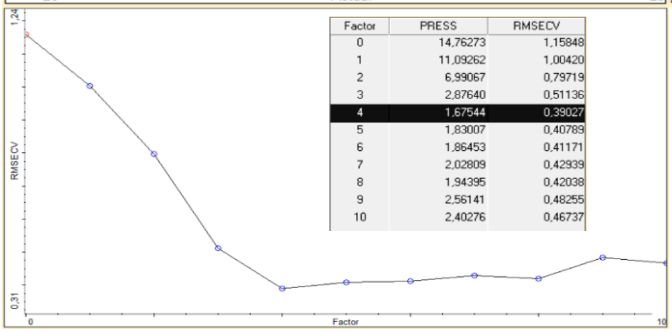


Transmittance PTO

Neutral substances 28

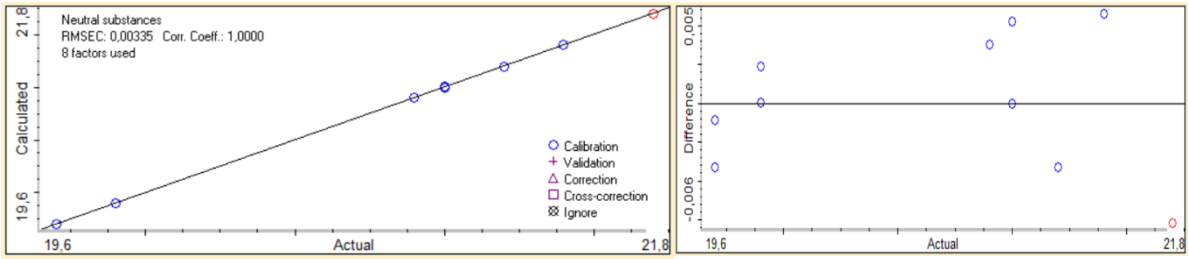


Factor	PRESS	RMSECV
0	14,76273	1,15848
1	11,09262	1,00420
2	6,99067	0,79719
3	2,87640	0,51136
4	1,67544	0,39027
5	1,83007	0,40789
6	1,86453	0,41171
7	2,02809	0,42939
8	1,94395	0,42038
9	2,56141	0,48255
10	2,40276	0,46737

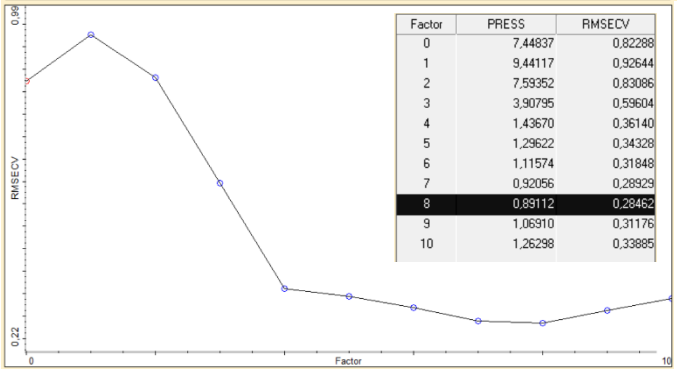


Transmittance PTO

Neutral substances 40



Factor	PRESS	RMSECV
0	7,44837	0,82288
1	9,44117	0,92644
2	7,59352	0,83086
3	3,90795	0,59604
4	1,43670	0,36140
5	1,29622	0,34328
6	1,11574	0,31848
7	0,92056	0,28929
8	0,89112	0,28462
9	1,06910	0,31176
10	1,26298	0,33885



Transmittance PTO

Neutral substances 60

Figure 19. Transmittance PTO neutral substances calibration charts.

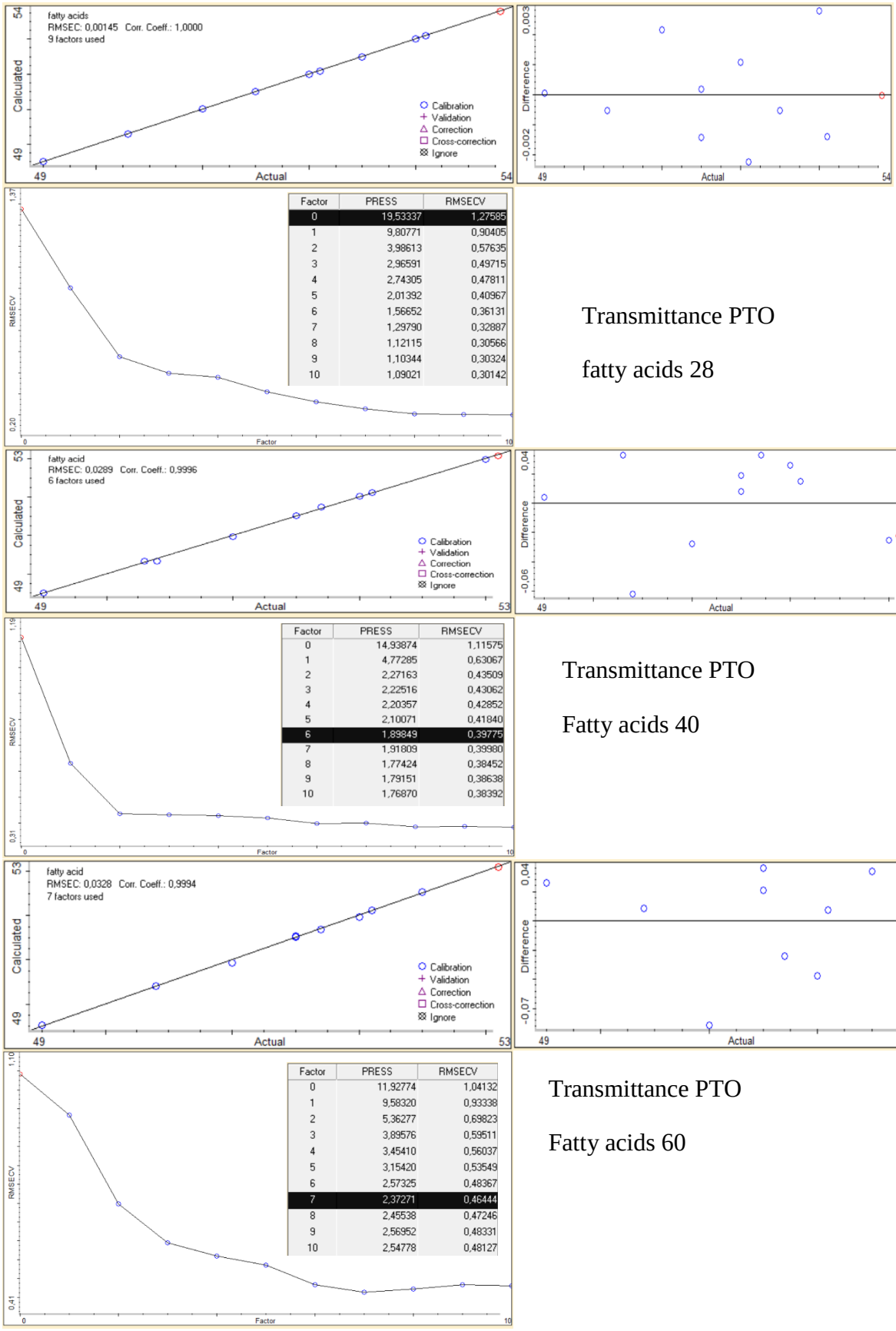


Figure 20. Transmittance PTO fatty acids calibration charts.

APPENDIX II Validation results

Reflectance CTO validation results

Table I. Reflectance CTO resin acid validation results.

CTO validation Resin acid, W-%				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-03328	15.1	15.53	-2.85
2	20-03329	15.2	15.96	-5.00
3	20-03330	14.7	16.51	-12.31
4	20-03331	15.4	15.67	-1.75
5	20-03332	14.8	16.03	-8.31
6	20-03333	14.6	14.43	1.16
7	20-03334	14.3	15.29	-6.92
8	20-03335	3.7	4.92	-32.97
9	20-03336	3.4	4.39	-29.12
10	20-03340	15.1	15.27	-1.13
11	20-03341	14.4	16.04	-11.39
12	20-03342	14.1	14.99	-6.31
13	20-03643	15.0	14.6	2.67
14	20-03916	26.9	29.74	-10.56
15	20-03917	10.6	10.68	-0.75
16	20-03918	22.3	22.51	-0.94
17	20-03930	19.1	22.05	-15.45
18	20-03931	24.8	26.41	-6.49
19	20-04061	4.0	5.96	-49.00
20	20-04062	14.6	17.58	-20.41
21	20-04149	27.5	24.28	11.71
22	20-04150	20.4	21.5	-5.39
23	20-04213	22.4	22.78	-1.70
24	20-04214	23.6	24.32	-3.05
25	20-04216	27.0	29.14	-7.93
26	20-04217	25.3	21.39	15.45
27	20-04218	19.1	22.17	-16.07
28	20-04219	26.3	28.05	-6.65
29	20-04220	23.1	24.43	-5.76
30	20-04221	24.4	25.58	-4.84
31	20-11111	23.7	22.69	4.26
32	20-11112	11.7	10.77	7.95
33	20-11114	26.9	27.08	-0.67
34	20-11540	24.1	21.29	11.66
35	20-11541	25.4	28.36	-11.65
36	20-11542	24.5	24.87	-1.51
37	20-11545	13.1	12.54	4.27

38	20-11546	21.1	21.55	-2.13
39	20-11548	14.8	15.46	-4.46
40	20-11801	28.60	27.09	5.28
41	20-11803	12.10	10.98	9.26
42	20-11811	20.10	20.78	-3.38
43	20-11812	20.00	15.97	20.15
44	20-11813	14.30	10.85	24.13
45	20-11814	35.30	34.03	3.60
46	20-11816	20.3	19.52	3.84
47	20-A00515	23.3	25.75	-10.52
48	20-A01108	25.20	25.07	0.52
49	20-A01145	27.70	28.82	-4.04

Table II. reflectance CTO acid number validation results

CTO validation acid number, mg/KOH/g				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-03328	121.1	131.06	-8.22
2	20-03329	121.0	126.17	-4.27
3	20-03330	120.9	127.81	-5.72
4	20-03331	123.3	126.96	-2.97
5	20-03332	121.6	131.15	-7.85
6	20-03333	121.5	129.46	-6.55
7	20-03334	120.8	130.31	-7.87
8	20-03335	101.9	126.97	-24.60
9	20-03336	105.6	125.63	-18.97
10	20-03340	121.7	124.62	-2.40
11	20-03341	121.7	130.90	-7.56
12	20-03342	122.4	130.89	-6.94
13	20-03643	112.6	128.69	-14.29
14	20-03916	148.3	140.56	5.22
15	20-03917	130.4	125.41	3.83
16	20-03918	136.6	132.69	2.86
17	20-03930	122.7	138.81	-13.13
18	20-03931	132.2	139.14	-5.25
19	20-04061	103.6	131.08	-26.53
20	20-04062	110.9	127.47	-14.94
21	20-04149	136.0	131.81	3.08
22	20-04150	133.4	136.31	-2.18
23	20-04213	134.3	130.28	2.99
24	20-04214	129.1	137.82	-6.75
25	20-04216	149.2	137.81	7.63
26	20-04217	137.7	124.85	9.33
27	20-04218	121.6	126.99	-4.43

28	20-04219	139.3	138.37	0.67
29	20-04220	132.7	139.88	-5.41
30	20-04221	140.3	134.6	4.06
31	20-11111	138.9	141.3	-1.73
32	20-11112	131.3	127.44	2.94
33	20-11114	139.1	140.7	-1.15
34	20-11540	137.1	134.74	1.72
35	20-11541	136.8	140.41	-2.64
36	20-11542	133.6	135.64	-1.53
37	20-11545	133.3	127.89	4.06
38	20-11546	134.4	132.51	1.41
39	20-11548	119.3	133.02	-11.50
40	20-11801	142.80	133.12	6.78
41	20-11803	115.50	131.30	-13.68
42	20-11811	135.80	132.42	2.49
43	20-11812	140.40	127.72	9.03
44	20-11813	136.70	127.57	6.68
45	20-11814	154.30	144.19	6.55
46	20-11816	134.7	134	0.52
47	20-A00515	129.8	135.26	-4.21
48	20-A01108	131.90	135.62	-2.82
49	20-A01145	138.70	140.56	-1.34

Table III. reflectance CTO neutral substances validation results.

CTO validation neutral substances, W-%				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-03328	38.1	39.06	-2.52
2	20-03329	38.2	38.69	-1.28
3	20-03330	38.2	38.69	-1.28
4	20-03331	37.0	39.39	-6.46
5	20-03332	37.9	39.79	-4.99
6	20-03333	38.0	40.44	-6.42
7	20-03334	38.4	38.54	-0.36
8	20-03335	48.6	40.81	16.03
9	20-03336	46.7	41.97	10.13
10	20-03340	37.8	39.26	-3.86
11	20-03341	37.9	35.99	5.04
12	20-03342	37.6	38.09	-1.30
13	20-03643	42.4	37.23	12.19
14	20-03916	23.7	29.68	-25.23
15	20-03917	33.7	33.7	0.00

16	20-03918	29.9	31.62	-5.75
17	20-03930	37.1	36.12	2.64
18	20-03931	31.9	30.55	4.23
19	20-04061	47.7	39	18.24
20	20-04062	43.3	39.54	8.68
21	20-04149	29.8	34.89	-17.08
22	20-04150	31.6	29.27	7.37
23	20-04213	31.0	31.91	-2.94
24	20-04214	33.5	32.82	2.03
25	20-04216	23.2	29.81	-28.49
26	20-04217	29.1	32.84	-12.85
27	20-04218	37.6	37.55	0.13
28	20-04219	28.2	31.84	-12.91
29	20-04220	31.8	34.62	-8.87
30	20-04221	27.9	28.56	-2.37
31	20-11111	28.6	31.83	-11.29
32	20-11112	33.2	34.42	-3.67
33	20-11114	28.3	31.12	-9.96
34	20-11540	29.5	30.48	-3.32
35	20-11541	29.5	30.39	-3.02
36	20-11542	31.2	30.15	3.37
37	20-11545	32.1	30.46	5.11
38	20-11546	31	29.13	6.03
39	20-11548	39	32.8	15.90
40	20-11801	26.30	31.15	-18.44
41	20-11803	41.10	37.78	8.08
42	20-11811	30.40	32.94	-8.36
43	20-11812	28.10	31.71	-12.85
44	20-11813	30.30	32.35	-6.77
45	20-11814	20.10	26.84	-33.53
46	20-11816	30.9	29.82	3.50
47	20-A00515	33.2	28.67	13.64
48	20-A01108	32.00	29.93	6.47
49	20-A01145	28.40	32.59	-14.75

Table. IV. Reflectance CTO fatty acids validation results.

CTO validation fatty acids, W-%				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-03328	46.8	48.78	-4.23
2	20-03329	46.6	48.62	-4.33
3	20-03330	47.1	49.23	-4.52
4	20-03331	47.6	48.90	-2.73
5	20-03332	47.3	48.10	-1.69

6	20-03333	47.4	48.82	-3.00
7	20-03334	47.3	49.09	-3.78
8	20-03335	47.7	51.20	-7.34
9	20-03336	49.9	52.79	-5.79
10	20-03340	47.1	48.55	-3.08
11	20-03341	47.7	48.42	-1.51
12	20-03342	48.3	48.72	-0.87
13	20-03643	42.6	48.02	-12.72
14	20-03916	49.4	46.41	6.05
15	20-03917	55.7	54.74	1.72
16	20-03918	47.8	47.96	-0.33
17	20-03930	43.8	45.66	-4.25
18	20-03931	43.3	45.45	-4.97
19	20-04061	48.3	52.33	-8.34
20	20-04062	42.1	48.57	-15.37
21	20-04149	42.7	47.24	-10.63
22	20-04150	48.0	48.93	-1.94
23	20-04213	46.6	46.43	0.36
24	20-04214	42.9	46.04	-7.32
25	20-04216	49.8	45.55	8.53
26	20-04217	45.6	50.09	-9.85
27	20-04218	43.3	46.9	-8.31
28	20-04219	45.5	45.92	-0.92
29	20-04220	45.1	48.06	-6.56
30	20-04221	47.7	45.63	4.34
31	20-11111	47.7	46.64	2.22
32	20-11112	55.1	52.3	5.08
33	20-11114	44.8	47.14	-5.22
34	20-11540	46.4	49.43	-6.53
35	20-11541	45.1	46.26	-2.57
36	20-11542	44.3	46.13	-4.13
37	20-11545	54.8	52.6	4.01
38	20-11546	47.9	48.71	-1.69
39	20-11548	46.2	49.36	-6.84
40	20-11801	45.10	46.00	-2.00
41	20-11803	46.80	49.52	-5.81
42	20-11811	49.50	47.40	4.24
43	20-11812	51.90	50.60	2.50
44	20-11813	55.40	53.67	3.12
45	20-11814	44.60	45.16	-1.26
46	20-11816	48.8	48.27	1.09
47	20-A00515	43.5	45.74	-5.15
48	20-A01108	42.80	46.02	-7.52
49	20-A01145	43.90	44.61	-1.62

Table V. Reflectance CTO residual soap validation results.

CTO validation residual soap				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-03328	0.06	0.08	-33.33
2	20-03329	0.13	0.06	53.85
3	20-03330	0.09	-0.01	111.11
4	20-03331	0.11	0.05	54.55
5	20-03332	0.11	0.04	63.64
6	20-03333	0.12	0.00	100.00
7	20-03334	0.05	0.07	-40.00
8	20-03335	0.14	0.16	-14.29
9	20-03336	0.19	0.08	57.89
10	20-03340	0.02	0.07	-250.00
11	20-03341	0.02	0.04	-100.00
12	20-03342	0.02	0.09	-350.00
13	20-03643	0.01	-0.03	400.00
14	20-03916	0.17	0.14	17.65
15	20-03917	0.17	0.2	-17.65
16	20-03918	0.12	0.14	-16.67
17	20-03930	0.32	-0.04	112.50
18	20-03931	1.32	0.19	85.61
19	20-04061	0.12	0.04	66.67
20	20-04062	0.08	-0.02	125.00
21	20-04149	0.44	0.03	93.18
22	20-04150	0.64	0.2	68.75
23	20-04213	0.10	0.14	-40.00
24	20-04214	1.06	0.15	85.85
25	20-04216	0.20	0.13	35.00
26	20-04217	0.21	0.08	61.90
27	20-04218	0.32	0.22	31.25
28	20-04219	0.00	0.05	-
29	20-04220	0.79	0.17	78.48
30	20-04221	0.04	0.14	-250.00
31	20-11111	0.06	0.19	-216.67
32	20-11112	0.11	0.07	36.36
33	20-11114	0.19	0.17	10.53
34	20-11540	0.13	0.15	-15.38
35	20-11541	0.22	0.15	31.82
36	20-11542	0	0.11	-
37	20-11545	0.07	0.17	-142.86
38	20-11546	0.15	0.06	60.00
39	20-11548	0.04	0.07	-75.00
40	20-11801	0.02	0.12	-500.00
41	20-11803	0.04	0.04	0.00

42	20-11811	0.27	0.22	18.52
43	20-11812	0.20	0.08	60.00
44	20-11813	0.12	0.25	-108.33
45	20-11814	0.10	0.21	-110.00
46	20-11816	0.47	0.12	74.47
47	20-A00515	0.00	0.08	-
48	20-A01108	0.01	0.15	-1400.00
49	20-A01145	0.11	0.15	-36.36

CTO validation water content,W-%				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-03328	1.99	2.07	-4.02
2	20-03329	1.79	1.89	-5.59
3	20-03330	3.45	3.80	-10.14
4	20-03331	2.70	2.59	4.07
5	20-03332	1.69	1.81	-7.10
6	20-03333	1.76	1.73	1.70
7	20-03334	1.77	1.90	-7.34
8	20-03335	2.29	1.77	22.71
9	20-03336	1.71	1.80	-5.26
10	20-03340	1.80	1.80	0.00
11	20-03341	1.72	1.71	0.58
12	20-03342	1.84	1.80	2.17
13	20-03643	1.61	1.73	-7.45
14	20-03916	0.97	1.21	-24.74
15	20-03917	0.33	-0.2	160.61
16	20-03918	1.54	1.82	-18.18
17	20-03930	0.93	1.28	-37.63
18	20-03931	1.21	1.45	-19.83
19	20-04061	1.86	1.69	9.14
20	20-04062	1.27	1.22	3.94
21	20-04149	5.96	6.48	-8.72
22	20-04150	0.32	0.07	78.13
23	20-04213	1.41	1.71	-21.28
24	20-04214	1.50	1.66	-10.67
25	20-04216	0.94	1.02	-8.51
26	20-04217	0.38	0.37	2.63
27	20-04218	0.93	0.86	7.53
28	20-04219	1.74	1.82	-4.60
29	20-04220	0.36	0.18	50.00
30	20-04221	0.93	1.39	-49.46
31	20-11111	0.97	1.35	-39.18

32	20-11112	0.57	0.36	36.84
33	20-11114	0.7	1.11	-58.57
34	20-11540	0.28	0.28	0.00
35	20-11541	0.63	0.8	-26.98
36	20-11542	1.53	1.51	1.31
37	20-11545	0.57	0.61	-7.02
38	20-11546	1.03	1.15	-11.65
39	20-11548	0.91	0.81	10.99
40	20-11801	1.05	1.30	-23.81
41	20-11803	1.01	0.74	26.73
42	20-11811	0.79	0.96	-21.52
43	20-11812	0.45	0.15	66.67
44	20-11813	0.43	0.06	86.05
45	20-11814	0.65	0.16	75.38
46	20-11816	1.11	0.75	32.43
47	20-A00515	0.91	1	-9.89
48	20-A01108	0.95	0.61	35.79
49	20-A01145	0.84	0.78	7.14

Reflectance PTO validation results

Table VI. Reflectance PTO resin acid validation results.

PTO validation Resin acid, W-%				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-10954	28.1	25.97	7.580
2	20-12099	24.1	26.01	-7.925
3	20-12068	26.5	26.66	-0.604
4	20-11981	26.2	26.33	-0.496
5	20-11903	25.8	26.95	-4.457
6	20-11830	25.5	26.41	-3.569
7	20-11631	26.2	26.39	-0.725
8	20-11607	25.8	26.33	-2.054
9	20-11588	25.7	26.18	-1.868
10	20-11565	25.8	26.5	-2.713
11	20-11497	26.1	26.64	-2.069
12	20-11434	25.9	27.07	-4.517
13	20-11350	26.2	27.28	-4.122
14	20-11165	25.5	26.44	-3.686

PTO validation Acid number, mg/KOH/g				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %

1	20-10954	155.6	153.24	1.517
2	20-12099	155.2	152.17	1.952
3	20-12068	158.1	153.43	2.954
4	20-11981	157.8	153.3	2.852
5	20-11903	157.6	152.62	3.160
6	20-11830	152.9	153.03	-0.085
7	20-11631	151.3	152.29	-0.654
8	20-11607	152.7	153.8	-0.720
9	20-11588	152.5	152	0.328
10	20-11565	151.9	152.73	-0.546
11	20-11497	152.9	152.76	0.092
12	20-11434	153.6	154.76	-0.755
13	20-11350	152.2	153.55	-0.887
14	20-11165	144.7	153.75	-6.254

Table VII. reflectance PTO neutral substances validation results.

PTO validation Neutral substances, W-%				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-10954	19.9	21.06	-5.829
2	20-12099	20.4	21.95	-7.598
3	20-12068	18.8	21.07	-12.074
4	20-11981	18.9	21.12	-11.746
5	20-11903	19.1	21.05	-10.209
6	20-11830	21.5	21	2.326
7	20-11631	22.2	21.84	1.622
8	20-11607	21.5	21.18	1.488
9	20-11588	21.7	22.17	-2.166
10	20-11565	21.9	21.4	2.283
11	20-11497	21.4	21.32	0.374
12	20-11434	21.1	20.15	4.502
13	20-11350	21.7	20.64	4.885
14	20-11165	25.6	20.64	19.375

Table. VII. Reflectance PTO fatty acid validation results.

PTO validation Fatty acid, W-%				
Sample number	Sample ID	Laboratory measurement	NIR measurement	Difference %
1	20-10954	52	51.41	1.135
2	20-12099	55.5	51.52	7.171
3	20-12068	54.7	51.68	5.521
4	20-11981	54.9	51.6	6.011
5	20-11903	55.1	51.21	7.060

6	20-11830	53.0	51.56	2.717
7	20-11631	51.6	51.97	-0.717
8	20-11607	52.7	52.7	0.000
9	20-11588	52.6	51.65	1.806
10	20-11565	52.3	51.47	1.587
11	20-11497	52.5	51.84	1.257
12	20-11434	53	51.44	2.943
13	20-11350	52.1	51.52	1.113
14	20-11165	48.9	51.7	-5.726

Table VIII. Transmittance CTO resin acid and acid number validation results.

Transmittance CTO resin acid and acid number validation results.							
Sample number	Sample ID-temperature	Resin acid, W-%			Acid number, mg/KOH/g		
		Laboratory measurement	NIR measurement	Difference, %	Laboratory measurement	NIR measurement	Difference, %
1	20-10303-28	25.6	26.76	-4.52	138.5	139.430	-0.67
2	20-10303-40	25.6	25.14	1.81	138.5	-	-
3	20-10303-60	25.6	25.65	-0.18	138.5	141.027	-1.83
16	20-10464-28	25.6	27.50	-7.42	141.5	140.3	0.85
17	20-10464-40	25.6	25.60	0.00	141.5	139.12	1.68
18	20-10464-60	25.6	25.69	-0.35	141.5	141.28	0.16
19	20-10539-28	26.5	28.47	-7.43	141.4	140.48	0.65
20	20-10539-40	26.5	25.70	3.02	141.4	139.96	1.02
21	20-10539-60	26.5	25.39	4.19	141.4	142.41	-0.71
4	20-11568-28	23.4	25.69	-9.80	136.7	137.04	-0.25
5	20-11568-40	23.4	23.84	-1.89	136.7	-	-
6	20-11568-60	23.4	24.43	-4.39	136.7	138.54	-1.35
25	20-11632-28	23.8	27.80	-16.81	136.3	135.9	0.29
26	20-11632-40	23.8	23.79	0.042	136.3	137.19	-0.65
27	20-11632-60	23.8	24.62	-3.45	136.3	136.74	-0.32
7	20-11832-28	23.2	24.05	-3.68	137.5	136.68	0.60
8	20-11832-40	23.2	23.39	-0.82	137.5	-	-
9	20-11832-60	23.2	23.82	-2.69	137.5	137.16	0.25
10	20-11904-28	23.1	23.58	-2.09	138.9	136.62	1.64
11	20-11904-40	23.1	23.50	-1.74	138.9	-	-
12	20-11904-60	23.1	23.85	-3.26	138.9	137.70	0.87
13	20-11984-28	23.3	24.70	-6.00	137.7	136.55	0.84
14	20-11984-40	23.3	23.49	-0.83	137.7	-	-
15	20-11984-60	23.3	24.00	-3.01	137.7	137.96	-0.19
22	20-12069-28	23.5	24.93	-6.09	137.7	135.64	1.50
23	20-12069-40	23.5	23.45	0.21	137.7	136.01	1.23
24	20-12069-60	23.5	23.25	1.06	137.7	137.76	-0.04

Table IX. Transmittance CTO neutral substances and fatty acids validation results

Transmittance CTO neutral substances and fatty acids validation results.							
Sample number	Sample ID-temperature	Neutral substances, W-%			Fatty acids, W-%		
		Laboratory measurement	NIR measurement	Difference, %	Laboratory measurement	NIR measurement	Difference, %
1	20-10303-28	28.7	28.69	0.03	45.70	47.30	-3.49
2	20-10303-40	28.7	28.69	0.04	45.70	47.50	-3.93
3	20-10303-60	28.7	27.57	3.93	45.70	47.22	-3.33
16	20-10464-28	27.2	27.59	-1.43	47.20	47.25	-0.11
17	20-10464-40	27.2	28.53	-4.89	47.20	46.99	0.44
18	20-10464-60	27.2	27.16	0.15	47.20	47.30	-0.21
19	20-10539-28	27.2	27.74	-1.99	46.30	46.81	-1.10
20	20-10539-40	27.2	28.28	-3.97	46.30	46.74	-0.95
21	20-10539-60	27.2	27.46	-0.96	46.30	47.88	-3.41
4	20-11568-28	29.7	29.69	0.04	46.90	47.73	-1.78
5	20-11568-40	29.7	29.36	1.15	46.90	47.99	-2.32
6	20-11568-60	29.7	28.27	4.80	46.90	47.29	-0.84
25	20-11632-28	29.9	29.76	0.47	46.30	47.61	-2.83
26	20-11632-40	29.9	29.43	1.57	46.30	47.46	-2.51
27	20-11632-60	29.9	28.29	5.38	46.30	46.18	0.26
7	20-11832-28	29.4	29.44	-0.15	47.40	47.41	-0.02
8	20-11832-40	29.4	29.35	0.18	47.40	47.92	-1.09
9	20-11832-60	29.4	28.57	2.81	47.40	47.15	0.54
10	20-11904-28	28.7	28.65	0.16	48.20	48.18	0.05
11	20-11904-40	28.7	29.31	-2.11	48.20	48.40	-0.41
12	20-11904-60	28.7	28.46	0.84	48.20	47.63	1.18
13	20-11984-28	29.3	29.27	0.09	47.40	47.40	0.00
14	20-11984-40	29.3	29.50	-0.69	47.40	47.76	-0.77
15	20-11984-60	29.3	28.88	1.43	47.40	47.18	0.46
22	20-12069-28	29.2	29.87	-2.29	47.30	48.36	-2.24
23	20-12069-40	29.2	29.80	-2.06	47.30	47.61	-0.66
24	20-12069-60	29.2	28.89	1.06	47.30	47.69	-0.82

Table X. Transmittance CTO residual soap and water content validation results.

Transmittance CTO resin acid and acid number validation results.							
Sample number	Sample ID-temperature	Residual soap			Water content, W-%		
		Laboratory measurement	NIR measurement	Difference, %	Laboratory measurement	NIR measurement	Difference, %
1	20-10303-28	0.14	0.15	-7.14	0.74	0.79	-6.76
2	20-10303-40	0.14	0.14	-2.86	0.74	0.74	0.14
3	20-10303-60	0.14	0.15	-5.71	0.74	0.75	-1.22
16	20-10464-28	0.14	0.15	-7.14	0.73	0.74	-1.37
17	20-10464-40	0.14	0.15	-7.14	0.73	0.68	6.85
18	20-10464-60	0.14	0.15	-7.14	0.73	0.69	5.48
19	20-10539-28	0.19	0.16	15.79	0.69	0.77	-11.59
20	20-10539-40	0.19	0.15	21.05	0.69	0.65	5.80
21	20-10539-60	0.19	0.15	21.05	0.69	0.69	0.00
4	20-11568-28	0.10	0.10	0.00	0.73	0.76	-4.11
5	20-11568-40	0.10	0.10	-1.00	0.73	0.73	0.55
6	20-11568-60	0.10	0.09	6.00	0.73	0.73	0.68
25	20-11632-28	0.10	0.10	0.00	0.80	0.68	15.00
26	20-11632-40	0.10	0.10	0.00	0.80	0.69	13.75
27	20-11632-60	0.10	0.08	20.00	0.80	0.75	6.25
7	20-11832-28	0.11	0.08	27.27	0.69	0.81	-17.39
8	20-11832-40	0.11	0.08	31.82	0.69	0.70	-1.45
9	20-11832-60	0.11	0.07	37.27	0.69	0.76	-10.43
10	20-11904-28	0.28	0.08	71.43	0.83	0.82	1.20
11	20-11904-40	0.28	0.08	71.79	0.83	0.62	25.66
12	20-11904-60	0.28	0.07	75.36	0.83	0.65	22.17
13	20-11984-28	0.11	0.08	27.27	0.77	0.79	-2.60
14	20-11984-40	0.11	0.08	30.91	0.77	0.76	1.56
15	20-11984-60	0.11	0.07	34.55	0.77	0.76	1.56
22	20-12069-28	0.11	0.07	36.36	0.73	0.65	10.96
23	20-12069-40	0.11	0.07	36.36	0.73	0.70	4.11
24	20-12069-60	0.11	0.06	45.45	0.73	0.75	-2.74

Table XI. Transmittance PTO resin acid and acid number validation results.

Transmittance PTO resin acid and acid number validation results.							
Sample number	Sample ID-temperature	Resin acid, W-%			Acid number, mg/KOH/g		
		Laboratory measurement	NIR measurement	Difference, %	Laboratory measurement	NIR measurement	Difference, %
25	20-10536-P28	28.70	26.81	6.59	154.80	151.60	2.07
26	20-10536-P40	28.70	26.28	8.43	154.80	150.40	2.84
27	20-10536-P60	28.70	28.53	0.59	154.80	153.06	1.12
1	20-11565-P28	25.80	25.40	1.54	151.90	151.36	0.35
2	20-11565-P40	25.80	24.78	3.94	151.90	150.98	0.61
3	20-11565-P60	25.80	25.84	-0.17	151.90	153.21	-0.86
4	20-11588-P28	25.70	25.33	1.42	152.50	151.71	0.52
5	20-11588-P40	25.70	25.92	-0.84	152.50	151.64	0.56
6	20-11588-P60	25.70	25.24	1.78	152.50	153.62	-0.74
7	20-11607-P28	25.80	26.17	-1.45	152.70	154.16	-0.95
8	20-11607-P40	25.80	24.92	3.41	152.70	152.71	0.00
9	20-11607-P60	25.80	26.06	-0.99	152.70	153.12	-0.28
10	20-11631-P28	26.20	26.20	0.01	151.30	152.81	-1.00
11	20-11631-P40	26.20	24.48	6.56	151.30	151.17	0.09
12	20-11631-P60	26.20	26.05	0.59	151.30	151.89	-0.39
16	20-11720-P28	26.00	25.17	3.19	154.20	152.13	1.34
17	20-11720-P40	26.00	23.14	11.00	154.20	146.57	4.95
18	20-11720-P60	26.00	25.51	1.88	154.20	150.98	2.09
13	20-11830-P28	25.50	25.05	1.77	152.90	152.44	0.30
14	20-11830-P40	25.50	24.33	4.59	152.90	151.45	0.95
15	20-11830-P60	25.50	25.70	-0.80	152.90	151.63	0.83
19	20-13021-P28	23.90	22.39	6.32	151.20	150.42	0.52
20	20-13021-P40	23.90	21.18	11.38	151.20	144.03	4.74
21	20-13021-P60	23.90	23.34	2.34	151.20	148.84	1.56
22	20-13105-P28	23.90	22.19	7.15	150.30	150.51	-0.14
23	20-13105-P40	23.90	20.91	12.51	150.30	143.82	4.31
24	20-13105-P60	23.90	23.77	0.54	150.30	149.89	0.27

TableXII. Transmittance PTO neutral substances and fatty acids validation results.

Transmittance PTO neutral substances and fatty acids validation results.							
Sample number	Sample ID-temperature	Neutral substances, W-%			Fatty acids, W-%		
		Laboratory measurement	NIR measurement	Difference, %	Laboratory measurement	NIR measurement	Difference, %
25	20-10536-P28	20.30	20.39	-0.44	51.00	51.07	-0.14
26	20-10536-P40	20.30	22.22	-9.46	51.00	48.69	4.53
27	20-10536-P60	20.30	21.24	-4.63	51.00	52.15	-2.25
1	20-11565-P28	21.90	21.51	1.79	52.30	52.43	-0.26
2	20-11565-P40	21.90	22.66	-3.48	52.30	51.81	0.94
3	20-11565-P60	21.90	21.46	2.03	52.30	52.62	-0.62
4	20-11588-P28	21.70	20.98	3.33	52.60	52.73	-0.24
5	20-11588-P40	21.70	22.74	-4.78	52.60	51.76	1.60
6	20-11588-P60	21.70	21.20	2.31	52.60	52.76	-0.30
7	20-11607-P28	21.50	20.04	6.77	52.70	53.30	-1.14
8	20-11607-P40	21.50	22.55	-4.87	52.70	52.59	0.20
9	20-11607-P60	21.50	21.10	1.86	52.70	53.10	-0.75
10	20-11631-P28	22.20	20.72	6.68	51.60	52.44	-1.63
11	20-11631-P40	22.20	22.67	-2.10	51.60	51.59	0.02
12	20-11631-P60	22.20	21.72	2.18	51.60	52.41	-1.57
16	20-11720-P28	20.80	21.51	-3.41	53.20	54.16	-1.80
17	20-11720-P40	20.80	22.23	-6.88	53.20	51.09	3.97
18	20-11720-P60	20.80	21.24	-2.12	53.20	53.74	-1.02
13	20-11830-P28	21.50	21.04	2.15	53.00	53.79	-1.49
14	20-11830-P40	21.50	21.95	-2.10	53.00	53.14	-0.27
15	20-11830-P60	21.50	21.50	0.01	53.00	53.55	-1.03
19	20-13021-P28	22.40	22.54	-0.63	53.70	56.26	-4.77
20	20-13021-P40	22.40	24.86	-10.98	53.70	51.27	4.53
21	20-13021-P60	22.40	21.29	4.96	53.70	54.10	-0.74
22	20-13105-P28	22.90	22.65	1.09	53.20	56.39	-6.00
23	20-13105-P40	22.90	24.97	-9.04	53.20	52.02	2.22
24	20-13105-P60	22.90	21.09	7.90	53.20	54.62	-2.67