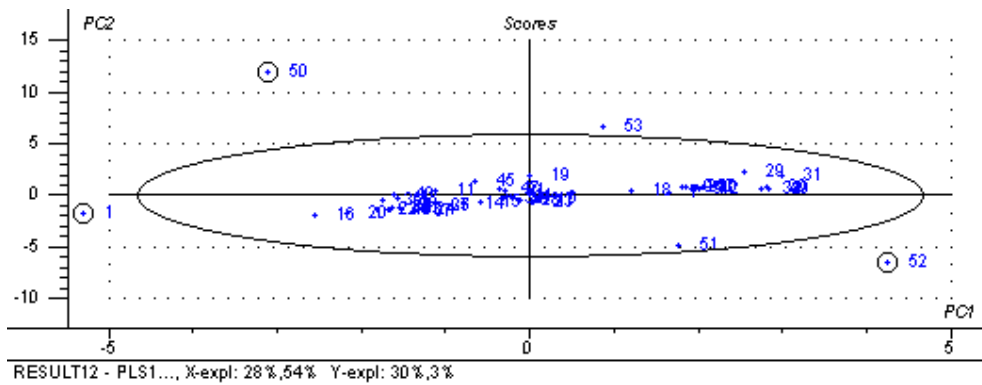# ANALYTICAL TOOL FOR RAPID ANALYSIS OF EDIBLE OIL

MSc THESIS PROJECT

CHEMICAL ENGINEERING
AALBORG UNIVERSITY ESBJERG

ANALYTICAL TOOL FOR RAPID ANALYSIS OF EDIBLE OIL

Supervised by
Lars Petersen Julius & Henrik Juhl.

MASTER THESIS
(MSC CHEMICAL ENGINEERING)
JUNE 2008

Author
RIZWANA SHAIK

ABSTRACT

Foss in Hillerød develops analytical instruments and in this project deals with the development of an analytical instrument for edible oils for rapid analysis purpose by replacing the time consuming conventional methods. In this project three parameters of edible oils i.e. Saponification number, Peroxide value & Free fatty acids are chosen for the analysis.

Different oils including Rape, palm, sunflower, sesame, maize, olive etc are collected and have been used for analysis using conventional methods and further by FTNIR, FTIR. A GC/MS method was developed using capillary column and ion trap technology for analyzing of free fatty acid. The Free fatty acids chosen as a parameter for analysis includes C12, C14, C16, C18, and C18:1, C18:2, C18:3.

These Fatty acids are modelled by developing Partial Least Squares calibration curves. The collected spectral data from both NIR &IR are used for development of multivariate methods using Unscrambler9.2.

PREFACE

The goal of this project is to develop a rapid analysis analytical tool for edible oils. This project material includes all FTIR, FTNIR, GC/MS & wet chemistry results that have been modelled using Unscrambler.

This project work has been carried with a view of analysing different edible oils using OenoFoss, an FTIR instrument which is presently used for wine quality analysis. And in this approach of finding a rapid analytical tool to replace the conventional time consuming methods, FTNIR has also been used to collect the spectra of these edible oils.

The wet chemistry part of this project along with FTNIR work has been undertaken at Aalborg University, Esbjerg. While the FTIR analysis was carried out at FOSS, Hillerod. The wet chemistry analysis which includes GC/MS method development took a lot of time.

The support of the supervisor Henrik Juhl at Foss contributes significantly to this project work and I am thankful to him for his support and discussions.

I would also like to acknowledge Dorte Spangsmark, Linda Birkebæk Madsen & Lisbet Skou laboratory staff at Aalborg University, Esbjerg.

I would as well like to mention in the list of my acknowledgements Henrik Thomsen at Foss who was always available at Foss laboratory to sort out the software issues while this project work was running.

Finally I would like to thank the foremost person who has given me the opportunity of doing this project Dorthe Kjær Pedersen, Foss.

RIZWANA SHAIK   ---------------------------

1. INTRODUCTION

Edible oils play an important role in the body as carriers of essential fatty acids (EFA). EFA are not synthesised in the body but are needed through the diet to maintain the integrity of cell membranes. They are also needed for the synthesis of prostaglandins which have many vital functions to perform in the body.

Based on maintenance of good diet and control health issues like diabetes it has been essential to consume authenticated edible oils. Therefore consumers need meaningful and honest information so that they can make an informed choice of their diet and the foods they purchase.

To meet the quality and composition standards, oil & food industries use certain oil parameters to maintain the quality. Protection against mislabelling and false description through legislation is an important part of food control.

Research to develop new analytical methods is likely to be an ongoing process but ever more difficult task as those who seek to gain financially, find increasingly sophisticated ways of food adulteration[1].

Authenticity of Edible oils and fats: According to UK and European Legislation under sections 14 and 15 of the Food Safety act, it is an offence to sell food that is not of the nature, substance or quality demanded by the consumer or to falsely or misleadingly describes or presents food[2].

Several factors affect the edible oil quality such as agronomic techniques, seasonal conditions, sanitary state of drupes, ripening stage, harvesting and carriage systems, method and duration of storage, and processing technology and it is determined by different analytical methods in order to assess the stability of oil and to avoid possible adulterations. The official methods for the determination of physico-chemical parameters of edible oils, such as acidity and peroxide content are based on titration methods that are time-consuming and laborious. On the other hand, oils and fats are only soluble in organic solvents and because of that classical analytical methods use big amounts of solvents, including chlorinated ones, thus resulting in increased costs and potential health and environmental hazards. Free fatty acid content (acidity value) is one of the most frequently determined quality indices during oil production, storage and marketing and it is often used to classify and/or evaluate oils. It is a measure of the extent to which hydrolysis has liberated fatty acids from their ester linkage with the parent triglyceride molecule. This determination is of a special importance for virgin olive oil since it is the only edible oil which has to be traded without any processing such as neutralization. Other edible oils undergo various processing steps, resulting in low free fatty acid content. European Commission Regulation No. 2568/91 (ECC, 1991) suggests classifying the olive oils according to their acidity value, reported as grams of oleic acid per 100 g of oil. Another important parameter to be considered in oil analysis is the peroxide index. Peroxides are indicators of oxidative

rancidity in foods. They are the so-called primary oxidation products and are used as indicators of oil quality and stability[3].

To meet the legislative demands, official methods are listed in the manual of the American oil chemists' society (AOCS). One such official method includes development of GC/MS for the free fatty acid analysis in oils.

Present traditional methods need to be replaced by rapid analysis method like FTIR and FTNIR. These traditional methods require longer time of analysis as well as a laboratory expert.

Of the different parameters that can be used for the quality analysis of oils, three parameters have been chosen for this project. Of the below mentioned parameters (chemical method) Peroxide value, Saponification number and Free fatty acids have been analyzed in this project by FTNIR and FTIR comparing with its reference analysis.

| MEASURE	Functional groups		Chemical method
Degree of Unsaturation	C=C		Iodine Value
Hydro Peroxides	OOH		Peroxide Value
Chain Length	CH		Saponification Number
Carboxylic Acids	COOH		Free Fatty Acids.
Moisture	OH		Moisture Content
Hydroxyl Groups Mono/Di glycerides	OH		Hydroxyl Number
Type of Unsaturation	Cis C=C, trans C=C, CH, C=C, C=C-C=C		Trans analysis Fatty acid profile.
CARBONYL Compounds	HC=O		Anisidine value.

Table1. IR oil Quality parameters.[4]

This project includes different types of oils and one such kind used is olive oil. Olive oil is an economically important product in the Mediterranean countries as it has a fine aroma and a pleasant taste, as well as it is also known for its health benefits. The quality of olive oil will range from the high quality extra virgin olive oil (EVOO) to the low-quality olive–pomace oil (OPO) (or raw residue oil). EVOO is actually obtained from the fruit of the olive tree (*Olea*

europaea L.) by mechanical press and without application of refining processes. Its acidity cannot be greater than 1%. The high quality of this oil made it as the most expensive type of olive oil. Sometimes this high quality oil is mislabelled or adulterated. And the process of adulteration involves addition of cheaper oils. The most common adulterants that are found in virgin olive oil are refined olive oil, seed oils (such as sunflower, soy, corn and rapeseed oils) and nut oils (such as hazelnut and peanut oils). In some cases, besides the economic fraud, adulteration may cause serious health problems, such as happened in 1981 in the case of the Spanish toxic oil syndrome, which affected about over 20000 peoples.

For the purpose of olive oil authentication more rapid, accurate, reliable and robust analytical methodologies are being requested that can be suitable for routine analysis.

The spectra in figure1 below show the FTIR spectra of olive oil, corn, sunflower and soyabean oils. The fingerprint region especially 1300-1000 cm-1 shows differences in the peak areas and peak heights of the absorption bands at 1163 cm-1 (assigned to -C-O stretching and CH - bending) and at 1118 and 1097 cm-1 (assigned to -C-O stretching).
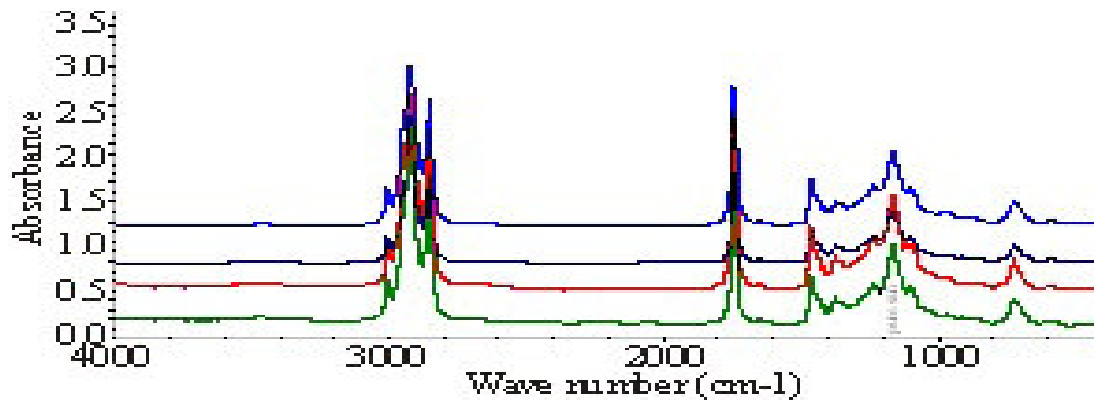
Figure1: FTIR spectra of oils[5]

The spectra in the Figure1 are typical FTIR spectra (4000-400 cm-1) of, pure soybean, corn, olive and sunflower oils observed from top to bottom.

[30]**Foss** is the company that is involved in this project. FOSS was founded in 1956 as N. Foss Electric A/S by Nils Foss. FOSS aim is to provide rapid, reliable and dedicated analytical solutions for routine control of quality and processing of agricultural, food, pharmaceutical and chemical product

Product range and technology

Foss offers a wide range of dedicated analytical solutions that are based on both indirect and reference methods. Solutions are provided for analysis and control throughout the production process, which is from raw material to finished product and from routine analysis to at-line, on-line and in-line process control.

Industries like Dairy, Meat , Wine, Edible oils, Confectionery, Pet food, Pharmaceuticals, Chemicals, Bio fuel, Grain, Beer, Feed and Forage and Milling use FOSS analytical solutions.

Solutions include Near Infrared spectroscopy, Infrared, FIA, X-ray technology etc. The recently developed solution for the quality control of wine is OenoFoss™, which provides information on the spot from just one simple test.

OenoFoss measures the quality parameters of grape under fermentation and wine from a single drop of a sample. Up to seven parameters are measured: sugar, pH, total acid, glucose, fructose, malic acid, ethanol, volatile acid and colour.

The OenoFoss instrument is simple-to-use and does not require the use of chemicals. A point and click software interface allows virtually anyone to use the instrument and record a valid analysis result.

Purpose of the project:

The main objective of the project is to develop rapid edible oils analysis tool based on Infrared spectroscopy. The **FTNIR** Bomem MB160 series is used for the NIR analysis. And OenoFoss that was developed for the wine quality parameter analysis is used for the **FTIR** oil analysis in this project. The instrument has 100μm path length range. And it functions based on the information obtained from two different path lengths i.e.; longer path length and shorter path length. During this project, analysis has been carried with different path length for the parameters – **Saponification number**, **Peroxide value**, **Free fatty acids** of about 50 oils of which few are of different kinds and few are from different brand names. In fact all the oils are different from each other.

The obtained IR & NIR spectra are compared with reference analysis results using chemometrics.

The developed instrument has a cuvette open and close lid and the time of analysis is 18sec for each path length.

Generally the Infrared analysis are carried by first taking the background spectrum using empty vial but for this instrument there is no need of background spectral recording instead it works based on two path length where the shorter path length spectrum is divided from the longer path length.

2. Experimental Approach

The project work was divided into five major parts:
1. Different edible oils collection.
2. Reference Analysis (Wet chemistry including GC/MS) at AAUE.
3. FTNIR using Bomem MB160 at Aalborg University, Esbjerg.
4. FTIR using OenoFoss at Foss, Hillerod.
5. Chemometrics using Unscrambler.

Edible oils collected for this project are placed in the appendices. All the oils are used to determine the three parameters – Peroxide value, Saponification number and free fatty acids

by reference analysis using laboratory procedures and by IR, NIR spectral analysis. All these results are modelled by PLS1 (Partial Least Square) regression.

2.1. Reference Analysis:

Free Fatty acids analysis was done by Gas chromatography/Mass spectrometer (GC/MS) by the synthesis of FAME (fatty acid methyl esters).

GC/MS method development was carried by making calibration curves for quantification purposes using standard fatty acids –lauric (C12) Myristic acid (C14), Palmitic (C16), Stearic (C18), Oleic (C18:1), Linoleic (18:2) and Linolenic acids (18:3).

The GC/MS with capillary column and Ion trap technology of 3800/2000 series was used for this part of work. With the help of calibration curves quantitative analysis of the fatty acids that are present in the edible oils were quantified which are separated by capillary column and MS (Mass spectrometer) by molecular weight. The FAME (fatty acid methyl esterification) synthesis and analysis using GC/MS for free fatty acid determination was done by general laboratory procedure which is a slightly modified procedure of AOAC official Method 969.33.[7.1.1.3. FREE FATTY ACIDS:]

Figure 2: GC/MS 3800/2000 instrument.

Peroxide Value: The peroxide value of the selected edible oils was determined by European Pharmocopoeia method (2.5.5 4th Edition 2001).[7.1.1.2. PEROXIDE VALUE:]

Saponification Number: The analysis of this parameter was done using titration method which is a commonly used practised in the laboratory.[7.1.1.1.Saponification Number:]

2.2 FTIR:

Foss OenoFoss apparatus has been used for infrared analysis, which works based on two different path lengths. So there is no need of background spectral collection for this apparatus.

The OenoFoss apparatus is currently used for the wine analysis and this is the apparatus used in this project for edible oil analysis.

Figure3: OenoFoss instrument

Main Principle:

OenoFoss instrument has a cuvette that opens and locks. Applying Beer's law at two different path lengths:

Absorbance $A = \log I_1/I_2$ --------------1
Where I_1 = longer path length
I_2 = Shorter path length

But $A = \varepsilon lc$ --------------------------2
Where:
A= Absorbance
ε = Absorptivity
l = path length
C= Concentration.
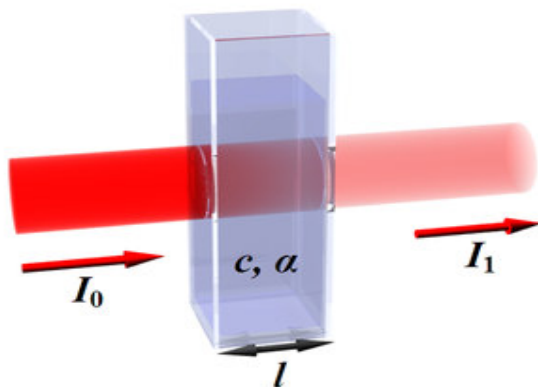
Figure 4: Light passing through Sample for analysis

$\varepsilon lc = \log I_1/I_2$

$I_1 = I_0 \exp(\varepsilon l_1 c)$ ---------------3

l_1= long path length

I_0= intensity of incoming light.

Similarly,

$I_2 = I_0 \exp(\varepsilon l_2 c)$ --------------4

l_2= short path length

I_0= intensity of incoming light

Then on combing equations 3&4 we get,

$I_1/I_2 = \exp \varepsilon c (l_1-l_2)$ --------5

$L_{effective} = l_1-l_2$

Oil in Cuvette measuring at two path length:

Today all the available IR measurements at Foss are based on a cuvette with fixed path length of ~40 μ and they need to go for zero setting of the instrument by the introduction of a water sample into the cuvette. In other words the measurements so far are relative to water

Disadvantages of instruments measuring relative to water:

1. Zero setting takes time.
2. Highly viscous samples are difficult to pump into the cuvette.
3. The cuvette can block up and be difficult to clean.
4. Zero setting using water makes no sense when measuring samples that are not based on water i.e. oils.

The project is based on the new idea to avoid zero setting relative to water, instead measure the samples at 2 different path lengths. By using the measurement at the shortest path length as a zero measurement the resulting spectrum look like a sample measured relative to air instead of water. The effective path length will be the difference between the two path lengths.

One way of implementing this technique will be in a simple cuvette that can be opened and closed for measurement at two path length. Here the sample is introduced manually and does not require flow system.

The OenoFoss will have the following **advantages:**

1. Zero setting is done automatically for every measurement.
2. Highly viscous samples can be measured more easily. (Of course they should be homogeneous).

3. The customer will never face the problem of blocking of the cuvette. One can simply open the cuvette and clean it.
4. Samples with low content of water can also be measured.
5. Due to automatic zero setting for every sample one will expect that it will be possible to measure at quite unstable conditions (ie: temperature of sample, temperature of instrument, relative humidity in air, detector sensitivity.)

The main disadvantage is that the effective path length ($L_{effective}$) will vary from one sample to another, but it is shown, based on simulations that the proper chemometric treatment of the spectra makes it possible to make a good calibration.

OenoFoss FTIR instrument was used to obtain resultant spectra based on two different path lengths. The time taken for the measurement of each sample is approximately 1 minute for path lengths 100µm & 60µm.

When the path lengths, that will be used for IR spectrum are decided, one measurement taken at path length for example 40µm, will not set it back to 40 while taking a second measurement instead will go to for example 41µm. Therefore the $L_{effective}$ is not the same and to overcome such multiplicative effect some transformations have to be carried which can be either SNV or MSC.

Measurements at 6 different path lengths (430, 220, 110, 60, 30, 10µm) are collected for every sample. But due to time limitation of the project chemometric analysis was done for 110-10 µm IR spectra.

3. Chemometric Analysis

.

Overall 53 oil samples have been used for analysis. Of them for peroxide value and Saponification number, 52 samples have been used by removing sample 26 i.e. Lab Wheat germ oil (Hvedes Kim oil) which is actually in the sample list but not considered for analysis as it is not an edible oil This project includes analysis of only edible oils. However for Free fatty acids 49 samples have been used for reference analysis. (Due to non availability of GC/MS apparatus after a certain time period the number of samples analysed by GC/MS are 49).

To avoid the spectral interference pre-treatment have been tried. The pre-treatments that have been tried for the FTNIR data are: 1) MSC 2) SNV 3) 1st Derivative and 4) 2nd Derivative. For the FTIR measurements at path lengths 100-60µm (it is written in the report as A100-60) MSC and SNV pre-treatments are tried as these spectra are expected to have multiplicative scatter effect. The FTIR measurements at path lengths 110 µm and 10 µm, in the report these path lengths are mentioned as A110-10 µm.

Segmented cross validation approach using random selection has been used for modelling with 6 segments of 8 (15%) samples in each segment for both NIR and IR data analysis. All the samples have been measured in replicates but have been reduced by a reduction factor of 2 thereby the samples have been reduced from 106 to 53 samples. Savitzky Golay differentiation is used for both 1st and 2nd derivatives pre-treatment. All

models that have very low correlation coefficient (R^2) value are considered as bad models and the one that was modelled using many components are over fitting.

3.1.FTNIR *Analysis*
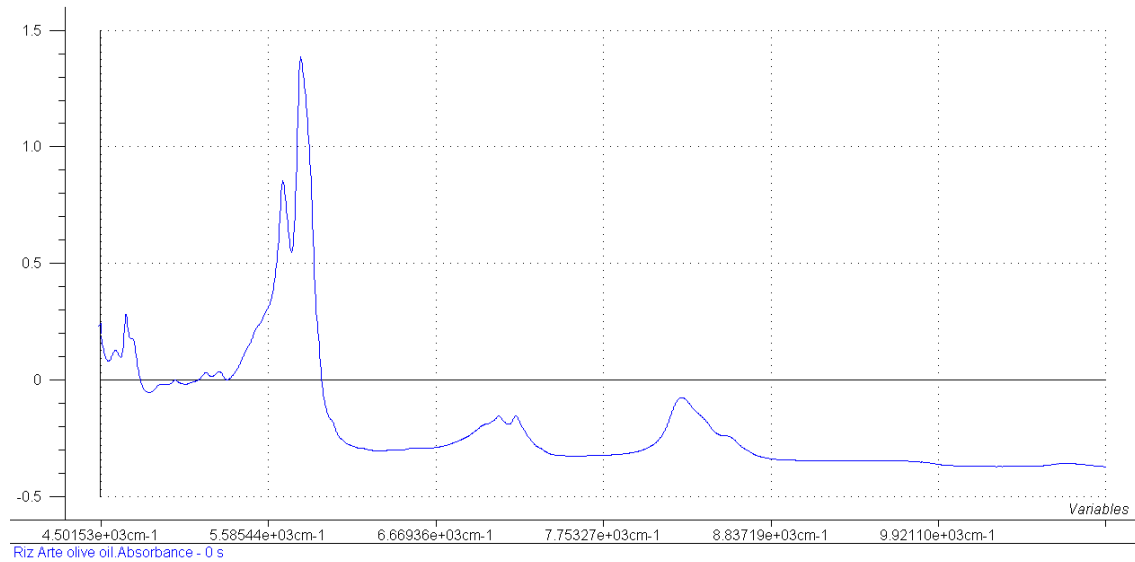
Figure 5: NIR Line plot of Olive oil.

The above is the line plot of arte olive oil to show the NIR spectrum. The FTNIR spectra were pretreated by different transformations i.e. MSC, SNV, 1st Derivative and 2nd derivative. The plots of these transformations are dicussed below. However of all the SNV correlations are found to be better compared to the others.

3.1.1. Pre-treatments:

3.1.1.1. FTNIR MSC: The NIR spectra are pre-treated by MSC transformation. The obtained MSC corrected data is subjected to PLS1 regression with cross validation for 49 samples. The obtained results are discussed for parameter C12 and C18:2.

3.1.1.1.1. C12:
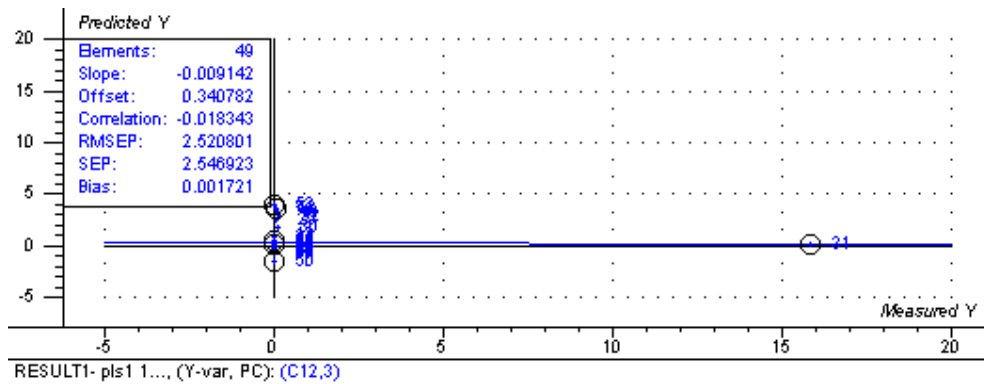
Figure 6: Predicted vs Measured Y plot.

RMSEP: 2.520801, Correlation: -0.018343, Slope: -0.009142, Bias: 0.001721.
RMSEC: 1.044644, Correlation: 0.884573, Slope: 0.782469, Bias: -2.006e-07.
Even when the numbers of components have been increased to 12, then also the correlation is very less of about 0.067480.
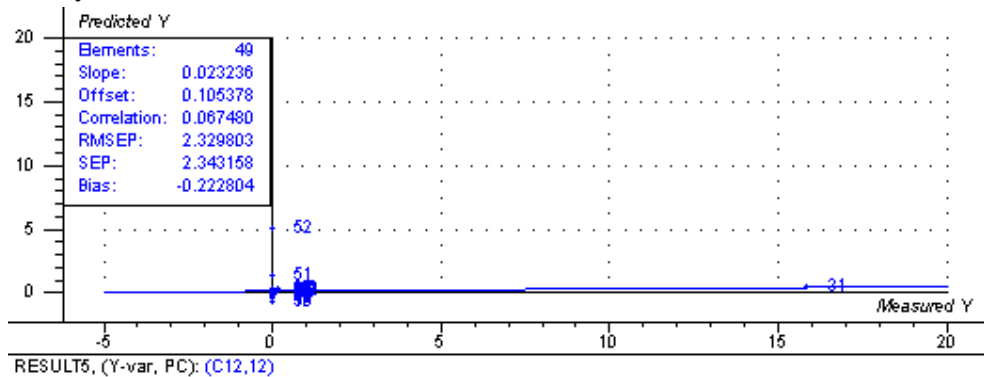
**Figure7**: Predicted Vs Measured plot for 12 components.
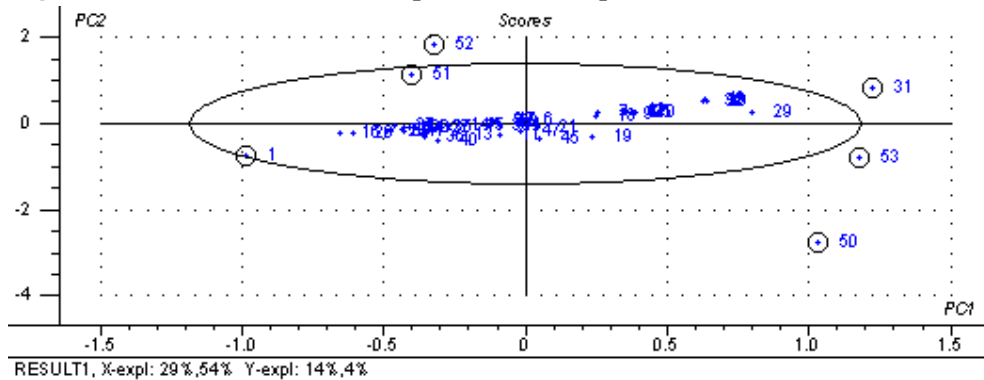

Figure8: Hoetelling T2 plot.

It is usual to choose the minimum variance PC (principal component) in the residual Y variance plot for modelling. But in this case the 12 PCs are found to be at minimal variance. And it is also observed that the X explanation is only 29% for PC1 and 54% for PC2 while Yexplanation is 14% for 1st component and only 4% for 2nd component. This explanation for both the components is very low. When the 1, 2 or 3 components are chosen for modelling

15

they turn out with negative correlation. In either way of choosing different components didnot turn out into a good fit.

The hoetelling T2 plot shows outliers 1, 31, 50, 51, 52, 53. These outliers are removed and recalculated without these to obtain the result below.
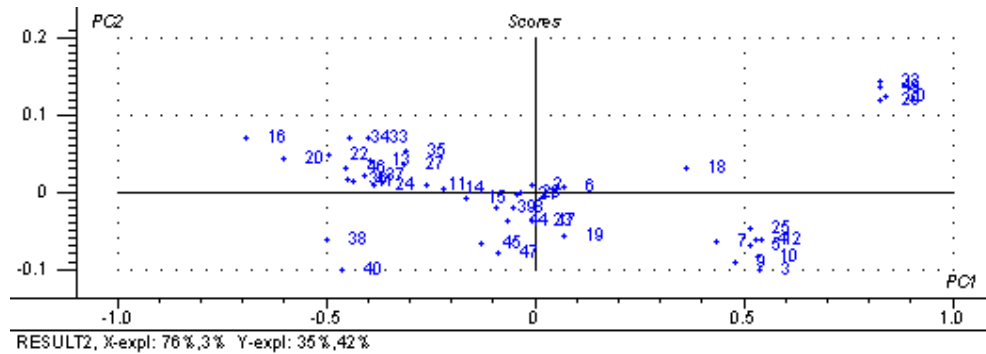
Figure 9: Hoetelling T2 plot.
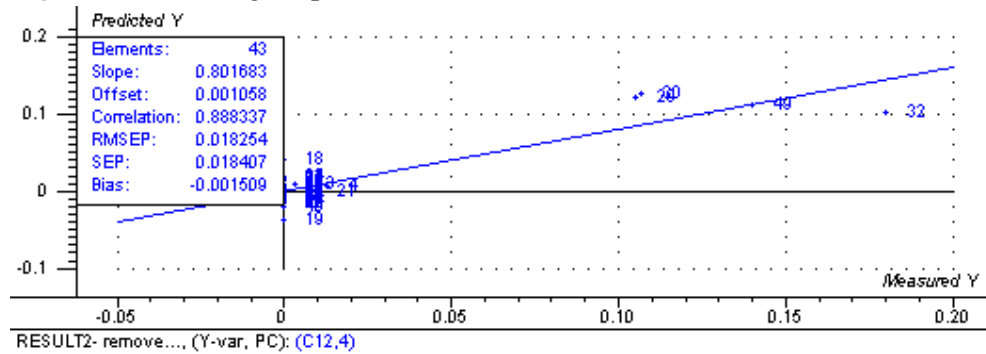
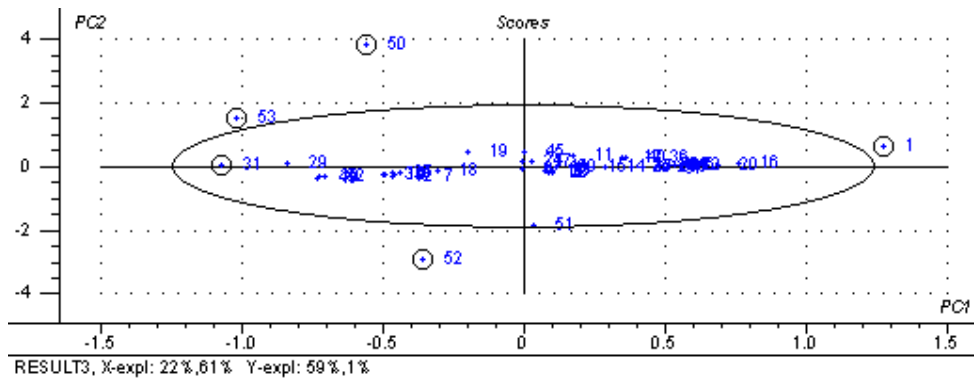**Figure 10:** Predicted vs Measured Y plot.

RMSEP: 0.018254, Correlation: 0.888337, Slope: 0.801683, Bias: -0.001509.
RMSEC: 0.013463, Correlation: 0.940423, Slope: 0.884396, Bias: -1.040e-08.

This model is with higher correlation lower RMSEP with slope which is close to 1. Even the X explanation has increase to 76% for 1st component but if it has been above at least 85% then might have been good. However the score plot shows some groups of similar type of oils. But still few samples are observed falling far away from the other samples being observed in the Predicted vs. Measured Y plot.

3.1.1.1.2. C18:2:

Figure11: Hoetelling T2 plot.
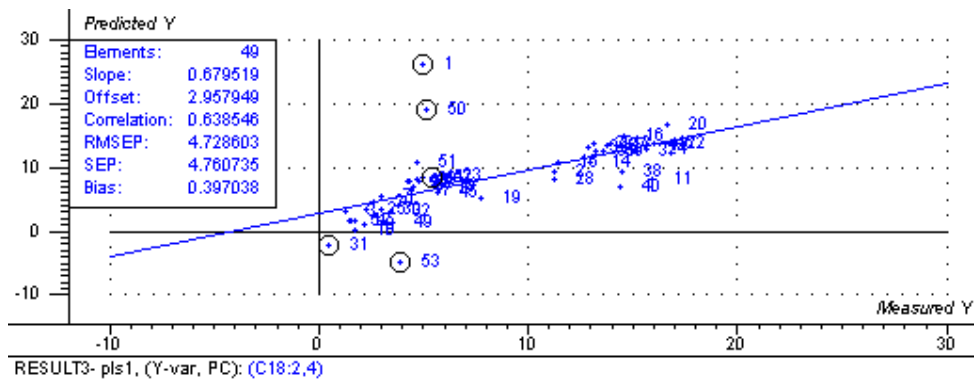
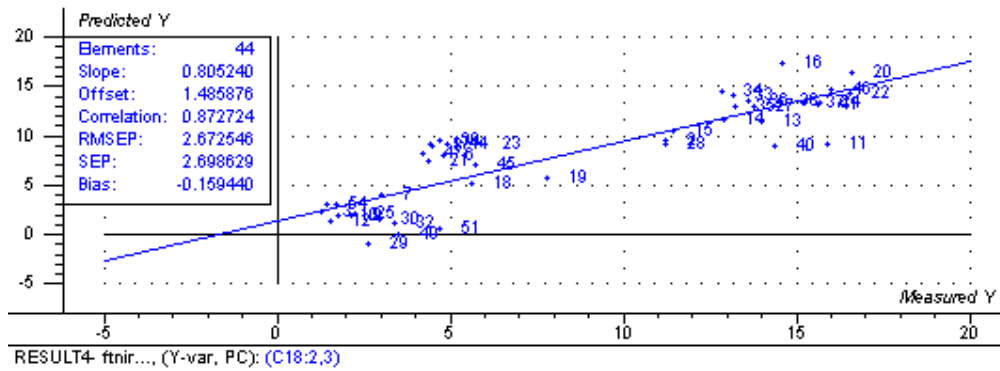**Figure12:** Predicted vs. Measured Plot.

RMSEP: 4.728603, Correlation: 0.638546, Slope: 0.679519, Bias: 0.397038.
RMSEC: 2.840570, Correlation: 0.847892, Slope: 0.718921, Bias: 1.277e-07

The PLS1 regression with segmented cross validation of full MSC pretreated FTIR data resulted in the model above with about 22% X explanation along 1st component while 51% along 2nd component. The Y explanation for 1st component is 59% and for 2nd component 1%. The R^2=0.638546 with slope around 0.67. This model showed 50, 1, 52, 53 and 31 as outliers which are marked and recalculated without the marked to get the below model with 44 samples and using 3 components. The R^2= 0.872724 and RMSEP has decreased in this model to 2.672546 from a value of 4.728603 which indicates a reduction in the error. The X explanation is 55% for 1st component and 17% for 2nd component while Y explanation is 77% and 2% for 1st and 2nd components respectively.
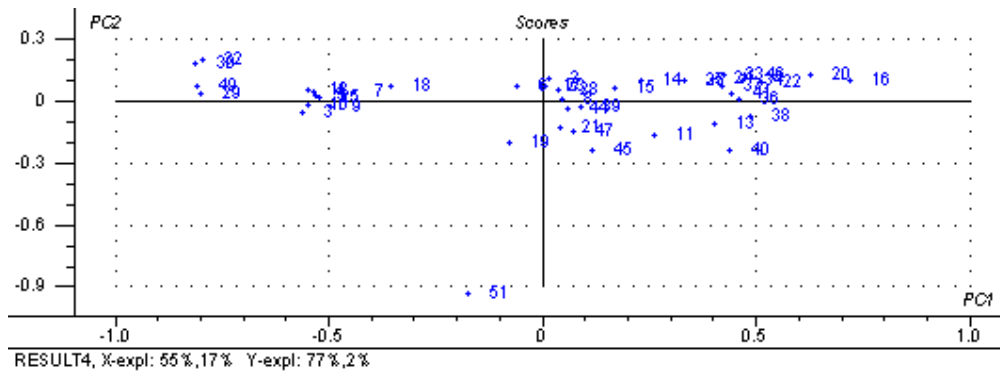
Figure 13: Score plot.

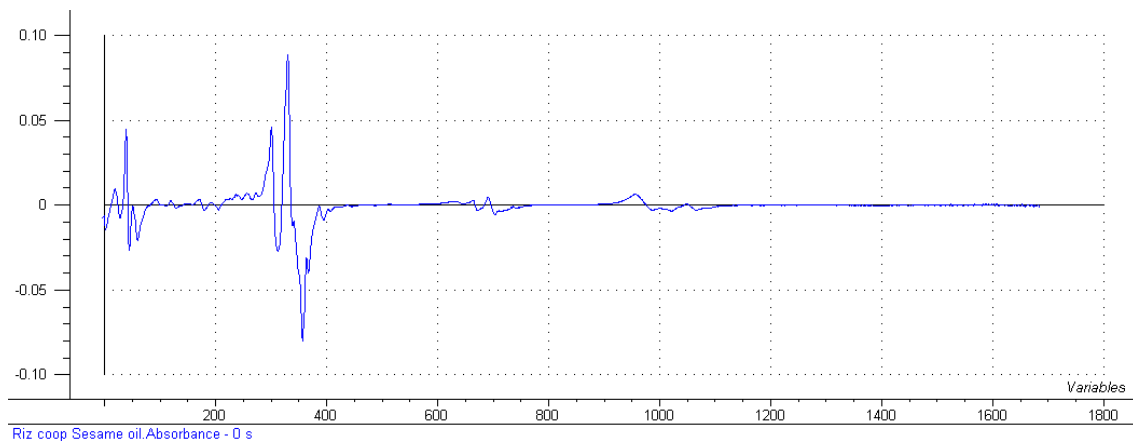
Figure14: Predicted vs. Measured plot.

RMSEP: 2.672546, Correlation: 0.872724, Slope: 0.805240, Bias: -0.159440.

RMSEC: 2.388503, Correlation: 0.898299, Slope: 0.806940, Bias: 1.707e-07.

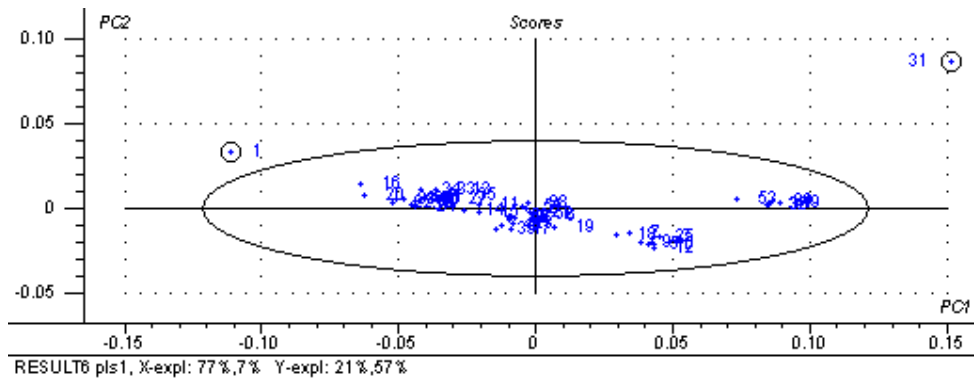Higher R^2 Values are obtained for SNV pretreatment than for the MSC in the case of C18:2..
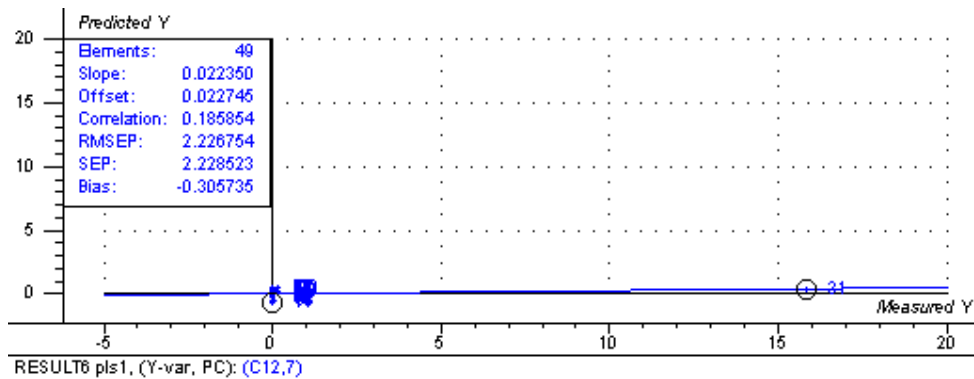
3.1.1.2. FTNIR 1ST DERIVATIVE:

Riz coop Sesame oil.Absorbance - 0 s

Figure 15: 1st Derivative Line plot of riz coop sesame oil.

3.1.1.2.1. C12:

First derivative for this parameter has been checked.

Figure16: Hoetelling T^2 plot

Figure17: Predicted vs. Measured Plot.

RMSEP: 2.226754, Correlation: 0.185854, Slope: 0.022350, Bias: -0.305735.
RMSEC: 0.200584, Correlation 0.995982, Slope: 0.991980, Bias: 6.451e-07.

The model obtained for 1st derivative pretreated data was with 77% x explanation for 1st component which is nearly good. But the R^2= 0.185854 which is very low and the number of components are 7 which makes the model a bit overfitting. The outliers 31 and 1 are removed and recalculated without these to obtain the following model with R^2= 0.833679 and lower RMSEP. The correlation is good with 79% X explanation and 41% Y explanation for 1st component. 2 components are used for the modelling purpose which make it a model with good fit (or slightly underfit). The score plot shows grouping of similar type of oils but the Predicted & Measured Y plot shows all the palm oils as being far away from the other oils.
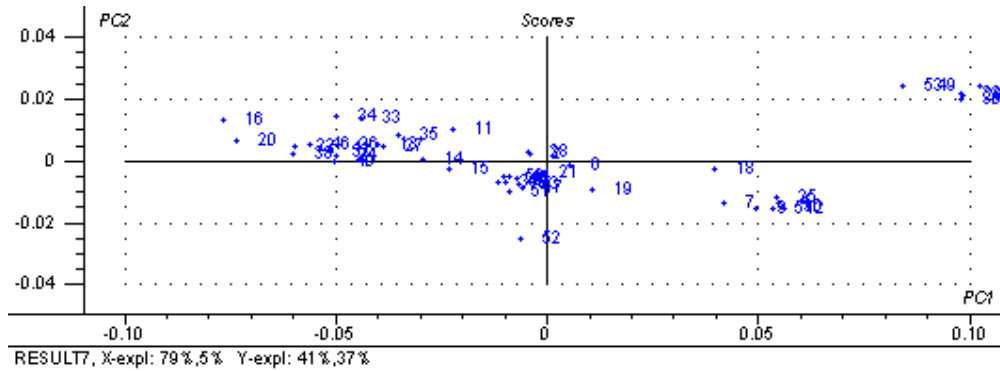
RESULT7, X-expl: 79%,5% Y-expl: 41%,37%

Figure18: Score plot.
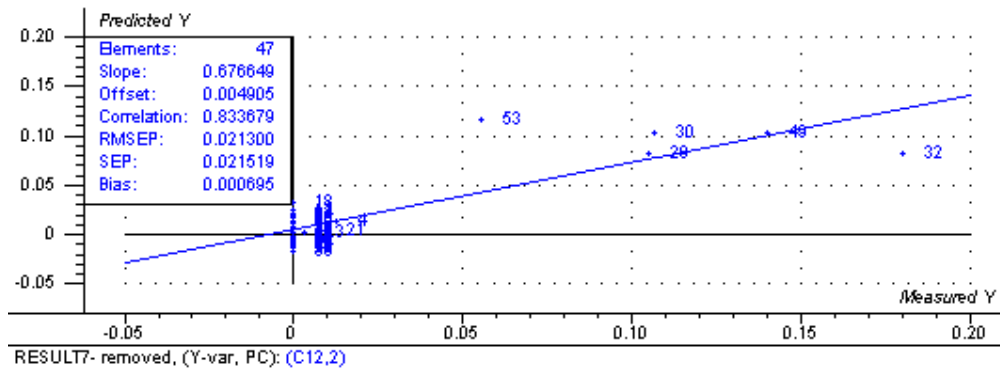

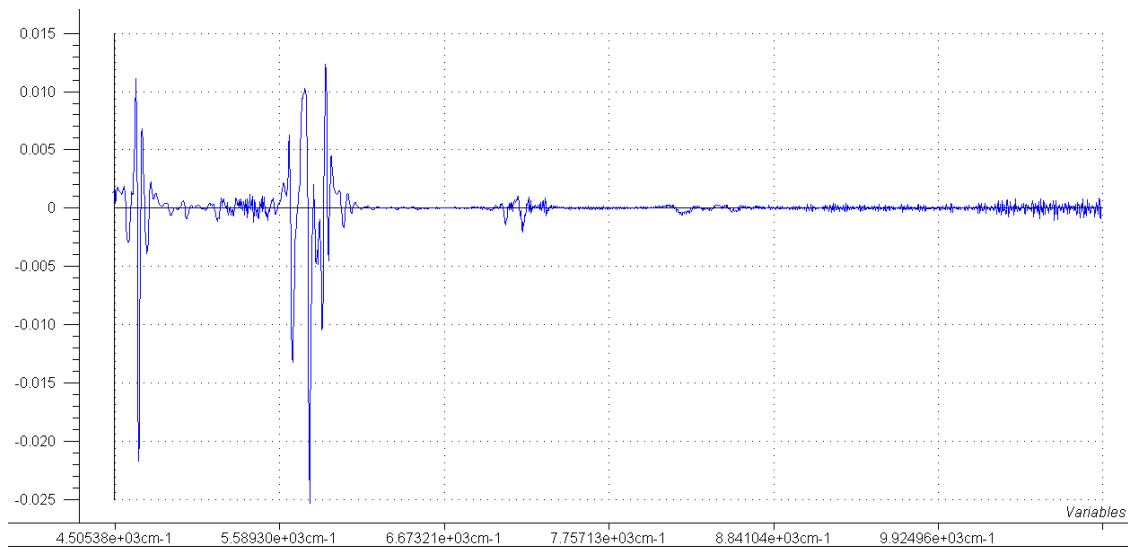RESULT7- removed, (Y-var, PC): (C12,2)

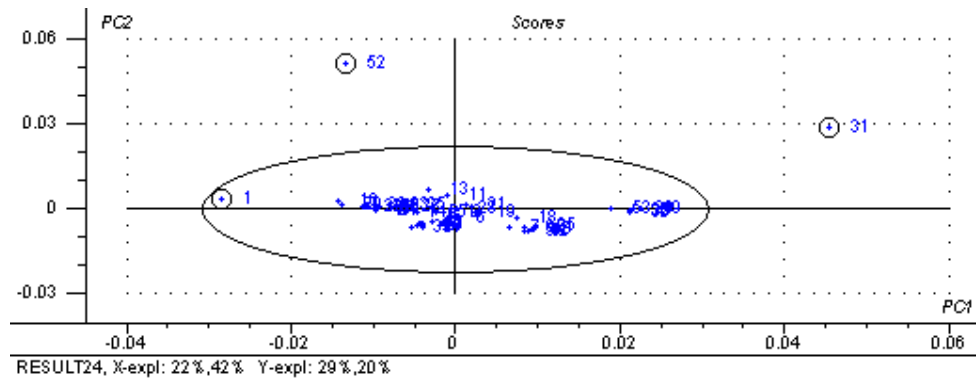**Figure 19:** Predicted vs. Measured Plot.

RMSEP: 0.021300, Correlation: 0.833679, Slope: 0.676649, Bias: 0.000695.
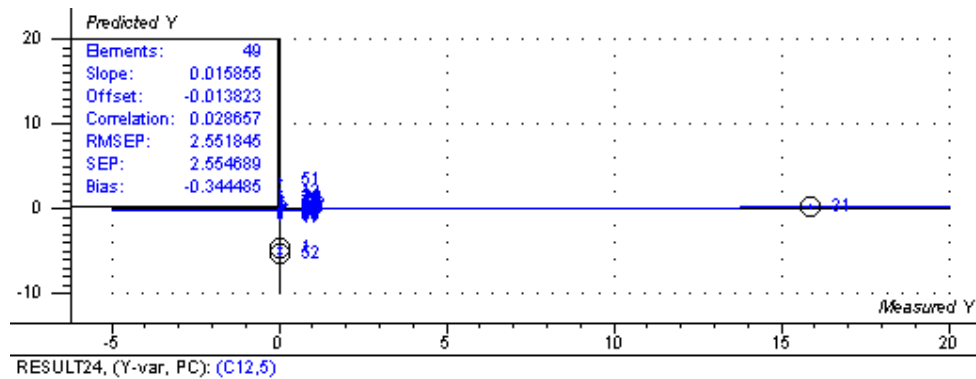RMSEC: 0.017794, Correlation: 0.886908, Slope: 0.786605, Bias: -1.556e-09.

3.1.1.3. FTNIR 2nd derivative.

Riz coop Sesame oil.Absorbance - 0 s

Figure 20: Line plot of 2nd derivative of coop sesame oil.

3.1.1.3.1. C12: Second derivative has been modelled for this parameter.
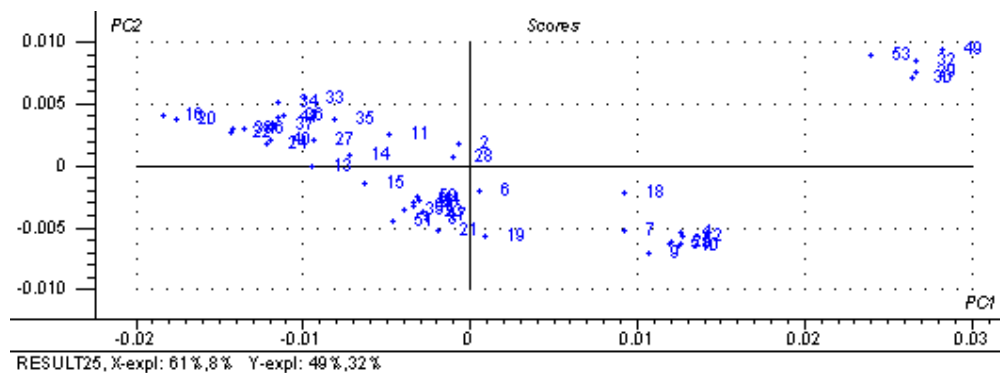

Figure 21: Score plot.

Figure22: Predicted vs. Measured Plot.

RMSEP: 2.551845, Correlation: 0.028657, Slope: 0.015855, Bias: -0.344485.
RMSEC: 0.488836, Correlation: 0.975893, Slope: 0.952367, Bias: 3.935e-08.

The PLS1 regression with segemented cross validation of second derivative pretreated FTNIR data gave the model with very very low correlation and only 22% X, 29% Y explanation for 1st component. The outliers 1, 31 and 52 samples are marked and recalculated without these marked which gave the following model with 51% X and 49% Y explanation as well as good correlation. 2 components are used for modelling 46 samples.

Figure 23: Score plot.
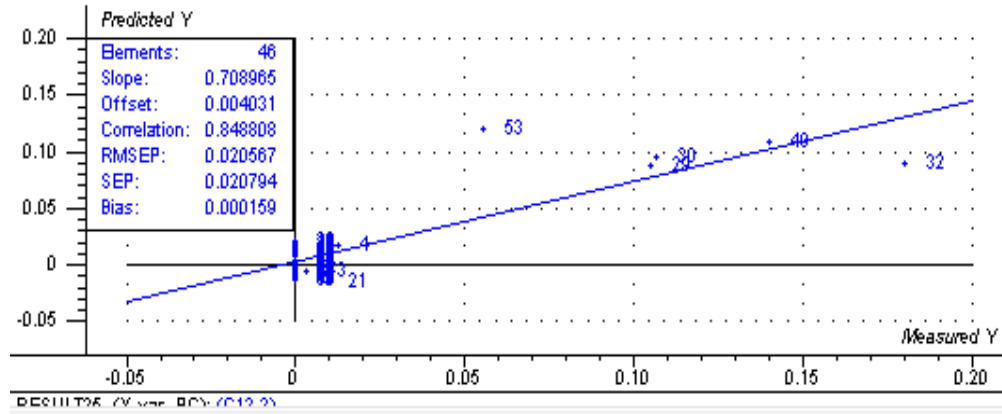
Figure 24: Predicted vs. Measured Plot.

RMSEP: 0.020794, Correlation: 0.848808, Slope: 0.708965, Bias: 0.000159.
RMSEC: 0.016758, Correlation: 0.902377, Slope: 0.814285, Bias: -2.733e-10.

3.1.2. FTNIR pretreatments Results.

Before removing outliers:

49 elements have been used for the below results regression.

C12	MSC	SNV	1ST Derivative	2nd Derivative
RMSEP	2.520801	2.177514	2.226754	2.551845
Correlation	-0.01834	0.410050	0.185854	0.028657
Slope	-0.00914	0.041418	0.022350	0.015855
Bias	0.001721	-0.29881	-0.305735	-0.344485
RMSEC	1.044644	0.127209	0.200584	0.488836
Correlation	0.884573	0.998386	0.995982	0.975893
components	12	12	7	5

Table2: Pre-treatment results.

SNV is good compared to MSC, 1st and 2nd derivatives in terms of correlation, lower RMSEP.

After removing outliers:

C12	MSC	SNV	1ST Derivative	2nd Derivative
RMSEP	0.018254	0.018457	0.021300	0.020794
Correlation	0.888337	0.889381	0.833679	0.848808
Slope	0.801683	0.831657	0.676649	0.708965
Bias	-0.001509	-0.001636	0.000695	0.000159
RMSEC	0.013463	0.013607	0.017794	0.016758
Correlation	0.940423	0.940403	0.886908	0.902377
Elements	43	42	47	46
components	4	4	2	2

Table 3: Pre-treatment results after modelling.

SNV is good compared to 1st and 2nd derivative in terms of lower RMSEP, higher correlation, and slope value higher. But when compared to MSC seems to be almost similar with slightly higher correlation and slope value. SNV pre-treated FTNIR data has been modelled for all the parameters.

3.1.3. FTNIR SNV:

#### 3.1.2.1. PV (Peroxide Value)

52 samples data of reference and NIR are collected and PLS1 regression, with cross validation of 6 segments with 8 samples in each segment, of all these samples gave the following results.

The regression results were with the following values.
RMSEP: 7.493101, Correlation: 0.693958, Bias: -1.046744, Slope: 0.462657.
RMSEC: 4.961398, correlation: 0.876276, slope: 0.767860, Bias: 8.505e-07.

Eight components are used for the modelling. The resulted model shows 36% X and 6% Y explanation by the 1st component. About 47% and 1% X & Y explanation respectively was resulted by the 2nd component.
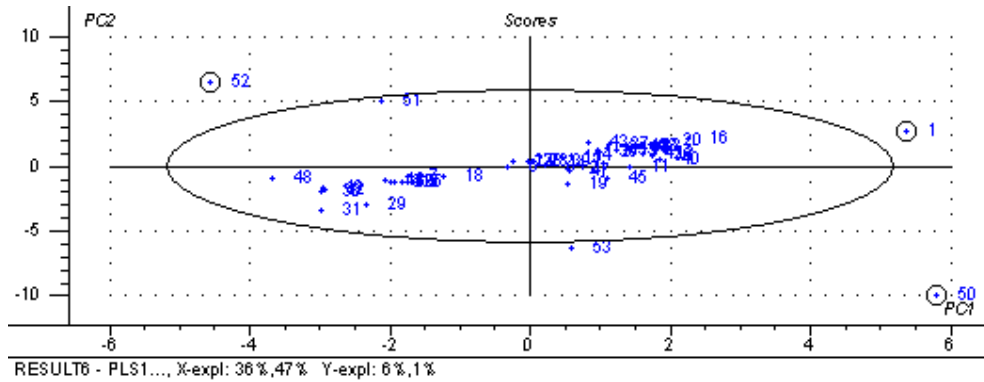
Figure 25: Score plot.
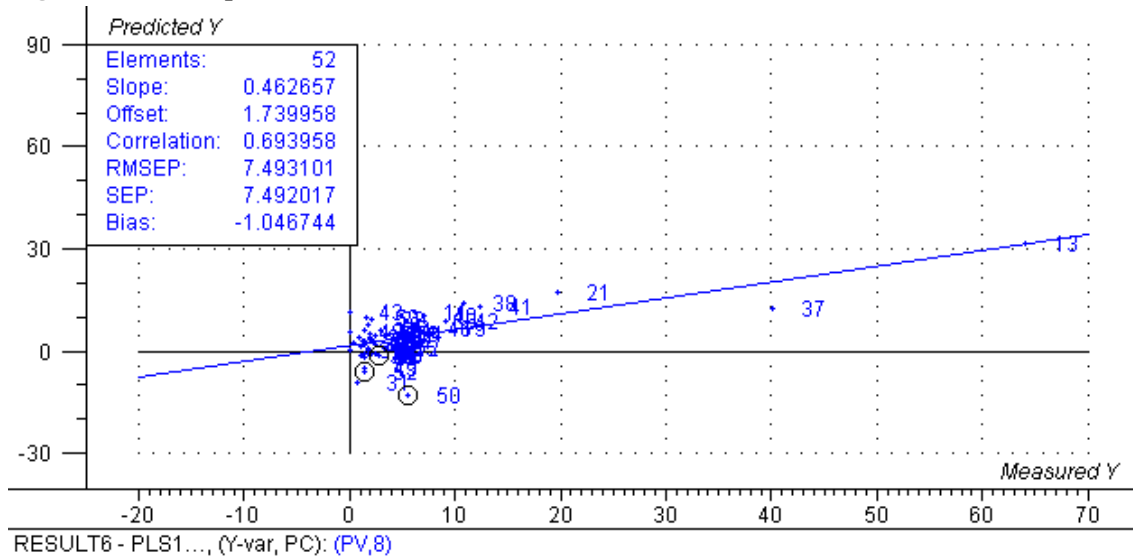
Figure 26: Predicted vs. Measured Plot.

1, 52, 50, 53, 31, 51, 21, and 37 outliers have been removed which resulted in the model with the following values. The number of elements now are 44.
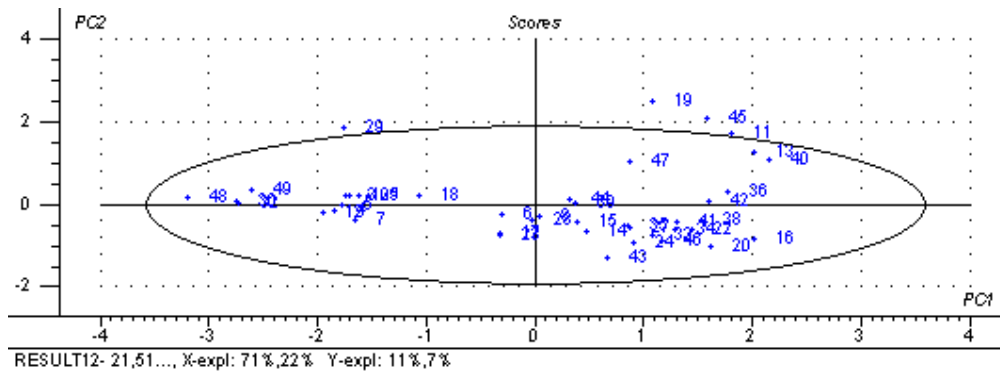
RMSEP: 4.965791, Correlation: 0.895829, Slope: 0.565089, Bias: -0.578939.
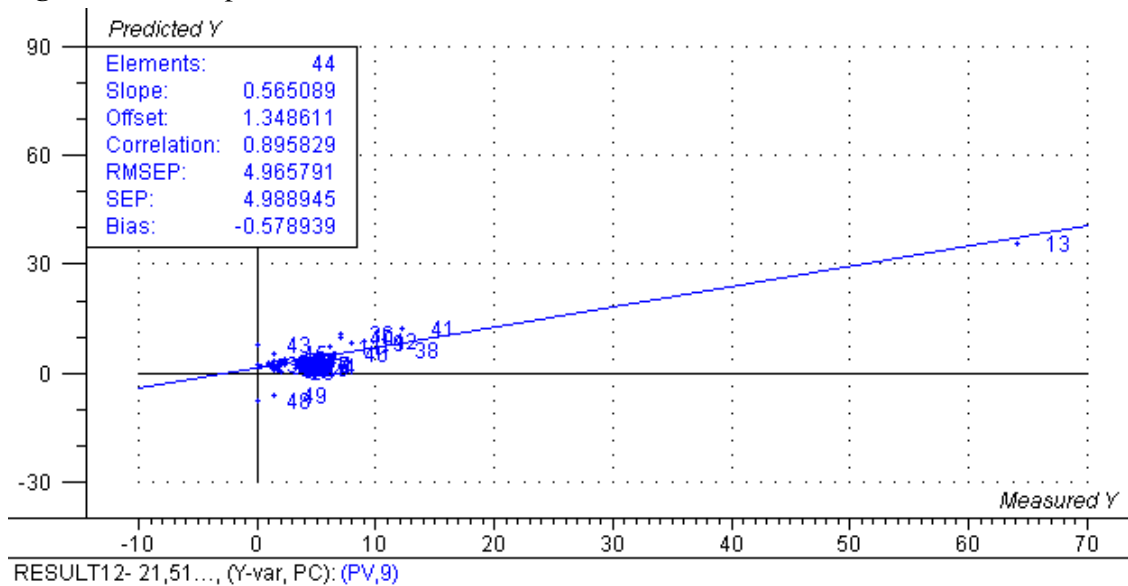RMSEC: 1.380218, Correlation: 0.989460, slope: 0.979032, Bias: 3.289e-06.

On the whole from a set of 52 samples 8 samples have been removed to make the model.
The 1st minimum is observed at 9 components in the variance validation plot.

Figure 27: Score plot.

Figure 28: Predicted vs. Measured plot.

It has been observed that sample number 13 is lying far away from other samples in the predicted vs measured Y plot. But this has not been observed as an outlier neither in hoetelling T^2 plot nor in other outlier detecting plots like X-Y relation plot. And when the model is checked after removing this outlier the correlation has decreased a lot thereby making a bad model. Therefore it has been not removed as an outlier. The sample number 13 is sunflower oil.

The final model obtained after the removal of all the above mentioned outliers was with good correlation which increased from 0.63958 to 0.895829 with 44 elements and nine components making the model slightly overfitting but with good correlation (R^2).

3.1.2.2. SN (Saponification number)

52 samples data of reference and NIR are collected and PLS1 regression, with cross validation of 6 segments with 8 samples in each segment, of all these samples gave the following results
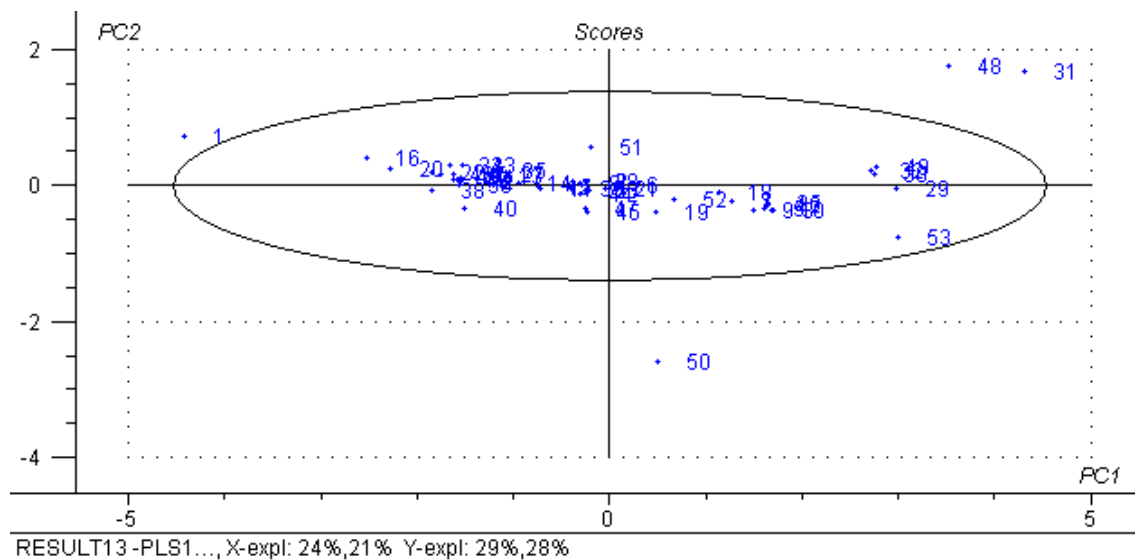
RMSEP: 7.315745, Correlation: 0.815031, Slope: 0.631223, Bias: -0.900452.

RMSEC: 5.332189, Correlation: 0.904443, Slope: 0.818017, Bias: 0.000
The number of components used for modelling are 5.
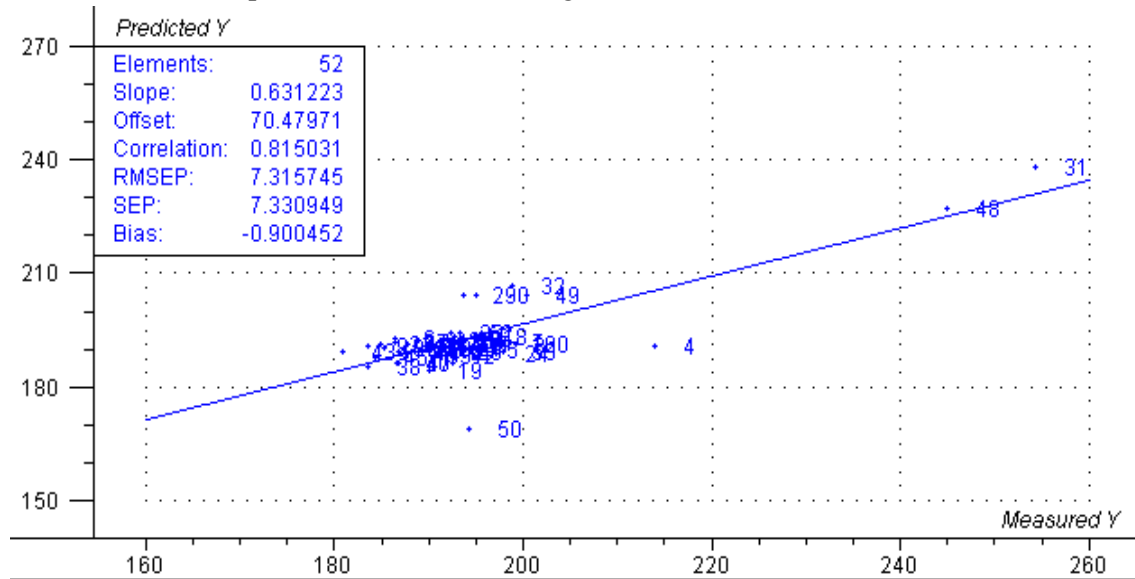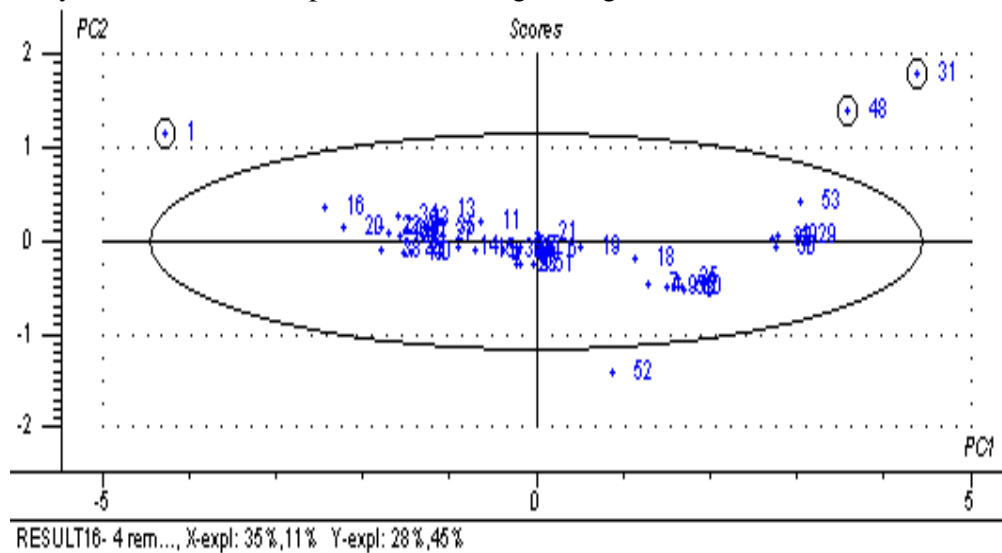

Figure 29: Predicted vs. Measured Plot.

Figure 30: Score plot.

After the removal of sample number 50 and 4 are removed as outliers observed in Predicted vs. Measured Y plot when this is removed the RMSEP and correlation values are good. But when 48, 31, 50 are removed the Correlation values are dropping down to 0.360659 and these are not seen as extreme outliers in other plots like predicted vs. Measured Y plot. The values of the results obtained when 50 &4 samples are removed as outliers.

RMSEP: 5.088634, Correlation: 0.919430, Slope: 0.739287, Bias: -0.207145.
RMSEC: 4.271522, Correlation: 0.938859, Slope: 0.881456, Bias: 6.104e-07.

The good model of 50 elements is obtained with high correlation using five components and decreased RMSEP value. But the resultant final model shows 48 and 31 as being very far away from the other samples but lies along the regression line.

Figure 31: Hoetelling T^2 plot.
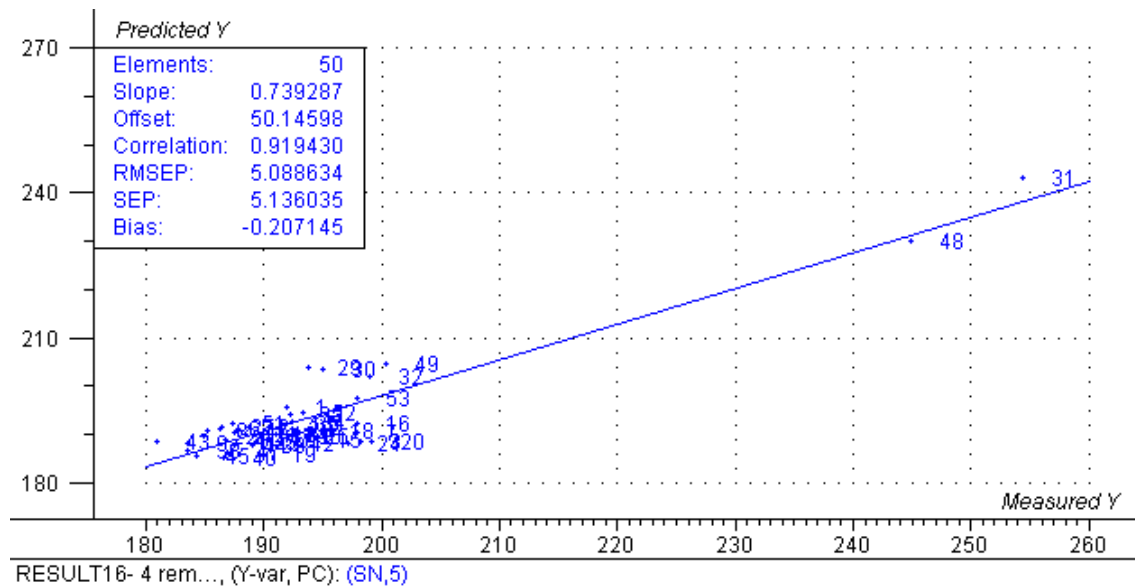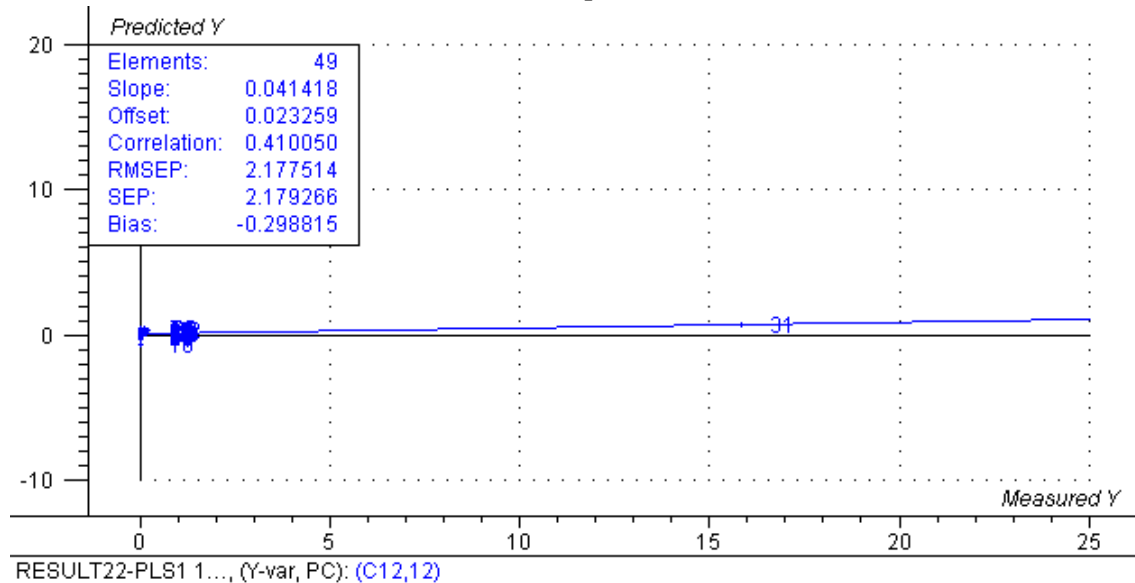

Figure 32: Predicted vs. Measured Y Plot.

3.1.2.3. C12:

The 49 samples are used forPLS1 regression with cross validation using 6 segments with 8 samples in each segment. The results obtained are discussed below.

The first model obtained as a result of regression of C12 free fatty acid gave the following values for 12 components making the model overfitting.
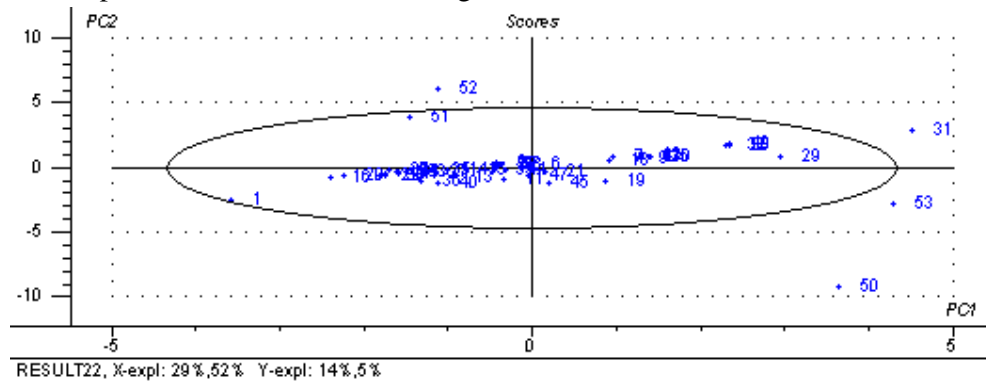
RMSEP: 2.177514, Correlation: 0.410050, Slope: 0.041418, Bias: -0.298815.
RMSEC: 0.127209, Correlation: 0.998386, Slope: 0.996774, Bias: 3.205e-08.

Figure 33: Predicted vs Measured Y plot.

12 components are used for modelling.

Figure 34: Hoetelling T^2 plot.

The samples 1, 40, 50, 52, 53, 31, 51 as per the observation from residual sample variance plot and Hoetelling T^2 plot it has been marked as outliers and these have been removed for obtaining a good model. The correlation and RMSEP values increased enormously. The number of principal components required were analysed to be 4 from the RMSEP vs PCs plot.

After the removal of all the above mentioned outliers the obtained model score plot shows clusters of oil samples. With palm oils being at the upper right corner while the raps oils at the centre, while soya and sunflower oils at the left and the olive oils are at the lower right corner.

Figure 35: Score plot.

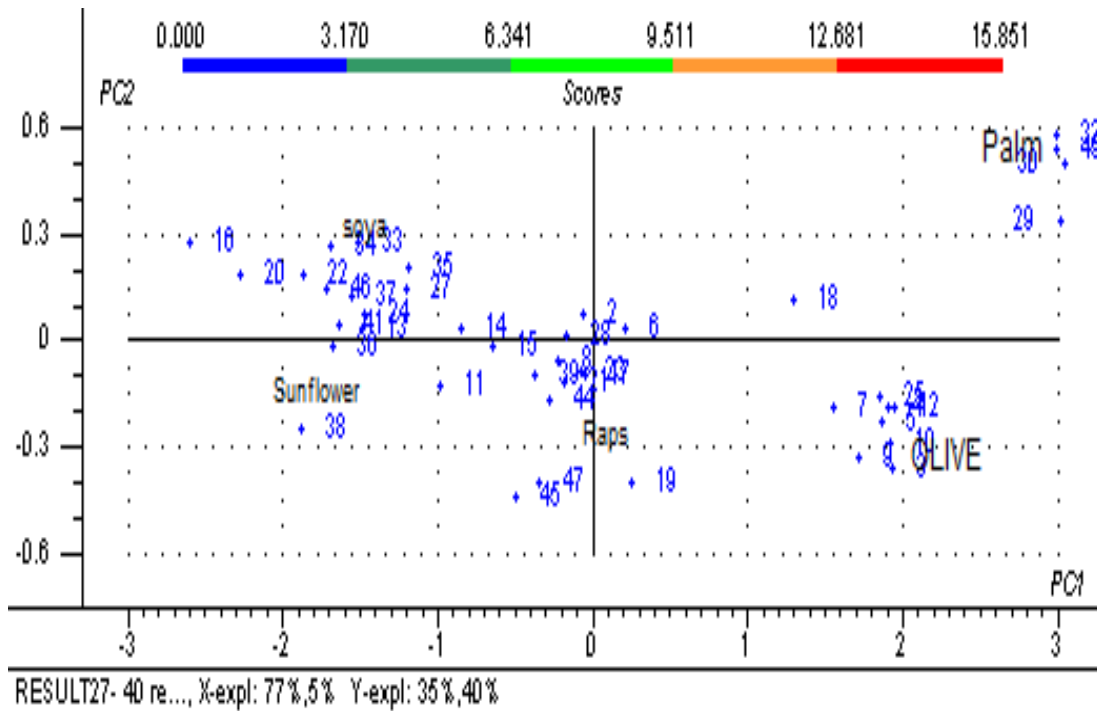Samples 30, 32, 29, 49 are palm oils; 33, 34 are Foss soya oils; 24, 37, 41, 36, 13 are sunflower oils as well as 38. 21, 17, 44, 8, 23 along with the below 45, 47 are Rape oils that has formed a cluster.
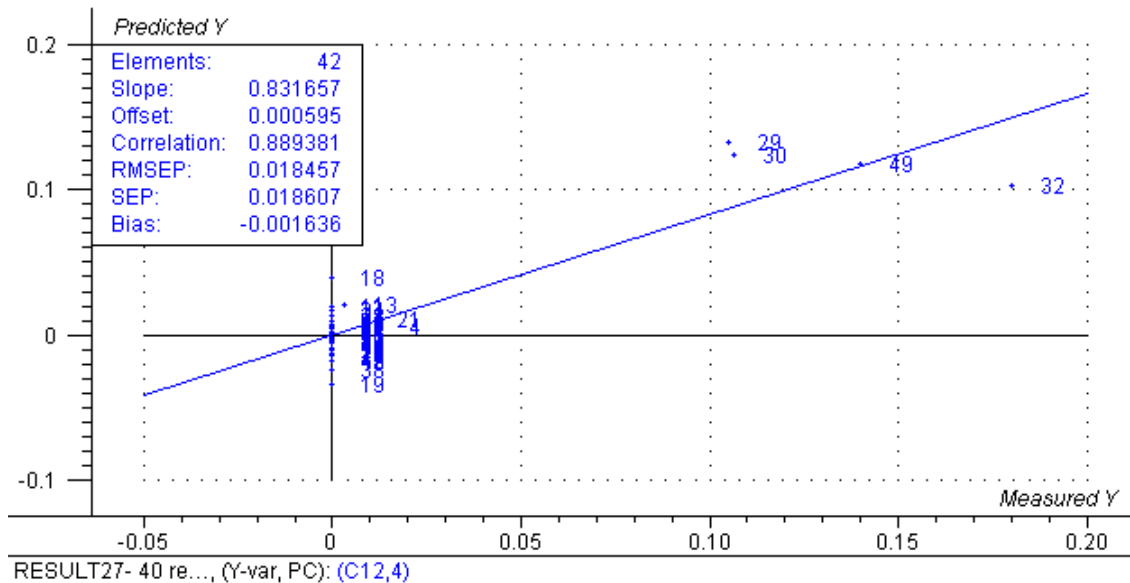
Figure 36: Predicted vs. Measured Plot.

RMSEP: 0.018457, Correlation: 0.889381, Slope: 0.831657, Bias: -0.001636.
RMSEC: 0.013607, Correlation: 0.940403, Slope: 0.884358, Bias: -2.051e-08.

3.1.2.4. C14:

Result 2 is obtained when 49 samples are subjected to PLS1 regression with cross validation of 6 segments and eight samples in each segment.
Based on the Y variance plot 8 components are used for modelling.

RMSEP: 0.331792, Correlation: 0.680462, Slope: 0.135037, Bias: -0.045881.
RMSEC: 0.031113, Correlation: 0.996546, Slope: 0.993103, Bias: 1.389e-08.
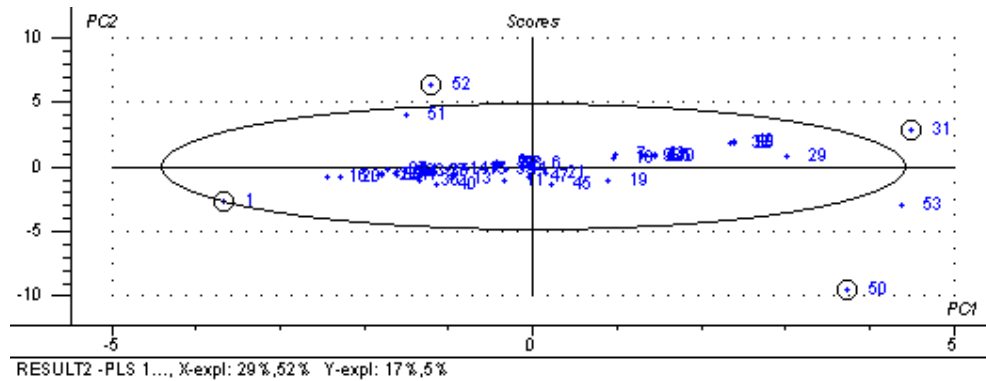
Figure 37: Hoetelling T^2 plot.
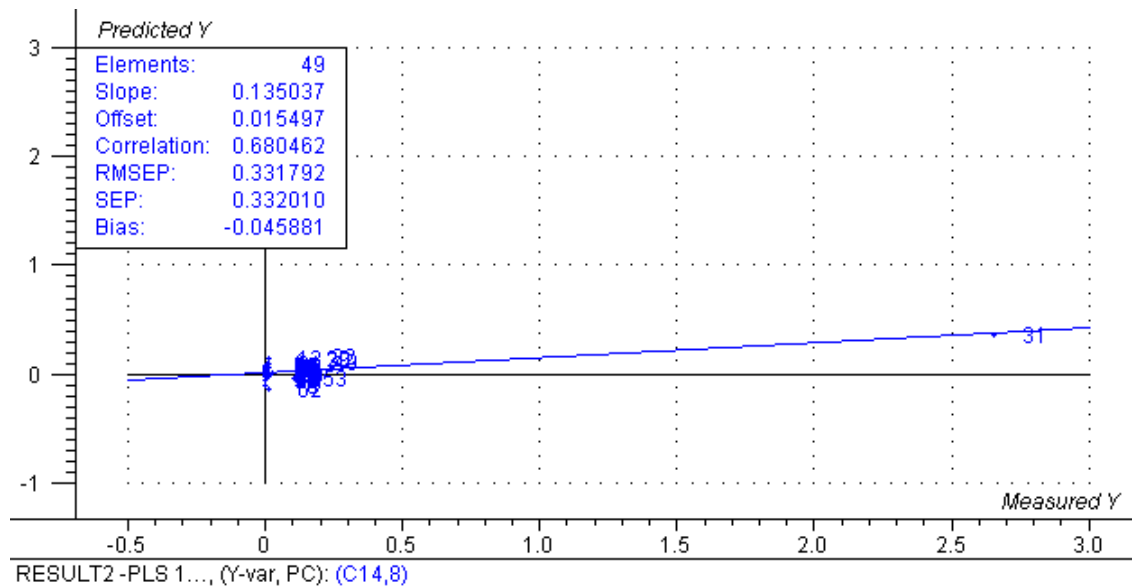
Figure 38: Predicted vs. Measured plot.
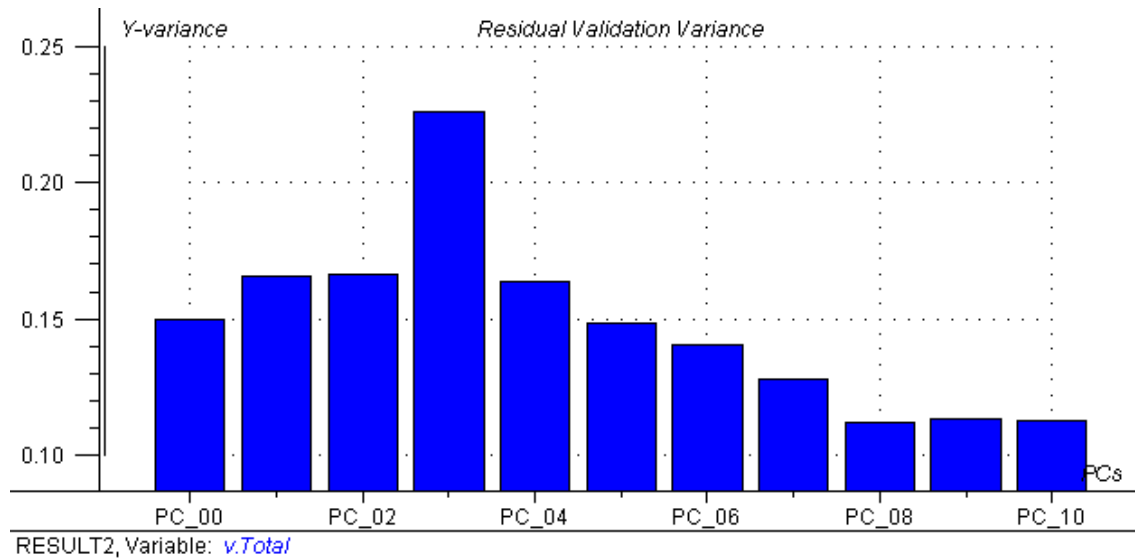
Figure 39: Residual Validation Variance plot.

When 52, 50, 31, 1, 53 and 51 are removed the obtained model is good with 43 elements and good correlation with 5 components.

RMSEP: 0.013471, Correlation: 0.931319, Slope: 0.847877, Bias: -0.000527.
RMSEC: 0.010564, Correlation: 0.958144, Slope: 0.918039, Bias: 8.274e-09.
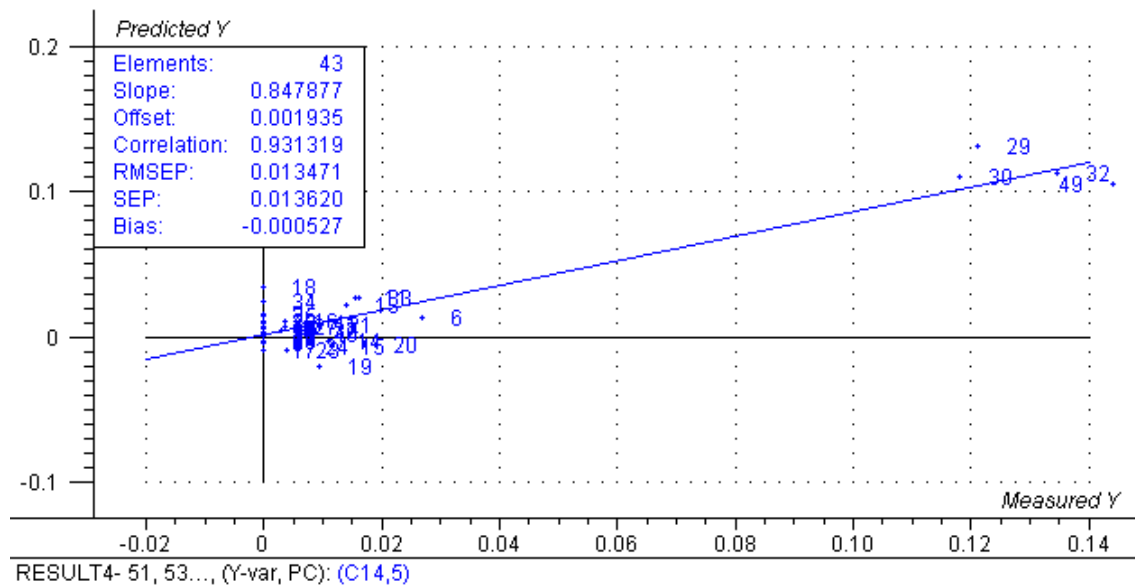
Figure 40: Predicted vs. Measured plot.

The palm oils (29, 30, 32, and 49) are observed as being present far away from other oil along the regression line in the predicted vs. measured Y plot.
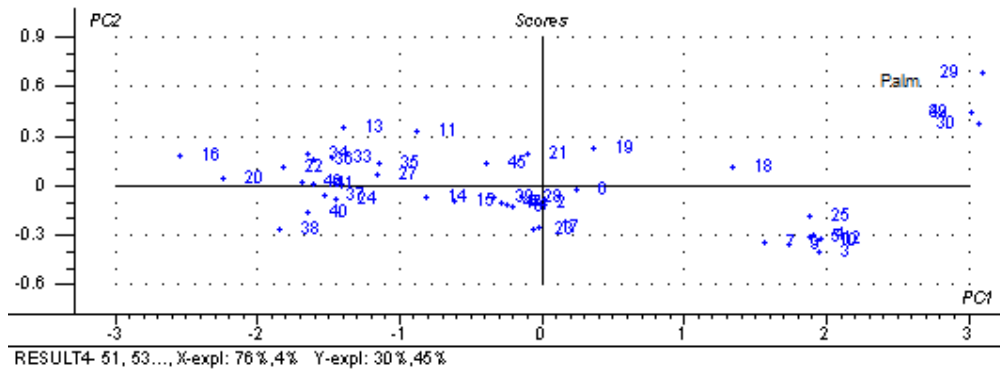
Figure 41: Score Plot.

All clusters are observed for olive, palm, soya, rape, maize oils separately.

3.1.2.5. C16:

PLS1 regression with cross validation of 49 samples with 6 segments of 8 sample in each segment resulted in the following.

RMSEP: 1.398795, Correlation: 0.524061, Slope: 0.374892, Bias: 0.096933.
RMSEC: 1.250843, Correlation: 0.622708, Slope: 0.387765, Bias: 1.484e-07
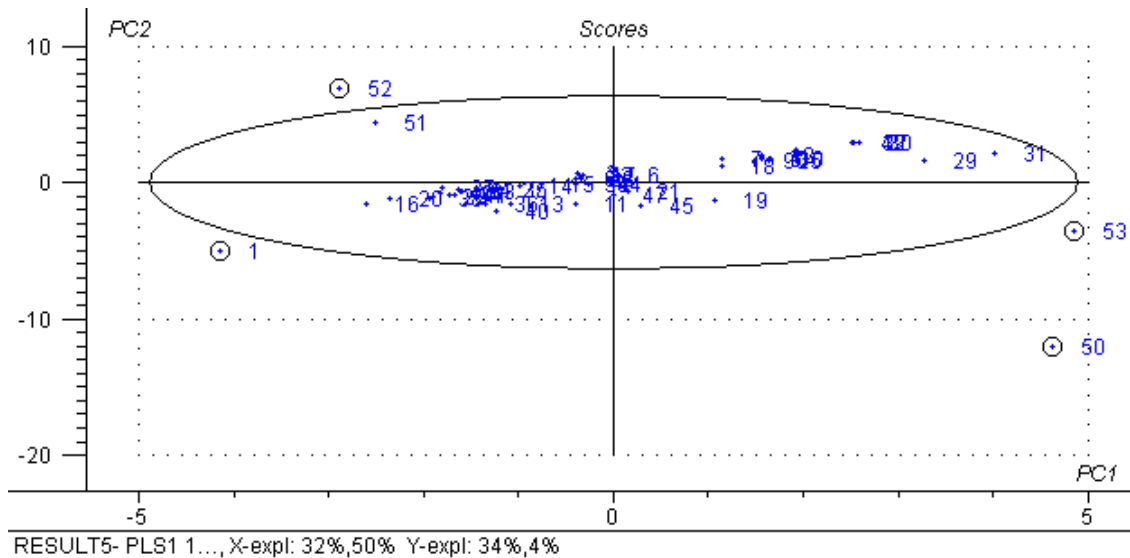
Figure 42: Hoetelling T^2 plot.
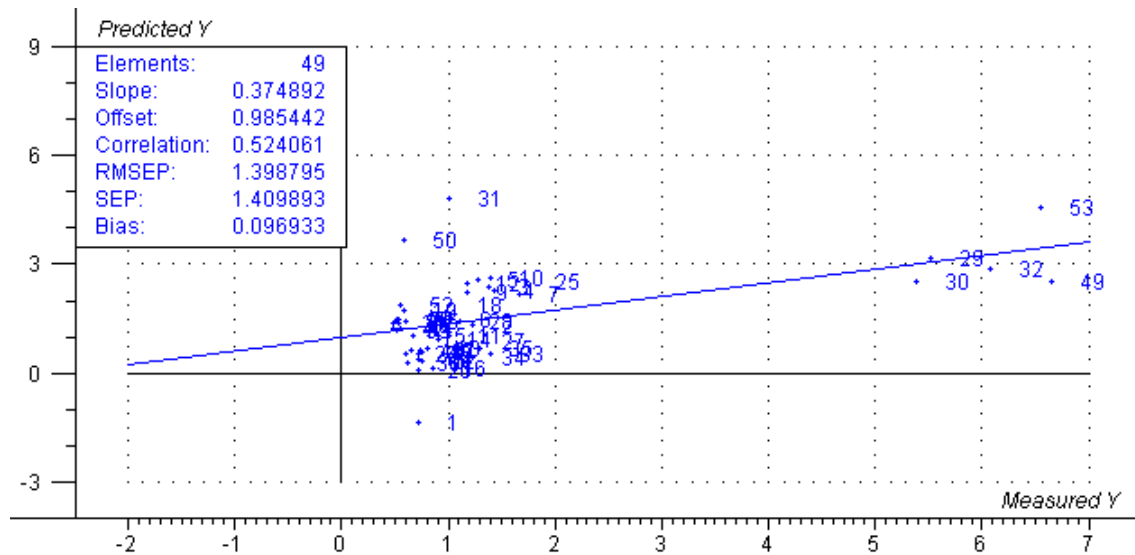
Figure 43: Predicted vs. Measured Y Plot.

The number of principal components required for modelling are 2 from the above y- variance plot. Based on the observations of Hoetelling T^2 plot, Residual sample variance plot and leverage plot the outliers 50, 52, 1, 53, 31, 51 are removed as outliers.

After the removal of the above outliers a good model is obtained with 43 elements, 5 principal components with the RMSEP and correlation values as below:

RMSEP: 0.384870, Correlation; 0.966306, Slope: 0.939398, Bias: -0.017539.
RMSEC: 0.265118, Correlation: 0.98414, Slope: 0.968481, Bias: 3.951e-07.
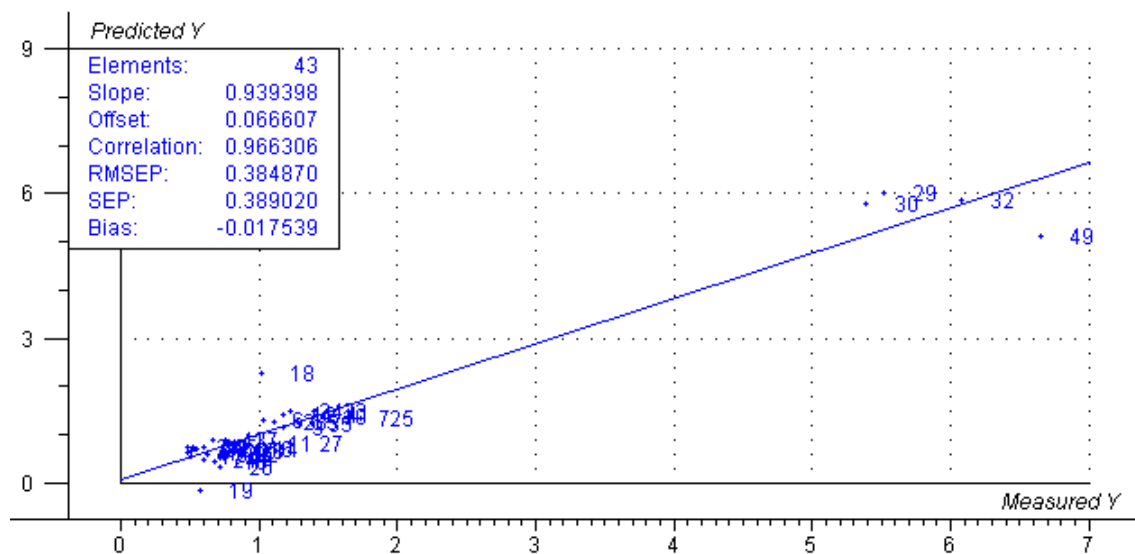
Figure 44: Predicted Vs Measured plot.
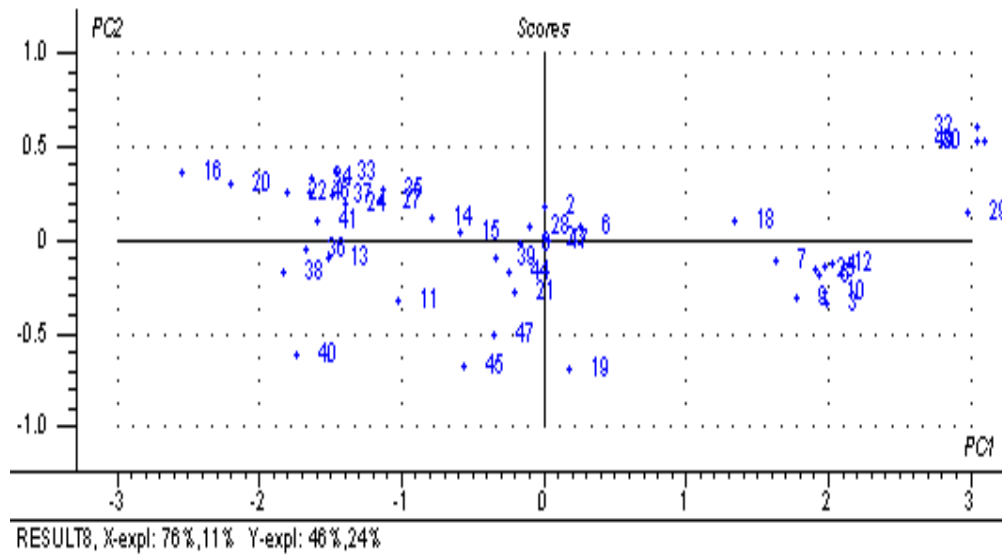
Figure 45: Score Plot.

3.1.2.6. C18:

PLS1 regression of 49 samples with 6 segments of 8 samples in each segment with C18 as y variable resulted with the following model with 2 components.

RMSEP: 0.78591, Correlation: -0.307742, Slope: -0.074927, Bias: 0.073430.
RMSEC: 0.690593, Correlation; 0.241050, Slope: 0.058105, Bias: -2.068e-08.
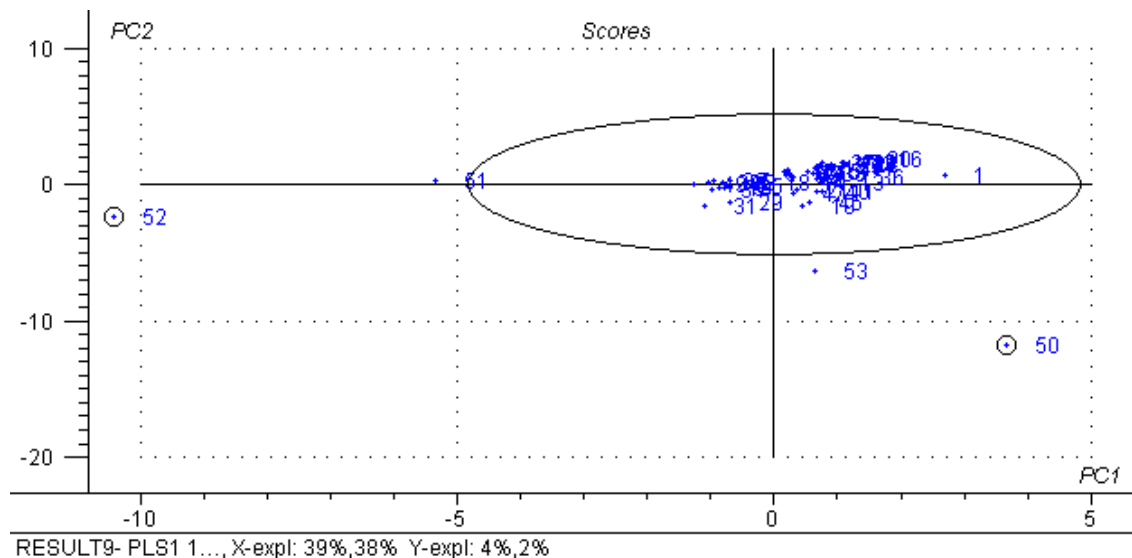
Figure 46: Hoetelling T^2 plot.
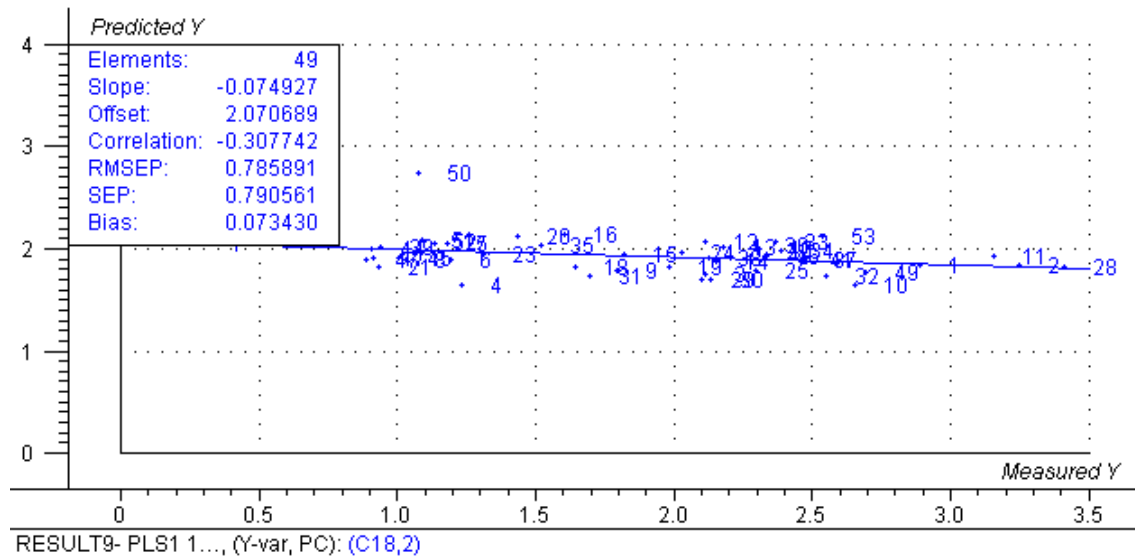
Figure 47: Predicted vs Measured Y plot.

Based on the analysis of the residual variance, score, leverage plots 52, 50, 1, 31, 53, 51 as outliers and these have removed and recalculated to obtain a model with 43 elements and 3 components with improved correlation. But the obtained correlation is 0.24166 which too low to conclude it as a good model. And in this model RMSEP value has been decreased.

RMSEP: 0,704180, Correlation: 0.24166, Slope: 0.113272, Bias: 0.009208.
RMSEC: 0.631256, Correlation: 0.448853, Slope: 0.201469, Bias: 3.604e-08.
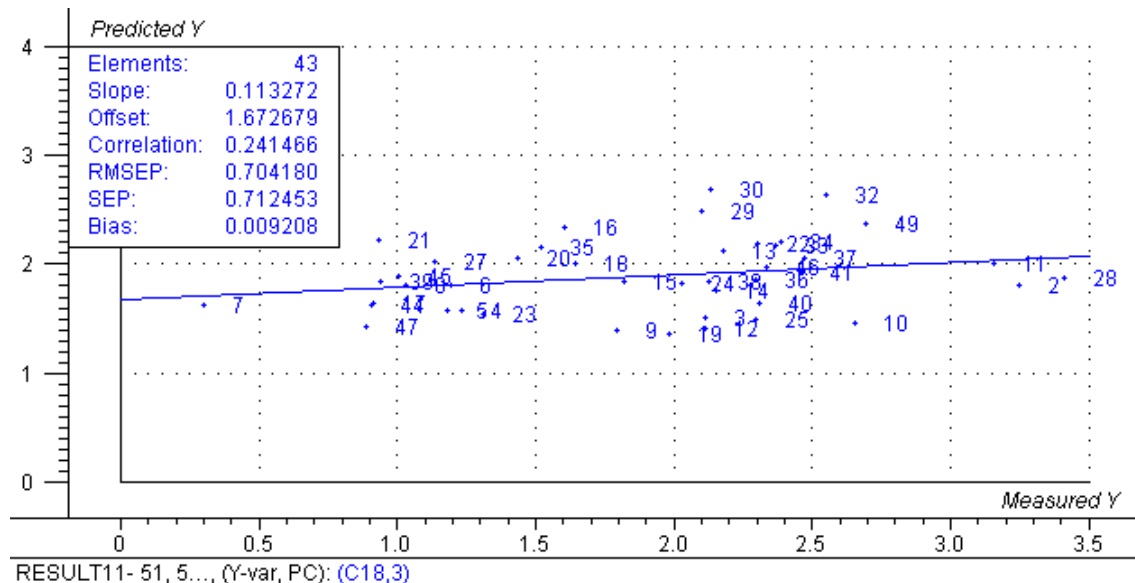
Figure 48: Predicted vs Measured Y Plot.
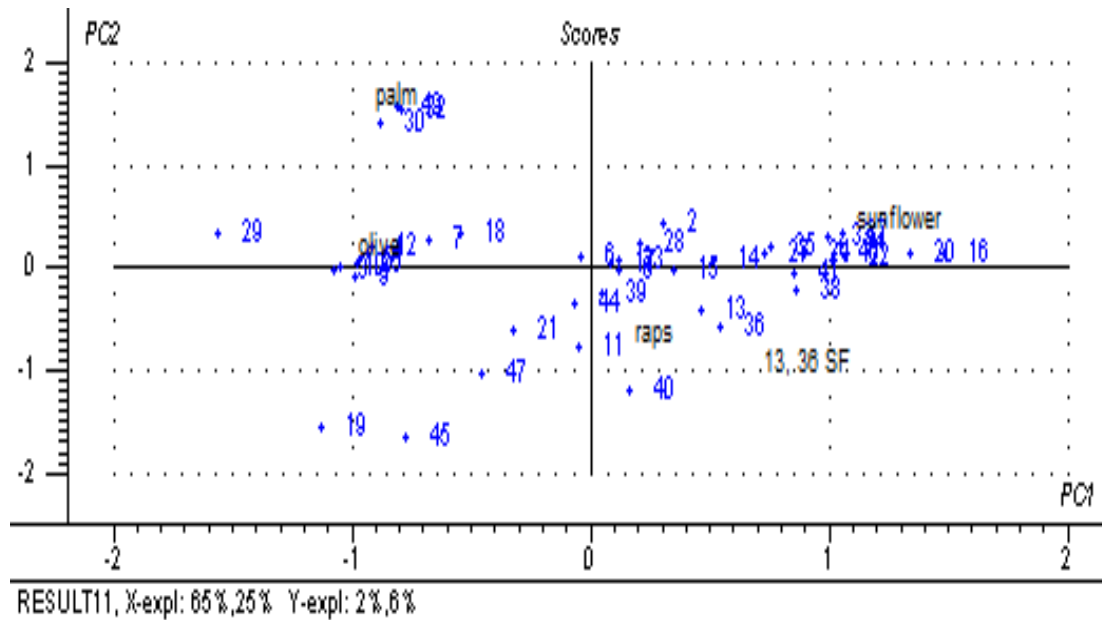
Figure 49: Score plot.

3.1.2.7. C18:1

PLS1 Regression with 49 samples with cross validation 6segments with 8 samples in each.
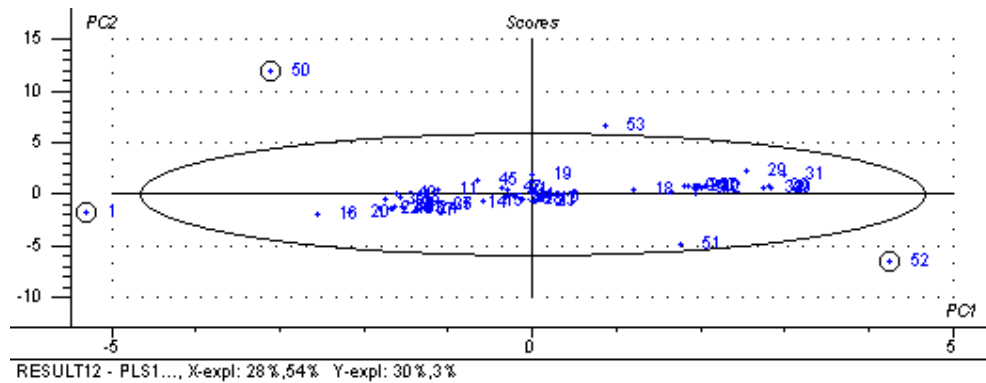
Figure 50: Hoetelling T^2 plot.
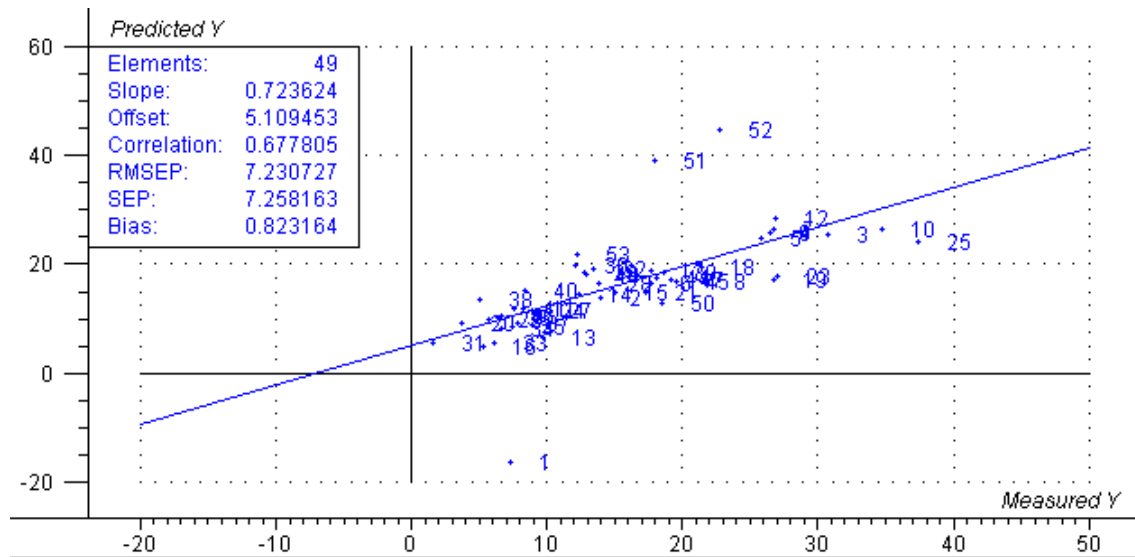
Figure 51: Predicted vs. Measured Y plot.

RMSEP: 7.230727, Correlation: 0.677805, Slope: 0.723624, Bias: 0.823164.
RMSEC: 4.672099, Correlation: 0.844424, Slope: 0.713952, Bias: -5.410e-07.

When the outliers 50, 1, 52, 51, 31 and 53 are also removed and recalculated a good model with high correlation for 43 elements and 5 components is obtained. Correlation and slope are good with lower RMSEP value. The X explanation is 76% and Y explanation is 49% for the 1st component whereas it is 15 and 14% respectively for the 2nd component. Overall the model is good showing groups of similar types of oils but there are still few samples like 23, 25 and 10 falling away from the regression line Predicted vs. measured plot.
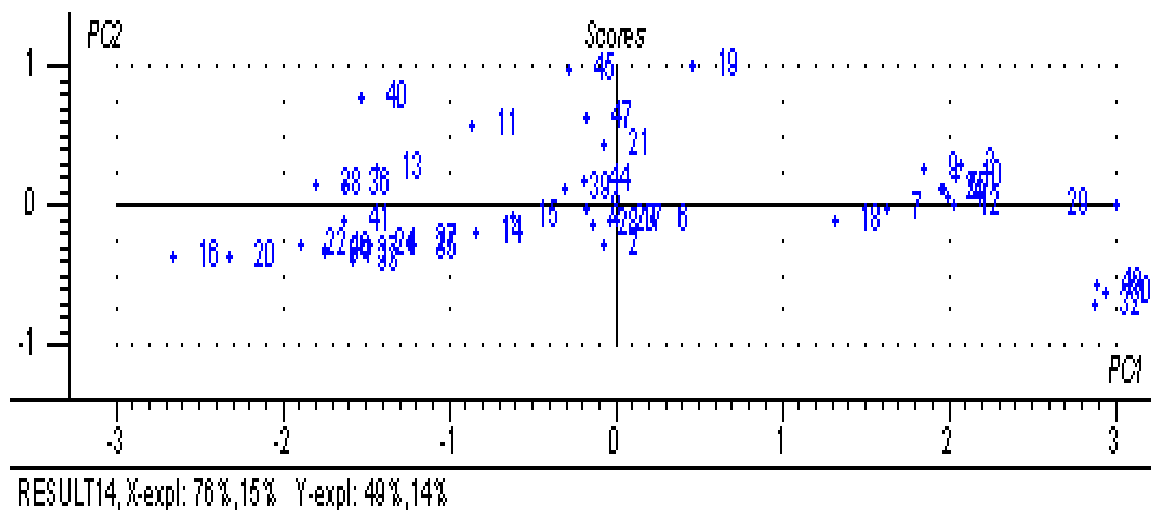
Figure 52: Score plot.
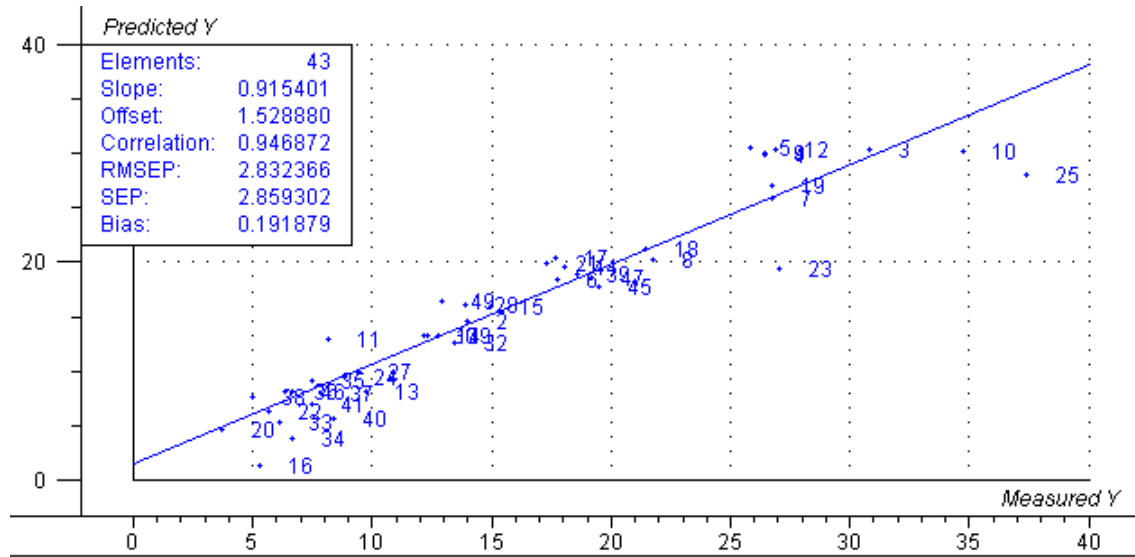
Figure 53: Predicted vs. Measure Y plot.

RMSEP: 2.832366, Correlation: 0.946872, Slope: 0.915401, Bias: 0.191879.
RMSEC: 2.428715, Correlation: 0.960887, Slope: 0.923304, Bias: 1.552e-07.

3.1.2.8. C18:2

PLS1 Regression with 49 samples with cross validation 6segments with 8 samples in each. The obtained regression model is with the following correlation:

RMSEP: 1.567827, Correlation: 0.958576, Slope: 0.979354, Bias: 0.121386.
RMSEC: 0.746791, Correlation: 0.990239, Slope: 0.980573, Bias: -5.263e-06.
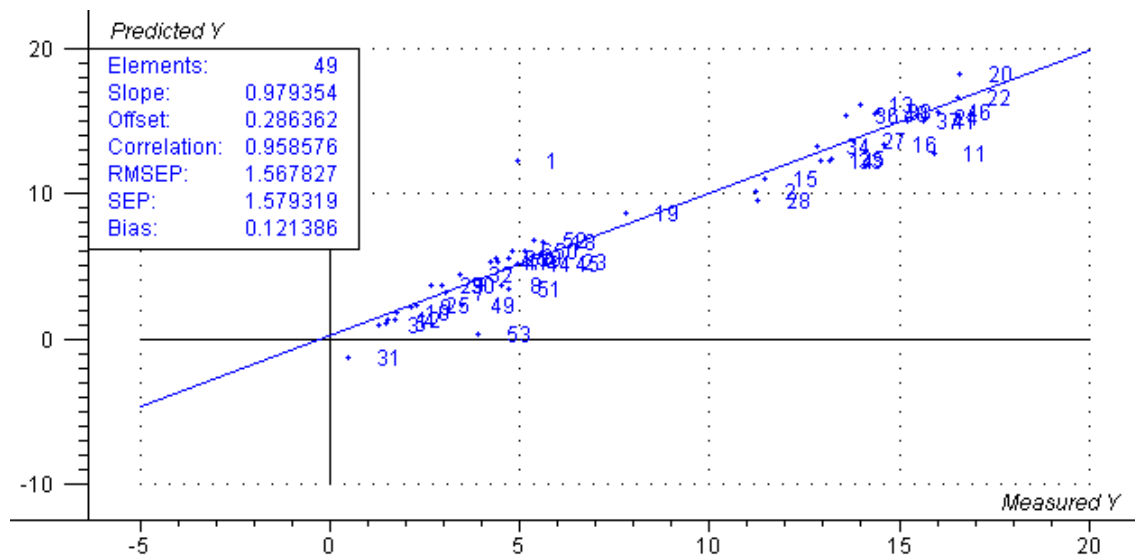
Figure 54: Predicted vs. Measured Y plot.

The number of principal components is found to be 9 as per the y- variance plot.

Based on the score plot observation it is observed that 50, 52, 1 and 53 are found as outliers in this Hoetelling T^2 plot, the Leverage plot also shows the same with all these outliers having values above 0.5. But in the residual sample variance plot the outliers observed are 31, 50, 52. Totally 50, 1, 31, 52, 53, 51 are removed as outliers and recalculated without these.
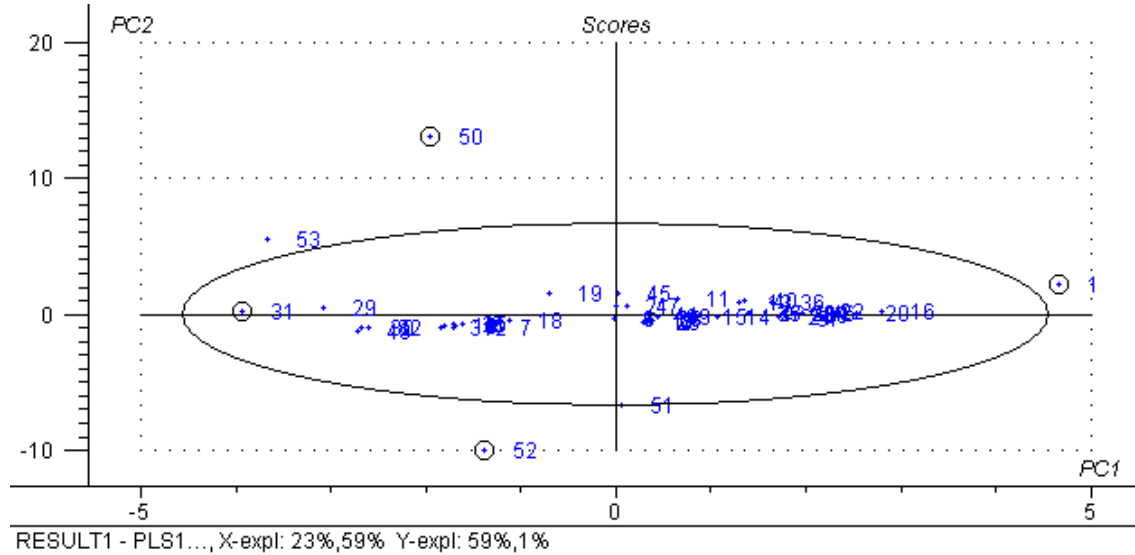
Figure 55: Hoetelling T^2 plot.

The final model has resulted as follows.

RMSEP: 1.049918, Correlation: 0.981412, Slope: 0.957959, Bias: 0.012895.

RMSEC: 0.739557, Correlation: 0.990812, Slope: 0.981709, Bias: -2.243e-06.
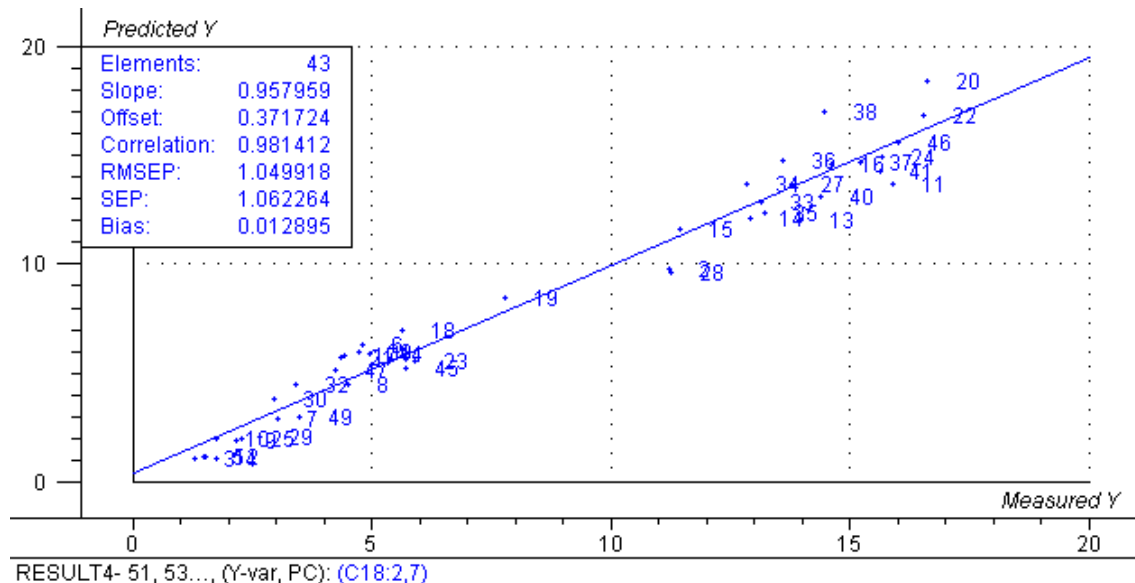
Figure 56: Predicte vs. Measure Y plot.
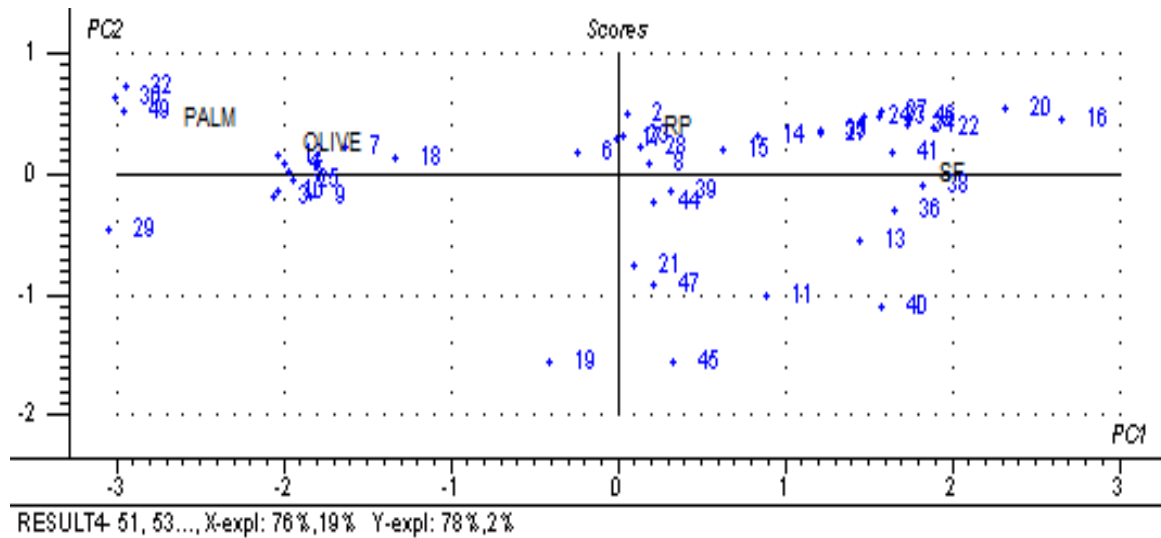
Figure 57: Score plot.

All the samples are found to be in clusters of olive oil, palm, Rape oil (RP) and sunflower oil (SF).

3.1.2.9. C18:3

PLS1 regression of 49 samples with 6 segments gave the following model with plots are discussed below.

RMSEP: 2.888012, Correlation: 0.251616, Slope: 0.142324, Bias: -0.0947483.
RMSEC: 2.382380, Correlation: 0.542845, Slope: 0.294681, Bias: -7.470e-08.
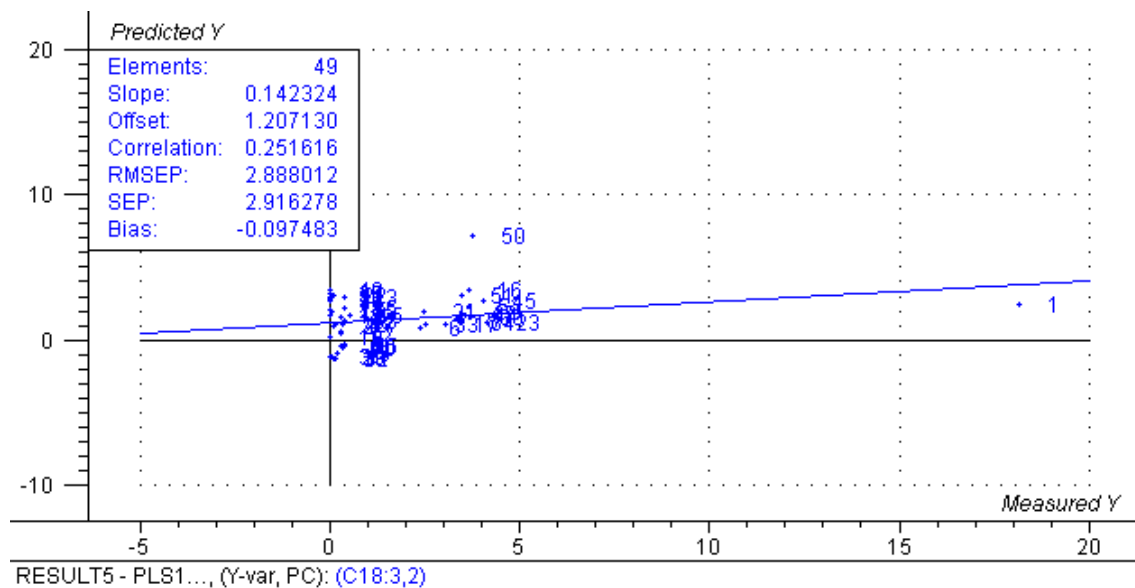
Figure 58: Predicte vs. Measure Y plot.

The outliers observed in this model are 50, 1, 51, and 52.
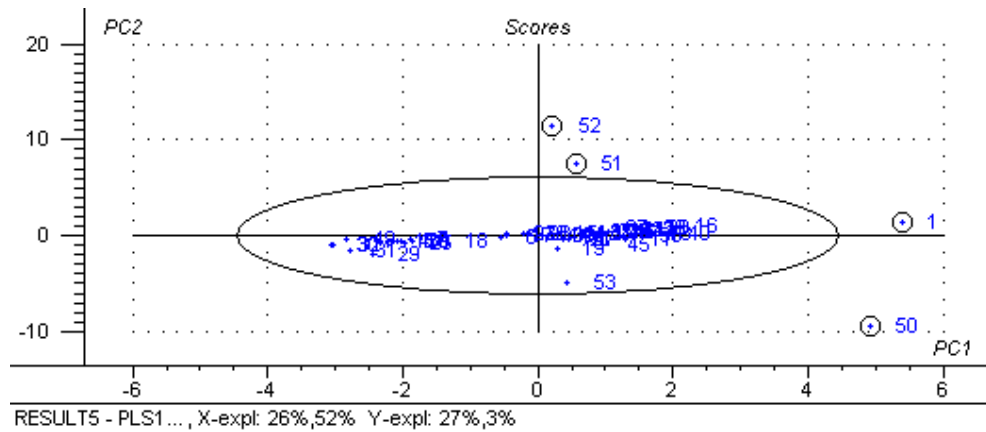
Figure 59: Hoetelling T^2 plot.

From the y- variance plot it is observed that two components can be used for modelling. When the outliers 50, 51, 52, 53, 31 and 1 are removed then the obtained model has the following correlation values which have increased compared to the previous model. Four components are used to model 43 elements. The predicted vs. measured plot shows lots of samples falling away from the regression line.

RMSEP: 1.171973, Correlation: 0.643847, Slope: 0.576191, Bias: 0.098912.
RMSEC: 0.878061, Correlation: 0.795846, Slope: 0.633370, Bias: 1.142e-06.
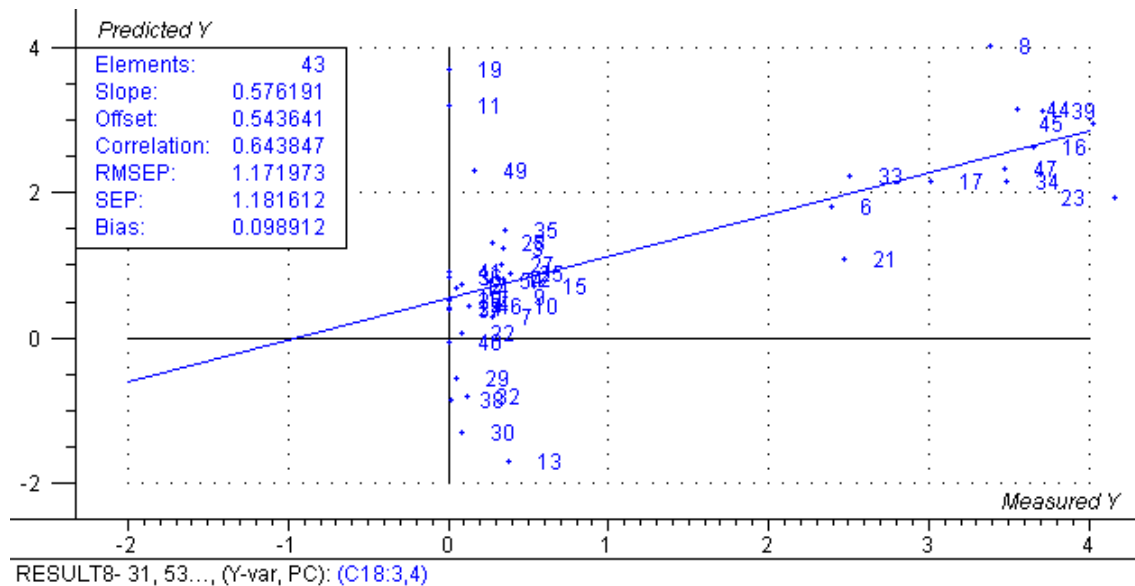
Figure 60: Predicte vs. Measured Y plot.
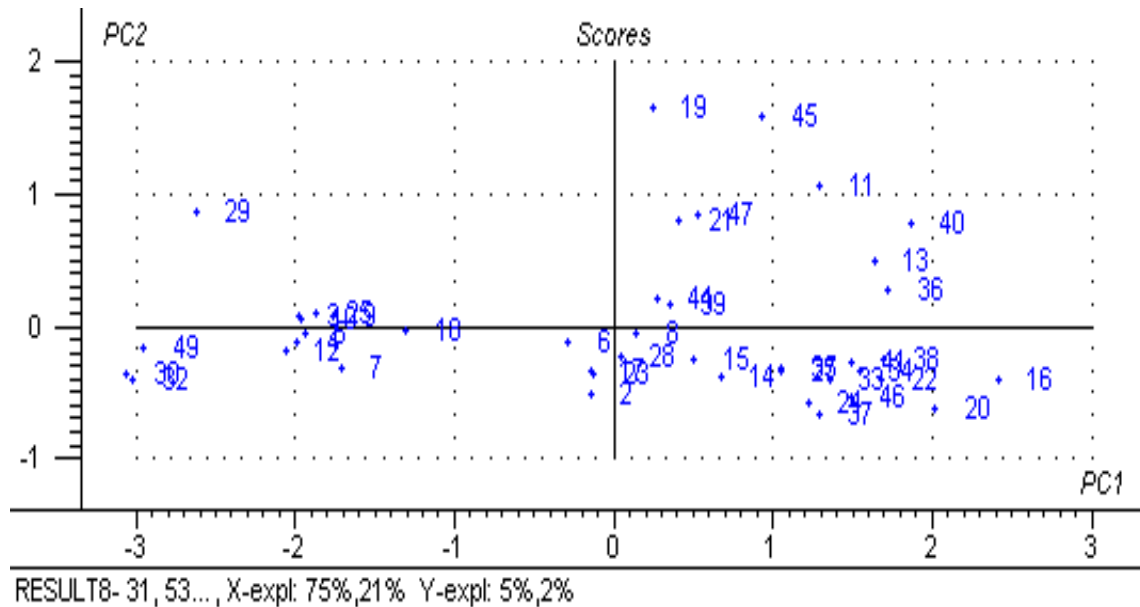
Figure 61: Score plot.

3.2. FOSS 01L – 100-60µm:

The entire 53 samples spectrum has been collected from OenoFoss, a Foss FTIR instrument at a shorter path length of 60µm and longer path length of 100µm. The instrument collects the spectral difference of these two path length and the collected single beam spectrum is observed for air bubbles.
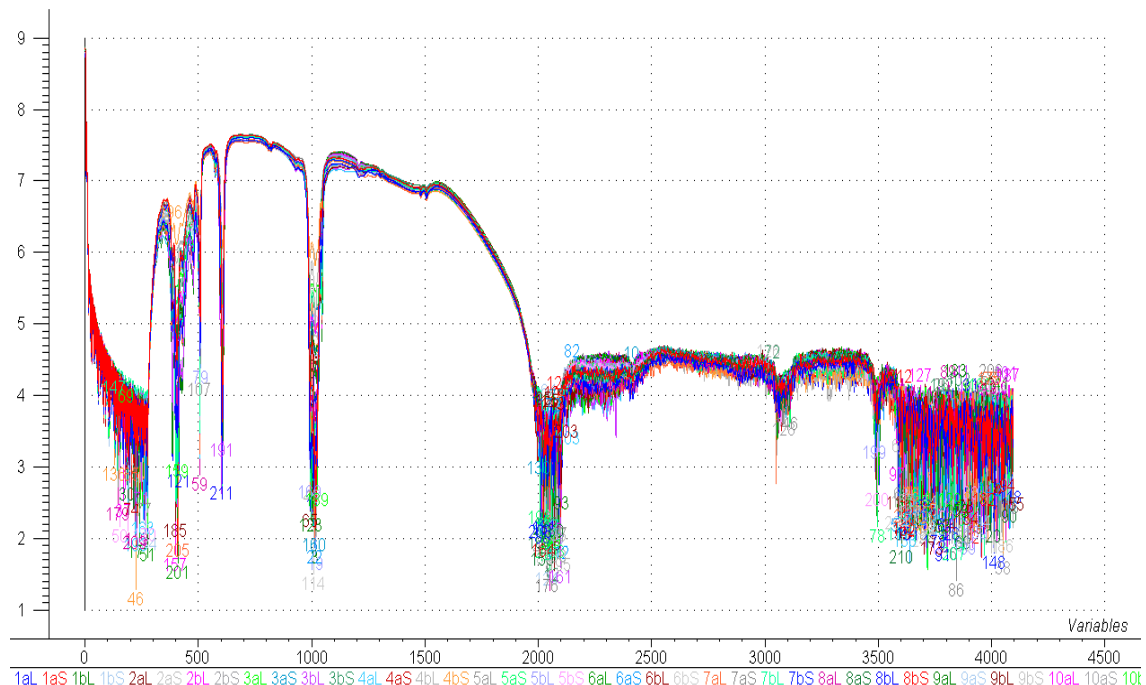
Figure 62: Line plot of single beam spectrum.

Based on the above plot of the single beam the variable number 1018 has been chosen and line plot of that shows samples. The samples above the level of approximately 4, a line has been drawn to consider it as a limit above which the samples are removed and are considered as having air . The removed samples are being replaced by copying its replicate however the samples that have both the replicates with air are completely removed. The replicates of 27 and 46 consists air so these have been kept out of analysis.
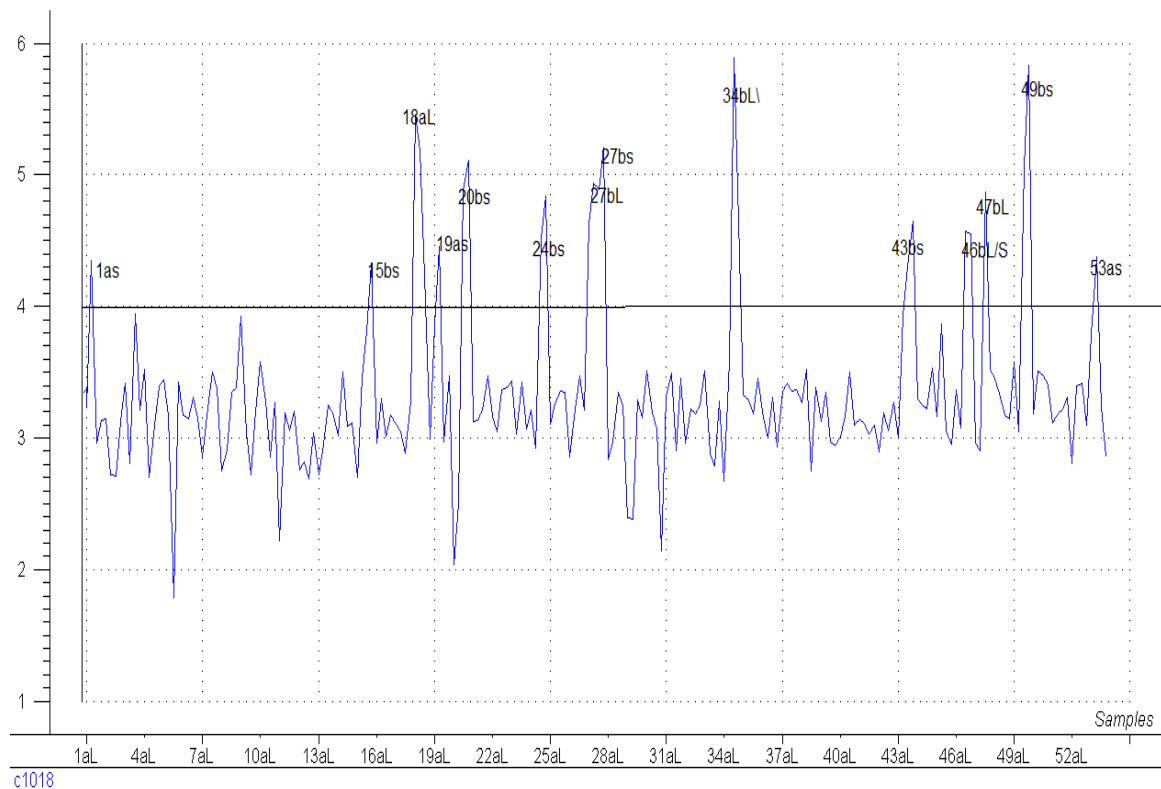
Figure 63: Line plot of 1018 variable.

Thereby the samples27 and 46 are kept out of calculation while regression. The regression PLS1 has been carried with 48 samples with cross validation of 6 segments with 8 samples in each segment.
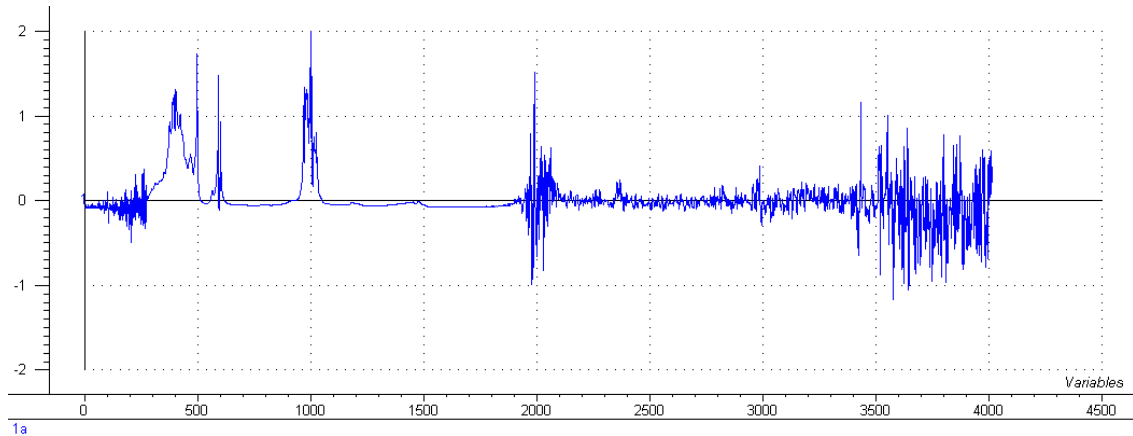
Figure 64: Line plot of raw spectra.

The line plot of 4096 variables is as above which is obtained when data of FTIR is imported and when all the noise has been deleted then it gave the line plot as below line plot with 1338 variables.
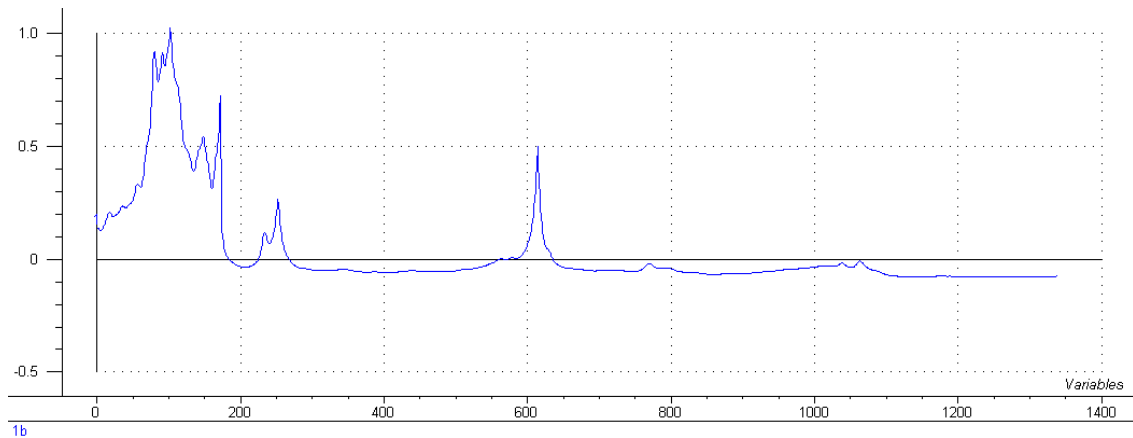
Figure 65: Line plot of A100-60

Spectra of the range 2743 cm^{-1}– 2964 cm^{-1}; 1492 cm^{-1}-1836 cm^{-1}; 1033 cm^{-1}-1377cm^{-1} are found to have meaningful information. Due to optical path length difference scattering effect can be present in these FTIR spectra that is collected so MSC pre-treatment is used to compare with SNV before regression. The X-axis values are multiplied by 2.87 to obtain the wavenumber units.

3.2.1. MSC Pre-treatment:

100-60µm pathlength spectra is pretreated using Multiplicative scatter correction (MSC). C18:1 parameter is used to check whether MSC or SNV is good for modelling.

3.2.1.1. C18:1
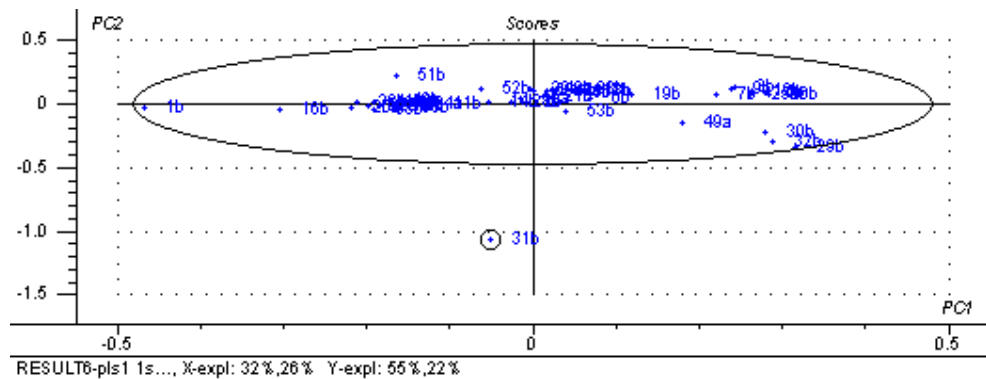
Figure 66: Hoetelling T^2 plot.
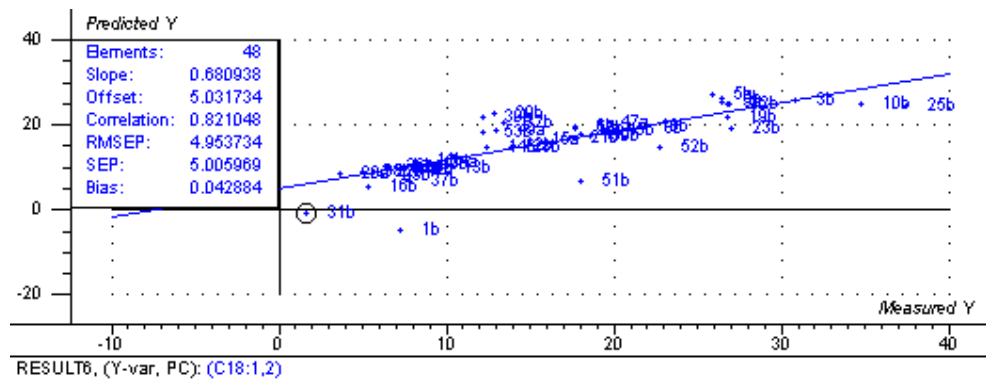
Figure 67: Predicted vs. Measured Y plot.

RMSEP: 4.953734, Correlation: 0.821048, Slope: 0.680938, Bias: 0.042884.
RMSEC: 4.176633, Correlation: 0.876514, Slope: 0.768277, Bias: 2.633e-07.
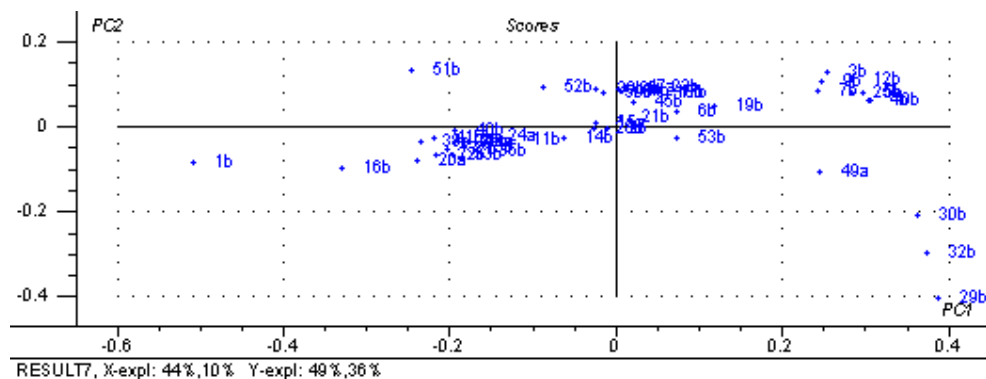
Figure 68: Score Plot.
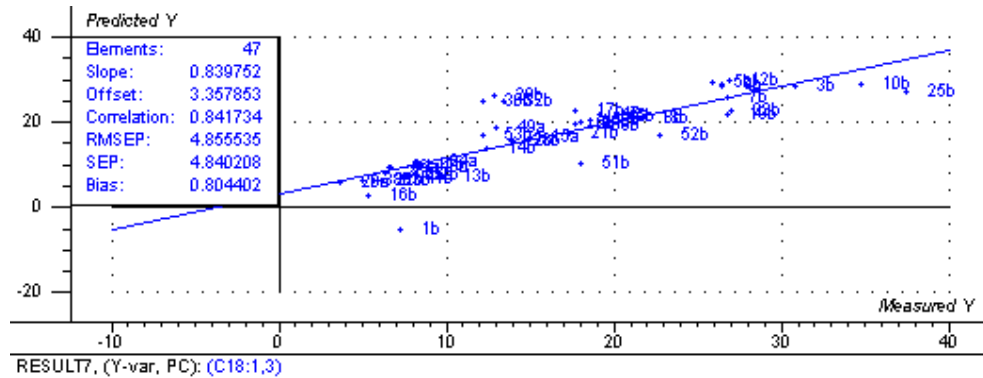
Figure 69: Predicted vs. Measured Y Plot.

RMSEP: 4.855535, Correlation: 0.841734, Slope: 0.839752, Bias: 0.804402.
RMSEC: 2.767005, Correlation: 0.945808, Slope: 0.894553, Bias: 6.189e-07.

The final model was obtained with little higher value of correlation, slope and slight decrease in the RMSEP value.

Before removal of outliers:

C18:1	SNV	MSC
RMSEP	191.5534	4.953734
Correlation	0.892013	0.821048
Slope	0.855771	0.680938
Bias	-4.259210	0.042884
RMSEC	160.9604	4.176633
Correlation	0.923288	0.876514
Elements	48	48
Components	3	2

Table 4: Pre-treatment Comparison.

After removal of outliers

C18:1	SNV	MSC
RMSEP	96.84937	4.855535
Correlation	0.966598	0.841734
Slope	0.935550	0.839752
Bias	-3.533085	0.804402
RMSEC	76.64680	2.767005
Correlation	0.979185	0.945808
Elements	36	47
Components	3	3

Table 5: Pre-teatment results comparison after removal of outliers.

The correlation and slope values are higher in SNV pretreatment than in MSC pretreated models.

3.2.2. 100-60µm SNV:

All the parameters have been modelled by PLS1 regression with cross validation of 6 segements with 8 samples in each segment. The results of regression are discussed for each parameter.

3.2.2.1. C12:

PLS1 regression of C12 as y variable with cross validation of 6 segments with 8 samples in each segment gave the following results.

RESULT1- c12 pl..., X-expl: 38%,21% Y-expl: 71%,23%
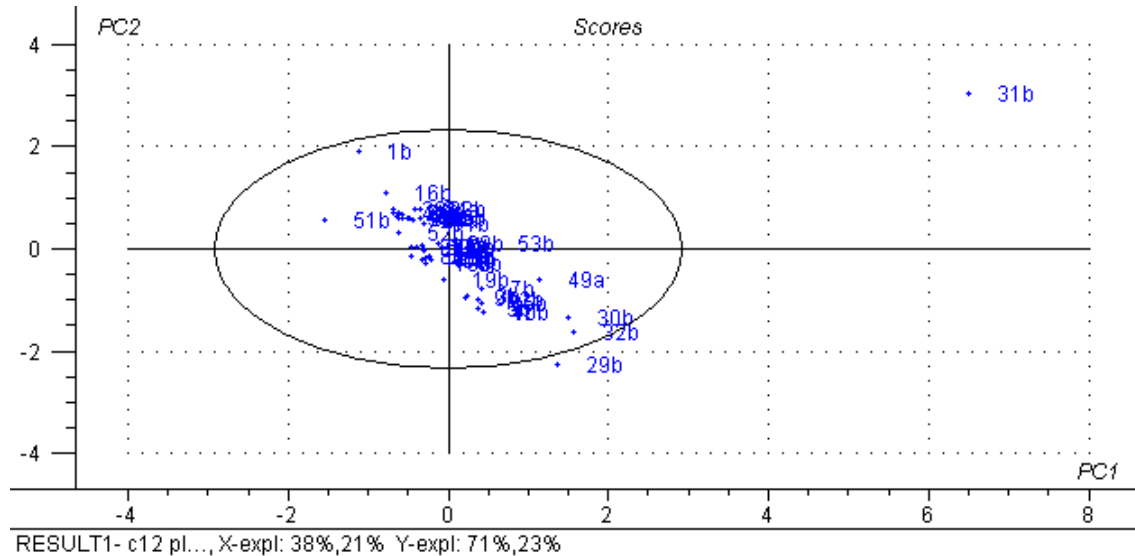
Figure 70: Hoetelling T^2 plot.

RESULT1- c12 pl..., (Y-var, PC): (C12,2)
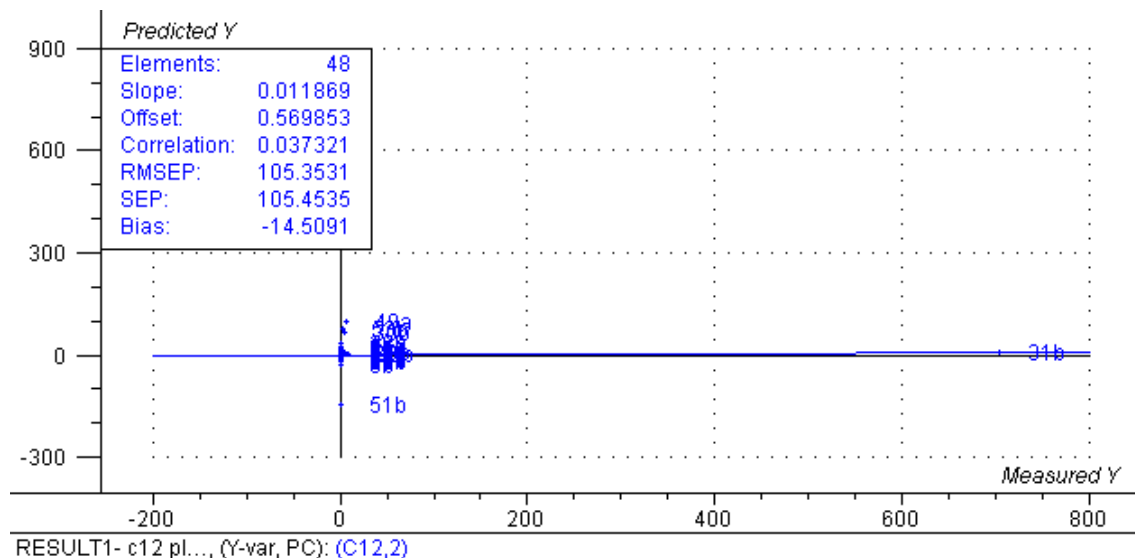
Figure 71: Predicted vs. Measured Y plot.

The Hoetelling T^2 plot shows 31b as an outlier which is also observed in the predicted vs. observed plot. 31b has been removed and recalculated without this sample, then the following model is obtained.

RMSEP: 105.3531, Correlation: 0.037321, Slope: 0.011869, Bias: -14.5091.
RMSEC: 24.04459, Correlation: 0.070976, Slope: 0.942795, Bias: 2.294e-05.

Based on the Residual validation variance plot 2 PCs (principal components) are found optimal in this model.
                        The obtained model when sample 31b is removed and recalculated is as shown in the below plots. The model shows few outliers which when removed gave very low correlation. So no outliers are removed any more. The correlation coefficient has increased for the model whose plots are shown below.
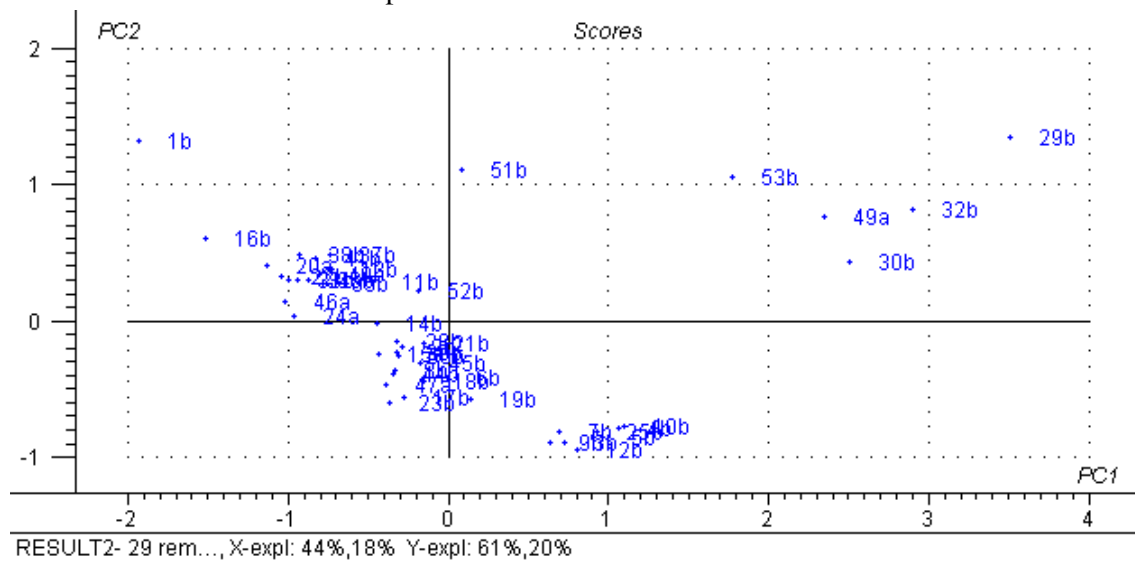
Figure 72: Hoetelling T^2 plot.

The samples 53b, 49a, 30b, 32b, 29b are palm oils and these are found to be extreme right in the Figure 72.
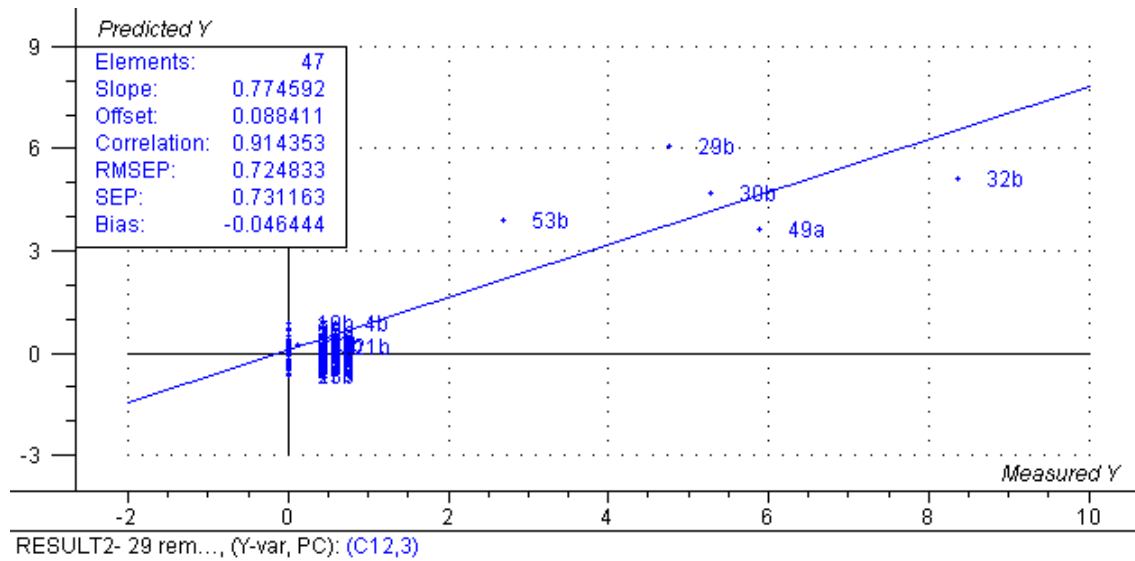
Figure 73: Predicted vs. Measured Y plot

The above observed outliers are actually not outliers as their removal from the model is actually decreasing the correlation and the resultant model came up with more outliers. So the above model is good with high correlation.

RMSEP: 0.724833, Correlation: 0.914353, Slope: 0.774592, Bias: -0.046444.
RMSEC: 0.672816, Correlation: 0.924251, Slope: 0.854239, Bias: -4.708e-08.

3.2.2.2. C14:

The PLS1 regression of the samples with C14 as Y variable has been modelled and the Result1 obtained has a correlation with sample number outlier
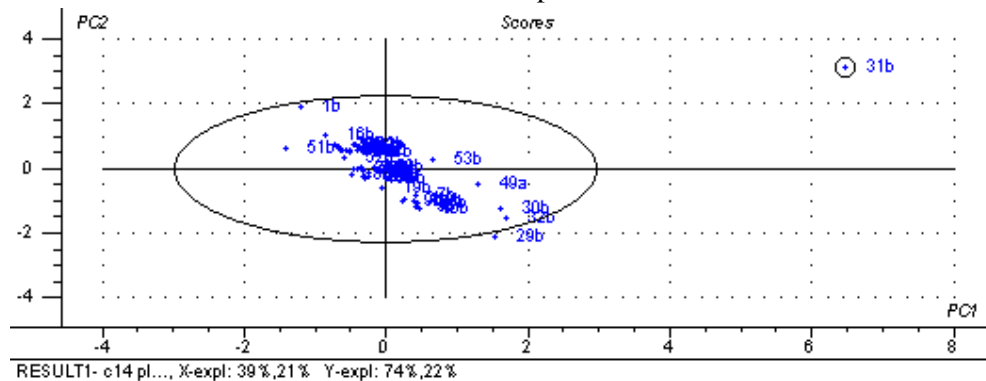
Figure 74: Hoetelling T^2 plot.
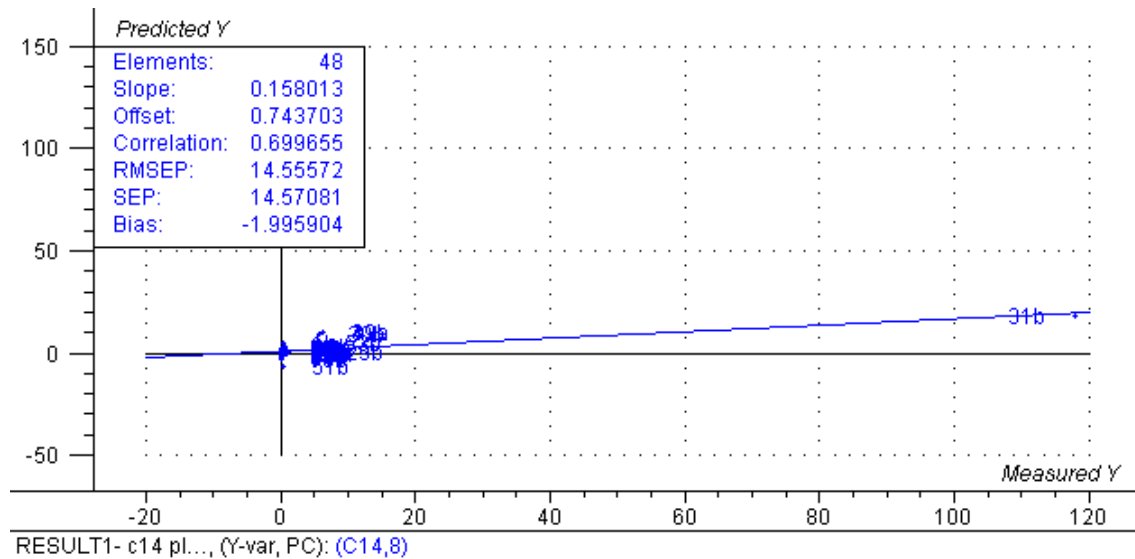
Figure 75: Predicted vs. Measured Y plot

RMSEP: 14.55572, Bias= -1.995904. Correlation: 0.699655 slope: 0.158013.
RMSEC: 0.721856, Bias: 3.907e-06, C0rrelation: 0.999078 slope: 0.998158.
When it is recalculated without 31b then the obtained model plots are as follows:
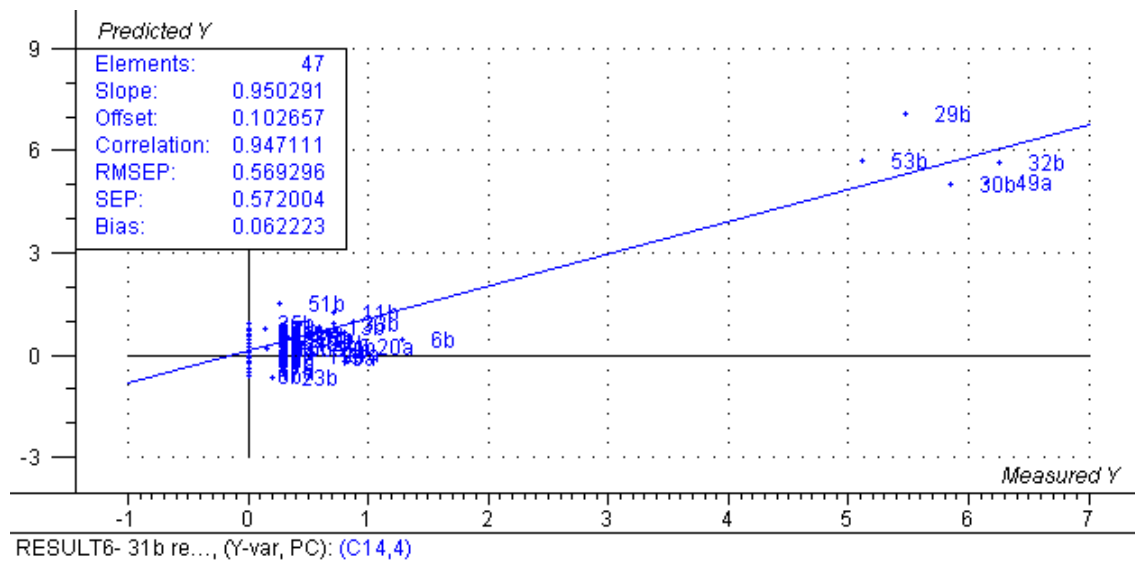
Figure 76: Predicted vs. Measured Y plot

The above plot shows few samples as a group which are actually palm oils.
The above plot has the following values.

RMSEP: 0.569296, Slope: 0.950291, Correlation: 0.947111, Bias: 0.062223.
RMSEC: 0.433534, Slope: 0.937700, Correlation: 0.968349, Bias: -1.664e-07.
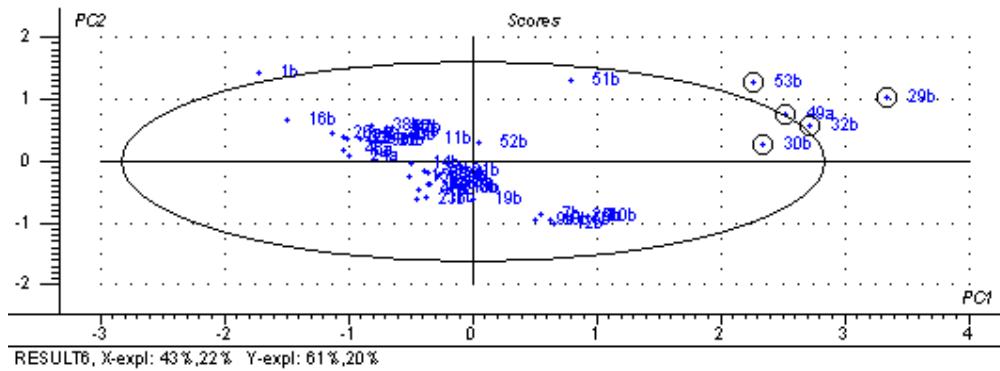
Figure 77: Hoetelling T^2 plot.

The Hoetelling T2 plot and Predicted vs Measured plot shows palm oils (53b, 29b, 49a, 32b, 30b) as outliers which when removed lowered the correlation values to a greater extent making a model bad. The plots below are related to the one when the above palm oil outliers are removed.
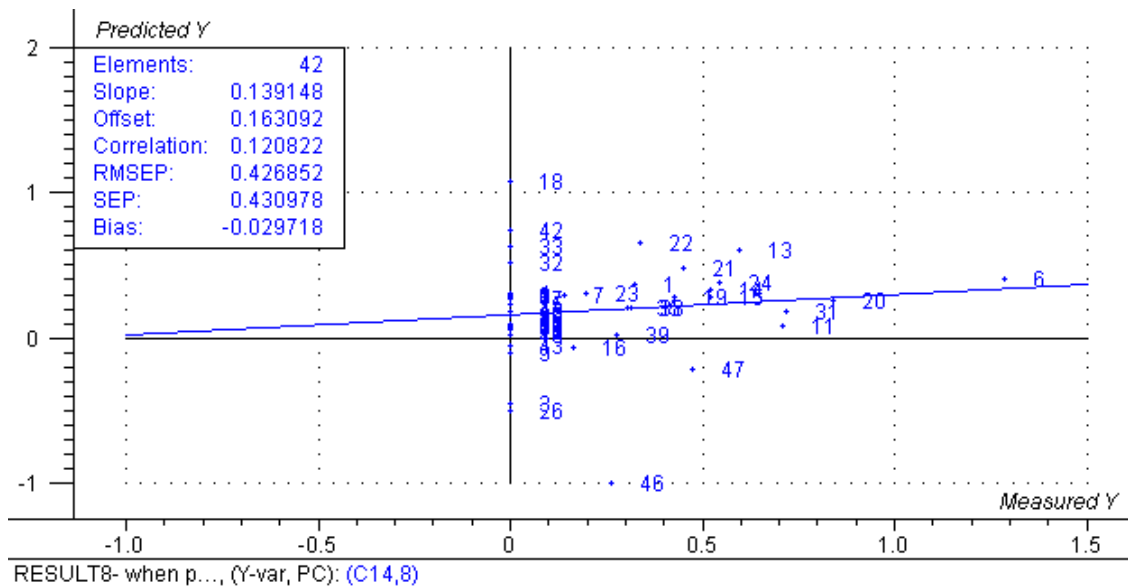
Figure 78: Predicted vs. Measured Y plot
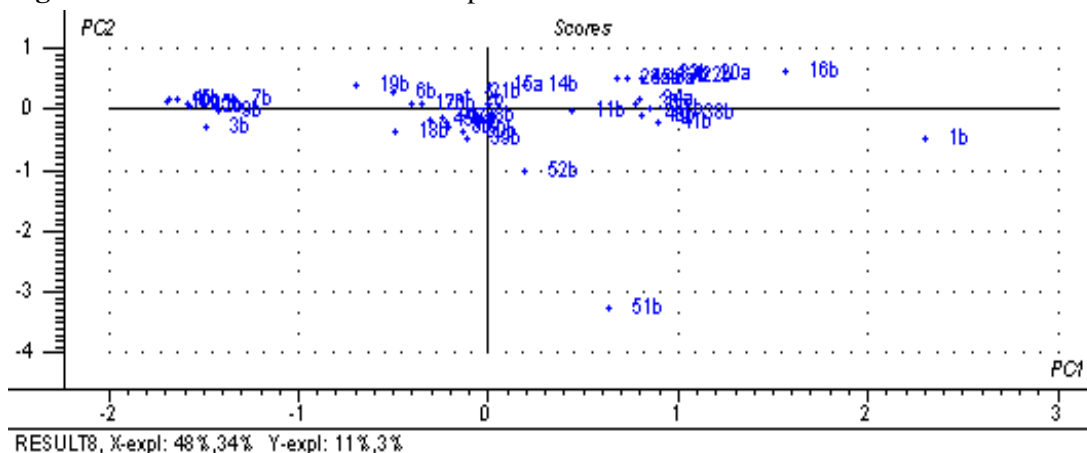
Figure 79: Score plot.

The correlation has decreased a lot.

RMSEP: 0.426852, Correlation: 0.120822, Bias=-0.029718, Slope: 0.139148.
RMSEC: 0.750446, correlation: 0.750446, Bias= 1.245e-07, Slope: 0.563170.

3.2.2.3. C16: 100-60

PLS1 regression of FTIR spectra of path length 100 &60 µm resulted in the following model
.
RMSEP: 111.2785, Correlation: 0.502285, Slope: 0.872277, Bias: 18.74160.
RMSEC: 11.63423, Correlation: 0.987141, Slope: 0.987141, Bias: 5.603e-06.
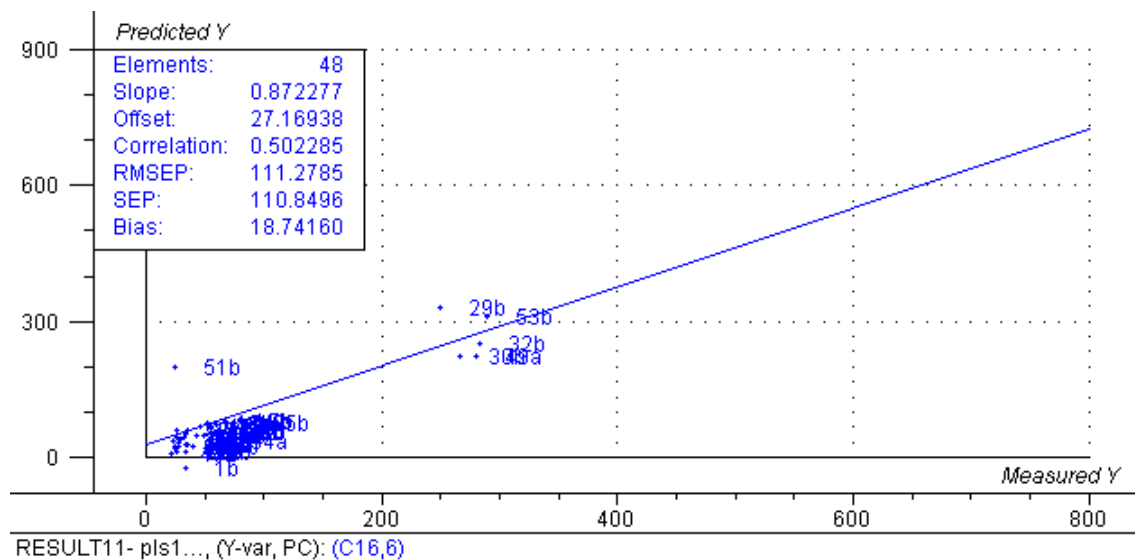
Figure 80: Predicted vs. Measured Y Plot.
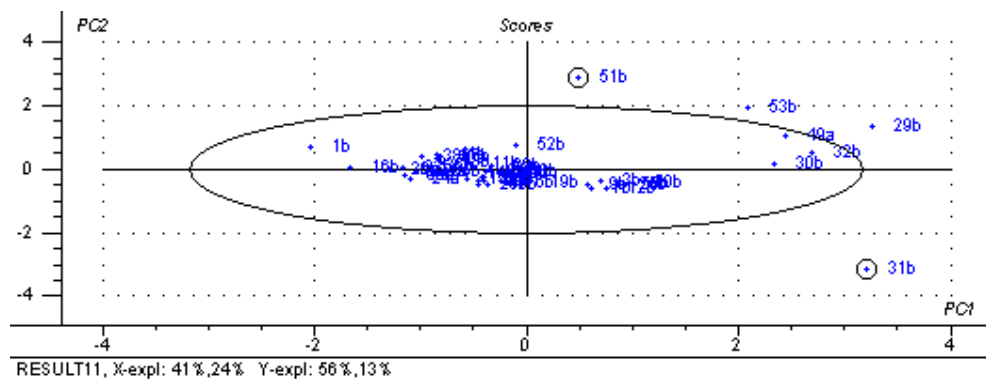
Figure 81: Hoetelling T^2 plot.

When 51b, 31b are removed from the model above, it resulted into the following correlation. The plots below are obtained as a result of performing regression after the removal of the above outliers.

RESULT12, X-expl: 53%,22% Y-expl: 72%,16%
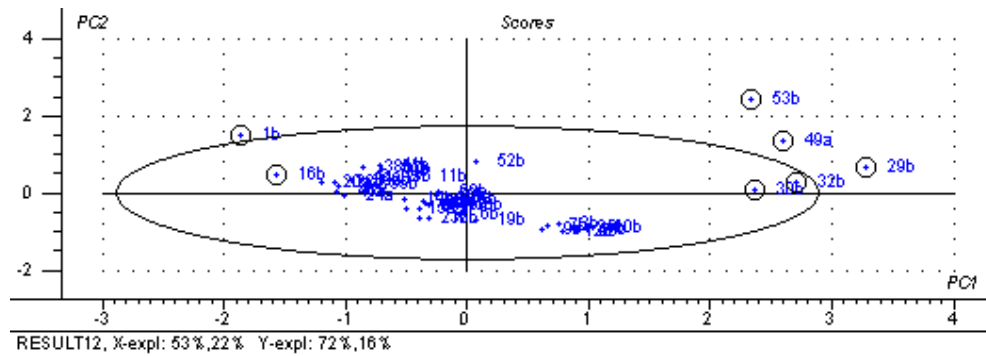
Figure 82: Hoetelling T^2 plot.

RESULT12- 51b,..., (Y-var, PC): (C16,6)
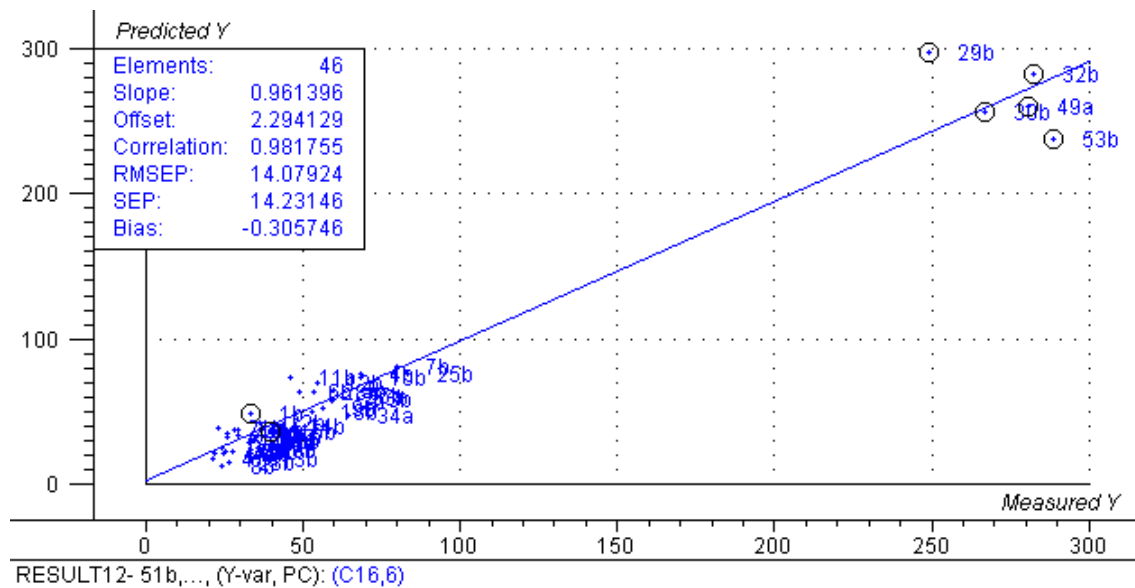
Figure 83: Predicted vs. Measured Y Plot.

RMSEP: 14.07924, Correlation: 0.981755, Slope: 0.961396, Bias: -0.305746.
RMSEC: 6.861839, Correlation: 0.995694, Slope: 0.991406, Bias: -2.488e-07.

The plot shows the outliers 1b, 16b, 53b, 49a, 30, 32b, 29b, 52b when removed and recalculated resulted in a good Hoetelling T2 plot with no outllers.

Removal of all the outliers observed in the above plots resulted into a model with 52b as an outlier and the removal of all these outliers resulted into a very good model with the following correlation.
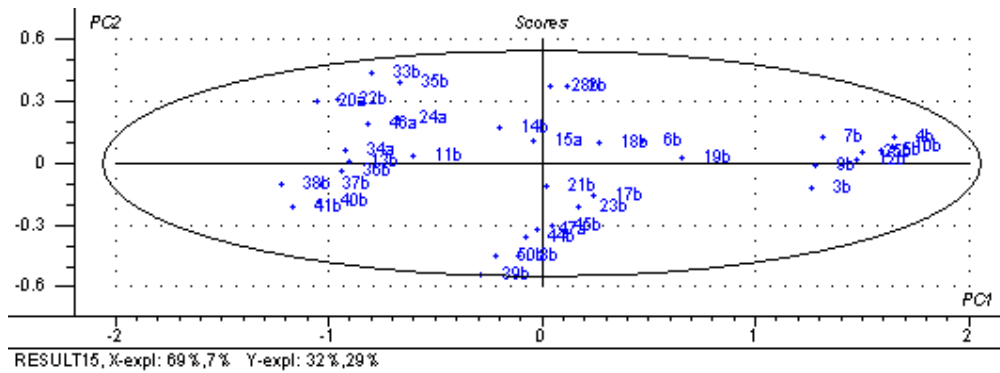
Figure 84: Hoetelling T^2 plot.
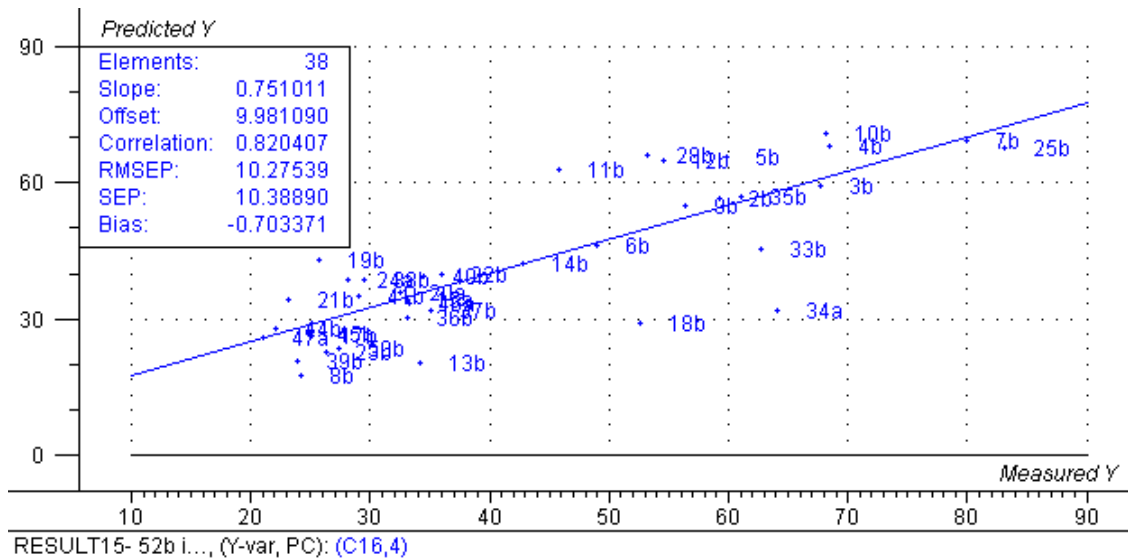
Figure 85: Predicted vs. Measured Y plot.

RMSEP: 10.27539, Correlation: 0.820407, Slope: 0.751011, Bias: -0.703371.
RMSEC: 7.088608, Correlation: 0.916167, Slope: 0.839361, Bias: 3.012e-07.

But this model has lower R^2 Value compared to the model with the palm oils.

3.2.2.4. C18:

PLS1 regression of the 48 samples with 6 segments and 8 samples in each resulted in the following model.
The model plots showed outliers 29b, 31b which are observed in the Hoetelling T^2 plot. The leverage plot also supports this. The prediction vs. measured plot shows the following correlation:

RMSEP: 42.60182, Correlation: 0.221633, Slope: 0.227728, Bias: 0.025163.
RMSEC: 20.65225, Correlation: 0.789886, Slope: 0.623920, Bias: 9.815e-06.
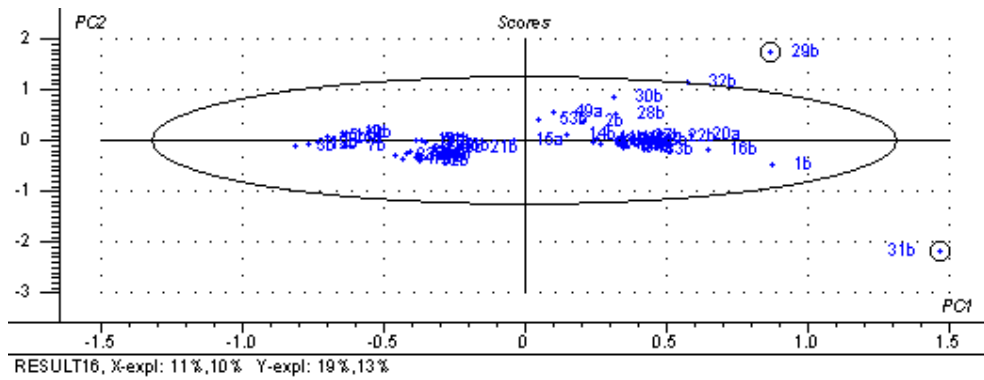
Figure 86: Hoetelling T^2 plot.
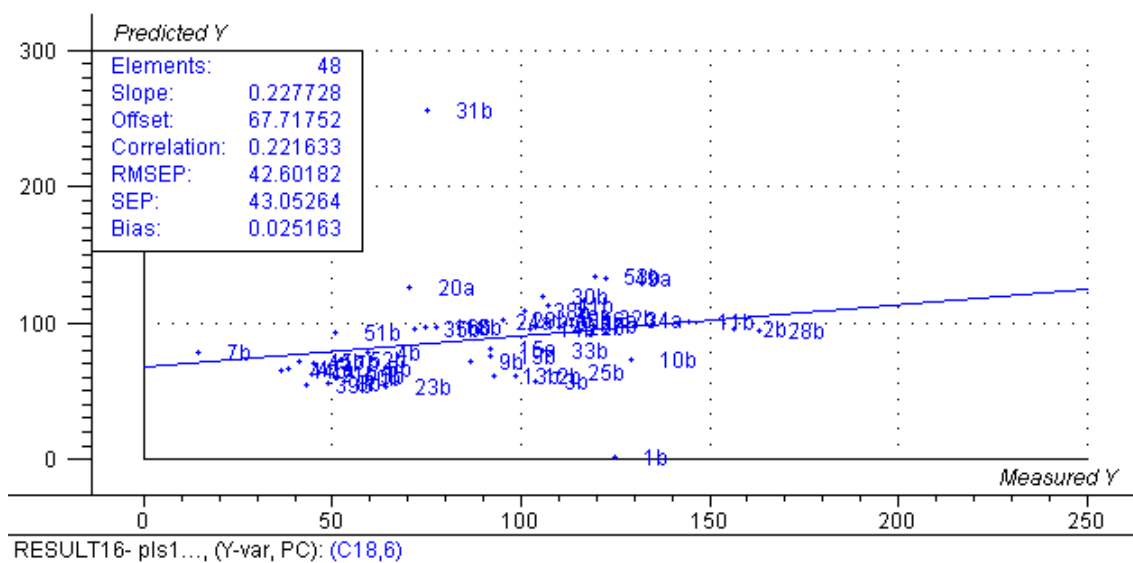
Figure 87: Predicted vs. Measured Y plot

The above plot also shows 31b as an extreme outlier. When the outliers 29b, 31b are removed then the outliers related to palm oils are observed as outliers. These outliers 49a, 30b, 32b as well as 7b, 1b, 51b, 53b are removed to obtain a final model.

The final model when all the above mentioned outliers are removed is as follows:
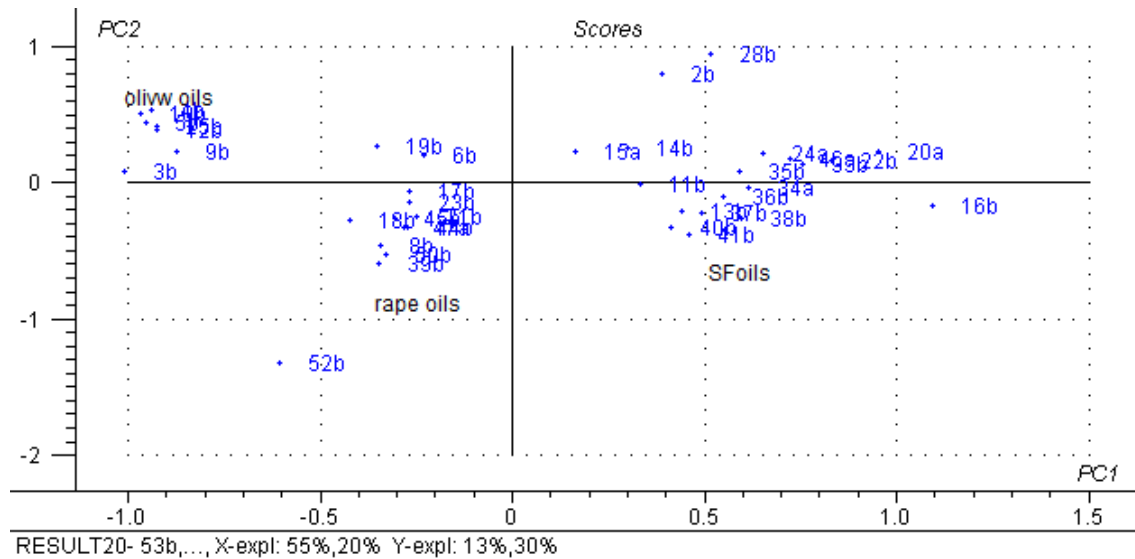
Figure 88: Score plot.

RMSEP: 21.08088, Correlation: 0.774063, Slope: 0.652380, Bias: -1.140582.
RMSEC: 15.45742, Correlation: 0.887087, Slope: 0.786924, Bias: - 6.847 e-06.
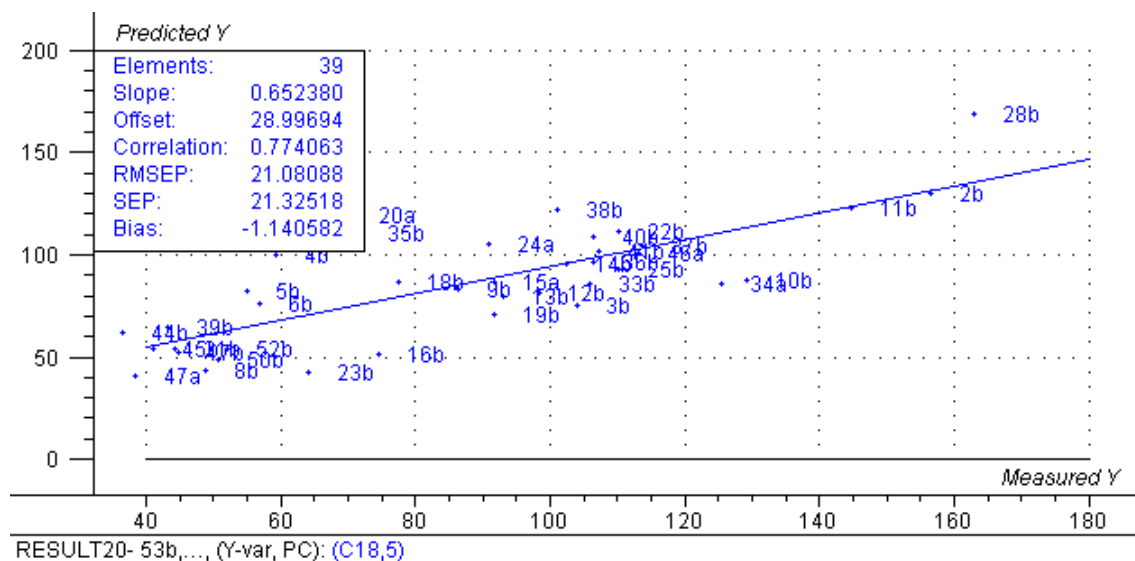
Figure 89: Predicted vs. Measured Y plot.
The model with good fit using 5 components for 39 samples is obtained for C18.

3.2.2.5. C18-1

When the samples have been performed PLS1 regression it resulted in the following model whose correlation and outliers are described below along with necessary plots. Cross validation was carried with 6 segments and 8 samples in each segment.

However in this model as well the few palm oils along with other outliers are observed whch are removed to make a better model.
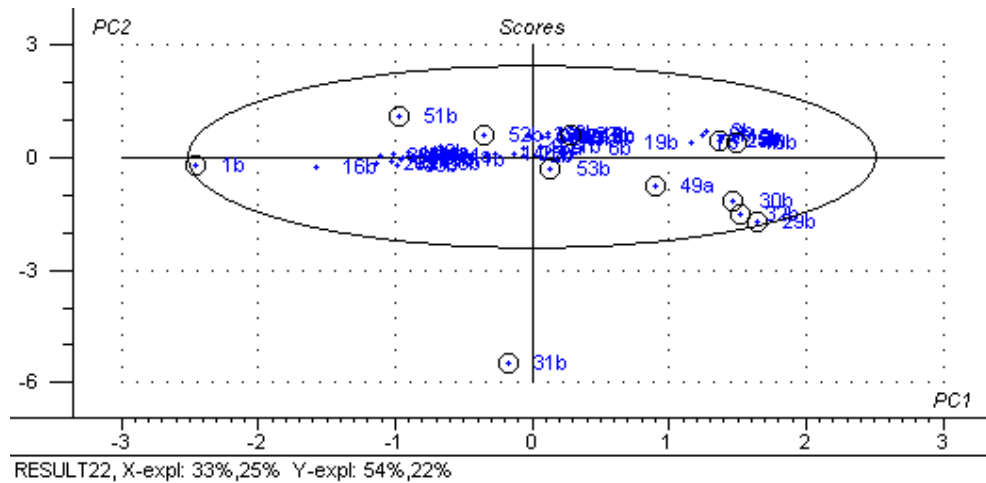
Figure 90: Hoetelling T^2 plot.

   The Hoetelling T2 plot shows only 31b as an outlier but the x-y relation plot shows many outliers which have been observed as outliers in the Hoetelling T2 plot after removing the 31b and these are supported as outliers by X-Y relation plot as well as predicted vs. Measured Y plot along with Hoetelling T2 plot. Based on these observations all the samples marked in the plots i.e. 1b, 31b, 51b, 52b, 53b, 49a, 30b, 32a, 29b, 10b, 25b are removed as outliers and then the obtained model results are discussed after the plots.
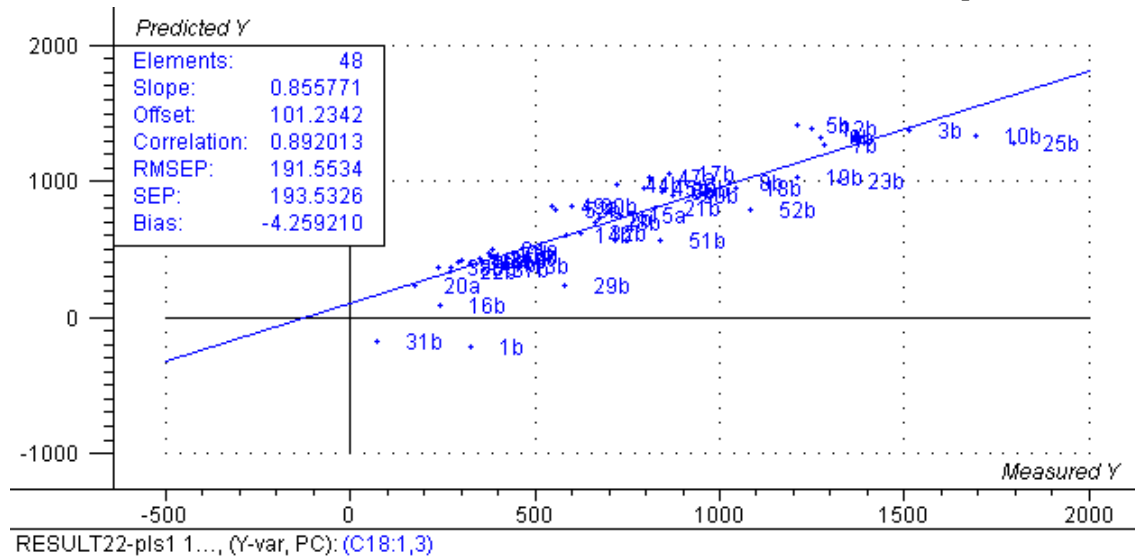
Figure 91: Predicted vs. measured plot.

RMSEP: 191.5534, Correlation: 0.892013, Slope: 0.855771, Bias: -4.259210.
RMSEC: 160.9604, Correlation: 0.923288, Slope: 0.852461, Bias: -0.000130.
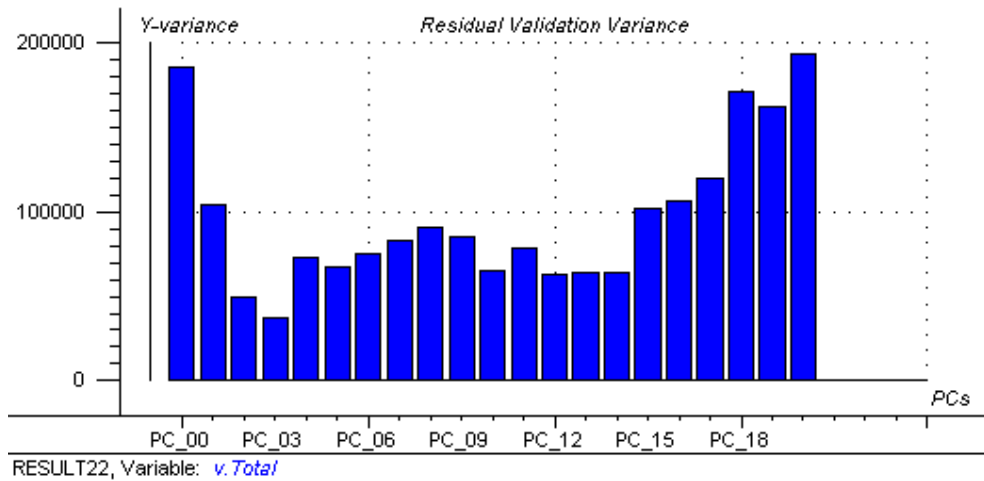
Figure 92: Residual validation variance plot.

The above plot shows 3 PCs are essential for modelling.

When all the above mentioned outliers are removed the following model is obtained. The results of which are discussed below.
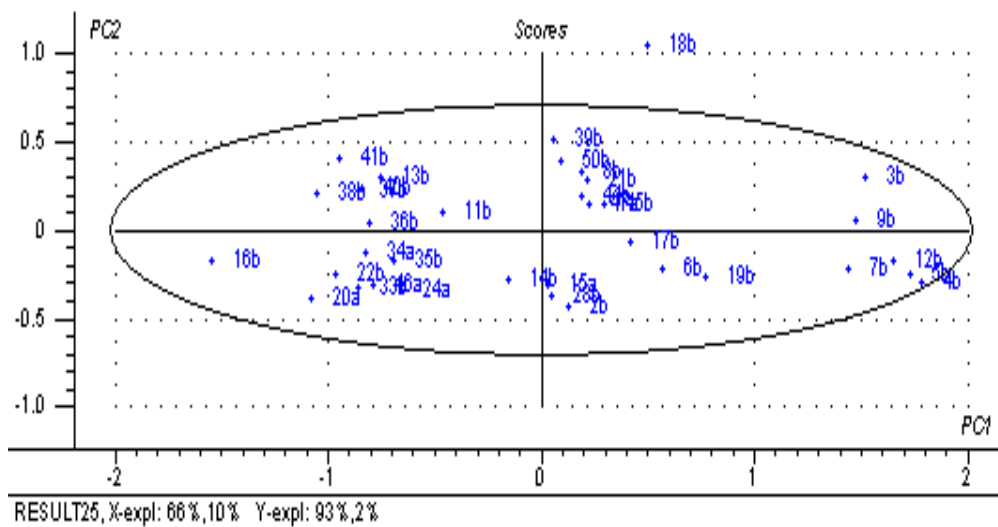
Figure 93: Hoetelling T^2 plot.
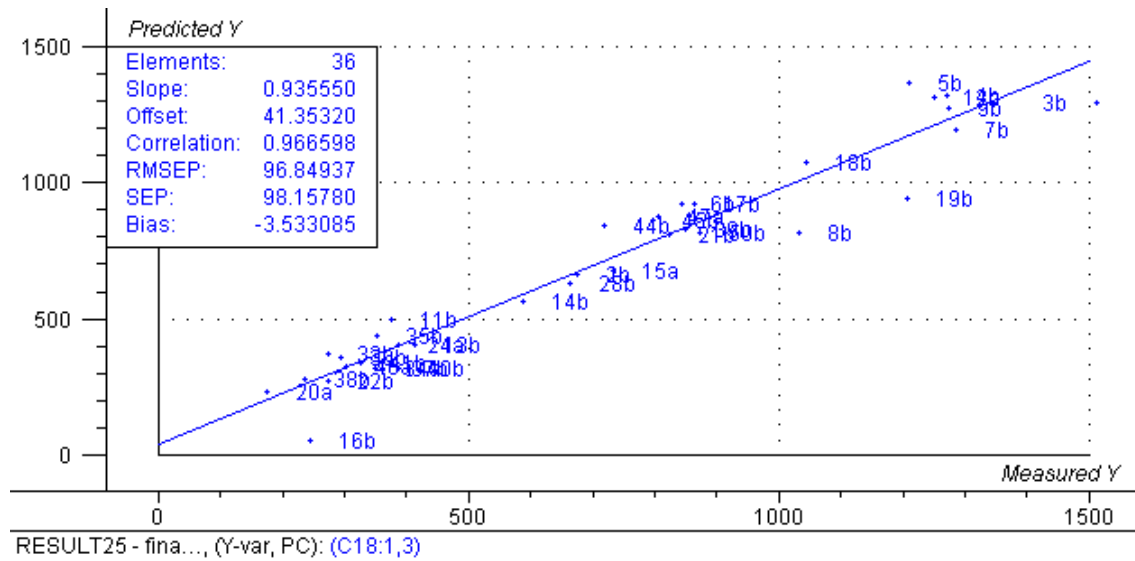
Figure 94: Predicted vs Measured Y Plot.

RMSEP: 96.84937, Correlation: 0.966598, Slope: 0.935550, Bias: -3.533085.
RMSEC: 76.64680, Correlation: 0.979185, Slope: 0.958802, Bias: -1.187e-05.

3.2.2.6. C18-2

PLS1 regression of all the samples resulted in the following model. The cross validation is carried with 6 segments of 8 samples in each.
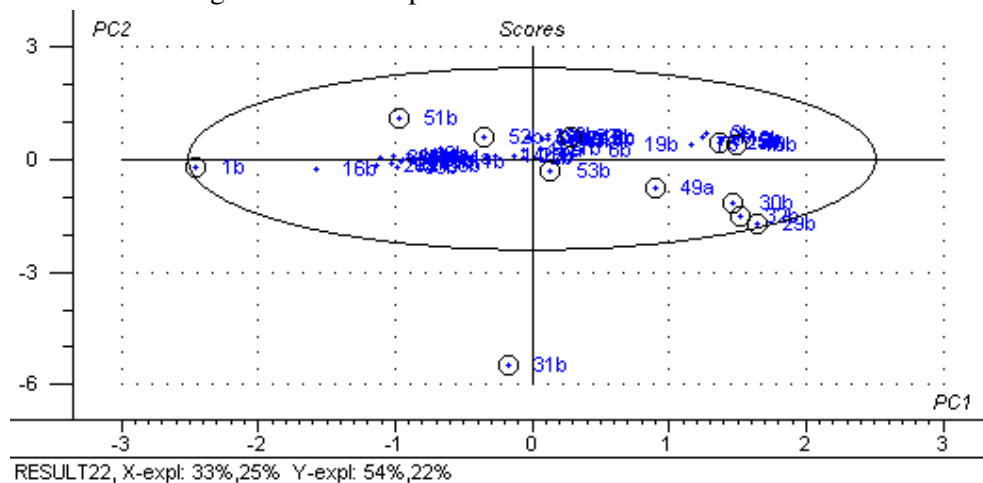
Figure 95: Hoetelling T^2 plot.

The plot shows 31b as an outlier but all the marked outliers are found in the models after removing the 31b and it is also supported by the X-Y relation plot, Predicted and Measured Y plot and Hoetelling T2 plot. After removal of the outliers 31b, 51b, 1b, 29b, 32b, 30b, 49a, 53b, 52b a good model has been obtained whose correlation is discussed after the plots.
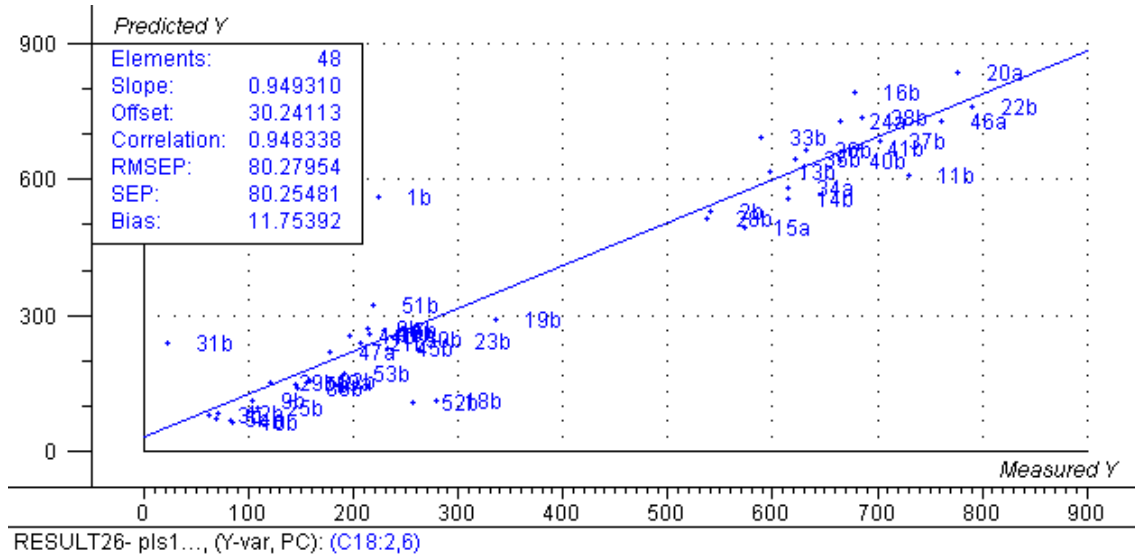
Figure 96: Predicted vs Measured plot.

RMSEP: 80.27954, Correlation: 0.948338, Slope: 0.949310, Bias: 11.75392.
RMSEC: 30.41475, Correlation: 0.992545, Slope: 0.985145, Bias: 8.631e-05
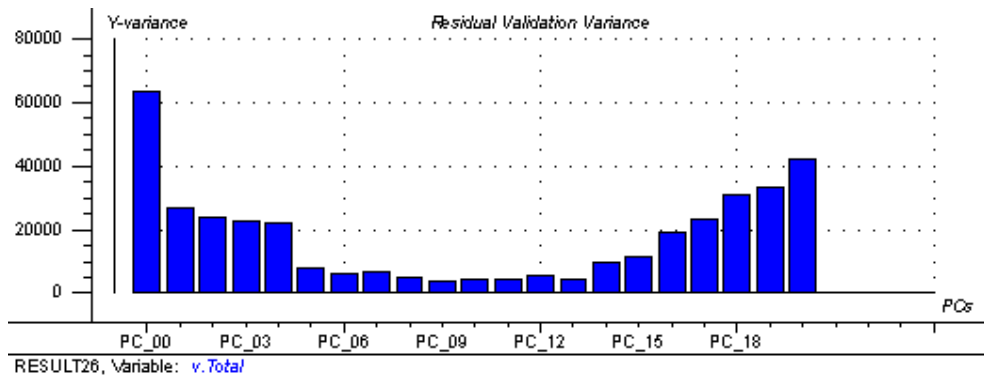
Figure 97: Residual Validation variance plot.

Based on the Residual Validation variance plot 6 PCs have been used for modelling. After the removal of all the above mentioned outliers the final model obtained is as follows:
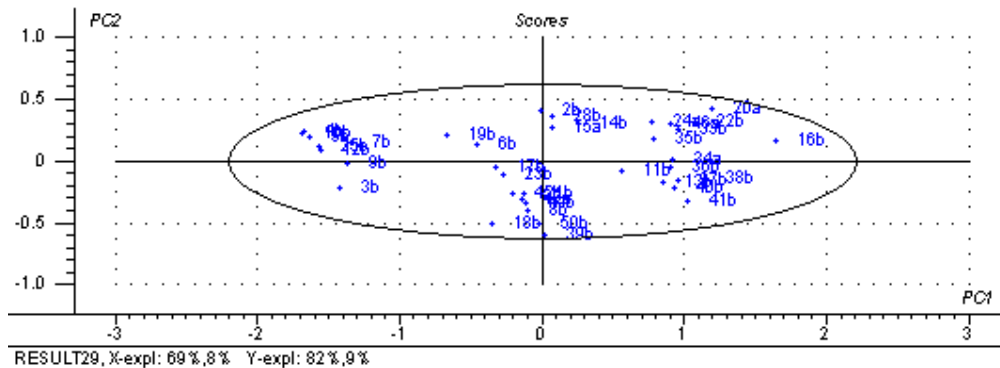
Figure 98: Hoetelling T^2 plot.
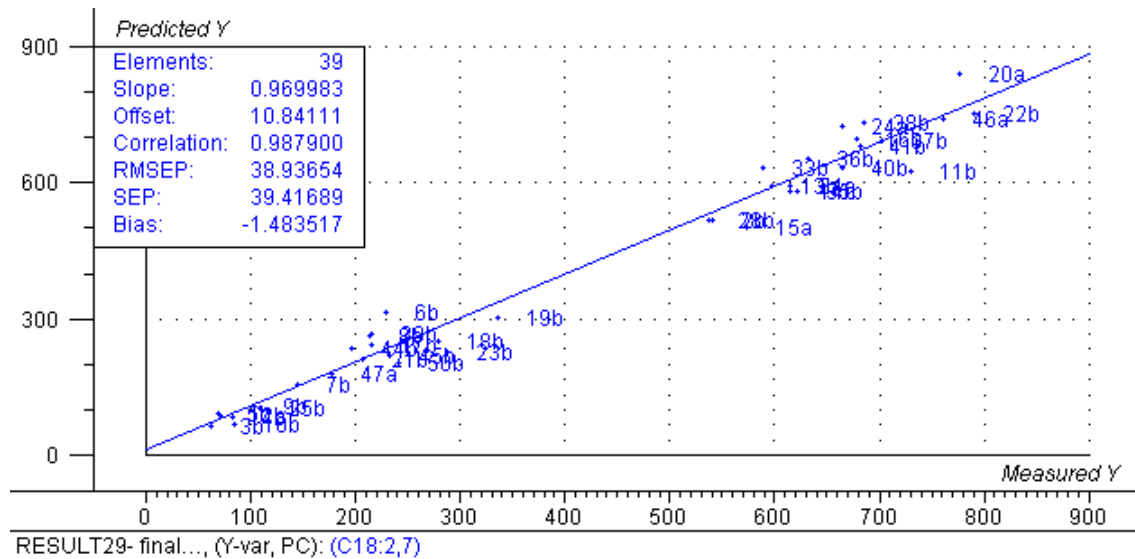
Figure 99: Predicted vs. Measured Y plot.

The Hoetelling T^2 plot is observed with no outliers and the correlation has increased compared to the first one.

RMSEP: 38.93654, Correlation: 0.987900, Slope: 0.969983, Bias:-1.483517.
RMSEC: 26.37453, Correlation: 0.994450, Slope: 0.988930, Bias: -5.517e-05.

3.2.2.7. C18:3

The 48 samples when subjected to PLS1 Regression which resulted in the following model. The obtained model showed the following results.
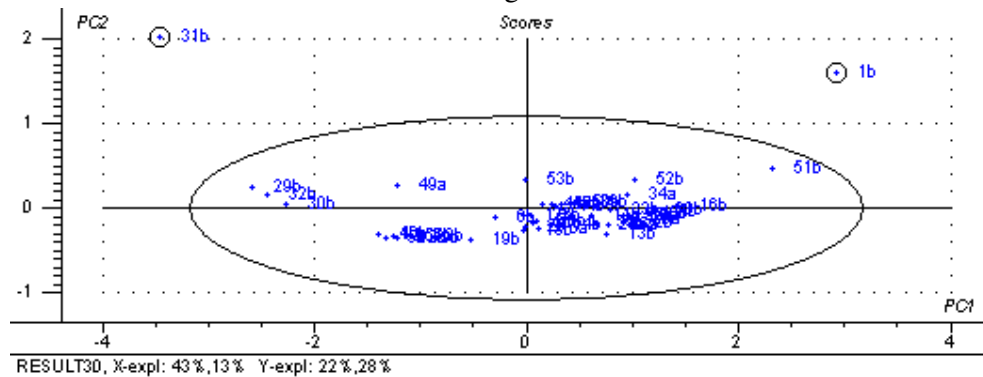
Figure 100: Hoetelling T^2 plot.
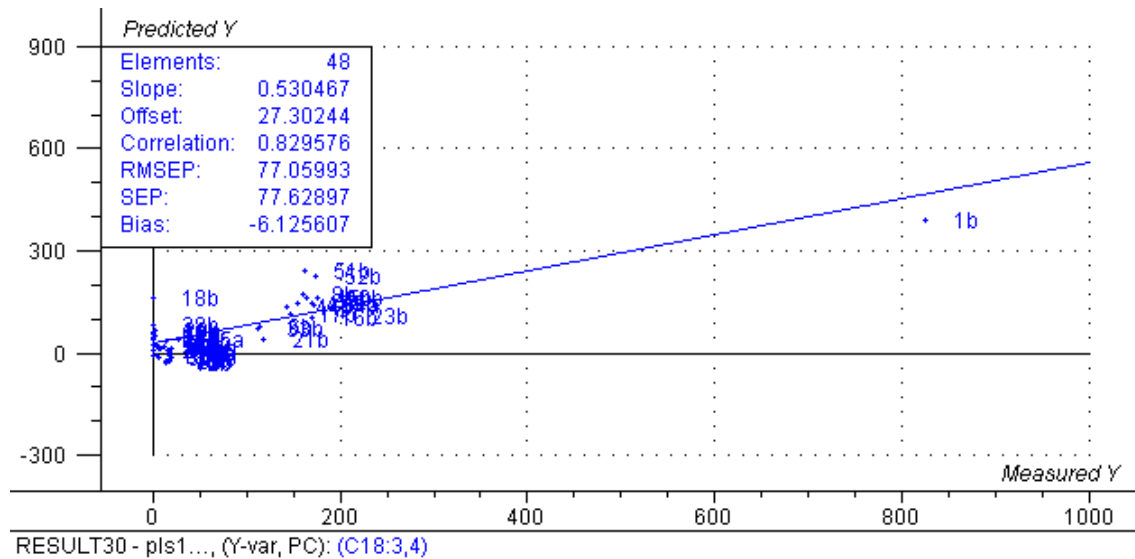
Figure 101: Predicted vs. Measured Y plot.

RMSEP: 77.05993, Correlation: 0.829576, Slope: 0.530467, Bias: -6.125607.
RMSEC: 35.19633, Correlation: 0.962783, Slope: 0.926951, Bias: -3.436e-05.
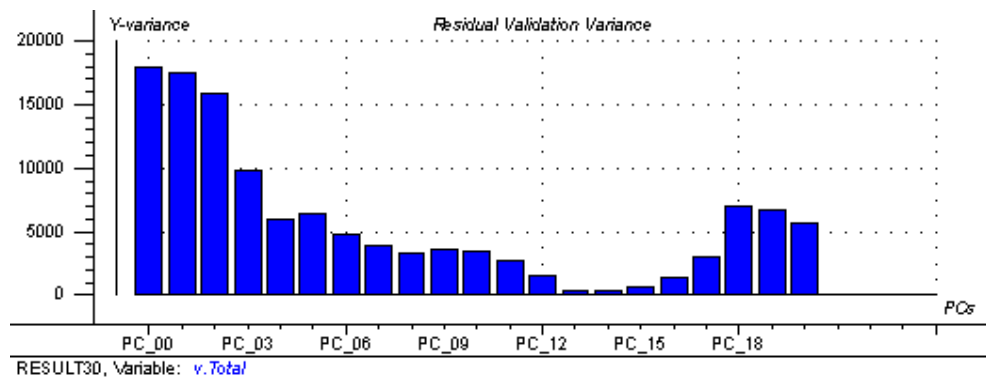
Figure 102: Residual validation variance plot.

The samples 31b, 1b, 51b, 29b, 32b, 30b, 53b, 23b, 34a, 44b, 47a are observed as outliers and their removal resulted in the model with good correlation. The final model obtained is as follows and the below plots are of that model.
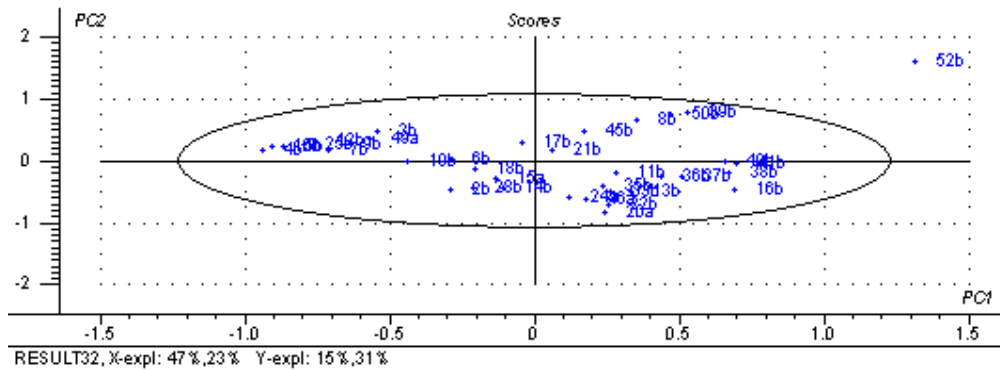
Figure 103: Hoetelling T^2 plot.
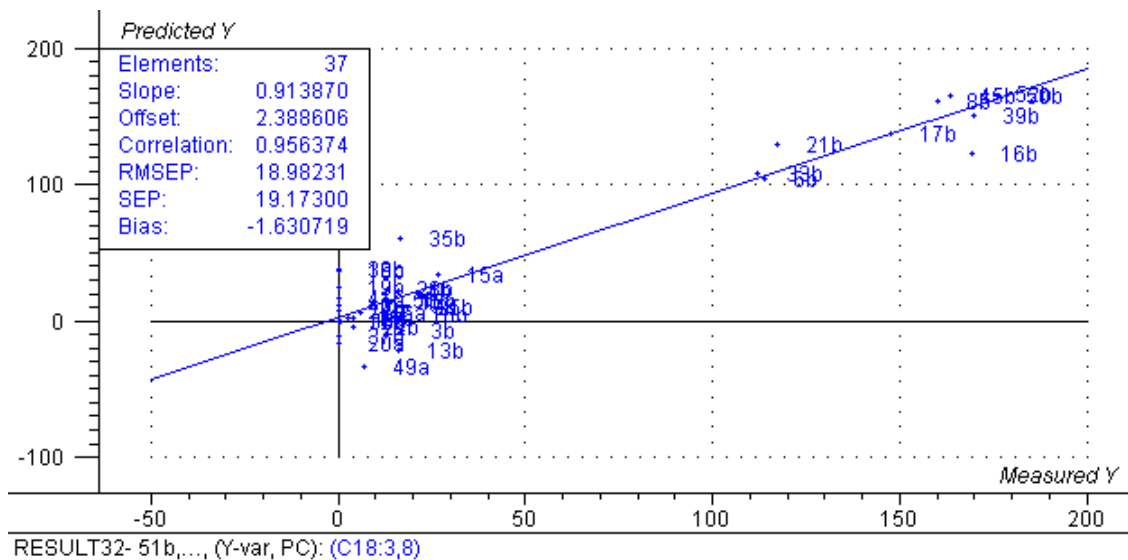
Figure 104: Predicted vs. Measured Y plot.

RMSEP: 18.98231, Correlation: 0.956374, Slope: 0.913870, Bias: -1.630719.
RMSEC: 8.809010, Correlation: 0.990698, Slope: 0.981483, Bias: -4.769e-05.
The model is with good correlation and slightly over fitting with 8 components.

3.2.2.8. PV 100-60

The 1338 variables with 48 samples are subjected to PLS1 regression and the results obtained are described below. The result obtained showed outliers in Hoetelling T^2 plot, X-Y relation, Predicted vs Measured plot and Residual sample variance plot. Based on these plots the outliers removed are 13b, 31b, 36b, 51b are removed as outliers. However with this parameter the outliers are found always in the subsequent models.
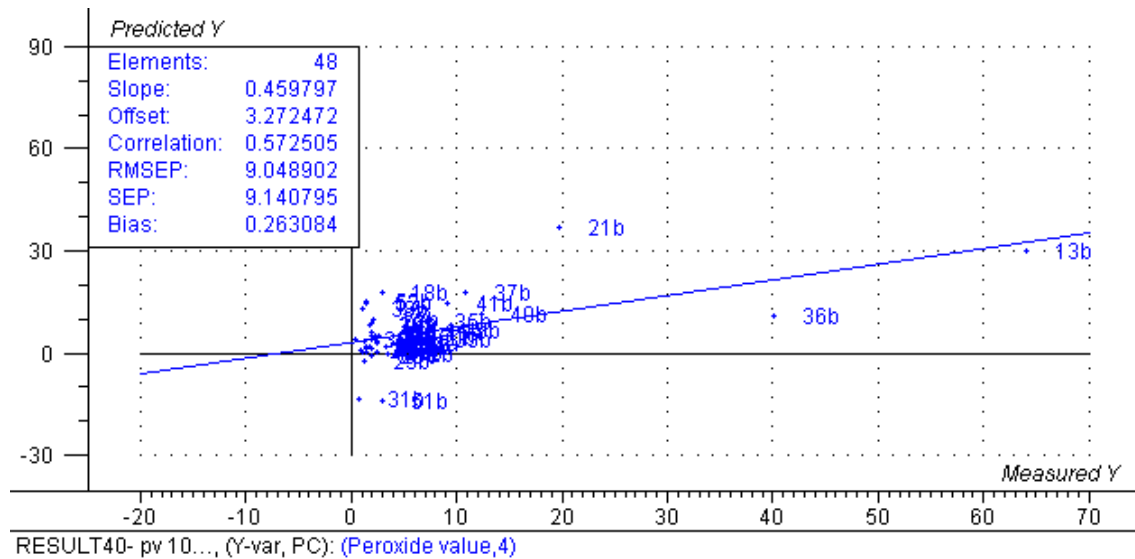
Figure 105: Predicted vs. Measured Y plot.

RMSEP: 9.048902, Correlation: 0.572505, Slope: 0.459797, Bias: 0.263084
RMSEC: 6.054508, Correlation: 0.821564, Slope: 0.674968, Bias: 3.278e-07.
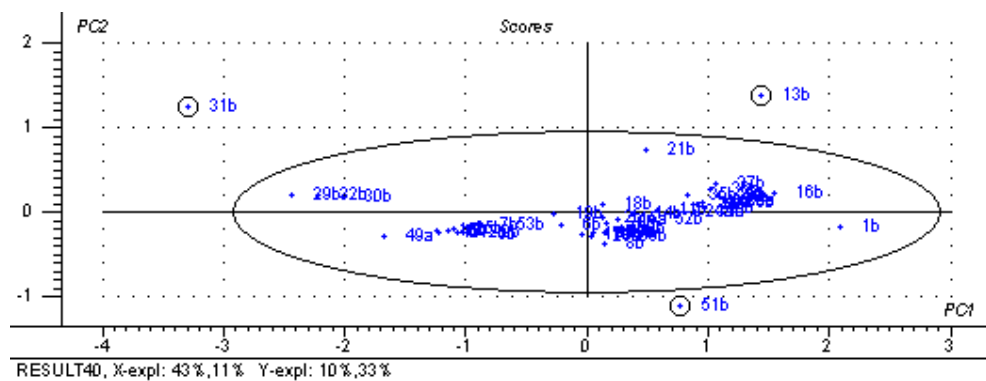
Figure 106: Hoetelling T^2 plot.

The obtained result is found as in the following plots with low correlation but with decreased RMSEP using 7 components.
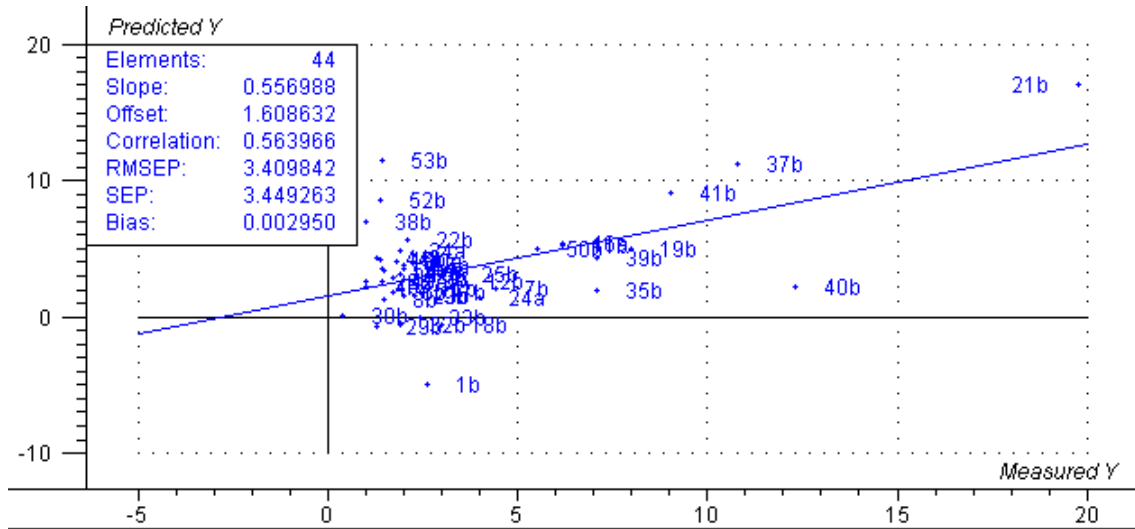
Figure 107: Predicted vs. Measured Y plot.

RMSEP: 3.409842, Correlation: 0.563966, Slope: 0.556988, Bias: 0.002950.
RMSEC: 1.713810, Correlation: 0.884528, Slope: 0.782390, Bias: -1.081e-06.
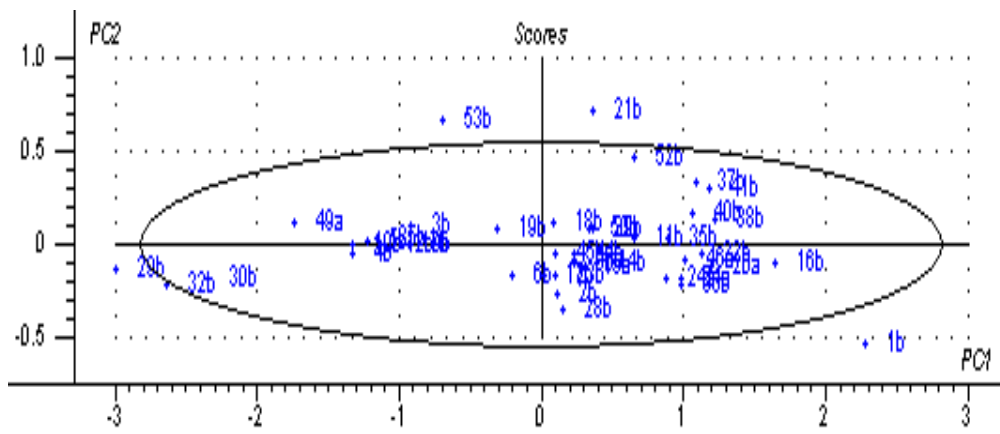
Figure 108: Hoetelling T^2 plot.

3.2.2.9. SN

RESULT51, X-expl: 41%,19%   Y-expl: 62%,16%
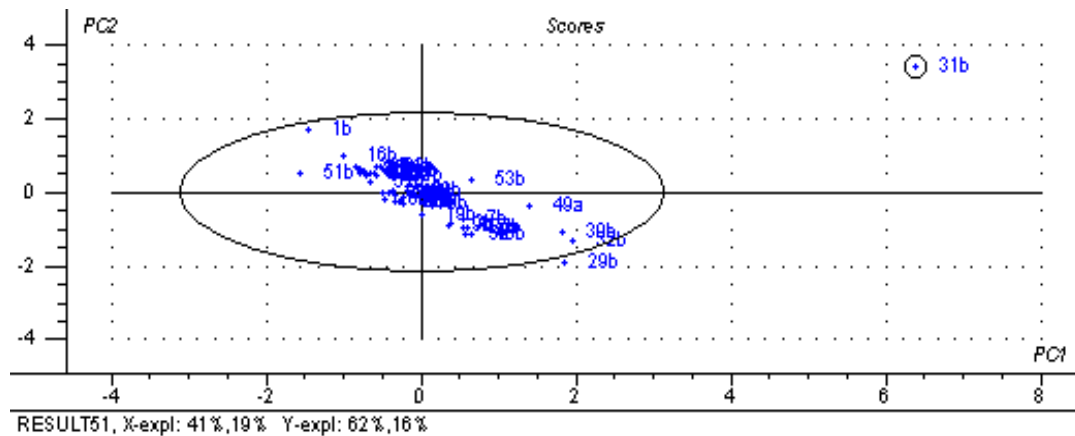
Figure 109: Hoetelling T^2 plot.

RESULT51- PLS1..., (Y-var, PC): (Avrg SN,5)
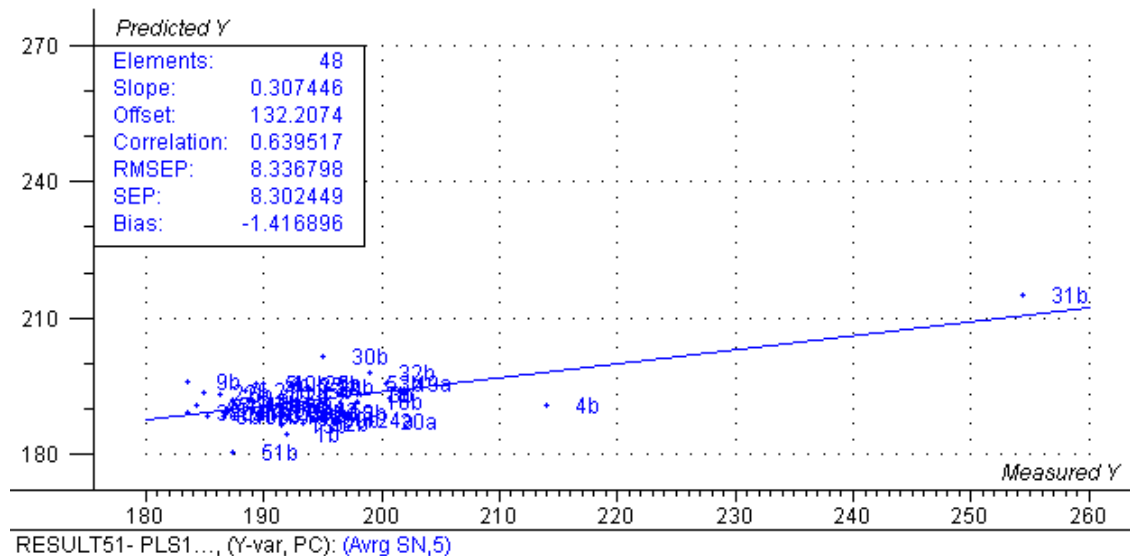
Figure 110: Predicted vs. Measured Y plot.

RMSEP: 8.336798, Correlation: 0.639517, Slope: 0.307446, Bias: -1.416896.
RMSEC: 4.552881, Correlation: 0.900415, Slope: 0.810747, Bias: 7.629e-06.

When the outlier 31b is removed the following model is obtained with very less correlation and 12 components are used for modelling 47 elements thereby the model is over fitted. When the outliers observed in the model after the removal of 31b, are recalibrated without the outliers then the models are obtained with very less correlation thereby making a still worse model with more outliers. Hence only 31b outlier is removed to get a model as below.
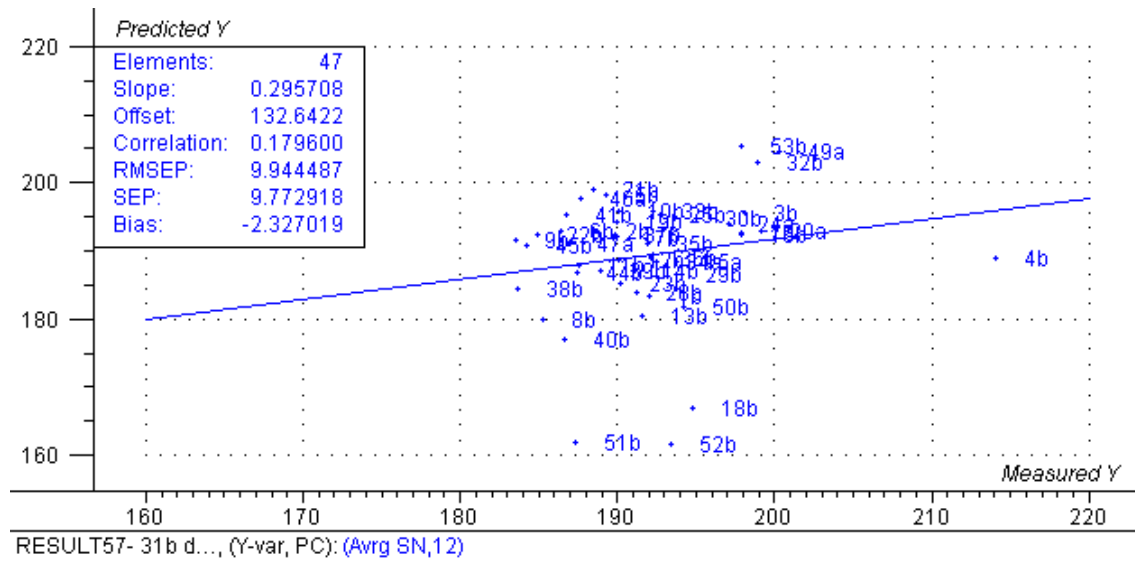
Figure 111: Predicted vs. Measured Y plot.

RMSEP: 9.944487, Correlation: 0.179600, Slope: 0.295708, Bias: -2.327019.
RMSEC: 3.555964, Correlation: 0.760279, Slope: 0.578024, Bias: 9.415e-06.
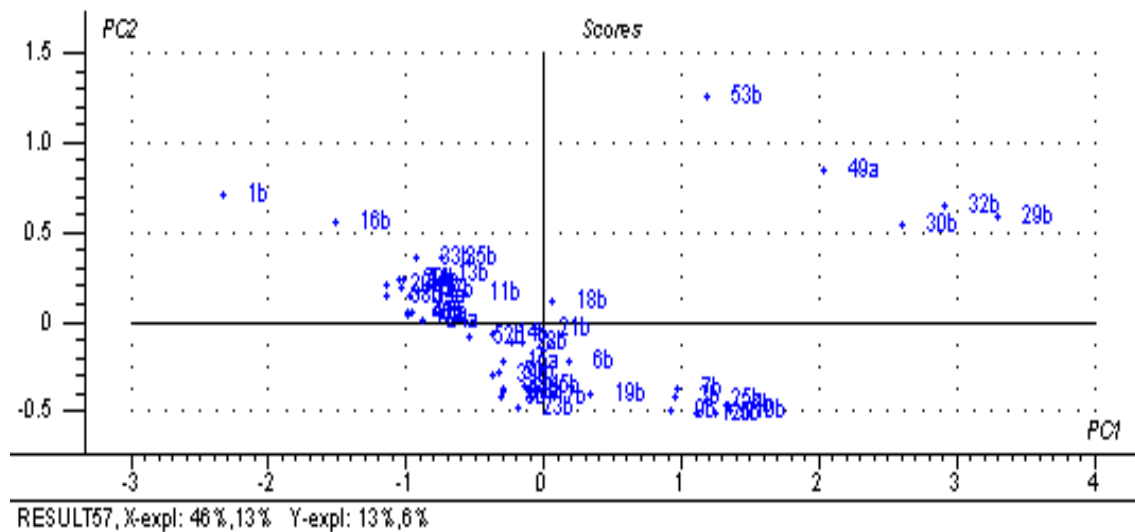
Figure 112: Score plot.

3.3.1 10-10µm IR spectrum:
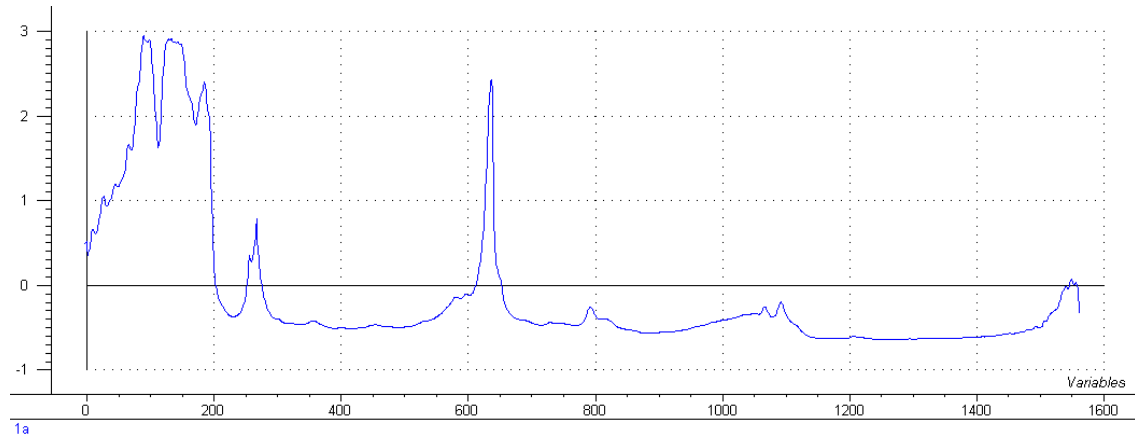
Figure 113: Lineplot of 110-10µm IR spectrum

3, 6 numbers of all single beams are plotted and then variable 1017 is selected as it shows saturation and line plot is drawn to check for samples that are really bad due to air bubbles. The samples above the line drawn at 5.0 are considered as having air. These sample numbers are multiplied by 3 to find the exact sample number in the list of 624 single beam spectra of samples. For example sample number 1 observed in the line plot of variable 1017 corresponds to the number 3 in the 624 single beam samples. In this 624 series of samples all the samples are measured once for 6 path lengths and then the replicate is measured for 6 path lengths and this continues for all 52 samples (Sample number 26 is not measured as it is not edible oil). The numbers marked in the line plot are that after multiplying with 3.
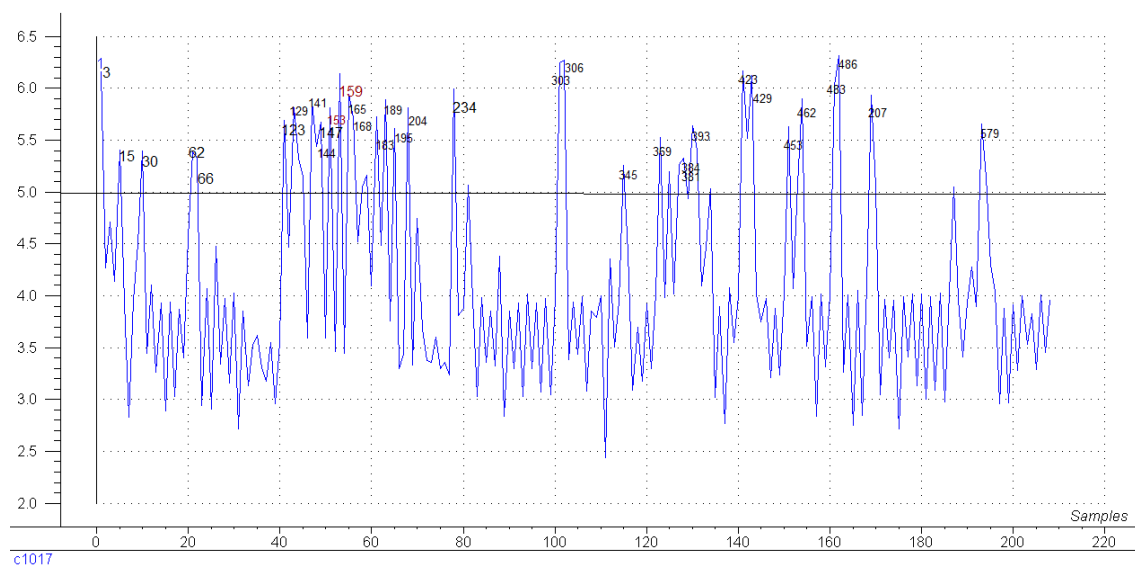
Figure 114: Line plot of 1017 variable.

Figure 115: Single Beam Line plot of 110&10µm IR spectrum

All these numbers are kept out of calculation of the samples list. The numbers above are equivalent to samples names 1a, 2a, 3a, 6a,11b,12b, 13b, 15a, 17b, 20a,27a,32b, 33b, 34a, 37a,37b, 39b, 40a, 42a, 50a. So all these samples have been deleted and its replicate has been copied in its place. However as both the replicates 37a and 37b are considered to have air bubbles this sample has been completely kept out of calculation.

3.3.1. 110-10 µm SNV

The 9 parameters with 51 samples that have been analyzed for PLS1 regression gave the following results with 6 segments and 8 samples in each segment.

3.3.1.1. PV

Figure 116: Hoetelling T^2 plot.

In the above plots it is observed that 13a, 21b, 34b, 33a, 11b are observed as outliers and when the regression is recalculated without these outliers gave the results as below with low correlation and X explaination.
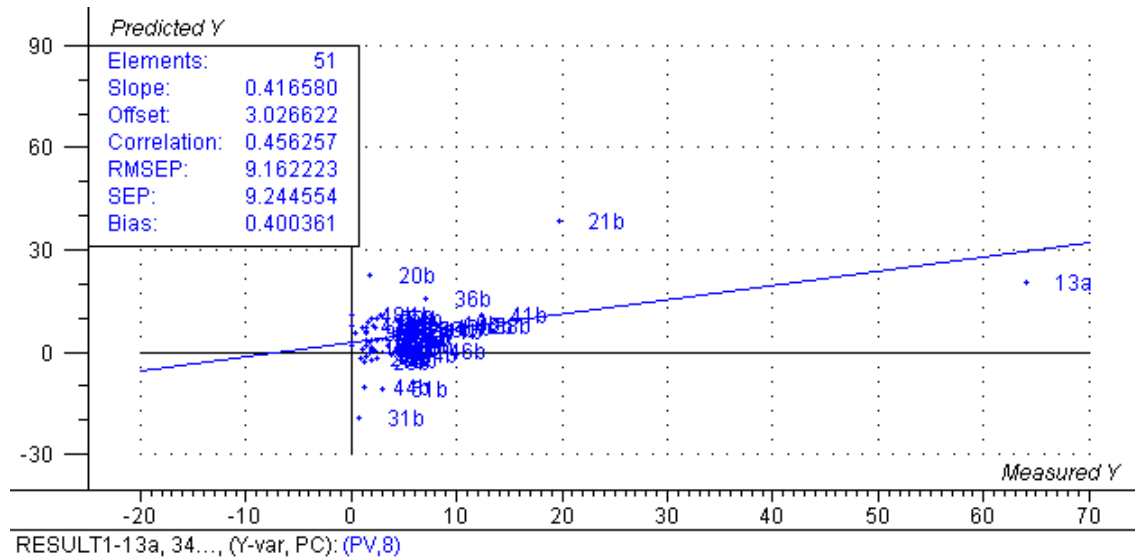
Figure 117: Predicted vs. Measured Y plot.

RMSEP: 9.162223, Correlation: 0.456257, Slope: 0.416580, Bias: 0.400361.
RMSEC: 3.109424, Correlation: 0.940505, Slope: 0.884550, Bias: 7.830e-07.

The residual validation variance plot shows 8 PCs can be used for modelling. When the above mentioned outliers are removed then the following model is obtained but with lesser correlation values.
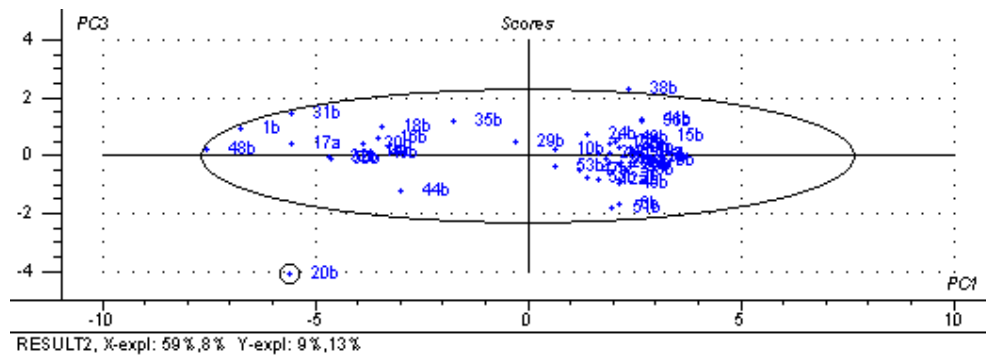
Figure 118: Hoetelling T^2 plot.
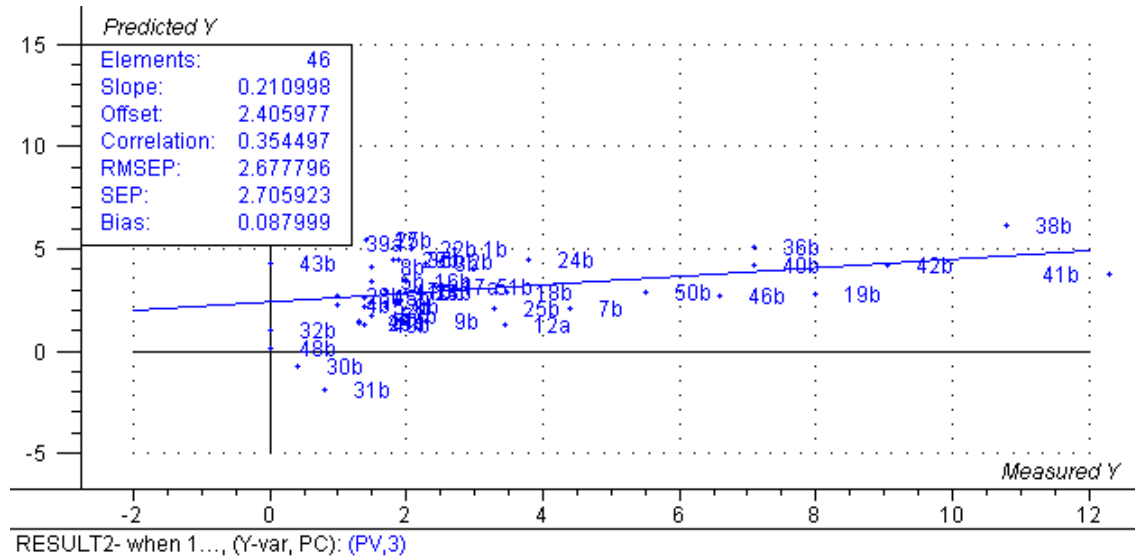
Figure 119: Predicted vs. Measured Y plot.

RMSEP: 2.677796, Correlation: 0.354497, Slope: 0.210998, Bias: 0.087999.
RMSEC: 2.150627, Correlation: 0.630882, Slope: 0.398012, Bias: -4.717e-07.

3.3.1.2. SN 110-10

The PLS1 regression with 6 segments of 8 samples in each segment gave the following models. The Hoetelling T^2 plot showed 48b, 31b, 34b, 4b are removed as outliers when these outliers are removed and recalculated the model resulted with lower correlation values.

RMSEP: 7.357358, Correlation: 0.820065, Slope: 0.763177, Bias: -0.331095.
RMSEC: 4.601713, Correlation: 0.931039, Slope: 0.866834, Bias: -2.992e-07.
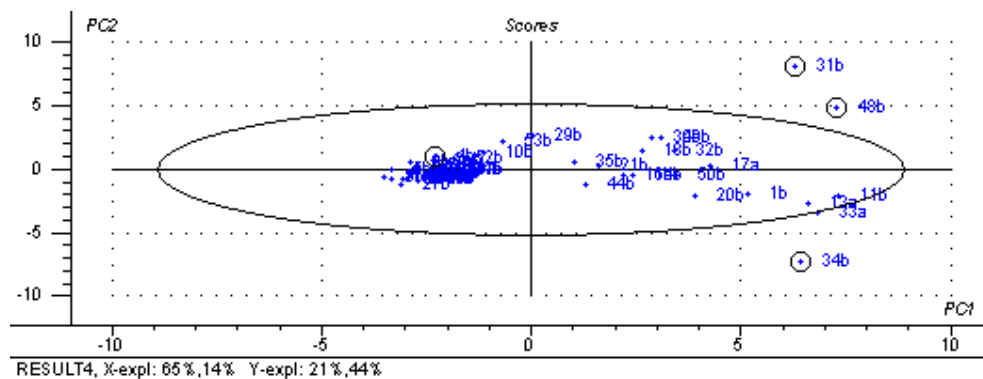
Figure 120: Hoetelling T^2 plot.
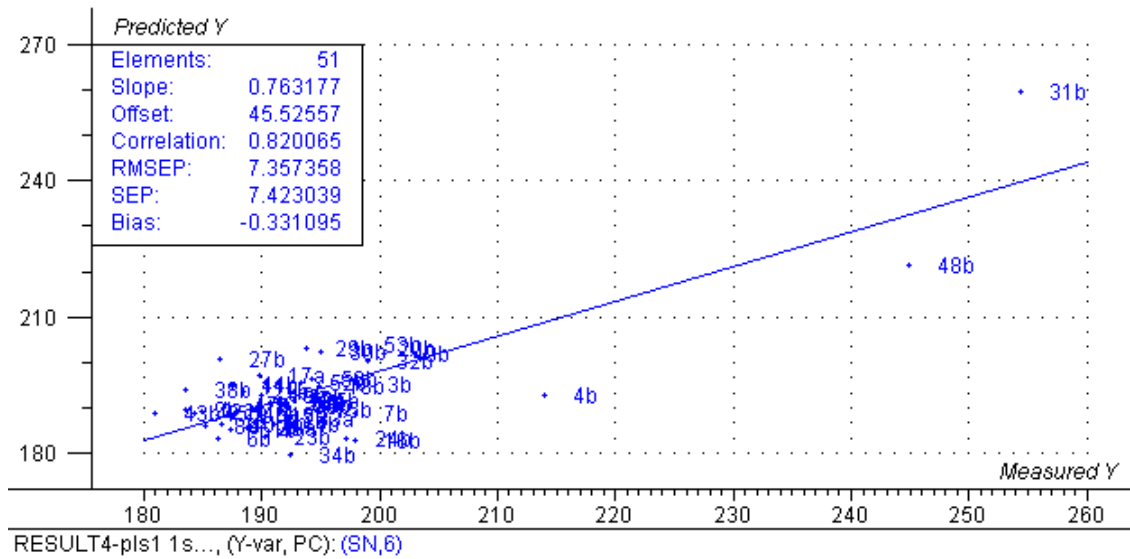
Figure 121: Predicted vs. Measured Y plot.
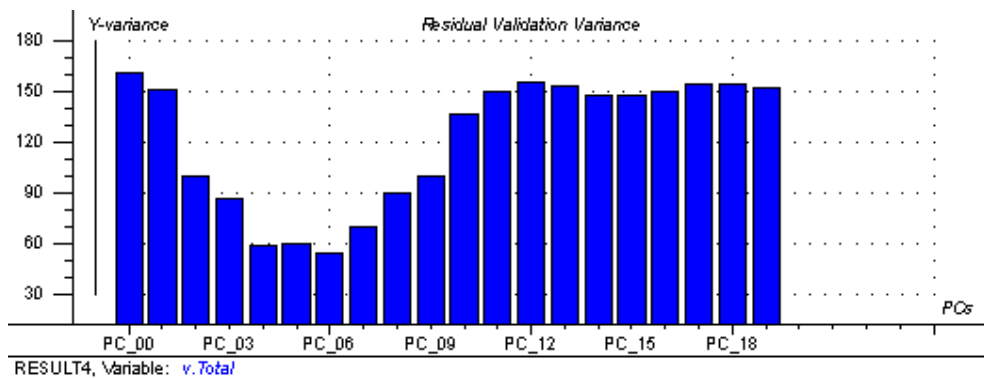
Figure 122: Residual Validation Variance.

When the above mentioned outliers are removed then it resulted in the following model.
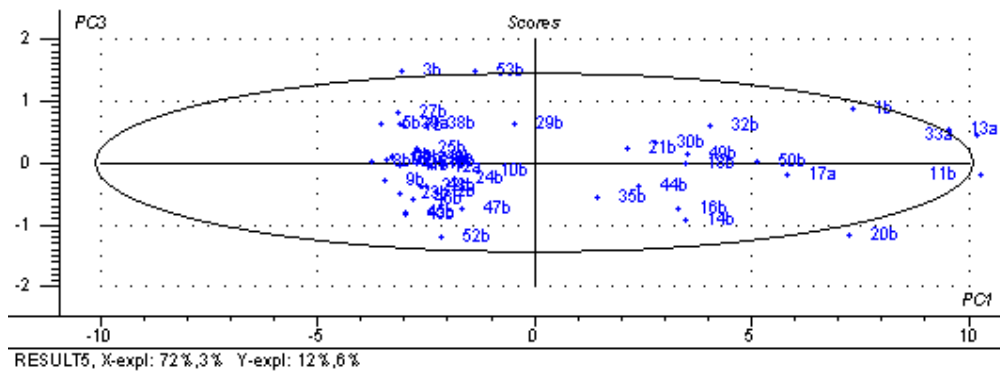
Figure 123: Hoetelling T^2 plot.
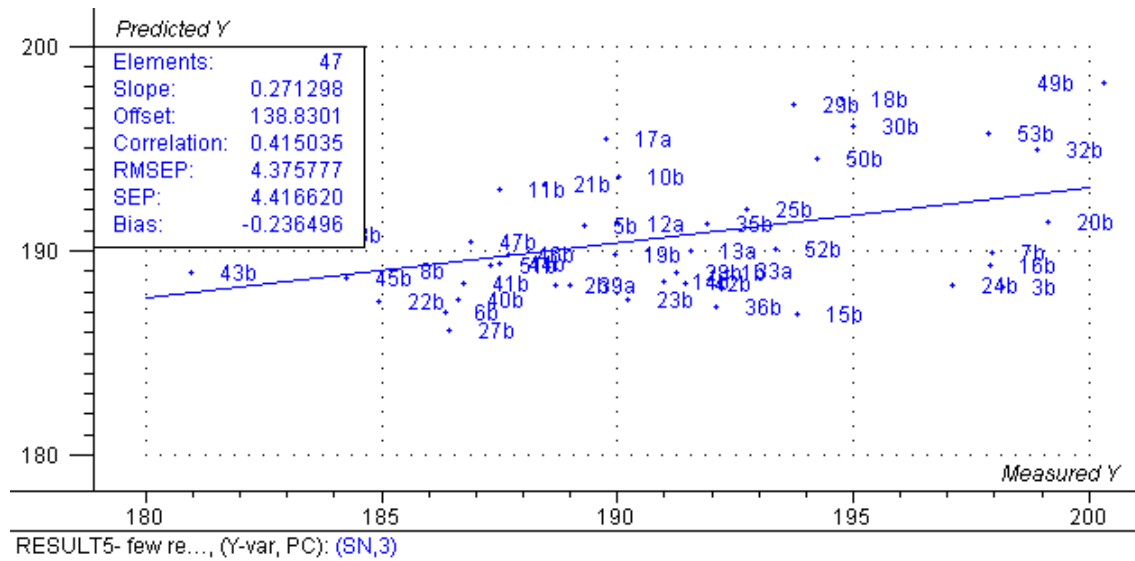
Figure 124: Predicted vs. Measured Y plot.

RMSEP: 4.375777, Correlation: 0.415035, Slope: 0.271298, Bias: -0.236496.
RMSEC: 3.424637, Correlation: 0.675664, Slope: 0.456521, Bias: 3.896e-06.

This model has actually decreased correlation from R^2 =0.820065 to R^2 = 0.415035 thereby the model obtained is a bad one but the RMSEP value has decreased a lot. The model used 3 components for modelling 47 elements. X explanation is 72% and Y is 12% for the 1st component.

3.3.1.3. C12

48 elements are used for regression which the following result. 31b, 34b are observed as outliers and the number of components is 5.
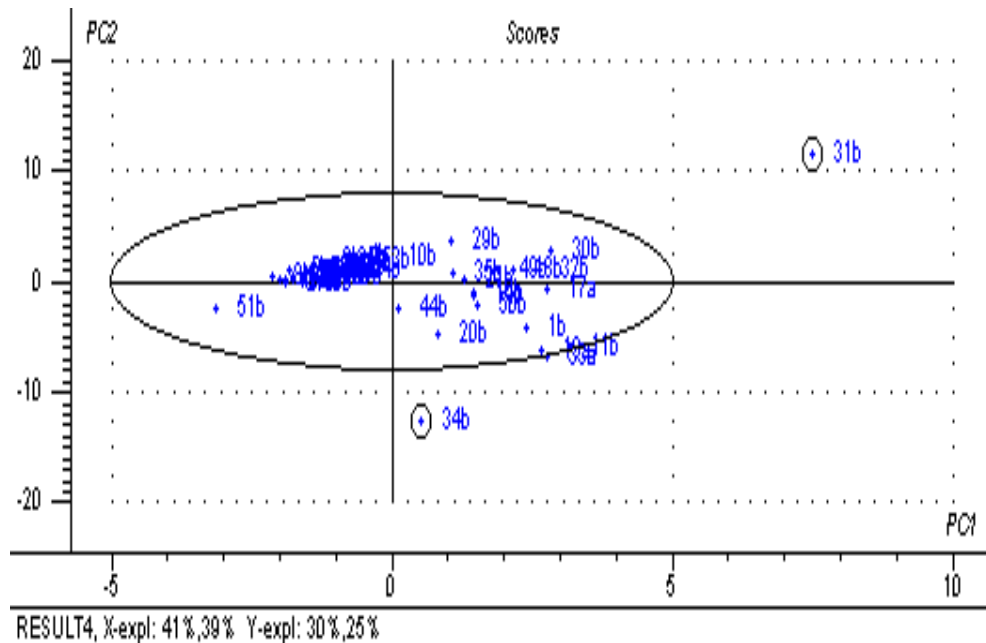

Figure 125: Hoetelling T^2 plot.
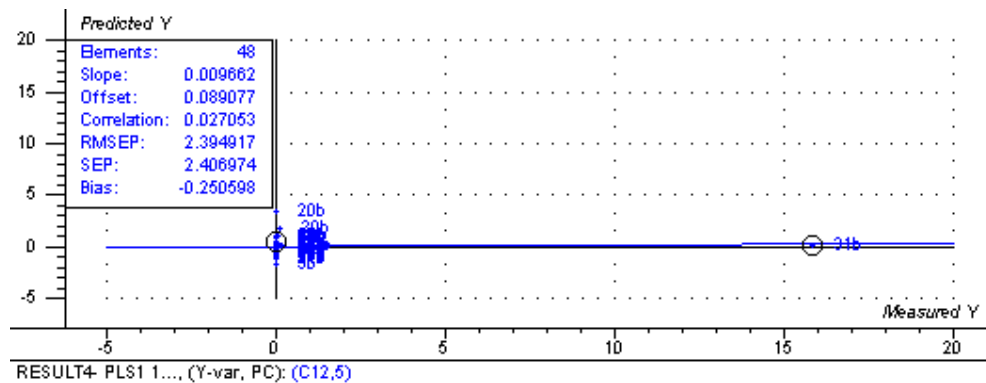
Figure 126: Predicted vs. Measured Y plot.

RMSEP: 2.595380, Correlation: 0.027053, Slope: 0.009662, Bias: -0.250598.
RMSEC: 0.537196, Correlation: 0.971403, Slope: 0.943623, Bias: -9.425e-08.
When the outliers are removed the model below is obtained with 3 components.
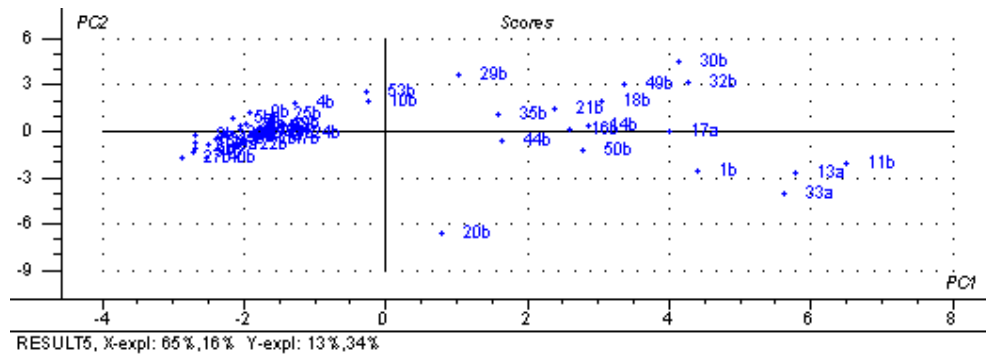
Figure 127: Score plot.
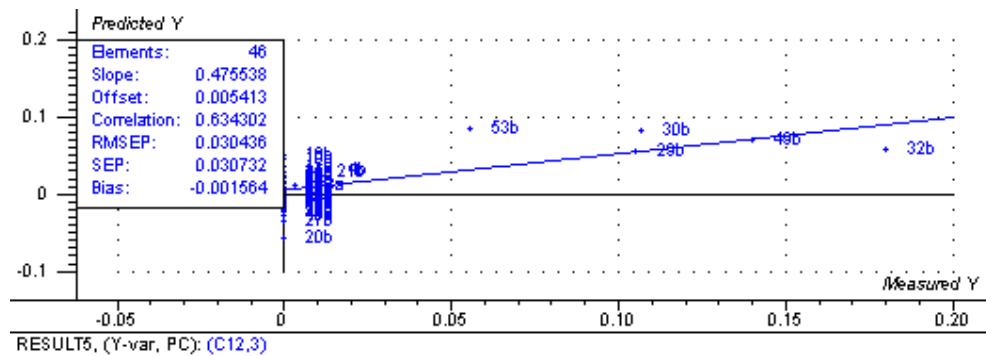
Figure 128: Predicted vs. Measured Y plot.

RMSEP: 0.030436, Correlation: 0.634302, Slope: 0.475538, Bias: -0.001564.

RMSEC: 0.024092, Correlation: 0.784964, Slope: 0.616168, Bias: 1.437e-09.

The model obtained has a correlation of 0.634302 which has increased from 0.027053 making it a good fit with 3 components and 46 elements. But the slope is 0.475538 which is far away from the value of 1.

3.3.1.4. C14

Regression of 48 elements was carried which gave the following result for C14 parameter
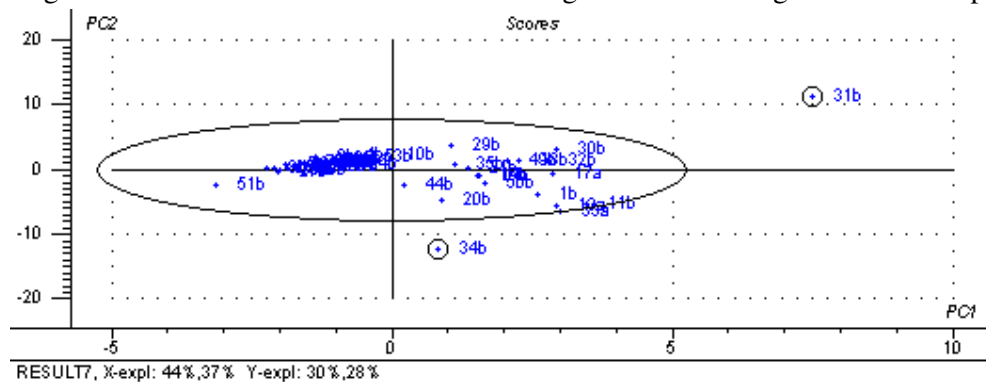
Figure 129: Hoetelling T^2 plot.
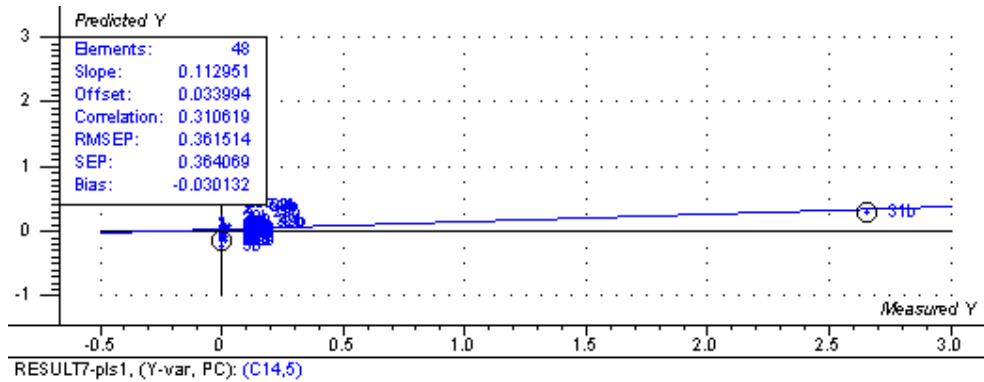
Figure 130: Predicted vs. Measured Y plot.

RMSEP: 0.361514, Correlation: 0.310619, Slope: 0.112951, Bias: -0.030132.
RMSEC: 0.079408, Correlation: 0.977735, Slope: 0.955966, Bias: -1.777e-08.
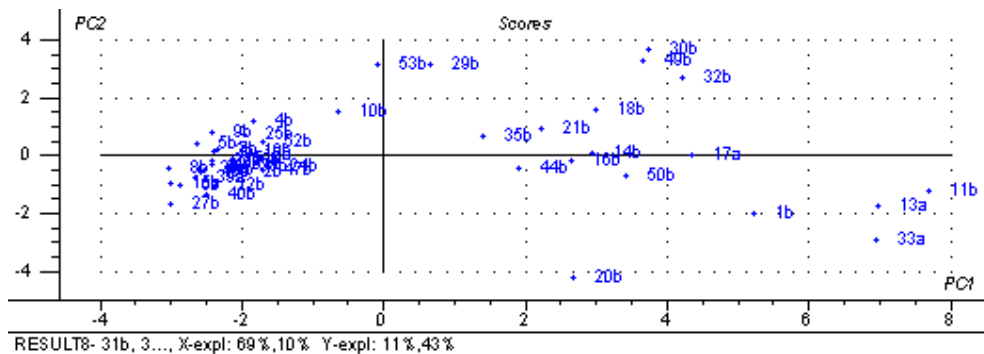
Figure 131: Score plot.
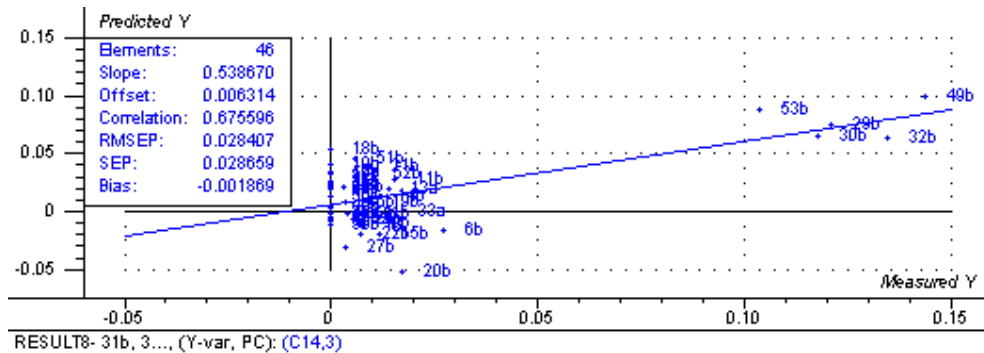
Figure 132: Predicted vs. Measured Y plot.

RMSEP: 0.028407, Correlation: 0.675596, Slope: 0.538670, Bias: -0.001869.
RMSEC: 0.22118, Correlation: 0.812414, Slope: 0.660017, Bias: 2.592e-09.
The obtained model is with a correlation coefficient of 0.67 for 46 elements and 3 components. The model has increased correlation compared to the previous ones but still the R^2 =0.675596 which is not enough for a model to be considered as a good one. X explanation is 69% while Y explanation is 11% for the 2nd component.

3.3.1.5. C16
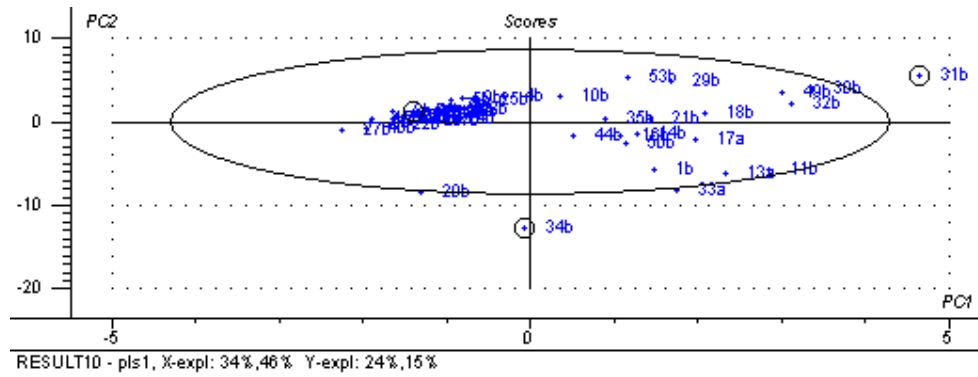
Figure 133: Hoetelling T^2 plot
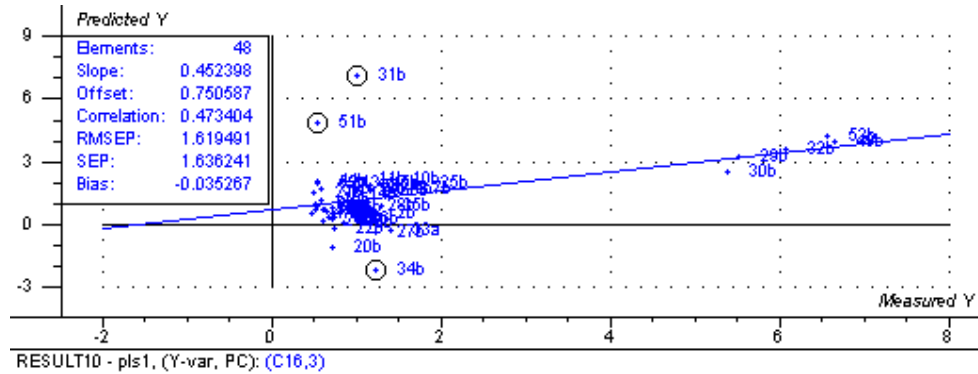
Figure 134: Predicted vs. Measured Y plot.

RMSEP: 1.619491, Correlation: 0.473404, Slope: 0.452398, Bias: -0.035267.
RMSEC: 1.010861, Correlation: 0.779052, Slope: 0.606923, Bias: 4.843e-08.
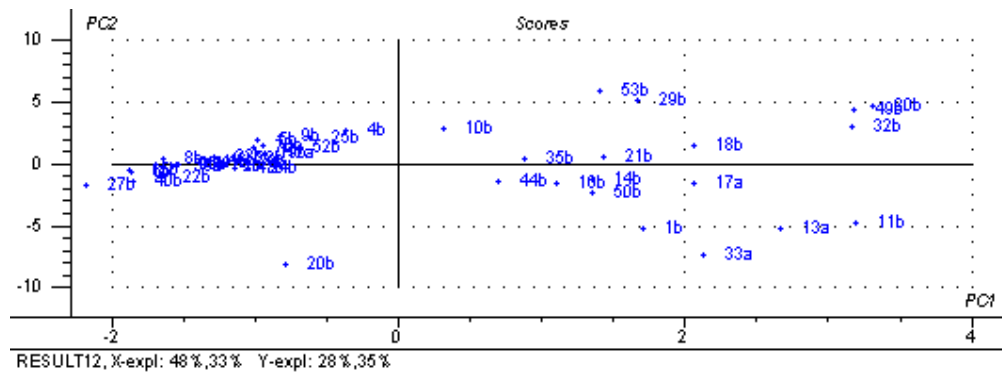
Figure 135: Score plot.
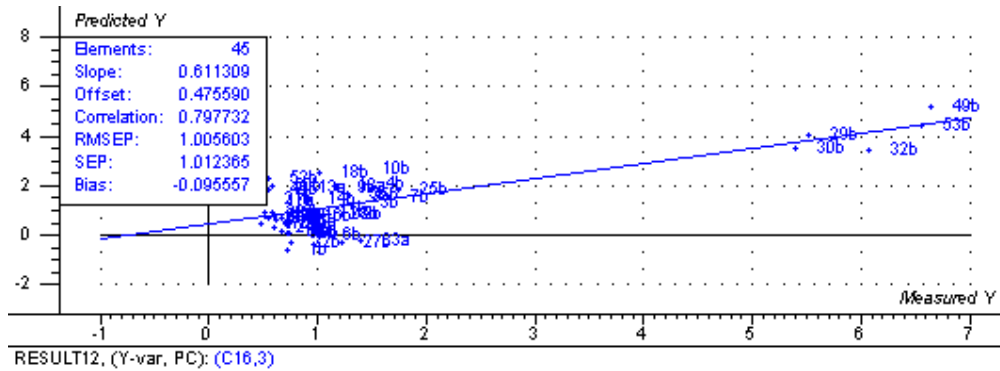
Figure 136: Predicted vs. Measured Y plot.

RMSEP: 1.005603, Correlation: 0.797732, Slope: 0.611309, Bias: -0.095557.
RMSEC: 0.774419, Correlation: 0.884191, Slope: 0.781793, Bias: 9.537e-08.

The final model obtained after the removal of the outliers 31b, 51b and 34b was with a R^2 = 0.797732 for 45 elements and 3 components is a good fit but the model is not with a good correlation coefficient even though it has increased from the previous models and the obtained model shows the palm oils as being present far away from the other oils.

The sample numbers 49b, 53b, 32b, 30b, 39b are palm oils and these are clearly observed as separated from other oils but lie along the regression line in the Predicted vs. Measured Y plot. The removal of these and recalibration resulted with lower correlation and the model was with still more outliers. So further more modelling was done.

3.3.1.6. C18
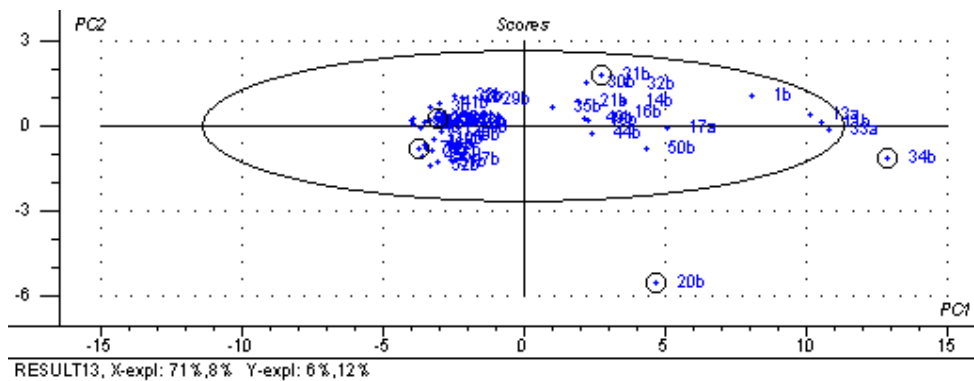
Figure 137: Hoetelling T^2 plot
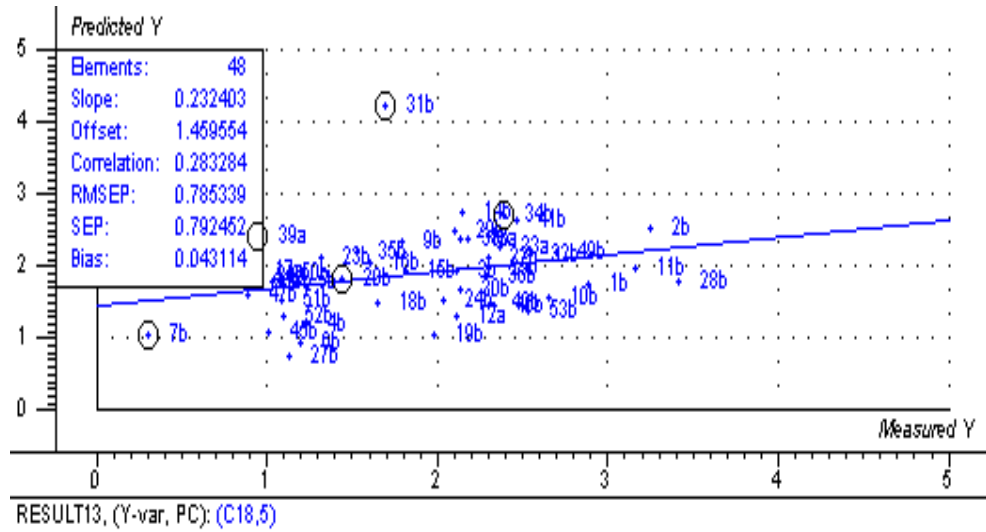
Figure 138: Predicted vs. Measured Y plot.

RMSEP: 0.785339, Correlation: 0.283284, Slope: 0.232403, Bias: 0.043114.
RMSEC: 0.482505, Correlation: 0.736574, Slope: 0.542542, Bias: -1.192e-07.
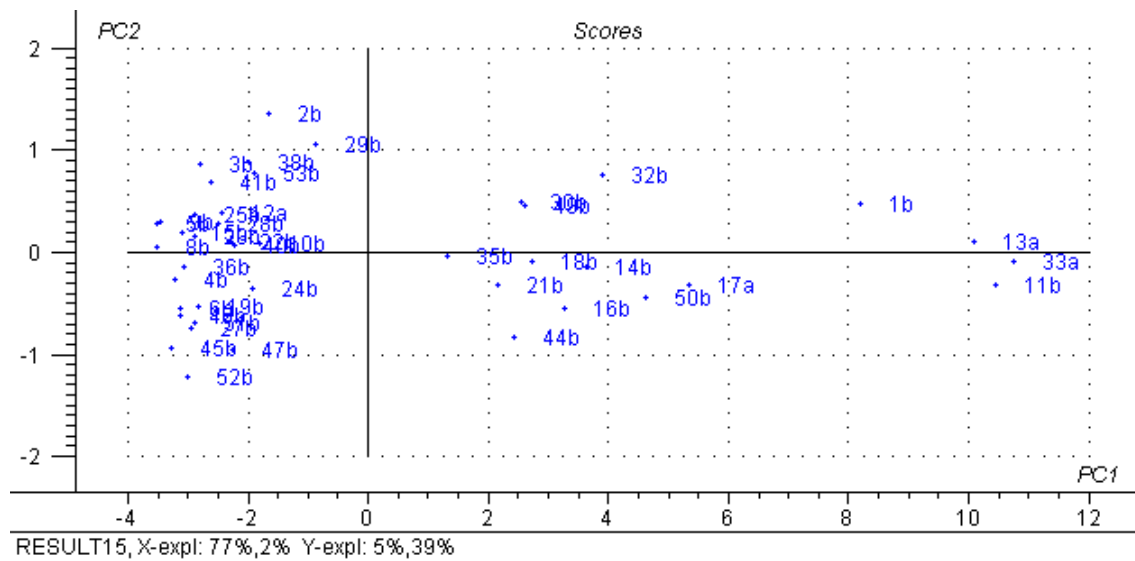
Figure 139: Score plot.

RESULT15, (Y-var, PC): (C18,5)
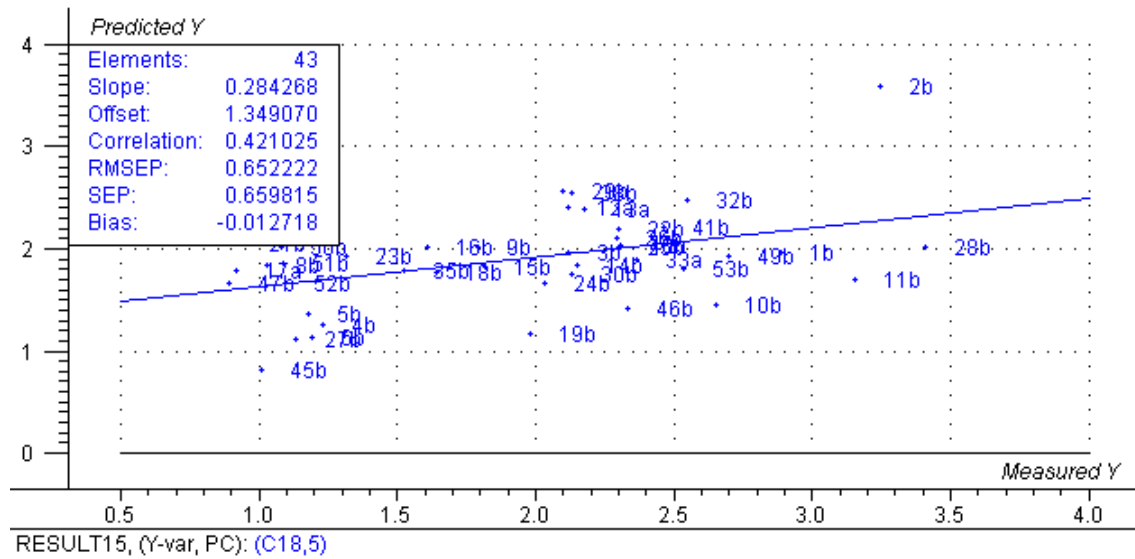
Figure 140: Predicted vs. Measured Y plot.

RMSEP: 0.652222, Correlation: 0.421025, Slope: 0.284268, Bias: -0.012718.
RMSEC: 0.379117, Correlation: 0.836706, Slope: 0.700078, Bias: 2.855e-07.

The R^2 = 0.421025 for the final model that is obtained after the removal of outliers 31b, 39a, 7b, 4a, 34b was very low to be considered as a good model. 43 elements are modelled with 5 components which is a good fit.

3.3.1.7. C18:1

RESULT16-pls1, X-expl: 72%,2% Y-expl: 22%,48%
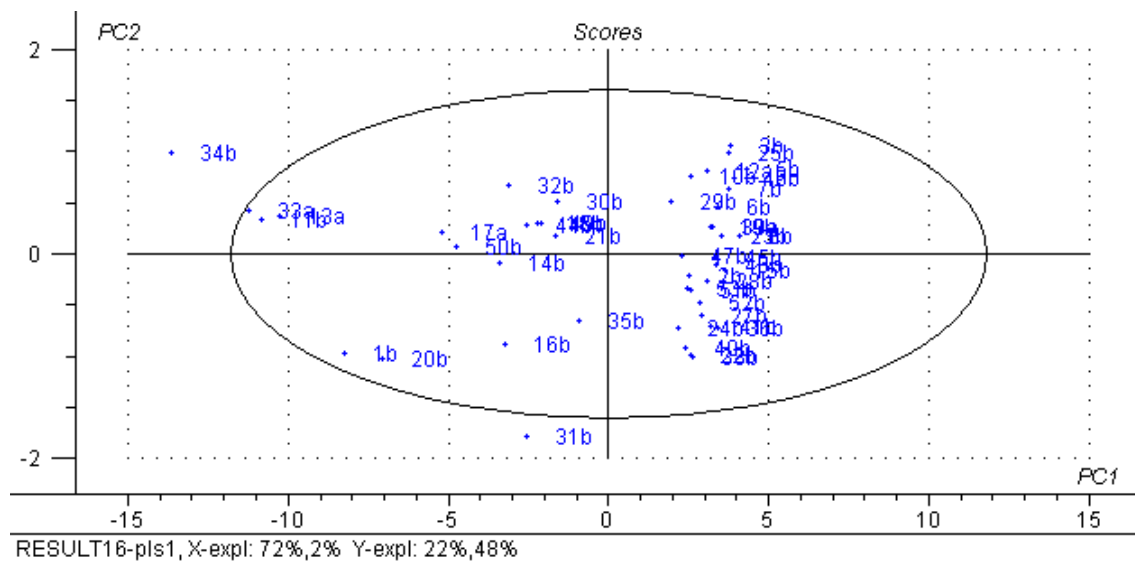
Figure 141: Hoetelling T^2 plot
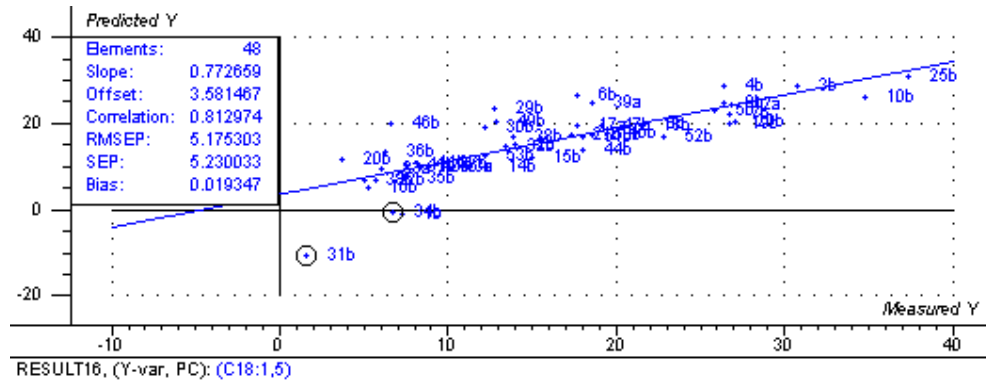
Figure 142: Predicted vs. Measured Y plot

RMSEP: 5.175303, Correlation: 0.812974, Slope: 0.772659, Bias: 0.019347.
RMSEC: 3.510917, Correlation: 0.913923, Slope: 0.835255, Bias: -9.437e-08.
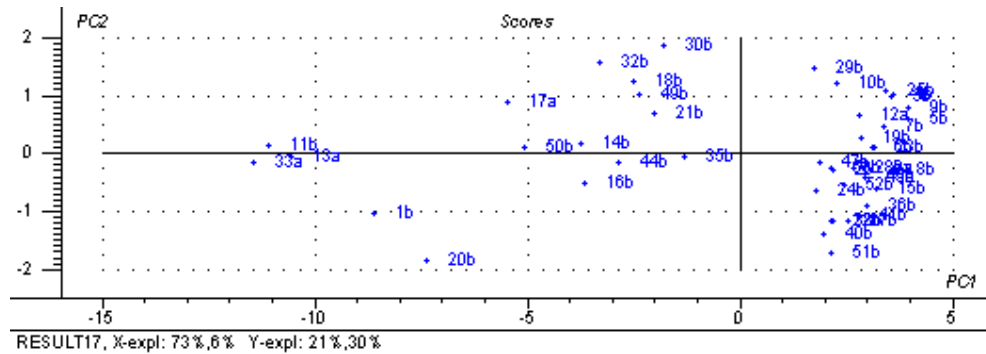
Figure 143: Score plot.
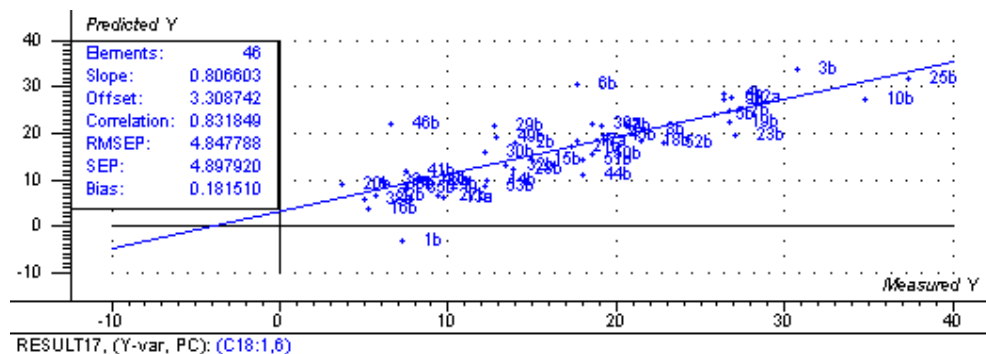
Figure 144: Predicted vs. Measured plot.

RMSEP: 4.847788, Correlation: 0.831849, Slope: 0.806603, Bias: 0.181510.
RMSEC: 3.103983, Correlation: 0.930454, Slope: 0.865745, Bias: -1.897e-06.
After the removal of outliers 31b and 34b outliers the obtained model is a good one with higher correlation, lower RMSEP and slope value approximately 0.86. The modelling of 46 elements with 6 components is a good fit.

3.3.1.8. C18:2
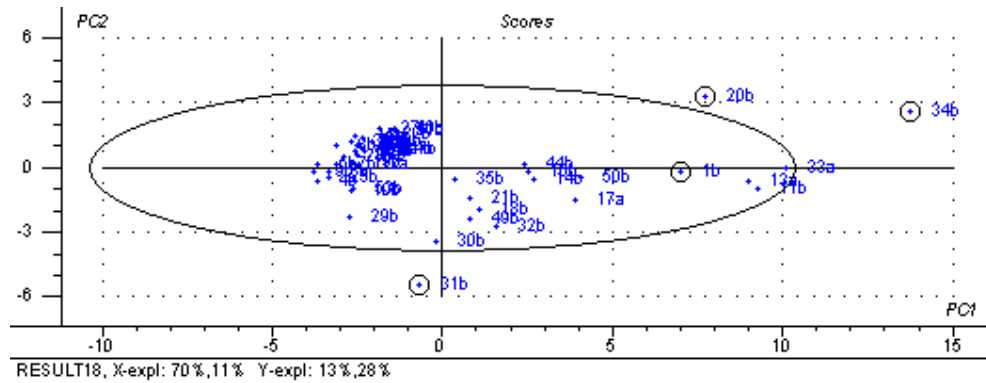
Figure 145: Hoetelling T^2 plot
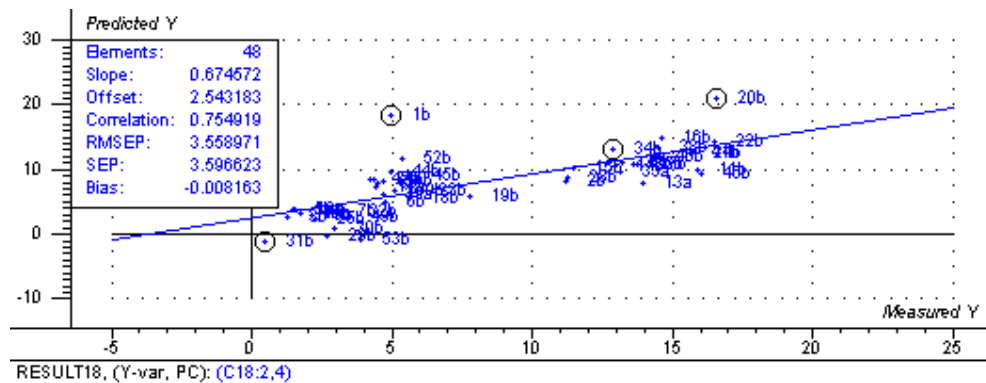
Figure 146: Predicted vs. Measured plot.

RMSEP: 3.558971, Correlation: 0.754919, Slope: 0.674572, Bias: -0.008163.
RMSEC: 2.718174, Correlation: 0.859012, Slope: 0.737902, Bias: -5.836e-08.

Figure 147: Score plot.

Figure 148: Predicted vs. Measured Y plot.

RMSEP: 2.713965, Correlation: 0.859135, Slope: 0.826082, Bias: 0.092412.
RMSEC: 1.747786, Correlation: 0.941758, Slope: 0.886907, Bias: -1.309e-06.
The R^2 =0.859135 and lower RMSEP and higher slope value makes a good model for 44 elements with 6 components after the removal of outliers 1b, 31b, 34b and 20b.

3.3.1.9. C18:3

PLS1 regression with 48 elements and segmented cross validation gave the following models.

Figure 149: Hoetelling T^2 plot

Figure 150: Predicted vs. Measured Y plot.

RMSEP: 2.965839, Correlation: 0.282753, Slope: 0.190639, Bias: -0.179458.

RMSEC: 1.862777, Correlation: 0.758319, Slope: 0.575048, Bias: -5.096e-07.

The outliers 1b, 34b, 31b, 20b are removed and recalculated without these samples and the result obtained is as follows:
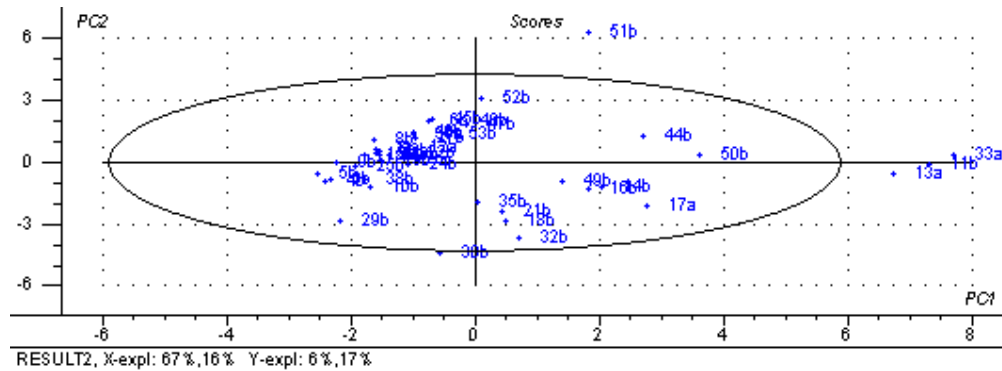
Figure 151: Hoetelling T^2 plot.
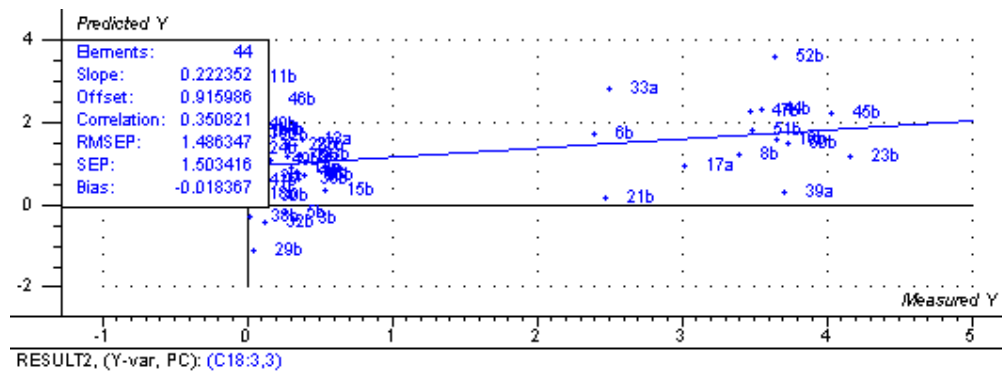
Figure 152: Predicted vs. Measured plot.

The obtained model has 44 elements and 3 components are used for modelling.

RMSEP: 1.486347, Correlation: 0.350821, Slope: 0.222352, Bias: -0.018367.

RMSEC: 1.172535, Correlation: 0.635885, Slope: 0.404350, Bias: 1.471e-07.

The R^2 Value is too low to be considered as a good model.

3.4 Summary Table:

	FTNIR	FTIR (100-60µm)	FTIR(110-10 µm)
PV	0.895829, 44, 9	0.563966, 44, 7	0.354497, 46, 3
SN	0.919430, 50, 5	0.179600, 47,12	0.415035, 47, 3
C12	0.889381, 42, 4	0.914353, 47, 3	0.634302, 46, 3
C14	0.931319, 43, 5	0.947111, 47, 4	0.675596, 46, 3
C16	0.966306, 43, 5	0.820407, 38, 4	0.797732, 45, 3
C18	0.241466, 43, 3	0.774063, 39, 5	0.421025, 43, 5
C18:1	0.946872, 43, 5	0.966598, 36, 3	0.831849, 46, 6
C18:2	0.981412, 43, 7	0.987900, 39, 7	0.826082, 44, 6
C18:3	0.643847, 43, 4	0.956374, 37, 8	0.350821, 44, 3

Table 6: Summary of all the results.

The above table shows R^2 value, Number of samples or elements and number of components. For example in the above table - 0.895829, 44, 9; the 0.895829 represents R^2 value, 44 represents number of samples and 9 represents the number of components used for modelling. From the above table it can be seen that there is an order of decrease in correlation from NIR >100-60µm> 110-10 µm. In all parameters NIR is found to have higher correlation compared to IR except in C18 and C18:3. While it is always found that IR100-60µm had good R^2 value compared to FTIR 110-10 µm except in case of SN. C18:1 and C18:2 are most found fatty acids in many oils and these are found to have good R^2 value for all the NIR and IR analytical instruments used in this project.

4. Discussion

About 53 samples have been used for the analysis of Peroxide value and saponification number while 49 samples for Free fatty acid parameter. The oils used for analysis are quite different from each other which include peanut, avocado, sunflower, olive, palm, sesame etc oils.

Full MSC, SNV, 1st & 2nd Derivatives pretreatments were done for the FTNIR data. 1st and 2nd derivative was done using Savitzky Golay for 53 samples and 1686 variables. The obtained results are compared and SNV is further used as pretreatment for multivariate analysis of all the parameters which includes Peroxide value, Saponification Number, C12, C14, C16, C18, C18:1, C18:2 &C18:3.

The outliers observed in the models have been removed based on the X-Y relation outliers plot, Hoetelling T^2 plot, Predicted vs Measured plot & Residual variances plot.

Peroxide value:

In case of FTNIR the correlation coefficient is 0.89 which is good but the slope value is less far away from the value of 1. After the removal of outliers the obtained model was with 9 components making the model over fitting. And overfitting models are considered

as bad ones. The first minimum validation variance curve was at 9 components and the components if chosen lower than this have very low slope, lower correlation and higher RMSEP values and for these reasons 9 components have been chosen for modelling.

This parameter has been observed with outliers as always being in the model even after the removal of few outliers. This is the case in A110-10, A100-60 & FTNIR data as well.

The validation Correlation coefficient has a value of 0.563966 which is very low and this is observed in the A100-60 path length FTIR model, thereby making the model bad. In the case A110-10 path length FTIR, the correlation coefficient value is 0.35 which indicates a bad model. Peroxide value is the stability parameter of oils which will change as the oxidation of oils takes place with time. However the oils are refrigerated all the time, it is possible to be oxidised when they are kept out for analysis. Before analysis the oils need to be brought to at least to room temperature which might have an effect on the oxidation.

Saponification Number:

This parameter resulted in a good model. FTNIR data, when analyzed for this parameter it showed a model with a very high correlation coefficient value of 0.919430 the slope value of 0.7 approximately makes the model very good with a good fit using 5 components for 50 samples. However the model still shows samples like 48 and31 being far away from other samples in the predicted vs. Measured Y plot. In the case of A110-10 path length FTIR measurements the model turned out to be very bad with a very low correlation coefficient of approximately 0.17 with very low slope value of 0.29 makes the model very bad with 47 samples and 12 components.

Free Fatty Acids:

Pre-treatment results of the C12 parameter are observed for the FTNIR data and the models show high correlation coefficient and of all, the SNV pre-treatment showed slightly higher R^2= 0.889381value than others.

C18:1 and C18:2 free fatty acid parameters showed good models of all other free fatty acids. They are observed to have the models with high correlation. In the A100-60 path length C18:2 is observed to have a very good model with no outliers.

For palm oils C16 & C18:1 is observed to be predominant fatty acids, for sunflower C18:2 and walnut oil also C18:2, for olive oils C18:1 is found to be present in abundance compared to other fatty acids according to **Table 7 &8**. For most of the oils C18:1 and C18:2 are found to be the predominant fatty acid. And this found to have good correlation values in this project IR and NIR data analysis.

Oleic acid with carboxyl C=O absorption band at 1711cm[-1] is important for all oils containing majorly this fatty acid. While for palm oils absorption region of 1715 to 1708 cm[-1] are found to be important to determine FFA[6]and these wave numbers are also found to be important in the FTIR spectra collected for this project.

FFA of edible oils when analyzed by FTIR using CaF_2 cells at [7]100µm gave R^2=0:999. Based on these results path length of 110 -10 µm have been chosen for this project analysis of oils.

[31]Palm oils contain a mixture of high and low melting triglycerides. At ambient temperatures, higher melting triglycerides will crystallize into a solid fraction called stearin, while the lower melting triglycerides will remain in a liquid form called olein. Hence these oils require higher temperatures for the analysis.

5. Conclusions

Edible oils are analysed by using FTNIR and FTIR and their spectra are compared with the reference analysis. The temperature limitations were set to 30°C in the NIR instrument due to the temperature limitations in OenoFoss instrument. Through the analysis of all the samples it is observed that the palm oils are present far away from the other oils in both NIR and IR data. Thereby it can be concluded that this temperature setting is not good enough for the palm oil analysis.

Of all the models C18:2 at A100-60µm path length resulted in a very good model. This parameter is observed with a good modelling even with NIR by high correlation coefficient values showing clusters of similar types of oils. The instrument design is causing a lot of air bubbles during the IR measurements, where the air bubbles are formed during the movement of the lid up and down while setting the path length. Approximately 20% of the measurements have been repeated to obtain the correct values and whenever the values set go below the limits that have to be observed, air bubbles are found all those times. Prevention of air bubbles are essential for proper analysis.

A change in the path length might solve the problem of good correlation between the IR and reference values. Oils of the same kind are found in clusters. All palm oils are found to be closer together while the rape oils formed separate group and likewise other oils as well.

Suggestions: Future work can be done by making calibration curves of individual standard fatty acids in IR as it was done for GC/MS for quantification purposes.

PV correlation has decreased from NIR to 100-60µm, which further decreases in 110-10 µm. This can be due to oxidation of the samples while measuring that included time gap between measurements IR and reference analysis. In further experiments care should be taken about the time gap.

If the analysis has to be done for palm oils then the temperature of the instrument has to be increased or else these oils are difficult to analyse.

Pathlength 100[7], 50 and 25[29] µm are found to be effective in oil parameter determination that has been used in literature and this might work out if further any research work is carried with these pathlengths.

6. References

1. http://www.rssdi.org/1985_july-sept/proc7.pdf
2. Michael Jee, *Oils and fats authentication*, Blackwell publishing, 2002.
3. S. Armenta, S. Garrigues, M. de la Guardia, '*Determination of edible oil parameters by near infrared spectrometry*', Analytica Chimica Acta 596 (2007) 330–337
4. http://www.lab-robotics.org/upstate/Archives/061026%20material/vandevoort.pdf
5. Mousa A. Allam and S.F. Hamed '*Application of FTIR Spectroscopy in the Assessment of Olive Oil Adulteration*'**,** Journal of Applied Sciences Research, 3(2): 102-108, 2007.
6. Y.B.Che Man, G.Setiowaty, 'Application of Fourier transform Infrared spectroscopy to determine free fatty acid contents in palm olein', Food chemistry 66, 1999, 109-114, Elsevier.
7. Ahmed Al-Alawi, Frederick R. van de Voort, Jacqueline Sedman, and Andrew Ghetler 'Automated FTIR Analysis of Free Fatty Acids or Moisture in Edible Oils'Journal of the Association For Laboratory Automation, 2006, Vol11; Number 1, Pages 23-29.
8. Kim H.Esbensen, '*Multivariate Data Analysis in practice'*, 5th Edition Camo 2002.
9. http://acpcommunity.acp.edu/Facultystaff/hass/oc1/exp/nabh4/nabh4des.html
10. Hu Li, F.R Van de voort, A.A. Ismail and R.Cox ; Determination of peroxide vaue by fourier transform Near – infrared spectroscopy, JAOCS, Vol.77 no.2 , 2000.
11. Yasushi Endo, misako Tagiri – Endo, and Kenichiro Kimura, Rapid determination of Iodine values and saponification value of Fish oils by Near infrared spectroscopy, Jounal of food science vol70, Nr2, 2005.
12. Gulgun Yildiz, Randy L.Wehling, and Susan L. Cuppett, 'Comparison of Four Analytical methods for the Determination of Peroxide value in Oxidized soybean oils', JAOCS, Vol.80, no.2, 2003.
13. http://www.chem.vt.edu/chem-ed/ms/quadrupo.html.
14. Alan J. Handley and Edward R.Adlard, '*Gas Chromatographic Techniques and Applications*', 2001 Shefield Academic Press.
15. Paul Worsfold, Alan Townshend, Colin Poole, 'Encyclopedia of analytical science', second edition, Elsevier 2005. Page 385.
16. http://www.international- agrophysics.org/artykuly/international_agrophysics/IntAgr_2002_16_2_83.pdf
17. Seamus P.J Higson, *Analytical chemistry*, oxford university press, 2004
18. Yukihiro ozaki, W. Fred Mc Clure and Alfred A. Christy. '*Near infrared spectroscopy in food science and technology*', Wiley interscience 2006.
19. http://books.google.co.in/books?id=E00CIb_MFl4C&pg=PA181&lpg=PA181&dq=what+is+fit,+reverse+fit+and+purity&source=web&ots=S4cccfZUik&sig=W8QYIDnOx4NmEbAlRAGXw-ZIzw8&hl=en#PPA181,M1
20. Richard G.Brereton, 'Chemometrics- Data Aanlysis for the Laboratory and Chemical Plant', 2003 Wiley.

21.   Marvin McMaster and Christopher McMaster, 'GC/MS A Practical User's Guide', 1998 Wiley-VCH.

22.   Raymond E. March and John F. J. Todd, *Practical Aspects of Ion Trap Mass Spectrometry*, Volume III CRC press 1995.

23.   Kenneth R. Beebe, Randy J. Pell and Mary Beth Seasholtz, *Chemometrics: A practical Guide*, 1998 John Wiley & Sons, Inc.

24.   http://www.scientificpsychic.com/fitness/fattyacids1.html

25.   L.Alonso, M.J. Fraga. M. Juarez and P. Carmona, 'Fatty acid composition of Spanish shortenings with special emlhasis on trans unsaturation content as Determined by Fourier transform Infrared spectroscopy and Gas chromatography', JAOCS, Vol. 79, no.1, 2002.

26.   William Horwitz **, '**Official Methods of Analysis of AOAC International, 17th Edition', AOAC International (2000).

27.   Tetsuo Sato, Sumio Kawano and Mutsuo lwarnoto, '*Near Infrared Spectral Patterns of Fatty Acid Analysis from Fats and Oils'* JAOCS, Vol, 68, No. 11, 1991.

28.   Enriqueta Bertran, Marcelo Blanco, Jordi Coello, Hortensia Iturriaga,Santiago Maspoch, and Ivan Montoliu, '*Determination of Olive Oil Free Fatty Acid by Fourier Transform Infrared Spectroscopy,'* JAOCS, Vol. 76, no. 5 (1999).

29.   J. Sedman, F.R. van de Voort, and A.A. Ismail, 'Upgrading the AOCS Infrared *trans* Method for Analysis of Neat Fats and Oils by Fourier Transform Infrared Spectroscopy', JAOCS, Vol. 74, no. 8 (1997).

30.   www.foss.dk.

31.   O. Zaliha, C. L. Chong, C. S. Cheow b, A. R. Norizzah and M. J. Kellens**, '**Crystallization properties of palm oil by dry fractionation' 2003 Elsevier Ltd.

32.   Yukihiro ozaki, W. Fred Mc Clure and Alfred A. Christy 'Near infrared spectroscopy in food science and technology', Wiley interscience 2006.

7. APPENDIX

7.1 Reference Analysis:

7.1.1.Laboratory procedures:

7.1.1.1.Saponification Number:

Principle: The saponification number is value that expresses the amount of mg (milligram) of potassium hydroxide (KOH) that is used to saponify the esters and to neutralize the acids in 1 gram of oil.

Reagents:

Alcoholic KOH 0.5M

30g KOH is dissolved in 50ml water and is diluted with ethanol. After 24 hours it is decanted and stored in dark.

HCL 0.5M: It is important to use HCL of exact molarity.

Procedure:

Weigh 1gm of oil into a clean and dry conical flask. Add 25ml alcoholic KOH. Whirl the conical flask vigorously, add a magnet and then boil it with a cooler on top. Titrate immediately after adding 5 drops of phenolphthalein with 0.5 M HCl until the colour changes from red back to the original oil colour (i.e.; colour before adding phenophthalein).

BLANK: Same procedure without sample.

Boiling Time:

½ hour: two replicates with sample and one blank.

Calculations:

Moles of KOH used to saponify the esters and neutralize the free fatty acids is equal to the moles HCl used to titrate the blank minus the average moles HCl used to titrate the samples. The ratio can be expressed by the formulas below:

$$mgKOH / gramoil = \frac{difference(mol) \times 56.11(g/mol) \times 1000(mg/g)}{a(g)}$$

Difference is the difference between the sample and the blank.

A is the weight of the sample.

7.1.1.2. PEROXIDE VALUE:

The peroxide value I_P is the number that expresses in mill equivalents of active oxygen the quantity of peroxide contained in 1000g of the substance, as determined by the methods described below.

Method A

Place 5g of the substance to be examined (*m* g) in a 250 ml conical flask fitted with a ground glass stopper. Add 30ml of a mixture of 2 volumes of chloroform and 3 volumes of glacial acetic acid. Shake to dissolve the substance and add 0.5ml of saturated potassium iodide solution. Shake for exactly 1 min then add 30ml of water. Titrate with 0.01M sodium thiosulphate, adding the titrant slowly with continuous vigorous shaking, until the yellow colour is almost discharged. Add 1ml of starch solution and continue the titration, shaking vigorously, until the colour is discharged (n_1 ml of 0.01M sodium thiosulphate).

Carry out a blank test under the same conditions (n_2 ml of 0.01M sodium thiosulphate). The volume of 0.01M sodium thiosulphate used in the blank titration must not exceed 0.1ml.

$$I_p = \frac{10\,(n_1 - n_2)}{m}$$

Peroxide value Ip is calculated using the above formula.

7.1.1.3. FREE FATTY ACIDS:

REAGENTS:

0.5 M Sodium hydroxide in methanol.

0.1% (w/v) Hydroquinone in methanol

Boron triflouride in methanol.

Internal Standard : 200mg of (c17) methyl heptadecanoate in 100ml of heptane.

Methyl Heptadeconoate (Fluka) – Mol. Formula $C_{18}H_{36}O_2$ Mol.Wt : 284.49

Saturated Sodium chloride (300gm of Nacl in 1lt of water).

Procedure:

Weigh 100mg of oil in a methylene tubes. Add 2ml of 0.5M methanolic sodium hydroxide, tighten the stopper and boil for 5min on a water bath. If oil is still observed at the bottom of the tube then heat it for 2min more.

Now add 3ml of borontriflouride reagent and 0.1 ml of hydroquinone solution. Tighten the stopper and boil for 5min in water bath. Then remove the tube from the water bath and cool to room temperature.

Add 10ml of saturated sodium chloride and 5ml of Internal standard. Then shake for 30 seconds and leave it to stand still until phase separation. Remove the upper phase with a pipette of approximately 1 ½ ml and place it in a vial.

Standards used for Calibration curve:

Palmitic acid, Myristic acid, stearic acid, Lauric acid, Linolenic acid, Linoleic acid, Oleic acid.

Weigh approximately 100mg of each of the above acids and follow the above procedure for free fatty acid determination. The phase exctracted is of concentration 20g/lt which is diluted with internal standard to 10g/lt (1:1), 5g/lt (1:3), 2.5g/lt (1:7), 1g/lt (1:19).

$$\text{TFFA} = \frac{result_{g/l} \times 5_{ml} \times 1000_{mg/g}}{1000_{ml} \times a_{g}}$$

Total free fatty acids= TFFA.

      a= weight of the sample in grams.

      Result= fatty acid quantity from the chromatogram result.

7.2 NIR:

BOMEM Near Infrared Spectrophotometer MB160 was used for the near infrared analysis. The procedure was carried according to the user instructions manual at the AAUE laboratory. InAs detector has been used along with 3.2% transmission filter and two windows. The temperature of the thermostat is maintained at 30°C; gain was adjusted to A for empty vial and for samples B gain. Alignment was above 70 - 80% for all samples with few exceptions. The resolution is 8cm^{-1} with 128 scans for background spectra and 64 scans for the samples. All the files are named by the sample names while the back ground spectra file name is 11022008A. The wave number range is 4500-11000 cm^{-1}. Back ground spectra was collected after every hour.

Figure 153: Bomem MB NIR instruement.

Lab Horfro oil B gain 49% Allign but with C gain allignment was ADC overflow. So set back to B gain and collected the absorbance.
Lab Almond oil Allign 52%.
Lab coconut oil Allign 55%
Lab Thisle oil Allign 52%
Lab sunflower oil Allign 53%
Lab Rape oil Allign 53%

7.3 FTIR Procedure:

While taking the IR spectra for 100 & 60 µm, just control +d has to be pressed after placing the oil in the cuvette and closing it. When taking the IR spectra at 6 different path length ie 430, 220, 110, 60, 30 µm the same - control +d has to be pressed. But for the first reading 430 µm three thin plates of 300 µm each are placed on the three elevating minute heads present near the cuvette. After closing and taking the spectra of this path length. The lid of the apparatus opens and then these three elevations have to be replaced by 150 µm and then taken the spectra for 220 µm. After this these three elevations are removed and press control +d then the system automatically collects the spectra of the other three path length in the decreasing order of path lengths. During this process there are some figures that have to be observed i.e. the values that are approximately equal to the path length numbers are observed on the screen which when goes too below the range of 420 for 430 µm path length, 250 for 220 µm path length, etc then the system is stopped and opened the lid to check for air bubbles in the cuvette. This parameter to observe the air bubbles quite matched with the numbers set for observation. And when the air bubbles are observed they are removed by placing new or fresh oil sample and taking the spectra again. Each path length spectral reading took ~18 sec while the whole reading for the 6 path length for one single time took about 3½ minutes, for two replicates it took 7 minutes. But due to some software errors during the process of IR measurements the total time of collecting the spectra has been increased by 1½minutes thereby making each sample measurement of 6 path length to 5 minutes. Most of the time the air bubble is observed on the upper side of the lid and it is observed that the air bubble is not observed in the longer path length but appears in the shorter path length thereby it is obvious that the lid when moving down for taking IR readings of lower path length is creating some air bubble as well as vice versa is also observed for about 20% of the total measurements. Even after taking the measurements repeatedly with fresh oil sample after observing the values on the computer screen which play a key role in expecting air bubbles, the spectral collection showed air bubbles and for this reason few oil samples have been not used for data analysis.
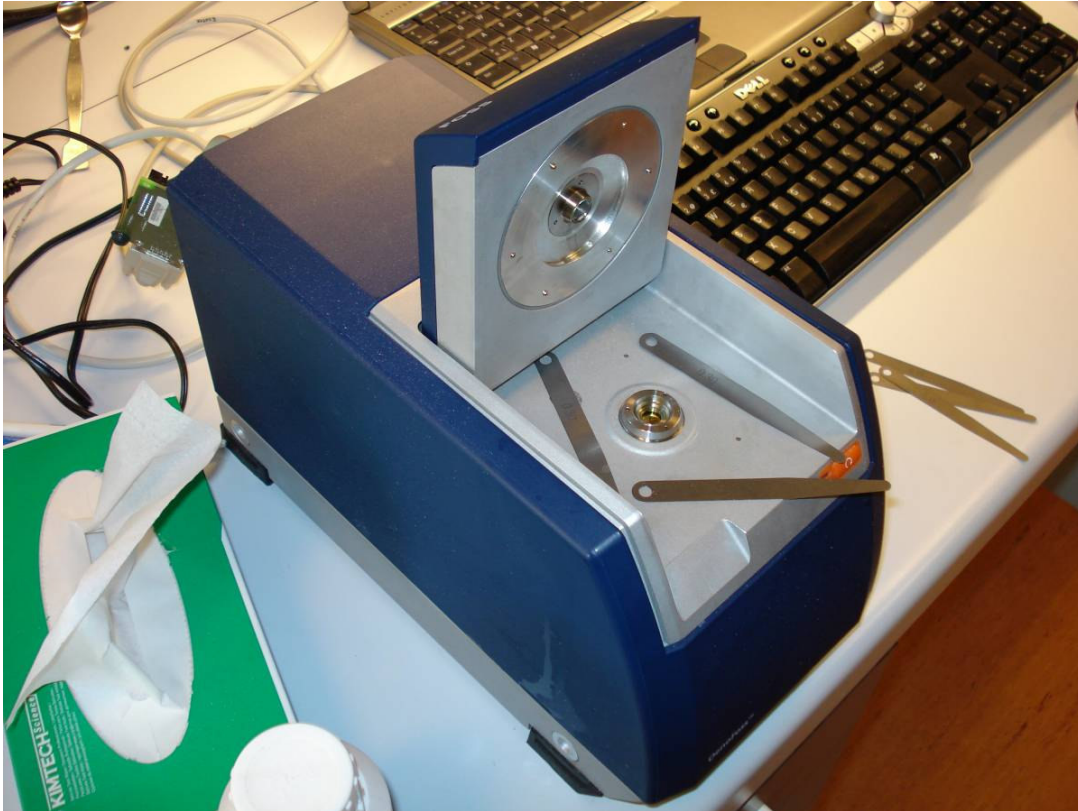
Figure 154: OenoFoss IR instruement.

7.4. Chemometrics:

Using Validation variance to detect problems:

In PLS any increase in the validation variance is a bad sign, usually because of the presence of outliers, noise, nonlinearities etc. When an increase in the first PCs is observed then it means X structure which is not relevant for Y has been modelled.

Generally a valid model describes only the systematic variations but not the random variations i.e. noise. Hence one has to choose a model that will give a low or minimum prediction error with as few components as possible.

If the model has noise then the resulting model will fail to predict new samples or objects thereby it will have too many components thereby making the model overfitting. Usually the overfitted models are bad[8].

Cross validation: Segmented cross validation is used when there is relative abundance of samples available for the training set and one is not sure to pick representative ones or when full cross validation is too time consuming.

In the segmented cross validation the calibration set of samples are divided into segments with for example 10% of the samples in each segment. When this cross validation is performed the samples present in one segment will be left out and the remaining are modelled. And then the other segment samples are left out of the model but the previous one is considered this time for modelling. So this way each time one whole segment samples will

be kept out of modelling. For a realistic approach of segmented cross validation the samples in a calibration set should be divided into few segments with a minimum of 10% of the samples in each segment[8].

Replicates: A means to quantify errors:

When the same samples are measured for several times to compare the results for repeatability then this will actually quantify the inaccuracies. In this project case even though the samples have been taken separately from the bottle of oils, the oils are not shaken or mixed every time to avoid air bubbles hence these are considered as sampling almost the same kind of sample and thereby resulting in quantifying the inaccuracies. To damp these inaccuracies average function will help. Hence all the samples have been averaged by a reduction factor of 2.

Bias: systematic difference between the predicted and measured values.

RMSEP: Root mean square error of prediction. RMSEP can be interpreted as the average prediction error, expressed in the same units as the original response values.

RMSEC: Root mean square error of calibration. A measurement of the average difference between predicted and measured values, at the calibration stage.

X-Y relation outliers plot: In a well behaved regression model usually all the samples will lie close to the straight regression line in the X-Y relation outliers plot.

Multiplicative scatter correction: The spectroscopic measurements of certain types of samples exhibit light scattering effects. This usually observed in NIR data, but it can also be observed in IR spectra by either background effects or varying optical path length or even by temperature and pressure variations.

Multiplicative scatter correction is a type of pretreatment or transformation method which compensates both multiplicative as well as additive effects

7.5 GCMS:

Introduction

Gas chromatography-mass spectroscopy (GC-MS) is a hyphenated analytical techniques. It is actually two techniques that are combined to form a single method of analyzing mixtures of chemicals. . Generally chromatography is used for the separation of a mixture of chemicals into individual components. The combination of gas chromatography (GC) for separation and mass spectrometry (MS) for the purpose of detection and identification of the components of a mixture of compounds has become the definitive analytical tool for both commercial as well as in the research laboratories. Combining these two techniques, it is possible to both qualitatively and quantitatively evaluate a solution containing a number of chemicals.

The GC-MS can extensively be used in the medical, pharmacological, environmental aspects.

Gas Chromatography

In general, chromatography is used to separate mixtures of chemicals into individual components. When once isolated, the components can be evaluated individually

Separation takes place when the sample mixture is introduced (injected) into a mobile phase. The mobile phase is an inert gas such like helium. The sample mixture is carried by the mobile phase through a stationary phase. The stationary phase is a usually chemical that can selectively attract components in a sample mixture. The stationary phase is usually contained in a column. The Columns can be glass or stainless steel of various dimensions.

The mixture of compounds in the mobile phase interacts with the stationary phase in the column. Each compound in the mixture will interact at a different rate, and those that interact the fastest will elute from the column first, while those that interact slowest will exit the column last. By changing the characteristics of the mobile phase and the stationary phase, different mixtures of chemicals can be separated. Further refinements to this separation process can be made by changing the temperature of the stationary phase or by changing the pressure of the mobile phase. The capillary column is held in an oven, the oven can be programmed to increase the temperature gradually and this helps separation. As the temperature increases, those compounds that have low boiling points elute from the column sooner than those that have higher boiling points.

As the compounds get separated, they will elute from the column and enter a detector. The detector is capable of creating an electronic signal whenever the presence of a compound is detected. As long as the GC conditions (oven temperature ramp, column type, etc.) are the same, a given compound will always elute from the column at nearly the same retention time. By knowing the RT for a given compound, we can make some assumptions about the identity of the compound. However, compounds that have similar properties often have the same retention times. Therefore, more information is usually required before one can make an identification of a compound in a sample containing unknown components.

Good results of GC/MS are based on the correct selection of the injector type, column support, carrier gas, oven temperature and properly designed interface feeding into the ion source can generally either make or break the mass spectrometric analysis.

There are various number of different possible GC/MS configurations, but all have similar components. The injector, gas chromatograph with its carrier gas source, temperature control oven & tubing to connect to the injector to the column as well as out to the mass spectrometer interface, a column with stationary phase, interface to mass spectrometer, Ionization source, mass analyzer, detector and data control system.

The heated transfer line maintains solutes in the vapour phase as they pass from the GC to the mass spectrometer. The ion source ionizes the solutes and propels them towards the mass analyser where they are separated according to their mass to charge ratio (m/z). The detector produces a voltage in response to the impact of the resolved ions, and this signal is then amplified and converted, via on board electronics and a data sytem, into mass spectral information. Helium is usually employed as the carrier gas in GC/MS systems.
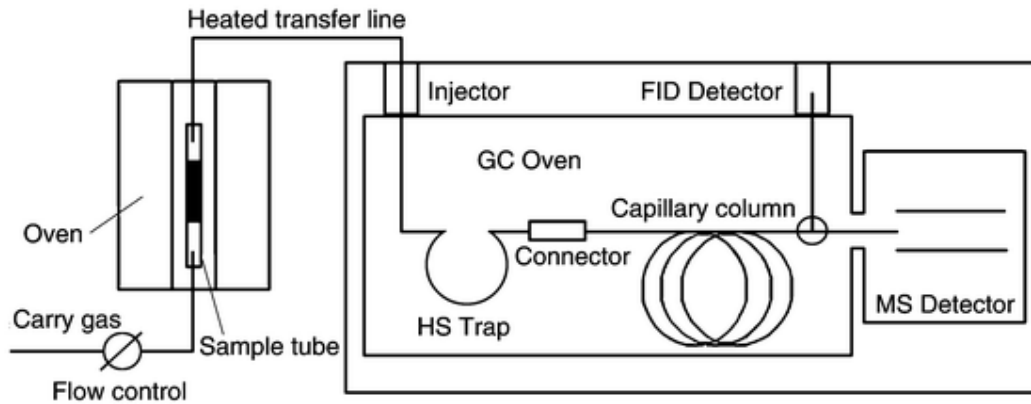
Figure 155: GC/MS diagram.

The higher the concentration in the sample, larger will be the signal. The signal will then be processed by a computer. The time from injection (time zero) to elution is known as the retention time (RT). The computer will generate a chromatogram (graph) from the signal as in figure 1. Each peak in the chromatogram represents the signal created when a compound elutes from the GC column into the detector. The x-axis shows the RT, and the y-axis shows the intensity (abundance) of the signal. In Figure 156, there are several peaks that are labelled with their RTs. Each peak represents an individual compound that was separated from a sample mixture. Dodecane showed a peak at 4.97 minutes, the peak at 6.36 minutes is from biphenyl, chlorobiphenyl peak is observed at 7.64 minutes, while the peak at 9.41 minutes is from hexadecanoic acid methyl ester.
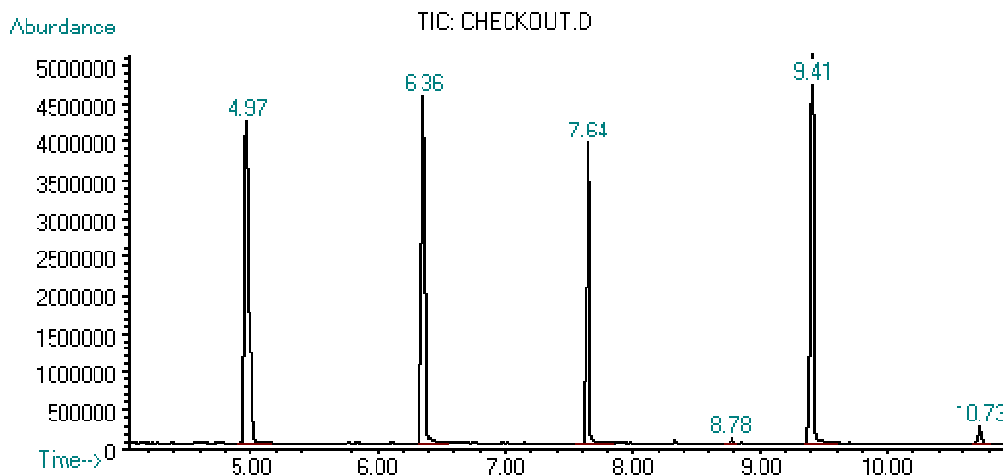
Figure 156: Chromatogram generated by a GC.

Ionisation techniques:
Electron Impact Ionisation b) Chemical Ionisation c) Field Ionisation.
Eletron Impact Ionisation (EI):
     In EI, the electrons that are produced by passing an electric current through a tungsten or rhenium filament are accelerated across the ion source by a potential of approximately

70V. These electrons will interact with gaseous analyte molecules that enter the ion source from the transfer line removing an electron to create positive ions. The two magnets that are present on eitherside of the ion source, collimate the ionising electrons into a narrow beam.

The EI is a relatively ineffecient process, ionizing only approximately 1% of the molecules that will enter the ion source. The positive ions produced are ejected by a positive voltage applied to source or by negative voltage on ion lenses mounted between the ion source and analyser region. The trap is mounted on the opposite side of the source to the filament. It is held at a positive potential and serves for the electrons removal that have traversed the ion source.

EI ionisation will produce extremely energetic molecular ions from the analyte molecules, and these molecular ions are formed by th eremoval of a single electron from a compound's molecular orbital, thereby resulting in the formation of a radical cation. The molecular ion cleaves to form fragment ions and nuetral species and these produced fragment ions give rise to an information rich EI mass spectrum.

EI is the most popular ionisation mode for routine analysis and its main advantage lies in the spectral reproducibility, relatively high sensitivity, abundance of fragment ions and its applicability to a wide range of compound types.
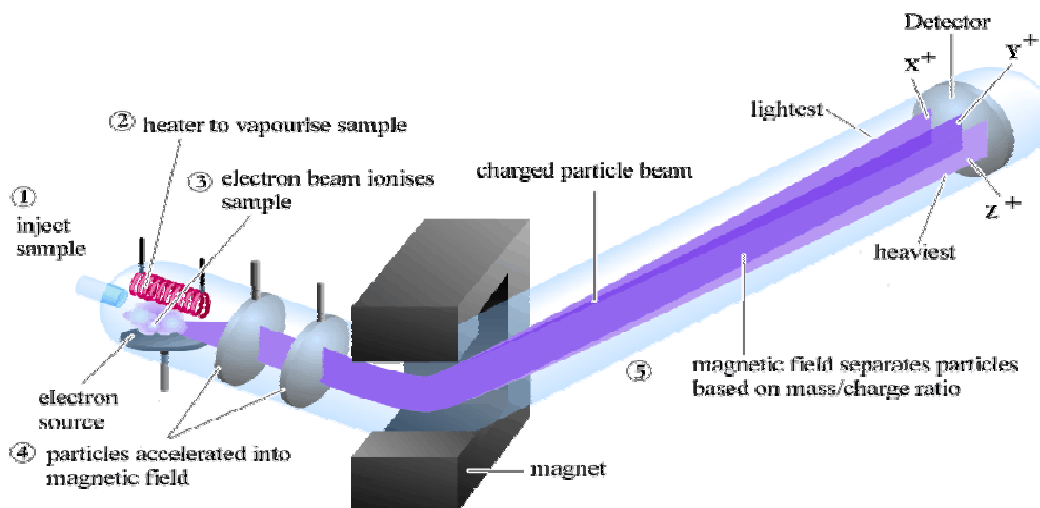
Figure 157: EI[9]

Chemical Ionisation (CI):

A reagent gas is used in CI for thepurpose of ionisation of analyte species. The ionisation process is less energetic than in EI. As a recult, the internal energy of the ion is

reduced, stabilising the molecular ion and thereby reducing the fragmentation. The ions loose some of the excess internal energy through collisions with the reagent gas molecules.

Both positive ions and negative ions can be produced by CI. For both the ions production a reagent gas will be introduced intot he source region of the mass spectrometer, raising the pressure to around 70-130 Pa (0.5-1 torr).

Chemical ionisation tends to be less reproducible when compared to electron impact ionisation, both in terms of appearance of mass spectra and the total ion yield. This is basically due to difficulty in maintaining stable pressure and temperature conditions in the ion source. A small increase in temperature will reduce the relative intensity of the molecular ion by increasing its internal energy, whereas the fluctuations in pressure will either increase or decrease the penetration of ionising electrons, thereby causing variations in the number of reagent gas ions formed.

The presence of small amount of impurities also influence the reactions in chemical ionisation. Therefore the reagent gases should be 99.9% pure. Problems can also be caused by the prolonged use of hydrocarbon based reagent gases like methane or isobutane which will form deposits that will polymerize on the metal surface of the ion source. This will result in the formation of insulating layer that is capable of supporting static charge. Symptoms of source contamination will range from altered tuning conditions to loss of sensitivity.

In positive CI, cations are produced form the reagent gas by electron bombardment and these ions produced will react with sample molecules to form a charged species by proton transfer.
Proton transfer reactions in positive CI will result in the formation of a pseudo molecular ion $(M + H^+)$, one mass unit heavier than the true molecular ion, often seen as a base peak.
Positive CI is commonly used to determine th emolecular masses of unknown compounds.
The Negative CI is the most selective ionisation method for GC-MS and is applicable mainly to compounds that are capable of stabilizing the negative charge. These usually includes halogenated species, phosphate esters and nitro compounds.The main advantages of negative CI are improved selectivity and sensitivity.

For the recording of these negative ions by an electron multiplier detector modification is necessary. A conversion dynode which is a metal plate mounted just before the multiplier. When a negative ion strikes the dynode it causes the emission of positiive ions, which are then detected by electron multiplier.

Analyser Types:

The ions produced from analyte molecules are resolved by mass analyser. Depending on the analyser's resolution the mass spectrometer separates ions with greater or poorer effeciency.

Quadrupole Mass Analyser:

This type of mass analyser comprises of four metal rods either of cylindrical or parabolic cross section that are arranged parallel to each other. Opposite rods are electrically connected to a voltage supply. The voltages that is applied to each set of rods have opposite signs but will have equal magnitude.

The ions that enter the mass analyser are made to oscillate between the four rods and during this process the ions with unstable oscillation paths are neutralised by colliding with the rods.
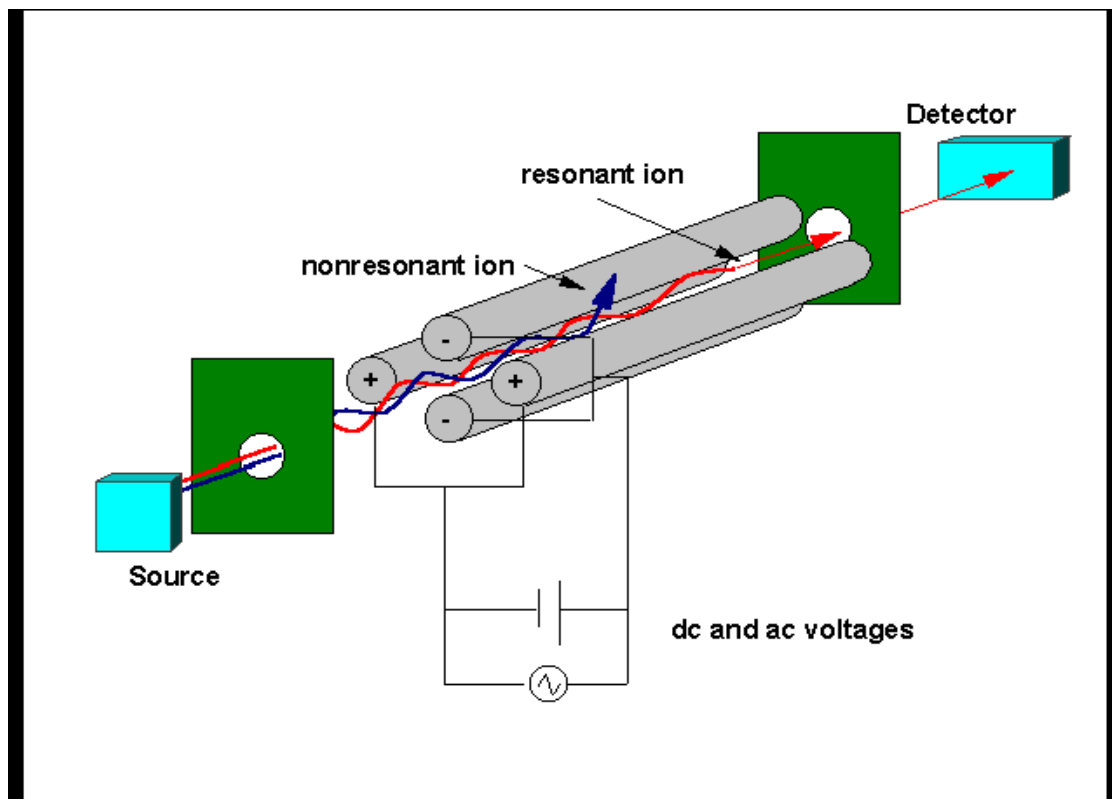
Figure 158: Quadrupole Analyser.

2) Ion Trap mass Analyser:

The Ion trap consists of a ring electrode to which r.f. voltage is applied which is sandwiched between two cylindrically symmetrical parabolic end cap electrodes. Ionisation of molecules that are introduced through the GC transfer line takes place in th eanalyser cavity, the region bounded by the endcap and ring electrodes. The electrons that are produced by an electrically heated filament are either permitted or denied access to the analyser cavity by the electron gate.

During the ionisation the ion gate will open the electron gate allowing the electrons into the analyser cavity after which the gate is closed and all the resulting ions are cofined to a stable trajectory inside the trap.

The kinetic energy spread of the ions can be reduced by a small pressure of helium carrier gas by collision, thereby improving th eresolution. As the r.f voltage is ramped the ions get destabilised in the order of increasing m/z. And these destabilized ions are moved towards the two end cap electrodes, while some of them travelling downwards pass through the holes in the bottom end cap electrode and reach the detector. Whereas th eions destabilised upwards towards th efilament are usually lost
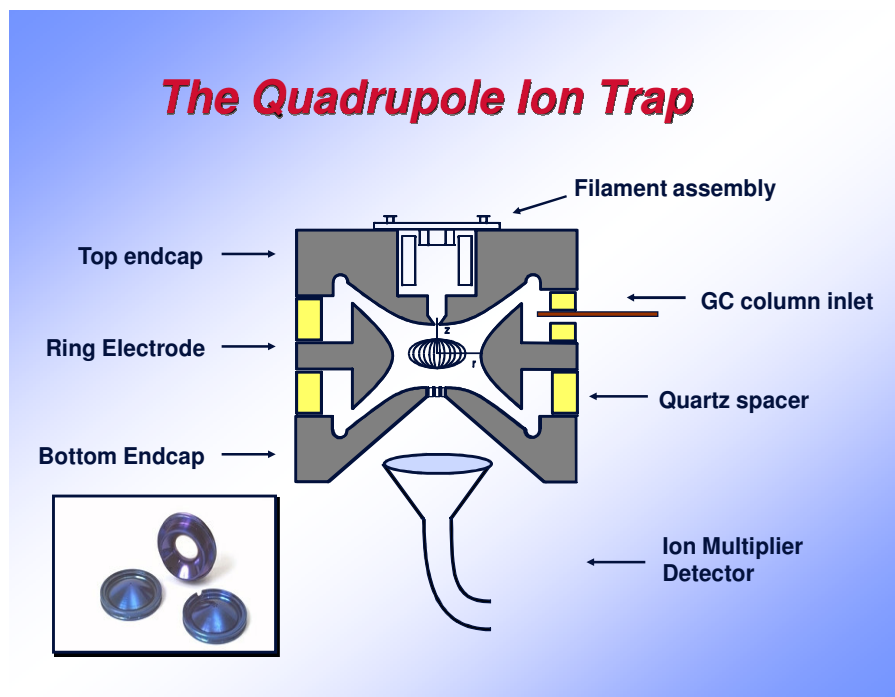
Figure 159: Ion trap.

Selected Ion Monitoring:

The selected ion monitoring (SIM) is used in target compound analysis to improve the limits of detection. This is particulary of interest for quantitative analysis of a particular compound in a sample to conform its presence or absence of this target compound in the sample. The compound characteristic ions that are of interest are selected and the instrument will be setup to scan only these ions.

The more the m/z values monitored in the simultaneously in SIM mode, the poorer will be th eresulting sensitivity.

SIM operation is most beneficial to quadrupole analysers where ions with unstable trajectories are not detected. Where as in TOF or Ion trap instruments aal ions will have atleast a nominal chance of being detected and SIM is an unnecessary tool. It is usual to monitor three or more different ions for each target compound to reduce the possibility of reporting a false postiive result[14].

Mass Spectroscopy

 The separated individual compounds will elute from the GC column to enter the electron ionization (mass spec) detector. There, they are bombarded with a stream of electrons causing them to break apart into fragments. These fragments can be large or small pieces of the original molecules.

The fragments are actually charged ions with a certain mass. The mass of the fragment divided by the charge is called the mass to charge ratio (m/z). Since most fragments have a charge of +1, the m/z usually represents the molecular weight of the fragment.

A group of 4 electromagnets known as quadrapole, focuses each of the fragments through a slit and into the detector. The quadropoles are programmed by the computer to direct only certain m/z fragments through the slit. The rest bounce away. The computer has the quadrapoles cycle through different m/z's one at a time until a range of m/z's are covered. This occurs many times per second. Each cycle of ranges is referred to as a scan.

The computer records a graph for each scan. The x-axis represents the m/z ratios. The y-axis represents the signal intensity (abundance) for each of the fragments detected during the scan. This graph is referred to as a mass spectrum.
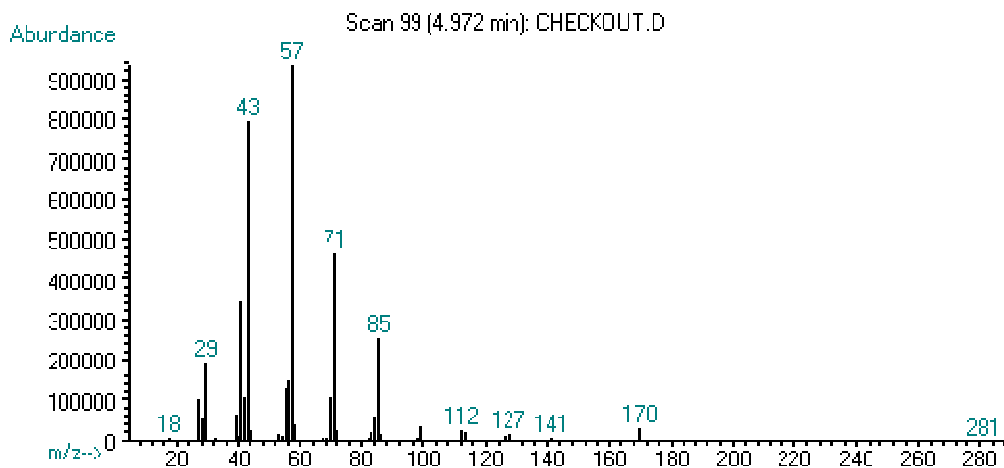
Figure 160: Mass-spectrum generated by an MS.

The mass spectrum produced by a given chemical compound is essentially the same every time. Therefore, the mass spectrum is essentially a fingerprint for the molecule. This fingerprint can be used to identify the compound. The mass spectrum in Figure 160 was produced by dodecane. The computer on our GC-MS has a library of spectra that can be used to identify an unknown chemical in the sample mixture. The library compares the mass

spectrum from a sample component and compares it to mass spectra in the library. It reports a list of likely identifications along with the statistical probability of the match.

7.6 INFRARED Spectroscopy

The infrared (IR) region of the electromagnetic spectrum lies between ~10 and 12800cm^{-1}. IR spectroscopy is the study of interaction of IR radiation with matter as a function of photon frequency. This interaction will take either of the form absorption, emission or reflection. IR spectroscopy is a fundamental analytical technique in order to obtain quantitative and qualitative information about the sample which is either in a solid, liquid or vapour state.
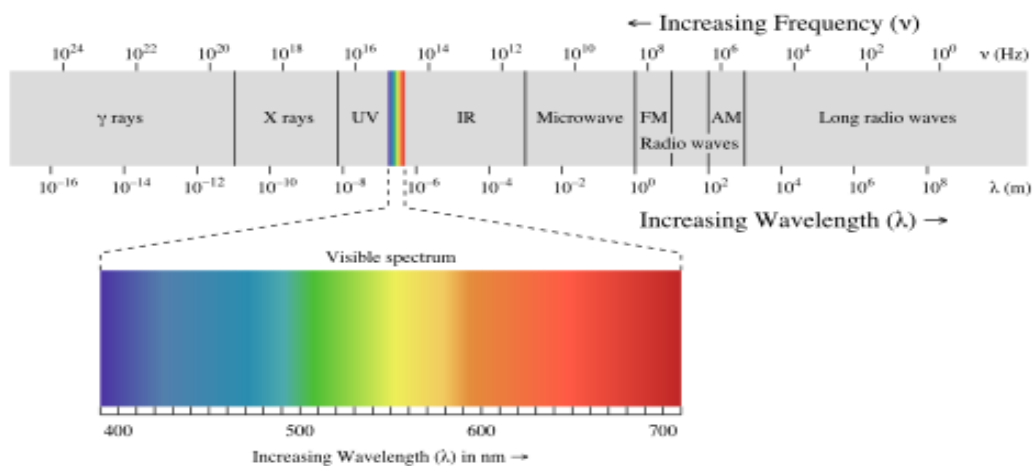
Figure 161: Infrared spectroscopy.

IR region is divided into far IR (10-200 cm^{-1}), mid IR (200-4000 cm^{-1}) AND near IR (4000-12800 cm^{-1}). Mid IR corresponds to fundamental transitions where one vibrational mode will be excited from its lowest energy state to its first excited state. The mid IR spectrum of a particular substance is a unique finger print that helps in its identification when compared with a reference spectrum. However when no reference is available then it can be used to identify the presence of certain structural characteristics.

Near IR spectroscopy corresponds to transitions in which a photon excites from the ground state to the second or higher excited states (overtones) or it arises due to transitions in which one photon simultaneously excites two or more vibrational modes (combination bands)[15].

NIR spectroscopy is particularly sensitive to the presence of molecules containing the C-H, O-H, and N-H groups. These bonds interact in a measurable way in the NIR portion of the spectrum, thus constituents such as starch and sugars (C-H), alcohols, moisture and acids (O-H), and protein (N-H) can be quantified in solids, liquids and slurries. In addition, an analysis of gases is possible. NIR is not a trace analysis technique and is generally used for measuring components that are present at concentrations greater than 0.1%. NIR creates a faster, safer working environment and does not require chemicals[16].

There are mainly four types of spectrometers that are routinely used for the analytical applications.
1.   Dispersive grating spectrometer.
2.   Fourier tansform Multiplet spectrometers.
3.   Non dispersive photometers and
4.   Reflectance photometers.

Dispersive grating spectrometer: These spectrometers are instruments in which the spectrum is analysed sequentially following the dispersion of multi wavelength radiation by means of either by a monochromatoer or by diffraction grating. These dispersive spectrometers have been replaced by the fourier transform instrumentation.

The dispersive spectrometers are normally double beam instruments in which double beam will be in conjunction with a beam chopper.
Michelson Interferometer:
Instead of monochromators, multiples systems based on fourier transform algorithm use Michelson interferometers for coding the polychromatic spectrum of the source in both the case of IR as well as for Raman spectrometers.
   Fourier transform IR (FTIR) spectrometers provide unrivalled speed, resolution, sensitivity as well as wavelength precision detecting and measuring all wavelengths simultaneously.
Normally in *Michelson interferometer,* the sample is first irradiated polychromatic IR radiation by means of two beams that are of nearly equal intensity. The radiation from the source is directed towards the beam splitter so that half of the radiation is transmitted and half of it is reflected. These two beams of radiation are again reflected back to the beam splitter by the mirrors, of which one is movable and the other immovable. This combined beam is now directed towards the sample and detector. Generally the beam splitter is placed at an angle of 45° relative to the beam.
   In the ideal case the light striking the beam splitter is split into 50% of light reflecting to the fixed mirror and 50% transmitted to the movable mirror. When the mirrors are placed equidistant then the two beams travel identical distance while reflecting back to the beam splitter and so they recombine resulting in a constructive interference. When both paths are equidistant then the moving mirror is placed at a location known as ZPD (zero path length difference).
   In situations when the moving mirror is displaced to a quarter of wavelength then the paths are no longer same and there is an optical path difference leading to the destructive interference upon recombination on beam splitter.[17]
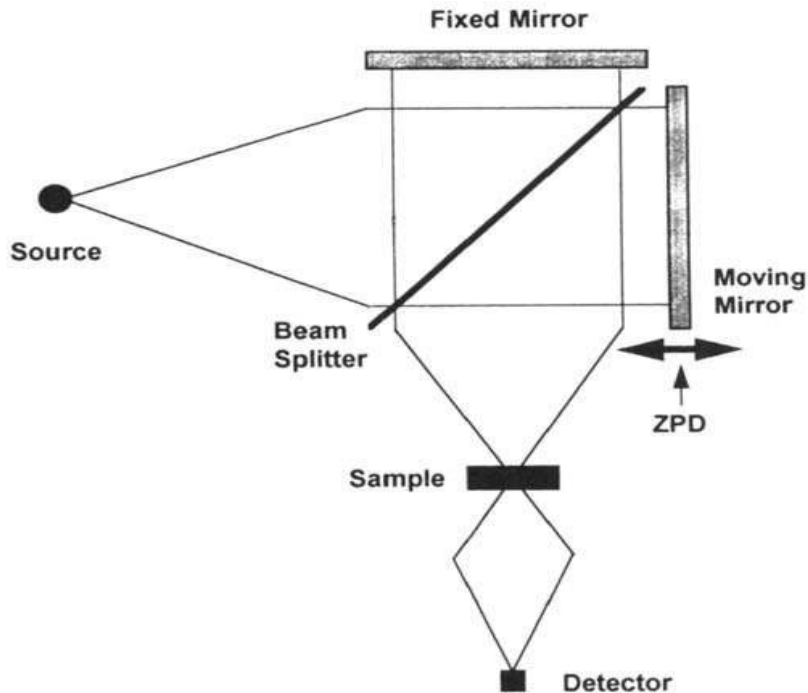
Figure 162: Michelson Interferometer.

The near-infrared transmittance (NIT) spectrometer sends infrared light through a sample and collects data after the sample of the transmitted light. To get the absorbance of the samples a background spectrum is reported for an empty vessel of the type used for the sample. The sample spectra are then calculated with the background spectrum subtracted.
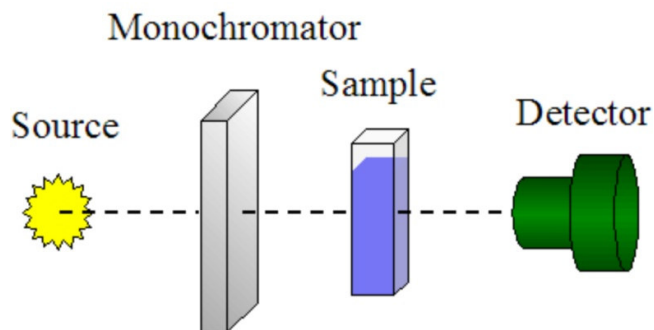
Figure 163: The principle of a transmittance spectrometer

Reflectance

Diffuse reflectance is used primarily for analysing solid samples. Instead of sending light through the sample it is sent to the sample surface from where it is reflected onto the detector which then measures the irradiance. This method was pioneered by Kubelka and Munk around 1920
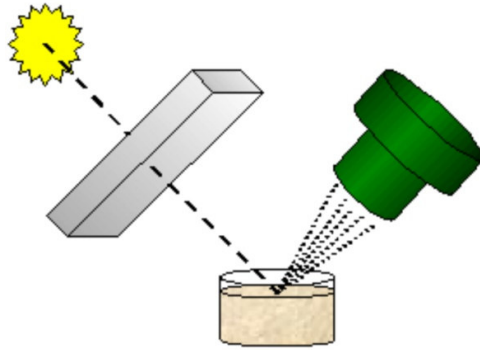
Figure164: Principle of a diffuse reflectance spectrophotometer. The yellow spot is the light source. The grey box is the monochromator. The light brown disc is the sample and the green thing is the detector.

Reflectance is normally measured as a relative reflectance given by the measured intensity, P_S, divided by the reflectance of a reference material, P_R. This reference material in NIR normally is a ceramic disc. This is done to ensure correction for base reflectance. Based on these values the reflectance, R, can be calculated:

$$R = \frac{P_S}{P_R}$$

This value is comparable to the transmittance in transmittance spectrometry, meaning that it is not itself proportional to the concentration. However the following correspondence between concentration and reflectance can be found:

$$\log \frac{1}{R} = \frac{a \cdot c}{s}$$

a is the absorptivity of the analyte and is directly comparable to ε in transmittance spectrometry. Concentration is c and s is the scattering coefficient. Of these values, a is constant for the same analyte and s is the same through the entire spectrum making c only dependent on the reflectance.

As mentioned s is an expression of the scattering primarily due to grain size. A general rule is that the smaller the grains in the sample the larger an irradiance is measured.

The near-infrared reflectance spectrophotometer works by sending a range of wavelengths of light in the near-infrared range into a sample. The light is then reflected on the sample and sent back to the detector is then measured.
The spectra acquired from NIR has the following four major advantages:
    1.   Speed, a spectrum can be acquired in as little as tenth of a second.
    2.   Little or no sample preparation is required.

3. Multiple analysis from a single scan i.e. it is not necessary to scan the sample for each constituent.
4. Non destructive measurement process allowing the analyzed subsample to be returned to the lot.

Fatty acid composition of some common edible fats and oils:

Percent by weight of total fatty acids.

| Oil or Fat | Unsat./Sat. ratio | Saturated | | | | | Mono unsaturated | Poly unsaturated | |
		Capric Acid C10:0	Lauric Acid C12:0	Myristic Acid C14:0	Palmitic Acid C16:0	Stearic Acid C18:0	Oleic Acid C18:1	Linoleic Acid (ω6) C18:2	Alpha Linolenic Acid (ω3) C18:3
Almond Oil	9.7	-	-	-	7	2	69	17	-
Beef Tallow	0.9	-	-	3	24	19	43	3	1
Butterfat (cow)	0.5	3	3	11	27	12	29	2	1
Butterfat (goat)	0.5	7	3	9	25	12	27	3	1
Butterfat (human)	1.0	2	5	8	25	8	35	9	1
Canola Oil	15.7	-	-	-	4	2	62	22	10
Cocoa Butter	0.6	-	-	-	25	38	32	3	-
Cod Liver Oil	2.9	-	-	8	17	-	22	5	-
Coconut Oil	0.1	6	47	18	9	3	6	2	-
Corn Oil (Maize Oil)	6.7	-	-	-	11	2	28	58	1
Cottonseed Oil	2.8	-	-	1	22	3	19	54	1
Flaxseed Oil	9.0	-	-	-	3	7	21	16	53
Grape seed Oil	7.3	-	-	-	8	4	15	73	-
Lard (Pork fat)	1.2	-	-	2	26	14	44	10	-
Olive Oil	4.6	-	-	-	13	3	71	10	1
Palm Oil	1.0	-	-	1	45	4	40	10	-
Palm Olein	1.3	-	-	1	37	4	46	11	-
Palm Kernel Oil	0.2	4	48	16	8	3	15	2	-
Peanut Oil	4.0	-	-	-	11	2	48	32	-
Safflower Oil*	10.1	-	-	-	7	2	13	78	-

Sesame Oil	6.6	-	-	-	9	4	41	45	-
Soybean Oil	5.7	-	-	-	11	4	24	54	7
Sunflower Oil*	7.3	-	-	-	7	5	19	68	1
Walnut Oil	5.3	-	-	-	11	5	28	51	5

* Not high-oleic variety.Percentages may not add to 100% due to rounding and other constituents not listed.Where percentages vary, average values are used.

Table 7: Fatty acid composition of few oils.

Fatty acids consist of the elements carbon (**C**), hydrogen (**H**) and oxygen (**O**) arranged as a carbon chain skeleton with a carboxyl group (**-COOH**) at one end. **Saturated fatty acids** (SFAs) have no double bonds between the carbons. **Monounsaturated fatty acids** (MUFAs) have only one double bond. **Polyunsaturated fatty acids** (PUFAs) have more than one double bond.
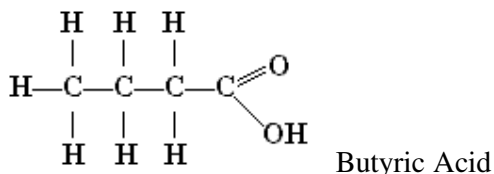
Butyric Acid

Butyric acid (butanoic acid) is one of the saturated short-chain fatty acids responsible for the characteristic flavor of butter. This image is a detailed structural formula explicitly showing four bonds for every carbon atom and can also be represented as the equivalent line formulas:

$CH_3CH_2CH_2COOH$ or $CH_3(CH_2)_2COOH$

The numbers at the beginning of the scientific names indicate the locations of the double bonds,for example "9,12-octadecadienoic acid" indicates that there is an 18-carbon chain (octa deca) with two double bonds (di en) located at carbons 9 and 12, with carbon 1 constituting a carboxyl group (oic acid). The structural formula corresponds to:

$CH_3CH_2CH_2CH_2CH_2CH=CHCH_2CH=CHCH_2CH_2CH_2CH_2CH_2CH_2COOH$

9,12-octadecadienoic acid   (Linoleic Acid)
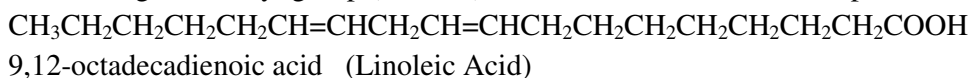
Fatty acids indicated as **C18:2** represents that the fatty acid consists of an 18-carbon chain and 2 double bonds. Although this could refer to any of several possible fatty acid isomers with this chemical composition, it implies the naturally-occurring fatty acid with these characteristics, i.e., linoleic acid. Double bonds are said to be "conjugated" when they are separated from each other by one single bond, e.g., (**-CH=CH-CH=CH-**). The term "conjugated linoleic acid" (CLA) refers to several C18:2 linoleic acid variants such as 9,11-CLA and 10,12-CLA which correspond to 9,11-octadecadienoic acid and 10,12-octadecadienoic acid. The principal dietary isomer of CLA is *cis*-9,*trans*-11 CLA, also known as rumenic acid. CLA is found naturally in meats, eggs, cheese, milk and yogurt.

$CH_3(CH_2)_5CH=CH-CH=CH(CH_2)_7COOH$

9,11-Conjugated Linoleic Acid
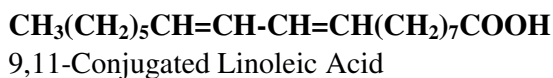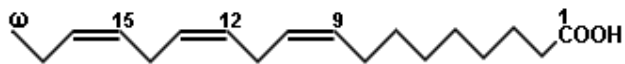
Alpha-Linolenic Acid (omega-3)
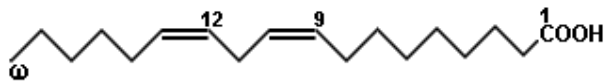
Linoleic Acid (omega-6)

Common Fatty Acids

Chemical Names and Descriptions of some Common Fatty Acids				
Common Name	Carbon Atoms	Double Bonds	Scientific Name	Sources
Butyric acid	4	0	butanoic acid	butterfat
Caproic Acid	6	0	hexanoic acid	butterfat
Caprylic Acid	8	0	octanoic acid	coconut oil
Capric Acid	10	0	decanoic acid	coconut oil
Lauric Acid	12	0	dodecanoic acid	coconut oil
Myristic Acid	14	0	tetradecanoic acid	palm kernel oil
Palmitic Acid	16	0	hexadecanoic acid	palm oil
Palmitoleic Acid	16	1	9-hexadecenoic acid	animal fats
Stearic Acid	18	0	octadecanoic acid	animal fats
Oleic Acid	18	1	9-octadecenoic acid	olive oil
Ricinoleic acid	18	1	12-hydroxy-9-octadecenoic acid	castor oil
Vaccenic Acid	18	1	11-octadecenoic acid	butterfat
Linoleic Acid	18	2	9,12-octadecadienoic acid	grape seed oil
Alpha-Linolenic Acid (ALA)	18	3	9,12,15-octadecatrienoic acid	flaxseed (linseed) oil
Gamma-Linolenic Acid (GLA)	18	3	6,9,12-octadecatrienoic acid	borage oil
Arachidic Acid	20	0	eicosanoic acid	peanut oil, fish oil
Gadoleic Acid	20	1	9-eicosenoic acid	fish oil
Arachidonic Acid (AA)	20	4	5,8,11,14-eicosatetraenoic acid	liver fats
EPA	20	5	5,8,11,14,17-eicosapentaenoic acid	fish oil

Behenic acid	22	0	docosanoic acid	rapeseed oil
Erucic acid	22	1	13-docosenoic acid	rapeseed oil
DHA	22	6	4,7,10,13,16,19-docosahexaenoic acid	fish oil
Lignoceric acid	24	0	tetracosanoic acid	small amounts in most fats

Table 8: Fatty acids.

7.7. List of sample names, the numbers after the sample names in brackets represent the sample numbers in NIR plots while the unscrambler order numbers ex: 1a are the names for plots in IR unscrambler.

Table 9: List of oil samples.

Samples list order	Unscrambler order	Samples list order	Unscrambler order
Lab Horfro Oil (1)	1a	Lab Coconut Oil (31)	31a
	1b		31b
Riz Coop Sesame oil (2)	2a	Foss Oleina de palma (32)	32a
	2b		32b
Riz Maeva Olive oil(3)	3a	Foss Soya Deodorizado (33)	33a
	3b		33b
Riz Florens Olive oil (4)	4a	Foss Soya Cruda (34)	34a
	4b		34b
Riz Toscano olive oil (5)	5a	Foss Maize Deodorizado (35)	35a
	5b		35b
Riz Fondue B (6)	6a	Foss sunflower oil 04.23.01 1/11(36)	36a
	6b		36b
Riz Avocado Oil (7)	7a	Foss Sunflower 05.31.01 1/19 (37)	37a
	7b		37b
Riz Citron Rapskim oil (8)	8a	Foss SF000605 HHZ990 (38)	38a
	8b		38b
Riz Arte olive oil (9)	9a	Foss RPKAB 583 10.11.00 (39)	39a
	9b		39b
Riz Rocchia Olive oil	10a	Foss SF000602	40a

(10)		NXJ685 (40)	
	10b		40b
Lab Thisle oil (11)	11a	Foss sunflower 05.31.01 1/17 (41)	41a
	11b		41b
Riz Tosca olive oil (12)	12a	Foss Sunflower 04.23.01 1/10(42)	42a
	12b		42b
Lab Sunflower oil (13)	13a	Foss Sunflower 04.23.01 1/14(43)	43a
	13b		43b
Riz Vegetal olive oil (14)	14a	Foss RP Kiel 583 4/11.00 (44)	44a
	14b		44b
Riz FondueA (15)	15a	Foss RP Kiel 583 24.11.00 (45)	45a
	15b		45b
Riz Walnut oil (16)	16a	Foss SV LEER 000728 (46)	46a
	16b		46b
Riz Basilkum Rapskim (17)	17a	Foss RP Alborg 585 19.11.00 (47)	47a
	17b		47b
RiZ Coop Peanut oil (18)	18a	Foss Palmiste cruda 09.05.01(48)	48a
	18b		48b
LAB MANDEL(Almond) OIL(19)	19a	Foss Palma cruda 09.05.01 (49)	49a
	19b		49b
RIZ COOP Thisle OIL (20)	20a	Foss Rape oil 05.17.01 2/22 (50)	50a
	20b		50b
Lab Rape oil (21)	21a	Foss Rape oil 09.13.01 3. (51)	51a
	21b		51b
Riz Grape oil (22)	22a	Foss Rape oil 09.13.01 4.(52)	52a
	22b		52b
Riz Rapskim (23)	23a	Riz Palm oil (53)	53a
	23b		53b
Riz sunflower oil (24)	24a		
	24b		

Riz Zara olive oil (25)	25a		
	25b		
Lab wheat germ oil (26)	26a		
	26b		
Riz coop Maize oil (27)	27a		
	27b		
Riz Sesame oil (28)	28a		
	28b		
Foss Palm Olein days 3 (29)	29a		
	29b		
Foss Palm Olein fresh (30)	30a		
	30b		

7.8 GC/MS library search:

An example of the library search of mass spectrum is shown in plots below:

Chromatogram Plot

File: c:\saturnws\data\rizwana\30-01-08\coop sesamea.sms
Sample: coop SesameA
Scan Range: 1 - 868 Time Range: 0.00 - 13.97 min.

Operator: Dorte
Date: 1/30/08 2:26 PM
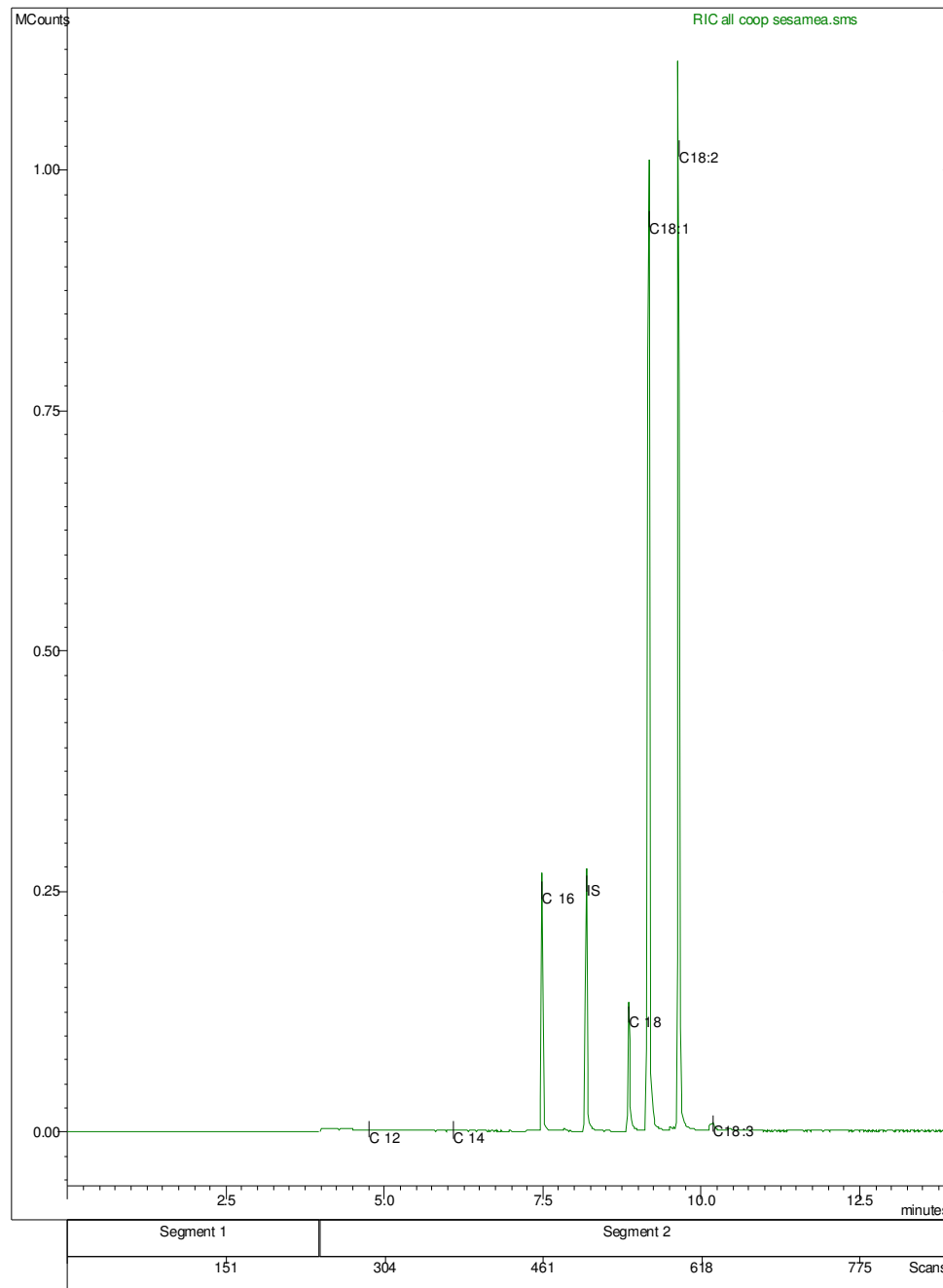
Figure 165: Coop Sesame oil chromatogram.

The Riz coop sesame oil chromatogram above shows peaks of different fatty acids at different retention times. The peak at retention time 9.636 min is searched for library matches. The spectrum matches with 9, 12 octadecadenoic acid Z, Z methyl ester with the following parameter values.

The above library search result show good fit with average purity. Purity decreases due to the presence of some impurities in the sample. Fit value is 740, Reverse fit is 506 and purity is 402.

Purity of the spectrum indicates the number of peaks which are same in both the spectra and how well their intensities match. A score of above 800 indicates very similar spectra.

Fit indicates how well the library spectrum peaks are found in the sample. The fit score will decrease when the relative intensities differ from the library spectrum.

 A high fit low purity indicates that the matched compound is there but there is also another compound present in the sample. A low fit score means that there are peaks in the library spectrum which didnot appear in the sample and this implies that the compound is not present.

Reverse fit measures the degree to which the peaks found in sample are also found in the library spectrum[19]

The document below shows the mass spectrum library search results:

Print Date: 19 Feb 2008 16:00:30

Hit 53 of Saturn Fit Search

Saturn Fit Search Results
Hits Found: 110
Pre-Search Hits Found: 1761

Saturn Fit Search Parameters
Threshold: 700
Target Ion Range: 35 - 650
Library MW Range: 0 - 2000
Library Ion Range: 35 - 650 (Target Mass Range)
Local Normalization: Off
Requested Pre-Search: 1000
Requested Final Search: 200
Search 3 Libraries: A. c:\saturnws\satlib\nist98.lbr
                                  B. c:\saturnws\satlib\nist98m.lbr
                                  C. c:\saturnws\satlib\nist98r.lbr

Hit 53 Fit: 740 RFit: 506 Purity: 402
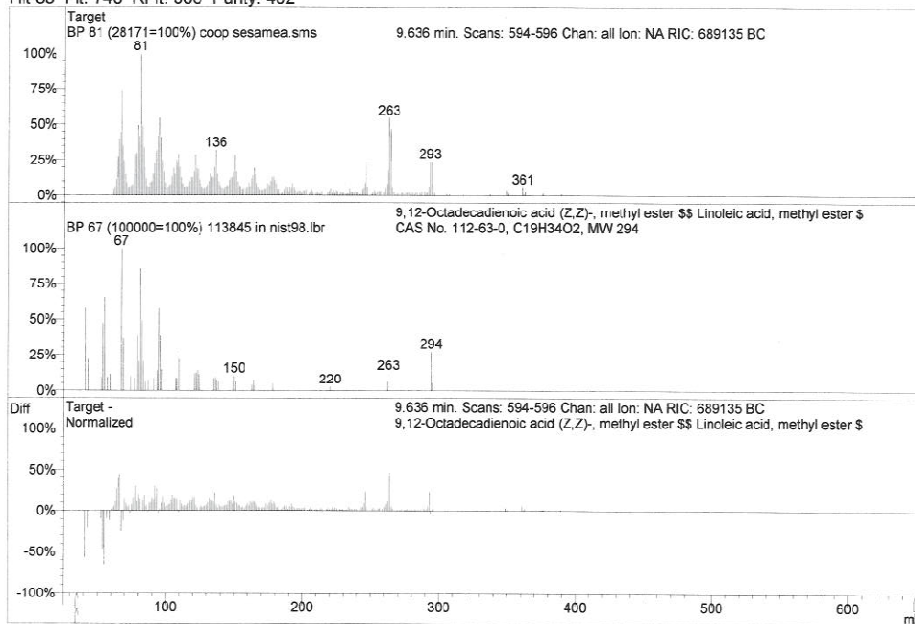
Target Spectrum from c:\... \data\rizwana\30-01-08\coop sesamea.sms
Scan No: 595, Time: 9.636 minutes
3 points averaged. Background corrected.
Comment: 9.636 min. Scans: 594-596 Chan: all Ion: NA RIC: 689135 BC
Pair Count: 245   MW: 0   Formula: None   CAS No: None Acquired Range: 60 - 650

Match Spectrum 113845 from nist98.lbr Library
Name: 9,12-Octadecadienoic acid (Z,Z)-, methyl ester $$ Linoleic
    methyl ester $$ Methyl cis,cis-9,12-octadecadienoate $$ Methyl l
    inoleate $$ Methyl octadecadienoate $$ Methyl 9-cis,12-cis-octad
    ecadienoate $$
Pair Count: 50   MW: 294   Formula: C19H34O2   CAS No: 112-63-0 Acquired Range: 41 - 295

116

Column Part No.	CP8824
Column Serial No.	5201438
Stationary Phase	VF-23ms
Film Thickness (µm)	0.25

L (m) x ID (mm) x OD (mm): 60 x 0.25 x 0.39

Approved By:

QUALITY CONTROL CERTIFICATE OF ANALYSIS

Maximum Allowable Operating Temperature Isothermal : 260 °C
Maximum Allowable Operating Temperature Programmed : 260 °C

Test mix : CP0048
Sample volume : 2.00 µl
Injector temperature : 270 °C
Detector temperature : 300 °C
Column temperature : 120.0 °C
Detector : FID
Carrier gas : Hydrogen
Pressure * : 140.0 kPa

* 100 kPa = 14.5 psi = 1 Bar

Column Performance:

	Rt [Min]	Component
	2.63	methane
	2.69	Solvent
	4.45	5-nonanone
	5.42	1-octanol
	7.06	n-heptadecane
	9.52	n-octadecane
	11.43	naphthalene
	13.32	n-nonadecane
	17.55	2,6-dimethylaniline
Total		

Specifications of the column:

Retention factor (k) 2,6-dimethylaniline	5.7
Number of theoretical plates 2,6-dimethylaniline	1692
Number of theoretical plates per/m 2,6-dimethylaniline	2820
Film thickness [µm]	0.27
Phase ratio	233.
Bleed, spc controlled at 260 °C [pA]	1.66
Linear gas velocity [cm/s]	38.0

VARIAN

This report is generated with Varian Galaxie software.
For more information about Galaxie Chromatography Data Systems,
visit our web site at www.varianinc.com

7.8. Document of Column Description.

Capillary column of dimensions 60 m in length, 0.25 inner diameter and 0.39 outer diameter with its quality control applications is shown in the above document. This column is used for the GC/MS FAME analysis.

7.9 Material Present in CD

1. Report PDF file.
2. GC/MS files – contains chromatograms, Calibration curves, Method.
3. NIR (Grams), IR (ASCII) files.
4. Unscrambler files- NIR, IR 100-60µm and 110-10 µm.
5. NIR laboratory report.
6. Reference Analysis – excel sheet.