



# Online monitoring of an organic synthesis reaction using Raman and NIR spectroscopy

By  
**Fasil Ayelegn Tassew**

A thesis submitted in partial fulfillment of the requirements for the  
Degree of European Master in Quality in  
Analytical Laboratories

Department of Chemistry  
University of Bergen  
March 2011

## Acknowledgment

First and foremost I would like to express my heartfelt appreciation to my supervisor Prof. Bjørn Grung for his support and valuable comments throughout this project. I would also like to extend my gratitude to Prof. Hans-Rene´ Bjørsvik, Dr. Egil Nodland, Raffaele Spaccini and Anna Tsoukala who have been supportive all along.

I am grateful to the financial support provided by the European commission to complete my Masters degree. I want to thank all the people involved in the European master in quality in analytical laboratories (EMQAL) particularly Prof. Isabel Cavaco for her continuous support from the beginning to the end of this masters program.

Finally, I have to acknowledge the support, advice and encouragement of my parents, sisters, brothers and friends. Thank you!

## Abstract

The synthesis of 10-Bromo-1-Decandiol was monitored using Raman and near infrared spectroscopy. Optical fiber Raman and near infrared probes were immersed in the reaction mixture. Raman spectra were recorded every 50 second whereas near infrared spectra were recorded every 30 second. The reaction was carried out for 150 minutes. Data obtained from the spectroscopic techniques went through statistical pretreatment methods. The pretreated data then subjected to multivariate data analysis techniques. Principal component analysis, multivariate curve resolution and partial least square regression were used to identify major trends in the reaction, to identify and monitor the concentration and spectral profiles of the components involved in the reaction and to develop a regression model for the reaction respectively. For the purpose comparison and model development, GC-MS analysis of the reaction mixture was done in 15 minutes interval during the course of the reaction.

Principal component analysis on Raman and near infrared data indicated that there were three main components in each data one component in decreasing trend and the other two in increasing trend. This was also confirmed by the GC-MS analysis as there were only three components in the reaction mixture. The decreasing component was assigned to 1,10-decanediol(reactant) while the other two components are assigned to 10-Bromo-1-Decandiol (Product) and 1,10-dibromodecane(byproduct).From concentration profiles obtained from multivariate curve resolution it was also shown that the three components in the reaction mixture followed the same trend as suggested by the principal component analysis and GC-MS analysis. Partial least square regression model was built for both Raman and near infrared data using ten variables from the GC-MS analysis and ten variables from Raman and near infrared. A good prediction was obtained for both data when the model was tested under a different set of variables. As a result, it was concluded that monitoring the synthesis of 10-Bromo-1-Decandiol using Raman and near infrared spectroscopy proved to be feasible.

## Abbreviations

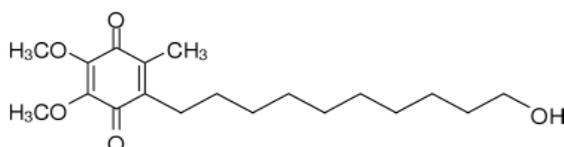
CCD	Charge coupled device
DT	De-trending
EMR	Electromagnetic radiation
GC-MS	Gas Chromatography Mass spectrometry
IR	Infrared
MCR-ALS	Multivariate curve resolution – alternate least squares
NIR	Near Infrared
PC's	Principal Components
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Square
SNV	Standard Normal Vitiate
TBAB	Tetrabutylammonium bromide
UV	Ultra violate

## Table of contents

<b>Acknowledgment</b> .....	2
<b>Abstract</b> .....	3
<b>Abbreviations</b> .....	4
<b>Table of contents</b> .....	5
<b>1. Introduction</b> .....	6
<b>2. Theory</b> .....	7
<b>2.1. Raman spectroscopy</b> .....	7
<b>2.2. NIR spectroscopy</b> .....	10
<b>2.3. Multivariate data analysis techniques</b> .....	12
2.3.1. Principal component analysis.....	12
2.3.2. Multivariate curve resolution-alternate least squares.....	16
2.3.3. Partial least square regression.....	18
2.3.4. Standard normal variate.....	20
<b>3. Experimental</b> .....	20
<b>3.1. Materials and methods</b> .....	20
<b>3.2. Experimental Procedures</b> .....	22
3.2.1. Raman and NIR measurement.....	22
3.2.2. GC-MS Analysis.....	22
<b>4. Result and discussion</b> .....	23
<b>4.1. Data description</b> .....	23
<b>4.2. Data pretreatment</b> .....	24
4.2.1. Raman data.....	24
4.2.2. NIR data.....	29
<b>4.3. PCA analysis</b> .....	30
<b>4.4. MCR–ALS analysis</b> .....	37
<b>4.5. GC-MS analysis</b> .....	40
<b>4.6. PLS models</b> .....	41
<b>5. Conclusion</b> .....	50
<b>6. Suggestion for future work</b> .....	50
<b>7. References</b> .....	52

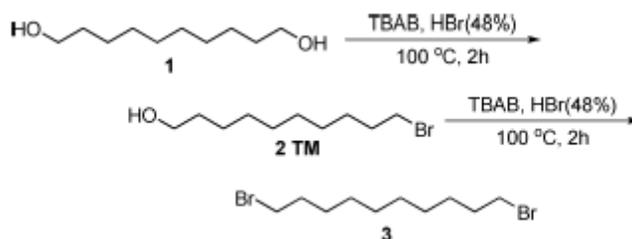
## 1. Introduction

This paper is part of a project which aims at finding an easy approach for the production of idebenone. Idebenone has several pharmaceutical applications including the treatment of Friedrich's ataxia [1], in skincare and anti-ageing treatments [2]. It is also under research for a possible application in the treatment of Alzheimer's disease [3].



*Fig.1 Idebenone*

10-Bromo-1-Decandiol is one of the derivatives for the synthesis of idebenone. An easy and cheap way to synthesize 10-Bromo-1-Decandiol is the ultimate goal of the project. This paper specifically focuses on monitoring of the reaction that leads to the synthesis of 10-Bromo-1-Decandiol.



*Fig 2-Schematic process for the production of 10-Bromo-1-Decandiol. 1: 1,10-decanediol, 2: 10-Bromo-1-Decandiol(Product), 3:1,10-dibromodecane(byproduct) and TBAB: Tetrabutylammonium bromide [4]*

The progress of the reaction can be monitored online using Raman and NIR spectroscopy. Chemometrics techniques such as PCA, PLS regression and MCR-ALS will be applied on the resulting data using Matlab 7.5.5 (R2007b) (MathWorks, Natick, MA), Sirius 8.0 (PRS, Bergen, Norway), Unscrambler X version 10.1 (CAMO software AS, Oslo, Norway) and Excel ver. 7.0 (Microsoft, Redmond, WA, USA). The result obtained from the analysis of data from Raman and NIR spectroscopy will be compared to a result obtained from GC-MS analysis.

Specific objectives in this thesis are to study the concentration profile of the reactant, product and byproduct of the reaction, to compare results obtained from Raman and NIR data with the result obtained from GC-MS analysis and to assess the feasibility of coupling Raman and NIR systems for the monitoring of the reaction in organic chemical syntheses of 10-Bromo-1-Decandiol.

## 2. Theory

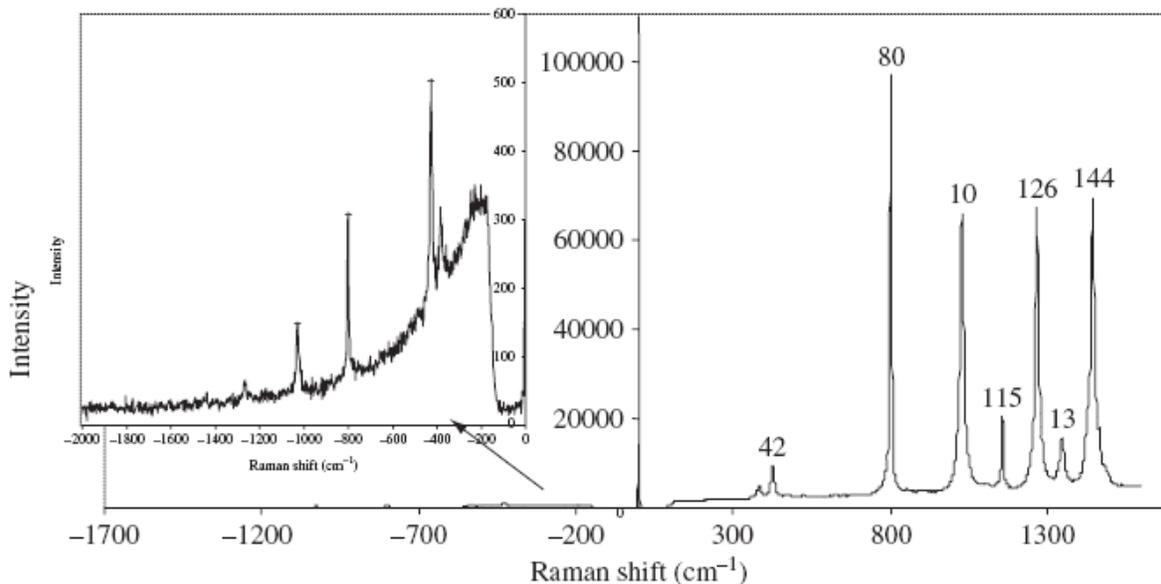
### 2.1. Raman spectroscopy

When an incident radiation of certain energy interacts with matter, there are two possibilities, either the radiation may not have any effect on the matter and passes through it or it causes various effects on the matter. One such effect is that the photons that make up the incident radiation may be absorbed by molecules of the matter (Absorption /emission spectroscopy). This happens when energy of the incident photon matches with the energy needed to excite molecules from ground state to excited state [5].

On the other hand the incident radiation may interact with the matter and scatter from it. If the energy of the scattered radiation remains the same as the incident radiation it is called elastic scattering (Rayleigh scattering). If the energy of the scattered radiation is weaker or stronger than the incident radiation, it is called inelastic scattering. It is this inelastic scattering which has come to be known as Raman scattering. Raman scattering occurs due to vibrational energy transfer between the radiation and the molecule. Rayleigh scattering is the dominant type of scattering with millions of degree of magnitude stronger than the Raman scattering [6]. Raman scattering

results the scattered radiation to be either higher (anti-Stokes line) or lower (Stokes line) in energy than the incident radiation.

Raman scattering is characterized by a shift from the frequency of the incident radiation. If  $\nu_1$  denotes wavenumber of a peak for a Raman band, then the wavenumber shift from frequency of the incident radiation  $\nu$  can be defined as:  $\Delta\nu = \nu - \nu_1$ . For Stokes-Raman scattering  $\nu_1 = \nu - \nu_M$  ( $\nu_M$  is characteristic frequency of the band) thus,  $\Delta\nu$  is positive but for anti-Stokes Raman scattering it is negative because  $\nu_1 = \nu + \nu_M$  [7]. Raman spectrum for cyclohexane in figure 3 shows the difference in Raman shift and intensity between Stokes and anti-Stokes scattering.



*Fig.3. Stokes and anti-Stokes scattering for cyclohexane. To show the weak anti-Stokes spectrum the y-axis has been extended in the inset [5]*

Anti-Stokes effect transfers energy from the system to the incident radiation. This means transition is carried out from a higher energy level to ground level. Such transitions are typically rare because according to Maxwell-Boltzmann distribution law, higher energy levels are less populated than the ground level. As a result, the anti-Stokes intensity is less than the Stokes intensity. Therefore, Stokes-Raman spectrum is detected in most cases [8, 9].

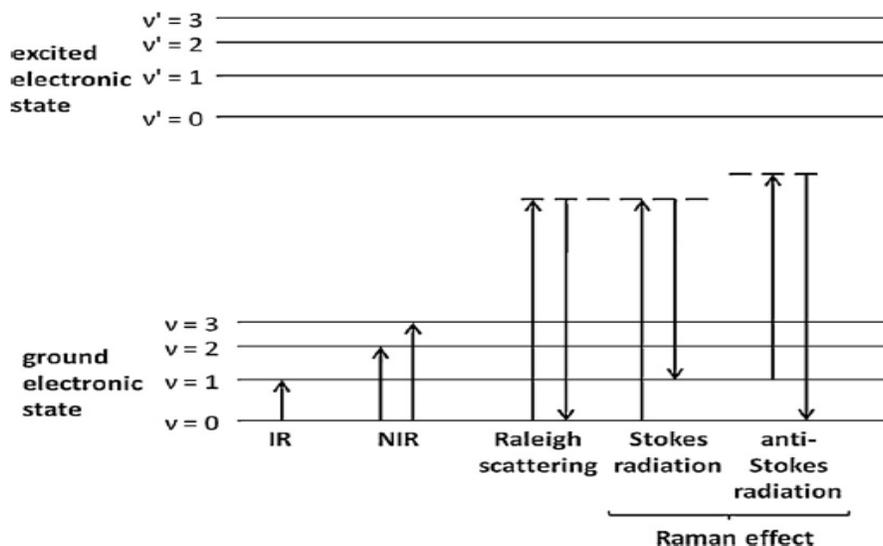


Fig.4. IR & NIR absorption and the Raman effect [10]

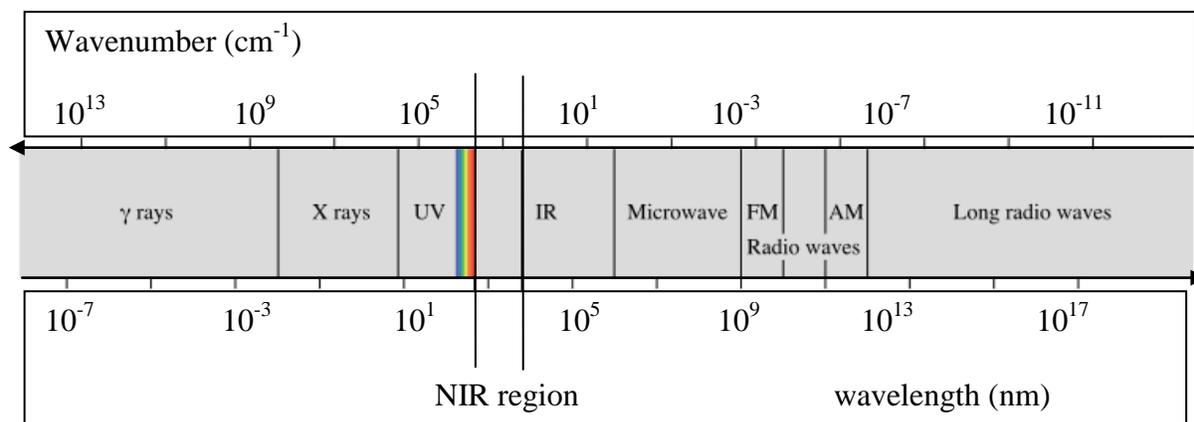
Raman spectroscopy (along with NIR spectroscopy) is becoming more and more crucial technique for process monitoring particularly in the pharmaceutical industry. This is because Raman spectroscopy allows rapid, real-time, non-destructive measurements without sample preparation. It also allows measurements to be carried out from relatively long distance by using probes attached to the instrument through optical fibers [10, 11].

For a molecular vibration to be Raman active there must be change in the polarizability in the molecule and this polarizability is intense in symmetric vibrations. Hence, unlike IR absorption where asymmetric vibrations result in larger change in dipole moment, intense Raman scattering is achieved in symmetric vibrations [5]. Fluorescence is a major source of interference in Raman spectroscopy. Even in a very small quantity, contaminants can be excited to emit fluorescence which is much stronger than the Raman intensity [12].

The chemical reaction investigated in this paper mainly involves three compounds that are Raman/NIR active. During the course of the reaction, the main structural molecular change is the conversion of  $-\text{CH}_2\text{-OH}$  into  $-\text{CH}_2\text{-Br}$ . As a result, the appearance and disappearance of peaks or change in intensities of the peaks associated with the above structural molecular changes will be utilized for reaction monitoring. C-O, O-H and C-Br bond vibrations will be mainly used for this purpose.

## 2.2. NIR Spectroscopy

NIR spectroscopy measures absorption of NIR radiation by organic molecules. All absorption bands are the result of overtones or combination of overtones from the mid IR region [13]. NIR region is found in between 800 nm and 2500 nm in the electromagnetic spectrum (figure 5).



*Fig.5. Location of NIR in the electromagnetic spectrum*

Interaction between electromagnetic radiations and molecules can be explained using the classical model and the quantum mechanical model. The classical model assumes interaction between EMR and the molecule occurs when the energy of EMR matches the energy associated with change in an electrical characteristic of the molecule. The Quantum model on the other hand says that interaction can happen when the energy difference between two energy levels of a molecule matches the energy of a photon of an EMR. This means various types of changes in electrical characteristic of a molecule can be a source of interaction with EMR (for example visible light interacts with the electron cloud surrounding the molecule whereas UV radiation interacts with electron cloud around an atom).

According to the classical model for molecular vibrations, vibrational modes corresponding to the transitions that are active in the near infrared are forbidden. However, in quantum mechanical model these vibrational modes are not forbidden. Moreover, unlike the classical model, the

energy levels that a molecule can attain are not equally spaced (figure 6). Instead the gap in energy levels decrease with increasing energy so that the gaps at higher energy levels are virtually nonexistent [14, 15]. As a result multilevel vibrational energy transitions are possible. Since the energy of such multilevel transitions corresponds to the photons in the NIR region, they result in NIR absorption bands. NIR absorption bands are often called ‘overtones’ and for each fundamental molecular vibration several overtones appear with decreasing intensity. Combination bands occur when absorbed photon excites two or more vibrations simultaneously. For this to happen, sum of energies of the coupling vibrations should be equal with the energy of the exciting photon [16].

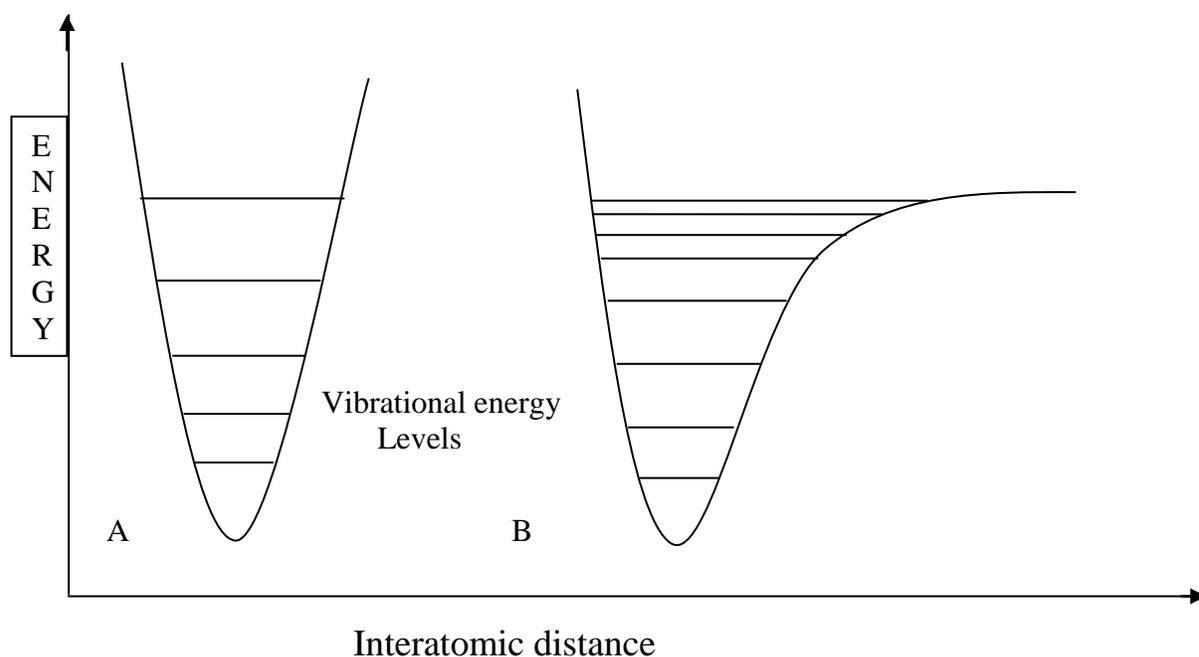


Fig.6. Energy diagram: A. classical model (harmonic oscillator) B. quantum model (anharmonic oscillator)

The most common absorption bands occurring in the NIR region are overtones and combination overtones of CH, NH and OH functional groups. Even if NIR ranges from 800 nm to 2500 nm,

practically most measurements are carried out between 1100 nm and 2500 nm because majority of first and second overtones appear in this region.

NIR bands are considerably weaker than their mid-IR counterparts. In addition, they are broad and susceptible to band overlap. Due to this, it is often difficult to assign a specific band to chemical components. This leads to a complex but information rich NIR spectra that requires multivariate data analysis techniques for interpretation.

Bond	Raman	NIR
O-H	Weak intensity 3650-3000cm <sup>-1</sup>	1900 nm stretching/deformation combination
C-Br	Strong intensity 700-500 cm <sup>-1</sup>	Non-specific
C-O	1030 cm <sup>-1</sup>	Non-specific

*Table 1 Bonds of interest and their absorption bands [17]*

## 2.3. Multivariate data analysis techniques

### 2.3.1. Principal component analysis

PCA analysis is a common data analysis tool used to identify patterns, compress huge data and provide evidence to explain relationships, between variables and samples in a dataset. It mainly works by avoiding variables that least explains major trends in the data without losing much information and by finding variables and combination of variables that describe the dataset well.

PCA extracts principal components which are far fewer in number than the original variables [11] (usually 2-3). If  $X$  is a data matrix with  $M$  rows and  $N$  columns, PCA decomposes  $X$  as the sum of  $r$   $t_i$  and  $p_i$ , where  $r$  is the rank of the matrix  $X$ :

$$X = t_1p_1^T + t_2p_2^T + \dots + t_kp_k^T + \dots + t_rp_r^T$$

The rank of the matrix should be less than or equal to the smaller dimension of the data matrix. The  $t_i$  vectors are called scores and contain information about relationship among samples. The  $p_i$  vectors are called loadings and contain information about relationship within variables. Scores have equal number of rows with the original data matrix whereas loadings have equal number of columns with the original data matrix [18, 19].

Generally, few numbers of principal components are chosen to represent the original data. If  $k$  is the number of principal components chosen, then the score matrix will have  $M$  rows and  $k$  columns and the loadings matrix will have  $k$  rows and  $N$  columns. After the number of principal components to use is decided the remaining little variance in the data is merged into a residual matrix;  $E$ . then the PCA model of data matrix can be represented as follows (figure 7 shows graphical representation of PCA decomposition):

$$X = t_1p_1^T + t_2p_2^T + \dots + t_kp_k^T + E$$

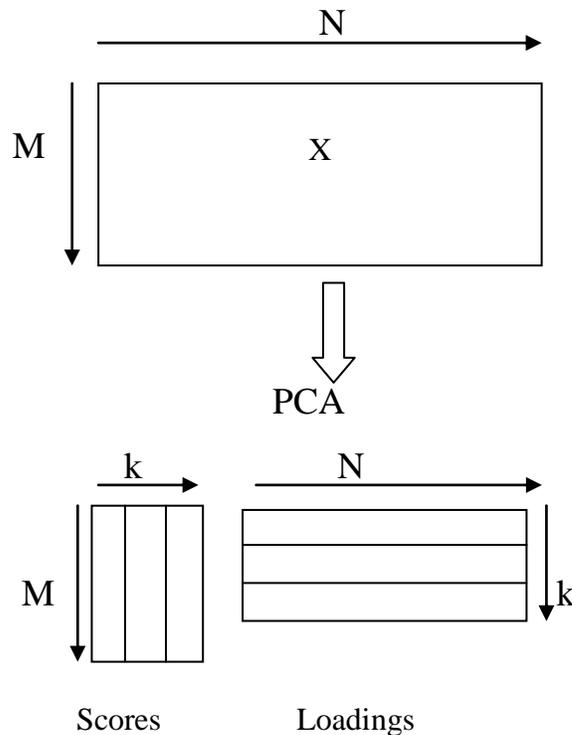


Fig.7. Graphical representation of PCA

In order to understand how PCA works mathematically it is good to start from covariance. For a data with a single variable, standard deviation and variance can describe the spread of the data. For a data with two variables covariance provides a similar measure to find out how the variables vary from the mean with respect to each other. For example covariance between variables A and B calculated as follows:

$$\text{cov}(A,B) = \frac{\sum_{i=1}^n (A_i - \bar{A})(B_i - \bar{B})}{n-1}$$

Where  $\bar{A}$  is mean of A values,  $\bar{B}$  is mean of B values and n is the number of samples in each variable. The sign of result of result of the covariance tells whether there is relationship between the variables or not. If the result is positive it means the variables increase or decrease together. But if the covariance value is negative it means there is opposite relationship. Zero covariance value means the variables are independent to each other.

When a data consists of a large number of variables, covariance values are calculated between each of the variables. This leads to a covariance matrix  $C_m$ . the covariance matrix is a square matrix. If a data set X consists of N variables the covariance matrix is given as:

$$C = \begin{pmatrix} \text{cov}(1,1) & \text{cov}(1,2) & \dots & \text{cov}(1,N) \\ \text{cov}(2,1) & \text{cov}(2,2) & \dots & \text{cov}(2,N) \\ \vdots & \vdots & & \vdots \\ \text{cov}(N,1) & \text{cov}(N,2) & \dots & \text{cov}(N,N) \end{pmatrix}$$

The covariance matrix contains  $\text{cov}(1,1)$ ,  $\text{cov}(2,2)$  up until  $\text{cov}(N,N)$  along the diagonal of the matrix. In such cases the covariance values are equal to the variance values. Moreover, since the covariance values for  $\text{cov}(A,B)$  and  $\text{cov}(B,A)$  are equal, the covariance matrix is symmetrical.

An important property of covariance matrices is that, since they are square matrices eigenvectors can be calculated. Eigenvectors are non-zero vectors which remain the same (or multiplied by a factor) after multiplied by a square matrix. It is these eigenvectors that have come to be known as loading vectors ( $p_i$ ). Eigenvectors of a matrix are perpendicular to each other. When an eigenvector is multiplied by the covariance matrix, the result will be the eigenvector itself multiplied by a factor. This factor is called eigenvalue ( $\lambda$ ). This can be denoted as follows:

$$C_m p_i = \lambda_i p_i$$

Where,  $C_m$  is the covariance matrix,  $\lambda_i$  is the eigenvalue associated with  $p_i$  [20]. The score vector  $t_i$  is a linear combination of the original variables in the data ( $X$ ) defined by  $p_i$ . Hence,  $t_i$  can be described mathematically as:

$$t_i = X p_i$$

Based on the eigenvalue associated with  $p_i$ , the  $t_i$ ,  $p_i$  pairs are arranged in decreasing order. The first  $t_i$ ,  $p_i$  pair holds the largest variance then the second pair holds the second largest variance and so on. PCA extracts components that hold much of the information about the data matrix and in doing so it relieves researchers from a great deal of burden of dealing with huge number of variables. The result is presented in a form of scores and loading vectors. Scores describe relationship among objects while loadings describe relationship among variables.

In this project the data obtained from NIR and Raman spectrometers were subjected to PCA analysis (wavenumbers as variables and number of spectra recorded as objects). Hence, information about relationship among spectra at different times; which can explain the yield and conversion of the reaction can be obtained from the score plots whereas the loading plot gives clues about the relationship among different wavelengths.

### 2.3.2. Multivariate curve resolution-alternate least squares

Monitoring organic chemical reactions can be a difficult task as such reactions usually involve a number of compounds whose individual contribution to the instrument response is complex to identify and interpret. This problem is hugely reduced by using a chemometrics technique called multivariate curve resolution. MCR is a very useful technique for spectral data obtained from chemical reactions. This is particularly manifested when monitoring chemical reactions [21] as it resolves the data into spectral and concentration profiles of pure components involved in the reaction.

Multivariate curve resolution decomposes a two-way data matrix into the product of two smaller matrices which are each related to one of the two orders of the data matrix (e.g. spectral or concentration) [22]. Multivariate curve resolution methods can be classified as iterative and non-iterative methods [23].

Multivariate curve resolution –alternate least square (MCR-ALS) is an iterative method which is based on optimization of initial estimates by using data structure and chemical constraints [24]. For example both concentration and spectra are necessarily non-negative. If there is sufficient selectivity in the data non-negativity is a sufficient constraint to extract the pure component spectra through alternate least square procedure [18]. However, analysis of single data matrix using curve resolution decomposition does not always guarantee that the true solutions are obtained. This is particularly the case when there is no selectivity in the data.

Curve resolution methods operate based on the assumption that the experimental data follows a linear model. For a data matrix  $X$ , with  $M$  number of rows and  $N$  number of columns, the model can be written as follows:

$$X=CS^T+E$$

Where,  $C$  is an  $M$  by  $n$  matrix called concentration profiles of the  $n$  components involved in the reaction mixture whereas  $S^T$  is an  $n$  by  $N$  matrix called spectral profiles of the  $n$  pure components

in the mixture and E describes the residual matrix [25]. This equation can be solved iteratively using alternate least square optimization. New estimates of C and S are obtained after each iteration of optimization.

Usually, after principal component analysis a preselected number of principal components will be truncated and used instead of the original data matrix, X in order to get a better result as the new matrix,  $X_{est}$  is relatively noise free. Then, an initial estimate of C will be available. As a result:

$$S^T = C^{+T} X_{est}$$

$$C = X_{est} S^+$$

Where,  $C^+ = C(C^T C)^{-1}$  and  $S^+ = S(S^T S)^{-1}$

However, ambiguity about the exact magnitude of the pure component spectra and concentration profiles remain. This ambiguity is called intensity ambiguity [18, 22]. The ambiguity can be resolved for a particular component if concentration of the component is known at some point within the data. Another type of ambiguity is called rotational ambiguity and it also affects the accuracy of the solutions for C and S. intensity and rotational ambiguities can be represented in the following equations respectively:

$$X_{est} = CS^T = CDD^{-1} S^T$$

$$X_{est} = CS^T = CRR^{-1} S^T$$

Where, D and R are arbitrary invertible diagonal matrices. Presence of the above ambiguities in the solution for C and S means the results from multivariate curve resolution methods have limitations and should not be overly interpreted. However, multivariate curve resolution methods can be quite effective in qualitative analysis. The method used by MCR-ALS is summarized in the following diagram.

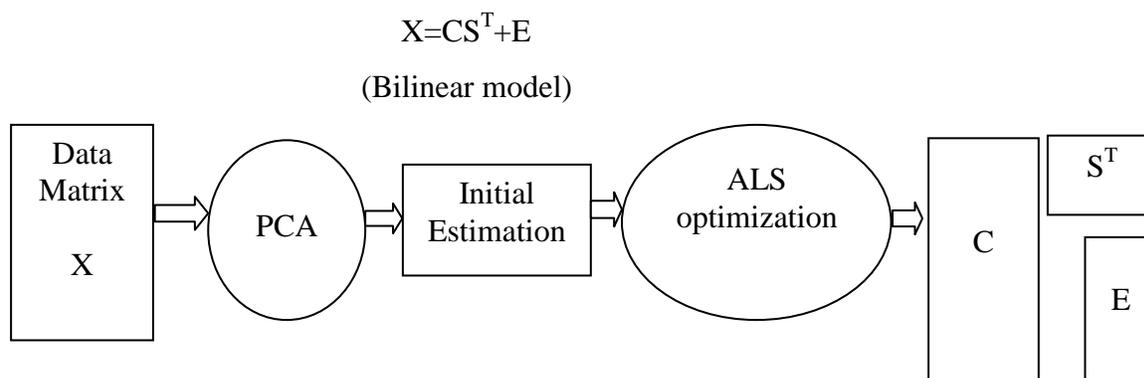


Fig.8. Diagrammatic representation of MCR-ALS [26]

MCR-ALS technique can be used without prior knowledge about the reaction mixture. In such cases determination number of components that cause majority of variability in the data matrix can be tricky. When there is prior knowledge about the reaction (for example knowledge about the number of major components involved in the reaction) MCR-ALS can be used to help build the reaction profile and further confirm results obtained from other methods like PCA.

### 2.3.3. Partial least square regression

PLS regression is a statistical method that comprises features of PCR (Principal component regression) and MLR (Multiple linear regressions). Like PCR, it decomposes the data so as to retain as much of variation as possible and like MLR it finds correlation between predicted and observed variables [27].

PLS regression tries to make relationship between predicted variables (dependent variables or X-block) and observed variables (independent variables Y-block) by finding a linear regression model. It is useful for constructing predictive models from a large number of variables that are highly collinear. PLS emphasizes on predicting the responses and not necessarily on trying to

understand the underlying relationship between the variables [28]. There are several ways to calculate PLS model parameters. The most common algorithms are SIMPLS and NIPALS algorithms. PLS model can be considered as consisting two outer relationships (for each X and Y blocks) and an inner relationship to link both blocks using a regression coefficient [29]. The two outer relationships can be represented as follows:

$$X = TP^T + E = \sum t_a p_a^T + E$$

$$Y = UQ^T + F = \sum u_a q_a^T + F$$

Where, T and U are score matrices,  $P^T$  and  $Q^T$  are loading matrices and E and F are residual matrices for X and Y blocks respectively. In PLS regression the predictor matrix X and the response matrix Y are modeled as linear combinations of a set of orthogonal components. The components are latent variables which are linear combinations of the predictor variables. These linear combinations are chosen successively in such a way to achieve maximum covariance with Y [30].

PLS regression first extracts small numbers of loading weights,  $W(X)$ , from the data matrix, X. Then the extracted loading weights will be stored in scores matrix, T, and used iteratively to model variables in X and Y. Variables in Y can be modeled through variables in X using a regression coefficient, B. the regression coefficient is estimated as a function of P and Q (loadings of X and Y).

$$B = W(P^T W)^{-1} Q^T$$

The regression coefficient enables a sample to be predicted without the need to resolve it into scores and loadings matrices. If  $x_i$  is a vector representing a spectrum in a dataset, then its concentration,  $y_i$ , can be estimated using the regression coefficient as follows:

$$y_i = x_i^T B$$

After PLS regression model is built an important step still remains and this step is validation of the model. There are several methods for model validation. Among the most common one is cross validation method. Cross validation is extensively used to test model stability and robustness. It involves using a sample dataset for training followed by test data set from the same population to evaluate how well the model predicts [31].

#### 2.3.4. Standard normal variate transformation

SNV transformation is used to reduce additive and multiplicative effects in spectra. The transformation is done for each spectrum individually by subtracting the spectrum mean and scaling with the spectrum standard deviation according to the following equation:

$$x_{i,\text{SNV}} = \frac{(x_i - \bar{x})}{\sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2}{k-1}}}$$

Where,  $x_{i,\text{SNV}}$  is the transformed Raman intensity for wavenumber  $x_i$ , and  $\bar{x}$  is the mean intensity of all the  $k$  wavenumbers in the spectrum. The above equation should be repeated for all  $k$  wavenumbers in the spectrum. SNV is effectively centering and normalizing the rows [32].

### 3. Experimental

#### 3.1. Materials and methods

##### Chemicals

Commercially available chemicals and reagents were used directly without further purification. The starting chemicals and reagents were: 1,10-decanediol, the phase transfer catalyst tetrabutylammonium bromide (TBAB) and HBr (48%). Methanol and silica were used to quench aliquots of the reaction mixture taken for GC-MS analysis.

## Equipments

An oil bath was used to keep the reaction mixture at maintained temperature of 105°C. The reaction was carried out in a 500 ml beaker. An electronic radial turbine (flat blades, 50 mm) was used to stir the reaction mixture at a speed of 200 rotations per minute. In addition, a thermometer, vials, small beakers and micropipettes were also used.

## Instrumentation

A RAMANRXN1 Raman spectrometer from Kaiser Optical Systems (Ann Arbor, MI, USA) was used. This spectrometer is equipped with a CCD detector, a holographic transmission grating and a probe attached to an optical fiber. The laser used was a 785 nm Invictus laser with a measurement range from -92 to 3543  $\text{cm}^{-1}$ . Spectral recording was carried out with exposure time of 1 second, 32 scans and with 4  $\text{cm}^{-1}$  resolution.

For NIR spectral measurement, NIRSystems model 6500 manufactured by PERSTORP Analytical (Perstorp analytical incorporated, MD, USA) was used. This spectrometer is equipped with a probe attached to an optical fiber. Spectral recording was done from 1100 nm to 2500 nm with an interval of 2 nm and 32 scans for each recording. The detector was in a transmission mode and a fiber optical module was selected.

GC-MS analysis was done using GC-8000 model gas chromatograph made by Fison instruments (Manchester, UK) and it is equipped with MD 800 mass spectrometer and fused-silica gas chromatographic capillary column with 25m X 0.25 mm internal diameter. Carrier gas was Helium. Ionization mode was electron impact with 70 eV. The injector and detector temperature were 250 °C and 300°C respectively.

## 3.2. Experimental Procedures

### 3.2.1. Raman & NIR measurements

An upscaled reaction protocol [4] was used for the reaction as a procedure with some modifications for proper instrumental measurement. 16.67 g of 1,10-Decandiol and 15.34 g of degranularized tetrabutylammonium bromide were added into a 500 ml beaker and placed into an oil bath of maintained temperature of 105 °C the mixture was melted for 10 minutes and 5 ml of aqueous HBr (48%) was added.

The reaction mixture was stirred at a rate of 200 turns per minute using a radial turbine. Both the Raman and NIR probes were inserted into the mixture and spectra were recorded after every 30 seconds for a period of 150 min. (Though sampling time was set to 30 seconds for both Raman and NIR spectrometers, Raman spectra were actually recorded approximately every 50 seconds. This results in different number of spectra in Raman and NIR measurements for the same reaction).

### 3.2.2. GC-MS analysis

The reaction started as soon as HBr was added to the reaction vessel. While Raman and NIR spectra were taken every few seconds, sampling for GC-MS analysis was done every 15 minutes as follows:

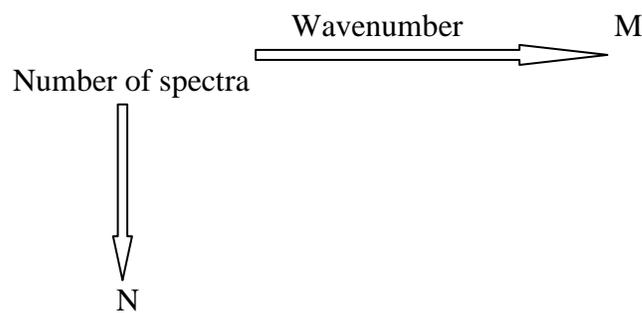
- First a micropipette was taken and a small amount of cotton was inserted in it as a filter.
- Then small amount of silica powder was added on top of the cotton.
- Then an aliquot of the reaction mixture was added on top of the silica powder and 5-10 drops of methanol was added. This step effectively stops the reaction from proceeding further.
- Then the contents in the micropipette were filtered into a vial assigned for GC-MS analysis using a small pipette pump.
- Finally the vial was sealed, labeled and taken to the GC-MS vial holder.

The above steps were repeated every 15 minutes for 2 and 1/2 hrs. A total of 10 samples were taken for GC-MS analysis.

## 4. Result and discussion

### 4.1. Data description

A spectrum recorded by a Raman or NIR instrument can be considered as a row vector consisting of  $M$  elements where  $M$  is the number of variables. In this case the variables are wavenumbers (3546 for Raman and 700 for NIR). This row vector represents only a single spectrum but during reaction monitoring several spectra are recorded in a certain time interval to follow the progress of a reaction. This results an  $N \times M$  data matrix where  $N$  is the number of spectra taken.



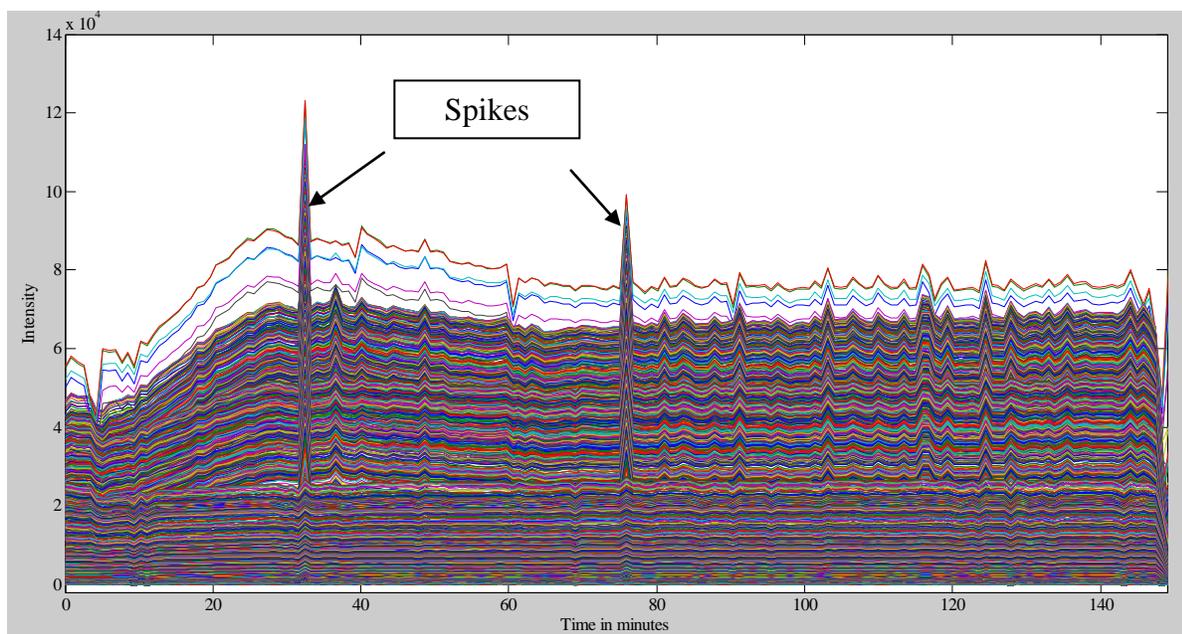
*Fig.9. Set up of the data matrix*

Therefore in the data matrix, every row represents a spectrum recorded at a specific time, while a column represents the absorption at a specific wavenumber (in NIR) and Raman intensity at a specific wavenumber (in Raman). Plots of the raw data for Raman and NIR data are shown later.

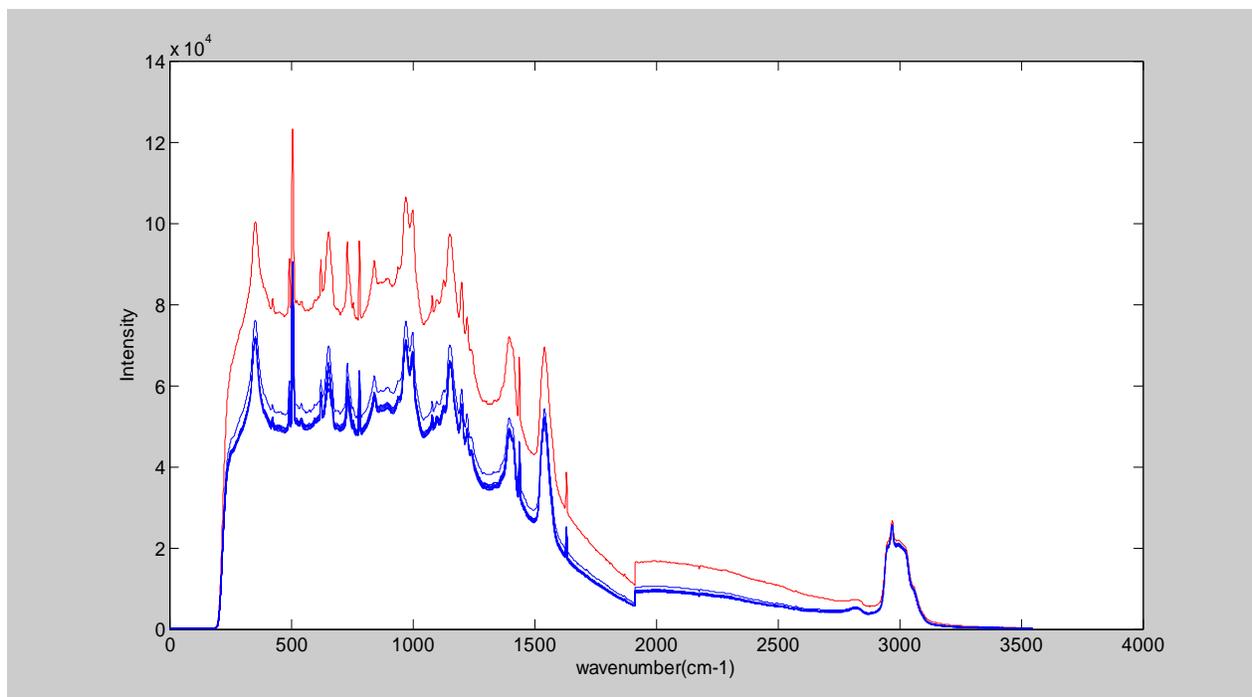
## 4.2. Data pretreatment

### 4.2.1. Raman spectroscopy

Four different Raman data sets were collected independently on four different experiments. However, the first three Raman data sets were collected without periodic GC-MS analysis of the reaction mixture (this applies for NIR data too). The last data set was collected while there was GC-MS analysis of aliquots of the reaction mixture and all the pretreatment, analysis and interpretation will be based on this data set. The raw data contains 176 rows and 3546 columns. However, a simple visual inspection shows that there are apparent outliers (spikes) characterized by sharp increase in intensity in some of the rows (figure 10). These spikes were removed from the data set. Usually the sources of such sudden increase in intensity are attributed to cosmic ray interferences which are quite common in Raman instruments that use charge coupled devices (CCD) as a detector. Such interferences are characterized by large intensities that usually eclipse the small Raman intensities by several degree of magnitude and should be removed prior to any data analysis method [33,34]. Two of the spikes that have been removed from the dataset are shown in the following figure.



*Fig.10. Plot of transposed raw data with spikes in the data set*



*Fig.11.Intensity comparison of one of the outlier spikes (red) with other spectra (blue)*

Furthermore, the score plot (figure 12) on the raw data shows that there are few more outliers that can be removed. Spectrum 39 and 90 are the spikes shown in figure 10, whereas spectrum 175 and 176 shows a steep decline in intensity. Perhaps this is because these are the last two spectra taken and the steep decline might be due to technical rather than chemical causes. Nevertheless, spectrum 39, 90, 175 and 176 were removed from the data. Spectrum 44, though the reason is unknown, shows a marked difference with other spectra around and it is also removed.

As shown in figure 13 the first 7 spectra (particularly spectrum 6) on the raw data seems to be relatively different than other spectra in close proximity, however this is because HBr, which kicks off the reaction as soon as it is added into the reactants, was added after the first few spectra were recorded (this helps to see what change addition HBr caused on the spectra of the reaction mixture).

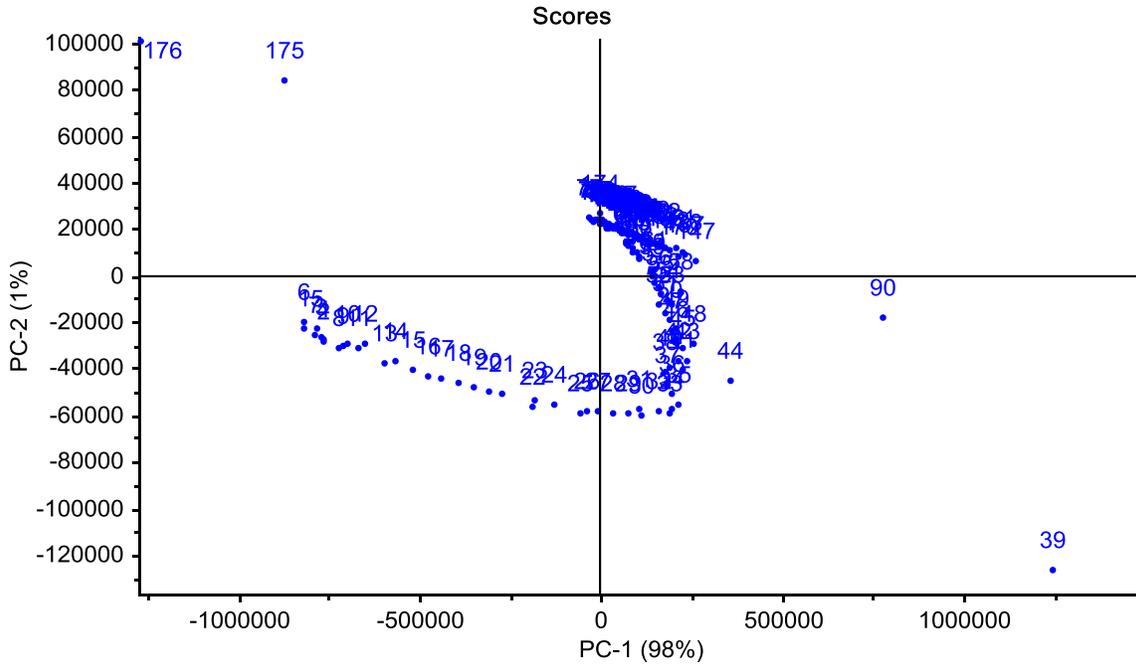


Fig.12. Score plot shows that sample 39, 44, 90,175 & 176 are outliers

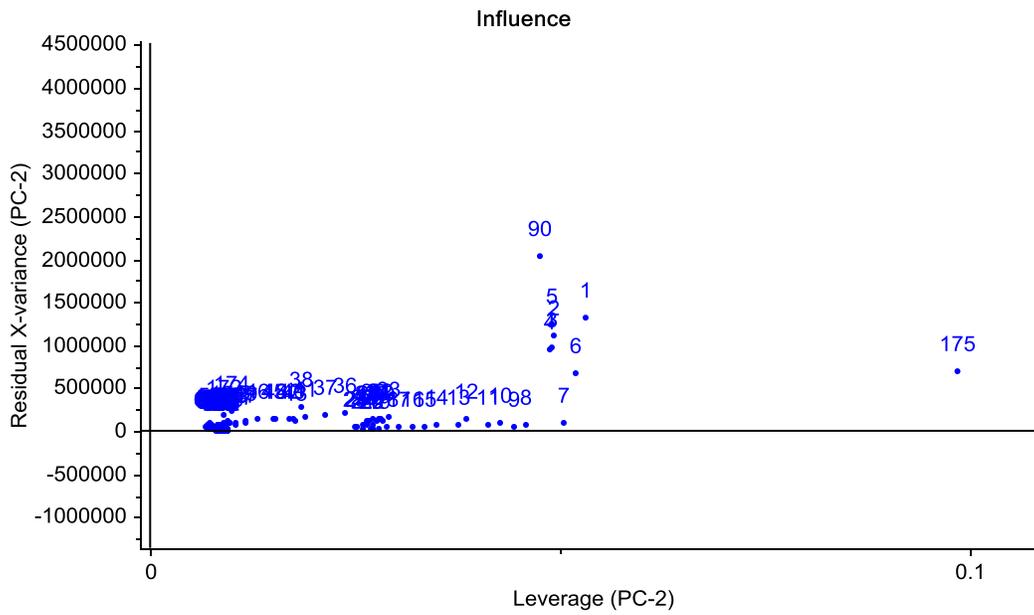
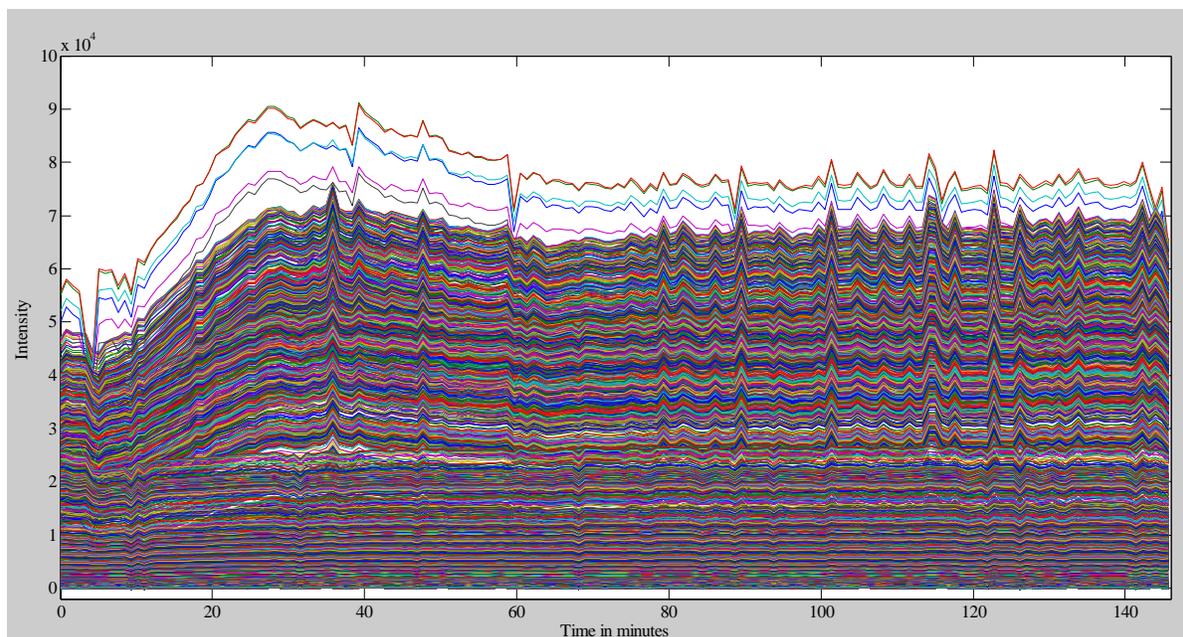


Fig.13. The first 7 samples can also be considered outliers



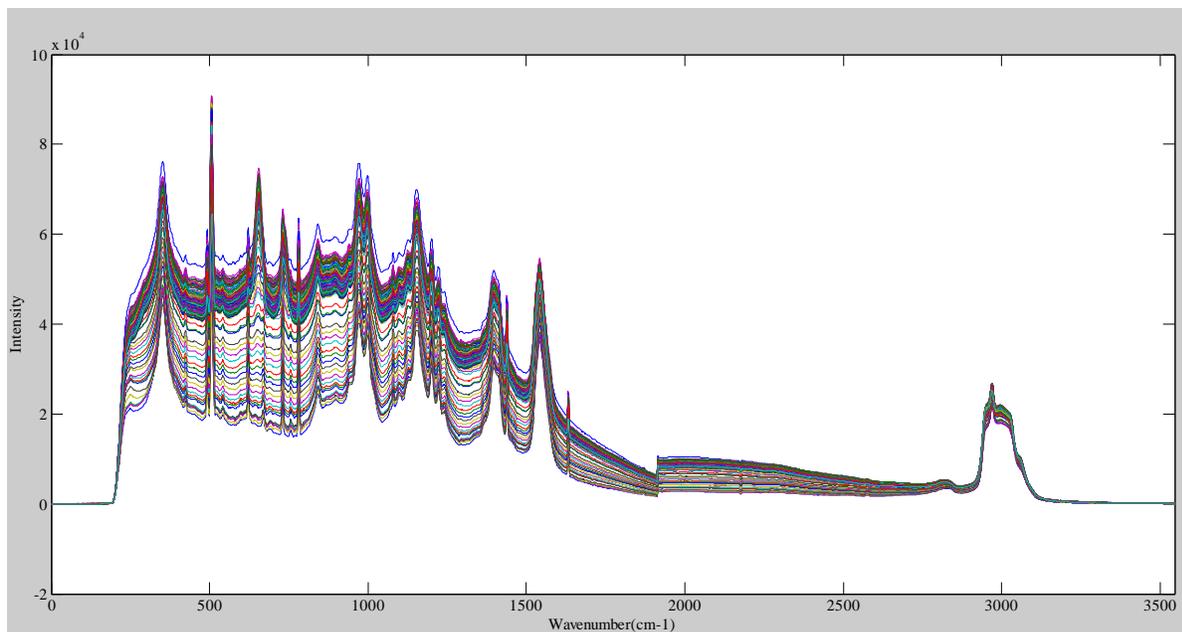
*Fig.14. Plot of transposed raw data after the spikes are removed*

After the spikes and other outliers are removed, very strong fluorescence background remains another issue to be resolved. Several background correction methods were tried to reduce or remove the effect of fluorescence using different methods ranging from polynomial fitting to Standard normal variate method (SNV) [35, 36, 37].

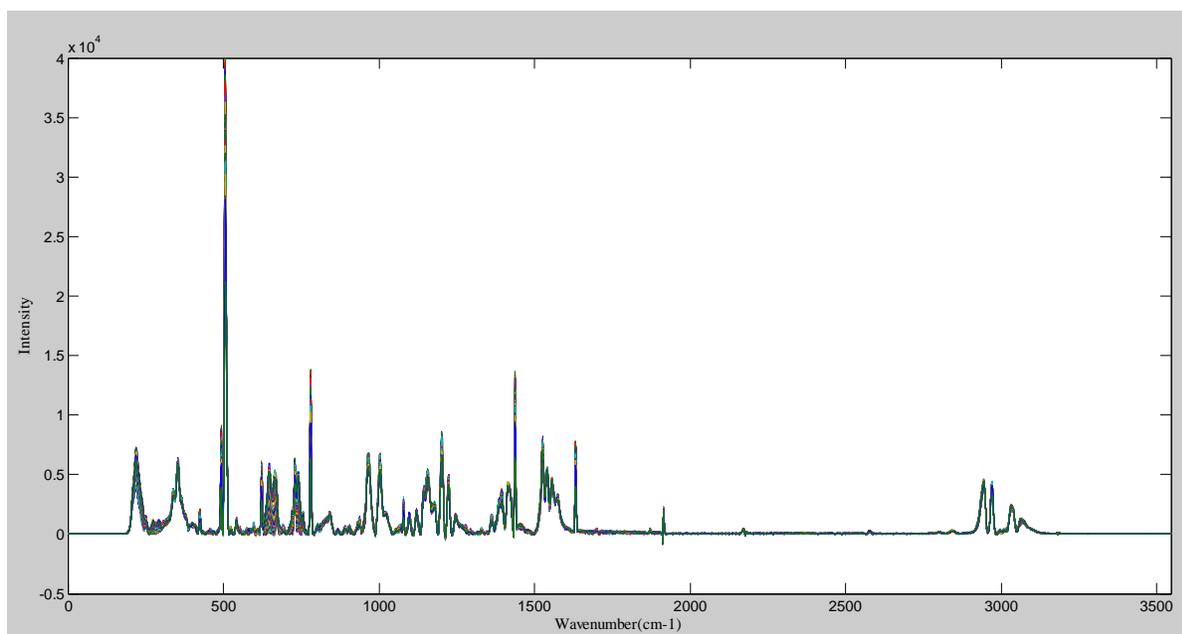
However, better result was obtained using a MATLAB function called MSBACKADJ (Mathworks bioinformatics toolbox) [38]. This function estimates baseline based on input values of window size (window size sets width of a shifting window in which a background point is estimated for windows of selected width) and step size (step size sets steps of a shifting window in which a background point is estimated for windows at every selected step size value) then it regresses the baseline using approximation methods such as linear interpolation or spline interpolation before it corrects the baseline of the spectrum.

Though this function was designed at first for mass spectral data, it can also be used for other spectral data [38] including Raman spectral data provided that the appropriate window and step sizes are selected. Default values of 200 window size and step sizes are too big for Raman data and doesn't work quit well because of bad peak resolution instead, by trial and error window and

step size values of 11 were found to give the best result. The effect of the pretreatment method can be seen in figure 15 and 16.



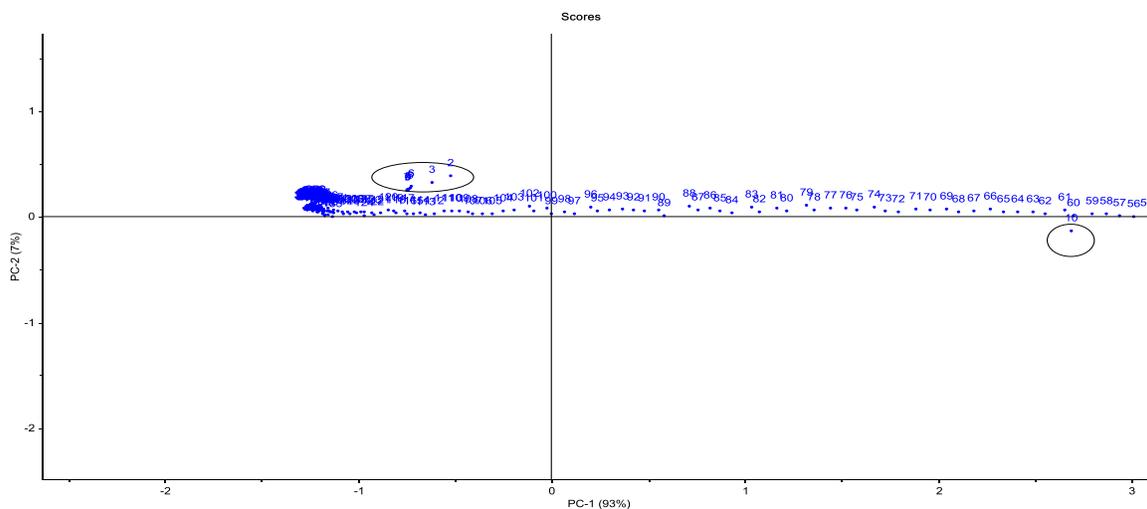
*Fig.15. Raman data before background correction*



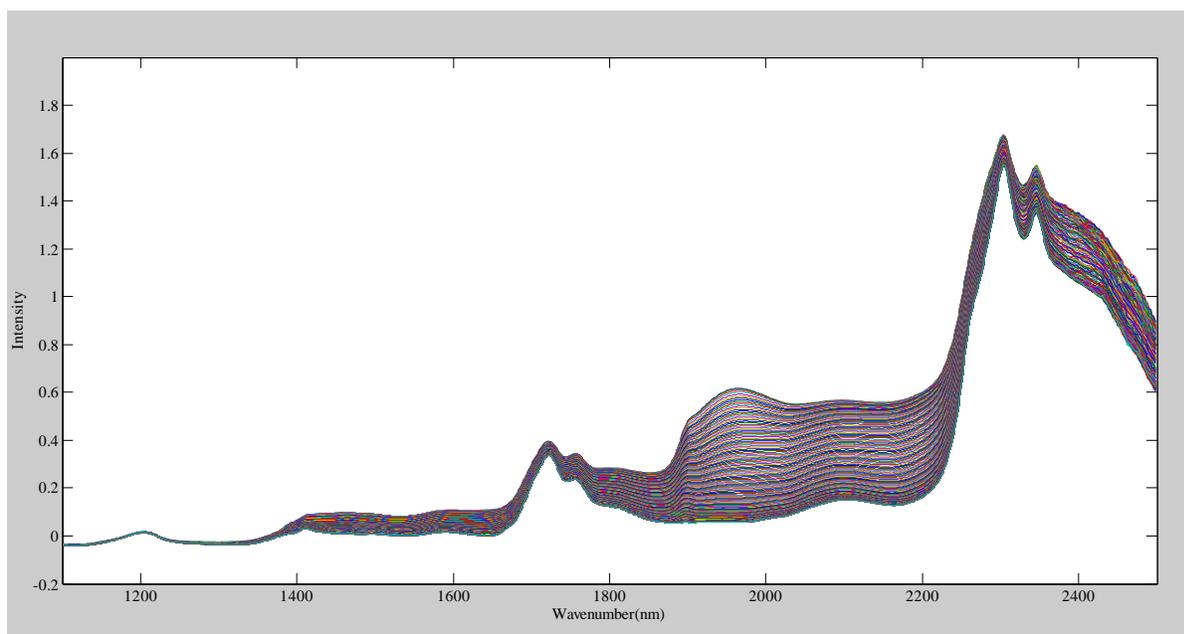
*Fig.16. Raman data after background correction*

### 4.2.2. NIR spectroscopy

Raw data for NIR spectroscopy contains 322 rows and 700 columns. The score plot below (figure 17) shows spectrum 1 and 10 to be clearly outliers. It also shows the effect of addition of HBr in to the reaction vessel. This is shown by a separate grouping of samples from 2-8 in score plot. 10 outliers were removed from the raw data and the remaining data consists of 312 rows and 700 columns.

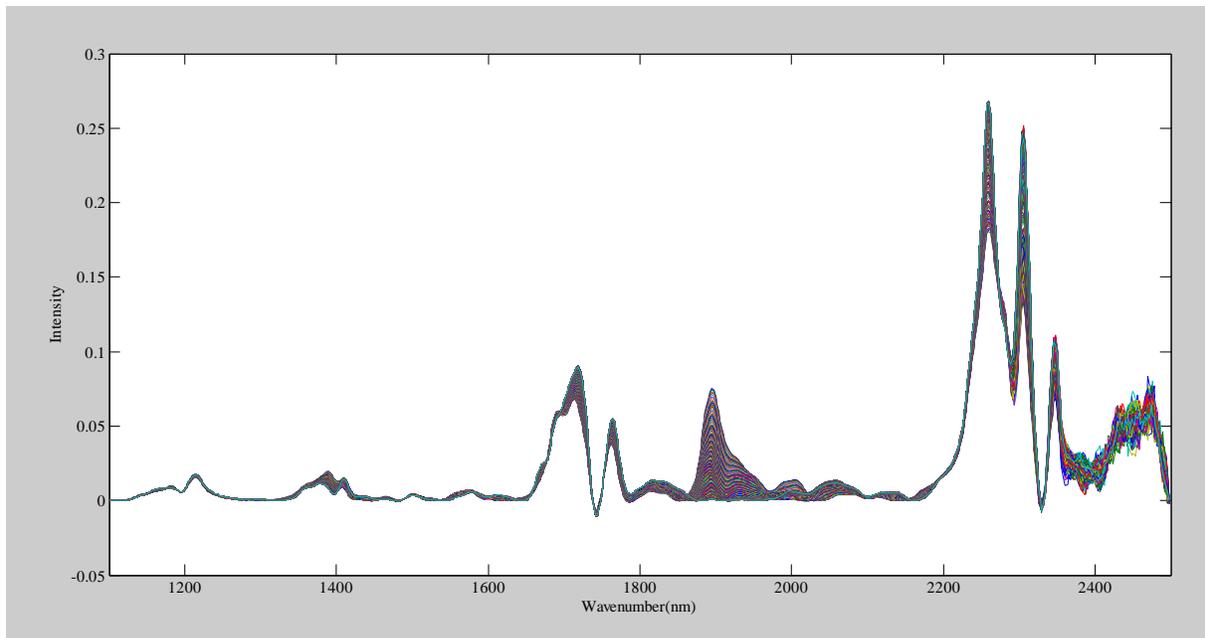


*Fig.17. Score plot of the raw NIR data*



*Fig.18. NIR spectroscopy raw data*

The raw dataset as shown in Figure 18 shows baseline variation and it was subjected to different spectral pretreatment methods like standard normal variate (SNV), detrending and others. However, once again MSBACKADJ function in MATLAB gives better result and this result is used for further interpretation.



*Fig.19.NIR data after baseline correction*

### **4.3. Principal component analysis**

#### **PCA on Raman Data**

Principal component analysis on the Raman data reveals that there are 3 major principal components that can explain a total of 96% of the variation in the data: PC1 74%, PC2 19% and PC3 3%.

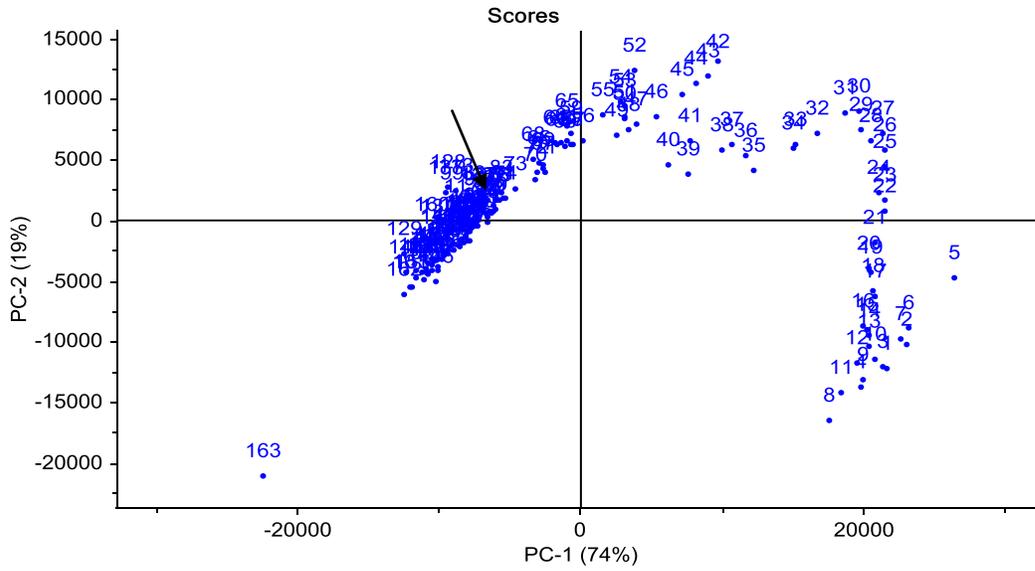


Fig.20A. Raman data score plot (PC1 Vs PC2)

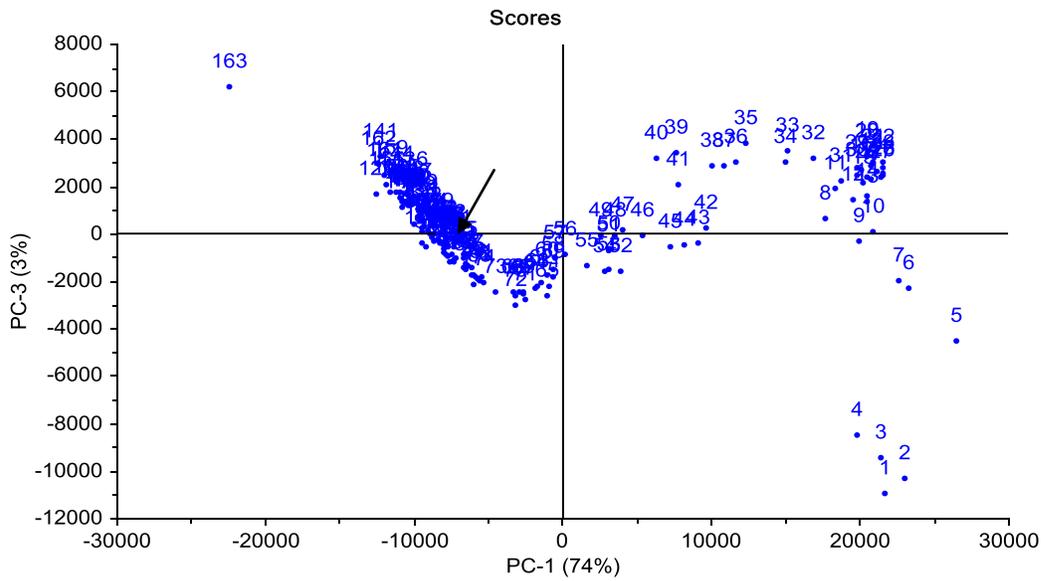
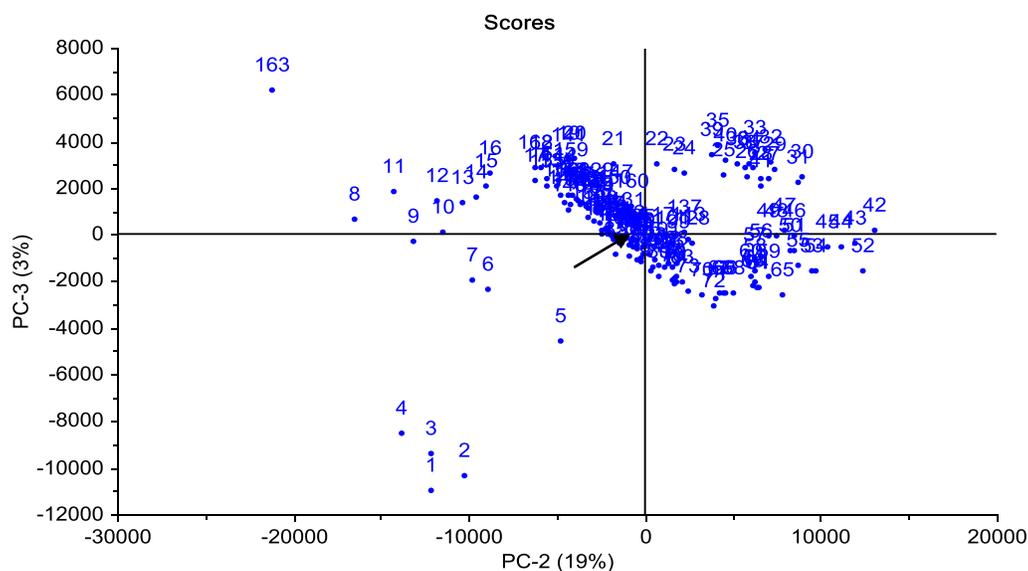


Fig.20B. Raman data score plot (PC1 Vs PC3)



*Fig.20C. Raman data score plot (PC2 Vs PC3)*

From the above score plots it is apparent that there isn't much variation in the objects after the 100<sup>th</sup> object (shown by the arrow in the above score plots) which corresponds to a reaction time of (about 90 min). This is also supported by the result from the NIR data. This indicates that shorter reaction time could be used instead of the one mentioned in the procedure [4]. Further increase in the reaction time increases the amount of the byproduct. This can be seen in the plot of score vectors (Figure 21 in red color).

The score vectors can explain the progress of the reaction. The following is plot of score vectors. It shows three distinct patterns which can be attributed to the three main reaction components. The blue one can be attributed to 1,10-decanediol as its amount is expected to decrease throughout the reaction while the green one is for the product and the red one is for the byproduct.

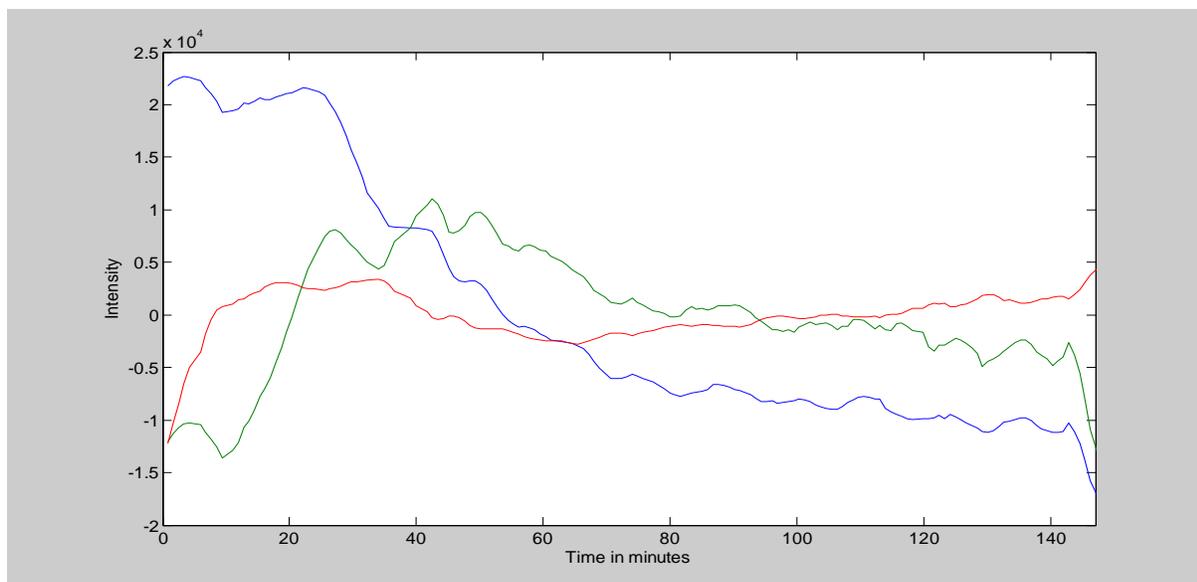


Fig.21. Plot of score vectors: Blue (reactant), green (product) and red (byproduct)

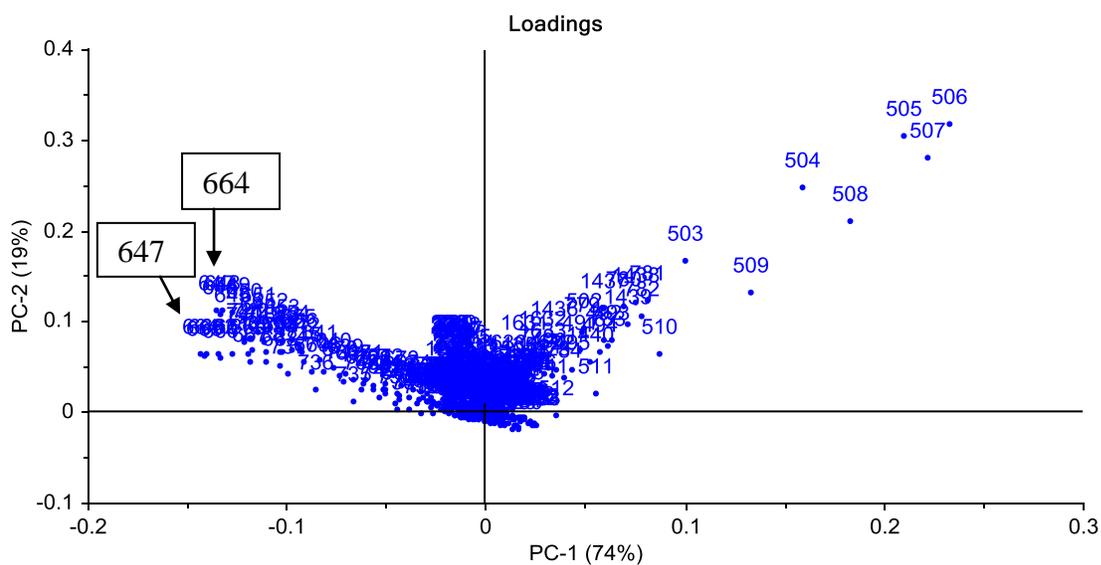


Fig.22. Raman data Loadings plot (PC1 Vs PC2)

From the loadings plot it is possible to see that there are three main variations in the variables: around variable 506 , around variable 647 and around variable 664. The peaks at 647 and 664 lie on the opposite end of variable 506 on the first PC indicating that they show opposite pattern (an increase in variable 506 means decrease in variable 664 and 647). This is fairly expected because

506  $\text{cm}^{-1}$  which is assigned to 1, 10-decanediol is a reactant which is decreasing throughout the reaction time and can account for the vast majority of the variation while peaks at 647  $\text{cm}^{-1}$  and 664  $\text{cm}^{-1}$  are due to the product and byproduct respectively which are in increasing pattern [39].

### PCA on the NIR data

PCA on the NIR data gives two PC'S that can explain a total of 95 % of variation in the data (PC1 93% and PC2 2%). Visual inspection of the score plot reveals that most of the variation in the objects happens until object number 178, which corresponds to a reaction time close to 90 minutes. This agrees with the reaction time indicated by the score plot from the Raman data.

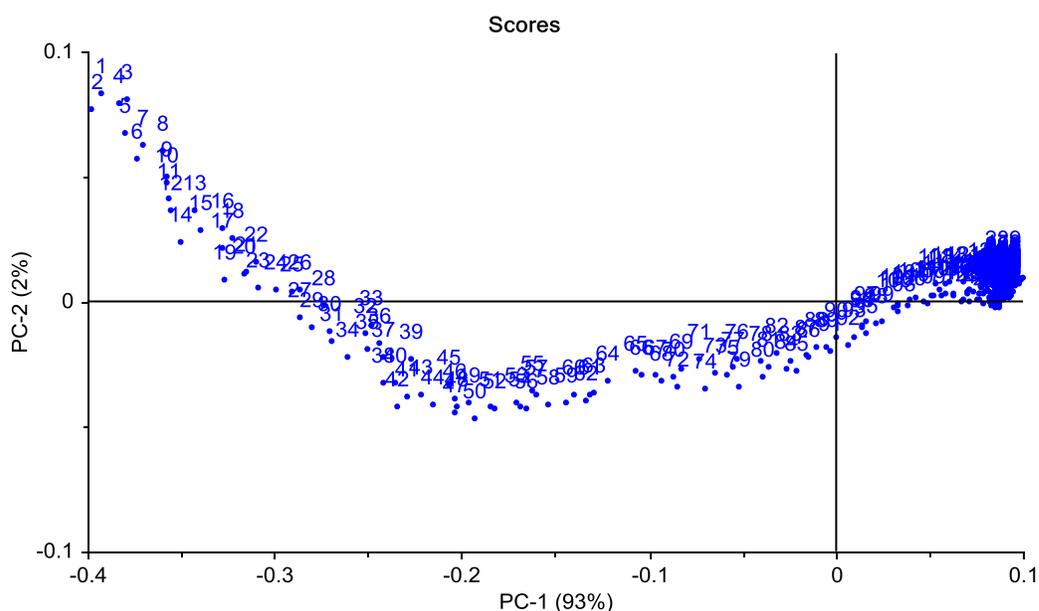


Fig.23. NIR data score plot (PC1 Vs PC2)



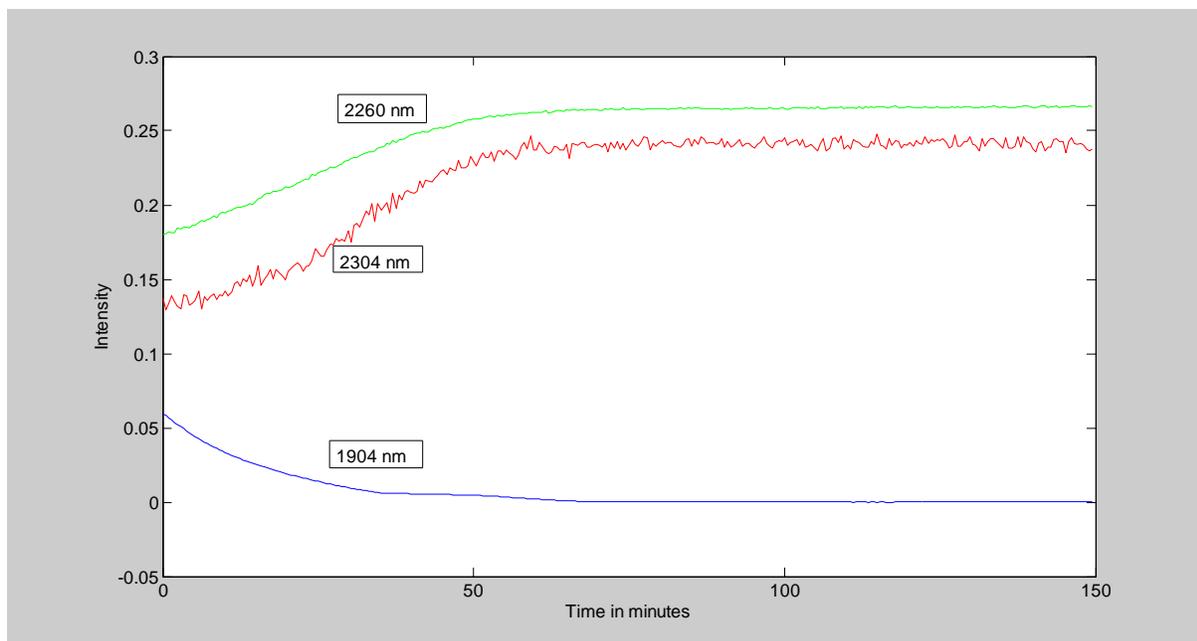


Fig.25. Intensity plot at 1902 nm, 2304 nm and 2260 nm

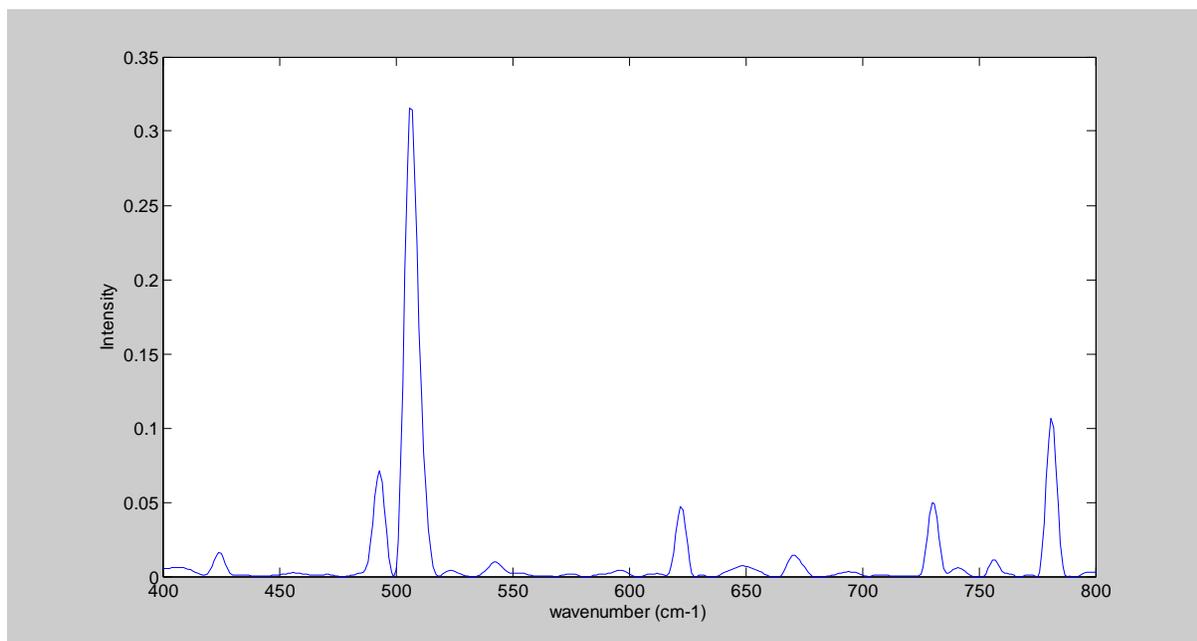


Fig.26. Pure component spectra for the reactant.

#### 4.4. MCR-ALS analysis

##### Application of MCR on the Raman data

Both for the Raman and NIR data three components are selected because at least three compounds are expected to appear in the reaction: the reactant, product and byproduct. The pure component spectrum for the reactant is given in figure 26.

The pure component spectrum for reactant clearly shows that the characteristic bands for C-Br bonds are not present in the regions ranging from 600-700  $\text{cm}^{-1}$ . Very high intensity band around 506  $\text{cm}^{-1}$  and lack of C-Br band is a common feature of this Raman data at the early stages of the reaction. As the reaction progresses the intense band at 506  $\text{cm}^{-1}$  decreases and bands due to C-Br emerges. This can be seen in the components spectra for the reactant, product and byproduct (Figure 27). The concentration profile obtained from the Raman data for the three main components is largely in agreement with the result obtained in the PCA analysis. The reactant once again shows a decreasing pattern while the product and byproduct shows increasing patterns (figure 28).

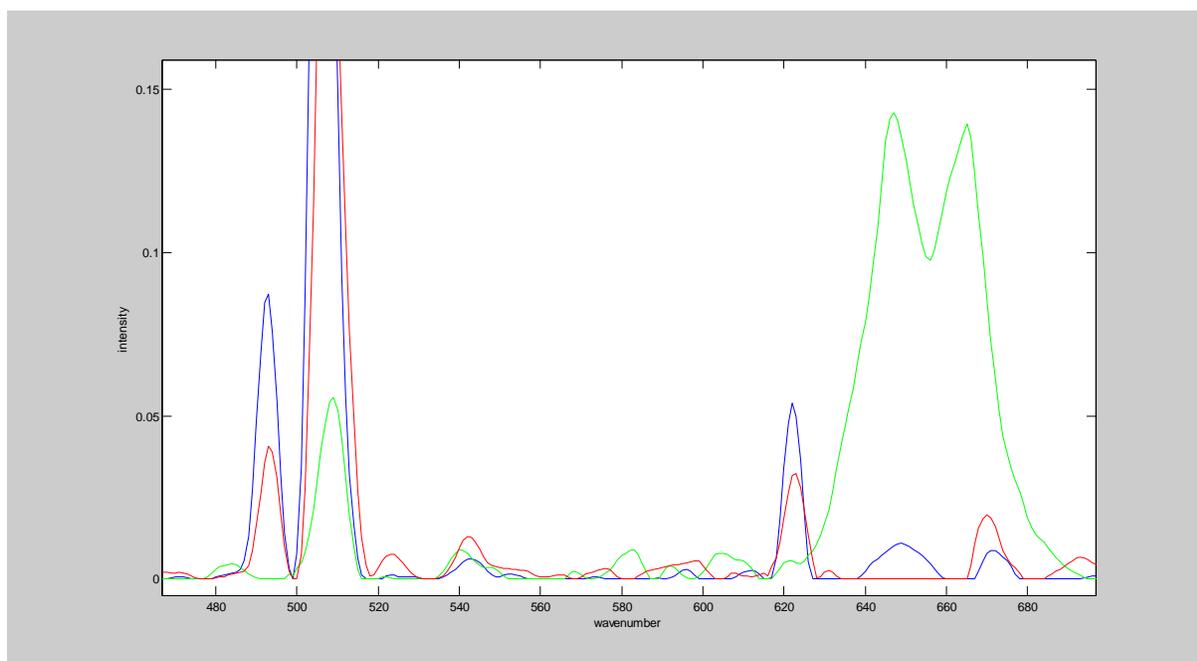


Fig.27. Comparison of spectral profiles: (Blue- reactant, Green - product and red- byproduct)

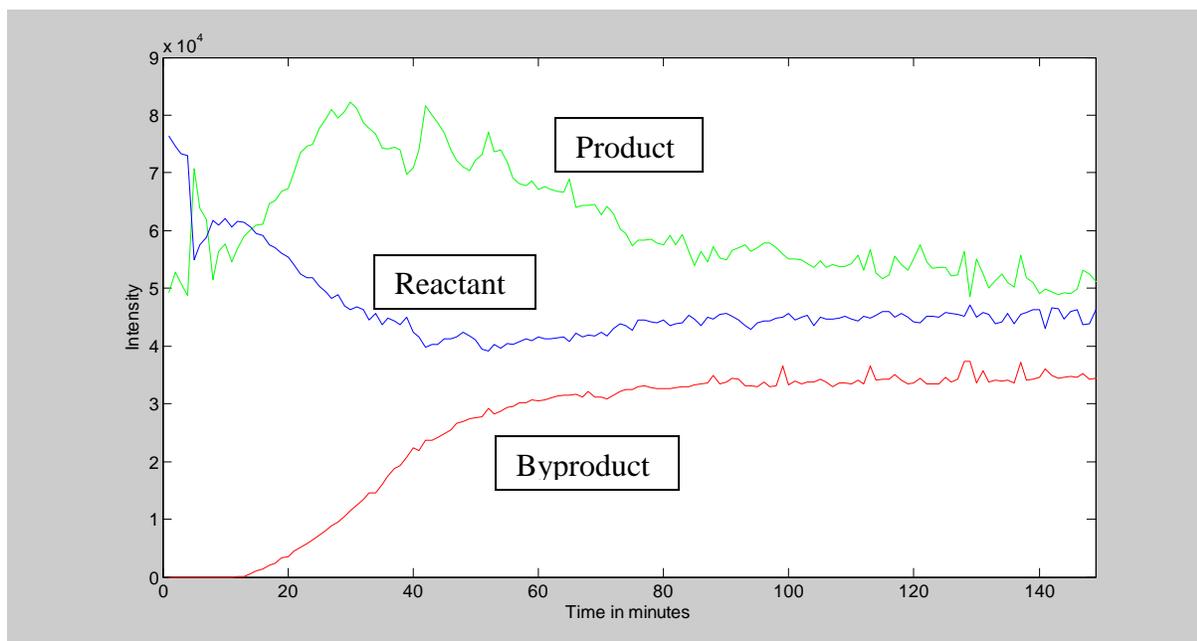
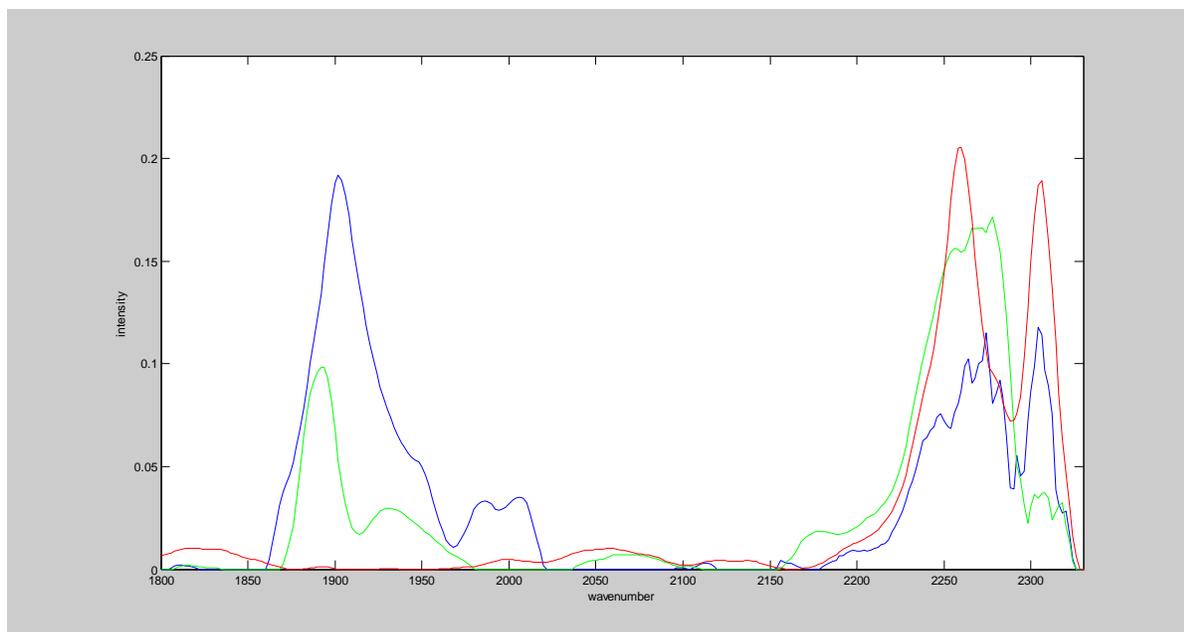


Fig.28. Pure component spectra of the three components (Raman)

### Application of MCR on the NIR data

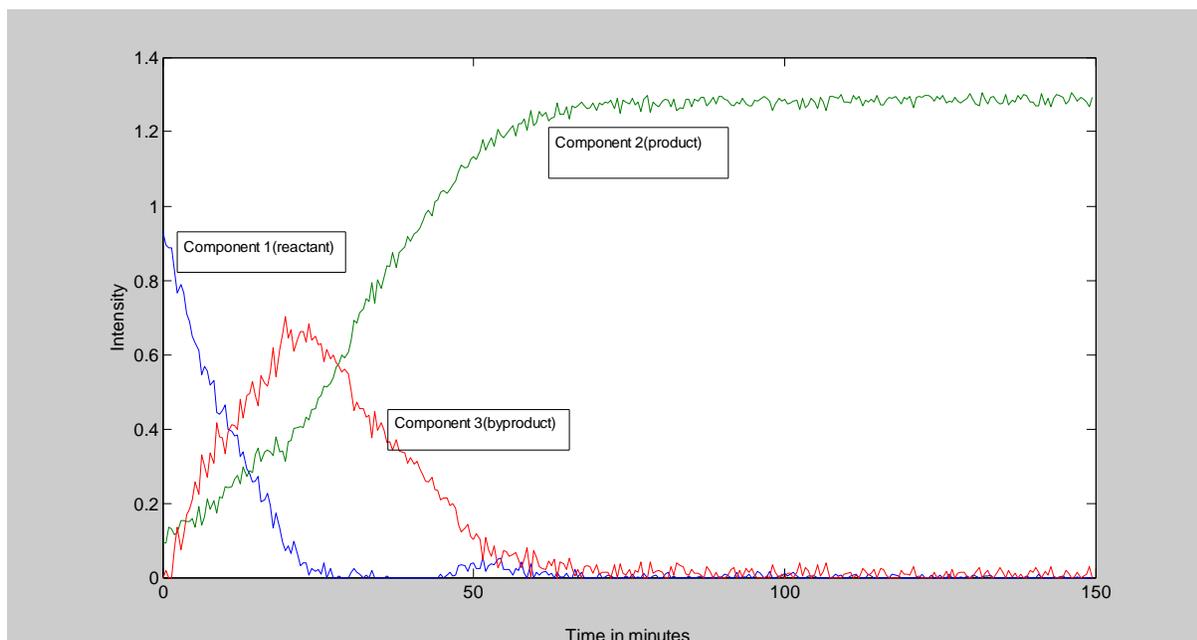
In order to see the three main reaction components, the NIR data was resolved into three components by MCR-ALS method. The plot of spectral profiles shows the common trends seen in the PCA score vector plot for NIR data.

The PCA loading plot shows that the majority of variance is due to variables 1902 nm, 2304 nm and 2260 nm. On the other hand figure 29 shows that the three spectral profiles behave differently around these variables. For example around 1902 nm which is assigned for O-H stretching-bending combination vibrations [40], the first spectral profile has intense peak while the second has medium peak and the third one has very weak intensity peak. From this it is possible to deduce that the pure component shown in the spectral profile number three (red color) doesn't have any O-H group and can be assigned to the byproduct (1,10-dibromodecane). This is further proved by the fact that there is a strong peak at 2304 nm and 2260 nm which are bands due to C-Br bonds.



*Fig.29. Pure component spectra of the three components (NIR)*

By following the same logic, it is possible to assign the spectrum in blue color in the above figure to the reactant and the green colored spectrum to the product. The corresponding concentration profile for the three components (figure 30) is in agreement with the result from PCA analysis.



*Fig.30. Reaction profile obtained by MCR-ALS method for NIR data*

The above plot is comparable with intensity profile plot of variables at 1902 nm, 2304 nm and 2260 nm (figure 25). These variables are the main sources of variability in the data and comparison of the two plots shows that, even though the similarity of the plots is not as good as expected, they still show the main trends happening in the reaction.

#### 4.5. GC-MS analysis

In order to see the variation in percentage composition of the reactant, product and byproduct in the reaction mixture over time, peak area of their corresponding peaks were taken and calculated as follows:

$$\% \text{ Composition of A} = \frac{(\text{Peak area of A}) 100\%}{\text{Total peak area}}$$

However, the above calculation is based on two assumptions. The first one is that the relative response factor for the three compounds in the reaction mixture is the same (this assumption is a fair assumption considering that all the three compounds have similar molecular structures and hence expected to cause more or less similar response from the detector).

The second assumption is that the reactant undergoes change only to the product and the byproduct as a result no other compound that might be formed during the reaction is accounted for in this calculation (there were no sign of peaks in the chromatograms apart from the three compounds of interest so, this assumption is also fair) [41].

Sample	Reactant	Product	Byproduct	Total area	% Reactant	% Product	% Byproduct
1	416184214	158147015	13853432	588184661	70.75741	26.887307	2.355286
2	767502196	311217933	110206990	1188927119	64.55418	26.176368	9.269449
3	772129916	372409680	157827437	1302367033	59.28666	28.594833	12.11851
4	845310217	514627656	227378301	1587316174	53.25405	32.421244	14.3247
5	866419934	721005526	340792667	1928218127	44.93371	37.392322	17.67397
6	890724904	943551176	502488088	2336764168	38.11788	40.378537	21.50359
7	637991903	733770496	439435144	1811197543	35.22487	40.513002	24.26213
8	594954318	646923318	387172916	1629050552	36.52154	39.711679	23.76678
9	617023471	782702103	479232586	1878958160	32.8386	41.656175	25.50523
10	624679519	788011992	463707117	1876398628	33.29141	41.99598	24.71261

Table .2. Composition of components calculated from chromatographic peak area

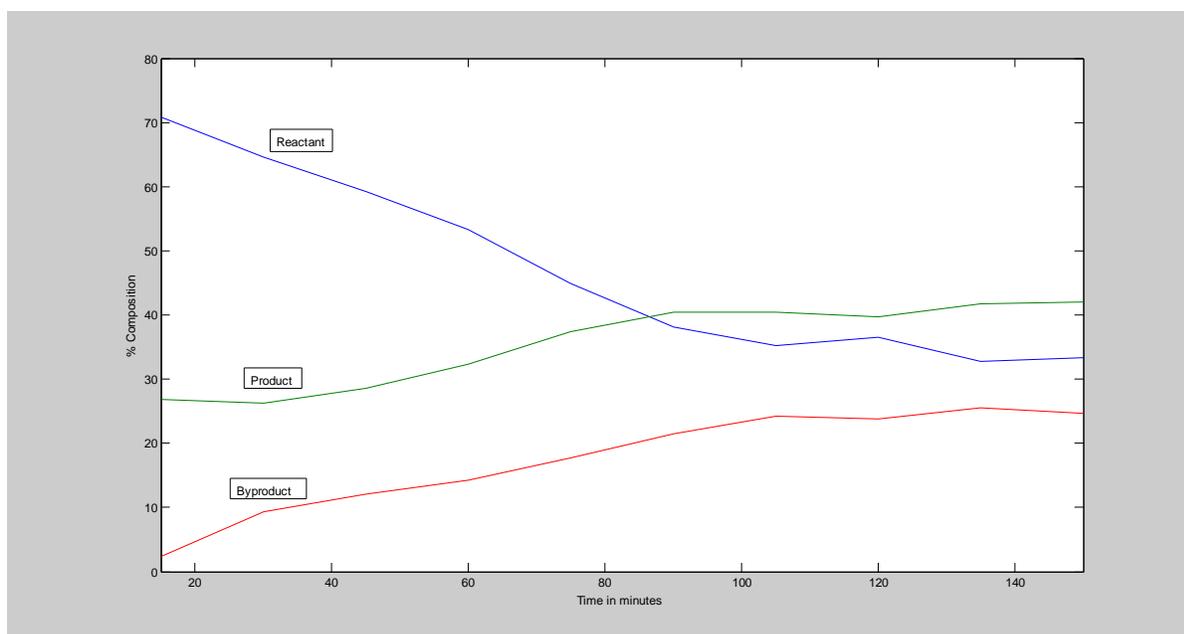


Fig.31. Plot of percentage composition of the reaction mixtures against time. This plot is similar to score vector plots for Raman and NIR data.

#### 4.6. PLS models

PLS models were built for Raman and NIR data separately by selecting 10 spectra from the preprocessed data. Each spectrum selected corresponds to a reaction time in which GC-MS

analysis was performed. This means, for the GC-MS analysis performed at the 15<sup>th</sup> minute in the reaction time, the corresponding Raman and NIR spectra will be selected. Since the reaction was carried out for 150 minutes, a total of ten spectra are selected. At the same time ten response values are also obtained from GC-MS analysis. Table 3 (Raman) and table 4 (NIR) shows values used in the PLS model.

No	Spectrum No.	Reaction time (min)	GC-MS result (% composition)		
			Reactant	Product	Byproduct
1	18	15	70.8	26.9	2.4
2	34	30	64.6	26.2	9.3
3	50	45	59.3	28.6	12.1
4	67	60	53.3	32.4	14.3
5	82	75	44.9	37.4	17.7
6	99	90	38.1	40.4	21.5
7	115	105	35.2	40.5	24.3
8	131	120	36.5	39.7	23.8
9	148	135	32.8	41.7	25.5
10	162	150	33.3	42	24.7

*Table .3. Spectra selected for Raman PLS model along with reaction time and % composition*

No	Spectrum No.	Reaction time (min)	GC-MS result (% composition)		
			Reactant	Product	Byproduct
1	32	15	70.8	26.9	2.4
2	64	30	64.6	26.2	9.3
3	94	45	59.3	28.6	12.1
4	125	60	53.3	32.4	14.3
5	157	75	44.9	37.4	17.7
6	190	90	38.1	40.4	21.5
7	220	105	35.2	40.5	24.3
8	250	120	36.5	39.7	23.8
9	282	135	32.8	41.7	25.5
10	311	150	33.3	42	24.7

*Table .4. Spectra selected for NIR PLS model along with reaction time and % composition*

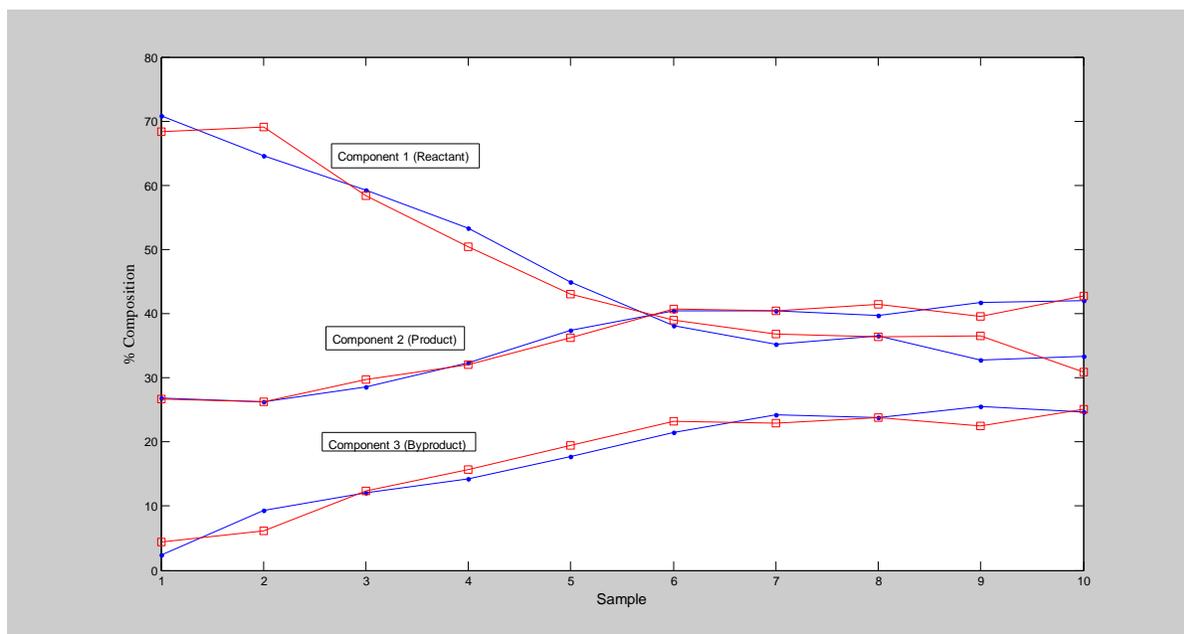
## Raman PLS model

A PLS model was established for Raman data using 3 components. The variance captured by the model is given in the following table.

Components	X-block		Y-block	
	Current	Total	Current	Total
1	99.01	99.01	93.55	93.55
2	0.86	99.87	5.74	99.29
3	0.08	99.95	0.45	99.74

*Table.5. Variation captured by Raman PLS model*

The prediction power of the model is compared by plotting the actual GC-MS analysis result along with the one predicted by the model (figure 32) and this seems to be a good prediction by the PLS model.



*Fig.32. Plot of measured (Blue) and predicted (red) composition*

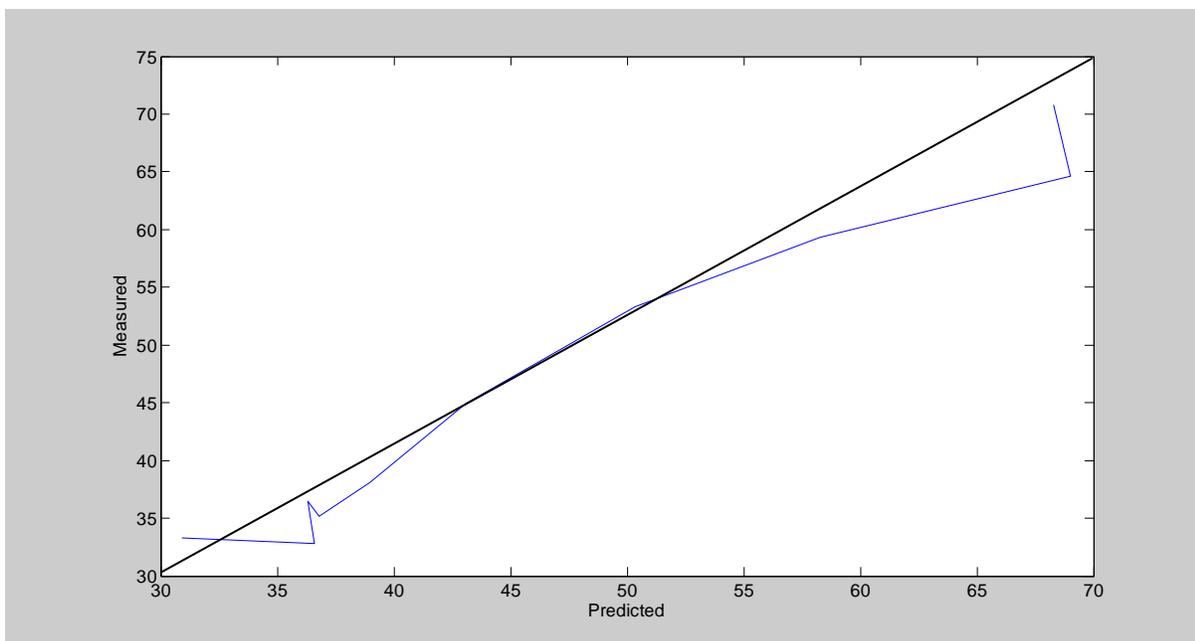


Fig.33A. Plot of predicted composition against measured composition for the reactant  
(Pearson's  $r = 0.9825$ )

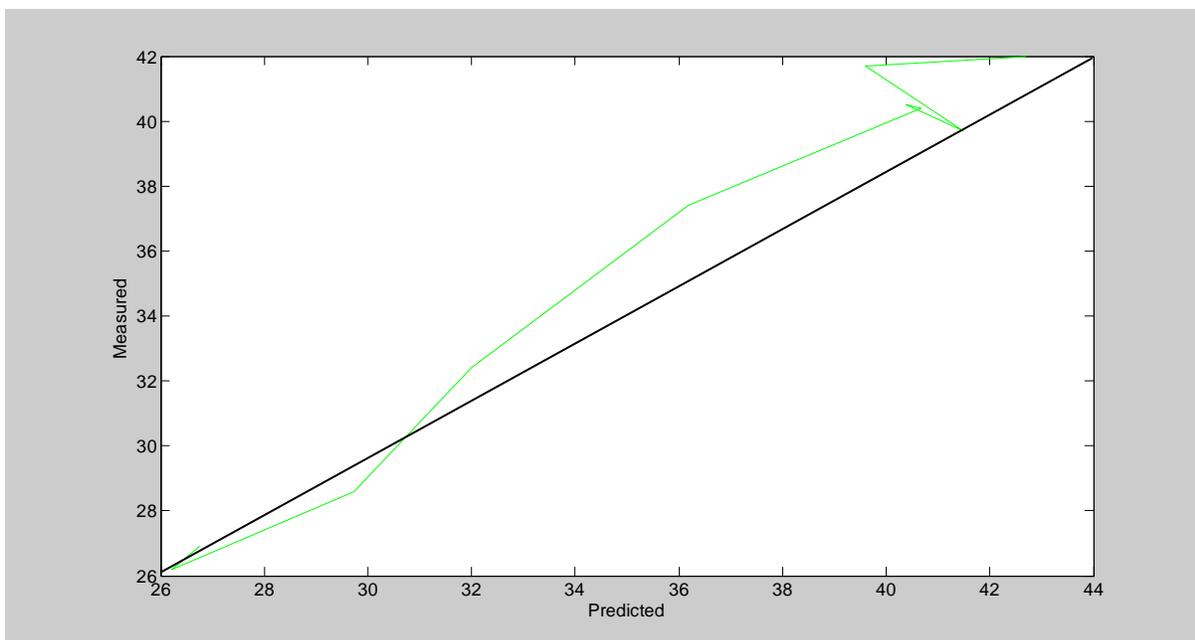
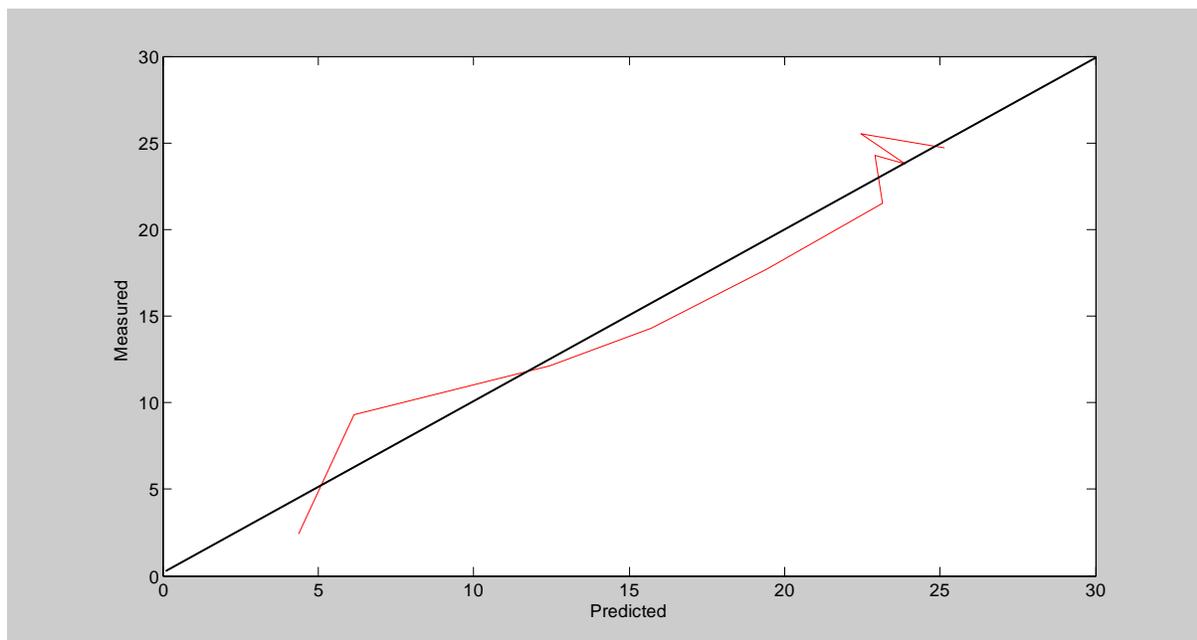


Fig.33B. Plot of predicted composition against measured composition for the product (Pearson's  
 $r = 0.9847$ )



*Fig.33C. Plot of predicted composition against measured composition for the byproduct (Pearson's  $r = 0.9696$ )*

### Raman model validation

Raman model validation was carried out by forming a new set of dependent variables (X-block 2) and performing the validation process using the previous Raman model. For this purpose 10 other spectra were selected from the pretreated Raman data consisting 5 spectra between 15<sup>th</sup> and 30<sup>th</sup> minute of reaction time and another 5 spectra between 135<sup>th</sup> and 150<sup>th</sup> minutes in the reaction time. In the table below, it is possible to see that the validation of the model is good as it predicts fairly well for a new X-block variables.

Since the validation spectra are taken at reaction times between two GC-MS analyses, the predicted composition are expected to be close to the two measured GC-MS values. For example spectra taken at 16<sup>th</sup>, 19<sup>th</sup>, 22<sup>nd</sup>, 25<sup>th</sup> and 28<sup>th</sup> minute in the reaction should have an estimated composition close to the measured GC-MS values at 15<sup>th</sup> and 30<sup>th</sup> minute for each component. As shown in the following table, the estimated values are within a reasonable range between the two GC-MS values.

Reaction time (min)	Estimated % Composition by the model		
	Reactant	Product	Byproduct
16	68.6	26.9	4.5
19	73.4	24.7	3.0
22	76.4	23.7	2.6
25	75.6	24.0	3.3
28	69.4	25.5	5.7
136	35.9	41.2	23.4
139	34.7	41.8	24.0
142	31.6	43.0	25.2
145	33.7	42.3	24.5
148	35.4	42.5	24.6

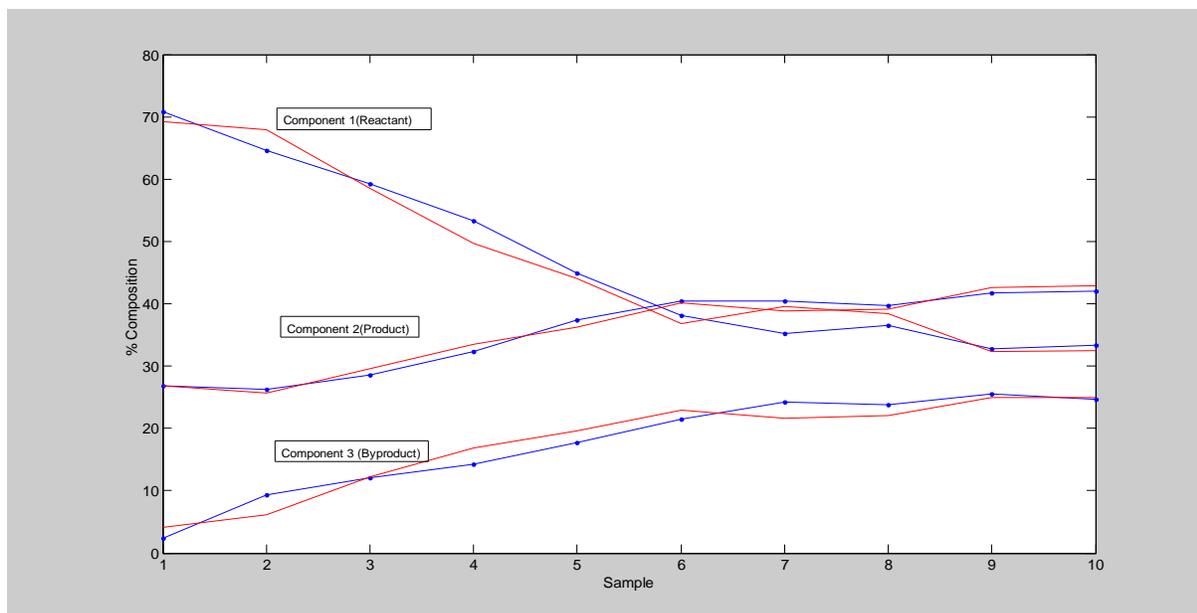
*Table.6. Prediction of composition using the Raman PLS model for new X-block variables*

### NIR PLS model

A PLS model was established for NIR data using three components. The variance captured by the model is given in the following table.

Components	X-block		Y-block	
	Current	Total	Current	Total
1	99.54	99.54	92.15	92.15
2	0.39	99.94	6.30	98.45
3	0.01	99.95	1.30	99.76

*Table.7. Variation captured by NIR PLS model*



*Fig.34. Plot of Measured (Blue) and Predicted (Red) % composition by NIR PLS model*

The above plot shows that the PLS model for the NIR data makes a good prediction. And it seems to agree with the one predicted by the Raman PLS model. The PLS model is further validated by selecting another set of samples from the NIR data and compare if it makes similar prediction. There is a good correlation between the predicted and measured composition for all components at different reaction times. Figure 35 A-C shows how the measured and predicted compositions are correlated.

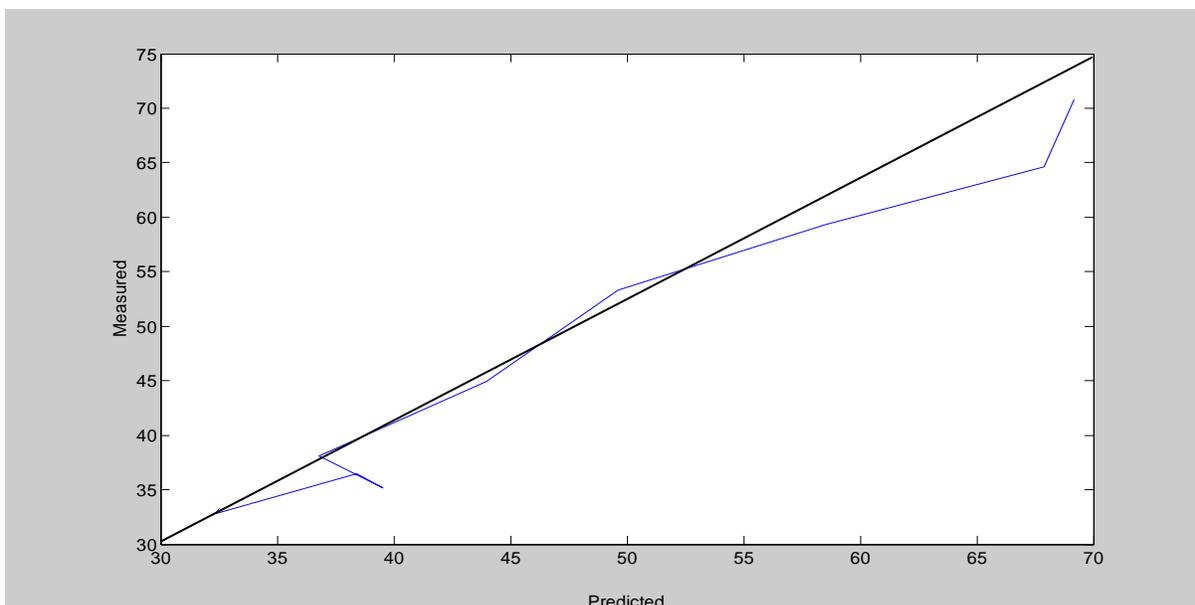


Fig.35A. Plot of predicted composition against measured composition for the reactant  
(Pearson's  $r = 0.9849$ )

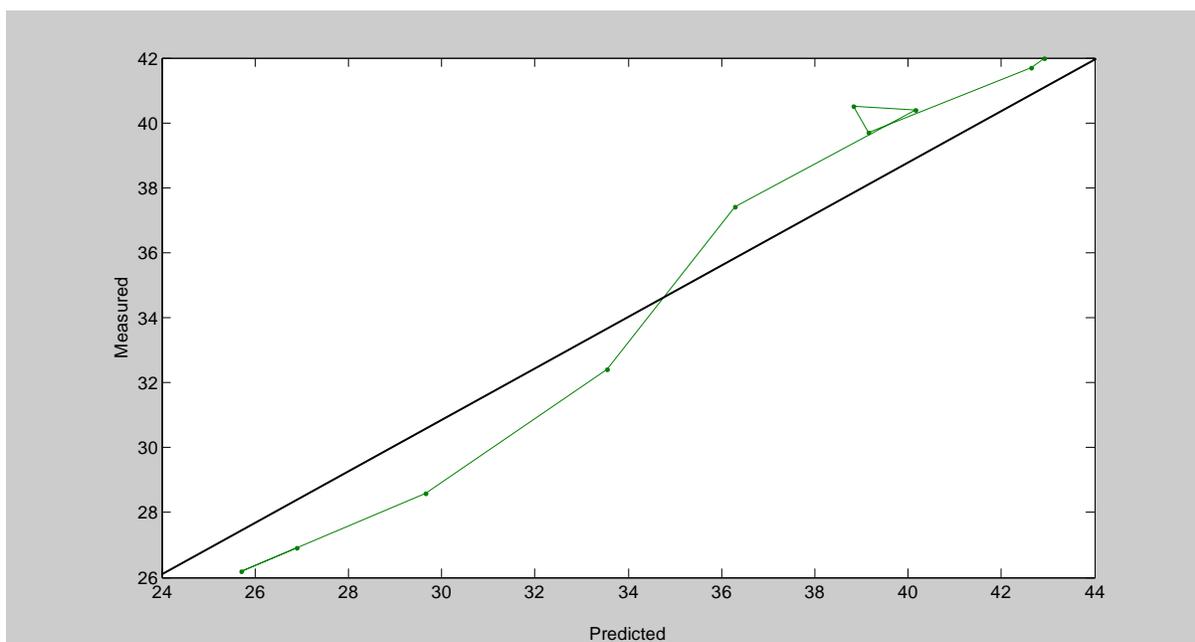
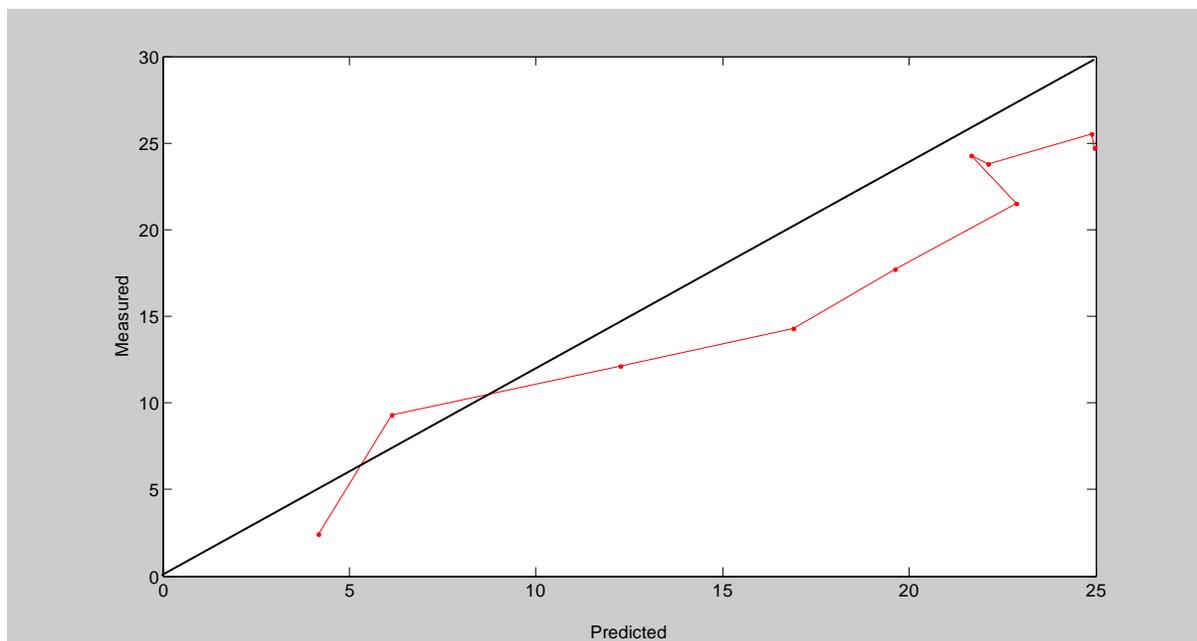


Fig.35B. Plot of predicted composition against measured composition for the product (Pearson's  
 $r = 0.9880$ )



*Fig.35C. Plot of predicted composition against measured composition for the byproduct (Pearson's  $r = 0.9672$ )*

### NIR model validation

For the validation process another ten samples were taken as new set of X-block variables. Three samples from 15<sup>th</sup> to 30<sup>th</sup> minute range, four samples from 75<sup>th</sup> to 90<sup>th</sup> minute range and three samples from 135<sup>th</sup> to 150<sup>th</sup> minute range in the reaction time were selected. Then they were subjected to the model and their respective % composition values were compared with the prediction made for the first set of samples. This comparison is shown in table 8.

Reaction time (min)	Estimated % Composition by the model		
	Reactant	Product	Byproduct
17	73.1	23.8	2.1
22	74.4	22.6	2.1
27	68.3	25.6	5.7
77	39.3	38.8	21.7
81	38.5	39.3	22.1
85	38.5	39.2	22.1

88	42.5	36.7	20.0
137	37.7	39.8	22.6
142	34.7	41.1	23.9
147	36.8	39.9	22.7

*Table.8. Prediction of composition using NIR PLS- model for new X-block variables*

## 5. Conclusion

Raman and NIR spectroscopy has been proven to be effective methods for online reaction monitoring of synthesis of 10-Bromo-1-Decandiol. GC-MS analysis of aliquots of the reaction mixture at different intervals of time during the reaction confirmed to be consistent with the result obtained from principal component analysis for Raman and NIR data. Moreover, PCA investigation of the score plots in the Raman and NIR data revealed that the reaction time can be shortened considerably without affecting the yield of the product.

MCR-ALS method was successfully employed in identifying pure component spectra of the reactant, product and byproduct. PLS- regression models for Raman and NIR data made very good predictions and the correlation between the predicted and observed values for all components involved in the reaction have been very good. Therefore, both Raman and NIR spectroscopy equipped with multivariate data analysis techniques proved to be adequate techniques to monitor the synthesis of 10-Bromo-1-Decandiol.

## 6. Suggestion for future work

For future studies on reaction monitoring using Raman and NIR spectroscopy, I suggest the following points to be taken into consideration. In order to avoid complications during data pretreatment and interpretation, the spectra recorded by Raman and NIR spectroscopy should be carried out at equal intervals of time. This leads to equal numbers of spectra that are more convenient to deal with.

For a better data pretreatment, particularly for background correction of spectral data, addition of Raman and NIR active internal standard to the reaction mixture is advisable. For reactions carried out in a solvent, the solvent itself can be used as an internal standard but for reactions carried out without solvent (like the reaction monitored in this thesis) selection and addition of appropriate internal standard to the reaction mixture will assist the background correction step of data pretreatment.

## 7. References

- [1] N. A. Di Prospero, A. Baker, N. Jeffries, K. H. Fischbeck, *The Lancet Neurol* 6 (2007) 878–886
- [2] D. H. McDaniel, B.A. Neudecker, J.C. DiNardo, J.A. Lewis II, H.I. Maibach, *Journal of Cosmetic Dermatology*, 4(2005)167–173
- [3] K. Schaffler, D. Hadler, M. Stark, *Arzneimittel-Forschung* 48 (1998) 720–726
- [4] R. Spaccini, A. Tsoukala, L. Liguori, C. Punta, H. R. Bjørsvik, *Org. Process Res. Dev.* 14 (2010) 1215–1220
- [5] E. Smith, G. Dent, *Modern Raman Spectroscopy-A Practical Approach*, Wiley, West Sussex, 2005
- [6] Z.Q.. Wen, *J. Pharm. Sci.* 96(2007) 2861-2878
- [7] D. A. Long, *The Raman effect: a unified treatment of the theory of Raman scattering by molecules*, Wiley, West Sussex, 2002.
- [8] M.Schmitt, J.Popp, *J. Raman Spec.* 37(2006) 20-28
- [9] R. Petry, M. Schmitt, J. Popp, *Chemphyschem* 4(2003) 14-30
- [10] T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J.P. Remon, C. Vervaet, *International Journal of Pharmaceutics* xxx (2010) xxx–xx
- [11] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jen, *Journal of Pharmaceutical and Biomedical Analysis* 44(2007) 683-700
- [12] T. Hasegawa, J. Nishijo, J. Umemura, *Chemical Physics Letters* 317(2000) 642–646
- [13] D. A. Burns, E. W. Ciurczak, *Handbook of Near-Infrared Analysis*, Marcel Decker, New York, 1992
- [14] R. Raghavachari, *Near-Infrared applications in biotechnology*, Marcel Decker, New York, 2001
- [15] G. Reich, *Advanced drug delivery reviews* 57(2005) 1109-1143
- [16] A. E. Cervera, N. Petersen, A. E. Lantz, A. Larsen, K. V. Gernaey, *Biotechnol. Prog.* 25(2009) 1561-1580
- [17] F. West ad, N.K. Afseth, R. Bro, *Analytica Chimica Acta* 595(2007) 323–327

- [18] B. M Wise, N. B. Gallagher, R. Bro, J. M. Shaver, W. Windig, R. S. Koch, *PLS\_Toolbox 4.0 for use with Matlab*, Eigenvector Research inc, Wenatchee, WA, USA, 2006
- [19] R. G. Brereton, *Chemometrics data analysis for the laboratory and chemical plant*, Wiley, West Sussex, England, 2003
- [20] L.I. Smith, *A tutorial on principal component analysis*, [http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf), 2002
- [21] V. del Río, M. P. Callao, M. S. Larrechi, L. M. de Espinosa, J.C. Ronda, V. Cádiz, *Anal. Chim. Acta.* 642(2009) 148–154
- [22] B. Kowalski, A. Smilde, R. Tauler, *J. Chemom.* 9(1995) 31-58
- [23] M. Shamsipur, B. Hemmateenejad, A. Babaei, L. F. Sharabiani, *J. elec. Anal. Chem.* 570(2004) 227-234
- [24] A. De Juan, S. C. Rutan, R. Tauler, D. L. Massart, *Chemom. Intell. Lab. Syst.* 40(1998) 19-32
- [25] M. Garrido, F. X. Rius, M. S. Larrechi, *Anal. Bioanal. Chem.* 390(2008)2059–2066
- [26] R. Tauler, *Multivariate curve resolution: theory and applications*, [http://www.ub.edu/mcr/web\\_mcr/download/pdf/mcr\\_2005.pdf](http://www.ub.edu/mcr/web_mcr/download/pdf/mcr_2005.pdf), 2005
- [27] [http://www.eigenvector.com/Docs/Wise\\_pls\\_properties.pdf](http://www.eigenvector.com/Docs/Wise_pls_properties.pdf)
- [28] R. D. Tobias, *An introduction to partial least squares regression*, <http://www.ats.ucla.edu/stat/sas/library/pls.pdf>, 1999
- [29] P. Geladi, B. Kowalski, *Anal. Chim. Acta.* 185(1986) 1-17
- [30] S. Devries, C. J. F. Ter Braak, *Chemom. Intell. Lab. Syst.* 30(1995) 239-245
- [31] V.E. Vinzi, W.W. Chin, J. Henseler, H. Wang, *Handbook of Partial Least Squares: Concepts, Methods and Applications*, Springer-verlag Berlin Heidelberg, 2010
- [32] O. Svensson, M. Josefson, F. W. Langkilde, *Chemom. Intell. Lab. Syst.* 49(1999) 49–66
- [33] U. B. Cappel, I. M. Bell, L. K. Pickar, *Society for applied spectroscopy* 64(2010) 195-200
- [34] G. R. Phillips, J. M. Harris, *Anal. Chem.* 62(1990) 2351-2357
- [35] O. Svensson, M. Josefson, F. W. Langkilde, *Chemom. Intell. Lab. Syst.* 49(1999) 49-66
- [36] O. Svensson, M. Josefson, F. W. Langkilde, *European Journal of Pharmaceutical Sciences* 11(2000) 141-155

- [37] S. Romero-Torres, J. D. Pérez-Ramos, K. R. Morris, E. R. Grant, *Journal of Pharmaceutical and Biomedical Analysis* 41(2006) 811-819
- [38] <http://www.mathworks.com/help/toolbox/bioinfo/ref/msbackadj.html>
- [39] G. Socrates, *Infrared and Raman characteristic group frequencies: tables and charts*, Wiley, West Sussex, 2006
- [40] H.W. Siesler, Y. Ozaki, S. Kawata, H. M. Heise, *Near-infrared spectroscopy: principles, instruments, applications*, Wiley-VCH, Weinheim, Germany, 2002
- [41] R. L. Grob, E. F. Barry, *Modern practice of Gas-Chromatography*, Wiley, Hoboken, New Jersey, USA, 2004