

Multivariate calibration of CO<sub>2</sub> Solubility in Diethanolamine (DEA) using Raman  
Spectroscopy

by

Nuralia Syairah Binti Osman

15669

Dissertation submitted in partial fulfilment of  
the requirements for the  
Bachelor of Engineering (Hons)  
(Chemical Engineering)

SEPTEMBER 2015

Universiti Teknologi PETRONAS  
32610, Bandar Seri Iskandar  
Perak

# CERTIFICATION OF APPROVAL

## **Multivariate calibration of CO<sub>2</sub> Solubility in Diethanolamine (DEA) using Raman Spectroscopy**

by

Nuralia Syairah Binti Osman

15669

A project dissertation submitted to the  
Chemical Engineering Programme  
Universiti Teknologi PETRONAS  
in partial fulfilment of the requirement for the  
BACHELOR OF ENGINEERING (Hons)  
(CHEMICAL ENGINEERING)

Approved by,

---

(Ir. Dr. Abd Halim Shah Bin Maulud)

UNIVERSITI TEKNOLOGI PETRONAS  
BANDAR SERI ISKANDAR, PERAK

September 2015

## CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the reference and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

---

NURALIA SYAIRAH BINTI OSMAN

## ABSTRACT

The emission of greenhouse gas which is mainly from carbon dioxide (CO<sub>2</sub>) have cause phenomena of global warming and climate changes. Thus, this issues has become main focus in worldwide nowadays. CO<sub>2</sub> removal process also is an essential step in many industrial processing especially in natural gas processing. Absorption based on alkanolamines like diethanolamine (DEA) have been widely applied in carbon capture plant. CO<sub>2</sub> solubility has become an important parameter in the absorption process for proper process control in plant operation. Measurement of CO<sub>2</sub> solubility in 10%, 20% and 30% DEA using Raman spectroscopy is the main focus in this research. Raman shift and intensity of different DEA concentration with respect to CO<sub>2</sub> loading (mol CO<sub>2</sub> / mol amine) are obtained by using Raman spectroscopy. Partial Least Square (PLS) regression approach will be utilized to develop a calibration model that relate the Raman spectroscopy data to CO<sub>2</sub> solubility in different concentration of DEA. From the result obtained, 10%, 20% and 30% model have shown great performance and demonstrate good prediction of CO<sub>2</sub> solubility. In addition, a combined concentration model and combined modified model is developed to predict CO<sub>2</sub> solubility at different DEA concentration. From the study, with the inverse of 900 – 1100 cm<sup>-1</sup> are added to predicting matrix, the validation R<sup>2</sup> have been increased from 0.9004 to 0.9136. Combination of analytical instrument and multivariate calibration tools will aid the process of online monitoring CO<sub>2</sub> solubility in DEA in process plant operation.

## **ACKNOWLEDGEMENTS**

First and foremost, I would like to express my praises to God for His blessing in completing this thesis. Special appreciation and sincere gratitude is extended to my supervisor, Ir Dr Abdul Halim Shah Maulud for his patience, motivation, support, and immense knowledge throughout the Final Year Project in order to complete my project. His invaluable guidance helped me in all the time of research and writing of this thesis. Next, my sincere appreciation to Muhammad Zubair Shahid, postgraduate student who helped a lot to understand and do this project. Special thanks to all my fellow coursemates also other lecturers who have always been accommodating and cooperative for their kindness in giving of ideas, opinion and moral support throughout the completion of this project report. I appreciate all parties that stay by my side throughout completing this project. Last but not least, I would like to acknowledge my family members for supporting me motivated and spiritually throughout the year.

## TABLE OF CONTENT

<b>CERTIFICATION OF APPROVAL</b>	<b>ii</b>
<b>CERTIFICATION OF ORIGINALITY</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>ACKNOWLEDGEMENTS</b>	<b>v</b>
<b>TABLE OF CONTENT</b>	<b>vi</b>
<b>LIST OF FIGURES</b>	<b>viii</b>
<b>LIST OF TABLE</b>	<b>xi</b>
<b>CHAPTER 1 INTRODUCTION</b>	<b>1</b>
1.1 Background Study	1
1.2 Problem Statement	3
1.3 Objective and Scope of Study	4
<b>CHAPTER 2 LITERATURE REVIEW</b>	<b>5</b>
2.1 Carbon Dioxide	5
2.2 Carbon Dioxide Removal Technique	7
2.3 Alkanolamines used as solvent in absorption	8
2.4 Diethanolamine (DEA) mechanism	9
2.5 Spectroscopic Technique for online measurement	11
2.6 Raman Spectroscopy	12
2.7 Raman Scattering	13
2.8 Multivariate Calibration	14
2.8.1 Multi Linear Regression (MLR)	15
2.8.2 Principle Component Analysis (PCA)	16
2.9 Partial Least Square (PLS) Regression	17
2.9.1 PLSR model	17
2.9.2 Model Evaluation	20
<b>CHAPTER 3 METHODOLOGY</b>	<b>21</b>
3.1 Project Methodology	21
3.2 Process Flow Chart	23
3.3 Key Milestones	25

	3.4	Gantt chart	26
<b>CHAPTER 4</b>		<b>RESULT AND DISCUSSION</b>	<b>27</b>
	4.1	Qualitative Analysis of Raman	27
	4.2	Partial Least Square Regression	28
	4.2.1	Model for 10% of DEA (D10)	28
	4.2.2	Model for 20% of DEA (D20)	30
	4.2.3	Model for 30% of DEA (D30)	31
	4.2.4	Combined Concentration Model (DC1)	32
	4.2.5	Combined Concentration Model (DC2)	34
	4.2.6	Modified Combined Model (DC3)	36
<b>CHAPTER 5</b>		<b>CONCLUSION</b>	<b>39</b>
		<b>REFERENCES</b>	<b>39</b>
		<b>APPENDICES</b>	<b>41</b>

## LIST OF FIGURES

Figure 2.1	Pressure-Temperature Phase Diagram	6
Figure 2.2	Typical Process Flow Diagram (PFD) of CO <sub>2</sub> Removing Process	7
Figure 2.3	Structural Formula of alkanolamines used in gas treating	8
Figure 2.4	CO <sub>2</sub> loading vs rate of absorbent	9
Figure 2.5	Chemical Structure of DEA	9
Figure 2.6	3 Different Forms of Scattering	14
Figure 2.7	Illustration of principle component analysis (PCA)	16
Figure 3.1	CO <sub>2</sub> Absorption Cell	22
Figure 3.2	Process Flow Chart	24
Figure 4.1	Raman Spectrum of DEA	27
Figure 4.2	Graph of actual and predicted CO <sub>2</sub> loading in calibration model	28
Figure 4.3	Graph of actual and predicted CO <sub>2</sub> loading in validation model	29
Figure.4.4	Graph of actual and predicted CO <sub>2</sub> loading in calibration model	30
Figure 4.5	Graph of actual and predicted CO <sub>2</sub> loading in validation model	30
Figure 4.6	Graph of actual and predicted CO <sub>2</sub> loading in calibration model	31
Figure 4.7	Graph of actual and predicted CO <sub>2</sub> loading in validation model	32
Figure 4.8	Graph of actual and predicted CO <sub>2</sub> loading in calibration model	33
Figure 4.9	Graph of actual and predicted CO <sub>2</sub> loading in validation model	33
Figure 4.10	Graph of actual and predicted CO <sub>2</sub> loading in calibration model	35
Figure 4.11	Graph of actual and predicted CO <sub>2</sub> loading in validation model	35
Figure 4.12	Graph of actual and predicted CO <sub>2</sub> loading in calibration model	38
Figure 4.13	Graph of actual and predicted CO <sub>2</sub> loading in validation model	38

## LIST OF TABLE

Table 2.1	Physical Properties of DEA.....	11
Table 2.2	PLS function in MATLAB.....	18
Table 4.1	Calibration and Validation Results in D10 .....	29
Table 4.2	Calibration and Validation Results in D20 .....	31
Table 4.3	Calibration and Validation Results in D30 .....	32
Table 4.4	Calibration and Validation Results in DC1.....	34
Table 4.5	Calibration and Validation Results in DC2.....	36
Table 4.6	Calibration and Validation Result in DC3 .....	37

## CHAPTER 1

### INTRODUCTION

#### 1.1 Background Study

Phenomenon of global warming and climate change has been recognized as a significant environmental problem by the worldwide scientific community [1]. In recent years, world energy supply is mainly dependent on fossil fuels combustion, accounting 80% of total energy consumed in all over the world. Due to global population growth, the world energy demand was expected to be increased and fossil fuels combustion is responsible for large amounts of CO<sub>2</sub> and other greenhouse gases [2]. According to the Climate Change 2014 synthesis report, since the pre-industrial era have started due to economic and population growth , greenhouse gas emissions have increased year by year and from 2000 to 2010 was reported to be the highest in history. The period from 1983 to 2012 was the warmest of 30-year period of the last 800 years and caused the risen of sea level as well as diminished of snow and ice [3]. Thus, great attention have been focus on the reduction of CO<sub>2</sub> emissions in the atmosphere.

Besides known as greenhouse gas, CO<sub>2</sub> also is an unwanted by-products in natural gas processing industries due to its properties as an acid gas. When CO<sub>2</sub> react with water, it will form carbonic acid that tend to cause corrosion and give major effects and damage to the oil field equipment and gas pipelines. Formation of carbonic acid and moisture will decrease pipeline flow capacities, even resulting in blockages, and potential harm to valves, filters and compressors that are being used throughout the process. Thus, CO<sub>2</sub> removal is essential process in order to increase heating value of natural gas and optimize pipeline capacity [4]

Various technologies have been developed to remove CO<sub>2</sub> from natural gas for instance absorption by chemical and physical solvents, cryogenic separation and membrane separation [5]. However, most attention have been focus on absorption and membrane separation due to its existing and maturity level in the process. Comparison of the technologies available for CO<sub>2</sub> capture have been done and conclude that absorption with aqueous amine solution is most favoured one.

Packed column are commonly used for gas absorption in gas processing industry. The efficiency of contact between gases and liquid can be increased due to high mass transfer efficiency and low pressure drop. Various research have been carried out in order to further investigate on efficiency of packed column in CO<sub>2</sub> removal. There are a lot of factors that affect the efficiency of packed column for instance CO<sub>2</sub> solubility in solution. CO<sub>2</sub> solubility give the strong influence to the mass transfer coefficient [6]. When the CO<sub>2</sub> loading in the lean solution was high, the driving force of mass transfer from gas to solution became small and the decrease in CO<sub>2</sub> removal efficiency could not be compensated even with the increase of solution flow rate [6]. The construction cost of entire process could also be reduced if the absorbent has high CO<sub>2</sub> capacity [1]. Increases of CO<sub>2</sub> loading in solvent will significantly lowered the availability of active absorbent concentration in the solution. Therefore, the solubility of CO<sub>2</sub> in aqueous amine solution has become one of the important parameter for the CO<sub>2</sub> absorption process.

Several amine solvents which are monoethanolamine (MEA), diethanolmine (DEA) and methyldiethanolamine (MDEA) have been mostly investigated and studied by the various researcher to determine CO<sub>2</sub> solubility at various conditions. Thermodynamics model such as Kent-Eisenberg model, modified Kent-Eisenberg (M-KE) [7] was developed based on vapor-liquid equilibrium (VLE). However, this model have some limitations on the accuracy and only suitable for certain amine solutions. Thus, empirical modelling technique was introduced to further investigated CO<sub>2</sub> solubility in amine solvent. Time is significantly reduced as well as it less complex compared to mathematic approach. Among the empirical models is based on partial least square (PLS) regression approach. PLS regression is one of the

powerful tools to predict dependent variable from a very large dataset of independent variable [8].

## 1.2 Problem Statement

During the absorption process, CO<sub>2</sub> loading measurement is a vital action for a proper process control. The earliest method to measure CO<sub>2</sub> loading in alkanolamines is titration. However, this method is quite tedious as the sample need to be prepared and the analysis cannot be done during the process [4]. It is not practical to perform experiments for each specific temperature, pressure and concentration and obtain CO<sub>2</sub> loading in aqueous solution. Besides that, it required a long time, hence it is not appropriate for process control.

Many studies have been conducted in order to find most suitable and efficient tools to measure CO<sub>2</sub> loading instantaneously. Infrared, NMR and Raman spectroscopy are some of the spectroscopic technique that have been used for instantaneous measurement of gas-liquid concentration. Raman spectroscopy is a powerful technique to be used for monitoring CO<sub>2</sub> loading in aqueous solution. It is a non-invasive, non-destructive and direct measurement method efficient for quantitative analysis for concentration as low as 0.1 up to 100%. It is relatively fast, suitable for measurements in aqueous system because of the weak Raman scattering of water molecules and application [4]

Raman spectroscopy are widely used in pharmaceutical and cosmetics industry as Raman has the ability to carry direct measurement on solids. The ability to identify raw material, determine active substances in different formulations and can support polymorphic screenings give advantages to pharmaceutical analysis. In this project, measurement of CO<sub>2</sub> loading in DEA using Raman spectroscopy is the main focus. To achieve this focus, Partial Least Square Regression (PLSR) approach will be utilized to develop a calibration model that relate the Raman spectroscopy to CO<sub>2</sub> loading. PLSR method have been successfully applied in many areas including multivariate statistics, analytical chemistry, medicine, face recognition etc.

### **1.3 Objective and Scope of Study**

- 1) To obtain the CO<sub>2</sub> solubility data in DEA with their respective Raman spectrum.
- 2) To develop an accurate multivariate calibration model based on PLSR to predict CO<sub>2</sub> solubility in aqueous solution of DEA using Raman spectroscopy for fixed DEA concentration.
- 3) To extend the calibration model to predict CO<sub>2</sub> solubility for different concentration of DEA.

The first phase of the research is focus on obtaining the CO<sub>2</sub> solubility data in DEA aqueous solution using Raman spectroscopy. Different concentration of DEA will be tested out to observe the CO<sub>2</sub> solubility. Then, the next phase is to focus on the development of calibration model based on PLSR in MATLAB software. This model will be used to predict CO<sub>2</sub> solubility in a fix DEA aqueous solution. In addition, the calibration model will be extended for different concentration of DEA.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Carbon Dioxide

Carbon dioxide ( $\text{CO}_2$ ) is a chemical molecule consist of one carbon atom covalently bonded with two oxygen atoms. It is a colourless, odourless, non-toxic gas and heavier than air. It is soluble in water, ethanol and acetone. The combustion of coal or hydrocarbon will produced  $\text{CO}_2$ . Besides that,  $\text{CO}_2$  is produced by the fermentation of liquids and the breathing of humans and animals. Found in small proportions in the atmosphere, it is assimilated by plants which in turn to produce oxygen.

Figure 2.1 shows the phase diagram of carbon dioxide.  $\text{CO}_2$  will not exist in liquid state at atmospheric pressure at any temperature. The liquid  $\text{CO}_2$  exists only after  $56.6^\circ\text{C}$  and have the lowest pressure is at 5.11 atm. The triple point is where liquid, gas and solid exist simultaneously in thermodynamic equilibrium. At the sublimation point,  $\text{CO}_2$  able to change from solid state directly into a gas. Physically, this boundary implies that the gas and solid can co-exist and transform back and forth without the presence of liquid as an intermediate phase at  $-78.5^\circ\text{C}$ .  $\text{CO}_2$  freezes at  $-78.5^\circ\text{C}$  to form carbon dioxide snow or dry ice. Above the critical point (73.8 atm and  $31.1^\circ\text{C}$ ), the liquid and gas phases cannot exist as separate phases, and liquid phase  $\text{CO}_2$  develops supercritical properties, where it has some characteristics of a gas and others of a liquid.

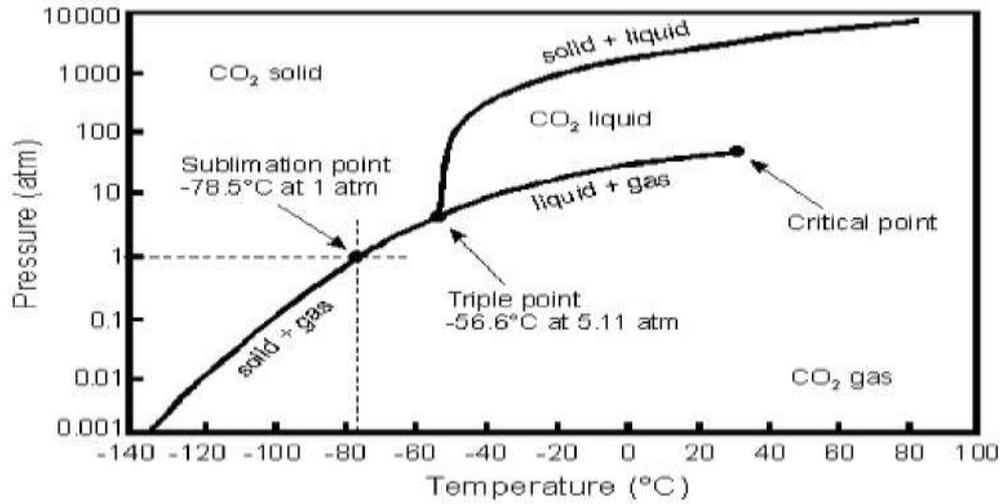


Figure 2.1 Pressure-Temperature Phase Diagram

CO<sub>2</sub> is widely used in many industries such as chemical, pharmaceutical, food and beverages industry. CO<sub>2</sub> is commercially used as in fire extinguishers, carbonate beverages and as a refrigerant [9]. Large quantities of solid carbon dioxide which in the form of dry ice are used in processes requiring large scale refrigeration. Besides that, CO<sub>2</sub> is used for environment protection. For instance, CO<sub>2</sub> is used for red fume suppression during scrap and carbon charging. It is also commonly used as the compressed gases for pneumatic (pressurized gas) systems in portable pressure tools.

Regardless of its wide applications, CO<sub>2</sub> is widely recognized as a major component of greenhouse gas contributing to global warming. It also known as acid gas that need to be removed from natural gas to avoid corrosion, blockages to the pipeline due to the formation of CO<sub>2</sub> in the low temperature system as well as to maintain the heating value of natural gas.

## 2.2 Carbon Dioxide Removal Technique

Removal of acid gas is a common unit process used in refineries, petrochemical plants, and also other industries to remove the contaminant in natural gas. The process of removing CO<sub>2</sub> and H<sub>2</sub>S from acid gas is called sweetening process as it removes the sour odour of the gas due to the presence of H<sub>2</sub>S. Chemical absorption by a solvent is the technique most commonly used to remove acid gas in natural gas flow. Absorption allows acidic gases to be dissolve in a solvent and will be released by regeneration.

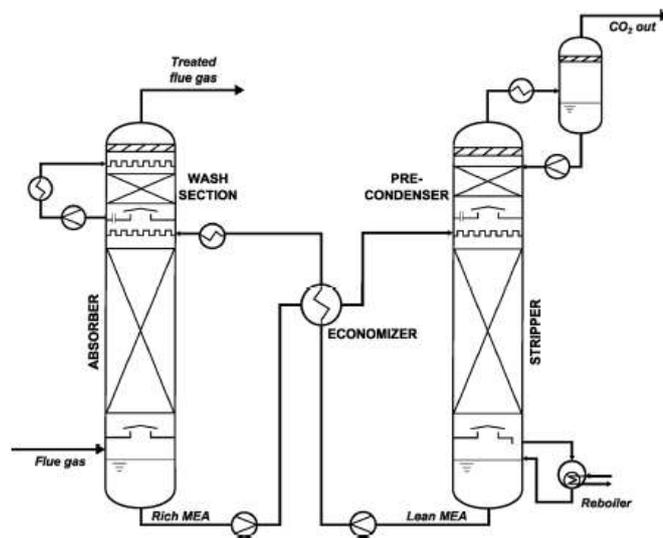


Figure 2.2 Typical Process Flow Diagram (PFD) of CO<sub>2</sub> Removing Process

A typical amine gas treating process consists of an absorber unit, a stripper unit and accessory equipment. In the absorber, flue gas CO<sub>2</sub> entered the bottom of absorber column and the 'lean' amine solution will absorb the flue gas counter-currently to produce sweet gas as a product [10]. Sweetened gas leaving the top of the absorber passes through an outlet separator and then flows to a dehydration unit (and compression unit, if necessary) before being considered ready for sale. Now, the 'rich' amine solution containing the absorbed acid gases is further heated in the regeneration column. Steam and acid gases separated from the rich amine are condensed and cooled.

### 2.3 Alkanolamines used as solvent in absorption

Alkanolamines are widely used in many application. They are versatile, polyfunctional molecule that combine the characteristics of amines and alcohols. Each alkanolamine has at least one hydroxyl group and one amino group. The hydroxyl group reduces the vapour pressure and increases the water solubility, while the amino group increases the pH of the solution which causes reaction with acidic gases such as CO<sub>2</sub> and H<sub>2</sub>S.

Alkanolamines can be classified as primary, secondary, or tertiary depending on the number of alkyl groups attached to the nitrogen atom. In primary amines, one hydrogen atom on the nitrogen is replaced with a functional group while in secondary amines two hydrogen atoms are replaced followed by all three atoms are replaced in tertiary amines.

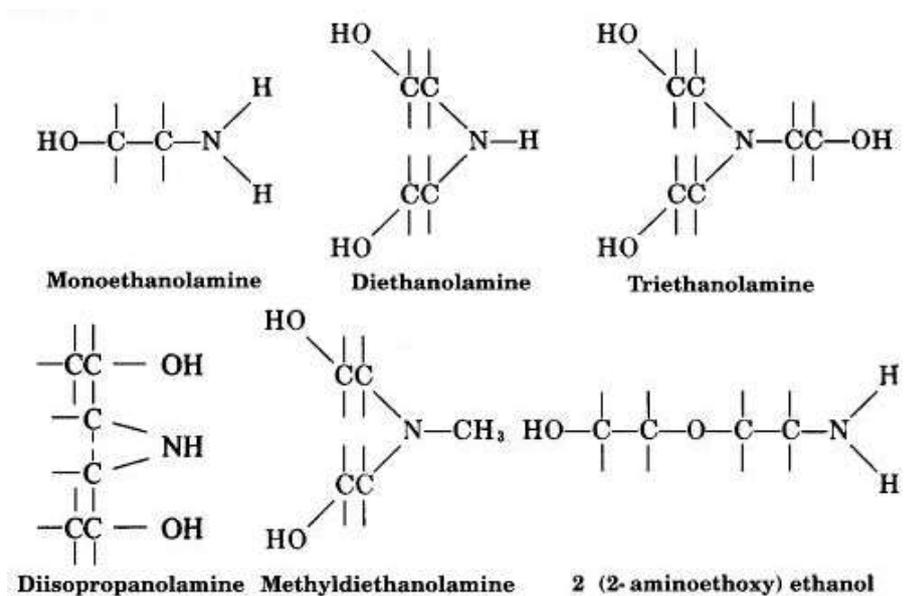


Figure 2.3 Structural Formula of alkanolamines used in gas treating

Different types of amines will have different properties when react with CO<sub>2</sub>. Most of the amines exist in liquid state in room temperature. In term of solubility, they are soluble in water. Their solubility will decrease as the hydrogen chain attached to the molecule getting longer. Both primary and secondary amines exhibit low CO<sub>2</sub> loading but they have high rate of absorption. However, tertiary amines showed

opposite behaviour. The selection of amine in CO<sub>2</sub> removal system not only depend on the absorption rate and maximum loadings achieved, but other factors such as regeneration energy, corrosion tendency and cost of solvent also are considered [7]

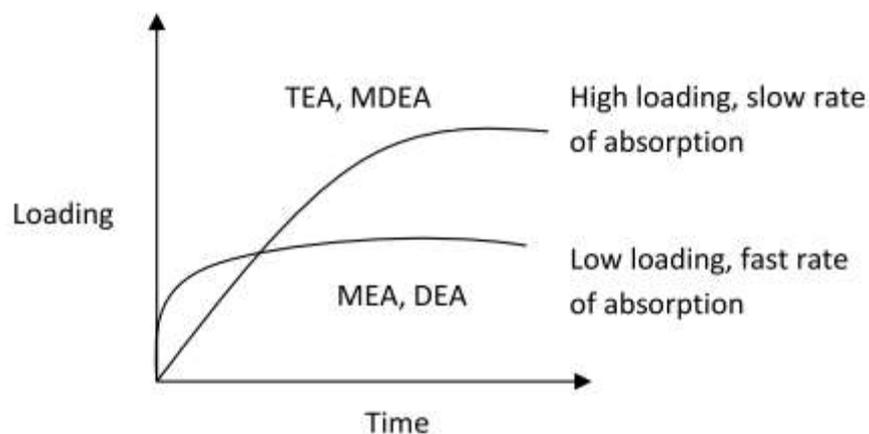


Figure 2.4 CO<sub>2</sub> loading vs rate of absorbent

#### 2.4 Diethanolamine (DEA) mechanism

Diethanolamine (DEA) is an organic compound with the molecular formula NH(CH<sub>2</sub>CH<sub>2</sub>OH)<sub>2</sub>. This colorless liquid is a secondary amine where two of the hydrogen group will be replaced by an alkyl group (C<sub>2</sub>H<sub>4</sub>OH). Aqueous solutions of diethanolamine (DEA) have been used for many years in gas treatment at oil refinery.

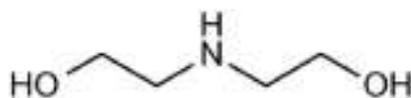


Figure 2.5 Chemical Structure of DEA

During the absorption of CO<sub>2</sub> in the DEA, carbonate (CO<sub>3</sub><sup>2-</sup>), bicarbonate (HCO<sub>3</sub><sup>-</sup>) and carbamate (R<sub>1</sub>R<sub>2</sub>NCOO<sup>-</sup>) will formed as the final product along the reaction. The equation of chemical reaction that take place can be written as below [11], [12]

Water ionization:



Dissociation of Carbon Dioxide:



Dissociation of Bicarbonate:



Carbamate Formation:



Overall Equation:



All the reaction occur in this process is reversible. When  $\text{CO}_2$  react with water, bicarbonate will be formed in the equation (2). Then in equation (3), bicarbonate react with water and formed carbonate. The reaction of DEA with bicarbonate will form carbamate as the final product. Equation (5) shows the overall reaction occurred in the process of absorption  $\text{CO}_2$ . All absorbed  $\text{CO}_2$  react with DEA to form carbamate. The formation of the carbamate is more stable compared to bicarbonate. Therefore, absorbents that formed carbamate required larger quantities of heat [1].

It has an advantage over a similar amine ethanolamine in that a higher concentration may be used for the same corrosion potential. It is less reactive than primary amines and make the reaction product less corrosive. This allows refiners to scrub acid gas at a lower circulating amine rate with less overall energy usage. The low vapour pressure of diethanolamine makes it suitable for low-pressure operations as vaporization losses are quite negligible [13]. Besides that, DEA has high reactivity and fast absorption rate.

The solvent apply (DEA) is considered to be chemically stable; DEA can be heated to its normal boiling point (269 °C at 760mmHg) before decomposition. Therefore reduce the solvent degradation during stripping and reduce solvent loss and

accumulation in the units. The heat of reaction of DEA with CO<sub>2</sub> is low compared to other amines hence the heat generated in the absorber during CO<sub>2</sub> absorption process is low which increases the solvent loading capacity in the absorber as solubility or loading of CO<sub>2</sub> increases at low temperature. It react rapidly with CO<sub>2</sub> to form carbamate with a stoichiometric loading of 0.5 mol of CO<sub>2</sub> per mole of amine [13]

Table 2.1 Physical Properties of DEA

Molecular formula	C <sub>4</sub> H <sub>11</sub> NO <sub>2</sub>
Molecular Weight	105.14 g/mol
Boiling Point	268.8 °C
Melting Point	28 °C
Flash Point	152 °C (306 oF) open cup
Vapor Density	3.65 (air = 1)
Density/Specific Gravity	1.09664 at 20/4 °C (water = 1)
Vapor Pressure	2.8 x 10 <sup>-4</sup> mm Hg at 25 °C
Log Octanol/Water Partition Coefficient	-1.43
Henry's Law Constant	5.35 x 10 <sup>-14</sup> atm-m <sup>3</sup> /mole at 25 °C
Conversion Factor	1 ppm = 4.3 mg/m <sup>3</sup>

## 2.5 Spectroscopic Technique for online measurement

Spectroscopy is a technique that use the interaction of energy with a sample to perform analysis. The data that obtained from spectroscopy is called a spectrum. A spectrum is a plot of the intensity of energy detected versus the wavelength, amplitude, frequency, etc. Information about atomic, molecular energy levels, interaction of molecules and other related process can be obtained via spectrum. It also can be used to identify the components of a sample (qualitative analysis) and measured the amount of material in a sample (quantitative analysis).

There are several instruments that are used to perform a spectroscopic analysis. As spectroscopy requires an energy source, laser, ion source or radiation source could be used. A spectrophotometer or interferometer are usually used as a device to measure the change in the energy source after it have been interacted with the sample. Several type of spectroscopy that are commonly used for analysis of material and chemicals such as X-Ray photoelectron spectroscopy (XPS), Ultraviolet and visible absorption spectroscopy (UV-VIS), Infrared absorption spectroscopy (IRS), Nuclear magnetic resonance spectroscopy (NMR) and Raman spectroscopy.

Infrared absorption spectroscopy is frequently used to identify materials and to measure the number of absorbing molecules. Recently, NIR- spectroscopy and Raman spectroscopy has been used for bioprocess monitoring. Both techniques is based on vibrational effects and suitable for measurement of small molecule in a chemical compounds. In addition, fluorescence spectroscopy also have been applied for monitoring several of biological processes. It was used to predict biomass, product, process and media characterization in biotechnology field especially in pharmaceutical and food process engineering [14].

## **2.6 Raman Spectroscopy**

Raman spectroscopy is a one of the spectroscopic technique that are available for analysis of material and chemicals. It also used to study the chemical components in gas, liquid or solid state phases through the vibration or rotation of a molecule. Raman spectroscopy is commonly used to characterize chemical components by providing a fingerprint by which the molecule can be identified.

It was based on Raman scattering of light by a material, where the light is scattered inelastically as opposed to the more prominent elastic Rayleigh scattering. Inelastic scattering means that the frequency of photons in monochromatic light changes upon interaction with a sample. This inelastic scattering causes shifts in wavelength, which can be used to figure out information about the material. Properties

of the material can be determined by analysis of the spectrum, or compared with a library of known spectra to identify a substance.

Raman spectroscopy is a non-destructive process [4], and can be used to monitor industrial processes. The speed of analysis means that it can give almost real-time information. Another advantage is that the light to be monitored can be sent down fibre-optics, so that the Raman equipment can be located some distance away from the actual processing. Raman spectroscopy is now an extremely powerful technique with many applications since the discovery of Raman scattering in the 1920s.

Raman spectroscopy was widely used in the pharmaceutical analysis. The ability of Raman spectroscopy to carry out direct measurement on solids give benefit to pharmaceutical and polymer industry [15]. Besides that, it also have been used in protein structural characterization especially in determining the changes in the secondary and primary structure of protein in different matrices [16]

## **2.7 Raman Scattering**

Scattered light in Raman consists of two types which are Rayleigh scattering and Raman scattering (stokes shift). Rayleigh scattering is an elastic effect, which means that the light does not gain or lose energy during the scattering. The amount of scattering is strongly dependent on the wavelength, being proportional to  $\lambda^{-4}$ . A photon interacts with a molecule, polarising the electron cloud and raising it to a “virtual” energy state and drops back down to its ground state, releasing a photon. Since the molecule is dropping back to the same state it started in, the energy released in the photon must be the same as the energy from the initial photon. Therefore the scattered light has the same wavelength.

Raman scattering is different in that it is inelastic. The light photons lose or gain energy during the scattering process, and therefore increase or decrease in wavelength respectively. If the molecule is promoted from a ground to a virtual state and then

drops back down to a vibrational state then the scattered photon has less energy than the incident photon, and therefore a longer wavelength. This is also called Stokes scattering. Anti-Stokes scattering occur when molecule is in a vibrational state and after scattering is in its ground state then the scattered photon has more energy, and therefore it has a shorter wavelength.

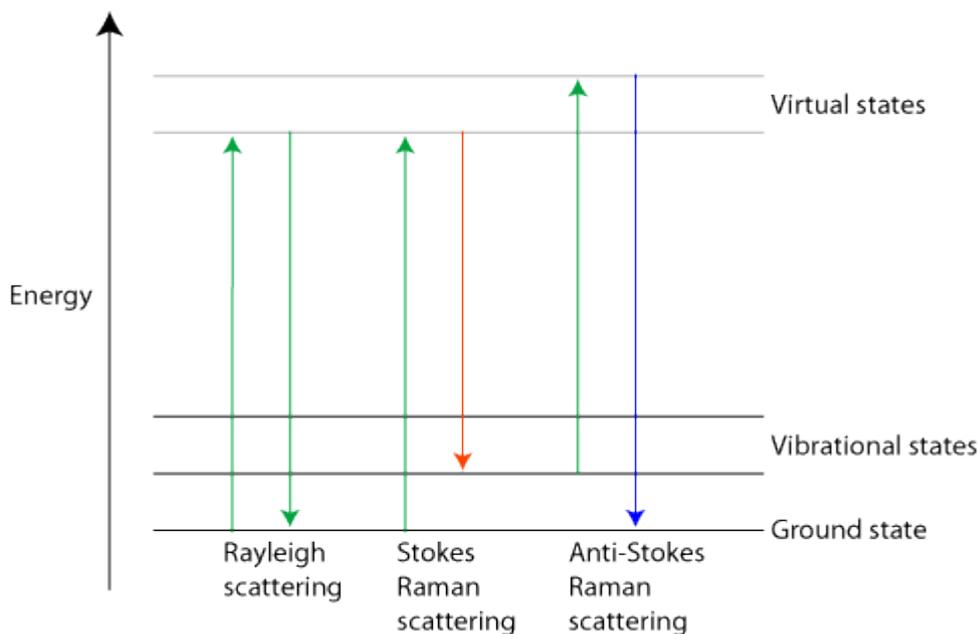


Figure 2.6 3 Different Forms of Scattering

Rayleigh scattering is about  $10^5$  times stronger than Raman scattering. Only about 1 in  $10^7$  photons undergo Stokes Raman scattering and so this is usually swamped by the far more prominent Rayleigh scattering. The amount of anti-Stokes scattering is even less than this. Raman scattering cross section is  $2 \times 10^{-14} \text{ cm}^2$  per molecule.

## 2.8 Multivariate Calibration

Calibration is the process of finding appropriate model parameters that lead from the spectrum to the desired information on the composition of the material. Nowadays, chemical problems and applications of chemometrics are involved with calibration. Chemometrics is the application in which mathematical and statistical method was used to analyse enormous amount of chemical data. Multivariate calibration methods

is more preferable to use in chemometrics compared to traditional method as it can be applied when more than one sample is acquired for each sample.

Multivariate calibration techniques such as partial-least squares regression, or principal component regression (and near countless other methods) are then used to construct a mathematical model that relates the multivariate response (spectrum) to the concentration of the analyte of interest, and such a model can be used to efficiently predict the concentrations of new samples.

The main advantages of the use of multivariate calibration techniques is it fast, cheap, and non-destructive analytical measurements as well as can be used to estimate sample properties which would otherwise require time-consuming, expensive or destructive testing. Multivariate calibration allows for accurate quantitative analysis in the presence of heavy interference by other analytes. The selectivity of the analytical method is provided as much by the mathematical calibration, as the analytical measurement modalities. For example infrared spectra, which are extremely broad and non-selective compared to other analytical techniques can often be used successfully in conjunction with carefully developed multivariate calibration methods to predict concentrations of analyte in very complex matrices.

### 2.8.1 Multi Linear Regression (MLR)

MLR is known as the best method from statistical point of view. It based on the simple linear regression based on the relationship between dependent variable  $y$  and the independent variable  $x$ . According to the multiple linear regression model the dependent variable is related to two or more independent variables [17].

$$y_i = \beta_0 + \beta_i x_i + e_i \quad i = 1,2,3, \dots n \quad (1)$$

It based on simple selection of variables for instance some variables are rather selective for the compounds or characteristics being determine. When we want to model one or several dependent variables,  $X$  with responses of  $Y$ , MLR are useful and can works well as long as  $X$ -variables are fairly uncorrelated. In addition, in modern measuring instrumentation including spectrometers, chromatographs and sensor batteries, the  $X$ -variables tend to be many and also strongly correlated [18]. MLR

attempts to model the relationship of  $X$  and  $Y$  by fitting a linear equation to observed data. This can be defined as

$$E(Y|X) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p,$$

Where  $\alpha$  is called intercepts and  $\beta_1$  are coefficient or slope.

### 2.8.2 Principle Component Analysis (PCA)

Due to disadvantages of MLR which required all the component must be known, PCA was introduced since it does not require details of all components. PCA is a way to describe multivariate data by maximizing variances. It is powerful tool that are abundantly used in all forms of analysis for data compression and information extraction. This technique reduce the dimensionality of the variables data, by forming a new set of orthogonal variables called principal components (PCs), which are a linear combination of the original measured variables and which explain the maximal amount of variability in the data.

It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data. PCA decomposes an  $X$  matrix into two smaller matrices, one of scores ( $T$ ) and the other of loadings ( $P$ ) as follows:

$$X = T * P$$

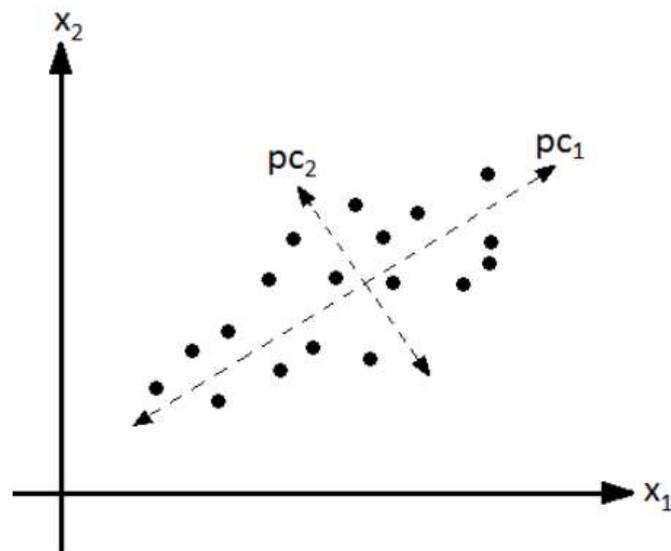


Figure 2.7 Illustration of principle component analysis (PCA)

## 2.9 Partial Least Square (PLS) Regression

Partial least-square (PLS) regression is a technique used to analyse the relationship between two data matrices,  $X$  and  $Y$  (predictor variable). This technique constructs new predictor variables, known as components, as linear combinations of the original predictor variables. PLS constructs the component by considering the observed response value. PLS has been used in analytical chemistry since the 1980s and increased steadily since then. It was first developed by Herman Wold in the late 1960s for economic purposes and after that it was introduced as a tool to analyse data from chemical applications [19].

This technique actually is an extension of multiple linear regression (MLR) analysis where the effect of linear combinations of several predictors on multiple response variables are analysed [19]. It is quite similar to PCA but it works in one step only. PCA works by finding the combination of the predictor with large variance and reducing correlations but neglects the response values. However, PLS works by finding the combination of the predictors that have large covariance as well as response values. In addition, PLS has the ability to analyse data with many, noise, collinear and even incomplete variables in both  $X$  and  $Y$ . Hence, a complex problem can be investigated and available data can be analysed in a more realistic way with PLS.

### 2.9.1 PLSR model

A linear model  $Y=XB+E$  will be built using PLS regression from a training set of  $X$  and  $Y$ .  $Y$  is a data matrix of  $n$  samples by  $m$  variable response,  $X$  is a data matrix of  $n$  samples by  $p$  variable and  $B$  is  $p$  by  $m$  regression coefficient matrix.  $E$  will be error or noise that have the same dimensions as  $Y$ . The predictions for new observations data are made based on  $X$ . The result from PLS regression will give the score matrix of  $X$  and  $Y$ , loadings of  $X$  and  $Y$ , weight ( $W$ ) and beta ( $B$ ). The result of  $X$ -scores are predictors of  $Y$  and model of  $X$  in an orthogonal way [18]. The score matrix is computed based on  $T=XW$ , and from the regression model  $Y=TQ+E$ , the loading  $Q$  for  $T$  can be known.

Once the loadings  $Q$  are computed, the regression model will be  $Y=XB+E$  where  $B=WQ$ . PLS regression produces the weight matrix  $W$  based on the covariance structure between the predictor and response variables. To complete the PLSR model

procedures is the  $p$  by  $c$  factor loading matrix  $P$  which gives a factor model  $X=TP+F$ , where  $F$  is the unexplained part of the  $X$  scores. To conclude that, PLS regression will decomposes both of  $X$  and  $Y$  as a product of a common set of orthogonal factors and a set of specific loadings [8]. To carry out PLS regression model in MATLAB, we will be used `plsregress` as the function. There are few function can be used in modelling as stated in Statistical Toolbox.

Table 2.2 PLS function in MATLAB [20]

Function	Description
<code>[XL, YL] = plsregress(X, Y, ncomp)</code>	Computes PLS regression of $Y$ on $X$ , using <code>ncomp</code> PLS components, and returns the predictor and response loadings in <code>XL</code> and <code>YL</code> , respectively. $X$ is an $n$ -by- $p$ matrix of predictor variables, with rows corresponding to observations and columns to variables. $Y$ is an $n$ -by- $m$ response matrix. <code>XL</code> is a $p$ -by- <code>ncomp</code> matrix of predictor loadings, where each row contains coefficients that define a linear combination of PLS components that approximate the original predictor variables. <code>YL</code> is an $m$ -by- <code>ncomp</code> matrix of response loadings, where each row contains coefficients that define a linear combination of PLS components that approximate the original response variables.
<code>[XL, YL, XS] = plsregress(X, Y, ncomp)</code>	Returns the predictor scores <code>XS</code> , that is, the PLS components that are linear combinations of the variables in $X$ . <code>XS</code> is an $n$ -by- <code>ncomp</code> orthonormal matrix with

		rows corresponding to observations and columns to components.
[XL, YL, XS, YS] plsregress(X, Y, ncomp)	=	Returns the response scores YS, that is, the linear combinations of the responses with which the PLS components XS have maximum covariance. YS is an <i>n</i> -by- <i>ncomp</i> matrix with rows corresponding to observations and columns to components. YS is neither orthogonal nor normalized.
[XL, YL, XS, YS, BETA, PCTVAR] plsregress(X, Y, ncomp)	=	Returns a 2-by- <i>ncomp</i> matrix PCTVAR containing the percentage of variance explained by the model. The first row of PCTVAR contains the percentage of variance explained in X by each PLS component, and the second row contains the percentage of variance explained in Y.
[XL, YL, XS, YS, BETA, PCTVAR, MSE] plsregress(X, Y, ncomp)	=	Contain matrix MSE, estimated mean-squared errors for PLS models with <i>ncomp</i> components. The first row of MSE contains mean-squared errors for the predictor variables in X, and the second row contains mean-squared errors for the response variable(s) in Y.

### 2.9.2 Model Evaluation

Any model needs to be evaluated first before it is used for predicting. Cross validation (CV) is one of the effective way for model validation as it can stimulate how well the model can predicts new data. The best validation of a model is that it able to predicts the Y-values of observations with new X-values precisely. Differences between actual and predicted Y-values are calculated and the sum of squares of these differences is computed to form the predictive residual sum of squares, PRESS, which estimates the predictive ability of the model. PRESS also often re-expressed as  $Q^2$  or  $R^2$ .

The formula to evaluate the  $R^2$

$$R^2 = 1 - \frac{\sum_{i=1}^n \left( Y_i - \left( \sum_{j=1}^p x_{ij} B_j \right) \right)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Where

$n$	# samples
$p$	# input variables
$B_j$	Regression Coefficient for feature j
$Y_i$	Vector of multidimensional output values for sample i
$x_{ij}$	Input value for feature j in sample I
$Y$	Sample mean of output data
$q$	# Components

## **CHAPTER 3**

### **METHODOLOGY**

#### **3.1 Project Methodology**

This project is mainly to develop a calibration model to analyse the solubility of CO<sub>2</sub> in the DEA solution using Raman spectroscopy. The development of the model is using PLS regression approach and was done in MATLAB. The first step was doing some study and literature review on every component related to the project such as carbon dioxide, absorption process, alkanolamines especially DEA and multivariate calibration.

Then, next phase is to obtain CO<sub>2</sub> solubility data in DEA using CO<sub>2</sub> absorption cell. CO<sub>2</sub> with purity of 99.8% was purchased from Air Product Malaysia Sdn. Bhd while DEA was purchased from the Merck Sdn. Bhd with purity of 99%. Distilled water was used for the preparation of solution. 10%, 20%, 30% of DEA aqueous solution was prepared and feed into the absorption vessel. CO<sub>2</sub> from feed tank will disperse into absorption vessel and was absorbed by DEA. CO<sub>2</sub> loading was measured by using pressure drop method.



Figure 3.1 CO<sub>2</sub> Absorption Cell

The portable Raman spectrometer was used to project the laser source to the sample. The laser source had wavelength of 785 nm and power of 500 mW. The spectrometer had resolution of 4cm<sup>-1</sup> and signal to noise ratio of 1000:1 [21]. To capture the spectra, the software SpectraWiz was used. The optical fiber probe of Raman spectrometer was connected with the CO<sub>2</sub> absorption cell.

After that, pre-processing of Raman spectra need to be done in order to get a good spectra. In the other typical measurement techniques, noise can cover the information hidden in a spectrum. Besides that, the information might contained in the high intensity and low intensity spectra. Hence, pre-processing is recommended to remove the noise effect and maximize the data obtained in spectrum. The Raman spectra were reduced by using Savitzky-Galoy filter.

Then, the last phase is started to do modelling in the MATLAB. Modelling approaches must be applied in order to convert the data contained in Raman spectra. The data will be divided into two types of data set, calibration and validation data. Calibration data set will be used to develop few models using PLS regression while validation data set is to test the performance of calibration model. The calibration and validation data set comprise Raman Intensities with respective Raman shift as X variable and CO<sub>2</sub> loading as Y variable.

Data matrix of X and Y need to be normalized first before do PLS regression in order to get accurate result. It will scaled up and centred the dataset and give mean equal to 0 and standard deviation equal to 1. This step only been done to calibration dataset. For validation data set, X and Y data matrix will be scaled up using mean and standard deviation obtained from calibration model.

Then, PLS regression is applied to the calibration dataset. It will generate XL, YL, XS, YS and Beta. XL is a predictor loading consist of coefficient between PLS components and original predictor variable. YL refer to response loading where each row consists of coefficients of PLS components that approximate the original response variables. XS and YS is refer to score of X and Y. Beta is coefficient which will be used in modelling.

After finish modelling calibration dataset, a new dataset which is validation dataset will be tested. When a process model is used for the prediction of process variables, the quality of that prediction must be judged. The prediction of calibration and validation has been evaluated by using coefficient of determination ( $R^2$ ) and mean square error (MSE). R-squared will be used to see how well the regression line approximated the data.

### **3.2 Process Flow Chart**

For this study, the methodology is divided into three main parts. The first part is obtaining CO<sub>2</sub> solubility data in DEA. Second part is developing a calibration model using PLS regression approach in MATLAB and lastly to evaluate the performance of model. This is the process flow for this project that will be carried out throughout this project so that the objective of the study can be successfully achieved.

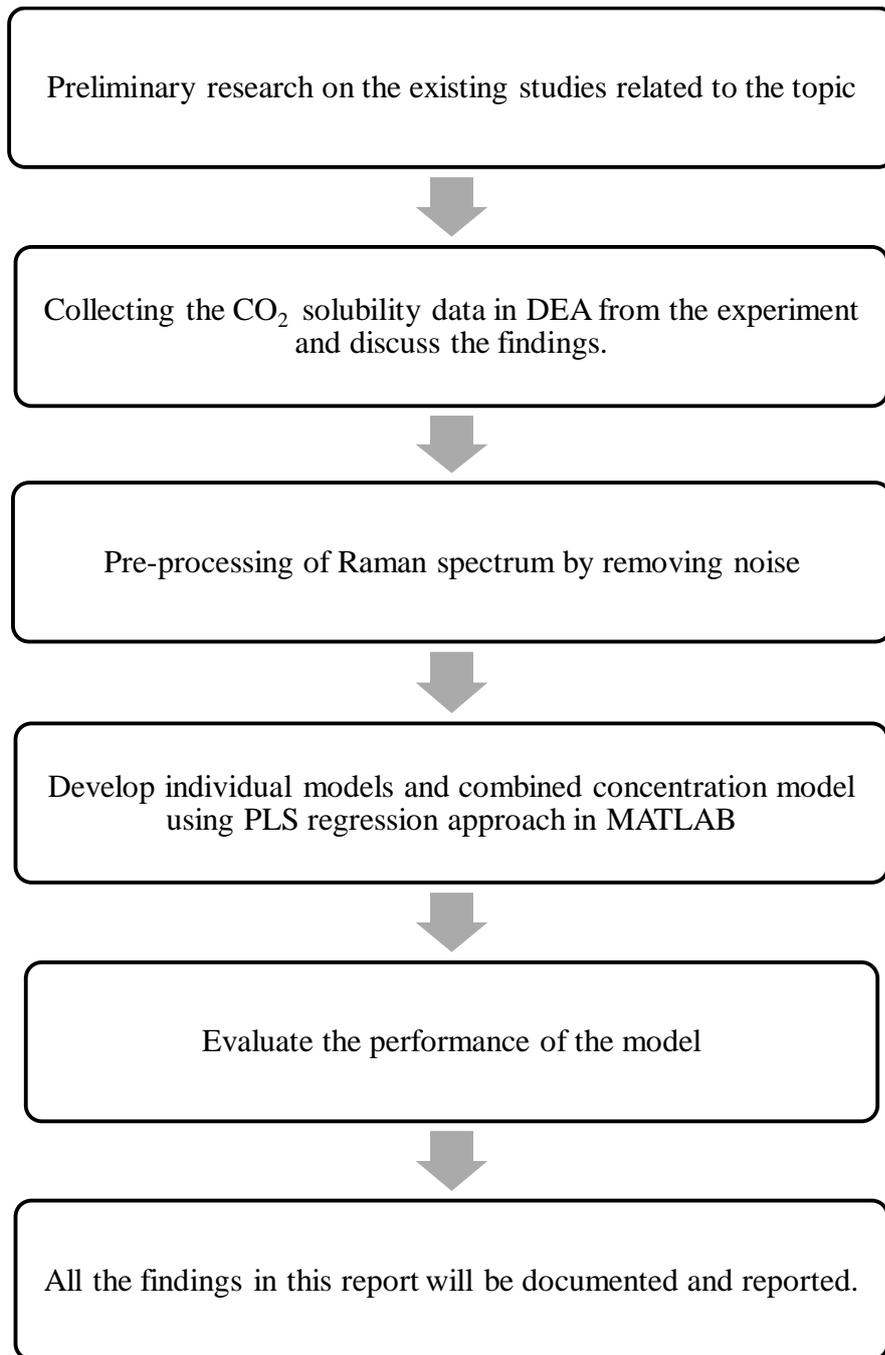


Figure 3.2 Process Flow Chart

### 3.3 Key Milestones

1. Obtained CO<sub>2</sub> solubility data using Raman Spectroscopy

2. Developed individual models based on 10%, 20%, and 30% DEA concentration

3. Developed combined concentration model to predict CO<sub>2</sub> loading at different concentration.

4. Evaluated the performance of the individual models and combined concentration model

### 3.4 Gantt chart

Activities/Plan	Final Year Project II (FYPII)													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Develop individual PLS models for each DEA concentration.	■	■	■											
Develop combined concentration models to predict CO2 loading of different concentration				■	■	■								
Evaluate the performance of the models							■	■	■					
Compile all data together							■	■	■	■				
Identify appropriate solution and recommendation for improving the result of the models									■	■	■	■		
Finalize all data and result											■	■		
Key Milestones	Final Year Project II (FYPII)													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Submission of the progress report							●							
Pre-SEDEX										●				
Submission of Technical Paper												●		
Submission of Dissertation (soft bound)												●		
Final Year Project Oral Presentation (Viva)													●	

## CHAPTER 4

### RESULT AND DISCUSSION

#### 4.1 Qualitative Analysis of Raman

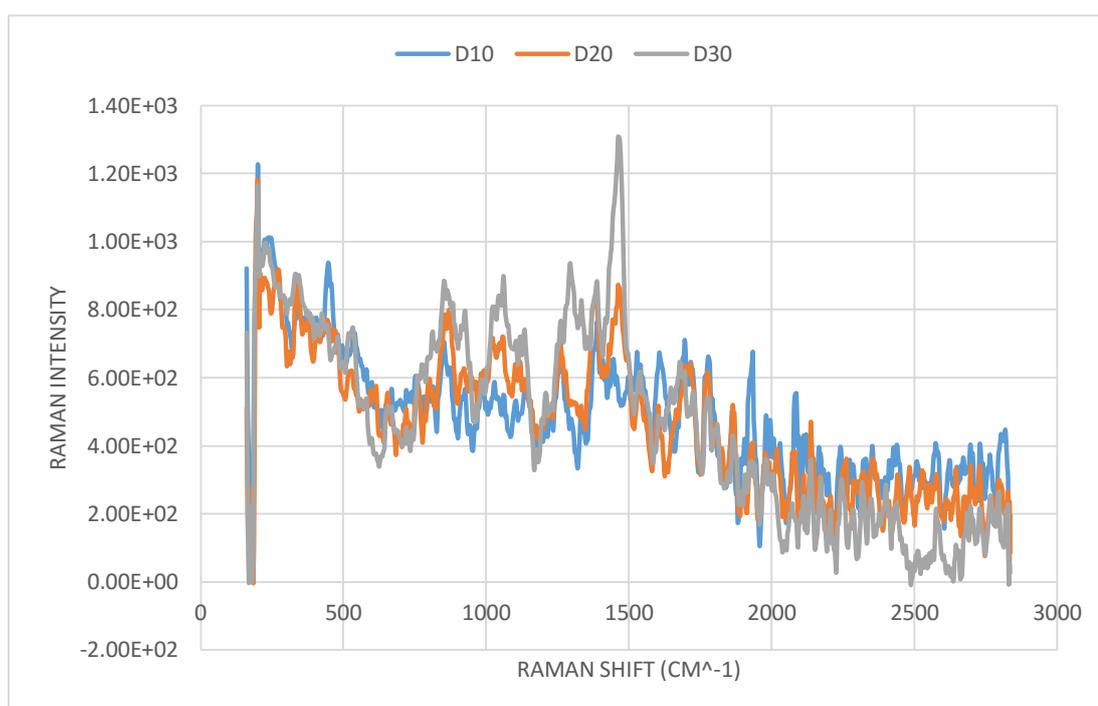


Figure 4.1 Raman Spectrum of DEA

Figure 4.1 shows the Raman spectrum of DEA at 10%, 20% and 30% concentration. It can be seen that the intensity of 30% concentration is higher at spectral region  $1200\text{cm}^{-1}$  to  $1500\text{cm}^{-1}$ . It might be because of formation carbamate when there is higher concentration of amine. The formation of carbonate and bicarbonate is identified between spectral regions of  $1000\text{ cm}^{-1}$  to  $1080\text{ cm}^{-1}$  it has been observed that peak intensities at that range is increasing during  $\text{CO}_2$  absorption. Then, the peaks declined due to the consumption of amine during absorption process.

## 4.2 Partial Least Square Regression

Calibration model of 10%, 20% and 30% of DEA concentration have been developed in the first stage of the project which named D10, D20 and D30. The data set used to load in MATLAB will be divided into calibration set and validation set. In the calibration set, each model consists of 45 samples of CO<sub>2</sub> loading ranging from 0.055 to 1.322 at 860 wavelengths whereby in validation set comprise of 25 samples. All the model were developed using PLS approach with 10 numbers of component.

### 4.2.1 Model for 10% of DEA (D10)

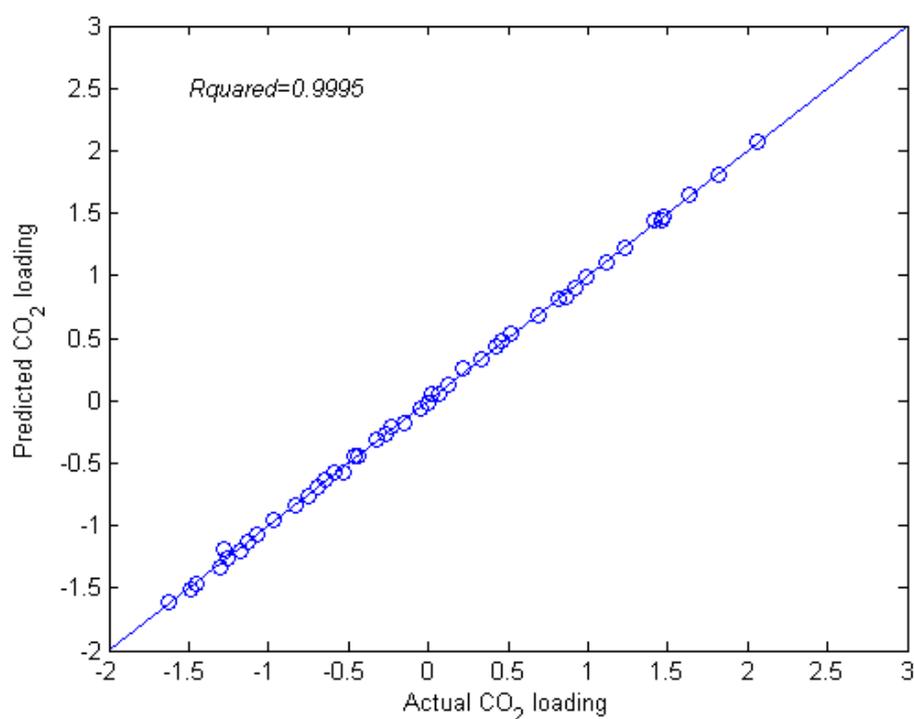


Figure 4.2 Graph of actual and predicted CO<sub>2</sub> loading in calibration model

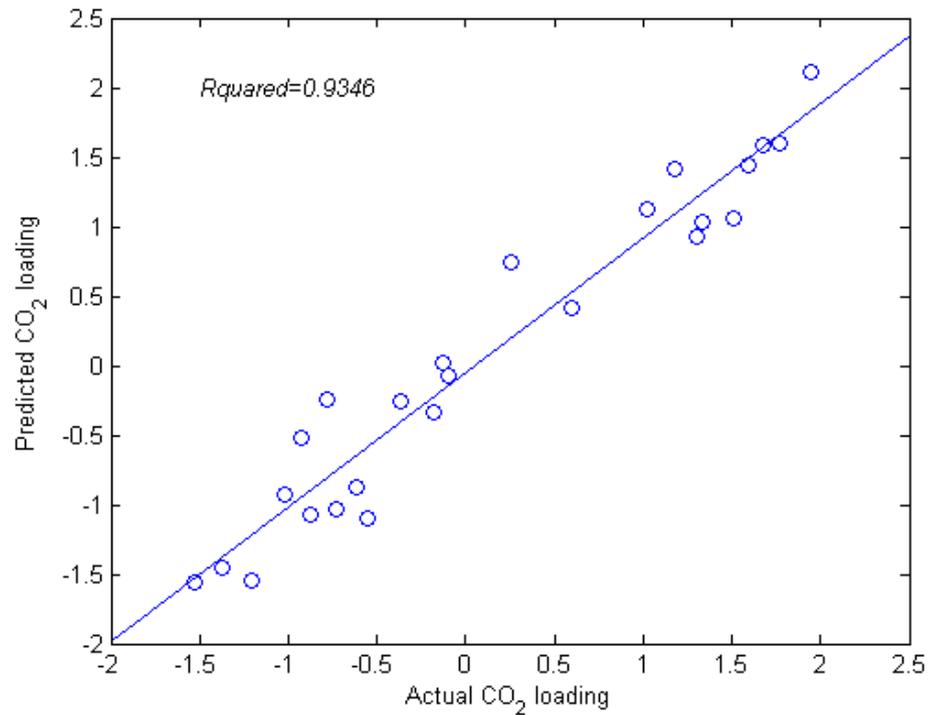


Figure 4.3 Graph of actual and predicted CO<sub>2</sub> loading in validation model

As we can see in Figure 4.1, D10 model give good prediction of CO<sub>2</sub> loading with excellent linearity while in Figure 4.2, the linearity of prediction CO<sub>2</sub> loading is quite low compared to Figure 4.1. The performance of the model will be evaluated by R<sup>2</sup> and MSE value in the table below

Table 4.1 Calibration and Validation Results in D10

	Calibration Model	Validation Model
Coefficient of Determination (R <sup>2</sup> )	0.9995	0.9346
Mean Square Error (MSE)	0.00048943	0.0802

Based on the figure above, we can see that using PLS regression approach give the good prediction with best linearity in the calibration and validation graph. The performances of models are determine based on the coefficient of determination (R<sup>2</sup>) and mean square error (MSE). Calibration model give R<sup>2</sup> = 0.9995 which is almost near to 1. Validation model give R<sup>2</sup>=0.9346 which is slightly low compared to calibration model as expected. However, the value is remain to be considered good.

#### 4.2.2 Model for 20% of DEA (D20)

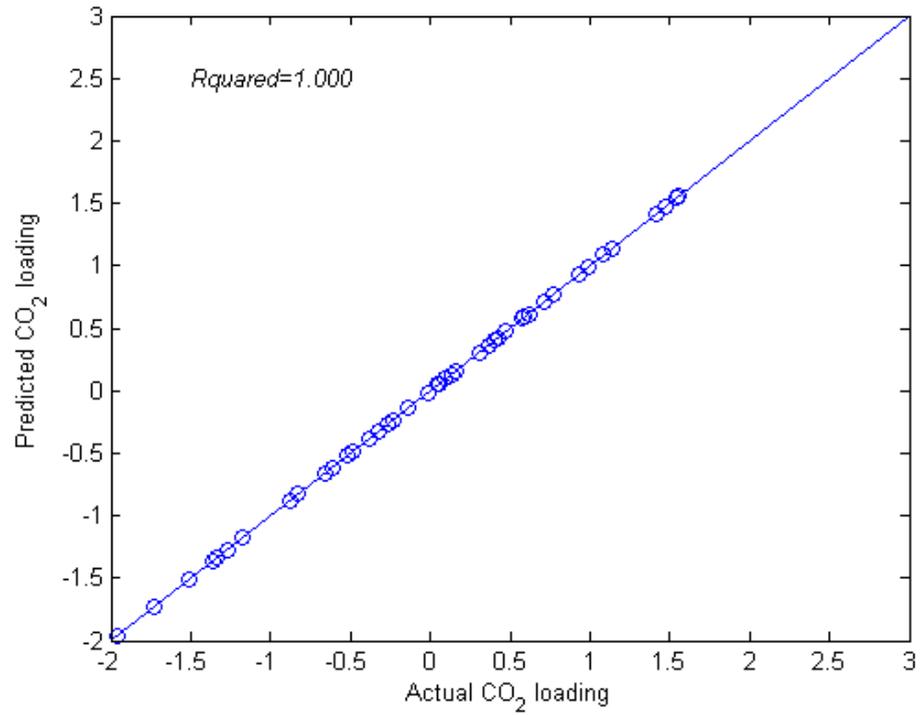


Figure.4.4 Graph of actual and predicted CO<sub>2</sub> loading in calibration model

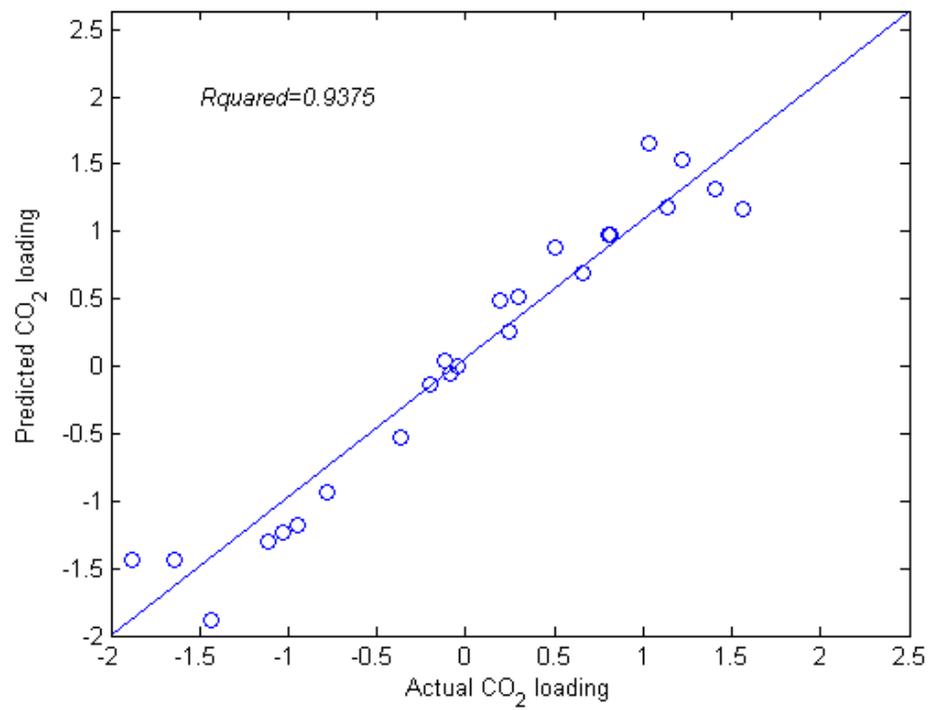


Figure 4.5 Graph of actual and predicted CO<sub>2</sub> loading in validation model

In D20, calibration model also shows good prediction of CO<sub>2</sub> loading with the good linearity. The performance of the model will be evaluated by R<sup>2</sup> and MSE value in table below.

Table 4.2 Calibration and Validation Results in D20

	Calibration Model	Validation Model
Coefficient of Determination (R <sup>2</sup> )	1.0000	0.9375
Mean Square Error (MSE)	0.0000064189	0.0680

The evaluation of the fit of D20 was done by correlate the actual values of CO<sub>2</sub> loading and predicted CO<sub>2</sub> loading values calculated by PLS. In the calibration model, it give the excellent result which is R<sup>2</sup>=1.0000 while in validation model give R<sup>2</sup>=0.9375. The mean square error of calibration model is very low that are approximate to 0. Calibration model of D20 shows the best performance with the highest R<sup>2</sup> value.

#### 4.2.3 Model for 30% of DEA (D30)

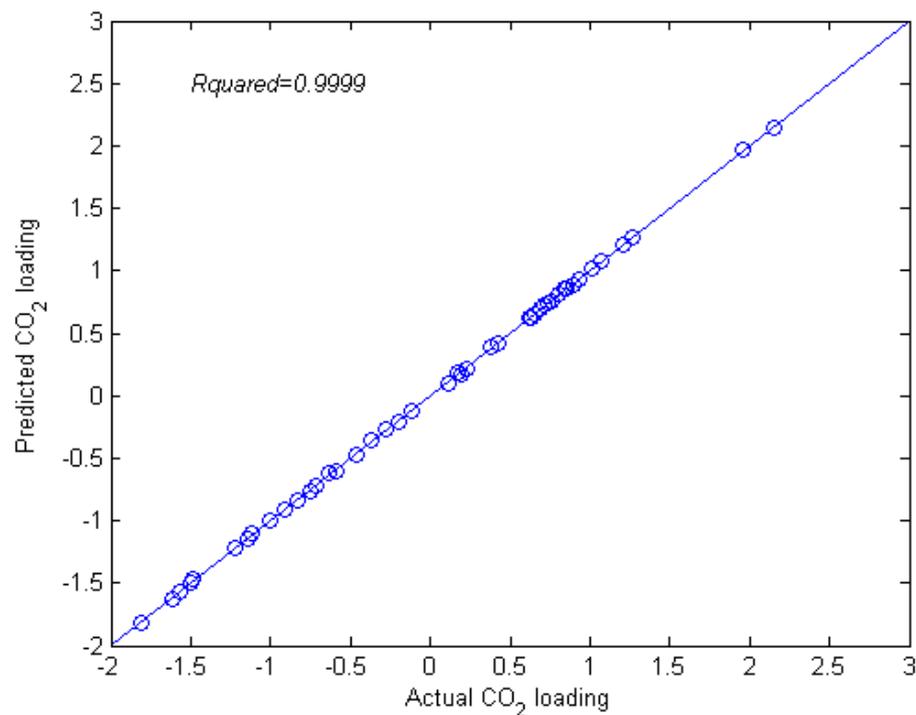


Figure 4.6 Graph of actual and predicted CO<sub>2</sub> loading in calibration model

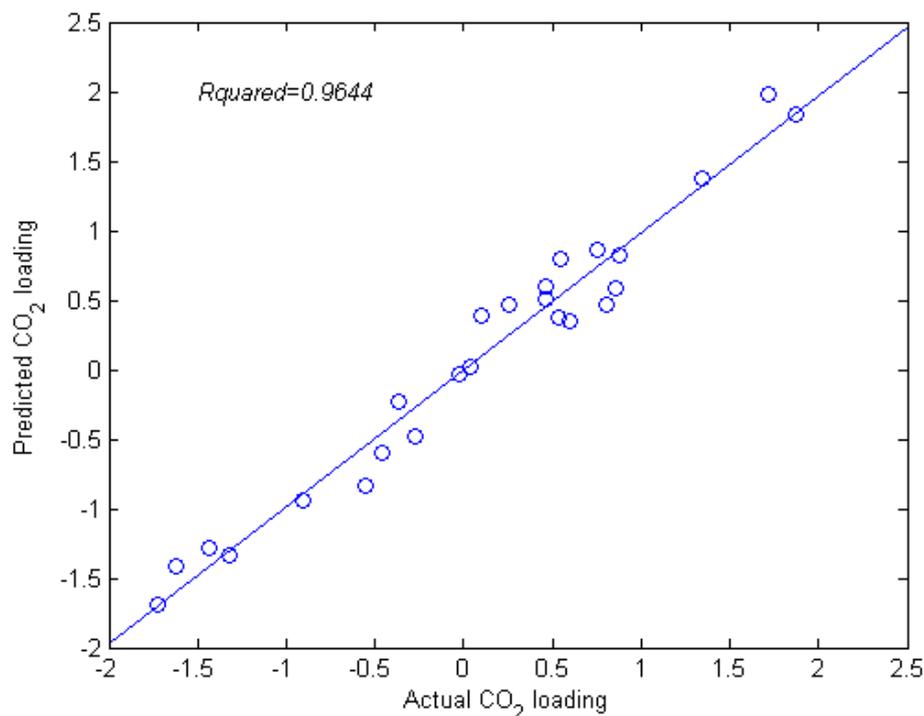


Figure 4.7 Graph of actual and predicted CO<sub>2</sub> loading in validation model

Table 4.3 Calibration and Validation Results in D30

	Calibration Model	Validation Model
Coefficient of Determination (R <sup>2</sup> )	0.9999	0.9644
Mean Square Error (MSE)	0.000096595	0.0329

The performance of D30 also quite good which give R<sup>2</sup> = 0.9999 and MSE = 0.000097. D30 model has the highest value of R<sup>2</sup> in the validation model compared to other two model which is R<sup>2</sup> = 0.9644. It might be due to high formation of carbamate when there is high concentration of DEA

#### 4.2.4 Combined Concentration Model - DC1

To fulfil the 3<sup>rd</sup> objective of this project which is to develop a model that able to predict CO<sub>2</sub> solubility at different concentration aqueous DEA, combined concentration modified model (DC1) is introduced in next phase. DC1 consist of combination dataset CO<sub>2</sub> loading and Raman intensities from D10, D20 and D30

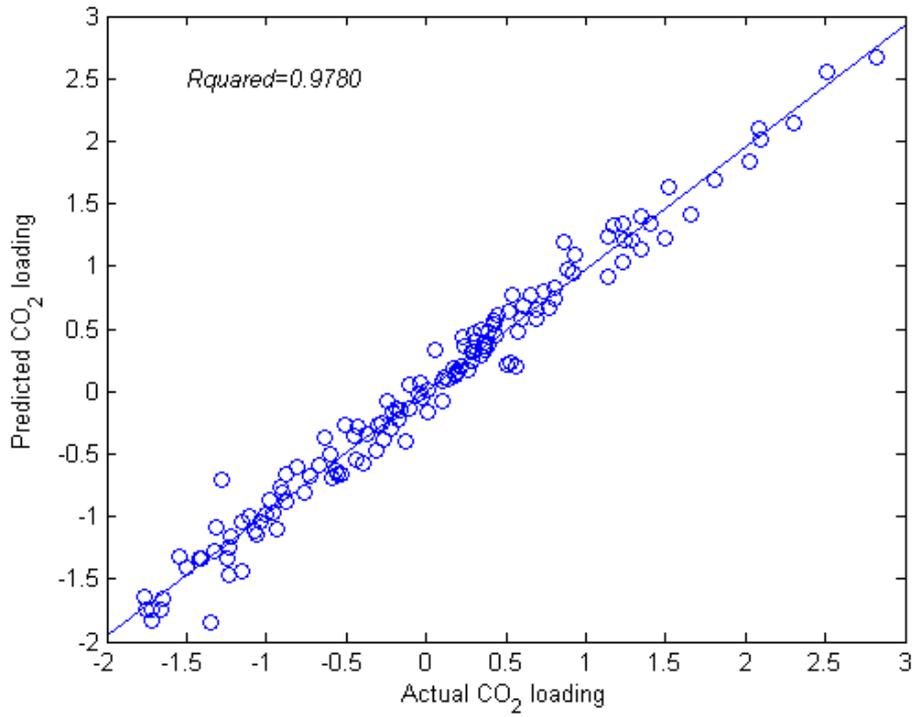


Figure 4.8 Graph of actual and predicted CO<sub>2</sub> loading in calibration model

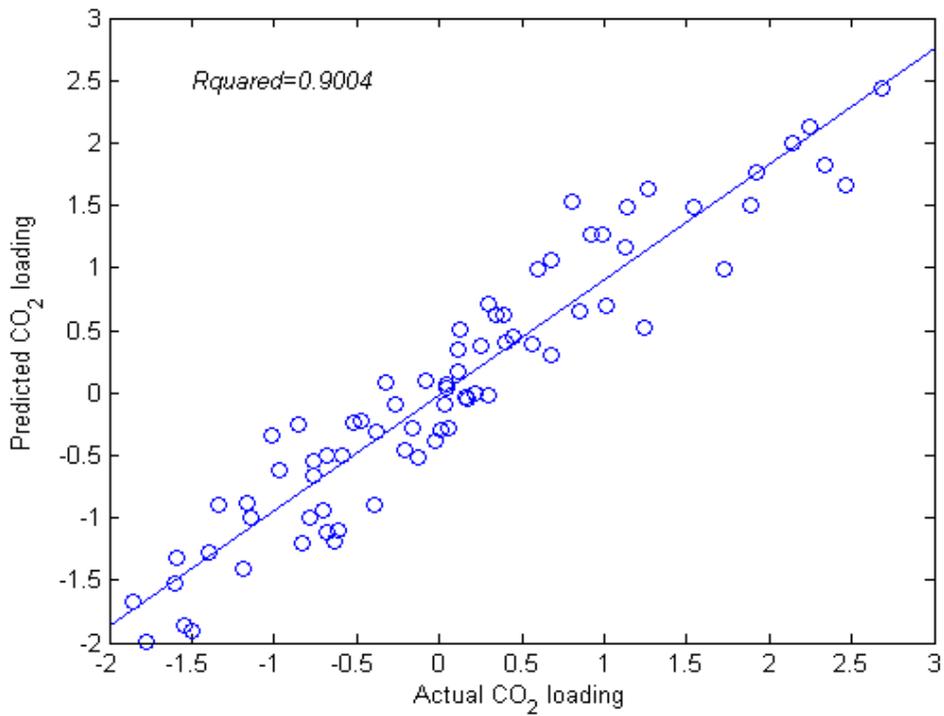


Figure 4.9 Graph of actual and predicted CO<sub>2</sub> loading in validation model

Table 4.4 Calibration and Validation Results in DC1

	Calibration Model	Validation Model
Coefficient of Determination (R <sup>2</sup> )	0.9780	0.9004
Mean Square Error (MSE)	0.0218	0.1143

By combining CO<sub>2</sub> loading and Raman intensities of 10%, 20% and 30% of DEA into 1 dataset, DC1 still give quite good prediction of CO<sub>2</sub> loading with good linearity in calibration model. However, in validation model, the linearity is not so good compared to calibration model. Calibration model in DC1 give R<sup>2</sup>=0.9780, slightly low compared to D10, D20 and D30 which give excellent result while in the validation model, the R<sup>2</sup> value is 0.9004 only. The error obtained also quite huge in validation model. Thus, few modified combined model will be introduced in order to further improve the model especially value of R<sup>2</sup> and MSE in validation.

#### 4.2.5 Modified Combined Concentration Model – DC2

DC2 model was developed to investigate the effect of concentration column in the prediction performances. In this case, additional column consists of the amine concentration is added to the X data matrix (observed variable).

$$\underbrace{\begin{pmatrix} 10 \\ 20 \\ 30 \end{pmatrix}}_{\text{Amine concentration}} + \underbrace{\begin{pmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{pmatrix}}_{\text{X data matrix which is refer to Raman intensities}}$$

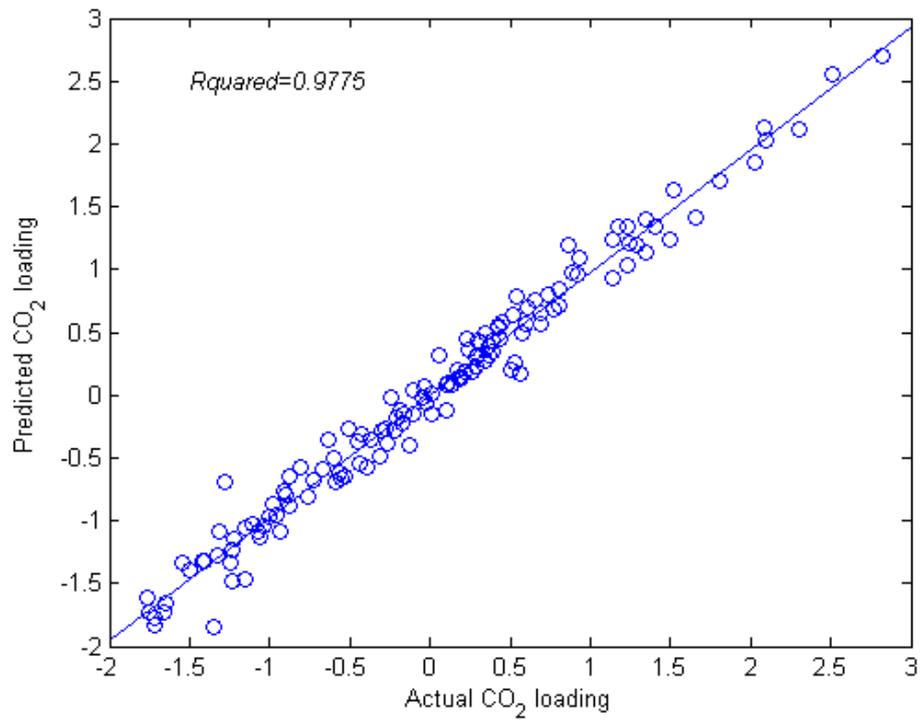


Figure 4.10 Graph of actual and predicted CO<sub>2</sub> loading in calibration model

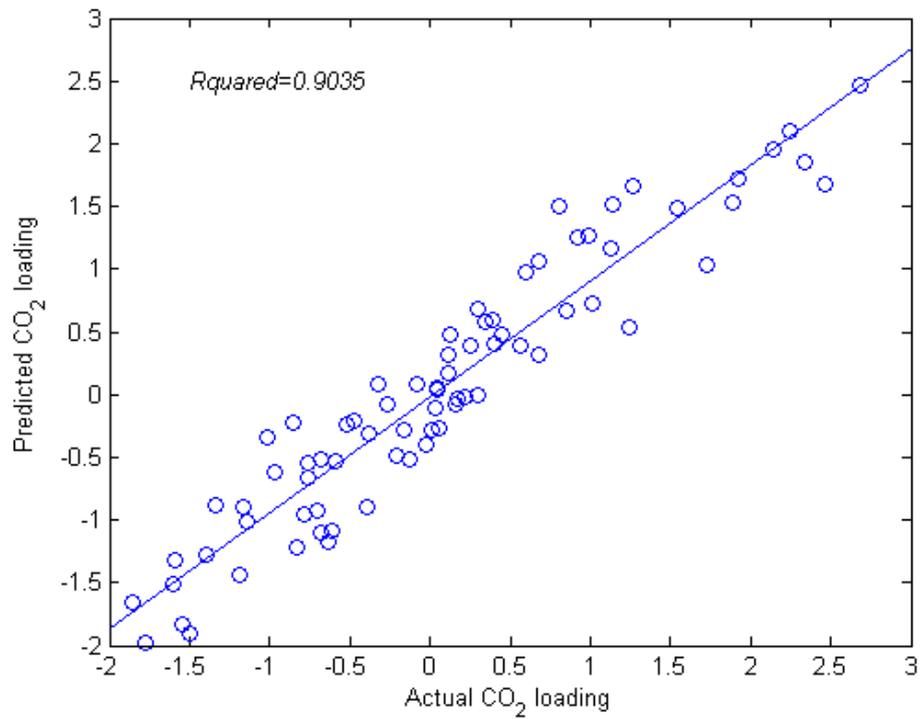


Figure 4.11 Graph of actual and predicted CO<sub>2</sub> loading in validation model

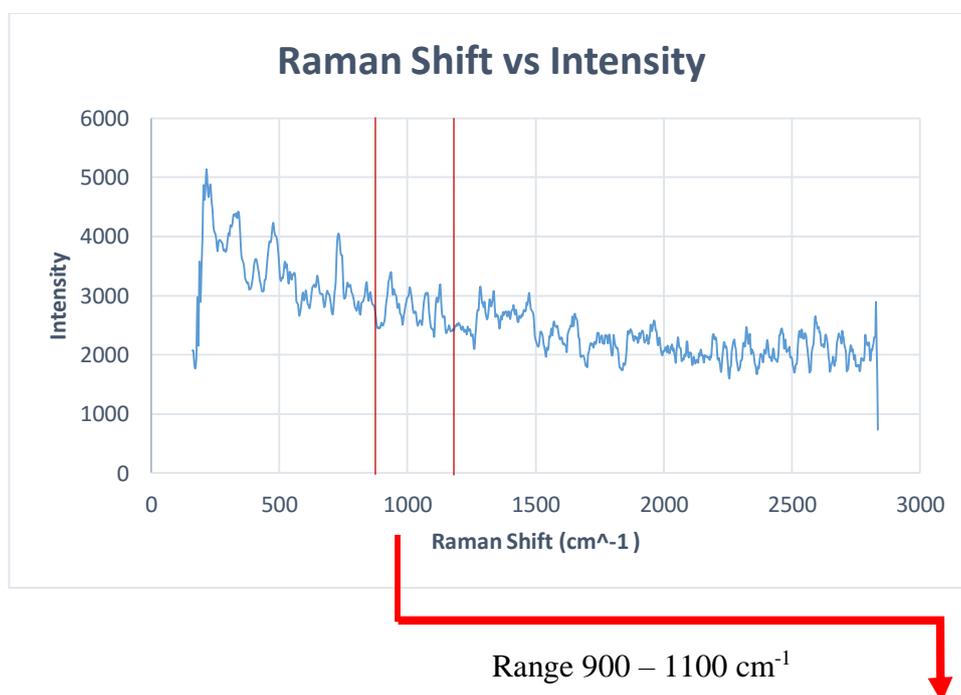
Table 4.5 Calibration and Validation Results in DC2

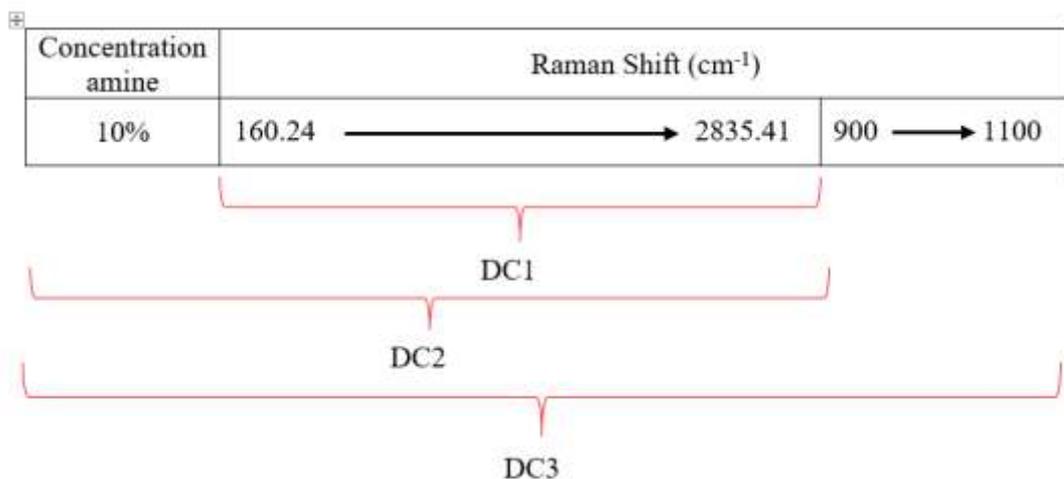
	Calibration Model	Validation Model
Coefficient of Determination ( $R^2$ )	0.9775	0.9035
Mean Square Error (MSE)	0.0224	0.1107

The result of  $R^2$  after add additional column consists of amine concentration is still not satisfying enough especially in the validation model. The difference of  $R^2$  in calibration and validation model is quite huge about 0.074. The error obtained in DC2 does not have much difference from DC1 model. Further study need to be done in order to improve validation model of DC1 and DC2. The information in DC1 and DC2 will act as a basis in the next modified combined model. Its means that the performance of next modified combined model must be greater than DC1 and DC2.

#### 4.2.6 Modified Combined Concentration Model - DC3

Another model, DC3 model is introduced to see the effect of formation carbonate, bicarbonate and carbamate in the DEA reaction with  $CO_2$ . According to the past research, the formation of carbonate, bicarbonate and carbamate is between Raman shifts  $900 - 1100\text{ cm}^{-1}$ . The subsequent study is to investigate how this range can be further exploited. The diagram below is to show how actually DC3 is developed





DC3 model is introduced to increase the validation  $R^2$  values. There are 3 sub models of DC3 which are DC3-A, DC3-B and DC3-C in developed to further improve the  $R^2$  value.

Models	Details
DC3-A	To see the influence of this range, we are duplicating by add additional column in X data matrix.
DC3-B	In DC3-B, the additional column will be squared.
DC3-C	In DC3-C, the additional column from DC3-B will be inversed. It will be $\frac{1}{range^2}$

Table 4.6 Calibration and Validation Result in DC3

Model	Coefficient of Determination ( $R^2$ )		Mean Square Error (MSE)	
	Calibration	Validation	Calibration	Validation
DC3 - A	0.9802	0.8943	0.0204	0.1188
DC3 - B	0.9756	0.8994	0.0242	0.1153
DC3 - C	0.9751	0.9136	0.0247	0.0991

The table below shows the result obtained after do PLS regression. We can see that calibration  $R^2$  values is kept decreasing but still in the good range while validation  $R^2$  values is kept increasing. DC3-C shows the highest validation  $R^2$  values compared to DC1, DC2, DC3-A and DC3-B which is 0.9136.

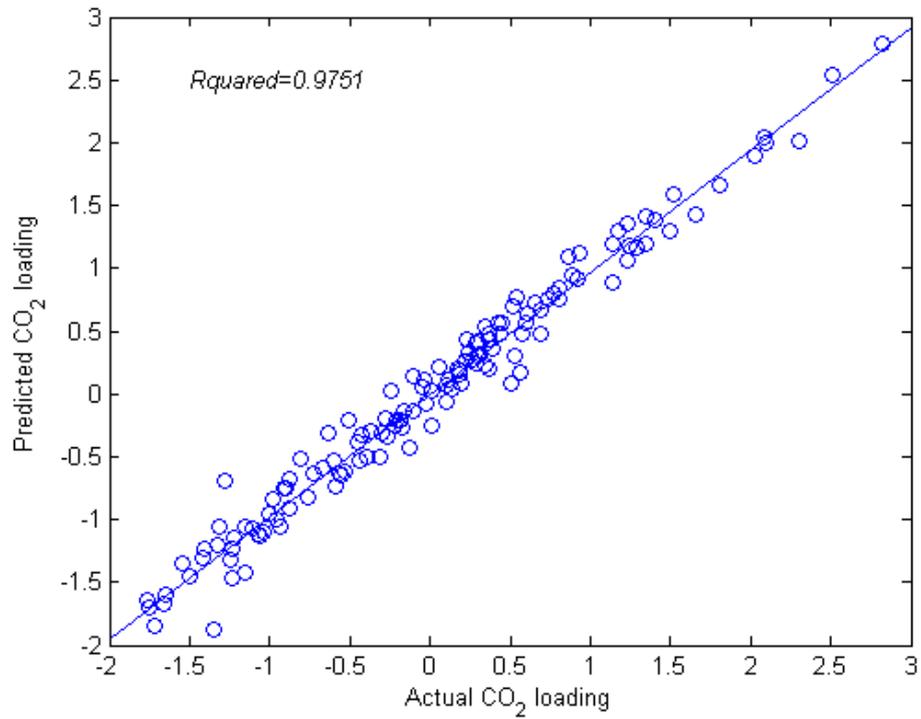


Figure 4.12 Graph of actual and predicted CO<sub>2</sub> loading in calibration model

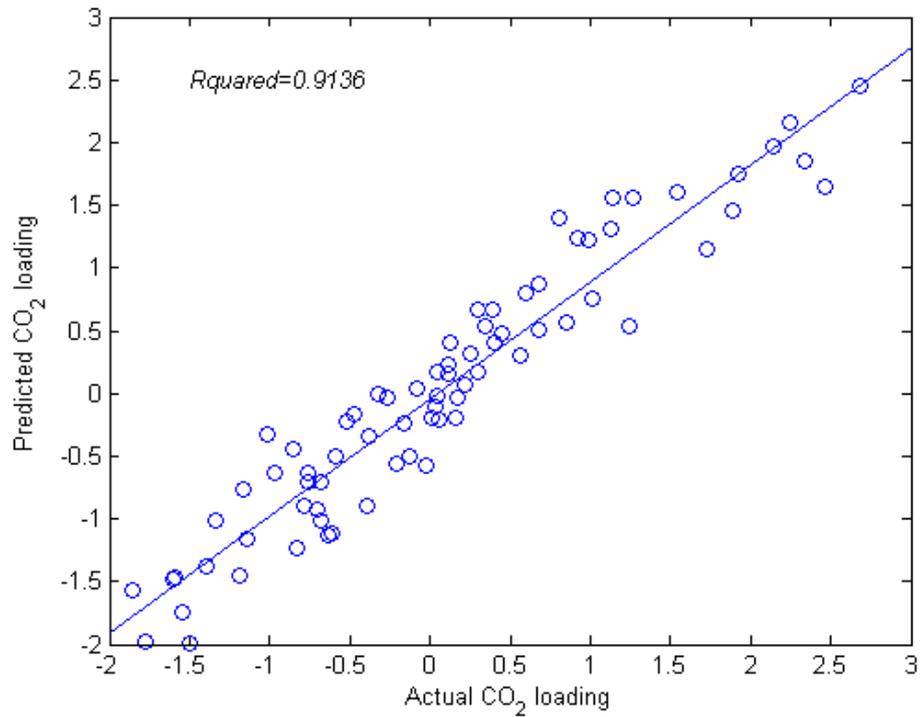


Figure 4.13 Graph of actual and predicted CO<sub>2</sub> loading in validation model

## CHAPTER 5

### CONCLUSION

This project is basically to establish a calibration model to predict CO<sub>2</sub> solubility in different concentration of aqueous DEA by using Raman spectroscopy for proper process control at CO<sub>2</sub> removing process plant. Based on the result obtained, it can conclude that individual models have shown good performance and great ability to predict CO<sub>2</sub> solubility at 10%, 20% and 30% concentration of DEA. This can be shown by the calibration R<sup>2</sup> values of this 3 models is nearly to unity and the validation R<sup>2</sup> values are more than 0.9. Besides that, combined concentration model, DC1 also have been developed in order to predict CO<sub>2</sub> solubility at different concentration DEA. Result shows that calibration R<sup>2</sup> value in DC1 is in the range of 0.97 while validation R<sup>2</sup> value is 0.90 only. Additional column consist of concentration DEA is introduced in combined concentration modified model, DC2 in order to enhance the validation set. However from the result obtained, concentration column did not have significant effect in increasing validation R<sup>2</sup> value. The final model which is DC3-C shows some improvement and has the highest validation R<sup>2</sup> value compared to DC1 and DC2. This model demonstrate great ability to predict CO<sub>2</sub> solubility. In conclusion, multivariate PLS regression has been successfully used to predict CO<sub>2</sub> solubility in different concentration of aqueous amine using Raman spectroscopy.

## REFERENCE

- [1] Y. E. Kim, J. A. Lim, S. K. Jeong, Y. Il Yoon, S. T. Bae, and S. C. Nam, "Comparison of carbon dioxide absorption in aqueous MEA, DEA, TEA, and AMP solutions," *Bull. Korean Chem. Soc.*, vol. 34, no. 3, pp. 783–787, 2013.
- [2] N. Barros, G. Oliveira, and M. Sousa, "Environmental Impact Assessment of Carbon Capture and Sequestration: General overview Introduction World Energy demand and supply GHG concentration in atmosphere and CCS CCS legal framework," no. June, pp. 1–6, 2012.
- [3] Ippcc, "Intergovernmental Panel on Climate Change. Climate Change 2014 Synthesis Report," p. 80, 2014.
- [4] M. K. Wong, M. A. Bustam, and A. M. Shariff, "International Journal of Greenhouse Gas Control Chemical speciation of CO<sub>2</sub> absorption in aqueous monoethanolamine investigated by in situ Raman spectroscopy," *Int. J. Greenh. Gas Control*, vol. 39, pp. 139–147, 2015.
- [5] S. Park, B. Min, J. Lee, and S. Nam, "Absorption Characteristic of Continuous CO<sub>2</sub> Absorption Process," *Fuel*, vol. 49, no. 1, pp. 249–250, 2004.
- [6] L. S. Tan, a. M. Shariff, K. K. Lau, and M. a. Bustam, "Factors affecting CO<sub>2</sub> absorption efficiency in packed column: A review," *J. Ind. Eng. Chem.*, vol. 18, no. 6, pp. 1874–1883, 2012.
- [7] Haji-Sulaiman M Z, M. K. Aroua, and A. Benamor, "Analysis of Equilibrium Data of CO<sub>2</sub> in Their Mixtures Using the Modified," vol. 76, no. November, 1998.
- [8] H. Abdi, "Partial Least Squares ( PLS ) Regression .," pp. 1–7, 2003.
- [9] Shakhashiri, "Carbon Dioxide , CO<sub>2</sub>," *Www.Scifun.Org*, vol. 3, pp. 2–3, 2008.
- [10] C.-H. Yu, "A Review of CO<sub>2</sub> Capture by Absorption and Adsorption," *Aerosol Air Qual. Res.*, pp. 745–769, 2012.
- [11] J. Gabrielsen, M. L. Michelsen, E. H. Stenby, and G. M. Kontogeorgis, "A Model for Estimating CO<sub>2</sub> Solubility in Aqueous Alkanolamines," *Ind. Eng. Chem. Res.*, vol. 44, no. 9, pp. 3348–3354, 2005.
- [12] T. Neveux, Y. Le Moullec, J. P. Corriou, and E. Favre, "Modeling CO<sub>2</sub> capture in amine solvents: Prediction of performance and insights on limiting phenomena," *Ind. Eng. Chem. Res.*, vol. 52, no. 11, pp. 4266–4279, 2013.
- [13] M. Kundu and S. S. Bandyopadhyay, "Solubility of CO<sub>2</sub> in water+diethanolamine+N-methyldiethanolamine," *Fluid Phase Equilib.*, vol. 248, no. 2, pp. 158–167, Oct. 2006.

- [14] S. Faassen and B. Hitzmann, "Fluorescence Spectroscopy and Chemometric Modeling for Bioprocess Monitoring," *Sensors*, vol. 15, no. 5, pp. 10271–10291, 2015.
- [15] T. Vankeirsbilck, A. Vercauteren, W. Baeyens, G. Van der Weken, F. Verpoort, G. Vergote, and J. P. Remon, "Applications of Raman spectroscopy in pharmaceutical analysis," *TrAC - Trends Anal. Chem.*, vol. 21, no. 12, pp. 869–877, 2002.
- [16] S. Pieters, Y. Vander Heyden, J. M. Roger, M. D'Hondt, L. Hansen, B. Palagos, B. De Spiegeleer, J. P. Remon, C. Vervaet, and T. De Beer, "Raman spectroscopy and multivariate analysis for the rapid discrimination between native-like and non-native states in freeze-dried protein formulations," *Eur. J. Pharm. Biopharm.*, vol. 85, no. 2, pp. 263–271, 2013.
- [17] S. H. Brown, "Multiple Linear Regression Analysis : A Matrix Approach with MATLAB," pp. 1–3, 2009.
- [18] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," *Chemom. Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, 2001.
- [19] L. M. Carrascal, I. Galván, and O. Gordo, "Partial least squares regression as an alternative to current regression methods used in ecology," *Oikos*, vol. 118, no. 5, pp. 681–690, 2009.
- [20] "Statistics Toolbox™ User 's Guide R 2014 a," 2014.
- [21] M. A. Bustam. M. Z. Shahid, A. S. Maulud, "Modelling of CO<sub>2</sub> loading in DEA by using peak ratio of Raman spectroscopy," *Appl. Mech. Mater.*, vol. 625, pp. 188–191, 2014.

## APPENDICES

### APPENDIX I

```
%CO2 loading calibration model in 10% DEA
%by Nuralia Syairah Osman (15669)
% input:
% X      data matrix of raman intensity
% y      data matrix of CO2 loading
load deal0calibration
y = deal0calibration(:,1);
X = deal0calibration(:,2:861);
%center and scale X and Y
[Xn,meanX,stdX]=zscore(X);
[yn,meany,stdy]=zscore(y);
% Outputs:
% XL     loading matrix of X
% YL     loading matrix of y
% XS     score matrix of X
% YS     score matrix of y
% beta   matrix of regression coefficient
% PCTVAR weight matrix of X
% STATS consists of Weight, T2, Xresidual and Yresidual
[XL,YL,XS,YS,beta,PCTVAR,MSE,STATS] = plsregress(Xn,yn,10);
[XL,YL,XS,YS,beta,PCTVAR,MSE,STATS] = plsregress(Xn,yn,10);
plot(1:10,cumsum(100*PCTVAR(2,:)),'-bo');
xlabel('Number of PLS components');
ylabel('Percent Variance Explained in y');
%obtained predicted CO2 Loading
ypredicted=Xn*beta(2:end,:);
%test the model performance
TSS=sum((yn-mean(yn)).^2);
RSS=sum((yn-ypredicted).^2);
Rsquared=1 - RSS/TSS;
MSEcal=RSS/45;
%Plot the Calibration Graph
figure()
plot(yn,ypredicted,'o');
xlabel('Actual CO_2 loading');
ylabel('Predicted CO_2 loading');
axis([-2 3 -2 3]);
lsline
%validation
load deal0validation
yval=deal0validation(:,1);
Xval=deal0validation(:,2:861);
%scale the validation data using mean and std
ynval=(yval-meany)/stdy;
Xnvalcenter=bsxfun(@minus,Xval,meanX);
Xnval=bsxfun(@rdivide,Xnvalcenter,stdX);
XSval=Xnval*STATS.W;
%obtained predicted CO2 Loading
yvalpredicted=Xnval*beta(2:end,:);
%test the model performance using validation data
```

```

TSSval=sum((ynval-mean(ynval)).^2);
RSSval=sum((ynval-yvalpredicted).^2);
Rsquaredval= 1 - RSSval/TSSval;
MSEval=RSSval/25;
%Plot the Validation Graph
figure()
plot(ynval,yvalpredicted,'o');
xlabel('Actual CO_2 loading');
ylabel('Predicted CO_2 loading');
lsline

```

## APPENDIX 2

```

%CO2 loading calibration model in 20% DEA
load dea20calib
y = dea20calib(:,1);
X = dea20calib(:,2:861);
%center and scale X and Y
[Xn,meanX,stdX]=zscore(X);
[yn,meany,stdy]=zscore(y);
% Outputs:
% XL      loading matrix of X
% YL      loading matrix of y
% XS      score matrix of X
% YS      score matrix of y
% beta    matrix of regression coefficient
% PCTVAR  weight matrix of X
% STATS consists of Weight, T2, Xresidual and Yresidual
[XL,YL,XS,YS,beta,PCTVAR,MSE,STATS] = plsregress(Xn,yn,10);
plot(1:10,cumsum(100*PCTVAR(2,:)),'-bo');
xlabel('Number of PLS components');
ylabel('Percent Variance Explained in y');
%obtained predicted CO2 Loading
ypredicted=Xn*beta(2:end,:);
%test the model performance
TSS=sum((yn-mean(yn)).^2);
RSS=sum((yn-ypredicted).^2);
Rsquared=1 - RSS/TSS;
MSEcal=RSS/45;
%Plot the Calibration Graph
figure()
plot(yn,ypredicted,'o');
xlabel('Actual CO_2 loading');
ylabel('Predicted CO_2 loading');
axis([-2 3 -2 3]);
lsline
%validation
load dea20valid
yval=dea20valid(:,1);
Xval=dea20valid(:,2:861);
%scale the validation data using mean and std
ynval=(yval-meany)/stdy;
Xnvalcenter=bsxfun(@minus,Xval,meanX);

```

```

Xnval=bsxfun(@rdivide,Xnvalcenter,stdX);
XSval=Xnval*STATS.W;
%obtained predicted CO2 Loading
yvalpredicted=Xnval*beta(2:end,:);
%test the model performance using validation data
TSSval=sum((ynval-mean(ynval)).^2);
RSSval=sum((ynval-yvalpredicted).^2);
Rsquaredval= 1 - RSSval/TSSval;
MSEval=RSSval/25;
%Plot the Validation Graph
figure()
plot(ynval,yvalpredicted,'o');
xlabel('Actual CO_2 loading');
ylabel('Predicted CO_2 loading');
axis([-2 2.5 -2 2.5]);
lsline

```

### APPENDIX 3

```

%CO2 loading calibration model in 30% DEA
load dea30calib
y = dea30calib(:,1);
X = dea30calib(:,2:861);
%center and scale X and Y
[Xn,meanX,stdX]=zscore(X);
[yn,meany,stdy]=zscore(y);
% Outputs:
% XL      loading matrix of X
% YL      loading matrix of y
% XS      score matrix of X
% YS      score matrix of y
% beta    matrix of regression coefficient
% PCTVAR  weight matrix of X
% STATS consists of Weight, T2, Xresidual and Yresidual
[XL,YL,XS,YS,beta,PCTVAR,MSE,STATS] = plsregress(Xn,yn,10);
plot(1:10,cumsum(100*PCTVAR(2,:)),'-bo');
xlabel('Number of PLS components');
ylabel('Percent Variance Explained in y');
%obtained predicted CO2 Loading
ypredicted=Xn*beta(2:end,:);
%test the model performance
TSS=sum((yn-mean(yn)).^2);
RSS=sum((yn-ypredicted).^2);
Rsquared=1 - RSS/TSS;
MSEcal=RSS/45;
%Plot the Calibration Graph
figure()
plot(yn,ypredicted,'o');
xlabel('Actual CO_2 loading');
ylabel('Predicted CO_2 loading');
axis([-2 3 -2 3]);
lsline
%validation

```

```

load dea30valid
yval=dea30valid(:,1);
Xval=dea30valid(:,2:861);
%scale the validation data using mean and std
ynval=(yval-meany)/stdy;
Xnvalcenter=bsxfun(@minus,Xval,meanX);
Xnval=bsxfun(@rdivide,Xnvalcenter,stdX);
XSval=Xnval*STATS.W;
%obtained predicted CO2 Loading
yvalpredicted=Xnval*beta(2:end,:);
%test the model performance using validation data
TSSval=sum((ynval-mean(ynval)).^2);
RSSval=sum((ynval-yvalpredicted).^2);
Rsquaredval= 1 - RSSval/TSSval;
MSEval=RSSval/25;
%Plot the Validation Graph
figure()
plot(ynval,yvalpredicted,'o');
xlabel('Actual CO_2 loading');
ylabel('Predicted CO_2 loading');
axis([-2 2.5 -2 2.5]);
lsline

```

## APPENDIX 4

```
%combined modified model DC3
% y...data matrix CO2 loading
% xa...concentration DEA
% xal..data matrix Raman Intensity
% xb...range carbonate, bicarbonate and carbamate
load calibconc
y=data(:,1);
xa=data(:,2);
xal=data(:,3:862);
xb=data(:,231:263).^2;
xb1=1./xb;
X=[xa,xal,xb1];
%center and scale X and Y
[Xn,meanX,stdX]=zscore(X);
[yn,meany,stdy]=zscore(y);
[XL,YL,XS,YS,beta,PCTVAR,MSE,STATS] = plsregress(Xn,yn,10);
%obtained predicted of Carbon Dioxide loading
ypredicted=Xn*beta(2:end,:);
%test the model performance
TSS=sum((yn-mean(yn)).^2);
RSS=sum((yn-ypredicted).^2);
Rsquared=1 - RSS/TSS;
MSEcal=RSS/135;
%Plot the Calibration Graph
figure()
plot(yn,ypredicted,'o');
xlabel('Actual CO_2 loading');
ylabel('Predicted CO_2 loading');
lsline
%validation
load validconc
yval = validconc(:,1);
xaval=validconc(:,2);
xavall=validconc(:,3:862);
xbval=validconc(:,231:263);
xbvall=1./xbval;
Xval=[xaval,xavall,xbvall];
%scale the validation data using mean and std
ynval=(yval-meany)/stdy;
Xnvalcenter=bsxfun(@minus,Xval,meanX);
Xnval=bsxfun(@divide,Xnvalcenter,stdX);
%obtained yval predicted
yvalpredicted=Xnval*beta(2:end,:);
%test the model performance using validation data
TSSval=sum((ynval-mean(ynval)).^2);
RSSval=sum((ynval-yvalpredicted).^2);
Rsquaredval= 1 - RSSval/TSSval;
MSEval=RSSval/75;
%Plot the Validation Graph
plot(ynval,yvalpredicted,'o');
xlabel('Actual CO_2 loading');
ylabel('Predicted CO_2 loading');
lsline
```