

Article



# Raman Calibration Models for Chemical Species Determination in CO<sub>2</sub>-Loaded Aqueous MEA Solutions Using PLS and ANN Techniques

Ahmad Syukri Hanafiah<sup>1</sup>, Abdulhalim Shah Maulud<sup>1,2,\*</sup>, Muhammad Zubair Shahid<sup>1</sup>, Humbul Suleman<sup>3</sup> and Azizul Buang<sup>1</sup>

- <sup>1</sup> Department of Chemical Engineering, Universiti Teknologi PETRONAS, Bandar Seri Iskandar 32610, Malaysia; ahmad\_0009857@utp.edu.my (A.S.H.); engrzubair86@gmail.com (M.Z.S.); azizul.buang@utp.edu.my (A.B.)
- <sup>2</sup> Centre of Contaminant Control and Utilisation, Universiti Teknologi PETRONAS, Bandar Seri Iskandar 32610, Malaysia
- <sup>3</sup> School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough TS1 3BX, UK; H.Suleman@tees.ac.uk
- \* Correspondence: halims@utp.edu.my; Tel.: +60-5-368-7615

Abstract: The improvement in energy efficiency is recognized as one of the significant parameters for achieving our net-zero emissions target by 2050. One exciting area for development is conventional carbon capture technologies. Current amine absorption-based systems for carbon capture operate at suboptimal conditions resulting in an efficiency loss, causing a high operational expenditure. Knowledge of qualitative and quantitative speciation of CO<sub>2</sub>-loaded alkanolamine systems and their interactions can improve the equipment design and define optimal operating conditions. This work investigates the potential of Raman spectroscopy as an in situ monitoring tool for determining chemical species concentration in the CO<sub>2</sub>-loaded aqueous monoethanolamine (MEA) solutions. Experimental information on chemical speciation and vapour-liquid equilibrium was collected at a range of process parameters. Then, partial least squares (PLS) regression and an artificial neural network (ANN) were applied separately to develop two Raman species calibration models where the Kent-Eisenberg model correlated the species concentrations. The data were paired and randomly distributed into calibration and test datasets. A quantitative analysis based on the coefficient of determination  $(R^2)$  and root mean squared error (RMSE) was performed to select the optimal model parameters for the PLS and ANN approach. The  $R^2$  values of above 0.90 are observed for both cases indicating that both regression techniques can satisfactorily predict species concentration. ANN models are slightly more accurate than PLS. However, PLS (being a white box model) allows the analysis of spectral variables using a weight plot.

**Keywords:** raman spectroscopy; carbon dioxide; monoethanolamine (MEA); partial least square (PLS); artificial neural network (ANN); modified kent eisenberg

# 1. Introduction

Carbon dioxide (CO<sub>2</sub>) is deemed a global environmental concern due to its role in global warming and climate change as a greenhouse gas [1,2]. Industrial activities such as oil and gas processing, power generation and steel and iron production contribute 64% of total CO<sub>2</sub> emissions [3,4]. In the natural gas production industry, CO<sub>2</sub> had been among the main contaminants that need to be removed. In addition to environmental concern, CO<sub>2</sub> is also corrosive to natural gas transmission pipelines with the presence of water leading to the formation of carbonic acid, posing a major safety risk and thus requiring stringent regulation and maintenance to be implemented during operation to manage it [5]. The gross heating value of sales gas is also affected with sale price per volume decreases as CO<sub>2</sub> volume increases [6]. As more low-CO<sub>2</sub> gas fields deplete rapidly and the demand



**Citation:** Hanafiah, A.S.; Maulud, A.S.; Shahid, M.Z.; Suleman, H.; Buang, A. Raman Calibration Models for Chemical Species Determination in CO<sub>2</sub>-Loaded Aqueous MEA Solutions Using PLS and ANN Techniques. *ChemEngineering* **2021**, *5*, 87. https://doi.org/10.3390/ chemengineering5040087

Academic Editors: Alírio E. Rodrigues and Andrew S. Paluch

Received: 21 August 2021 Accepted: 2 November 2021 Published: 14 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). for cleaner energy increases, the need to monetize the high-CO<sub>2</sub> containing natural gas fields increases.

Chemical absorption of  $CO_2$  from gas streams using aqueous amine solvent is considered to be a robust separation technology that had found commercial prevalence in the natural gas processing industry and will remain relevant for years to come [7]. Amine absorption is feasible for  $CO_2$  removal due to its technology maturity, operational robustness and process flexibility. Amine absorption technology also has relatively fast absorption rate and higher capacity compared to other  $CO_2$  removal technologies [8]. Amine reacts with  $CO_2$  through acid-base reaction to form ionic species as well as carbamate in primary and secondary amines due to its basic nature [9]. Due to the plateau growth trends in the development of other  $CO_2$  removal technologies as well as more stringent regulation for  $CO_2$  emission around the globe, it can be predicted that the total application of amine solvent for  $CO_2$  removal applications will continue to grow.

Monoethanolamine (MEA) is an industrial benchmark for  $CO_2$  removal applications and a widely used amine in the industry [10]. Compared to secondary and tertiary amine, the higher reaction rate of MEA with  $CO_2$  proves to be the main advantages of its selection in  $CO_2$  removal applications [11]. MEA is also capable of reducing  $CO_2$  content in a gas stream to a trace level [9].

Increasing the efficiency of an MEA process for CO<sub>2</sub> removal depends considerably upon the understanding of the thermodynamics and kinetics of the system. This could not be properly established unless all the species of the system and their interactions are well known and understood through constant observation and analysis of the system itself. Raman spectroscopy is an analytical technique that is applicable for quantitative analysis of chemical species in a system. Understanding that a Raman spectrum from a CO<sub>2</sub>-loaded aqueous MEA system uniquely characterizes the chemical species present in the system. The information can be utilized for gaining an understanding of the dynamics and kinetics of the system towards optimizing the process [12]. Raman spectroscopy is non-destructive and non-invasive for analysing liquid or gas system as well as capable to analyse wide range of physical state. The technique had been utilized to quantify phase equilibria, transport properties and species' concentration in liquid and gas sample [13–15]. Raman spectroscopy possesses lower water absorptivity compared to Fourier Transform Infrared (FT-IR) technique resulting in less distortion for aqueous system application [15,16].

Raman spectra carry information regarding the chemical species within the monitored system. The information was contained within a specific range of variable within the Raman spectrum which was defined as a characteristic band. The peak (which was contained within the characteristic bands) was formed due to the inelastic Raman scattering as a result of the collision between ions or molecules with the laser beam. The characteristic bands are unique for each chemical species in the CO<sub>2</sub>-loaded MEA system. The observed peak changes within the characteristic bands are proportional to the changes in the respective species' concentration in accordance with Beer-Lambert's Law [17].

Interpreting the spectrum from Raman spectroscopy requires the development of a calibration model specific to the chemical species of interest. The development of such a model requires data in the form of Raman spectra with the respective species concentration determined. The commonly applied method for acquiring species concentration data includes the wet chemistry method and application of the thermodynamic model. The wet chemistry method was the earliest technique to be utilized for quantifying species concentration in a system. It is often used due to the unavailability of an accurate thermodynamic models [18]. However, the method is prone to human errors due to dependency on personnel for properly conducting the laboratory work. In addition, there are also chemical species that are difficult to synthesize in the laboratory, thus preparing standard solution for the species is difficult [12] and requires a large number of man and machine hours for overall analysis [4,13].

Application of a thermodynamic model for  $CO_2$ -loaded amine speciation analysis can minimize human errors and overall time required as well as provides means to estimate species concentration of chemical species which are difficult to synthesize. The Kent–Eisenberg model is a thermodynamic model developed for acid gases-loaded amine systems [19]. The model is well-known for its simplicity that is achieved by lumping all the non-idealities into the equilibrium constants which are then fitted to the experimental data, resulting in quick estimation while maintaining good correlative accuracy [20]. Other complex models use excess Gibbs energy or a combination of Equation of State and Gibbs energy approach, requiring a substantial number of regression parameters for iteration compared to the much simpler Kent–Eisenberg model [21]. The measurement accuracy of a modified version of the Kent–Eisenberg model was proven to be good within the range of 0.2 to 0.8 mol of CO<sub>2</sub> per mol of amine [22–24].

Developing a Raman calibration model requires the application of a multivariate calibration technique. The requirement is due to the multivariate nature of the Raman spectrum. Each Raman spectrum is a linear combination of the pure compound spectra from chemical species in the system [12]. A Raman calibration model can be developed by first identifying the spectral peak of the species of interest. The spectral peak is identified by comparing the spectra with that from pure compound. The peak height or area of the identified spectra is then quantified against the measured species' concentration [6]. For a complex system with various chemical species, multiple spectral peaks may overlap causing difficulties in identifying and measuring the respective peak area. Each variable within the Raman spectrum is also highly correlated to each other and needs to be uncorrelated before the species calibration model can be constructed. This requires the application of a dimensionality reduction technique.

Principal component analysis (PCA) is a dimensionality reduction technique that transforms the original variables into new, orthogonal variables called principal components. Each principal component is a linear combination of the original variables [25,26]. Construction of principal components maximizes the percentage of explained variance for the first component. The percentage subsequently reduced as additional principal components are added. Such a hierarchical structure is attained as principal components are constructed to be orthogonal to each other. The orthogonality and hierarchical structure of principal components allow relevant components with a high percentage of explained variance to be identified. The components can then be used for developing the Raman calibration model by applying a regression technique such as principal component regression (PCR) [27].

The concept of variable transformation towards orthogonality in PCA was wellintegrated in the more advanced regression technique, called partial least squares (PLS) regression. The PLS regression model was developed using components that were constructed to maximize the correlation between predictor and response variables while also maximizing the explained variance within each predictor and response variables [28]. The approach in the PLS technique allows the construction of a simpler model compared to PCR. However, modelling a non-linear relationship using a linear regression technique such as PLS could impact accuracy. A non-linear relationship can exist in Raman spectral data due to overlapping signals, inconsistency in particle sizes within the sample, contaminants from external or ambient light and use of non-homogeneous or light-absorbing samples.

Modelling such a non-linear relationship might require the application of a non-linear regression technique. With the advancements made in the field of machine learning, the application of an artificial neural network (ANN) in chemometrics is also gaining traction [29]. ANN is a non-linear regression technique by virtue of utilizing a non-linear activation function. This capability can potentially improve measurement accuracy made by a process analyser compared to PCR and even PLS-based calibration model in real-world application [30–32].

Raman spectroscopy has been applied previously to quantify species concentration in a  $CO_2$ -loaded aqueous MEA system. Souchon et al. studied species distribution in various amine types using Raman spectroscopy. Species concentration was derived from the reference solutions that were measured by a <sup>13</sup>C nuclear-magnetic resonance (NMR) spectrometer. The work concluded that Raman spectroscopy can achieve lower analysis time compared to NMR due to the latter requires longer relaxation times per spectrum. The work also highlighted the importance of pairing a thermodynamic model with spectral measurement tools in estimating the quantities of carbamate ion, which are not measurable through the wet chemistry method [12]. Wong et al. investigated the chemical speciation in CO<sub>2</sub>-loaded aqueous MEA at varying pressure and temperature ranges using Raman spectroscopy. The author developed calibration models by manually plotting peak area ratio from the spectrum against species concentration, which was derived through the wet chemistry method. The models can predict species concentration with good accuracy [6]. Jinadasa et al. investigated the application of Raman spectroscopy for CO<sub>2</sub>-loaded, 30 wt% MEA solutions at room pressure and temperature. Using the PLS regression technique combined with offline titration analysis and Kent–Eisenberg thermodynamic models, the Raman calibration models are developed and the in situ monitoring technique was tested with good results [33,34]. Zubair et al. utilized Raman spectroscopy for measuring CO<sub>2</sub> concentration in aqueous diethanolamine (DEA), methyldiethanolamine (MDEA) as well as a combination of both. The development of a calibration model in their work utilized PLS regression and a model was developed to include  $CO_2$  loading range above 0.5 for high-pressure applications [35]. Similarly, the author also investigated the application of Raman spectroscopy for monitoring species concentration in CO<sub>2</sub>-loaded aqueous DEA systems with a  $CO_2$  loading range of up to 0.98, utilizing the PLS regression technique for developing a calibration model [36]. It can be established that previous research work within the area focused mainly on the application of PLS regression methods, while other multivariate regression methods remained unexplored.

In this work, Raman species calibration models for CO<sub>2</sub>-loaded aqueous MEA system at 3, 4 and 5 molar MEA concentrations were developed using PLS regression and ANN. To develop the calibration models, Raman spectra of the system were first acquired using a vapour-liquid equilibrium (VLE) experimental setup with Raman spectroscopy installed. Next, CO<sub>2</sub> loading values were calculated using a pressure drop method with ideal system assumed. The experimental VLE data together with CO<sub>2</sub> loading were used on the regressed modified Kent–Eisenberg model to calculate the species concentration. The smoothed and centred Raman spectra together with the respective species concentration data were then utilized for developing the Raman calibration models using PLS regression and ANN techniques. Comparisons were made between the developed models based on the predictive performance by evaluating the coefficient of determination (R<sup>2</sup>) and root mean squared error (RMSE).

#### 2. Materials and Methods

# 2.1. Materials

Materials required for conducting the VLE experiment are detailed in Table 1. Monoethanolamine was sourced from Merck Sdn. Bhd, while deionized water was obtained in-house. Carbon dioxide gas was procured from Linde Malaysia Sdn. Bhd.

14010 11	Tuble	or materix	10 101 11	eexperim	citt.	

Table 1 Table of materials for the experiment

Materials	Abbreviation	CAS No.	Purity
Monoethanolamine	MEA	141-43-5	0.995 (mass fraction)
Carbon dioxide	CO <sub>2</sub>	124-38-9	0.9990 (volume fraction)
Water	H <sub>2</sub> O	7732-18-5	~1.000 (mass fraction)

Aqueous MEA solutions at 3, 4 and 5 molar strength were prepared by first weighing the required quantity of MEA on a Sartorius BSA2245S-BW mass balance. The weighted MEA was added to a 250 mL volumetric flask. Deionized water was added to the volumet-

ric flask to prepare the solution at the required concentration. Each solution was prepared at a controlled temperature of 298.15 K.

# 2.2. Experimental Setup and Method

The VLE experimental setup was constructed as depicted in Figure 1. The setup consists of three sub-units which are CO<sub>2</sub>-amine loading cell, Raman spectroscopy unit and data acquisition unit. The CO<sub>2</sub>-amine loading cell consists of a 435 mL feed vessel connected to a 460 mL stainless steel reaction vessel. Both vessels are built with integrated pressure and temperature sensors to measure conditions and gathering experimental data. A CO<sub>2</sub> feed line was connected to the feed vessel to supply CO<sub>2</sub> gas into the experimental setup. A ceramic heater was installed around the reaction vessel to regulate system temperature. A CO<sub>2</sub> vent line was installed at the top side of the reaction vessel to allow vessel degassing. A magnetic stirrer was used inside the reaction vessel throughout the experimental work to maintain mass transfer between the gas and liquid phase.



Figure 1. Experimental setup to collect vapour-liquid equilibrium (VLE) data and Raman spectrum.

A Raman spectroscopy unit was procured from StellarNet. The unit consists of a Raman spectrometer and laser source. The Raman spectrometer has a resolution of  $4 \text{ cm}^{-1}$  with signal to noise ratio of 1000:1 while the laser source uses 500 W of power-producing laser beam with a wavelength of 785 nm. A probe connected to the Raman spectroscopy unit by an optical fibre cable was mounted at the side of the reaction vessel where a sapphire-glass window was built. The probe directed the laser beam towards the system and retrieved the resulting Raman scattering for the Raman spectrometer. The data acquisition unit was pre-installed with SpectraWiz software and serves to process the data received from the Raman spectroscopy unit.

In a typical run, the prepared aqueous MEA solution was firstly loaded into the reaction vessel. The temperature for each run was set at 303.15 K.  $CO_2$  gases were then introduced into the reaction vessel and the aqueous MEA solution was continuously stirred until an equilibrium state was reached. The temperature and pressure of the system were continuously recorded using built-in temperature and pressure transmitters throughout the experimental run for  $CO_2$  loading calculation using Equation (1) [15,35].

$$\alpha = \frac{\left[V_{FV}\left(P_i - P_f\right) - \left(P_{RV}\left(V_{RV} - V_{amine}\right)\right)\right]}{RTn_{amine}} \tag{1}$$

After the equilibrium state was reached,  $CO_2$  was further introduced into the reaction vessel and the process was repeated until the targeted  $CO_2$  loading in the aqueous MEA solution was reached inside the reaction vessel. The recorded Raman spectra were within the range of 0 to 2800 cm<sup>-1</sup>. Forty (40), 67 and 66 datapoints for the Raman spectra were gathered for 3, 4 and 5 molar aqueous MEA solutions, respectively.

### 2.3. Thermodynamic Framework and Kent–Eisenberg Model

The modified Kent–Eisenberg model was used to estimate the species concentration for the  $CO_2$ -loaded aqueous MEA solutions in equilibrium [20]. The model is an improved version based on the work by Kent and Eisenberg [19]. Aqueous MEA reaction with  $CO_2$ follows an acid-base reaction and produced numerous ions. Equations (2)–(6) describe  $CO_2$ dissolution into MEA.

Dissociation of a protonated alkanolamine:

$$AlkH^+ \leftrightarrow Alk + H^+; K_1 \tag{2}$$

Reversion of carbamate to bicarbonate:

$$AlkCOO^{-} + H_2O \leftrightarrow Alk + HCO_3^{-}; K_2$$
(3)

Hydrolysis and ionization of dissolved CO<sub>2</sub>:

$$CO_2 + H_2O \leftrightarrow H^+ + HCO_3^-; K_3$$
(4)

Dissociation of bicarbonate:

$$HCO_3^- \leftrightarrow H^+ + + CO_3^{2-}; K_4$$
(5)

Ionization of water:

$$H_2O \leftrightarrow H^+ + OH^-; K_5$$
 (6)

Based on the defined chemical reactions, the equilibrium constants for each reaction are as follows:

$$K_1 = \frac{\left[\text{Alk}\right]\left[\text{H}^+\right]}{\left[\text{AlkH}^+\right]} \tag{7}$$

$$K_2 = \frac{[\text{Alk}][\text{HCO}_3^-]}{[\text{AlkCOO}^-]}$$
(8)

$$K_{3} = \frac{[\text{HCO}_{3}^{-}][\text{H}^{+}]}{[\text{CO}_{2}]}$$
(9)

$$K_4 = \frac{[\mathrm{CO}_3^{2-}][\mathrm{H}^+]}{[\mathrm{HCO}_3^-]} \tag{10}$$

$$K_5 = \left[ \mathbf{H}^+ \right] \left[ \mathbf{O} \mathbf{H}^- \right] \tag{11}$$

The value for equilibrium constants  $K_2$ ,  $K_3$ ,  $K_4$  and  $K_5$  were calculated based on the work of Edwards et al. [37], whereas the value of  $K_1$  was treated as a fitting parameter to the experimental values. To fit the experimental loading data into the model, a parameter  $k_1^m$  is defined where:

$$k_1^m = k_1 f \tag{12}$$

As for *f*:

$$f = qln(P_{CO_2}) + rln(Alk_{total})$$
(13)

The purpose of fitting the model with experimental loading data is so that accurate calculation of speciation data is acquired. The defined  $k_1^m$  serve as an adjustable parameter to improve the prediction accuracy of the model towards the experiment parameters [6,20,38]. The unknowns were found by simultaneously solving the five reaction equations and three balance equations detailed below:

Amine balance:

$$[Alk_{total}] = [Alk] + [AlkH^+] + [AlkCOO^-]$$
(14)

CO<sub>2</sub> balance:

$$\alpha[\text{Alk}_{\text{total}}] = [\text{HCO}_3^-] + [\text{AlkCOO}^-] + [\text{CO}_3^{2-}] + [\text{CO}_2]$$
(15)

Charge/electroneutrality balance:

$$[AlkH^{+}] + [H^{+}] = [OH^{-}] + [HCO_{3}^{-}] + [AlkCOO^{-}] + 2[CO_{3}^{2-}]$$
(16)

Henry's law is applied to quantify the concentration of carbon dioxide dissolved in the liquid phase:

ſ

$$CO_2] = \frac{P_{CO_2}}{H_{CO_2}}$$
 (17)

The equations were solved in a polynomial form:

$$A[H^{+}]^{5} + B[H^{+}]^{4} + C[H^{+}]^{3} + D[H^{+}]^{2} + E[H^{+}]^{1} + F = 0$$
(18)

where:

$$A = 1$$
  

$$B = [Alk_{total}] + k_1^m$$
  

$$C = -k_5 - k_3[CO_2] + \frac{k_1^m k_3[CO_2]}{k_2}$$
  

$$D = -\frac{[Alk_{total}]k_1^m k_3[CO_2]}{k_2} - k_1^m k_5 - k_1^m k_3[CO_2] - 2k_3k_4[CO_2]$$
  

$$E = -\frac{k_1^m k_3 k_5[CO_2]}{k_2} - \frac{k_1^m k_3^2[CO_2]^2}{k_2} - 2k_1^m k_3 k_4[CO_2]$$
  

$$F = -\frac{2k_1^m k_3^2 k_4[CO_2]^2}{k_2}$$

Equation  $(18)^{5}$  was solved using MATLAB to find the roots. Five numerical values were obtained, in which one of them is the concentration of hydrogen ion. The valid hydrogen ion concentration is within the range of  $10^{-7}$  to  $10^{-12}$  molar with respect to the pH value of between 7 and 11 in the loaded and unloaded MEA solution [22]. The concentration of each chemical species was then calculated using Equations (7)–(11).

#### 2.4. Calibration Models Development and Evaluation

The gathered Raman spectra and the species concentration estimated using the regressed modified Kent–Eisenberg model were paired and classified according to the total MEA concentration in the system from which the data originated. The classification of the dataset is presented in Table 2. Each dataset was then randomly divided into training and test datasets. The training dataset was used to develop a calibration model of the chemical species. The test dataset was then used to validate the performance of the developed calibration model externally, thus the test dataset was excluded from the model training. As for ANN, the training data were further divided randomly into a validation dataset which was used during model training to pre-validate model performance for early stopping implementation. Generally, 60% to 70% of the data were randomly assigned as training data while the remaining 30% to 40% were assigned as test data [39–41].

MEA Concentration (Molar)	Number of Datapoints	CO <sub>2</sub> Loading Range (Moles of CO <sub>2</sub> /Moles of MEA)	
3	40	0.2593-1.2838	
4	67	0.0261-1.0418	
5	66	0.1754-1.0232	

**Table 2.** Table of materials for the experiment.

Before model development, Raman spectra were preprocessed through the application of Savitzky-Golay smoothing, followed by mean centring and scaling of the spectra. Savitzky-Golay smoothing was performed to reduce noise while preserving information contained in characteristic peaks within the spectrum. The application was through a builtin filter in MATLAB with smoothing degree and range of filter span adjustment dictates the smoothing effects on the individual spectrum. The optimal smoothing parameters were determined by observing the filter effects on the performance of the developed models. Mean centring was then performed on each Raman spectrum by subtracting the mean value of a variable from the respective variable in each observation in the MATLAB environment.

Raman calibration models were developed by applying PLS and ANN techniques on the preprocessed Raman spectra and the estimated species concentration, which served as a predictor and response data, respectively. While the form of response data used for model developments was similar for both techniques, the form of the predictor data differs between both techniques. For the PLS model, preprocessed Raman spectra were used directly while for ANN, the PCA score from the transformed Raman spectra was used. Full Raman spectrum range (160 to 2800 cm<sup>-1</sup>) were utilized for the model development.

PLS-based Raman species calibration models were developed using the PLSREGRESS function in MATLAB, which utilized the SIMPLS algorithm for fast iterations [28,42]. During the execution of the PLSREGRESS function, both Raman spectra and species concentration data were each transformed into a PLS component, which is a linear combination of each variable within the data. The PLS component was constructed such that the variance explained by the component as well as the correlation between the predictor and the response variables were maximized. Each of the constructed PLS components was also orthogonal to the others. The orthogonality resulted from the deflation step performed at the end of each component from the error matrix while preserving the unexplained by the current PLS components were then constructed using the error matrix. Initial PLS-based models were developed using 15 components. Such a large number of components is to ensure that all-important information was considered for developing a calibration model for a particular response variable. Diagnostics from the developed model were then used to select an optimal number of components for the final model.

Application of the PLSREGRESS function on the dataset produces several outputs, which include the loadings and scores for the predictor and response data, the PLS regression coefficients, and a matrix containing the percentage of variance explained by the developed model. Using the percentage-explained variance data, a PLS variance plot (which is a plot of cumulative percent explained variance against the PLS component) was drawn. The PLS plot was used to estimate the potential number of PLS components to be used in developing the calibration model. These PLS components were determined up to the point where the variance profile flattened. Beyond that point, the components do not contain significant information towards the model and will lead to overfitting [27]. An example of a scree plot is depicted in Figure 2 revealing that mean squared error (MSE) reduced as the number of components increased. However, the plotted trend flattens after 6 components. The area before the 6th component was defined as scree region and was retained in the final model. A similar method was conducted in selecting the number of components for PLS-based calibration models.



Figure 2. Scree plot for partial least squares (PLS)-based CO<sub>2</sub> loading Raman calibration model.

ANN-based Raman species calibration models were developed in MATLAB using the built-in Neural Network Toolbox combined with manual parameter adjustments. An untrained neural network architecture was initiated with three layers as default consisting of one input layer, one hidden layer and one output layer. A single hidden layer can act as a universal approximator capable of modelling any functions. Although there is a possibility of attaining quicker training time, additional hidden layers rarely improve the performance of calibration models versus using a single hidden layer [43–45].

For the input layer, raw Raman spectra were first compressed into principal components (PC) using principal component analysis (PCA). The score was then used as input for ANN. The step was implemented to reduce the dimension and extract only the relevant features from the data for lower memory requirement and faster iteration time [26,46,47]. The number of input nodes was set based on the number of principal components extracted using PCA. The number of output nodes was set to one to account for the respective chemical species concentration. Similar to the consideration made for PLS models, 15 LVs was selected for the initial training. The number is then further lowered during optimization. Additionally, the iterative optimization process on neural network structure during training will minimize the significance from less relevant input towards the resulting output [48]. One hidden layer was set for the initial design with one node to constraint model complexity and minimize overfitting. Additional nodes improve the prediction performance in the presence of slight non-linearity, which is common in multivariate datasets [48]. The number of nodes was subsequently adjusted based on the observed model performance. For the output layer, one node was used similar to the number of targeted outputs. The tan-sigmoid function was set as an activation function to allow flexibility for modelling non-linearity during training. MSE was implemented as the cost function to measure the performance and as a learning parameter for the optimization algorithm during ANN training. Random weights were assigned with the values constrained within -1 and 1 corresponding to the assigned activation function for the network.

The initiated neural network architecture was then trained using the Levenberg– Marquadt optimization algorithm which employed a backpropagation algorithm. Early stopping was implemented by using 10% to 15% of randomly selected training data as a monitoring set during training. The training stopped after six iterations if no improvement in MSE value was observed for the training dataset compared to the monitoring set to minimize overfitting. The training progression after each iteration is depicted in Figure 3 from the training calibration model for  $CO_2$  loading using three molar total MEA datasets. The MSE values for validation and test datasets was observed to decrease from the start before increasing after the third iteration. The training was stopped based on the trends of MSE values for validation data which do not improve after the third iteration. Network parameters at the third iteration were then used for the resulting model.



Figure 3. Plot of mean squared error (MSE) values after each iteration during training.

Ten training runs were conducted per model using randomized initial weights to improve model parameters towards achieving the best possible solution. The trained models were then evaluated based on the calculated MSE values. Adjustments were made on the neural network parameters based on the training results. The model was then retrained with the implemented adjustment and the training process was monitored using validation data. The overall training process for the model concluded once the MSE values stopped improving.

The performance of each developed model was evaluated by assessing the RMSE and coefficient of determination ( $R^2$ ) of the model on the training dataset.  $R^2$  value provides an overview of the tightness of the predicted and actual data towards the regression line, while RMSE evaluates the closeness of the predicted value to the actual value. A model with  $R^2$  value of above 0.90 indicates excellent correlations between Raman spectrum and species concentration. Also, models with an  $R^2$  value of 0.83 to 0.90 are still applicable within a limited scope of use, typically in research work [49].

Selection of the final model parameters was conducted through external validation using the previously split test datasets. Externally validating the model with an independent test dataset allows for unbiased evaluation of the developed calibration model. RMSE for test datasets was compared to the similar RMSE plot for training datasets as observed in the examples in Figures 2 and 3 for PLS and ANN, respectively. The plot shows that the quality of the model does not improve with an additional number of components beyond six, which was consistent with the finding from the scree plot. Further evaluation of  $R^2$  was conducted using a test dataset. Additionally, regression plots of the training and test dataset for each calibration model were constructed to analyse the trends in datapoints distribution along the regression line. A total of 21 Raman calibration models were developed using each technique, with seven models each at 3, 4 and 5 molar MEA concentration. The seven models were for protonated MEA, bicarbonate, carbonate, unreacted MEA, carbamate, unreacted CO<sub>2</sub>, and CO<sub>2</sub> loading.

# 3. Results and Discussion

# 3.1. Characteristic Bands of Chemical Species in CO<sub>2</sub>-Loaded Aqueous Monoethanolamine (MEA) System

Comparison between the smoothed spectra of the unloaded and the CO<sub>2</sub>-loaded MEA solution were made as exemplified in Figure 4. Significant peaks were observed at Raman shift range between 800 cm<sup>-1</sup> and 1100 cm<sup>-1</sup> as well as 1250 cm<sup>-1</sup> and 1400 cm<sup>-1</sup>. The observed peaks were due to changes in species composition caused by CO<sub>2</sub> absorption

and chemical reactions between species as described in Equations (7)–(11). The ionic species present in the  $CO_2$ -loaded aqueous MEA solution were protonated MEA, bicarbonate, carbonate, and carbamate while non-ionic species include unreacted MEA and dissolved  $CO_2$ .



Figure 4. Plots of Raman spectra captured from unloaded and  $CO_2$ -loaded aqueous monoethanolamine (MEA) solution. Each are for (a) combined plot, (b) unloaded MEA, and (c) loaded MEA.

For the ionic species, the characteristic band for bicarbonate ion is visible at  $1017 \text{ cm}^{-1}$  (C–OH stretching) while the characteristic band for carbonate is located at  $1065 \text{ cm}^{-1}$  (C–O bond stretching). The characteristic band for carbamate is observed at  $1162 \text{ cm}^{-1}$  (C–O bond stretching) and for protonated MEA at  $1011 \text{ cm}^{-1}$  (CH<sub>2</sub> twisting),  $1274 \text{ cm}^{-1}$  (C–NH stretch) and  $1320 \text{ cm}^{-1}$  (C–C stretch). For non-ionic species, the dissolved CO<sub>2</sub> characteristic bands are observed at  $1274 \text{ cm}^{-1}$  (CO<sub>2</sub> symmetric stretch) and  $1383 \text{ cm}^{-1}$  (CO<sub>2</sub> bend), while the band for unreacted MEA is observable at 845 cm<sup>-1</sup> (CH<sub>2</sub> rocking) and 873 cm<sup>-1</sup> (CN stretching) [4,6,12,50,51].

It was observed that peak intensity varies at the spectral ranges containing the characteristic bands for the chemical species in the system. In contrast, minimal variations were observed at other spectral ranges. As an example, the Raman spectra in Figure 4 show that spectral intensity at characteristic bands for bicarbonate (1017 cm<sup>-1</sup>), carbonate (1065 cm<sup>-1</sup>) and aqueous CO<sub>2</sub> (1274 cm<sup>-1</sup>) increases. The observation agrees with the understanding that ionic species of bicarbonate and carbonate increases in concentration as CO<sub>2</sub> loading increases. As the pressure inside the reaction vessel increases, CO<sub>2</sub> began to dissolve physically into the solution. The physical dissolution increases aqueous CO<sub>2</sub> concentration in the liquid phase. For carbamate, the concentration increases up to a certain point before decreasing as the pressure increases. This follows the theory that carbamate production from the MEA reaction with CO<sub>2</sub> limits the capacity for absorption to 0.5 mol of  $CO_2$  to mol of amine. As pressure increase, hydrolysis of carbamate forms free amine capable to react with  $CO_2$  leading to loading of above 0.5 [23,24].

# 3.2. Evaluation of the Raman Species Calibration Models

Evaluation on the final Raman calibration models was performed by comparing the coefficient of determination ( $R^2$ ) and RMSE values from the training dataset against the values of the test dataset. The closeness of the  $R^2$  and RMSE values between training and test datasets indicate that the model had good generalization towards external data and overfitting is minimal. The overall results for the performance evaluation and validation for PLS and ANN-based calibration models are summarized in Tables 3 and 4, respectively.

Table 3. Summary of performance evaluation results for PLS-based Raman species calibration models.

Species/CO <sub>2</sub> Loading	Amine Concentration (mol/L)	Number of LV Components	R <sup>2</sup> Calibration	R <sup>2</sup> Validation	RMSE Calibration	RMSE Validation
	3	6	0.9636	0.9319	0.0510	0.0762
CO <sub>2</sub> Loading	4	6	0.9438	0.9075	0.0771	0.0790
- 0	5	6	0.9562	0.9092	0.0652	0.0758
	3	6	0.9437	0.9122	0.1497	0.2110
MEAH <sup>+</sup>	4	6	0.9400	0.9031	0.2653	0.3564
	5	6	0.9460	0.9137	0.3256	0.3655
	3	5	0.9466	0.9139	0.1539	0.2381
$HCO_3^-$	4	6	0.9421	0.9027	0.2727	0.3407
Ũ	5	5	0.9485	0.9039	0.3181	0.4228
	3	6	0.9614	0.9133	0.0067	0.0072
$CO_{3}^{2-}$	4	6	0.9361	0.9031	0.0027	0.0035
0	5	6	0.9321	0.9254	0.0052	0.0059
	3	6	0.9571	0.9201	0.1136	0.1095
MEA	4	7	0.9304	0.9074	0.2820	0.3163
	5	6	0.9211	0.9017	0.3736	0.3963
	3	6	0.9755	0.9361	0.0224	0.0436
MEACOO <sup>-</sup>	4	9	0.9497	0.9082	0.0364	0.0518
	5	8	0.9282	0.9003	0.0565	0.0661
	3	6	0.9396	0.9383	0.0616	0.0548
CO <sub>2</sub>	4	6	0.9358	0.9147	0.0439	0.0580
_	5	6	0.9311	0.9050	0.0497	0.0460

It was observed that the  $R^2$  values for all calibration models are above 0.90 indicating good fit [49]. For PLS models, the average  $R^2$  value is 0.9442 for calibration and 0.9050 for validation while the average RMSE value is 0.1301 for calibration and 0.1583 for validation. As for ANN models, the average  $R^2$  value is 0.9505 for calibration and 0.9355 for validation while the average RMSE value is 0.0965 for calibration and 0.1225 for validation. The average value of RMSE for validation (RMSEV) is observed to be higher than RMSE for calibration (RMSEC) while the average value for  $R^2$  calibration is higher than  $R^2$  validation on both types of model. The observations were due to the use of test data for external validation which carries variance that the model never observed during development. Such an approach was to ensure that the validated model is parsimonious [13,52–54]. The differences between the average values were also small, indicating that both PLS and ANN-based calibration models can predict the species concentration in the CO<sub>2</sub>-MEA-H<sub>2</sub>O system with good accuracy.

Species/CO <sub>2</sub> Loading	Amine Concentration (mol/L)	Number of Principal Components	Number of Hidden Nodes	R <sup>2</sup> Calibration	R <sup>2</sup> Validation	RMSE Calibration	RMSE Validation
<u> </u>	3	2	2	0.9527	0.9582	0.0469	0.0616
Londing	4	2	2	0.9305	0.9377	0.0600	0.0938
Loading	5	2	2	0.9589	0.9069	0.0539	0.0592
	3	1	1	0.9758	0.9584	0.1072	0.1439
MEAH <sup>+</sup>	4	1	1	0.9225	0.9250	0.2956	0.3038
	5	2	2	0.9521	0.9543	0.2377	0.2421
	3	1	1	0.9779	0.9528	0.1105	0.1425
$HCO_3^-$	4	2	2	0.9453	0.9476	0.2352	0.2975
	5	3	3	0.9232	0.9010	0.1559	0.2124
	3	3	3	0.9788	0.9796	0.0035	0.0042
$CO_{3}^{2-}$	4	2	2	0.9397	0.9032	0.0023	0.0028
0	5	3	3	0.9833	0.9490	0.0024	0.0038
	3	1	1	0.9215	0.9041	0.1315	0.1712
MEA	4	2	2	0.9363	0.9062	0.1881	0.2715
	5	2	2	0.9517	0.9574	0.2191	0.2900
	3	2	2	0.9783	0.9400	0.0111	0.0238
MEACOO-	4	3	3	0.9421	0.9394	0.0265	0.0616
	5	3	3	0.8835	0.8621	0.0424	0.0693
	3	2	2	0.9652	0.9423	0.0412	0.0574
CO <sub>2</sub>	4	2	2	0.9612	0.9516	0.0332	0.0387
	5	2	2	0.9803	0.9692	0.0220	0.0223

Table 4. Summary of performance evaluation results for artificial neural network (ANN)-based Raman species calibration models.

3.3. Comparative Analysis between Partial Least Squares (PLS) and Artificial Neural Network (ANN) Technique for the Development of Raman Calibration Model

Comparisons were made between the final PLS and ANN-based calibration models to determine the technique that produced better performing models. Comparisons were based on the average values of the performance parameters for calibration data. The results are summarized in Table 5.

**Table 5.** Comparison between average values of the performance parameters between PLS and ANN models.

Species/CO <sub>2</sub> Loading	PLS R <sup>2</sup> Validation	ANN R <sup>2</sup> Validation	PLS RMSE Validation	ANN RMSE Validation
CO <sub>2</sub> Loading	0.9162	0.9343	0.0770	0.0715
MEAH <sup>+</sup>	0.9097	0.9459	0.3110	0.2299
HCO <sub>3</sub> -	0.9068	0.9338	0.3339	0.2175
$CO_{3}^{2-}$	0.9139	0.9439	0.0055	0.0036
MEA	0.9097	0.9226	0.2740	0.2442
MEACOO <sup>-</sup>	0.9149	0.9138	0.0538	0.0516
CO <sub>2</sub>	0.9193	0.9544	0.0564	0.0395

Comparing the R<sup>2</sup> values, both techniques resulted in comparable calibration models with R<sup>2</sup> values above 0.9 indicating a good fit. As for the average RMSE values, ANN-based calibration models have better prediction accuracy compared to PLS-based calibration models. This was indicated by the overall lower average RMSE values observed for all ANN models compared to PLS models. The better performance recorded by ANN is due to the application of the non-linear transfer function that transforms original input into non-linear projection at lower-dimensional subspace. Compared to the linear projection in PLS, the non-linear projection in ANN is more flexible, allowing the ANN model to better

fit the datasets. The better fit translates to lower RMSE values for ANN models compared to PLS models [55].

An advantage of applying PLS over ANN is the ability to analyse the model parameters to understand the significance of each Raman variables within the modelled system. It was previously discussed that Raman spectra carry information regarding the chemical species within the monitored system. The information was contained within a specific range of the Raman spectrum called the characteristic band. The characteristic band for a particular chemical species contains spectral variables that are highly correlated with the chemical species. The spectral variables thus contributed highly towards explaining the variance of the species correlated with them in the model. For a PLS calibration model, the contribution was reflected in the PLS weight plot. An example of a weight plot is depicted in Figure 5. The plot was generated from the calibration model of bicarbonate at 4 molar total MEA concentration The plot shows multiple peaks appearing between 500 and  $1500 \text{ cm}^{-1}$ , indicating that variables within that range are highly correlated to the response variables. The large PLS weight values of the highly correlated variables are evidence that they contribute the most towards the variance explained in the model. The observation is reflective of the literature finding where the characteristic band for bicarbonate is located at 1017 cm<sup>-1</sup>. Other peaks are from weak intensity signals at 632 cm<sup>-1</sup>, 1360 cm<sup>-1</sup> and 1630 cm<sup>-1</sup> due to (OH)CO band, symmetric CO stretching and anti-symmetric CO<sub>2</sub> stretch, respectively [56,57].



Figure 5. Weight plots from bicarbonate calibration model (4 molar MEA) for (a) PLS Component 1 and (b) PLS Component 2.

The results from performance evaluation are further supported by observing the regression plots. A good model shall have the datapoints distributed along the regression line. Examples of regression plots are depicted in Figure 6, taken from calibration models for 3 molar MEA concentration. It was observed that the datapoints are well-distributed along the straight line on all models. There was no noticeable trending observed in the distribution of datapoints, indicating that the current model is sufficiently parametrized in explaining the variance in the data. A small deviation was observed between actual and predicted datapoints which was reflected by the resulting RMSEC and RMSEV values.



**Figure 6.** Regression plots from 3 molar MEA calibration models developed using PLS and ANN techniques.

It was also noted that during the development of the calibration model, the whole spectrum range was utilized. However, the characteristic bands for a chemical species are present only at a specific spectral range within a Raman spectrum based on literature. Other variables outside the spectral range are considered as noises. Developing calibration models by utilizing specific spectral ranges that were correlated with the particular species could reduce noises that interfere during model development thus improving performance [58,59].

#### 4. Conclusions

Development of Raman calibration models utilizing two different multivariate regression techniques, which are partial least squares (PLS) regression and an artificial neural network (ANN), were explored. The Raman spectra and respective species concentration were acquired through the VLE experiment of the CO<sub>2</sub>-MEA-H<sub>2</sub>O system at varying total MEA concentration. The VLE data from the experiment were then further used to predict species concentration inside the CO<sub>2</sub>-MEA-H<sub>2</sub>O system using the regressed Kent-Eisenberg thermodynamic model. A total of 21 Raman species calibration models were developed using each technique utilizing smoothed Raman spectra and the estimated species concentration. The performance of each model was evaluated using performance parameters of R<sup>2</sup> and RMSE. The R<sup>2</sup> values from the developed calibration models were above 0.90 while the RMSE values improved throughout model optimization. Comparison of both techniques shows that ANN models were more accurate with lower RMSE values observed compared to the PLS model. However, the small differences between the RMSE values indicate that PLS models were comparatively good in predicting species concentration. Additionally, the PLS model allows studies on the significance of each variable to be conducted through analysing the weight plots, which was not possible in ANN models, this being a black box technique. The regression plots show close data distribution along the regression line with no trending observed from the data distribution for both calibration and test datasets. The result signifies that both techniques are applicable for developing Raman species calibration models to predict species concentration with good accuracy in CO<sub>2</sub>-loaded aqueous MEA system. Notably, PLS provides added advantages over ANN in terms of analysis on the significance of each spectral variable on the developed models.

**Author Contributions:** Conceptualization, A.S.H. and A.S.M.; methodology, A.S.H., M.Z.S. and H.S.; validation, A.S.H., A.S.M. and M.Z.S.; formal analysis, A.S.H. and M.Z.S.; investigation, A.S.H.; data curation, A.S.H. and M.Z.S.; writing—original draft preparation, A.S.H.; writing—review and editing, A.S.H., A.S.M., M.Z.S. and H.S.; supervision, A.S.M. and A.B.; funding acquisition, A.S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by PETRONAS Research Sdn Bhd, grant number 0153B2-E76. Any opinions and findings as well as conclusions and recommendations made in this material are those of the author(s) and do not necessarily reflect the view of PETRONAS Research Sdn Bhd.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- Luis, P. Use of monoethanolamine (MEA) for CO<sub>2</sub> capture in a global scenario: Consequences and alternatives. *Desalination* 2016, 380, 93–99. [CrossRef]
- Fan, G.J.; Wee, A.G.H.; Idem, R.; Tontiwachwuthikul, P. NMR studies of amine species in MEA-CO<sub>2</sub>-H<sub>2</sub>O system: Modification of the model of vapor-liquid equilibrium (VLE). *Ind. Eng. Chem. Res.* 2009, 48, 2717–2720. [CrossRef]
- 3. Rogelj, J.; Shindell, D.; Jiang, K.; Fifita, S. Mitigation Pathways Compatible with 1.5 °C in the Context of Sustainable Development. In *Global Warming of 1.5* °C; An IPCC Special Report; IPCC: Geneva, Switzerland, 2018.
- Kachko, A.; van der Ham, L.V.; Bardow, A.; Vlugt, T.J.H.; Goetheer, E.L.V. Comparison of Raman, NIR, and ATR FTIR spectroscopy as analytical tools for in-line monitoring of CO<sub>2</sub> concentration in an amine gas treating process. *Int. J. Greenh. Gas Control* 2016, 47, 17–24. [CrossRef]
- 5. Cole, I.S.; Corrigan, P.; Sim, S.; Birbilis, N. Corrosion of pipelines used for CO<sub>2</sub> transport in CCS: Is it a real problem? *Int. J. Greenh. Gas Control.* **2011**, *5*, 749–756. [CrossRef]
- 6. Wong, M.K.; Bustam, M.A.; Shariff, A.M. Chemical speciation of CO<sub>2</sub> absorption in aqueous monoethanolamine investigated by in situ Raman spectroscopy. *Int. J. Greenh. Gas Control* **2015**, *39*, 139–147. [CrossRef]
- 7. Rochelle, G.T. Amine Scrubbing for CO<sub>2</sub> Capture. Science 2009, 325, 1652–1654. [CrossRef] [PubMed]
- Puxty, G.; Maeder, M. The fundamentals of post-combustion capture. In *Absorption-Based Post-Combustion Capture of Carbon Dioxide*; Woodhead Publishing: Sawston, UK, 2016; pp. 13–33. ISBN 9780081005156.
- Böttinger, W.; Maiwald, M.; Hasse, H. Online NMR spectroscopic study of species distribution in MEA-H<sub>2</sub>O-CO<sub>2</sub> and DEA-H<sub>2</sub>O-CO<sub>2</sub>. *Fluid Phase Equilib.* 2008, 263, 131–143. [CrossRef]
- 10. Kontogeorgis, G.M.; Folas, G.K. Thermodynamic Models for Industrial Applications: From Classical and Advanced Mixing Rules to Association Theories; John Wiley & Sons: Hoboken, NJ, USA, 2009; ISBN 9780470747537.
- 11. Rinker, E.B.; Ashour, S.S.; Sandall, O.C. Kinetics and Modeling of Carbon Dioxide Absorption into Aqueous Solutions of Diethanolamine. *Ind. Eng. Chem. Res.* **1996**, *35*, 1107–1114. [CrossRef]

- 12. Souchon, V.; Aleixo, M.D.O.; Delpoux, O.; Sagnard, C.; Mougin, P.; Wender, A.; Raynal, L. In situ determination of species distribution in alkanolamine-H<sub>2</sub>O-CO<sub>2</sub> systems by Raman spectroscopy. *Energy Procedia* **2011**, *4*, 554–561. [CrossRef]
- Beumers, P.; Brands, T.; Koss, H.J.; Bardow, A. Model-free calibration of Raman measurements of reactive systems: Application to monoethanolamine/water/CO<sub>2</sub>. *Fluid Phase Equilib.* 2016, 424, 52–57. [CrossRef]
- 14. Smith, E.; Dent, G. Modern Raman Spectroscopy—A Practical Approach; John Wiley & Sons: Hoboken, NJ, USA, 2005; ISBN 9780471497943.
- 15. Vogt, M.; Pasel, C.; Bathen, D. Characterisation of CO<sub>2</sub> absorption in various solvents for PCC applications by Raman spectroscopy. *Energy Procedia* **2011**, *4*, 1520–1525. [CrossRef]
- 16. Samarakoon, P.A.G.L.; Andersen, N.H.; Perinu, C.; Jens, K.J. Equilibria of MEA, DEA and AMP with bicarbonate and carbamate: A Raman study. *Energy Procedia* **2013**, *37*, 2002–2010. [CrossRef]
- 17. Swinehart, D.F. The Beer-Lambert Law. J. Chem. Educ. 1962, 39, 333. [CrossRef]
- Matin, N.S.; Remias, J.E.; Neathery, J.K.; Liu, K. Facile method for determination of amine speciation in CO<sub>2</sub> capture solutions. *Ind. Eng. Chem. Res.* 2012, *51*, 6613–6618. [CrossRef]
- 19. Kent, R.L.; Elsenberg, B. Better Data for Amine Treating. Hydrocarb. Process. 1976, 55, 87–90.
- 20. Haji-Sulaiman, M.Z.; Aroua, M.K.; Benamor, A. Analysis of equilibrium data of CO<sub>2</sub> in aqueous solutions of DEA, MDEA and their mixtures using the modified Kent Eisenberg Model. *Trans. Chem. E* **1998**, *76*, 1–8. [CrossRef]
- 21. Suleman, H.; Maulud, A.S.; Man, Z. Carbon Dioxide Solubility in Aqueous Potassium Lysinate Solutions: High Pressure Data and Thermodynamic Modeling. *Procedia Eng.* **2016**, *148*, 1303–1311. [CrossRef]
- 22. Suleman, H.; Maulud, A.S.; Man, Z. Review and selection criteria of classical thermodynamic models for acid gas absorption in aqueous alkanolamines. *Rev. Chem. Eng.* 2015, *31*, 599–639. [CrossRef]
- 23. Sartori, G.; Savage, D.W. Sterically hindered amines for carbon dioxide removal from gases. *Ind. Eng. Chem. Fundam.* **1983**, *22*, 239–249. [CrossRef]
- 24. Haji-Sulaiman, M.Z.; Aroua, M.K.; Pervez, M.I. Equilibrium concentration profiles of species in CO<sub>2</sub>—alkanolamine—water systems. *Gas Sep. Purif.* **1996**, *10*, 13–18. [CrossRef]
- 25. Van Der Maaten, L.; Postma, E.; Van den Herik, J. Dimensionality Reduction: A Comparative Review. J. Mach. Learn. Res. 2009, 10, 66–71.
- 26. Hotelling, H. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 1933. [CrossRef]
- 27. Varmuza, K.; Filzmoser, P. Introduction to Multivariate Statistical Analysis in Chemometrics; CRC Press: Boca Raton, FL, USA, 2009; ISBN 9781420059472.
- Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 2001, 58, 109–130. [CrossRef]
- Marini, F.; Bucci, R.; Magri, A.L.; Magri, A.D. Artificial Neural Networks in Chemometrics: History, Examples and Perspectives. *Microchem. J.* 2008, *88*, 178–185. [CrossRef]
- Mouazen, A.M.; Kuang, B.; De Baerdemaeker, J.; Ramon, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* 2010, 158, 23–31. [CrossRef]
- Kuang, B.; Tekin, Y.; Mouazen, A.M. Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. *Soil Tillage Res.* 2015, 146, 243–252. [CrossRef]
- 32. Ramadan, Z.; Hopke, P.K.; Johnson, M.J.; Scow, K.M. Application of PLS and Back-Propagation Neural Networks for the estimation of soil properties. *Chemom. Intell. Lab. Syst.* 2005, 75, 23–30. [CrossRef]
- Jinadasa, M.H.W.N.; Chandra, K.A.; Halstensen, M. System Development for On-line Monitoring using Raman Spectroscopy for CO<sub>2</sub> Absorption by MEA. In Proceedings of the 59th Conference on Simulation and Modelling (SIMS 59), 26–28 September 2018; Oslo Metropolitan University: Oslo, Norway, 2018.
- 34. Jinadasa, M.H.W.N.; Jens, K.; Øi, L.E.; Halstensen, M. Raman Spectroscopy as an Online Monitoring Tool for CO<sub>2</sub> Capture Process: Demonstration Using a Laboratory Rig. *Energy Procedia* **2017**, *114*, 1179–1194. [CrossRef]
- Shahid, M.Z.; Maulud, A.S.; Bustam, M.A. Non-invasive monitoring of CO<sub>2</sub>concentration in aqueous diethanolamine (DEA), methyldiethanolamine (MDEA) and their blends in high CO2loading region using Raman spectroscopy and partial least square regression (PLSR). *Int. J. Greenh. Gas Control* 2018, *68*, 42–48. [CrossRef]
- Shahid, M.Z.; Suleman, H.; Maulud, A.S.; Khalil, M.A.B.; Man, Z. Monitoring of Chemical Speciation of DEA-CO<sub>2</sub>-Water System by Raman Spectroscopy. *Adv. Mater. Res.* 2015, 1113, 358–363. [CrossRef]
- Edwards, T.J.; Maurer, G.; Newman, J.; Prausnitz, J.M. Vapor-liquid equilibria in multicomponent aqueous solutions of volatile weak electrolytes. *AIChE J.* 1978, 24, 966–976. [CrossRef]
- Aroua, M.K.; Salleh, R.M. Solubility of CO<sub>2</sub> in aqueous piperazine and its modeling using the Kent-Eisenberg approach. *Chem. Eng. Technol.* 2004, 27, 65–70. [CrossRef]
- 39. Vasques, G.M.; Grunwald, S.; Sickman, J.O. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* **2008**, *146*, 14–25. [CrossRef]

- Silva, M.A.M.; Ferreira, M.H.; Braga, J.W.B.; Sena, M.M. Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. *Talanta* 2012, *89*, 342–351. [CrossRef] [PubMed]
- 41. Uysal, R.S.; Boyaci, I.H.; Genis, H.E.; Tamer, U. Determination of butter adulteration with margarine using Raman spectroscopy. *Food Chem.* **2013**, *141*, 4397–4403. [CrossRef] [PubMed]
- 42. de Jong, S. SIMPLS: An alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 251–263. [CrossRef]
- 43. Smits, J.R.M.; Melssen, W.J.; Buydens, L.M.C.; Kateman, G. Using artificial neural networks for solving chemical problems. Part I. Multi-layer feed-forward networks. *Chemom. Intell. Lab. Syst.* **1994**, 22, 165–189. [CrossRef]
- 44. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
- 45. Svozil, D.; Kvasnička, V.; Pospíchal, J. Introduction to multi-layer feed-forward neural networks. *Chemom. Intell. Lab. Syst.* **1997**, 39, 43–62. [CrossRef]
- 46. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
- Janik, L.J.; Cozzolino, D.; Dambergs, R.; Cynkar, W.; Gishen, M. The prediction of total anthocyanin concentration in red-grape homogenates using visible-near-infrared spectroscopy and artificial neural networks. *Anal. Chim. Acta* 2007, 594, 107–118. [CrossRef] [PubMed]
- 48. Despagne, F.; Massart, D.L. Neural networks in multivariate calibration. Analyst 1998, 123, 157R–178R. [CrossRef] [PubMed]
- 49. Williams, K. Near-infrared technology in the agricultural and food industries. In *Near-Infrared Technology in the Agricultural and Food Industries*; American Association of Cereal Chemists, Inc.: Saint Paul, MI, USA, 2001; ISBN 1-891127-24-1.
- 50. Batista de Carvalho, L.A.E. Raman spectra, conformational stability and normal coordinate analysis of ethylmethylamine. *J. Raman Spectrosc.* **1995**, *26*, 653–661. [CrossRef]
- 51. Larkin, P. IR and Raman Spectra-Structure Correlations. In *Infrared and Raman Spectroscopy*; Elsevier: Amsterdam, The Netherlands, 2011; ISBN 9780128041628.
- 52. Salciccioli, J.D.; Crutain, Y.; Komorowski, M.; Marshall, D.C. Sensitivity analysis and model validation. In *Secondary Analysis of Electronic Health Records*; Spring: Berlin/Heidelberg, Germany, 2016; ISBN 9783319437422.
- 53. Alpaydin, E. Introduction to Machine Learning; MIT Press: Cambridge, MA, USA, 2014; ISBN 0262325748.
- 54. Posada, D.; Buckley, T.R. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst. Biol.* **2004**, *53*, 793–808. [CrossRef]
- 55. Bakeev, K.A. Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries: Second Edition; John Wiley & Sons: Hoboken, NJ, USA, 2010; ISBN 9780470722077.
- 56. Davis, A.R.; Oliver, B.G. A vibrational-spectroscopic study of the species present in the CO<sub>2</sub>-H<sub>2</sub>O system. *J. Solut. Chem.* **1972**, *1*, 329–339. [CrossRef]
- 57. Wen, N.; Brooker, M.H. Ammonium carbonate, ammonium bicarbonate, and ammonium carbamate equilibria: A raman study. *J. Phys. Chem.* **1995**, *99*, 359–368. [CrossRef]
- Axrup, L.; Markides, K.; Nilsson, T. Using miniature diode array NIR spectrometers for analysing wood chips and bark samples in motion. J. Chemom. 2000, 14, 561–572. [CrossRef]
- 59. Xu, F.; Wang, D. Rapid determination of sugar content in corn stover hydrolysates using near infrared spectroscopy. *Bioresour. Technol.* **2013**, 147, 293–298. [CrossRef]