

# *In situ* Real-Time Multivariate Analysis of Methanolysis Monitoring of Sunflower Oil Using FTIR

Pascal Mwenge, Tumisang Seodigeng

**Abstract**—The combination of world population and the third industrial revolution led to high demand for fuels. On the other hand, the decrease of global fossil fuels deposits and the environmental air pollution caused by these fuels has compounded the challenges the world faces due to its need for energy. Therefore, new forms of environmentally friendly and renewable fuels such as biodiesel are needed. The primary analytical techniques for methanolysis yield monitoring have been chromatography and spectroscopy, these methods have been proven reliable but are more demanding, costly and do not provide real-time monitoring. In this work, the *in situ* monitoring of biodiesel from sunflower oil using FTIR (Fourier Transform Infrared) has been studied; the study was performed using EasyMax Mettler Toledo reactor equipped with a DiComp (Diamond) probe. The quantitative monitoring of methanolysis was performed by building a quantitative model with multivariate calibration using iC Quant module from iC IR 7.0 software. 15 samples of known concentrations were used for the modelling which were taken in duplicate for model calibration and cross-validation, data were pre-processed using mean centering and variance scale, spectrum math square root and solvent subtraction. These pre-processing methods improved the performance indexes from 7.98 to 0.0096, 11.2 to 3.41, 6.32 to 2.72, 0.9416 to 0.9999, RMSEC, RMSECV, RMSEP and R2Cum, respectively. The  $R^2$  value of 1 (training), 0.9918 (test), 0.9946 (cross-validation) indicated the fitness of the model built. The model was tested against univariate model; small discrepancies were observed at low concentration due to unmodelled intermediates but were quite close at concentrations above 18%. The software eliminated the complexity of the Partial Least Square (PLS) chemometrics. It was concluded that the model obtained could be used to monitor methanol of sunflower oil at industrial and lab scale.

**Keywords**—Biodiesel, calibration, chemometrics, FTIR, methanolysis, multivariate analysis, transesterification

## I. INTRODUCTION

GLOBAL energy supply such as heating, transportation and power plants generally produced from fossil sources. The increase in worldwide population and the industrialisation of the 21<sup>st</sup> century have caused a high demand for fuels. This increase in the consumption of the fossil fuel resulted in the depletion of global fossil fuels deposits and the increase of environmental air pollution caused by these fuels. Therefore

P. Mwenge is with the Vaal University of Technology, Andries Potgieter boulevard, Vanderbijlpark, South Africa, (corresponding author, phone: +27 16 950 7722; e-mail: pascalmwenge@gmail.com).

T. Seodigeng is with the Vaal University of Technology, Andries Potgieter boulevard, Vanderbijlpark, South Africa, (phone: +27 16 950 9655; e-mail: tumisangs@vut.ac.za).

research has to focus on developing new forms of environmentally friendly and renewable forms of fuels [1]-[10]. Global energy consumption has doubled from 1973 to 2013; this is expected to increase by 34% from 2014 to 2035 for which the transportation sector is the main CO<sub>2</sub> emitter accounting for 33.6%. Renewable energy comes from sources such as geothermal energy, solar, biomass and hydropower, for which solar is the most growing renewable energy. However, biomass is classified as the only green carbon [11].

Biodiesel is defined as an alkyl ester; it is produced by reacting triglycerides with short chain alcohols. The short chain alcohols that are mostly used are methanol and ethanol [2], [3], [12]-[16]. Among the advantages of biodiesel compared to its counterpart petroleum diesel are: low environmental impact, reduction of greenhouse gases effects due to less SO<sub>x</sub>, NO<sub>x</sub> and CO<sub>2</sub> gases emissions, biodegradability, high cetane number, high combustion efficiency; these advantages made biodiesel to be classified as a renewable fuel [1], [16]-[19]. According to estimations, each kilogram of biodiesel produced reduces the CO<sub>2</sub> emission by 3.2 kilograms as compared to petroleum diesel [20], [21].

The transesterification reaction is the reaction of triglycerides and short chain alcohol such as methanol or ethanol at a specific temperature in the presence of a catalyst; glycerol is also produced as a by-product. Transesterification is a reversible reaction; hence the alcohol has to be in excess to favour the forward reaction [18], [22]-[27]. The reaction conditions, feedstock compositional limits and post-separation requirements are determined by the nature of the catalyst [18], [22]-[27].

Generally, the catalysts used in biodiesel production are characterized in three categories namely: alkalis (heterogeneous and homogeneous), acids (heterogeneous and homogeneous) and enzymes. In comparison with enzyme catalysts, homogenous alkali and acid catalysts are the most used in biodiesel production [28]-[34]. The mostly used homogenous base catalysts are: sodium hydroxide, potassium hydroxide; the mostly used homogenous acid catalyst are: Sulphuric acid, hydrochloric acid and phosphoric acid. Alkali catalysts are preferred due to their advantages such as low cost, the fast reactions and less energy requirement (60 °C is enough for high yield, while acid catalysts require more than 100 °C) [16], [27]. Heterogeneous alkali catalysts used in methanolysis are: magnesium oxide (MgO), calcium oxide (CaO) and strontium oxide (SrO). Liquid and immobilized enzymes have attracted researchers' attention as they prevent

soap formation and the product obtained is of good quality thus, reducing the purification processes [16], [28], [29], [31].

To ensure biodiesel quality, fuel properties and reaction monitoring transesterification, analytical methods have been used and reported by many scholars. The primary analytical method used to quantitatively and qualitatively monitor biodiesel yield has been Gas Chromatography (GC), this has the advantage to determine intermediate and other components. These components are; monoglycerides, diglycerides, glycerol moiety and unreacted triglyceride [35]-[38]. Flame ionisation detection (FID) has been a widely used detector; the introduction of mass spectrometer has increased the efficiency of the method [35]-[37]. Numerous other methods have been employed such: Capillary Gas Chromatography (CGC) [15], [38], High Pressure Liquid Chromatography (HPLC) [15], [35], [37], [38], Gel Permeation Chromatography (GPC) [37]-[39], FTIR Spectroscopy [14], [15], [37], [38], [40]-[45], Ramon Spectroscopy [14], [38], Size Exclusion Chromatography [15], [37], [38], Nuclear Magnetic Resonance (NMR) [15], [37], [38], [46], Ultra-sonication assisted spray ionization (UASI) spectroscopy, fluorescence spectroscopy, [15], [37], [38], [41], [48], UV spectroscopy [14], Thin Layer Chromatography (TLC), thermogravimetric analysis (TGA) [37], Viscometry, [37], [38].

Chromatography and Spectroscopy being the most used methods of qualitative and quantitative methods of biodiesel are reputed for their accurateness. Nevertheless, they have disadvantages such as time-consuming (require extensive sample preparation), requiring highly qualified personnel. Their usage for *in-situ* transesterifications (fast reversible reactions) has not been feasible. FTIR spectroscopy has been reported as an accurate, fast and reliable method for quantitative and qualitative analysis of *in situ* transesterification monitoring [15], [37], [38], [41], [48]. Many authors have reported the use Infrared Spectroscopy such as NIR (Near Infrared) spectroscopy, FTIR for real-time transesterification [14], [38], [46], [49], [50].

FTIR analysis uses infrared spectroscopy technique, in which the electromagnetic energy of radiation spectrum is used; the energy used in spectroscopy has wavelengths between the visible light and radio waves [36]. The infrared region used in infrared spectroscopy is divided into three regions which are: Near-infrared (NIR) wavenumber ranging 12500-400  $\text{cm}^{-1}$ , mid-infrared (MIR) wavenumber ranging 4000-400  $\text{cm}^{-1}$  and far infrared (FIR) wavenumber ranging 400 to 33  $\text{cm}^{-1}$  [36]. MIR has been the most used region and used in this study. Infrared spectroscopy has been in use for analysis for the past 70 years; infrared spectrums are classified as fingerprints of materials or samples, these fingerprints are constituted of absorption peaks that correspond to the frequencies of vibrations of atomic bonds constituting the material [51], [52].

Attenuated total reflection Fourier Transform Infrared (ATR-FTIR) coupled with PLS regression has been reported [41] to be used in the multivariate calibration models for determining the content of biodiesel in its blend with

petroleum diesel. The designed model was shown to be reliable for determining biodiesel-diesel blends. This same approach was used to determine the vegetable oil used in the transesterification by using FTIR in combination with soft independent modelling of class analogy (SIMCA), hierarchical cluster analysis (HCA), interval principal component analysis (iPCA) and principal component analysis (PCA). PLS combined with interval partial least squares (iPLS) and synergy partial least squares (siPLS) regressions was further used to determine the quality variables (amount, flash point, sulfur content and specific gravity) of diesel/biodiesel blends. These were further used to identify the most suitable region for each studied property [41] Quantification of biodiesel yield using ATR-FTIR coupled with PLS regression was also used in microwave-assisted methanolysis of soybean oil by [41]. Reference [15] worked on the transesterification of degummed oil and ethanol, applying FTIR and multivariate analyses. They concluded that using FTIR to monitor transesterification was a fast and reliable analytical method. Those above were confirmed by [14] in their work on transesterification of high oleic sunflower oil with ethanol.

Transesterification monitoring of vegetable oils using FTIR being a fast and accurate analytical method needs the application of multivariate analysis to be successfully reliable and industrial quality control [15], [37], [38], [42]. FTIR, coupled with multivariate analysis for complex reaction, attracted researchers since the late 90s [15]. PLS is a mathematical method based on Beer's law principle for quantification [38].

FTIR uses fibre-optic which was developed to work in the mid and NIR regions [15]. Reference [53] used fibre optic and PLS to quantify methyl ester from transesterification of soybean oil and they made use of the difference in NIR spectra between methyl esters at 6005  $\text{cm}^{-1}$  and 4428  $\text{cm}^{-1}$ . A cut band on a shoulder band of the oil (triglycerides) was observed. Loss of  $\text{OCH}_2$  groups in glycerol moiety was observed, which resulted in a decrease in the 1378  $\text{cm}^{-1}$  peak, as per FTIR spectrum [39]. On the other hand, [54] made use of the sample principle at a region of 1300-1060  $\text{cm}^{-1}$ .

Most of the multivariate calibration methods that have been developed use mainly PLS, Principal Component Regression (PCR) and PCA techniques. These methods have been reported to be tiresome as they necessitate simultaneous decomposition of absorption matrices and concentration in computing the spectral loading. The spectral loadings are afterwards used to determine the most impacting spectra to the calibration model. The multivariate calibration methods can be improved by increasing the number of calibration and validation samples [37]. The regression calibration models can be improved by first using pre-processing methods such as mean centring technique, [42], [37] and baseline correction has been often used in as pre-processing method [42]. Mean centring has been the mostly used pre-processing technique. Executing mean centring over numerous factors brings about the removal of the mean sample vector from all sample vectors in the data collection. Afterwards, the relative contrasts of the spectrum intensity at different wavelengths are

more straightforward to recognise [42]. After a calibration model has been pre-processed, performance index algorithms are used to access the performance and accuracy of the model. These algorithms are Root Mean Square Error (RMSE), Root Mean Square Error of Calibration (RMSEC), Root Mean Square Error of Prediction (RMSEP) and Root Mean Square Error of Cross Validation (RMSECV) [37]. Though the use of PLS and PCA generate a reliable and robust model, they rely on Beer's law, which has high complex, mathematics [37]. Therefore the use of software becomes essential, as it was the case in this study, in which iC Quant Module Multivariate modelling was employed.

The most used analytical methods of biodiesel monitoring, namely Chromatography and Spectroscopy, offer information on the composition of the mixture. Though spectroscopy can be used for *in-situ* monitoring, they all offer some limitation for the fact that they can only be performed when the reaction had occurred or by an off-taking sample from the reaction mixture. FTIR spectroscopy offered the possibility of *in-situ* real-time monitoring of the process and had been found to be non-destructive, user-friendly and a rapid analytical technique. Therefore, this study aimed at developing *in-situ* real-time monitoring of the transesterification reaction using EasyMax 102 and the iC Quant module from iC IR 7.0 software, by simplifying the mathematical complexity and thereby reducing time of multivariate analysis. Sample validation and performance index were performed to check the accuracy of the model built.

## II. MATERIALS AND METHODS

### A. Materials and reagents

Refined sunflower oil (triglyceride) was sourced from Golden Fry a local cooking oil supplier. Sodium hydroxide (99.5%) used as a catalyst for the transesterification reaction and methanol (99%) used as an acyl acceptor for the transesterification reaction, hexane (99.5%) used as GC solvent and phosphoric acid (99.5%) used for washing biodiesel, were sourced from ACE (Associate Chemical Enterprises) a local laboratory chemicals supplier.

### B. Standard Preparation

The biodiesel standard used in this work was obtained by transesterification reaction of sunflower with methanol in a laboratory scale reactor composed of flat bottom flask mount with a reflux condenser and hot plate with magnetic stirrer equipped with a temperature controller. Sunflower oil was heated for 60 minutes at 105 °C for moisture removal. The transesterification reaction occurred under the following conditions: Atmospheric pressure, 60 °C, 12 methanol to oil mole ratio, 60 minutes reaction time, 1wt% (compared to oil) catalyst ratio and 600RPM stirring speed. The product obtained was separated by density difference using separating funnel, the bottom layer containing glycerol was then removed, and the remaining contained an excess of methanol. The excess methanol was removed by distillation. The biodiesel was vigorously washing with acidic water containing

3% phosphoric acid until the water was clear, the purpose of the acidic water was to remove any soap and catalyst from the product [54]. The washed biodiesel was heated at 105 °C for moisture removal, cooled and stored. A sample was collected for characterisation with GC-FID. The biodiesel standard produced was used to prepare 15 standard samples, for which one was methanol, and 14 were a mixture of biodiesel and sunflower oil at different concentrations for FTIR/PLS calibration using ReactIR 15 and data processing using iC IR software 7.0 and iC Quant Module from Mettler Toledo EasyMax 102.

### C. In-Situ Real-Time Transesterification

The *in situ* transesterification methanolysis or transesterification of sunflower oil was performed in a Mettler Toledo EasyMax 102 reactor. Mettler Toledo EasyMax 102, as shown in Fig. 2 is equipped with two reactors of 100 ml each. The reactor was mounted with a reflux condenser and a mechanical stirrer. EasyMax reactors have the advantage to simultaneously and independently control two reaction mixtures. The dosing mechanism has a purpose to allow more accurate reactant addition to the system. EasyMax 102 has the advantage to control the parameters such as temperature and pH in the specified range. The temperature is controlled by adjusting the heat of reaction and mixture by the heating and cooling system installed. All results timeously recorded for analysis at a further stage. EasyMax 102 is equipped with incorporated real-time *in situ* instruments; this enables a depth process understanding and makes the determination of the kinetics easier. This instrument further allows the correction of parameters as they are quickly analytically detected [55].

A batch set up was used. Approximately 50 g of sunflower oil was weighed and transferred in the reactor; approximately 17 g (approximating 9 moles) of methanol was weighed and mixed with 0.5 g (1wt % compared to oil) of sodium hydroxide and pumped in the reactor. A new batch transesterification reaction experiment was set using iControl software 5.0, with 55 °C, 200 RPM, 60 minutes; reaction temperature, stirring speed and reaction time, respectively. The reaction timing and stirring were set to start once the reaction mixture reached the set temperature, then using FTIR probe of the ReactIR spectrums were collected every 15 seconds by inserting the probe in the reaction mixture, this further produced a surface plot on the iCIR software for online quantification with the calibrated model. The overhead impeller was used as a stirring element.



Fig. 1 Mettler Toledo EasyMax 102 equipped with a dosing system

#### D. GC Analysis of the Standard

The GC analysis of the standard was performed using flame ionisation (FID) on Trace 1310 Thermo Scientific (USA) Gas Chromatograph (GC) instrument, equipped with an automatic sampler and injector AI 1310. The separation was carried out in a BPX70-FAME optimised capillary column (Phase: BPX70, 0.25  $\mu\text{m}$  film, Column: 25 m x 0.32 mm ID) from SGE (USA).

A constant flow of 2.2 ml/l of the carrier (helium) was used, and the column head pressure was adjusted to 11.8 psi. The injection temperature was 250  $^{\circ}\text{C}$ ; the injection mode was split with a ratio of 8:1, the detector temperature was 300  $^{\circ}\text{C}$ , and the injection volume was 1  $\mu\text{L}$ . The initial temperature was 80  $^{\circ}\text{C}$  held for 2 minutes, firstly ramped to 130  $^{\circ}\text{C}$  at a rate of 50  $^{\circ}\text{C}/\text{minute}$  held for 10 min, secondly ramped to 172 at a rate of 2  $^{\circ}\text{C}/\text{minute}$  and held for 6 minutes. The total run time was 33 minutes. Hexane was used as a solvent to prepare standard solutions. The methyl ester percentage was calculated by averaging the sum of surface areas of every methyl ester in the

sample from the data of GC, this was done in triplicate and was found to be 99.57% and was later used as standards.

#### E. FTIR Spectra Collection

FTIR spectra were collected on the 15 samples, for which one was pure methanol which was used for solvent subtraction on the multivariate data analysis. The probe was introduced in all the 14 biodiesel/sunflower oil mixture samples to collect samples in duplicate for multivariate calibration; these are shown in Fig. 2.

For the real-time analysis, the probe was inserted in the reaction mixture; spectra were collected every 15 seconds which was transferred to the iC IR software for interpretation. The spectrums recorded by averaging 128 scans in the wave number ranging from 2500 to 650  $\text{cm}^{-1}$  using ReactIR15 spectrometer from Mettler Toledo. The ReactIR15 spectrometer instrument operates with an MCT detector, 8  $\text{cm}^{-1}$  resolution, Happ-Genzel apodization DiComp (Diamond) probe and AgX 6mm x 2m Fiber (Silver Halide) interface.

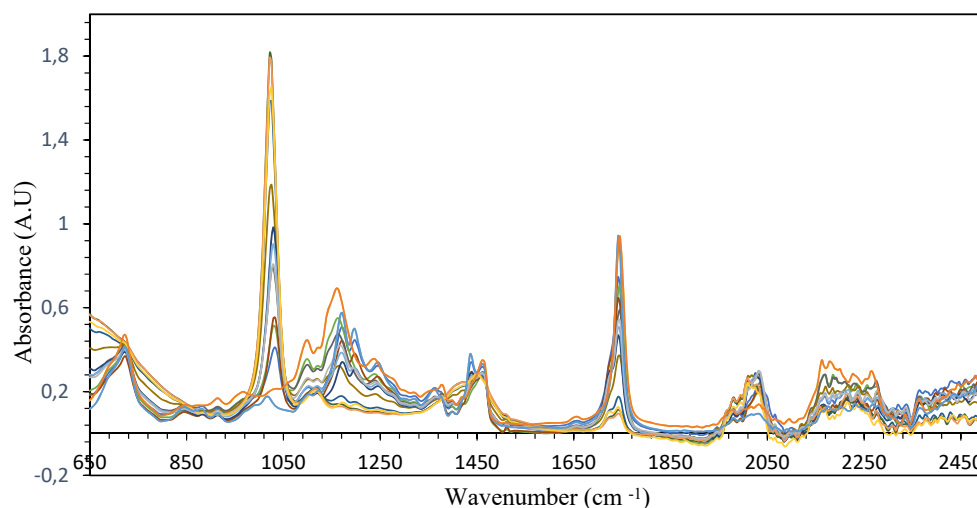


Fig. 2 Spectra for multivariate analysis

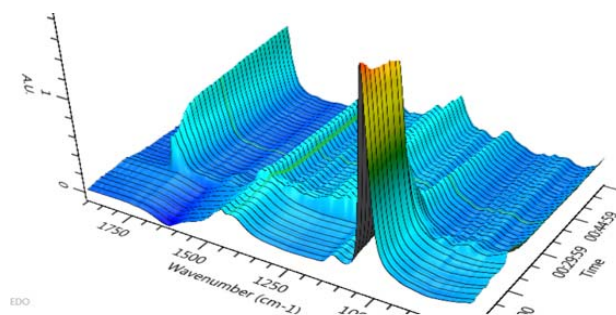


Fig. 3 Surface plot of the online monitoring of transesterification

#### F. Multivariate Analysis

The multivariate analyses were performed on 14 samples of standards mixture containing different concentrations; the mixture concentration was in a way that the entire range of

conversion was covered. The multivariate analyses of the FTIR spectra of the 14 samples were performed using PLS regression from iC Quant module of the iC IR 7.0 software of Mettler Toledo. The region from 701 to 1822  $\text{cm}^{-1}$  was selected for data treatment. Uncontrolled sources of variation can lead to inter-batch differences. Therefore all spectral data were pre-processed before performing multivariate analyses. Pre-processing methods such as mean centring, variance scaling, and spectrum math square root were used. Cross-validation by a leave-one-out procedure was performed during the validation step to define the optimal number of factors that should be kept in the model to detect outliers. The spectra were collected in duplicated and divided into two sets for calibration and validation of the PLS regression model.

Performance index was used to evaluate the performance of the method. It indicates how well the calibrated method can classify the validation standards. As the performance index

algorithm is set to “RMSE”, the root mean square error is reported. The closer the RMSE value is to zero, the smaller are the differences between the calculated concentrations values and the actual values [14]. RMSEC, RMSECV and RMSEP have been calculated. Detection and elimination of outliers were carried out using score and leverage plots. The predictive ability of the model was evaluated according to the RMSEP obtained for the external validation set. An F-test at a confidence level of 95% was used to assess the statistical significance of differences between RMSE values. The F-values were calculated as the ratio of the squares of the largest and smallest RMSE values. Detection and elimination of outliers were performed using score, residual (X and Y) and leverage plots.

TABLE I  
BIODIESEL AND SUNFLOWER OIL MIXTURE FOR MULTIVARIATE ANALYSES

Samples	Sunflower Biodiesel (%)	Sunflower Oil (%)
1	100	0
2	89.5	10.5
3	86.7	13.3
4	83.3	16.7
5	69.2	30.8
6	64.7	35.3
7	60	40
8	43.8	56.2
9	36.4	73.6
10	33.3	66.7
11	28.6	71.4
12	11.8	88.2
13	6.3	93.7
14	0	100

### III. RESULTS AND DISCUSSION

#### A. Multivariate Calibration

It was observed from the results obtained that they are similarities in the spectra of sunflower oil and the biodiesel, this is due to the fact of the high chemical similarity between triglycerides and methyl esters, as it can be seen in Fig. 4. Nevertheless, some differences can be observed at the region of wave number from 900 to 1500  $\text{cm}^{-1}$ ; this region is called the fingerprint region of biodiesel [15], [41]. As shown in Figs. 5 and 6, another spectral region around 1740  $\text{cm}^{-1}$  showed a difference; this includes the stretching vibration of carbonyl groups and is due to the extent of glycerol substitution in the fatty acids by methoxy radicals.

There was an O-CH<sub>2</sub>-C axial asymmetric stretching observed at the wavenumber of 1110  $\text{cm}^{-1}$  in the oil spectra which was not observed in the methyl esters. It was also observed an O-CH<sub>3</sub> stretching at 1195  $\text{cm}^{-1}$  in the biodiesel spectra which was absent in the one of oil. These were other factors considered for discrimination of oil and biodiesel; these were found to be similar to what [37] and [41] reported. A region of wavenumber ranging from 1370 to 1400  $\text{cm}^{-1}$  showed O-CH<sub>2</sub> groups glycerol moiety of triglycerides, diglycerides and monoglycerides present in the oil but absent in the biodiesel [41], [37] but diglycerides and monoglycerides

were very low to be considered for the multivariate modelling. Another region (see Fig. 5) that clearly showed distinction and was also used as a crucial factor for differentiation was from 1425  $\text{cm}^{-1}$  to 1450  $\text{cm}^{-1}$ , which represent an asymmetric bending of CH<sub>3</sub> [15], [37], [41].

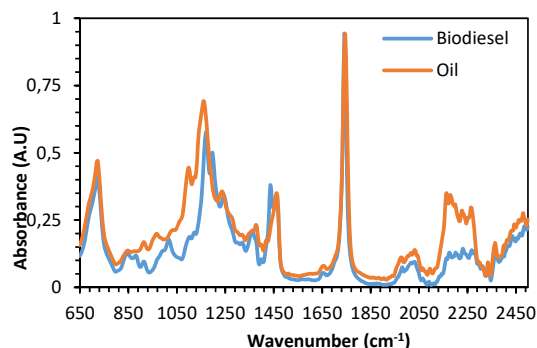


Fig. 4 FTIR spectra sunflower oil and biodiesel from 650 to 2500  $\text{cm}^{-1}$

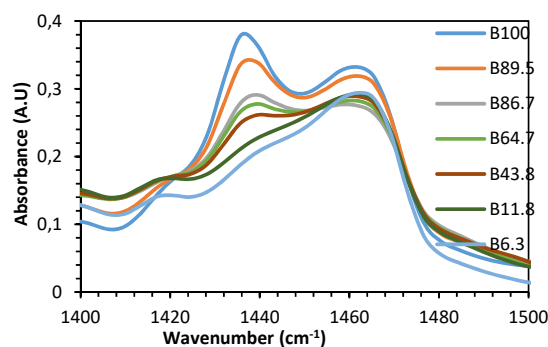


Fig. 5 FTIR spectra sunflower oil and biodiesel from 1400 to 1500  $\text{cm}^{-1}$

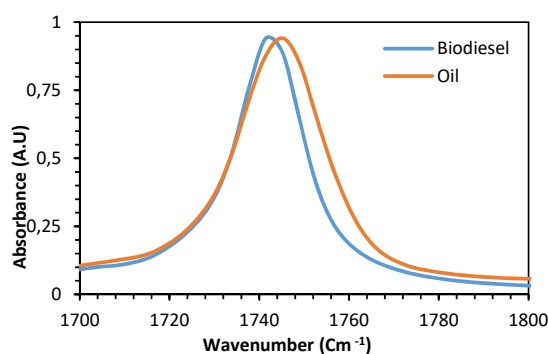


Fig. 6 FTIR spectra sunflower oil and biodiesel from 1700 to 1800  $\text{cm}^{-1}$

Figs. 6 and 7 showed that the stretching vibration of carbonyl groups was one of the critical factors for distinction among the set standard samples which contained different percentages of sunflower and biodiesel. The obtained results confirmed the observation that the dislocation of the carbonyl signal, between the variables, was responsible for the spectral



differentiation that occurred as a result of a gradual increase of sunflower concentration. That is, for the seven standard mixtures described in Table I, there was a gradual migration of the C=O stretching vibration from  $1746\text{ cm}^{-1}$  in triglycerides to  $1738\text{ cm}^{-1}$ . Authors [15] reported a similar observation, in the work they reported ethanolsis of degummed soybean oil in which the C=O stretching vibrations were observed at  $1746.2\text{ cm}^{-1}$  and  $1735.2\text{ cm}^{-1}$  for degummed soybean oil and ethyl esters, respectively.

Methanolysis of sunflower influences changes in the FTIR spectra of the reaction mixture, therefore to allow the in situ real-time monitoring of reaction yield, a multivariate model was developed using PLS from iC Quant module from iC IR 7.0 software.

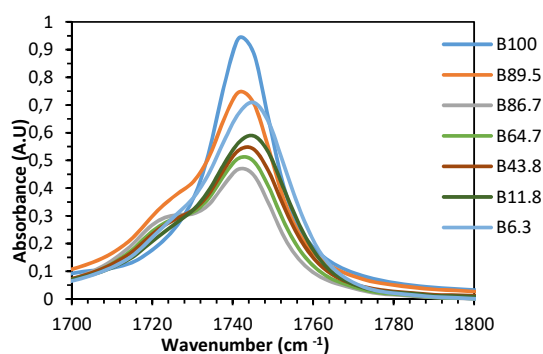


Fig. 7 FTIR spectra sunflower oil and biodiesel mixtures from  $1700$  to  $1800\text{ cm}^{-1}$

The FTIR/PLS calibration model was developed from the 14 samples (see Table I), the region ranging from  $701$  to  $1822$

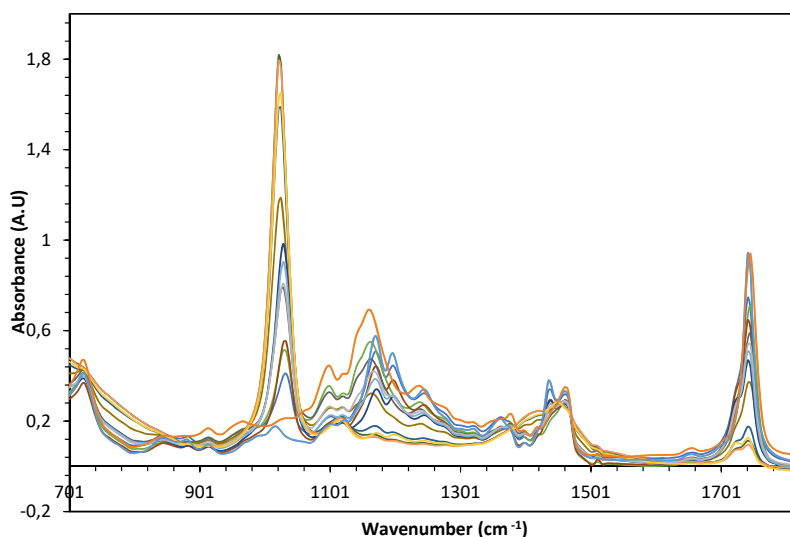


Fig. 8 Spectra range use for PLS/FTIR multivariate analysis

The actual Predicted (Cross Validation) graph (see Fig. 10) represents cross-validation of the actual versus the predicted values in the model. In this process, the standards samples

$\text{cm}^{-1}$  (see Fig. 8) was selected; this was motivated by the fact the most discrimination of the spectrum was found in that range, and the cumulative root mean square increased from  $0.82$  (on full spectrum) to  $0.89$  (on selected range).

The same 14 samples were used for calibration (training data) and validation (test data); therefore after calibration spectrums were taken, the validation spectrums were taken as well. Pre-processing methods (Variance Scale, Mean Centring, spectrum math square root and solvent subtraction) were used. These resulted in the improvement of the performance index; this shows the predictive capability of the model (see Table II). The improvement of the RMSEC, RMSECV, RMSEP and  $R^2\text{Cum}$  showed the effect of perfecting the model by applying the pre-processing method. Authors [42], [37] reported the use of the performance index, as crucial tools to improve the performance of the model.

The predicted versus actual value of the model (see Fig. 9) illustrated how close the predicted results are to the actual values. The system calculated the best fit for the training set samples to the actual values, and the result of the training data is displayed in the chart. Since the model was adequately calibrated, almost all the training set samples fell on the  $45^\circ$  line. The chart also displays the number of factors used in calibrating the training set samples.

Cross-validation by a leave-one-out procedure was performed during the validation step to define the optimal number of factors that should be kept in the model to detect outliers; this gave 11 optimum factors that were used in the model. By using score and leverage plot, outliers were detected and eliminated. Detection and elimination of outliers were carried out using score and leverage plots.

were taken as 28 samples each sample had two spectrums, one training set sample (standard) was removed from the calibration step at a time. For each standard removed, a new

model was calculated, and this model was used to estimate the concentration of the omitted standard. For each standard, the maximum available data are included in the calibration model, but the model is not influenced by the inclusion of the standard to be estimated. This method provides a reliable estimate of the expected errors in using the calibration to estimate concentration in the unknowns.

TABLE II  
PERFORMANCE INDEX BEFORE AND AFTER APPLYING PRE-PROCESSING  
METHODS FOR MULTIVARIATE CALIBRATION

Index	before	after
RMSEC	7.98	0.0096
RMSECV	11.2	3.41
RMSEP	6.32	2.72
R <sup>2</sup> Cum	0.9416	0.9999

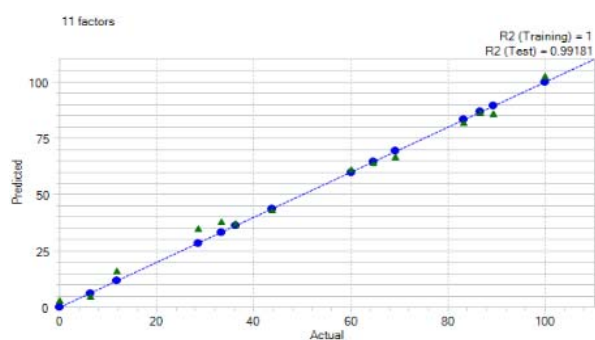


Fig. 9 The predictive capability of the multivariate model for oil and biodiesel blend: Actual versus Predicted value of the model (blue training data and green test data)

The  $R^2$  describes the fraction of the explained variations by each factor for each response variable.  $R^2$  cumulative is cumulative explained variance by each factor for each response variable. The  $R^2$  value of 1 (training), 0.9918 (test), 0.9946 (cross-validation) indicated the degree to which the model fits the data. As seen in Fig. 11, more optimum factors are used in the PLS model, the cumulative  $R^2$  approached 1; this indicated the number of optimum factors influences the cumulative  $R^2$  positively. Hence the model became perfect.

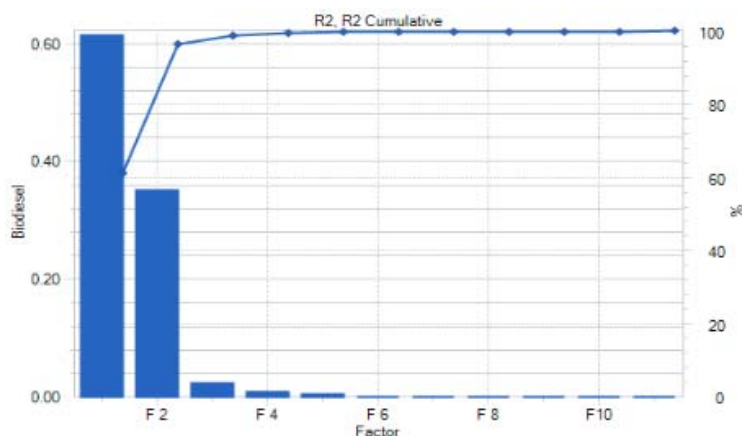


Fig. 11  $R^2$ ,  $R^2$  Cumulative graph of the model

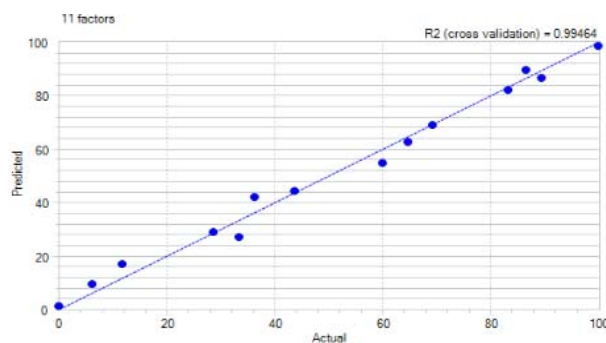


Fig. 10 Cross-Validation Predictive capability of the multivariate model for oil and biodiesel blend: Actual versus Predicted value of the model

The F Test Statistic graph (see Fig. 12) plots the F-Ratio calculated for residual variance of each data sample in the predictor variable space. The F-Ratio is the ratio of residual variance for a sample versus the total variance of the training sample set. This ratio is an indication of how one sample is statistically similar to the training dataset of the model. The 95% confidence limit is displayed on the chart as a reference line. The sample's F-Ratio being far below the 95% confidence limit showed that there is 95% possibility that the sample should not be considered an outlier.

The scores are the projection of each spectrum onto the various PLS factors. The projection tells how much of a specific factor is used to model the given spectrum. The PLS score was used in conjunction with F-Test to detect outliers to classify samples and to obtain qualitative information about the calibration set. The circle describes the boundary using the two factors, this was done with 95% confidence limit and helps identify if there any potential outliers in the data set. Fig. 13 shows PLS scores graphs of factor 1 vs factor 2 (a), factor 2 vs factor 2 (b), factor 6 vs factor 2 (c), and factor 11 vs factor 2 (d). From Fig. 13 it was observed that the training data are very close to the test data and no point was outside the ellipsoid, therefore, no potential outliers.

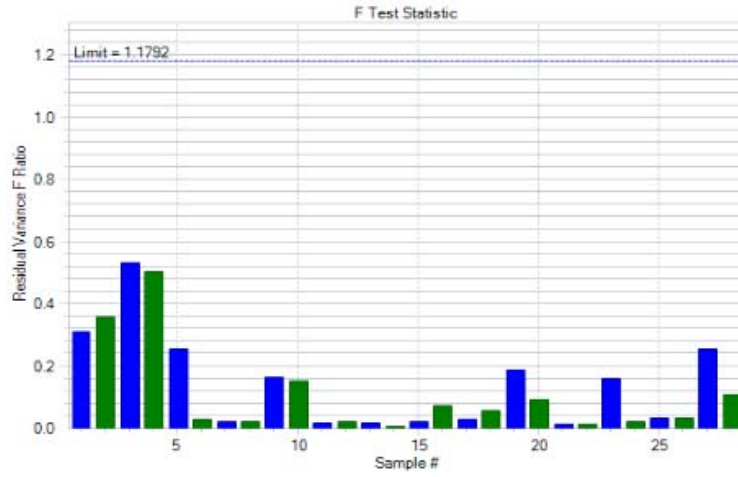
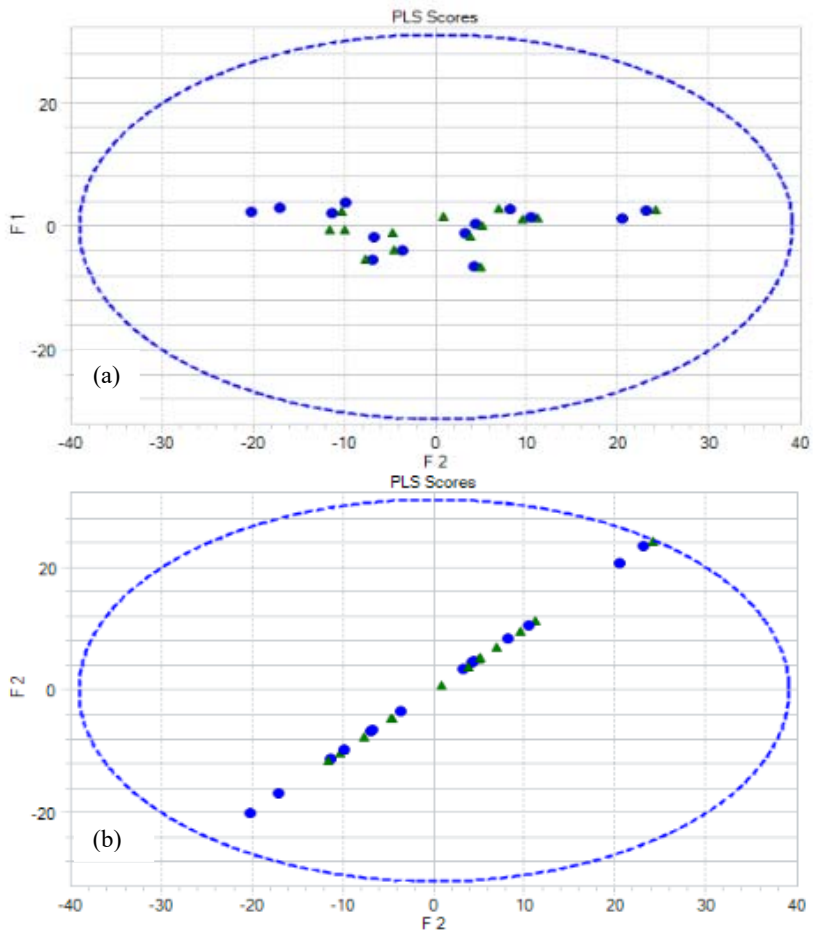


Fig. 12 F Test Statistic (blue training and green test)





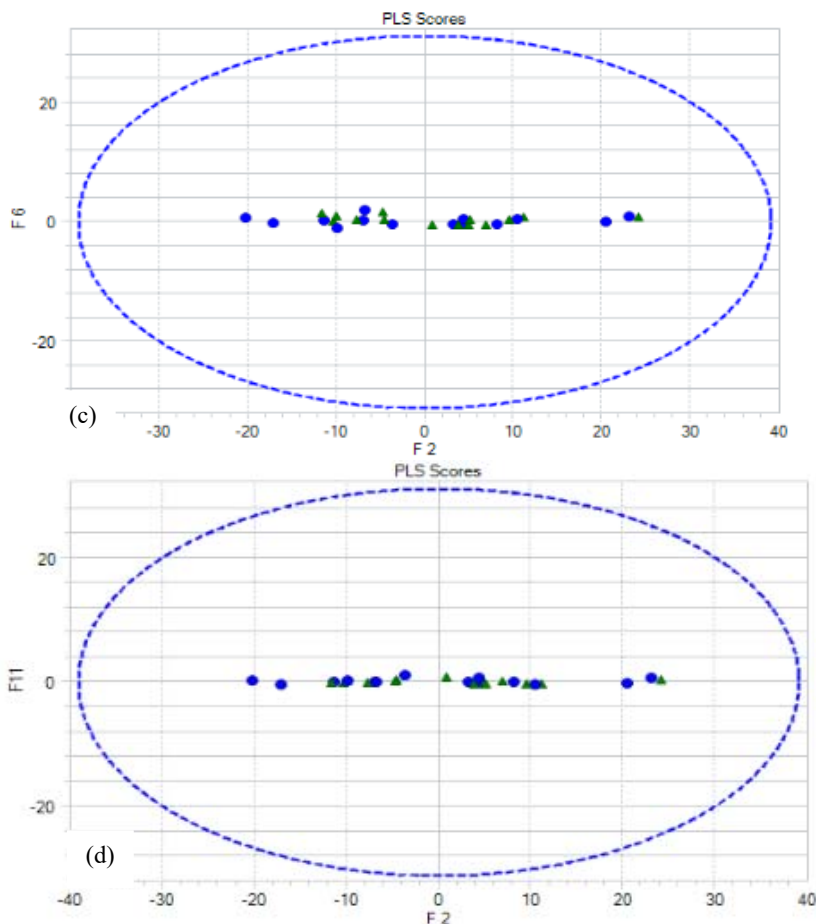


Fig. 13 PLS scores graphs: factor 1 vs factor 2 (a), factor 2 vs factor 2 (b), factor 6 vs factor 2 (c), and factor 11 vs factor 2 (d)

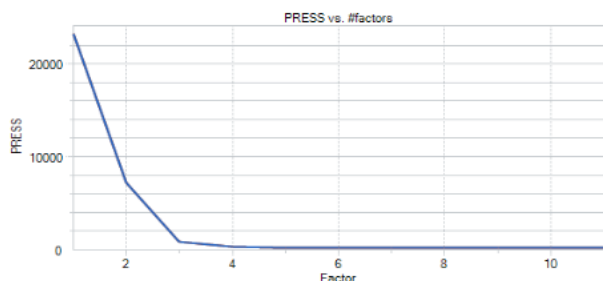


Fig. 14 PRESS vs number of Factors

Fig. 14 shows how the PRESS is affected by the increase in the number of factors; PRESS is the PRediction Error Sum of Squares versus the number of factors. PRESS is defined as the sum of the squared differences between actual and predicted Y values for the training data points that are left out in the Leave-N-out cross-validation process [56]. PRESS values are proper measures of the predictive power of the model. The system picks the number of factors that give the minimum PRESS value as the optimum number of factors. PRESS and RMSECV are closely related. RMSECV is the Root Mean Square of the PRESS.

Fig. 15 plots the value of RMSEC versus the number of

factors used in the model. RMSEC values are calculated for all the training data points used in the model calibration and therefore are direct estimates of the modelling error. Each Y response variable has its own RMSEC series. Typically, the RMSEC value decreases when more factors are used in the model. PRESS and RMSEC are closely related.

Fig. 16 plots the value of RMSECV versus the number of factors used in the model. RMSECV are calculated for the data points that are left out in the Leave-N-Out cross-validation during the model calibration. Each Y response variable has its own RMSECV series. Typically, RMSECV value decreases when more factors are used in the model.

Fig. 17 plots the value of RMSEP versus the number of factors used in the model. RMSEP is calculated for all the test data points using the calibrated model. RMSEP expresses the average error to be expected with the future predictions. Typically, RMSEP values decrease when more factors are used in the model. RMSEP is a useful measure when the user compares different models, regardless of how the models were calibrated.

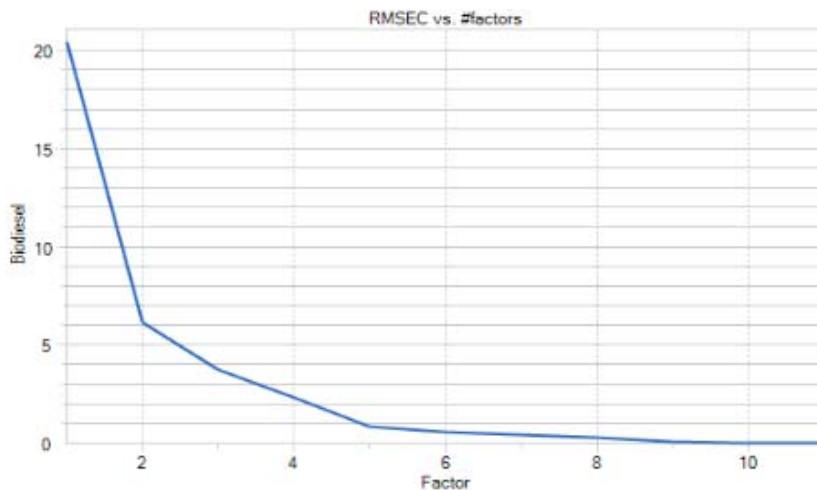


Fig. 15 RMSEC vs number of Factors

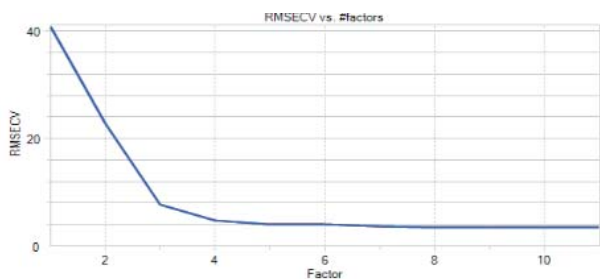


Fig. 16 RMSECV vs number of Factors

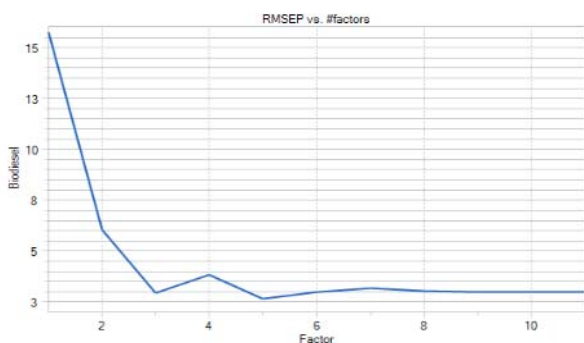


Fig. 17 RMSEP vs number of Factors

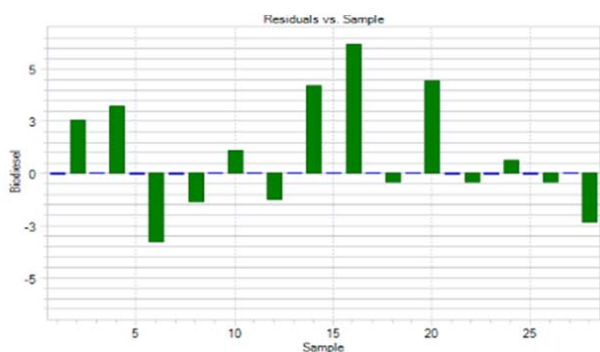


Fig. 18 Residual vs sample

Fig. 18 plots the modelling residuals for the selected response variable for all samples. Residuals are the differences between the actual values and the predicted values. A large residual for a sample indicates that this sample has not been well modelled (it may even be an outlier). The residuals should evenly be distributed meaning that the remaining unexplained variations in the data should be similar to white noise. A systematic pattern in the residuals indicates that there might be some systematic variation in the data remains.

Fig. 19 plots the studentised residual versus leverage and can be used to detect possible outliers. An outlier is a standard which, due to either concentration error, is not similar to the rest of the calibration set and may be detrimental to the calculation. Ideally, no single standard should exert undue influence (leverage) on the calibration. Residuals are also ideally small. In practice, a high value for one or the other, particularly leverage, does not automatically indicate an outlier. It may merely indicate an extreme standard in the calibration space, either in concentration or its spectrum. Those points with a significant residual and leverage may be outliers and therefore, should be removed from the calibration set before recalibration.

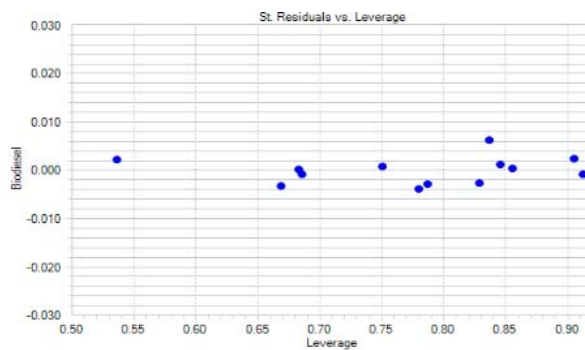


Fig. 19 St Residuals vs leverage

*B. In Situ Real-Time Monitoring of Transesterification Yield*  
The model obtained was then used to predict the biodiesel

yield in real time from an experiment run on an EasyMax 102 Mettler Toledo reactor.

Using iC IR software 7.0, an *in-situ* real-time monitoring was performed. Fig. 20 showed the results obtained; it can be seen that the model can be used to predict the biodiesel yield in the transesterification of sunflower oil. It was also observed that peak present in biodiesel ( $1437\text{ cm}^{-1}$  and  $1744\text{ cm}^{-1}$ ) was increasing with time and could be used to monitor the online transesterification of sunflower oil. As shown in Fig. 20, it was clearly seen that there is a proportional increase of the

peaks identified as peaks representing biodiesel, hence the model obtained is in line with the with what reported by [49], stating that single peaks can be used to monitor the biodiesel transesterification reaction. Considering a closed system used in this work, where no loss of methanol was assumed, methanol (peak at  $1023\text{ cm}^{-1}$ ) consumption could as well be used as well to predict the biodiesel yield. The reaction monitoring of the biodiesel synthesis using methanol peak agrees with the work reported by [49].

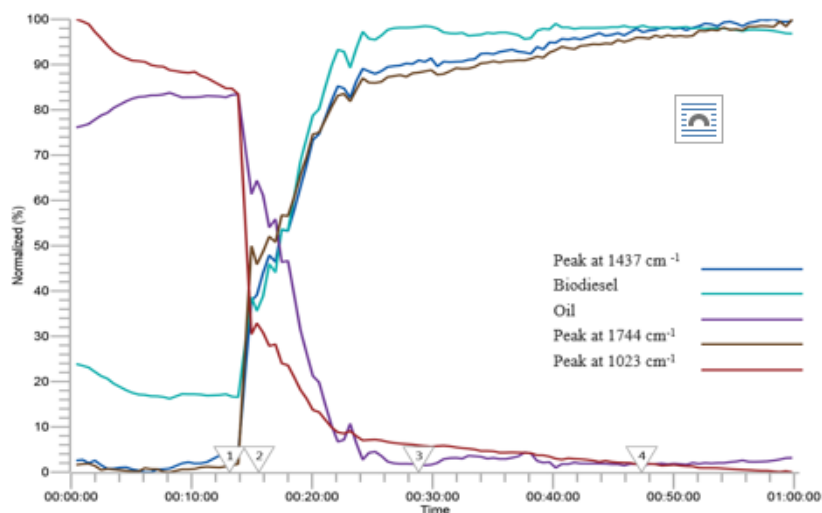


Fig. 20 In Situ Real-time monitoring of transesterification yield

Fig. 20 indicates that the model obtained from iC IR software was quite reliable for predicting reaction yields at advanced reaction times in which conversions higher than 18% were obtained. Small discrepancies were observed along the linear regression, particularly at low conversion yields. This apparent uncertainty of the model was attributed to the interference caused by unmodelled reaction intermediates (unreacted mono and diglycerides), which are particularly crucial at low reacting times. The matrix was significantly simplified towards the end of the reaction. Therefore, deviations at short reaction times were expected due to the occurrence of a higher amount of unmodelled reaction intermediates.

#### IV. CONCLUSION

FTIR spectroscopy has been proven to potentially offer a direct, non-destructive, rapid and real-time transesterification monitoring. The effectiveness of multivariate variation technique using FTIR technique lies in carefully performing calibration and model validation to be done with quite a high number of samples. Once this is done, there is no need for further validation as this can be used for years as long as the materials remain the same. This will give much more easiness to the operators, quality controllers and quality assurance operating the biodiesel plant to use the FTIR to determine the yield of biodiesel and to monitor the transesterification. The

FTIR/PLS model was built to predict more advanced reaction yields of methanolysis sunflower oil, which is what matters to indicate the time in which the reaction should be stopped. The obtained model can be considered suitable for the quantification of fatty acid methyl ester in biodiesel produced in EasyMax 102 Mettler Toledo reactor and iC Quant module from iC IR 7.0 software. A more robust model can be built if non-converted mono and diglycerides are to be considered in the samples used for both calibration and validation. EasyMax Mettler Toledo equipped with ReactIR FTIR probe does not require sampling for offline FTIR analysis, neither sample preparation for the analysis. As opposed to the GC reference method adopted by most of the official methods from regulating and monitoring agencies in the world, it can be concluded that it is a promising candidate for becoming a reference technique to be upgraded for industrial scale.

#### ACKNOWLEDGEMENT

The authors would like to acknowledge the Department of Chemical Engineering at Vaal University of Technology and the Biochemistry Laboratory of Resolution Circle for technical support and assistance. The authors would also like to thank the Research and Higher Degree Unit at the Vaal University of Technology for financial support and assistance

## REFERENCES

- [1] J. M. Marchetti, V. U. Miguel, and A. F. Errazu, "Possible methods for biodiesel production," *Renew. Sustain. Energy Rev.*, vol. 11, no. 6, pp. 1300–1311, 2007.
- [2] M. R. Avhad, M. Sanchez, A. Bouaid, M. Martinez, J. Aracil, and J. M. Marchetti, "Modeling chemical kinetics of avocado oil ethanolysis catalyzed by solid glycerol-enriched calcium oxide," *Energy Convers. Manag.*, vol. 126, pp. 1168–1177, 2016.
- [3] D. Y. C. Leung, X. Wu, and M. K. H. Leung, "A review on biodiesel production using catalyzed transesterification," *Appl. Energy*, vol. 87, no. 4, pp. 1083–1095, 2010.
- [4] A. E. Atabani, A. S. Silitonga, H. C. Ong, T. M. I. Mahlia, H. H. Masjuki, I. A. Badruddin, and H. Fayaz, "Non-edible vegetable oils: A critical evaluation of oil extraction, fatty acid compositions, biodiesel production, characteristics, engine performance and emissions production," *Renew. Sustain. Energy Rev.*, vol. 18, pp. 211–245, 2013.
- [5] H. Pourzolfaghar, F. Abnisa, W. Mohd, and A. Wan, "A review of the enzymatic hydroesterification process for biodiesel production," *Renew. Sustain. Energy Rev.*, vol. 61, pp. 245–257, 2016.
- [6] M. Yusoff, J. Brask, P. Munk, and Z. Guo, "Journal of Molecular Catalysis B: Enzymatic Kinetic model of biodiesel production catalyzed by free liquid lipase from *Thermomyces lanuginosus*," *Journal Mol. Catal. B, Enzym.*, vol. 133, pp. 55–64, 2016.
- [7] J. Amoah, E. Quayson, S. Hama, A. Yoshida, T. Hasunuma, C. Ogino, and A. Kondo, "Simultaneous conversion of free fatty acids and triglycerides to biodiesel by immobilized *Aspergillus oryzae* expressing *Fusarium heterosporum* lipase," *Biotechnol. J.*, vol. 12, no. 3, 2017.
- [8] G. Baskar, A. Guruguladevi, T. Nishanthini, R. Aiswarya, and K. Tamilarasan, "Optimization and kinetics of biodiesel production from Mahua oil using manganese doped zinc oxide nanocatalyst," *Renew. Energy*, vol. 103, pp. 641–646, 2017.
- [9] C. Muthukumar, R. Praniash, P. Navamani, R. Swathi, G. Sharmila, and N. Manoj, "Process optimization and kinetic modeling of biodiesel production using non-edible *Madhuca indica* oil," *Fuel*, vol. 195, pp. 217–225, 2017.
- [10] P. Verma and M. P. Sharma, "Comparative analysis of effect of methanol and ethanol on Karanja biodiesel production and its optimisation," *Fuel*, vol. 180, pp. 164–174, 2016.
- [11] M. Ilmi, A. Hommes, J. G. M. Winkelmana, C. Hidayatc, and H. J. Heeres, "Enzymatic biodiesel synthesis using novel process intensification principles," *Biochem. Eng. J.*, vol. 114, pp. 110–118, 2017.
- [12] A. Demirbas, "Biofuels securing the planet's future energy needs," *Energy Convers. Manag.*, vol. 50, no. 9, pp. 2239–2249, 2009.
- [13] A. P. S. Chouhan and A. K. Sarma, "Modern heterogeneous catalysts for biodiesel production: A comprehensive review," *Renew. Sustain. Energy Rev.*, vol. 15, no. 9, pp. 4378–4399, 2011.
- [14] R. Richard, Y. Li, B. Dubreuil, S. Thiebaut-Roux, and L. Prat, "On-line monitoring of the transesterification reaction between triglycerides and ethanol using near infrared spectroscopy combined with gas chromatography," *Bioresour. Technol.*, vol. 102, no. 12, pp. 6702–6709, 2011.
- [15] G. F. Zagonel, P. Peralta-Zamora, and L. P. Ramos, "Multivariate monitoring of soybean oil ethanolysis by FTIR," *Talanta*, vol. 63, no. 4, pp. 1021–1025, 2004.
- [16] M. R. Avhad and J. M. Marchetti, "A review on recent advancement in catalytic materials for biodiesel production," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 696–718, 2015.
- [17] N. Lukovi, Z. Knezevi, and D. Bezbradica, "Biodiesel Fuel Production by Enzymatic Transesterification of Oils: Recent Trends, Challenges and Future Perspectives," *Dr. Maximi.*, InTech, 2011, pp. 47–72.
- [18] Y. M. Sani, W. M. A. Daud, and A. R. Abdul Aziz, "Biodiesel Feedstock and Production Technologies: Successes, Challenges and Prospects," *Biodiesel - Feed. Prod. Appl.*, pp. 77–101, 2012.
- [19] H. H. Mardhiah, H. C. Ong, H. H. Masjuki, S. Lim, and H. V. Lee, "A review on latest developments and future prospects of heterogeneous catalyst in biodiesel production from non-edible oils," *Renew. Sustain. Energy Rev.*, vol. 67, pp. 1225–1236, 2017.
- [20] H. J. Kim, B. S. Kang, M. J. Kim, Y. M. Park, D. K. Kim, J. S. Lee, and K. Y. Lee, "Transesterification of vegetable oil to biodiesel using heterogeneous base catalyst," *Catal. Today*, vol. 93–95, pp. 315–320, 2004.
- [21] P. D. Patil and S. Deng, "Optimization of biodiesel production from edible and non-edible vegetable oils," *Fuel*, vol. 88, no. 7, pp. 1302–1306, 2009.
- [22] J. Zhang, S. Chen, R. Yang, and Y. Yan, "Biodiesel production from vegetable oil using heterogenous acid and alkali catalyst," *Fuel*, vol. 89, no. 10, pp. 2939–2944, 2010.
- [23] K. Thinnakorn and J. Tscheikuna, "Transesterification of palm olein using sodium phosphate impregnated on an alumina support," *Appl. Catal. A Gen.*, vol. 484, pp. 122–133, 2014.
- [24] J. Amoah, S. H. Ho, S. Hama, A. Yoshida, A. Nakanishi, T. Hasunuma, C. Ogino, and A. Kondo, "Lipase cocktail for efficient conversion of oils containing phospholipids to biodiesel," *Bioresour. Technol.*, vol. 211, no. April, pp. 224–230, 2016.
- [25] M. Y. Firdaus, Z. Guo, and S. N. Fedosov, "Development of kinetic model for biodiesel production using liquid lipase as a biocatalyst, esterification step," *Biochem. Eng. J.*, vol. 105, pp. 52–61, 2016.
- [26] G. Knothe and L. F. Razon, "Biodiesel fuels," *Prog. Energy Combust. Sci.*, vol. 58, pp. 36–59, 2017.
- [27] M. H. M. Killner, J. J. R. Rohwedder, and C. Pasquini, "A PLS regression model using NIR spectroscopy for on-line monitoring of the biodiesel production reaction," *Fuel*, vol. 90, no. 11, pp. 3268–3273, 2011.
- [28] S. Al-zuhair, F. Wei, and L. Song, "Proposed kinetic mechanism of the production of biodiesel from palm oil using lipase," *Process Biochem.*, vol. 42, pp. 951–960, 2007.
- [29] M. K. Lam, K. T. Lee, and A. R. Mohamed, "Homogeneous, heterogeneous and enzymatic catalysis for transesterification of high free fatty acid oil (waste cooking oil) to biodiesel: A review," *Biotechnol. Adv.*, vol. 28, no. 4, pp. 500–518, 2010.
- [30] Samir Najem Aldeen Khurshid, "Biodiesel Production by Using Heterogeneous Catalysts," *R. Inst. Technol.*, no. March, pp. 1–63, 2014.
- [31] N. S. Talha and S. Sulaiman, "Overview of Catalysts in Biodiesel Production," *ARPN J. Eng. Appl. Sci.*, vol. 11, no. 1, pp. 439–448, 2016.
- [32] P. M. Ejikeme, I. D. Anyaogu, C. L. Ejikeme, N. P. Nwafor, C. A. C. Egbuonu, K. Ukogu, J. A. Ibemesi, I. Chemistry, and F. Polytechnic, "Catalysis in Biodiesel Production by Transesterification Processes-An Insight," vol. 7, no. 4, pp. 1120–1132, 2010.
- [33] M. L. T. Cossio, L. F. Giesen, G. Araya, M. L. S. Pérez-Cotapos, R. L. VERGARA, M. Manca, R. A. Tohme, S. D. Holmberg, T. Bressmann, D. R. Lirio, J. S. Román, R. G. Solís, S. Thakur, S. N. Rao, E. L. Modelado, A. D. E. La, C. Durante, U. N. A. Tradición, M. En, E. L. Espejo, D. E. L. A. S. Fuentes, U. A. De Yucatán, C. M. Lenin, L. F. Cian, M. J. Douglas, L. Plata, and F. Héritier, *New and Future Developments in Catalysis: Catalytic Biomass Conversion*, vol. XXXIII, no. 2, 2012.
- [34] A. Guldhe, B. Singh, T. Mutanda, K. Permaul, and F. Bux, "Advances in synthesis of biodiesel via enzyme catalysis: Novel and sustainable approaches," *Renew. Sustain. Energy Rev.*, vol. 41, pp. 1447–1464, 2015.
- [35] M. R. Monteiro, A. R. P. Ambrozini, L. M. Lião, and A. G. Ferreira, "Critical review on analytical methods for biodiesel characterization," *Talanta*, vol. 77, no. 2, pp. 593–605, 2008.
- [36] W. B. Zhang, "Review on analysis of biodiesel with infrared spectroscopy," *Renew. Sustain. Energy Rev.*, vol. 16, no. 8, pp. 6048–6058, 2012.
- [37] N. Mahamuni, I. Devices, Y. Adewuyi, N. Carolina, and T. St, "Fourier Transform Infrared Spectroscopy ( FTIR ) Method To Monitor Soy Biodiesel and Soybean Oil in Transesterification ... Fourier Transform Infrared Spectroscopy ( FTIR ) Method To Monitor Soy Biodiesel and Soybean Oil in Transesterification Reactions," vol. 76, no. April, pp. 3773–3782, 2017.
- [38] T. Yuan, E. Akochi-Koble, D. Pinchuk, and F. de Voort, "FTIR On-line Monitoring of Biodiesel Transesterification," *Int. J. Renew. Energy Biofuels*, vol. 2014, pp. 1–13, 2014.
- [39] M. A. Dubé, S. Zheng, D. D. McLean, and M. Kates, "A comparison of attenuated total reflectance-FTIR spectroscopy and GPC for monitoring biodiesel production," *JAOCs, J. Am. Oil Chem. Soc.*, vol. 81, no. 6, pp. 599–603, 2004.
- [40] G. Knothe, "Analyzing Biodiesel: Standards and Other Methods," vol. 83, no. 10, pp. 823–833, 2006.
- [41] S. N. Rabelo, V. P. Ferraz, L. S. Oliveira, and A. S. Franca, "FTIR Analysis for Quantification of Fatty Acid Methyl Esters in Biodiesel Produced by Microwave-Assisted Transesterification," *Int. J. Environ. Sci. Dev.*, vol. 6, no. 12, pp. 964–969, 2015.
- [42] C. L. Cunha, A. S. Luna, R. C. G. Oliveira, G. M. Xavier, M. L. L. Paredes, and A. R. Torres, "Predicting the properties of biodiesel and its

- blends using mid-FT-IR spectroscopy and first-order multivariate calibration,” *Fuel*, vol. 204, pp. 185–194, 2017.
- [43] A. Purandaradas, T. Silambarasan, K. Murugan, R. Babujanathanam, A. D. Gandhi, K. V. Dhandapani, D. Anbumani, and P. Kavitha, “Development and quantification of biodiesel production from chicken feather meal as a cost-effective feedstock by using green technology,” *Biochem. Biophys. Reports*, vol. 14, no. January, pp. 133–139, 2018.
- [44] O. Mowla, E. Kennedy, and M. Stockenhuber, “In-situ FTIR study on the mechanism of both steps of zeolite-catalysed hydroesterification reaction in the context of biodiesel manufacturing,” *Fuel*, vol. 232, no. January, pp. 12–26, 2018.
- [45] P. Mohamed Shameer and K. Ramesh, “FTIR assessment and investigation of synthetic antioxidant on the fuel stability of *Calophyllum inophyllum* biodiesel,” *Fuel*, vol. 209, no. January, pp. 411–416, 2017.
- [46] S. M. De Lima, B. F. A. Silva, D. V. Pontes, C. F. Pereira, L. Stragevitch, and M. F. Pimentel, “In-line monitoring of the transesterification reactions for biodiesel production using NIR spectroscopy,” *Fuel*, vol. 115, pp. 46–53, 2014.
- [47] P. Chand, C. V. Reddy, J. G. Verkade, T. Wang, and D. A. Grewell, “Novel Characterization Method of Biodiesel Produced from Soybean Oil using Thermogravimetric Analysis,” 2008.
- [48] A. Oleszko, J. Hartwich, A. Wójtowicz, G. Marlina, and H. Huras, “Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy Comparison of FTIR-ATR and Raman spectroscopy in determination of VLDL triglycerides in blood serum with PLS regression,” vol. 183, pp. 239–246, 2017.
- [49] J. F. de Souza, Adriana Velloso A. Cajaiba da Silva, “Biodiesel Synthesis Evaluated by Using Real-Time ATR-FTIR,” *Org. Process Res. Dev.*, 2012.
- [50] D. Mueller, M. F. Ferrão, L. Marder, A. Ben da Costa, and R. de C. de S. Schneider, “Fourier transform infrared spectroscopy (FTIR) and multivariate analysis for identification of different vegetable oils used in biodiesel production,” *Sensors (Basel)*, vol. 13, no. 4, pp. 4258–4271, 2013.
- [51] P. Singh, H. C. Andola, M. S. M. Rawat, G. J. N. Pant, and V. K. Purohit, “Fourier Transform Infrared (FT-IR) Spectroscopy in An-Overview,” *Research Journal of Medicinal Plants*, vol. 5, no. 2, pp. 127–135, 2011.
- [52] G. Knothe, “Rapid monitoring of transesterification and assessing biodiesel fuel quality by near-infrared spectroscopy using a fiber-optic probe,” *JAOCS, J. Am. Oil Chem. Soc.*, vol. 76, no. 7, pp. 795–800, 1999.
- [53] M. G. Siatis, N. G., Kimbaris, A. C., Pappas, C. S., Tarantilis, P. A. & Polissiou, “Improvement of Biodiesel Production Based on the Application of Ultrasound: Monitoring of the Procedure by FTIR Spectroscopy,” *J. Am. Oil Chem. Soc.*, no. 83, pp. 853–857, 2006.
- [54] G. N. and T. Y. A. Abbaszadeh, B. Ghobadian, “an Experimental Investigation of the Effective,” *Int. J. Automot. Mech. Eng.*, vol. 9, no. June, pp. 1525–1537, 2014.
- [55] U. Guth, “Fast and Effective DoE Studies For Innovative Chemical Development,” White Paper, 2016. (Online) Available: <http://www.mt.com/easymax>. (Accessed: 19-Jun-2017).
- [56] M. J. Simon, “Concrete Mixture Optimization Using Statistical Methods : Final Report.”