ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

A Multivariate Regression Model for Estimation of Biodiesel Production from Jatropha Seeds

Eman Tora *, Hawash S, Abdel Kader E, and El Diwani G.

Department of Chemical Engineering and Pilot Plant, National Research Centre, El Dokki, El Giza, 12311, Egypt

ABSTRACT

This paper presents a statistical technique to characterize the biodiesel production process from Jatropha seeds and predict the conversion efficiency of the oil into biodiesel. A multivariate regression model is proposed herein to simultaneously capture the joint effect of the variable operating conditions. It is a general technique and can be used with different production methods, but this paper uses experimental results of the in- situ heterogeneous extraction and transesterification process. The results imply the success of the introduced technique to assess how much the operating variables affect the conversion. Also, the joint effect of the time, catalyst dose, methanol to oil ratio, and hexane to oil ratio has been defined. Experimental measurements have been used to perform this analysis. A linear multiple regression model with coefficient of determination (R^2) of 0.999 has been developed whereby the four investigated operation variables are related with the conversion into biodiesel has been developed. Hence, utilizing the coefficients of the developed model have been used to rank their significance, which is the most important variable and to what extent. Accordingly the most important factors are the time and the hexane ratio.

Keywords: Multivariate analysis, modeling, biodiesel production, operating variables, Statistics.

**** Corresponding author:***

INTRODUCTION

Biodiesel as a renewable energy source has attained significant attraction worldwide. Its production growing significantly; USA alone produced 121 million gallons in July 2015. This is motivated by the promising energetic and economical value of biodiesels [1] [2]; biodiesel production cost from soybean was typically \$0.55/liter at the same time the diesel wholesale price was around \$0.46/liter [3]. Likewise, environmental favorable impact in addition to its fuel value is another considerable merit [4]. It is an alternative fuel that can be used as a transportation fuel whereby it can be mixed with other materials to give appropriate fuel mix [5]. There are different processes of biodiesel production, and each process has its own special operating conditions of temperature, pressure, catalyst, etc [6]. Most often, investigating the impact of these variables takes place via fixing all the operating variables at predefined values, except one variable. That latter variable is investigated and treated as the independent variable of the process. After that, another variable is investigated via repeating the procedure. Nevertheless, studying the impact of the different operating conditions simultaneously may give further understanding of the process, which may open the door for more enhancements.

To conduct this type of analysis - investigating many independent variables and their effect on an independent variable simultaneously- a multivariate analysis sounds as an appropriate technique [7] [8] [9]. This study applies multivariate analysis in order to define the significance of the different operating conditions of biodiesel production considering the heterogeneous in situ extraction and tranesterification process. Moreover, these variables are ranked to show which of them has the most important role. A linear relation has been developed using experimental results of our previous work [10]. This statistically drawn relation between the different investigated operating conditions and the conversion into biodiesel has been used to identify the optimum combination of these different operating conditions in order to maximize the conversion into biodiesel.

BIODIESEL PRODUCTION

At the present, there are a variety of methods to produce biodiesel from several materials. Base-catalyzed transesterification is one of these methods; it uses hexane, methanol, and a CaO as a base catalyst in order to produce biodiesel from Jatropha Curcas (JCL) seeds. The production process includes insuito oil extraction and transesterification, which makes controlling the operation a critical issue especially with existence of different variable operating conditions. In our previous work [6], the effect of different operating conditions on the biodiesel production and conversion ratio has been investigated, namely the reaction time, catalyst concentration, methanol to oil ratio, hexane to oil ratio. The effect of these variables has been analyzed via manipulating their values over considerable range. The results of this study have been used herein as the input data to perform a multivariate analysis.

MULTIVARIATE ANALYSIS

Multivariate analysis can be classified into sub-techniques; however, all can deal with a set of data such as measurements on more than one variable. Its significant advantage is the capacity to deal with more multiple dependent variables simultaneously. There are many appreciated applications employing multivariate analysis, for instance in genetic science huge data sets can be analyzed in order to look for and define new genomes [11].

Among those broad applications of multivariate, exploratory data analysis is the goal of this study. Particularly looking for a hidden joint relation between the different operating conditions of biodiesel production from Jatropha via drawing this relation from considerable amount of experimental measurements is tackled and implemented here. Furthermore, multivariate analysis can reveal whether there are any of the variables much more important than the others and which extent, and vise versa. Figure 1 shows how multivariate regression is to be explored here in order to develop a model relating the different operating variables of biodiesel production with the conversion of oil to the biodiesel. In addition Fig. 1 acts as a visual comparison between the multivariate regression proposed here and the conventional regression techniques.

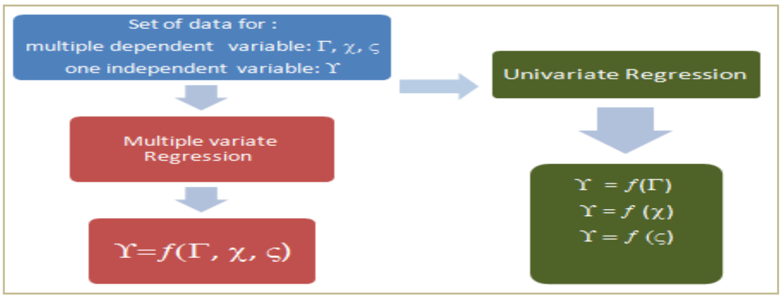

Fig. 1: A comparison between multivariate regression (left) and univarite regression (right).

UNIVARIATE REGRESSION OF EXPERIMENTAL MEASUREMENTS

With reference to section 2, the measurements of our previous work have been used here as the data set. Listing these measurements may be a tedious task and this does not fit here; its graphical representation sounds appropriate, especially this may have a twofold value. In addition to visualizing these data, the relation between each pair can be perceived.

As Fig.2 presents each of the figures represents the relation between only two of the working variables, with fixing the remaining process variables at pre-defined values. Thus, this type of regression model cannot give a comprehensible reflection on the real relation and interaction between the process different operating conditions, and consequently a potential to enhancement may be lost. Fulfilling this gap and overcome this shortage is the reason and motivation for this work to develop multivariate regression model for the process.

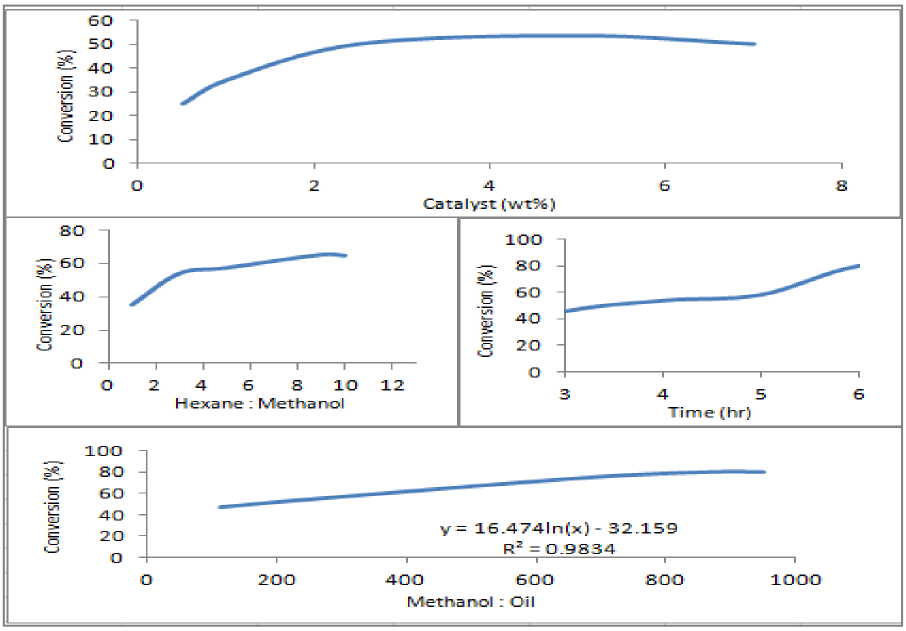

Fig. 2: Univariate Regression of operating variables, adapted from [10]

MULTIVARIATE REGRESSION

This paper, five operating condition are selected as the independent variables of the multivariate analysis, and one dependent variables is the conversion of Jatropha oil into biodiesel. These operation variables are the time, the hexane to methanol ratio, the hexane to oil ratio, catalyst dose, and the methanol to oil ratio.

Multivariate analysis is applied to define what operating variables affect most on the process. Then a relationship between these variables and the targeted objective (conversion of oil into biodiesel) is drawn based on experimental measurements.

The developed multiple regression model is a linear relation including all the investigated variables. Therefore, it can be used to identify the optimum combination of the different operating variables. The optimum here indicates the maximum conversion of Jatropha oil under the considered operating condition ranges into biodiesel. The ranges of the operating conditions have been selected to well fit the technical acceptable values, the minimum and the maximum value of each variable according to the data available in literature.

RESULTS AND DISCUSSION

First, the correlation matrix has been developed to check whether there is a relationship between the investigated different variables. The matrix is given as Table 1. It is evident that there is a strong relationship between the different investigated variables and the conversion.

From Table 2, the developed multiple regression model can be written as follow:

$$Conversion = -30.9388 + 8.7762\left(\frac{Hexane}{Methanol}\right) - 0.3993\left(\frac{Methanol}{Oil}\right) + 37.9052\ wt.\%\ catalyst + 32.46\ time$$

Where: $\frac{Hexane}{Methanol} \in \{1,10\}; \frac{Methanol}{Oil} \in \{100,1000\}; catalyst\ wt.\% \in \{0.5,9\}; time \in \{1,7\}$

The results listed in Table 2 read that the R Square value is 0.9999 which is almost one. This indicates that the variation in the conversion ratio is almost controlled by these four operating variables. With 95% confidence level, the significant F is much less than F value, therefore the model well repetitive of the experimental results.

The coefficient of each variable gives good clue on its influence extent. For instance, the coefficient of hexane to methanol ratio is a high positive value, thus this variable has a significant role in the conversion and its increase significantly promotes the conversion [6]. On the other side, methanol to oil ratio has a relatively negative value which may refer to the fact that methanol is not efficient solvent for oil extraction. Another likely explanation for that negative coefficient is that the considered ratio range was large, so maybe this ratio exceeds the optimum at some points- maybe there is a peak and after that the influence of the alcohol changes. Nonetheless, this point still weird and needs to be studied furthermore to come up with an elaborate reason but this is out of scope of this work; it may be useful to select the appropriate range of the measurements. Catalyst shall be used in small quantities as its coefficient is very high so its effect shall be remarkable. According to these coefficients, the most important factors affecting the conversion can be ranked starting from the most important as follows: time, catalyst, Hexane/Methanol, methanol/oil.

In case of the catalyst, the p- value is 0.3341 is higher than the acceptable range, which means catalyst shall be unpredictable. Thus to improve the model prediction, the model has been re-driven via including three dependent variables and excluded the catalyst dose. Table 2 has presents the attained model after omitting the catalyst concentration. For that, the regression model can be re-written as:

$$Conversion = -31.364 + 9.127\frac{Hexane}{Methanol} - 0.411\frac{Methanol}{Oil} + 32.809\ t$$

Table- 1. Correlation matrix.

	Hexane : Methanol	*Catalyst wt%*	*Time*	*Methanol : Oil*
Hexane : Methanol	1			
Catalyst wt%	0.996057	1		
Time	0.910488	0.873541	1	
Methanol : Oil	0.994087	0.980687	0.948528	1

Table– 2. Summary of the multiple regression relating four operating variables with the conversion ratio into biodiesel.

Regression Statistics								
Multiple R	0.99998							
R Square	0.99996							
Adjusted R Square	0.99993							
Standard Error	0.13457							
Observations	9							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	4	1999.928	499.982	27610.117	3.93498E-09			
Residual	4	0.072	0.018					
Total	8	2000						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-30.9389	1.3144	-23.5385	0.0000	-34.5883	-27.2895	-34.5883	-27.2895
Hexane : Methanol	8.7762	0.9172	9.5687	0.0007	6.2297	11.3227	6.2297	11.3227
Catalyst wt%	37.9052	34.5391	1.0975	0.3341	-57.9907	133.8010	-57.9907	133.8010
Time	32.4600	0.9801	33.1177	0.0000	29.7387	35.1814	29.7387	35.1814
Methanol : Oil	-0.3993	0.0323	-12.3472	0.0002	-0.4890	-0.3095	-0.4890	-0.3095

ISSN: 0975-8585

Table- 3. Summary of the multiple regression relating three operating variables with the conversion ratio into biodiesel.

Regression Statistics								
Multiple R	0.99998							
R Square	0.99995							
Adjusted R Square	0.99992							
Standard Error	0.13729							
Observations	9							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	1999.906	666.635	35367.190	0.000000			
Residual	5	0.094	0.019					
Total	8	2000.000						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	
Intercept	-31.364	1.281	-24.476	0.000	-34.658	-28.070	-34.658	
Hexane : Methanol	9.127	0.877	10.405	0.000	6.872	11.382	6.872	
Time	32.809	0.946	34.680	0.000	30.377	35.240	30.377	
Methanol : Oil	-0.411	0.031	-13.183	0.000	-0.491	-0.331	-0.491	

ISSN: 0975-8585

CONCLUSIONS

In this paper, a multivariate model is developed to represent the joint effect of multiple operating variables affecting on Jatropha oil conversion into biodiesel. The model developed using experimental measurements attained from our previous work on biodiesel production from Egyptian Jatropha seeds via a heterogeneous in situ extraction and transesterification. The model accuracy has been evaluated based on the coefficient of determination (R^2) and the error probability (*P-value*). A multivariate analysis has been applied to determine which factors are the most important and to what extent. A multiple regression model has been developed to relate four operating variables with the conversion. This model can be used to assess the conversion with any proposed combinations of the operating conditions, which could be helpful for different stakeholders whether in the academia or the industry in evaluation of the process feasibility in terms of conversion efficiency under different operating conditions. Applying multivariate analysis to develop multiple regression model relating Jatropha oil conversion into biodiesel with operating variables is a main contribution here, yet we consider it a primary step toward a more elaborate model once more experimental measurements become available.

REFERENCES

[1] USA Department of Energy, USA Energy Information Administration, www.eia.gov/biofuels/biodiesel/production/, accessed on October 5, 2015.
[2] Hill J, Nelson E, David T, Polasky S, Tiffany D. Proceeding of National Academy of Sciences of the United States of America;103: 11206- 11210
[3] Fortenbery TR, Department of Agricultural and Applied Economics Staff Paper No. 481 (2005) Biodiesel Feasibility Study: An Evaluation of Biodiesel Feasibility, Univ. of Wisconsin,Madison, WI.
[4] Tora EA, El-Halwagi MM. Computer and Chemical Engineering 2011;35: 96-76
[5] Tora, E. A, A. El-Baz, D. El-Monayeri, and M. M. El-Halwagi. Design for Energy and the Environment: Proceedings of the 7th International Conference on the Foundations of Computer-Aided Process Design (FOCAPD), CRC Press, Taylor & Francis, 2009
[6] Marchetti, JM, Miguel VU, Errazu, A, Renewable and Sustainable Energy Reviews; 2007; 11: 1300-1311
[7] Timm, NH. Applied Multivariate Analysis, Springer, NY, 2002.
[8] Moore D, McCabe G. Introduction to the Practice of Statistics, W.H.Freeman and Company, New York, 1993
[9] Mertler GA, Vannatta RA, Advanced and Multivariate Statistical Methods: Practical Application and Interpretation, 2nd ed., 2002
[10] Hawash SI, Abdel Kader E, EL Diwani, G. AFINIDA; 2015; LXXII, 570
[11] Introduction to multivariate analysis: Applications to genomic data. Mar 2014, http://adegenet.r-forge.r-project.org/files/simGWAS/lecture-MSc-MVA.1.2.pdf