ISSN NO. 2320-5407

Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/3719
DOI URL: http://dx.doi.org/10.21474/IJAR01/3719

RESEARCH ARTICLE

APPLICATION OF DESIGN EXPERT IN THE ANALYSIS OF RESPONSE TRANSFORMATION OF PROCESSES – A CASE STUDY OF BIOETHANOL PRODUCTION PROCESS FROM CORN-STOVER.

***Ohimor Evuensiri Onoghwarite[1], Ndirika Victor Ifeanyichukwu Obiora[2] and Eke Akachukwu Ben[2].**

1. Department of Chemical Engineering, Federal University of Petroleum Resources, Effurun, Delta State, Nigeria.
2. Department of Agricultural and Bioresources Engineering, Michael Okpara University of Agriculture, Umudike Abia State, Nigeria.

……………………………………………………………………………………………………....

Manuscript Info

Manuscript History

Received: 15 January 2017
Final Accepted: 05 February 2017
Published: March 2017

Abstract

Statistical tools and features of the Design Expert software (7.0.0 Trial version) were used in experimental design and analyses of bioethanol production process from corn stover. The Box-Behnken design (BBD) was used in the design of the experiments and it yielded a total of 54 experimental runs. The response variable, Bioethanol (mg/l) was represented with Y, while the six independent variables; Sulphuric acid concentration, Hydrolysis time, Fermentation time, Concentration of yeast, Fermentation temperature and pH of Hydrolysate samples were represented by A, B, C, D, E & F respectively. A model equation was developed from the experimental results using the analysis of variance (ANOVA) feature of the Design Expert software. From the analyses of variance, only the terms A, F, AF, CD, A^2 were significant model terms (having P-value ≤ 0.05), which resulted in a reduced Model Equation (i.e. Adjusted) in terms of actual factors, as follows: $Y = 552.60398 - 0.31052*A + 3.87830*F - 5.76556*A*F - 0.26625*C*D + 5.38883*A^2$. The optimum process parameters for the predicted optimum bioethanol yield of 149.41 mg/l were 1.08%, 3.32hrs, 14.32hrs, 6.43g/l, 39.34°C and 7.64 for A, B, C, D, E & F respectively. However, an experimental bioethanol yield of 143.15 mg/l was obtained from three experimental replicates with the optimum process parameters. The experimental bioethanol yield obtained was 95.81% close to the predicted optimum bioethanol yield. The values of R, R^2, adjusted R^2 and Adeq Precision were 0.8747, 0.7651, 0.5212 and 8.309 respectively. The value of R^2 indicates a good degree of correlation between the experimental and predicted bioethanol yields. The model is adjudged to be adequate since the "Adeq Precision" is greater than 4.

……………………………………………………………………………………………………....

Introduction:-

The bioethanol production from a lignocellulosic material, corn-stover using the separate hydrolysis and fermentation (SHF) process where hydrolysis (saccharification) of the cellulose to glucose occurs first, separate

Corresponding Author:- Ohimor Evuensiri Onoghwarite.
Address:- Department of Chemical Engineering, Federal University of Petroleum Resources, Effurun, Delta State, Nigeria.

from the fermentation of the resulting glucose and other sugars (from the pretreatment of hemicellulose,) to ethanol. The separate saccharification step enables operation of the saccharification step at an elevated temperature. For fermentation, the *Saccharomyces cerevisiae* is used as the biocatalyst which will ferment glucose and other sugars to ethanol. Some important factors that affect the performance of separate hydrolysis and fermentation (SHF) and thus bioethanol yield are: concentration of the hydrolysis solution, hydrolysis time, pH of hydrolysate, concentration of yeast, fermentation temperature and time. The feasibility of these two operations is determined by the effects of each factor as well as the interactions between factors at varying values of all the factors.

Thus to design the experiment involving multivariable factors, a statistical software package like the Design-Expert, that is specifically dedicated to performing comparative tests, screening, characterization, optimization, robust parameter design, mixture designs and combined designs is highly recommended (Tanco *et al.*, 2008). Design-Expert offers test matrices for screening up to 50 factors. To keep pace with this, therefore, a power calculator which helps to establish the number of test runs needed is required. Statistical significance of these factors is established with analysis of variance (ANOVA). Based on the validated predictive models, a numerical optimizer helps the user to determine the ideal values for each of the factors in the experiment. Graphical tools help identify the impact of each factor on the desired outcomes and reveal abnormalities in the data (Cornley, 2009). Design-Expert provides 11 graphs in addition to text output to analyze the residuals (Plant, 2013). The software determines the main effects of each factor as well as the interactions between factors by varying the values of all factors in parallel (Black, 2013). Design-Expert provides the user with a broad range of possible response transformations. The appropriate choice depends on subject matter and/or statistical considerations. The software provides extensive diagnostic capabilities to validate statistical assumptions. This feature of the Design-Expert offers a helpful plot, called the Box-Cox, which recommends the appropriate power transformation (including the no transformation option). Design-Expert also has an option of plotting the responses in terms of the original response data by calculating the surface matrix of data points and then applying the inverse transform before making the plot.

In the present study, the application of statistical tools of the design expert software to determine statistical significance of the various process factors in bioethanol production from corn-stover is presented as a case study. The impact of these factors on the response (i.e. bioethanol yield), diagnostic capacities to validate statistical assumptions, possible response transformations and the criteria for the choices of transformation were clarified.

Materials and Methods:-

2.1 Design of Experiment:-

Box-Behnken design (BBD), a statistical tool, was used in the design of the experiment in this study. In BBD, the six independent variables; sulphuric acid concentration X_1 (A), hydrolysis time X_2 (B), fermentation time X_3 (C), concentration of yeast X_4 (D), fermentation temperature X_5 (E), and pH of hydrolysate X_6 (F), were all set at three levels; minimum (-1), centre (0) and maximum (1). The BBD was used to determine the number of runs or sets of experiments that were needed to be carried out. Each experiment was carried out in triplicates and the mean values outcome, taken as response, Y, which is bioethanol yield (mg/l).

For statistical calculations, the variables X_1, X_2, $X_{3\,\ldots\ldots}$ X_n were coded as x_1, x_2, $x_{3\ldots\ldots Xn}$ respectively, according to Equation 1 Nuran (2007):

$$x_i = \frac{(X_i - \bar{x}_i)}{\Delta x_i} \quad i = 1, 2, 3, ., n \qquad (1)$$

Where x_i is a dimensionless value of the independent variable, x_i is the mean value of the independent variable at the center point and Δx_i is the step change.

The design matrix used for the six independent variables showing the real values of the variables in terms of the three levels is presented in Table 1.

Table 1:- Design matrix employed for the six independent variables

Factors	Code	Range		
		-1	0	+1
Sulphuric Acid Concentration (%)	A	1.00	2.50	4.00
Hydrolysis Time (Hrs)	B	2.00	4.00	6.00
Fermentation Time (Hrs)	C	12.00	30.00	48.00
Concentration of Yeast (g/l)	D	3.00	6.00	9.00
Fermentation Temperature (^{o}C)	E	30.00	35.00	40.00
pH of Hydrolysate	F	5.00	6.50	8.00

The outcome of the BBD experimental design of the process (Separate Hydrolysis and Fermentation, SHF) using the Design Expert software, resulted in a total of 54 experimental runs as sets of coded variables.

Pretreatment, Hydrolysis and Fermentation of corn stover:-

The pretreatment, hydrolysis and fermentation of the corn stover were carried out in line with standard procedure (Ohimor *et al.,* 2016). The milled corn stover was alkaline pretreated with dilute sodium hydroxide (2% w/w NaOH) summing to a solid to liquid weight ratio of 1:8. The hydrolysis of the alkaline pretreated corn stover was with dilute-acid (H_2SO_4) of varying concentrations (1, 2.5 and 4%), such that a solid to liquid ratio of 1:10 is maintained in a 250ml round bottom flask and then refluxed. Hydrolysate samples were retained at intervals of 2, 4, and 6 hours for subsequent fermentation. The hydrolysate sample for fermentation was adjusted to the various pH of 5, 6.5 and 8 by adding concentrated sulphuric acid and 2N sodium hydroxide as may be appropriate.

Varying quantities of yeast (*Saccharomyces cerevisiae*) equivalent to 3, 6 and 9 g/l respectively where added to each hydrolysate samples contained in 250 ml Erlenmeyer flask, then incubated at various temperatures (30, 35 and 40 °C) for fermentation. The bioethanol content was determined by gas chromatography at various fermentation times of 12, 30, and 48 hours.

Statistical Analyses:-

Upon completion of the laboratory analyses, the results (bioethanol yield) of all 54 experimental runs were entered into a table in the file already created in the Design Expert software for subsequent analysis. The results were statistically analyzed in order to have an understanding of the interactions between variables and the bioethanol yield. These results were also used for the optimization of the process.

A second order polynomial model, similar to that of Dasgupta *et al*. (2013), was obtained from the Design Expert software indicating linear, interaction and quadratic effect of variables on the response variable (*Y*).

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{11}X_1 + b_{22}X_2 + b_{33}X_3 + b_{12}X_1X_2 + b_{23}X_2X_3 + b_{13}X_1X_3 \quad (2.2)$$

Where Y is the predicted response; X_1, X_2, X_3 are the independent variables; b_o is the offset term; b_1, b_2, b_3 are the linear effects; b_{11}, b_{22}, b_{33} are the square effects; and b_{12}, b_{23}, b_{13} are the cross effects of the interaction terms.
Using regression analysis, the significance of the coefficients of the terms in the model equation was determined by computing the standard errors, T-values and P-values. Analysis of variance (ANOVA) was also used to determine the level of confidence (α) as well as linear, interaction and quadratic effects. The quality of fit of the BBD model was estimated by comparing the variance (R^2) of the adjusted model and that of the predicted Model.

Finally, the experiment was repeated in triplicate using the variables that gave the optimum bioethanol yield and the average of the results was compared to the optimum value. The validity of the model was adjudged to be good since the result is comparable to the optimum bioethanol yield to an extent. Point Prediction was carried out which allows the entering of each factor or component into the current model at different levels. The software then calculates the expected responses and associated confidence intervals based on the model equation that is shown in the ANOVA output. The predicted values are updated as the levels are changed. The **95%** confidence interval (**CI**) is the range in which the process average is expected to fall into 95% of the time. The 95% prediction interval (PI) is the range in which any individual value is expected to fall into 95% of the time. The prediction interval will normally have a wider spread than the confidence interval because of the random nature of the individual values when compared to the averages.

Results and Discussion:-

Experimental Design of the Bioethanol Production Process and Bioethanol Yield:-

The results of the design of experiments and experimental bioethanol yields (mg/l) for each runs are presented in Table 2; it shows the experimental design of the six (6) independent variables for the bioethanol production process in terms of their coded levels (i.e. -1, 0, +1).

Table 2:- Experimental design and bioethanol yield

Run Order	Coded Variables						Experimental	Std
	A	B	C	D	E	F	Bioethanol Yield (mg/l) (Y)	
1	0	0	0	0	0	0	58.14	51
2	+1	0	+1	0	0	-1	52.77	44
3	+1	0	-1	0	0	+1	53.28	46
4	0	0	-1	+1	0	+1	87.44	23
5	-1	+1	0	+1	0	0	63.89	7
6	0	+1	-1	0	+1	0	65.74	14
7	-1	-1	0	+1	0	0	86.19	5
8	+1	0	0	+1	+1	0	54.52	32
9	0	0	0	0	0	0	66.89	52
10	0	+1	0	0	+1	-1	81.88	36
11	0	-1	0	0	+1	-1	53.36	35
12	0	-1	0	0	-1	+1	37.48	37
13	+1	0	+1	0	0	+1	87.21	48
14	-1	0	0	-1	+1	0	88.32	29
15	0	0	+1	-1	0	-1	69.56	18
16	0	0	+1	+1	0	+1	39.34	24
17	0	0	0	0	0	0	67.33	49
18	0	0	+1	-1	0	+1	86.64	22
19	0	+1	-1	0	-1	0	72.86	10
20	-1	0	-1	0	0	-1	79.25	41
21	+1	-1	0	-1	0	0	61.26	2
22	+1	0	0	-1	-1	0	68.29	26
23	0	-1	0	0	+1	+1	109.23	39
24	0	-1	+1	0	-1	0	66.77	11
25	-1	0	+1	0	0	-1	68.14	43
26	0	0	+1	+1	0	-1	57.33	20
27	0	0	0	0	0	0	88.76	50
28	0	-1	-1	0	+1	0	88.68	13
29	+1	+1	0	-1	0	0	41.44	4
30	+1	-1	0	+1	0	0	39.88	6
31	0	0	-1	+1	0	-1	66.73	19
32	-1	0	0	+1	+1	0	134.88	31
33	-1	0	0	-1	-1	0	111.87	25
34	-1	0	-1	0	0	+1	139.56	45
35	0	0	-1	-1	0	+1	65.56	21
36	0	-1	+1	0	+1	0	68.22	15
37	0	0	0	0	0	0	67.33	53
38	-1	-1	0	-1	0	0	87.93	1
39	-1	0	+1	0	0	+1	147.64	47
40	+1	0	-1	0	0	-1	51.69	42
41	0	+1	0	0	+1	+1	63.42	40
42	0	+1	+1	0	-1	0	83.02	12
43	+1	+1	0	+1	0	0	62.24	8
44	-1	+1	0	-1	0	0	76.38	3

45	+1	0	0	-1	+1	0	28.62	30
46	+1	0	0	+1	-1	0	39.74	28
47	0	-1	0	0	-1	-1	66.13	33
48	-1	0	0	+1	-1	0	114.23	27
49	0	-1	-1	0	-1	0	68.01	9
50	0	+1	0	0	-1	+1	68.06	38
51	0	+1	0	0	-1	-1	73.76	34
52	0	0	-1	-1	0	-1	33.12	17
53	0	0	0	0	0	0	67.33	54
54	0	+1	+1	0	+1	0	91.45	16

The analytical experiments were then carried out according to the assignment of variables in Table 2. The experimental runs were performed randomly to avoid systemic error.

Transformation Equation:-

The response (i.e Bioethanol Yield) ranged from 28.62 mg/l to 147.64 mg/l. The ratio of the maximum to minimum response determines whether a transformation equation is needed or not for the statistical analysis. A ratio greater than 10 usually indicates a transformation is required. For ratios less than 3, the power transform will have little effect. For ratios ranging from 3 to 10, a transform is not required.

Transformation types can either be Square Root, Natural Logarithm, Base 10 Logarithm, Inverse Square Root, Inverse, Power, Arcsin Square Root or None.

From the experimental results, the ratio of maximum to minimum bioethanol response is 1 : 5.15863, hence a transform was not required before we proceeded with the statistical analysis.

Model type selection:-

From the statistical analysis using Design Expert, the information on *Sequential Model Sum of Squares [Type I*] and *Lack of Fit Tests*, shown in Tables 3 and 4 respectively, were obtained.

Table 3:- Sequential model sum of squares [Type I]

Source	Sum of Squares	df	Mean Square	F- Value	p-value Prob > F	Comment
Mean vs Total	2.844E+005	1	2.844E+005			
Linear vs Mean	15434.74	6	2572.46	6.82	< 0.0001	Suggested
2FI vs Linear	6864.30	15	457.62	1.35	0.2321	
Quadratic vs 2FI	3074.07	6	512.35	1.71	0.1584	
Cubic vs Quadratic	6223.58	18	345.75	1.77	0.2080	Aliased
Residual	1565.33	8	195.67			
Total	3.175E+005	54	5880.54			

Table 4:- Lack of fit tests

Source	Sum of Squares	Df	Mean Square	F- Value	p-value Prob > F	Comment
Linear	17206.59	42	409.68	3.93	0.0647	Suggested
2FI	10342.29	27	383.05	3.68	0.0755	
Quadratic	7268.22	21	346.11	3.32	0.0932	
Cubic	1044.64	3	348.21	3.34	0.1134	Aliased
Pure Error	520.69	5	104.14			

For the *"Sequential Model Sum of Squares [Type I]"*, it is expected that the model equation that was not aliased is to be selected provided the model is shown to be siginificant in terms of its "Model F-value". In this case the Cubic vs Quadratic Model was aliased, meaning that the Cubic Model cannot be resolved. The test suggested Linear vs Mean, meaning that a Linear Model is preferred compared to a Mean. Nevertheless, the Quadratic Model was preferred over 2FI (i.e two – factor interaction) because it was not Aliased. The choice of the Quadratic Model was to account for interactions between two or more variables.

For the *"Lack of Fit Tests"*, the selected model is to have an insignificant lack-of-fit, which is a “P-value ” that is > 0.05. In this case the software suggested Linear, however, the Quadratic model is preferrable in as much as it was not aliased, with a P-value of 0.0932. The choice of the quadratic over linear model is to be able to account for interactions between two or more variables.

The "Lack of Fit F-value" of 3.32 implies there is a 3.32% chance that a "Lack of Fit F- value" as much as this could occur due to noise. Lack of fit is bad -- we want the model to fit. This relatively low probabiity (<10%) is worrisome. What this means is that there are many insignificant model terms and as such, there is need to consider model reduction in order to improve the model.

Regression analysis:-

The result of the regression analysis is shown in Table 5 below. If the T-value is above threshold and P-value is lower than 0.05 for any given term, then that term is said to have a meaningful contribution to bioethanol production. Terms that have no contribution were removed from the predicted model and a new model referred to as adjusted model was obtained.

Table 5:- Regression analysis

Source	Coefficient Estimate	Sum of Squares	DF	Mean Square	F-value	P-value (Prob. >F)	
Model	69.30	25373.11	27	939.74	3.14	0.0023	Significant
A-H_2SO_4	23.22	12942.83	1	12942.83	43.20	< 0.0001	
B-Hydrolysis Time	0.46	5.04	1	5.04	0.017	0.8978	
C-Ferm. Time	1.92	88.82	1	88.82	0.30	0.5907	
D-Conc. of Yeast	1.14	31.33	1	31.33	0.10	0.7490	
E-Ferm. Temp	2.42	140.65	1	140.65	0.47	0.4993	
F-pH	9.63	2226.07	1	2226.07	7.43	0.0113	
AB	4.55	165.53	1	165.53	0.55	0.4639	
AC	4.76	180.88	1	180.88	0.60	0.4441	
AD	-2.37	89.87	1	89.87	0.30	0.5886	
AE	-2.75	60.45	1	60.45	0.20	0.6570	
AF	-12.97	1346.29	1	1346.29	4.49	0.0437	
BC	7.20	414.29	1	414.29	1.38	0.2503	
BD	3.93	123.48	1	123.48	0.41	0.5265	
BE	-4.77	363.95	1	363.95	1.21	0.2805	
BF	-6.42	329.99	1	329.99	1.10	0.3036	
CD	-14.38	1653.70	1	1653.70	5.52	0.0267	
CE	-0.46	1.68	1	1.68	5.620E-03	0.9408	
CF	-0.13	0.26	1	0.26	8.513E-04	0.9769	
DE	12.33	1216.48	1	1216.48	4.06	0.0543	
DF	-5.85	273.78	1	273.78	0.91	0.3479	
EF	8.97	643.69	1	643.69	2.15	0.1547	
A^2	12.12	1512.13	1	1512.13	5.05	0.0334	
B^2	-6.92	492.21	1	492.21	1.64	0.2112	
C^2	4.97	254.56	1	254.56	0.85	0.3651	
D^2	-9.60	948.45	1	948.45	3.17	0.0869	
E^2	8.24	698.35	1	698.35	2.33	0.1389	
F^2	-1.45	21.74	1	21.74	0.073	0.7897	
Residual		7788.91	26	*299.57*			
Lack of Fit		7268.22	21	346.11	3.32	0.0932	Not significant
Pure Error		520.69	5	104.14			
Cor Total		33162.02	53				

Model summary statistics:-

The summary of the statistical analyses are shown in Tables 6 and 7 below.

Table 6:- Model summary statistics

Source	Std. Dev.	R-Squared	Adjusted R-Squared	Predicted R-Squared	PRESS	
Linear	19.42	0.4654	0.3972	0.2770	23975.89	Suggested
2FI	18.42	0.6724	0.4575	-0.0657	35340.10	
Quadratic	17.31	0.7651	0.5212	-0.1677	38722.52	
Cubic	13.99	0.9528	0.6873	-7.0869	2.682×10^5	Aliased

From Table 6, the choice of model was based on the Predicted R-Squared value. A positive value of Predicted R-Squared is more preferable to a negative term, hence the Cubic model was aliased (i.e ruled out) while the Linear model was suggested. However, the Two-Factor Interaction (2FI) and Quadratic models were not aliased because their Predicted R-Squared values were not too negative, hence the Quadratic model was selected.

Table 7:- Adeq Precision and various R– Squared values for quadratic model

Std. Dev.	17.31	R-Squared	0.7651
Mean	72.57	Adj R-Squared	0.5212
C.V. %	23.85	Pred R-Squared	-0.1677
PRESS	38722.52	Adeq Precision	8.309

The adequate precision and various R–Squared values for the quadratic model are given in Table 8. The "Adeq Precision" is a measure of the signal to noise ratio and a ratio greater than 4 is desirable. For the quadratic model, the "Adeq Precision" is 8.309. Since it is greater than 4, the model can be used to navigate the design space, but there is need for improvement, through the removal of insigificant model terms.

Confidence interval:-

The confidence interval (CI) depicted in Table 8 is another statistical data that can be used to determine whether a model term has effect (i.e. significant) or not on the model). Hence, the terms in the columns "**95% CI Low and 95% CI High"** represent the range that the true coefficient should be found in 95% of the time. If the range spans zero (i.e. one limit is positive and the other negative) then the coefficient would not be true, indicating that the corresponding term or factor has no effect. However, if the range does not span zero (i.e. both limits having the same sign, either positive or negative) then the coefficient would be true, meaning that the corresponding term or factor has effect. For example, the term B has a confidence interval which spans 0 since it is between -6.80 and 7.72, hence term B has no effect. Whereas term A has a confidence interval which does not span 0 since it is between -30.48 and -15.96, as such it has significant effect on the model equation. Similarly, the confidence intervals of the following terms A, F, AF, CD, A^2 as well as the intercept do not span 0 and therefore have significant effects on the model equation.

Table 8:- Confidence interval, degree of freedom and variance inflation factor

Factor	Coefficient Estimate	Degree of Freedom	Standard Error	95% CI Low	95% CI High	VIF
Intercept	69.30	1	7.07	54.77	83.82	
A-H_2SO_4	-23.22	1	3.53	-30.48	-15.96	1.00
B-Hydrolysis Time	0.46	1	3.53	-6.80	7.72	1.00
C-Fermentation Time	1.92	1	3.53	-5.34	9.19	1.00
D-Conc. of Yeast	1.14	1	3.53	-6.12	8.40	1.00
E-Fermentation Temp	2.42	1	3.53	-4.84	9.68	1.00
F-pH	9.63	1	3.53	2.37	16.89	1.00
AB	4.55	1	6.12	-8.03	17.13	1.00
AC	4.76	1	6.12	-7.82	17.33	1.00
AD	-2.37	1	4.33	-11.26	6.52	1.00
AE	-2.75	1	6.12	-15.33	9.83	1.00
AF	-12.97	1	6.12	-25.55	-0.39	1.00

BC	7.20	1	6.12	-5.38	19.77	1.00
BD	3.93	1	6.12	-8.65	16.31	1.00
BE	-4.77	1	4.33	-13.66	4.12	1.00
BF	-6.42	1	6.12	-19.00	6.16	1.00
CD	-14.38	1	6.12	-26.96	-1.80	1.00
CE	-0.46	1	6.12	-13.04	12.12	1.00
CF	-0.13	1	4.33	-9.02	8.77	1.00
DE	12.33	1	6.12	-0.25	24.91	1.00
DF	-5.85	1	6.12	-18.43	6.73	1.00
EF	8.97	1	6.12	-3.61	21.55	1.00
A^2	12.12	1	5.40	1.03	23.22	1.30
B^2	-6.92	1	5.40	-18.01	4.18	1.30
C^2	4.97	1	5.40	-6.12	16.07	1.30
D^2	-9.60	1	5.40	-20.70	1.49	1.30
E^2	8.24	1	5.40	-2.85	19.33	1.30
F^2	-1.45	1	5.40	-12.55	9.64	1.30

VIF: Variance Inflation Factor in Table 8, measures how much the variance of the model is inflated by the lack of orthogonality in the design. If the factor is orthogonal to all other factors in the model, the VIF is one. Values greater than 10 indicate that the factors are too correlated together (they are not independent.). Thus, since the values of Variance Inflation Factors range from 1.00 to 1.30, it indicates that the factors or terms are independent.

Standard Error: The Standard Error in Table 8 is associated with the calculation of the mean. It comes from the standard deviation of the data divided by the square root of the number of repetitions in a sample. The larger the value of the Standard Error of a term the more insignificant is the term.

Diagnostics Case Statistics:-

The actual value, predicted value of bioethanol and their residuals are given in Table 9, which informed the plots in Figures 1-5.

Table 9:- Internally and Externally Studentized Residual and Influence on Fitted Value

Std Order	Actual Value	Predicted Value	Residual	Leverage	Internally Studentized Residual	Externally Studentized Residual	Influence on Fitted Value	Cook's Distance	Run Order
1	87.93	92.63	-4.70	0.563	-0.411	-0.404	-0.458	0.008	38
2	61.26	41.83	19.43	0.563	1.697	1.765	* 2.00	0.132	21
3	76.38	76.59	-0.21	0.563	-0.019	-0.018	-0.021	0.000	44
4	41.44	43.98	-2.54	0.563	-0.222	-0.218	-0.247	0.002	29
5	86.19	91.80	-5.61	0.563	-0.490	-0.483	-0.547	0.01	17
6	39.88	31.52	8.36	0.563	0.731	0.724	0.821	0.025	30
7	63.89	91.47	-27.58	0.563	-2.409	-2.681	* -3.04	0.267	5
8	62.24	49.39	12.85	0.563	1.123	1.129	1.280	0.058	43
9	68.01	72.76	-4.75	0.562	-0.415	-0.408	-0.463	0.008	49
10	72.86	68.82	4.04	0.562	0.353	0.347	0.393	0.006	19
11	66.77	63.13	3.64	0.562	0.318	0.312	0.354	0.005	24
12	83.02	87.98	-4.96	0.562	0.433	-0.426	-0.483	0.009	42
13	88.68	88.06	0.62	0.562	0.054	0.053	0.061	0.000	28
14	65.74	65.04	0.70	0.562	0.061	0.060	0.068	0.000	6
15	68.22	76.59	-8.37	0.562	-0.731	-0.725	-0.822	0.025	36
16	91.45	82.36	9.09	0.562	0.794	0.788	0.893	0.029	54
17	33.12	30.16	2.96	0.563	0.258	0.254	0.287	0.003	52
18	69.56	63.02	6.54	0.563	0.571	0.564	0.639	0.015	15
19	66.73	72.90	-6.17	0.563	-0.539	-0.532	-0.603	0.013	31

20	57.33	48.25	9.08	0.563	0.793	0.787	0.893	0.029	26
21	65.56	61.38	4.18	0.563	0.365	0.359	0.407	0.006	35
22	86.64	93.73	-7.09	0.563	-0.619	-0.612	-0.694	0.018	18
23	87.44	80.72	6.72	0.563	0.587	0.580	0.657	0.016	4
24	39.34	55.56	-16.22	0.563	-1.417	-1.446	-1.640	0.092	16
25	111.87	106.93	4.94	0.563	0.431	0.425	0.481	0.009	33
26	68.29	70.72	-2.43	0.563	-0.213	-0.209	-0.236	0.002	22
27	114.23	89.29	24.94	0.563	2.178	2.362	* 2.68	0.218	48
28	39.74	43.61	-3.87	0.563	-0.338	-0.332	-0.376	0.005	46
29	88.32	92.61	-4.29	0.563	-0.374	-0.368	-0.417	0.006	14
30	28.62	45.40	-16.78	0.563	-1.466	-1.501	-1.702	0.099	45
31	134.88	124.29	10.59	0.563	0.925	0.922	1.045	0.039	32
32	54.52	67.61	-13.09	0.563	-1.144	-1.151	-1.305	0.060	8
33	66.13	54.43	11.70	0.563	1.022	1.023	1.160	0.048	47
34	73.76	77.73	-3.97	0.563	-0.347	-0.341	-0.387	0.006	51
35	53.36	50.87	2.49	0.563	0.217	0.213	0.242	0.002	11
36	81.88	55.10	26.78	0.563	2.340	2.582	* 2.93	0.251	10
37	37.48	68.60	-31.12	0.563	-2.718	-3.151	* -3.57	0.339	12
38	68.06	66.21	1.85	0.563	0.162	0.159	0.180	0.001	50
39	109.23	100.92	8.31	0.563	0.726	0.719	0.815	0.024	23
40	63.42	79.45	-16.03	0.563	-1.400	-1.428	-1.619	0.090	41
41	79.25	88.27	-9.02	0.563	-0.788	-0.782	-0.886	0.028	20
42	51.69	58.26	-6.57	0.563	-0.574	-0.566	-0.642	0.015	40
43	68.14	82.86	-14.72	0.563	-1.285	-1.303	-1.477	0.076	25
44	52.77	71.87	-19.10	0.563	-1.668	-1.731	-1.963	0.128	2
45	139.56	133.73	5.83	0.563	0.510	0.502	0.569	0.012	34
46	53.28	51.83	1.45	0.563	0.127	0.125	0.141	0.001	3
47	147.64	127.81	19.83	0.563	1.732	1.806	* 2.05	0.138	39
48	87.21	64.93	22.28	0.563	1.946	2.065	* 2.34	0.174	13
49	67.33	69.30	-1.97	0.167	-0.124	-0.122	-0.055	0.000	17
50	88.76	69.30	19.46	0.167	1.232	1.245	0.557	0.011	27
51	58.14	69.30	-11.16	0.167	-0.706	-0.699	-0.313	0.004	1
52	66.89	69.30	-2.41	0.167	-0.152	-0.149	-0.067	0.000	9
53	67.33	69.30	-1.97	0.167	-0.124	-0.122	-0.055	0.000	37
54	67.33	69.30	-1.97	0.167	-0.124	-0.122	-0.055	0.000	53

* Exceeds limits

Diagnostic plots:-

Important statistical plots derived from the analyses of the data are shown in Figures 1-5 below:

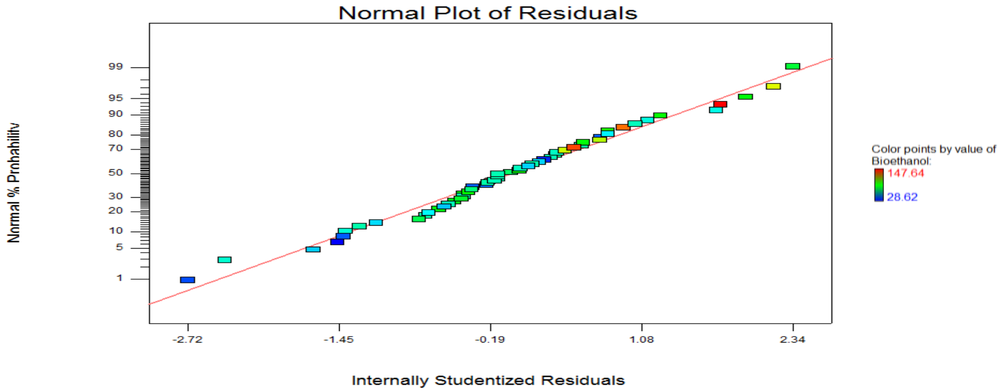

Figure 1:- Normal probability plot of the studentized residuals.

The normal probability plot of the studentized residuals in Figure 1, follows a linear pattern, which indicates a normal distribution of the residuals, hence a transformation is not required. This is in line with Hothorn, and Everitt (2014), who reported that a non-linear pattern indicates that the residuals did not follow a normal distribution, in which case a transformation of the response may provide a better analysis.

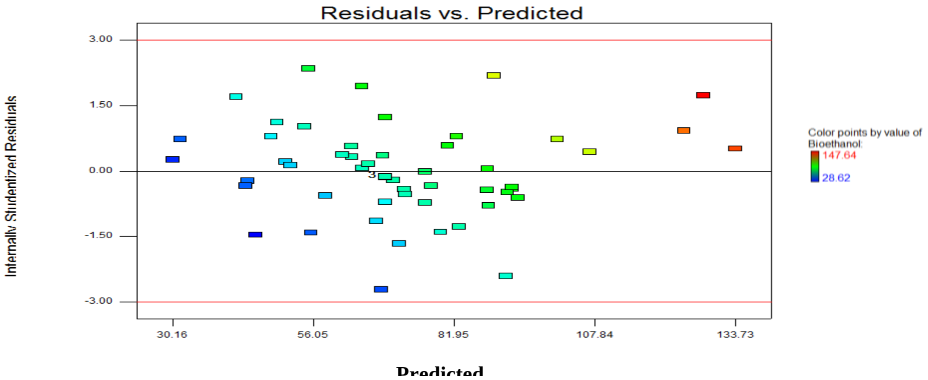

Figure 2:- Studentized residuals versus predicted values plot**.**

The studentized residuals versus predicted values plot, Figure 2, is a plot of the residuals versus the ascending predicted response values. It tests the assumption of constant variance. A random scatter (constant range of residuals across the graph.) is acceptable while an expanding variance ("megaphone pattern <") indicates the need for a transformation (Engle and Sheppard 2001). Since the plot showed a random scatter, it indicates that there was no assumption of a constant variance, thus, transformation is not required in the data so analyzed.

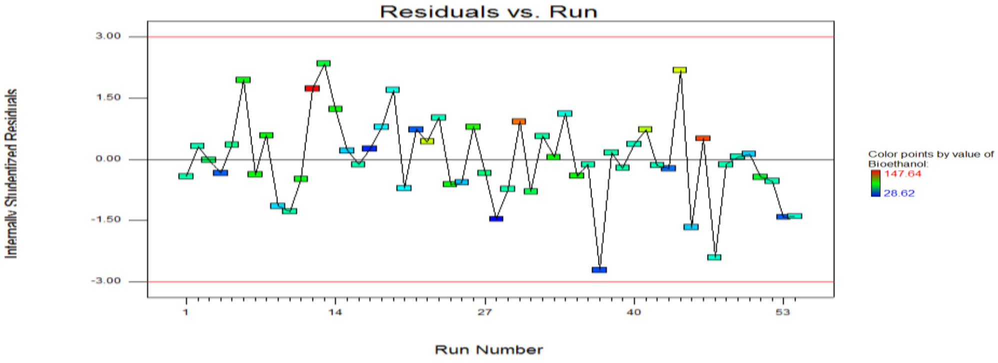

Figure 3:- Internally studentized residuals versus run number

The internally studentized residuals versus run plot, Figure 3, show that there were no lurking variables that may have influenced the bioethanol responses, since it followed a random scatter. According to Engle and Sheppard (2001), internally studentized residuals versus run plot which results in a trend will normally indicate the presence of lurking variables which may have influenced the responses and as such a transformation would be required.

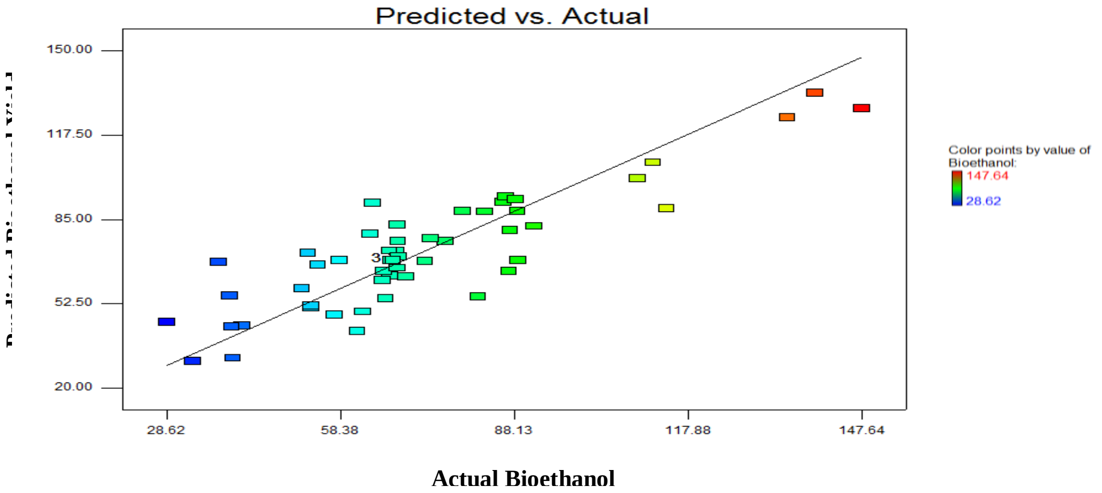

Figure 4:- Predicted bioethanol yield versus actual bioethanol yield plot.

The predicted bioethanol yield versus actual bioethanol yield plot, Figure 4, shows points clustering along a linear graph which indicates that there are no values that would not be easily predicted by the model. A predicted responnse versus actual response plot helps detect a value, or group of values, that are not easily predictable by a model. Thus, the clustering of the points around a line indicates parity between the predicted and actual data being analyzed; hence the model can easily be used to predict a value of a response when the actual value is known (Cao *et al.*, 2009).

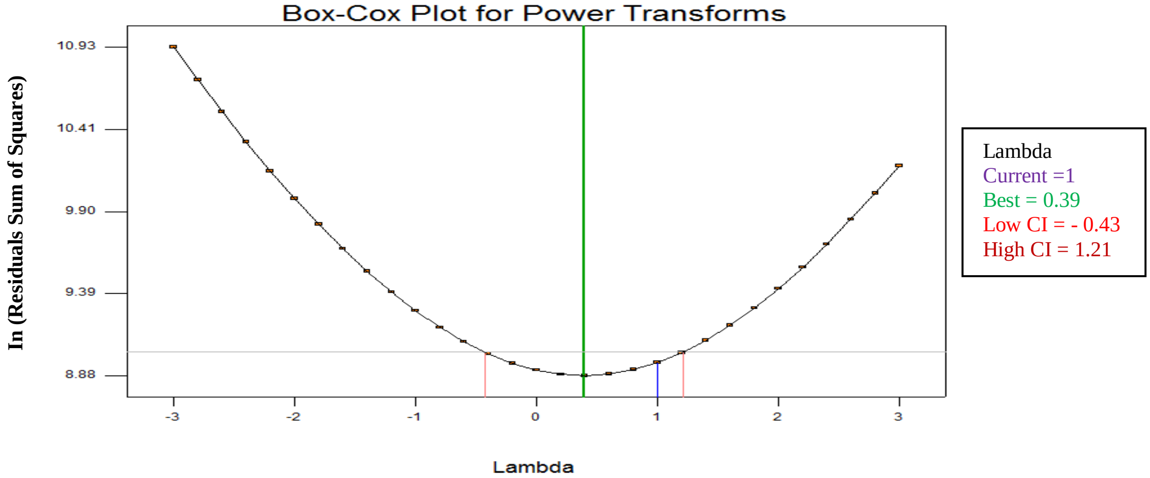

Figure 5:- Box-Cox plot for power transformations.

The Box Cox Plot, Figure 5, gave the best lambda value of 0.39 corresponding to the minimum point of the curve generated by the natural log of the sum of squares of the residuals. It was also observed that the 95% confidence interval around this best lambda value, includes 1, therefore no specific transformation is recommended for the data that was analyzed, based on the recommended transformation list in Box and Cox, (1964).

Conclusion:-

The Design Expert software and its several statistical tools are dedicated to designing experiments involving multivariable factors and analyzing the main effects of each factor as well as the interactions between factors. Thus in this paper it has been employed to determine statistical significance of the various process factors in bioethanol production from corn-stover, their impact on the response variable (bioethanol yield), diagnostic capacities to validate statistical assumptions, models and possible response transformations.

References:-

1. Black, S. (2013). "Cure for cratering: Gel coat perfected with design of experiments software," *CompositesWorld*.
2. Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society*, Series B, *26*, 211-252.
3. Cao, G., Ren, N., Wang, A., Lee, D. J., Guo, W., Liu, B., and Zhao, Q. (2009). Acid hydrolysis of corn stover for biohydrogen production using Thermoanaerobacterium thermosaccharolyticum W16. *international journal of hydrogen energy*, *34*(17), 7182-7188.
4. Comley, J. (2009). "Design Of Experiments: useful statistical tool in assay development or vendor disconnect!", *Drug Discovery World*, Winter.
5. Dasgupta, D., Suman, S.K., Pandey, D., Ghosh, D., Khan, R.M., Agrawal, D.M., Jain, R.K., Vadde, V.T., and Adhikari, D.K. (2013). Design and optimization of ethanol production from bagasse pith hydrolysate by a thermotolerant yeast *Kluyveromyces* sp. IIPE453 using response surface methodology. *Springer Plus*, *2*:159
6. Engle, R. F., and Sheppard, K. (2001). *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH* (No. w8554). National Bureau of Economic Research.
7. Hothorn, T., and Everitt, B. S. (2014). *A handbook of statistical analyses using R*. CRC press.
8. Montgomery, D.C. (2005). Design and Analysis of Experiments: Response surface method and designs. *New Jersey: John Wiley and Sons, Inc.*
9. Nuran, B. (2007). Response Surface Methodology; *M.Sc. Project, Department of Mathematical Sciences, Unpublished, Indiana University of South Bend.*
10. Ndirika, V.I.O. and Ikejiofor, M.C. (2015)."Mathematical modeling and optimization of post-harvest processes". Paper Presented at a Three day Training Workshop on Basic Mathematical Modeling, *Organized by the National Centre for Equipment Maintenance and Development (NCEMD), University of Nigeria, Nsukka.*
11. Plant, A. (2013). "Be Sure to Check the Residuals When Analyzing the Outcome of your Designed Experiments," *SQT blog*.
12. Tanco, M., Viles, E., Ilzarbe, L. and Jesus Alvarez, M. (2008). "Dissecting DoE Software," *Six Sigma Forum Magazine*.