EMPIRICAL MODELING AND OPTIMIZATION OF CLAY-CATALYSED ESTERIFICATION USING RESPONSE SURFACE METHODOLOGY

Kamen, F.L.[*]
Igbokwe, P.K[**]
Nwakaudu, M.S[*]
Ndukwe, O.C.[*]

Abstract

The clay-catalysed esterification reaction between acetic acid and ethanol was modeled using Box-Behnken Response Surface Methodology (RSM) design. The reaction was catalysed by acid-treated montmorillonite clay from Udi stream in Enugu State of Nigeria. The input variables modeled were the catalyst weight, reaction temperature and reactants mole ratio while the output variable was the percentage conversion of acetic acid. The RSM result showed that the temperature of the reaction had the highest effect with coefficient of 15.611, followed by the reactants mole ratio with coefficient of 9.895 and then the catalyst weight with coefficient of 3.674. The RSM model was statistically tested and found to be adequate and accurate in describing the reaction in the domain of the specified variables. Optimization of the model was achieved using the Constrained Optimizer of MATLAB software. The model optimization predicated maximum conversion of 88.12% while the experimental validation obtained maximum conversion of 86.11%. These are in good agreement, thus validating the model.

Key Words: Clay catalyst, esterification, empirical modeling, optimization.

[*] Department of Chemical Engineering, Federal University of Technology, Owerri, Imo State, Nigeria.

[**] Department of Chemical Engineering, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

Introduction

The search for a cleaner process is one of the major challenges in modern chemical industries; however, the versatility, low cost and gain in yield and/or selectivity render clays very attractive catalysts from the green chemistry point of view. They are reusable and present minimal environmental impact. Within this context, the use of clays and clay minerals in synthetic organic chemistry has increased enormously in the last years (Lazslo, 1990). Different clay types abound in Africa and the world today. In Nigeria, areas with major clay mineral deposits include Ukpor, Udi and Ahoko, etc found in Anambra, Enugu and Kogi States respectively. These clays have different properties which make some more reactive than the others (Nwajagu, 2003). Acid-treated clays from Udi and Ukpor from Nigeria have been shown to catalyse some organic reactions like dehydrogenation and esterification (Igbokwe, et al, 2008, 2009; Olebunne, et al, 2011). Response Surface Methodology is a collection of statistical techniques for designing experiments, building models, evaluating the effects of factors and searching for the optimum condition by varying several factors simultaneously. The multivariate approach reduces the number of experiments, improves interpretation and evaluates the significance of several factors even in the presence of complex interactions (Shrivastava et al, 2008) The main objective of the RSM is to find a desirable location in the design space which could be a maximum, minimum or an area where the response is stable over a range of factors. It is a mathematical tool model that can represent curvature unlike the linear models and a tool for understanding the quantitative relationship between multiple input variables and one output variable. The two most common designs used in response surface modeling are Central Composite designs and Box-Behnken designs. In these designs, the inputs take three or five distinct levels but not all combinations of these levels appear in the design. These statistical designs provide empirical models that adequately predict the response within the design space and obtain unambiguous results with the least expense (Mathworks, 2004; Edrissi et al, 2008).Optimization techniques are very important in industrial planning, resource allocation and laboratory processes. The classic optimization is done by varying one process parameter at a time while keeping the others constant. When multiple variables are involved, it becomes difficult to study the system using the common approach of varying one factor at a time while holding the others constant. Statistical designs such as Box-Behnken design consider all the factors simultaneously and hence provide the

evaluation of all the effects at once. These designs are regarded as the most favourable techniques for obtaining unambiguous results with least expense (Edrissi et al, 2007). From literature, lots of works have been done in the area of kinetics and mechanistic modeling of esterification reactions. The results from these researches provide the necessary kinetic parameters for designing suitable reactors and efficient operation of the process. Empirical modeling of the reaction is equally important because it produces models that will adequately predict the response within the design space. (Mathworks, 2004)

Materials and Method

The clay sample obtained from Udi in Enugu State of Nigeria which has been characterized to be montmorilllonite (Okafor and Nwajagu, 1989) was crushed and screened with micrometer sieve. It was then activated by treating with 1MH2SO4 to form slurry which was heated in oven at 373K for 4hours. The dried acidified clay was pulverized and stored in air-tight container. The activated clay sample was used to catalyse liquid-phase esterification reaction. A typical reaction was run by pipetting 2.5ml of acetic acid into a 10ml stoppered bottle; 0.15g of the clay catalyst was added to it first before 2.5ml of ethanol was pipetted into it to ensure that the active sites of the catalyst were not blocked by the alcohol. The container was tightly closed, the contents shaken vigorously and immersed in a water bath maintained at temperature of 323K for 6hours after which the content was titrated with 1MNaOH.

. The summary of the reaction equation is

$CH_3COOH + C_2H_5OH \rightleftharpoons CH_3COOC_2H_5 + H_2O$ ……(1)

Acetic acid Ethanol Ethyl acetate Water

The Box-Behnken response surface design matrix was constructed and the experiments run accordingly. The response, Y which are the percentage conversions were calculated and tabulated. The natural and the coded values of the variables are shown on Table 1 and the results are shown on Table 2. The matrix of the experimental plan is orthogonal hence the coefficients of the RSM model were obtained with equation 3.

Table 1: Natural and Coded Values of the Independent Variables

VARIABLES	NATURAL VALUES			CODED VALUES		
	Low level	Mid point	High level	Low level	Mid point	High point
Catalyst	0.2	0.3	0.4	-1	0	+1

wt;(g), x_1						
Reaction temp;(K), x_2	323	343	363	-1	0	+1
Alcohol/Acid mole ratio, x_3	1	3	5	-1	0	+1

For the three factor inputs of X_1, X_2 and X_3 representing the catalyst weight, reaction temperature and reactant mole ratio, the equation of the quadratic response is given by:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{11}X_1^2 + b_{22}X_2^2 + b_{33}X_3^2 \quad .....(2)$$

$$b = (F^TF)^{-1}F^TY = CF^TY \quad .....(3)$$

Where F is response surface matrix plan. The MATLAB program was written and run for Equation 3. The coefficients of the model obtained for the linear, interaction and the second-order terms are shown on Table 3 while the model equation obtained is given by Equation 17.

Statistical Analysis of the Response Surface Model

Some standard statistical analyses were performed on the model obtained to ensure its acceptability. These include: test for the reproducibility of the data, test for the significance of the coefficients of the model, test for the adequacy of the model and test for accuracy of the model.

i) Reproducibility of data: This was checked using the variance due to the experiment given by Equation 4 and another factor, G given by Equation 5.

$$S_u^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(y_{uk} - \bar{y}_u\right)^2 \quad .....(4)$$

$$G = \frac{\max S_u^2}{\sum S_u^2} \quad .....(5)$$

If the factor, G obtained for the experiment is less than that obtained from Cochrain's distribution table at alpha level of 0.05, degrees of freedom n-1and N(n-1), the data are said to be reproducible, if otherwise, they are not.

Significance of the model's coefficients: To check for the significance of the model's coefficients, the variance due to error was first evaluated with additional experiments at the centre of the RSM design where the independent variables are at the mid-point (0 0 0) with Equation 6 and 7 given below.

$$S_E^{\,2} = \frac{1}{n-1}\sum_{g=1}^{n}\left(yug - \overline{yg}\right)^2 \quad \ldots\ldots(6)$$

$$\overline{yg} = \frac{1}{n}\sum_{g=1}^{n} yug \quad \ldots\ldots(7)$$

Where n is the number experiment replication at the centre. Next, the information matrix of Box-Behnken RSM design given by (F^TF) was inverted to give another matrix $C = (F^TF)^{-1}$ shown below.

Table 2: The Matrix obtained from the Inversion of $C = (F^TF)^{-1}$

A	0	0	0	0	0	0	p	p	p
0	e	0	0	0	0	0	0	0	0
0	0	e	0	0	0	0	0	0	0
0	0	0	e	0	0	0	0	0	0
0	0	0	0	g	0	0	0	0	0
0	0	0	0	0	g	0	0	0	0
0	0	0	0	0	0	g	0	0	0
P	0	0	0	0	0	0	c	d	d
P	0	0	0	0	0	0	d	c	d
P	0	0	0	0	0	0	d	d	c

Next, a constant, D was evaluated for the linear, interaction and the second-order terms of the RSM model using the diagonal values of the inverted matrix shown as Table 2.

For linear terms,

$$D_{Li} = \sqrt{e} \times S_E^{\,2} \times t_{(\alpha,\, n-1)} \quad \ldots\ldots (8)$$

Where $t_{(\alpha, n-1)}$ is the Students' test distribution at the level of significance, α and n-1 degree of freedom. All the coefficients of the linear terms whose absolute values are greater or equal to D_{Li} are significant.

For the interaction terms:

$$D_{In} = \sqrt{g} \times S_E^2 \times t_{(\alpha, n-1)} \quad \text{..... (9)}$$

All the coefficients of the interaction terms whose absolute values are greater or equal to D_{In} are significant.

For the second-order terms:

$$D_{Se} = \sqrt{c} \times S_E^2 \times t_{(\alpha, n-1)} \quad \text{..... (10)}$$

All the coefficients of the second-order terms whose absolute values are greater or equal to D_{Se} are significant.

ii) Adequacy of the model:

The adequacy of the model got after the elimination of the insignificant coefficients was checked, first, by forming another matrix, F, containing only the significant coefficients. The responses obtained from the model, Y_{mod} were evaluated using the matrix equation given below.

$$Y_{mod} = (F * b^1) \quad \text{..... (11)}$$

Next, the residual variance associated with the model was obtained using Equation 12.

$$S_{res}^2 = \frac{1}{N - (k+1)} \sum_1^N \left(y_{\exp t} - \overline{y_{\mathrm{mod}}} \right)^2 \quad \text{.....(12)}$$

Where N is the total number of the experiments and k is the number of the significant coefficients. The ratio of the residual variance to the error variance was evaluated and checked against the value obtained from the Fisher's distribution table at alpha level of significance and the degrees of freedom of residual variance and the error variance respectively.

$$F = \frac{S_{res}^2}{S_E^2} \quad \text{.....(13)}$$

If the F_{expt} is less than F_{table}, the model is adequate.

iii) Accuracy of the model:

Accuracy of the model was tested using the correlation coefficient, R which is the measure of the accuracy of the approximation of the model.

$$R = \sqrt{1 - \frac{Q_{res}}{V_{res}}} \quad \text{.....(14)}$$

$$Q_{res} = \sum_{1}^{N} \left(y_{\exp t} - \overline{y_{\mathrm{mod}}} \right)^2 \quad \text{..... (15)}$$

$$V_{res} = \frac{1}{N - (\text{illegible} + 1)} \quad \text{..... (16)}$$

The closer R is to 1.0, the more accurate the model describes the real situation.

The Graphical User Interface (GUI) response surface tool *(rstool)* of MATLAB was used in the modeling to give intuitive visualization of the response surface. The final model equation given by Equation 18 was optimized using the non-linear constrained optimization tool (Constrained Optimizer) of MATLAB software. The results are shown on Table 8. The equation and the optimization results were finally validated by performing experiment with the input variables values obtained for the optimized model which were converted to their natural values as shown on Table12. The validating experimental results obtained after six hours are shown on Table 14.

$$Y = 63.478 + 3.674X_1 + 15.611X_2 + 9.895X_3 - 0.87X_1X_2 - 2.275X_1X_3 - 0.326X_2X_3 + 0.506X_1^2 - 3.539X_2^2 - 3.877X_3^2 \quad \text{.....(17)}$$

Table 3: Result of the Box-Behnken Response Surface Design

N0	X_0	X_1	X_2	X_3	X_1x_2	X_1x_3	X_2x_3	X_1^2	X_2^2	X_3^2	Y_1	Y_2	Y	Su^2
1	1	-1	-1	0	1	0	0	1	1	0	42.00	42.50	42.25	0.13
2	1	-1	1	0	-1	0	0	1	1	0	74.40	73.75	74.08	0.21
3	1	1	-1	0	-1	0	0	1	1	0	48.20	48.90	48.55	0.25
4	1	1	1	0	1	0	0	1	1	0	67.00	76.80	76.90	0.02
5	1	-1	0	-1	0	1	0	1	0	1	43.50	42.42	42.96	0.58
6	1	-1	0	1	0	-1	0	1	0	1	67.4	66.84	67.12	0.16
7	1	1	0	-1	0	-1	0	1	0	1	59.00	58.20	58.60	0.32
8	1	1	0	1	0	1	0	1	0	1	71.20	72.30	71.75	0.61
9	1	0	-1	-1	0	0	1	0	1	1	28.80	29.40	29.10	0.18

10	1	0	-1	1	0	0	-1	0	1	1	50.14	51.20	50.67	0.56
11	1	0	1	-1	0	0	-1	0	1	1	62.44	61.75	62.10	0.24
12	1	0	1	1	0	0	1	0	1	1	82.85	81.70	82.38	0.45
13	1	0	0	0	0	0	0	0	0	0	64.34	62.97	63.65	0.91
14	1	0	0	0	0	0	0	0	0	0	63.94	63.60	63.77	0.06
15	1	0	0	0	0	0	0	0	0	0	62.85	61.72	62.28	0.64
16	1	0	0	0	0	0	0	0	0	0	64.60	64.50	64.55	0.01
17	1	0	0	0	0	0	0	0	0	0	63.50	64.86	64.18	0.93
18	1	0	0	0	0	0	0	0	0	0	61.96	62.92	62.44	0.46

Table 4: Coefficients of the RSM Model obtained with MATLAB Software

LINEAR TERMS				INTERACTION TERMS			SECOND-ORDER TERMS		
b_0	b_1	b_2	b_3	b_1b_2	b_1b_3	b_2b_3	b_1^2	b_2^2	b_3^2
63.478	3.674	15.611	9.895	-0.87	-2.275	-0.326	0.506	-3.595	-3.877
$D_{Li} = 0.78$				$D_{In} = 1.10$			$D_{Se} = 1.05$		

Table 5: Diagonal Values of the Inverted B-B's Information Matrix

a	c	d	e	g	p
0.1667	0.2292	-0.0208	0.1250	0.2500	-0.833

$$Y = 63.478 + 3.674X_1 + 15.611X_2 + 9.895X_3 - 2.275X_1X_3 - 3.539X_2^2 - 3.877X_3^2 \quad \text{.....(4.2)}$$

Table 6: Matrix of the Significant Coefficients

N0	X_0	X_1	X_2	X_3	X_1x_3	X_2^2	X_3^2	Y_{expt}	Y_{mod}	$(Y_{expt} - Y_{mod})^2$
1	1	-1	-1	0	0	1	0	42.25	40.65	2.56
2	1	-1	1	0	0	1	0	74.08	71.88	4.84
3	1	1	-1	0	0	1	0	48.55	48.00	0.303
4	1	1	1	0	0	1	0	76.90	79.22	5.382
5	1	-1	0	-1	1	0	1	42.96	43.86	0.81

6	1	-1	0	1	-1	0	1	67.12	68.10	0.96
7	1	1	0	-1	-1	0	1	58.60	55.66	8.644
8	1	1	0	1	1	0	1	71.75	70.90	0.723
9	1	0	-1	-1	0	1	1	29.10	30.56	2.132
10	1	0	-1	1	0	1	1	50.67	50.35	0.102
11	1	0	1	-1	0	1	1	62.10	61.78	0.048
12	1	0	1	1	0	1	1	82.38	81.57	0.656
13	1	0	0	0	0	0	0	63.65	63.48	0.029
14	1	0	0	0	0	0	0	63.77	63.48	0.084
15	1	0	0	0	0	0	0	62.28	63.48	1.44
16	1	0	0	0	0	0	0	64.55	63.48	1.145
17	1	0	0	0	0	0	0	64.18	63.48	0.49
18	1	0	0	0	0	0	0	62.44	63.48	1.082

Table 7: Summary of the Results from the Statistical Analyses

PARAMETER TESTED	TEST USED	RESULT OBTAINED
Homogeneity of data	Cochrain's test (G-test)	$G_{expt.} = 0.138$ $G_{table} = 0.43$
Significance of equation coefficients	Students' test (t-test)	i) All coefficients of linear terms whose absolute values are greater than 0.78 are significant. ii) All coefficients of interaction terms whose absolute values are greater than 1.10 are significant. iii) All coefficients of 2nd-order terms whose absolute values are greater than 1.05 are significant
Adequacy of the model	Fisher's test (F-test)	$F_{expt.} = 3.68$ $F_{table} = 4.75$
Accuracy of the model	Correlation coefficient, R^2	$R^2 = 0.9895$

Check on the correlation coefficient, R	Fisher's test (F-test)	$F_R = 133.7$ $F_{table} = 3.14$

Table 8: Natural and Coded Values for Upper Limit Constraints for Optimization

Variables	X_1 (Catalyst Weight), g	X_2 (Temperature), ^oC	X_3 (Mole Ratio)
Natural Values	0.5	100	12
Coded Values	2.0	1.5	5.0

Table 9: Results of the Response Surface Model Optimization

INPUT VARIABLES			RESPONSE
X_1 (Catalyst Weight)	X_2 (Temperature)	X_3 (Mole Ratio)	Y (%Conversion)
2.0	1.5	0.689	88.123

Table 10: Results of the Experimental Validation of the RSM Model

OUTPUT VARIABLES, Y (% CONVERSION)			
Y_1	Y_2	Y_3	Y_{AV}
85.84	86.35	86.14	86.11

Discussion

Table 3 shows the result of the predicted conversions of the acetic acid replicated (Y_1, Y_2), the average conversion (Y) and the variance due to experiment (Su^2). The D-factor evaluated for the linear, interaction and second-order terms of the RSM model using the diagonal values of the inverted matrix of Box-Behnken's design are also shown on Table 4. Table 5 shows the values of the diagonal of the inverted Box-Behnken information matrix which was used to determine the significant coefficients using Equations 8 – 10. From the result, all the coefficients of the linear terms whose absolute values are greater than or equal to 0.78 are significant. For the interaction

terms, all coefficients whose absolute values are greater than or equal to 1.10 are significant. While for the second-order terms, coefficients with absolute values greater than or equal to 1.05 are significant. After eliminating the insignificant coefficients, the final model is given by Equation 18 which was used to generate matrix of the significant coefficients and the error between the experiment and the model ($Y_{expt} - Y_{mod}$) shown on Table 6. The response surface model showed that temperature had the highest effect on the reaction with coefficient of 15.611, followed by the reactant mole ratio having coefficient of 9.895 and then catalyst weight with the coefficient of 3.674.The adequacy of the model was tested using the Fishers distribution table and found to be adequate. The accuracy of the model was tested using the correlation coefficient, R^2 which was found to be 0.9895. This shows that the correlation between the model and experimental data is accurate up to 98.95%. The level of the accuracy of the model is acceptable. The correlation coefficient is a random quantity; hence a check was also conducted on it using the Fisher's distribution table to confirm that it is significantly different from zero. The summary of the results of the statistical analyses are shown on Table7. The result obtained experimentally is 86.11% conversion while that predicted by the optimized response surface model is 88.12% conversion. These are in good agreement because the correlation coefficient is 0.9895 which is saying that the accuracy of the prediction of the correlation between the experimental data and the RSM model is 98.95%. The difference between the model's accuracy predicted by the correlation coefficient and the experimental result is 1.23% which is acceptable. The response surface model and the optimization results were thus validated.

Conclusion

The empirical model of clay-catalysed esterification reaction was obtained in terms of three key parameters of the reaction: catalyst weight, reaction temperature and reactant mole ratio. The response surface model obtained with Box-Behnken's design was proven to be adequate and accurate statistically in describing the non-linear correlation between the input and the output variables of the reaction. The final empirical equation was optimized and validated experimentally, hence it is reliable for predicting the percentage conversion in the domain of the specified variables.

References

Edrissi, M.; Asl, N.R.; Madjidi, B.; (2008) 'Interaction of Mefenamic Acid with Cobalt11 Ions in Aquous Media' Turkish Chemistry Journal. Vol.32, Page 505-519.

Igbokwe, P.K; Ugonabo, V.I; Iwegbu, N.A; Akachukwu, P.C; Olisa, C.J. (2008) 'Kinetics of the Catalytic Esterification of Propanol with Ethanoic Acid using Catalyst obtained from Nigerian Clays' International Journal of Engineering Sciences, Ghana. Vol.1, No.2, Pp 132-137.

Igbokwe, P.K; Olebunne, F.L; Opebiyi, S.O; Uzondu, F.N. (2009) 'Effect of Silica gel Support on Clay-catalysed Vapour-phase Dehydrogenation of Ethanol' International Journal of Engineering Science, Ghana. Vol.1, No.2, Pp 132-137.

Laszlo P., (1990) 'Catalysis of Organic Reactions by Inorganic Solids', Pure & Applied Chemistry, Vol.62, (10), Pp 2027- 2030.
MathsWorks (2004) 'MATLAB: The Language of Technical Computing' MathWorks Inc. USA.

Nwajagu, C.O; Aneke, L.E. (1986) 'Characteristics of Ukpor Clay Deposits' Journal of Nigerian Society of Chemical Engineers' Vol.5 No.1.

Okafor B. F and Nwajagu C. O (1989); "Characterization of Udi clay Deposits" Journal of Nigerian Society of Chemical Engineers'(NSChE) Vol. 2 No. 1, pp 32-35.

Olebunne, F.L; Igbokwe, P.K; Onyelucheya, O.E.; Osoka, E.C; Ekeke, I.C. (2011) 'Mechanistic Modeling of Clay-Catalysed Liquid-phase Esterification Reaction' Journal of Emerging Trends in Engineering and Applied Sciences, UK. Volume 2, No.4, 2011, Pp 631-635.

Shrivastava, A; Saudagar, P; Singhal, R. (2008) 'Media Optimization for the Production of *y*-Linolenic Acid by *Cunninghamella echinulata var. elegans* MTCC 552 Using Response Methodology' International Journal of Food Engineering. Vol.4, Issue 2, Article 10.