OPTIMIZATION OF PROCESS VARIABLES USING RESPONSE SURFACE METHODOLOGY (RSM) FOR THE YIELD AND PHOSPHATIDYLCHOLINE ENRICHMENT (*PCE*) OF DEOILED RAPESEED LECITHIN WHEN FRACTIONATED WITH ETHANOL

Yisa, Yakubu
Mathematics and Statistics Department,
Federal University of Technology, Minna, Niger State, Nigeria
E-mail: yisa_yakubu@yahoo.com

Abstract

Response surface methodology is a collection of Statistical and Mathematical techniques useful for developing, improving, and optimizing processes. In this paper, the central composite design (CCD) method for fitting a second – order response surface model was used with the data extracted from Ipadeola (1990) as found in Illimiese (2008), to optimize the process variables (extraction time (t), solvent volume (v), ethanol concentration (c), and temperature (T)) and examine their effects on the yield and phosphatidylcholine enrichment (PCE) of deoiled rapeseed lecithin when fractionated with ethanol. The significance of the linear, quadratic and interaction terms were first examined and it was discovered that each of them significantly contributes to the response model at $\alpha = 0.05$ level, which implies that the fitted second-order model significantly explains the response surface. Then the analysis proceeds to locate the set of the levels of the factors (the stationary point) that yield optimum value of the response variable. The canonical analysis was carried out and it was detected that the stationary point is a saddlepoint.

Keywords: Response surface, second-order models, central composite design, yield, stationary point

Introduction

The statistical design of experiments approach to process development offers several key advantages over the traditional one-variable-at-a-time approach. Box et al., (2005) reported that "the practice of a single factor optimization by maintaining other factors involved at an unspecified constant level do not portray the combined effect of interactions of factors involved. The method is tedious, time consuming and expensive, especially for a large number of variables. Moreover it does not guarantee the determination of optimum conditions among the variables".

Response surface methodology (RSM) is a powerful mathematical model with a collection of statistical techniques where in, interactions between multiple process variables can be identified with fewer experimental trials (Myers *et al*, 2009). It is widely used to examine and optimize the operational variables for experiment designing, model developing and factors and conditions optimization. There are various advantages in using statistical methodologies in terms of rapid and reliable short listing of process conditions, understanding interactions among them and tremendous reduction in total number of experiments. The classical method of studying one variable at a time can be effective in some cases but it is useful to consider the combined effects of all the factors involved (Box, 1952, Box and Hunter, 1957). The Response Surface Methodology (RSM), based on statistical principles, can be employed as an interesting strategy to implement process conditions that drive to optimal yield of deoiled rapeseed by performing a minimum number of experiments.

Response Surface Methodology (RSM) is a set of techniques that includes setting up a series of experiments that will yield adequate and reliable measurements of the response of interest, determine a model that best fits the data collected from the design chosen and determine the optimal settings of the experimental factors that produce the maximum (or minimum) value of the

response (Montgomery, 2001).These designs provide information on direct effects, pair wise interaction effects and curvilinear variable effects.
Response Surface Design allows for evaluation of statistical significance of fitted mathematical models, the contribution of individual process parameters, as well as that of the interaction between factors, which is not possible using the one-variable-at-a-time approach. The mathematical models can then be utilized to find the predicted optimum system response within the experimental region of the study. The optimized set of conditions can then be verified experimentally to validate the model prediction.

In practice, the form of the relationship is unknown but can be approximated, within the experimental region, by a low degree polynomial model of the form

$$\boldsymbol{y} \quad \boldsymbol{X\beta} \qquad .$$

y is an (*n* X 1) vector of observations, *X* is an (*n* X *k*) matrix of levels of independent variables, *ß* is a (*k* X 1) vector of regression coefficients, and *e* is an (*n* X 1) vector ofrandom errors (Montgomery 2001).

If *X* is a (*k* X *k*) matrix, then the linear system $y = Xß + e$ has a unique *least squares* solution given by

$$\boldsymbol{\beta} \quad \boldsymbol{X}\,\boldsymbol{X} \quad \boldsymbol{X}\,\boldsymbol{y} \qquad .$$

The estimated regression equation is$\boldsymbol{y} \quad \boldsymbol{X\beta}$ □□lack-of-fit techniques are then applied in order to check whether the model is appropriate.
In this work, optimization of process conditions (extraction time, solvent volume, ethanol concentration, and temperature) using RSM for the yield and phosphatidylcholine enrichment (*PCE*) of deoiled rapeseed lecithin when fractionated with ethanol have been carried out and the influence of these variables was well studied using CCD experiments.

Materials and Methods

In this paper, the central composite design (CCD) method for fitting a second –order model was used and illustrated with the data extracted from Ipadeola (1990) as found in Illimiese (2008), to examine the effects of extraction time (*t*), solvent volume (*V*), ethanol concentration (*C*), and temperature (*T*) on the yield and phosphatidylcholine enrichment (*PCE*) of deoiled rapeseed lecithin when fractionated with ethanol. First –order model have earlier been illustrated on the data (Audu et al., 2009).

Central composite design (CCD) as the most popular methodfor fitting a second–order model was introduced by Box and Wilson (1951). It consists of factorial points (from a 2^k *factorial* design or a 2^{k-p} *fractional factorial design*), central points, and axial points.

A uniform-precision 2 ($k = 4$) factorial central composite experimental design with eight star points, one center point, resulting in a total of 25 experimental runs was used to optimize the chosen key variable yield of deoiled rapeseed lecithin.

Table 2.1 shows the four independent variables (extraction time, solvent volume, ethanol concentration, and temperature) and their concentrations at different coded and actual levels of the variables employed in the design matrix. The natural variables (ξ_i) were transformed to coded variables (x_i) using the relation

$$x_{iu} = \frac{\left(\xi_{iu} - \overline{\xi}_i\right)}{S_i}, \ \textit{where} \ \ S_i = \left\{ \sum_{u=1}^{N} \frac{\left(\xi_{iu} - \overline{\xi}_i\right)^2}{N/c} \right\}^{1/2} \qquad (2.1)$$

This gives

$$\sum_{u=1}^{N} x_{iu} = 0 \quad \text{and} \quad \sum_{u=1}^{N} x_{iu}^2 = N \qquad (2.2)$$

Table 2.1: Codes and actual levels of the independent variables

Independent variable	Coded levels				
	-1.414	-1	0	1	1.414
Extraction time (t)	2.93	5	10	15	17.07
Solvent volume (v)	3.965	5	7.5	10	11.035
Ethanol concentration (c)	90.758	92	95	98	99.242
Temperature (T)	12.93	15	20	25	27.07

Using these coded values for the natural variables, the proposed second –order model is of the form:

$$y_u = \beta_0 + \sum_{i=1}^{k} \beta_i x_{iu} + \sum_{i=1}^{k} \beta_{ii} x_{iu}^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_{iu} x_{ju} + e_u \qquad (2.3)$$

Where y_u (u = 1, 2, ...,N) represents the *u*th response value (yieldof deoiled rapeseed lecithin) obtained as a result of applying the *u*th design setting (or *u*th treatment combination), x_{iu} is the level of the *i*th factor in the *u*th treatment combination, (*i = 1, 2, ..., t*), β_i is the *i*th factor coefficient, e_u is the random error associated with the *u*th observation that is independently and normally distributed with mean zero and common variance σ^2.

Analysis of the second-order model

The second-order model is flexible, because it can take a variety of functional forms and approximates the response surface locally. Therefore, this model is usually a good estimation of the true response surface. Besides, the method of least squares can be applied to estimate the coefficients β_j in the model.TheANOVA and regression analysis for the response variable *Yield,* as generated by the Minitab14 software, are shown in Tables 3.1and 3.2below.

The estimated regression equation is

$$\begin{aligned} yield = {} & 21.46 + 1.34A + 2.67B + 2.13C + 1.28D + 0.41A^2 - 1.59B^2 - 1.54C^2 - 0.94D^2 \\ & + 0.775AB + 0.275AC + 0.150AD + 0.625BC + 0.500BD - 0.100CD \end{aligned} \qquad (3.1)$$

Response Surface Regression: yield versus A, B, C, D

Table 3.1: Estimated Regression Coefficients for Yield

Model Term	Regression Coefficient	Standard Error(Coef.)	*t*-statistics	*P*-value
Constant	21.4632	0.4338	49.480	0.000
A	1.3380	0.1617	8.275	0.000

B	2.6706	0.1617	16.516	0.000
C	2.1336	0.1617	13.195	0.000
D	1.2805	0.1617	7.919	0.000
A*A	0.4106	0.2557	1.606	0.139
B*B	-1.5900	0.2557	-6.218	0.000
C*C	-1.5400	0.2557	-6.022	0.000
D*D	-0.9398	0.2557	-3.675	0.004
A*B	0.7750	0.1808	4.287	0.002
A*C	0.2750	0.1808	1.521	0.159
A*D	0.1500	0.1808	0.830	0.426
B*C	0.6250	0.1808	3.457	0.006
B*D	0.5000	0.1808	2.766	0.020
C*D	-0.1000	0.1808	-0.553	0.592
S	0.7231			
R-Sq	98.6%			
R-Sq(adj)	96.7%			

Table 3.2: ANOVA Table

Source	DF	SS	MS	F	P
Regression	14	371.469	26.5335	50.74	0.000
Linear	4	302.270	75.5675	144.52	0.000
Square	4	47.609	11.9022	22.76	0.000
Interaction	6	21.590	3.5983	6.88	0.004
Residual Error	10	5.229	0.5229		
Total	24	376.698			

As can be observed from table3.1,the regression coefficients of the linear terms *A, B, C,* and *D* are significant at both 0.01 and 0.05 levels. The quadratic terms, B^2, C^2, D^2 and interaction terms *AB,BC* and *BD* significantly contribute to the response model at α = 0.05.This table also shows a high value (98.6%) of R^2 (coefficient of determination), which indicates that the yield and phosphatidylcholine enrichment (*PCE*) of deoiled rapeseed lecithin fitted the second- order polynomial equation well.

The Analysis of Variance table (Table 3.2) summarizes the linear terms, the squared terms, and the interactions.This table demonstrated that the model is significant due to the small p-value. The linear, square and interaction effects in the model are all significant due to the exhibited small p-values. This indicate that the contributions of these effects to the model are significant. Since there are no replicated center points, thesoftware cannot obtain a lack-of- fit. But, small *p*-values for the interactions and thesquared terms suggest there is curvature in the response surface.

In addition, the package draws four residual plots (Figure 3.1) - *Histogram of residuals*, which is an exploratory tool to show general characteristics of the data, *Normal plot of residuals*, to show if the data obey the normality assumption, *Residuals versus fits,* which shows a random pattern of residuals on both sides of 0, and *Residuals versus order,* which is a plot of all residuals in the order that the data were collected.

We can see that the residual plots do not indicate any problems with the model.
As a result, the final model for the response variable *Yield,* based on these significant terms, is given as:

$$\text{Yield} = 21.5 + 1.34A + 2.67B + 2.13C + 1.28D - 1.59B^2 - 1.54C^2 - 0.94D^2 + 0.77AB + 0.62BC + 0.50BD \quad (3.2)$$

Locating the Stationary Point

The second-order models illustrate quadratic surfaces such as minimum, maximum, ridge, and saddle. If there exits an optimum then this point is a stationary point. The stationary point is the combination of design variables where the surface is at either a maximum or a minimum in all directions. If the stationary point is a maximum in some direction and minimum in another direction, then the stationary point is a saddlepoint. When the surface is curved in one direction but is fairly constant in another direction, then this type of surface is called ridge system (Montgomery, 2001). The stationary point can be found by using matrix algebra. The fitted second-order model of Eq. (2.3) above can be expressed in matrix form as follows:

$$\hat{\mathbf{y}} = \hat{\boldsymbol{\beta}}_0 + \mathbf{x'b} + \mathbf{x'Bx} \quad (4.1)$$

where

$$\mathrm{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_k \end{bmatrix} \quad \mathrm{b} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ . \\ . \\ . \\ \hat{\beta}_k \end{bmatrix} \quad \text{and} \quad \mathrm{B} = \begin{bmatrix} \hat{\beta}_{11}, & \hat{\beta}_{12}/2, & . & . & ., & \hat{\beta}_{1k}/2 \\ & \hat{\beta}_{22}, & . & . & ., & \hat{\beta}_{2k}/2 \\ & & . & & & \\ & & & . & & \\ & & & & . & \\ \text{sym.} & & & & & \hat{\beta}_{kk} \end{bmatrix}$$

That is, $\mathbf{b}$ is a (kx1) vector of the first order regression coefficients and $\mathbf{B}$ is a (kxk)symmetric matrix whose main diagonal elements are the *pure* quadratic coefficients (β_{ii}) and whose off - diagonal elements are one-half the *mixed* quadratic coefficients ($\hat{\beta}_{ij}, i \neq j$) (Montgomery, 2001).

The derivative of $\hat{\mathbf{y}}$ with respect to the elements of the vector x equated to zero is

$$\frac{D\hat{y}^2}{D\mathrm{x}} = \mathrm{b} + 2\mathrm{Bx} = 0 \quad (4.2)$$

The stationary point is the solution to equation (4.2), that is

$$\mathrm{x}_0 = -\tfrac{1}{2}\mathrm{B}^{-1}\mathrm{b} \quad (4.3)$$

By substituting equation (4.3) into equation (4.1), we can find the predicted response at the stationary point as

$$\hat{y}_0 = \hat{\beta}_0 + \tfrac{1}{2}\mathrm{x}_0^{'}\mathrm{b} \quad (4.4)$$

Using Excel package, the calculations for locating the stationary point for the response *Yield* are as follows.

Now from the fitted regression equationabove, we have our B in equation (4.2) to be:

$$\mathbf{B} = \begin{bmatrix} 0.41 & 0.387 & 0.137 & 0.075 \\ 0.387 & -1.59 & 0.312 & 0.25 \\ 0.137 & 0.312 & -1.54 & -0.05 \end{bmatrix}$$

0.075	0.25	-0.05	-0.94

The inverse of B is B^{-1}, given as

B^{-1} =

1.800809	0.531824	0.259138	0.27134
0.531824	-0.52531	-0.05605	-0.0943
0.259138	-0.05605	-0.63894	0.039754
0.27134	-0.0943	0.039754	-1.06937

The *kx1* vector of the first order regression coefficients, $\mathbf{b}$, is given by $\mathbf{b} = (X'X)^{-1}X'y$. That is,

b=

1.338041
2.670601
2.133629
1.280457

And from equation (4.3), the stationary point is given by $x_0 = -\frac{1}{2}B^{-1}b$ .Thatis,

$$x_0 = -\frac{1}{2}\begin{bmatrix} 1.800809 & 0.531824 & 0.259138 & 0.27134 \\ 0.531824 & -0.52531 & -0.05605 & -0.0943 \\ 0.259138 & -0.05605 & -0.63894 & 0.039754 \\ 0.27134 & -0.0943 & 0.039754 & -1.06937 \end{bmatrix}\begin{bmatrix} 1.338041 \\ 2.670601 \\ 2.133629 \\ 1.280457 \end{bmatrix} = \begin{bmatrix} -2.36509 \\ 0.46582 \\ 0.557663 \\ 0.586615 \end{bmatrix}$$

That is, $x_{10} = -2.36509$, $x_{20} = 0.46582$, $x_{30} = 0.557663$, $x_{40} = 0.586615$, approximately.

In terms of the natural variables (time, volume, concentration and temperature), the stationary point is

$$-2.36509 = \frac{t-10}{5},\ 0.46582 = \frac{v-7.5}{2.5},\ 0.557663 = \frac{Con-95}{3},\ 0.586615 = \frac{T-20}{5}$$

which yields t = 1.82545 = 1.83 minutes of reaction time, V = 8.66455 = 8.66 liter solvent volume, *Conc* = 96.67299 = 96.67 percent of ethanol concentration, and T = 22.93308 = 22.93 ºC temperature.

We can see that the stationary point is within the region of exploration for fitting the second-order model.

The predicted response at the stationary point is given by equation (4.4) as $\hat{y}_0 = \hat{\beta}_0 + \frac{1}{2}x_0'b$. That is,

$$\hat{y}_{yield} = 21.46 + \tfrac{1}{2}\begin{bmatrix} -2.36509 & 0.46582 & 0.557663 & 0.586615 \end{bmatrix}\begin{bmatrix} 1.338041 \\ 2.670601 \\ 2.133629 \\ 1.280457 \end{bmatrix}$$

$$= 21.46 + 0.010204$$

$$\therefore \quad \hat{y}_{yield} = 21.470207 \approx 21.47.$$

Therefore, the predicted response at the stationary point is $\hat{y}_{yield} = 21.47$.

Characterizing the Response Surface

The simplest way to characterize a response surface is to construct a contour plot of the response as a function of a pair of the variables. Figure 3.2 below gives the two-dimensional contour plot of the yield as a function of a pair of each of the four input variables. Since we have more than three process variables, the interpretation of the contour plot is a little bit complicated here. However, it is clear from examining figure 3.2 that each of the main factors is related to the response variable *Yield* at their high levels.

At this point we adopt a more formal analysis to determine whether the stationary point above is a point of maximum or minimum response or a saddle point. Transformingthe model into a new coordinate system with the origin at the stationary point x_0, we have the fitted model

$$\hat{y} = \hat{y}_0 + \lambda_1\omega_1^2 + \lambda_2\omega_2^2 + \lambda_3\omega_3^2 + \lambda_4\omega_4^2 \qquad (5.1)$$

called the *canonicalform* of the model. Where the $\{\omega_i\}$ are the transformed independent variables and the $\{\lambda_i\}$ are constants, which are just the *eigenvalues* or *characteristic roots* of the matrix B (Montgomery, 2001).

Now, the eigenvalues λ_1, λ_2, λ_3, and λ_4 are the roots of the determinantal equation

$$|B - \lambda I| = 0$$

That is,

$$\begin{vmatrix} 0.41-\lambda & 0.387 & 0.137 & 0.075 \\ 0.387 & -1.59-\lambda & 0.312 & 0.25 \\ 0.137 & 0.312 & -1.54-\lambda & -0.05 \\ 0.075 & 0.25 & -0.05 & -0.94-\lambda \end{vmatrix} = 0$$

Then by means of computer software (Mapple 13), this gives us

$$\lambda^4 + 3.66\lambda^3 + 3.385593\lambda^2 - 0.508999116\lambda - 1.176002089 = 0$$

The roots of this equation are

$$\lambda_1 = 0.5102911689,\ \lambda_2 = -0.8811205241,\ \lambda_3 = -1.345987115, \text{ and } \lambda_4 = -1.943183530$$

And the canonical form of the fitted model is

$$\hat{y} = 21.47 + 0.51029\omega_1^2 - 0.88112\omega_2^2 - 1.34599\omega_3^2 - 1.94318\omega_4^2$$

Since the $\{\lambda_i\}$ have different signs we conclude here that the stationary point x_0 is a saddle point.

Figure 3.1: Residual Plot of Purified Lecithin *Yield*

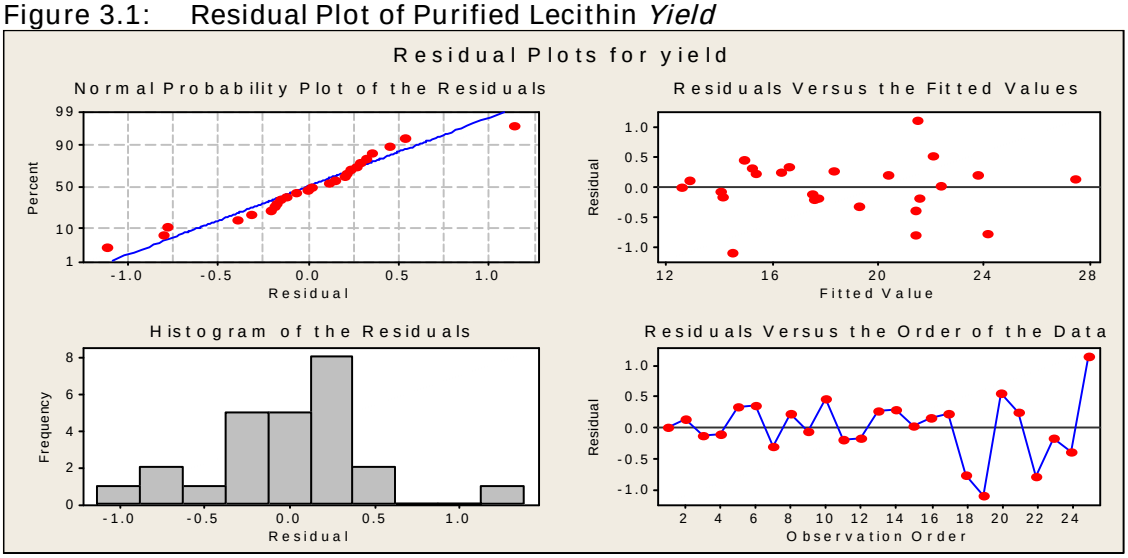
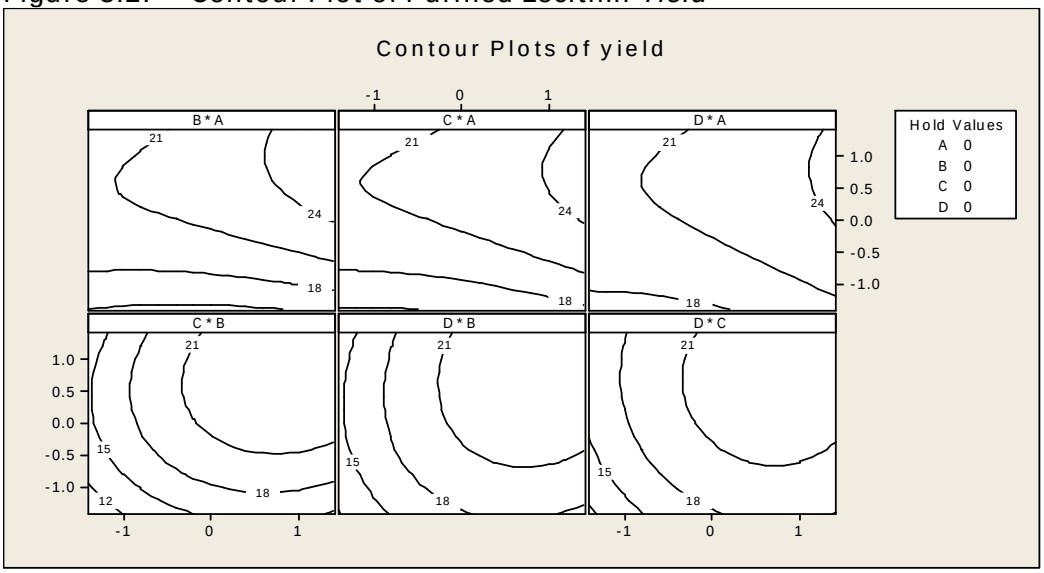


Figure 3.2: Contour Plot of Purified Lecithin *Yield*



Conclusion

Statistically designed experiments are highly efficient in that they give a fixed amount of information with much less effort than the classical one-variable at-a-time approach and many of them give additional information about interaction as a bonus. Response surface methods (RSM) provide statistically-validated predictive models that can then be manipulated for finding optimal process configurations. Second-order model describes quadratic surfaces, and this kind of surface can take many shapes. Therefore, response surface can represent maximum, minimum, ridge or saddle point.

Our analysis results show that each of the four main effects and the quadratic terms, B^2, C^2, D^2 and interaction terms *AB,BC* and *BD* significantly contribute to the response model at α = 0.05. The Analysis of Variance table indicates that there are significant interactions between the factors at 0.05 level of significance. The small *p*-values for the interactions and the squared terms suggest there is curvature in the response surface.

A Central Composite Design (CCD) was used for the optimization of process conditions. From the present study, it is evident that the use of statistical process condition optimization approach, response surface methodology has helped to locate the most significant conditions with minimum effort and time. In addition, it has also proved to be useful in increasing yield. Only 25 experiments were necessary and the obtained model was adequate (P < 0.001). By solving the regression equation, the optimum process conditions were determined.Itwas found that the estimated optimum response (Yield) of deoiled rapeseed lecithin is $\hat{Y} = 21.47$ at the stationary point (*t* = 1.82545 = 1.83 minutes of reaction time, *V* = 8.66455 = 8.66 liter solvent volume, *Conc* = 96.67299 = 96.67 percent of ethanol concentration, and *T* = 22.93308 = 22.93ºC temperature). This is the maximum yield obtained at the optimized process conditions.

Also, the residual plots drawn do not indicate any problems with the model. The two-dimensional contour plots of the yield as a function of pairs of the four input variables indicate that each of the main factors is related to the response variable *Yield* at their high levels. The located stationary point is within the region of exploration for fitting the second-order model. The canonical analysis performed shows that the located stationary point is a saddle point. Therefore RSM is a critical technology in developing new processes and optimizing their performance. The objectives of quality improvement, including reduction of variability and improved process and product performance, can often be accomplished directly using RSM.

References

Audu, I., Nafiu, L. A., & Ilimiese, P. K. (2009). Matrix Approach to the experimental design of response surface methodology. *Journal of Science, Education and Technology, 2, (1)*.

Anderson, Mark J. and Whitcomb, Patrick J. 2004. *Design solutions from concept through manufacture: Response surface methods for process optimization. desktop engineering.* http://www.deskeng.com/ (accessed Oct.16, 2011).

Box, G. E. P., Hunter, W. G. & Hunter, J. S. (2005), *Statistics for experimenters*. New York: John Wiley & Sons, Inc.

Box, G.E.P. & Wilson, K.B. (1951). On the experimental attainment of optimum conditions", *Journal of Royal Statistical Society*, *Series B, Vol. 31, Pp 1-45.*

Box, G.E.P. (1952). Multifactor designs of first order. *Biometrika*, 39, *49-57.*

Box, G.E.P. & Hunter, J. S. (1957). Multi-factor experimental designs for exploring response surfaces. *Annals of Mathematical Statistics*, *28, 195-241.*

Ilimiese, P. K. (2008). *Statistical analysis of the experimental design using response surface methodology.* Unpublished undergraduate project, Federal University of Technology, Minna, Nigeria.

Montgomery, D. C. (2001): *Design and Analysis of Experiments* (6th ed). New York: Wiley.

Raymond, H. M., Douglas, C. M. & Christine, M. A. (2009). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*(3rd edition). New York: John Wiley and Sons inc.

Shi, Y. & Weimer, P. J. (1992). Response surface analysis of the effects of pH and dilution rate on ruminococcus flavefaciens FD-1 in cellulose-fed continues culture*. American Society for Microbiology 58 (8), 2583-2591.*

Verseput, R. (2000). Digging into DOE: Selecting the right central composite design for response surface methodology applications. *Quality Digest*. http://www.qualitydigest.com/june01/html/doe.html (accessed Nov. 7, 2011).

APPENDIX: The Design Matrix

factor / run no.	*t*	*V*	*C*	*T*	*Yield*
1	-1	-1	-1	-1	12.6
2	1	-1	-1	-1	13
3	-1	1	-1	-1	14
4	1	1	-1	-1	17.4
5	-1	-1	1	-1	15.6
6	1	-1	1	-1	17
7	-1	1	1	-1	19
8	1	1	1	-1	24
9	-1	-1	-1	1	14
10	1	-1	-1	1	15.4
11	-1	1	-1	1	17.4
12	1	1	-1	1	21.4
13	-1	-1	1	1	16.6
14	1	-1	1	1	18.6
15	-1	1	1	1	22.4
16	1	1	1	1	27.6
17	-1.414	0	0	0	20.6
18	1.414	0	0	0	23.4
19	0	-1.414	0	0	13.4
20	0	1.414	0	0	22.6
21	0	0	-1.414	0	15.6
22	0	0	1.414	0	20.6
23	0	0	0	-1.414	17.6
24	0	0	0	1.414	21
25	0	0	0	0	22.6