IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 5, Ser. 2 (Sept. – October. 2023), PP 01-12
www.iosrjournals.org

Optimization Of Oil Production From Mango Seed Using Response Surface Methodology And Central Composite Design

Daniel Mwangi Muriithi[1], Dennis K. Muriithi[2], Fidelis Ngugi[3]and Eric Kibagendi Osoro[4]
[1,3,4] Department of Basic Sciences, Faculty of Science and Technology, Tharaka University, Kenya
[2]Department of Physical Sciences, Faculty of Science, Engineering and Technology, Chuka University, Kenya

Abstract
Response Surface Methodology (RSM) and Central Composite Design (CCD) is an important statistical tool for modeling and analysis of statistical problems where the response of interest is influenced by several variables with the objective of optimizing the response. The purpose of this study was to optimize oil extraction from mango seed using Response Surface Methodology. The experimental design for this study had three factors at five levels. Central Composite Design, is more useful methodology for modelling a second order model for a response variable in a full factorial design of experiments. This research optimized extraction of oil from mango seeds which are readily available and treated as a waste, and left to litter everywhere. n-hexane solvent and Soxhlex apparatus was used to extract oil from the powdered mango seeds. This may lead to reduced environmental pollution which has contributed immensely to climate change. The study sought to determine the effect of controlled variables (reaction Temperature, Solvent to Solute ratio and reaction Time) on the Yield of the Oil. The experimental data was analyzed using R-studio statistical software. The findings from the experimental data revealed that the optimum conditions for maximum oil yield (9.056%) obtained from Mango seed, are a temperature of 72.8 ^{0}C, seed powder to solvent ratio of 1:7 and a time of 138 minutes. The regression equation obtained for the model having a coefficient of correlation (R^2), and adjusted coefficient of correlation (R^2adj) are 0.9549 and 0.9143 respectively which shows the goodness of fit for the model. From the results optimization using CCD gave satisfactory results almost 82.33% of mango seed oil composition. A further research on determining edibility and quantitative composition of the mango seed oil are recommended.
***Key Word**: Oil Production; Central Composite Design; Mango Seeds Oil, Optimization*

Date of Submission: 27-09-2023 Date of Acceptance: 07-10-2023

I. Introduction

Once the mango fruits are consumed, only few seeds are used for propagation but most of them are left to decompose. The seed inside the fruit, represent 10 - 25% of the total fruit weight. The kernel inside the seed represents 45-75% of the total seed weight [4]. Depending on the variety, mango seeds contain 9–13% oil [7]. The oil has a high percentage of unsaturated fatty acids, such as oleic acid, at 46.22%, which is greater than palm oil's 15.4%, sunflower oil's 21.1%, and soybean oil's 23.4% as well as the oil from the Jatropha Curcas Plant, which has 44.7% oleic acid [3]. This shows that, mango is a good source of oil since it is non-edible and has no effect on food chain. While extracting oil from the mango seed, there was a need to get the optimum set of factors so that we get a maximum amount of oil. The study used RSM as a tool to achieve this optimum conditions. The goal of response surface methodology (RSM) is to optimize the response. It is a combination of mathematical and statistical techniques used for modelling and analysis of issues when a response of interest is influenced by numerous variables [6]. The CCD is the most commonly used in factorial design used in the RSM [2]. CCD is capable of accommodating up to five levels per every factor and can gather data from a properly planned factorial experiment. When the design permits sequential experimentation, the design is often used. A set of axial points known as star points are added to the center points in this design. Having this design, the first-order and second-order polynomial terms can be approximated. CCD is a design widely and successfully used for estimating second order response surfaces [1]. This study applied the CCD to layout the experimental runs with an aim to optimize the production of the oil from mango seeds. The control variables under study include; reaction time, reaction temperature, solute to solvent ratio and the response variable was Oil yield from mango seeds.

There are several studies on extraction of oil from various seeds and nuts among them Jatropha, Canola, Sunflower, macadamia nuts. Limited information is provided on optimization of the production of this oils specifically by use of RSM and CCD. In the study on Optimization and extraction of oil from mango seed kernel (*Mangifera Indica*), [9] investigated the influence of various parameters such as temperature, time, volume of n-hexane (solvent) and particle size on oil yield from mango seeds in India. The results revealed that oil yield increased with increase in temperature, time and volume of solvent but decreased with the increase in the particle size. Maximum oil yield of 15.20% was obtained at 1 mm size grinded kernel particles with an extraction time of 90 minutes using 25g of mango kernel seed and 250 mL of n-hexane. The study did not mention on any design of optimization.

The general objective of study was to optimize the oil production from mango seeds using Response Surface Methodology and Central Composite Design

The study was guided by the specific objectives given;

i. To investigate the appropriate mathematical model of second order polynomial that best fits the experimental data
ii. To optimize oil extraction from mango seed using Response Surface Methodology

II. Material and Methods

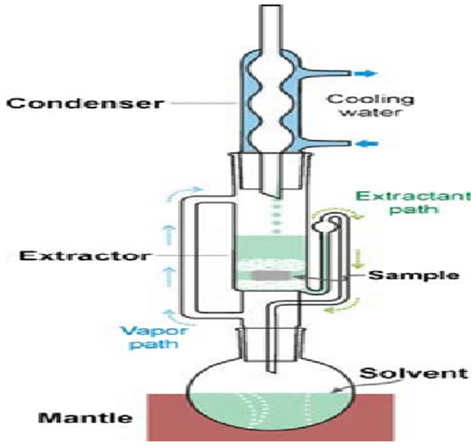
Mango seeds were obtained through physicl collection from market places. Hexane solvent was obtained through purchasing. Equipments such as soxhlet, rotary evaporator, thermometer, conical flasks, heating mantles weighing balance and measuring cyliders were obtained from Tharaka University Teaching and Research Laboratory Kenya.

Oil Extraction

The initial stage in the oil extraction was to collect the mango seeds from Gatunga Marimanti and Chuka market places, where they were in plenty. The outer seed coat was removed and the remaining seed dried, crushed and stored for oil extraction. Figure 1, represent the structure of a mango seed.

Figure 1: Structure of a Mango Seed

The study used the solvent extraction method, where a round bottom flask was filled with different amounts of *n*-hexane according to the experimental design. Then a 50g of the mango seed powder, was added into the extractor's center. The soxhlet was heated to different temperatures as outlined on the design. Once the solvent boils, the vapor rose through the vertical tube and into the condenser at the top. The condensate liquid dripped into the mango powder. The powder mango seed will be contained in a thimble at the center of the extractor. The extract trickled back down into the flat-bottom flask after passing through the thimble's pores and filling the siphon tube. This went on for a stipulated period of time after which the condenser was taken out at the conclusion of the extraction. The resulting mixture of oil and *n*-hexane, was placed in a rotary evaporator, with an aim of separating the oil and *n*-hexane. The solution in a round bottomed flask, on a rotary motion, was heated on water bath at a temperature of 70^0C. Within this time, the *n*-hexane evaporated off and was recovered in a separate round bottomed flask. The volume of the oil extracted was recorded. Figure 2 a shows the soxhlet apparatus and figure 2 b shows the rotary evaporator used to separate the oil from the solvent.



(a)

(b)

Figure 2: (a) Soxhlet Extractor and (b) Rotary Evaporator

The researcher applied RSM during solvent extraction process to establish the region of optimal production. During the solvent extraction process three controlled variables (P_1 reaction time, P_2 reaction temperature and P_3 amount of solvent used) were studied. Different levels of the controlled variables were varied with an aim of determining the region of optimum yield of the oil.

Three Factor Central Composite Design for oil Extraction

In this experiment a set of twenty experimental runs was carried out in a random manner where the runs included, 2^3 factorial design, six star points and six center points. This was repeated three times to assure reliability. During statistical calculations, the independent variables were coded using the formula $P_1 = \frac{(p_1-65)}{5}$ $P_2 = \frac{(P_2-120)}{30}$ $P_3 = \frac{(P_3-6)}{1}$

Where p_i was the coded variable of the i^{th} variable, p_0 was the mean of the variable in high and low level, P was the difference between high level and low level of a particular variable divided by 2 and P_i was an encoded value of the i^{th} particular variables. Table 1 shows the three variables at five level actual and coded values.

Table 1: Three Factors at Five Levels Estimated Values

Variables	Actual and coded values				
	-1.682	-1	0	1	1.682
P_1 =Reaction temperature(^{0}C)	55	60	65	70	75
P_2 =Reaction time (minutes)	60	90	120	150	180
P_3 =Ratio of solute: solvent(g:ml)	1:4	1:5	1:6	1:7	1:8

The yield of the oil was determined by the formula given below [8]

$$oil\ yield(\%) = \frac{weight\ of\ oil}{weight\ of\ sample}\ X\ 100$$

Second order polynomial model was used to analyze the CCD data as given in equation below

$$y = \beta_0 + \sum_{i=1}^{3}\beta_i p_i + \sum_{i=1}^{3}\beta_{ii}\ p_i^2 + \sum_{i=1}^{2}\sum_{j=i+1}^{3}\beta_{ij}p_i p_j$$

where, y represented the yield Oil, β_0 is the constant coefficient, β_i is the linear coefficient, β_{ii} is the quadratic coefficient, β_{ij}, is the interaction coefficient. p_i is the i^{th} independent variable.

Determination of the Optimal Parameter-Setting

The optimal parameter setting was achieved analytically by use of the following function;

$$\hat{Y} = \hat{\beta}_0 + \boldsymbol{P'b} + \boldsymbol{P'BP}$$

This function can be written in matrix form as

$$\boldsymbol{P} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix},\ \boldsymbol{b} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} \text{ and } \boldsymbol{B} = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/_2 & \hat{\beta}_{13}/_2 \\ \hat{\beta}_{21}/_2 & \hat{\beta}_{22} & \hat{\beta}_{23}/_2 \\ \hat{\beta}_{31}/_2 & \hat{\beta}_{32}/_2 & \hat{\beta}_{33} \end{bmatrix}$$

Where **b** represents a 3×1 vector of regression coefficients and B represents a 3×3 symmetric matrix with the elements in the main diagonal as quadratic coefficients β_{ii}, and the off-diagonal elements are half the mixed quadratic coefficients, β_{ij} $(i \neq j)$ and $i = 1,2,3,4$. To get the optimum point we differentiate $\hat{Y}$ with respect to X as follows.

$$\frac{\partial \hat{Y}}{\partial X} = \boldsymbol{b} + 2\boldsymbol{BP} = 0$$

The stationary point is given as

$$\boldsymbol{P_S} = -\frac{1}{2}\boldsymbol{B^{-1}b}$$

And the estimated response is

$$\hat{Y} = \hat{\beta}_0 + \frac{1}{2}\boldsymbol{P'_S b}$$

After establishing the stationary point, we sought to determine if it was a maximum, minimum or saddle point. The study used partial derivatives to calculate the coordinates and characteristics of each stationary point, which will then be visualized using contour plots and response surface. The maximum, minimum, saddle point, or stationary ridge are typically taken into account by the second order model.

Canonical Analysis

The canonical analysis is very important in systems of maximum and minima as it is used, to define the type and nature of the stationary points. Using the sign on the eigenvalues from matrix $\boldsymbol{B}$, we can determine whether the stationary point is a maximum or minimum response surface. If all the eigenvalues have a negative sign the stationary point is maximum and if the eigenvalues values have positive signs, the stationary point is a minimum. If we have mixed signs on the eigenvalues, the stationary point is a saddle point.

The canonical equation is given by

$$\hat{y} = \hat{y}_s + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \cdots + \lambda_k w_k^2$$

Where $\{\lambda_i\}$ are the eigenvalues of the matrix $\boldsymbol{B}$, while the $\{w_i\}$ are the transformed explanatory variables. The response surface is sloppiest in the $\{w_i\}$ direction, corresponding to the largest absolute eigenvalue.

III. Result

RESULTS AND DISCUSSION

Optimization of Oil Yield from Mango Seeds using RSM and ANN

The Central Composite Design (CCD) comprised of a total of 20 experimental runs: with a full factorial design of eight experimental points, six axial points, and central points replicated six times.

3.1.1 Experimental Results of Oil Yield from Mango Seeds

The experimental, predicted and deviations results for the oil extraction from the mango seed are presented in Table 2. The difference between the actual value and predicted value ranged between -1.3 to 0.6 with a percentage error ranging between -14.13 to 22.82. pared t-test was carried out to establish whether there was any statistical significant difference between the actual value and predicted value. The t-statistical value for the test was found to be 0.473381 with two sided critical value of $\pm$2.12 i.e. (-2.12 < 0.473381 < 2.12). this shows that there was no statistical significant difference between the actual value and predicted value suggesting that the response surface approach was appropriate for predicting the oil production from mango seeds.

Table 2: Full Factorial Central Composite Design Matrix and Experimental Results for the Yield of Oil

R	Coded Values			Actual Values			Amount of Oil (%)			
	P_1	P_2	P_3	Temperature	Time	Ratio	Actual	Predicted	Deviation	Error (%)
1	-1	-1	-1	60	90	1:5	6.7	7.3	0.6	8.96
2	+1	-1	-1	70	90	1:5	3.5	3.2	-0.3	-8.57
3	-1	+1	-1	60	150	1:5	7.8	8.1	0.3	3.85
4	+1	+1	-1	70	150	1:5	3.3	3.2	0.1	3.03
5	-1	-1	+1	60	90	1:7	7.0	6.7	-0.3	-4.29
6	+1	-1	+1	70	90	1:7	9.2	7.9	-1.3	-14.13
7	-1	+1	+1	60	180	1:7	5.6	5.2	-0.4	-7.14
8	+1	+1	+1	70	180	1:7	6.9	7.3	0.4	5.80
9	0	0	0	65	120	1:6	4.4	4.5	1	22.73
10	0	0	0	65	120	1:6	8.0	7.3	-0.7	-8.75
11	0	0	0	65	120	1:6	5.5	5.4	-0.1	-1.82
12	0	0	0	65	120	1:6	7.8	7.9	0.1	1.28
13	-1.682	0	0	55	120	1:6	5.8	6.8	1	17.24
14	1.682	0	0	75	120	1:6	6.5	6.6	0.1	1.54
15	0	-1.682	0	65	60	1:6	3.9	4.4	0.5	12.82
16	0	1.682	0	65	180	1:6	5.6	5.7	0.2	3.57
17	0	0	-1.682	65	120	1:4	7.3	7.3	0	0.00
18	0	0	1.682	65	120	1:8	7.3	7.3	0	0.00
19	0	0	0	65	120	1:6	7.3	7.3	0	0.00
20	0	0	0	65	120	1:6	7.0	7.3	0.3	4.29

Model Fitting of Oil Yield from Mango Seed

In this section two models were used to show the relationship between the response variable and the control variables. The two models are the first order and the second order model.

First Order Model of Oil Yield from Mango Seeds

The first order model was fitted to the experimental data to determine its appropriateness in predicting the yield of the oil from mango seed. The results on the estimates of the co-efficient, standard error (Std. error), t-statistic and probability value, are given in Table 3.

Table 3: First Order Model Parameter Estimates for the Yield of Oil

Parameters	Estimate	Std. error	t-statistic	Pr (> | t |)
(Intercept)	6.471	0.279	23.173	0.000
P_1	1.575	0.369	4.263	0.002
P_2	0.375	0.369	1.015	0.334
P_3	0.750	0.369	2.030	0.070

The results in Table 3 indicate that the reaction temperature (P_1) had the highest effect (1.575) followed by Ratio of Solute to Solvent (P_3) effect (0.75) whereas the Time had the lowest effect (0.375) on the Yield of oil. Moreover, it was noted that only reaction temperature (P_1) was statistically significant on the yield of Oil since the p-value was less than 0.05 (0.002 < 0.05). This implies that reaction temperature (P_1) is very critical in the production of Oil from mango seed. Reaction time and the solute to solvent variables were statistically insignificant at 5% significance level. The first order model is expressed as follows;

$$\hat{y} = 6.471 + 1.575P_1 + 0.375P_2 + 0.75P_3$$

A constant of 6.471 represent the estimated amount of oil from mango seed oil when the control variables are set zero. This is a coded model useful for identifying the relative impact of the control factors by comparing their coefficients.

Model Summary Statistic

Table 4 presents the model summary statistic which gives the values of R square, Adjusted R squared F-statistic and the p-value. These statistic values especially, adjusted R-squared, helps in determining the goodness of fit for the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Y_{\exp(i)} - Y_{\text{pred}(i)})^2}{\sum_{i=1}^{N}(Y_{\text{pred}(i)} - Y_{\exp(i)})^2} = 0.6999$$

$$Adjusted\ R^2 = 1 - \left[(R^2) \times \frac{N-1}{N-P-1}\right] = 0.6099$$

Table 4: Model Summary Statistic

Statistic	Value
R-squared	0.6999
Adjusted R-squared	0.6099
F-statistic	7.7760
p-value:	0.0057

The results in Table 4, indicate that 69.9% of variation in oil production is explained by the controlled factors in the model. This implies that the first order model is slightly suitable for prediction purpose. This could have been attributed by the fact that two of the main control factors were statistically insignificant but crucial in oil production.

Analysis of Variance

It was important to check for the adequacy of the fitted first order model and the findings are displayed in Table 5

Table 5: Analysis of Variance

Source of variation	df	SS	MSS	F	Pr(>F)	Remarks
Model (FO)	3	25.470	8.490	7.776	0.006	Significant
Residuals	10	10.919	1.092			
Lack of fit	5	9.985	1.997	10.699	0.011	Significant
Pure error	5	0.933	0.187			

SS (FO) = 25.47, SS (Residuals) = 10.919, SS (Lack of fit) = 9.985 SS (Pure error) = 0.933

$MSS = \frac{SS}{Df}$, MSS (FO) = 8.490. MSS (Residual) =1.092, MSS (Lack of fit) = 1.997

MSS (Pure error) = 0.187

$$F\ (FO) = \frac{MSS\ (FO)}{MSS\ (RESIDUALS)} = \frac{8.490}{1.092} = 7.776$$

$$F\ (Lack\ of\ fit) = \frac{MSS\ (Lack\ of\ fit)}{MSS\ (Pure\ error)} = \frac{1.997}{0.187} = 7.776$$

The adequacy of the model was tested by use of lack of fit test. A lack of fit test is a test used to determine whether or not a suggested model statistical model offers a significantly better fit to experimental data than the full polynomial model. The decision of the lack of fit test is based on statistical significance of the F-statistic at 5% level of significance.

If the p-value is greater than 0.05 level of significance, then there is no evidence that the model does not fit the data (the relationship assumed in the model is reasonable) and therefore there is no lack of fit of the model and vice versa. In other words, the predicted model gives accurate relationship between study variables.

The lack of fit F-statistic is calculated by dividing the lack of fit mean sum of square (MSS) by the pure error mean sum of square (MSS). In order to determine the lack of fit for the hypothesis was set as;

H_0: There is no evidence that the model does not fit the data, i.e., there is no lack of fit.

H_1: There is evidence that the model does not fit the data, i.e., there is lack of fit

Table 5 shows the lack of fit F-statistic value of 10.699 and p-value 0.011 which is less than 0.05 level of significance. Hence, the null hypothesis is not supported (rejected) suggesting that there is lack of fit meaning that there is evidence that the reduced model is unreasonable for prediction purpose. This meant that there was a need to use a higher order model in predicting the oil yield.

The Second Order Polynomial Model for the Yield of Oil

This section presents the analysis of the second order polynomial model showing the relationship between the response variable and the control variables.

Fitting a Second Order Polynomial Model for the Yield of Oil

The study sought to fit a second order model to the experimental data. A second order rotatable design was adopted and the following results (Table 6) were obtained. The regression estimates, standard error of estimate, t-statistic and probability value associated with the estimate of linear, interaction and quadratic effects were presented.

Table 6: Parameter Estimates for Second Order Model for the Yield of Oil

Parameters	Estimate	Std. Error	t – statistic	Pr (> | t |)
(Intercept)	7.267	0.191	38.129	0.000
P_1	1.452	0.126	11.484	0.000
P_2	0.343	0.126	2.711	0.022
P_3	0.599	0.126	4.740	0.001
$P_1 P_2$	-0.325	0.165	-1.967	0.078
$P_1 P_3$	0.300	0.165	1.816	0.099
$P_2 P_3$	0.400	0.165	2.421	0.036
P_1^2	-0.574	0.123	-4.666	0.001
P_2^2	-0.451	0.123	-3.660	0.004
P_3^2	-0.362	0.123	-2.942	0.015

The result indicated that all main effects had a significant effect on oil yield. For instance, for a positive change in reaction temperature, it influences oil yield by a factor of 1.452. Indeed, for an increase in reaction time, correspond to an increase of oil yield by a factor of 0.343. An increase in P_3 by one unit, the yield of oil would increase by a factor 0.599. Furthermore, the results revealed that only interaction effects between reaction time and solute to solvent ratio were statistically significant at 5% significance value (0.036< 0.05). This implies that for a unit increase in the interaction between the two variables, correspond to a positive effect on the yield of oil by a factor of 0.4. The results found that there was statistically significant quadratic effect in all the controlled variables at 5% significance level although the impacts were negative. This suggest that too much of reaction temperature, time and solute to solvent ratio had resulted to reduced oil yield. This finding is in line with a study by [5] who found that too much temperature impact oil production negatively. The empirical second order model can be expressed as follows;

$$\hat{y} = 7.267 + 1.452\, P_1 + 0.343\, P_2 + 0.599\, P_3 - 0.325\, P_1\, P_2 + 0.300\, P_1\, P_3 + 0.400\, P_2\, P_3 - 0.574 P_1^2 - 0.451 P_2^2 - 0.362 P_3^2$$

After dropping the effects which are not statistically significant, the reduced model becomes,

$$\hat{y} = 7.267 + 1.452\, P_1 + 0.343\, P_2 + 0.599\, P_3 - 0.325\, P_1\, P_2 - 0.574\, P_1^2 - 0.451 P_2^2 - 0.362 P_3^2$$

Where $\hat{y}$ represents the oil yield from mango seeds

P_1 represents reaction Temperature

P_2 represents reaction Time

P_3 represents Solute to Solvent ratio

This is the coded model suggested for the prediction of the oil yield in the presence of the controlled variables.

Model Summary Statistic

Table 7: presents the model summary statistic which gives the values of R square, Adjusted R squared F Value and the p-value. These statistic values especially, adjusted R-squared, helps in determining the goodness of fit for the fitted model.

Table 7: Model Summary Statistic

Statistic	Value
R-squared	0.9549
Adjusted R-squared	0.9143
F – statistic	23.52
p – value	1.396×10^{-5}

The results in Table 7 indicate that 91.4% of the variation in oil yield is accounted for by reaction temperature, reaction time and solute to solvent ratio in the second order model. This implies that the fitted second order polynomial model is suitable for prediction purpose.

Analysis of Variance

It was important to check the adequacy of the second order model by conducting analysis of variance. Results from ANOVA were presented in Table 8

Table 8: Analysis of Variance

Source of Variation	df	SS	MSS	F-statistic	Pr(>F)
FO (P_1, P_2, P_3)	3	35.31	11.77	53.8964	1.77E-06
TWI (P_1, P_2, P_3)	3	2.845	0.9483	4.3425	0.03336
PQ $(P_1, P_2, P_3))$	3	8.073	2.691	12.3226	0.001074
Residuals	10	2.184	0.2184		
Lack of fit	5	1.25	0.2501	1.3398	0.37802
Pure error	5	0.933	0.1867		

The results in Table 8 shows the significance of each component (main effects, two way interactive and pure quadratic) as well as lack of fit for the second order polynomial model. The results revealed that the main effects components are statistically significant at 5 % significance level (F-statistic is 53,59, p-value < 0.05) also the two-way interactive effect between the controlled variable in the model was also statistically significant at 5% significance level, although it is not possible to identify which two interactive factors are significant. What it means is that at least one two-way interaction factors are significant at 5% significant level. Similarly, the study found that the pure quadratic components had significant effects on the yield of oil at 5%significant level (F-statistic is 12.32 and p-value 0.001 < 0.05).

If the p-value is greater than 0.05 level of significance, then there is no evidence that the model does not fit the data (the relationship assumed in the model is reasonable) and therefore there is no lack of fit of the model and vice versa. In other words, the predicted model gives accurate relationship between study variables.

The lack of fit F-statistic is calculated by dividing the lack of fit mean sum of square (MSS) by the pure error mean sum of square (MSS). In order to determine the lack of fit for the hypothesis was set as;

H_0: There is no evidence that the model does not fit the data, i.e., there is no lack of fit.

H_1: There is evidence that the model does not fit the data, i.e., there is lack of fit

In this case, the lack of fit of the model was found to be statistically insignificant at 5% significant level (F-statistic is 1.24 and p- value 0.37802 > 0.05). Hence, the null hypothesis was supported (fail to reject) and therefore there is no lack of fit meaning that there is no evidence that the model does not fit the data (reasonable). This implies that the fitted model is capable of predicting the oil yield on the basis of controlled variables under study.

Validation of the Model

It was necessary to assess the validity of the model and for easy understanding and clarity. Graphical representation of predicted values using the model together with the corresponding measured values of all the responses has been made in Figures 3. One of the ways of validating the fitted models is by plotting a line that join the points of the predicted and experimental on a scatter chart. The y-axis shows the model's predicted values, while the x-axis shows the experimental actual values. The estimated regression line is the diagonal line in the center of the plot. From the figure, each data point is quite close to the projected regression line, thus we conclude that the regression model slightly fits the data equitably well.

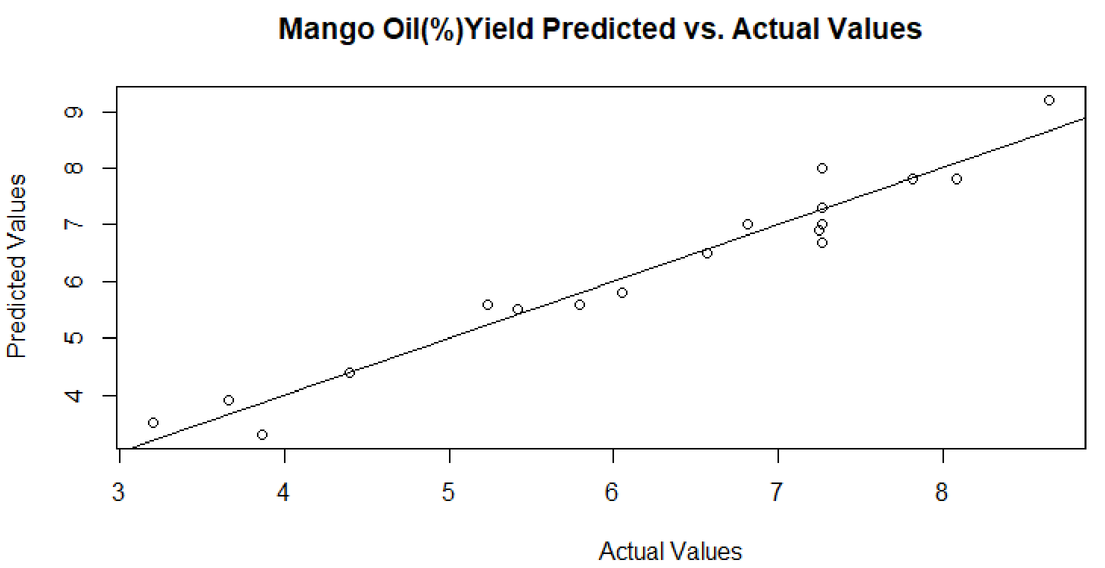


Figure 3: Predicted Values versus Experimental Values of the Oil yield

Optimization Analysis using RSM

The operating optimal levels of reaction temperature, time and solute to solvent ratio that maximizes oil yield from mango seed were determined through RSM and ANN.

Three Factor Central Composite Design for oil Extraction

Since second-order model could be used to approximate the response, then it was important to determine the required levels of the three controlled variables that could guarantee the maximum yield of oil. A three-dimensional (3D) surface plot was constructed to investigate the interactive effects of the two factors on the yield of oil while the other factor is held a constant within the experimental ranges. The following response surface graphs display the controlled variables on the x- and y-axes and then a continuous surface that represents the fitted response values on the z-axis. Response Surface and Contour Plots are a 3D representation of the response surfaces produced by the model for oil yield, the relationship between the response variable and the control variables is depicted. In order to create these charts, two variables had to be kept within the experimental range, while the third variable's value remained constant.

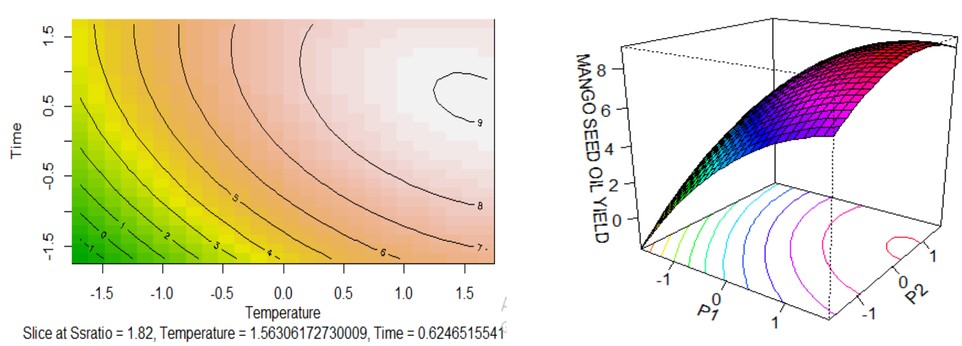


Figure 4: Contour Plot and Response Surface for Oil Yield as a Function of Temperature (P1) and Time (P2) at a fixed S/S Ratio

Figure 4 shows the 3D response surface and contour plots. It shows that a slight adjustment of temperature and time, at a fixed solute to solvent ratio, affect the yield of oil positively. It gives the visualization nature of optimal points for oil production. It is evident from the figure that the yield of oil had a maximum point, and it appears that the results of the optimization conditions of oil extraction to obtain maximum yield indicate an increase. The maximum percentage yield is achieved when temperature assumes a coded value of 1.56, at 0.62 time when solute to solvent ratio is fixed at 1.82.

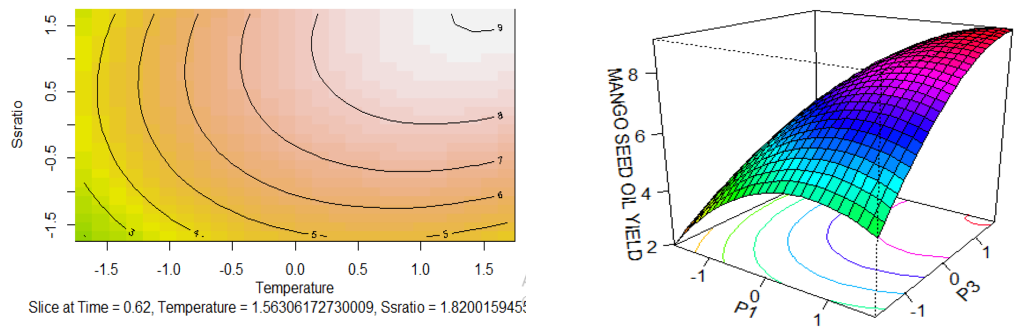


Figure 5: Contour and Response Surface Plot for Oil yield as a Function of Temperature and S/S Ratio at a fixed Time

Figure 5 shows the contour and 3D response surface plot of the yield of oil as a function of Temperature and solute to solvent ratio at a constant Time. Temperature and solute to solvent ratio were shown to have a direct effect on the yield of oil. The point was a maximum which was achieved at fixed time coded value of 0.62, temperature of 1.56 and solute to solvent ratio of 1.82.

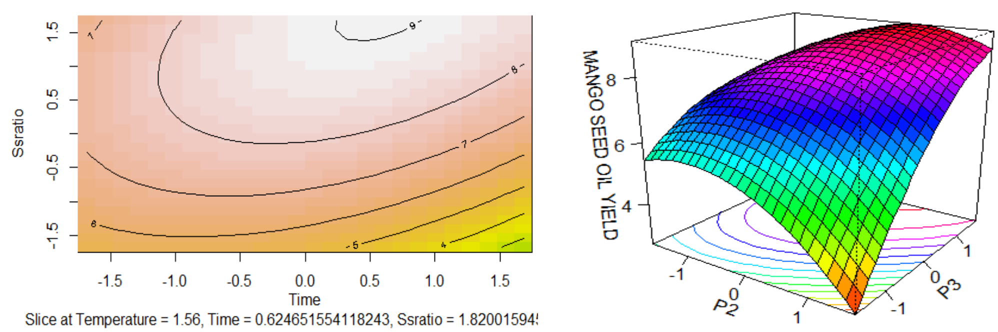


Figure 6: Contour and Response Surface Plot for Oil yield as a Function of Time and S/S Ratio at a fixed Temperature

Figure 6 shows the 3D response surface and contour plots. It shows that a slight adjustment of time and s/s ratio at a fixed temperature increases the yield of oil. It gives the visualization nature of optimal points for oil production. It is evident from the figure that the yield of oil had a maximum point, and it appears that the results of the optimization conditions of oil extraction to obtain maximum yield indicate an increase. The maximum percentage yield is achieved when time assumes a coded value of 0.62, at 1.82 solute to solvent ratio when temperature is fixed at 1.56.

Optimal Conditions for Maximum Mango Seeds Oil Production

In order to determine the optimal settings of the controlled variables, a canonical analysis was performed by obtaining the stationary point. The response surface stationary points output was in coded form, which could be transformed into their natural values using the equations as stated below;

$$Reaction\ Temperature = (5 \times 1.5630617) + 65 = 72.09^{0}c$$
$$Reaction\ Time = (0.5 \times 0.6246516) + 2 = 2.14\ hrs$$
$$solute\ to\ solvent\ ratio = 1{:}(50 \times 1.8200159) + 300 = 391 = 1{:}7$$

This set of conditions gave an Oil yield of 9.05 %. The summary of stationary points and optimal oil yield are tabula ted in Table 9

Table 9: Optimal Conditions for Maximum Mango Seeds Oil Production

Variable	Optimal Values (coded)	Optimal Values (Actual)
Temperature	1.563	72.09
Time	0.625	2.14
Solute to solvent ratio	1.820	1:7
Optimal oil yield		9.05%

The study revealed that to obtain 9.05% of oil yield, it is recommended that a temperature of 72.09 ºC, a time of 2.14 hours and a solute to solvent ratio of 1:7 is crucial. This finding is in line to a study by [4] who conducted an oil extraction on three varying time lengths, three- hour, four- hour and five hours and yield of the oil recorded. At three-hour extraction period. They reported that hexane had the highest yield of 9.85. At four-hour extraction period. They reported that hexane had the highest yield of 9.15%. At five-hour extraction period. They reported that hexane had the highest yield of 9.12%.

Eigen values and Eigen vectors

Eigen values and Eigen vectors were realized from canonical analysis of the fitted model and the results are shown in Table 10

Table 10: Eigen values and vectors

		Eigen values	
	λ_1 = - 0.2005875	λ_2 = - 0.3720088	λ_3 = - 0.8142427
		Eigen vectors	
P_1	0.05941199	0.7333987	-0.6771976
P_2	0.59989440	- 0.5684613	-0.5630084
P_3	0.79787024	0.3727976	0.4737352

The nature of the stationary points is determined by the signs of the eigenvalues. Table 10 shows that all Eigen values had negative signs implying that the stationary point was a maximum point. The canonical equivalence form for the fitted model can be stated as follows;

$$\hat{y} = 9.05 - 0.201w_1 - 0.372w_2 - 0.814w_3$$

IV. Conclusion

A second order polynomial model was found to be reasonable meaning that there is no evidence that the model does not fit the data. This implies that the fitted second order polynomial model is suitable for prediction purpose. In order to obtain optimal oil yield of 9.05% from mango seeds, it recommended a temperature of 72.09 ºC, time of 2.14 hours and a solute to solvent ratio of 1:7 at optimal levels using response surface methodology. The study has established that the mango seed has a substantial amount of oil which can be extracted using the optimal conditions set on controlled variables. Value addition on mango seeds could guarantee improvement of economic status of farmers and the country. Further the pollution attributed to littering the environment by mango seeds would be minimized.

Acknowledgement

The author acknowledges the support from Tharaka University management for Funds granted and provision of the Laboratory and equipment's. The purchase of equipment's and chemicals used, analysis of the samples was success because of the funds provided.

References

[1]. Box, G. E. P., And K. B. Wilson. "On The Experimental Designs For Exploring Response Surfaces." Ann Math Stat 13 (1951): 1-45.
[2]. Box, G.E.P And Hunter, J.S (1957). Multi-Factorial Experimental Designs For Exploring Response Surfaces. Ann.Math.Statist.28 195-241
[3]. Emil Akbar, Zahira Yaakob; Siti Kartom Kamarudin, Manal Ismail And Jum'at Salimon (2009) Characteristic And Composition Of Jatropha Curcas Oil Seed From Malaysia And Its Potential As Biodiesel Feedstock.

[4]. Karunanithi, B., Bogeshwaran, K., Tripuraneni, M., & Krishna Reddy, S. (2015). Extraction Of Mango Seed Oil From Mango Kernel. International Journal Of Engineering Research And Development, 11(11), 32-41.

[5]. Mas'ud, F., Mahendradatta, M., Laga, A., & Zainal, Z. (2017). Optimization Of Mango Seed Kernel Oil Extraction Using Response Surface Methodology. OCL, 24(5), D503.

[6]. Montgomery, C.D (2001). Designs And Analysis Of The Experiments, 5th Edition, Arizon State University.

[7]. Nzikou, J. M., Kimbonguila, A., Matos, L., Loumouamou, B.,Pambou-Tobi, N. P. G., Ndangui, C. B., Abena, A. A., Silou, Th.,Scher, J. And Desobry, S. (2010) "Extraction And Characteristics Ofseed Kernel Oil From Mango (Mangifera Indica)" Research Journal Of Environmental And Earth Sciences 2(1): 31-35

[8]. Sani I. (2014) Soxhlet Extraction And Physicochemical Characterization Of Mangifera Indica L. Seed Kernel Oil. Research And Reviews: J Food Dairy Tech 2(1): 20–24.

[9]. Yadav, K., Garg, N., Verma, A., Kumar, S., & Trivedi, M. (2017). Optimization And Extraction Of Oil From Mango Seed Kernel (Mangifera Indica). Indian Journal Of Agricultural Sciences, 87(7), 943-946.