International Journal of Statistics and Applied Mathematics 2016; 1(4): 24-33

ISSN: 2456-1452
Maths 2016; 1(4): 24-33

www.mathsjournal.com
Received: xx-09-2016
Accepted: xx-10-2016

Edwine B Atitwa
Department of Computer and Statistics, Moi University, Eldoret, and Kenya

Joseph K Koske
Department of Computer and Statistics, Moi University, Eldoret, and Kenya

John M Mutiso
Department of Computer and Statistics, Moi University, Eldoret, and Kenya

Correspondence
Edwine B Atitwa
Department of Computer and Statistics, Moi University, Eldoret, and Kenya

The optimum settings of culture conditions for optimum growth of kefir grains for nutrition and health using RSM with BBD

Edwine B Atitwa, Joseph K Koske and John M Mutiso

Abstract
This thesis provides the process used to optimize the growth of kefir grains for nutrition and health to yield the optimal size of the kefir grain. This combined process was successfully modeled and optimized by response surface methodology method with a Box–Behnken design. In this work, Box-Behnken Design of response surface methodology was used to optimize growth of kefir grains using fat content in milk, incubation time and the number of rotations using milk as culture liquid. The contour plots were generated to study the optimum yield and levels of culture conditions. The second-order model, was used to locate the optimum growth of kefir grains. Testing the prediction of the second order and its adequacy was verified using the I-optimality and G -optimality which was used to predict the growth of kefir gains at nearly all values of culture conditions. The statistical analysis was done using a two-way analysis of variance for fitting the data. The confidence level to check the growth of kefir grain were at 95%.Design expert was used to formulate the BBD design which yielded a class of three incomplete factorial designs which are useful for estimating the co-efficient in a second degree graduating polynomial. The optimal settings of the culture conditions were evaluated using stationery points to determine the possibilities of the response for the growth to optimal at stationery points levels. The data was analyzed using R statistical software and excel. Data was presented using tables and graphs.

Keywords: optimum settings, culture, conditions, growth, nutrition, health

1. Introduction
1.1 Background of the study
Kefir grains are a kind of yoghurt starter which are white to yellow-white, gelatinous and variable in size and consist of a complex microbial symbiotic mixture of lactic acid bacteria, yeasts and few acetic acid bacteria which stick to a complex-protein matrix and carbohydrates [6] These benefits of Kefir grains contain probiotics.it is a fantastic source of other nutrients, it lowers blood pressure, lowers blood sugar, good for cancer patients especially when going through chemotherapy, it can help people with allergies, helps the detoxification process by binding to some mutagens, agents that can literally change your DNA**,** they are wonderful for digestion, it is a whole food for nutrients and health, Its better than yoghurt since kefir usually contains about 30 strains of bacteria, whereas yoghurt products usually only contain about 10**.** It can help prevent side effects and damage done by antibiotics; it may help with lactose intolerance [5] To optimize growth, higher degree polynomial such as second-order model is used and this leads to the approximating function with squared variables and its interaction.In factorial experiments, different levels of multiple factors are investigated simultaneously and one factor can be examined at different levels of the other factor or factors [12].

1.2 Problem Statement of the study
Presently, there are many research articles about optimization methods; the typical ones are based on calculus, numerical methods, and random methods. The calculus based methods have been intensely studied and are subdivided in two main classes: 1) the direct search methods that find a local maximum moving on a function over the relative local gradient directions and 2) the indirect methods that usually find the local ends by solving a set of non-linear

equations, resultant of equaling the gradient from the object function to zero, i.e., by means of multidimensional generalization of the notion of the function's extreme points from elementary calculus give a smooth function without restrictions to find a possible maximum which is to be restricted to those points whose slope is zero in all directions. Considering the diversity in composition of the natural biomass in kefir grain production, as well as the structure and physicochemical properties of these probiotics, a universal processing protocol is not conceivable, and specific processes must be designed and optimized for each growth size depending on different factors and the culture composition used [10]. For such programmes to be effective, it is essential that the models are accurate. Once process variability from these sources is identified and optimized, an accurate extraction process model can be derived. Available optimization tools make it possible to optimize the production process design variables and the one of the best tool is the response surface methodology (RSM). Thus the main problem make the growth of kefir grains that yields grains that contain high level of probiotics using RSM with BBD techniques with high level of probiotics that are healthy and nutritious.

1.3 Purpose of the study

The study was carried out to understand the optimum setting of fat content, number of rotation and time that will yield optimum growth of kefir grains using milk. It also involved modeling and optimizing the level of the controllable independent variable.

2. Literature Review

The first goal for Response Surface Method is to find the optimum response. When there is more than one response then it is important to find the compromise optimum that does not optimize only one response [16]. This design was developed by Box and Behnken [7] it provides three levels for each factor and consists of a particular subset of the factorial combinations from the 3^k factorial design. The actual construction of such a design is described in the three RSM books [4]. For this study fitting a second-order model was done using the Box-Behnken design (BBD) which is a response surface methodology (RSM) design that requires only three levels to run an experiment. It is a special 3-level design because it does not contain any points at the vertices of the experiment region. This could be advantageous when the points on the corners of the cube represent level combinations that are prohibitively expensive or impossible to test because of physical process constraints [9]. To attain this [2], defined a variance function, i.e., the scaled prediction variance. The SPV provides a measure of the precision of the estimated response at any point in the design space is optimal using G-optimal criteria. G-optimality protects the experimenter against the worst case scenario being too undesirable. An interesting and an important result is that the lower bound for the maximum SPV is equal to *p*, the number of parameters in the model [13]. The G- and I-criteria are prediction oriented criteria, so they are used for second-order models, as second-order models are often used for Optimization and good prediction properties are essential for optimization [12]. I-optimality seeks designs that minimize the average variance of prediction over the experimental region χ by integrating the variance surface $x'M^{-1}x$. [17] argue in favor of I optimality that the attempt of determining I-optimality is to generate a single measure of prediction performance through an averaging process; that is, $v(x)$ is averaged over some region of interest χ. The rotatability of the design was determined to check the variance of the predicted response at any point is a function of the distance from the central point alone [11]. A second -order response surface design is slope rotatable if the variance of the estimate of the first derivative $\frac{\delta yu}{\delta xi}$, is only a function of the distance $(\sum_{i=1}^{k} x^2_{iu})$ of the points $(x_{1u}, x_{2u}, ...x_{ku})$ from the origin(center) of the design [18]

3. Material and Methodology

Response surface methodology was used to optimize the growth of kefir grains using milk fat content, number of rotations and time since the response was influenced by different parameters [16]. The optimization entailed the basic strategies involved use of four steps [7].

3.1 Second-order model

For this study fitting a second-order model was done using the Box-Behnken design (BBD). The second order model used to approximate the response was expressed as

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \beta_{ii} x^2_i + \varepsilon \quad (1)$$

The stationary points in the design [2] were located by expressing the model in matrix notation,

$$\hat{y} = \beta_0 + X'b + X'BX \quad (2)$$

To find stationery points, partial derivatives were taken as follows

$$\frac{\delta y}{\delta x1s} = \frac{\delta y}{\delta x2s} = ...\frac{\delta y}{\delta xks} = 0 \quad (3)$$

Where x_{1s}, $x_2s....x_{ks}$ are stationary points. This resulted in to the following possibilities of the response

i) Apoint of maximum response
ii) Apoint of minimum response
iii) The saddle point

The contour plots were also generated using R for further response analysis.

The derivatives w.r.t. elements of vector x= $\begin{bmatrix} x1 \\ x2 \end{bmatrix}$ for (3) was taken to determine the stationery points as follows

$$\frac{\delta \hat{y}}{\delta x} = b + 2BX = 0$$

$$Xs = -\frac{1}{2} B^{-1} b \quad (4)$$

and the predicted response at the stationary points were determined by

$$\hat{y}s = b_0 + X'sb + X'sBXs$$

$$= b_0 + X'sb + (-\frac{1}{2} bB^{-1}) BX's$$

$$= b_0 + \frac{1}{2} X'sb \quad (5)$$

And the predicted maximum response was given by

$$\hat{y} = b_0 + \frac{1}{2} y_s \quad (6)$$

Optimizing was done by considering the second –degree model in written in matrix for as

$$y = \beta_0 + x'\beta_* + x'Bx + \varepsilon \quad (7)$$

where $x = (x_1, x_2, ...x_k)'$, $\beta* = (\beta_1, \beta_2, ...\beta_k)'$ and B is a symmetric matrix of order k × k whose $i^{th,}$ diagonal element is β_{ii} (i = 1, 2... k), and its $(i, j)^{th}$ off-diagonal element is $\frac{1}{2}\beta_{ij}(i, j = 1, 2, ..., k;\ \ i \neq j)$

And the method of Lagrange multipliers was used to search for the optimum considering the function

$$H = \hat{\beta}_0 + x'\hat{\beta}* + x'\hat{\beta}x - \lambda(x'x - r^2) \quad (8)$$

Differentiating H w.r.t x, we have

$$\hat{\beta}* + 2(\hat{\beta}x - \lambda x) = 0 \quad (9)$$

Solving for **X**, we obtain

$$X = \frac{1}{2}(\hat{\beta} - \lambda I_n)^{-1}\hat{\beta}* \quad (10)$$

Which gives a solution of stationery point of $\hat{y}(x)$ A maximum (minimum) is achieved at this point if the Hessian matrix, that is, the matrix $\frac{\partial}{\partial x}\left[\frac{\partial H}{\partial x'}\right]$ of second-order partial derivatives of H with respect to $\boldsymbol{x}$ is negative definite (positive definite). This was expressed in matrix form by

$$\frac{\partial}{\partial x}\left[\frac{\partial H}{\partial x'}\right] = 2(\hat{\beta} - \lambda I_n) \quad (11)$$

Therefore, to achieve a maximum, we applied [9] concept who suggested that λ be chosen larger than the largest eigenvalue of $\hat{\beta}$. Such a choice causes $\hat{\beta} - \lambda I_n$ to be negative definite. Choosing λ smaller than the smallest eigenvalue of $\hat{\beta}$ causes $\hat{\beta} - \lambda I_n$ to be positive definite, which results in a minimum.

To test the prediction of the second model through the region of interest was done by testing the ratatability of the model to provide good predictions throughout the *region* of interest. At some point x, the variance is given by the formula,

$$Var\ [\hat{y}(X)] = \sigma^2 X'(X'X)^{-1}X \quad (12)$$

To measure the degree of rotatability for a given response surface design.We determine the scaled predictive probability as

$$SPV = \frac{N}{\sigma^2}\operatorname{var}[\hat{y}(x)] \quad (13)$$

A design

$$X = \begin{pmatrix} x_{12} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{x1} & \cdots & x_{nk} \end{pmatrix}$$

is said to be rotatable if the prediction variance in [13] is constant at all points that are equidistant from the design center, this, by a proper coding of the control variables, can be chosen to be the point at the origin of the k-dimensional coordinates [8].

3.2 Optimal -Criterion

The design prediction criteria of response surface experiment for this model was determined by using G-optimality criterion and I-optimality criterion. G-optimality seeks to minimize the maximum entry in the diagonal of the hat matrix $X'(X'X)^{-1}X$. The design is G-optimal if it minimizes the maximum scaled prediction variance over the design region i.e. is the maximum value of G-optimality was determined by

$$G = \frac{NV[\hat{y}(x)]}{\sigma^2} \quad (14)$$

Over the design region is a minimum, where N is the number of points in the design. Then efficiency of the design was determined by

$$Ge = \frac{p}{\max \frac{Nv[\hat{y}(x)]}{\sigma^2}} \quad (15)$$

Where Ge is the G-efficiency.

Assume that we wish to construct a design for fitting a full quadratic polynomial response surface on a *k*-dimensional design space

$$y(x) = \beta_0 + \sum_{i=1}^{k}\beta_i x_i + \sum_{i=1}^{k}\beta_{ii}x^2_i + \sum_{i=1}^{k-1}\sum_{i+1}^{k}\beta_{ij}x_i x_j + \varepsilon \quad (16)$$

Where y **(x)** is the response variable, x_1... x_k are the parameters, are the errors of the quadratic model which are independent, with zero mean value, and σ^2 variance.β are the unknown coefficients given by $p = \frac{(k+1)(k+2)}{2}$ (17)

Assuming that the design consists of N ≥ p samples.

For I –optimal, the predicted variance at arbitrary point x was expressed as

$$Var[y(x)] = \frac{\sigma^2}{N} f(x)' M_x^{-1} f(x) \quad (18)$$

And I-optimal

$$I = \frac{n}{\sigma^2}\int_R \mathrm{var}[y(x)]dx = Trace(MM_x^{-1}) \tag{19}$$

Where R is the design space and

$$M = \int_R f(x)f(x)'dx \text{ (cuvazzuti M.2013)} \tag{20}$$

The I-optimal design to determine the I –optimal criteria efficiency was then defined by moment matrix M as shown below (Giovannitti-Jensen and Myers, 1989);

M=

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \sigma_2 & \sigma_2 & \sigma_2 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_2 & 0 & 0 & 0 & \sigma_4 & \sigma_{22} & \sigma_{22} & 0 & 0 & 0 \\ \sigma_2 & 0 & 0 & 0 & \sigma_{22} & \sigma_4 & \sigma_{22} & 0 & 0 & 0 \\ \sigma_2 & 0 & 0 & 0 & \sigma_{22} & \sigma_{22} & \sigma_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{22} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{22} \end{bmatrix}$$

Fig 1

Where

$$\left.\begin{aligned} \sigma_2 &= \int_R x_i^2 dx = \frac{r^2}{k} \\ \sigma_4 &= \int_R x_i^4 dx = \frac{3r^4}{k(k+2)} \\ \sigma_{22} &= \int_R x_i^2 x_j^2 dx = \frac{r^4}{k(k+2)} \\ r &= \left\{\sum_{i=1}^{k} X_i^2\right\}^{\frac{1}{2}} = k^{\frac{1}{2}} = \alpha \end{aligned}\right\} \tag{21}$$

r is the radius of a ball or sphere [15]. Hence it seeks to minimize the average prediction variance over the design space [1] which leads to a nice efficiency for the design which is given by;

$$I_{eff} = \frac{Min(APV(X))}{APV(X)} \tag{22}$$

Where APV= $x(X'X)^{-1}x'$

4. Results

Determination of the optimum levels of milk fat content, number of rotations and time that lead to optimum growth of kefir grains Since there was presence of significant lack of fit for the first order model for main and interaction effects, hence there was a curvature in the response surface which makes the first-order model to be insufficient especially when all the factors interactions. A second-order model was developed to help in approximating a portion of the true response surface with parabolic curvature or quadratic effects**.** This model was designed using BBD and The data in appendix 1 used to fit the second order model for optimization of kefir grains using milk From appendix II, the second order equation model was fitted as follows

$Yyield=65.670+5.116A+4.374B-7.100C-2.911A^2-1.256B^2+4.011C^2-5.513AB-2.560AC-3.425BC$

From appendix II all other parameters are significant at p-value<.05 except B^2, AC and ABC which had non- significant influence on the growth of kefir grains. Thus the refined best model to predict growth of kefir grains using second order model was determined as follows;

$Yyield=65.670 +5.116A+4.374B-7.100C-2.911A^2+4.011C^2-5.513AB-3.425BC$

The respective a nova table of the second order model was displayed as shown in Appendix III, From a nova table, coefficients of all the factors were significant factor at p-value<.05, except the coefficient of B(speed of rotations) in the quadratic component and the interaction of AC for the growth of kefir grain, this results to the equation below that is best suit for fitting second order model. Thus fitting values of factors A, B, and C in using coded values, the optimum yield of kefir grains using milk was predicted as follows

$Yyield=65.670+5.116(1)+4.374(1)-7.100(-1)-2.911(1)+4.011(-1)^2-5.513(1)(1)-3.425(1)(-1) =80.602$ grams

The R^2 and adjusted R^2 were very high(>.9) and close to 1 hence it indicated that equation (4.20) is a good model to estimate the optimum growth of kefir grains using second order model. The R^2 and adjusted R^2 were also significant at a p-value<.05.

4.1 Contours of relationship of factors for growth of kefir grains using milk

The contours plots for lines and area were drawn to explain the relationship between the culture conditions and the response as shown in the figure below.

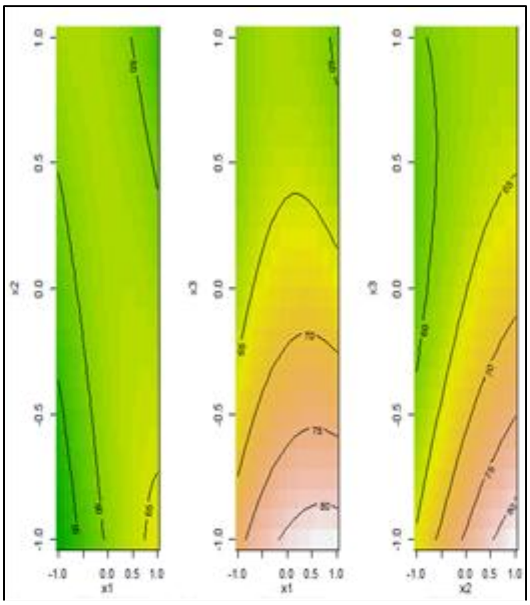

Fig 2: Relationship of Fat (x1), Rotations(x2) and Time(x3) on growth of kefir grains

From figure 2, the response surface is a plane, the contour plots are parallel straight lines. The analysis of the contour plot for x2*x1: This plot indicates how number of rotations and milk fat content, are related to the growth of kefir grains while incubation time is held constant at low level (-1). The response is at its highest at 65 grams of the kefir grains obtained from the graph (lower right corner). The second is x3 * x1: This plot indicates how Time and milk fat content, are related to the *growth* of kefir grains while number of rotation is held constant at high level (1). The response is at its highest at 80 grams of kefir grains obtained from the graph (lower right corner).The third is x3 * x2: This plot indicates how Time and milk fat content, are related to the *growth of kefir grains* while number of rotations is held constant at high level (1). The response is at its highest at 80 grams of kefir grains obtained from the graph (lower right corner).Thus in order to maximize the growth of kefir grains, we can choose low level settings of time and high level settings for milk fat content and number of rotations. The final estimated regression model using the coded variables is thus expressed and predicted as follows;

$Yyield=65.670 +5.116(1) +4.374(1)-7.100(-1)-2.911(1) +4.011(-1)^2-5.513(1)(1)-3.425(1)(-1) =80.602$ grams

The growth rate was thus given by;

$$\left(\frac{80.602-20}{20}\right)*100=303.01\%$$

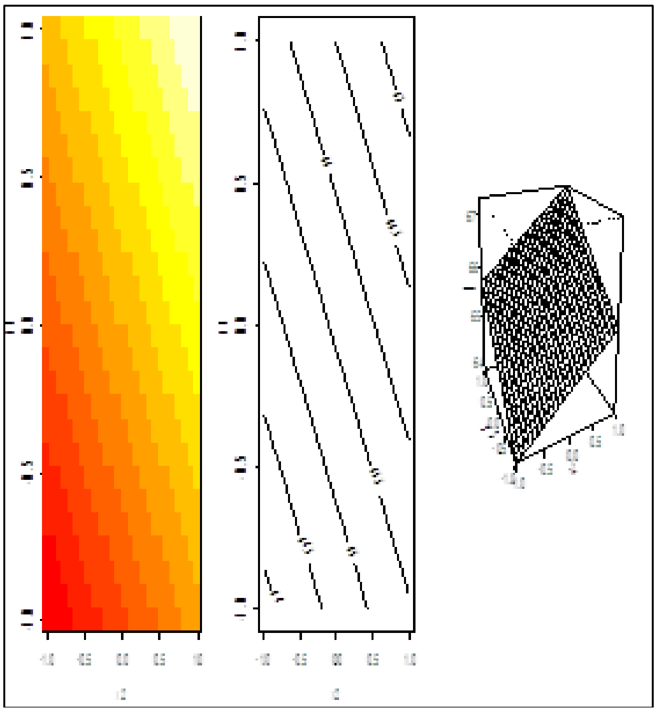

Fig 3: the area only, line only and the wireframe plots for relationship between fat and rotations on growth of kefir grains.

From figure 3, the relationship of fat content and speed of rotations indicates an optimal growth for kefir grains which is > 67 grams. This occurs at high levels (1) of the variables. It also indicates that increases in growth of kefir grains will increases with increase in both fat content in milk and speed of rotation of solution for the growth of kefir grains.

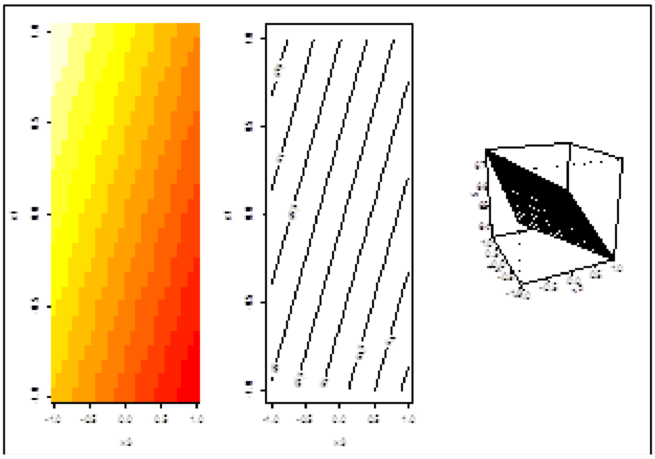

Fig 4: the area only, line only and the wireframe plots for relationship between fat and time on growth of kefir grains.

From figure 4, the relationship of fat content and time indicates an optimal growth for kefir grains which is > 67.5 grams. This occurs at high level (1) of the fat content and low level (-1) of time. It also indicates that increases in growth of kefir grains will increases with increase in fat content in milk and decrease in incubation time when growing kefir grains.

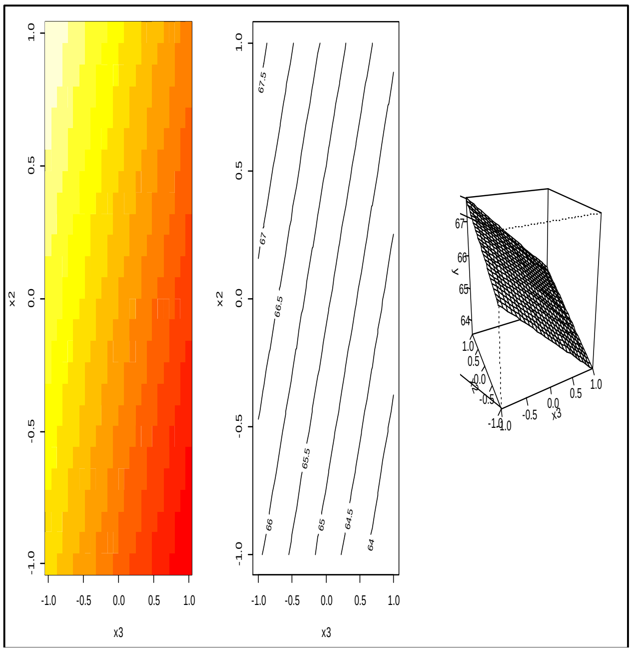

Fig 5: the area only, line only and the wireframe plots for relationship between rotations and time on growth of kefir grains.

From figure 5, the relationship of fat content and time indicates an optimal growth for kefir grains which is > 67.5 grams. This occurs at high level (1) of the number of rotations and low level (-1) of time. It also indicates that increases in growth of kefir grains will increases with increase speed of rotations and decrease in incubation time when growing kefir grains.

4.2 The 3D Surface plots for operating conditions of growing kefir grains using milk

They were useful for establishing the growth of kefir grains based on milk fat content, speed of rotations and time using milk as the culture liquid. The 3D surface plots provided a clearer concept of the response surface than contour plots. For this study they were displayed as follows

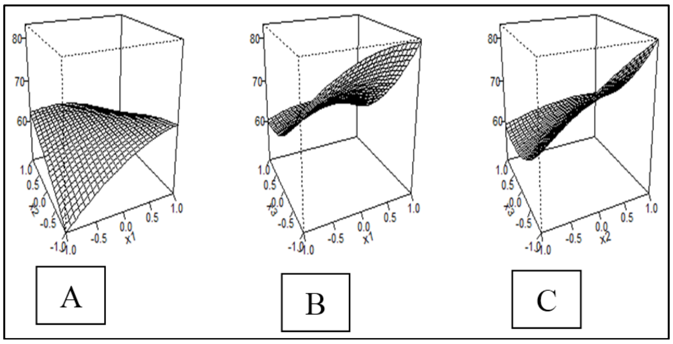

Fig 6: Three design surface plots for a Box-behnken Design.

From figure 6, the 3D for the BBD is a simple maximum 3D surface plot since the response values of the kefir grains are based on regression model. As the color gets darker, the response increases i.e. the growth of kefir grains increases. The increases in fat content and speed of rotation increases the growth of kefir grains using milk. For (B) and (C) are both minimax patterns of the 3D surface plots of the response. As the color gets darker, the response increases. From the stationary point (saddle point), increasing either factor while decreasing the other leads to an increase in the response. Hence increases in speed of rotations and fat content while decreasing time yields increase in growth of kefir grains based on regression model.

4.3 Analyzing the Stationary Point

The second-order model illustrates quadratic surfaces such as minimum, maximum, ridge and saddle such that if there exits an optimum then this point is a stationary point. When the surface is curved in one direction but is fairly constant in another direction, then this type of surface is ridge system. The stationary point for this design was found by using matrix algebra. Recall

Yyield=65.670+5.116A+4.374B-7.100C-2.911A^2-1.256B^2+4.011C^2-5.513AB-2.560AC-3.425BC

Then to find the location of the stationary point for *Yield given that;*

$$b=\begin{bmatrix} 5.116 \\ 4.374 \\ -7.100 \end{bmatrix}$$ and

$$\beta=\begin{pmatrix} -2.911 & -2.757 & -1.280 \\ -2.757 & -1.256 & 1.7125 \\ -1.280 & 1.7125 & 4.011 \end{pmatrix}$$

Then stationary points were obtained using (3.35)

$$X_s=\begin{bmatrix} 3.125 \\ -5.339 \\ 2.392 \end{bmatrix}$$

The stationary points were also found in terms of the natural variables for milk fat content, number of rotations and time as follows;

$$3.125=\frac{f-7}{3.5}$$;f=17.938 which is a minimum point of reaction by using 18 grams of fat content in growing kefir grains using milk no change is expected on the growth within the levels of settings for time and speed of rotations.

$$-5.339=\frac{r-100}{50}$$; r=-166.95 which is a maximum point of reaction and it takes 167 rotations to grow kefir grains using milk and no change is expected on the growth within the levels of settings for milk fat content and time.

$$2.392=\frac{t-36}{12}$$;t=64.704 which is a minimum point of reaction and it takes 65 hours to grow kefir grains using milk and no change is expected on the growth within the levels of settings for milk fat content and speed of rotation.

Thus the estimated maximum response *Yield* of kefir grains was obtained as follows $Y_s = 24.34567$

And the predicted maximum response was

$\hat{y}_s$ =65.67+(1/2*Ys)= 77.84 *grams*

This was a significant optimum growth of kefir grains at stationery points. Since f and t were positive hence minimum point and r was negative hence maximum, then the stationery points of this experiment resulted to a saddle point.

4.4 Application of Ridge analysis to optimize the design

Since the stationery points resulted to a saddle point, the ridge analysis for saddle point was applied to second-order model. This helped to maximize the response yield of kefir grain under stationary points. Recall

Yyield=65.670+5.116A+4.374B-7.100C-2.911A^2-1.256B^2+4.011C^2-5.513AB-2.560AC-3.425BC

Then
$$\beta=\begin{pmatrix} -2.911 & -2.757 & -1.280 \\ -2.757 & -1.256 & 1.7125 \\ -1.280 & 1.7125 & 4.011 \end{pmatrix}$$

And λ being the eigen value of β was obtained as λ =(5.0781081, -0.2725766, -4.9615315) and let $\lambda = \mu$ then

$$\mu I=\begin{pmatrix} -5.078 & .000 & .000 \\ .000 & -0.273 & .000 \\ .000 & .000 & -4.961 \end{pmatrix}$$ and $$b=\begin{bmatrix} 5.116 \\ 4.374 \\ -7.100 \end{bmatrix}$$

The stationary points were obtained as

$$X_s=\begin{bmatrix} -1.027 \\ 4.156 \\ -.544 \end{bmatrix}$$

To achieve the maximum or minimum, we let μ be =6.078 which is more than the largest value of μ to obtain

$$\begin{pmatrix} -17.978 & -5.514 & -2.56 \\ -5.514 & -14.668 & 3.425 \\ -2.56 & 3.425 & -4.134 \end{pmatrix}$$

That is all diagonals (-17.978,-14.668,-4.134)<0 which is maximum point

And letting μ =-6.078

$$= \begin{pmatrix} 6.334 & -5.513 & -2.56 \\ -5.514 & 9.644 & 3.425 \\ -2.56 & 3.425 & 20.178 \end{pmatrix}$$

This leads to all diagonals (6.334, 9.644, 20.178)>0 which is minimum point.

But the Eigen values of $\hat{\beta}$ determined by R program were 5.078,-0.273 and-4.962, indicating a minimum point which is positive and two maximum points which are all are negatives and this yields saddle point of operation. This indicates that a ridge analysis reveals reasonable operating conditions, with the implied constraints of growing kefir grains using milk.

4.5 Testing the prediction of the second model through the region of interest

This was done by determining the rotatability of the second model to provide a good prediction throughout the region of interest to grow kefir grains using milk. The variance was determined as follows

$$X = \begin{vmatrix} 1 & -1 & -1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 1 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & -1 & 1 & 0 & 1 & 0 & -` & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 & 0 & 1 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

Fig 7

The generalized inverse of X'X

$(X'X)^{-1} =$

$$\begin{bmatrix} 0.33 & 0.00 & 0.00 & 0.00 & -0.17 & -0.17 & -0.17 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.00 & -0.00 & -0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.17 & -0.00 & 0.00 & 0.00 & 0.27 & 0.02 & 0.02 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.17 & 0.00 & 0.00 & 0.00 & 0.02 & 0.27 & 0.02 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.17 & 0.00 & 0.00 & 0.00 & 0.02 & 0.02 & 0.27 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.02 & 0.00 & 0.00 & 0.25 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.25 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.25 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

Fig 8

Let σ^2 =MSE=4.67 from appendix IV

Hence variance of the predicted response at a point of interest

X was obtained using (3.42) and the results were displayed in table below

Appendix 1

Run	A	B	C	A*A	B*B	C*C	AB	AC	BC	ABC	Yield
1	-1	-1	0	1	1	0	1	0	0	0	48.33
2	-1	1	0	1	1	0	-1	0	0	0	67.7
3	1	-1	0	1	1	0	-1	0	0	0	66.33
4	1	1	0	1	1	0	1	0	0	0	63.65
5	-1	0	-1	1	0	1	0	1	0	0	64.09
6	-1	0	1	1	0	1	0	-1	0	0	55.96
7	1	0	-1	1	0	1	0	-1	0	0	82.7
8	1	0	1	1	0	1	0	1	0	0	64.33
9	0	-1	-1	0	1	1	0	0	1	0	68
10	0	-1	1	0	1	1	0	0	-1	0	59.7
11	0	1	-1	0	1	1	0	0	-1	0	84
12	0	1	1	0	1	1	0	0	1	0	62
13	0	0	0	0	0	0	0	0	0	0	65.67
14	0	0	0	0	0	0	0	0	0	0	65.67
15	0	0	0	0	0	0	0	0	0	0	65.67

Appendix II

The data was processed using R program and results were displayed as shown in the tables below.

Table 1: Regression estimates for second order model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	**65.6700**	**1.2477**	**52.633**	**4.68e-08 *****
A	5.1163	0.7641	6.696	0.001124 **
B	4.3738	0.7641	5.724	0.002276 **
C	-7.1000	0.7641	-9.292	0.000243 ***
A^2	-2.9113	1.1247	-2.589	0.048923 *
B^2	-1.2563	1.1247	-1.117	0.314776
C^2	4.0113	1.1247	3.567	0.016101 *
AB	-5.5125	1.0805	-5.102	0.003765 **
AC	-2.5600	1.0805	-2.369	0.064018.
BC	-3.4250	1.0805	-3.170	0.024824 *
ABC	NA	NA	NA	NA

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.161 on 5 degrees of freedom Multiple R-squared: 0.9785, Adjusted R-squared: 0.9399 F-statistic: 25.34 on 9 and 5 DF, p-value: 0.00119

Appendix III: ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	209.41	209.41	44.8380	0.0011237 **
B	1	153.04	153.04	32.7681	0.0022762 **
C	1	403.28	403.28	86.3495	0.0002428 ***
A^2	1	36.06	36.06	7.7219	0.0389563 *
B^2	1	9.09	9.09	1.9474	0.2216783
C^2	1	59.41	59.41	12.7207	0.0161015 *
AB	1	121.55	121.55	26.0262	0.0037651 **
AC	1	26.21	26.21	5.6130	0.0640185.
BC	1	46.92	46.92 1	0.0469	0.0248235 *
Residuals	5	23.35	4.67		

---Signif. Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 2: Predictive variances

Run	$x(X'X)^{-1}x'$	$Var[y^{\wedge}(x)] = \sigma^2 x(X'X)^{-1}x'$; **Where** σ^2 **=4.67**
1	0.75	3.502
⋮	⋮	⋮
⋮	⋮	⋮
12	0.75	3.502
13	0.33	1.541
⋮	⋮	⋮
15	0.33	1.541

Hence prediction variance is constant at all points that are equidistant from the design center($Var[\hat{y}(x)] = 3.502)$, this was achieved by a proper coding of the control variables, which were to be chosen at the origin of the k-dimensional coordinates of the design. This resulted to predictive variance which is the variance of the predicted response at points of interest. The variance of the stationery points was 1.541 as shown in the above table. Thus for all points that are equidistance from the design center the variance=4*0.75=3.502.This shows that the design is rotatable because the variance of the points that are equidistance from the design center is the same for all non-stationery runs. The scaling predictive variance (SPV) was computed to enable comparison of various design points. The SPV was used to provide a measure of the precision of the estimated response at any point in the design space and it was used to compare the design. It was obtained and the results were displayed in the table below

Table 3: Unscaled and scaled predictive variances.

Run	$x(X'X)^{-1}x'$	$Nx(X'X)^{-1}x'$ **Where N=15**
1	0.75	11.25
⋮	⋮	⋮
⋮	⋮	⋮
12	0.75	11.25
13	0.33	4.95
⋮	⋮	⋮
15	0.33	4.95

From table 3, the results of the SPV of the designs are stable for equidistant points and stationery points which have a predictive variance of 11.25 and 4.95 respectively. Hence it’s a good response design since it has a good profile of the unscaled predictive variance and scaled predictive variances and also has very few runs, hence the second –order model had a good prediction of the region of interest. The design is also rotatable since it had same values of SPV=11.25 for any two points in non-stationery points and SPV=4.95 for any two points in stationery points.

4.6 Design Prediction Criteria of Response Surface experiment for the second –order model

This was determined by use of G-optimality criterion and I-optimality creiterion. G-optimality seeks to minimize the maximum SPV throughout the region of the design i.e. min [max (V(x))] value which gives the G-optimal design of the experiment. This lead to a good prediction at a particular location in the design space. G-optimal criterion was obtained as G-optimal =11.25

And the corresponding $G-effiency = \frac{10}{11.25}$ =.89

Where p=10, given that k=3.

Hence 89% efficiency which ≈90% was achieved and this means that the maximum SPV=11.25 for this design is not that much different from its optimum value. This makes the maximin G-optimal design cost effective and highly recommendable since it’s more effective and hence reduces cost which involved in designing and growing kefir grains using milk by box-behnken design.

I-optimality was also computed to minimize the normal average integrated prediction variance and M was determined as follows by

Obtaining $\sigma_2 = \frac{1.73}{3}$ =0.58, $\sigma_4 = \frac{27}{15}$ =1.8, $\sigma_{22} = \frac{9}{15}$ =0.6 and $\alpha = \sqrt{3}$ =1.73

Then substituting in M we obtaine

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & .58 & .58 & .58 & 0 & 0 & 0 \\ 0 & .58 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .58 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .58 & 0 & 0 & 0 & 0 & 0 & 0 \\ .58 & 0 & 0 & 0 & 1.8 & .6 & .6 & 0 & 0 & 0 \\ .58 & 0 & 0 & 0 & .6 & 1.8 & .6 & 0 & 0 & 0 \\ .58 & 0 & 0 & 0 & .6 & .6 & 1.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .6 \end{bmatrix}$$

Fig 9

And

$(M'M)^{-} =$

$$\begin{bmatrix} 1.88 & 0 & 0 & -1.09 & 1.09 & 1.09 & 2.18 & 0 & 0 & 0 \\ 0 & 1.09 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.09 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.63 & -0.54 & -.544 & 0 & 0 & 0 & 0 \\ 1.09 & 0 & 0 & -2.25 & 4.50 & 0 & 2.25 & 0 & 0 & 0 \\ 1.09 & 0 & 0 & -2.25 & -1.55 & 4.5 & 2.25 & 0 & 0 & 0 \\ 1.09 & 0 & 0 & -1.13 & 1.13 & 1.13 & 6.75 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Fig 10

Using (3.50) we obtain I-optimal =25.94
(4.29)
The I-optimality criterion seeks the design that minimizes the integrated (or average) variance of the estimated response over the design space R and this was achieved by determining efficiency of I-optimal by substituting the minimum and average predictive variances in to obtain

$$I-eff = \frac{.33}{(.33+.75)/2} = .611$$

This yielded 61.1% of the efficiency in growing kefir grains using milk. This is a good efficiency on average and it suggests that there was uniformity in running experiments to grow kefir grains using milk. Thus the average differences in running the experiments for the box-behnken designed were minimized by 61.1% and hence more efficiency was realized in operations during the growing process.

5. Conclusion

The second order model was studied to determine the optimum growth of kefir grains which yielded an optimal growth of kefir during the process. The best results for predictive response were obtained at stationery points. Second-order model described quadratic surfaces, which represented minimum point using aridge analysis where all the stationery points were all positive and hence minimum point was achieved from this study.

6. Acknowledgement

I would like to thank Almighty God who gave me strength, good health, and wisdom during period of my research. Thanks to my supervisor, Professors Koske J.K and John Muindi Mutiso, whose tolerance and help has been invaluable.

7. References

1. Atkinson AC. Fast very robust methods for the detection of multiple multivariate data, Computational Statistics and Data Analysis. Oxford: Oxford University Press. 2007.
2. Box GEP, Hunter WG, Hunter JS. Statistics for Experimenters: An Introduction toDesign, Data Analysis, and Model Building. 2nd edition. John Wiley & Sons, New York. 2005.
3. Box, Leu. Multifactor experimental designs for exploring response surfaces. Ann MathStat. 1999, 1957, 28:195-241.
4. Box GEP, Draper NR. Response Surfaces, Mixtures, and Ridge Analyses. 2nd ed. Hoboken, New Jersey: John Wiley & Sons. 2007.
5. Dominic N Anfiteatro. How to make kefir and Recipes Created, published, maintained and Copyrighted © by. All rights reserved. 2015.
6. Farnworth ER. Kefir–A complex probiotic. Food Sci Technol. Bulletin: Functional Foods: 2000; 2(1)1-17. doi: 10.1616/1476-2137.13938
7. Box GEP, Hunter JS. Multifactor experimental designs for exploring response surfaces, Annals of Mathematical Statistics. 2010; 28:195-241.
8. Huda S, Al-Shingiti A. Rotatable generalized central composite designs: A – minimax efficiencies for estimating slopes. Pakistani Journal of Statistics. 2004; 20(3):397-407.
9. Jun S, Irudayaraj J, Demirci A, Geiser D. "Pulsed UV-light treatment of corn meal for inactivation of Aspergillusniger spores, International Journal of Food Science and Technology. 2003; 38(8):883-888.
10. Molina S, Moran-Valero MI, Martin D, Vázquez L, Vargas T, Torres CF. Antiproliferative effect of alkylglycerols as vehicles of butyric acid on colon cancer cells. ChemPhys Lipids. 2013; 175-176:50-56. doi: 10.1016/j.chemphyslip.
11. Cavazzuti M. Optimization Methods: From Theory to Design, 2013, 13 DOI: 10.1007/978-3-642-31187-1_2, © Springer-Verlag Berlin Heidelberg
12. Montgomery DC. *Design and analysis of experiments*, Wiley, Hoboken, NJ, 5th ed. 2009.
13. Myers RH, Montgomery DC. Response Surface Methodology: Product and Process Optimization Using Designed Experiments. 2nd edition. John Wiley & Sons, New York. 2002.
14. Nam-Ky Nguyena, John J Borkowskib. New 3-level response surface designs construc- ted from incomplete block designs, Journal of Statistical Planning and Inference. 2008, 294-305.
15. Onukogu IB, Chigbu PE. (eds.) Super Convergent Line Series in Optimal Design of Experiments and Mathematical Programming. Nsukka: AP Express Publishers. 2002.
16. Park SH, Kim HJ, Cho J. Optimal central composite designs for fitting Second order response surface regression models. Recent advances in Linear Models and Related Areas. Physica – Verlag HD, 2008, 323-339.
17. Rodriguez M, Jones B, Borror CM, Montgomery DC. Generating and Assessing Exact G-Optimal Designs, Journal of Quality Technology, 2010; 42(1):329.

18. Victorbabu, Ch Surekha VVS. a note on measure of rotatability for second order response surface designs using incomplete block designs. 2013.