JMCS

Journal of Mathematics & Computing Science

Vol 1. Issue 1, May 2016

UNIVERSITI TEKNOLOGI MARA
Unit Penerbitan UiTM Kelantan

e-ISSN: 0128-2697

JOURNAL OF MATHEMATICS AND COMPUTING SCIENCE, 2018,
Volume 1, Number 1, 39-54

Modified Desirability Function For Optimization of Multiple Responses

Nasuhar Abd Aziz[1]*, Nurul Awanis AS[2], Siti Nurliana Noraziman[3]

[1,2,3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Kelantan, Kota Bharu, Kelantan, Malaysia
*[1]nasuha978@kelantan.uitm.edu.my
[2]nurulawaniiss@gmail.com
[3]nurliananrzmn@gmail.com

Abstract: Harrington's desirability function approach is frequently used to overcome the problem of optimization of multiple responses simultaneously. However, this method will give a huge impact in the presence of outliers. Hence, it is not reliable to use Harrington's desirability function method to find the optimum responses in this case because it is not resistant to outliers. As an alternative, Modified Geometric Mean (MGM) approach is proposed to estimate the parameter since this approach is resistant to outliers. Numerical example study is carried out to compare the performance of the proposed method with existing procedures. Based on the value of the overall desirability function, D ,MGM is better compared with Harrington's desirability function as it clearly shows that the value of D is larger and the standard error of the MGM approach is smaller. In overall, it is evident that the MGM approach can be an alternative method in dealing with the presence of outliers.

Keywords: Harrington's desirability approach, modified geometric mean, multiple responses, outliers.

1 Introduction

Montgomery [1] stated that the response of interest in modeling and analyzing problem can be obtained by applying Response Surface Methodology (RSM). The purpose of this technique is to optimize the response. For instance, assume that a chemical engineer wants to discover the level of temperature (x_1) and time (x_2) that expand the yield of the procedure. The process yield is a function of the levels of temperature and time, say

$$y = f(x_1, x_2) + \varepsilon \tag{1.1}$$

where ε represents the error observed or can be called as noise in the response *y*. If it be denoted as the expected response by $E(y) = f(x_1, x_2) = \eta$, then the surface represented by

$$\eta = f(x_1, x_2) \tag{1.2}$$

can be called as a response surface.

Apart from that, the applications of RSM comprise the experiment for investigating the space of the procedure or controlled factors, empirical statistical modeling to build up a proper approximating relationship between the yield and process variables. The optimization technique is useful in determining the process variables that generate desirable value of the responses.

RSM is extremely valuable in engineering and manufacturing field since it considers finding and investigating how a few factors possibly impact some execution measures of a procedure and product. In addition, it can help the industrial manager or worker to manage their production such as maximizing products and minimizing cost of production. Montgomery [1] also added that RSM is a sequential method. It means that when an optimum region has been determined, a second-order model may be applied and analyzed to locate the optimum points.

Besides, another purpose of RSM is to find the optimum operating conditions of a process and a region of a factor level to get a satisfied operating requirement. Thus, optimizing the response and determining a combination that gives the highest response are the main objectives of RSM. Overall, RSM can be defined as an optimization process in finding the best set of value of the level of factors in order to get the optimum target goal.

RSM usually associates with experimental design, regression model and optimization of more than one response. There are different ways and techniques used to determine the performance of the response in the optimization process, since some of the characteristics of response variables in the model are different. Therefore, the goal is to find a suitable solution for explanatory variables which will result in the best possible value for each response.

Contour plot is a relatively straightforward and traditional way to approach. It optimizes a few responses that works well for each when there are a few variables. This method is very effective for two or three explanatory variables. However, it will lose its efficiency when it has greater dimension. Overlaying contour plot can be plotted to find the best possible value for each response in a particular area. Besides, a popular approach to find the best optimization is by formulating and overcoming a problem as a constrained optimizing problem. Sometimes these techniques are referred to as a nonlinear programming method.

Another useful approach to deal with optimization of multiple responses is to use the simultaneous optimization technique popularized by Derringer and Suich [2]. The technique uses desirability function. The general approach is to convert each response into an individual desirability function, *d* that varies over the range

$$0 \le d \le 1 \tag{1.3}$$

where if the response $\hat{y}$ is at its goal or target, then $d_i = 1$ and if the response is outside an acceptable region, $d_i = 0$ and later, the individual desirability is combined into an overall desirability. If any of the individual desirability is undesirable, the value of overall desirability will be zero.

This study aims to achieve its objectives which is first, to determine the best estimated model for multiple responses. Next, the objectives are to modify the desirability function for multiple responses based on geometric median as well as to locate the best optimum point for each response.

2 Literature Review

2.1 Multiple Responses Experiment

Box and Draper [3] mentioned that RSM has evolved to a model of experiment responses and later on, to a model of numerical experiments. Response surface methodology (RSM) is a collection of statistical and mathematical techniques which are useful in developing, improving and optimizing processes as explained in Myers & Montgomery [4]. Meanwhile, according to Khuri and Mukhopadhyay [5], response surface methodology (RSM) consists of a group of mathematical and statistical techniques used in the development of an adequate functional relationship between a response of interest, y and a number of associated control variables or independent variables denoted by $x_1, x_2, x_3, \ldots x_k$.

RSM is an important part in experimental designs because it plays a vital role in designing, formulating, developing as well as analyzing new scientific studies and products as stated in Malik [6]. Besides, it is also helpful in improving the existing studies and products. Malik [6] also mentioned that applications of RSM are mostly found in Chemical Industrial, Biological, Food Science and Engineering Science.

According to Oehlert [7], RSM is a design and model for working with continuous treatments when the goal is to find the optimum or describe the responses. Oehlert [7] also added that finding the optimum response of the problem is the first and most important goal in RSM. It must be remembered that it is important to find the best and compromising optimum that does not optimize only one response when there are more than one response. The main objective of RSM is to determine the optimum operational conditions of the process. The RSM usually contains three steps that are design and experiments, response surface modeling through regression and optimization.

RSM usually associates with experimental design, regression models and optimization for more than one response. There are different techniques and ways in the optimization process to determine the performance of the response when it is in use, since response variables in the model are different in some characteristics. The optimization analysis is more complex in the presence of multiple responses than in the one response case as reported by Khuri and Cornell [8]. As a result, Khuri and Cornell [8] stated that it is rare for the entire response variable to achieve the respective optimal in the same conditions. There are many types of process optimization problem to apply in RSM based on the multiple objectives of the optimization in multiple responses experiment. Lind, Goldin and Hickman[9] developed a graphical approach called contour plot. The contour plots of all the responses were superimposed on each other and then the optimal point for all the responses were found. Harrington[10] developed the desirability function approach to multiple responses optimization that the transformations of exponential for each of the responses were used into desirability functions. Later, Derringer and Suich [2] modified the Harrington's desirability approach.

Besides, a research has been done by Yusof, Talib, Mohamed and Bakar [11] in Malaysia, a determination of optimum pH, temperature and 'Brix to produce guava concentration by using RSM in their research. The factors that were chosen to be the explanatory variables in this research were optimum pH, temperature and 'Brix. Meanwhile, the exploratory aspects in this research were colour and viscosity of the concentration, titratability acidity, flavour, body (mouthfeel) as well as overall acceptability of the diluted juice of guava. The data of this study was analysed by using multiple regression analysis. Thus, from the analysis, the optimum value to obtain suitable colour were pH 4.0, 87-95℃ and 46" Brix. For flavour, the optimum values were pH 3.3-3.6, 78436℃ and 50P"Brix whereas for overall acceptability, pH3.3-3.9, 79.1℃ and 35-55" Brix were its optimum values. The result generated found that pH factor was the crucial factor that contributed to the characteristics of a product. It also shows high significant influence on the colour of the concentration and on titratable acidity, flavour as well as on overall acceptability of the juice of guava. Meanwhile, the 'Brix factor only affected the colour, titratable acidity and overall acceptability score. However, the temperature factor only affected the colour.

Besides, a study done by Hu, Cai and Liang [12] in Guangzhuo, China was conducted to examine the optimization of Microwave-Assisted Extraction (MAE) of Saikosaponins from Radix Bupleuri. The purpose of optimization process was to determine the MAE condition that gives maximum extraction yields of each saikosaponins simultaneously. The optimization method was analysed by using RSM with Central Composite Rotatable Design (CCRD). There are four explanatory variables that have been used which are microwave power (x_1), irradiated time (x_2), extraction temperature (x_3) and ethanol concentration (x_4). Meanwhile there are three exploratory variables namely extraction yields of saikosaponin a(Y_1), saikosaponin c$(Y_2)$and saikosaponin d(Y_3). The effects of explanatory variables on the respective exploratory variables were tested by using ANOVA. By applying desirability function approach, the optimum MAE conditions to achieve desirable extraction yield for all saikosaponins were found at the microwave power of 360-400 W, irradiated time of 5.8-6.0 min, temperature of 73-74 and ethanol of 47-50%. Meanwhile, the yield of saikosaponin a, c and d are 96.18-96.91%, 95.05-95.71% and 97.05-97.25% respectively.

Besides that, a study by Islam, Alam and Hannan [13] in Bangladesh was conducted to study the particle board production by using multiple response optimization process. There are seven factors used in the experiment that are flake thickness, flake length, dried chip moisture content, amount of adhesive, pressing time, pressure and press temperature. The data was analyzed by ANOVA and the second-order polynomial model was developed using multiple regression analysis. An optimization by Derringer's desirability function was performed and the best optimized conditions were found to be flake thickness of 0.15mm, temperature $182^{\circ}C$ and 3.5% of dried chip moisture content. The result from the study shows that flake thickness, dried chip moisture content and press temperature were found to have a significant effect on particle board properties production.

Moreover, a research done by Fitrianto and Midi [14] in Malaysia was conducted to examine the advanced oxidation of the black liquor effluent obtained from the pulp and paper industry using the dark Fenton reaction. The data for this experiment came from an experiment conducted by Torrades, Saiz and Garcia-Hortal [15]. The factors used in the experiment were temperature, H_2O_2 concentration and Fe (II) concentration and the three response variables studied in the experiment were COD removal after 90min, UV254 removal after 90min and UV280 removal after 90min. The data was analysed by ANOVA and the polynomial models were developed using multiple regression analysis. An optimization study using Derringer's desirability function approach was performed to optimize the responses simultaneously at one best setting of factors. The optimal setting was found to be 46.84 mM and 6.771 mM of H_2O_2 concentration and Fe (II) concentration, respectively with total desirability function of 0.782. The results from the study show that H_2O_2 concentration and Fe (II) concentration significantly contribute to the quadratic model.

In addition, a study done by Maran, Manikan and Mekala [16] in India inspected the extraction of betalain pigments and colour extraction from prickly pear fruits. The interactive effect of the process variables used in the experiment which are temperature, time, mass and pH was optimized by using Box-Bexnken response surface design. Experimental data obtained from 29 experiments was analysed by ANOVA and the second-order polynomial models were developed using multiple regression analysis. Since the goal of the experiment was to maximize the extraction of betalain pigments and colour extraction, an optimization study using Derringer's desirability function methodology was performed. Under the optimized conditions, the optimal extraction conditions were found to be temperature of $42^{\circ}C$, time of 115 min, mass of 1.2 g and pH of 6.9 with total desirability value of 0.936. The result from the study shows that temperature, mass and time had a significant effect on the extraction of betalain pigments and colour extraction from prickly pear fruits.

Next, a research done by Rafieian, Keramat and Kadivar [17] in India was conducted to examine the extraction from chicken deboner residue. This study applied optimization process by using Central Composite Design. The independent variables for this study were HCI concentration, extraction temperature and extraction time. Meanwhile, the dependent variables were extraction yield, gel strength, viscosity and lightness. This study was analysed by using Minitab statistical software to obtain the regression and graphical analysis. Minitab software was also used to generate desirability function to find the maximum values of the dependent variables. The analysis of this data found that the optimum values of HCI concentration was 6.73%, extraction temperature was 86.6℃ and extraction time was 1.95h.

2.2 Desirability Function

The desirability function was introduced by Harrington [10] as an approach to multiple response optimizations and it has been widely used to simultaneously optimize several responses. Harrington [10] used exponential transformation to transform each of the responses $\hat{y}_i$ into desirability functions, d_i that can be shown as follow:

For a one-side transformation,

$$d_i = \exp(-\exp(-\hat{y}_i)) \tag{2.1}$$

For a two-side transformation,

$$d_i = \exp(-|\hat{y}_i|^r) \tag{2.2}$$

where r is a user-selected shape parameter. According to Chen et al. [18] desirability function approach consists of three stages that are model building, transformation into individual desirability function and combination into an overall desirability function.

Later, Derringer and Suich [2] modified Harrington's transformation and classified them into three forms which were the larger-the-better, the smaller-the-better and the nominal-the-better The desirability function purpose is to transform each of the *m* predicted responses $\hat{y}_1, \hat{y}_2, ..., \hat{y}_m$ to an individual desirability function, d_i, where $0 \le d_i \le 1$ with 0 indicates an undesirable value of $\hat{y}_i$ meanwhile 1 indicates a desirable value of $\hat{y}_i$. Next, the *m* individual desirability values $d = (d_1, d_2, ..., d_m)$ are combined into an overall desirability function, *D,* where $0 \le D \le 1$. The value of d_i will increase when the desirability of the corresponding response increases. However, this approach has its own weakness in which the variability of each predicted response is not explicitly considered in the algorithm of obtaining the optimal response.

Moreover, Chen et al.[18] mentioned that if the transformation into desirability does not cover the prediction interval, the optimal solution will not be acceptable for practical implementation. Therefore, Chen et al. [18] developed Augmented Desirability Function approach to determine the factors settings and optimum mean response to make the optimal solution more practical. Thus, the Augmented Desirability Function incorporates the desirability approach of the secondary information into the overall desirability function via combination. Suppose the *m* secondary information variables $s = (s_1, s_2, ..., s_m)$ are called as the informative variables that affect the process of optimization. Later, the *s* is transformed into secondary individual desirability function $d_s = (d_{s1}, d_{s2}, ..., d_{sm})$ that ranges between 0 and 1, where 0 indicates an undesirable value while 1 indicates desirable value. All the d_s are combined into a secondary overall desirability function denoted as *S* using geometric mean where $0 \le S \le 1$. Lastly, these two overall desirability, *D* and *S* are combined into the DS_λ which is called as Augmented Overall Desirability Function where $0 \le DS_\lambda \le 1$.

2.3 Identification of outliers

A variety of identification procedures for outliers have been suggested in the statistical literature. The residual plots based on different types of residuals have been suggested for the identification of outliers as explained in Atkinson [19]. In regression, observations corresponding to residuals which show an unusual pattern are usually flagged as outliers. A good number of analytical detection procedures of outlier are available based on the scale estimate of the residuals.

Vertical outliers are the observations with large residuals. To identify these outliers in rent diagnostic methods such as residual plots based on different types of residuals have been suggested in the statistical literature [20]. A good number of analytical detection procedures of these outliers are available based on the scale estimate of the residuals such as Standardized residuals which are defined as $d_i = e_i/\hat{\sigma}$, $i = 1,2,...,n$, where e_i is the OLS residual for the i^{th} case and $\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2$. Any observation with absolute standardized residual value larger than 2.5 is considered as a vertical outlier as stated in Rocke et al. [21]. Srikantan [22] defined another outlier diagnostic method which is called studentized residual denoted as $r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$, $i = 1,2,...,n$ where h_{ii} are the diagonal elements of the hat matrix H . Ellenberg [23] suggested another detection method as deletion studentized (also known as externally studentized or R-Student) residual for the identification of outliers. The i^{th} deletion studentized residual is defined as $t_i = \frac{y_i - x_i' \hat{\beta}_{(i)}}{\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}}$, $i = 1, 2,..., n$ where $\hat{\beta}_{(i)}$ and $\hat{\sigma}_{(i)}$ are the respective OLS parameter estimates and the MSE based on a data set without the observation that have outlier.

3 Methodology

3.1 Sources of Data

In order to apply the concept of Response Surface Methodology, Microwave-Assisted Extraction (MAE) of Saikosaponins data was chosen which was conducted by Hu, Cai and Liang [12]. The main purpose of the experiment is to analyse the optimization of the Microwave-Assisted Extraction of Saikosaponins from *Radix Bupleuri*. The purpose of optimization process was to determine the MAE condition that gives maximum extraction yields of each saikosaponins simultaneously. The optimization method was analysed by using RSM with Central Composite Rotatable Design (CCRD). There are four explanatory variables used which are microwave power (X_1), irradiated time (X_2), extraction temperature (X_3) and ethanol concentration (X_4). Meanwhile there are three exploratory variables namely, extraction yields of saikosaponin a (Y_1), saikosaponin c (Y_2) and saikosaponind (Y_3). Table 3.1 shows the Microwave-Assisted Extraction of Saikosaponins with coded and actual values of the experiment.

Table 3.1: Microwave-Assisted Extraction of Saikosaponins data with coded and actual values

Power X_1 (W)	Time X_2 (min)	Temp, X_3 (°C)[a]	Ethanol X_4 (%)	Relative Extraction Yield(%)[b]		
				Saikosaponin a, Y_1	Saikosaponin c, Y_2	Saikosaponin d, Y_3
-1 (200)	-1 (3)	-1 (65)	-1 (35)	87.43	81.79	84.97
-1 (200)	-1 (3)	-1 (65)	1 (65)	85.74	81.26	83.27
-1 (200)	-1 (3)	1 (75)	-1 (35)	87.49	84.41	90.09
-1 (200)	-1 (3)	1 (75)	1 (65)	84.91	84.1	85.7
-1 (200)	1 (5)	-1 (65)	-1 (35)	91.16	89.4	92.82
-1 (200)	1 (5)	-1 (65)	1 (65)	88.36	90.94	92.25
-1 (200)	1 (5)	1 (75)	-1 (35)	92.58	90.2	93.39
-1 (200)	1 (5)	1 (75)	1 (65)	88.08	88.43	91.25
1 (400)	-1 (3)	-1 (65)	-1 (35)	87.3	88.15	86.21
1 (400)	-1 (3)	-1 (65)	1 (65)	84.17	86.61	85.58
1 (400)	-1 (3)	1 (75)	-1 (35)	90.49	91.71	91.08
1 (400)	-1 (3)	1 (75)	1 (65)	87.35	89.46	89.31

1 (400)	1 (5)	-1 (65)	-1 (35)	93.94	90.83	93.54
1 (400)	1 (5)	-1 (65)	1 (65)	87.34	90.3	92.28
1 (400)	1 (5)	1 (75)	-1 (35)	94.29	92.33	94.7
1 (400)	1 (5)	1 (75)	1 (65)	93.25	93.35	92.82
-2 (100)	0 (4)	0 (70)	0 (50)	90.2	86.93	90.32
2 (500)	0 (4)	0 (70)	0 (50)	92.26	92.48	92.91
0 (300)	-2 (2)	0 (70)	0 (50)	88.64	83.49	89.68
0 (300)	2 (6)	0 (70)	0 (50)	94.23	94.37	95.43
0 (300)	0 (4)	-2 (60)	0 (50)	88.95	82.68	85.65
0 (300)	0 (4)	2 (80)	0 (50)	93.53	93.33	93.81
0 (300)	0 (4)	0 (70)	-2 (20)	86.07	74.08	81.25
0 (300)	0 (4)	0 (70)	2 (80)	84.72	74.35	80.52
0 (300)	0 (4)	0 (70)	0 (50)	92.25	91.34	92.69
0 (300)	0 (4)	0 (70)	0 (50)	94.02	92.18	93.29
0 (300)	0 (4)	0 (70)	0 (50)	93.21	92.24	93.72
0 (300)	0 (4)	0 (70)	0 (50)	93.78**(1.3567)**	92.57	94.4
0 (300)	0 (4)	0 (70)	0 (50)	93.87	94.91	94.98
0 (300)	0 (4)	0 (70)	0 (50)	94.39	94.02	93.95

3.2 Method of Analysis

3.2.1 Parameter Estimates for Fitting Second-Order Model

If there is curvature in the system of RSM, then a polynomial of higher degree must be used, such as the second-order model. A second-order model is appropriate in approximating the parabolic curvature. This model incorporates all terms in the first order model, all quadratic terms as well as all cross product terms. The equation can be expressed as follows:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_1 x_1 + \sum_{i=1}^{k} \beta_{ii} x_i^2 + . \sum_{i<j} \sum \beta_{ij} x_i x_j + \varepsilon \tag{3.1}$$

where $x_i = (x_{i1}, x_{i2} \dots, x_{ik})'$ are input variables and $\beta_i = (\beta_1, \beta_2, \dots, \beta_k)'$ are regression coefficients. The second order model is flexible and easily accommodated via the use of a wide, variety of experimental designs. Hence, this model can be used to find a good estimation for the response surface. In addition, the second-order model applies a method of ordinary least square to determine the coefficients of $\beta's$.

Thus, ordinary least square (OLS) can be defined as a method to generalize linear modelling technique that may be applied to either single or multiple regressor variables as well as categorical regressor variables that have been properly coded as mentioned in Moutinho and Hutcheson [24]. Ordinary Least Square method is commonly used by experimenters in order to obtain the "good" estimators of the regression parameters. The method of least square of high degree of polynomial is built using a matrix approach and it is defined by

$$\hat{\beta} = (X'X)^{-1}(X'y) \tag{3.2}$$

One of the important properties of the least squares estimators is the Gauss-Markov theorem. This theorem states that for the regression model, the least squares estimators are unbiased and have minimum variance when compared with all other unbiased linear estimators. They are called the Best

Linear Unbiased Estimators (BLUE) as stated in Montgomery [25]. Thus, the OLS estimates are optimal with assumptions of error are normal. Moreover, these estimators are more precise than any other estimators belonging to the class of unbiased estimators that are linear functions of the observations.

3.2.2 Classical Desirability Function

Desirability function analysis (DFA) is widely used for the optimization of multiple responses problems as mentioned in Derringer and Suich [2]. The aim of desirability function is to overcome the problem of multiple responses to become a single response problem. During the desirability approach process, each response will transform into individual desirability value (*d*) and the geometric mean of the individual desirability value is computed and optimised, which is known as the overall desirability function (*D*). As the response approaches the target, the desirability value becomes closer to 1.

In the first step, an individual desirability function for each response $\hat{y}_i(k)$ must be created by using the fitted models and establishing the optimization criteria. Derringer and Suich [2] mentioned that the corresponding responses of individual desirability index can be measured by using their formula. Desirability always takes value between 0 and 1 where $d_i(\hat{y})_i$ is equal to zero. This shows that it is an undesirable response since $d_i(\hat{y})_i$ which equals to 1 represents a completely desirable value. Derringer and Suich [2] modified Harrington's transformation and classified them into three forms.

There are three types of desirability functions regarding to its response characteristics such as the larger the better (LTB), the smaller-the-better (STB), and the nominal-the-better (NTB), depending on whether the response has to be maximized, minimized or obtained a target value respectively. The desirability function of the larger-the-better can be written as the term in (3.3).

$$d_i = \begin{cases} 0 & \hat{y} \le L_i \\ (\frac{\hat{y} - L_i}{U_i - L_i}), & L_i \le \hat{y} \le U_i \quad r \ge 0 \\ 1 & \hat{y} \ge U_i \end{cases} \tag{3.3}$$

where U_i is the upper acceptable value for the response and L_i is the lower acceptable value. Meanwhile *r* represents the weight, set by the experimenter to determine how important it is for $\hat{y}$ to be close to the maximum. Correspondingly, the individual desirability for the smaller-the-better can be written as the term in (3.4).

$$d_i = \begin{cases} 1 & \hat{y} \le L_i \\ (\frac{U_i - \hat{y}}{U_i - L_i}), & L_i \le \hat{y} \le U_i \quad r \ge 0 \\ 0 & \hat{y} \ge U_i \end{cases} \tag{3.4}$$

The desirability function of the nominal-the-better can be written as the term in (3.5).

$$d_i = \begin{cases} 0 & \hat{y} \le L_i \\ (\frac{\hat{y} - L_i}{T_i - L_i}) & L_i \le \hat{y} \le T_i \\ 1 \quad , & \hat{y} = T_i \quad r \ge 0 \\ (\frac{\hat{y} - U_i}{T_i - U_i}) & T_i \le \hat{y} \le U_i \\ 0 & \hat{y} \ge U_i \end{cases} \tag{3.5}$$

For instance, if the desirability value is equal to 1, it means that the value of $\hat{y}$ will be equal to T. However, if the desirability value equals to 0, it means that the value of $\hat{y}$ exceeds the range of target requirement. This scenario indicates the worst case for the nominal-the-better function.

The *r* value used in equation (3.3), equation (3.4) and equation (3.5) is defined according to the requirement of the analyst. If the corresponding response is expected to be closer to the target, the weight can be set to a larger value, otherwise, the weight can be set to a smaller value.

Then, the individual desirability values are combined into an overall desirability function, D where $0 \leq D \leq 1$. The overall desirability function can be calculated as shown in equation (3.6).

$$D = (d_1.d_2........d_m)^{\frac{1}{m}} \tag{3.6}$$

After all the value is obtained, the predicted optimum conditions need to be calculated. Once the optimal level of the designed parameters has been selected, the final step is to predict and verify the quality characteristics using the optimal level of the designed parameters.

3.2.3 Modified Desirability Function: Modified Geometric Mean (MGM)

The ordinary least square (OLS) method is often used to estimate the parameters of a second order polynomial RSM model. According to Habshah, Mohd and Anwar [26], the OLS method gives good parameter estimates when the responses are normally distributed and no outliers in the data sets. Nevertheless, in real situations many distributions of exploratory variables are considered not normal due to the presence of outliers. Outliers arise in many different forms and due to many various reasons as stated in Simpson and Montgomery [27]. In addition, according to Yohai [28], a small fraction of outlier or one outlier may have significant effects on the OLS estimates.

Outliers can lead to misinterpretations of the regression result. This is because the presence of outliers can pull the regression line towards themselves, which can make the solution more accurate for the outliers but less accurate for other cases in the data set. Subsequently, the determination of the optimum response is not reliable because it is based on the OLS which is not resistance to outliers. The outliers can wrongly show optimum responses which are not reliable and may produce inefficient results.

The overall desirability function introduced by Derringer and Suich [2] is as follows:

$$D = (d_1.d_2........d_m)^{\frac{1}{m}} \tag{3.7}$$

However, this function can affect the optimum solution by the presence of outliers. Hence, the Modified Geometric Mean (MGM) is introduced to remedy this problem because the MGM is resistant to outliers.

Firstly, for a set of positive observation, the Geometric Mean (GM) can be defined as:

$$GM = \sqrt[m]{d_1.d_2...d_m} \tag{3.8}$$

Then, by taking the logarithm on both sides, equation 3.8 can be written as

$$\log GM = \frac{1}{m}\sum_{i=1}^{m} \log d_i \tag{3.9}$$

Next, conventionally, change the logarithm into natural logarithm with any arbitrary base. Thus, equation 3.9 can be shown as

$$ln GM = \frac{1}{m} \sum_{i=1}^{m} ln\, d_i \tag{3.10}$$

Thus, to modify the formula in equation 3.10, a new formula to find the MGM can be expressed as

$$ln MGM = \{median \lfloor ln(d_i) \rfloor\} \tag{3.11}$$

Therefore, the MGM should be as follows:

$$MGM = exp\{median \lfloor ln(d_i) \rfloor\} \tag{3.12}$$

Hence, in cases where the geometric mean is known as not resistant to outliers, the Modified geometric mean is proposed because it uses median instead of mean and median is resistant to outliers. The MGM approach also takes significant variables only into consideration to increase the accuracy of the model.

On top of that, Classical Desirability function and Modified Desirability function approaches will be analysed by using R-package software. Since an appropriate coding transformation is a crucial step in response surface analysis, this coding method can make all coded variables in the experiment vary over the same range in order to get the optimal setting. Besides, R-package is the most widely used as statistical analysis because it considers all standard statistical tests, models and analyses as well as the language in analysing and manipulating data.

4 Data Analysis

4.1 Microwave-assisted Extraction of Saikosaponins

The optimization of microwave-assisted extraction (MAE) conditions in 30 analyzed runs was obtained from a test conducted by Hu, Cai and Liang [12]. Four operating component settings or explanatory variables were considered and they are microwave power (x_1), irradiated time (x_2), extraction temperature (x_3) and ethanol concentration (x_4). The purpose was to boost the response variable of the extraction yield, which were the saikosaponin a (y_1), saikosaponin c (y_2) and saikosaponin d (y_3) by using a central composite rotatable design (CCRD). A second-order polynomial model was fitted to each response variable:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_1 x_1 + \sum_{i=1}^{k} \beta_{ii} x_i^2 + . \sum_{i<j} \sum \beta_{ij} x_i x_j + \varepsilon \tag{4.1}$$

In order to boost the response variable, the minimum and maximum values of the responses which are $[y_i^{min}, y_i^{max}] = [93{,}100]$ was chosen for the optimization and $r = 0.3$ as imposed by Hu et al. [12] for all three individual desirability functions in equation (3.3). Hu et al. [12] stated that the value of the constant $r = 0.3$ was chosen as in practice, 100% yield is difficult to be obtained, thus it would be desirable if the yield deviates moderately from y_i^{min}. The outliers were introduced to observation 28 in saikosaponin a (y_1). The value of responses was 93.78 and changed to 127.23 when the outliers were present. The usual analysis was done using the coded values. For each of the independent variable, the coded value level was taken from -2 to 2 with a 0.2 interval and then the data was analysed by using the Harrington's desirability-OLS and Modified Geometric Mean (MGM). This analysis method was applied to the data with and without outlier values. The results are displayed in the following tables:

a) Data without outliers

Table 4.1: The optimum responses for Microwave-assisted Extraction using the OLS-based method with no outlier data.

Optimization results	d_i	$\ln(d_i)$
$\hat{y}_1 = 93.5867 + 0.6875\beta_1 + 1.8875\beta_2 + 0.9233\beta_3 - 1.1742\beta_4$ (0.5511) (0.2755) (0.2755) (0.2755) (0.2755) $-0.7279\beta_5 - 0.6767\beta_6 - 0.7254\beta_7 - 2.1867\beta_8 + 0.3062\beta_9$ (0.2577) (0.2577) (0.2577) (0.2577) (0.3375) $+0.7662\beta_{10} - 0.1463\beta_{11} + 0.1125\beta_{12} - 0.2750\beta_{13} + 0.1850\beta_{14}$ (0.3375) (0.3375) (0.3375) (0.3375) (0.3375)	0.7831	
$\hat{y}_2 = 92.8767 + 1.8046\beta_1 + 2.5021\beta_2 + 1.5004\beta_3 - 0.1596\beta_4$ (0.9764) (0.4882) (0.4882) (0.4882) (0.4882) $-0.2736\beta_5 - 0.4674\beta_6 - 0.6986\beta_7 - 4.1461\beta_8 - 1.0331\beta_9$ (0.4567) (0.4567) (0.4567) (0.4567) (0.5979) $+0.4506\beta_{10} - 0.1394\beta_{11} - 0.5644\beta_{12} + 0.3056\beta_{13} - 0.1406\beta_{14}$ (0.5979) (0.5979) (0.5979) (0.5979) (0.5979)	0.7679	
$\hat{y}_3 = 93.8383 + 0.7067\beta_1 + 2.4308\beta_2 + 1.4058\beta_3 - 0.6583\beta_4$ (0.6041) (0.3020) (0.3020) (0.3020) (0.3020) $-0.3463\beta_5 - 0.1112\beta_6 - 0.8175\beta_7 - 3.0288\beta_8 - 0.2825\beta_9$ (0.2825) (0.2825) (0.2825) (0.2825) (0.3699) $+0.1987\beta_{10} + 0.2037\beta_{11} - 0.9300\beta_{12} + 0.1650\beta_{13} - 0.3763\beta_{14}$ (0.3699) (0.3699) (0.3699) (0.3699 (0.3699)	0.8718	
Harrington's desirability $x^* = [0.8, 1.8, 0.6, 0]$ $D = 0.8063$		

**The value in the brackets ( ) indicates standard error.

Table 4.2: The optimum responses for Microwave-assisted Extraction using the OLS-based method with no outlier data using significant variables.

Optimization results	d_i	$\ln(d_i)$
$\hat{y}_1 = 93.5867 + 0.6875\beta_1 + 1.8875\beta_2 + 0.9233\beta_3 - 1.1742\beta_4$ (0.5094) (0.2547) (0.2547) (0.2547) (0.2547) $-0.7279\beta_5 - 0.6767\beta_6 - 0.7254\beta_7 - 2.1867\beta_8 + 0.7662\beta_9$ (0.2382) (0.2382) (0.2382) (0.2382) (0.3119)	0.7317	-0.3124
$\hat{y}_2 = 91.5969 + 1.8046\beta_1 + 2.5021\beta_2 + 1.5004\beta_3 - 3.9862\beta_4$ (0.5408) (0.4683) (0.4683) (0.4683 (0.4275)	1	0
$\hat{y}_3 = 93.3808 + 0.7067\beta_1 + 2.4308\beta_2 + 1.4058\beta_3 - 0.6583\beta_4$ (0.3965) (0.2804) (0.2804) (0.2804) (0.2804) $-0.7603\beta_5 - 2.9716\beta_6 - 0.9300\beta_7$ (0.2575) (0.2575) (0.3434)	0.9248	-0.07818
Modified Geometric Mean $x^* = [1.4, 2.0, 0.8, -0.2]$ $D(MGM) = 0.9248$ $D = 0.8779$		

**The value in the brackets ( ) indicates standard error.

b) Data with outliers

Table 4.3: The optimum responses for Microwave-assisted Extraction using the OLS-based method with outlier data.

Optimization results	d_i	$\ln(d_i)$
$\hat{y}_1 = 99.1617 + 0.6875\beta_1 + 1.8875\beta_2 + 0.9233\beta_3 - 1.1742\beta_4$ (3.2875) (1.6437) (1.6437) (1.6437) (1.6437) $- 2.1217\beta_5 - 2.0704\beta_6 - 2.1192\beta_7 - 3.5804\beta_8 + 0.3063\beta_9$ (1.5376) (1.5376) (1.5376) (1.5376) (2.0132) $+ 0.7663\beta_{10} - 0.1462\beta_{11} + 0.1125\beta_{12} - 0.2750\beta_{13} + 0.1850\beta_{14}$ (2.0132) (2.0132) (2.0132) (2.0132) (2.0132) (0.2575) (0.2575) (0.3434)	0.9344	
$\hat{y}_2 = 92.8767 + 1.8046\beta_1 + 2.5021\beta_2 + 1.5004\beta_3 - 0.1596\beta_4$ (0.9764) (0.4882) (0.4882) (0.4882) (0.4882) $- 0.2736\beta_5 - 0.4674\beta_6 - 0.6986\beta_7 - 4.1461\beta_8 - 1.0331\beta_9$ (0.4567) (0.4567) (0.4567) (0.4567) (0.5979) $+ 0.4506\beta_{10} - 0.1394\beta_{11} - 0.5644\beta_{12} + 0.3056\beta_{13} - 0.1406\beta_{14}$ (0.5979) (0.5979) (0.5979) (0.5979) (0.5979)	0.7574	
$\hat{y}_3 = 93.8383 + 0.7067\beta_1 + 2.4308\beta_2 + 1.4058\beta_3 - 0.6583\beta_4$ (0.6041) (0.3020) (0.3020) (0.3020) (0.3020) $- 0.3463\beta_5 - 0.1112\beta_6 - 0.8175\beta_7 - 3.0288\beta_8 - 0.2825\beta_9$ (0.2825) (0.2825) (0.2825) (0.2825) (0.3699) $+ 0.1987\beta_{10} + 0.2037\beta_{11} - 0.9300\beta_{12} + 0.1650\beta_{13} - 0.3763\beta_{14}$ (0.3699) (0.3699) (0.3699) (0.3699 (0.3699) (0.5979) (0.5979) (0.5979) (0.5979) (0.5979)	0.8275	
Harrington's desirability $x^* = [0.6, 1.2, 0.4, 0]$ $D = 0.8366$		

**The value in the brackets ( ) indicates standard error.

Table 4.4: The optimum responses for Microwave-assisted Extraction using the OLS-based method with outlier data using significant variables.

Optimization results	d_i	$\ln(d_i)$
$\hat{y}_1 = 93.5520 - 2.8790\beta_1$ (1.6780) (1.3270)	0.4666	-0.7623
$\hat{y}_2 = 91.5969 + 1.8046\beta_1 + 2.5021\beta_2 + 1.5004\beta_3 - 3.9862\beta_4$ (0.5408) (0.4683) (0.4683) (0.4683 (0.4275)	1	0
$\hat{y}_3 = 93.3808 + 0.7067\beta_1 + 2.4308\beta_2 + 1.4058\beta_3 - 0.6583\beta_4$ (0.3965) (0.2804) (0.2804) (0.2804) (0.2804) $-0.7603\beta_5 - 2.9716\beta_6 - 0.9300\beta_7$ (0.2575) (0.2575) (0.3434)	0.9850	-0.01511
Modified Geometric Mean $x^* = (2,2,0,0)$ $D(MGM) = 0.9850$ $D = 0.7717$		

**The value in the brackets ( ) indicates standard error.

For the data without outliers and with the use of Harrington's desirability function-OLS approach, the optimal factor setting obtained was $(x_1, x_2, x_3, x_4) = (0.8, 1.8, 0.6, 0)$ and the individual desirability function $(d_1, d_2, d_3) = (0.7831, 0.7679, 0.8718)$ with $D = 0.8063$. Meanwhile, the proposed MGM approach shows that the results obtained $(x_1, x_2, x_3, x_4) = (1.4, 2.0, 0.8, -0.2)$ and the desirability function $(d_1, d_2, d_3) = (0.7317, 1, 0.9248)$ with $D(MGM) = 0.9248$. In order to get a genuine result, the D value of Harrington's desirability function-OLS approach is compared between the D value of Harrington's desirability function and MGM approach for significant variables. Therefore, from this result, MGM approach is better compared with Harrington's desirability function-OLS approach.

Next, for the data with outliers and with the use of Harrington's desirability function-OLS approach, the optimal factor setting obtained was $(x_1, x_2, x_3, x_4) = (0.6, 1.2, 0.4, 0)$ and the individual desirability function $(d_1, d_2, d_3) = (0.9344, 0.7574, 0.8275)$ with $D = 0.8366$. However, the proposed MGM approach shows that the results obtained $(x_1, x_2, x_3, x_4) = (2, 2, 0, 0)$ and the desirability function $(d_1, d_2, d_3) = (0.4666, 1, 0.9850)$ with $D(MGM) = 0.9850$. Thus, it shows that MGM approach is better as compared to Harrington's desirability function-OLS approach. In addition, the standard error of the MGM approach is smaller.

5 Conclusion

The weakness of desirability function approach is that the variability of each predicted response can be generated although it is not explicitly included in the optimization procedure. In the optimal factor, this variability can influence the range of the prediction interval for each response. Thus, to obtain each accurate prediction, it is advisable to minimize this variability. At first, the result is obtained by using Harrington's desirability function approach. However, this approach can affect the optimum solution by the presence of outliers. Hence, an outlier resistance approach based on Modified Geometric Median was proposed as the approach is resistant to outliers. In the numerical application, it can be clearly seen that the desirability function based on MGM approach is more efficient, more desirable and more practical in reducing the standard error of the predicted responses. Based on the value of the overall desirability function, D, it clearly shows that the approach based on MGM is better since the value of D is larger and the standard error of the MGM approach is smaller as compared to Harrington's desirability function,

Thereupon, the result shows that the proposed method which is the MGM approach has produced smaller standard error and larger desirability value. Consequently, the MGM approach can be an alternative method in dealing with the presence of outliers.

References

[1] Montgomery, D. C. (2009). *Design and Analysis of Experiments*. 7th Edition. John Wiley and Sons.
[2] Derringer, G., & Suich R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology*, *12*(4), 214-219.
[3] Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces* (Vol. 424). New York: Wiley.
[4] Myers, R. H., & Montgomery, D. C. (2001). Response surface methodology, (2002). *Willey, New York*.
[5] Khuri, A. I., & Mukhopadhyay, S. (2010). Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(2), 128-149.
[6] Malik, N. R. (2016). Optimization of Process Parameters in Extraction of Thyme Oil Using Response Surface Methodology (RSM). *Science, Engineering and Technology*, 16.
[7] Oehlert, G. W. (2000). Design and analysis of experiments: Response surface design. *New York*.
[8] Khuri, A. I., & Cornell, J. A. (1996). *Response surfaces: designs and analyses* (Vol. 152). CRC press.
[9] Lind, E. E., Goldin, J., & Hickman, J. B. (1960). Fitting yield and cost response surfaces. *Chemical Engineering Progress*, *56* (11), 62-68.
[10] Harrington, E. C. (1965). The desirability function. *Industrial quality control*, *21*(10), 494-498.
[11] Yusof, S., Talib, Z., Mohamed, S., & Bakar, A. A. (1988). Use of response surface methodology in the development of guava concentrate. *Journal of the Science of Food and Agriculture*, *43*(2), 173-186.
[12] Hu, Z., Cai, M., & Liang, H. H. (2008). Desirability function approach for the optimization of microwave-assisted extraction of saikosaponins from Radix Bupleuri. Separation and Purification Technology, 61(3), 266-275.
[13] Islam, M. A., Alam, M. R., & Hannan, M. O. (2012). Multiresponse optimization based on statistical response surface methodology and desirability function for the production of particleboard. *Composites Part B: Engineering*, *43*(3), 861-868.
[14] Fitrianto, A., & Midi, H. (2012). Multi-Response Optimization via Desirability Function for the Black Liquor DATA. *Journal of Science and Technology, 4*(1).
[15] Torrades, F., Saiz, S., & García-Hortal, J. A. (2011). Using central composite experimental design to optimize the degradation of black liquor by Fenton reagent. *Desalination*, *268*(1), 97-102.
[16] Maran, J. P., Manikandan, S., & Mekala, V. (2013). Modeling and optimization of betalain extraction from Opuntia ficus-indica using Box–Behnken design with desirability function. *Industrial crops and products*, *49*, 304-311.
[17] Rafieian, F., Keramat, J., & Kadivar, M. (2013). Optimization of gelatin extraction from chicken deboner residue using RSM method. *Journal of Food Science and Technology*, *50*(2), 374-380.

[18] Chen, H. W., Wong, W. K., & Xu, H. (2012). An augmented approach to the desirability function. *Journal of Applied Statistics*, *39*(3), 599-613.

[19] Atkinson, A. C. (1994). Fast very robust methods for the detection of multiple outliers. *Journal of the American Statistical Association*, *89*(428), 1329-1339.

[20] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2001). *Introduction to linear regression analysis*. 3rd edition. John Wiley & Sons.

[21] Rocke, D. M., & Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association*, *91*(435), 1047-1061.

[22] Srikantan, K. S. (1961). Testing for the single outlier in a regression model. *Sankhyā: The Indian Journal of Statistics, Series A*, 251-260.

[23] Ellenberg, J. H. (1976). Testing for a single outlier from a general line regression. *Biometrics*, 637-645.

[24] Moutinho, L., & Hutcheson, G. D. (Eds.). (2011). *The SAGE dictionary of quantitative management research*. Sage.

[25] Montgomery, D. C. (2014). *Design and analysis of experiments*. John Wiley & Sons.

[26] Habshah, M., Mohd, S. M., Anwar, F. (2012). The Performance of Optimum Response Surface Methodology Based on MM-Estimator. *International Journal of Mathematical Models and Methods in Applied Sciences*. 6(6), 757-764.

[27] Simpson, J. R., & Montgomery, D. C. (1998). A performance-based assessment of robust regression methods. *Communications in Statistics-Simulation and Computation*, *27*(4), 1031-1049.

[28] Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 642-656.