Overview on the Response Surface Methodology (RSM) in Extraction Processes

Khairul Anwar Mohamad Said\textsuperscript{1} and Mohamed Afizal Mohamed Amin\textsuperscript{2}

\textsuperscript{1,2}Department of Chemical Engineering and Energy Sustainability, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia.
\textsuperscript{1}Email: mskkanwar@feng.unimas.my, \textsuperscript{2}Email: mamafizal@feng.unimas.my

\section*{Abstract}
Response Surface Methodology (RSM) is an optimization tool that can identify interrelationship between variables as being adopted by experiment/research studies in food and herbal plants extraction niche area. This review discusses the optimization approach through utilization of research surface methodology either using central composite design or Box-Behnken method specifically in extraction processes. The use of analysis of variance (ANOVA) to evaluate the degree of accuracy held by the derived model is based on several responses. RSM helps to determine the best experimental design in order to identify the relationship between variables. This paper also discusses on the utilization of RSM to derive a model equation that later can be applied for response prediction and the determination of optimal conditions.

\textbf{Keywords:} RSM, response surface methodology, box-behnken, central composite design, regression.

\section*{1. Introduction}
Response surface methodology (RSM) can be described as a technique that involves complex calculation for optimization process. This approach develops a suitable experimental design that integrates all of the independent variables and uses the data input from the experiment to finally come up with a set of equations that can give theoretical value of an output. The outputs are obtained from a well-designed regression analysis that is based on the controlled values of independent variables. Thereafter, the dependent variable can be predicted based on the new values of independent variables \cite{4,5}. Back in 1951 when RSM was first introduced by Box and Wilson \cite{6}, the experimental runs were reduced enormously compared to the number of runs determined using full factorial design. Therefore, it has been adapted in many research including food technology where the technical steps can be found explained in details such that by Henika \cite{1,2} and Giovanni \cite{3}. Beside the reduction of experimental runs, the results obtained from RSM are claimed to be statistically acceptable \cite{7}. By applying RSM method in the optimization process, only a short period of time is required to test all of the variables pertaining to the consumer evaluation, making the laboratory test stage more efficient \cite{8}. In addition, parameters estimation can identify the variables that are largely affecting the model which then helps researcher to focus on those particular variables that contribute to the product acceptance \cite{9}.

Generally, one factor or process variable can depend on or be depended by another variable in a set of experimental design. The knowledge of the interaction between the factors is crucial in order to find the output-input relationship. This is the reason that the interactions are hardly determined using one-factor-at-a-time approach \cite{10}. By establishing a model equation, RSM can evaluate the
relationship as well as interactions among the multiple parameters using quantitative data. There are three steps in RSM implementation; (1) design of experiment i.e Box Behnken and Central Composite Design (CCD); (2) statistical and regression analysis to develop model equations that represent the response surface modeling; and (3) parameters/variables optimization carried out through model equation [11].

As discussed earlier, RSM has been applied in various experimental designs involving extraction process [12], food preservation [13][14], fermentation [15] as well as other discipline of engineering. For example, Lui et al. [16] has chosen to extract anthocyanin from purple sweet potato according to four-factor parameters which were temperature, ratio of ethanol to ammonium sulfate, time of extraction and pH value. They designed the experiment using Box Behnken approach where there were 27 runs consisted of 24 factorial experiments and 3 repeated runs at the center point. The experimental results were fitted to polynomial equation specifically second order regression whereas the significance of the derived model equation was determined using the analysis of variance (ANOVA) that evaluates the goodness-of-fit of the regression model and the significance of each parameter that affects the model. Finally, the result validation is as important as the ANOVA where the predicted model is subjected for comparison with actual experimental value. Lui et al. [16] found that the actual experimental values were almost identical to the theoretical values derived using the model equation. This paper aims to introduce researcher with alternative method of analyzing data and provide general ideas to two important methods of optimization; central composite design and Box-Behnken. Moreover, both of these methods are exclusive and provide specific design of experiment to address the different approaches in analyzing data.

2. Optimization using central composite design (CCD)

Response surface methodology comprised of several methods to design the experimental procedures and one of them is Central Composite Design (CCD). Optimization carried out with CCD can allow screening of a broad range of parameters as well as the role of each factor [17]. In addition, CCD is also able to evaluate a single variable or the cumulative effect of the variables to the response. Although this ability is shared with the other types of experimental design such as full factorial and partial factorial method, it differs in a way that the experimental runs are reduced. For instance, with just four independent variables, full factorial method will suggest at least 81 experimental runs plus replication [6]. Otherwise when using CCD method, only 31 experimental points (16 factorial points, 8 axial points and 7 center points) are needed [18].

Before any variables can be carried into experimental phase, the variables must be coded according to (1);

$$x_i = (X_i - \bar{X}_i)/\Delta X_i$$  \hspace{1cm} (1)

where $x_i$ is the coded level; $X_i$ is the natural level for the independent variable; $\bar{X}_i$ is the mean for the natural level of the independent variables; and $\Delta X_i$ is the step change value. In optimization of bioactive alkaloid from rhizome coptidis, Teng and Choi [19] decided to use CCD where three variables were investigated namely ethanol concentration, extraction time and extraction temperature. Table 1 shows the coded and natural variables where the codes are represented by -1, 0 and 1 whereas natural variables are the actual value for the coded variables.
Table 1. An example of a partial experimental design in coded term [19]

<table>
<thead>
<tr>
<th>Test runs</th>
<th>Coded variable level</th>
<th>Natural variable levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X₁(%)</td>
<td>X₂(min)</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>-1.682</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1.682</td>
<td>0</td>
</tr>
</tbody>
</table>

2.1. Regression

All responses received after several experimental procedures are analyzed using regression method which involves fitting the response into polynomial model, for instance the second order polynomial model presented in (2):

\[ y = \beta_0 + \sum_{i=1}^{4} \beta_i X_i + \sum_{i=1}^{4} \sum_{j=1}^{4} \beta_{ij} X_i X_j + \sum_{i=1}^{4} \beta_{ii} X_i^2 \]  

(2)

where \( y \) represents the independent responses; \( \beta_0, \beta_i, \beta_{ii} \) and \( \beta_{ij} \) represent the regression coefficient of the process variables for the intercept, linear, quadratic, and cross product terms, respectively [20]. For the reliability test, analysis of variance (ANOVA) must be carried out where it evaluates the fitness of the model through the response by coefficient of R square and also F-test for lack of fit.

2.2. Current optimization research using CCD

The responses obtained from extraction experiments are usually the antioxidant activity, and the yield. However, the factor affecting the response involves several variables which resulted a high number of experimental runs to test all the possible sequences. In a recent study by Yolmeh et al. (2014) to optimize an extraction of essential oil from Annato seed, four distinct variables were studied; temperature (20–80 °C), sonication time (2–10 min), duty cycle (0.2–0.8s) and the ratio of seeds to the solvent (5–20%) in which two models were presented based on the extraction yield and antioxidant activity as described in (3) and (4) [21]:

\[ \text{Extraction yield} = -9.535 + 0.091X_1 + 0.635X_2 + 131.547X_4 - 0X_1^2 - 0.021X_2^2 - 337.587X_4^2 - 0.357X_1X_4 - 1.408X_2X_4 \]  

(3)
Antioxidant activity = 0.0053X_1 + 0.0442X_2 + 6.6342X_4 - 0.0020X_2^2 + 0.4161X_3^2 - 
16.0090X_4^2 \tag{4}

where X1, X2, X3 and X4 are referring to the temperature, sonication time, duty cycle and seed-to solvent ratio respectively. The extraction yield were calculated based on the weight difference between annatto seed and post-extraction annatto powder whereas antioxidant activity were quantified through absorbance test of reacted DPPH (1,1-diphenyl-2-picylhydrazyl) assay in extracted compound. In order to evaluate the significance of both models, analysis of variance (ANOVA) has been performed which indicated that duty cycle (X3) as not a significant variable in determining the extraction yield and the highest effect to extraction yield were from seed-to-solvent ratio (X4) and quadratic term of seed-to-solvent ratio (X_3^2). The highest effect among the variables in contributing to the extraction yield can be identified using P-value (<0.05). In contrast, antioxidant activity model showed that all of the variables are significant in estimating the predicted value except for duty cycle (X3) that showed insignificant effect on the model. However, the effect of duty cycle (X3) has been replaced by the quadratic term of duty cycle (X_3^2) where it showed moderate impact on the model.

A study on extraction of bioactive compound from Nephelium lappaceum L. fruit peel using central composite face-centered response surface designed by Prakash Maran et al. [20] explored in detail the usage of CCD as a tool for optimization. They listed all possible models that could represent the responses (total phenolic content, total anthocyanin content and total flavonoid content) and compared them in terms of R2, R, p-value, mean square and F-value. As a result, all responses were best represented in a quadratic model which showed the least p-value at <0.0001 as well as value of R2 closest to 1. The models are presented as follows [20].

\[ \text{TAC} = 7.06 + 1.91X_1-0.2X_2-0.018X_3 + 0.11X_4 + 0.085X_1X_2-0.051X_1X_3-0.065X_1X_4 + 0.41X_2X_3-0.78X_2X_4+0.65X_3X_4 + 0.92X_2^2 + 8.772E-003X_2^2-2.04X_3^2 + 0.2X_4^2 \tag{5} \]

\[ \text{TPC} = 374.91 + 102.41X_1 - 10.87X_2 - 1.03X_3 + 6.88X_4 + 3.33X_1X_2 - 2.16X_1X_3 - 3.77X_1X_4 + 18.61X_2X_3 - 41.47X_2X_4 - 34.14X_3X_4 + 52.24X_1^2 + 1.58X_2^2 - \]

\[ 111.67X_3^2 + 11.45X_4^2 \tag{6} \]

\[ \text{TFC} = 70.66 + 18.78X_1 - 1.43X_2 - 0.24X_3 + 0.93X_4 + 0.91X_1X_2 - 0.7X_1X_3 - 1.21X_1X_4 + 3.64X_2X_3 - 7.57X_2X_4 - 6.84X_3X_4 + 9.41X_1^2 + 0.26X_2^2 - 2.019X_3^2 + 2.21X_4^2 \tag{7} \]

In order to verify the reliability of the derived model, the predicted response must be compared with the experimental result at fixed variables condition. Prakash Maran et al. tested the model at 50 oC extraction temperature, 20 W ultrasonic power, 20 min extraction time and 1:18.6 w/v solid-to-liquid ratio while the responses were; total phenolic content, total anthocyanin content and total flavonoid content. The comparison on the results are shown in Table 2.
Table 2. Central composite design predicted and experimental values for the responses [20]

<table>
<thead>
<tr>
<th>Process parameter</th>
<th>Predicted values (model)</th>
<th>Experimental values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TAC (mg/100g)</td>
<td>TPC (mgGAE/100g)</td>
</tr>
<tr>
<td></td>
<td>TFC (mgRE/100g)</td>
<td>TAC (mg/100g)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TPC (mgGAE/100g)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TFC (mg/100g)</td>
</tr>
<tr>
<td>X1 (50°C)</td>
<td>10.17</td>
<td>10.26±0.39</td>
</tr>
<tr>
<td>X2 (20W)</td>
<td>546.98</td>
<td>552.64±1.57</td>
</tr>
<tr>
<td>X3 (20min)</td>
<td>100.93</td>
<td>104±1.13</td>
</tr>
<tr>
<td>X4 (1:18.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Optimization using Box Behken (BB)

Box-Behnken (BB) is another method in response surface methodology where the final outcome is to find the optimal variables to produce optimum response/ output. According to Maran et al. (2013), Box-Behnken design is considered as a design without the embedded factorial or fractional factorial point that could be identified as the variable condition located at the midpoint edges of the variables space as well as at the center [22]. The number of experiments using Box Behnken method can be calculated by:

\[ N = 2k(k - 1) + C_0 \]  

where \( k \) is the factorial number; whereas \( C_0 \) is the replicate number of the central point [23][24][25]. In addition, the central point actually helps to estimate the pure error as well as enabling intermediate levels calculation of the response function. Hence, a system performance estimation within the studied range is possible following the replication of the central point [26].

3.1. Regression

Linear or multiple regressions are regularly carried out as the last stage to study the interaction between variables. This useful information can be obtained through a polynomial equation that can either be a combination of the first or the second order with analysis of variance (ANOVA) to provide statistical measurement. Roosta et al. (2014) studied the interaction between four variables and use the following polynomial relation as a medium of interaction [27]:

\[ y = \beta_0 + \sum_{i=1}^{4} \beta_i X_i + \sum_{i=1}^{4} \sum_{j=1}^{4} \beta_{ij} X_i X_j + \sum_{i=1}^{4} \beta_{ii} X_i^2 \]  

where \( y \) is the predicted response quantified by an extraction recovery; \( X_i \) represents the independent variables (four variables); \( \beta_0 \) is model constant; \( \beta_i \) is the linear coefficient; \( \beta_{ii} \) is the quadratic coefficient; and \( \beta_{ij} \) is the cross-product coefficient.

Another study by Elksibi et al. (2014) investigated the effect of five variables mainly sodium hydroxide concentration, extraction time, temperature, and mass of the waste on total phenolic content and relative color strength. In order to analyze the interaction between the variables, they adopted the same polynomial model as Roosta et al [10]. This has proven the reliability of the polynomial model to study interaction between four variables.

Khajeh (2011) in experimenting on three variables to study the effect of pH, concentration of dithizone and solvent volume on the lead extraction from food samples used the same polynomial model that has been expanded as shown in (10). According to the number of variables showing the applicability of the model on studies that have less than four variables [28];
\[ Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_{12} AB + \beta_{13} AC + \beta_{23} AC + \beta_{11} A^2 + \beta_{22} B^2 + \beta_{33} C^2 \quad (10) \]

### 3.2. Current optimization research using BB

Recently, Aybastier et al. (2013) completed an optimization study on four variables to extract antioxidant out of blackberry leaves. The optimization process were carried out using Box-Behken method where 30 experiments were employed based on four variables consisting of hydrochloric acid and methanol concentration as well as extraction temperature and time [29]. Based on the polynomial model, \( y \) represents a response or output of the study. Hence, three responses in this study were total phenolic content (TPC), ABTS to analyze antioxidant activity and finally CUPRAC to study cupric ion reducing antioxidant activity. As mentioned previously, \( Y \) only accounts to one response which tells us that this study has three different models and can be individually evaluated using analysis of variance (ANOVA). The following equations are the polynomial model based on each response:

\[
Y_1 = 68.31 - 2.23x_1 + 3.82x_2 + 11.22x_3 + 10.55x_4 - 10.26x_2^2 - 6.53x_4^2 \quad (11)
\]

\[
Y_2 = 80.85 - 5.12x_1 + 8.11x_3 + 8.87x_4 + 4.61x_1x_3 + 3.55x_2x_3 - 6.01x_3x_4 - 8.23x_2^2 - 4.68x_4^2 \quad (12)
\]

\[
Y_3 = 285.00 - 15.91x_1 + 55.93x_3 + 78.11x_4 - 24.94x_2x_3 - 22.52x_2x_4 - 56.71x_2^2 - 26.92x_4^2 \quad (13)
\]

where \( Y_1 \), \( Y_2 \) and \( Y_3 \) are total phenolic content, ABTS value and CUPRAC respectively. All listed equations have different variable interaction. For instance, response \( y_1 \) has no cross product interaction while response \( Y_2 \) and \( Y_3 \) both have the interaction. It is not necessary for every model to have the same interaction as the non-significant interaction can be eliminated based on the analysis of variance to increase the accuracy of the response prediction. Thereafter, all models must undergo reliability test to measure the accuracy of the response which took into account for the differences between the predicted and the experimental values. The predicted values can be calculated using the derived model while the experimental values are obtained as average values from an experiment performed in at least triplicates with assigned values of independent variables. Aybastier et al. (2013), compared the values predicted using equations (11) to (13) with the experimental values as shown in Table 3 and found that the deviation was only in the range of 0.2 – 3 %.
Table 3. Box-Behnken prediction and experimental values of the responses [29]

<table>
<thead>
<tr>
<th>Process parameter</th>
<th>Predicted values (model)</th>
<th>Experimental values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total phenolic content</td>
<td>ABTS</td>
</tr>
<tr>
<td></td>
<td>(mg GAE/g dried plant)</td>
<td>(mg TE/g dried plant)</td>
</tr>
<tr>
<td>X₁ (0.45M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₂ (64% v/v)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₃ (67°C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₄ (112min)</td>
<td>82.12</td>
<td>84.80±2.69</td>
</tr>
<tr>
<td>X₁ (0.42M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₂ (64% v/v)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₃ (66°C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₄ (105min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₁ (0.41M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₂ (61% v/v)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₃ (68°C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₄ (117min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Similarly, Tian et al. (2013) investigated the interaction between four variables in producing the highest fatty acids methyl ester. In order to optimize the extraction, they utilized Box-Behnken to design the experiment as well as to analyze the interaction between the ultrasonic power, extraction temperature, time and solvent/seed ratio with a total of 29 experiments and 5 replicates at the center point [30]. Based on the polynomial model, they produced the following model based on one response - fatty acids methyl ester:

\[ Y = 23.59 - 1.52x₁ + 0.18x₂ + 0.087x₃ - 1.00x₄ - 0.14x₁x₂ - 0.51x₁x₃ - 0.17x₁x₄ - 0.27x₂x₃ - 0.36x₂x₄ + 0.65x₃x₄ - 0.0598x₁x₃ - 2.03x₂x₄ - 0.55x₃x₃ - 0.55x₄x₄ \] (14)

where \( x_n \) refers to the different individual variables throughout the experiment. In a statistical point of view, \( R \) and \( R^2 \) values of the model were reported to be 0.9987 and 0.9975 respectively which indicate a good agreement between the experimented and predicted results evaluated using the model.
4. Conclusion

The discussion presented here has clearly established the importance of choosing the right optimization tools such as response surface methodology (RSM). Major achievement of response surface methodology compared to conventional methods is the reduction of experimental runs for the same objective which is to obtain optimal variables condition/value for the highest output/response. Besides, the model derived can be used to predict the response prior to experimentation phase. Moreover, this step can help researcher or industries to focus on certain variables/aspect that contribute to the highest effect on process output. The use of either central composite design or Box-Behnken for extraction purposes especially with high-cost raw material is an economical alternative to traditional optimization approach through one-factor method. Combining extraction process with RSM can cause significant degree of accuracy in model prediction.

Acknowledgement

The author gratefully acknowledges Ministry of Education Malaysia for financial support through Research Acculturation Grant Scheme (RAGS) no: RAGS/c(9)/943/2012(44).

References


Authors

Khairul Anwar Mohamad Said

Khairul Anwar Mohamad Said is a lecturer from the Department of Chemical Engineering and Energy Sustainability, Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS). He obtained his Masters of Engineering from Universiti Teknologi Malaysia (UTM) and his research interest is separation process including extraction and membrane.

Mohamed Afizal Mohamed Amin

Mohamed Afizal Mohamed Amin is a lecturer from the Department of Chemical Engineering and Energy Sustainability, Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS). He obtained his Masters of Engineering from Universiti Teknologi Malaysia (UTM) and his research interest is membrane and fuel cell technology.