



ARTICLE

## Optimizing Food Processing through a New Approach to Response Surface Methodology

Sungsue Rheem\*

Division of Big Data Science, Korea University, Sejong 30019, Korea

OPEN ACCESS

Received January 28, 2023

Revised February 8, 2023

Accepted February 8, 2023

\*Corresponding author : Sungsue Rheem  
Division of Big Data Science,  
Korea University, Sejong 30019, Korea  
Tel: +82-44-860-1550  
Fax: +82-44-860-1551  
E-mail: [rheem@korea.ac.kr](mailto:rheem@korea.ac.kr)

\*ORCID  
Sungsue Rheem  
<https://orcid.org/0000-0001-7009-3343>

**Abstract** In a previous study, ‘response surface methodology (RSM) using a fullest balanced model’ was proposed to improve the optimization of food processing when a standard second-order model has a significant lack of fit. However, that methodology can be used when each factor of the experimental design has five levels. In response surface experiments for optimization, not only five-level designs, but also three-level designs are used. Therefore, the present study aimed to improve the optimization of food processing when the experimental factors have three levels through a new approach to RSM. This approach employs three-step modeling based on a second-order model, a balanced higher-order model, and a balanced highest-order model. The dataset from the experimental data in a three-level, two-factor central composite design in a previous research was used to illustrate three-step modeling and the subsequent optimization. The proposed approach to RSM predicted improved results of optimization, which are different from the predicted optimization results in the previous research.

**Keywords** response surface methodology, lack of fit, three-step modeling, balanced higher-order model, balanced highest-order model

### Introduction

Response surface methodology (RSM) refers to a set of statistical methods for the design of experiments, modeling of responses, and optimization of factor levels (Myers et al., 2009). RSM has extensively been used in experimental sciences, including the food science of animal resources. In RSM, a satisfactory modeling of responses is crucial for an accurate understanding of the response system and a precise optimization of explanatory factors. A model can be said to be satisfactory when: (1) the model is significant at the 5% level (the p-value of the model  $\leq 0.05$ ), (2) the lack of fit is non-significant at the 5% level (the p-value of the lack of fit  $> 0.05$ ) or the model has no lack of fit, and (3) the adjusted r-square is as large as at least 0.8 ( $\geq 0.8$ ).

In RSM, after obtaining data from an experiment, we fit a standard second-order model to the data and then check whether the above three criteria are met. If not all three criteria are met, the second-order model is unsatisfactory. Subsequently, we fit a

“balanced higher-order model” to the data and check whether the above three criteria are met. If still not all three criteria are met, the balanced higher-order model too is unsatisfactory, and we fit a “balanced highest-order model” to the data. This three-step sequential modeling can be summarized as follows:

- [1] Step 1: Fit a second-order model to the data. If this model is satisfactory, modeling stops here. Otherwise, go to step 2.
- [2] Step 2: Fit a balanced higher-order model to the data. If this model is satisfactory, modeling stops here. Otherwise, go to step 3.
- [3] Step 3: Fit a balanced highest-order model to the data.

When each factor of the experimental design has five levels, a balanced higher-order model in step 2 is a third-order model, and a balanced highest-order model in step 3 is the fullest balanced model proposed by Rheem et al. (2017). These models used in Rheem et al. (2017) contain cubic terms in the form of  $X_j^3$ . When each factor of the design has three levels that are coded as  $-1$ ,  $0$ , and  $1$ , these models cannot be used, as cubic terms cannot be entered into the model because they are the same as linear terms [ $(-1)^3=-1$ ,  $0^3=0$ , and  $1^3=1$ , that is,  $X_j^3=X_j$ ]. Thus, when each factor has three levels, the balanced higher-order and balanced highest-order models that have no cubic terms in them are required. The construction of such models will be explained later.

The present study proposes a three-step modeling strategy for experiments where each factor of the design has three levels, and subsequently optimizes factor levels for maximizing the response based on the balanced highest-order model. The study was conducted using the data from a three-level, two-factor experiment on a coffee-supplemented milk beverage (Ahn et al., 2017; Rheem et al., 2019).

## Materials and Methods

### Dataset to be analyzed

The construction of a balanced highest-order model will be explained through re-analysis of a dataset from the experimental data in the article titled “Improving the Quality of Response Surface Analysis of an Experiment for Coffee-supplemented Milk Beverage: II. Heterogeneous Third-order Models and Multi-response Optimization” authored by Rheem et al. (2019). In that study, two three-level factors were used in an experiment. Among the responses in that experiment, the Zeta-potential, for which the objective of optimization is maximization, is the Y variable in this study. To illustrate the model for the data from a three-level, two-factor design, the factors (the  $X_1$  and  $X_2$  variables) in this experiment and their coded and actual levels are presented in Table 1.

The dataset used for re-analysis was the data in Rheem et al. (2019) from which the first center run was deleted. The data in Rheem et al. (2019) was obtained through the screening (Rheem and Oh, 2019) of the original data in Ahn et al. (2017). The dataset that was analyzed in this study is presented in Table 2. In this dataset, the experimental design is the central composite design by Box and Wilson (1951) for two factors having  $-1$ ,  $0$ , and  $1$  as their coded levels. With this dataset, we performed a response surface analysis using balanced higher-order and balanced highest-order models.

### Statistical analysis tool

Data were analyzed using the SAS software. SAS (2016b) procedures were used for the modeling of the response and the

**Table 1. Response, actual, and coded factors and the levels of the factors**

Response=Y	Actual factor	Coded factor	Actual factor level corresponding to the coded factor level of		
			-1	0	1
Y=Zeta-potential	F <sub>1</sub> =Speed of primary homogenization (rpm)	X <sub>1</sub>	5,000	10,000	15,000
	F <sub>2</sub> =Concentration of emulsifier (%)	X <sub>2</sub>	0.1	0.2	0.3

**Table 2. Experimental design in coded factors and responses**

Standard order	Design point	X <sub>1</sub>	X <sub>2</sub>	Y
1	1	-1	-1	27.5
2	2	-1	1	29.9667
3	3	1	-1	24.3
4	4	1	1	32.5666
5	5	-1	0	36.1
6	6	1	0	28.2667
7	7	0	-1	29.1
8	8	0	1	28.2
9	9	0	0	29.0667
10	9	0	0	29.6
11	9	0	0	29.1

drawing of a contour plot. The optimization of factor levels was performed by SAS data-step programming. A three-dimensional plot was generated using SAS (2016a).

## Results and Discussion

### Fitting the second-order model to the data (Step 1 in modeling)

First, the second-order polynomial regression model containing two linear, two quadratic, and one interaction term was fitted to the data by using the RSREG procedure of SAS/STAT. Results of the analysis of variance for the second-order model are shown in Table 3.

Table 3 says that: (1) the model is non-significant at the 5% level (the p-value of the model=0.3338>0.05), (2) the lack of fit is significant at the 5% level (the p-value of the lack of fit=0.0071<0.05), and (3) the adjusted r-square=0.1388<0.8; none of the three criteria were satisfied, and thus, this model was unsatisfactory. As the second-order model had a poor fit, we proceeded to build a balanced higher-order model.

### Fitting a balanced higher-order model to the data (Step 2 in modeling)

In a multifactor polynomial regression model, if each factor appears the same number of times in all the terms in the model, such a model is said to be “balanced.” In a balanced response surface model, each factor has an equal opportunity of being evaluated through the coefficients of the model terms that contain it.

**Table 3. Analysis of variance for the second-order model**

Model terms: $X_1, X_2; X_1^2, X_2^2; X_1X_2$					
Source	Degrees of freedom	Sum of squares	Mean square	F-value	p-value
Model	5	50.05	10.01	1.32	0.3833
Error	5	37.85	7.57		
Total	10	87.90			
Root MSE=2.75		r-square=0.5694		Adjusted r-square=0.1388	
Test of lack of fit					
Source	Degrees of freedom	Sum of squares	Mean square	F-value	p-value
Lack of fit	3	37.67	12.56	140.69	0.0071
Pure error	2	0.18	0.09		

MSE, mean square error.

When there are two factors, a second-order model, which consists of  $X_1, X_2; X_1^2, X_2^2; X_1X_2$ , is balanced, because for  $j=1$  and  $2$ , each  $X_j$  appears three times in all the terms (1 time in  $X_1, X_2$ ; 2 times in  $X_1^2, X_2^2$ ; and 1 time in  $X_1X_2$ ) in the model. Thus, the balanced higher-order model consists of  $X_1, X_2; X_1^2, X_2^2; X_1X_2; X_1X_2^2, X_1^2X_2$ , where for  $j=1$  and  $2$ , each  $X_j$  appears 7 times (1 time in  $X_1, X_2$ , 2 times in  $X_1^2, X_2^2$ , and 1 time in  $X_1X_2$ , and 3 times in  $X_1X_2^2, X_1^2X_2$ ). Table 4 shows the results of the analysis of variance for the balanced higher-order model.

Table 4 says that: (1) the model is non-significant at the 5% level (the p-value of the model=0.0841>0.05), (2) the lack of fit is significant at the 5% level (the p-value of the lack of fit=0.0154<0.05), and (3) the adjusted r-square=0.7785<0.8; none of the three criteria were satisfied, and thus, the balanced higher-order model was also unsatisfactory. Therefore, for the next step of model fitting, a balanced highest-order model was fitted to the data.

### Fitting a balanced highest-order model to the data (Step 3 in modeling)

In Table 4, the lack of fit had one degree of freedom, implying that we can add one more term to the model. For the model to be balanced, this additional term needs to contain all of  $X_1$  and  $X_2$ . As the last terms in the model are  $X_1X_2^2$  and  $X_1^2X_2$ , the

**Table 4. Analysis of variance for the balanced higher-order model**

Model terms: $X_1, X_2; X_1^2, X_2^2; X_1X_2; X_1X_2^2, X_1^2X_2$					
Source	Degrees of freedom	Sum of squares	Mean square	F-value	p-value
Model	7	82.06	11.72	6.02	0.0841
Error	3	5.84	1.95		
Total	10	87.90			
Root MSE=1.40		r-square=0.9335		Adjusted r-square=0.7785	
Test of lack of fit					
Source	Degrees of freedom	Sum of squares	Mean square	F-value	p-value
Lack of fit	1	5.66	5.66	63.45	0.0154
Pure error	2	0.18	0.09		

MSE, mean square error.

next term to be entered into the model should be  $X_1^2X_2^2$ . When we add this term to the model, the model becomes a balanced highest-order model, which has no lack of fit, because the number of coefficients in the model is the same as the number of design points in the experimental design. Design points are indicated in Table 2. Results of the analysis of variance for the balanced highest-order model are presented in Table 5.

Table 5 says that: (1) the model is significant at the 5% level (the p-value of the model=0.0081<0.05), (2) the model has no lack of fit, and (3) the adjusted r-square=0.9898>0.8; all three criteria were satisfied, and the r-square was 0.9980, almost equal to 1. Therefore, the balanced highest-order model was satisfactory and accepted as our final model. With  $\hat{Y}$  denoting the predicted value of Y, this model can be specified as:

$$\hat{Y} = 29.25557 + (-3.91665) X_1 + (-0.45) X_2 + 2.92778 X_1^2 + (-0.60557) X_2^2 + 1.44998 X_1X_2 + 3.76662 X_1X_2^2 + 3.13333 X_1^2X_2 + (-2.99446) X_1^2X_2^2. \tag{1}$$

Table 6 shows the coefficient estimates and their SEs, t-values, and p-values of the balanced higher-order model.

The p-values in Table 6 say that the  $X_1$ ,  $X_1^2$ ,  $X_1X_2$ ,  $X_1X_2^2$ ,  $X_1^2X_2$ , and  $X_1^2X_2^2$  terms, which all involve the  $X_1$  factor, are all significant at the 5% level, whereas the  $X_2$  and  $X_2^2$  terms, which involve only the  $X_2$  factor, are non-significant at the 5% level. From this, we see that the  $X_1$  factor plays a more important role than the  $X_2$  factor in the model.

### Finding the optimum point of the factors

According to Ahn et al. (2017), the objective of optimization for Y was maximization. Thus, using a search on a grid (Oh et al., 1995), we optimized the model for maximization in the form of the above equation. In this experiment, the bounds were  $-1 \leq X_j \leq 1$  for  $j=1, 2$ . Thus, within these bounds, we conducted a search on a grid using the SAS data step programming. A search for the maximum on a grid was performed by calculating the  $\hat{Y}$  function over a grid for the values of  $X_1$  and  $X_2$ , with an increment of 0.01 within the bounds  $-1 \leq X_j \leq 1$  for  $j=1, 2$ , and then sorting the calculated function values in descending order. The optimum point at which  $\hat{Y}$  was maximized was found this way and is presented in Table 7.

In Rheem et al. (2019), the predicted maximum Zeta-potential based on their model was 35.2957, which is lower than 36.1515, our predicted maximum Zeta-potential in Table 7. And, according to the data and the model in Rheem et al. (2019),

**Table 5. Analysis of variance for the balanced highest-order model**

Model terms: $X_1, X_2; X_1^2, X_2^2; X_1X_2; X_1X_2^2, X_1^2X_2; X_1^2X_2^2$					
Source	Degrees of freedom	Sum of squares	Mean square	F-value	p-value
Model	8	87.72	10.97	112.86	0.0081
Error	2	0.18	0.09		
Total	10	87.90			
Root MSE=0.30		r-square=0.9980		Adjusted r-square=0.9898	
Test of lack of fit					
Source	Degrees of freedom	Sum of squares	Mean square	F-value	p-value
Lack of fit	0	0			
Pure error	2	0.18	0.09		

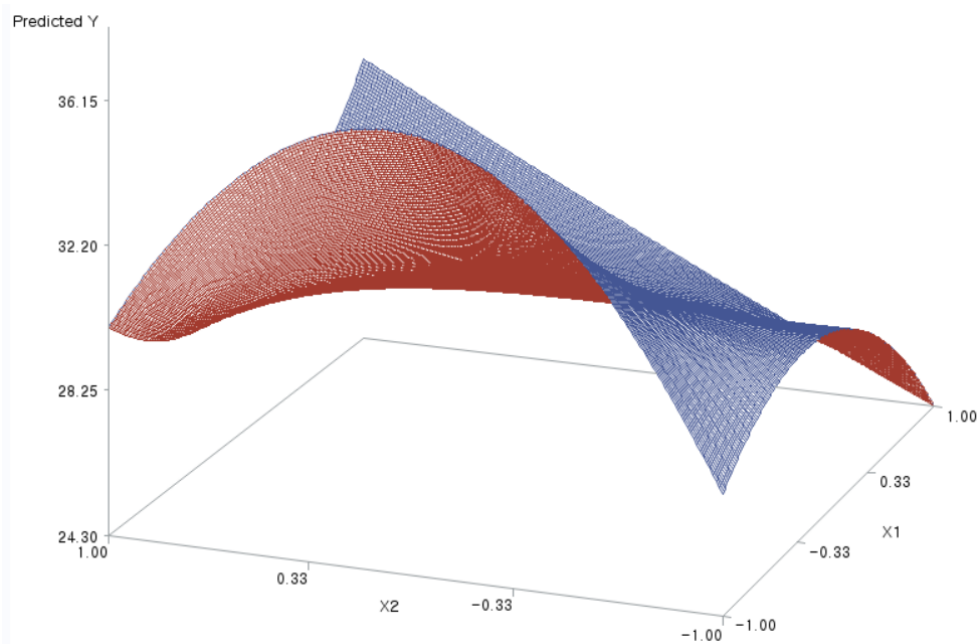
MSE, mean square error.

**Table 6. Coefficient estimates and related statistics in the balanced highest-order model**

Coefficient	Estimate	SE	t-value	p-value
Intercept	29.25557	0.1725	169.61	<0.0001
X <sub>1</sub>	-3.91665	0.2112	-18.54	0.0029
X <sub>2</sub>	-0.45000	0.2112	-2.13	0.1669
X <sub>1</sub> <sup>2</sup>	2.92778	0.2727	10.74	0.0086
X <sub>2</sub> <sup>2</sup>	-0.60557	0.2727	-2.22	0.1565
X <sub>1</sub> X <sub>2</sub>	1.44998	0.1494	9.71	0.0104
X <sub>1</sub> X <sub>2</sub> <sup>2</sup>	3.76662	0.2587	14.56	0.0047
X <sub>1</sub> <sup>2</sup> X <sub>2</sub>	3.13333	0.2587	12.11	0.0067
X <sub>1</sub> <sup>2</sup> X <sub>2</sub> <sup>2</sup>	-2.99446	0.3759	-7.97	0.0154

**Table 7. Optimization results**

X <sub>1</sub>	X <sub>2</sub>	F <sub>1</sub> =Speed of primary homogenization (rpm)	F <sub>2</sub> =Concentration of emulsifier (%)	Predicted maximum of Y (=Zeta-potential)
-1	0.08	5,000	0.208	36.1515

**Fig. 1. Three-dimensional plot for the effects of X<sub>1</sub> and X<sub>2</sub>.**

for their predicted maximum (35.2957) of Zeta-potential, their predicted optimum points of (X<sub>1</sub>, X<sub>2</sub>) in coded levels and (F<sub>1</sub>, F<sub>2</sub>) in actual levels are calculated as (X<sub>1</sub>=-1, X<sub>2</sub>=0.10) and (F<sub>1</sub>=5.000, F<sub>2</sub>=0.210), which are different from (X<sub>1</sub>=-1, X<sub>2</sub>=0.08) and (F<sub>1</sub>=5.000, F<sub>2</sub>=0.208), our predicted optimum points in coded and actual levels in Table 7. Thus, it can be said that our approach has predicted improved results of optimization, which are different from the predicted optimization results in the previous research. A validation experiment is needed after the optimum point is predicted.

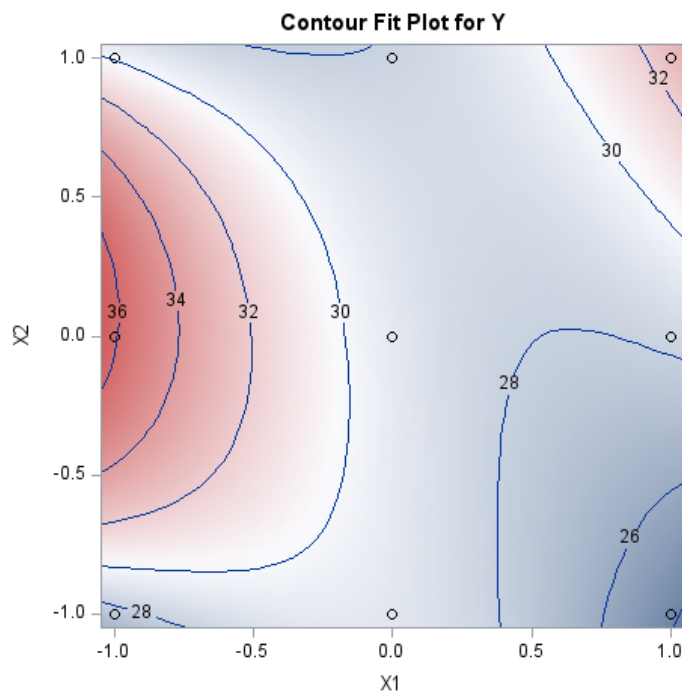


Fig. 2. Contour plot for the effects of  $X_1$  and  $X_2$ .

### Drawing the 3D and contour plots of the response surface

For the  $X_1$  and  $X_2$  factors, the three-dimensional (3D) response surface plot was drawn, with the vertical axis representing the predicted response and two horizontal axes indicating the coded levels of the  $X_1$  and  $X_2$  factors. Fig. 1 is such a plot.

The two-dimensional contour plot of the response surface was also drawn, with two axes indicating two coded factors. Fig. 2 is such a plot.

These figures visualize the optimization results shown in Table 7; they show that the predicted response was maximized at the point of ( $X_1=-1$ ,  $X_2=0.08$ ).

## Conflicts of Interest

The authors declare no potential conflicts of interest.

## Acknowledgements

This work was financially supported by the College of Public Policy at Korea University in 2022.

## Author Contributions

The article is prepared by a single author.

## Ethics Approval

This article does not require IRB/IACUC approval because there are no human and animal participants.

## References

- Ahn SI, Park JH, Kim JH, Oh DG, Kim M, Chung D, Jhoo JW, Kim GY. 2017. Optimization of manufacturing conditions for improving storage stability of coffee-supplemented milk beverage using response surface methodology. *Korean J Food Sci Anim Resour* 37:87-97.
- Box GEP, Wilson KB. 1951. On the experimental attainment of optimum conditions. *J R Stat Soc* 13:1-38.
- Myers RH, Montgomery DC, Anderson-Cook CM. 2009. *Response surface methodology: Process and product optimization using designed experiments*. 3<sup>rd</sup> ed. John Wiley & Sons, Hoboken, NJ, USA.
- Oh S, Rheem S, Sim J, Kim S, Baek Y. 1995. Optimizing conditions for the growth of *Lactobacillus casei* YIT 9018 in tryptone-yeast extract-glucose medium by using response surface methodology. *Appl Environ Microbiol* 61:3809-3814.
- Rheem S, Oh S. 2019. Improving the quality of response surface analysis of an experiment for coffee-supplemented milk beverage: I. Data screening at the center point and maximum possible r-square. *Food Sci Anim Resour* 39:114-120.
- Rheem S, Rheem I, Oh S. 2017. Response surface methodology using a fullest balanced model: A re-analysis of a dataset in the *Korean Journal for Food Science of Animal Resources*. *Korean J Food Sci Anim Resour* 37:139-146.
- Rheem S, Rheem I, Oh S. 2019. Improving the quality of response surface analysis of an experiment for coffee-supplemented milk beverage: II. Heterogeneous third-order models and multi-response optimization. *Food Sci Anim Resour* 39:222-228.
- SAS. 2016a. *SAS/GRAPH user's guide*. Release 9.4. SAS Institute, Cary, NC, USA.
- SAS. 2016b. *SAS/STAT user's guide*. Release 9.4. SAS Institute, Cary, NC, USA.