Korean Journal for Food Science of Animal Resources

pISSN 1225-8563
eISSN 2234-246X

Korean J. Food Sci. An. 37(1): 139~146 (2017)
DOI https://doi.org/10.5851/kosfa.2017.37.1.139

ARTICLE

Response Surface Methodology Using a Fullest Balanced Model: A Re-Analysis of a Dataset in the *Korean Journal for Food Science of Animal Resources*

Sungsue Rheem[1]*, Insoo Rheem[2], and Sejong Oh*

Division of Animal Science, Chonnam National University, Gwangju 61186, Korea
[1]Department of Applied Statistics, Korea University, Sejong 30019, Korea
[2]Department of Laboratory Medicine, Dankook University Hospital, Cheonan 31116, Korea

Abstract

Response surface methodology (RSM) is a useful set of statistical techniques for modeling and optimizing responses in research studies of food science. In the analysis of response surface data, a second-order polynomial regression model is usually used. However, sometimes we encounter situations where the fit of the second-order model is poor. If the model fitted to the data has a poor fit including a lack of fit, the modeling and optimization results might not be accurate. In such a case, using a fullest balanced model, which has no lack of fit, can fix such problem, enhancing the accuracy of the response surface modeling and optimization. This article presents how to develop and use such a model for the better modeling and optimizing of the response through an illustrative re-analysis of a dataset in Park *et al.* (2014) published in the *Korean Journal for Food Science of Animal Resources*.

Keywords response surface methodology, lack of fit, second-order model, fullest balanced model, optimization, search on a grid

Received February 3, 2017
Accepted February 8, 2017

Corresponding authors
Sungsue Rheem
Department of Applied Statistics, Korea University, Sejong 30019, Korea
Tel: +82-10-8725-7124
E-mail: rheem@korea.ac.kr

Sejong Oh
Division of Animal Science, Chonnam National University, Gwangju 61186, Korea
Tel: +82-62-530-2116
E-mail: soh@jnu.ac.kr

Introduction

The “change-one-factor-at-a-time” method has traditionally been used in experiments with multiple factors. This is a method in which one factor is varied while all other factors are fixed under certain conditions (Logothetis and Wynn, 1989). However, this method does not take all factors into account at the same time. This can lead to unreliable results and incorrect conclusions. Considering all factors simultaneously, response surface methodology (RSM) can better handle experiments for modeling and optimization. RSM is a set of statistical techniques for designing experiments, creating models, evaluating the impacts of factors, and exploring optimal conditions for desirable responses (Myers *et al*., 2009).

Regarding experimental designs in RSM, central composite designs (CCD; Box and Wilson, 1951) have been used most frequently. A CCD is a three- or five-level design that can fit a second-order polynomial model to data within a cubic or spherical experimental region. For a second-order model to be a good predictive model, it should satisfy some criteria that the p-value of the model ≤ 0.05, the p-value of the lack of fit > 0.1, and the adjusted R-square ≥ 0.8 (Myers *et al*., 2009). If the model fitted to the data does not meet these criteria, modeling and optimiza-

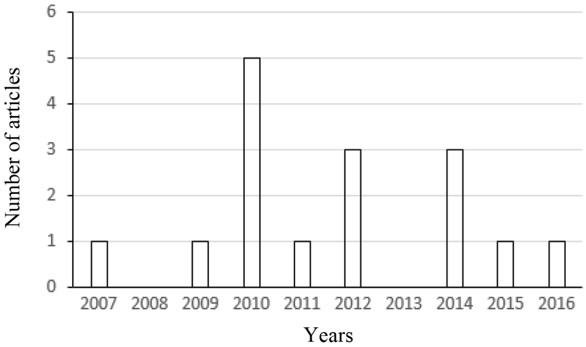

Fig. 1. Number of articles published each year using CCD in Korean Journal for Food Science of Animal Resources.

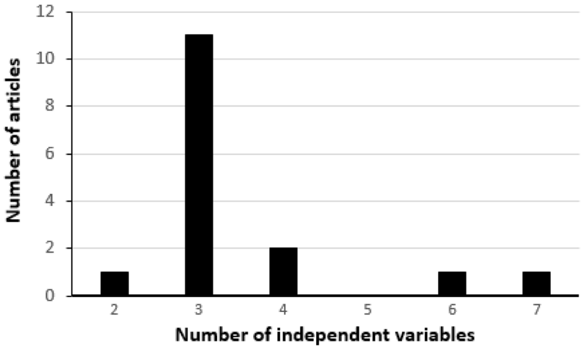

Fig. 2. Number of articles per the number of factors in a CCD.

tion results might not be accurate.

However, in reality, it is observed that the models that do not satisfy the above criteria are used in the analyses of response surface experiments. This seems to be because researchers have little knowledge of what to do in such a situation. A remedy in this case is to use a third-order model. Rheem and Rheem (2012) improved a second-order model with a significant lack of fit by adding cubic terms to it. However, cases can exist where a third-order model still falls short of such criteria for a good predictive model. In these cases, there arises a need to use a fullest model that has no lack of fit.

When a spherical CCD is used as an experimental design, fullest models with no lack of fit exist, but they are not unique. However, among them, a balanced model is unique. This article proposes such a model, which is called a fullest balanced model, and how to use it.

For the last ten years (from 2007 to 2016), sixteen articles using a CCD in RSM were published in the *Korean Journal for Food Science of Animal Resources*. The number of such articles published each year during that period is shown in Fig. 1. After examining these 16 articles, we found that three independent variables (three factors) were most frequently used in such articles (Fig. 2). Thus, a dataset with three factors, which is in Park *et al*. (2014) published in the *Korean Journal for Food Science of Animal Resources*, will be re-analyzed for the illustration of the method suggested in this article.

Materials and Methods

Dataset to be re-analyzed

How to use a fullest balanced model will be explained through re-analysis of a dataset described in the article entitled "Application of Response Surface Methodology (RSM) for Optimization of Anti-Obesity Effect in Fermented Milk by *Lactobacillus plantarum* Q180" authored by Park *et al*. (2014). In this article, three factors were used in an experiment to model three responses. Among them, the third response, which was anti-adipogenetic activity (%), had the poorest fit of a second-order model. Thus, this response is used as the Y variable in this article. Factors (X variables) in this experiment and their coded and actual levels are given in Table 1.

The dataset to be re-analyzed is shown in Table 2. In this dataset, the experimental design is the CCD for three factors with an axial value of 1.68179 and three center runs. Using this design, to the data, we can fit a second-order model, a third-order model, and a fullest balanced model.

Statistical analysis

Data were analyzed using SAS software. SAS/STAT (2013) procedures were used for regression modeling. Optimum conditions were found through SAS data-step

Table 1. Response and factors

Response = Y	Actual factor	Coded factor	Actual factor level at the coded factor level of				
			-1.68179	-1	0	1	1.68179
Anti-adipogenetic activity (%)	Skim milk powder (%)	X_1	8.318	9	10	11	11.682
	Incubation temp. (°C)	X_2	31.955	34	37	40	42.045
	Incubation time (h)	X_3	12.841	20	30.5	41	48.159

Table 2. Experimental design in coded levels and responses

Standard Order	Design point	X_1	X_2	X_3	Y
1	1	-1	-1	-1	19.17
2	2	1	-1	-1	-2.39
3	3	-1	1	-1	13.73
4	4	1	1	-1	5.94
5	5	-1	-1	1	10.29
6	6	1	-1	1	-4.02
7	7	-1	1	1	12.28
8	8	1	1	1	5.58
9	9	-1.68179	0	0	26.78
10	10	1.68179	0	0	-2.57
11	11	0	-1.68179	0	13.91
12	12	0	1.68179	0	5.76
13	13	0	0	-1.68179	30.04
14	14	0	0	1.68179	10.11
15	15	0	0	0	18.44
16	15	0	0	0	16.45
17	15	0	0	0	15.00

programming. Plots were generated using SAS/GRAPH (2013).

Results and Discussion

Developing a regression model

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

In Table 3, the p-value of the model = 0.0642 > 0.05, the *p*-value of the lack of fit = 0.0526 < 0.1, and the adjusted R-square = 0.5654 < 0.8; none of the three criteria are satisfied. Since this second-order model has a poor fit, next we will fit to the data a third-order model that

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

Model terms: X_1, X_2, X_3; X_1^2, X_2^2, X_3^2; X_1X_2, X_1X_3, X_2X_3					
Source	Degrees of freedom	Sum of squares	Mean square	*F*-value	*p*-value
Model	9	1187.5291	131.9477	3.31	0.0642
Error	7	278.8277	39.8325	-	-
Total	16	1466.3568	-	-	-
Root MSE = 6.3113		R-square = 0.8099		Adjusted R-square = 0.5654	
Test of lack of fit					
Source	Degrees of freedom	Sum of squares	Mean square	*F*-value	*p*-value
Lack of fit	5	272.8623	54.5725	18.3	0.0526
Pure Error	2	5.9654	2.9827	-	-

Table 4. Analysis of variance for the third-order model

Model terms: X_1, X_2, X_3; X_1^2, X_2^2, X_3^2; X_1X_2, X_1X_3, X_2X_3; X_1^3, X_2^3, X_3^3; $X_1X_2X_3$					
Source	Degrees of freedom	Sum of squares	Mean square	*F*-value	*p*-value
Model	13	1334.9522	102.6886	2.34	0.2627
Error	3	131.4045	43.8015	-	-
Total	16	1466.3568	-	-	-
Root MSE = 6.6183		R-square = 0.9104		Adjusted R-square = **0.5221**	
Test of lack of fit					
Source	Degrees of freedom	Sum of squares	Mean square	*F*-value	*p*-value
Lack of fit	1	125.4391	125.4391	42.06	0.0230
Pure Error	2	5.9654	2.9827	-	-

consists of linear, quadratic, cubic, and two-way and three-way interaction terms, anticipating a possible improvement in modeling. Table 4 shows the results of analysis of variance for this third-order model.

In Table 4, the p-value of the model = 0.2627 > 0.05, the p-value of the lack of fit = 0.0230 < 0.1, and the adjusted R-square = 0.5221 < 0.8; none of the three criteria are satisfied. This third-order model is worse than the previous second-order model, let alone better. Now, the lack-of-fit part has 1 degree of freedom, which means 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, X_2, and X_3. Then, since the latest term in the model is $X_1 X_2X_3$, the next term to enter the model should be $X_1^2X_2^2 X_3^2$. Now, we add this term to the model, expecting a possible improvement in modeling. Results of analysis of variance for this fullest balanced model are given in Table 5.

In Table 5, the p-value of the model = 0.0281 < 0.05, and the adjusted R-square = 0.9675 > 0.8; two criteria are satisfied. The lack-of-fit part has 0 degree of freedom, which means that this model has no lack of fit. And, the R-square is 0.9959, almost 1. Finally, we have obtained the improved model that will be used for optimization. Letting $\hat{Y}$ denote the predicted value of Y, we specify this model as

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{11}X_1^2 + b_{22}X_2^2 + b_{33}X_3^2 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{111}X_1^3 + b_{222}X_2^3 + b_{333}X_3^3 + b_{123}X_1X_2X_3 + b_{112233}X_1^2X_2^2X_3^2$$

where the coefficients b_1, b_2, …, b_{112233} are given in Table 6, which says that $X_1^2 X_2^2 X_3^2$ is the most significant term among the model terms.

Finding the optimum point of the factors

According to Park *et al.* (2014), the optimization objective for Y was maximization. Thus, through a search on a grid (Oh *et al.*, 1995), we maximized the model with the coefficients in Table 5. In this experiment, the bounds are $-1.682 \le X_j \le 1.682$ for j = 1, 2, 3. In the CCD in Table 2, every design point is under the constraint $X_1^2 + X_2^2 + X_3^2 \le (\pm1)^2 + (\pm1)^2 + (\pm1)^2 = 3$, which makes the design region

Table 5. Analysis of variance for the fullest balanced model

Model terms: $X_1, X_2, X_3; X_1^2, X_2^2, X_3^2; X_1X_2, X_1X_3, X_2X_3; X_1^3, X_2^3, X_3^3; X_1X_2X_3; X_1^2 X_2^2 X_3^2$					
Source	Degrees of freedom	Sum of squares	Mean square	*F*-value	*p*-value
Model	14	1460.3914	104.3137	34.97	0.0281
Error	2	5.9654	2.9827	-	-
Total	16	1466.3568	-	-	-
Root MSE = 1.7271		R-square = 0.9959		Adjusted R-square = **0.9675**	
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	5.9654	2.9827	-	-

Table 6. Coefficient estimates in the fullest balanced model

Term	Parameter Estimate	Standard Error	*t*-value	*p*-value
Intercept	b_0 = 16.63000	0.99711	16.68	0.0036
X_1	b_1 = -4.96553	1.02465	-4.85	0.0400
X_2	b_2 = 4.12512	1.02465	4.03	0.0565
X_3	b_3 = 0.85838	1.02465	0.84	0.4903
X_1^2	b_{11} = -1.59983	0.55740	-2.87	0.1030
X_2^2	b_{22} = -2.40240	0.55740	-4.31	0.0498
X_3^2	b_{33} = 1.21800	0.55740	2.19	0.1605
X_1X_2	b_{12} = 2.67250	0.61060	4.38	0.0484
X_1X_3	b_{13} = 1.04250	0.61060	1.71	0.2299
X_2X_3	b_{23} = 1.08750	0.61060	1.78	0.2169
X_1^3	b_{111} = -1.32947	0.51889	-2.56	0.1245
X_2^3	b_{222} = -2.31512	0.51889	-4.46	0.0467
X_3^3	b_{333} = -2.39838	0.51889	-4.62	0.0438
$X_1X_2X_3$	b_{123} = -0.77000	0.61060	-1.26	0.3345
$X_1^2 X_2^2 X_3^2$	b_{112233} = -6.27326	0.96735	-6.49	0.0230

Table 7. Optimization results

X_1	X_2	X_3	Distance from the origin	Skim milk powder (%)	Incubation temp. (°C)	Incubation time (h)	Anti-adipogenetic activity (%)
-0.42	0.03	-1.68	1.73196	9.58	37.09	12.86	32.6492

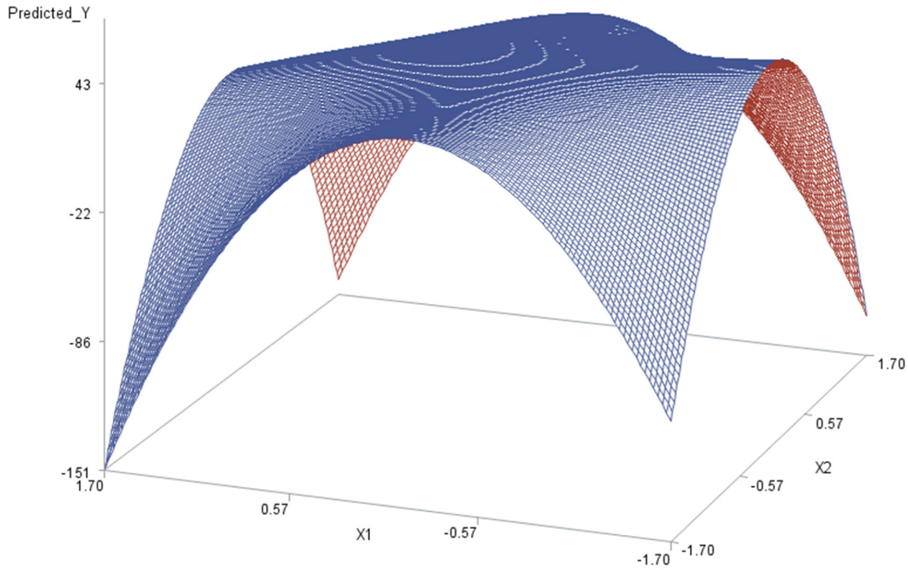

Fig. 3. Response surface for the effects of X_1 and X_2 on the predicted Y at X_3 = -1.68.

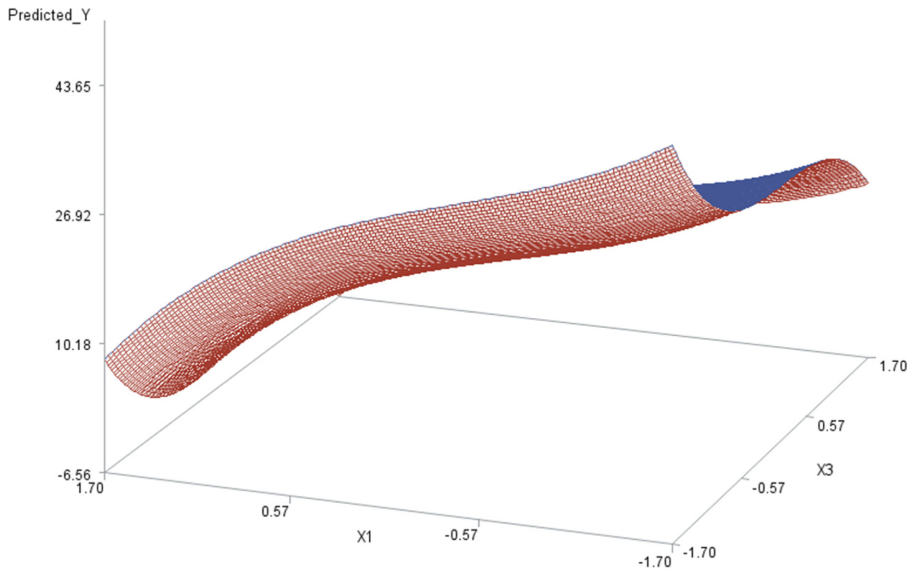

Fig. 4. Response surface for the effects of X_1 and X_3 on the predicted Y at X_2 = 0.03.

spherical with the radius $\sqrt{3}$ = 1.732. Thus, satisfying these bounds and constraint, we conducted a search on a grid using the SAS data step programming. Here, a search for the maximum on a grid was performed by calculating the $\hat{Y}$ function over a grid of the values of X_1, X_2, and X_3 with an increment of 0.01 under the bounds $-1.682 \leq X_j \leq 1.682$ for j = 1, 2, 3 and the constraint $X_1^2 + X_2^2 + X_3^2 \leq 3$, and then sorting the calculated function values in desc-

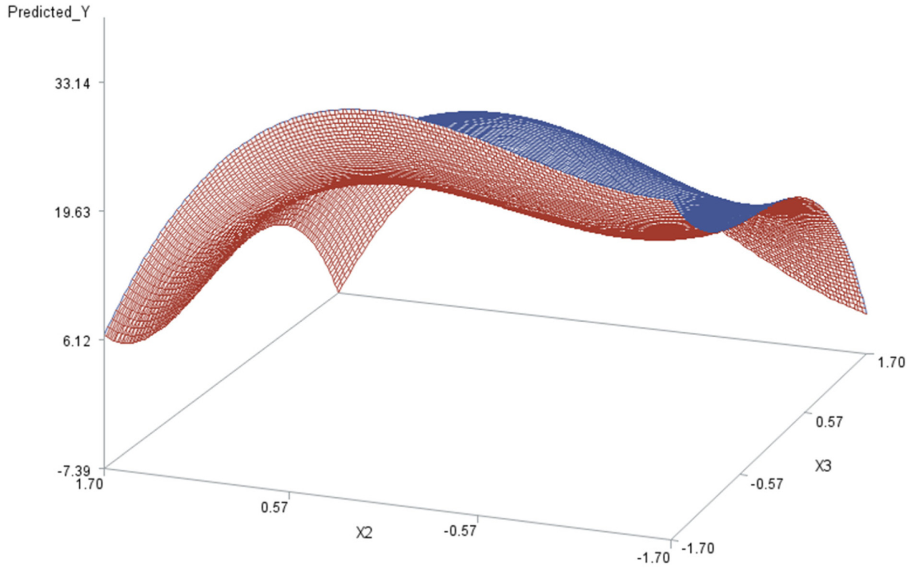

Fig. 5. Response surface for the effects of X_2 and X_3 on the predicted Y at X_1 = -0.42.

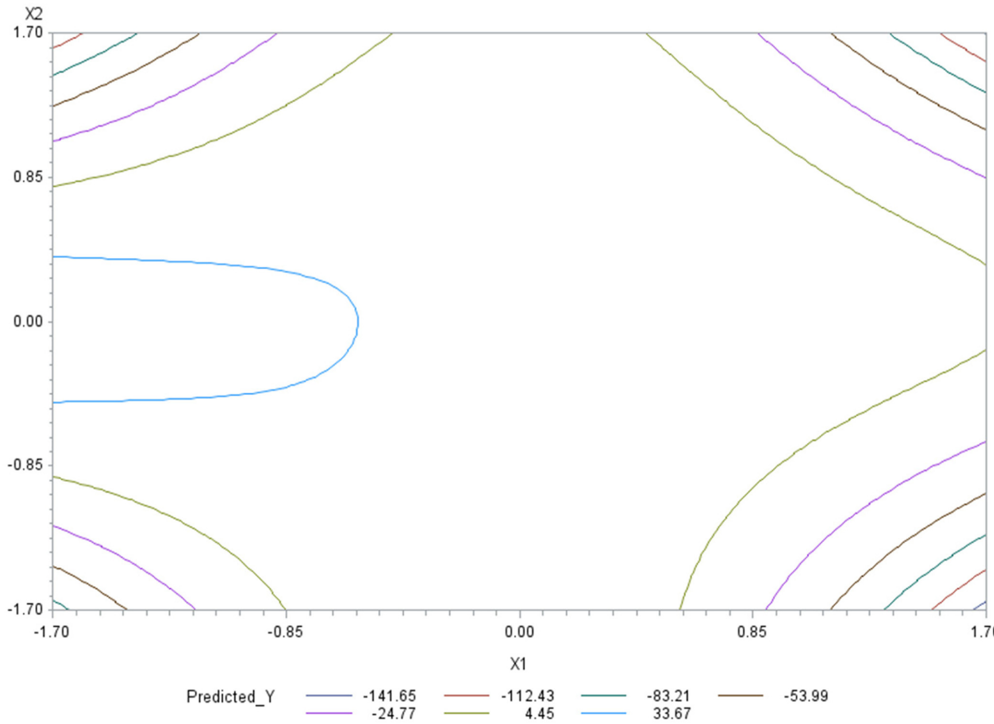

Fig. 6. Response contour for the effects of X_1 and X_2 on the predicted Y at X_3 = -1.68.

ending order. The optimum point at which $\hat{Y}$ is maximized was found this way and presented in Table 7.

In Park *et al.* (2014), the predicted maximum anti-adipogenetic activity was 31%. Their optimum conditions for this maximum were skim milk powder = 8.4677%, incubation temperature = 65.3815ºC, and incubation time = 12.8412 h. These maximum and optimum conditions are different from our optimization results. Our predicted maximum was 32.6492%, which was greater than their predicted maximum 31%. A validation experiment is nee-

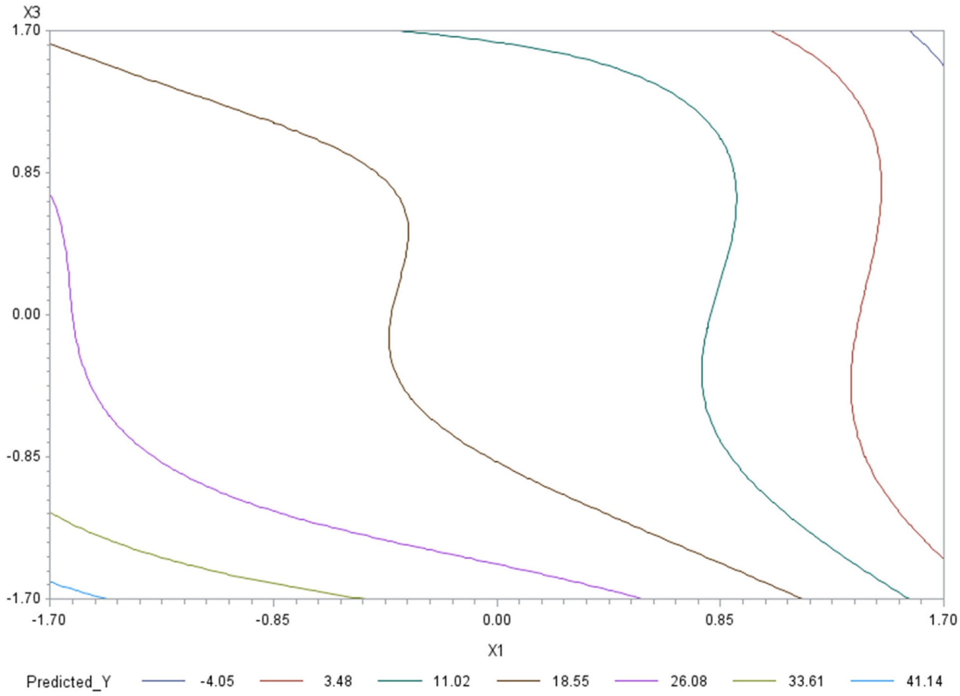

Fig. 7. Response contour for the effects of X_1 and X_3 on the predicted Y at X_2 = 0.03.

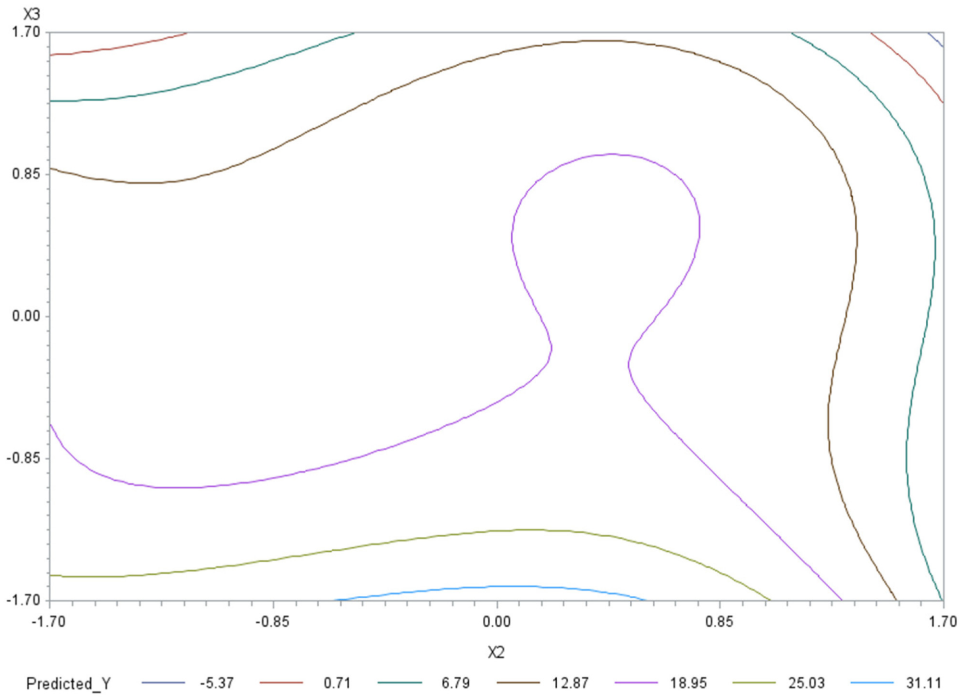

Fig. 8. Response contour for the effects of X_2 and X_3 on the predicted Y at X_1 = -0.42.

ded to verify the optimization results obtained by this methodology.

Drawing 3D and contour plots of response surfaces

Like in Oh *et al.* (1995), for any two of the three factors, a three-dimensional (3D) response surface plot was drawn with the vertical axis representing the predicted response and two horizontal axes representing the coded levels of two explanatory factors. In each 3D plot, the factor not represented by the two horizontal axes is fixed at its optimum level. All three 3D plots were produced. Figs. 3 through 5 are such plots.

Two-dimensional contour plots of response surfaces were also drawn with two axes indicating two coded factors. In each contour plot, the factor not represented by the two axes is fixed at its optimum level. All three contour plots were produced. Figs. 6 through 8 are such plots.

Acknowledgements

This research was supported by a Korea University Grant.

References

1. Box, G. E. P. and Wilson, K. B. (1951) On the experimental attainment of optimum conditions. *J. Royal Stat. Soc. Series B* **13**, 1-45.
2. Logothetis, N. and Wynn, H. P. (1989) Quality through design: Experimental design, off-line quality control, and Taguchi's contributions. Oxford University Press.
3. Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. (2009) Response Surface Methodology: Process and Product Optimization Using Designed Experiments. 3rd edition, John Wiley & Sons.
4. Oh, S., Rheem, S., Sim, J., Kim, S., and Baek, Y. (1995) Optimizing conditions for the growth of *Lactobacillus casei* YIT 9018 in tryptone-glucose medium by using response surface methodology. *Appl. Environ. Microbiol.* **61**, 3809-3814.
5. Park, S. Y., Cho, S. A., and Lim, S. D. (2014) Application of response surface methodology (RSM) for optimization of anti-obesity effect in fermented milk by *Lactobacillus plantarum* Q180. *Korean J. Food Sci. Anim. Resour.* **34**, 836-843.
6. Rheem, I. and Rheem, S. (2012) Response surface analysis in the presence of the lack of fit of the second-order polynomial regression model. *J. Korean Data Anal. Soc.* **14**, 2995-3002.
7. SAS Institute Inc. (2013) SAS/STAT User's Guide, Release 9.4. SAS Institute, Inc., Cary, NC, USA.
8. SAS Institute Inc. (2013) SAS/GRAPH User's Guide, Release 9.4. SAS Institute, Inc., Cary, NC, USA.