

Techniques for Optimizing Response Surfaces

Randall D. Tobias, SAS Institute Inc., Cary, NC
Jason Brown, University of Missouri at Columbia, Columbia, MO

Abstract

Typically, the effects of various factors on certain characteristics of a process are studied, with an eye toward identifying those factors that significantly affect the process and selecting settings for these factors that optimize the process characteristics. The present methods of optimization are either computationally intensive (nonlinear optimization) or graphically confusing (overlayed contour plots). In this paper, we present simple techniques to optimize a process: a new graphical method called the profile ridge plot and some sorting methods. Both of these techniques assume the analyst has estimated a model of the process and created a set of predicted values over the design space.

Introduction

The importance of experimentation for industrial quality improvement applications has received renewed attention in recent years (Taguchi 1978, Taguchi and Wu 1980, Box and Bisgaard 1987). Typically, the effects of various factors on certain characteristics of a process are studied, with an eye toward identifying those factors that significantly affect the process and selecting settings for these factors that optimize the process characteristics.

The first stage here involves the statistical problem of fitting a model to the experimental data and comparing the estimated effect of each term to the background level of noise in the process. In many cases this step can be viewed as honing the model for the second stage; however, it can also disclose factors whose levels are not necessary to control, leading to cost reductions.

By contrast, the second stage is more of a mathematical problem than a statistical one: given the model for the process determined by the first stage, the goal is to find a combination of the factors that optimizes the process. Optimization involves minimizing or maximizing the response, or setting it as close as possible

to a target value. Statistical considerations prohibit extrapolating too far beyond the design space, but otherwise the problem is one of exploring a given functional relationship within a specified region.

In this paper, we present simple sorting methods as well as a new graphical technique to allow the analyst to find the optimum response. Both of these techniques assume that the first stage of the preceding process has been performed and that a set of predicted values has been created over the entire design space.

Our graphical technique, the *profile ridge plot*, is based on a ridge analysis of the data. This plot enhances the ridge analysis by looking at the "path" to the maximum or minimum as it is superimposed on a plot of each factor and response over the entire design space. This type of graphic gives an indication of the movement of the optimal response ridge over the multidimensional response surface.

In the next section, we give the groundwork for the general problem to be discussed. Then we describe and explain the profile ridge plots through an example with a single response. Finally, we present and explain sorting methods through an example with multiple responses.

General Problem

Suppose we have p factors $X_1, X_2, \dots, X_p$ thought to affect the value of a response Y . For example, the response might be the yield of a chemical reaction process, and two factors of interest might be the reaction temperature and the reaction time. We assume that the relationship between the factors and the response is described by a known functional form:

$$Y = f(X_1, \dots, X_p) \quad (1)$$

where the factors $X_1, X_2, \dots, X_p$ take values in a given region of interest; their composite range defines the design space. For all our applications, we assume

it to be intrinsically linear; that is,

$$Y = \sum_{i=1}^k \beta_i g_i(X_1, \dots, X_p) \quad (2)$$

for some known functions $g_i(\cdot)$. For example, a quadratic model includes an overall intercept term, linear and quadratic terms for each factor, and crossproduct terms, so that we have in (2)

$$\begin{aligned} g_1 &= 1 \\ g_{1+i} &= X_i \\ g_{1+p+i} &= X_i^2 \\ g_{1+2p+(i-1)(i-2)/2+j} &= X_i X_j \end{aligned}$$

Equation (2) usually arises from fitting a model to data from a designed experiment and thus is, in some sense, approximate and subject to random error. Usually, we ignore this feature of the relationship between the factors and the responses, except to note that the approximation is apt to be particularly imprecise near the edges of the experimental region determined by the factors.

The general problem, then, is to find the functional form of the relationship between the response and the factors, and to use this relationship to find settings of the factors that which optimize the response. Many statistical packages are able to estimate the parameters, β_i , which in turn can be used to create a set of predicted values of the response over the entire design space.

Profile Ridge Plots

In response surface experiments, the factors are treated as continuous variables, although they are only observed at a finite number of values. Using the range of each factor, we interpolate over the design space in order to optimize the response. For regression analysis of response surface experiments, canonical analysis of the shape of the surface and response ridge optimization are two common numerical techniques.

We illustrate these concepts with an example by Khuri and Cornell (1987) on optimizing crop yield of snap beans in an agricultural experiment. The objective of the experiment is to find the combination of levels of fertilizer components that maximizes snap bean yield. Previous experimentation has isolated three components---Nitro Chalk (NC), Super Phosphate (SP), and Potash Muriate (PM)---thought to affect the crop yield, as well as likely ranges of each

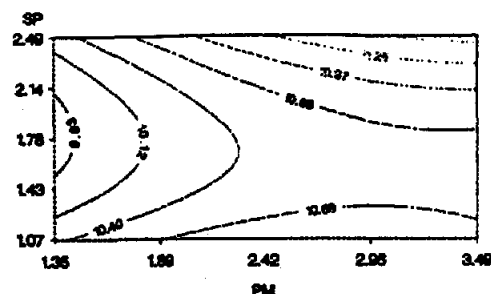


Figure 2: Contour Plot of Snap Bean Data

in which to find the optimal levels. The data for this experiment are given in the appendix.

Khuri and Cornell (1987) used a uniform precision central composite design (CCD-U) with 20 runs in their analysis. A response surface regression was performed on the data and the following estimated model was found.

$$\begin{aligned} \hat{Y} &= 12.45 \\ &- 1.48 \cdot PM - 2.30 \cdot SP + 0.97 \cdot NC \\ &- 0.24 \cdot PM^2 + 1.12 \cdot SP^2 - 0.27 \cdot NC^2 \\ &+ 0.31 \cdot PM \cdot SP + 0.70 \cdot NC \cdot PM \\ &- 0.60 \cdot NC \cdot SP \end{aligned}$$

The canonical analysis of the shape of the fitted surface results in a saddle point, a point of inflection on the general response surface that is neither a minimum nor a maximum.

A common technique in response surface experimentation is to examine a plot of the predicted response as a function of two factors, usually represented two-dimensionally as response contours. Figure 2 is a plot of predicted yield as a function of the amounts of superphosphate and potash muriate. This plot shows the saddle nature of the predicted response surface. Since all three factors affect the response, it is not possible to use a plot such as this to identify a region of maximum yield.

As an alternative, the method of ridge analysis can be used to locate the optimum response within the design space. This technique computes the estimated ridge of optimum response for increasing radii from the center of the original design. The optimum values start at the center point and follow the steepest descent or ascent, respectively. Table 2 presents the numerical results of this iterative process for finding the maximum. As this table shows, within the region of experimentation the expected maximum is near the point $(PM, SP, NC) = (2.504, 2.904, 2.748)$. This suggests that a relatively high level of superphosphate, a low level of nitrochalk, and a level of potash in the middle of its range will give maximum bean yield. Note, however, that the standard error increases as

Table 2: Ridge of maximum response

PM	SP	NC	Predicted Value	Standard Error
2.410	1.780	3.615	10.46	0.404
2.496	1.851	3.444	10.57	0.402
2.527	1.969	3.342	10.69	0.399
2.536	2.089	3.260	10.84	0.397
2.537	2.207	3.184	11.02	0.399
2.535	2.324	3.110	11.24	0.411
2.530	2.441	3.036	11.50	0.439
2.525	2.557	2.964	11.79	0.488
2.518	2.673	2.891	12.12	0.560
2.511	2.788	2.819	12.48	0.654
2.504	2.904	2.748	12.88	0.771

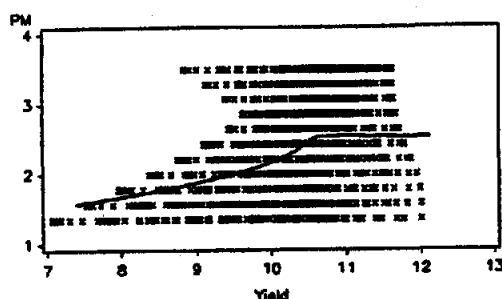


Figure 3: Potash Muriate Profile Ridge Plot

the optimum response ridge moves away from the center point of the design and toward the boundary of the design space, indicating that predictions near the edge are inherently less reliable.

The ridge analysis is just a table of numbers. To better understand these numbers, consider a plot of the ridge values superimposed on the profile plot for each factor versus response. This is a *profile ridge plot*. Each profile plot is simply a scatter plot of one factor's values against response predictions made over the entire region of experimentation. Such a plot shows the relationship of the factor to the response, as well as giving an indication of interactions with other factors. Superimposing the ridge analysis gives an indication of the movement of the optimal response ridge over the multidimensional response surface.

The profile ridge plot is constructed as follows. First, we predict the yield over a grid in the design space using the estimated model. Then we overlay the ridge information on each plot. Figures 3 -- 5 give profile ridge plots for each factor in the fertilizer example. Notice that while the absolute maximum yield is attained at high values of SP, some quite high yields are also predicted for low values of this factor. Thus, for example, if SP is an expensive fertilizer component, the profile ridge plots indicate that low levels of it can

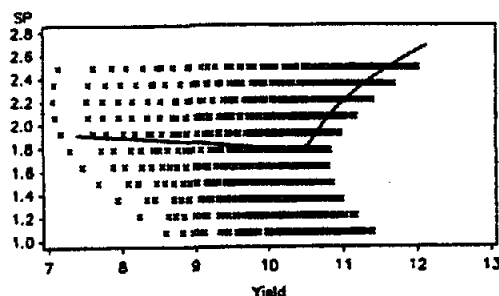


Figure 4: Super Phosphate Profile Ridge Plot

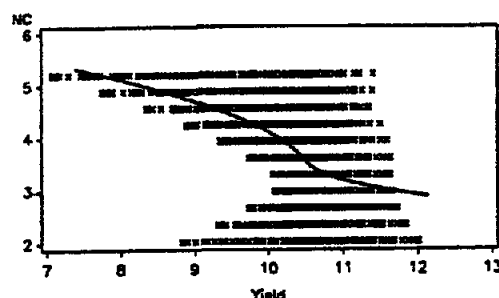


Figure 5: Nitro Chalk Profile Ridge Plot

be compensated for by appropriate settings for the other two components. Also note that near a PM level of 3, yield lies in a narrow band around 10.5, while at low levels of this factor the predicted yield is much more spread out over its range. This indicates that the other two factors are more important for determining predicted yield at low levels of potash.

Sorting Methods

Very often an experiment has multiple responses, and the objective is to explore how the factors affect all the responses simultaneously; ultimately, we wish to optimize as many of these responses as possible. In this general framework, the multiresponse problem is unsolvable; it is rarely the case that any settings of the factors result in optimum values for all the responses simultaneously. However, by studying the dependence, and the co-dependence, of the responses on the factors you can often see how to make the best compromise—for example, optimizing the yield while holding the cost under a certain level, or optimizing some weighted average of yield and cost.

The correct statistical approach for any particular multiresponse problem tends to be dependent on the context. Any of the following optimization techniques might be appropriate:

- View simultaneous/overlaid graphics.

- Optimize a *desirability* index.
- Use a grid search.

In the last section, we examined some of the uses of simultaneous graphics. While such plots can be informative for problems with few variables, they become extremely difficult to interpret as the number of factors and responses grows. Desirability indices are a technique for reducing a problem with multiple responses down to a single response problem, so that the usual techniques for optimizing a single response can be employed.

Another approach to exploring a multiresponse problem is simply to make a large number of predictions throughout the design region and employ standard sorting tools for data manipulation. Questions that can be answered using this method include the following:

- What settings of the factors optimize each response?
- What settings of the factors optimize a given response while assuring that other responses satisfy certain constraints?
- What are the ranges of the factor values that result in each response attaining a given percentage of its optimum level?

The experiment used to demonstrate these techniques is concerned with the development of a tire tread compound [Derringer and Suich (1980)]. There are three factors (silica, silane, and sulfur) and four responses (abrasion, modulus, elongation, and hardness). The data for this experiment are given in the appendix. Abrasion and modulus are to be maximized, while elongation and hardness are to lie as close as possible to target values of 500 and 67.5, respectively. The design is a standard central composite (Box-Wilson) response surface design with a modified axial extreme value. We fit a quadratic model to each response factor, individually, to create four quadratic models. We then predict the responses over an $11 \times 11 \times 11$ grid in the factors, transforming predicted elongation and hardness to produce variates that should be maximized in order to put these responses as close as possible to their respective targets. These transformations were chosen as

$$E' = -|E - 500|, \quad H' = -|H - 67.5|$$

We can determine which factor values optimize each response by simply sorting the predicted values and examining the factor values of the first (for a minimum)

or last (for a maximum) observation. By this method we find that the maximum predicted Abrasion is 231.3, achieved at (Silica, Silane, Sulfur) = (2, 65, 3); whereas Elongation is predicted to be closest to 500 when (Silica, Silane, Sulfur) = (1.1, 44, 1.95). Evidently these two objectives—maximizing abrasion and keeping elongation as close as possible to its target value—conflict with one another. For example, whereas silane should be set high to maximize abrasion, the predicted value of elongation closest to 500 has it set near the middle of its range. Furthermore, sulfur should be set high to maximize abrasion but low to bring elongation near its target value.

One solution to this conflict is to set constraints on elongation (for instance, to constrain it to lie between 475 and 525) and to optimize abrasion subject to these constraints. The result of this approach finds Abrasion = 130.9 and Elongation = 490 at (Silica, Silane, Sulfur) = (0.95, 65, 1.95). Notice that the compromise has significantly affected the optimum for abrasion.

Another type of compromise is to make sure that each response is within a certain percentage of its optimum value. Table 3 lists all of the predictions in the design space according to this type of compromise plan, keeping all of the responses within 40% of their optimum value.

One difficulty with optimizing multiple responses is that the mathematical problem is not well-defined; in general, it is not possible to find values of the factors that simultaneously optimize the predicted values of more than one response. A solution is to replace the multiplicity of responses with a single desirability index that combines all of the responses. The idea is first to construct a dimensionless measure of how far each response is from its optimum, and then to find those factor levels that optimize the average of the measure for all the responses. Derringer and Suich (1980) discuss methods for constructing such indices. Given high (H) and low (L) values for a response Y , they define a one-sided desirability transformation as

$$d = \begin{cases} 0 & Y \leq L \\ \left[\frac{Y-L}{H-L} \right]^r & L \leq Y \leq H \\ 1 & Y \geq H \end{cases}$$

and when a target value T for the response is given, they define a two-sided desirability transformation as

$$d = \begin{cases} \left[\frac{Y-L}{T-L} \right]^r & L \leq Y \leq T \\ \left[\frac{H-Y}{H-T} \right]^s & T \leq Y \leq H \\ 0 & Y \leq L \text{ or } Y \geq H \end{cases}$$

The values of r , s , and u are set to reflect the penalty for the response near the boundary. For example, the

Abrasion	Modulus	Elongation	Hardness	Silica	Silane	Sulfur
133.894	1489.04	418.291	70.2024	1.55	62	1.65
137.805	1428.86	421.426	70.1672	1.40	59	1.80
135.739	1434.06	422.382	68.9653	1.40	56	1.80
133.053	1416.80	426.453	67.7739	1.40	53	1.80

Table 3: Sorting Technique -- Percentage Constraint

Response	Low	Target	High
Abrasion	120		170
Modulus	1000		1300
Elongation	400	500	600
Hardness	60	67.5	75

Table 4: Target Values for Tire Tread Data

larger the value of r , the more undesirable it is to have the predicted response near the lower bound; while large values of s and u force a desirable predicted response to be nearer to the target. The overall desirability index is defined to be the geometric mean of the desirability indices d_i for the k responses

$$D = (d_1 \times d_2 \times \dots \times d_k)^{1/k}$$

Since the objective of the experiment at hand is to maximize abrasion and modulus while having elongation and hardness as close to their respective target values, we choose one-sided transformations for abrasion and modulus and two-sided transformations for elongation and hardness; here, r , s , and u are all set to 1. The respective high, low, and target values are given in Table 4.

We explore the desirability D not as a function of observed responses but as a function of predicted responses for given values of the factors---that is, as a continuous function of the factors themselves. The easiest way to do this is to create a grid of factor values and compute the predicted responses as well as the desirability at each one. When we consider the responses over an $11 \times 11 \times 11$ grid in the factors and sort to find the values of the factors that yield the optimal desirability, we find $D=0.571$ at (Silica, Silane, Sulfur) = (1.25, 53, 1.8), with Abrasion=130.38, Modulus=1370.88, Elongation=456.25, and Hardness=68.15. Note that while this is a crude way to search for the optimum, it compares quite well with the results of the numerical optimization algorithm employed by Derringer and Suich (1980).

Summary

We have discussed various ways to optimize responses in the context of designed experimentation. For multiple responses, we have shown how to go beyond simple techniques with more complex methods, such as desirability functions. While it is straightforward to use these techniques to optimize a single response, it is necessary to use new graphical techniques, such as the profile ridge plots, to understand the interrelationships among many variables.

References

- Box, G.E.P., and Bisgaard, S. (1987), "The Scientific Context of Quality Improvement," *Quality Progress*, 20(6), 54-61.
- Derringer, G. and Suich, R. (1980), "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology*, 12(4), 214-219.
- Khuri, A. and Cornell, J. (1987), *Response Surfaces: Design and Analysis*, in *Statistics: Textbooks and Monographs*, 81, New York: Marcel Dekker.
- Taguchi, G. (1978), "Off-line and On-line Quality Control Systems," *Proceedings of International Conference on Quality Control*, Tokyo, Japan.
- Taguchi, G. and Wu, Y. (1980), "Introduction to Off-line Quality Control," Central Japan Quality Control Association. (Available from American Supplier Institute, 32100 Detroit Industrial Expressway, Romulus, MI 48174.)

Appendix: Data Sets

Fertilizer Example
Khuri and Cornell (1987)

Y	PM	SP	NC
7.98	0.60	1.78	3.62
11.28	1.35	1.07	2.03
8.44	1.35	1.07	5.21
13.19	1.35	2.49	2.03
7.71	1.35	2.49	5.21
12.08	2.42	0.59	3.62
8.26	2.42	1.78	0.94
10.14	2.42	1.78	3.62
10.22	2.42	1.78	3.62
10.53	2.42	1.78	3.62
9.50	2.42	1.78	3.62
11.53	2.42	1.78	3.62
11.02	2.42	1.78	3.62
7.87	2.42	1.78	6.29
11.06	2.42	2.97	3.62
8.94	3.49	1.07	2.03
10.90	3.49	1.07	5.21
11.85	3.49	2.49	2.03
11.03	3.49	2.49	5.21
10.43	4.22	1.78	3.62

Y = Yield

PM = Potash muriate

SP = Super phosphate

NC = Nitro chalk

Tire Composition Example
Derringer and Suich (1980)

A	M	E	H	Sc	Sn	Su
102	900	470	67.5	0.7	40	2.8
120	860	410	65	1.7	40	1.8
117	800	570	77.5	0.7	60	1.8
198	2294	240	74.5	1.7	60	2.8
103	490	640	62.5	0.7	40	1.8
132	1289	270	67	1.7	40	2.8
132	1270	410	78	0.7	60	2.8
139	1090	380	70	1.7	60	1.8
102	770	590	76	0.3835	50	2.3
154	1690	260	70	2.0165	50	2.3
96	700	520	63	1.2	33.67	2.3
163	1540	380	75	1.2	66.33	2.3
116	2184	520	65	1.2	50	1.4835
153	1784	290	71	1.2	50	3.1165
133	1300	380	70	1.2	50	2.3
133	1300	380	68.5	1.2	50	2.3
140	1145	430	68	1.2	50	2.3
142	1090	430	68	1.2	50	2.3
145	1260	390	69	1.2	50	2.3
142	1344	390	70	1.2	50	2.3

A = Abrasion, M = Modulus

E = Elongation, H = Hardness

Sc = Silica, Sn = Silane

Su = Sulfur