World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:3, No:1, 2009

Developing New Processes and Optimizing Performance Using Response Surface Methodology

S. Raissi*

***Abstract*—** Response surface methodology (RSM) is a very efficient tool to provide a good practical insight into developing new process and optimizing them. This methodology could help engineers to raise a mathematical model to represent the behavior of system as a convincing function of process parameters.
Through this paper the sequential nature of the RSM surveyed for process engineers and its relationship to design of experiments (DOE), regression analysis and robust design reviewed. The proposed four-step procedure in two different phases could help system analyst to resolve the parameter design problem involving responses. In order to check accuracy of the designed model, residual analysis and prediction error sum of squares (PRESS) described.
It is believed that the proposed procedure in this study can resolve a complex parameter design problem with one or more responses. It can be applied to those areas where there are large data sets and a number of responses are to be optimized simultaneously. In addition, the proposed procedure is relatively simple and can be implemented easily by using ready-made standard statistical packages.

***Keywords*—** Response Surface Methodology (RSM), Design of Experiments (DOE), Process modeling, Process setting, Process optimization.

I. Introduction

Experimentation is made to determine the effect of the independent variable (factor) on the dependent variable say response of a process and a relation between them usually illustrated through a regression model by using experimental data. Statistical design of experiment (DOE) is a well known efficient experimentation technique and has been applied in a broad range of fields such as automobile, food, drug, textile, composites and so on industries, to produce high quality products, to operate them more economically, to ensure more stable and reliable process [1], [2].

Theodore T. Allen [3] called DOE as the jewel of quality engineering in his valuable book in the field of the Six Sigma. The studies including application of DOE methods have been made for more than 40 years and the advance of DOE applications has been increasingly assisted by the developments in the field of computer science.

Carr et al. [4], applied statistical program planning for process improvement to reduce the process development time by applying fractional factorial design. Lind et al. [5] applied response surface methodology (RSM) and full two-level factorial design to a chemical process in which antibiotic was produced. Xu et al. [6] used statistically based experimental designs for the medium optimization of an important medical microorganism. Andersons [7], applied design of experiments technique to the problem of preparing microwave popcorn. More studies are available in e-journals [8], [9].

There are a lot of DOE methods and the selection of them is made according to objectives and the number of examined factors [2]. The objectives of the experiment can be classified as screening, comparing and applying RSM. Screening experiment is applied to determine the most effective factors on the process response. RSM is generally used to find the optimal condition by using usually quadratic polynomial model and it is applied in consequence of a screening experiment.

RSM is primarily relevant when the decision-maker desires (1) to create a relatively accurate prediction of engineered system input output relationships and (2) to “tune” or optimize thoroughly of the system being designed. Since these methods require more runs for a given number of factors than screening using fractional factorials, they are generally reserved for cases in which the importance of all factors is assumed, perhaps because of previous experimentation.

RSM is widely used and the prediction models generated by them can yield 3D surface plots. The methods are based on three types of design of experiments matrices. First, “central composite designs” (CCDs) are matrices corresponding to (at most) five level experimental plans from Box and Wilson [3]. Second, “Box Behnken Designs” (BBDs) are matrices corresponding to three level experimental plans from Box, Behnken [3]. Third, Allen et al. [3] proposed methods based on so-called “expected integrated mean squared error optimal” (EIMSE-optimal) designs. EIMSE-optimal designs are one type of experimental plan that results from the solution of an optimization problem.

II. Response Surface Methodology

Often engineering experimenters wish to find the conditions under which a certain process attains the optimal results. That is, they want to determine the levels of the design parameters at which the response reaches its optimum. The optimum could be either a maximum or a minimum of a function of the design parameters. One of methodologies for obtaining the optimum is response surface technique.

Response surface methodology is a collection of statistical and mathematical methods that are useful for the modeling and analyzing engineering problems. In this technique, the main objective is to optimize the response surface that is influenced by various process parameters. Response surface

* - Sadigh Raissi, Islamic Azad University, South Tehran Branch, Industrial Eng. Dept., Raissi@azad.ac.ir

methodology also quantifies the relationship between the controllable input parameters and the obtained response surfaces.

The design procedure of response surface methodology is as follows [10]:

(i) Designing of a series of experiments for adequate and reliable measurement of the response of interest.

(ii) Developing a mathematical model of the second order response surface with the best fittings.

(iii) Finding the optimal set of experimental parameters that produce a maximum or minimum value of response.

(iv) Representing the direct and interactive effects of process parameters through two and three dimensional plots.

If all variables are assumed to be measurable, the response surface can be expressed as follows: .

$$y = f(x_1, x_2, ...x_k) \tag{1}$$

The goal is to optimize the response variable y. It is assumed that the independent variables are continuous and controllable by experiments with negligible errors. It is required to find a suitable approximation for the true functional relationship between independent variables and the response surface. Usually a second-order model is utilized in response surface methodology.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \beta_{ij} x_i x_j + \varepsilon \tag{2}$$

where ε is a random error. The β coefficients, which should be determined in the second-order model, are obtained by the least square method. In general (2) can be written in matrix form.

$$\mathbf{Y} = \mathbf{bX} + \mathbf{E} \tag{3}$$

where Y is defined to be a matrix of measured values, **X** to be a matrix of independent variables. The matrixes **b** and **E** consist of coefficients and errors, respectively. The solution of (3) can be obtained by the matrix approach.

$$\mathbf{b} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathbf{T}}\mathbf{Y} \tag{4}$$

where $\mathbf{X}^{\mathbf{T}}$ is the transpose of the matrix **X** and $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$ is the inverse of the matrix $\mathbf{X}^{\mathbf{T}}\mathbf{X}$.

The mathematical models were evaluated for each response by means of multiple linear regression analysis. As said previous, modeling was started with a quadratic model including linear, squared and interaction terms. The significant terms in the model were found by analysis of variance (ANOVA) for each response. Significance was judged by determining the probability level that the F-statistic calculated from the data is less than 5%. The model adequacies were checked by R^2, adjusted-R^2, predicted-R^2 and prediction error sum of squares (PRESS) [10]. A good model will have a large predicted R^2, and a low PRESS. After model fitting was performed, residual analysis was conducted to validate the assumptions used in the ANOVA. This analysis included calculating case statistics to identify outliers and examining diagnostic plots such as normal probability plots and residual plots.

Maximization and minimization of the polynomials thus fitted was usually performed by desirability function method, and mapping of the fitted responses was achieved using computer software such as Design Expert.

III. Response Surface Methodology and Robust Design

RSM as an important branch of experimental design is one of the most important technique in developing new processes and optimizing their performance. The objectives of quality improvement, including reduction of variability and improved process and product performance, can often be accomplished directly using RSM. It is well known that variation in key performance characteristics can result in poor process and product quality.

During the 1980s considerable attention was given to process quality, and methodology was developed for using experimental design, specifically for the following:

1. For designing or developing products and processes so that they are robust to component variation.

2. For minimizing variability in the output response of a product or a process around a target value.

3. For designing products and processes so that they are robust to environment conditions.

By robust means that the product or process performs consistently on target and is relatively insensitive to factors that are difficult to control. Professor Genichi Taguchi used the term robust parameter design (RPD) to describe his approach to this important problem. Essentially, robust parameter design methodology prefers to reduce process or product variation by choosing levels of controllable factors (or parameters) that make the system insensitive (or robust) to changes in a set of uncontrollable factors that represent most of the sources of variability. Taguchi referred to these uncontrollable factors as noise factors. RSM assumes that these noise factors are uncontrollable in the field, but can be controlled during process development for purposes of a designed experiment.

IV. The Sequential Nature of the Response Surface Methodology

Most applications of RSM are sequential in nature and can be carried out based on the following phases.

Phase 0: At first some ideas are generated concerning which factors or variables are likely to be important in response surface study. It is usually called a screening experiment. The objective of factor screening is to reduce the list of candidate variables to a relatively few so that subsequent experiments will be more efficient and require fewer runs or tests. The purpose of this phase is the identification of the important independent variables.

Phase 1: The experimenter's objective is to determine if the current settings of the independent variables result in a value of the response that is near the optimum. If the current settings or levels of the independent variables are not consistent with optimum performance, then the experimenter must determine a set of adjustments to the process variables that will move the process toward the optimum. This phase of RSM makes considerable use of the first-order model and an optimization technique called the method of steepest ascent (descent).

Phase 2: Phase 2 begins when the process is near the optimum. At this point the experimenter usually wants a model that will accurately approximate the true response function within a relatively small region around the optimum. Because the true response surface usually exhibits curvature near the optimum, a second-order model (or perhaps some higher-order polynomial) should be used. Once an appropriate approximating model has been obtained, this model may be analyzed to determine the optimum conditions for the process. This sequential experimental process is usually performed within some region of the independent variable space called the operability region or experimentation region or region of interest.

V. Multi-response surface method

In practical cases, there are many situations where the researchers encounter to multi-responses. In such cases surveying two or more response variables are critical.

Over the last few years in many manufacturing organizations, multiple response optimization problems were resolved using the past experience and engineering judgment, which leads to increase in uncertainty during the decision-making process.

Myers and Carter [11] proposed an algorithm for obtaining the optimal solutions of the dual-response surface method (DRSM). Their method assumed that the DRSM includes a primary response, y_p and a constraint response, y_s . Both y_p and y_s can be respectively fitted as a quadratic model as follows:

$$\begin{aligned} y_p &= \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \beta_{ij} x_i x_j + \varepsilon_p \\ y_s &= \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \beta_{ij} x_i x_j + \varepsilon_s \end{aligned} \tag{5}$$

where the β 's and γ 's represent the unknown coefficients, and ε_P and ε_s denote the random errors, respectively. The random errors are assumed to possess a normal distribution with mean 0 and variance σ^2.

The DRSM attempts to obtain a set of *X*, which can optimize $\hat{y}_p$ subjected to the constraint $\hat{y}_s = c$, where *C* is a constant.

The desirability function was originally developed by Harrington [12] to simultaneously optimize the multiple responses and was later modified by Derringer and Such [13] to improve its practicability. The desirability function approach is one of the most frequently used multi-response optimization techniques in practice. The desirability lies between 0 and 1 and it represents the closeness of a response to its ideal value. If a response falls within the unacceptable intervals, the desirability is 0, and if a response falls within the ideal intervals or the response reaches its ideal value, the desirability is 1.

Meanwhile, when a response falls within the tolerance intervals but not the ideal interval, or when it fails to reach its ideal value, the desirability lies between 0 and 1. The more closely the response approaches the ideal intervals or ideal values, the closer the desirability is to 1. According to the objective properties of a desirability function, the desirability function can be categorized into the following three forms:

1- Nominal-the-best (NB)
2- Larger-the-better (LB)
3- Smaller-the-better (SB).

The total desirability is defined as a geometric mean of the individual desirability:

$$D = \left(d_1 \times d_2 \times \ldots \times d_k\right)^{1/k} \tag{6}$$

where *D* is the total desirability and d_i is the i^th^ desirability, $i = 1, 2, \ldots, k$. If all of the quality characteristics reach their ideal values, the desirability d_i is 1 for all *i*. Consequently, the total desirability is also 1. If any one of the responses does not reach its ideal value, the desirability d_i is below 1 for that response and the total desirability is below 1. If any one of the responses cannot meet the quality requirements, the desirability d_i is 0 for that response. Total desirability will then be 0. The desirability function is a scale-invariant index which enables quality characteristics to be compared to various units. Therefore, the desirability function is an effective means of simultaneously optimizing a multi-response problem.

VI. Model Adequacy Checking

Model adequacy is always necessary to:

1. Examine the fitted model to ensure that it provides an adequate approximation to the true system;

2. Verify that none of the least squares regression assumptions are violated. There are several techniques for checking model adequacy.

Residual Analysis: The residuals from the least squares fit, defined by $e_i = y_i - \hat{y}_i$, i = 1, 2,…, n, play an important role in judging model adequacy. Many response surface analysts prefer to work with scaled residuals, in contrast to the ordinary least squares residuals. These scaled residuals often convey more information than do the ordinary residuals.

The standardizing process scales the residuals by dividing them by their average standard deviation. In some data sets, residuals may have standard deviations that differ greatly. There is some other way of scaling that takes this into account. Let's consider this.

The vector of fitted values $\hat{y}_i$ corresponding to the observed values y_i is

$$\hat{\mathbf{y}} = \mathbf{Xb} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{Hy} \quad (7)$$

The n x n matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is usually called the hat matrix because it maps the vector of observed values into a vector of fitted values. The hat matrix and its properties play a central role in regression analysis.

Since $e_i = y_i - \hat{y}_i$, there are several other useful ways to express the vector of residuals

$$\mathbf{e} = \mathbf{y} - \mathbf{Xb} = \mathbf{y} - \mathbf{Hy} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (8)$$

The "prediction error sum of squares" (PRESS) proposed in [6, 7], provides a useful residual scaling

$$PRESS = \sum_{i=1}^{n}\left(\frac{e_i}{1-h_{ii}}\right)^2 \quad (9)$$

From (7), it is easy to see that the PRESS residual is just the ordinary residual weighted according to the diagonal elements of the hat matrix h_{ii}. Generally, a large difference between the ordinary residual and the $PRESS$ residual will indicate a point where the model fits the data well, but a model built without that point predicts poorly.

VII. Conclusion

RSM is one of the most important tools in developing new processes and optimizing their performance. This approach utilizes data gathered from process through statistically design of experiments method. In the practical application of RSM, it is necessary to develop an approximating model for the true response surface. The underlying true response surface is typically driven by some unknown physical mechanism. The approximating model is based on observed data from the process or system and is an empirical model. Multiple regression as a collection of statistical techniques useful for building the types of empirical models required in RSM.

There are some approaches to select the appropriate subset of variables for a regression model. The most important ones which could be followed by most standard statistical packages are all possible regression technique, the stepwise regression methods, the forward selection method and the backward elimination routine.

Through this paper, readers could be familiarize to RSM method technically and they could track their model buildings effectively. The residual analysis method and the prediction error sum of squares proposed for evaluating the capability of the designed models. Researcher could follow standard optimization techniques such as the differentiation, the operation research method to set their process in optimum conditions.

In this article we are concentrated mostly on building the empirical models and practically hah not focused on the details of experimental design and optimizing the models.

References

[1] D.C. Montgomery, *Design and Analysis of Experiments*, 3rd ed., John Wiley & Sons, New York, 1991, pp. 270–569.
[2] NIST/SEMATECH e-Handbook of Statistical Methods, 2005. http://www.itl.nist.gov/div898/handbook/.
[3] Theodore T. Allen, *Introduction to Engineering Statistics and Six Sigma - Statistical Quality Control and Design of Experiments and Systems*, London: Springer-Verlag, 2006, pp 241.
[4] J.M. Carr, E.A. McCracken, Statistical program planning for process development, Chemical Engineering Progress. Vol. 56, no. 11, 1960, pp. 56–61.
[5] E.E. Lind, J. Goldin, J.B. Hickman, Fitting yield and cost response surfaces, Chemical Engineering Progress. Vol. 56, no. 11, 1960, pp. 62–68.
[6] C.-P. Xu, S.-W. Kim, H.-J. Hwang, J.-W. Yun, Application of statistically based experimental designs for the optimization of exopolysaccharide production by Cordyceps milltaris NG3, Biotechnol. Appl. Biochem, vol. 36, 2002, pp. 127–131.
[7] M.J. Anderson, H.P. Anderson, Applying DOE to microwave popcorn, Process Ind. Quality, 1993, pp. 30–32.
[8] S.E. Kruger, I.C. Silva, J.M.A. Rebello, Factorial design of experiments applied to reliability assessment in discontinuity mapping by ultrasound, NDT Net, vol. 3, no. 11, 1998.
[9] B.V. Mehta, H. Ghulman, R. Gerth, Extrusion die design: a new methodology of using design of experiments as a precursor to neural networks, JOM-e , vol. 51, no. 9, 1999.
[10] Jae-Seob Kwak, Application of Taguchi and response surface methodologies for geometric error in surface grinding process, International Journal of Machine Tools & Manufacture, vol.45, 2005, pp. 327–334
[11] Myers, R.H.; Carter, W.H., Jr. Response Surface Techniques for Dual Response Systems. Technometrics 1973, vol. 15, no. 2, pp. 301–317.
[12] Harrington, E.C., Jr. The Desirability Function. Industrial Quality Control, 1965, vol. 21, no. 10, pp. 494 – 498.
[13] Derringer, G.; Suich, R. Simultaneous Optimization of Several Response Variables. Journal of Quality. Technology, 1980, vol. 12, no. 4, pp. 214 –219.