Discussion

DENNIS K. J. LIN

Pennsylvania State University, University Park, PA 16802-1913

Box's and Myers's papers provide two important discussions of response surface methodology (RSM). Dr. Box and Patrick Liu's paper (Part I) introduces the basic concept of RSM through a clever example that can be used in almost any classroom teaching situation. This paper is followed by a general discussion on scientific learning and robustness to demonstrate the value of RSM (Part II). Dr. Myers's paper discusses potential future directions for RSM. Needless to say, both are a must for all practitioners and researchers to read.

Box and Liu's paper illustrates the fundamental philosophy and thinking process of RSM and related methodologies to be used in each stage. Besides the important messages behind it, the beauty of the paper is its cleanness. Dr. Box is probably one of the very few people who can introduce RSM so clearly. RSM can be roughly classified into the following steps: screening, steepest ascent, factorial experiment, composite design, canonical analysis, ridge analysis, and optimization. Basically, RSM emphasizes the important characteristics of the sequential nature of scientific discovery.

- In the screening stage, the goal is to detect the most relevant factors (those which may contribute to main or interaction effects). Rather than depending solely on so-called professional knowledge, experimenters may run a small size experiment to confirm (or discover) the key factors for the next stage. Detecting relevant factors has received a great deal of attention recently (see, e.g., Lin (1993a, 1993b, 1995) and references therein). That a simple first-order model fits well for the "survived variables" more or less indicates that the current region is far away from the optimum. A steepest ascent (or descent) step is then recommended.
- In the steepest ascent stage, the goal is merely to find out the direction for improvement and then to determine a new center point for future investigation.
- The next step is to run another set of factorial experiments. If the data present curvature (significant interaction or quadratic effects), then an assembled composite design is suggested. In Box's paper, the miscalculation complicates the findings, but this is not uncommon in reality.
- Data analysis for the composite design is then performed. Specifically, canonical analysis, as well as ridge analysis, is commonly used to detect the optimal setting which yields the optimum response.

As Box has stressed in many of his publications, experimentation should be undertaken with dual goals, namely, optimization and advancement of scientific knowledge. If the process is understood very well, then optimization is the over-riding goal of the study. However, in most situations, the behavior of the process under varying conditions is not fully understood. Thus, RSM requires the emphasis of sequential experimentation with knowledge gained at each step. Conversely, a one-step optimization procedure requires a priori extensive professional knowledge and adds little to the knowledge base.

Unlike the classical examples in Box and Wilson (1951) and Box, Hunter, and Hunter (1978), the helicopter example has two objectives to be optimized—the location and the dispersion effects. Recently much research work has been done in this area. Notably, many articles have appeared in the *Journal of Quality Technology* on the subject of "dual response surface" (see, e.g., Vining and Myers (1990) and, recently, Kim and Lin (1998) and references therein). A related and broader point that needs to be carefully addressed is multiple response surface problems. While Box's paper clearly indicates when and how to optimize for the univariate response, it is not easy to extend those ideas to multivariate cases, especially when the responses are correlated or even contracted. This is one of the most important and timely

Dr. Lin is a Professor in the Department of Management Science and Informational Systems and in the Department of Statistics. He is a Senior Member of ASQ. His email address is lin@net12pc248.smeal.psu.edu.

research subjects for investigation. Some work by André Khuri may be a good starting point (see, e.g., his review paper Khuri (1996, Ch. 12)).

"Classical" research is still challenged by basic assumptions such as normality, independence, and equal variance. Instead of making other assumption systems, as is done in many other publications, it is perhaps more important to study how *robust* those assumptions are in practice. In other words, if those common assumptions do not match 100% with reality, then we must question how effective is the general methodology. The issue of robustness certainly should receive much more attention.

Box has introduced the key ideas of RSM, followed by the concept of scientific leaning and robustness. The concept of scientific learning is an important and difficult subject. There is much research in this area, mainly in education and philosophy programs. Box has introduced the key concept in a rather simple manner. Moreover, scientific learning relates directly to the central issues of RSM. In the rest of this discussion, I will mainly comment on the last portion of the paper that has to do with robustness and optimality.

Robustness and Optimality

Robustness and optimality are two important and useful concepts in statistical procedures. Optimality of a procedure requires specific assumptions, and more likely, a specific model. On the other hand, robustness implies insensitivity to models and assumptions. Often, something that is optimal is not robust and vice-versa. Box states that the effect of a departure from an assumption depends on the magnitude of the deviation from the assumption and a measure of the insensitivity of the response to such deviations. He calls this measure a *robustness factor*. He also states that many so-called robust procedures have been developed that are insensitive to certain assumptions, but quite sensitive to others. A truly robust procedure will be insensitive to all assumptions. This is a very important point that many overlook. We should not assume that a procedure is robust to all assumptions solely because it is "distribution-free" or non-parametric in nature.

Experimental designs may also be assessed in terms of optimality and robustness. There are several different methods for measuring the optimality of a design based on a given model. Two of the most popular methods of alphabetical optimality are A- and D-optimality (see Box and Draper (1987)). Thus, what constitutes an optimal design depends on the chosen optimality criterion and the "correct" form of model for the response. Different optimality criteria and different models may lead to different designs. Therefore, it seems that designs that are "optimal" are usually not robust.

In assessing the robustness of an experimental design, we have a similar situation. We can only speak of robustness if we define the characteristic to which we wish to be robust. Moreover, we must define a measure of robustness. As with optimality criteria, there does not appear to be one "best" robustness measure of an experimental design. Thus, in designing an experiment, we face the same tradeoffs as in designing a product or process.

In product design, we want to optimize the mean response and minimize the variance of the response. Similarly, in experimental design, we want to select a design that is optimal in some sense, but also insensitive to deviations from the assumptions used to determine such optimality. In product design, optimization of the mean response has been of primary interest. While there was some early work on designing robust products, little attention was paid to variation reduction until the 1980's. Analogously, many papers have been published on optimal designs, but there does not seem to be much activity in developing experimental designs that are robust themselves. Many screening designs study factors at two levels each in order to conserve precious experimental resources. However, these designs may provide misleading results if there is substantial curvature in the response over the experimental region studied. Augmenting these two-level designs with center points or running a composite design makes the design more robust to the assumption of strictly linear effects. Thus, a two-level design may be optimal in terms of minimizing experimental resources, but the composite design is more robust. In other words, RSM, as developed by Box and Wilson (1951) are, in some sense, robust experimental procedures. Box recommends sequential experimentation and using the results to simultaneously improve the results and the experimenter's knowledge. We should focus on the optimality and robustness of a *sequential* series of experimental designs, not of individual designs.

So, we have some unanswered questions. How do we measure robustness of experimental designs? Assuming we can develop reasonable measures of robustness, how can we make appropriate tradeoffs between robustness and optimality of designs? Box provides some insight with his discussion of robustness factors. It is my opinion that more research should be focused in this area. It may be that robustness can not be studied as rigorously (in mathematical terms) as optimality. However, this lack of structure does not lessen its importance.

We should pay attention to Box's point that statisticians need to work on more unstructured problems. We, as statisticians, seem to believe there is always a "best" way to do things. Here, at Penn State, our second-year students take a consulting practicum where they help graduate students from other disciplines with their research. There they learn that the textbook approaches are not always feasible. Most students find this lack of structure a bit intimidating compared to the other courses they take. However, this work on problems that are often unstructured provides valuable experience for the students and solves real-world problems.

Dr. Myers's paper discusses how to correctly apply RSM in non-standard situations. I have to agree with his comment that "the type of problems that are coming to the table have required that statistics research be moved to a new level." These situations include (in the order that appears in Myers's paper): Taguchi's robust design (for variation reduction); computer generated design; robust design; Bayesian design; generalized linear models; nonparametric/semiparametric; multiple response; and restriction in randomization.

Taguchi's Robust Design (for Variation Reduction)

One may or may not agree with the "Taguchi method," but it seems that we all agree on the importance of variation reduction (this was implicitly indicated in Box's paper). From an RSM perspective, the major task is variation modeling while simultaneously optimizing mean and variation responses. Simultaneously (empirically) modeling the mean and the variation of responses remains a problem for researchers because the common assumptions (such as equal variance and independence) are no longer valid. Most techniques for modeling dispersion effects assume no location effects and vice-versa, which results in a chicken and egg problem. However, the problem of simultaneously optimizing mean and variation responses, known as the dual response problem, has received a great deal of attention recently. See Kim and Lin (1998) for recent developments.

Computer Generated Designs

There are some design situations where a computer can perform a job better than the human brain and other situations where the opposite holds true. The successful experimenter needs the ability to identify which situation occurs when. Computers need clear guidelines in terms of a set of criteria. Unfortunately, all criteria require a prior assumption. When much is known, that is, when the assumptions or model is close to reality, computer generated designs can be useful. When not much is known, for example, when the optimality criterion is not well defined, standard designs may be more appealing. If you are a true believer in alphabetic optimality (such as D-optimality), then computer generated designs can be effective. Research in this area focuses on finding better algorithms to optimize a given criterion.

Robust Design

This is one of the key concepts in the original RSM designs for empirical model building when knowledge of the functional form of the model is lacking. It is interesting to note that the standard response surface design works well here. The research direction here is to define a good measurement for robustness of this kind so that the *optimal* robust design can be found.

Bayesian Design

Unlike robust design, the Bayesian design requires a clear prior and a clear likelihood (model) function at the beginning. There are situations, such as many applications in the biological sciences, where these designs are appreciated. These prior assumptions need to be handled with care. Inappropriate likelihood functions and priors may lead to unappealing design points. Note that strong prior knowledge always helps. If it can be written in a functional form, Bayesian design is definitely recommended.

Generalized Linear Models for Non-Normal Cases

Although it is believed that the normality assumption is a robust assumption, in reality, there are many cases where it does not apply. A typical approach to induce normality is to use the Box-Cox transformation in such situations, as suggested in many of Box's publications. Another possibility is to adapt the approach of generalized linear modeling. Generalized linear models are most often used when the usual identical independent normal distribution assumption is violated. Due to today's computing power, many theoretical results from the past can now be physically performed. From a data analysis perspective, there is not much theoretical work beyond the original McCullugh and Nelder (1989) book. There is, however, a lot of room for research on *experimental design* in this area.

Nonparametrics/Semiparametrics

Modeling is one of the key issues in RSM. The concept of empirical model building, along with Taylor expansion, always leads to the use of low-order polynomial fitting. This is, of course, not always appropriate. Will nonparametric (or even semiparametric) methods work better here? There are two major problems with these methods when used in RSM. First, nonparametrics require more data points to obtain a reliable fitting. This may not be feasible in most RSM applications. Second, nonparametric methods may yield a better fitting model, but are, in general, difficult to use for optimization. More research needs to be done in this area.

Multiple Response

The multiple response problem is clearly one of my top concerns for the future of RSM. As mentioned in Box's paper, this common problem has not received the attention it deserves. The whole procedure of RSM (such as screening, steepest ascent, composite design, and canonical analysis) works excellently for the univariate cases, but may not work at all for multivariate cases. How to appropriately apply these concepts and methodologies to multivariate cases deserves careful study. The use of weighted average on the response variables to reduce the multiple response problem to a single response problem may not be feasible, as noted in many articles cited in Myers's paper. Most of the recent work in this area has focused on the optimization issue. I believe that this is a good beginning and anticipate more work will be done in the near future.

Restriction in Randomization

Restriction in randomization is a "classical" problem, particularly in agricultural applications. When applied to RSM, many other assumptions need to be added (see also the comments on Bayesian design as well as generalized linear models). RSM has been used rather extensively in the chemical and processes industries. In other fields, it may be, for example, that the time between runs is substantially long, requiring some form of blocking to be used in the RSM experiments.

In his paper, Dr. Myers, using his vast knowledge and experience, provides a large set of research problems in RSM, including what has been done and what should be done. I want to express my personal gratitude to him for sending us such an important message.

General Comments

I think that RSM can be roughly described by Figure 1. The experimenter begins with a "simple" model (relatively inexpensive in data collection and data analysis). If it fits well, then the current experimental region is probably away from the optimal point. The objective is then to find the *direction* for improvement and then find a new center point to start all over again. This is the fundamental idea of steepest ascent/descent. If the "simple" model does not fit well, then either we have chosen an insufficient set of candidate variables or the current setting is close to the target. For the first situation, we add or drop some variables. For the second situation, a more "complicated" model is employed, which may require a few more experimental results (observations). Finally, we optimize the fitted final model to find the best setting for experimental variables. These steps form the entire RSM procedure of screening (add/drop variables), experimental region searching, model building, and optimization. Here, based on the Taylor expansion idea, we typically use a first-order polynomial model for the simple model and a second-order polynomial model for the complicated model. Note that the entire procedure works well for a smooth surface. When the (true) response surface is not "even," the methodology may not be as powerful as we expect. This limitation of RSM deserves notice.

There are three major stages of RSM: Data Collection (mainly design of experiments), Data Analysis (mainly model building), and Optimization. The key issues at each stage are:

Data Collection:	how to collect useful information,
Data Analysis:	how to model the observations, and
Optimization:	how to find the best combination(s) of the input variables that will optimize the response variable(s).

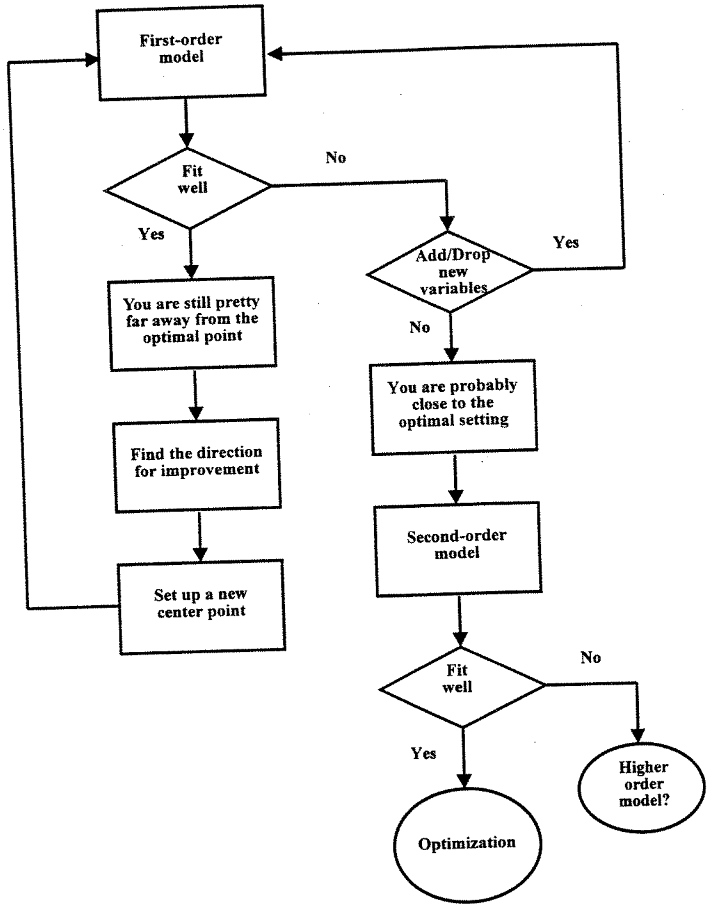

FIGURE 1. A Flow Chart for Response Surface Methodology.

For the data collection stage, the classical designs still dominate in practice for many reasons as discussed before. Recent research, such as that in optimal design, Bayesian design, and computer generated designs, all attempt to produce a better way for collecting data when the term "information" is clearly defined (i.e., the optimization criteria and, most likely, the underlying model are pre-specified). This is discussed in detail in Myers's paper.

For the data analysis stage, while standard low-order polynomials are still popular, many researchers have made important contributions to either (1) assumption failure in linear regression, such as normality, independence, and equal variance or (2) nonlinear fitting, such as nonparametric/semiparametric, kernel methods, and neural networks. There is lots of room for improvement. The use of today's computing power is certainly a good start. In particular, finding a "good" method for modeling variance (dispersion) is an important task. Some investigations of applying neural networks in response surface models have been conducted by the author. It is clear that some modifications are needed to implement neural networks in analyzing small data sets that are typical in RSM.

For the optimization stage, many classical methods are adequate for univariate response problems. The main research problem in the next few years will be to simultaneously optimize (or compromise) several variables. Many techniques in the optimization literature can be useful, although most of them do not consider the noise (uncertainty) issue. This leaves much room for new research.

It has been almost 50 years since the very original work of Box and Wilson's (1951) response surface paper. Since then, science has made many significant accomplishments in various disciplines. In statistics, most of the research has focused on using the power of today's computer in basically two ways: in research with a well-built theoretical foundation (bootstrapping, EM algorithm, generalized linear models, wavelet, Bayesian method computation, etc.) and in research whose theoretical base is yet to be built (fuzzy method, neural network, data mining, etc.). What about developments in RSM? As mentioned by Myers, the environment is different than it was 50 years ago; now we need to serve a wide variety of researchers, not only those working in the chemical processes. What are the new methodologies or philosophies beyond Box and Wilson's 1951 work? Hopefully, Drs. Box and Myers can give us more advice in this direction.

Acknowledgments

I would like to thank Mr. Richard N. McGrath (Pennsylvania State University) for helpful comments when preparing this paper.

Additional Reference

KHURI, A. (1996). "Design and Analysis of Experiment" in *Handbook of Statistics* edited by S. Ghosh and C. R. Rao. Elsevier, Amsterdam.