Title: Response Surface Methodology--Current Status and Future Directions. By: Myers, Raymond H., Journal of Quality Technology, 00224065, Jan1999, Vol. 31, Issue 1

Database: Business Source Complete

Response Surface Methodology--Current Status and Future Directions

AUTHOR:RAYMOND H. MYERS

TITLE:Response Surface Methodology--Current Status and Future Directions

SOURCE:Journal of Quality Technology 31 no1 30-74 Ja '99

ABSTRACT

This paper is a reflection on where response surface methodology (RSM) is at this point and what will likely be future directions. The emphasis in the last two decades on robust parameter design has brought attention to RSM as an alternative methodology for variance reduction and process improvement. While computer generated design technology has been beneficial to those who are interested in constructing RSM designs, changes are needed in this area to allow consideration of design robustness rather than design optimality. RSM is moving into areas involving the use of generalized linear models (GLM's), and optimal RS designs for these areas are either difficult or impossible to implement by the user. Example applications of GLM's include logistic and Poisson regression. Other RSM areas that will enjoy use by practitioners in the twenty-first century include multiple responses and nonparametric and semiparametric methods. In addition, design and analysis techniques for cases where natural restrictions in randomization occur need to be addressed further and communicated to users.

Key Words: Generalized Linear Models, Hard To Change Design Variables, Multiple Responses, Robust Design.

INTRODUCTION

THIS paper is not a comprehensive historical review of response surface methodology (RSM). Three of those have been written. See Hill and Hunter (1966); Mead and Pike (1975); and Myers, Khuri, and Carter (1989). Rather, the paper is a reflection on where RSM is at this point and what may be important future directions. The field of RSM has grown in recent decades and has been the beneficiary of advances in other fields of statistics. The range of types of practitioners interested in RSM is much greater than ever before, and, of course, as most professionals are aware, RSM cannot survive without practitioners. The clientele using RSM has broadened. It is no longer merely chemicals, foods, and other manufacturing processes that reach out for RSM tools. Professionals in biological, biomedical, and the rapidly growing biopharmaceutical area are quickly becoming intrigued with response surface ideas. And clearly we must be there to be able to solve their problems.

I became interested in RSM as a chemical engineering student in the later 1950's. As a graduate student in statistics I read the work by Box and Wilson (1951) and have remained enthusiastic about the field. I think we learn a great deal by doing retrospective thinking in any field. We learn both by past mistakes and good decisions that are made.

WHAT IS RSM? WHAT HAS CHANGED?

One usually views RSM in the context of design of experiments (DOE), model fitting, and process optimization. Recently, robust methods, nonparametric methods, Bayesian design, and generalized linear models (GLM's) have become a part of the total package. As a result, there have been substantial changes over the years, and much of this has been healthy, interesting, and exciting. However, the goals of RSM for the practitioner remains the same as described in the seminal paper by Box and Wilson. Namely, one is interested in designing an experiment on a set of quantitative process factors (or mixture of quantitative and qualitative factors) and analyzing the resulting data with the goal of determining conditions on the design variables that provide process improvement or perhaps even process optimization. Consequently, we can never lose sight of George Box's total contribution which goes far beyond the Box and Wilson work. A retrospective look at RSM and advice for new practitioners is not complete without a mention of the works by Box (1952, 1954) on first order designs and the groundbreaking work on rotatability in Box and Hunter (1957). A discussion of design robustness begins in the fifth section and remains a dominant theme throughout this paper. But it must be emphasized that the importance of design robustness was exemplified in Box and Draper (1959, 1963) and Box and Draper (1975), the former two papers suggesting a need to deal with model uncertainty in design and the latter paper dealing with the thesis that general robustness is preferable to optimal design. The multiple stage approach to experimental design is considered throughout this text, yet it is well known that augmentation of "first stage" designs is discussed in much of Box's work. Of course the space is not available here to cite all of his contributions. But it is clear that any solutions that we suggest for complicated problems in RSM in the twenty-first century will be influenced by his philosophy and genius. The changes that have come about reflect the facts that the practitioner is more optimistic and ambitious and that the problems have become more sophisticated and difficult to solve. To point to examples, we have multiple responses, more complicated models, and some scenarios unsuitable for polynomial approximations. We have incredibly large numbers of design variables and clearly nonnormal responses that question the suitability of classical designs. In other words, the types of problems that are coming to the table have required that statistics research be moved to a new level. This exciting prospect is, of course, greatly enhanced by the incredible changes that have occurred in computer technology.

It is interesting that the attention drawn to RSM has been very intense in the last 15-20 years, whereas this was not the case in the 1970's. I remember vividly a telephone conversation I had with Norman Draper in the late 1970's in which he said, "We who are working in the response surface area have to stick together. There aren't many of us left." He was correct, but since then many changes have occurred. Progress in the last 15 years has surpassed that of the previous 20-25 years as far as the enticement of practitioners is concerned. Of course, much of this had to do with the fact that computing capabilities lagged behind until the 1980's. It is healthy and fun to look at how any field has progressed. What directions did we take and why? As far as the future of RSM is concerned, directions that could be taken are abundant. In addition to biomedical areas, the number of problems that can be solved with RSM in traditional manufacturing have increased greatly due to developments in fields such as nonlinear optimization, Bayesian experimental design, nonparametric regression, estimation using GLM's, mixed model analysis, and many others. In what follows, we look back and discuss what areas have had a strong impact, positive or negative, on the progress in RSM in recent decades.

TAGUCHI'S PARAMETER DESIGN

I feel that sometimes we do not stop and reflect on the so-called "Taguchi era" and its effect on current directions in experimental design and in particular RSM. Much of the criticism of Taguchi's suggested methodology led to an impressive array of "alternative technology" produced by those whose motivation was to solve important problems using sound statistical practice. Few criticize Taguchi's engineering ideology concerning the need for reduction in process variance. But there is an endless list of works describing how the Taguchi ideas are flawed and what alternative approaches are desirable. For some detail and a written panel discussion on robust parameter design (RPD) alternatives to Taguchi, the reader should consult the paper edited by Nair (1992) in which many authors review some of the approaches.

The Taguchi era brought considerably more attention to experimental design, general quality improvement methods, and especially RSM. Please note that I am not implying that Taguchi himself brought attention to RSM. Rather, RSM was such an obvious improvement on the Taguchi approach for process optimization that this placed focus on RSM!

My own interest in RPD was induced by a colleague's statement "The goal of robust parameter design is to find optimum process conditions on a set of standard control (or design) variables." My immediate reaction was "That sounds a great deal like response surface methodology." Indeed, RSM became a viable and popular tool in dealing with parameter design problems. In fact, in the last ten years many research papers have been written that suggest the use of RSM in solving parameter design problems; for just a few examples, consider Lucas (1994); Vining and Myers (1990); Myers, Khuri, and Vining (1992); Engel and Huele (1996); Lin and Tu (1995); Myers, Kim, and Griffiths (1997); and Vining and Schaub (1996).

The Taguchi era has matured now. But variance reduction techniques are used extensively in the manufacturing segment of American industry. And RSM has become a richer collection of techniques because of it. An example of this is the inclination by quality practitioners to make use of techniques in variance modeling. At times, these models are a result of active interaction between control and noise variables or perhaps the characterization of replication variance produced in a designed experiment. It is encouraging to see these concepts taught and used in many companies in this country. This is a result of the realization that for too long textbooks, statistical curricula, and experimental design culture and folklore have placed all the emphasis on modeling the mean. This point is made notwithstanding the enormous amount of attention over the last several decades given to random effects and mixed effects modeling. By variance modeling here we mean the use of the process variance or experimental error variance as a response in a regression or response surface model, thereby allowing a "dual response surface approach" to process optimization. Recently, Box and Jones (1989), Vining and Myers (1990), Lucas (1989), Del Castillo and Montgomery (1993), Carroll and Ruppert (1988), Lin and Tu (1995), and Engel and Huele (1996) are but a few of the researchers who have pointed out the virtues of variance modeling. Interestingly, in nearly all of these cases the papers or conference talks made reference to the approach as an alternative to "Taguchi methods." It also should be pointed out that variance modeling, as we discuss it here, first surfaced in a paper by Bartlett and Kendall (1946) in which they discuss ln(s[sup2]) as a response in analysis of variance (ANOVA) situations. This fact seems to have been largely undiscovered. Engineers will continue to embrace the concept of variance modeling.

The focus on the RPD era has allowed many scientists and engineers to discover and pursue RSM for the first time, much like one would by reading Box and Wilson (1951) for the first time. In the section that follows, we discuss the role that optimal design theory has played in developments in RSM. In addition, we address the important task of identifying directions optimal design and computer generated design should take in the future.

COMPUTER GENERATED DESIGN--A VAST WASTELAND???

This is a provocative topic. It is important that all practitioners or potential users of computer generated design computer packages view both sides of the issue. For example, one should read Kiefer (1975); Box (1982); and Box, Hunter, and Hunter (1978). A historical review article that highlights many of these issues is given in Myers, Khuri, and Carter (1989). Much has been written about the utility and/or overuse of computer generated design in RSM. There is no intention here to give a complete bibliography. Different opinions reside, but two things are certain. The first is that in the past too many users have relied on design optimality when the use of standard designs, such as central composite designs (ccd's) or Box-Behnken designs would be preferable. This occurred because practitioners often used them primarily to produce designs with fewer runs than many of the standard designs. The second is that if, historically, computer software companies had embraced design robustness more vigorously than design optimality, then more flexible tools would be available for use in RSM problems. I remember very well a symposium on experimental design in the early 1980's at DuPont in Wilmington, Delaware. At a dinner on the eve of the conference, considerable conversation centered around the "expert system" concept for designing experiments. DuPont statisticians had their ideas, and other industrial and academic types talked about their own attempts at constructing expert systems in which the user would input certain information and out would come an "optimal" design with D-optimality being the focal point. Of course, the DET MAX algorithm developed by Mitchell (1974) was to be the basis of the methodology in which designs were constructed that produced a maximum value of $|X'X|$, where X is the so-called "model" matrix. At the time, I was quite naive about all this, and my planned effort at putting the history of DOE in context in the keynote address at the conference was to include very little on design optimality. I couldn't help but think that the ideas I was hearing sounded too "plastic" in a practical world. It seemed that in the short run, an expert system would be interesting (to say the least), but in the long run it may be damaging because what I was hearing did not allow the computer to solicit answers to many questions regarding the scientists' or engineers' most agonizing uncertainties. These uncertainties center around choice of model, realism regarding assumptions, and even handling multiple goals in the experiment.

The concern over the unbridled use of computer generated design for dealing with response surface problems is not a question of a waste of computing tools. Rather, the larger issue is merely this: the inertia that resulted from the preoccupation with D-optimality (and the resulting popularity with users) prevented further developments that would eventually find value today and in the future, developments that have design robustness as a center piece rather than alphabetic optimality. By alphabetic optimality we mean the family of optimality criteria that are characterized by a letter of the alphabet, for example, A, D, E, and G. Consult Atkinson (1982); Ash and Hedayat (1978); and Myers, Khuri, and Carter (1989) for reviews of these criteria. Here, I speak not merely about robustness to the foregoing uncertainties (which themselves are often reason for concern) but also about robustness ideas that speak to the directions being taken by RSM currently and in the future. There are so many RSM design problems that are couched in scenarios from which no optimal designs are appropriate, for example, cases of nonlinear models and/or nonnormal error distributions. Certainly, quality improvement problems with responses that are "proportion defective" or counts of number of defects come to mind, not to mention the vast array of applications in the environmental and biomedical fields in which survival models and GLM's are used. No foundation exists with the software companies that allows a transition to construction of good or robust designs (not necessarily optimal) for these kinds of problems. And indeed, this is the direction RSM should be moving. That is not to say that no research is being done in this area. See, for example, Sitter (1992), Abdelbasit and Plackett (1983), Myers et al. (1996), and many others. Many new and exciting RSM design problems exist, and the software companies will be able to accommodate new clientele if there is a stronger investment in robust designs or Bayesian designs rather than total investment in alphabetic optimality. In later sections of this paper, more attention will be put on design robustness and Bayesian design.

The foregoing is not meant to imply that computer generated design is a vast wasteland. Far from it! There are certainly many cases which feature constraints on the design region, mixture designs, strange blocking requirements, or a mixture of qualitative and quantitative factors, that is, cases where standard designs are difficult to come by. Indeed, design optimality has rendered itself of practical use. However, in many RSM situations, standard designs are more attractive because they appeal to the notions of design robustness discussed subsequently in this paper. While I have not carried out a scientific polling on this issue, it seems that standard RSM designs have recently enjoyed more and more use in preference to computer generated D-optimal designs when the scenario allows. Much of this comes from increased awareness on the part of practitioners. Indeed, many practitioners have been exposed to industrial short courses that emphasize the use of response surface designs.

ROBUST EXPERIMENTAL DESIGNS

Let me begin by saying that this is not robust "parameter design," but rather robust "experimental design." I feel that the preoccupation with making design optimality practical in the 1970's and 1980's vis à vis computer generated designs diverted proper attention from robust experimental design. In addition, I firmly believe that the future directions of RSM will be well served if design robustness, sequential design, and Bayesian design take their rightful place in the type of development work that is needed to produce practical RSM designs in relatively new areas.

Box and Draper (1975) wrote a paper containing philosophical as well as substantive motivation for robust designs. They list several properties that designs might require, properties that are related to detection of lack of fit, estimation of error variance, insensitivity to outliers and missing data points, insensitivity to errors in control, and many others. In any given response surface problem it is likely that at least three or four of these properties are needed, and in many cases alphabetic optimality is thus not appropriate. Box and Draper point out that what is often required are good designs, not optimal designs. It seems that for a long period of time the "warning" rendered by Box and Draper had fallen on deaf ears. But if one accounts for the renewed interest in RSM in industry as well as its discovery in environmental, biological, and biopharmaceutical areas, the problems encountered will more and more be couched in the kind of assumptions that render optimal designs not only difficult to find, but impractical to use. For example, consider a dose-response situation in which the response is binary, as in the "proportion that survive" in a biomedical example or the proportion defective in a manufacturing scenario. The design variables enter the model as drug combinations for the biomedical problem or standard processing variables in the latter case. As a result, we may have a logistic regression model

p[subi] = [1 + exp [-f(x[subi])'beta]][sup-1] (i = 1, 2, ..., n), (1)

where p[subi] is the probability of survival (or defective) and f(x[subi])'beta contains standard response surface terms. Apart from single variable cases, optimal design theory is not used here. Perhaps standard response surface designs, for example, factorial designs, ccd's, or Box-Behnken designs work well here. How would one know? The use of any alphabetic optimality here requires prior knowledge of model parameters due to the nonlinearity of the model. This suggests different approaches to this and similar problems, such as Bayesian and sequential designs. Obviously, one very important approach here is the search for robust designs, that is, those designs that are efficient despite the lack of knowledge of parameters or even in the face of model uncertainty. It is interesting to note that Box and Lucas (1959) discussed designs for nonlinear models and the idea of a sequential approach.

BAYESIAN AND TWO-STAGE DESIGN

It would seem that much of what has been said in the foregoing would serve to motivate the notion of both Bayesian and two-stage design. The Bayesian design approach may be indicated to incorporate any available prior information on the parameter in the choice of a design. The multiple stage approach to design is certainly obvious. When information about parameters is necessary in order to design an efficient experiment, the use of a preliminary experiment for the purpose of obtaining parameter guesses followed by an appropriate (perhaps optimal) augmentation is more than reasonable. It is important to emphasize that both of these approaches require intensive use of the computer. It may seem unorthodox to think of Bayesian methods together with notions of RSM. Clearly the notion of Bayesian design surfaces in a setting in which there are uncertainties. In classical RSM the uncertainty resides with the choice of model. This problem is particularly acute in the case of mixture designs where it is not uncommon that constraints on the mixture components produces a dilemma for the researcher. Candidates are first order, second order, special cubic, and cubic models. On the other hand, as we indicated previously, in biomedical and other biological and industrial areas where nonlinear models are used, uncertainties are of a more extensive nature. Not only is there a model uncertainty, but optimal or even "good" designs are found through the use of guesses on model parameters. It should be emphasized that practitioners in these fields sorely need RSM tools for problem solving.

While little work appears in RSM literature or classic RSM applications, there is a rich literature to borrow from in Bayesian design. A very nice overview appears in Chaloner and Verdinelli (1995) that explains the philosophy of Bayes design and cites many applications. Most of the application roots of Bayesian designs are in biological type problems in which a Bayes D-optimal design is sought. For example, if the logistic regression (or logistic RSM) model in Equation (1) is used with parameter vector, beta, and criterion function, R({Begin Greek}d, b), {End Greek}then a Bayes design is one in which one chooses the design, delta, that maximizes the function,

[Graphic Character Omitted] R({Begin Greek}d, b)p(b){End Greek}dbeta.

For example, a Bayes D-optimal design would involve the use of ln|I|, where I is the information matrix generated by the logistic model. In order that the reader have a clearer idea of why knowledge of parameters or guesses/priors are necessary, let us consider a special case of Equation (1), namely, a single variable logistic regression model:

p[subi] = 1 / 1 + exp[-(beta[sub0] + beta[sub1]x[subi])], (i = 1, 2, ... n). (2)

Here, p[subi] is the probability of a success (survival or nondefective item), and x[subi] is the dose of a drug or perhaps a processing variable. The information matrix on which any design optimality criterion is based is given by

[Formula Omitted]

where x[subi]'s are design levels and p[subi] is given by the model in Equation (2). Clearly then, the information matrix is very much a function of beta[sub0] and beta[sub1] above.

The thrust of Bayesian design is to produce an experimental plan that is robust to parameter guesses. As an illustration, let us consider the single variable model in Equation (2). The D-optimal design requires knowledge of parameters through the choice of design levels that target a certain effective dose (ED). ED[sub20], for example, is the dose that produces 0.2 probability of success. It turns out that the D-optimal design is a saturated design (which is often the case) with two levels: namely, N/2 experimental units are dosed with the ED[sub17.6], while the additional N/2 units are dosed with ED[sub82.4]. Clearly, two levels will not be sufficient for most researchers. In addition, the design is known to be nonrobust to the mistakes in the researchers' guess of the location of these two ED's. Consider now a Bayesian design. It is best seen with a reparameterization of the model in Equation (2) as

pi = 1 / (1 + exp[-beta[sub1](x - μ)]). (3)

Here, μ is the very important ED[sub50]. So the problem requires prior knowledge of the slope parameter and the ED[sub50]. Suppose it is felt that beta[sub1] [element of] [0.1, 0.3], while the ED[sub50] [element of] [210, 280], where the units are, say, in grams. Also suppose a normal prior is used with interpretation that the boundaries are +/-2sigma. A Bayesian D-optimal design is given by

```
 ED[sub5] = 230.13 gms ED[sub27] = 240.04 gms
ED[sub73] = 249.96 gms ED[sub95] = 259.87 gms.
```

Here, the actual dosage values are based on the center of the boundaries or the mean of the prior. Clearly, the design is more robust than the two-level design. Indeed, it may be very practical for the user. It turns out, as one would expect, that a uniform prior on the parameters results in even more levels. Further details regarding these designs can be found in Chaloner and Larntz (1989) and Jia and Myers (1998). More attention will be devoted to Bayesian designs in similar models and in settings that are closer to a true RSM situation later in this paper.

Multistage response surface experiments are certainly not new. They play a large part in the rich heritage of RSM. George Box said time and time again that DOE, in general, and RSM, specifically, are iterative processes. Indeed, the Box-Wilson work reflects this philosophy through the idea of using 2-level factorial designs as a first stage along with center run augmentation. This is followed with a lack of fit test and a second stage that involves an additional augmentation with axial runs to produce a ccd. Certainly anyone reading this paper who has designed experiments knows that, almost without exception, a retrospective view results in changes in the design. Indeed, the whole idea of using computer generated design in the light of model uncertainty suggests the need for augmentation in light of a rethinking of the model. Again, the need for multistage designs becomes even more important as we note current and future directions in RSM. Nonlinear modeling and modeling with nonnormal errors produces a need for reasonable guesses of parameters if one hopes for a design which is efficient. Certainly, designing experiments in two stages is a natural avenue for future research and practice.

As an illustration of multistage design, consider once again the logistic regression model in a single variable as described by Equation (3). The first stage involves an attempt at placing dosage levels at the D-optimal design, using guesses at ED[sub17] and ED[sub83]. For this example, 1/3 of the experimental runs are placed in the first stage with N/6 placed at each of those two "guessed dosages." Then the parameter estimates found from the data in the first stage are used as a basis for the optimum augmentation. Suppose it is estimated from the data that the first stage guesses occurred at ED[sub7] and ED[sub50]. The optimum augmentation is designed to be "conditionally optimum" using any criterion of interest (such as D- or F-optimality). The F-optimality criterion is used to minimize the squared width of a Feiller interval around a certain "estimated ED" (see Sitter (1992)). This may be important in either engineering or biomedical applications. Illustrations of these two-stage designs are shown in Myers et al. (1996). In our example, suppose it is determined that our criterion of interest results in the balance of the experimental units being placed at ED[sub94]. We use the 1/3-2/3 ratio since it was recommended in Myers et al. (1996). Figure 1 shows a display of the dose response curve in which the ED[sub94] point "balances" the experiment. Two important points should be made. The major virtue of the two-stage design is robustness to the initial guesses; the chance to "patch up" the design in a sense helps to correct the mistakes made in the first stage. Also, one must understand that even though an alphabetic optimality criterion is used at each stage, the result is not an optimal design in any sense, but rather a robust design. An optimal design can only be claimed in a case where the guesses of the location of optimal design levels are made with no error. Studies suggest that the multistage robust approach is more efficient than the "attempted" optimal design even when minor errors are made in the guessed parameter.

The Bayesian and multistage designs are discussed here in the context of a single design variable, which is certainly not in the spirit of classical RSM. However, applications are abundant in more than one design variable in both biological and industrial applications. Certainly, any scenario that involves binary responses or count responses is a candidate for the use of logistic or Poisson regression, respectively, and nonlinear models are suggested, requiring preliminary information on parameters for efficient choice of design. Myers and Montgomery (1995, p. 129) discuss an experiment in a semiconductor fabrication plant in which 5 factors were used and the response might involve the proportion of good chips in a wafer. Poisson responses are abundant in the textile industry when the number of defects in a bolt of cloth are modeled as a function of process variables. While not much technology is available, it is a natural area for future research, particularly since RSM and GLM's are two disciplines that are being combined by practitioners. Much more attention is given to GLM's in a later section. In addition, designs for specific GLM's will also be discussed.

Let us return to the use of Bayesian principles in RSM. Let us also turn attention back to more traditional RSM in which empirical polynomial models are used. Of interest are RSM designs that are robust in some sense to model uncertainty. Some very interesting work based on a simple idea by Dumouchel and Jones (1994) illustrates a practical use of Bayesian methods for selection in which model uncertainty does not allow the use of alphabetic optimality. They assume that in practical usage there are two kinds of model terms, certain terms and questionable terms. They conceive a prior on model coefficients that takes this assumption into account and then produce a practical way of locating a Bayesian D-optimal design that maximizes the determinant of the posterior information matrix. Research into application of this idea is readily accessible with the use of candidate designs. Andere-Renden, Montgomery, and Roller (1997) make use of the idea in developing mixture designs. Here, the need for robustness in the model is particularly acute since there are often many alternatives involved in the model choice. In addition, Neff and Myers (1998) demonstrate the advantage of the method and illustrate a two-stage approach in which the prior information is updated at the completion of the first stage.

An additional useful application of Bayesian principles in classical RSM was introduced by Vining and Schaub (1996). They use a Bayesian approach to evaluate ccd's in which replication is induced in the design at well chosen locations, candidates for replication being axial points and a fraction of the factorial component of the design. The purpose is to find appropriate designs for dual modeling of the mean and variance. Many companies in the U.S. have begun a standard practice of this type of dual modeling. The result is that extra information about variability is available to better achieve process improvement. Dual modeling is illustrated in examples in Myers and Montgomery (1995). In one illustration the number of solder defects per million joints is modeled against processing variables involved in a circuit board assembly operation. Mean and variance models are fit and compromise conditions are determined that allow a small number of defects with reasonable consistency. Bayesian principles and RSM designs represent fruitful partners for research and application in the future. The kinds of applications that exist certainly make it a promising partnership.

GLM AND OTHER NONNORMAL TYPES OF MODELING

In much of what has preceded we have alluded to the increased use of nonnormal modeling in manufacturing applications that involve, say, count data, binary responses, or time to failure type data. For many years statisticians in biomedical applications have been involved in survival, toxicity, binomial, and count models and have recently become interested in using response surface methods. As a result, there is a natural inclination to expect that nonnormal models and GLM's will become a major application area for response surface analysis and design (see McCullagh and Nelder (1989)). The design variables are processing variables in the manufacturing area and dosage type variables in the biomedical area. Another subject matter field where RSM should and hopefully will make stronger inroads is environmetrics, where pollution problems often involve either binary or count data and where logistic or Poisson models are used. See, for example, Oris and Bailer (1993). Here, the response represents counts (organisms) that survive when they are treated with certain types of pollutants. Poisson regression with a log link results in

E(y | x) = lambda(x) = exp[f(x)'beta],

where f(x) contains response surface terms involving dosage of one or more pollutants. The goals of the experiment are to fit the response function using maximum likelihood procedures, to predict the proportion of impairment of reproduction at selected levels, and to estimate points or contours of effective concentration (EC), that is, concentration that results in a certain EC, say, 80% impairment. This is just one type of standard toxicity study in which response surface principles apply, but standard normal theory polynomial regression is not used by the practitioner. Carter, Wampler, and Stablein (1983) describe toxicity studies in which logistic models were used to describe dosage regions and optimum conditions were found. In addition, Poisson and Gamma models are used to build response surfaces in practical illustrations in Myers and Montgomery (1997) and Bisgaard and Fuller (1994). While existing applications are not yet abundant, this remains one of the most important areas for the future if GLM's are to meet their potential in reaching practitioners.

The need for applications of RSM in these areas is clear, and the curiosity exhibited by practitioners is encouraging. Nevertheless, there remains the issue of experimental design. Analysis of data can be carried out using procedures in SAS, S-PLUS, and ECHIP. However, no software for design optimality exists, and indeed, even if it did, it is not clear that it would be useful since it often requires parameter knowledge or guesses. It is a certainty that the software for developing second stages of two-stage designs may be useful. Also, some creative robustness ideas are needed in order to deal with the design problems here. Optimal designs have been found in many situations (See Sitter and Torsney (1995), Jia and Myers (1998), and Atkinson and Haines (1996)). However, a more robust approach may include Bayesian ideas or even the involvement of standard designs in some fashion. The concept involved in finding good designs makes use of new and interesting ideas. Some of this will be reviewed in a later section.

Some may say that the use of GLM's in a formal fashion in response surface applications may be avoided by making use of transformations to stabilize variance. A good discussion and illustration is found in Bisgaard and Fuller (1994). In this case, of course, standard designs would be quite reasonable. However, Hamada and Nelder (1997) outline a strong case for the use of the GLM model in formal fashion without transformation. Their arguments center around the notion that transformations may indeed solve a single problem, such as variance heterogeneity, but that they cannot handle several problems simultaneously. They also present a convincing argument that natural nonhomogeneity brought about by, say, a binary or Poisson response should be attacked through knowledge of that distribution and the accompanying likelihood structure. Myers and Montgomery (1997) illustrate with examples that the GLM approach appears to outperform transformed models.

SEMIPARAMETRIC AND NONPARAMETRIC RSM'S

Semiparametric response surface procedures seem to have their roots in computer experiments, though many applications exist in other areas, particularly biopharmaceutical and biotechnology areas. This is a very important area that seems to fill a void that has existed with standard RSM procedures for years. As a result, it should be a subject of increased research and consideration by practitioners. For years, some practitioners have wondered about the use of RSM when the function is not accommodated by polynomial models in the operability region. Indeed, perhaps there are multiple models that function inside the region of operability. Among the scenarios that suggest the use of nonparametric RSM (NPRSM) are when

(i) optimum conditions are required;

(ii) there is less interest in an interpretive function and more in the appearance of the response surface;

(iii) the function is highly nonlinear, but not necessarily well behaved; or

(iv) there is not a need for the designs constructed to honor "model form," but rather for designs to be from a space-filling grid.

In the case of scenario (iv), the advantage of a space-filling grid is obvious since these nonparametric methods, which assume no graduating polynomial model, produce predicted values at a specific location that are very much dependent on the data points in close proximity. It is natural then for the subject matter problem not to allow region-seeking gradient procedures like steepest ascent for iterative movement. Rather, the design variable ranges are quite wide--wider than standard RSM tools allow. A global view of the process takes precedence rather than a local impression, which is sought in more classical response surface analysis and design. Three major techniques are considered here, namely, thin plate splines, Gaussian stochastic processes, and neural networks.

The abundant NPRSM applications include optimization of reaction buffers for amplification of DNA evidence, injection molding experiments, and many others (see Haaland et al. (1994)). While use of NPRSM is on the increase and rightly so, there are still some intriguing, unanswered questions which will probably be solved through many experiences by practitioners. For example: Since all successful RSM is iterative, how does one combine experiments using these methods? Can sequential experimentation be made formal with these methods? How is the search for optimum conditions directed? How do we deal with multiple responses? Are space-filling designs the most effective to use if one is only interested in optimum conditions? Incidentally, perhaps one of the applications of NPRSM with the greatest potential is in the area of variance modeling, where standard empirical models are difficult to fit and ill behavior is expected (see, e.g., Vining and Bohn (1998)).

MULTIPLE RESPONSES

Software companies dealing in RSM have made great strides in recent years in applying multiple response optimization. The advancements in RSM have benefited from no other field more than that of mathematical programming and optimization theory. As a result, constrained optimization involving multiple responses has served users extremely well in recent years. As a trivial illustration, if a problem contains three responses, y[sub1], y[sub2], and y[sub3], and interest centers around the maximization of ^y[sub1] (the primary response) subject to ^y[sub2] > A and B < ^y[sub3] < C, conditions are easily found using software packages. The Derringer-Suich method is also simple to use (see Box and Draper (1987), Myers and Montgomery (1995), and Derringer and Suich (1980)). It allows for the use of a geometric mean of desirabilities that are defined according to power functions that are based on the impact priorities of the experiments regarding which specifications on the responses are the "tightest." The final univariate response that is maximized is

[Formula Omitted]

which is an "overall" desirability. This tool is very appealing to practitioners, and it is my experience that it enjoys considerable use.

There is one caution light flashing here, and it applies not merely with the Derringer-Suich procedure but other multiple response optimization procedures as well. It is tempting to treat the optima or constrained optima as if they were based on deterministic functions. One must always remember that optima are stochastic in nature and use of any of these procedures should be followed by rather extensive confirmatory experiments. We must remember that the response surface models involve ^y values that have, at times, considerable variance. As a result, the "location" of optimal x-values have variance that is unknown. We learned even in the single response case from Box and Hunter (1954) that confidence statements on the points of optimum conditions may be very pessimistic. I fear that using confirmatory experiments may not be standard practice, and this could lead to disenchantment by practitioners and/or supervisors. In addition, not enough attention is given by practitioners to the consideration of correlation structures among responses (see, e.g., Khuri and Conlon (1981)).

RESTRICTIONS IN RANDOMIZATION

Practitioners using RSM have been making a certain type of mistake for decades. There has been a "sweeping under the rug" process that prevails. Here is the setting. Among the design variables is one or more that is either very difficult or very costly to control. In much of what we do in analysis and estimation of model coefficients, it is assumed that the experiment was designed in a completely randomized fashion. In the scenario described here, it is very unlikely that the practitioner will allow or do complete randomization. In almost all problems involving a temperature variable, the engineer will not allow temperature to be bounced around according to a "randomization prescription." Often we have assumed (and hoped) that it didn't make any difference in the analysis.

The attention put on this situation in the current decade can again be attributed to tangential fallout from the Taguchi parameter design era. Noise factors are often difficult to control in a process setting. It became apparent that in some practical situations noise factors are also difficult to control in an experimental situation. As a result, focus was placed on the problem; Box and Jones (1992) pointed out that the analysis should be that of a split plot design. Now, everyone remembers the split plot experiment from the graduate design course. But the response surface literature is hardly filled with information as to how this affects estimation of model coefficients. It is interesting that Cornell (1988), in a discussion of a mixture experiment involving fish patties, recognized the need to take the split plot nature of the design into account in estimating model coefficients. Lucas and Ju (1992) shed light on this problem and how it influences choice of 2-level designs. Letsinger, Myers, and Lentner (1996) dealt with the problem in the context of second order models where standard ccd's or Box-Behnken designs are used. Clearly, the factors are not crossed; thus, the classic split plot structure does not apply, and alterations in analysis should be considered.

Consider a 3-factor ccd with factors z, x[sub1], and x[sub2], where z is a hard to control factor. Thus, we assign it to be a whole plot factor. The design layout (assuming 3 center runs and axial level alpha) is given in Table 1. Note that the use of this ccd involves an unequal whole plot size. In addition, whereas a standard split plot experiment involves replication of the entire design, the standard ccd in a split plot structure contains only the standard repeat runs at the center. This does not mean that replication of the entire design should not be done. Indeed it should. But many practitioners like to "have their cake and eat it too," namely, to be able to account for the randomization structure and still be able to use a standard economic design.

Given the expected randomization structure for the ccd listed above, a reasonable model is the mixed model

```
y[subij] = f(x[subi])'beta + delta[subi] + [element of][subij], for i = 1,2,..., 5
and j = 1,..., n[subi],
```

where n[subi] is the number of subplots in the ith whole plot and delta[subi] and [element of][subij] represent the whole plot and subplot error, respectively. The term f(x[subi])'beta represents the response surface portion of the model. This model then produces a variance covariance matrix for the vector y = y[subij] as

[Formula Omitted]

Where each off-diagonal element in V[subi] is sigma[sup2[sub[subdelta] (the whole plot error variance component), where each diagonal element is sigma[sup2[sub[sub[element of]] + sigma[sup2[sub[subdelta], and where sigma[sup2[sub[sub[element of]] is the subplot error variance component. As a result, the structure of the analysis naturally involves estimation by

^beta = (X'V[sup-1]X)[sup-1] X'V[sup-1]y, (4)

and SAS PROC MIXED is certainly a candidate for supplying the coefficients and standard error. The pertinent estimation procedure for sigma[sup2[sub[sub[element of]] and sigma[sup2[sub[subdelta] is the method of restricted maximum likelihood (REML).

So why discuss this problem in this paper where reflections about the future are and should be pertinent? First, because practitioners need to become aware of the problem. Also, depending on the design and the nature of replication, there are times in which it is better to assume independence and use ordinary least squares. (Indeed, for some designs/models, generalized least squares in Equation (4) is equivalent to ordinary least squares.) In other words, when estimation of sigma[sup2[sub[subdelta] is done with relatively little information it may be best to ignore the problem, as many have done for decades. This needs to be sorted out, and in addition, more needs to be learned regarding the role of design and the need for replication of the complete experiment. As a result, it seems as if this topic should require attention at several levels.

One very intriguing personal observation on this topic springs to mind and is a good example of how research workers in different fields of statistics do not keep abreast of what others are doing. If they did, they would find that problems they encounter are very similar and that ideas on solutions should be shared. The problem encountered here is very much like that discussed in biostatistics in the analysis of longitudinal data. Generalized estimating equations or the mixed model approach is often used. The whole plot is the subject, and repeated observations are taken on the subject across time--thus the assumption of a correlation structure. Liang and Zeger (1986), Zeger and Liang (1988), and many others have carved out a sizeable literature. And it seems to me as if the response surface people can learn from it. Obviously there are differences, the primary one being the design problem and the need for fitting polynomial models in RSM. The greatest common ground here exists when the response surface split plot type problem involves nonnormal errors, and the use of GLM's becomes necessary. It may be advisable for some readers to see a book by Diggle, Liang, and Zeger (1994). It would seem to me that a good review article for practitioners should be written about the topic in this section.

MORE ON DESIGNS FOR NONNORMAL MODELS

We indicated in previous sections that much interest centers around experimental designs when the user is interested in building a model with nonnormal responses, such as logistic regression, Poisson regression, and a plethora of other models for which any notion of design optimality or design quality necessitates that the user supply guesses on the parameters. The use of binary, count, time to event, rate, and other obviously nonnormal responses occur frequently in practice and certainly suggest the use of GLM's. As a result, these parameter guesses are necessary in order that efficient designs can be found. Any study of design efficiency or design robustness or multistage design certainly must be preceded by a study on design optimality, even though it is unlikely that anyone will ever use a "truly optimal" design in such a scenario. Also, the study of what makes a design efficient or near optimal is fascinating when one compares the design to that which might be used in a standard linear model with homogeneous variance. The relatively simplistic notions of optimal response surface design in standard conditions makes use of the homogeneous variance assumption, whereas in nonnormal response situations, such as the family of GLM's, the variance is a function of the mean. Many questions arise: Are standard RSM designs efficient or robust? Can one alter the replication designed into the experiment in such a way as to create an efficient design? Can a two-stage design be a practical design in these cases? One can gain a picture from all this that strict optimal design must give way to robustness and that robustness has claimed a new element--robustness to parameter guesses. This is added to the all important robustness to choice of model (link function for GLM) in these cases.

As one moves from one to two design variables for, say, a logistic regression problem, the complexity changes significantly. In addition, most scholars and practitioners would feel that the study is not truly RSM unless there are at least two design variables. So, let us provide some insight into the extension of the optimal design for logistic regression to two design variables. Consider two drug doses or two process variables, x[sub1] and x[sub2], with a binary response and thus the model

E(y) = p = 1 / (1 + exp[-[beta[sub0] + beta[sub1]x[sub1] + beta[sub2]x[sub2]]]).

In the case of a single regressor, the D-optimal design requires the placement of half of the experimental units on the ED[sub17.6] and the other half on the ED[sub82.4]. Here, however, the points must be placed on ED contours. Indeed, quite analogous to a factorial design, the D-optimal design here is to place four points according to Figure 2. Any parallelogram with two points on the ED[sub22.7] and two points on the ED[sub77.3] is D-optimal. It is interesting that the D criterion is invariant to the location and angle in the parallelogram. However, if the figure gets larger because of wider ranges, the D criterion becomes larger. See Figures 3 and 4. Even here we have an analog to factorial experiments in which ranges increase and hence variances of parameter estimates grow smaller. The key theoretical and interesting element here is that points must be at certain ED levels in order to accommodate the variance covariance structure. Bayesian designs result in more ED levels to choose and hence are more robust, at least in principal. See Jia and Myers (1998). Obviously multiple stage designs are likely necessary here.

Poisson regression applications also give rise to interesting design problems. Consider the model with Poisson responses and log link in an experiment in which doses x[sub1] and x[sub2] of two pollutants impact the reproduction of an organism. Or we may be dealing with two drugs that impair growth of cancer cell colonies. Or we may refer to applications where the response is number of defects in an industrial application and where the design variables represent components in a mixture. No matter the application, the model for an experiment is given by

lambda[subi] = exp[beta[sub0] + beta[sub1]x[sub1i] + beta12]x[sub2i]] (i = 1, 2, ..., n)

or, for the case where interaction is to be studied,

lambda[subi] = exp[beta[sub0] + beta[sub1]x[sub1i] + beta[sub2]x[sub2i] + beta[sub12]x[sub1i]x[sub2i]] (i = 1, 2, ... n).

Of course the variance of y[subi] is equal to lambda[subi] also. Here, lambda[subi] is the expected counts at the ith data point. At x[sub1i] = x[sub2i] = 0, namely, the "control" where there is no pollutant (or drug), the expected value is its largest, and growth is impaired when pollutant is used. Hence, the expected number of organisms or cancer colonies grows smaller. Minkin (1993) and Oris and Bailer (1993) analyzed data and studied the design problem. It turns out that the optimal designs in x[sub1] and x[sub2] are determined by appropriate selection of a specific EC = lambda / (lambdac), where lambdac is the mean number of events at the control, that is, when x[sub1] = x[sub2] = 0. Here, EC[sub50] is the dosage combination at which the expected number of counts is half that of the control. The single toxin optimal designs, both traditional and Bayesian, have been well studied. As an example, a D-optimal two-level design with a single toxin requires N/2 experimental units at the control and N/2 at the EC[sub13.5]. Again, note the need for "guessing" parameters in order to execute the optimal design. Bayesian three and four level designs have also been studied.

As in the case of logistic regression discussed previously, for more than a single toxin, the optimal design involves the selection of points on a "guessed" EC contour in some configuration. For example, for the no interaction case, the design involves the assignment of N/3 runs at each of the three points 0, 1, and 2 depicted in Figure 5. Here, the design points include the control, (x[sub1] = 0, x[sub2] = EC[sub13.5]), and (x[sub1] = EC[sub13.5], x[sub2] = 0), where the EC[sub13.5] represents that of a single component model. It turns out that if it is not practical to use the single component points 1 and 2 and all points (apart from the control) must contain some of both x[sub1] and x[sub2] according to the restricted region defined by the dotted outer two rays in Figure 5, then the D-optimal design represents the points on the extremes of the region with the EC[sub13.5] being the boundary. As a result, the appropriate design contains the points 0, 3, and 4.

In the case of the interaction model, the D-optimal design takes on a factorial configuration. For example, consider Figure 6. Here, points 0, 1, 2, and 3 have a factorial structure. Again, a similar factorial can be constructed in which 1 and 3 are not on the x[sub1] and x[sub2] axis, respectively, and in which point 3 completes the factorial. This restricted region design is also D-optimal. This can be extended to k design variables, and indeed fractional factorial designs can be constructed depending on the number of and which interactions are of interest. See Huffman and Myers (1998) for details.

Despite the amount of research that is dedicated to design selection for the models in the GLM family, there remains the work of rendering the designs useful from a pragmatic point of view. The need for guessing parameters and approximate locations of EC contours results in frustration, and yet it serves as a springboard for interesting and exciting response surface research and applications of the future.

CONCLUSION

Much excitement surrounds the present and future of RSM. Just as George Box and his coworkers undoubtedly felt excitement in the solution of industrial chemical problems at the inception of RSM in the late 1940's, more tools for even more complex problems involving varied types of practitioners will be discovered in the future. In this paper, we emphasized the way RSM applications are changing in both classic industrial and biological or biomedical applications. It should be strongly emphasized that RSM applications as well as other applications do not differ significantly across areas of applications, although the "bottom line" may be different. Survival models in biomedical applications may deal with the welfare of patients that are being treated for a disease, whereas a similar analysis in industry may deal with the survival of an electronic component in a larger system. However, as we have implied throughout this paper, statistical researchers or practitioners in one area can learn a great deal from statistical practices in another area. Industrial practitioners can learn a great deal from RSM successes in biomedical applications. It is important to learn from diverse applications rather than only focusing on our own. In fact, it is surprising that so little borrowing of technology occurs. In future applications in RSM, learning from the other side will be beneficial. Thus, the emphasis on biological applications in this paper.

ADDED MATERIAL

RAYMOND H. MYERS Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Dr. Myers is Professor Emeritus in the Department of Statistics. He is a Member of ASQ. His email address is rmyers@vt.edu.

This paper was presented at the Journal of Quality Technology Session at the 42nd Annual Fall Technical Conference of the Chemical and Process Industries Division and Statistics Division of the American Society for Quality and the Section on Physical & Engineering Sciences of the American Statistical Association in Corning, New York, October 22-23, 1998.

TABLE 1. Design Layout for a ccd With 3 Center Runs, Axial Level alpha, and One Hard to Change Factor, z

Z -- Whole Plot									
-1		0		1		-{Begin Greek}a		a{End Greek}	
x[sub1]	x[sub2]	x[sub1]	x[sub2]	x[sub1]	x[sub2]	[sub2]	x[sub2]	x[sub1]	x[sub2]
-1	-1	0	alpha	-1	-1	0	0	0	0
-1	1	0	-alpha	-1	1				
1	-1	alpha	0	1	-1				
1	1	-alpha	0	1	1				
		0	0						
		0	0						
		0	0						

FIGURE 1. Two Stage Illustration for Logistic Regression.

FIGURE 2. Parallel Line Design.

FIGURE 3. Logistic Regression--Invariance To Location.

FIGURE 4. Logistic Regression--Design Is Improved With Larger Range.

FIGURE 5. D-optimal Design for Poisson Reproduction Impairment (No Interaction).

FIGURE 6. D-optimal Design for Poisson Reproduction Impairment (Interaction Case).

REFERENCES

ABDELBASIT, K. M. and PLACKETT, R. L. (1983). "Experimental Design for Binary Data". Journal of the American Statistical Association 79, pp. 90-98.

ANDERE-RENDEN, J.; MONTGOMERY, D. C.; and ROLLER, D. A. (1997). "Design of Mixture Experiments Using Bayesian D-optimality". Journal of Quality Technology 29, pp. 451-464.

ASH, A. and HEDAYAT, A. (1978). "An Introduction to Design Optimality With an Overview of the Literature". Communications in Statistics--Theory and Methods 7, pp. 1295-1325.

ATKINSON, A. C. (1982). "Developments in the Design of Experiments". International Statistical Review 50, pp. 161-177.

ATKINSON, A. C. and HAINES, L. M. (1996). "Designs for Nonlinear and Generalized Linear Models" in Handbook of Statistics edited by S. Gosh and C. R. Rao. Elsevier, Amsterdam. pp. 437-475.

BARTLETT, M. S. and KENDALL, D. G. (1946). "The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation". Journal of the Royal Statistical Society B 8, pp 128-150.

BISGAARD, S. and FULLER, H. T. (1994). "Analysis of Factorial Experiments With Defects or Defectives as the Response". Quality Engineering 7, pp. 429-443.

Box, G. E. P. (1952). "Multifactor Designs of First Order". Biometrika 39, pp. 49-57.

BOX, G. E. P. (1954). "The Exploration and Exploitation of Response Surface: Some General Considerations and Examples". Biometrics 10, pp. 16-60.

BOX, G. E. P. (1982). "Choice of Response Surface Design and Alphabetic Optimality". Utilitas Mathematica 21, pp. 11-55.

BOX, G. E. P. and DRAPER, N. R. (1959). "A Basis for the Selection of a Response Surface Design". Journal of the American Statistical Association 54, pp. 622-654.

BOX, G. E. P. and DRAPER, N. R. (1963). "The Choice of a Second Order Rotatable Design". Biometrika 50, pp. 335-352.

BOX, G. E. P. and DRAPER, N. R. (1975). "Robust Designs". Biometrika 67, pp 347-352.

BOX, G. E. P. and DRAPER, N. R. (1987). Empirical Model-Building and Response Surfaces. John Wiley & Sons, New York, NY.

BOX, G. E. P. and HUNTER, J. S. (1954). "A Confidence Region for the Solution of a Set of Simultaneous Equations With an Application to Experimental Design". Biometrika 41, pp. 190-199.

BOX, G. E. P. and HUNTER, J. S. (1957). "Multifactor Experimental Designs for Exploring Response Surfaces". The Annals of Mathematical Statistics 28, pp. 195-241.

BOX, G. E. P.; HUNTER, W. G.; and HUNTER, J. S. (1978). Statistics for Experimenters. John Wiley & Sons, New York, NY.

BOX, G. E. P. and JONES, S. (1989). "Designing Products that are Robust to the Environment". Presented at the American Statistical Association Conference, Washington, DC.

BOX, G. E. P. and JONES, S. (1992). "Split-Plot Designs for Robust Product Experimentation". Journal of Applied Statistics 19, pp. 3-26.

BOX, G. E. P. and LUCAS, H. L. (1959). "Design of Experiments in Non-linear Situations". Biometrika 46, pp. 77-90.

BOX, G. E. P. and WILSON, K. B. (1951). "On the Experimental Attainment of Optimum Conditions". Journal of the Royal Statistical Society B 13, pp. 1-45.

CARROLL, R. J. and RUPPERT, D. (1988). Transformation and Weighting in Regression. Chapman & Hall, New York, NY.

CARTER, W. H.; WAMPLER, G. L.; and STABLEIN, D. M. (1983). Regression Analysis of Survival Data in Cancer Chemotherapy. Marcel Dekker, New York, NY.

CHALONER, K. and LARNTZ, K. (1989). "Optimal Bayesian Design Applied to Logistic Regression Experiments". Journal of

Statistical Planning and Inference 21, pp. 191-208.

CHALONER, K. and VERDINELLI, I. (1995). "Bayesian Experimental Design: A Review". Statistical Science 10, pp. 273-304.

CORNELL, J. A. (1988). "Analyzing Data from Mixture Experiments Containing Process Variables: A Split-Plot Approach". Journal of Quality Technology 20, pp. 2-23.

DEL CASTILLO, E. and MONTGOMERY, D. C. (1993). "A Nonlinear Programming Solution to the Dual Response Problem". Journal of Quality Technology 25, pp. 199-204.

DERRINGER, G. and SUICH, R. (1980). "Simultaneous Optimization of Several Response Variables". Journal of Quality Technology 12, pp. 214-219.

DIGGLE, P. J.; LIANG, K. Y.; and ZEGER, S. L. (1994). Analysis of Longitudinal Data. Clarendon Press, Oxford, UK.

DUMOUCHEL, W. and JONES, B. (1994). "A Simple Bayesian Modification of D-optimal Designs to Reduce Dependence on an Assumed Model". Technometrics 36, pp. 37-47.

ENGEL, J. and HUELE, A. F. (1996). "A Generalized Linear Modeling Approach to Robust Design". Technometrics 38, pp. 365-373.

HAALAND, P. D.; MCMILLAN, N.; NYCHKA, D.; and WELCH, W. (1994). "Analysis of Space-filling Designs". Computing Science and Statistics 26, pp. 111-120.

HAMADA, M. and NELDER, J. A. (1997). "Generalized Linear Models for Quality Improvement Experiments". Journal of Quality Technology 29, pp. 292-305.

HILL, W. J. and HUNTER, W. G. (1966). "A Review of Response Surface Methodology: A Literature Review". Technometrics 8, pp. 571-590.

HUFFMAN, J. W. and MYERS, R. H. (1998). "Experimental Designs for Impaired Reproduction and Standard Industrial Applications Using Poisson Regression Models". Proceedings of the Meetings of American Statistical Association, August 1998 (to appear).

JIA, Y. and MYERS, R. H. (1998). "Bayesian Experimental Designs For Logistic Regression Models". Technical Report No. 98-22, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA.

KHURI, A. I. and CONLON, M. (1981). "Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions". Technometrics 23, pp. 363-375.

KIEFER, J. (1975). "Optimal Design: Variation in Structure and Performance Under Change of Criterion". Biometrika 62, pp. 277-288.

LETSINGER, J. D.; MYERS, R. H.; and LENTNER, M. (1996). "Response Surface Methods for Bi-Randomization Structures". Journal of Quality Technology 28, pp. 381-397.

LIANG, K. Y. and ZEGER, S. L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models". Biometrika 73, pp. 13-22.

LIN, D. K. J. and TU, W. (1995). "Dual Response Surface Optimization". Journal of Quality Technology 27, pp. 34-39.

LUCAS, J. M. (1989). "Achieving a Robust Process Using Response Surface Methodology". Presented at the American Statistical Association Conference, Washington, DC.

LUCAS, J. M. (1994). "How to Achieve a Robust Process Using Response Surface Methodology". Journal of Quality Technology 26, pp. 248-260.

LUCAS, J. M. and JU, H. L. (1992). "Split Plotting and Randomization in Industrial Experiments". ASQC Quality Congress Transactions. American Society for Quality Control, Milwaukee, WI.

MCCULLAGH, P. and NELDER, J. A. (1989). Generalized Linear Models, 2nd ed. Chapman & Hall, London, UK.

MEAD, R. and PIKE, D. J. (1975). "A Review of Response Surface Methodology From a Biometric Viewpoint". Biometrics 31, pp. 803-851.

MINKIN, S. (1993). "Experimental Designs For Clonogenic Assays In Chemotherapy". Journal of the American Statistical Association 88, pp. 410-420.

MITCHELL, T. J. (1974). "An Algorithm for the Construction of D-Optimal Experimental Designs". Technometrics 16, pp. 203-210.

MYERS, R. H.; KHURI, A. I.; and CARTER, W. H., JR. (1989). "Response Surface Methodology: 1966-1988". Technometrics 31, pp. 137-157.

MYERS, R. H.; KHURI, A. I.; and VINING, G. G. (1992). "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach". American Statistician 46, pp. 131-139.

MYERS, R. H.; KIM, Y.; and GRIFFITHS, K. (1997). "Response Surface Methods and the Use of Noise Variables". Journal of Quality Technology 29, pp. 429-441.

MYERS, R. H. and MONTGOMERY, D. C. (1995). Response Surface Methodology. John Wiley & Sons, New York, NY.

MYERS, R. H. and MONTGOMERY, D. C. (1997). "A Tutorial on Generalized Linear Models". Journal of Quality Technology 29, pp. 274-291.

MYERS, W. R.; MYERS, R. H.; CARTER, W. H., JR.; and WHITE, K. L. (1996). "Two Stage Designs for the Logistic Regression Model In Single Agent Biassays". Journal of Biopharmaceutical Statistics 9, pp. 121-128.

NAIR, V. N. (ED.) (1992). "Taguchi's Parameter Design: A Panel Discussion". Technometrics 34, pp. 127-161.

NEFF, A. and MYERS, R. H. (1998). "Recent Developments in Response Surface Methodology and Its Application in Industry" in Statistical Process Control for Quality and Productivity Improvement edited by S. Park and G. G. Vining. Marcel Dekker, New York,

NY (to appear).
ORIS, J. T. and BAILER, H. J. (1993). "Statistical Analysis of The Cerodaphnia Toxicity Test: Sample Size Determination for Reproductive Effects". Environmental Toxicology and Chemistry 12, pp. 85-90.
SITTER, R. R. (1992). "Robust Designs for Binary Data". Biometrics 48, pp. 1145-1155.
SITTER, R. S. and TORSNEY, B. (1995). "D-optimal Designs for Generalized Linear Models" in Advances in Model Oriented Data Analysis edited by C. P. Kitsos and W. G. Miller. Physica-Verlag, Heidelberg.
VINING, G. G. and BOHN, L. (1998). "Response Surfaces for the Mean and the Variance Using a Nonparametric Approach". Journal of Quality Technology 30, pp. 282-291.
VINING, G. G. and MYERS, R. H. (1990). "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach". Journal of Quality Technology 22, pp. 38-45.
VINING, G. G. and SCHAUB, D. (1996). "Experimental Designs for Estimating Both Mean and Variance Functions". Journal of Quality Technology 28, pp. 135-147.
ZEGER, S. L. and LIANG, K. Y. (1988). "Longitudinal Data Analysis for Discrete and Continuous Outcomes". Biometrics 42, pp. 121-130.

DISCUSSION

DOUGLAS C. MONTGOMERY
Arizona State University, Tempe, AZ 85287-5906

SOME COMMENTS ON FUTURE DIRECTIONS IN RSM

I would like to express my appreciation to the authors for their insightful and illuminating presentations. From my experiences as both a practicing engineer and an engineering educator, the topics presented here by the authors are of enormous significance. The response surface framework has always been the most useful, direct approach to process and product characterization and optimization. I first realized this in a course on RSM that I took from Professor Myers at Virginia Tech in 1967. That course initiated an interest in designed experiments that has continued for 30 years. These papers are landmark works for the Journal. I will confine my remarks to observations on four eras in the development of modern statistical design and some comments on future directions for RSM.

The first or agricultural era of statistical design was led by the pioneering work of R. A. Fisher, beginning around 1920. Fisher recognized that the analysis of data from systems (in this case agricultural systems) was often hampered by flaws in the way the experiment that generated the data had been performed. He developed the insights that led to the three basic principles of experimental design: randomization, replication, and blocking. Fisher systematically introduced statistical thinking and principles into designing experimental investigations, including the factorial design concept, the basic building block of much subsequent work in the field.

Applications of statistical design in industrial settings certainly began no later than the 1930's, perhaps with the work of L. H. C. Tippett. However, these early activities had no broad impact on manufacturing industry. The second or industrial era was catalyzed by the development of RSM in the late 1940's and the publication of the seminal paper by Box and Wilson (1951). Over the next 30 years, RSM and other design techniques spread throughout the chemical and process industries, although most often the applications were in research and development work. Anytime one looks at the literature of this field, the profound impact made by Professor Box is understood. His genius has truly touched every aspect of the field. During this 30-year period, the application of statistical design at the plant or manufacturing process level was still not extremely widespread. Some of the reasons for this include inadequate training in basic statistical concepts and methods for engineers, scientists, and other process specialists and the lack of computing resources and software to support widespread applications.

The third era of statistical design was ushered in by the increasing interest of Western industry in quality improvement that began in the late 1970's. As Professor Myers has indicated, Genichi Taguchi's approach to what he called the robust parameter design problem was the first view of statistical design that many engineers had seen. Because the problem was a real and important one, this had a significant impact on expanding the interest in and use of designed experiments.

Taguchi's methodology generated much discussion and controversy. Part of the controversy arose because the methodology was advocated in the West initially (and primarily) by entrepreneurs, and the underlying statistical science had not been adequately peer-reviewed. By the late 1980's the results of peer review indicated that while Taguchi's engineering concepts and objectives were well-founded, there were substantial problems with his experimental strategy and methods of data analysis.

There were two positive outcomes of the Taguchi controversy. First, designed experiments became more widely used in industrial settings that previously made little use of the technique, such as the discrete parts industries, including automotive and aerospace manufacturing, electronics, and semiconductors. Second, the fourth era of statistical design began. This era has included a renewed interest in statistical design by both researchers and practitioners.

RSM will continue to expand to new and technically-challenging environments. Experiments with computer models is an area presenting many opportunities. There are several ways to classify these problems, such as deterministic simulation models versus stochastic simulation models. Deterministic simulations, such as circuit design tools, finite element analysis models, and other CAD/CAM tools are widely used in engineering design and analysis. There has been little work in the statistical design field that transfers naturally and easily to the engineering design community, as the initial RSM tools so successfully did. There have been many applications of RSM to stochastic simulators--I have done some of this work myself. However, many problems remain. For example, increased computer capability has led to simulators with massive numbers of variables, such as the factory-scale models used in the semiconductor industry and models used to study the behavior of very complex systems, such as aircraft engines. Many of these models produce time series output. These are rich opportunities for us.

As Professor Myers has noted, the multiple response problem, including parameter estimation techniques, optimization procedures and designs, will be of increasing importance. It is not unusual to find RSM applications with 20 or more responses, many of which are correlated. Much work remains to be done in this area. For example, when there are several responses, it is not unusual to find that some formulations of the overall objective have multiple optima. Some of the solution procedures currently used in practice and implemented in widely-used computer software either do not recognize this issue or do not deal with it very effectively. More interaction with the operations research community to develop and exploit optimization techniques that are effective in finding the "right" solution is likely to be profitable.

As applications of RSM and designed experiments in general continue to expand, responses that do not satisfy the usual "nice" assumptions will be more frequently encountered--such as proportions, counts, and continuous responses that are dramatically non-normal. An analysis framework is needed for these cases based on the generalized linear model, as suggested by Professor Myers, that parallels the normaltheory case in ease of use so that practitioners can apply it easily. Much more research on the selection of appropriate choice of designs for these situations is needed.

Professor Box has elegantly and persuasively discussed the powerful inductive capabilities that arise from factorial designs. Increased emphasis on education on these principles is still needed--many people still think that statistics is something you do to your data once it's been collected. Unfortunately, many of these individuals are my faculty colleagues in engineering and the sciences. Not long ago I read a column in the New York Times extolling the joy of finding the defective bulb in a string of Christmas tree lights by a process that the author referred to as "a form of controlled experimentation utilizing the sound engineering principle of changing the factors one-at-a-time." Imagine my horror at discovering that this column was written, not by a journalism major, but a famous engineering professor. We have much work to do.

Dr. Montgomery is a Professor in the Department of Industrial and Management Systems Engineering. He is a Fellow of ASQ. His email address is doug.montgomery@asu.edu.

DISCUSSION

NECIP DOGANAKSOY and GERALD J. HAHN GE Corporate Research and Development, Schenectady, NY 12301

THERE are few instances in life when one can predict the future with confidence. This paper, written by George Box, one of the founding fathers of industrial applications of the design of experiments (DOE) and one of the world's most eminent statisticians, and his associate, Patrick Liu, is a unique exception; we can safely predict that it will be widely read and highly appreciated by industrial statisticians and quality technology professionals. It will serve as one of the first comprehensive case studies for many statistics (and hopefully engineering and science) students. Their views on how DOE's can be integrated into important applications will be significantly shaped by this paper. As part of their DOE course requirements, students will engage in experiments similar to the helicopter example (some short courses already do this). In fact, we would like to see this example presented early on in DOE courses and used throughout to illustrate important points and practical considerations. Students will persistently seek (and most likely achieve) better results. It is conceivable that some day, someone will come up with an innovation which will enable flight times (exceeding 600 centiseconds?) currently considered unimaginable! In fact, it might be instructive to ask students to design their helicopters at the beginning and then again at the end of the course and compare the results. The "helicopter data" will be analyzed and reanalyzed--and, perhaps, in time, become a classic similar to Fisher's famous iris example. We feel, indeed, privileged to have been asked to comment on this paper.

Many discoveries have been credited to the folkloric notion of "discovery by accident" (remember the curious cat in the GE commercial who discovers a high grade polymer by accident?). This is most unfortunate. Even though this is a widely promoted perspective on how discoveries come about (even from first years of schooling), it is hardly appropriate for today's environment in which companies compete ferociously to achieve valuable patent rights and to be first on the market. Unquestionably, this concept must be replaced by the understanding that most discoveries will require continuous and highly disciplined experimentation, closely integrated with physical understanding and assessment. This, of course, is the essence of the iterative approach that Box and his associates have advocated for years! In fact, we feel that gaining an understanding of the basic concepts of DOE, although not necessarily many of its details, is so important that all engineering and science students should have some exposure to it--something that is, somewhat belatedly, happening in our company now. It should certainly be a part of all introductory courses in statistics (which, in our judgment, frequently focus much too heavily on the details of data analysis, without providing an understanding of the importance of obtaining statistically valid data in the first place) and be built upon in engineering and science lab projects.

Would we have taken the same path in the design and analysis of the helicopter study? Most probably not. However, this is one of the authors' key points: Scientific learning is self-correcting.

We agree fully with the authors' main points and approach. Specifics will obviously vary with the situation. In light of our expectation that the helicopter example will become a classic for academic and industrial training, we felt an even further discussion of this example from an industrial perspective would be useful. Thus, our comments will follow from the viewpoint of effectively adopting the proposed approach in industrial situations for process improvement (e.g., chemical process design) and product design (e.g., new thermoplastic resin or new locomotive engine design).

SOME PRACTICAL CONSIDERATIONS

An investigation of this type will progress iteratively, along with improved physical understanding, as the authors make very clear. For the proposed approaches to become reality, significant changes need to occur in the training of statisticians and practitioners and in the development of soft skills to operate effectively in an investigative team environment. As already suggested, similar changes must also occur in the engineering and science curricula. The authors make this point several times, but we don't think it can be overemphasized. It relates closely to furthering work in our contributions to mechanistic modeling, DOE's on computer models, and so forth, which are now widely used to enhance scientific learning.

The details, however, will vary with the situation at hand and depend upon how speedily experimental responses, referred to by the authors as the "immediacy," can be realized. The situations encountered in industry still vary a great deal in this regard. In many cases, and ever increasingly so in our highly computerized environment, one can obtain instantaneous, or near instantaneous, responses. These cases lend themselves especially well to an iterative approach. In other situations, complete responses are delayed. As pointed out in the paper, this is often the case in agricultural statistics--the birthplace of modern experimental design--and may explain why an iterative approach did not get appropriate attention in the early days. It is also so for some manufacturing processes that require many steps over a long period of time, such as semi-conductor manufacturing or batch polymerization processes. And it is most certainly the case in many product life applications--although there we do have the added interesting aspects of censored data. This results in what is often the most important information, that is, premature failures, occurring earliest. Similar considerations, undoubtedly, also apply in many medical applications.

In the actual design of a new helicopter, most of the testing and DOE's will, undoubtedly, occur at the component level (i.e., tails, wings, engine, etc.), rather than on the full system, as in the example. This is due to both cost considerations and logistical factors (i.e., the components are generally available long before the system design has been frozen). In fact, often only a very limited number of full (prototype) units would be available for testing, generally at the optimization and validation stages. Integration of the component data with the system results, whether or not obtained via a DOE, is a key challenge.

In their specific application, just as in real life situations, the authors had to make a number of somewhat subjective decisions such as dropping the "paper clip" option early in design, not pursuing the two-way interactions from the first study, and selection of factors and their levels. These examples illustrate the critical decisions that need be made at various stages of the experiment that impact its eventual success and general validity. The ability of the "DOE expert" to integrate with the investigative team in the thought process behind such decisions is crucial to contributing to the investigation in the broad sense advocated by the authors.

SOME TECHNICAL QUANDARIES

The helicopter example suggests a number of key concepts that students might raise or instructors might wish to present.

VARIANCE COMPONENTS AND THEIR ESTIMATION

Implementation of the helicopter experiment involves two phases: building helicopters and conducting flights. The first phase might be much more expensive than the second. Industrial experimenters would take this fact into consideration by using the same helicopter, or parts of it, over and over. This suggests the possibility of conducting repeat runs or "repetitions" on the same helicopter, as suggested briefly by the authors. This, in fact, was done in the first experiment in which there were four repeat runs on the same helicopter. In contrast, in the third experiment, there were four center point "replicates," each involving a different helicopter, that is, distinct helicopters built under the same conditions, although these also involved multiple runs. There are, thus, clearly two components of variance. The first, measured by repeat runs on the same helicopter, is that induced by variations in method of launch, projection and flight, wind conditions, and so on for a particular helicopter. The second, potentially measured by building replicate helicopters at the same target design and manufacturing conditions, is the variation due to variability in material, construction technique, and so forth. Clearly, both sources of variation are of interest and need to be measured. Moreover, as the authors point out, the repeat variance would itself likely be a function of the helicopter design. In fact, one of our goals is to identify design conditions where both the repeat and replicate variance are minimized, that is, identify design conditions that are robust to both sources of variation.

If one conducted only replicate runs (i.e., single runs on different helicopters) the residual variance would provide a combined estimate of the repeat and replicate variance, and the two components would not be distinguishable. Thus, we would want to conduct repeat runs merely to be able to estimate separately each of these two sources of variation and to study how this variation differs from one condition to the next. (Note the relationship to Taguchi inner and outer array designs--whose development, in our opinion, was, at least in part, generated by practical cost considerations such as those in this example). In addition, cost considerations would also lead one to conducting repeat runs on the same helicopter versus building a large number of helicopters with few runs on each. Of course, this leads to the question of the appropriate number of repeat runs--the subject of a tutorial paper all in itself. However, we suspect that knowledgeable experimenters might have a pretty good "gut feeling" about this number.

Of course, once we decide to do repeat runs, the next question is to decide on the sequence of such repeats. At one extreme we might conduct the repeats one right after another. We would advise against that for obvious reasons, including that our skill at launching these flights might change over time. At another extreme, we might do repeat runs in blocks or randomize the sequence altogether. However, this might not be very practical. A possible compromise would be to do two repeat runs right after each other and then repeat in, say, three (or whatever the appropriate number might seem to be) blocks. As always, cost considerations become relevant in making this judgment.

The possibility of changes over time in our experimental technique needs to also be addressed in the experimental plan in its own right. For example, a possible source of concern might be damage to the helicopter induced by previous flights and landings that impact its subsequent performance. If this is, indeed, an important consideration it could add another possible can of worms--the potential of identifying those helicopter design conditions that are most robust to flight damage! All of this again provides the opportunity to stress two key points: (1) the importance of making the experimental plan maximally robust against such concerns and (2) taking such concerns into consideration in our analysis of the experiment, for example, by doing residual plots versus time. Needless to say, the differentiation between repeat runs and replicates also needs to be factored into the formal analysis of the experimental data.

SPLIT PLOT CONSIDERATIONS

There is even more to the likelihood that the major costs of this experiment are those associated with building helicopters, rather than those costs associated with flying them. Specifically, consider a 2[sup3] full factorial design involving the factors wing length, body length, and paper clip (this is not one of the situations considered in the paper but is used here to make a point). Implicit in the authors' discussion is that this would require 8 distinct helicopters, with randomized run order. In a real life situation, due to cost considerations, undoubtedly some of the helicopters would be "reworked" and "reused" in the experiments to cut cost. Suppose, for example, it is possible (and in the helicopter example it seems that it is) to go from a long wing (body) length to a short wing (body) length by just "chopping off" part of the helicopter. (However, presumably, this change can not be reversed). Then, let's take a closer look at this 2[sup3] full factorial design (l denotes wing length and L denotes body length):

Run Order	l	L	Paper Clip
1	long	long	yes
2	long	long	no
3	long	short	yes
4	long	short	no
5	short	short	yes
6	short	short	no
7	short	long	yes
8	short	long	no

The preceding plan can be run with 2 helicopters only. The first 6 runs will use the same helicopter by reworking its wing and body lengths and using (or removing) the paper clip. The last two runs will require a new helicopter due to irreversible change that occurred in the body length of the first one. Cost considerations set aside, this sequence might also be appealing to experimenters due to the fact that length is a "harder to change" factor than paper clip. Obviously, this has implications on randomization, design replication, etc. and thus analysis of data, all of which warrant classroom discussion. For example, the two helicopter plan clearly does not allow randomization and can therefore result in biased estimates if, for example, there are learning factors involved in conducting the experiment or if the plane itself is impacted by previous flights. Therefore, in practice, one might opt for a compromise between the minimum possible construction of two helicopters and the maximum of eight.

HOW DO WE HANDLE THESE QUALITATIVE VARIABLES?

It was conveniently found, in the first phase of the experiment, that the qualitative variables did not appear important--and, therefore, we were able to exclude them in subsequent phases. A reasonable subject for classroom discussion would be how we would have proceeded if this had not been the case.

"STEEPEST ASCENT" STRATEGY

The authors illustrate the "steepest ascent" strategy in attempting to continually improve the helicopter design. This is most appropriate in this application, in which there appears to be just one response variable (distance of flight) to maximize. Matters become more complex when there are two or more response variables--especially if their desirable values tend to be negatively correlated--or when a major goal is to quantify the response surface, rather than to seek a single optimum. See our discussion of the accompanying Myers paper.

ANALYTIC STUDIES

Box appropriately makes reference to the concept of analytic studies, popularized by Deming (1953, 1975). Most scientific and industrial experiments are, indeed, analytic, not enumerative, studies. This means that that we are not taking a sample from a well defined frame (as, e.g., in a public opinion survey), but, instead, are dealing with a dynamic ever-changing process. Moreover, we are generally interested, not in the operating environment of the experiment, but in the future environment in which we will be building the product. Thus, for example, the prototype helicopters that we build in our experiment might be made under much more restrictive and better controlled conditions with regard to materials, manufacturing technique, etc. than the conditions we can expect in actual large scale production. This has, at least, two significant consequences.

The consequence of lesser importance is the fact that formal statistical evaluations need to be used with much caution, if at all (e.g., formal confidence intervals provide only a lower bound on the total uncertainty). However, most critical, as emphasized by Box, is that, like Student and his associates in their early brewery investigations, we make the experiment as broad as possible. This means introducing, to the degree possible and in an appropriately controlled manner, a variety of material batches, operators, manufacturing conditions, etc. with the goal of making the resulting experiment maximally robust. This concept again might run against the classical misconception of wanting to hold all variables, except one, constant!

This paper contains a wealth of wonderful insights. Our comments focus mainly on the helicopter example and scratch only the surface of this thought-provoking work. We congratulate the authors for their "instant-classic" contribution.

Ray Myers is one of our foremost authorities and researchers on response surface methodology (RSM). His paper is a thoughtful and forward-looking assessment of the field. It provides a beacon to the future and suggests some wonderful opportunities and challenges to many generations of Ph.D. students and researchers. His emphasis on the need for robustness of designs is especially appropriate.

Our comments deal, principally, with the relevance to industrial applications of Myers's comments. The context is a corporate environment in which the design of experiments, including RSM, is no longer left principally to the professional statistician. Instead, it is a powerful tool for gaining understanding and improving processes and operations in active use, throughout our organization, by folk with only limited statistical training. In recent years, we have seen a rapid increase in the use of RSM and related approaches in our company. Even though the majority of applications have been to chemical and other manufacturing operations, we are beginning to see competent applications in less traditional areas, such as service, TV broadcasting, and financial operations. In fact, we continue to marvel at the fact that while only a few years ago the term DOE was associated with the U. S. Department of Energy; today--as a consequence of the Six Sigma initiative--just about everybody, from our CEO down, knows that it refers to the design of experiments (see Hahn and Hoerl (1998)). As a consequence, in addition to "robust parameter design" and "robust experimental design," we need to make DOE robust against potential misapplication by practitioners who do not have the time and training to recognize all the nuances that may be obvious to people with extensive training in statistics. These three types of robustness, of course, are not independent.

THE GOALS AND CONSEQUENCES OF RSM

Myers defines the goal of RSM to be that "... of determining conditions on the design variables that provide process improvement or perhaps even process optimization." We believe that Myers will agree that, in addition, a broader objective is often that of just gaining process understanding, that is, obtaining a quantitative assessment of the impact of changes in the process variables (the X's) on the response variables (the Y's). Sometimes, such understanding will lead directly to improvement by, say, setting target values and allowable ranges of variation for the process conditions in our factory. However, in many other cases, although the ultimate goal may be improvement, the process is not quite as direct. For example, we may, as Myers has pointed out, have multiple Y's--some of which might even be deterministic. Thus, including more impact modifier in a chemical formulation might improve the strength of our product in a manner estimated from the results of our experiment, but it may also increase cost in a known manner. Now, different customers have different strength requirements, which they wish to achieve consistently and at a minimum cost. Hence, "optimum" conditions vary from one customer, and possibly even from one application, to the next. However, the common thread that ties all of this together is the estimated functional relationships (for different Y's) that result from the designed experiment.

None of this, of course, contradicts Myers's definition--it just expands on it. However, it does impact the experimental strategy. In particular, an iterative design strategy seeking, for example, the path of deepest ascent, is more appropriate if the goal is to find some single optimum condition--even if this optimum is a function of multiple Y's--than if the major purpose is to get a quantitative representation of a response surface.

SOME REFLECTIONS ON TAGUCHI

Myers's discussion of Taguchi's parameter design led us to reflect on the impact of what he correctly refers to as the "Taguchi era." About ten to fifteen years ago (memory fades with age), we were besieged with calls requesting "Taguchi" experiments (a decade or two earlier, some of us received similar requests for a "Latin Square" experiment!). There was, often contentious, debate about "classical experimental designs" versus "Taguchi designs." This was not helped by the sometimes aggressive approach of some of the Taguchi promoters.

We are glad to report that the dust has now settled. We are no longer being pressured to use a "Taguchi approach," or to apply some of the less than optimal (to put it kindly) approaches to data analysis that had, in the past, been associated with Taguchi--and whose deficiencies have been well documented in, and appropriately improved upon, in the statistical literature. Moreover, Box, Hunter, and others again receive their just recognition as the prime movers in bringing DOE to industry.

However, Taguchi has had a profound and, in our opinion, highly beneficial impact. As Myers suggests, this has been on two major fronts:

* Bringing "more attention to experimental design" as a key tool for gaining understanding. Taguchi's engineering background, flair for promotion, and proven successes helped convince Western industry of the importance of DOE--something from which we have all

benefited.

* Emphatically moving to the forefront the concept of parameter design, in general, and the importance of reducing process variance, in particular. Although addressed by Box and his associates long ago, this concept did not really catch on until it was forcibly driven home by Taguchi. We have found Taguchi's concepts to be right on the mark in some key applications. For example, in fabricating thermoplastic resin pellets that are subsequently to be molded into plastic parts, the average viscosity of the material supplied to a particular customer may be anywhere within a fairly generous range. However, it is essential that, whatever that average may be, we maintain it consistently from batch to batch--so that our customers do not have to adjust their machine settings for each batch.

A POSSIBLE GENERALIZATION OF ROBUST PARAMETER DESIGN

Myers indicates that robust parameter design leads to "variance modeling," in addition to the classical mean modeling. This certainly seems appropriate in our pellet fabrication example. However, there are other applications in which, although the mean response is not the key criterion of interest, the variance--even when taken together with the mean--does not fully capture what we are after, either. This is especially so in dealing with non-normally distributed responses, particularly if the key item of interest is an extreme distribution percentile. For example, the major measure of interest may be the first percentile of the time to failure distribution of a product (or the proportion of product that fails prior to a specified time) or the 95th percentile of the distribution of the waiting time to respond to a request for product service. In these and many other real life examples in which a key quantity of interest lies in one of the tails of the distribution, the underlying distribution is not normal. Moreover, as we all learned in our first course in statistics, the impact of non-normality tends to be most profound in estimating the distribution tails. Also, we note that frequently there are unique process factors that explain these extreme results that differ from those that have a major impact on the distribution mean. Thus, in our time to failure example, the first percentile of the distribution might be the result of infant mortalities (or even "dead on arrival" product) due to a manufacturing flaw. Moreover, this flaw might impact only some of our product--for example, those units built on one of 15 machines. In such cases, mean modeling and variance modeling are not enough. The challenge is to develop appropriate design and analysis strategies that allow us to understand the impact of the process variables on, say, a specified percentile or other distribution characteristic of a response variable beyond just the mean and variance--especially in dealing with non-normal responses.

EXTENSIONS TO PLANNING RELIABILITY AND LIFE TESTS

In the spirit of this paper, we will discuss another situation, not traditionally considered part of standard RSM, in which the multistage approach to planning, as outlined by Myers, seems to provide a natural fit. This is the area of life testing and, in particular, accelerated life testing of products. In this context, key planning decisions involve determination of stress levels and allocation of test specimens to these levels. Often such testing serves multiple purposes: extrapolation to user stress, product comparisons, identification of failure modes, and so forth. Statistically optimum test plans, typically designed to estimate a specified low percentile of the life distribution with high precision, have been shown to be highly sensitive (nonrobust) to even modest deviations from model assumptions. See, for example, Nelson (1990). (Such plans, nevertheless, provide useful benchmarks for comparison.) The "censoring" aspect of life data (i.e., unfailed units), however, provides some flexibility to revise experimental conditions while the investigation is still underway. This has long been recognized and exploited by practitioners to achieve "robustness" of test programs. For example, Meeker and Escobar (1998, p. 511) describe a situation in which the engineers increased the number of test specimens at a high test stress after concluding from the initial results that occurrence of any failures at the lower test stress during the planned duration of the test would be unlikely. This intuitively based, dynamic approach to planning can be formalized to provide further efficiency and discipline. As Myers points out, though, a knowledgeable person "being there" is a key element in effectively identifying and addressing evolving needs of the experiment.

CHALLENGES OF DOE IN THE COMPLEX WORLD

Myers reminds us of the many important elements that enter into the skillful planning of an experiment and why the design of an experiment is as much an art as it is a science. This also is one reason why computer generated designs can get one only so far, even when embedded with appropriate "expert knowledge." In addition to robust parameter design, these complexities include

* Designs that are maximally robust to deviations from model assumptions, recognizing the various ways such deviations can occur and their potential impact.

* Binary responses, leading potentially to logistic regression models and other non-normal response situations. Such models, although originated in biomedical applications, also have significant applicability to industrial situations when one has a go-no-go response, as, for example, when product has been classified as defective or non-defective. (In most situations, it is more informative to obtain a measurement, rather than a dichotomous response. However, in some cases, a dichotomy may be unavoidable--these include situations where one is dealing with historical data for which the measurements have been lost; products such as switches or flash bulbs that either operate or do not; or situations for which there are multiple responses that need be combined into one criterion, in addition to being analyzed separately).

* The utilization of relevant, and sometimes subjective, prior knowledge that might lead one to a Bayesian type design.

* The frequent need to deal with multiple responses (and, possibly, to try to combine these).

* Restriction on the design space for reasons that might range from getting essentially uninteresting (or previously known)

information to the danger of blowing up the plant.

* The all pervasive nature of split plot situations in practical application: All variables are generally not created equal; some are easier to vary than others. Similarly, it might often be reasonably inexpensive (and statistically efficient) to obtain replicate observations, rather than to conduct repeat runs. Of course, in such cases, the resulting data need to be analyzed appropriately.

* The need to implement the robust approaches to test planning in commercial software.

Myers provides many useful comments and insights on these and other situations to which we have little to add--other than our congratulations.

Dr. Doganaksoy is a Statistician in the Applied Statistics Program. He is a Member of ASQ

Dr. Hahn is Manager of the Applied Statistics Program. He is a Fellow of ASQ. His email address is hahn@crd.ge.com.

ACKNOWLEDGMENTS

We would like to thank our colleague Martha Gardner for helpful comments that we have incorporated into the discussion.

ADDITIONAL REFERENCES

DEMING, W. E. (1953). "On the Distinction Between Enumerative and Analytic Surveys". Journal of the American Statistical Association 48, pp. 244-255.

DEMING, W. E. (1975). "On Probability As a Basis for Action". The American Statistician 29, pp. 146-152.

HAHN, G. J. and HOERL, R. (1998). "Key Challenges for Statisticians in Business and Industry" (with discussion). Technometrics 40, pp. 195-210.

MEEKER, W. Q. and ESCOBAR, L. A. (1998). Statistical Methods for Reliability Data. John Wiley & Sons, New York, NY.

NELSON, W. (1990). Accelerated Testing: Statistical Models, Data Analyses, and Test Plans. John Wiley & Sons, New York, NY.

DISCUSSION

J. STUART HUNTER 503 Lake Drive, Princeton, NJ 08540

OVERVIEW

BOTH of George Box's papers discuss the importance of statistics as an integral part of the scientific method. To elucidate the argument, the first paper, written in collaboration with P. Y. T. Liu, employs the scientific method repeatedly in the pursuit of a longer flight time response for a paper helicopter.

The example is simple yet sophisticated. We find ourselves locked into a sequential learning process, alternating between the generation of hypotheses based on present knowledge and crucial feedback from data resulting from surprises and failures provided by the experiments. We are compelled to design new experiments and again to complete a learning feedback loop through data analysis. The responses change; the variables and their ranges change. Back and forth the process goes until a level of knowledge adequate to satisfy the investigator is achieved. The final analysis raises still more questions and opportunities. The linkage to Deming's edict of never ending improvement is clear.

Dr. Box's approach throughout is the scientific method. In the helicopter example an empirical approach is employed which is now widely identified as response surface methodology (RSM). The model postulated is a Taylor series approximation of a smooth unknown function, a polynomial. If a theoretical model had been known, say, based on the theory of air flows over thin foils, it would likely have been used. The early experimental design chosen allowed for several simple (first order) empirical models to be fitted in any or all of the variables for each response along with measures of experimental error variance. If a theoretical model were in hand, then only a single experiment might have been designed to entertain the model. Response surface designs and analyses are demonstrations of enlightened empiricism. Dr. Box's message is not simply RSM, but rather its use within a larger iterative learning process called the scientific method.

Dr. Myers's extensive paper discusses how the earlier simple empirical attributes of RSM have dramatically changed over the past few decades due mainly to the role of the computer. He acknowledges Professor Genichi Taguchi's resourcefulness in adapting experimental design to the design, manufacture, and the environmental hardiness of industrial products. Both Myers and Box discuss "optimal" designs and the important concept of robustness. Under an expanded RSM umbrella, Dr. Myers goes on to discuss generalized linear models, nonparametric regression, Bayesian methods, and non-linear problems. The papers by Box, Liu, and Myers span a vast number of topics. Additional ones whose origins can be found in early RSM work are studies of the linkages between linear and non-linear models (Box and Youle (1955)) and the discrimination between competing models (Hunter and Reiner (1965) and Atkinson and Fedorov (1975)).

OPTIMAL DESIGNS

OPTIMAL EXPERIMENTAL DESIGN

The use of computer software to construct "optimal" experimental designs receives a note of caution from both Dr. Myers and Dr. Box. Designing a good experimental program always requires the balancing of competing objectives. Small numbers of trials are the enemy of precise estimates. Replication at only one point argues that experimental variance is homogeneous over the entire studied region. Aliases endanger every saturated and near-saturated design. Lack of fit measures challenge every postulated model.

Blocking combats variance but can lead to split plot arrangements amongst the variables. The convenience of using designs with variables at only two or three levels can take precedence over other more "optimum" designs which allow settings of variables to occur anywhere within their range. Even granting that an optimum design can be constructed, it becomes a fleeting accomplishment: the moment the first few days have passed, or observations are in hand, the analyst begins to wish the design had been just a little bit different. As someone once said, "The only time to design an experiment is after it has been run."

A designed experiment should be robust. It should allow the experimenter to confirm his assumptions, identify aberrant observations, and provide the opportunity for surprise. The purpose of an experimental design is not decision making but enhancing the learning process. Further, the conduct of the experiments must be such that, should the trials be repeated at another time or place, the new results will confirm those of the present experiment. Simple designs that provide for replication with randomization most often prove robust.

In RSM there is little need for a computer generated design. Standard designs, the two level factorial or fractionals, the Box-Behnken, or the central composite design will almost always do quite well. Of course conditions do exist in empirical investigations in which constraints on the variables or novel experimental regions occur that require the construction of a design unique to the situation. Computers are important in these unusual situations and certainly necessary in the design of experiments for models non-linear in their parameters.

ROBUSTNESS

The first time the word "robustness" entered the published statistical literature was in Dr. Box's 1953 paper in Biometrika. This early paper concerns tests of hypotheses for means and variances. In the paper Dr. Box comments, "there is abundant evidence that comparative tests on means are remarkably insensitive to general non-normality of the parent populations ... this remarkable property of 'robustness' to non-normality ... is not necessarily shared by other statistical tests, and in particular is not shared by the tests for equality of variances, ..." (p. 318). Nevertheless, even though the point was made in 1953 we continue to see in today's beginning textbooks and software programs the recommendation that before using two sample averages to test the hypothesis of the equality of the two population means, the two associated estimates of variance first should be tested for homogeneity. A t-test with unpooled variances and adjusted degrees of freedom is then proposed, or a suggestion to employ some non-parametric procedure is offered. Concerning such activity Box commented (p. 333), "To make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!" Tests on variances are not robust to departures from normality; tests on means are robust.

Dr. Box goes on to comment that "assumptions have sometimes ... been interpreted rather too literally. For this reason it is most important that derived criteria should be studied for robustness" (p. 333). The same is true today when one considers efforts for constructing "optimum" experimental designs. Concentrating on obtaining a special variance-covariance matrix can easily divert an analyst from what is really important in designing an experiment. To keep from getting wet in Dr. Box's rowboat, whether using Student's t or designing an experiment, one is tempted to recommend the acronym "KISS": keep it simple statistician.

NON-PARAMETRICS

Of interest is Dr. Myers's description of non-parametric response surface designs (NPRSM) wherein the primary objective is the construction and use of a "space-filling grid" over a region that may have considerable non-linearity. The methodology is very different from the use of RSM wherein an approximating polynomial models is employed over a small region. NPRSM designs are constructed with the objective of filling a large region with sufficient points to allow the use of splines or similar interpolants. Often, every point in a large cartesian grid comprises the design. These designs can be expensive in the number of experiments, most particularly when the number of variables exceeds two. Randomization is seldom mentioned as part of the investigation, nor is the idea of blocking ever considered. If one is not careful, the tempting words "non-parametric" used in combination with "experimental design" become an oxymoron.

NPRSM investigations can of course serve as the stimulus for subsequent standard RSM approaches. At best, however, NPRSM methods are a distant numerical cousin to RSM. The objective of RSM is not mimicry of historical data over an experimental region, but the construction of an empirical model that may prove a useful forecast for the future. Further, the empirical model should be both simple and parsimonious, that is, limited in the number of fitted parameters. Leaps of faith linking empirical evidence to knowledge require useful models.

THE 2^k FACTORIAL

The robustness and usefulness of the 2^k factorial designs and fractionals is an important message. Dr. Box makes the argument that the 2^k designs allow for numerous "same or different" comparisons, each clear of the other. Further, following Fisher, he goes on to note that inferences derived from such studies are valid across all the variables entering the study. Projections of the designs into lower dimensionality continue to reinforce their usefulness. They are robust.

An attribute of the 2^k factorials I find interesting is their capacity to illustrate "coupling" (non-additivity). To explain, consider the Taylor's series second order approximation of a bivariate smooth deterministic function $\eta = f(x_1, x_2)$, that is,

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2.$$

Now ask an engineer or researcher to list in order importance the various coefficients. Usually the first to be chosen are beta[sub1] and beta[sub2], the "linear" terms, followed by beta[sub1] and beta[sub22], the "curvature" terms, with the crossproduct or interaction coefficient beta[sub12] relegated to last place. In most minds curvature naturally follows linearity. But the function is bivariate, and the only coefficient possessing information on this important bivariate quality is the beta[sub12] coefficient. Should beta[sub12] equal zero the model would simply separate into two uncoupled univariate functions.

The bad habit of contemplating curvature without simultaneously considering interaction persists, as illustrated by the widespread use of the four variable 3[sup4-2] fractional factorial (the L9 design or Graeco-Latin square). These and other similar three-level designs are frequently touted as "optimal" in terms of number of runs required and ease of analysis. Unhappily, they are not robust to the presence of coupling (two factor interactions) and serious misinterpretations of data can easily occur (see Hunter (1989)). Of course one can no more interpret the meaning of a two factor interaction in the absence of its associated quadratic effects than one can judge the importance of a covariance without knowledge of the associated variances. RSM directs us to consider the entire second order model.

MULTIPLE RESPONSES

Both Dr. Box and Dr. Myers mention the problems associated with handling multiple responses. In the early work on RSM an investigator was advised to superimpose contour maps of the different responses and to locate, visually, "optimum" operating conditions amongst the variables. The approach was informal and, should no immediate solution arise, allowed the investigator to trade off various alternatives amongst the responses and variables. The feedback from data to hypothesis was thus encouraged.

Another more formal approach is to convert the multivariate responses into a single univariate response index, fit a model, and then locate the settings of the independent variables that optimize the index. Of course, weighting the responses to form the index is highly subjective, and the variance of an index can be disturbing. On occasion, two or more indices can be formed and their contour maps studied, or indices constructed containing responses that were not part of the experiments. With computers, the use of response indices can quickly become fun and games. Alternatively, various mathematical programming techniques are available for locating best conditions subject to constraints among both the multivariate responses and the variables. Further study is needed to investigate the influences of standard errors in the parameters comprising the fitted models used in finding these best regions. In recent years "chemometric" data analysis techniques, in particular PLS (projected latent structures), have been used in the study of data involving many variables and multivariate responses. These analyses offer experimenters the opportunity to design experiments in the essential multivariate sub-spaces of both responses and variables.

THE INVERSE PROBLEM

The classical direct problem is to find the output of a system given the inputs to the system model. The inverse problem has the desired output known, while it is the system input that is to be determined. Obtaining regions in variables, x, that produce desired response characteristics, y, offers serious design of experiments challenges not recognized by many statisticians. In its simpliest statistical form one is provided with the fitted straight line y = b[sub0] + b[sub1]x[sub1], where b[sub0] and b[sub1] are least squares estimates with variances and covariances. Given a specific value for the response, produce a confidence interval estimate for the variable x[sub1]. Alternatively, imagine the fitted model is second order in two variables x[sub1] and x[sub2] and find the confidence region in the space of x[sub1] and x[sub2] associated with the estimated maximum response. One will often find the confidence region not closed. Closed confidence regions about a solution for x[sub1] and x[sub2] given a specific response would seem a desirable objective. (Remember too, the desired response might provide a contour.) In an early RSM example, Box and Hunter (1954), after fitting a second order model and solving for the location of the maximum response, added a series of experimental runs to close the confidence region. The size and shape of the confidence regions in the variables that can provide a desired response offer a challenge to experimental design.

In recent years the "inverse problem" has attracted the attention of many engineers. An example would be the manufacture of an item that must have a particular temperature-time profile at a given spot. Given a model for heat transfer and the initial boundary conditions, the time path of temperature at any point in or on the item can be determined. But consider the inverse problem. Given a required temperature-time path at some spot in or on the item, what are the initial conditions at the source? Given temperature-time measurements on the external surface of a rotating kiln, what are the originating temperature gradients within the kiln? Given the time dependent trace of a pollutant concentration at downstream position on a river, determine the location and magnitude of a possible addition of the pollutant up-stream. In every case, where to locate sensors and how frequently to record the data become the problems of experimental design.

Engineers and scientists often have an important advantage over most statisticians in that they have a "theoretical" model for the phenomenon under study. Their models are dynamic with both temporal and spatial characteristics, ordinary or partial differential equations which in integrated form are commonly non-linear in their parameters. Like statisticians, they too are locked in an iterative learning process and concerned with the problems of data planning and analysis leading to hypothesis testing and generation.

An interesting early text co-authored by an engineer and a statistician on estimation and analysis of non-linear models is by Beck and Arnold (1977). Recent texts on the inverse problem are by Tarantola (1987), Hensel (1991), and a proceedings has been published by the American Society of Mechanical Engineers (Zabaras, Woodbury, and Raynaud (1993)). There is even an e-mail Inverse Problems Network Newsletter published by Prof. Patrica Lamm at Michigan State University (lamm@math.msu.edu).

Statisticians anxious to extend the arts of experimental design and analysis could profitably begin to interface with these engineers. As in the case of chemometrics and biometrics, one can perceive here an emerging new branch of statistical application.

FINAL COMMENTS

As undoubtedly intended by our editor, the papers by Box, Liu, and Myers compel us to contemplate where we have been, where we are now, and to what future we tend. The subject of RSM proved an attractive entrée to a vast smorgasbord of statistical topics. But the message I gathered from reading these papers went far beyond a discussion of current theory and application of RSM. Discovered anew was the challenge to pass along the message of the importance of the learning process and iterative nature of the scientific method.

How can the fraternity of quality engineers and statisticians best meet this challenge? Answer: Recognize ourselves as a profession. This requires the establishment of accredited undergraduate programs in applied statistics and quality engineering. Also needed are extensions of present certification procedures to identify individuals skilled in the combined arts of applied statistics and problem solving. Already available are the journals and technical meetings that define a profession. We have Dr. Box to thank for making so obvious the philosophy and practice of iterative learning and the scientific method. Needed only is a greater sense of self-awareness.

Dr. Hunter's email address is stu@ceor.princeton.edu.

ADDITIONAL REFERENCES

ATKINSON, A. C. and FEDOROV, V. V. (1975). "Optimal Designs: Experiments for Discriminating Between Several Models". Biometrika 62, pp. 289-303.

BECK, J. V. and ARNOLD, K. J. (1977). Parameter Estimation in Engineering and Science. John Wiley & Sons, New York, NY.

Box, G. E. P. (1953). "Non-Normality and Tests on Variances". Biometrika 40, pp. 318-335.

HENSEL, E. (1991). Inverse Theory and Applications for Engineers. Prentice-Hall, Englewood Cliffs, NJ.

HUNTER, J. S. (1989). "Let's All Beware the Latin Square". Quality Engineering 1, pp. 453-465.

HUNTER, W. G. and REINER, A. M. (1965). "Designs for Discriminating Between Two Rival Models". Technometrics 7, pp. 145-160.

TARENTOLA, A. (1987). Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation. Elsevier, New York, NY.

ZABARAS, N.; WOODBURY, K. A.; and RAYNAUD, M. (Eds.) (1993). Inverse Problems in Engineering Theory and Practice. Engineering Foundation, American Society of Mechanical Engineers, New York, NY.

DISCUSSION

ANDRÉ I. KHURI
University of Florida, Gainesville, FL 32611

OVERVIEW

GEORGE Box, Raymond Myers, and Patrick Liu are to be commended for their timely and insightful discussions concerning the current status of response surface methodology (RSM). For many years, both Box and Myers have been major contributors to the development of RSM. Their assessment of the evolution of RSM and its future directions should therefore be very informative.

I believe that research work in the area of RSM has reached a critical turning point since the introduction of generalized linear models (GLM's) in the statistical literature. The rapidly growing field of biostatistics and the pharmaceutical industry are the main beneficiaries of GLM applications. Historically, the key ideas in RSM were developed using classical linear models with continuous responses and fully controllable input variables. Some or all of these ingredients might be missing in situations where GLM's are used. The need to re-evaluate and/or possibly modify existing RSM techniques within the framework of GLM's is therefore urgent and of utmost importance. Unfortunately, there have been no concerted efforts on the part of veteran RSM researchers to meet the new challenges brought about by the introduction of GLM's.

It is true that the "clientele using RSM has broadened." This, however, may be attributed to the availability of commercial software packages with some RSM capability. Thus while the pool of users of RSM has widened, the pool of RSM researchers has not. As a result, core research in RSM has lagged behind. It is distressing that only very few departments of statistics teach RSM as a separate course. What Norman Draper related to Myers in the late 1970's, that "we who are working in the response surface area have to stick together. There are not many of us left," is still, unfortunately, very true today.

VARIANCE MODELING

The interest in variance modeling has increased in recent years. However, in spite of the many published articles in this area, there is still no clear consensus as to how to effectively model both the mean and variance of a response in a given experimental situation. Some model forms have been suggested for the variance, but no mechanism is available for testing lack of fit of such models. It is also not clear whether to model the mean and variance separately, as was the case in the dual response approach by Vining and Myers (1990), or perhaps assume some relationship between the mean and variance (see Engel (1992) for a discussion on modeling variation in industrial experiments). Furthermore, little is known about the selection of designs for the efficient estimation of both the mean response and the variance function. Vining and Schaub (1996) have done some work in this direction. More work is needed.

Another concern here is how the noise variables are to be treated in a given experiment. These variables contribute to the variance of the response. As Khattree (1996, p. 197) pointed out, "while during the experimentation, the underlying linear model is a fixed effects model, for the estimation of the variances, the model needs to be treated as a mixed effects model." Khattree provided a method for assessing the effect of noise variables in the context of robust parameter designs in situations where the noise factors cannot be studied within a single experimental layout. Tuck, Lewis, and Cottrell (1993, p. 673) stated that "if it is difficult and time consuming to change the levels of the noise factors, it would be preferable that the experiments are run so that the design factor levels are changed within each combination of noise factor levels. This is in contrast with many published case studies in which the noise factor levels are varied within each combination of the design factor levels."

DESIGNS FOR GENERALIZED LINEAR MODELS

The majority of designs for GLM's are for logistic models with one control variable. The use of the D-optimality criterion is overwhelming. This is hardly a RSM problem. I find the reliance on a single criterion function for choosing a design to be too simplistic and artificial. Dette and Sahm (1997) demonstrated that some of the commonly used optimality criteria can yield rather awkward designs, in the sense that the optimal designs degenerate into one-point designs, which are useless.

It should be remembered that D-optimality for GLM's is really based on a linearization of the link function using Taylor's series approximation, which is then followed by applying the same methodology as in linear models. The main difference is that with GLM's we have the unsettling problem of dependency of the design on the unknown parameters in the model. This is not a problem afflicting only GLM's. The same difficulty occurs in nonlinear models of which the class of GLM's is a special case. Several years ago I addressed the dependency problem with regard to general nonlinear models. I introduced a design criterion that did not require specification of initial values of the model's parameters. This criterion is based on approximating the mean response function using Lagrange interpolating polynomials (see Khuri and Cornell (1996, Ch. 10)). However, the proposed methodology applies only to nonlinear models with a single control variable and may require a large number of design points depending on the desired approximation accuracy associated with Lagrange interpolation.

There is a need for response surface designs that apply to GLM's with several control variables and which are efficient with respect to several criteria, such as protecting against lack of fit of the model for the linear predictor or perhaps the wrong choice of the link function (see Khuri (1993)). Another possibility is extending the use of the well known integrated mean squared error criterion to GLM's.

There are two alternatives to assigning "guessed values" for the unknown parameters when selecting a GLM design: the Bayesian approach and the sequential approach. These approaches are not free of faults. In the former, the selection of design depends on the choice of the prior distribution of the unknown parameters. Sun, Tsutakawa, and Lu (1996) discussed some problems associated with Bayesian designs for quantal responses. In the latter approach, time trends or seasonal effects may be introduced if the experimental conditions change over time (see Chaloner and Verdinelli (1995, p. 291)). Perhaps a combination of these two approaches may produce better results whereby the choice of prior distributions is based on information acquired in previous stages of the experiment.

Graphical techniques may prove effective in comparing designs for GLM's. Such techniques were used with standard response surface models as in Giovannitti-Jensen and Myers (1989); Myers et al. (1992); and Khuri, Kim, and Um (1996). The advantage of these techniques is that they provide information concerning the adequacy of the design under consideration over the whole experimental region. They can also provide additional information concerning the robustness of a design to the choice of initial values of the model's parameters. The use of a particular graphical approach in conjunction with nonlinear models was described in Khuri and Lee (1998).

MULTIRESPONSE EXPERIMENTS

Undoubtedly, the most visible aspect of multire-sponse analysis is optimization. There are other aspects that have received some attention, such as the choice of design and lack of fit testing (see Ch. 7 in Khuri and Cornell (1996) and the review article in Khuri (1996a)). Unfortunately, as was pointed out by Myers, "not enough attention is given by practitioners to the consideration of correlation structures among responses." Observing multiresponse data in experimental work is now a commonplace. In fact, it is rare that only one response variable is considered in a given situation. As Box succinctly stated, "in most real examples there would be several responses." More often than not, the responses are correlated. Ignoring this fact amounts to discarding information that may be vital to a proper exploration of the multiresponse system.

More research work is sorely needed in the multiresponse area, particularly with regard to design selection and GLM's. The known multiresponse methods apply mainly to continuous responses. Extending these methods to responses having discrete distributions is essential. Design considerations and determination of optimum conditions under the more general GLM setting can be quite challenging problems.

It should be noted here that little is known about optimization of the mean response using GLM's. The paper by Brinkley, Meyer, and Lu (1996) is one of very few articles that address this issue. The authors used Poisson regression techniques combined with mathematical programming to identify operating conditions that minimize the predicted mean and variance of the occurrence of defects in a manufacturing process.

RESPONSE MODELS WITH RANDOM EFFECTS

The analysis of response surface models that include random effects, such as batch effects, has been receiving more attention in recent years. This represents a significant departure from the traditional form of a response surface model, which contains only fixed polynomial effects. The inclusion of random effects broadens the scope of a response surface investigation. It also brings RSM closer to other fields of statistics that are currently receiving great attention, such as mixed models and longitudinal data analysis. The use of random effects in a response surface model was described in Khuri (1992) and was later extended in Khuri (1996b) to models containing interactions between random effects and the fixed polynomial effects. Optimization of the predicted response in the presence of random effects was discussed in the latter paper.

CONCLUSION

There are many challenging problems facing RSM investigators. Tackling these problems requires the development of new tools and the initiation of closer collaborations between academic and industrial RSM workers. This is particularly true given that "the opportunities for the use of statistics and particularly experimental design in industrial investigations were growing at an unprecedented pace," as Box pointed out. It is comforting to know that the key ideas of RSM that were outlined in Box and Wilson (1951) are still viable today and can provide guidance in meeting the challenges in the 21st Century.

Dr. Khuri is a Professor in the Department of Statistics. His email address is ufakhuri@stat.ufl.edu.

ADDITIONAL REFERENCES

BRINKLEY, P. A.; MEYER, K. P.; and LU, J. C. (1996). "Combined Generalized Linear Modelling--Non-linear Programming Approach to Robust Process Design--A Case Study in Circuit Board Quality Improvement". Applied Statistics 45, pp. 99-110.

DETTE, H. and SAHM, M. (1997). "Standardized Optimal Designs for Binary Response Experiments". South African Statistical Journal 31, pp. 271-298.

ENGEL, J. (1992). "Modelling Variation in Industrial Experiments". Applied Statistics 41, pp. 579-593.

GIOVANNITTI-JENSEN, A. and MYERS, R. H. (1989). "Graphical Assessment of the Prediction Capability of Response Surface Designs". Technometrics 31, pp. 159-171.

KHATTREE, R. (1996). "Robust Parameter Design: A Response Surface Approach". Journal of Quality Technology 28, pp. 187-198.

KHURI, A. I. (1992). "Response Surface Models with Random Block Effects". Technometrics 34, pp. 26-37.

KHURI, A. I. (1993). "Response Surface Metodology Within the Framework of GLM". Journal of Combinatorics, Information and System Sciences 18, pp. 193-202.

KHURI, A. I. (1996a). "Multiresponse Surface Methodology" in Handbook of Statistics edited by S. Ghosh and C. R. Rao. Elsevier, Amsterdam. pp. 377-406.

KHURI, A. I. (1996b). "Response Surface Models with Mixed Effects". Journal of Quality Technology 28, pp. 177-186.

KHURI, A. I. and LEE, J. (1998). "A Graphical Approach for Evaluating and Comparing Designs for Nonlinear Models". Computational Statistics and Data Analysis 27, pp. 433-443.

KHURI, A. I., KIM, H. J., and UM, Y. (1996). "Quantile Plots of the Prediction Variance for Response Surface Designs". Computational Statistics and Data Analysis 22, pp. 395-407.

MYERS, R. H.; VINING, G.; GIOVANNITTI-JENSEN, A.; and MYERS, S. L. (1992). "Variance Dispersion Properties of Second-order Response Surface Designs". Journal of Quality Technology 24, pp. 1-11.

SUN, D.; TSUTAKAWA, R. K.; and LU, W. S. (1996). "Bayesian Design of Experiment for Quantal Responses: What Is Promised Versus What Is Delivered". Journal of Statistical Planning and Inference 52, pp. 289-306.

TUCK, M. G.; LEWIS, S. M.; and COTTRELL, J. I. L. (1993). "Response Surface Methodology and Taguchi: A Quality Improvement Study from the Milling Industry". Applied Statistics 42, pp. 671-681.

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 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).

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 (boot-strapping, 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.

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.

ACKNOWLEDGMENTS

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

FIGURE 1. A Flow Chart for Response Surface Methodology.

ADDITIONAL REFERENCE

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

RESPONSE

GEORGE E. P. BOX

I am grateful and indebted to the discussants for the warmth with which the papers by Mr. Liu and myself have been received and for the many cogent and thought-provoking comments.

They will perhaps sympathize with me when I confess that over the years the sessions that I have attended, allegedly about response surface methodology (RSM), have frequently left me wondering whether the speaker and I were talking about the same subject. As indicated by their titles, our papers are not primarily about RSM, but about innovation and discovery with statistics. To try to avoid misunderstanding, however, the papers use an example (the helicopter) within an example (RSM) which in turn is used to exemplify some general principles for statistical methods of use in scientific investigation.

I think we have spent too much time on one-shot statistical procedures designed to test rather than to learn. I have explained how I think this has come about, largely because of the idea that we can develop statistics from mathematical ideas conceived at our desks. In particular, for optimal experimental design, the experimenter is supposed to know in advance which factors, which region of the factor space, which function, and which constraints should be considered. The experimenter is thus credited, on one hand, with the prescience of the Oracle of Delphi, and on the other with sufficient naiveté to accept the simplistic solutions that may be offered.

When an experimental design is run the most relevant question is "So now what do we do?" I have been saying this for a very long time (see, e.g., Box (1957)). I had hoped that we had seen the end of the obscene tribal habit practiced by statisticians of continually exhuming and massaging dead "data sets" after their purpose in life has long since been forgotten and there was no possibility of doing anything useful as a result of this treatment.

But in today's mail I see the fourth such disinterment from a 16 run fractional factorial design originally obtained at Chrysler (Hsieh and Goodwin (1986), Bisgaard and Fuller (1994), Myers and Montgomery (1997), Nelder (1998)). These data have now been subjected to Taguchi methods, various kinds of transformations, and to generalized linear models (twice) with and without consideration of Chipman's "weak and strong heredity principles." The later authors believed that they should produce "models." These differ somewhat, which is hardly surprising since there are nine main effects and thirty-six two factor interactions, but only sixteen runs. But why (see again Daniel (1961) and Bisgaard (1999)) must we think in terms of factorial effects? If you look at the results they have analyzed, there are just three runs that are perfect with no defects (0, 0, 0). By contrast, the three worst runs give large numbers of defects (56, 17, 50). The remaining ten runs are not very bad, but not very good--with defects (2, 4, 3, 4, 2, 1, 3, 12, 3, 4). For the three very good runs, factors D and F are at the plus levels; and for the three very bad runs, D and F are at their minus levels (and this is for a number of different combinations of the other factors). So what I would do, I think, is to tell the management that I believe the problem can be solved by increasing mold pressure (D) and using method II for priming (F) and that we need a few more runs to check this out. This is not very different from what the original authors did (although using biased Taguchi style "analysis of variance" additional factors were declared significant and changed, but this was probably harmless). The important thing is they eventually saved their employers an estimated $900,000 per year. (Notice that in the original paper two results were given at each set of conditions, but that only the first of these has been used by subsequent authors.)

Now that Patrick Liu and I have published all this helicopter data I fear that the grave robbers will be out in full force. I would like to restrain them and try to persuade them to move away from this preoccupation with morbid anatomy and instead to concentrate their attention on researching, teaching, and developing the art of keeping a living investigation healthy. We ought to be developing further statistical methods for scientific inquiry (not necessarily RSM). This is certainly a very open field, though surprisingly little has been done.

When a new topic in statistics pops up, the standard procedure is to formalize, generalize, and present alternative solutions. But one may ask, where do such new topics (in some instances breakthroughs) pop up from? I have argued elsewhere, quoting some 34 examples extending from Gauss to Efron and beyond, that they have most often arisen from practical need (Box (1984)). In particular, when a model that at first looks sensible does not lead to good practice, we know we have work to do. I think I can best make my point in terms of a few examples. The first may be well known to you, but is perhaps worth repeating.

I. During World War II, sampling schemes were used in which, say, 30 items were submitted to an accelerated test and if three or more failed then the batch was rejected. Given the assumptions, this was an optimal procedure. However, a certain officer in the U.S. Navy raised the question, "Suppose the first three I test turn out to be bad, do I have to inspect the other 27? And if I do not, then how could the optimal schemes you have recommended be optimal? It was a good question, but Wallis (1980) has described how the mathematical statisticians who protected Abraham Wald believed that to look for such "super-powerful" tests would be a waste of his time. When at last Wald was consulted, however, he quickly saw that the optimality in question assumed a fixed sample size and he (and also George Bernard at about the same time) developed sequential procedures based not on the proportion of rejects in a sample of fixed size, but on the size of the likelihood ratio. There was no fixed sample size, and indeed such tests were in general more economical. Thus, sequential hypothesis testing came to be invented.

II. In 1960, when Gwilym Jenkins and I first started to think about feedback adjustment methods suitable for statistical process control, it seemed natural to develop schemes for which the deviations from the target had minimum mean square error (MMSE). We worked out schemes that did this. But not long afterwards, Dr. Park Reilly, who had been one of Gwilym's Ph.D. students and had taken an important job as an industrial statistician in Canada, told us how this "optimal" theory applied to one of his processes had led to a control algorithm which was totally non-sensical. It went something like this: observations were being taken, say, every 30 minutes with temperature as a control variable. The algorithm which would, theoretically at least, have provided MMSE at the output was telling him at one point to increase the temperature by 2000° C and 30 minutes later to reduce it by 1500° C! We quickly realized that our MMSE criterion needed to be modified.

Practical schemes requiring remarkably little manipulation at the cost of only a slight increase in the output mean square error above the minimal level were soon discovered (Box and Jenkins (1968, 1970), Aström (1970), and Aström and Wittenmark (1984)). Such schemes could be complicated, however; and in more recent work it has been possible (again, by modifying the criterion), to obtain almost equally good schemes that are very simple to design and to use (Box and Luceño (1997)).

III. Several discussants have mentioned the use of nonlinear models, the need to consider multiple responses, and the problems caused by correlation and by constraints. I have always preferred to use theoretical models when from physical, chemical, or engineering considerations approximate mechanisms could be conjectured. When available such models can lead to deeper understanding of the system being studied and to much faster progress. Frequently, however, for want of something better, we have to begin with empirical models hoping always that they might provide insight that can lead to a more satisfactory mechanistic model. In my later years at I.C.I. and after I joined the Statistical Techniques Research Group at Princeton in 1956, my colleagues and I were particularly interested in nonlinear estimation required for mechanistic models (see, e.g., Box (1960)). However, in such studies it is common to measure not one, but a number of responses simultaneously (in particular the amounts of various substituents at different reaction times). In what can be regarded as a generalization of least squares it was shown that estimation with multiple correlated responses required the minimization of the determinant of sums of squares and products of deviations from expectation which automatically took care of correlation between such deviations. (Box and Draper (1965))

Again practice quickly enlightened theory. A correspondent, this time from DuPont, told us that he had applied our method but the computer could find no unique minimum. As I remember it, each of three substituents, say, y[sub1], y[sub2], and y[sub3], supplied information about one or more of a set of physical constants to be estimated. Because chemical balance required that y[sub1] + y[sub2] + y[sub3] = 100%, it turned out that the experimenter had not actually measured y[sub3] but had obtained it by taking 100 - y[sub1] - y[sub2]. The determinant was then of course zero everywhere because of this linear constraint, and the computer was searching on the basis of a rounding error. By calculating the determinant only for those responses which were actually measured, satisfactory answers were obtained. My Monday Night Beer Seminar was regularly attended by a number of graduate students from chemical engineering who said that data of this kind were often published with no indication that such dependencies had been introduced. They found a publication where experimental data, on 5 responses, were again modeled in terms of kinetics and again the determinant was zero everywhere as before. However, after eliminating the (1, 1, 1, 1, 1) constraint, the 4 × 4 determinant was still hovering close to zero. The engineers searched the library and unearthed the original investigation. From this it was clear that a second dependency had been introduced because one of the substituents had been calculated by a chemical balance formula involving two of the others. Clearly, dependence of this kind was quite likely to be missed, so we devised an eigenvalue analysis that could check empirically for dependencies. A paper about this problem was later published by four of the regular Monday nighters (Box et al. (1973)). Notice that, as commonly occurs, the nature of the constraints was one more thing that was not known a priori and had to be learned about. A somewhat similar problem arises in re-estimating parameters from data collected from a system which is undergoing feedback control. (Box and MacGregor (1974, 1976)).

IV. When I first began to look at response surfaces, I think I assumed that maxima would be reasonably symmetric. For example, in three dimensions contours would surround the maximum like the skins of an onion. But the systems I studied almost always had ridgey maxima or consisted of stationary or rising ridges. (Ridges, of course, relate directly to interactions and to the fact that changes in variables like temperature, concentration, and pressure can often compensate each other.) Such ridges were again identified by eigenvalue analysis. In one example where only three factors were being studied, I was somewhat surprised to find two of the eigenvalues close to zero. The contour picture in three dimensions was not like an onion or like a jelly roll, but like a sandwich with the ham being the maximum and the slices of bread being contour planes on which lower yields were obtained. (I'm not sure that industry is sufficiently aware of the economic possibilities that such ridge systems offer. In particular, they raise the possibility of maximizing more than one response and so of using alternative, cheaper processes.)

My physical-chemist friend, Dr. Phillip Youle, was fascinated by this ham sandwich maximum and conjectured it could occur as the result of a particular type of consecutive reaction. You can see the idea if you think of the molecules in the reaction initially as black and white billiard balls, initially with a large preponderance of black balls, all in rapid motion on the billiard table. Now, suppose that whenever a black ball collides with a white ball, it will, with a certain probability, produce a red ball (the desired product), but, similarly, whenever one of these newly created red balls collides with a black ball, a green ball (an undesired product) will be produced. If we set the system going, the number of the desired red balls on the table will first increase then decrease until finally only black and green balls remain. To get the best yield you must stop the system at some time t--the optimal reaction time--when there is a maximum of red balls.

Now running the experiment again at a higher temperature can be simulated by imagining the speed of all the balls on the table to be increased. If the probabilities of all collisions is increased proportionately, then precisely the same sequence will occur, with the same maximum yield of red balls, but in a shorter time. Thus reaction time and reaction temperature will compensate each other. A similar effect will be produced by increasing the initial number of black balls (corresponding to increasing the concentration). So it is easy to see that a stationary ridge system, in reaction time, temperature, and concentration, of exactly the kind we found will result. This will be so, however, only if the speed of all the balls is increased by the same amount. The increase of speed produced by the temperature change is measured by a constant called the activation energy. For this example, the activation energies of the two reactions were evidently about equal. Consequently we got the same yield at many different combinations of temperature, reaction time, and concentration. If the activation energies were unequal, then higher yields could be obtained by changing temperature, and we would obtain a rising ridge (see Box and Youle (1955)). Also, if the "orders" of the two reactions were different, the maximum yield could also be changed by changing concentration. In our analogy this would occur, for example, if a red ball had to collide with two black balls simultaneously to produce a green ball while the rest of the system remained unchanged.

In this investigation then, an interesting empirical response surface study had shown us alternative ways of getting the same yield. Within this framework we could minimize cost and satisfy certain constraints revealed by analysis of other responses. But this analysis did not tell us how to increas the yield. However, the subject matter knowledge which led to physical understanding of the system allowed the chemical engineers to figure out how they might get higher yields and in particular to consider different catalysts which would speed up the first reaction more than the second. (It showed what new dimensions to investigate.)

V. Concerning multiple measurements and the occurrence of constraints, each problem, I think, has to be considered scientifically on its merits. In particular, it is the responses themselves that frequently set the constraints. A priori we don't know where they will be, but must learn about them. For over 20 years I worked with a major manufacturer of packaged foods. We always had multiple responses, and as Stu Hunter has mentioned, we found the superimposition of contour maps very helpful. In particular, they helped us to learn what these constraints were and where they were. The kind of constraint we might be up against was that outside a region defined by certain specific factors, for instance, we would find that the "crumbliness" of the cake was too high. (Initially, we might not even know that "crumbliness" could be a problem.)

I don't think we should dismiss this simple idea of overlapping contours too quickly. We are blessed with three-dimensional eyesight, and nowadays, the computer allows us to use this greatly increased capability. So the same kind of thing can be done with any choice of three factors and any number of responses indicated by contour surfaces of different colors. Several 3-D graphics of this kind can be shown on the computer screen at the same time. It is impossible to over-emphasize the catalytic value of appropriate graphics of this kind which can blend with subject matter knowledge to produce lateral thinking and new ideas. Another interesting example concerned the formulation of a paint used for automobiles. In one particular study, after over one hundred runs "ad hoc" had been made and many responses looked at, I was consulted (I think as a last resort). Going through the notebooks, they said things like, "We can get these responses right here, but then these aren't right" and so forth. The appearance of a newly painted automobile is extremely important to the salesman. It is almost equally important that the paint job should not be spoiled by minor abrasions. Some designed experiments and the careful analysis of what was going on pointed to two important factors that affected the glossiness and the abrasion resistance. Unfortunately, they showed that in these dimensions there was a diagonal band in the two-dimensional factor space where adequate abrasion resistance could be obtained, but that this band was exactly parallel to the limit of the region which gave adequate glossiness. Since the regions did not overlap there was no way of using these variables to obtain both properties simultaneously; this was the root cause for the fruitless ad hoc runs. I showed my two-dimensional diagram to the paint technologists, and I said what you need to do, using your specialist knowledge, is to try to think of a new variable and consider it as constituting a third dimension. This variable must be such that when you travel through the additional dimension either the abrasion resistance band will move down or the glossiness region will move up. Or, the new variable will make these constraint limits go at different angles so the two regions will overlap. They thought about it for some time and tried a few things and finally produced a variable which did exactly what was wanted. These examples all illustrate the ability of statistics, experimental design, RSM, and graphical representation to inspire lateral thinking (see De Bono (1967)).

VI. So far we have talked about the theoretical/mechanistic model, but frequently there are a number of sensible models that might explain the data. A problem important to chemical engineers with whom I've worked is how to discriminate between rival models (see, e.g., Box and Hill (1967)). Discussion of the problem of comparing models that contain a different number of parameters is given in Box and Henson (1969, 1970). Recently, work with chemical engineers at Madison has led to further developments both for univariate and multivariate data (Stewart, Henson, and Box (1996) and Stewart, Shon, and Box (1998)). In these papers a number of examples are analyzed. Note that in the second paper we are once more talking about multivariate data, but multivariate data arising in this particular context.

VII. Concerning dispersion effects I think that, just as is the case for interactions, it is important to consider transformable and non-transformable dispersion effects. There seems to be some feeling that we should be reluctant to transform data, but it surely must be rare that the quantity most easily measured is also that in which it is appropriate and simplest to model. Often the importance of transformation tends to be obscured because, for standard transformations and if the data cover only a narrow range, the transformed data can be (almost) a linear recoding of the original. When the data cover wider ranges, however, large location effects can, of course, produce larger interaction and dispersion effects simply because the choice of metric is inappropriate (see, e.g., examples in Box and Cox (1964) and re-analyses in Box and Fung (1995)).

VIII. Correlation in data is of enormous importance and can occur in various ways. In particular, it is good to see that these days serial dependence is being taken more seriously. But even for contemporaneous correlation we need to distinguish carefully between, on the one hand, relationships between errors and, on the other, relationships between mean values. Again, an example is enlightening.

I remember a small designed experiment on a rotary baker; the main object was to increase yield. The baker contained "canon balls" to break up the product, and the operation could be very noisy. The process workers told me that "they already knew when to expect a high yielding batch because it banged more." So, I arranged a scale of banging with them--if you had to shout 12 inches away from your co-workers ear in order to be heard that was a "ten" and so forth. So, one of several responses reported from the experiment was the banging intensity.

The analysis showed indeed that certain factors produced large main effects and interactions for yield, but effects very similar in sign and magnitude were also found for banging. So, I temporarily forgot about the factorial analysis and plotted banging against yield to get a very good straight line. It seemed possible, therefore, that the factors we had tested were just changing the "gooey-ness" of the product and it was "gooey-ness" that determined yield.

We consulted a rheologist who thought about it and then said that for this particular process this idea might well be correct. He also suggested other simpler and cheaper ways of changing "gooey-ness." These were later tested and produced the desired result.

Notice that, unlike my previous examples, what was important was not the correlation between errors, but the correlation between mean values for yield and banging. Also notice that the solution to the problem came about as the result of a quite unexpected phenomenon which was suggested by a simple but unorthodox analysis.

Questions concerning correlations, multiple responses, and multiple constraints have been around for many years, but there have been very few answers. It is as if we hoped that some purely mathematical solution could be applied to every multiple response problem. I don't believe this is possible, and my examples above are intended to illustrate this. I think the researchers must grit their teeth and also become practitioners. Only then will they become the respected colleagues of investigators, and only then will they discover where the real problems are. If you don't know whether you are a good statistician or not, here is a simple test: If your name is Joe Blow and there is a very tough investigation coming up, and the engineers and scientists say, "We've got to have Joe on the team," then you will have arrived.

From all of this I conclude that whether statistics is taught by statistics departments, industrial engineering departments, or quality engineering departments, we need to reconsider a number of things. Mathematical capability is important, particularly if necessary training is directed towards use in engineering and science, but mathematical thinking and scientific thinking are very different. The prerequisites for a student studying scientific statistics should include experience in an experimental science, and where this is lacking, remedial courses should be required and supplied. Furthermore, Ph.D. theses should not be judged on the amount of mathematics they contain (this might be small or large depending on what is needed to solve the problem), but they must demonstrate the student's ability to catalyze learning and to develop new methods which do this. Promotions, including promotions to tenure, should have similar requirements. The curriculum should be re-examined in the light of these considerations and should place much more emphasis on the usefulness to scientific inquiry of the subjects that are taught.

ADDITIONAL REFERENCES

ASTRÖM, K. J. (1970). Introduction to Stochastic Control Theory. Academic Press, New York, NY.

ASTRÖM, K. J. and WITTENMARK B. (1984). Computer Controlled Systems: Theory and Design. Prentice-Hall, Englewood Cliffs, NJ.

BISGAARD, S. (1991). Letter to the Editor. The American Statistician (to appear).

BOX, G. (1957) "Integration of Techniques in Process Development". Transactions of the American Society for Quality Control. American Society for Quality Control, Milwaukee, WI. Reprinted in 1991 in Quality Engineering 3, pp. 9-26.

BOX, G. (1960). "Fitting Empirical Data". Annals of the New York Academy of Sciences 86, pp. 792-816.

BOX, G. (1984). "The Importance of Practice in the Development of Statistics". Technometrics 26, pp. 1-8.

BOX, G. and DRAPER, N. R. (1965). "The Bayesian Estimation of Common Parameters from Several Responses". Biometrika 52, pp. 355-365.

Box, G. and FUNG, C. (1995). "The Importance of Data Transformation in Designed Experiments for Life Testing." Quality Engineering 7, pp. 625-838.

BOX, G. and HENSON, T. L. (1969). "Model Fitting and Discrimination". Technical Report No. 211, Department of Statistics, University of Wisconsin, Madison, WI.

BOX, G. and HENSON, T. L. (1970). "Some Aspects of Mathematical Modeling in Chemical Engineering". Proceedings of the Inaugural Conference of the Scientific Computation Centre and the Institute of Statistical Studies and Research. Cairo University Press, Cairo, Egypt. p. 548.

BOX, G., and HILL, W. J. (1967). "Discrimination among Mechanistic Models". Technometrics 9 pp. 57-71.

BOX, G.; HUNTER, W. G.; ERJAVEC, J.; and MACGREGOR, J. F. (1973). "Some Problems Associated with the Analysis of Multiresponsive Data". Technometrics 15, pp. 33-51.

BOX, G. E. P. and JENKINS, G. M. (1968). "Discrete Models for Feedback and Feed-forward Control" in The Future of Statistics edited by D. G. Watts. Academic Press, New York, NY. pp. 201-240.

BOX, G. E. P. and JENKINS, G. M. (1970). Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco, CA. (2nd ed. printed in 1976; 3rd. ed. by Box, G. E. P.; Jenkins, G. M.; and Reinsel, G. C. printed by Prentice-Hall, Englewood Cliffs, NJ in 1994.)

BOX, G. and MACGREGOR, J. F. (1974). "The Analysis of Closed-loop Dynamic-stochastic Systems". Technometrics 16, pp. 391-398.

BOX, G. and MACGREGOR, J. F. (1976). "Parameter Estimation for Dynamic-stochastic Models Using Closed-loop Operating Data". Technometrics 18, pp. 371-380.

DE BONO, E. (1967). The Use of Lateral Thinking. Penguin Books, London, UK.

HSIEH, P. I. and GOODWIN, D. E. (1986). "Sheet Molded Compound Process Improvement". Fourth Symposium on Taguchi Methods. American Supplier Institute, Dearborn, MI. pp. 13-21.

NELDER, J. A. (1998). "The Selection of Terms in Response Surfaces Models--How Strong is the Weak-Heredity Principle?". The American Statistician 52, pp. 315-318.

STEWART, W. E.; HENSON, T. L.; and BOX, G. E. P. (1996). "Model Discrimination and Criticism with Single-Response Data". AIChE Journal 42, pp. 3055-3062.

STEWART, W. E.; SHON, Y.; and BOX, G. E. P. (1998). "Discrimination and Goodness of Fit of Multiresponse Mechanistic Models". AIChE Journal 44, pp. 1404-1412.

WALLIS, W. A. (1980). "The Statistical Research Group 1942-45". Journal of the American Statistical Association 75, pp. 320-335.

RESPONSE

RAYMOND H. MYERS

THE discussants are to be congratulated. My only regret is that some of their points did not occur to me as I was organizing my thoughts for the paper. Many thought-provoking comments were made, and they displayed a sense of very clear understanding of the genesis and roots of RSM, as well as a clear vision of needs for the future. I prefer to not organize responses on a discussant basis but rather by topical area.

MULTIPLE RESPONSES

Each of the discussants commented about the important role of analysis and/or design when the problem involves multiple responses. It is interesting that reactions reflected each of the discussants' own interest and experience. And this is as it should be. I was particularly intrigued with a reference made by Drs. Doganaksoy and Hahn who include the need to consider (in addition to modeling mean and variance) the percentile of the distributions in certain process optimization exercises. It not only touches on multiple response ideas but also robust parameter design. Surely in reliability problems this is crucial, and we must remember that this crosses engineering as well as biomedical disciplines. As a result one may replace the dual modeling idea with point estimates of percentiles of, say, the Weibull or gamma distribution. Of interest then are, perhaps, tolerance interval calculations in an RSM setting. Myers, Kim, and Griffiths (1997) attempt to deal with the problem for the normal error case that is complicated with noise variables. However, this is merely a beginning and much more needs to be done. I dare to emphasize once again that we should look over our shoulder at work being done by the biomedical statistician here.

Dr. Lin expands on a point that I made on multiple responses and it requires a bit more discussion. We have made wondrous advancements in multiple response optimization in the 1990's. This is thanks in large part to the borrowing of technology from operations researchers and numerical analysts. However, there has been next to nothing that accounts for "noise" or the joint sampling distribution of the "point of optima." This should not be surprising. We are still struggling with this problem for the univariate case. By "struggling" I do not mean there is a dearth of solutions, (solutions begin, of course, with Box and Hunter (1954)). Rather, I mean that dealing with constraints on design variables and the use of noise variables renders this an imposing problem. Then there is the practitioner. How many times do you see a practitioner follow up optimization calculations with a joint confidence region on the coordinates of the optimum? While the use of confirmatory runs aids in determining the validity of its optimum point, it is helpful to do other analyses that express "how confident" we are in the selected point. This often provides a guide for the use of future experiments. Extension of this problem to the multiple response case is clearly difficult (as many other multiple response problems are). Perhaps bootstrapping or other resampling approaches may be useful.

TAGUCHI'S PARAMETER DESIGN

I noticed with great fascination the responses on Taguchi's parameter design and its role in the progress of promoting DOE and RSM. Professor Montgomery and Drs. Doganaksoy and Hahn enthusiastically acknowledge the positive role of Taguchi. In fact, Montgomery goes so far as to list the "Taguchi era" as one of those influential eras in DOE, following classic work of Fisher with agricultural experiments and George Box with his brilliant introduction of notions of RSM. This historical analysis by Dr. Montgomery is right on the mark in my opinion. I only wish I had thought of it first. Drs. Doganaksoy and Hahn remind us of the contentious debates in which most of the authors and discussants here took part. But as Stu Hunter so cleverly put it in a telephone conversation with me a few weeks ago, the backlash and fallout filled the air and when we shook the trees a great number of positive things fell to the ground.

COMPUTER GENERATED DESIGN AND DESIGN ROBUSTNESS

It would be pleasing if we (rather than Dr. Montgomery) could have listed a "fourth positive influential era" of DOE as being the era of the introduction of computer generated design. He properly left it out. But I haven't given up on it. If the ideas of design robustness progress in the fashion aptly described by several discussants then computer generated design as a RSM tool will play a vital role. Even now, Bayesian designs and two-stage designs as described in my paper can for the most part only be found through rather lengthy computations. The very important transition that must be made at some point is the movement from design optimality to design robustness by the software companies. Obviously the time is not right yet. More research and demonstration of positive results must come first. I feel that we can progress when we set minds favorably toward robustness and away from strict optimality. The specific comments made by Dr. Lin and Dr. Khuri should help inspire us. In addition Dr. Hunter's comments about the negative aspects of design optimality should be very helpful to practitioners.

GENERALIZED LINEAR MODELS

Generalized linear models (GLM's) is a relatively new area for industrial statisticians, though it is a staple for biostatisticians. But why hasn't the design problem already been solved? Biostatisticians generally use GLM's with designed experiments in single variable cases (dose response situations). Only recently have biomedical and environmental statisticians become interested in multivariable design problems in GLM's (drug combinations and mixtures of pollutants). Dr. Khuri writes passionately about the need for designs for GLM's. I can almost hear the words come off the paper. I am becoming convinced that considerable progress can be made in this area, and I alluded to small illustrations of this for the case of multiple x's in my paper. In addition, I don't feel as if we should necessarily avoid the unfortunate requirement of parameter knowledge. Most problems can be parameterized in such a fashion that what is required can be "guessed" by practitioners (ED values, EC values) or can be estimated with "first stage" designs. Dr. Khuri's idea of mixing a Bayesian approach with a sequential philosophy also has promise. However, it must be made clear that this is an area that should be dealt with in collaboration with practitioners.

MISCELLANEOUS COMMENTS

* Dr. Hunter's eloquent description of and the history of the term "robustness" was a very welcome component in his comments. I think many practitioners are confused about the concept, and they should be because we often throw this term around without a description of "robustness to what." This is particularly true in the current context since we are still trying to formally determine appropriate robustness criteria. JQT readers will benefit from the illustration and the history.

* In many RSM problems process optimization must give way to studies that increase one's knowledge of the process. This is extremely important. In textbooks and other descriptions of classical RSM this is often not discussed. Thanks to Dr. Lin and Drs. Doganaksoy and Hahn for pointing this out.

* Dr. Khuri laments that few statistics researchers are working in the area of RSM. It is true that a relatively small number of active researchers have made contributions in the historic mainstream of the topic as Dr. Khuri has done so successfully. However, unlike the 1970's, there are many competent researchers who have made some inroads in the interesting and relatively new areas such as semiparametric RSM, uses of RSM in robust parameter design, design robustness, variance modeling, and others. Also the number of interested practitioners is increasing rapidly. I truly think that the future does not look so grim. The realization that real world problems exist in new areas (for RSM) such as pharmaceutics, biology, and the environment will invite solutions. My concern is not the number of researchers but rather the nature of our ability to communicate solutions to practitioners. Research papers written in journals never seen by practitioners will not result in implementation of methodology. Communication to practitioners has always been a problem in all areas of statistics. For much of this we must rely on software package professionals. For example, any creative ideas generated regarding development of robust design will not be implemented without available software and proper communication regarding what is involved.

Source: Journal of Quality Technology, Jan1999, Vol. 31 Issue 1, p30, 13p
Item: 1513771