One-Factor-at-a-Time Versus Designed Experiments

Veronica CZITROM

Many engineers and scientists perform one-factor-at-a-time (OFAT) experiments. They will continue to do so until they understand the advantages of designed experiments over OFAT experiments, and until they learn to recognize OFAT experiments so they can avoid them. A very effective way to illustrate the advantages of designed experiments, and to show ways in which OFAT experiments present themselves in real life, is to introduce real examples of OFAT experiments and then demonstrate why a designed experiment would have been better. Three engineering examples of OFAT experiments are presented, as well as designed experiments that would have been better. The three examples have been successfully used in an industrial workshop and can also be used in academic courses.

KEY WORDS: Teaching statistics.

1. INTRODUCTION

Engineers and scientists often perform one-factor-at-a-time (OFAT) experiments, which vary only one factor or variable at a time while keeping others fixed. However, statistically designed experiments that vary several factors simultaneously are more efficient when studying two or more factors.

That is what statisticians know. But in industry, they need to be able to convince adult, practicing engineers that what they have been doing for years can be improved upon. This is particularly true because engineers usually have higher standing in the company than statisticians from the Quality Assurance Department, and hence may be inclined to discount the statisticians' advice unless they understand it. Also, engineers need to learn to recognize OFAT experiments in order to avoid them. When teaching an academic course, it is important to convince engineering and science students that designed experiments are relevant to *their* applications, and to give statistics students (some of whom will work in industry) a better understanding of practical considerations.

In teaching a three-day design of experiments workshop for engineers in industry, the author has found it extremely helpful to use examples of real engineering OFAT experiments, and to compare the OFATs to designed experiments to illustrate why the latter would have been better. It is important to describe the disadvantages of OFAT experimentation early in an industrial workshop, so students do not drop out of the course to do "more important" things. College textbooks, which present a fair amount of statistics up front, often introduce OFAT experiments later: Box, Hunter, and Hunter (1978, pp. 312 and 510); Montgomery (1997, p. 201); and Mason, Gunst, and Hess (1989, p. 101).

The first morning of the three-day design of experiments industrial workshop is an overview. The overview starts with a brief description of what designed experiments are, what they are used for, and how the rest of the industry uses them. The core of the overview is a complete, real example (2^3 with center points) that is used to introduce the basic concepts, including description of the process, planning the experiment, conducting the experiment, analyzing the data with main effect and interaction plots, and reaching conclusions and implementing recommendations. The example is followed by a section on "Why DOE Works—or Why it is Possible to Study Several Factors Simultaneously and Still Get Useful Information." The overview ends with a section on the advantages of designed experiments over OFAT experiments, which will be described in this article.

The student reaction to the overview is very positive. The material is stripped down to bare essentials, and is illustrated by real-life examples they can relate to. In the author's experience, this goes a long way toward convincing engineers (and managers) to use designed experiments.

Section 2 describes advantages of designed experiments over OFAT experiments, and Section 3 gives three examples that illustrate these advantages. Section 4 is a summary. The OFAT examples can be used in both academic and industrial design of experiments courses. The examples are semiconductor industry experiments, and they can easily be adapted for use in other areas.

2. ADVANTAGES OF DOE OVER OFAT EXPERIMENTS

A designed experiment is a more effective way to determine the impact of two or more factors on a response than a OFAT experiment, where only one factor is changed at one time while the other factors are kept fixed, because:

- It requires less resources (experiments, time, material, etc.) for the amount of information obtained. This can be of major importance in industry, where experiments can be very expensive and time consuming.

Veronica Czitrom is a Distinguished Member of Technical Staff at Bell Laboratories, Lucent Technologies, 9333 S. John Young Parkway, Orlando, FL 32819 (Email: czitrom@lucent.com). The author thanks Osvaldo Rodriguez, Juan Becerro, and Charles Storey for their contributions to the article.

Table 1. OFAT experiment in two factors in three runs, with 16 of the 48 wafers at each run

Pressure	Temperature: Standard	Temperature: New
Standard	16 wafers	16 wafers
New		16 wafers

Table 2. Full factorial designed experiment in two factors at two levels each in four runs, with 12 of the 48 wafers at each run

Pressure	Temperature: Standard	Temperature: New
Standard	12 wafers	12 wafers
New	12 wafers	12 wafers

- The estimates of the effects of each factor are more precise. Using more observations to estimate an effect results in higher precision (reduced variability). For example, for full and fractional factorial designs, all the observations are used to estimate the effect of each factor and each interaction (property of hidden replication), while typically only two of the observations in a OFAT experiment are used to estimate the effect of each factor.
- The interaction between factors can be estimated systematically. Interactions are not estimable from OFAT experiments. Engineers who are not using designed experiments often perform a hit-and-miss scattershot sequence of experiments from which it may be possible to estimate interactions, but they usually do not estimate them.
- There is experimental information in a larger region of the factor space. This improves the prediction of the response in the factor space by reducing the variability of the estimates of the response in the factor space, and makes process optimization more efficient because the optimal solution is searched for over the entire factor space

These concepts are now illustrated using three examples.

3. EXAMPLES

3.1 Two Factors in Three Runs

An engineer planned an experiment to compare pressure and temperature for a standard gas anneal process and a new gas anneal process using three experimental runs:

1. Standard pressure and standard temperature;
2. standard pressure and new temperature; and
3. new pressure and new temperature.

The engineer planned to use one lot of 48 wafers, with 16 wafers for each run, for the experiment.

The engineer's experiment in two factors—temperature and pressure—in three runs is shown in Table 1. Sixteen of the 48 wafers are used at each one of the three experimental runs. When the experimental runs are presented in a table, it is clear that there is no information at the new pressure with standard temperature. The experiment studies one factor at a time: starting from the standard pressure and new temperature, only one factor is changed to obtain the other two runs. The standard temperature is compared to the new temperature using the $16 + 16 = 32$ wafers at the standard pressure. The standard pressure is compared to the new pressure using the $16 + 16 = 32$ wafers at the new temperature. Figure 1 is an attempt to draw an interaction graph for the OFAT experiment. The (invented) values of the response at the three experimental conditions is shown as a function of the two factors, temperature and pressure. The interaction between temperature and pressure (difference between the effect of temperature on the response at the standard pressure and the effect of temperature on the response at the new pressure) cannot be estimated because there is no information at the new pressure with standard temperature.

Table 2 shows a designed experiment that could have been performed, a 2^2 full-factorial with two factors (temperature and pressure) at two levels each (standard and new) in four runs. Twelve of the 48 wafers are used for each run, which allows 12 replications of the four-run 2^2 full factorial experiment. To study the effect of temperature, the standard temperature is compared to the new temperature using the $12+12 = 24$ wafers at the standard pressure, and the standard temperature is compared to the new temperature using the $12 + 12 = 24$ wafers at the new pressure. The average of the two comparisons is the main effect of temperature, and the difference between the two comparisons is the interaction between temperature and pressure. The interaction graph between temperature and pressure is shown in Figure 2. All 48 wafers are used to study the effect of temperature, and to estimate the interaction between temperature and pressure.

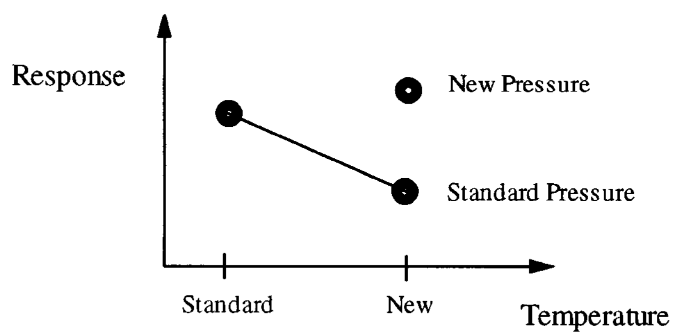

Figure 1. The interaction graph for temperature and pressure cannot be drawn for the OFAT experiment.

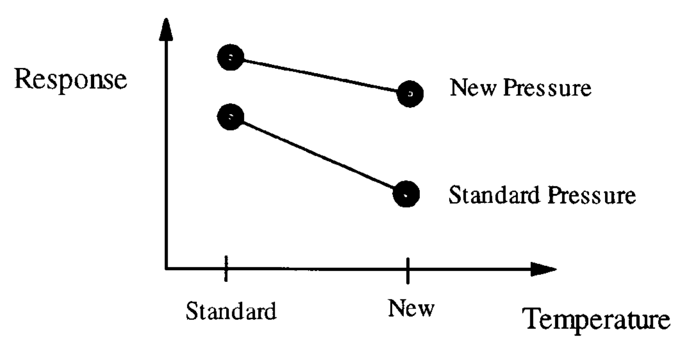

Figure 2. The interaction graph for temperature and pressure can be drawn for the designed experiment.

Table 3. OFAT experiment in two factors in six runs

Time (sec)	Temperature (°C)
10	980
10	1000
10	1020
9	1000
10	1000
11	1000

The designed experiment is better than the OFAT experiment because, using the same 48 wafers (resources):

- The estimates of the effects of each factor are more precise (all 48 wafers are used to estimate the effects in the designed experiment, while only 32 wafers are used in the OFAT experiment). The variance of the estimate of each effect (which is the difference of two averages) is $\mathrm{V(effect)} = 2\sigma^2/24 = \sigma^2/12$ for the designed experiment, and $\mathrm{V(effect)} = \sigma^2/8$ (which is 50% larger) for the OFAT experiment. Here σ^2 is the variance of one observation.
- The interaction between the factors can be estimated (the interaction *cannot* be estimated in the OFAT experiment).
- There is experimental information over a broader factor space in temperature and pressure (there is no information at the new location with standard temperature for the OFAT experiment).

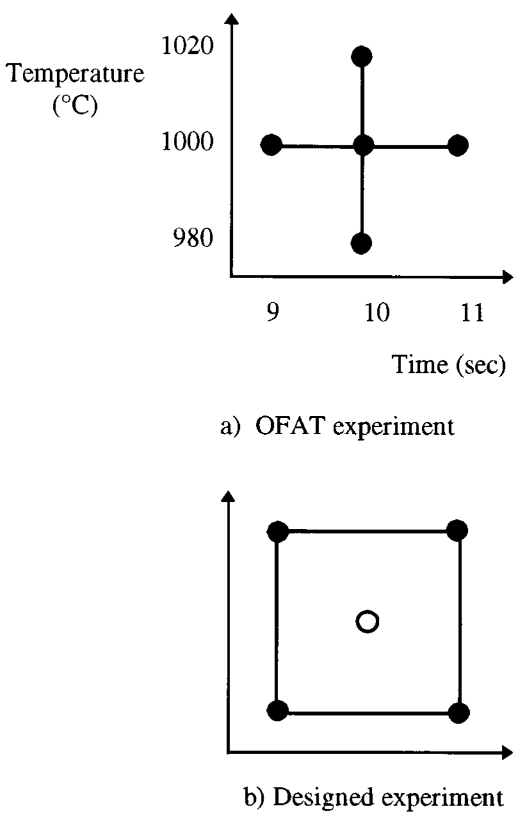

Figure 3. (a) OFAT experiment in five different runs. (b) Designed experiment in four or five different runs.

Table 4. Full factorial designed experiment in two factors at two levels each in four runs

Time (sec)	Temperature (°C)
9	980
9	1020
11	980
11	1020

- The average variance of the estimates of the response at the four experimental conditions is 13% higher for the OFAT than for the designed experiment. For both designs, the estimate of the response at each experimental condition is the average of the observations at that condition. For the designed experiment, the variance of the estimate of the response at each one of the four experimental conditions is $\sigma^2/12$, which gives an average variance of $\sigma^2/12$. For the OFAT experiment, the variance of the estimate of the response at the three experimental conditions is $\sigma^2/16$, and at the new location with standard temperature it can be shown to be $3\sigma^2/16$, which gives an average at the four points of $3\sigma^2/32$ (13% larger than for the designed experiment).

The advantage of the OFAT experiment over the designed experiment is that it requires three runs instead of four (less resources), although in this experiment it is easy to perform the additional run using the same number of wafers.

The full factorial design in Table 2 has 12 wafers at each experimental condition. It would be advisable to use some of those wafers to run center points, at the average level setting for each continuous factor. Center points can be used to check for curvature in the response as a function of the factors, and if curvature is present, the design can be augmented into a central composite design to determine which of the two factors contributes to the curvature in the response. If the center point is replicated, it can be used to estimate natural variability.

3.2 Two Factors in Six Runs

Before taking a design of experiments class, two engineers planned an experiment for a rapid thermal anneal process. They wanted to study the sensitivity of the response sheet resistance to two factors—time and temperature—using the OFAT experiment in six runs given in Table 3 and illustrated in Figure 3a. The effect of temperature is studied using three different temperatures, at the current processing time of 10 seconds. The effect of time is studied using three different times, at the current processing temperature of 1000°C. The interaction between time and temperature cannot be estimated.

After taking a design of experiments workshop, the engineers performed the 2^2 full factorial designed experiment with two factors at two levels each in four runs shown in Table 4 and illustrated graphically by the four full circles in Figure 3b. As in the 2^2 full factorial designed experiment described in Section 3.1, all four runs are used to estimate the effect of time, the effect of temperature, and the interaction between time and temperature.

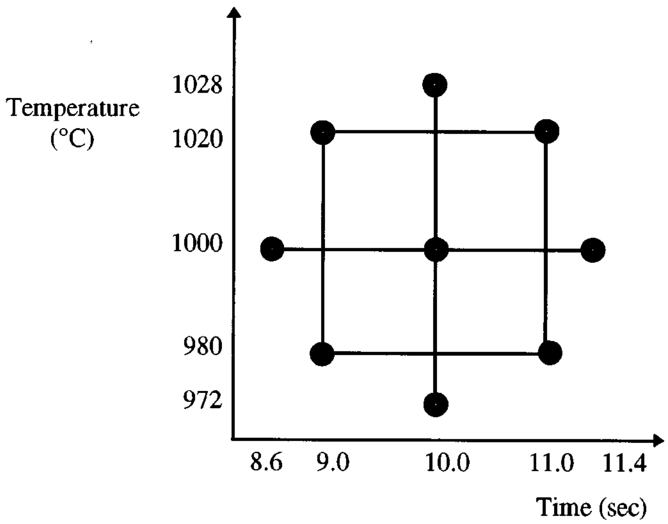

Figure 4. Central composite experimental design in two factors in nine different runs: four runs of full factorial, four axial points, and one center point.

The 2^2 designed experiment is better than the OFAT experiment because:

- It requires less resources (four runs instead of six).
- The estimates of the effects of each factor are more precise (four runs are used to estimate each effect in the designed experiment, three runs are used to estimate each effect in the OFAT experiment).
- The interaction between the factors can be estimated (the interaction *cannot* be estimated for the OFAT experiment).
- There is experimental information in a larger region of the factor space. For example, the effect of dose is studied at two temperatures (980°C and 1020°C) in the designed experiment, but only at one temperature (1000°C) in the OFAT experiment.

One advantage of the OFAT experiment over the 2^2 full factorial designed experiment is that it can be used to estimate curvature in the factors, namely curvature in the response as a function of temperature when time is 10 seconds (along the vertical line of circles in Figure 3a), and curvature in the response as a function of time at a temperature of 1000 °C (along the horizontal line of circles in Figure 3a). Another advantage of the OFAT experiment is that the center point is replicated twice, and can be used to estimate natural variability.

If the engineers want to determine whether there is curvature before running an experiment to estimate it, the 2^2 full factorial in Table 4 can be augmented with one or more center points, illustrated by the empty circle in Figure 3b. If the 2^2 full factorial with two centerpoints is run, it is better than the OFAT experiment because, with the same resources (six runs),

- The estimates of factor effects are more precise.
- The interaction between the factors can be estimated.
- There is experimental information in a larger region of the factor space.

The advantage of the OFAT is that it can be used to estimate curvature along the two lines of circles in Figure 3a, although the designed experiment can be used to determine *whether* there is curvature. If there *is* curvature, it can be estimated by augmenting (with blocking) the 2^2 full factorial with center points, into the central composite design shown in Figure 4.

The central composite design in Figure 4 has nine different runs. If the center point is replicated four times, the central composite design is rotatable (equal precision of estimation at all points equidistant from the center point), and the replicates can be used to estimate natural variability with more degrees of freedom than in the OFAT experiment. The central composite design is better than the OFAT experiment because:

- The estimates of the factor effects are more precise.
- The interaction between the factors can be estimated.
- The central composite design estimates curvature in the entire factor space, and allows optimization in the entire factor space. The central composite design allows estimation of curvature in the response as a function of temperature for all times between 9 and 11 seconds (not just at a time of 10 seconds as in the OFAT experiment), and estimation of curvature in the response as a function of time for all temperatures between 980°C and 1020°C (and not just at a temperature of 1000°C as in the OFAT experiment). This means that, for the OFAT, the response can only be "optimized" in temperature for a time of 10 seconds, and it can only be "optimized" in time for a temperature of 1000°C. On the other hand, for the designed experiment, the response can be optimized in the entire factor region, namely for all times between 9 and 11 seconds, and for all temperatures between 980°C and 1020°C.

The advantage of the OFAT is that it requires less resources (six runs) compared to the central composite design (nine different runs, with possibly replicated center points).

Table 5. OFAT experiment in three factors in 15 runs

Exhaust time (sec)	*Resist temperature (°C)*	*Environmental temperature (°C)*
2	23	21
4	23	21
8	23	21
12	23	21
16	23	21
2	21	21
2	22	21
2	23	21
2	24	21
2	25	21
2	23	19
2	23	20
2	23	21
2	23	22
2	23	23

3.3 Three Factors in 15 Runs

An engineer performed an experiment on a new piece of equipment used in a photolithographic process. The objective was to minimize the response, within-wafer standard deviation of resist thickness, as a function of three factors, exhaust "on" time, resist temperature, and environmental temperature. The engineer expected curvature in the response as a function of each factor, and he expected interactions between the factors.

The engineer performed the OFAT experiment for three factors in 15 runs shown in Table 5, and illustrated graphically in Figure 5. Curvature can be estimated along each one of the three lines of circles shown in Figure 5, but interactions cannot be estimated.

The Box–Behnken designed experiment shown in Figure 6 and in Table 6 could have been performed instead. Both the OFAT and the designed experiments have 15 runs, if three center points are used in the Box–Behnken design to make the design rotatable and to provide an estimate of natural variability.

The designed experiment is better than the OFAT experiment because, using the same resources (15 runs):

• The interaction between the factors can be estimated for the designed experiment, but it *cannot* be estimated for the OFAT experiment.

• The experimental runs are more evenly spread out in the factor space for the designed experiment shown in Figure 6 than for the OFAT experiment shown in Figure 5, so the designed experiment gives a better prediction of the response over the entire factor space.

• The OFAT experiment can be used to estimate curvature along the three strings of circles shown in Figure 5, while the designed experiment can be used to estimate curvature in the entire experimental region shown in Figure 6. For example, the first five runs of the OFAT experiment given in Table 5 (vertical string of circles in Figure 5) can be used to estimate curvature in resist thickness standard deviation as a function of exhaust time, at a constant resist temperature of 23°C and at a constant environmental temperature of 21 °C. The designed experiment can be used to estimate curvature in resist thickness standard deviation as a function of exhaust time, for all resist temperatures between 21°C and 25°C, and for all environmental temperatures between 19°C and 23°C.

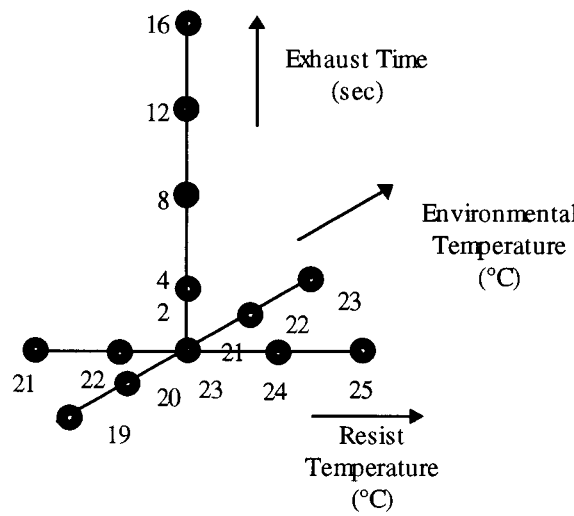

Figure 5. OFAT experiment in three factors in 15 different runs.

• In the OFAT experiment the response can be "optimized" only along the three lines of circles shown in Figure 5, while for the designed experiment the response can be optimized in the entire factor space shown in Figure 6. For example, the first five runs of the OFAT experiment given in Table 5 can be used to find the value of exhaust time that minimizes resist thickness standard deviation, when resist temperature is 23°C and when environmental temperature is 21°C. The designed experiment can be used to find the minimum value of resist thickness standard deviation over the entire region, for all exhaust times between 2 and 16 seconds, all resist temperatures between 21°C and 25°C, and all environmental temperatures between 19°C and 23°C.

Table 6. Box–Behnken designed experiment in three factors in 15 runs

Exhaust time (sec)	Resist temperature (°C)	Environmental temperature (°C)
2	21	21
2	25	21
16	21	21
16	25	21
2	23	19
2	23	23
16	23	19
16	23	23
9	21	19
9	21	23
9	25	19
9	25	23
9	23	21
9	23	21
9	23	21

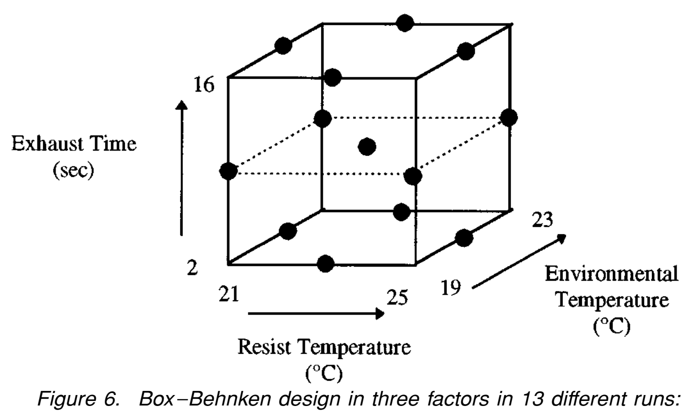

Figure 6. Box–Behnken design in three factors in 13 different runs: 12 runs at edge midpoints, and 1 run at center point.

The optimization will be illustrated using Figure 7, which shows a contour plot of the response resist thickness standard deviation as a function of resist temperature and exhaust time, for a constant environmental temperature of 21°C. The contour lines of resist thickness standard deviation were drawn to be consistent with the results of the OFAT experiment (shown inside the circles), and could have been obtained if the Box–Behnken designed experiment had

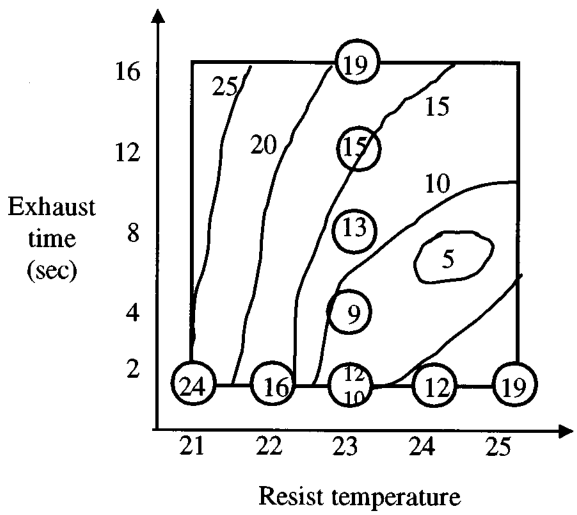

Figure 7. Invented contour plot of the response for the central composite designed experiment. Real response values are shown inside the circles. The response is not minimized along the two strings of circles of the OFAT experiment, but is minimized using the contour lines for the designed experiment.

been performed. The contour plot is a prediction of the response, and is obtained from a model for the response as a function of the three factors. In the contour plot, the minimum value of resist thickness standard deviation is 5 Å when resist temperature is 24.5°C and exhaust time is 8 seconds. This "predicted" resist thickness standard deviation of 5 Å is almost half the minimum value of 9 Å obtained from the OFAT experiment. A smaller "minimum" value was found by searching the entire area inside the square using the designed experiment, than by searching the two strings of circles using the OFAT experiment. The contour plot can be used to study curvature in resist temperature along any horizontal line at a specified exhaust time, and curvature in exhaust time along any vertical line at a specified resist temperature.

If it is important for the factors to be at five levels in the designed experiment as they were in the OFAT experiment, then it is possible to use a three-factor central composite design (the three-dimensional version of Figure 4) instead of the Box–Behnken design.

4. SUMMARY

The advantages of designed experiments over OFAT experiments are illustrated using three real engineering OFAT experiments, and showing how in each case a designed experiment would have been better. This topic is important because many scientists and engineers continue to perform OFAT experiments. The examples can be used in academic and industrial design of experiments classes.

[Received January 1997. Revised March 1998.]

REFERENCES

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, New York: Wiley.

Mason, R. L., Gunst, R. F., and Hess, J. L. (1989), *Statistical Design & Analysis of Experiments*, New York: Wiley.

Montgomery, D. C. (1997), *Design and Analysis of Experiments* (4th ed.), New York: Wiley.