32

EXPERIMENTAL DESIGN FOR PHARMACEUTICAL DEVELOPMENT

Gregory S. Steeno
Pfizer Global Research & Development, Groton, CT, USA

32.1 INTRODUCTION

In the pharmaceutical industry and in today's regulatory environment, process understanding in terms of characterization or optimization is critical in developing and manufacturing new medicines that ensure patient safety and drug efficacy. Experimentation is the key component in building that knowledge base, whether those activities are physical experiments or trials run *in silico*. This understanding comes from relating changes in observed response(s) back to changes, both intended and observational, in independent factors. If the factor changes are deliberate, then the scientist is testing hypotheses about factor effects and qualifying or quantifying their impact. However, what can be often neglected is how important the *quality* of the experimental plan is, and how that connects to making inferences from those data. This chapter focuses on both of these aspects, the experimental design and the data modeling.

For engineers characterizing a chemical reaction, experimental inputs span catalyst load, reaction concentration, jacket temperature, reagent amount, and other factors that are continuous in nature, as well as types of solvents, bases, catalysts, and other factors that are discrete in nature. These are examples of *controllable factors* and are represented as $x_1, x_2, \ldots, x_p$. Additional elements that can influence reaction outputs, such as analysts, instruments, and laboratory humidity, are examples of *uncontrollable factors* and are represented as $z_1, z_2, \ldots, z_q$. Figure 32.1, as shown by Montgomery [1], depicts a general process where both factors types impact the output.

The set of trials to develop relationships between factors and responses plus the structure of how the trials are executed, comprise the *experimental design*. What follows are strategies for sound statistical design and analysis.

As a motivating example, consider a process where the output is yield (g) and is expected to be a function of two controllable factors, reaction time (min) and reaction temperature (°C). That is, $\text{yield} = f(\text{time}, \text{temperature})$. As a first step in understanding and optimizing this process, experiments were executed by fixing time at 30 min and observing yield across a range of reaction temperatures. Figure 32.2 shows the data and it is concluded that 35°C is a reasonable choice as an optimal temperature. For the second step, experiments were executed by now fixing temperature at 35°C and observing yield across a range of reaction times. Figure 32.3 shows the data and it is concluded that the optimal time is about 40 min. The combined information from the totality of experiments produces an optimal setting of (temperature, time) = (35°C, 40 min) with a predicted yield around 70 g.

This experiment was conducted using a *one factor at a time* (OFAAT) approach—while easy to implement and instinctively sensible, there are a few shortcomings when compared to a statistically designed experiment. In general, OFAAT studies (1) are not as precise in estimating individual factor effects, (2) cannot estimate multivariate factor effects, such as linear × linear interactions, and (3) as a by-product are not as efficient at locating an optimum. Figure 32.4 illustrates the experimental path (• · · · •), the chosen optimum ("*XX*"), and the actual underlying relationship between time

Chemical Engineering in the Pharmaceutical Industry: R&D to Manufacturing Edited by David J. am Ende

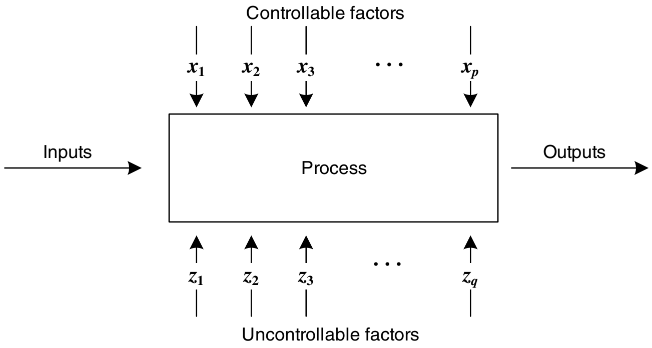

FIGURE 32.1 General process diagram.

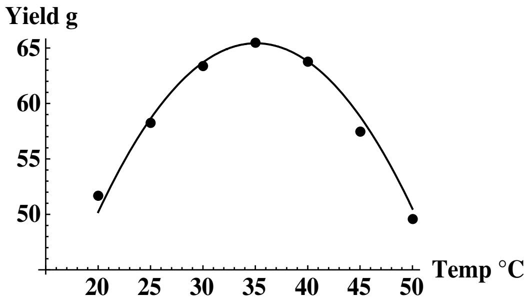

FIGURE 32.2 Scatterplot of yield against temperature with empirical fit.

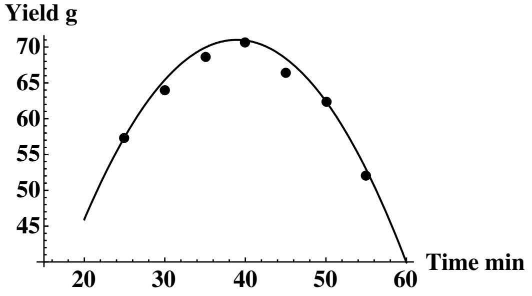

FIGURE 32.3 Scatterplot of yield versus time with empirical fit.

and temperature on yield. Since the joint relationship between time and temperature was not fully investigated, the true optimum around (temperature, time) ≈ (60°C, 60 min) was missed.

This example illustrates that even with only two variables, the underlying mechanistic relationship between the factors and response(s) can be complex enough to easily misjudge. This is especially true of many pharmaceutical processes, where functional relationships are dynamic and nonlinear. Because of this complexity, experiments that produce the most complete information in the least amount of resource (time, material, and cost) are vital. This is a major aspect of *statistical experimental design*, which is an efficient method for evaluating process inputs in a systematic and multivariate way. The integration of experimental design and model building is generally known as *response surface methodology*, first introduced by Box and Wilson [2]. Data resulting from such a structured experimental plan coupled with regression analysis easily lend to deeper understanding of where process sensitivities exist, as well as how to improve process performance in terms of speed, quality, or optimality.

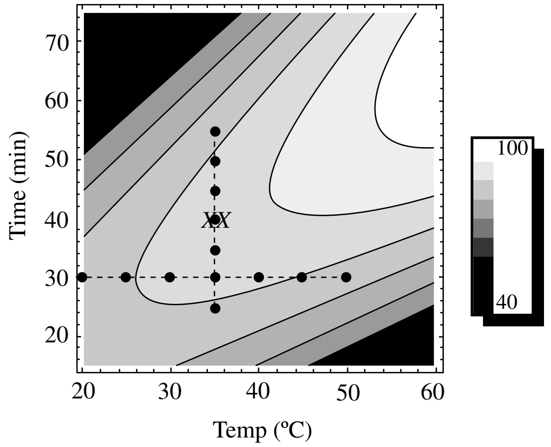

FIGURE 32.4 Contour plot of underlying mechanistic relationship of temperature and time on yield, along with experimental path (● · · · ●) and chosen optimum (*XX*).

The experimental design and analysis procedure is straightforward and intuitive, but is described below for completeness to ensure that all information is collected and effectively used.

1. Formulate a research plan with purpose and scope
2. Brainstorm explanatory factors denoted $X_1, X_2, X_3, \ldots,$ that could impact the response(s). Discuss ranges and omit factors that have little scientific value.
3. Determine the responses to be measured, denoted Y_1, Y_2, Y_3, ..., and consider resource implications.
4. Select appropriate experimental design in conjunction with purpose and scope. Consider the randomization sequence. Hypothesize process models.
5. Execute the experiment. Measurement systems should be accurate and precise. The randomization sequence is key to balancing effects of random influences and propagation of error.
6. Appropriately analyze the data
7. Draw inference and formulate next steps

The statistical experimental designs discussed in this chapter are used to help estimate approximating response functions for chemical process modeling. To begin, the functional relationship between the response, y, and input variables, $\xi_1, \xi_2, \ldots, \xi_k$, is expressed as

$$y = f(\xi_1, \xi_2, \ldots, \xi_k) + \varepsilon$$

which is unknown and potentially intricate. The inputs, $\xi_1, \xi_2, \ldots, \xi_k$, are called the natural variables, as they represent the actual values and units of each input factor. The ε term represents the variability not explicitly accounted for in the model, which could include the analytical component, the laboratory environment, and other natural sources of noise. For mathematical convenience, the natural variables are centered and scaled so that coded variables, $x_1, x_2, \ldots, x_k$ have mean zero and standard deviation one. This does not change the response function, but it is now expressed as

$$y = f(x_1, x_2, \ldots, x_k) + \varepsilon$$

If the experimental region is small enough, $f(\cdot)$ can be empirically estimated by lower order polynomials. The motivation comes from Taylor's theorem that asserts any sufficiently smooth function can locally be approximated by polynomials. In particular, first-order and second-order polynomials are heavily utilized in response modeling from designed experiments.

A first-order polynomial is referred to as a *main effects model*, due to containing only the primary factors in the model. A two-factor main effects model is expressed as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where β_1 and β_2 are coefficients for each factor and β_0 is the overall intercept, and represents a plane through the (x_1, x_2) space. As an example consider an estimated model

$$\hat{y} = 100 - 10x_1 + 5x_2$$

Figure 32.5 shows a 3D view of that planar response function, also called a surface plot. Figure 32.6 represents the 2D analogue called a contour plot. The contour plots are often easier to read and interpret since the response function height is projected down onto the (x_1, x_2) space. If there is an interaction between the factors, it is easily added to the model as follows:

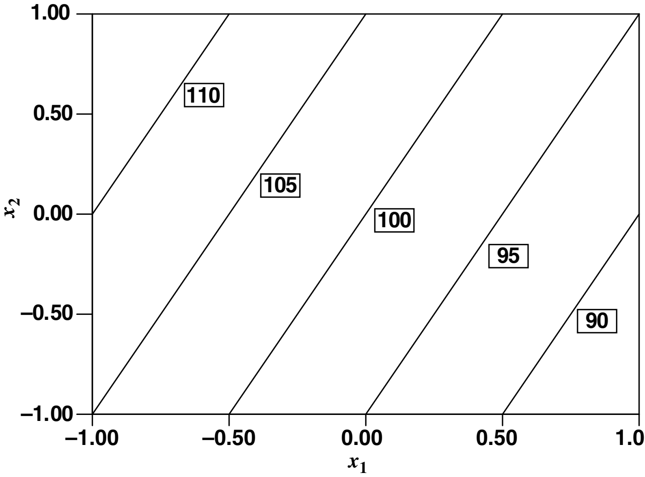

FIGURE 32.6 Contour plot of $\hat{y} = 100 - 10x_1 + 5x_2$.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

This is called a first-order model with interaction. To continue with the example, let the estimated model be

$$\hat{y} = 100 - 10x_1 + 5x_2 - 5x_1 x_2$$

The additional term $-5x_1x_2$ introduces curvature in the response function, which is displayed on the surface plot in Figure 32.7 and the corresponding contour plot in Figure 32.8. Occasionally, the curvature in the true underlying response function is strong enough that a first-order plus interaction model is inadequate for prediction. In this case, a second-order (quadratic) model would be useful to approximate $f(\cdot)$ and takes the form

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

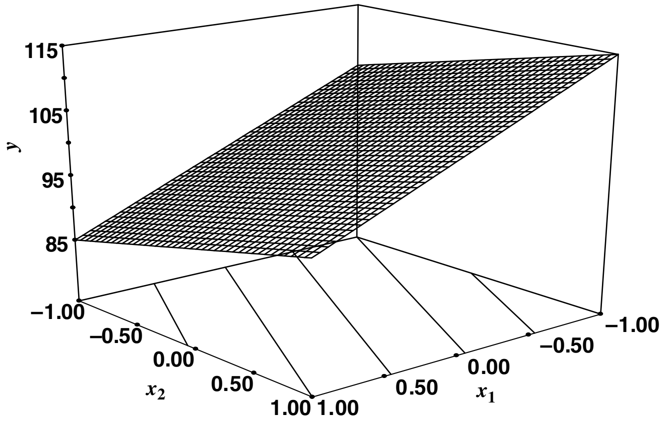

FIGURE 32.5 Surface plot of $\hat{y} = 100 - 10x_1 + 5x_2$.

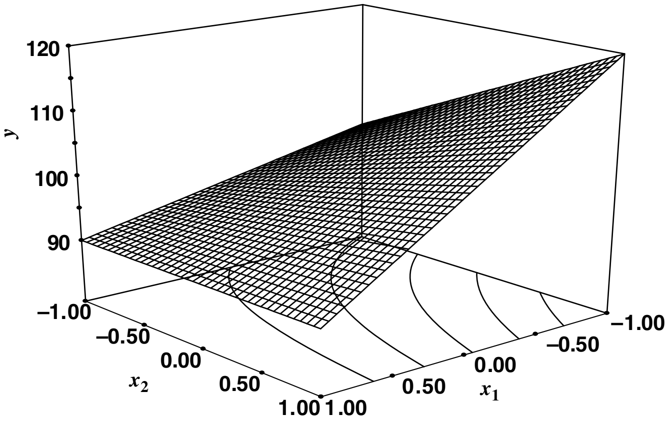

FIGURE 32.7 Surface plot of $\hat{y} = 100 - 10x_1 + 5x_2 - 5x_1x_2$.

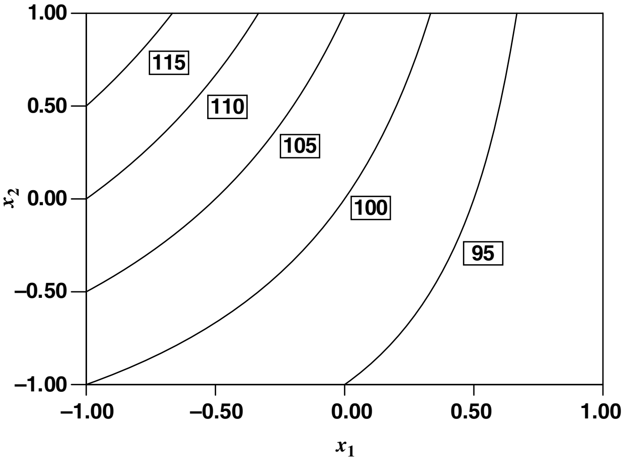

FIGURE 32.8 Contour plot of $\hat{y} = 100-10x_1+5x_2-5x_1x_2$.

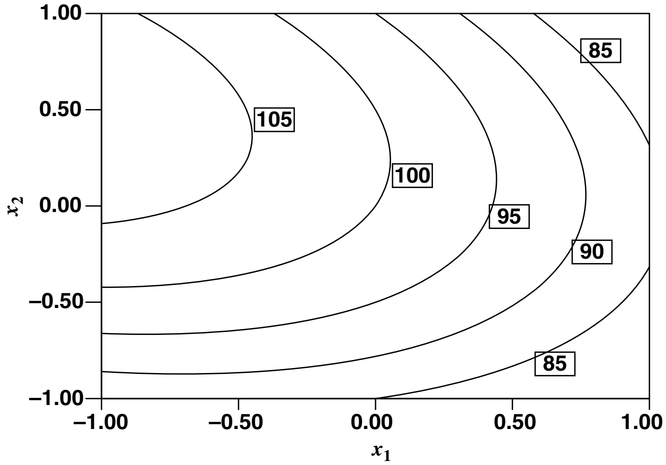

FIGURE 32.10 Contour plot of $\hat{y} = 100-10x_1+5x_2-5x_1x_2-4x_1^2-10x_2^2$.

To finish the example, let the estimated model be

$$\hat{y} = 100 - 10x_1 + 5x_2 - 5x_1x_2 - 4x_1^2 - 10x_2^2$$

Figure 32.9 shows a parabolic relationship between y and x_1, x_2, while Figure 32.10 displays the typical elliptical contours generated by this model.

There is an iterative, sequential nature to understanding and optimizing the performance of chemical processes. If the goal is to first identify the most important factors for further study, a screening design may be carried out. This is sometimes referred to as *phase zero* of the study. Once this is complete, the next objective is to determine if the optimum lies within current experimental region, or if the factors need adjustment to locate a more desirable one, say by using methods of steepest ascent/descent. This is referred to as *phase one* of the study, also known as *region seeking*. Finally, once the region of desirable response is established, the goal becomes to precisely model that area and identify optimal factor settings. For this case, higher order models are employed to capture likely curvature about the optimum point.

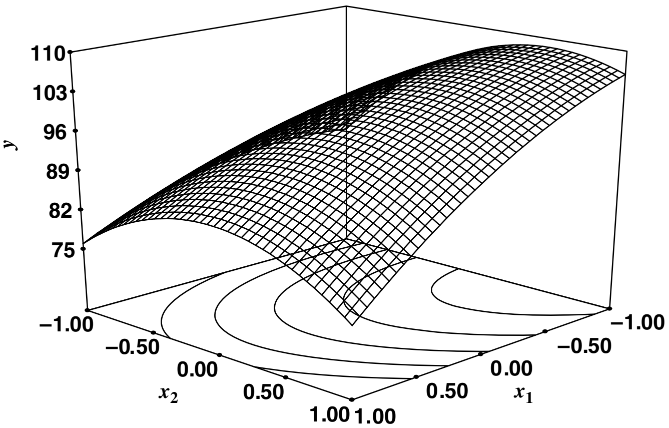

FIGURE 32.9 Surface plot of $\hat{y} = 100-10x_1+5x_2-5x_1x_2-4x_1^2-10x_2^2$.

32.2 THE TWO-LEVEL FACTORIAL DESIGN

Factorial designs are experimental plans that consist of all possible combinations of factor settings. As an example, a factorial design with three different catalysts, two different solvents, and four different temperatures produces a design with $3 \times 2 \times 4 = 24$ unique experimental conditions. The advantage of these designs is that all joint effects of factors can be investigated. The disadvantage is that these designs become prohibitively large and impractical when factors contain more than just a few levels or the number of factors under investigation is extensive.

The simplest and most widely used factorial designs for industrial experiments are those that contain two levels per factor, called 2^k factorial designs, where k is the number of factors under investigation. The two levels for each factor are usually chosen to span a practical range to investigate. These designs can be augmented into fuller designs and are very effective in terms of time, resource, and interpretability. The class of 2^k factorial designs can be used as building blocks in process modeling by:

- Screening the most important variables from a set of many;
- Fitting a first-order equation used for steepest ascent/descent;
- Identifying synergistic/antagonistic multifactor effects; and
- Forming a base for an optimization design, such as a central composite (to be introduced).

TABLE 32.1 Example 2^2 Experimental Design with $n = 2$ Replicates Per Design Point

	Design		Temperature × Time (AB)	Response	
Treatment	Temperature (A)	Time (B)		Rep 1	Rep 2
(1)	−1	−1	1	y_{11}	y_{12}
a	1	−1	−1	y_{21}	y_{22}
b	−1	1	−1	y_{31}	y_{32}
ab	1	1	1	y_{41}	y_{42}

Factors can either be continuous in nature or discrete. For the rest of the chapter it is assumed that the factors are continuous. This allows for predictive model building and regression analysis that includes the linear and interaction terms, and subsequently the quadratic/second-order terms.

To illustrate a two-level factorial design, consider the previous case where there are $k = 2$ factors, temperature (Factor A) and time (Factor B), each having an initial range under investigation. In coded units, the low level of the range for each factor is scaled to -1, and the high level of the range is scaled to $+1$. The 2^2 experimental design with all four treatment combinations is shown in Table 32.1 and graphically depicted via circles in Figure 32.11. Notice that each treatment condition occurs at a vertex of the experimental space. For notation, the four treatment combinations are usually represented by lowercase letters. Specifically, a represents the combination of factor levels with A at the high level and B at the low level, b represents A at the low level and B at the high level and ab represents both factors being run at the high level. By convention, (1) is used to denote A and B each run at the low level.

Two-level designs are used to estimate two types of effects, main effects and interaction effects, and these are estimated by a single degree of freedom *contrast* that partitions the design points into two groups: the low level (-1) and the high level ($+1$). The contrast coefficients are shown in Table 32.1 for each of the main factors of temperature and time, as well as the temperature × time interaction obtained through pairwise multiplication of the main factor contrast coefficients.

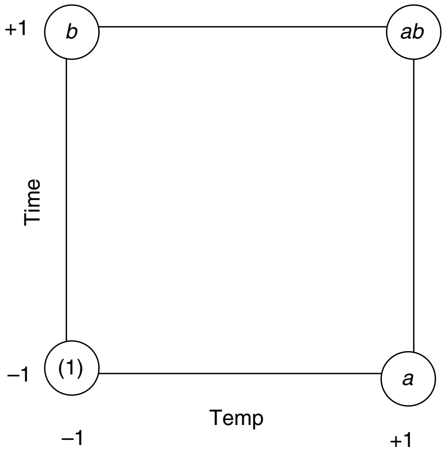

FIGURE 32.11 Factor space of 2^2 experimental design.

A main effect of a factor is defined as the average change in response over the range of that factor and is calculated from the average difference between data collected at the high level ($+1$) and data collected at the low level (-1). For the 2^2 design above and using the contrast coefficients in Table 32.1, the temperature (A) main effect is estimated as

$$A=\bar{y}_{\text{Temp}+}-\bar{y}_{\text{Temp}-}=\frac{ab+a}{2n}-\frac{b+(1)}{2n}=\frac{1}{2n}[ab+a-b-(1)]$$

where (1), a, b, and ab are the respective sum total of responses across the n replicates at each design point ($n = 2$ in Table 32.1). Geometrically this is a comparison of data on average from the right side to the left side of the experimental space in Figure 32.11. If the estimated effect is positive, the interpretation is that average response increases as the factor level increases. Similarly, the time (B) main effect is estimated as

$$B=\bar{y}_{\text{Time}+}-\bar{y}_{\text{Time}-}=\frac{ab+b}{2n}-\frac{a+(1)}{2n}=\frac{1}{2n}[ab+b-a-(1)]$$

which is a comparison of data on average from the top side to the bottom side in Figure 32.11.

An interaction between factors implies that the individual factor effects are not additive and that the effect of one factor depends on the level of another factor(s). As with the main effects, the interaction is estimated by partitioning the data into two groups and comparing the average difference. The contrast coefficients in Table 32.1 show that the temperature × time (AB) interaction effect is estimated as

$$AB=\bar{y}_{\text{Temp}\times\text{Time}+}-\bar{y}_{\text{Temp}\times\text{Time}-}=\frac{ab+(1)}{2n}-\frac{a+b}{2n}$$
$$=\frac{1}{2n}[ab+(1)-a-b]$$

which is a comparison on average of data on the right diagonal against the left diagonal in Figure 32.11.

The sum of squares for each effect are mathematically related to their corresponding contrast. Specifically, the sum of squares for an effect is calculated by the squared contrast divided by the total number of observations in that contrast. For the example above, the sums of squares for temperature, time, and the temperature × time interaction are

$$SS_{\text{Temp}}=\frac{[a+ab-b-(1)]^2}{4n}$$

$$SS_{\text{Time}}=\frac{[b+ab-a-(1)]^2}{4n}$$

TABLE 32.2 ANOVA Table for Completely Randomized 2^2 Design with n Replicates Per Design Point

Source	SS	DF	MS	F	p-value
Temp	SS_{Temp}	1	MS_{Temp}	MS_{Temp}/MS_E	p_{Temp}
Time	SS_{Time}	1	MS_{Time}	MS_{Time}/MS_E	p_{Time}
Temp × Time	$SS_{Temp\times Time}$	1	$MS_{Temp\times Time}$	$MS_{Temp\times Time}/MS_E$	$p_{Temp\times Time}$
Error	SS_E	$4(n-1)$	MS_E		
Total	SS_T	$4n-1$			

$$SS_{Temp\times Time} = \frac{[ab+(1)-a-b]^2}{4n}$$

In 2^k designs, the contrasts are orthogonal, thus additive. The total sum of squares, SS_T, is the usual sum of squared deviations of each observation from the overall mean of the data set. Because the contrasts are orthogonal, the error sum of squares, denoted SS_E, can be calculated as the difference between the total sum of squares, SS_T, and the sums of squares of all effects. For the 2^2 example, $SS_E = SS_T - SS_{Temp} - SS_{Time} - SS_{Temp\times Time}$. With this information, the analysis of variance (ANOVA) table is constructed, as shown in Table 32.2.

The ANOVA table contains all numerical information in determining which factor effects are important in modeling the response. The hypothesis test on individual factor effects is conducted through the F-ratio of MS_{Factor} against MS_{Error}. If this ratio of "signal" against "noise" is large, this implies that the factor explains some of the observed variation in response across the experimental design region and should be included in the process model. If the ratio is not large, then the inference is that the factor is unimportant and should be deleted from the model. All statistical evidence of model inclusion comes via the p-value. Large F-ratios imply low p-values, and a common cutoff for model inclusion of a factor is $p \leq 0.05$, although this should be appropriately tailored with experimental objectives, such as factor screening where the critical p-value is normally a little higher. Another interpretation of the p-value is in terms of the confidence level, equal to $(1-p) \times 100\%$. Thus, a factor with a p-value less than 0.05 implies there is greater than 95% confidence that the observed factor effect is real and not due to noise.

Clearly it is important to identify all significant factors for modeling change in response back to change in factor level. An *underspecified* model, one that does not contain all the important variables, could lead to bias in regression coefficients and bias in prediction. One approach to mitigate this issue is to not model edit but rather incorporate all factor terms in the process model, including those that contribute very little or nothing of value in predicting. However, an *overspecified* model, one that contains insignificant terms, produces results that lead to higher variances in coefficients and in prediction. Thus, a proper model will be a compromise of the two. This can be completed manually, say, by investigating the full model ANOVA and then deleting insignificant effects one at a time, all while updating the ANOVA after every step. Yet for models that could contain many effects, this exercise becomes cumbersome. There are several variable selection procedures that can aid in helping identify smaller sized candidate models. The most common algorithms used in standard software packages entail either sequentially bringing in significant factors to build the model up (called forward selection), sequentially eliminating regressors from a full model (called backward elimination), or a hybrid of the two (called stepwise regression). The procedures typically involve defining a critical p-value for factor inclusion/exclusion in the model building process. Once a term enters or leaves, factor significance is recalculated and the process is repeated for the next step. The engineer should use these tools not as a panacea to the model building process, but rather as an exercise to see how various models perform. For more information on the process and issues of model selection, the reader is instructed to see Myers [3].

Once the significant factors are selected, the estimated regression coefficients in the linear predictive model are functionally derived from the factor's effect size. To estimate the regression coefficient β_i for factor i, its effect is divided by two. The rationale being that by definition the regression coefficient represents the change in y per unit change in x. Since each factor effect is calculated as a change in response over a span two coded units (-1 to 1), division by 2 is needed to obtain the per-unit basis. Finally, the model intercept, β_0, is calculated as the grand average of all the data.

EXAMPLE 32.1 2^3 FACTORIAL DESIGN

A factorial experiment is carried out to investigate the effect of three factors on percent reaction conversion, catalyst load (Factor A), ligand load (Factor B), and temperature (Factor C). Each experimental condition is completely randomized and independently replicated ($n = 2$). The design in coded units, full model, and data are listed in Table 32.3, and depicted in Figure 32.12. Note that the run sequence in Table 32.3 is in *standard order*, as opposed to a randomized order for the actual experiment.

As previously remarked, the estimated effects and associated sums of squares are functions of their respective

TABLE 32.3 2^3 Reaction Conversion Experimental Design, Model, and Data

Treatment	Design			*AB*	*AC*	*BC*	*ABC*	Data	
								Conversion	
	Catalyst *A*	Ligand *B*	Temperature *C*					Rep 1	Rep 2
(1)	−1	−1	−1	1	1	1	−1	63.8	64.1
a	1	−1	−1	−1	−1	1	1	76.4	74.9
b	−1	1	−1	−1	1	−1	1	66.5	63.7
ab	1	1	−1	1	−1	−1	−1	76	76.3
c	−1	−1	1	1	−1	−1	1	78.7	77.7
ac	1	−1	1	−1	1	−1	−1	77.6	80.4
bc	−1	1	1	−1	−1	1	−1	78	81.3
abc	1	1	1	1	1	1	1	79	77.2

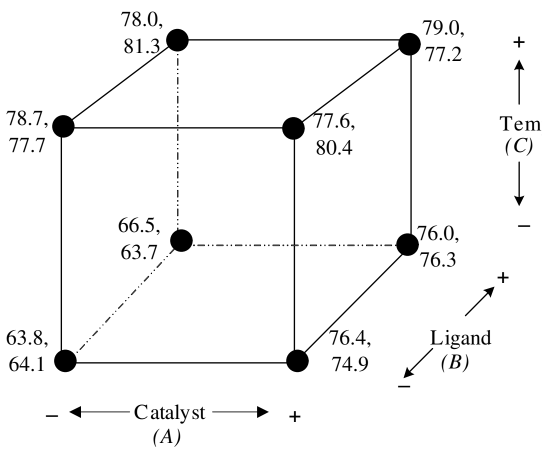

FIGURE 32.12 2^3 reaction conversion experimental space and corresponding data.

contrasts. Using catalyst load (*A*) as an example, the contrast coefficients shown in Table 32.3 represent the comparison between data on the right side of the cube (+) in Figure 32.12 to the data on the left side of the cube (−). The data where catalyst load is high sum to

$$\begin{aligned} a+ab+ac+abc &= 76.4+74.9+76.3+76.0+77.6 \\ &\quad +80.4+79.0+77.2 = 617.8 \end{aligned}$$

while the data where catalyst load is low sum to

$$\begin{aligned} b+c+bc+(1) &= 66.5+63.7+78.7+77.7+78.0 \\ &\quad +81.3+63.8+64.1 = 573.8 \end{aligned}$$

The estimated effect of catalyst load is calculated as

$$\begin{aligned} A = \bar{y}_{\mathrm{Cat}+} - \bar{y}_{\mathrm{Cat}-} &= \frac{a+ab+ac+abc}{4n} - \frac{b+c+bc+(1)}{4n} \\ &= \frac{617.8}{4\times 2} - \frac{573.8}{4\times 2} = 5.5 \end{aligned}$$

TABLE 32.4 Reaction Conversion Experiment ANOVA Table with Full Model

Source	SS	DF	MS	*F*	*p*-value
A (catalyst)	121.00	1	121.00	58.24	<0.0001
B (ligand)	1.21	1	1.21	0.58	0.4673
C (temperature)	290.70	1	290.70	139.93	<0.0001
AB	2.25	1	2.25	1.08	0.3284
AC	138.06	1	138.06	66.46	<0.0001
BC	0.30	1	0.30	0.15	0.7127
ABC	0.72	1	0.72	0.35	0.5717
Error	16.62	8	2.08		
Total	570.87	15			

The interpretation is that the average yield increases by 5.5% as the catalyst load increases from low to high. The corresponding sum of squares for the catalyst load effect is calculated as

$$\begin{aligned} \mathrm{SS}_A &= \frac{[a+ab+ac+abc-b-c-bc-(1)]^2}{8n} \\ &= \frac{(617.8-573.8)^2}{8\times 2} = 121 \end{aligned}$$

The other estimated factor effects and sums of squares follow the same logic as above and are trivial to calculate. The full model ANOVA table is shown in Table 32.4. Based on the information, there are three highly significant effects ($p < 0.0001$): catalyst load, temperature, and their corresponding interaction. All other effects are insignificant at the 95% confidence level ($p > 0.05$). The final model and ANOVA table after sequential model editing are shown in Table 32.5.

This model accounts for ~96.3% of the observed variability in reaction completion, as determined by the coefficient of determination, R^2.

TABLE 32.5 Reaction Conversion Experiment ANOVA Table After Model Editing

Source	SS	DF	MS	F	p-value
A (catalyst)	121.00	1	121.00	68.80	<0.0001
C (temperature)	290.70	1	290.70	165.29	<0.0001
AC	138.06	1	138.06	78.50	<0.0001
Error	21.10	12	1.76		
Total	570.87	15			

$$R^2 = \frac{\mathrm{SS}_{\mathrm{Model}}}{\mathrm{SS}_{\mathrm{Total}}} = \frac{\mathrm{SS}_A + \mathrm{SS}_C + \mathrm{SS}_{AC}}{\mathrm{SS}_{\mathrm{Total}}} = \frac{549.76}{570.87} = 0.9630$$

The final regression model expressed in coded units is estimated as

$$\hat{y} = 74.48 + \left(\frac{5.5}{2}\right)x_1 + \left(\frac{8.26}{2}\right)x_3 + \left(\frac{-5.87}{2}\right)x_1x_3$$

$$\Rightarrow \hat{y} = 74.48 + 2.75x_1 + 4.26x_3 - 2.94x_1x_3$$

where x_1 and x_3 represent catalyst load and temperature, respectively.

Because the factors are all centered and scaled and all the effects are orthogonal, one can compare which effects are the most dominant by the size of the coefficient. Using the reaction conversion model, temperature has the largest effect, followed by the catalyst × temperature interaction and then catalyst. However, because of the interaction between catalyst and temperature, the main effects of those individual factors have lost some interpretability. Specifically, the conclusion from the temperature effect is that for every unit change in temperature, the conversion increases 4.26% via the coefficient on the temperature term. But this estimate is a pooled average over the other factors. That is, it smoothes over the significant joint effect between temperature and catalyst. The information contained in the interaction will need to be visually explored in greater detail.

After obtaining the final equation, residual analysis and other model diagnostics are carried out. This is critical step in validating the process model and having trust in its ability to accurately predict over the experimental region. Standard residual analyses consist of inspecting the normality assumption, checking for constant variance, identifying outliers relative to the model, and observing patterns in residuals over time. There are many flavors of model diagnostic information, both numerical and graphical. With the aid of Design Expert [4] software package, highlighted below are two visuals that, in this author's view, capture a significant snapshot of model performance. First, the normal probability plot is an effective graphical tool to verify the normality assumption of the errors as well as for outlier detection. Figure 32.13 shows this plot for the reaction conversion residuals using the final regression model. If the residuals fall along a straight line, then the normality assumption is valid. Any large errors or outliers from the fitted model would be visually apparent by significantly falling off the line. The interpretation of Figure 32.13 is that there is no severe problem with the normality assumption.

Second, another very effective plot is the model-based predictions against the observed data, often referred to as "Predicted versus Actual." This reveals how well the model predicts back the original data and is a graphical depiction of the calculated R^2 value. Additionally, it will aid in identifying (sets of) data that are not well captured by the model, as well as indicate any trends of nonconstant variance across the prediction range. Figure 32.14 is an example using the reaction conversion data and final model. The interpretation from this graph is that the regression model is performing well and the spread of about the 45° line is relatively constant across the range.

Once the engineer is satisfied with the diagnostics information, visualizing the change in response across the

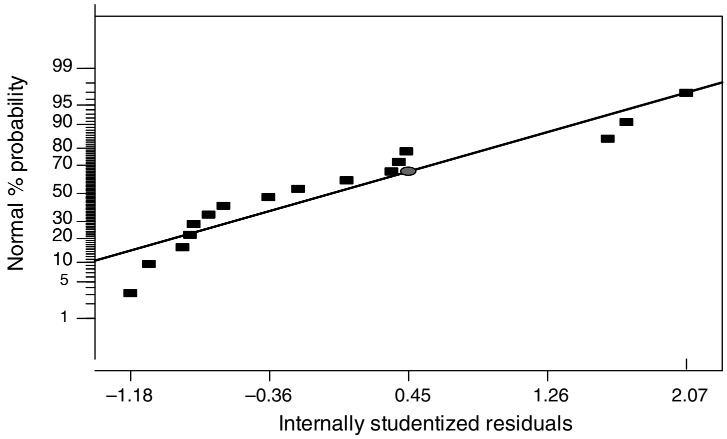

FIGURE 32.13 Normal probability plot of residuals from reaction conversion model.

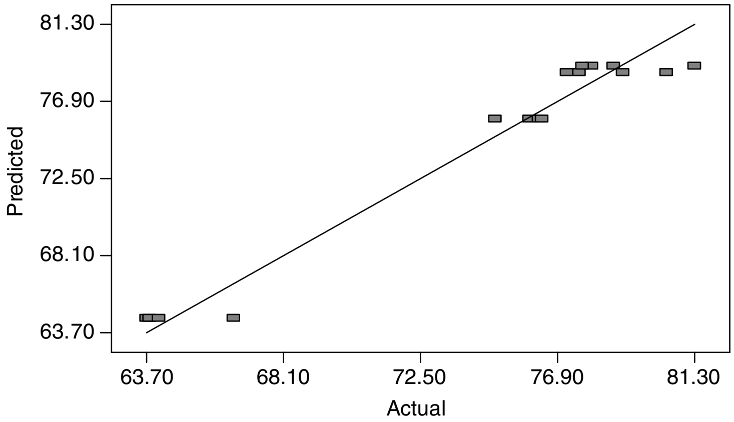

FIGURE 32.14 Predicted versus actual plot using reaction conversion example.

subspace of significant factors is the next step. This is usually accomplished by main effect and interaction plots. If factors are continuous, then contour and surface plots are very informative and more descriptive in illustrating bivariate factor effects on response. As previously shown with the reaction conversion example, the interaction between catalyst and temperature is significant and therefore the synergistic relationship between these two factors needs further inspection. Figure 32.15 displays an interaction plot for these two factors. Notice that when temperature is at the high (+ 1) level, there is no observed effect of catalyst load on the reaction conversion. That is, it is robust to changes in catalyst. However, when temperature is at the low (−1) level, reaction conversion is now a function of catalyst load, and higher load leads to higher predicted conversion. Figure 32.16 shows the corresponding contour plot that also illustrates the predicted change in conversion across the temperature and catalyst levels, but takes advantage of the continuous nature of the factors. Regardless, the inference is the same: At high temperatures, the prediction is virtually constant across catalyst (~78% conversion), while at lower temperatures, the catalyst effect is present.

Two-level designs are very intuitive, comprehensive, and powerful in identifying main and interaction effects on responses of interest. However, even at two levels they can become impractically large as the number of factors increases. For example, if $k = 6$ factors are under investigation, then a full factorial experiment with no replication would consist of $2^6 = 64$ runs, which is not a cost-effective experiment. If this design were executed, the full regression model with all possible main effects and interactions would leave zero degrees of freedom to estimate variability for statistical inferences on factor effects. There are strategies and tools that combat the issues of unreplicated designs and variance estimation. One such strategy takes advantage of the *sparsity-of-effects principle* (or similarly, the Pareto principle) that states that a process is usually dominated by only the vital few effects, such as main effects and two-factor interactions, from the trivially many. More specifically, observing effects from higher order terms such

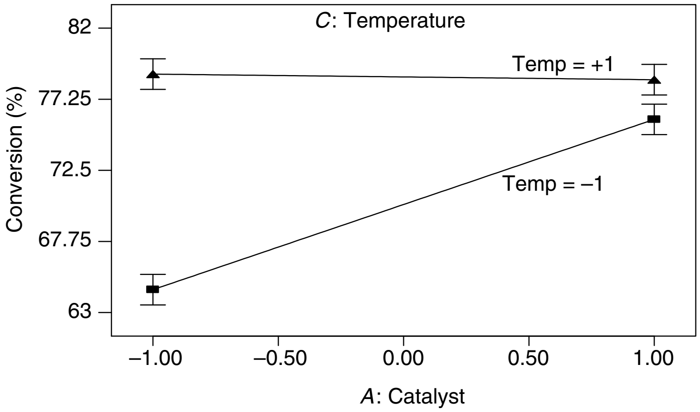

FIGURE 32.15 Interaction plot of catalyst and temperature on reaction conversion.

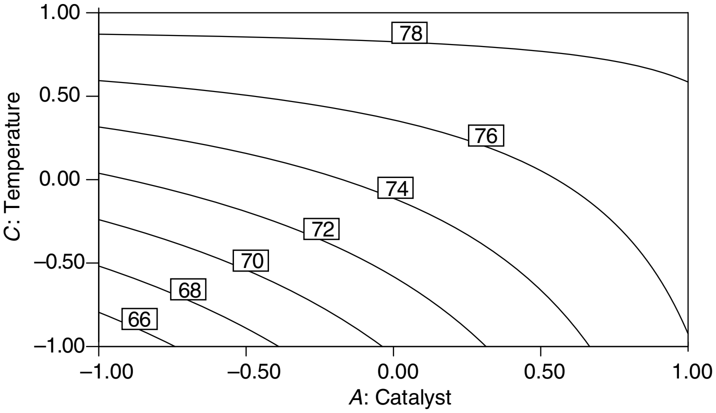

FIGURE 32.16 Contour plot of predicted reaction conversion across catalyst and temperature levels.

as three-factor interactions and beyond is rare in practice. If it is reasonable to assume that effects from higher order interactions are negligible, then those terms can be pooled to estimate variability through the mean squared error. Another approach is to use a normal probability plot or half-normal probability plot to determine the significant effects. Proposed by Daniel [5], these are effective graphical tools that use the estimated effects to highlight which factors are important in explaining response variation relative to those that do not. Negligible effects are assumed to be normally distributed with mean zero and variance σ^2, while significant effects are assumed normally distributed at their true effect size different from zero with variance σ^2. Normal and half-normal effect plots are standard in most statistical software packages.

32.3 BLOCKING

There are situations that call for the design to be run in groups or clusters of experiments. This often occurs when the number of factors is large, in which case the design would be broken down into smaller *blocks* of more homogeneous experimental units. These are referred to as incomplete blocks, as not all treatment combinations occur in these smaller sets. This situation also occurs with equipment set up or limitations, where groups of experiments are executed at the same time. As one example, consider a replicated 2^2 design that investigates temperature and solvent concentration on reaction completion. The equipment chosen for the experiment is a conventional process chemistry workstation that has four independent reactors. A natural and appropriate way to execute this design is to run each replicate of the 2^2 design on the workstation with random assignment of the reactor vessels to each treatment combination. By doing this, any block-to-block variability is accounted for and does not influence the analysis on factor effects. As another example, consider a 2^3 design using the same four-reactor workstation that investigates temperature, solvent concentration, and catalyst loading on reaction completion. To execute this study, the design should be split into two groups of four, since it is not possible to execute all experiments in one block and impractical to execute the experiments one at a time.

Both of these examples result in observations within the same block to be more homogeneous than those in another block. For situations where the experiment is subdivided, appropriate blocking can help in optimally constructing a design based on the assumption or knowledge that certain (higher order) interactions are negligible. This design technique is called *confounding* or *aliasing*, where information on treatment effects is indistinguishable from information on block effects. The number of blocks for the two-level design is usually a multiple of two, implying designs are run in blocks of two, four, eight, and so on.

To illustrate, recall the previous 2^3 design with temperature (A), solvent concentration (B), and catalyst loading (C) as the factors, with the design executed in two blocks of four on the workstation. The design and full model with associated contrast coefficients are shown in Table 32.6.

One candidate design partitioned into two blocks of four is to group all combinations where ABC is at the low level (-1) into one block, and all combinations where ABC is at the high

TABLE 32.6 2^3 Design Full Model Contrast Coefficients

Treatment	*A*	*B*	*C*	*AB*	*AC*	*BC*	*ABC*
(1)	−1	−1	−1	1	1	1	−1
a	1	−1	−1	−1	−1	1	1
b	−1	1	−1	−1	1	−1	1
ab	1	1	−1	1	−1	−1	−1
c	−1	−1	1	1	−1	−1	1
ac	1	−1	1	−1	1	−1	−1
bc	−1	1	1	−1	−1	1	−1
abc	1	1	1	1	1	1	1

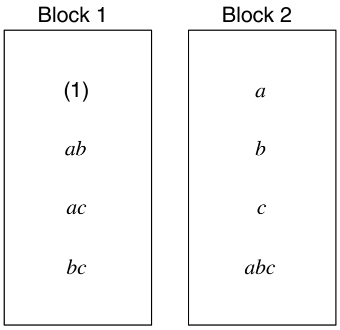

FIGURE 32.17 Schematic of 2^3 factorial divided into two blocks of size four.

level $(+1)$ into the other block. Schematically, each block would appear as in Figure 32.17.

The contrast to estimate the effect of A (temperature) assuming n=1 replicate is written as

$$A = \tfrac{1}{4}[a + ab + ac + abc - (1) - b - c - bc]$$

Any block effect between experiments conducted in Block 1 and those conducted in Block 2 is canceled out in this contrast, as the overall effect of A is actually a pooled sum of the *simple within-block effects* of A. That is, let A_1 be the comparison of A at the high level versus A at the low level within Block 1, and A_2 be the corresponding comparison within Block 2. The overall effect of A is calculated as

$$\begin{aligned} A_1 &= [ab + ac - (1) - bc] \\ A_2 &= [a + abc - b - c] \\ \Rightarrow A &= \tfrac{1}{4}[A_1 + A_2] \end{aligned}$$

This holds for all other effects except the ABC effect, where its corresponding contrast from Table 32.6 is

$$ABC = \tfrac{1}{4}[a + b + c + abc - (1) - ab - ac - bc]$$

By design, the difference in those treatment combinations corresponds exactly to the design partition into Blocks 1 and 2. That is, the effect of ABC is not estimable; it is confounded with blocks. Assuming the ABC effect is negligible, this is an optimally constructed design, as the ABC effect was intentionally confounded to preserve inferences on lower order effects in the presence of any block-to-block variation.

The last example showed that for a design partitioned into two blocks, one effect (ABC) is chosen to be confounded with the block effect. Another way of stating this is that the block effect and the ABC effect *share a degree of freedom* in the analysis. For the case of four blocks that use three degrees of freedom in the analysis, the general procedure is to independently select two effects to be confounded with blocks, and then a third confounded effect is determined by the generalized interaction. This will be described in a more detail with fractional factorial designs. For more information on constructing blocks in 2^k designs, see Ref. 6.

32.4 FRACTIONAL FACTORIALS

The unreplicated 2^6 design contains 64 unique treatment combinations of the 6 factors and therefore 63 degrees of freedom for effects. Of those 63 degrees of freedom, only 6 are used for main effects (Factors $A, B, \ldots, F$) and 15 for two-factor interactions ($AB, AC, \ldots, EF$). Assuming the sparsity-of-effects principle holds, only a subset of the total degrees of freedom are used for the vital few effects that should adequately model the process. This discrepancy gets bigger as k gets larger, making full factorial designs an inefficient choice for experimentation. As with blocking, it is possible to optimally construct designs based on the assumption or knowledge that higher order interactions are negligible, which are smaller in size yet preserve critical information about likely effects of interest. These are called *fractional factorial* designs and they are widely utilized for any study involving, say, five or more factors. In particular, these are highly effective plans for *factor screening*, the exercise to whittle down to only the crucial process factors to be subsequently studied in greater detail. As with the full factorial designs, the fractional factorials are balanced and the estimated effects are orthogonal. Two-level fractional factorial designs are denoted 2^{k-p}, where k still represents the number factors in the study, and p represents fraction level. A 2^{k-1} design is called a one-half fraction of the 2^k, a 2^{k-2} design is called a one-quarter fraction of the 2^k, and so on. A 2^{k-p} design is a study in k factors, but executed with 2^{k-p} unique treatment combinations.

Consider the 2^3 design, but due to limited resources only four of the eight treatment combinations can be studied. The candidate design is a one-half fraction of a 2^3 factorial, denoted 2^{3-1}. The primary questions in design construction are similar to those encountered in blocking, which center on how the four treatment combinations should be chosen and what information is contained in those experiments. Under the sparsity-of-effects principle, the ABC effect is negligible. Thus, one choice of design is to choose those treatment combinations that are all positive in the ABC contrast coefficients. Equivalently one could choose the set that are all negative in ABC. Table 32.7 shows the experimental design for those coefficients positive in ABC.

TABLE 32.7 One-Half Fraction of the 2^3 Full Factorial Design

Treatment	A	B	C	AB	AC	BC	ABC
a	1	−1	−1	−1	−1	1	1
b	−1	1	−1	−1	1	−1	1
c	−1	−1	1	1	−1	−1	1
abc	1	1	1	1	1	1	1

It should be clear to the reader from the contrast coefficients in Table 32.7 that (1) all information for the ABC term is sacrificed in creating this design and (2) the contrast coefficients that estimate one effect exactly match those for another. For instance, the contrast to estimate the A effect simultaneously estimates the BC effect. That is,

$$\tfrac{1}{2}[a + abc - b - c] = A + BC$$

As previously discussed with the blocks, the effect of A is said to be confounded (aliased) with BC. Similarly, the B effect is confounded with the AC effect, and the C effect is confounded with the AB effect. There is no way to individually estimate those effects, only their linear combinations. This pooling of effect information is a by-product of fractional factorial designs.

In the previous example, the ABC term was assumed the least important effect and used as the basis for constructing the 2^{3-1} experimental design. Formally, ABC is called the *design generator* and is algebraically expressed in the relation

$$I = ABC$$

This is known as the defining relation, where I stands for identity, and implies that the ABC effect is confounded with the overall mean. Knowing this relationship helps determine details about the alias structure. This is accomplished by multiplying each side of the defining relation by an effect of interest and deleting any letter raised to the power 2 (i.e., via modulo 2 arithmetic). Any effect multiplied by I gives the effect back. Below demonstrates how to determine which effects are confounded with the main effect of A.

$$\begin{aligned} A \cdot I &= A \cdot ABC \\ \Rightarrow A &= A^2BC \\ \Rightarrow A &= BC \end{aligned}$$

The interpretation is that the estimated A effect is confounded with the BC effect ($A = BC$), which was previously observed via the contrast coefficients in Table 32.7. Likewise it is trivial to show $B = AC$ and $C = AB$.

The defining relation for the chosen fraction above is more descriptively expressed as $I = +ABC$, since all contrast coefficients were positive in ABC. This is called the principle fraction, and will always contain the treatment combination with all levels at their high setting. Alternatively, the complementary fraction could have been selected such that the defining relation would be expressed as $I = -ABC$. As a consequence, the linear contrasts would estimate $A - BC$, $B - AC$, and $C - AB$. Irrespective of sign, both fractions are statistically equivalent as main effects are confounded with two-factor interactions, although there may be a practical difference between the two.

Appropriately, one-half fraction designs are always constructed with the highest order interaction in the defining relation. For instance, a 2^{5-1} experiment uses $I = ABCDE$ to create the fraction and to investigate the alias structure, as this interaction is assumed the least likely effect to significantly explaining response variation. However, the one-half fraction may still be too large to feasibly execute and therefore fractions of higher degrees should be considered. The quarter-fraction design, denoted 2^{k-2}, is the next highest degree fraction from the half-factorials and comprises of a fourth of the original 2^k factorial runs. These designs require two defining relations, call them $I = E_1$ and $I = E_2$, where the first designates the half-fraction based on the " + " or "−" sign on the E_1 interaction, and the second divides it further into a quarter fraction based on the " + " or "−" sign on the E_2 interaction. Note that all four possible fractions using $\pm E_1$ and $\pm E_2$ are statistically equivalent, with the principle fraction corresponding to choosing $I = +E_1$ and $I = +E_2$ in the defining relation. In addition, the generalized interaction $E_3 = E_1 \cdot E_2$ using modulo 2 arithmetic is also included. To investigate the alias structure for a 2^{k-2} fractional factorial design, the complete defining relation is written as $I = E_1 = E_2 = E_3$. These interactions need to be chosen carefully to obtain a reasonable alias structure.

As an example, consider the 2^{6-2} design for factors A through F, and let $E_1 = ABCE$ and $E_2 = BCDF$. The generalized interaction, E_3, is computed by multiplying the two interactions together and deleting any letter with a power of 2. That is,

$$E_3 = (ABCE) \cdot (BCDF) = AB^2C^2DEF = ADEF$$

and hence the complete defining relation is expressed as

$$I = ABCE = BCDF = ADEF$$

It is easy to show for this design that main effects are confounded with three-factor interactions and higher (e.g., $A = BCE = ABCDF = DEF$) and that two-factor interactions are confounded with two-factor interactions and higher ($AB = CE = ACDF = BDEF$). Again, individual effects are not estimable, only linear combinations.

To succinctly describe the alias structure of fractional factorials, *design resolution* is introduced. The resolution of a fractional factorial design is summarized by the length of the shortest effect (often referred to as the *shortest word*) in the defining relation and is represented by a Roman numeral subscript. The one-half fraction of a 2^5 factorial with defining relation $I = ABCDE$ is called a resolution V design and is formally denoted 2_{V}^{5-1}. Similarly, the one-quarter fraction of a 2^6 factorial with defining relation $I = ABCE = BCDF = ADEF$ is a resolution IV design and is denoted 2_{IV}^{6-2}. The design resolutions of greatest interest are described below.

- *Resolution III*: There exist main effects aliased with two-factor interactions. These designs are primarily used for screening many factors to identify which are the most influential in process modeling.

- *Resolution IV*: There exist main effects confounded with three-factor interactions and two-factor interactions confounded with each other. Assuming the sparsity-of-effects principle, main effects are said to be estimated free and clear, since three-factor interactions and higher are considered negligible.
- *Resolution V*: There exist main effects confounded with four-factor interaction and two-factor interactions confounded with three-factor interactions. Assuming the sparsity-of-effects principle, main effects and two-factor interactions are said to be estimated free and clear, since three-factor interactions and higher are considered negligible.

32.5 DESIGN PROJECTION

One of the major benefits to using the two-level factorials and fractional factorials is to take advantage of the *design projection property*. This states that factorial and fractional factorial designs can be projected into stronger designs in a subset of the significant factors. In the case of unreplicated full factorials, those designs project into full factorials with replicates. For example, by disregarding one insignificant factor from a 2^4 full factorial the design becomes a full $2^{4-1} = 2^3$ factorial with 2^1 replicates at each point. In the case of fractional factorial designs of resolution R, those designs project into full factorials in any of the $R - 1$ factors, possibly with replicates. It may be possible to project to a fuller design with more parameters than the $R - 1$ rule dictates, but this is not guaranteed.

As a specific example, consider the 2_{III}^{3-1} fractional factorial design with defining relation $I = +ABC$ that investigates catalyst equivalents (A), ligand equivalents (B), and solvent volume (C). The four treatment combinations are depicted on the cube in the middle of Figure 32.18. As this is a resolution III design, it projects into a full factorial in any two of the three factors, also displayed in Figure 32.18. The projection property of two-level designs is very important, simply due to its usefulness in obtaining full modeling information on a subset of factors, and its implicit use in sequential experimentation.

32.6 STEEPEST ASCENT

The content so far has focused on employing experimental designs with the sole purpose of identifying magnitude of effects plus two-parameter synergies. Often though, the process models are used for optimization or improvement. The combination of experimental design, model building, and *sequential experimentation* used in searching for a

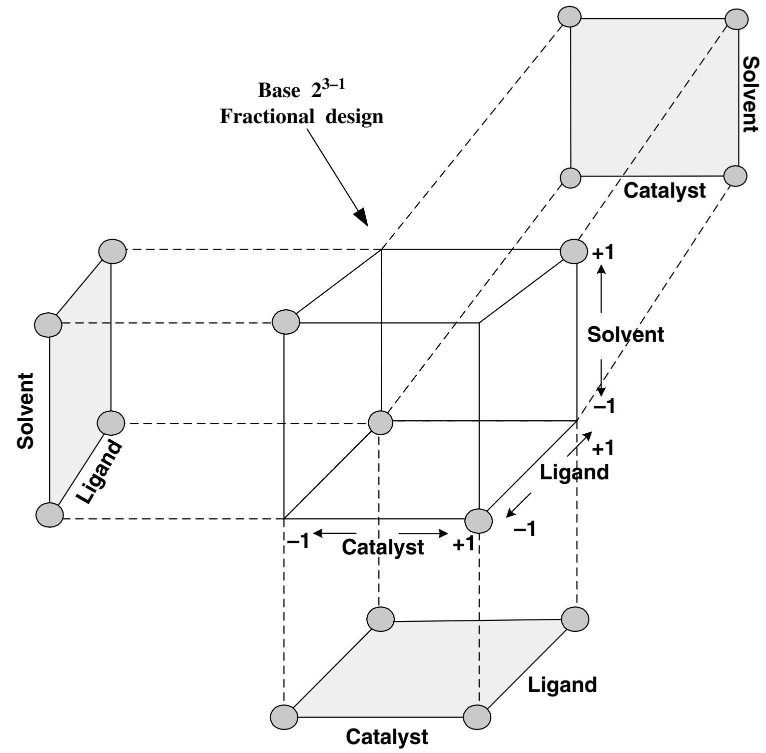

FIGURE 32.18 Illustration of projecting a 2^3 fraction onto a full 2^2 factorial in a subspace of the experimental region.

region of improved response is called the method of steepest ascent (or "descent" if the goal is to explicitly minimize). The goal is to effectively and efficiently move from one region in the factor space to another. Thus, model simplicity as well as design economy are very important. The general algorithm consists of the following:

- Fitting a first-order (main effects) model with an efficient two-level design.
- Computing the path of steepest ascent (descent), where there is an expected maximum increase (decrease) in response.
- Conduct experiments along the path. Eventually response improvement will slow or start to decline.
- Carry out another factorial/fractional design.
- Recompute new path, or augment into optimization design.

Constructing the path of steepest ascent is straightforward. Consider all points that are a fixed distance from the design region center (i.e., radius r) with the desire to seek the parameter combination that maximizes the response. Mathematically, one uses the method of Lagrange multipliers to find where the maximum response lies, constrained to the radius r. Intuitively, the path of steepest ascent is proportional to the size and sign of the coefficients for the first-order model in coded units. For example, let fitted equation be $2+3x_1-1.5x_2$. As shown in Figure 32.19, the path of steepest ascent will have x_1 moving in a positive direction and x_2 in a negative direction. More specifically, the path is such that for every 3.0 units of increase in x_1, there will correspondingly be 1.5 units of decrease in x_2. For steepest decent, the path is chosen using the opposite sign of the coefficients.

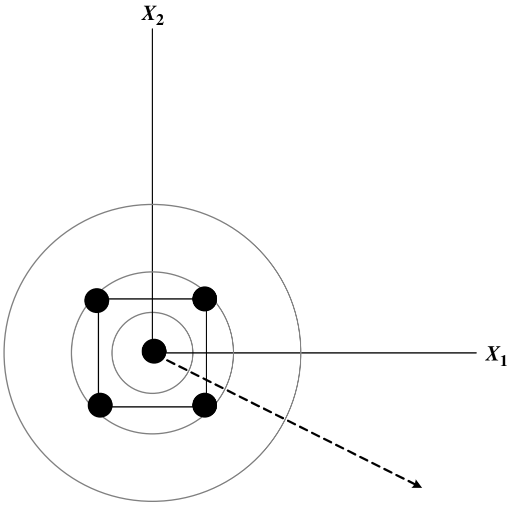

FIGURE 32.19 Path of steepest ascent for the model $\hat{y}=2+3x_1-1.5x_2$.

The success of the steepest ascent method rests on whether the region where the path is constructed is main-effect driven. Steepest ascent should still be successful in the presence of curvature (interaction or quadratic), as long as it is small relative to size of the main effects. If curvature is large, then this exercise is self-defeating. In addition, the success of the path is dependent on the overall process model. Models that are poor and have high uncertainty lead to paths with high uncertainty. Finally, modifying the steepest ascent path with linear constraints are both mathematically and practically easy to incorporate. For more information on process improvement with steepest ascent, see Ref. 6.

32.7 CENTER RUNS

One of the model assumptions in using a two-level design is linearity across the experimental region. If the region is small enough, this is a fair assumption. If the region spans a somewhat broader space and/or the region contains the optimal process condition, then it would not be surprising if nonlinearity exists. Unfortunately, two-level designs by themselves cannot even detect any curvilinear relationship across the design region, much less model it. A cost-effective strategy to initially identify curvature and also have an independent estimate of variance is to add *center runs* to the experimental design. This second point is critical as in practice most 2^k designs are unreplicated. Using the standard ±1 scaling of factor levels, center runs are replicated n_c times at the design point $x_i=0$, $i=1, 2, \ldots, k$. Note that adding center runs produces no impact on the usual effect estimates in the 2^k design. The pure error variance is estimated with n_c-1 degrees of freedom and the test for nonlinearity is via a single degree of freedom contrast that compares the average response at the center to the average response from the factorial points. If nonlinearity is nonexistent across the design region, these averages should be comparable. Specifically, let $\bar{y}_c$ be the average of the n_c center points and let $\bar{y}_f$ be the average of the n_f factorial points. The formal hypothesis test of nonlinearity is conducted by comparing the sum of squares for curvature,

$$\mathrm{SS}_C=\frac{n_f n_c(\bar{y}_f-\bar{y}_c)^2}{n_f+n_c}$$

against the mean square error. This test does not give any information on which factors contribute the sources of curvature, only whether curvature exists or not. If $\bar{y}_f-\bar{y}_c$ is large, then curvature is present across the design region. The implication is that the linear model with main effects and interactions is inadequate for prediction and additional design points or a more advanced design is necessary to identify which specific factors are contributing to the nonlinearity in order to accurately predict across the experimental region.

32.8 RESPONSE SURFACE DESIGNS

Previous discussion has centered on fitting first-order and first-order plus interaction models. However, a higher order model is necessary when in the neighborhood of optimal response. In this case, second-order models are very good approximations to the true underlying functional relationship when curvature exists. These take the form

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \\ &\quad + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \cdots + \beta_{k-1,k} x_{k-1} x_k \\ &\quad + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \cdots + \beta_{kk} x_k^2 + \varepsilon \end{aligned}$$

To estimate the second-order model each factor must have at least three levels. Implicitly there has to be at least as many unique design points as model terms. Many efficient designs are available that accommodate the above model. The most common is the central composite design, abbreviated CCD [2]. The k-factor CCD is comprised of three components, (1) a full 2^k factorial or resolution V fraction, (2) center runs, and (3) $2 \cdot k$ axial points. One compelling feature of CCD designs is that the axial points are a natural augmentation to the standard 2^k or 2_{V}^{k-p} plus center run designs.

As the name suggests, the axial points lie on the axes of each factor in the experimental space. In coded units they are set at a distance $\pm\alpha$ from the center of the design region. The axial value, α, can take on any value, which speaks to the flexibility of these designs. In practice they are usually taken at either $\alpha = 1$ for a face-centered design, $\alpha = \sqrt{k}$ for a spherical design, or $\alpha = f^{1/4}$ for a rotatable design, where f is the size of factorial or fractional used in the CCD. Table 32.8 is an example of a two-factor CCD, and Figure 32.20 displays the experimental region.

Referring to Figure 32.20, if the axial value is set at $\alpha = 1$, then all of the experimental conditions except the center runs lie on the surface of the cube. Similarly, if the axial value is set at $\alpha = \sqrt{k}$, all of the experimental conditions except the center runs lie on the surface of a sphere.

TABLE 32.8 General Two-Factor Central Composite Design

X_1	X_2	
−1	−1	Factorial runs
1	−1	
−1	1	
1	1	
0	*0*	Center runs (≥1)
$-\alpha$	0	Axial runs
α	0	
0	$-\alpha$	
0	α	

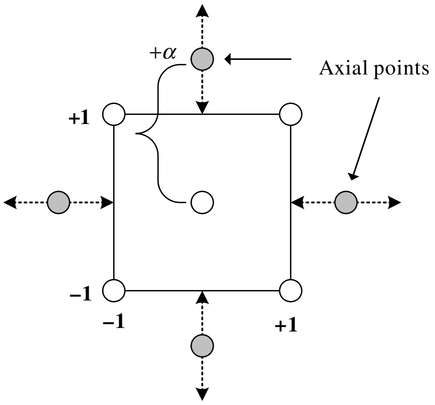

FIGURE 32.20 Illustration of general two-factor central composite design.

For rotatable designs, the precision on the model prediction is a function of only the distance from the design center and the error variance, σ^2. This is illustrated through the two-factor CCD in Table 32.8. The number of factorial points is $n_f = 4$, which yields $\alpha = 4^{1/4} = \sqrt{2}$ as the axial value for rotatability. Assume the design contains $n_c = 3$ center runs. Using this experimental plan, Figure 32.21 displays how the scaled standard error of prediction, a quantity proportional to the size of the confidence interval on the model prediction, varies across the design region. First, the prediction error increases toward the boundary of the design region, a behavior typically seen with confidence bands about a simple linear regression fit. Second, as the design is rotatable, the standard error of prediction is constant on spheres of radius r. Consider two model predictions in Figure 32.21, $\hat{y}(\underline{x}_1)$ and $\hat{y}(\underline{x}_2)$, that are at different coordinates in the experimental region. While the model-based predictions should be different, the precision of $\hat{y}(\underline{x}_1)$ and $\hat{y}(\underline{x}_2)$ is the same as both are equidistant from the center of the design region.

An alternative to the class of central composite designs are the Box–Behnken [7] designs (BBD). These experimental plans are very efficient in fitting second-order models, are nearly rotatable, and have a potentially practical advantage of experimenting with three equally spaced levels over the experimental region. These designs are constructed by incorporating 2^2 or 2^3 factorial arrays in a balanced incomplete block fashion, with the other factors set at their center value. Table 32.9 is an example of a three-factor BBD and Figure 32.22 displays the design in graphical form. The BBDs are spherical designs and there are no factorial "corner points" or face points. For the $k = 3$ design shown in Figure 32.22, all conditions except the center runs are at $\sqrt{2}$ distance from the center. This should not deter the engineer from using this design, especially if predicting at the corners is not of interest, potentially due to cost, impracticality, feasibility, or other issues.

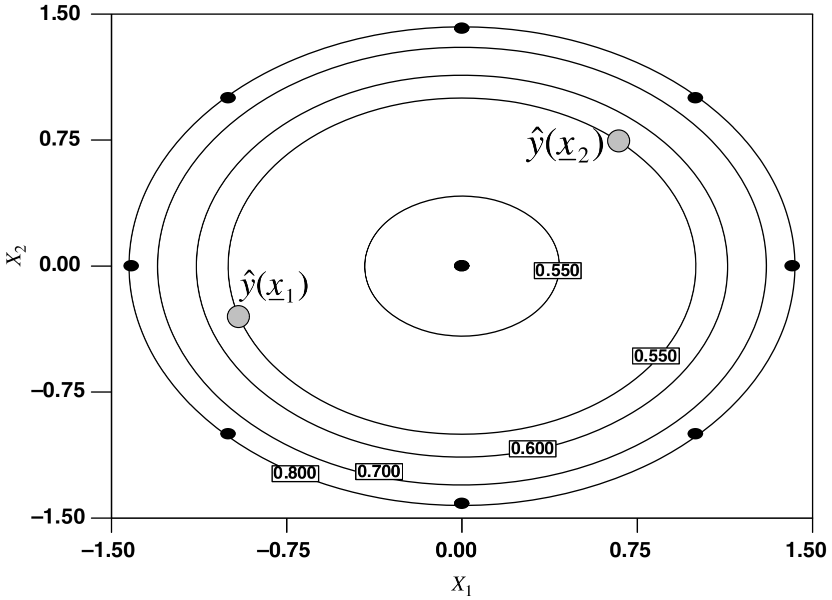

FIGURE 32.21 Illustration of rotatability using the standard error of prediction across factor space.

TABLE 32.9 Three-factor Box–Behnken Design

X_1	X_2	X_3
−1	−1	0
1	−1	0
−1	1	0
1	1	0
−1	0	−1
1	0	−1
−1	0	1
1	0	1
0	−1	−1
0	1	−1
0	−1	1
0	1	1
0	*0*	*0*

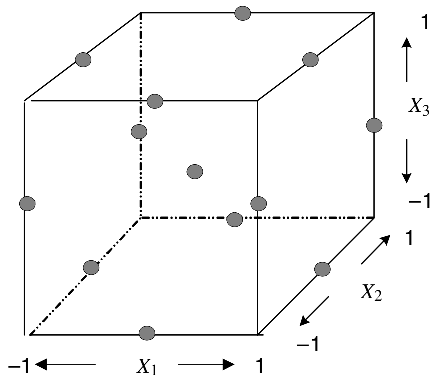

FIGURE 32.22 Illustration of three-factor Box–Behnken design.

32.9 COMPUTER GENERATED DESIGNS

Classical experimental designs such as those discussed so far, may not be appropriate for a practical situation, due to one or more constraints. These could include:

- Sample size limitations due to time, budget, or material, which may yield a nonstandard number of runs, say 11 or 13;
- Nonviable factor settings, thereby impacting the experimental region's geometry. Examples include solubility limitations, safety concerns, and small-scale mixing sensitivities;
- Desired factor levels are nonstandard, say 4 or 6;
- Factors are both qualitative and quantitative;
- Proposed model may be more complicated than first- or second-order polynomial, either higher in order or nonlinear.

For such cases, experimental designs should be tailored to accommodate any constraints, yet still preserve properties that the more classical designs typically possess, such as those based on model precision and prediction precision. The design construction is accomplished with computer assistance and falls under the class of *computer-generated designs*. The computer is a vital tool to construct appropriate designs that meet certain objectives, but unfortunately can be viewed as a black box and misused because of gaps in understanding exactly what the computer algorithms are doing.

Computer-generated designs, and the area of optimal design theory, can be attributed to Kiefer [8, 9]˙ and

Kiefer and Wolfowitz [10]. The general algorithm consists of the engineer providing an objective function that reflects the design property of interest, a hypothesized model, the sample size for the study, and other design elements potentially corresponding to blocks, center runs, and lack-of-fit. The algorithm then uses a fine grid of candidate experimental conditions and searches for the design that optimizes the objective function.

There are several objective functions that speak to design properties of interest and are referred to as alphabetic optimality criteria. As background to the objective functions, consider the linear model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}$$

where $\underline{y}$ is the $(N \times 1)$ vector of observations, X is the $(N \times p)$ model matrix, $\underline{\beta}$ is the $(p \times 1)$ vector of model coefficients, and $\underline{\varepsilon}$ is the $(N \times 1)$ vector of errors assumed to be independent and normally distributed with mean zero and variance σ^2. The ordinary least squares estimate of model coefficients is

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$$

and the covariance matrix of those estimates is given by

$$Var(\hat{\underline{\beta}}) = (X'X)^{-1}\sigma^2 \qquad (32.1)$$

In addition, the variance of a predicted mean response, $\hat{y}_0$, at coordinates $\underline{x}_0$ is given by

$$\mathrm{Var}(\hat{y}_0) = \underline{x}'_0(X'X)^{-1}\underline{x}_0\sigma^2 \qquad (32.2)$$

It should be apparent to the reader from (32.1) and (32.2) the importance of good experimental design on the model and prediction precision, as demonstrated through the $(X'X)^{-1}$ matrix embedded in both of those quantities.

The most common optimality criterion is D-optimality, which *minimizes the joint confidence region* on the regression model coefficients. A-optimality is a similar and common criterion that *minimizes the average size* of a confidence interval on the regression coefficients. Both D- and A-optimality use functions of (32.1) to obtain the appropriate design, and are defined by the scaled moment matrix, M, expressed as

$$M = \frac{X'X}{N}$$

for completely randomized designs. The scaling takes away any dependence on σ^2, a constant independent of the design, and the sample size, N, which allows for comparisons across designs of different size. The algorithm finds that set of design points that maximizes the determinant of M for D-optimality, and minimizes the trace of M^{-1} for A-optimality.

Two other criteria are G-optimality and IV-optimality, which use functions of the prediction variance in (32.2) to obtain a candidate design. A G-optimal design minimizes *the maximum size* of a confidence interval on a prediction over the entire experimental region, whereas IV-optimality minimizes *the average size* of a confidence interval on a prediction over the entire experimental region. Similar to the scaling done with the D-criterion above, these criteria are defined by the scaled prediction variance given by $\upsilon(\underline{x}) = N\underline{x}'(X'X)^{-1}\underline{x}$

Research and software application in the area of optimal design theory has grown immensely over the recent past. Clearly the flexibility is appealing and at times invaluable. It allows the engineer to generate an experimental design for any sample size, number of factors (both discrete and continuous), type of model (linear or nonlinear), and randomization restrictions. Here are some additional notes and cautions regarding computer-generated designs:

- They are *model-dependent* optimal. Occasionally, the engineer will proposed a mechanistic model that represents the true relationship between y and $\underline{x}$. Often though, empirical models are proposed and inevitably edited after collecting data. Optimal designs constructed for one model could be fairly suboptimal with respect to an edited model, thereby impacting process modeling performance. There are strategies and graphical tools available to help generate and assess model robustness, such as those proposed by Heredia-Langner [11].
- The optimal design for one criterion is usually robust/near optimal across other criteria [12]. This is not overly surprising due to the importance of the $(X'X)^{-1}$ matrix. However, this is not guaranteed as it is possible to generate a D-optimal design that has poor prediction variance.
- Slight variations in algorithms and software packages could lead to generated designs that are statistically equivalent but different experimentally. As a matter of good scientific practice, the engineer needs to scrutinize the design for merit and practicality.
- Two-level designs for main effects only and main effects plus two-factor interaction models are all A-, D-, G-, and IV-optimal.
- Classical CCD and BBD designs for second-order models are near optimal. That is, they are highly efficient relative to the most optimal designs.

32.10 MULTIPLE RESPONSES

Up until now any discussion involving effect identification and regression analysis has focused on a single response, and the process model for that response can be used to hone in on a region of *desirability* in the factor space, as defined by the engineer. However, rarely in practice are experiments conducted when only a single response is collected. With chemical reaction trials, natural outputs could include various impurities levels, completion time, and yield, just to name a few. Standard analysis practice would involve (1) modeling each of those outcomes separately, (2) defining their respective region of acceptable response over the factor space, and (3) identifying the intersection of individual regions where all responses are deemed acceptable. Some software packages include this feature of *overlaying* contour plots, which is an effective approach in locating an optimal process-operating region. However, when the number of responses and/or factors gets somewhat large, this exercise can become quite cumbersome. In addition, it is not surprising to have competing responses, meaning the optimal region for one response is suboptimal for one or more of the other responses. This commonly occurs with crystallization processes, where maximum impurity purge is often at the sacrifice of higher yield, due to similar solubility properties of the chemical species. The question then becomes how to effectively merge process model information to identify conditions that are optimally balanced across multiple criteria.

The idea of desirability functions introduced by Derringer and Suich [13] addresses this problem. This is a formula scaled between [0,1] inclusively, where the researchers own priorities and requirements are built into the optimization procedure. To illustrate, consider minimizing the total impurity level (%) from a chemical reaction and assume that any reaction that produces $\leq 0.5\%$ is highly desirable. On the other hand, assume a reaction with $>3\%$ is unacceptable. The desirability function, d, for that response is expressed as

$$d = \begin{cases} 1, & \hat{y} \leq 0.5 \\ \left(\dfrac{3.0-\hat{y}}{3.0-0.5}\right)^S, & 0.5 < \hat{y} < 3.0 \\ 0, & \hat{y} \geq 3.0 \end{cases}$$

Model predictions less than 0.5% get the highest desirability score of 1, whereas model predictions higher than 3.0% get the lowest desirability score of 0. For cases in between those levels, there exist a gradient of desirability scores that are a function of both the model prediction and a weight, S. This weight is chosen by the engineer and it determines the severity of not achieving the most desirable goal, which in this case is 0.5% or less. Figure 32.23 gives a visual of how that desirability function behaves for various values of S.

For a given factor setting, each of the m responses has its own desirability score. That is, response i gets desirability, d_i, $i = 1, 2, \ldots, m$. To obtain the overall desirability across m responses at any experimental condition, the overall desirability score, D, is calculated as the geometric mean of each individual d_i, expressed as $D = \{d_1, d_2, \ldots, d_m\}^{1/m}$. This overall score is easily modified when responses vary in importance. For example, impurity responses that affect drug product quality (and therefore affect the patient) are considered more important versus, say, yield that affects a sponsor's bottom line. The final objective is to locate parameter conditions that make D largest. This is normally accomplished via response surface modeling of D across experimental space and/or numerical techniques. Many software packages include this functionality as part of optimization. Note that any identified conditions should be confirmed for acceptability.

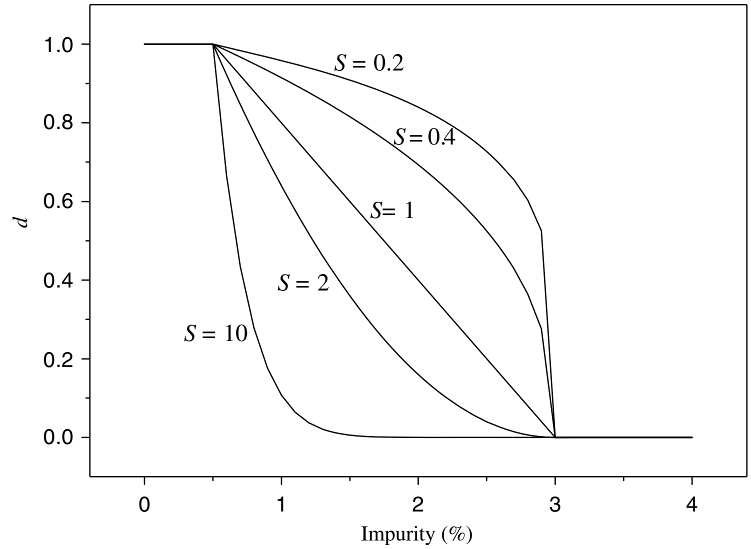

FIGURE 32.23 Example total impurity desirability function across various weights of S.

EXAMPLE 32.2 THREE-FACTOR CENTRAL COMPOSITE DESIGN

A three-factor face-centered central composite design is carried out to identify factor combinations that simultaneously minimize total impurities (%) and maximize yield (g). The three factors are catalyst (Factor *A*), concentration (Factor *B*), and temperature (Factor *C*). Each experimental condition is completely randomized and the center runs are replicated three times ($n_c = 3$). Total size of the study is 17 runs. The randomized design in coded units and data are listed in Table 32.10, and depicted in Figure 32.24. A full second-order model was fit to each response. The final ANOVA, fit statistics, predicted model equation, and relevant plots are presented below.

TABLE 32.10 Three-Factor CCD Design to Optimize Total Impurities and Yield

Design			Data	
Catalyst	Concentration	Temperature	Total Impurities	Yield
−1	1	−1	23.54	33.95
−1	−1	−1	10.57	15.27
0	0	0	12.20	35.10
1	1	−1	19.74	15.43
−1	1	1	5.20	18.68
1	−1	−1	8.61	2.50
0	0	0	11.10	37.20
1	−1	1	0.79	24.00
0	0	1	8.63	33.40
−1	−1	1	2.59	8.30
0	0	0	10.10	42.20
1	0	0	11.96	32.16
1	1	1	4.40	42.10
0	−1	0	5.47	31.90
0	0	−1	15.98	21.60
0	1	0	17.07	47.80
−1	0	0	12.82	31.17

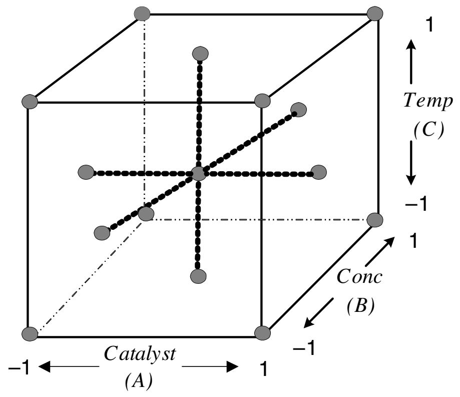

FIGURE 32.24 Illustration of three-factor face-centered CCD design.

Figure 32.25 gives an example snapshot of all relevant information in modeling total impurities. The table shows that total impurities are jointly impacted by concentration and temperature, as shown by the significant *p*-value on the interaction term, and independent of catalyst level. The model explains approximately 90.7% of the variation in the data as calculated by the R^2 value, and the final predicted model equation in coded units is embedded as well. Two previously highlighted diagnostic plots are included in this snapshot, although there could be others of interest that highlight or confirm aspects of the data/model. The normal probability plot of residuals shows no deviation from the normality assumption, and the predicted versus actual plot indicates that the model is performing well with constant variability across the range of prediction. Finally, a model-based contour plot of predicted total impurities across concentration and temperature is shown along with its 3D analogue surface plot. From this graphs, total impurities are minimized at the combination lower concentration and higher temperatures levels, independent of catalyst. That is, (concentration, temperature) = (−1, 1).

Figure 32.26 shows all relevant information in modeling yield. The embedded table shows that yield is impacted by all three factors, including second-order catalyst and temperature effects. Notice also that the main effect of catalyst is included even though the *p*-value is insignificant. This is to preserve *model hierarchy*, as catalyst does explain some of the variation in response, but in conjunction with higher order terms (either interactions and/or as a second-order effect). The model explains approximately 97.4% of the variation in the data as calculated by the R^2 value, and the final predicted model equation in coded units is shown. The normal probability plot of residuals demonstrates that the residuals can be assumed normally distributed, and the predicted versus actual plot indicates that the model is performing quite well with no departures from the constant variability assumption. And again, a model-based contour plot of predicted yield across catalyst and temperature levels at the high concentration is displayed along with its corresponding surface plot. The concentration level is set to high because that factor comes in the model *only* as a positive main effect, implying higher concentration predicts higher yield. Therefore setting concentration at its high level is in the optimal direction. From this information, yield is maximized at (catalyst, concentration, temperature) ≈ (0.25, 1.0, 0.25) in coded units.

In identifying a candidate process-operating region that is optimal over both responses, criteria that define acceptable performance are established. For this example, assume that the process is acceptable if the predicted total impurities are below 10% and the yield is above 40 g. Figure 32.27 displays

Source	SS	DF	MS	F	p-value
B - Conc	175.7286	1	175.73	41.46	< 0.0001
C- Temperature	322.9649	1	322.96	76.19	< 0.0001
BC	39.9618	1	39.96	9.43	0.0089
Error	55.10706	13	4.24		
Total	593.7624	16			

$R^2 = 90.7\%$ $\hat{y}_{\text{imp}} = 10.63 + 4.19(\text{Conc}) - 5.68(\text{Temp}) - 2.24(\text{Conc})(\text{Temp})$

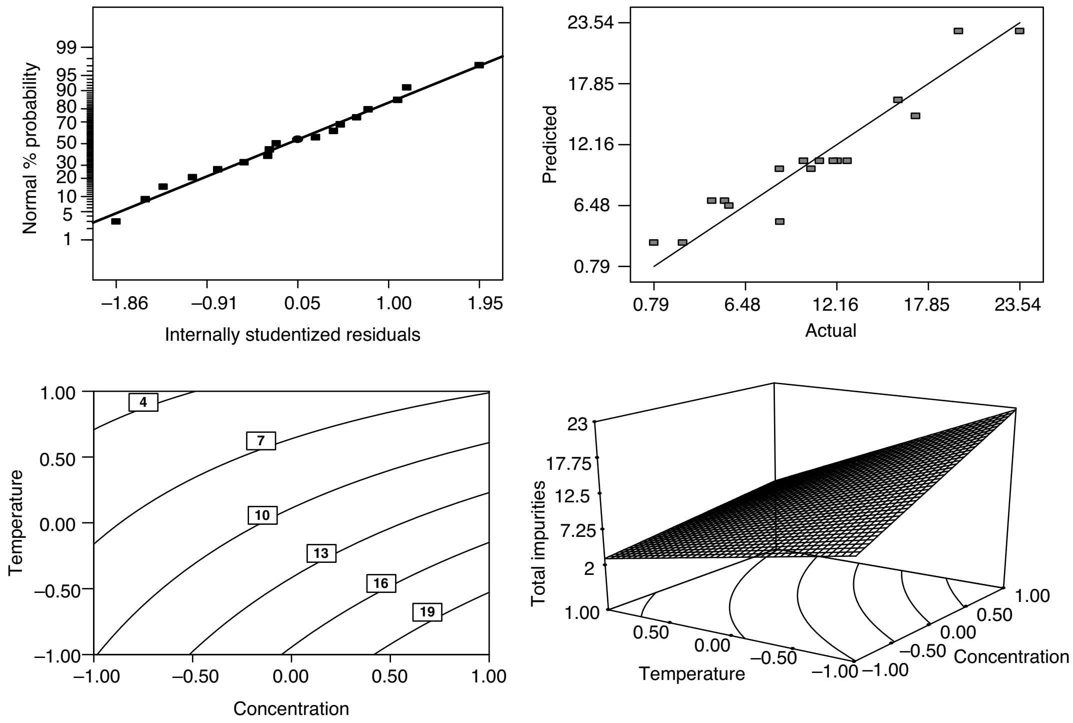

FIGURE 32.25 Summary of total impurities modeling.

at the high (+ 1) concentration level the subspace (in white) of catalyst–temperature combinations where simultaneously both responses meet the specified process performance criteria. Similarly, one solution from using the desirability function predicts a candidate optimal setting at (catalyst, concentration, temperature) ≈ (0.67, 1, 1). This is denoted in the upper part of the white area by (●). This condition would be verified experimentally.

32.11 ADVANCED TOPICS

32.11.1 Industrial Split-Plot Designs

One assumption made throughout the chapter is that all experimental designs have complete randomization of the treatment combinations, which are generally called *completely randomized designs*. However, for experiments run at larger scale and/or with equipment limitations, complete randomization of the experiments is arduous. This happens with factors that are difficult to independently change for each design run, or impractical when certain factors can and should be held constant for some duration of the experiment due to resource or budget constraints. Temperature and pressure are examples that immediately come to mind of hard-to-change factors. When this situation occurs, part of the design is executed in "batch-mode." That is, certain treatment combinations are fixed across a sequence of experiments, without resetting the actual treatment combination. This type of execution results in nested sources of variation and is called a *split-plot design*. These designs were originally developed for agronomic experiments, but its applicability easily spans all fields of science, even as the agricultural naming conventions have endured.

The basic split-plot experiment can be viewed as two experiments that are superimposed on each other. The first corresponds to a randomization of hard-to-change factors to

Source	SS	DF	MS	F	p-value
A-Catalyst	7.78	1	7.8	1.2	0.3047
B - Conc	577.45	1	577.4	86.9	< 0.0001
C-Temp	142.36	1	142.4	21.4	0.0009
AC	619.70	1	619.7	93.2	< 0.0001
A^2	162.79	1	162.8	24.5	0.0006
C^2	400.29	1	400.3	60.2	< 0.0001
Error	66.46	10	6.6		
Total	2536.27	16			

$R^2 = 97.4\%$

$$\hat{y}_{\text{yield}} = 38.88 + 0.88(\text{Catalyst}) + 7.6(\text{Conc}) + 3.77(\text{Temp}) + 8.8(\text{Catalyst})(\text{Temp}) - 7.33(\text{Catalyst}^2) - 11.50(\text{Temp}^2)$$

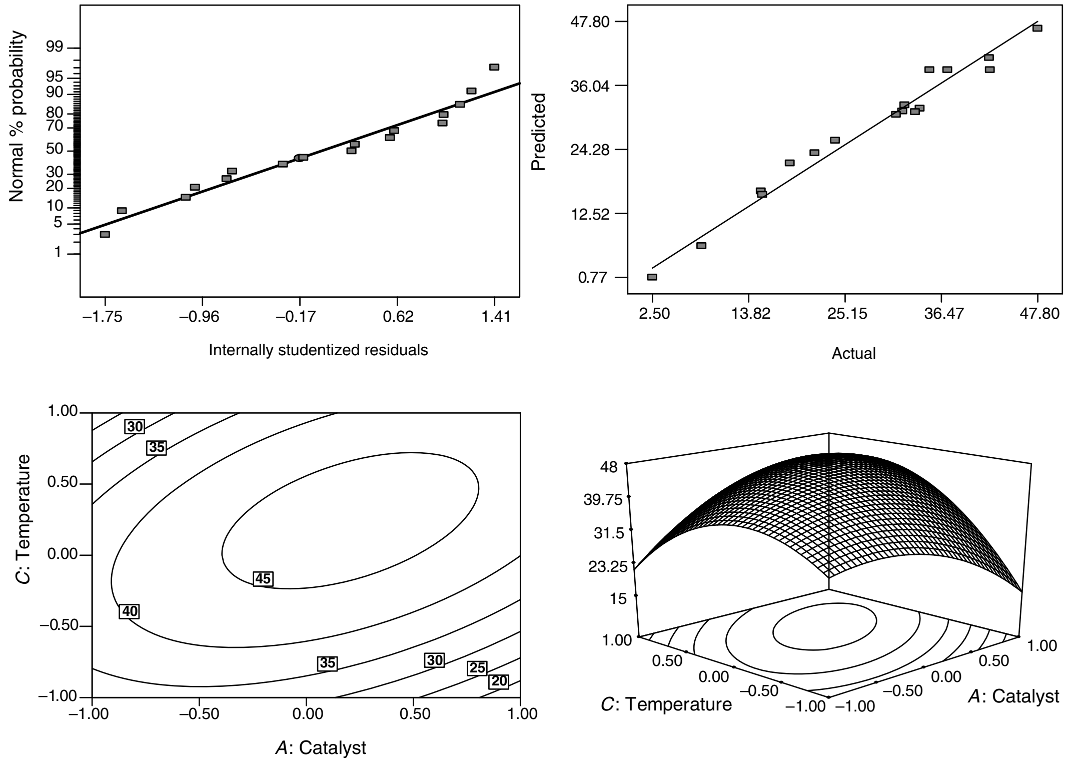

FIGURE 32.26 Summary of yield modeling.

experimental units called *whole plots*, while the second corresponds to a separate randomization of the easy-to-change factors within each whole plot, called *subplots*. This is different than completely randomized designs that have unrestricted randomization factor combinations across all experimental units. The blocking of the hard-to-change factors along with the two separate randomization sequences creates a correlation structure among data collected within the same whole plot. Inferentially, the associated ANOVA needs to reflect that the experiment was executed in a split-plot fashion. If data from a split-plot structure were analyzed as if the experiments were completely randomized, then it is possible to erroneously conclude significance of hard-to-change effects when they are not, while conclude insignificance of easy-to-change effects when they are. A good discussion on classical split-plot designs can be found in Hinkelmann and Kempthorne [14], as well as in Box and Jones [15] regarding response surface methodology.

In the recent past considerable attention has been given to constructing and evaluating optimal split-plot designs, especially with the computational horsepower of today's computers. Topics span algorithms for D-optimal split-plot designs [16, 17], to comparing the performance between classical response surface designs in a split-plot structure [18], to graphical techniques for comparing competing split-plot designs [19]. This is an important area of research as many industrial experiments are conducted with restricted randomization, whether deliberately planned or not. For those cases when it is planned, more software tools are becoming widely available so that the split-plot experiment are both powered sufficiently and analyzed appropriately.

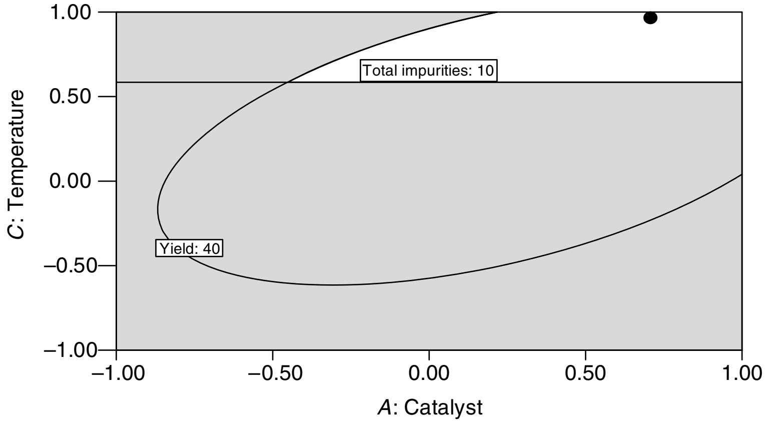

FIGURE 32.27 Overlay plot of catalyst–temperature combinations that meet specified performance criteria for total impurities and yield, with candidate optimal setting (●).

32.11.2 Nonstandard Conditions

Model diagnostics and the analysis of residuals are a crucial part of any model building exercise and are standard with any software package with a regression component. The information can numerically or visually identify any conditions from the standard assumption of normal and independent residuals with mean zero and common variance σ^2. Examples of nonstandard conditions are heterogeneous variance, data transformations, outliers, and nonnormal errors.

In practice, the assumption of homogeneous variance is most likely violated. For many scientific applications, it is natural that variability increases with either response or regressor. From a theoretical perspective, this is simple to accommodate as the ordinary least squares estimate of regression coefficients, $\hat{\beta}$, is slightly altered to a weighted least squares estimate, where the weights are the inverse of the variances. The implication is that residuals with large variances (small weights) should not count as much in the model fit as those with small variances (large weights). For a given matrix of weights, W, the expression for the weighted least squares estimate of β is given by

$$\underline{\hat{\beta}}_{WLS} = (X'WX)^{-1}X'W\underline{y}$$

The immediate practical issue is how to obtain the weights. One approach is to collect multiple data at each experimental condition and use the corresponding estimate of variance. However, this is problematic if the sample variances are based on a small number of data, and therefore potentially unreliable as a solution. A loose rule of thumb is that any estimated weights should be based on no less than a sample of nine [20]. Less than that can result in very poor performance of the weighted least squares solution, and ignoring the weights would be the better course of action.

One effective approach to the issue of increasing variance relative to increasing response is through a response transformation. As a by-product, the reexpression of data also helps with error normality assumption, as well as with both model prediction and model selection. The most common data transformation is the log transformation ($y^* = \ln(y)$ or $y^* = \log_{10}(y)$), where the mechanism of change in y across $\underline{x}$ is exponential. Other less common transformations include the square root ($y^* = \sqrt{y}$) and reciprocal ($y^* = y^{-1}$). All of these examples fall under the class of power transformations, where $y^* = y^{\lambda}$. The Box–Cox method [21] is one particular and powerful technique that aids in properly transforming y into y^* and takes the form

$$y^*(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y}\ln(y) & \lambda = 0 \end{cases}$$

where $\dot{y}$ is the geometric mean of the original data. The assignment of ln(y) when $\lambda = 0$ comes from the limit of $(y^{\lambda} - 1)/\lambda$ as λ approaches zero, which allows for continuity in λ. The use of $\dot{y}$ rescales the response to the same units so that the error sum of squares can be comparable across different values of λ. As mentioned, common power transformation values of λ are -1, 0, 0.5, and 1, the last corresponding to no response transformation. In practice, after an appropriate model is fit to the data, an estimate of λ that minimizes the error sum of squares is estimated. If this value is significantly different than 1.0, a transformation on the data is recommended to produce a better fit, with potentially a different model. Most software packages that contain the Box–Cox procedure as part of model diagnostics provide a confidence interval on λ that simultaneously tests the hypothesis of no response transformation needed and puts a bound on a recommended one.

32.11.3 Designs for Nonlinear Models

Except for a brief mention in the section on computer generated designs, all the discussion and examples in this chapter have only considered models that are linear in their parameters, which are then (successfully) used to approximate underlying nonlinear behavior over a small region. By definition, a linear model is one where the partial derivatives with respect to each model parameter are only a function of the factor/regressor. A simple example is the two-factor interaction model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$$

It is trivial to show that $\partial y/\partial \beta_i$ ($i = 0$, 1, 2, and 12), does not depend on β_i. Conversely, a nonlinear model is defined as one where at least one of the partial derivatives is a function of a model parameter. An example of a nonlinear model would be

$$y = \beta_0 \, \mathrm{e}^{-\beta_1 x}$$

where both $\partial y/\partial \beta_0$ and $\partial y/\partial \beta_1$ are still functions of the unknown β parameters. Of course, chemical engineers are fully aware that the majority of kinetic models are nonlinear and not even closed form but expressed as differentials.

As opposed to linear models, in most cases classical experimental designs are not appropriate or optimal for nonlinear models. (The notable exception is in some applications of generalized linear models, as discussed by Myers et al. [22].) Thus, practitioners typically rely on computer generated designs that are optimal in some respect, such as by D- or G-optimality criteria previously described. However, this is problematic and circular in terms of design construction, since *the optimal design for a nonlinear model is a function of the unknown parameters*. To circumvent that issue, initial parameter estimates are used that are often the results of previous studies and/or scientific knowledge. These designs are called *locally optimal*, and Box and Lucas [23] discuss locally D-optimal designs for nonlinear models. Intuitively, if the initial model parameter estimates are poor, then the locally optimal design will suffer in performance. One avenue that does not rely on initial parameter point-estimates is to use Bayesian approach to experimental design [24], where the scientist would postulate a *prior distribution* on the parameters plus specify a utility function, which is similar in spirit to the objective functions discussed in classical alphabetic optimal criteria. Regardless, optimal designs for nonlinear models are inherently sequential, which may be an obstacle in adoption. In addition, commonly available technology has not caught up to the contemporary thinking in this field, adding another very tangible barrier. Nevertheless, independent of model-type it should be apparent that the process model fit is only as good as the design behind the data.

REFERENCES

1. Montgomery DC. *Design and Analysis of Experiments*, 6th edition, Wiley, New York, 2005.
2. Box GEP, Wilson KB. On the experimental attainment of optimum conditions. *J. Royal Stat. Soc., Ser. B* 1951;13: 1–45.
3. Myers RH. *Classical and Modern Regression with Applications*, 2nd edition, Duxbury Press, Belmont, CA, 1990.
4. *Design Expert version 7.1.6*, Stat-Ease, Inc, Minneapolis, MN November, 2008.
5. Daniel C. Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics*, 1959;1:311–342.
6. Myers RH, Montgomery DC. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edition, Wiley, New York, 2002.
7. Box GEP, Behnken DW. Some new three-level designs for the study of quantitative variables. *Technometrics* 1960;2: 455–475.
8. Kiefer J. Optimum experimental designs. *J. Royal Stat. Soc., Ser. B* 1959;21:272–304.
9. Kiefer J. Optimum designs in regression problems. *Ann. Math. Stat.* 1961;32:298–325.
10. Kiefer J, Wolfowitz J. Optimum designs in regression problems. *Ann. Math. Stat.* 1959;30:271–294.
11. Heredia-Langner A, Montgomery DC, Carlyle WM, Borror CM. Model robust designs: a genetic algorithm approach. *J. Quality Technol.* 2004;36:263–279.
12. Cornell JA. *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, 3rd edition, Wiley, New York, 2002.
13. Derringer G, Suich R. Simultaneous optimization of several response variables. *J. Quality Technol.* 1980;12:214–219.
14. Hinkelmann K, Kempthorne O. *Design and Analysis of Experiments Volume I: Introduction to Experimental Design*, Wiley, New York, 1994.
15. Box GEP, Jones S. Split-plot designs for robust product experimentation. *J. Appl. Stat.* 1992;19:3–26.
16. Goos P, Vandebroek M. Optimal split plot designs. *J. Quality Technol.* 2001;33:436–450.
17. Goos P, Vandebroek M. D-optimal split plot designs with given numbers and sizes of whole plots. *Technometrics* 2003;45: 235–245.
18. Letsinger JD, Myers RH, Lentner M. Response surface methods for bi-randomization structure. *J. Quality Technol.* 1996;28: 381–397.
19. Liang L, Anderson-Cook CM, Robinson TJ, Myers RH. Three dimensional variance dispersion graphs for split-plot designs. *J. Comput. Graph. Stat.* 2006;15:757–778.

20. Deaton ML, Reynolds MR, Jr., Myers RH. Estimation and hypothesis testing in regression in the presence of nonhomogenous error variances. *Commun. Stat.* 1983;B12(1): 45–66.
21. Box GEP, Cox DR. An analysis of transformations (with discussion). *J. Royal Stat. Soc., Ser. B* 1964;26: 211–246.
22. Myers RH, Montgomery DC, Vining GG. *Generalized Linear Models With Applications in Engineering and the Sciences*, Wiley, New York, 2002.
23. Box GEP, Lucas HL. Design of experiments in non-linear situations. *Biometrika* 1959;46:77–90.
24. Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Stat. Sci.* 1995;10:273–304.