

Practical Aspects for Designing Statistically Optimal Experiments

Mark J. Anderson, Patrick J. Whitcomb
Stat-Ease, Inc., Minneapolis, MN USA

Due to operational or physical considerations, standard factorial and response surface method (RSM) design of experiments (DOE) often prove to be unsuitable. In such cases a computer-generated statistically-optimal design fills the breach. This article explores vital mathematical properties for evaluating alternative designs with a focus on what is really important for industrial experimenters. To assess “goodness of design” such evaluations must consider the model choice, specific optimality criteria (in particular D and I), precision of estimation based on the fraction of design space (FDS), the number of runs to achieve required precision, lack-of-fit testing, and so forth. With a focus on RSM, all these issues are considered at a practical level, keeping engineers and scientists in mind. This brings to the forefront such considerations as subject-matter knowledge from first principles and experience, factor choice and the feasibility of the experiment design.

Key words: design of experiments, optimal design, response surface methods, fraction of design space.

Introduction

Statistically optimal designs emerged over a half century ago (Kiefer, 1959) to provide these advantages over classical templates for factorials and RSM:

- Efficiently filling out an irregularly shaped experimental region such as that shown in Figure 1 (Anderson and Whitcomb, 2005),

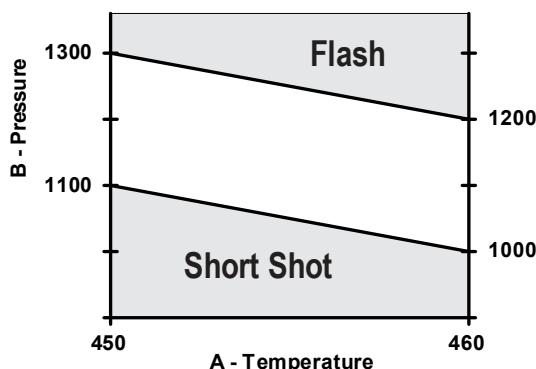


Figure 1. Example of irregularly-shaped experimental region (a molding process).

- Minimizing the runs to just what is needed to fit the assumed polynomial model,
- Accommodating unusual requirements concerning either the number of blocks or the number of runs per

block,

- Handling a combination of factor types, such as continuous, discrete, categorical and mixture.

Over time a number of criteria labeled alphabetically became favored by industrial experiments for optimal designs (Box and Draper, 2007), of which two will be covered in this article: I-optimal and D-optimal. Our focus will be kept to RSM.

What is a “Good” Experiment Design?

To answer this important question, let’s start with a wish-list for choosing a “suitable” experiment design derived from one provided by George Box (Box, 1982)—the co-inventor of RSM (Box and Wilson, 1951):

- (1) Allow the chosen polynomial to be estimated well.
- (2) Give sufficient information to allow a test for lack of fit by
 - a. Having more unique design points than model coefficients, and
 - b. Providing an estimate of “pure” error, i.e., replicates.
- (3) Remain insensitive to outliers, influential values and bias from model misspecification.
- (4) Be robust to errors in control of the factor levels.
- (5) Permit blocking and sequential experimentation.
- (6) Provide a check on variance assumptions, e.g., studentized residuals are normal with a mean of zero and constant variance.
- (7) Generate useful information throughout the region of interest, i.e., provide a good distribution of standard error of predictions.
- (8) Do not contain an excessively large number of trials.

When applying RSM, industrial experimenters generally choose as step one a quadratic polynomial, which are remarkably versatile for empirical modeling. For this purpose, the central composite design (CCD), also known as the Box-Wilson in honor of the developers, scores well on all the desired attributes. However, standard layouts like the CCD are not a good fit for non-cuboidal regions such as that illustrated in Figure 1. For situations like this or others spelled out in the Introduction, optimal designs are seemingly the panacea. However, as we will discuss further, you had best keep in mind that *“designing an experiment should involve balancing multiple objectives, not just focusing on a single characteristic”* (Myers, Montgomery and Anderson-Cook, 2009).

Purely Optimal Designs—Comparing I Versus D as the Criterion

Although there are many variations on the theme of optimal design, the following two criteria are the ones primarily used in industrial experimentation:

- optimal (also known as “IV”) to minimize the integral of the prediction variance
- D-optimal to minimize the volume of the confidence ellipsoid for the coefficients and thus maximize information on the polynomial coefficients.

Rather than getting mired down in the mathematical details for these two criteria and the multitude of algorithms for applying them (readily available from a vast array of references—some already cited here), we will focus on how they differ in actual application to RSM experiments designed to fit quadratic polynomials such as the equation show above. For this purpose a good tool for comparison is the standard error (SE) plot, such as those shown in Figure 2 for a 12-run RSM design on one factor where the points are picked I- versus D-optimally.

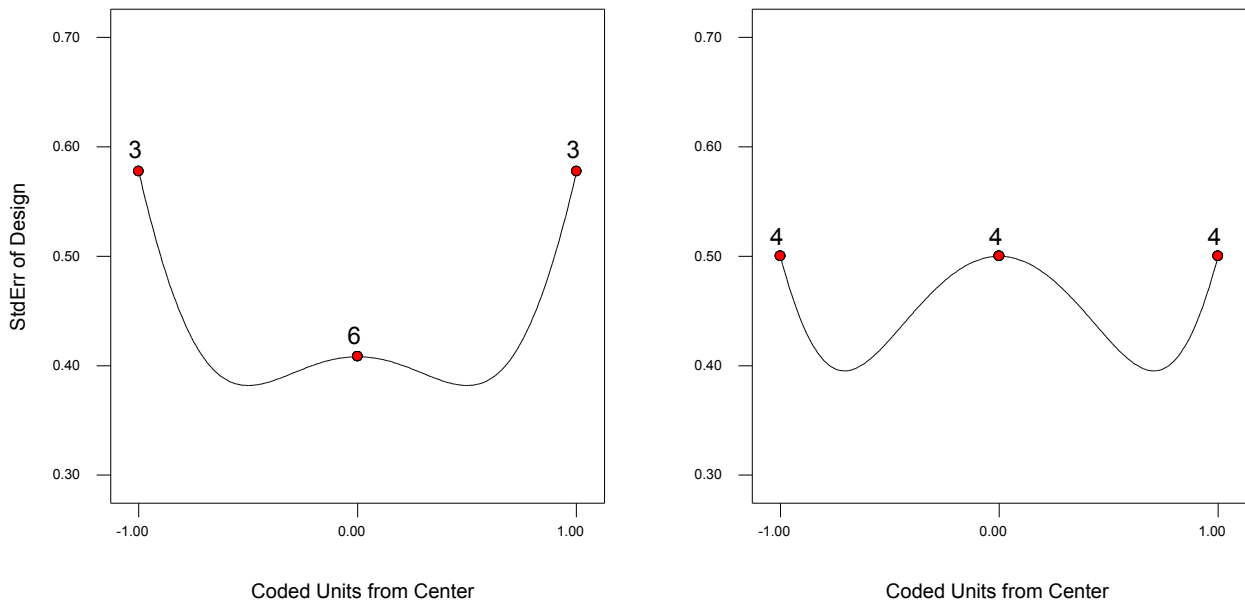


Figure 2. Standard error plots for I (left) versus D optimal (right) for a 12-run, one-factor RSM design.

Comparing these plots side-by-side provides a picture of the relative quality of predicted response at various locations spanning the experimental region (shown here in terms of coded units from the center). The greater replication of points by the I-criterion is desirable for RSM because it lowers the standard error of prediction at the center—the point of greatest interest—and provides a fairly flat profile for a broader (relative to the D-criterion) range in the middle of the experimental region.

Another way to compare designs is via the fraction of design space (FDS) plot (Anderson-Cook, Borror and Montgomery, 2009), which consists of a single line for a given design, thus allowing display of the prediction variance for several designs at once. Figure 3 lays out the FDS curves for the aforementioned 12-run RSM design done by the competing criterion—I versus D.

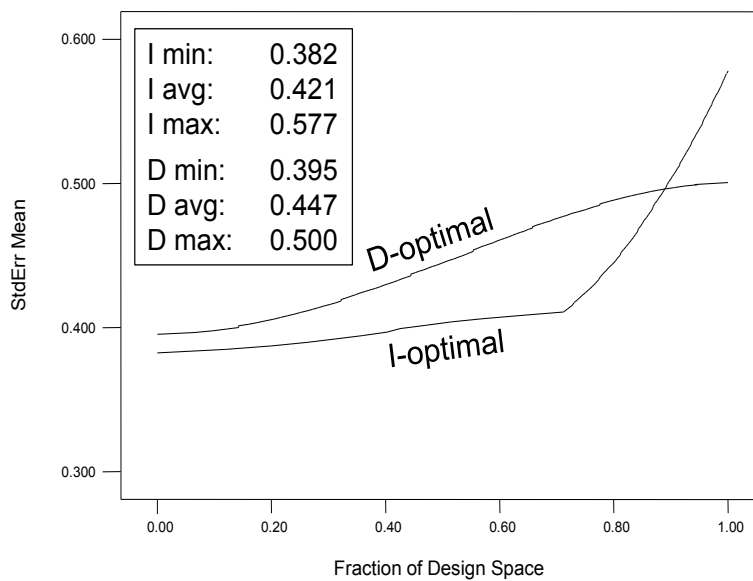


Figure 3. FDS plots for I versus D optimal for a 12-run, one-factor RSM design.

The legend in Figure 3 provides benchmarks on the standard error minimums, averages and maximums. For RSM purposes, the I-optimal provides a desirable tradeoff of being higher at its maximum (not good) but lower on average (good) than the D-optimal design.

At this stage we put I-optimal at the forefront. However, neither of the two optimal designs under consideration provide for a lack-of-fit (LOF) test. Therefore an experimenter cannot assess whether the model they chose provides an adequate approximation of the true response. It's vital to keep in mind that *“no postulated model can ever be assumed to be fully correct [therefore] the basic assumptions underlying the alphabetic-optimality approach are often unrealistic from the practical viewpoint of actually designing real experiments”* (Draper & Guttman, 1988).

Modifying Optimal Designs to Make Them More Robust to Model Misspecification

We will now do another comparison of point selection using I- vs D-optimal on a one-factor design for a quadratic model, but this time the designs will be modified with LOF points to check for model misspecification. We recommend 4 runs for these ‘check-points’—chosen to maximize the minimum distance from existing design points; thus filling ‘holes’ in the experimental space. This is known as the “distance” criterion. Figure 4 provides a side-by-side view, by way of standard error plots, of how this modification affects the spread of points compared to the purely optimal selection shown in Figure 2.

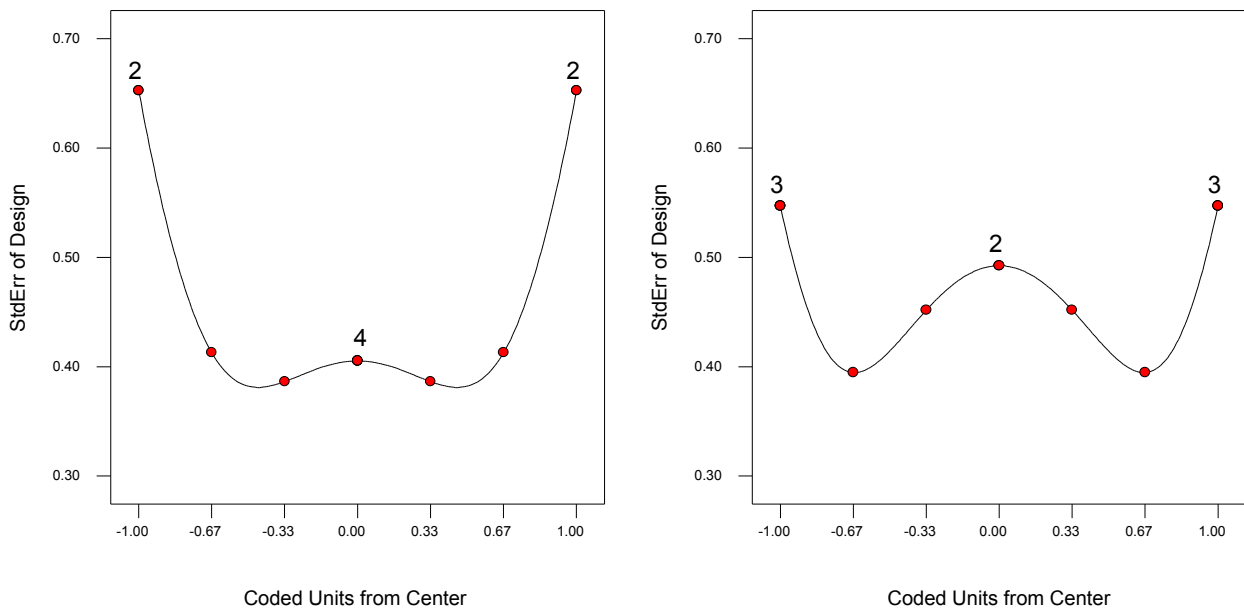


Figure 4. SE plots for LOF-modified I-optimal (left) versus LOF-modified D-optimal (right) for a 12-run, one-factor RSM design.

Observe how gaps in the experimental range have been plugged.

Next let's recalculate the FDS for the two alternative optimal criterion.

Compare and contrast the curves and data on Figures 3 and 5—the first being purely optimal and the second modified with lack-of-fit points. Notice that by all statistical measures and the curves themselves that not much differs. In this case the advantage of having a check for model misspecification outweighs the minor loss in optimality and slight degradation in FDS quality.

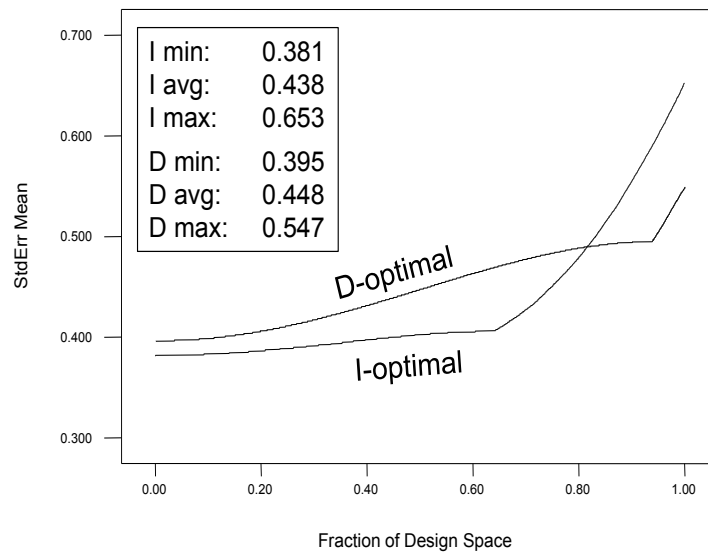


Figure 5. FDS plots for LOF-modified I versus D optimal for a 12-run, one-factor RSM design

Extending These Findings to Two Factors

A similar case can be made for two factors and, by extension, beyond—once the minimum points needed to fit the chosen polynomial model are selected via an optimal criterion, adding LOF points causes little harm and they create a lot of good. Figure 6 lays out the points on SE plot for a 14-run, I-optimal, RSM-quadratic design with zero versus four LOF points.

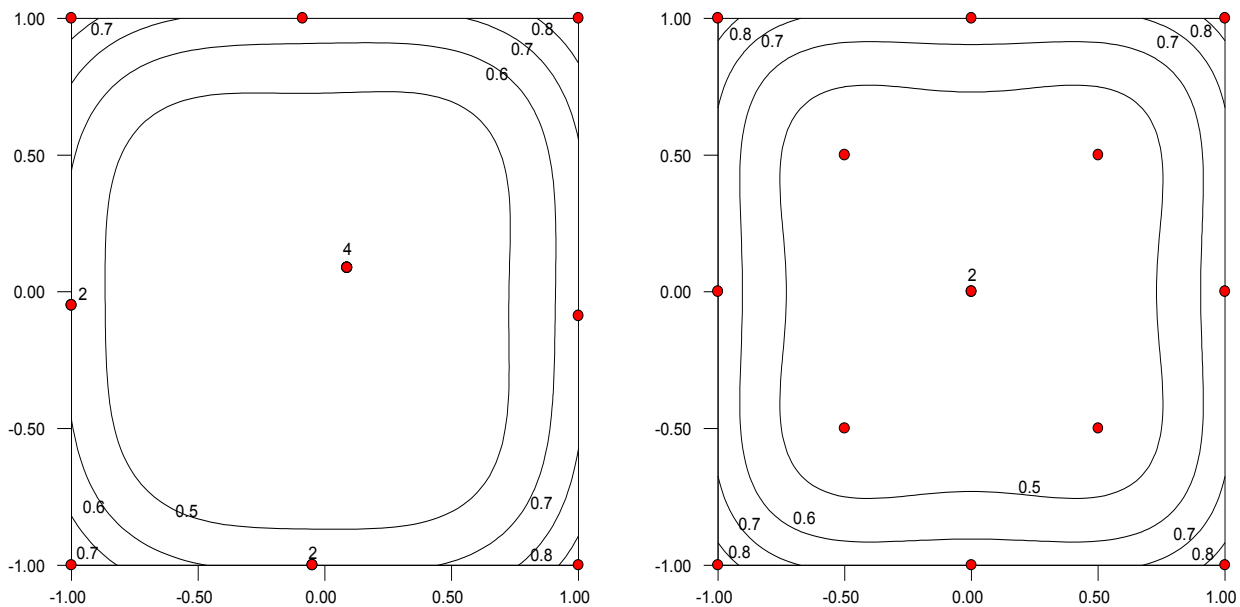


Figure 6. SE plots for purely I-optimal (left) versus LOF-modified (right) for a 14-run, two-factor RSM design.

The space-filling effect of shifting some points from optimal to distance-based criterion (for LOF) is very apparent when comparing these plots from left to right. This comes at little cost to the I-optimal design quality as evidenced by the FDS curve and properties laid out in Figure 7.

However, when we make the same comparison for D-optimal designs, the trade-off of 4 purely optimal points for ones chosen by distance (layouts shown in Figure 8) to provide LOF does not go quite as well—see how the FDS curve shifts in Figure 9 and note the degradation in D-optimality as evidenced by the determinant results being increased.

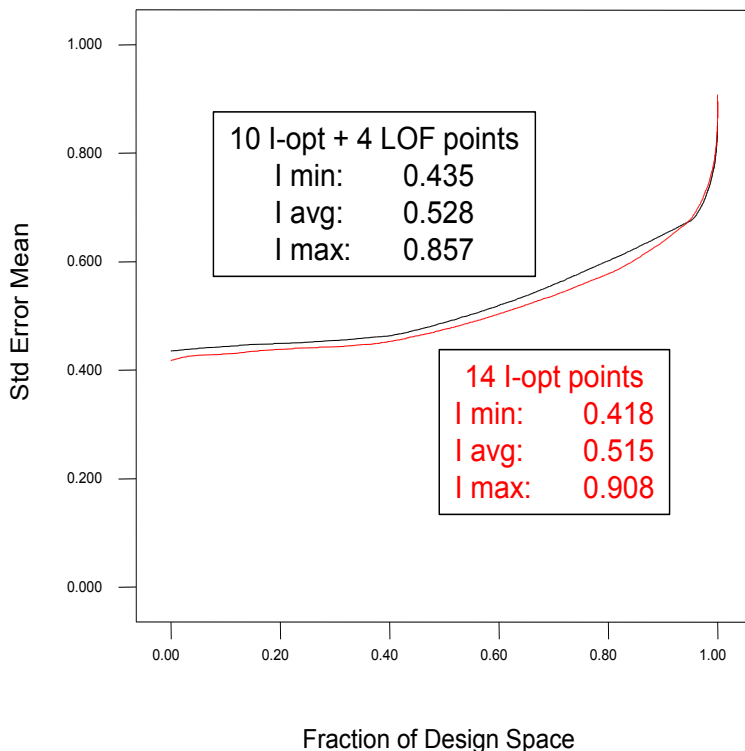


Figure 7. FDS plots for purely I-optimal versus LOF-modified I-optimal for a 14-run, two-factor RSM design.

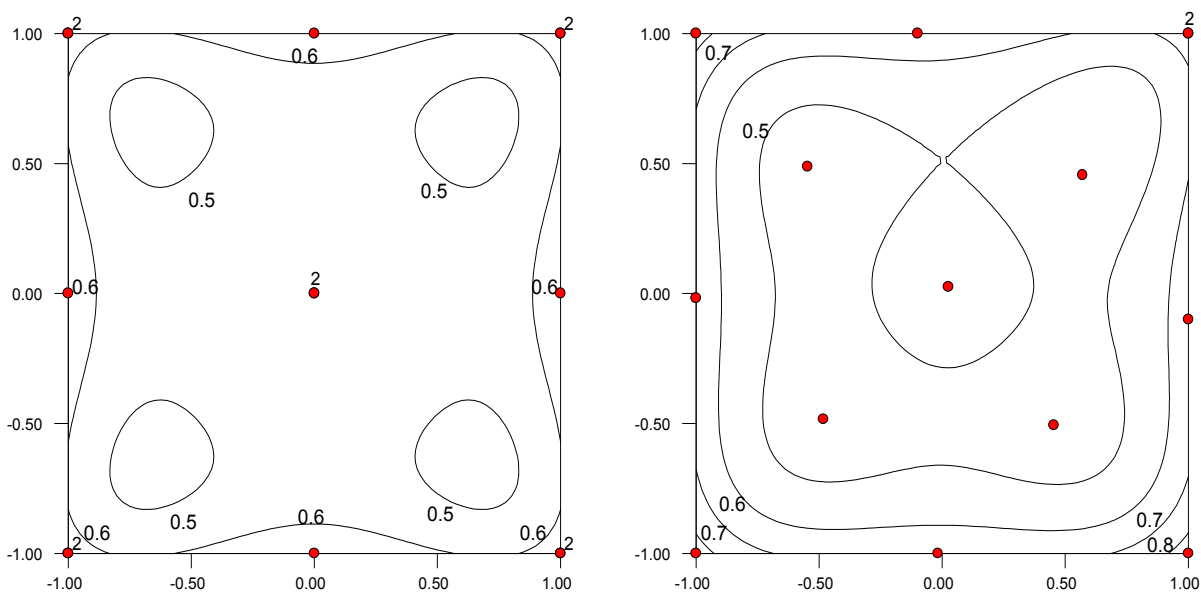


Figure 8. SE plots for purely \underline{D} -optimal (left) versus LOF-modified D-optimal (right) for a 14-run, two-factor RSM design.

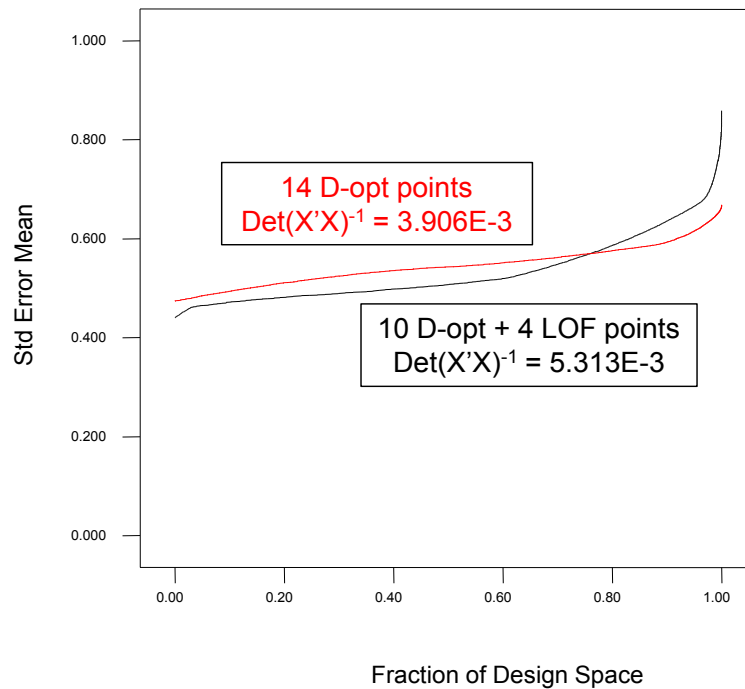


Figure 9. FDS plots for purely D-optimal versus LOF-modified D-optimal for a 14-run, two-factor RSM design.

This lends further support for the use of I over D-optimal designs for response surface method optimization.

Bolstering the Design with Replicates

Replicates are needed to provide the pure error needed for a LOF test. In the cases presented above they came with the optimal portion of the designs. A better way to build a practical experiment is to detail the number of replicates and choose them by the optimal criterion—we recommend a minimum of 4 points being selected in this manner. This provides enough degrees of freedom for a reasonably-powerful LOF test.

Conclusion

In consideration of the practical aspects of algorithmic design for RSM we recommend adding by distance-base at least 4 points for testing lack of fit, even though this makes the experiment less alphabetically optimal. This a good trade off. Furthermore, in physical experiments it is desirable build in an estimate of experimental error based on at least 4 degrees of freedom via replicated design point(s). For choice of optimal criterion we advise using I- for empirical modeling (RSM)—reserving D-optimal (because of their more-precise estimation of model coefficients) for factorial screening.

However, even with these practical aspects incorporated, optimal design cannot make up for:

- Choosing the factors that are not among the vital few.
- Designing for an inadequate empirical model.
- Going out of the region of operability.
- Measuring the wrong responses.

“The exact functional relationship is usually unknown and possibly unknowable. We have only to think of the flight of a bird, the fall of a leaf, or the flow of water through a valve to realize that we are likely to be able to approximate only the main features of such a relationship.” – Box & Draper

References

- Anderson, M. J. and Whitcomb, P.J. (2007) RSM Simplified, *Optimizing Processes Using Response Surface Methods for Designs of Experiments*, Computer-Generated Optimal Designs,” 7: 149-150, New York: Productivity Press.
- Anderson-Cook, C.M., Borror, C. M. and Montgomery, D. C. (2009) “Response surface design evaluation and comparison,” *Journal of Statistical Planning and Inference*, 139: 629–641.
- Box, G. E. P., and Wilson, K B. (1951) “On the experimental attainment of optimum conditions,” *Journal of the Royal Statistical Society, Series B* 13:1-45.
- Box, G. E. P. (1982) “Choice of Response Surface Design and Alphabetic Optimality,” *Utilitas Mathematica*, 21B: 11-55.
- Box, G. E. P., and Draper, N. R. (2007) *Response Surfaces, Mixtures, and Ridge Analyses*, “The Alphabetic Optimality Approach,” 14.4: 471-481, New York: Wiley.
- Draper, N. R. and Guttman, I. (1988) “Rationalization of the “Alphabetic-Optimal” and “Variance Plus Bias” Approaches to Experimental Design,” Technical Report 841, Department of Statistics, University of Wisconsin.
- Kiefer, J. (1959). “Optimum Experimental Designs,” *Journal of the Royal Statistical Society, Series B* 21: 272–319.
- Myers, R. H., Montgomery, D. C. and Anderson-Cook, C.M. (2009) *Response Surface Methodology, 3rd Edition*, New York: Wiley.