DOE

The First Fork in The Road

The choice between model-based or space-filling designed experiments

by Christine M. Anderson-Cook and Lu Lu

Suppose you're planning to run a designed experiment and want to know how to get started. This is a common theme for many studies and applications. The benefits are well-documented for strategically manipulating the inputs to collect the right data and be able to establish causal connections between changes in the responses and the input settings.

Compared to observational studies in which you get what you get, a designed experiment provides user control, allows more informative data to be collected and facilitates an intentional analysis tailored to match the study's objectives.

But when you look through introductory design of experiments (DoE) textbooks[1, 2] or software,[3, 4] there are so many different choices for design construction. How do you choose which type of design is right for your experiment?

An initial decision that divides different designs into two major categories is whether to select the design based on an assumed underlying model, or whether to assume less and choose a space-filling design. The two choices—model-based designs or space-filling designs—take fundamentally different approaches to how the input combinations are chosen and the resulting geometry in the input space.

- **Model-based designs** use design construction criteria based on a user-selected assumed model form. They might seek to have good estimation of model parameters (D- and A-optimality) or good prediction throughout the input space (I- and G-optimality) for the assumed model.
- **Space-filling designs** take a simpler approach and make fewer assumptions about the relationship form. They seek to have good performance with the spacing or coverage of the experimental runs throughout the input space, with no built-in connection to the response. Common choices of criteria for these designs include minimizing the worst (biggest) distance between a run and any point in the input space (minimax), or maximizing the distance between the closest two runs in the design (maximin).[5]

Figure 1 (p. 50) shows a sample of each of the model-based and space-filling designs in which the budget for the experiment is 10 runs, and there are two inputs to manipulate.

- Figure 1(a) shows a response surface design[6] based on assuming a second-order polynomial will be adequate to describe the relationship between the inputs and the response(s) throughout the input space of interest.
- Figure 1(b) shows a Latin hypercube space-filling design[7] of the same size in the same input space.

Even a casual examination of these two designs shows different priorities and strategies for where the experimental runs are placed. The model-based design emphasizes runs at the edge of the input space and often has fewer distinct levels of each input. The space-filling design has many more levels of each input and spreads the points evenly with inclusion of interior and edge points. How do you choose?

Decisions, decisions

Both designs can be the right choice, depending on what you assume the nature of the relationship is between the inputs and responses. As noted in an earlier column,[8] if there's information or knowledge about the process under study, it is helpful to use it. Model-based designs are predicated on the belief that the input-response relationship will be "well-behaved" throughout the region of interest.

This means that you expect the curve or surface between the inputs and response will be continuous and smooth, and it can be approximated well by a simple curve of the assumed functional form. In response surface method literature, the most common assumption is that a second-order Taylor series approximation will be adequate to capture the feature of the curve throughout the input region of interest. Note that this does not suggest that this approximation also will necessarily be adequate outside the region of the experiment—the assumption is just for the design region selected. This more-local focus takes pressure off such a simple approximation working well globally.

There are many relationships across broad classes of applications for which this assumption is reasonable and works extremely well. For those cases, model-based designs are ideal for efficiently getting the data to characterize the underlying relationship. Another assumption implicit with the model-based design is that you are quite confident in obtaining data at each of the input combinations selected. The goodness of the design hinges on having complete data for the experiment because the calculation of the properties of the design are highly dependent on the interdependence of the runs.

A different approach

Space-filling designs take a different approach, with fewer assumptions being made. These designs work well for broader input-response relationships, possibly with discontinuities and unknown degrees of wiggliness. They also tend to be more robust to lost observations if a certain region of the input space isn't viable.

Figure 2 (p. 51) shows simple examples of the response surfaces suitable for model-based and space-filling designs, respectively, in the simple scenario with a single input and one response.

- The three blue curves in Figure 2(a) show that the simple curves for which model-based designs are intended still represent a flexible, adaptable class of relationships.
- Figure 2(b) shows an example of a discontinuous curve (pink) and one with elaborate wiggles in some portions of the input range (blue).

Depending on which types of input-response curves you think are possible for the experiment, this should guide you toward one or the other class of designs.

Time to link

Now, link what you see in the designs of Figure 1 with the relationship you want to model in Figure 2. For model-based designs, the initial assumption that you know what class of curves are suitable for modeling the input-response relationship is a great simplification. You can structure the choice of what data to collect based on this choice. If you're right, you can gain great efficiency.

Hence, in Figure 1(a), each of the two inputs are explored at only three levels, and the points are placed as far apart in the input space as possible to maximize the quality of your model parameter estimates for the assumed second-order model. The experimenter also has implied that experimental runs can be placed confidently at the edges of the input space, with eight of the 10 runs on a corner or edge of the input space in this design.

Using a space-filling design allows the experimenter to make fewer assumptions about what to expect for the response, and so the exploration of the input space is much more deliberate.

To not miss a feature, space-filling designs try to place points throughout the input space and allow local connections between observed responses to be made.

In addition, with the fewer assumptions, the experimenter is not placing as many design points on the edges of the input space for optimizing estimation efficiency and instead saving more points for a thorough exploration of the interior space. Only four of the 10 runs in Figure 1(b) are on an edge, with none being placed in a corner. To allow extra flexibility on the model forms that could be explored, the space-filling designs tend to sacrifice on the precision of the estimated response surface for any particular model.

What if you're wrong?

What happens if you're wrong about what you assumed about the underlying model form? Clearly, the two approaches tackle creating the design for the experiment with different priorities and are based on different assumptions. Suppose that you assume the input-response relationship can be well-characterized with a simple polynomial, and that later turns out to be incorrect. In this case, the model-based design with its smaller number of levels for each input might lead you to miss an interesting feature (a wiggle or a discontinuity) and get no warning that it has happened.

Online Figure 1, which can be found on this column's webpage at qualityprogress.com, shows a sample of what information would be obtained by looking only at three levels across the input range. The pink and blue dots show what the experimenter would learn about the values of the response from the data. Clearly, there is more going on with each of the relationships than seen from the obtained data.

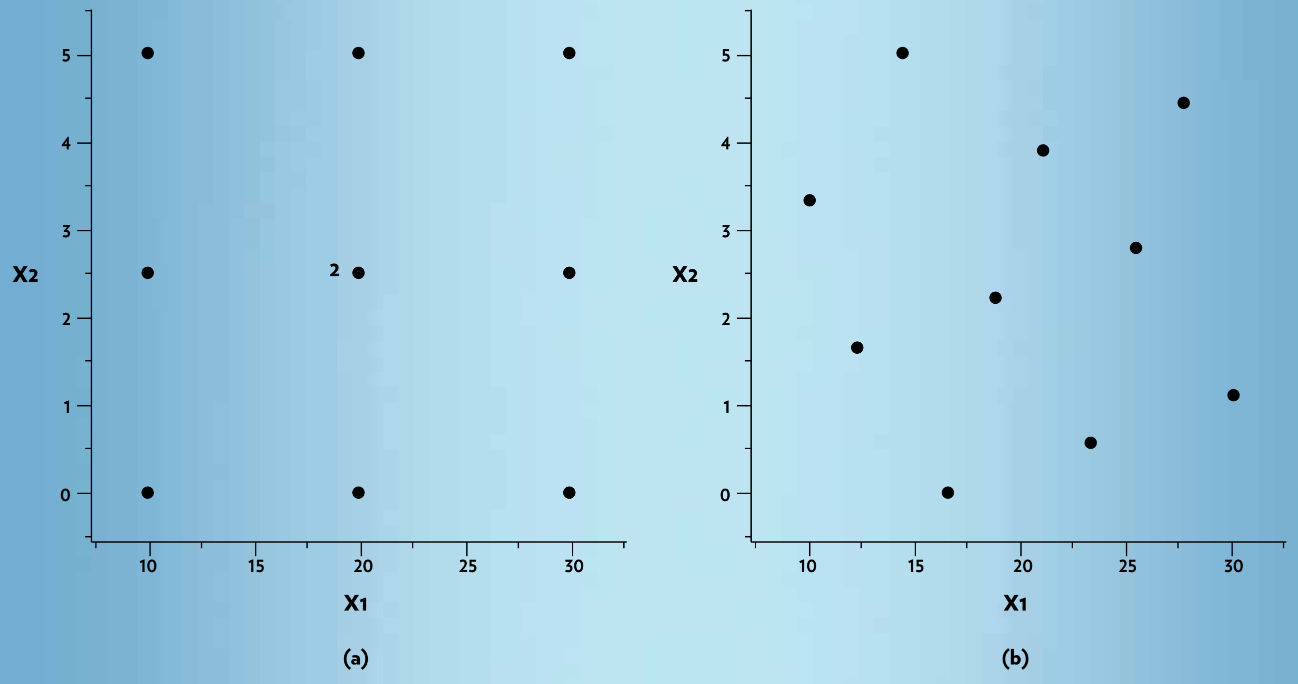
Instead of assuming a known form for the input-response relationship, suppose you're more cautious with a space-filling design. In this case, what's lost is efficiency. You're sampling different levels of the input across its range, some of them not providing as much information for estimating the model parameters if the model is in fact of the simpler form.

Perhaps a good analogy is to think about trying to travel from point A to point B in the dark (you're usually making these design choices before having a lot of information).

Scenario one (the parallel to the model-based design) assumes the path is smooth and well-behaved. You stride out confidently toward the destination. If there are no unexpected obstacles, you can reach point B quickly. If there are

FIGURE 1

Two 2-factor 10-run designs

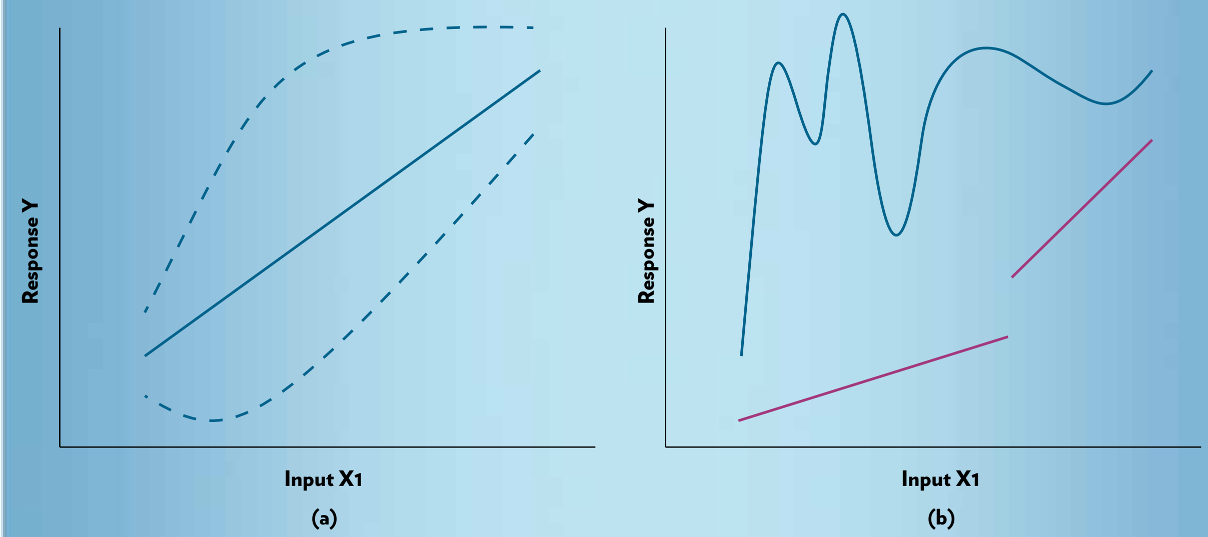


(a) represents a central composite model-based design.

(b) represents a Latin hypercube space-filling design.

FIGURE 2

Types of relationships connecting a single input, X1, with a response, Y



(a) shows smooth, continuous curves suitable for model-based designs using a second-order polynomial.
(b) shows discontinuous (pink) or wiggly (blue) curves that are better suited for space-filling designs.

unexpected obstacles, you might trip or bash your shins. Scenario two (the parallel to space-filling designs) assumes that the path is unknown. You inch along cautiously, expecting something unpredictable. You reach your destination more slowly, but also have lower risk of injury or a nasty surprise.

Is there a compromise that can help you reach the destination smoothly and efficiently, given that the process often is started with quite incomplete knowledge of the underlying input-response relationship? As noted in a column last year,[9] sequential DoEs allow you to divide your experimental budget into smaller increments. This facilitates building knowledge piece-by-piece and leveraging what's learned at each stage. If you're really starting the experiment with little knowledge about what to expect from the responses, a small space-filling design can help build understanding about what the relationship looks like.

If results look well-behaved, maybe you can switch to a model-based design in the second phase of the experiment to increase efficiency. If things don't seem to match the assumption of a simple curve that is well-modeled by a low-order polynomial, augment the first design with another space-filling design, maybe more focused on newly identified regions of interest based on what was learned from the earlier stage.

Remember that the choice between model-based and space-filling designs is just the first of many important choices that must be made when constructing a design. Within each of these classes, there are many choices that can be tailored to match the goals of the experiment. But ensuring the journey's first steps are in the right direction makes a big difference for the overall success of the trip. **QP**

EDITOR'S NOTE
References listed can be found on the column's webpage at qualityprogress.com.

Christine M. Anderson-Cook is a retired research scientist in the Statistical Sciences Group at Los Alamos National Laboratory. She is a fellow of both ASQ and the American Statistical Association.

Lu Lu is an associate professor in the department of mathematics and statistics at the University of South Florida in Tampa. She is a member of ASQ and the American Statistical Association.