DESIGNING SIMULATION EXPERIMENTS

Russell R. Barton

The Smeal College of Business Administration
The Pennsylvania State University
University Park, PA 16802, U.S.A.

ABSTRACT

Simulation models provide relatively fast and inexpensive estimates of the performance of alternative system configurations and/or alternative operating procedures. This tutorial provides some techniques for planning a set of simulation model runs, in order to gain insight on system behavior. There is an emphasis on graphical methods for planning the experiment and displaying the results.

1   INTRODUCTION

Discrete-event simulation modeling is a popular method for predicting the performance of complex systems, particularly systems that include random phenomena. Simulation projects can fall short of their intended goals, however, unless the simulation model is exercised intelligently to gain useful understanding of the likely performance of the real system.

This is where the design of simulation experiments plays a key role. Usually, simulation projects are conducted within time and budget limits. Often the bulk of time and resources are spent on building and validating the model, with little remaining to exercise the model for decision-making insight. This is risky, since poorly planned simulation runs can result in a significant loss of information, or worse, provide misleading results.

This tutorial presents a five-step process for the design of a simulation experiment. Graphical methods are emphasized for the first four steps, drawing largely from Barton (1999). A hypothetical simulation project for a die-making machine shop will help to illustrate each step.

The tutorial largely follows that in Barton (2002). The next section describes the limits of the topics covered, defines the five-step process, and describes the machine shop scenario. Sections 3-7 describe each step in more detail and illustrate the activities for the machine shop simulation. The next section shows how to use the graphical framework to present results. In some cases the graphical display provides more insight than an ANOVA table or regression analysis. Section 9 covers some remaining issues.

2   FOCUS OF THE TUTORIAL

Exactly what activities are considered as part of the design of an experiment? Consider the overall process of scientific investigation. Generally, it is a repeating, cyclic process which can be broken down into several activities:

A.   State a hypothesis to be evaluated.
B.   Plan an experiment to test the hypothesis.
C.   Conduct the experiment.
D.   Analyze the data from the experiment. This will likely lead to modification of the original hypothesis, and a return to activity one for the next cycle.

This tutorial focuses on activity B. Typically, simulationists (and the simulation methodology literature) emphasize C and D, but careful planning in B can simplify the remaining activities. Activity B is what we will call the design of the simulation experiment (DOE), although many texts consider B, C, and D together under this topic.

For people without statistical training, it can be difficult to organize information about the system under study in a way that aids the design of the experiment. To help clarify this process, we break the design task B into five separate steps.

1.   Define the goals of the experiment.
2.   Identify and classify independent and dependent variables.
3.   Choose a probability model for the behavior of the simulation model.
4.   Choose an experiment design.
5.   Validate the properties of the chosen design.

The next sections describe each step in detail. To make these activities more concrete, they will be described in the context of a hypothetical simulation project.

Imagine that you have constructed a simulation model of a machine shop that specializes in making dies for stamping parts. You are particularly interested in the operation of the milling machines, and in how jobs are released to the shop floor. You would like to choose policies

that allow the shop to operate efficiently. You want to examine how shop operation is affected by release policies, the schedule for preventative maintenance, the speed/feed of the milling machines for a particular family of dies, and the production lot size. The example is described in more detail as we cover the five steps of the DOE process.

3   DEFINE GOALS

Of course, the selection of what conditions to run in a simulation experiment depends on the goals of the experimenter. Why was the simulation model constructed? What particular issues are being examined during the current cycle of experimentation? These goals generally fall at a middle level in the hierarchy of goals within the organization that is sponsoring the development and use of the simulation model. It is good to place these goals in perspective, to gain support for the effort that will be required, and to make sure that the short-term objectives are consistent with the overall goals of the organization.

Goal hierarchy plots provide a graphical means to show an experiment's impact on higher-level goals, and to identify resources that will be needed to conduct the simulation study. These plots were developed as part of a special decision-making procedure called the Analytic Hierarchy Process (Saaty 1980).

Figure 1 shows a goal hierarchy plot for the machine shop study. At the highest level, one goal of the company is to grow its share of the die market. Two sub-goals to help accomplish this are to lower the cost of the dies, and to lower the average cycle time for producing a lot of dies. In order to achieve lower cost, the experimenters need to understand factors that affect throughput and operating costs. To lower the cycle time, the experimenters need to understand the impact that various factors have on cycle time. Both of these needs can be met through experimental studies using a simulation model of the shop. Completing such studies requires a validated simulation model and a planned set of model runs. The validated simulation model requires construction of a model and a planned set of runs to validate its performance.

The figure highlights the repeated cycle of experimentation: the simulation model must be validated through a preliminary experiment before conducting the experiment to examine the impact of lot size, release policy, etc. on cost, throughput and cycle time. Table 1 classifies the experiment goals in one direction: the stage of the scientific process. At the earliest stage, activities focus on validation. Next, one often seeks to identify the most important design or policy variables affecting system performance. The next cycle often involves experiments to understand in a quantitative and predictive way how design or policy variables affect system performance. In some cases this level of understanding is sufficient for decision-making. In other cases the simulation is exercised repeatedly to
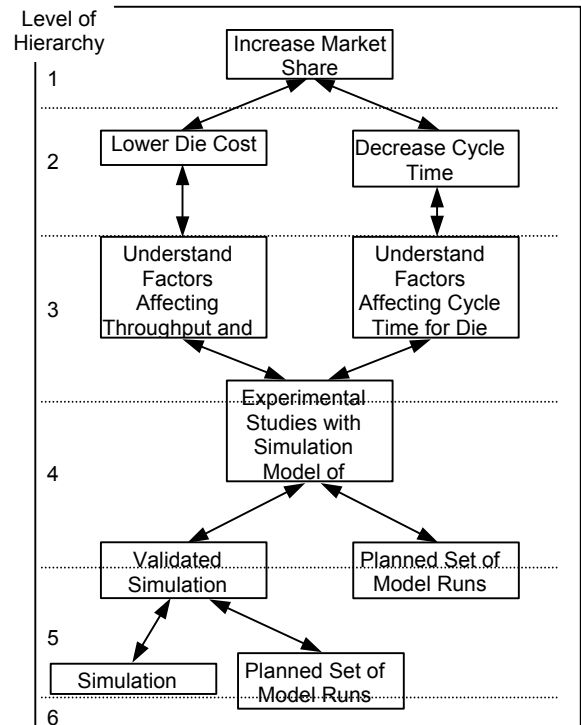

Figure 1: Goal Hierarchy Plot for the Machine Shop Study

optimize some measure of system performance. See Kleijnen and Sargent (2000) for a similar structuring of goals.

This tutorial focuses on the third and fourth cycles of goals in Table 1. Specific activities for validation are described in the Verification, Validation and Accreditation Track of the *Proceedings of the 2000 Winter Simulation Conference* (Joines et al. 2000) and the references cited in these papers. Screening designs are described in standard texts on the design of experiments such as Montgomery (1997), and in papers by Lin (1995), Trocine and Malone (2000) and Wan, Ankenman and Nelson (2003) and references therein. Optimization and robust design are frequently topics of tutorials at the Winter Simulation Conference. See the papers by Fu et al. (2000) and Sanchez (2000) to find additional information and references.

Table 1: Goals by Cycle of the Investigation

Cycle	Goal
1. Early	Validation
2. Early	Screening
3. Middle	Sensitivity Analysis, Understanding
4. Middle	Predictive Models
5. Late	Optimization, Robust Design

For our machine shop example, we will focus on the fifth level of the goal hierarchy plot. The goal for the experiment is to construct predictive models of throughput, cy-

cle time and operating costs, the fourth cycle in Table 1. The form of the predictive probability model that will be fitted is

$$Y = \beta_0 + \beta_1 g_1(x_1, x_2, ..., x_d) + ... + \beta_p g_p(x_1, x_2, ..., x_d) + \varepsilon, \quad (1)$$

where ε are independent, normal random quantities with mean zero and unknown variance. The x's are independent variables that are identified using the techniques in the next section. The appropriate g functions can be determined using the techniques in Section 5. For some simulation outputs, it may not be reasonable to assume that the random variation in the output performance measure will be normally distributed. When the output measure is an average or cumulative quantity over time, a form of the Central Limit Theorem often applies, making a normal distribution a reasonable approximation for ε.

The goal of the experiment (at hierarchy level 4) is to provide estimates of the unknown β_i coefficients in the model (1), as well as an estimate of the variance of ε. For the machine shop study, there are three such models: one for cycle time, one for operating costs, and one for throughput. They do not necessarily have the same factors or same kinds of terms.

4 IDENTIFY AND CLASSIFY VARIABLES

The second step in the experiment design process is to identify quantities in the simulation that can be set to desired values (*independent* variables) and the resulting system performance measures that are of interest (*dependent* variables).

There are two other classes of variables to be considered when designing the experiment. *Nuisance* variables are known to affect the behavior of the system, but cannot be controlled directly. These are rarely present in simulation, where all factors are generally under the user's control.

The fourth type of variable is an *intermediate* variable. Intermediate variables cannot be controlled independently: they are affected by the settings of the independent variables. They are not considered dependent variables, however, if there is no interest in their value except as it affects an important performance measure. For example, the average number of unscheduled maintenance operations per month will affect operating costs for our shop. This quantity cannot be set independently: it will depend on the speed and feed rates used for the milling machines and on the preventative maintenance policy, among other things. It is not of direct interest, since we care ultimately about cycle time, throughput, and operating costs.

It is important to identify all variables of all four types before planning the set of runs. Intermediate variables must be recognized so that they are not mistakenly included as independent variables. Nuisance variables must be monitored so that random variation in the experiment results can be understood. *ALL* independent variables should be identified, not just the ones that will be varied in

the experiment. In order for you (or others) to be able to reproduce your results at a later date, you must record the (fixed) values of any independent variables that you did not adjust, as well as the values of ones that were varied. The *held-fixed* variables have a way of changing over time, as the simulation model is run and rerun for different purposes. Independent variables whose values are actually changed during the experiment will be called *factors*.

Dependent variables are determined by the objectives of the study. For our example, they appear in the goal hierarchy plot in Figure 1: cycle time, throughput, and operating costs. Independent variables are harder to identify. Process diagrams (IDEF0) and cause-effect diagrams can be used to identify them. We will illustrate the cause-effect diagram here. For examples of IDEF0 process diagrams, see Barton (1999).

Figure 2 shows a cause-effect diagram for throughput, one of the dependent variables in the study. A similar diagram must be constructed for each dependent variable. The diagram shows a chain of cause-effect relations. At the end of each chain is the dependent variable. At the beginning of each chain is a root cause, which may be either an independent variable (if it can be controlled) or a nuisance variable.
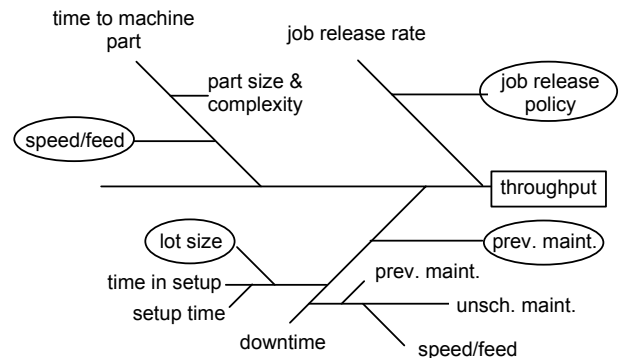

Figure 2: Cause-Effect Diagram for the Machine Shop Study

Independent variables and nuisance variables appear as lines with no lines impinging on them: if the diagram is a tree, they are the leaves. The independent variables in ovals are the ones that will be varied in this study. *Setup time* is assumed to be fixed at a particular value for this study. Only *part size and complexity* remains. It might be considered either a nuisance variable, if we wish to model a random mix of die types, or a held-fixed independent variable, if we wish to consider only certain fixed mixes of parts in our production schedule.

Intermediate variables appear as branches. Time in unscheduled maintenance will depend on the speed/feed of the tools and on the preventative maintenance policy, for example.

Table 2 shows the dependent variables for our machine shop study, and the independent variables associated

with each of them. In parentheses are the ranges of interest for each independent variable. These ranges are determined by practical limits or by policy decisions.

Table 2: Dependent Variables and Associated Independent Variables and Ranges for the Machine Shop Study

Dependent	Independent
throughput	job release policy (1, 2), lot size (10-30 pieces), prev. maint.(2-8 hours), speed/feed (0.1 - 0.5 inch/second)
cycle time	job release policy (1, 2), lot size (10-30 pieces), prev. maint.(2-8 hours), speed/feed (0.1 - 0.5 inch/second)
operating costs	prev. maint.(2-8 hours), speed/feed (0.1 – 0.5 inch/second)

Next, we need to identify, in a qualitative way, whether we expect a linear or nonlinear relationship. This will determine what g functions we will need in the probability model (1).

5 CONSTRUCT A PROBABILITY MODEL

This step is closely linked to step one of the overall process of scientific investigation: define the hypothesis to be tested. Before we can choose a set of simulation runs, we need to know the form of the model (1) that will be fitted and tested. That means we need to know not only which x's but also which g's, as well as something about the variance of the ε's. Often the g functions are just power and cross-product terms of the form x_i, $x_i x_j$, x_i^2, $x_i x_j x_k$, $x_i x_j^2$, and so forth. These forms are supported to some extent by Taylor's Theorem, which shows that polynomial functions (power and cross-product terms) provide good *local* approximations to any smooth response function.

Two kinds of graphs can help us identify the kind of terms to be included in the probability model. A-priori main effect plots help to identify g function power terms of the form x_i, x_i^2, x_i^3, and so forth. A-priori interaction plots help to identify terms of the form $x_i x_j$, $x_i x_j x_k$, $x_i x_j^2$, and so forth. Space and time limitations restrict this presentation to main effect plots. See Barton (1999) for details on constructing and interpreting a-priori interaction plots.

Figure 3 shows a set of four a-priori main effect plots for the hypothesized effect of each factor on throughput. Plots that are roughly linear *over the range of interest* require only an x_i term. Plots with curvature may require an x_i^2 term in addition, and plots with changing curvature *over the range of interest* may require x_i^3 and higher terms as well.

Since we are only considering two candidate job release policies, there are only two discrete choices. We hypothesize that job release policy 2 will provide greater throughput, although this aspect of the hypothesis is not critical to the design chosen in Section 6. The speed/feed variable can be expected to have a linear impact on through-
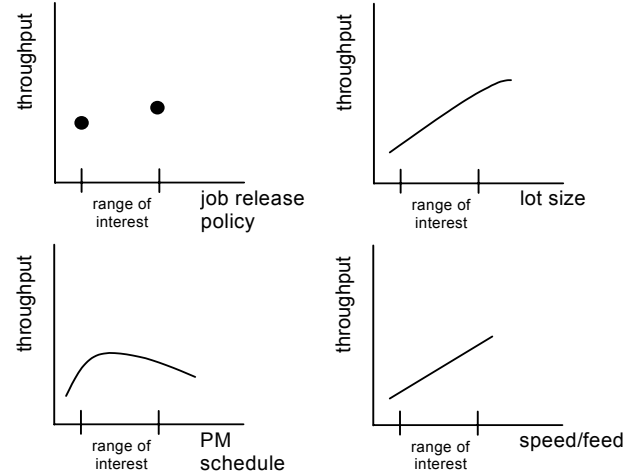

Figure 3: A-Priori Main Effect Plots for Throughput

put. The figure shows an expected increase in throughput as lot size is increased, with diminishing returns. Over the region of interest, however, the figure suggests that a linear approximation should be adequate. The preventative maintenance schedule's impact on throughput is to reduce throughput for too-frequent PM times, and to again reduce throughput (due to frequent unscheduled maintenance) if the time between PMs is too long. In this case, curvature occurs in the region of interest, and so linear and quadratic terms for the impact of PM will be included in the model.

Two comments about these plots are in order. First, there are no scales on the vertical axes. These plots are qualitative, not quantitative. We do not know the quantitative relationships in advance: that is why we are conducting the simulation experiment! Second, the qualitative forms need not be correct. After all, they are only our guesses. These plots provide a description of the hypotheses that we will test in this cycle of the scientific investigation.

An assessment of these figures, along with the a-priori interaction plots (not shown) leads to the hypothesized model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_3^2 + \beta_5 x_4 + \beta_6 x_3 x_4 + \varepsilon, \quad (2)$$

where x_1 is job release policy (1 or 2), x_2 is lot size (units), x_3 is PM schedule (hours), and x_4 is speed/feed (in./second), and ε has a normal distribution with unknown variance. The only interaction term that appears is due to a change in the impact of the PM schedule depending on speed/feed. Similar plots would have to be constructed for the dependent variables cycle time and operating costs.

6 CHOOSE AN EXPERIMENT DESIGN

In this activity, one determines the number of distinct model settings to be run, and the specific values of the factors for each of these runs. There are many strategies for selecting the number of runs and the factor settings for each run. These include random designs, optimal designs,

combinatorial designs, mixture designs, sequential designs, and factorial designs.

Factorial designs are based on a grid, with each factor tested in combination with every level of every other factor. Factorial designs are attractive for three reasons: i) the number of levels that are required for each factor are one greater than the highest-order power of that variable in the model, and the resulting design permits the estimation of coefficients for all cross-product terms ii) they are probably the most commonly used class of designs, and iii) the resulting set of run conditions are easy to visualize graphically for as many as nine factors.

The disadvantage of factorial designs is that they require a large number of distinct runs when the number of factors and/or the number of levels of the factors are large. In this case, fractional-factorials are often employed. This section focuses on factorial and fractional-factorial designs.

Figure 4 shows geometric representations for three and five factors each with two levels. The five-factor design shows how additional factors can be incorporated by hierarchically using rectangular frames.
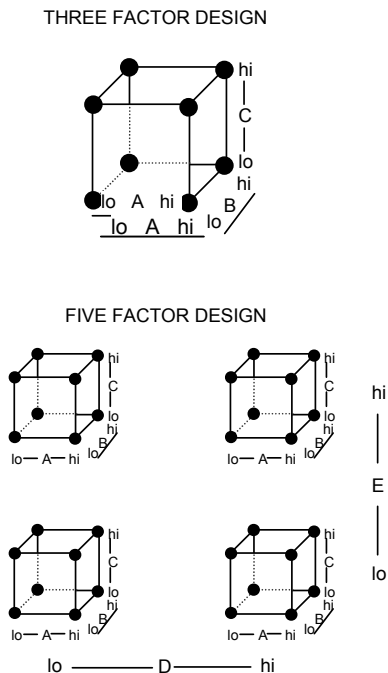

Figure 4: Factorial Designs for Three and Five Factors

Figure 5 shows a candidate design for the machine shop study. Three levels are used for the preventative maintenance schedule, to allow estimation of the quadratic term in (2). There are seven β coefficients in the model (2) and the variance of ε to estimate, so we need at least seven different run conditions, eight if there are no replications (repeated runs with the same factor settings). Replications allow us to check the adequacy of the model, so rather than run each factorial point for a total of 24 runs, we have

chosen 1/2 of the factorial points, and replicated four of those, for a total of 15 runs. Barton (1999) describes geometric characteristics that can be used to guide the selection of a fraction of the full factorial design.
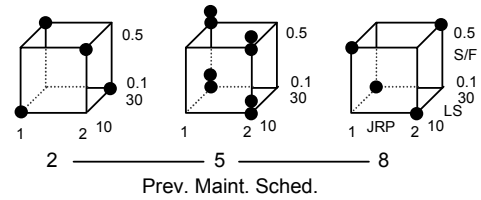

Figure 5: Fractional Factorial Design for the Machine Shop Study

7 VALIDATE THE PROPERTIES OF THE DESIGN

Because this design was selected based on geometric properties, there is no guarantee that it will allow the estimation of all of the terms in the model. A mathematical check is necessary. The mathematical requirements can be found in a design of experiments text such as Montgomery (1997) or in Barton (1999).

The simplest check, however, is to create a random artificial response for *Y* before running the simulation model and proceed with the statistical analysis. If the design is inadequate, the statistical package will inform you that the parameters cannot be estimated.

You can also use this approach to get an idea about whether the number of runs will be sufficient to estimate the coefficients in (1) with adequate precision. In this case, guesses for all of the β's and an estimate of the variance of ε can be used to generate artificial *Y*'s using a spreadsheet and the hypothesized model. If you find a lack of significance for the model terms when you analyze the statistics for the artificial data, you will need to increase the number of replications, or increase the magnitude of the β coefficients that you will be able to detect.

Because we have two other dependent variables, and we would like a single experiment to allow us to fit all three models, we need further checks on the design. It must also be validated for fitting the hypothesized models for cycle time and operating costs.

8 GRAPHICAL DISPLAY OF RESULTS

A graphical design framework provides an added bonus: it can be used to display the results of the experiment. Figure 6 shows the results of the experiment presented in Figure 5, using the same framework. The size of the circle corresponds to the throughput. We see that lot size increases throughput (the estimate for β_2 is positive) and increasing the frequency of preventative maintenance increases throughput (the estimate for β_3 is negative). No other effects are apparent (except for β_0, all other β's are approximately zero).
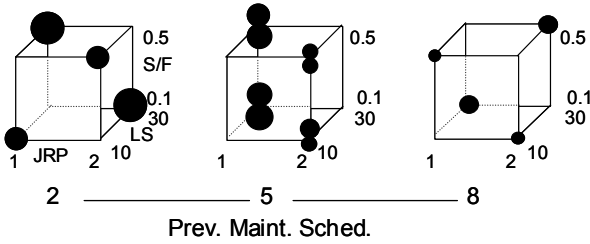
Figure 6: Results of the Machine Shop Experiment

In some cases, this graphical presentation can be more informative than the estimated coefficients of models like (1). Imagine a simple simulation of the percent of on-time completions for a mortgage application processing center. The simulation study involves three factors: the arrival rate of applications (x_A), whether the processing personnel also have telephone-answering responsibilities (x_B) and whether the processing personnel also have copier maintenance responsibilities (x_C).

Figure 7 shows the results of the simulation experiment: the percentage of on-time processing of applications is shown by the diameter of the circle. The results are easy to interpret: processing personnel can handle telephone and/or copier maintenance without affecting performance when the arrival rate of applications is low; performance is degraded only when arrival rate is high and the personnel must also answer phones *and* maintain the copier.
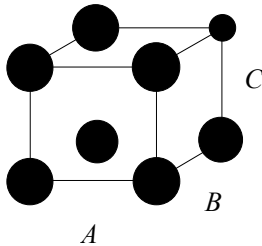

Figure 7: Mortgage Process-
ing Example

Table 3 shows the fitted regression model. Without Figure 7, these results would be hard to interpret. Every possible term has a significant coefficient! In this case the model would have an intercept, three linear terms, three two-factor cross-product terms, and a three-factor interaction term. The insight would be: "Gee, this is a complex response function."

Table 3: Regression Coefficients for the Mortgage Example

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEPT	1	8.054951	0.11733721	68.648	0.0001
XA	1	-0.341090	0.11733721	-2.907	0.0050
XB	1	-0.706990	0.11733721	-6.025	0.0001
XC	1	-0.266921	0.11733721	-2.275	0.0263
XAXB	1	-0.418560	0.11733721	-3.567	0.0007
XAXC	1	-0.521779	0.11733721	-4.447	0.0001
XBXC	1	-0.302551	0.11733721	-2.578	0.0122
XAXBXC	1	-0.557293	0.11733721	-4.749	0.0001

9 ADDITIONAL ISSUES IN PLANNING AND CONDUCTING SIMULATION EXPERIMENTS

This tutorial focused on the planning of run conditions for fitting a probability model. There are a number of other issues that the simulationist faces when planning an experiment.

1. If the simulation model is non-terminating (as was the case for our machine shop example), then the run length (in hours, days, or weeks) must be determined.
2. There is often a tradeoff between run length and replications. Depending on the estimated variance of ε, and the desired precision of the β estimates, multiple replications may be needed.
3. The variability of the performance measure may differ from one set of experimental conditions to another. This may be handled by adjusting run lengths, transforming the dependent variable, or using a weighted least squares method for analysis.
4. One must determine whether an initial transient period must be deleted from each run (and its length) before computing the performance measure.
5. Random number streams must be allocated to different components of the model. This may be done in a non-independent fashion as a means to reduce the variability of the estimates of the model coefficients (β's) in (1).
6. The results of the experiment must be analyzed: statistical techniques must be used to fit the probability model and test hypotheses about its adequacy.
7. Experiments may focus on optimization, requiring a sequential approach to the design and analysis of simulation experiments.

These issues must be addressed to conduct a successful simulation study. For in-depth coverage of these issues, see simulation texts such as Law and Kelton (2000), Kleijnen (1987), Banks et al. (2001), and Banks (1998).

REFERENCES

Banks, J. 1998. *Handbook of simulation: principles, methodology, advances, applications and practice.* ed. J. Banks. New York: John Wiley and Sons.

Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2001. *Discrete-event system simulation.* Upper Saddle River, New Jersey: Prentice-Hall.

Barton, R. R. 1999. *Graphical methods for the design of experiments.* New York: Springer-Verlag.

Barton, R. R. 2002. Designing simulation experiments. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon and J. M. Charnes, 45-51. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Fu, M. C., S. Andradóttir, J. S. Carson, F. Glover, C. R. Harrell, Y.-C. Ho, J. P. Kelley, and S. M. Robinson. 2000. Integrating optimization and simulation: research and practice. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang and P. A. Fishwick, 610-616. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Joines, J. A., R. R. Barton, K. Kang, and P. A. Fishwick. 2000. *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang and P. A. Fishwick. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Kleijnen, J. P. C. 1987. *Statistical tools for simulation practitioners.* New York: Marcel Dekker.

Kleijnen, J. P. C., and R. G. Sargent. 2000. A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research* 120: 14-29.

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling and analysis*, 3rd ed. New York: McGraw-Hill.

Lin, D. K. J. 1995. Generating systematic supersaturated designs. *Technometrics* 37: 213-225.

Montgomery, D. C. 1997. *The design and analysis of experiments,* 4th ed. New York: John Wiley and Sons.

Saaty, T. L. 1980. *The analytic hierarchy process.* New York: McGraw-Hill.

Sanchez, S. M. 2000. Robust design: seeking the best of all possible worlds. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang and P. A. Fishwick, 69-76. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Trocine, L., and L. C. Malone. 2000. Finding important independent variables through screening designs: a comparison of methods. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang and P. A. Fishwick, 749-754. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Wan, H., B. Ankenman, and B. L. Nelson. 2003. Controlled sequential bifurcation: a new factor-screening method for discrete-event simulation. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. J. Sánchez, D. Ferrin and D. J. Morrice, 565-573. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

AUTHOR BIOGRAPHY

RUSSELL R. BARTON is a Professor and Associate Dean for Research and Ph.D./M.S. Programs in the Smeal College of Business Administration at the Pennsylvania State University. He has a B.S. degree in Electrical Engineering from Princeton and M.S. and Ph.D. degrees in Operations Research from Cornell. Before entering academia, he spent twelve years in industry. He is President for the INFORMS College on Simulation. His research interests include applications of statistical and simulation methods to system design and assessment and to product design, manufacturing and delivery. His email address is <rbarton@psu.edu>.