

DESIGNING SIMULATION EXPERIMENTS

Russell R. Barton

Department of Supply Chain and Information Systems
The Pennsylvania State University
University Park, PA 16802, USA

ABSTRACT

Much effort goes into building and validating simulation models. Similar care is necessary in the use of the model to support decision making. Good simulation experiment design is important to get useful and valid insights for specific management questions. This introductory tutorial gives an overview of experiment design techniques for justifying and planning a series of simulation runs to uncover the impact of system design parameters on simulation output performance. Graphical methods are emphasized.

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.

The design of simulation experiments plays a key role. Simulation projects are conducted within time and budget limits, and often the bulk of time and resources are spent on building and validating the model, with little time or budget in the schedule to exercise the model for decision-making insight. This is risky, since poorly planned simulation runs can give results that provide little insight, or worse, provide misleading results. Further, the kinds of decisions the simulation model will support should be decided up-front, since model construction, verification and validation depend on this information (Sargent 2011).

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 is an updated version of that in Barton (2010) with editorial changes and discussion of an experiment conducted by Beeler, Aleman and Carter (2012). Introduction to the design of simulation experiments is presented from different perspectives in the WSC papers by Sanchez and Wan (2012) and in Kleijnen (2008b). 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. Section 8 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. A number of other simulation experiment issues are summarized in Section 9.

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 (e.g., the appropriateness of a particular model).
- 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 A for the next cycle.

This tutorial focuses on activities A and B. Typically, simulationists (and the simulation methodology literature) focus on C and D, but careful conduct of A and 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 tasks A and B into five 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 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 operating policies that allow the shop to perform 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. This 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 do this, and to simultaneously 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 optimize some measure of system performance. See Kleijnen and Sargent (2000) for a similar structuring of goals.

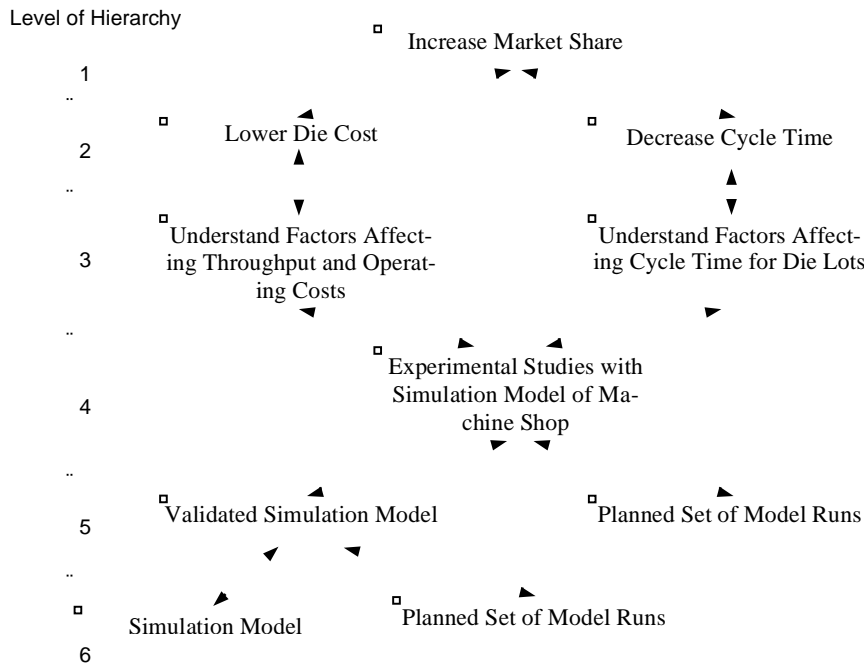


Figure 1: Goal hierarchy plot for the machine shop study

Specific activities for validation are described in Law (2009) and Sargent (2011) and the references cited in these papers. Screening designs are described in standard texts on the design of experiments such as Montgomery (2009), and in papers by Lin (1995) and Sanchez, Wan and Lucas (2009) and references therein. Optimization and robust design are frequently topics of tutorials at the Winter Simulation Conference, but are too advanced for this tutorial. This tutorial has its focus on the third and fourth cycles of Table 1.

For our machine shop example, experiment goals connect the third through fifth levels of the goal hierarchy plot in Figure 1. The experiment will enable predictive models of throughput, cycle time and operating costs, the fourth cycle in Table 1. The form of the predictive probability model that will be fitted is

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

where ε are independent, 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.

Table 1: Goals by cycle of the investigation

Cycle	Goal
1. Early	Validation
2. Early	Screening Variables
3. Middle	Sensitivity Analysis, Understanding
4. Middle	Predictive Models
5. Middle	Selecting the Best Configuration
6. Late	Optimization, Robust Design

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 our example experiment 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 independent variables 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 (1997, 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 paths of cause-effect relations. At the end of each path (on the far right) is the dependent variable, in this case, *throughput*. At the beginning of each path is a root cause, which may be either an independent variable (if it can be controlled) or a nuisance variable.

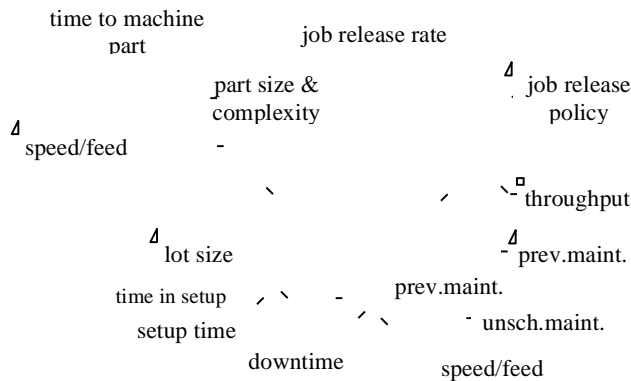


Figure 2: Cause-effect diagram for the machine shop study

Independent variables and nuisance variables appear as lines with no lines impinging on them: if the diagram were a tree, they would be leaves. The independent variables in ovals in Figure 2 are the ones that will be varied in this study. *Preventative maintenance* appears twice in the diagram but only need be circled once. *Setup time* is another independent variable but is not circled since it will be fixed at a particular value for this study. The leaf *part size and complexity* remains to be specified. 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 a certain fixed mix 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.

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

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)
-----------------	---

5 CONSTRUCT A PROBABILITY MODEL

This step is closely linked to step one of the overall process of scientific investigation: state a hypothesis to be evaluated. 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 (1997, 1999) for details on constructing and interpreting a priori interaction plots. The main idea is to qualitatively imagine the a priori expected relationship between each independent variable and a dependent variable, e.g, throughput.

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 throughput. 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.

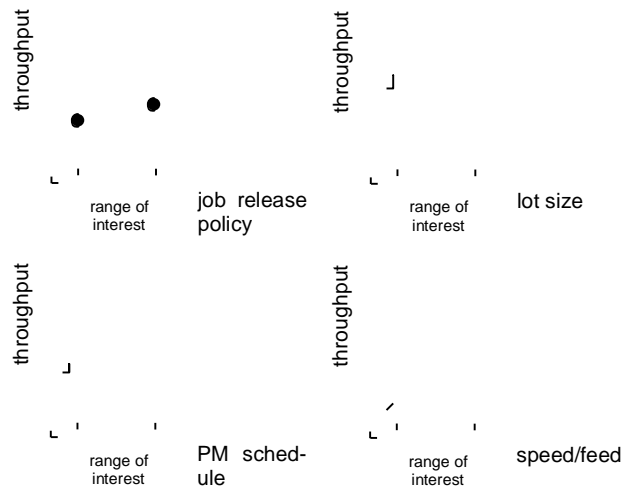


Figure 3: A priori main effect plots for throughput

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 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 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 (1998, 1999) describes geometric characteristics that can be used to guide the selection of a fraction of the full factorial design.

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 (2009).

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 not adequate, 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 (2) 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. Finally, given that things can go wrong in conducting the various experiments, the design should be robust to missing observations, unequal variance across runs, and unexpected terms in the model. These topics are beyond the level of this introductory tutorial.

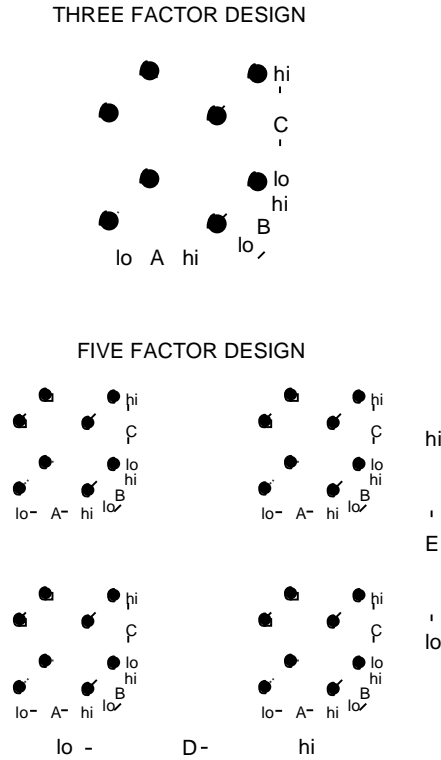


Figure 4: Factorial designs for three and five factors

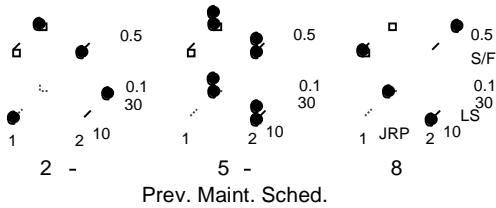


Figure 5: Fractional factorial design for the machine shop study

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).

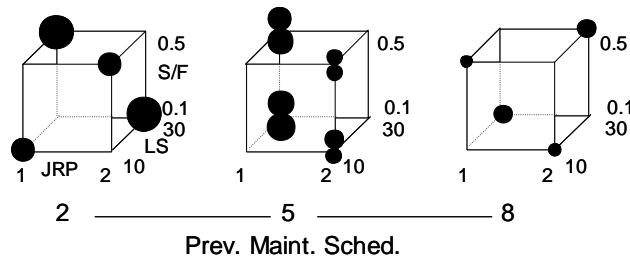


Figure 6: Results of the machine shop experiment

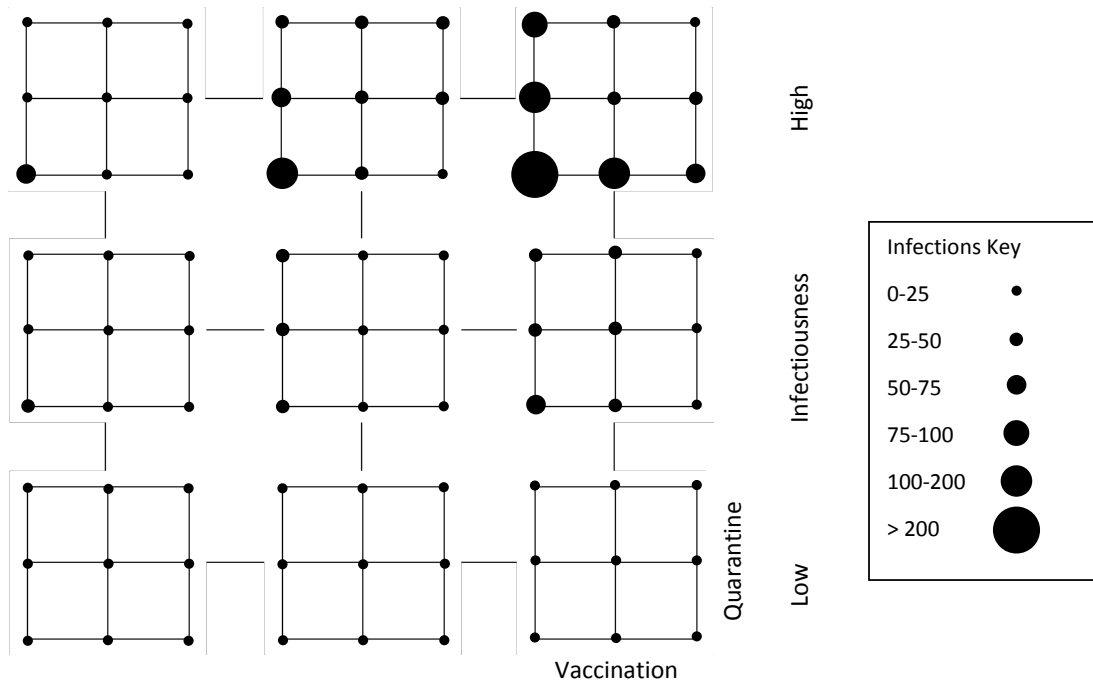


Figure 7: Influenza study in Beeler, Aleman and Carter (2012)

This graphical presentation can be more informative than the estimated coefficients of models like (1). Consider the simulation experiment on influenza pandemic behavior in Beeler, Aleman and Carter (2012). The results of the full factorial design are not available, but can be approximated from the significant main effect and two- and three-factor interaction coefficients given in the paper. Figure 7 shows the results in the graphical framework. The results are easy to interpret: no vaccination and no self-quarantine

is especially bad, particularly when infectiousness is high, and a high level of vaccination results in a low number of infections when infectiousness is low or medium, regardless of the number of contagious days. High levels of vaccination and self-quarantine are necessary to curb infection if both infectiousness and days contagious are high. Compare this with the coefficients in Table 1 of their paper – 37 significant coefficients! Without Figure 7, these results would be hard to interpret.

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.
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.
5. Random number streams must be allocated to different components of the model.
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, so a sequential design approach may be preferred.

These issues must be addressed to conduct a successful simulation study. For in-depth coverage of these issues, see simulation texts such as Law (2007), Kleijnen (2008a), and Banks et al. (2009).

ACKNOWLEDGMENTS

This tutorial is based on the prior work of many researchers. I have borrowed significantly from the work, advice, and/or philosophies of Stu Hunter, David Kelton, Jack Kleijnen, Barry Nelson, Susan Sanchez, Bob Sargent, Lee Schruben, and Jim Wilson. I thank Russell Cheng and Barry Nelson helpful reviews of this tutorial.

REFERENCES

- Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2009. *Discrete-Event System Simulation*, 5th ed. Upper Saddle River, New Jersey: Prentice-Hall.
- Barton, R. R. 1997. Pre-experiment planning for designed experiments: graphical methods. *Journal of Quality Technology* 29, 307-316.
- Barton, R. R. 1998. Design-plots for factorial and fractional-factorial designs. *Journal of Quality Technology* 30, 40-54.
- Barton, R. R. 1999. *Graphical Methods for the Design of Experiments*. New York: Springer-Verlag.
- Barton, R. R. 2010. Simulation experiment design. In *Proceedings of the 2010 Winter Simulation Conference*, ed. B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 75-86. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Beeler, M. F., D. M. Aleman and M. W. Carter. 2012. A large simulation experiment to test influenza pandemic behavior. In *Proceedings of the 2012 Winter Simulation Conference*, ed. C. Laroque, J.

- Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Kleijnen, J. P. C. 2008a. *Design and Analysis of Simulation Experiments*. New York: Springer.
- Kleijnen, J. P. C. 2008b. Design of experiments: overview. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, 479-488. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- 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. 2009. How to build valid and credible simulation models. In *Proceedings of the 2009 Winter Simulation Conference*, ed. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, 24-33. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Law, A. M. 2007. *Simulation Modeling and Analysis*, 4th ed. New York: McGraw-Hill.
- Lin, D. K. J. 1995. Generating systematic supersaturated designs. *Technometrics* 37: 213-225.
- Montgomery, D. C. 2009. *The Design and Analysis of Experiments*, 7th ed. New York: John Wiley and Sons.
- Saaty, T. L. 1980. *The Analytic Hierarchy Process*. New York: McGraw-Hill.
- Sanchez, S. M. and H. Wan. 2012. Work smarter, not harder: a tutorial on designing and conducting simulation experiments. In *Proceedings of the 2012 Winter Simulation Conference*, ed. C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Sanchez, S. M., H. Wan and T. W. Lucas. 2009. Two-phase screening procedure for simulation experiments. *ACM Transactions on Modeling and Computer Simulation* 19: 1-24.
- Sargent, R. G. 2011. Verification and validation of simulation models. In *Proceedings of the 2011 Winter Simulation Conference*, ed. S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, 183-198. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

AUTHOR BIOGRAPHY

RUSSELL R. BARTON is Professor of Supply Chain and Information Systems and Professor of Industrial Engineering at the Pennsylvania State University. He currently serves as senior associate dean for research and faculty in the Smeal College of Business and as associate director of its Center for the Management of Technological and Organizational Change. He received a B.S. degree in Electrical Engineering from Princeton University and M.S. and Ph.D. degrees in Operations Research from Cornell University. Before entering academia, he spent twelve years in industry. He is a past president of the INFORMS Simulation Society, and currently serves as the I-Sim representative on the INFORMS Subdivisions Council. He is a senior member of IIE and IEEE and a Certified Analytics Professional. His research interests include applications of statistical and simulation methods to system design and to product design, manufacturing and delivery. His email address is <rbarton@psu.edu>.