

HOW TO BUILD VALID AND CREDIBLE SIMULATION MODELS

Averill M. Law
Michael G. McComas

Averill M. Law and Associates, Inc.
P.O. Box 40996
Tucson, AZ 85717, U.S.A.

ABSTRACT

In this tutorial we present techniques for building valid and credible simulation models. Ideas to be discussed include the importance of a definitive problem formulation, discussions with subject-matter experts, interacting with the decision-maker on a regular basis, development of a written conceptual model, structured walk-through of the conceptual model, use of sensitivity analysis to determine important model factors, and comparison of model and system performance measures for an existing system (if any). Each idea will be illustrated by one or more real-world examples. We will also discuss the difficulty in using formal statistical techniques (e.g., confidence intervals) to validate simulation models.

1 WHAT IS MODEL VALIDATION

Use of a simulation model is a surrogate for experimentation with the actual system (existing or proposed), which is usually disruptive, not cost-effective, or simply impossible. Thus, if the model is not a “close” approximation to the actual system, any conclusions derived from the model are likely to be erroneous and may result in costly decisions being made. Validation should and can be done for *all* models, regardless of whether the corresponding system exists in some form or whether it will be built in the future.

We now give definitions of validation and credibility. *Validation* is the *process* of determining whether a simulation model is an accurate representation of the system, *for the particular objectives of the study*. The following are some general perspectives on validation:

- Conceptually, if a simulation model is “valid,” then it can be used to make decisions about the system similar to those that would be made if it were feasible and cost-effective to experiment with the system itself.
- The ease or difficulty of the validation process depends on the complexity of the system being mod-

eled and on *whether a version of the system currently exists* (see Section 2.8). For example, a model of a neighborhood bank would be relatively easy to validate since it could be closely observed. On the other hand, a model of the effectiveness of a naval weapons system in the year 2025 would be virtually impossible to validate completely, since the location of the battle and the nature of the enemy weapons would be unknown. Also, it is often possible to collect data on an existing system that can be used for building and validating a model.

- A simulation model of a complex system can only be an *approximation* to the actual system, no matter how much time and money is spent on model building. There is no such thing as absolute model validity, nor is it even desired. Indeed, a model is supposed to be an abstraction and simplification of reality. The more time (and hence money) that is spent on model development, the more valid the model should be in general. However, the most valid model is not necessarily the most cost-effective. For example, increasing the validity of a model beyond a certain level might be quite expensive, since extensive data collection may be required, but might not lead to significantly better insight or decisions.
- A simulation model should always be developed for a particular set of objectives. In fact, a model that is valid for one objective may not be for another.
- Validation is not something to be attempted after the simulation model has already been developed, and only if there is time and money remaining. Unfortunately, our experience indicates that this recommendation is often not followed.

Example 1. An organization paid a consulting company \$500,000 to perform a “simulation study” that had a six-month duration. After the study was supposedly completed, a person from the client organiza-

tion called us and asked, “Can you tell me in *five minutes* on the phone how to validate our model?”

A simulation model and its results have *credibility* if the decision-maker and other key project personnel accept them as “correct.” Note that a credible model is not necessarily valid, and vice versa. The following things help establish credibility for a model:

- The decision-maker’s understanding and agreement with the model’s assumptions
- Demonstration that the model has been validated and verified (i.e., that the model computer program has been debugged)
- The decision-maker’s ownership of and involvement with the project
- Reputation of the model developers

Note that many of the ideas and examples presented in this paper are based on the chapter “Building Valid, Credible, and Appropriately Detailed Models” in Law and Kelton (2000) and also on the simulation short courses presented by the first author. Other references on model validation are Balci (1998), Banks, Carson, Nelson, and Nicol (2001), Carson (1986), Sargent (1996), and Shannon (1975).

The remainder of this paper is organized as follows. We present in Section 1.1 a seven-step approach for conducting a successful simulate study. In Section 2 we discuss techniques for developing a more valid and credible simulation model. Guidelines for obtaining good model data are given in Section 3. Finally, Section 4 provides a summary of the most important validation ideas.

1.1 A Seven-Step Approach for Conducting a Successful Simulation Study

In Figure 1 we present a seven-step approach for conducting a successful simulation study. Having a definitive approach for conducting a simulation study is critical to the study’s success in general and to developing a valid model in particular. In Section 2 we will relate each of our validation/credibility enhancement techniques to one or more of these steps.

We now discuss important activities that take place in each of the seven steps.

1.1.1 Step 1. Formulate the Problem

- Problem of interest is stated by the decision-maker
- A kickoff meeting(s) for the simulation project is (are) conducted, with the project manager, the simulation analysts, and subject-matter experts (SMEs) in atten-

dance. The following things are discussed at this meeting:

- The overall objectives of the study
- The *specific* questions to be answered by the study (without such specificity it is impossible to determine the appropriate level of model detail)
- The performance measures that will be used to evaluate the efficacy of different system configurations
- The scope of the model
- The system configurations to be modeled
- The time frame for the study and the required resources

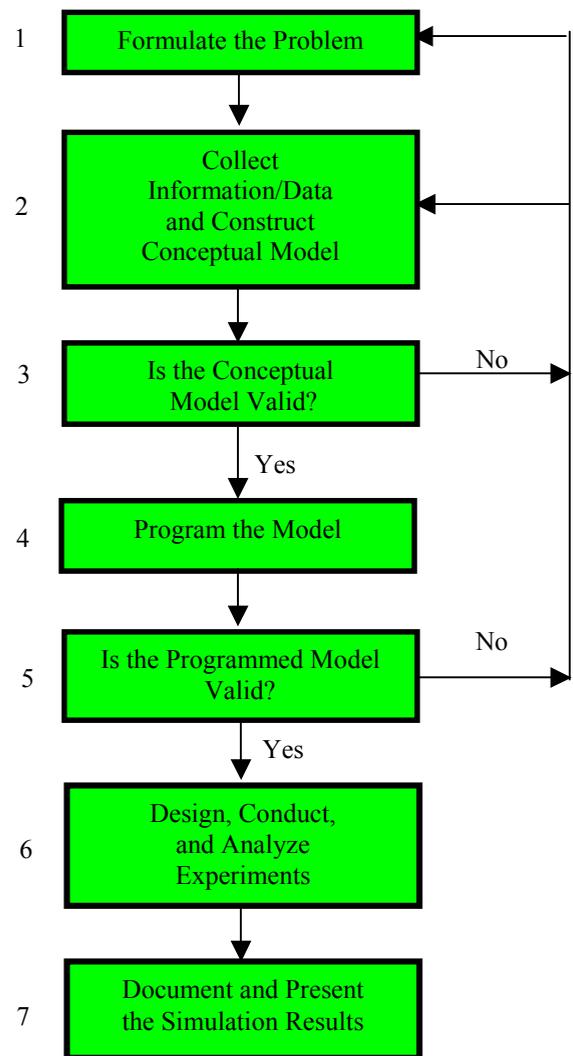


Figure 1: A Seven-Step Approach for Conducting a Successful Simulation Study

1.1.2 Step 2. Collect Information/Data and Construct a Conceptual Model

- Collect information on the system layout and operating procedures.
- Collect data to specify model parameters and probability distributions (e.g., for the time to failure and the time to repair of a machine).
- Document the model assumptions, algorithms, and data summaries in a written *conceptual model*.
- The level of model detail should depend on the following:
 - Project objectives
 - Performance measures of interest
 - Data availability
 - Credibility concerns
 - Computer constraints
 - Opinions of SMEs
 - Time and money constraints
- There should *not* be a one-to-one correspondence between the model and the system.
- Collect performance data from the existing system (if any) to use for model validation in Step 5.

1.1.3 Step 3. Is the Conceptual Model Valid?

- Perform a structured walk-through of the conceptual model before an audience that includes the project manager, analysts, and SMEs. This is called *conceptual-model validation*.
- If errors or omissions are discovered in the conceptual model, which is almost always the case, then the conceptual model must be updated before proceeding to programming in Step 4.

1.1.4 Step 4. Program the Model

- Program the conceptual model in either a commercial simulation-software product or in a general-purpose programming language (e.g., C or C++).
- Verify (debug) the computer program.

1.1.5 Step 5. Is the Programmed Model Valid?

- If there is an existing system, then compare model performance measures with the comparable performance measures collected from the actual system (see Step 2). This is called *results validation*.
- Regardless of whether there is an existing system, the simulation analysts and SMEs should review the simulation results for reasonableness. If the results are consistent with how they perceive the system should

operate, then the simulation model is said to have *face validity*.

- Sensitivity analyses should be performed on the programmed model to see which model factors have the greatest effect on the performance measures and, thus, have to be modeled carefully.

1.1.6 Step 6. Design, Make, and Analyze Simulation Experiments

- For each system configuration of interest, decide on tactical issues such as run length, warmup period, and the number of independent model replications.
- Analyze the results and decide if additional experiments are required.

1.1.7 Step 7. Document and Present the Simulation Results

- The documentation for the model (and the associated simulation study) should include the conceptual model (critical for future reuse of the model), a detailed description of the computer program, and the results of the current study.
- The final presentation for the simulation study should include animations and a discussion of the model building/validation process to promote model credibility.

2 TECHNIQUES FOR DEVELOPING VALID AND CREDIBLE MODELS

In this section we present *practical* techniques for developing valid and credible models. At the end of each subsection title, we state in square brackets (“[]”) in which of the seven steps (at a minimum) the technique should be applied.

2.1 Formulating the Problem Precisely [1]

It is critical to formulate the problem of interest in a precise manner. This should include an overall statement of the problem to be solved, a list of the specific questions that the model is to answer, and the performance measures that will be used to evaluate the efficacy of particular system configurations. Without a definitive statement of the specific questions of interest, it is impossible to decide on an appropriate level of model detail. Performance measures must also be clearly stated since different measures may dictate different levels of model detail [see Law and Kelton (2000, pp. 678-679) for an example].

When the decision-maker first initiates a simulation study, the exact problem to be solved is sometimes not precisely stated or even completely understood. Thus, as the study proceeds and a better understanding is obtained, this information should be communicated to the decision-maker who may reformulate the problem.

2.2 Interviewing Subject-Matter Experts [1, 2]

There will never be a single person that knows all of the information necessary to build a simulation model. Thus, it will be necessary for the simulation analysts to talk to many different SMEs to gain a complete understanding of the system to be modeled. Note that some of the information supplied by the SMEs will invariably be incorrect – if a certain part of the system is particularly important, then at least two SMEs should be queried. In Section 2.6, we will discuss a technique that helps ensure that a model’s assumptions are correct and complete – this technique is also useful for resolving differences of opinion among SMEs.

2.3 Interacting with the Decision-Maker on a Regular Basis [1-7]

One of the most important ideas for developing a valid and credible model is for the analyst to interact with the decision-maker and other members of the project team on a regular basis. This approach has the following key benefits:

- Helps ensure that the correct problem is solved
 - The exact nature of the problem may not be initially known.
 - The decision-maker may change his/her objectives during the course of the study.
- The decision-maker’s interest and involvement in the study are maintained.
- The model is more credible because the decision-maker understands and agrees with the model’s assumptions.

Example 2. A military analyst worked on a simulation project for several months without interacting with the general who requested it. At the final Pentagon briefing for the study, the general walked out after five minutes stating, “That’s not the problem I’m interested in.”

2.4 Using Quantitative Techniques to Validate Components of the Model [2]

The simulation analyst should use quantitative techniques whenever possible to test the validity of various components of the overall model. We now give examples of techniques that have been used for this purpose.

If one has fit a theoretical probability distribution (e.g., exponential or normal) to a set of observed data, then the adequacy of the representation can be assessed by using graphical plots and goodness-of-fit tests [see Law and Kelton (2000, Chapter 6)].

As will be discussed in Section 3, it is important to use appropriate data in building a model; however, it is equally important to exercise care when structuring these data. For example, if several sets of data have been observed for the

“same” random phenomenon, then the correctness of merging these data sets can be assessed by using the Kruskal-Wallis test of homogeneity of populations [see Law and Kelton (2000, pp. 394-395)]. If the data sets appear to be homogeneous, they can be merged and the combined data set used for some purpose in the simulation model.

Example 3. Consider a manufacturing system for which time-to-failure and time-to-repair data were collected for two “identical” machines made by the same vendor. However, the Kruskal-Wallis test showed that the two distributions were, in fact, different for the two machines. Thus, each machine was given its own time-to-failure and time-to-repair distributions in the simulation model.

2.5 Documenting the Conceptual Model [2]

Communication errors are a major reason why simulation models very often contain invalid assumptions. Documenting all assumptions, algorithms, and data summaries can lessen this problem. This report is the major documentation for the model and should be readable by analysts, SMEs, and decision-makers alike. The following are some of the things that should be included in the conceptual model:

- An overview section that discusses overall project goals, specific issues to be addressed by the model, and relevant performance measures
- A diagram showing the layout for the system
- Detailed descriptions of each subsystem (in *bullet format* for easy reading) and how they interact
- What simplifying assumptions were made and why
- Summaries of model input data (technical analyses should be put in appendices to promote report readability by decision-makers)
- Sources of important or controversial information

The conceptual model should contain enough detail so that it is a “blueprint” for creating the simulation computer program (in Step 4).

2.6 Performing a Structured Walk-Through of the Conceptual Model [3]

As previously discussed, the simulation analyst will need to collect system information from many different SMEs. Furthermore, these people are typically very busy dealing with the daily problems that occur within their organization, often resulting in their giving something less than their undivided attention to the questions posed by the simulation analyst. As a result, there is a considerable danger that the analyst will not obtain a complete and cor-

rect description of the system. An effective way of dealing with this potential problem is to conduct a *structured walk-through of the conceptual model* before an audience of SMEs and decision-makers. Using a projection device, the analyst goes through the conceptual model bullet-by-bullet, but not proceeding from one bullet to the next until everybody in the room is convinced that a particular bullet is correct and at an appropriate level of detail. A structured walk-through will increase both the validity and credibility of the simulation model. (As stated above, this exercise is called conceptual-model validation.)

The structured walk-through should ideally be held at a remote site (e.g., a hotel meeting room), so that people give the meeting their full attention. Furthermore, it should be held prior to the beginning of programming in case major problems are uncovered at the meeting. The conceptual model should be sent to participants prior to the meeting and their comments requested. We do not, however, consider this a substitute for the structured walk-through itself, since people may not have the time or motivation to review the document carefully on their own. Furthermore, the interactions that take place at the actual meeting are invaluable.

Example 4. At a structured walk-through of a transportation system, a significant percentage of the assumptions given to us by our corporate sponsor were found to be wrong by the SMEs present. (Due to the large geographic distances between the home offices of the sponsor and the SMEs, it was not possible for the SMEs to be present at the kickoff meeting for the project.) As a result, various people were assigned responsibilities to collect information on different parts of the system. The collected information was used to update the conceptual model, and a second walk-through was successfully performed.

2.7 Performing Sensitivity Analyses to Determine Important Model Factors [5]

An important technique for determining which model factors have a significant impact on the desired measures of performance is *sensitivity analysis*. If a particular factor appears to be important, then it needs to be modeled carefully. The following are examples of factors that could be investigated by a sensitivity analysis:

- The value of a parameter (see Example 5)
- The choice of a probability distribution
- The entity moving through the simulated system
- The level of detail for a subsystem

Example 5. In a simulation study of a new system, suppose that the value of a probability is estimated to be 0.75 as a result of conversations with SMEs. The importance of getting the value of this probability “ex-

actly” correct can be determined by running the simulation with 0.75 and, for example, by running it with each of the values 0.70 and 0.80. If the three simulation runs produce approximately the same results, then the output is not sensitive to the choice of the probability *over the range 0.70 to 0.80*. Otherwise, a better specification of the probability is needed. (Strictly speaking, to determine the effect of the probability on the model’s results, we should make several independent replications of the model using different random numbers for each of the three cases.)

If one is trying to determine the sensitivity of the simulation output to changes in two or more factors of interest, then it is not correct, in general, to vary one factor at a time while setting the other factors at some arbitrary values. A more correct approach is to use statistical experimental design, which is discussed in Law and Kelton (2000, Chapter 12) and in Montgomery (2000). The effect of each factor can be formally estimated and, if the number of factors is not too large, interactions between factors can also be detected.

2.8 Validating the Output from the Overall Simulation Model [5]

The most definitive test of a simulation model’s validity is establishing that its output data closely resemble the output data that would be observed from the actual system. If a system similar to the proposed one now exists, then a simulation model of the existing system is developed and its output data are compared to those from the existing system itself. If the two sets of data compare “closely,” then the model of the *existing* system is considered “valid.” (The accuracy required from the model will depend on its intended use and the utility function of the decision-maker.) The model is then modified so that it represents the proposed system. The greater the commonality between the existing and proposed systems, the greater our confidence in the model of the proposed system. There is no completely definitive approach for validating the model of the proposed system. If there were, then there might be no need for a simulation model in the first place. If the above comparison is successful, then it has the additional benefit of providing credibility for the use of simulation. (As stated above, the idea of comparing the model and system output data for the existing system is called results validation.)

Example 6. A U.S. Air Force test agency performed a simulation study for a bomb wing of bombers using the Logistics Composite Model (LCOM). The ultimate goal of the study was to evaluate the effect of various proposed logistics policies on the availability of the bombers, i.e., the proportion of time that the bombers were available to fly missions. Data were

available from the actual operations of the bomb wing over a 9-month period, and included both failure data for various aircraft components and a bomb-wing availability of 0.9. To validate the model, the Air Force first simulated the 9-month period with the existing logistics policy and obtained a model availability of 0.873, which is 3 percent different than the historical availability. This difference was considered acceptable because an availability of 0.873 would still allow enough bombers to be available for the Air Force to meet its mission requirements.

Example 7. A manufacturer of heat-treated aluminum products was thinking of replacing its existing batch furnace by a new continuous furnace in order to increase its production capacity [see Law (1991)]. We first simulated the existing system and found that the model monthly throughput differed from the historical monthly throughput by less than one percent. Thus, it appeared that the model of the existing system was reasonably “valid.”

A number of statistical tests (t , Mann-Whitney, etc.) have been suggested in the validation literature for comparing the output data from a simulation model with those from the corresponding real-world system [see, for example, Shannon (1975, p. 208)]. However, the comparison is not as simple as it might appear, since the output processes of almost all real-world systems and simulations are *non-stationary* (the distributions of the successive observations change over time) and *autocorrelated* (the observations in the process are correlated with each other). Thus, classical statistical tests based on independent, identically distributed (IID) observations are not *directly* applicable. Furthermore, we question whether hypothesis tests, as compared with constructing confidence intervals for differences, are even the appropriate statistical approach. Since the model is only an approximation to the actual system, a null hypothesis that the system and model are the “same” is clearly false. We believe that it is more useful to ask whether or not the differences between the system and the model are significant enough to affect any conclusions derived from the model. For a brief discussion of statistical procedures that can be used to compare model and system output data, see Section 2.10.

Whether or not there is an existing system, analysts and SMEs should review simulation output (numerical results, animations, etc.) for reasonableness. (Care must be taken in performing this exercise, since if one knew exactly what output to expect, then there would be no need for a model.) If the simulation results are consistent with perceived system behavior, then, as stated above, the model is said to have face validity.

Example 8. The above idea was put to good use in the development of a simulation model of the U.S. Air Force manpower and personnel system. (This model was designed to provide Air Force policy analysts with a system-wide view of the effects of various proposed personnel policies.) The model was run under the baseline personnel policy, and the results were shown to Air Force analysts and decision-makers, who subsequently identified some discrepancies between the model and perceived system behavior. This information was used to improve the model, and after several additional evaluations and improvements, a model was obtained that appeared to approximate current Air Force policy closely. This exercise improved not only the validity of the model, but also its credibility.

2.9 Using Graphical Plots and Animations of the Simulation Output Data [5-7]

Graphical plots (static or dynamic) and animations (dynamic) are useful for showing that a simulation model is *not* valid and for promoting model credibility. The following are some examples of graphical plots:

- Histogram (a graphical estimate of the underlying probability density or mass function)
- Correlation plot (shows if the output data are autocorrelated)
- Time plot (one or more model variables are plotted over the length of the simulation run to show the *long-run* dynamic behavior of the system)
- Bar charts and pie charts

An animation, which shows the *short-term* dynamic behavior of a system, is useful for communicating the essence of a model to decision-makers and other people who do not understand or care about the technical details of the model. Thus, it is a great way to enhance the credibility of a model. Animations are also useful for verification of the simulation computer program, for suggesting improved operational procedures, and for training.

2.10 Statistical Techniques for Comparing Model and System Output Data [5]

In this section we discuss the possible use of statistical procedures for carrying out the comparison of model and system output data discussed in Section 2.8.

Suppose that R_1, R_2, \dots, R_k are observations from a real-world system and that M_1, M_2, \dots, M_l are output data from a corresponding simulation model (see Example 9 below). We would like to compare these data sets in some way to determine whether the model is an accurate representation of the real-world system. However, most classi-

cal statistical approaches such as confidence intervals and hypothesis tests assume that the real-world data and the model data are each IID data sets, which is generally not the case (see the discussion in Section 2.8). Thus, these classical statistical approaches are not *directly* applicable to our comparison problem.

Example 9. Consider a manufacturing system where the data of interest are the times in system of successively completed parts. These data are not independent for the actual system (nor for a corresponding simulation model). For example, if the system is busy at a particular point in time, then all of the parts being processed will tend to have large times in system (i.e., the times are *positively correlated*).

Law and Kelton (2000, pp. 283-290) discuss inspection, confidence-interval, and time-series approaches that might possibly be used for comparing model and system output data.

3 GUIDELINES FOR OBTAINING GOOD MODEL DATA

A model is only valid for a particular application if its logic is correct *and* if it uses appropriate data. In this section we provide some suggestions on how to obtain good model data.

3.1 Two Basic Principles

If a system similar to the one of interest exists, then data should be obtained from it for use in building the model. These data may be available from historical records or may have to be collected during a time study. Since the people who provide the data might be different from the simulation analysts, it is important that the following two principles be followed:

- The analysts need to make sure that the data requirements (type, format, amount, why needed, etc.) are specified precisely to the people who provide the data.
- The analysts need to understand the process that produced the data, rather than treating the observations as just abstract numbers. For example, suppose that data are available on the time to load a ship, but there are a few observations that are significantly larger than the rest (called *outliers*). Without a good understanding of the underlying process, it is impossible to know whether these large observations are the result of measuring or recording errors, or are just legitimate values that occur with small probability.

3.2 Common Difficulties

The following are four potential difficulties with data:

- Data are not representative of what one really wants to model

Example 10. The data that have been collected during a military field test may not be representative of actual combat conditions due to differences in troop behavior and to lack of battlefield smoke.

- Data are not of the appropriate type or format

Example 11. In modeling a manufacturing system, the largest source of randomness is usually random down-times of a machine. Ideally, we would like data on time to failure (in terms of actual machine busy time) and time to repair of a machine. Sometimes data are available on machine breakdowns, but quite often they are not in the proper format. For example, the times to failure might be based on wall-clock time and include periods that the machine was idle or off-shift.

- Data may contain measuring, recording, or rounding errors

Example 12. Data representing the time to perform some task are sometimes rounded to the closest 5 or 10 minutes. This may make it difficult to fit a continuous theoretical probability distribution to the data, since the data are now discrete.

- Data may be “biased” because of self-interest

Example 13. The maintenance department in an automotive factory reported the reliability of certain machines to be greater than reality to make themselves look good.

4 SUMMARY

All simulation models need to be validated or any decisions made with the model may be erroneous. The following are the ideas that we believe are the most important for developing a valid and credible model:

- Formulating the problem precisely
- Interviewing appropriate subject matter experts
- Interacting with the decision-maker on a regular basis throughout the simulation project to ensure that the correct problem is being solved and to promote model credibility
- Developing a written conceptual model

- Performing a structured walk-through of the conceptual model – for a nonexistent system, this may be the single most-important validation technique
- Performing sensitivity analyses to determine important model factors
- Comparing model and system results for an existing system (if any) – this is, in general, the most definitive validation technique available
- Reviewing of model results and animations to see if they appear to be reasonable

Many of the above ideas would seem to be just common sense. However, our experience indicates that they are very often not applied.

REFERENCES

- Balci, O. 1998. "Verification, Validation and Testing," in *The Handbook of Simulation*, J. Banks, ed., Chapter 10, John Wiley, New York.
- Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2001. *Discrete-Event System Simulation*, Third Edition, Prentice-Hall, Upper Saddle River, N. J.
- Carson J. S. 1986. "Convincing Users of Model's Validity Is Challenging Aspect of Modeler's Job," *Industrial Engineering*, 18: 74-85.
- Law, A. M. 1991. "Simulation Study Puts the Right Heat On at Kaiser Aluminum," *Industrial Engineering*, 23: 16-17.
- Law, A. M. and W. D. Kelton. 2000. *Simulation Modeling and Analysis*, Third Edition, McGraw-Hill, New York.
- Montgomery, D.C. 2000. *Design and Analysis of Experiments*, 5th Edition, John Wiley, New York.
- Sargent, R. G. 1996. "Verifying and Validating Simulation Models," *Proceedings of the 1996 Winter Simulation Conference*, San Diego, 55-64.
- Shannon, R. E. 1975. *Systems Simulation: The Art and Science*, Prentice-Hall, Englewood Cliffs, N.J.

AUTHOR BIOGRAPHIES

AVERILL M. LAW is President of Averill M. Law & Associates, a company specializing in simulation consulting, training, and software. He has been a simulation consultant to numerous organizations including Accenture, ARCO, Boeing, Compaq, Defense Modeling and Simulation Office, Kimberly-Clark, M&M/Mars, 3M, U.S. Air Force, and U.S. Army. He has presented more than 335 simulation short courses in 17 countries. He has written or coauthored numerous papers and books on simulation, operations research, statistics, and manufacturing including the book *Simulation Modeling and Analysis* that is used by more than 75,000 people. He developed the ExpertFit distribution-fitting software and also several videotapes on

simulation modeling. He has been the keynote speaker at simulation conferences worldwide. He wrote a regular column on simulation for *Industrial Engineering* magazine. He has been a tenured faculty member at the University of Wisconsin-Madison and the University of Arizona. He has a Ph.D. in industrial engineering and operations research from the University of California at Berkeley. His E-mail address is averill@ix.netcom.com and his Web site is www.averill-law.com.

MICHAEL G. MCCOMAS is Vice President of Averill M. Law & Associates for Consulting Services. He has considerable simulation modeling experience in application areas such as manufacturing, oil and gas distribution, transportation, defense, and communications networks. His educational background includes an M.S. in systems and industrial engineering from the University of Arizona. He is the coauthor of seven published papers on applications of simulation.