HOW TO BUILD VALID AND CREDIBLE SIMULATION MODELS

Averill M. Law

Averill M. Law & Associates, Inc.
4729 E. Sunrise Drive, #462
Tucson, AZ 85718, USA

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 assumptions document, structured walk-through of the assumptions document, use of sensitivity analysis to determine important model factors, and comparison of model and system output data 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:

- A "valid" model can be used to make decisions 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 modeled 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 can be closely observed. On the other hand, a model of the effectiveness of a naval weapons system in the year 2035 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 the model beyond a certain level might be quite expensive, since extensive data collection may be required, but might not lead to

significantly better insights or decisions. A famous quote by Professor George Box states, "All models are wrong, but some models are useful."

- 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 organization called us and asked, "Can you tell me in five minutes on the phone how to validate our model?"

- Each time a simulation model is being considered for a new application its validity needs to be reexamined. The current purpose may be substantially different from the original purpose or the passage of time may have invalidated certain model assumptions or parameters.

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. In particular, a "valid" model might not be credible if the model's key assumptions are not understood and agreed with by the decision-maker. Conversely, a model may be credible because of an impressive 3-D animation, but not valid because of incorrect model assumptions that are not apparent from the animation. The following things help establish credibility for a model:

- The decision-maker's understanding of 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
- A compelling animation

The U.S. Department of Defense (DoD) is a large user of simulation models, and there is considerable interest in verification, validation, and a concept known as accreditation (VV&A). *Accreditation* – see DMSE (2022) – is the official certification (by the project sponsor) that a simulation model is acceptable for a specific purpose. The main reason that accreditation is mandated within DoD is that someone must take the responsibility for the decision to use a model for a particular application, since a large amount of money and people's lives may be at stake.

Note that many of the ideas and examples presented in this paper are based on the chapter "Building Valid, Credible, and Appropriately Detailed Simulation Models" in Law (2015) and also on the simulation short courses presented by the author. Other references on model validation are Carson (1986); Banks et al. (2010); MSCO (2012); Sargent (2015); Sargent et al. (2016); Robinson (2017); Sargent and Balci (2017); and Sargent (2020), with the second-to-last paper giving a history of validation (and verification) from pre-1970 to the present.

The remainder of this paper is organized as follows. We present in Section 1.1 a seven-step approach for conducting a successful simulation 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. This tutorial is an enhancement of one on the same subject that was presented at the 2019 Winter Simulation Conference.

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.

Step 1. Formulate the Problem

- Problem of interest is stated by the decision-maker
  - It may not be stated precisely or in quantitative terms.
  - An iterative process is often necessary.
- A kickoff meeting(s) for the simulation project is (are) conducted, with the project manager, the simulation analysts, and subject-matter experts (SMEs) in attendance. The following things are discussed at this meeting:
  - The overall objectives for 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 (people, computers, etc.)

Step 2. Collect Information/Data and Construct an Assumptions Document

- Collect information on the system structure and operating procedures.
- Collect data (if possible) to specify model parameters and input 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 *assumption document* (see Section 2.5).
- Collect data on the performance of the existing system (if any) for use in Step 5.
- Choosing the level of model detail, which is an art, 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.

Step 3. Is the Assumptions Document Valid?

- Perform a structured walk-through of the assumptions document (see Section 2.6) before an audience that includes the project manager, analysts, and SMEs.
- If errors or omissions are discovered in the assumptions document, which is almost always the case, then the assumptions document must be updated before proceeding to programming in Step 4.

Step 4. Program the Model

- Program the assumptions document in a commercial simulation-software package (AnyLogic, Arena, ExtendSim, FlexSim, Simio, SIMUL8, etc.) or in a general-purpose programming language (e.g., Python or Java).
- Verify (debug) the computer program.

Step 5. Is the Programmed Model Valid?

- If there is an existing system, then compare simulation model output data for the system with the comparable output data 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 factors have the greatest impact on the performance measures and, thus, have to be modeled carefully.
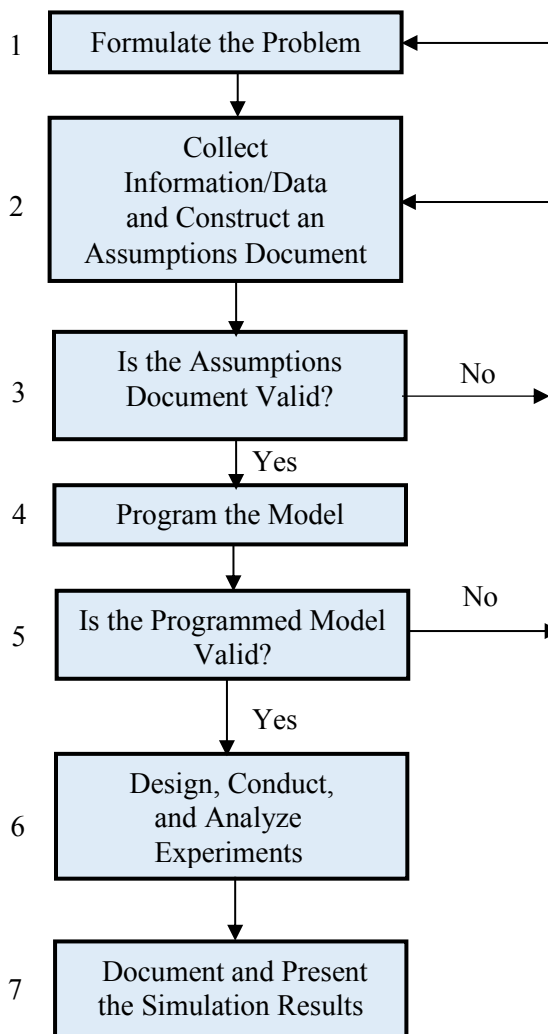
Figure 1: A seven-step approach for conducting a successful simulation study.

Step 6. Design, Conduct, and Analyze Experiments

- For each system configuration of interest, decide on tactical issues such as run length, warmup period, and the number of independent model replications (see chapter 9 in Law 2015)
- Analyze the results and decide if additional experiments are required.

Step 7. Document and Present the Results

- The documentation for the model (and the associated simulation study) should include the assumptions document (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 generally include an animation 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 be clearly stated since different measures may dictate different levels of model detail; see chapter 14 in Law (2015) 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 Utilizing Subject-Matter Experts [1-7]

There will never be a single person who 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. Ideally, an SME should have some knowledge of simulation modeling, so that they can supply relevant information. 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.

Example 2. For a manufacturing system, relevant people to talk to might include machine operators and their foreman, manufacturing and industrial engineers, maintenance personnel, schedulers, quality-assurance people, managers, and vendors.

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 may change during 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 3. 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 lognormal) 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 chapter 6 in Law (2015).

Example 4. In Figure 2 is a histogram of 219 inter-arrival times of cars to a drive-up banking facility. Superimposed over the histogram is the density function of the best-fitting exponential distribution, as determined by the ExpertFit® distribution-fitting software (Averill M. Law & Associates 2022). Visually, the exponential distribution appears to be a reasonable (but not fantastic) model for the interarrival times. If one performs a Kolmogorov-Smirnov goodness-of-fit test to determine whether the data can be considered an independent sample from an exponential distribution, then this hypothesis cannot be rejected at level 0.05. The chi-square test also does not reject the exponential distribution at level 0.05. A complete analysis of this data set can be found in chapter 6 of Law (2015).

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 (2015). 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 5. 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 Developing the Assumptions Document [2-3]

Communication errors are a major reason why simulation models very often contain invalid assumptions. The documentation of all concepts, assumptions, algorithms, and data summaries can lessen this problem. It will also increase the credibility of the model. An assumptions document should not be an "exact" description of how the system works, but rather a description of how it works relative to the particular issues that the model is to address. This report is the *major documentation for the model* and should be readable by analysts, SMEs, and technical managers. The following are some of the things that should be included in the assumptions document:

- An overview section that discusses overall project goals, specific issues to be addressed by the model, model inputs, and relevant performance measures (might be called *model requirements*)
- A process-flow/system-layout diagram (if appropriate)
- Detailed descriptions of each subsystem (in *bullet format* for easy reading), and how these subsystems interact
- What simplifying assumptions were made and why?
- Limitations of the model
- 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 (people, books, technical papers, etc.)

The assumptions document should contain enough detail so that it is a "blueprint" for creating the simulation computer program (in Step 4). In the defense community and elsewhere, there is a document with *some* commonality to the assumptions document, which is called a *conceptual model* (Robinson 2015). However, in many cases it represents the model developers' *initial thoughts* on the form the model will eventually take. The format for an assumptions document presented here was developed for small- to medium-sized models, and it might have to be modified to satisfy the needs of other situations.
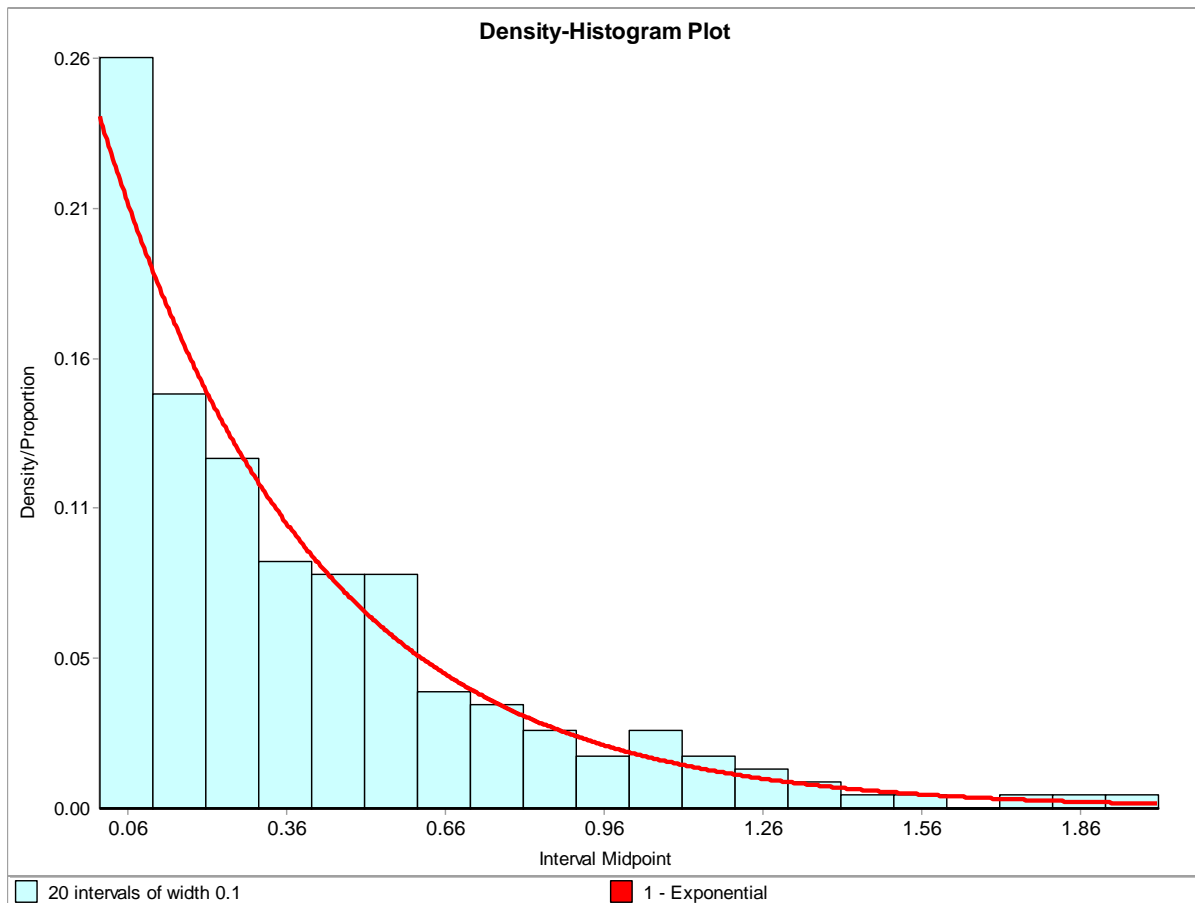

Figure 2: Histogram of 219 inter-arrival times with the density function of the best-fitting exponential distribution superimposed.

2.6 Performing a Structured Walk-through of the Assumptions Document [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 correct description of the system. An effective way of dealing with this potential problem is to conduct a *structured walk-through of the assumptions document* before an audience of SMEs and decision-makers. Using a projection device, the analyst goes through the assumptions document 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.

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 assumptions document 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 6. 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 assumptions document, and a second walk-through was successfully performed. The purported savings due to the simulation study was 52 million dollars.

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 7)
- The choice of a probability distribution
- The entity moving through the simulated system
- The level of detail for a subsystem

Example 7. 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 "exactly" 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, because factors often interact with each other. (This dangerous practice is

sometimes called the *one-factor-at-a-time appro*ach.) A more correct approach is to use statistical experimental design, which is discussed in chapter 12 of Law (2015) and in Montgomery (2017). The effect of each factor can be formally estimated and, if the number of factors is not too large, interactions among the 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 8. 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 9. A system for producing paper pulp had two machines, but local management wanted to buy a third machine. A model was first built for the existing system, and model and system throughputs for the two machines differed by only 0.4 and 1.1 percent. (The utilization differed by 2 and 11 percent, due to the second machine operator not following company policy.) A third machine was then added to the model, but it was found not to be necessary to meet system performance requirements. However, the key to the actual use of the model's results was the "buy-in" by the corporate VP of manufacturing, resulting from the success of the validation exercise.

Several 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, page 208 in Shannon 1975 and Sargent et al. 2016). However, the comparison is not as simple as it might appear, since the output processes of almost all real-world systems and simulations are *nonstationary* (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 model and the system are significant enough to affect any conclusions derived from the model. For a brief discussion of statistical procedures that *potentially* could be used to compare model and system output data, see Section 2.10.

In addition to statistical procedures, one can use a *Turing test* (see Schruben 1980) to compare the output data from the model to those from the system. People knowledgeable about the system (e.g., SMEs) are asked to examine one or more sets of system data as well as one or more sets of model data without knowing which data sets are which. If these SMEs can differentiate between the system and model data, their explanation of how they were able to do so is used to improve the model.

Example 10. An animation version of the Turing test was used in validating a simulation model of microscopic vehicle flow on a freeway. An animation of traffic flow from the simulation was displayed simultaneously on a large-screen monitor with an animation produced from data collected from the actual freeway. The data from the freeway were collected by a video camera mounted on an airplane.

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 11. 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.

Suppose that another model was developed for the same system and for a similar purpose, and that it is thought to be a "valid" representation. Then numerical statistics and graphical plots can be compared for the two models. It should be noted that just because two models produce similar results doesn't necessarily mean that either model is valid, since both models could contain a similar error. This idea seems to be most-widely used for defense applications, since there are sometimes several models for a particular system with a similar purpose.

Example 12. A defense supply center was building a new simulation model called the Performance and Requirements Impact Simulation to replace an existing model. One of the purposes of both models is to decide when to order and how much to order for each stock number. To validate the old model, the total dollar amount of all orders placed by the model for fiscal year 1996 was compared with the total dollar amount for the actual system for the same time period. Since these dollar amounts differed by less than 3 percent, there was a fair amount of confidence in the validity of the old model. To validate the new model, the two models were used to predict the total dollar amount of all orders for fiscal year 1998, and the results differed by less than 6 percent. Thus, there was reasonable confidence in the validity of the new model.

2.9   Animation [5-7]

An animation is useful for showing that a simulation model is *not* valid and for promoting model credibility. It is also valuable for verification of the simulation computer program, for suggesting improved operational procedures, and for training.

Example 13. A simulation model was developed for a candy packaging system. A newly promoted operations manager, who had no familiarity with the simulation model, declared, "That's my system!" upon seeing an animation of the system for the first time – the model gained instantaneous credibility.

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 $R_1, R_2, \ldots, R_k$ are observations from a real-world system and that $M_1, M_2, \ldots, M_l$ are output data from a corresponding simulation model (see Example 14 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 classical 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 *rarely* applicable to our comparison problem!

Example 14. Consider a manufacturing system where the output 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).

Chapter 5 in Law (2015) discusses 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, may have to be collected during a time study, or may be obtained electronically. 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, conditions under which it should be collected, 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.

Example 15. An oil company supplied us with 857 tanker-loading times, one of which was appreciably larger than the rest. A person who was involved in the actual loading process told us that the largest observation must have been a recording error. Thus, only the 856 smallest observations were used in our analysis.

3.2.   Common Difficulties

The following are five potential difficulties with data:

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

Example 16. 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 17. In modeling a manufacturing system, the largest source of randomness is usually random downtimes 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 18. Data representing the time to perform some task are sometimes rounded to the closest hour or day. 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 19. The maintenance department in an automotive factory reported the reliability of certain machines to be greater than reality to make themselves look good.

- Data were collected under unknown conditions.

Example 20. It was not known how many people performed the repair of a particular component for the U.S. space shuttle.

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 SMEs
- 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
- Using quantitative techniques to validate components of the model
- Developing a *written* assumptions document – extremely important
- Performing a structured walk-through of the assumptions document – 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
- Using a Turing test to compare model and system output data
- Reviewing model results and animations to see if they appear to be reasonable (i.e., face validity)
- Comparing model output data with the comparable output data for another model that is thought to be "valid"
- Obtaining "representative" data for use in building and validating a model

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

REFERENCES

Averill M. Law & Associates. 2022. "ExpertFit Distribution-Fitting Software". Version 8. Tucson, AZ.

Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2010. *Discrete-event System Simulation*. 5th ed. Upper Saddle River, N.J.: Prentice-Hall.

Carson J. S. 1986. "Convincing Users of Model's Validity is Challenging Aspect of Modeler's Job". *Industrial Engineering* 18(6):74-85.

Law, A. M. 2015. *Simulation Modeling and Analysis*. 5th ed. New York, NY: McGraw-Hill.

MSCO. 2012. *Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations (MIL-STD-3022)*. Alexandria, VA: Modeling and Simulation Coordination Office.

DMSE. 2022. *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide*. https://vva.msco.mil, accessed 7th July 2022. Alexandria, Virginia: Defense Modeling and Simulation Enterprise.

Montgomery, D.C. 2017. *Design and Analysis of Experiments*. 9th ed. New York, NY: John Wiley.

Robinson, S. 2017. "A Tutorial on Simulation Conceptual Modeling". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 565-579. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sargent, R. G. 2015. "An Introductory Tutorial on Verification and Validation of Simulation Models". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 1729-1740. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sargent, R. G. 2020. "Verification and Validation of Simulation Models: An Advanced Tutorial". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 16-29. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sargent, R. G., and O. Balci. 2017. "A History of Verification and Validation of Simulation Models". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 292–307. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sargent, R. G., D. M. Goldsman, and T. Yaacoub. 2016. "A Tutorial on the Operational Validation of Simulation Models". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 163–177. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schruben, L.W. 1980. "Establishing the Credibility of Simulations". *Simulation* 34(3):101-105.

Shannon, R. E. 1975. *Systems Simulation: The Art and Science*. Englewood Cliffs, N.J.: Prentice-Hall.

AUTHOR BIOGRAPHY

AVERILL M. LAW is President of Averill M. Law & Associates, Inc., a company specializing in simulation seminars, simulation consulting, and software. He has presented more than 580 simulation and statistics short courses in 20 countries, including onsite seminars for AT&T, Boeing, Caterpillar, Coca-Cola, Defence Research and Development Canada, Defence Science and Technology Group (Australia), GE, GM, IBM, Intel, Lockheed Martin, Los Alamos National Lab, NASA, NATO (Netherlands), Norwegian Defence Research Establishment, NSA, Sasol Technology (South Africa), 3M, UPS, U.S. Air Force, U.S. Army, and U.S. Navy. Dr. Law has been a simulation and statistics consultant to more than 50 organizations including Booz Allen & Hamilton, Conoco/Phillips, Defense Modeling and Simulation Office, Kimberly-Clark, M&M/Mars, Oak Ridge National Lab, U.S. Air Force, U.S. Army, U.S. Marine Corps, and U.S. Navy. He has written or coauthored numerous papers and books on simulation, operations research, statistics, manufacturing, and communications networks, including the book *Simulation Modeling and Analysis* that has been cited more than 23,300 times. He developed the ExpertFit® distribution-fitting software and also several videotapes on simulation modeling. He was awarded the INFORMS Simulation Society Lifetime Professional Achievement Award in 2009. Dr. Law 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@simulation.ws and his website is www.averill-law.com.