VERIFICATION AND VALIDATION OF SIMULATION MODELS*

Robert G. Sargent

Simulation Research Group
Department of Electrical Engineering and Computer Science
College of Engineering and Computer Science
Syracuse University
Syracuse, New York 13244

ABSTRACT

This paper discusses verification and validation of simulation models. The different approaches to deciding model validity are presented; how model verification and validation relate to the model development process are discussed; various validation techniques are defined; conceptual model validity, model verification, operational validity, and data validity are described; ways to document results are given; and a recommended procedure is presented.

1 INTRODUCTION

Simulation models are increasingly being used in problem solving and in decision making. The developers and users of these models, the decision makers using information derived from the results of the models, and people affected by decisions based on such models are all rightly concerned with whether a model and its results are "correct." This concern is addressed through model verification and validation. Model validation is usually defined to mean "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" (Schlesinger et al. 1979) and is the definition used here. Model verification is often defined as "ensuring that the computer program of the computerized model and its implementation are correct," and is the definition adopted here. A model sometimes becomes accredited through model accreditation. Model accreditation determines if a model satisfies a specified model accreditation criteria according to a specified process. A related topic is model credibility, which is concerned with sufficiently developing the confidence that (potential) users have in a model and in the information derived from the model that they are willing to use the model and the derived information.

A model should be developed for a specific purpose (or application) and its validity determined with respect to that purpose. If the purpose of a model is to answer a variety of questions, the validity of the model needs to be determined with respect to each question. Several sets of experimental conditions are usually required to define the domain of a model's intended applicability. A model may be valid for one set of experimental conditions and invalid in another. A model is considered valid for a set of experimental conditions if its accuracy is within its acceptable range, which is the amount of accuracy required for the model's intended purpose. This generally *requires* that the model's output variables of interest (i.e., the model variables used in answering the questions that the model is being developed to answer) be identified and that their required amount of accuracy be specified. The amount of accuracy required should be specified prior to starting the development of the model or very early in the model development process. If the variables of interest are random variables, then properties and functions of the random variables such as means and variances are usually what is of primary interest and are what is used in determining model validity. Several versions of a model are usually developed prior to obtaining a satisfactory valid model. The substantiation that a model is valid, i.e., model verification and validation, is generally considered to be a process and is usually part of the model development process.

It is often too costly and time consuming to determine that a model is *absolutely* valid over the complete domain of its intended applicability. Instead, tests and evaluations are conducted until sufficient confidence is obtained that a model can be considered valid for its intended application (Sargent 1982, 1984 and Shannon 1975). The relationships of cost (a similar relationship holds for the amount of time) of performing model validation and the value of the model to the user as a function of model confidence are illustrated in Figure 1. The cost of model validation is usually quite significant, particularly when extremely high model confidence is required.

*This paper is a modified version of Sargent (1996b).

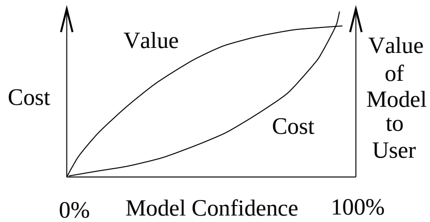

Figure 1: Model Confidence

The remainder of this paper is organized as follows: Section 2 discusses the basic approaches used in deciding model validity; Section 3 defines validation techniques; Sections 4, 5, 6, and 7 contain descriptions of data validity, conceptual model validity, model verification, and operational validity, respectively; Section 8 describes ways of presenting results; Section 9 contains a recommended validation procedure; and Section 10 gives the conclusions.

2 VALIDATION PROCESS

Three basic approaches are used in deciding whether a simulation model is valid or invalid. Each of the approaches requires the model development team to conduct verification and validation as part of the model development process, which is discussed below. The most common approach is for the development team to make the decision as to whether the model is valid. This is a subjective decision based on the results of the various tests and evaluations conducted as part of the model development process.

Another approach, often called "independent verification and validation" (IV&V), uses a third (independent) party to decide whether the model is valid. The third party is independent of both the model development team and the model sponsor/user(s). After the model is developed, the third party conducts an evaluation to determine its validity. Based upon this validation, the third party makes a subjective decision on the validity of the model. This approach is usually used when a large cost is associated with the problem the simulation model is being used for and/or to help in model credibility. (A third party is also usually used for model accreditation.)

The evaluation performed in the IV&V approach ranges from simply reviewing the verification and validation conducted by the model development team to a complete verification and validation effort. Wood (1986) describes experiences over this range of evaluation by a third party on energy models. One conclusion that Wood makes is that a complete IV&V evaluation is extremely costly and time consuming for what is obtained. This author's view is that if a third party is used, it should be *during* the model development process. If the model has already been developed, this author believes that usually a third party should evaluate only the verification and validation that has already been performed.

The last approach for determining whether a model is valid is to use a scoring model (see, e.g., Balci 1989, Gass 1979, and Gass and Joel 1987). Scores (or weights) are determined subjectively when conducting various aspects of the validation process and then combined to determine category scores and an overall score for the simulation model. A simulation model is considered valid if its overall and category scores are greater than some passing score(s). This approach is infrequently used in practice.

This author does not believe in the use of a scoring model for determining validity, because (1) the subjectiveness of this approach tends to be hidden and thus appears to be objective, (2) the passing scores must be decided in some (usually subjective) way, (3) a model may receive a passing score and yet have a defect that needs correction, and (4) the score(s) may cause overconfidence in a model or be used to argue that one model is better than another.

We now discuss how model verification and validation relate to the model development process. There are two common ways to view this relationship. One uses a detailed model development process, and the other uses a simple model development process. Banks et al. (1988) reviewed work using both of these ways and concluded that the simple way more clearly illuminates model verification and validation. This author recommends the use of a simple way (see, e.g., Sargent 1982), which is presented next.

Consider the simplified version of the modeling process in Figure 2. The *problem entity* is the system (real or proposed), idea, situation, policy, or phenomena to be modeled; the *conceptual model* is the mathematical/logical/verbal representation (mimic) of the problem entity developed for a particular study; and the *computerized model* is the conceptual model implemented on a computer. The conceptual model is developed through an *analysis and modeling phase*, the computerized model is developed through a *computer programming and implementation phase*, and inferences about the problem entity are obtained by conducting computer experiments on the computerized model in the *experimentation phase*.

We now relate model validation and verification to this simplified version of the modeling process (see Figure 2). *Conceptual model validity* is defined as determining that the theories and assumptions underlying the conceptual model are correct and that the model representation of the problem entity is "reasonable" for the intended purpose of the model. *Computerized model verification* is defined as ensuring that the computer programming and implementation of

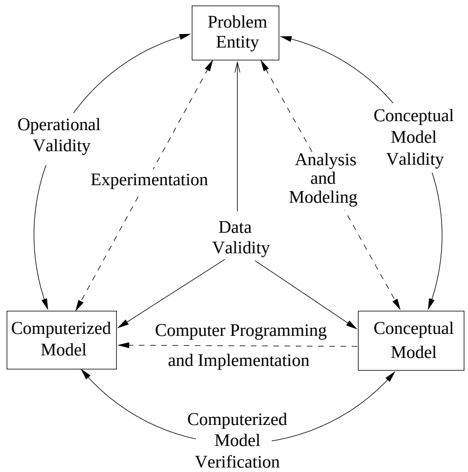

Figure 2: Simplified Version of the Modeling Process

the conceptual model is correct. *Operational validity* is defined as determining that the model's output behavior has sufficient accuracy for the model's intended purpose over the domain of the model's intended applicability. *Data validity* is defined as ensuring that the data necessary for model building, model evaluation and testing, and conducting the model experiments to solve the problem are adequate and correct.

Several versions of a model are usually developed in the modeling process prior to obtaining a satisfactory valid model. During each model iteration, model verification and validation are performed (Sargent 1984). A variety of (validation) techniques are used, which are described below. No algorithm or procedure exists to select which techniques to use. Some attributes that affect which techniques to use are discussed in Sargent (1984).

3 VALIDATION TECHNIQUES

This section describes various validation techniques (and tests) used in model verification and validation. Most of the techniques described here are found in the literature (see Balci and Sargent (1984a) for a detailed bibliography), although some may be described slightly differently. They can be used either subjectively or objectively. By "objectively," we mean using some type of statistical test or mathematical procedure, e.g., hypothesis tests and confidence intervals. A combination of techniques is generally used. These techniques are used for validating and verifying the submodels and overall model.

Animation: The model's operational behavior is displayed graphically as the model moves through time. For example, the movements of parts through a factory during a simulation are shown graphically.

Comparison to Other Models: Various results (e.g., outputs) of the simulation model being validated are compared to results of other (valid) models. For example, (1) simple cases of a simulation model may be compared to known results of analytic modes, and (2) the simulation model may be compared to other simulation models that have been validated.

Degenerate Tests: The degeneracy of the model's behavior is tested by appropriate selection of values of the input and internal parameters. For example, does the average number in the queue of a single server continue to increase with respect to time when the arrival rate is larger than the service rate?

Event Validity: The "events" of occurrences of the simulation model are compared to those of the real system to determine if they are similar. An example of events is deaths in a fire department simulation.

Extreme Condition Tests: The model structure and output should be plausible for any extreme and unlikely combination of levels of factors in the system; e.g., if in-process inventories are zero, production output should be zero.

Face Validity: "Face validity" is asking people knowledgeable about the system whether the model and/or its behavior are reasonable. This technique can be used in determining if the logic in the conceptual model is correct and if a model's input-output relationships are reasonable.

Fixed Values: Fixed values (e.g., constants) are used for various model input and internal variables and parameters. This should allow the checking of model results against easily calculated values.

Historical Data Validation: If historical data exist (or if data are collected on a system for building or testing the model), part of the data is used to build the model and the remaining data are used to determine (test) whether the model behaves as the system does. (This testing is conducted by driving the simulation model with either distributions or traces (Balci and Sargent 1982a, 1982b, 1984b).)

Historical Methods: The three historical methods of validation are *rationalism, empiricism*, and *positive economics.* Rationalism assumes that everyone knows whether the underlying assumptions of a model are true. Logic deductions are used from these assumptions to develop the correct (valid) model. Empiricism requires every assumption and outcome to be empirically validated. Positive economics requires only that the model be able to predict the future and is not concerned with a model's assumptions or structure (causal relationships or mechanism).

Internal Validity: Several replications (runs) of a stochastic model are made to determine the amount of (internal) stochastic variability in the model. A high amount of variability (lack of consistency) may cause the model's results to be questionable and, if typical of the problem entity, may question the appropriateness of the policy or system being investigated.

Multistage Validation: Naylor and Finger (1967) proposed combining the three historical methods of rationalism, empiricism, and positive economics into a multistage process of validation. This validation method consists of (1) developing the model's assumptions on theory, observations, general knowledge, and function, (2) validating the model's assumptions where possible by empirically testing them, and (3) comparing (testing) the input-output relationships of the model to the real system.

Operational Graphics: Values of various performance measures, e.g., number in queue and percentage of servers busy, are shown graphically as the model moves through time; i.e., the dynamic behaviors of performance indicators are visually displayed as the simulation model moves through time.

Parameter Variability–Sensitivity Analysis: This technique consists of changing the values of the input and internal parameters of a model to determine the effect upon the model's behavior and its output. The same relationships should occur in the model as in the real system. Those parameters that are sensitive, i.e., cause significant changes in the model's behavior or output, should be made sufficiently accurate prior to using the model. (This may require iterations in model development.)

Predictive Validation: The model is used to predict (forecast) the system behavior, and then comparisons are made between the system's behavior and the model's forecast to determine if they are the same. The system data may come from an operational system or from experiments performed on the system.

Traces: The behavior of different types of specific entities in the model are traced (followed) through the model to determine if the model's logic is correct and if the necessary accuracy is obtained.

Turing Tests: People who are knowledgeable about the operations of a system are asked if they can discriminate between system and model outputs. (Schruben (1980) contains statistical tests for use with Turing tests.)

4 DATA VALIDITY

Even though data validity is usually not considered to be part of model validation, we discuss it because it is usually difficult, time consuming, and costly to obtain sufficient, accurate, and appropriate data, and is frequently the reason that attempts to validate a model fail. Data are needed for three purposes: for building the conceptual model, for validating the model, and for performing experiments with the validated model. In model validation we are concerned only with the first two types of data.

To build a conceptual model we must have sufficient data on the problem entity to develop theories that can be used in building the model, to develop the mathematical and logical relationships in the model that will allow it to adequately represent the problem identity for its intended purpose, and to test the model's underlying assumptions. In addition, behavioral data is needed on the problem entity to be used in the operational validity step of comparing the problem entity's behavior with the model's behavior. (Usually, these data are system input/output data.) If these data are not available, high model confidence usually cannot be obtained, because sufficient operational validity cannot be achieved.

The concern with data is that appropriate, accurate, and sufficient data are available, and if any data transformations are made, such as disaggregation, they are correctly performed. Unfortunately, there is not much that can be done to ensure that the data are correct. The best that can be done is to develop good procedures for collecting and maintaining it, test the collected data using techniques such as internal consistency checks, and screen for outliers and determine if they are correct. If the amount of data is large, a data base should be developed and maintained.

5 CONCEPTUAL MODEL VALIDATION

Conceptual model validity is determining that (1) the theories and assumptions underlying the conceptual model are correct, and (2) the model representation of the problem entity and the model's structure, logic, and mathematical and causal relationships are "reasonable" for the intended purpose of the model. The theories and assumptions underlying the model should be tested using mathematical analysis and statistical methods on problem entity data. Examples of theories and assumptions are linearity, independence, stationary, and Poisson arrivals. Examples of applicable statistical methods are fitting distributions to data, estimating parameter values from the data, and plotting the data to determine if they are stationary. In addition, all theories used should be reviewed to ensure they were applied correctly; for example, if a Markov chain is used, does the system have the Markov property, and are the states and transition probabilities correct?

Next, each submodel and the overall model must be evaluated to determine if they are reasonable and correct for the intended purpose of the model. This should include determining if the appropriate detail and aggregate relationships have been used for the model's intended purpose, and if the appropriate structure, logic, and mathematical and causal relationships have been used. The primary validation techniques used for these evaluations are face validation and traces. Face validation has experts on the problem entity evaluate the conceptual model to determine if it is correct and reasonable for its purpose. This usually requires examining the flowchart or graphical model, or the set of model equations. The use of traces is the tracking of entities through each submodel and the overall model to determine if the logic is correct and if the necessary accuracy is maintained. If errors are found in the conceptual model, it must be revised and conceptual model validation performed again.

6 MODEL VERIFICATION

Computerized model verification ensures that the computer programming and implementation of the conceptual model are correct. To help ensure that a correct computer program is obtained, program design and development procedures found in the field of software engineering should be used in developing and implementing the computer program. These include object-oriented design, top-down design, structured programming, and program modularity. A separate program module or object should be used for each submodel, the overall model, and for each simulation function (e.g., time-flow mechanism, random number and random variate generators, and integration routines) when using general purpose higher-order languages, e.g., FORTRAN, PASCAL, C, or C++, and where possible when using simulation languages.

One should be aware that the type of computer language used affects the probability of having a correct program. The use of a special-purpose simulation language generally will result in having fewer errors than if a general-purpose simulation language is used, and using a general purpose simulation language will generally result in having fewer errors than if a general purpose higher-order language is used. Not only does the use of simulation languages increase the probability of having a correct program, programming time is usually reduced significantly. (However, flexibility is usually reduced also.)

After the computer program has been developed, implemented, and—optimistically—most of the programming "bugs" removed, the program must be tested for correctness and accuracy. First, the simulation functions should be tested to see if they are correct. Usually, straightforward tests can be used here to determine if they are working properly. Next, each submodel and the overall model should be tested to see if they are correct. Here the testing is more difficult.

There are two basic approaches to testing—static and dynamic testing (analysis) (Fairley 1976). In static testing the computer program of the computerized model is analyzed to determine if it is correct by using such techniques as correctness proofs, structured walk-through, and examining the structure properties of the program. The commonly used structured walk-through technique consists of each program developer explaining his or her computer program code statement-by-statement to other members of the modeling team until all are convinced it is correct.

In dynamic testing the computerized model is executed under different conditions and the resulting values are used to determine if the computer program and its implementations are correct. This includes both the values obtained during the program execution and the final values obtained. There are three different strategies used in dynamic testing: (1) bottom-up testing, which means, e.g., testing the submodels first and then the overall model; (2) top-down testing, which means, e.g., testing the overall model first using programming stubs (sets of data) for each of the submodels and then testing the submodels; and (3) mixed testing, which uses a combination of bottom-up and top-down testing (Fairly 1976). The techniques commonly used in dynamic testing are traces, investigations of input-output relations using different validation techniques, internal consistency checks, and reprogramming critical components to determine if the same results are obtained. If there are a large number of variables, one might aggregate some of the variables to reduce the number of tests needed or use certain types of design of experiments (Kleijnen 1987), e.g., use factor screening experiments to identify the key variables in order to reduce the number of experimental conditions that need to be tested.

It is necessary to be aware while checking the correctness of the computer program and its implementation that errors may be caused by the data, the conceptual model, the computer program, or the computer implementation.

For a more detailed discussion on model verification, see Whitner and Balci (1989).

7 OPERATIONAL VALIDITY

Operational validity is concerned with determining that the model's output behavior has the accuracy required for the model's intended purpose over the domain of its intended applicability. This is where most of the validation testing and evaluation takes place. The computerized model is used in operational validity, and thus any deficiencies found may be due

Table 1: Operational Validity Classification

	OBSERVABLE SYSTEM	NON-OBSERVABLE SYSTEM
SUBJECTIVE APPROACH	• COMPARISON USING GRAPHICAL DISPLAYS • EXPLORE MODEL BEHAVIOR	• EXPLORE MODEL BEHAVIOR • COMPARISON TO OTHER MODELS
OBJECTIVE APPROACH	• COMPARISON USING STATISTICAL TESTS AND PROCEDURES	• COMPARISON TO OTHER MODELS USING STATISTICAL TESTS AND PROCEDURES

to an inadequate conceptual model, an improperly programmed or implemented conceptual model (e.g., due to programming errors or insufficient numerical accuracy), or due to invalid data.

All of the validation techniques discussed in Section 3 are applicable to operational validity. Which techniques and whether to use them objectively or subjectively must be decided by the model development team and other interested parties. The major attribute affecting operational validity is whether the problem entity (or system) is observable, where observable means it is possible to collect data on the operational behavior of the program entity. Table 1 gives a classification of the validation approaches for operational validity. "Comparison" means comparing/testing the model and system input-out behaviors, and "explore model behavior" means to examine the output behavior of the model using appropriate validation techniques and usually includes parameter variability-sensitivity analysis. Various sets of experimental conditions from the domain of the model's intended applicability should be used for both comparison and exploring model behavior.

To obtain a *high* degree of confidence in a model and its results, comparison of the model's and system's input-output behaviors for *at least* two different sets of experimental conditions is usually required. There are three basic comparison approaches used: (1) graphs of the model and system behavior data, (2) confidence intervals, and (3) hypothesis tests. Graphs are the most commonly used approach, and confidence intervals are next.

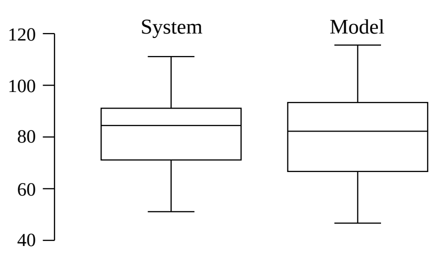

Figure 3: Box Plot

7.1 Graphical Comparison of Data

The behavior data of the model and the system are graphed for various sets of experimental conditions to determine if the model's output behavior has sufficient accuracy for its intended purpose. Three types of graphs are used: histograms, box (and whisker) plots, and behavior graphs using scatter plots. (See Sargent (1996a) for a thorough discussion on the use of these for model validation.) An example of a box plot is given in Figure 3, and examples of behavior graphs are shown in Figures 4 and 5. A variety of graphs using different types of (1) measures such as the mean, variance, maximum, distribution, and time series of a variable, and (2) relationships between two measures of a single variable (see Figure 4) and between measures of two variables (see Figure 5) are required. It is important that appropriate measures and relationships be used in validating a model and that they be determined with respect to the model's intended purpose. See Anderson and Sargent (1974) for an example of a set of graphs used in the validation of a simulation model.

These graphs can be used in model validation in different ways. First, the model development team can use the graphs in the model development process to make a subjective judgment on whether a model possesses sufficient accuracy for its intended purpose. Second, they can be used in the face validity technique where experts are asked to make subjective judgments on whether a model possesses sufficient accuracy for its intended purpose. Third, the graphs can be used is in Turing tests. Another way they can be used is in IV&V.

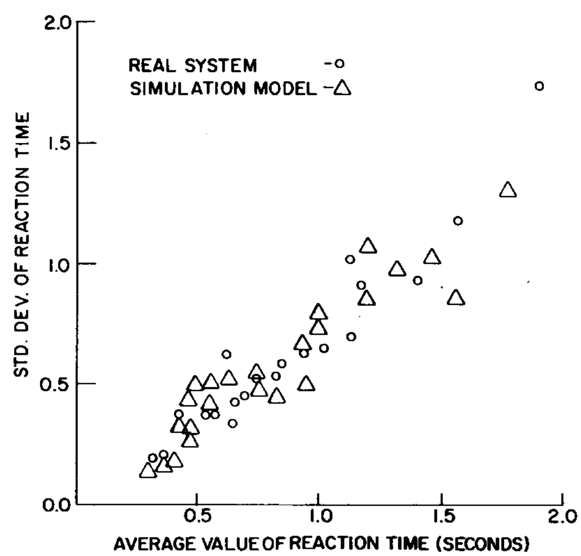

Figure 4: Reaction Time

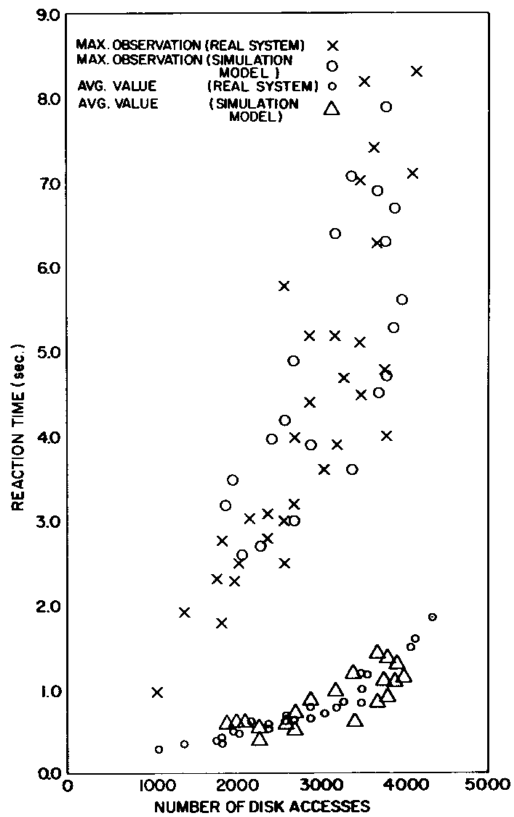

Figure 5: Disk Access

7.2 Confidence Intervals

Confidence intervals (c.i.), simultaneous confidence intervals (s.c.i.), and joint confidence regions (j.c.r.) can be obtained for the differences between the means, variances, and distributions of different model and system output variables for each set of experimental conditions. These c.i., s.c.i., and j.c.r. can be used as the model range of accuracy for model validation.

To construct the model range of accuracy, a statistical procedure containing a statistical technique and a method of data collection must be developed for each set of experimental conditions and for each variable of interest. The statistical techniques used can be divided into two groups: (1) univariate statistical techniques, and (2) multivariate statistical techniques. The univariate techniques can be used to develop c.i., and with the use of the Bonferroni inequality (Law and Kelton 1991), s.c.i. The multivariate techniques can be used to develop s.c.i. and j.c.r. Both parametric and nonparametric techniques can be used.

The method of data collection must satisfy the underlying assumptions of the statistical technique being used. The standard statistical techniques and data collection methods used in simulation output analysis (Banks, Carson, and Nelson 1996, Law and Kelton 1991) can be used for developing the model range of accuracy, e.g., the methods of replication and (nonoverlapping) batch means.

It is usually desirable to construct the model range of accuracy with the lengths of the c.i. and s.c.i. and the sizes of the j.c.r. as small as possible. The shorter the lengths or the smaller the sizes, the more useful and meaningful the model range of accuracy will usually be. The lengths and the sizes (1) are affected by the values of confidence levels, variances of the model and system response variables, and sample sizes, and (2) can be made smaller by decreasing the confidence levels or increasing the sample sizes. A tradeoff needs to be made among the sample sizes, confidence levels, and estimates of the length or sizes of the model range of accuracy, i.e., c.i., s.c.i., or j.c.r. Tradeoff curves can be constructed to aid in the tradeoff analysis.

Details on the use of c.i., s.c.i., and j.c.r. for operational validity, including a general methodology, are contained in Balci and Sargent (1984b). A brief discussion on the use of c.i. for model validation is also contained in Law and Kelton (1991).

7.3 Hypothesis Tests

Hypothesis tests can be used in the comparison of means, variances, distributions, and time series of the output variables of a model and a system for each set of experimental conditions to determine if the model's output behavior has an acceptable range of accuracy. An acceptable range of accuracy is the amount of accuracy that is required of a model to be valid for its intended purpose.

The first step in hypothesis testing is to state the hypotheses to be tested:

H_0: Model is valid for the acceptable range of accuracy under the set of experimental conditions.

H_1: Model is invalid for the acceptable range of accuracy under the set of experimental conditions.

Two types of errors are possible in testing hypotheses. The first, or type I error, is rejecting the validity of a valid model and the second, or type II error, is accepting the validity of an invalid model. The probability of a type error I, α, is called *model builder's risk*, and the probability of the type II error, β, is called *model user's risk* (Balci and Sargent 1981). In model validation, the model user's risk is extremely important and must be kept small. Thus *both* type I and type II errors must be carefully considered when using hypothesis testing for model validation.

The amount of agreement between a model and a system can be measured by a validity measure, λ, which is chosen such that the model accuracy or

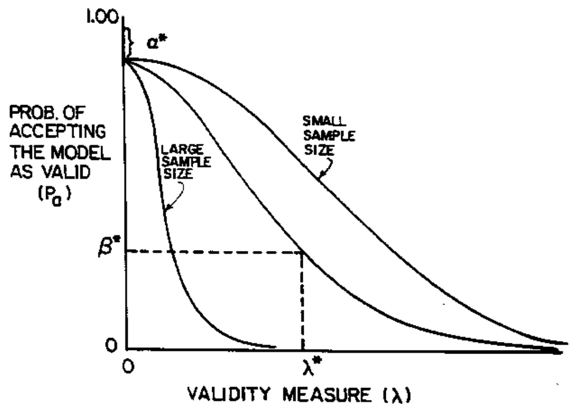

Figure 6: Operating Characteristic Curves

the amount of agreement between the model and the system decreases as the value of the validity measure increases. The acceptable range of accuracy can be used to determine an acceptable validity range, $0 \leq \lambda \leq \lambda^*$.

The probability of acceptance of a model being valid, P_a, can be examined as a function of the validity measure by using an Operating Characteristic Curve (Johnson 1994). Figure 6 contains three different operating characteristic curves to illustrate how the sample size of observations affect P_a as a function of λ. As can be seen, an inaccurate model has a high probability of being accepted if a small sample size of observations is used, and an accurate model has a low probability of being accepted if a large sample size of observations is used.

The location and shape of the operating characteristic curves are a function of the statistical technique being used, the value of α chosen for $\lambda = 0$, i.e., α^*, and the sample size of observations. Once the operating characteristic curves are constructed, the intervals for the model user's risk $\beta(\lambda)$ and the model builders risk α can be determined for a given λ^* as follows:

$\alpha^* \leq$ model builder's risk $\alpha \leq (1 - \beta^*)$

$0 \leq$ model user's risk $\beta(\lambda) \leq \beta^*$.

Thus there is a direct relationship among the builder's risk, model user's risk, acceptable validity range, and the sample size of observations. A tradeoff among these must be made in using hypothesis tests in model validation.

Details of the methodology for using hypothesis tests in comparing the model's and system's output data for model validations are given in Balci and Sargent (1981). Examples of the application of this methodology in the testing of output means for model validation are given in Balci and Sargent (1982a, 1982b, 1983). Also, see Banks et al. (1996).

8 DOCUMENTATION

Documentation on model verification and validation is usually critical in convincing users of the "correctness" of a model and its results, and should be included in the simulation model documentation. (For a general discussion on documentation of computer-based models, see Gass (1984).) Both detailed and summary documentation are desired. The detailed documentation should include specifics on the tests, evaluations made, data, results, etc. The summary documentation should contain a separate evaluation table for data validity, conceptual model validity, computer model verification, operational validity, and an overall summary. See Table 2 for an example of an evaluation table of conceptual model validity. (See Sargent (1994, 1996b) for examples of two of the other evaluation tables.) The columns of the table are self-explanatory except for the last column, which refers to the confidence the evaluators have in the results or conclusions, and this is often expressed as low, medium, or high.

9 RECOMMENDED PROCEDURE

This author recommends that, as a minimum, the following steps be performed in model validation:

1. Have an agreement made *prior* to developing the model between (a) the model development team and (b) the model sponsors and (if possible) the users, specifying the basic validation approach and a minimum set of specific validation techniques to be used in the validation process.
2. Specify the amount of accuracy required of the model's output variables of interest for the model's intended application prior to starting the development of the model or very early in the model development process.
3. Test, wherever possible, the assumptions and theories underlying the model.
4. In each model iteration, perform at least face validity on the conceptual model.
5. In each model iteration, at least explore the model's behavior using the computerized model.
6. In at least the last model iteration, make comparisons, if possible, between the model and system behavior (output) data for *at least* two sets of experimental conditions.
7. Develop validation documentation for inclusion in the simulation model documentation.
8. If the model is to be used over a period of time, develop a schedule for periodic review of the model's validity.

Table 2: Evaluation Table for Conceptual Model Validity

Category/Item	Technique(s) Used	Justification for Technique Used	Reference to Supporting Report	Result/ Conclusion	Confidence In Result
• Theories • Assumptions • Model representation	• Face validity • Historical • Accepted approach • Derived from empirical data • Theoretical derivation				

Strengths	
Weaknesses	

Overall evaluation for Computer Model Verification	Overall Conclusion	Justification for Conclusion	Confidence In Conclusion

Models occasionally are developed to be used more than once. A procedure for reviewing the validity of these models over their life cycles needs to be developed, as specified by step 8. No general procedure can be given, as each situation is different. For example, if no data were available on the system when a model was initially developed and validated, then revalidation of the model should take place prior to each usage of the model if new data or system understanding has occurred since its last validation.

10 SUMMARY

Model verification and validation are critical in the development of a simulation model. Unfortunately, there is no set of specific tests that can easily be applied to determine the "correctness" of the model. Furthermore, no algorithm exists to determine what techniques or procedures to use. Every new simulation project presents a new and unique challenge.

There is considerable literature on verification and validation. Articles given in the limited bibliography can be used as a starting point for furthering your knowledge on model verification and validation. For a fairly recent bibliography, see the following UHL on the WWW: http://manta.cs.vt.edu/biblio/.

LIMITED BIBLIOGRAPHY

Anderson, H. A. and R. G. Sargent. 1974. An Investigation into Scheduling for an Interactive Computer System, *IBM Journal of Research and Development*, 18, 2, pp. 125–137.

Balci, O. 1989. How to Assess the Acceptability and Credibility of Simulation Results, *Proc. of the 1989 Winter Simulation Conf.*, pp. 62–71.

Balci, O. 1995. Principles and Techniques of Simulation Validation, Verification, and Testing, *Proc. of the 1995 Winter Simulation Conf.*, pp. 147–154.

Balci, O. and R. G. Sargent. 1981. A Methodology for Cost-Risk Analysis in the Statistical Validation of Simulation Models, *Comm. of the ACM*, 24, 4, pp. 190–197.

Balci, O. and R. G. Sargent. 1982a. Validation of Multivariate Response Simulation Models by Using Hotelling's Two-Sample T^2 Test, *Simulation*, 39, 6, pp. 185–192.

Balci, O. and R. G. Sargent. 1982b. Some Examples of Simulation Model Validation Using Hypothesis Testing, *Proc. of the 1982 Winter Simulation Conf.*, pp. 620–629.

Balci, O. and R. G. Sargent. 1983. Validation of Multivariate Response Trace-Driven Simulation Models, *Performance 83*, ed. Agrawada and Tripathi, North Holland, pp. 309–323.

Balci, O. and R. G. Sargent. 1984a. A Bibliography on the Credibility Assessment and Validation of Simulation and Mathematical Models, *Simuletter*, 15, 3, pp. 15–27.

Balci, O. and R. G. Sargent. 1984b. Validation of Simulation Models via Simultaneous Confidence Intervals, *American Journal of Mathematical and Management Science*, 4, 3, pp. 375–406.

Banks, J., J. S. Carson II, and B. L. Nelson. 1996. *Discrete-Event System Simulation*, 2nd Ed., Prentice-Hall, Englewood Cliffs, N.J.

Banks, J., D. Gerstein, and S. P. Searles. 1988. Modeling Processes, Validation, and Verification of Complex Simulations: A Survey, *Methodology and Validation*, Simulation Series, Vol. 19, No. 1, The Society for Computer Simulation, pp. 13–18.

DOD Simulations: Improved Assessment Procedures Would Increase the Credibility of Results. 1987. U. S. General Accounting Office, PEMD-88-3.

Fairley, R. E. 1976. Dynamic Testing of Simulation Software, *Proc. of the 1976 Summer Computer Simulation Conf.*, Washington, D.C., pp. 40–46.

Gass, S. I. 1977. A Procedure for Evaluation of Complex Models, *Proc. of the First International Conf. on Math. Modeling,* Univ. of MO, pp. 247–257.

Gass, S. I. 1983. Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis, *Operations Research*, 31, 4, pp. 601–663.

Gass, S. I. 1984. Documenting a Computer-Based Model, *Interfaces,* Vol. 14, No. 3, pp. 84–93.

Gass, S. I. and L. Joel. 1987. Concepts of Model Confidence, *Computers and Operations Research*, 8, 4, pp. 341–346.

Gass, S. I. and B. W. Thompson. 1980. Guidelines for Model Evaluation: An Abridged Version of the U.S. General Accounting Office Exposure Draft, *Operations Research*, 28, 2, pp. 431–479.

Johnson, R. A. 1994. *Miller and Freund's Probability and Statistics for Engineers*, 5th Ed., Prentice-Hall, Englewood Cliffs, N.J.

Kleijnen, J. P. C. 1987. *Statistical Tools for Simulation Practitioners*, Marcel Dekker, New York.

Kleindorfer, G. B. and R. Ganeshan. 1993. The Philosophy of Science and Validation in Simulation, *Proc. of 1993 Winter Simulation Conf.*, 50–57.

Knepell, P. L. and D. C. Arangno. 1993. *Simulation Validation: A Confidence Assessment Methodology,* IEEE Computer Society Press.

Law, A. M. and W. D. Kelton. 1991. *Simulation Modeling and Analysis*, 2nd Ed., McGraw-Hill.

Naylor, T. H. and J. M. Finger. 1967. Verification of Computer Simulation Models, *Management Science*, 14, 2, pp. B92–B101.

Oren, T. 1981. Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference, *Comm. of the ACM*, 24, 4, pp. 180–189.

Rao, M. J. and R. G. Sargent. 1988. An advisory System for Operational Validity, *Artificial Intelligence and Simulation: The Diversity of Applications*, ed. T. Hensen, Society for Computer Simulation, San Diego, CA, pp. 245–2250.

Sargent, R. G. 1979. Validation of Simulation Models, *Proc. of the 1979 Winter Simulation Conf.*, San Diego, CA, pp. 497–503.

Sargent, R. G. 1981. An Assessment Procedure and a Set of Criteria for Use in the Evaluation of Computerized Models and Computer-Based Modelling Tools, Final Technical Report RADC-TR-80-409.

Sargent, R. G. 1982. Verification and Validation of Simulation Models, Chapter IX in *Progress in Modelling and Simulation*, ed. F. E. Cellier, Academic Press, London, pp. 159–169.

Sargent, R. G. 1984. Simulation Model Validation, *Simulation and Model-Based Methodologies: An Integrative View,* ed. Oren, et al., Springer-Verlag.

Sargent, R. G. 1985. An Expository on Verification and Validation of Simulation Models, *Proc. of the 1985 Winter Simulation Conf.*, pp. 15-22.

Sargent, R. G. 1986. The Use of Graphic Models in Model Validation, *Proc. of the 1986 Winter Simulation Conf.*, Washington, D.C., pp. 237–241.

Sargent, R. G. 1988. A Tutorial on Validation and Verification of Simulation Models, *Proc. of 1988 Winter Simulation Conf.*, pp. 33–39.

Sargent, R. G. 1990. Validation of Mathematical Models, *Proc. of Geoval-90: Symposium on Validation of Geosphere Flow and Transport Models*, Stockholm, Sweden, pp. 571–579.

Sargent, R. G. 1991. Simulation Model Verification and Validation, *Proc. of 1991 Winter Simulation Conf.*, Phoenix, AZ, pp. 37–47.

Sargent, R. G. 1992. Validation and Verification of Simulation Models, *Proc. of 1992 Winter Simulation Conf.*, Arlington, VA, pp. 104–114.

Sargent, R. G. 1994. Verification and Validation of Simulation Models, *Proc. of 1994 Winter Simulation Conf.*, Lake Buena Vista, FL, pp. 77–87.

Sargent, R. G. 1996a. Some Subjective Validation Methods Using Graphical Displays of Data, *Proc. of 1996 Winter Simulation Conf.*, pp. 345–351.

Sargent, R. G. 1996b. Verifying and Validating Simulation Models, *Proc. of 1996 Winter Simulation Conf.*, pp. 55–64.

Schlesinger, et al. 1979. Terminology for Model Credibility, *Simulation*, 32, 3,, pp. 103–104.

Schruben, L. W. 1980. Establishing the Credibility of Simulations, *Simulation*, 34, 3, pp. 101–105.

Shannon, R. E. 1975. *Systems Simulation: The Art and the Science*, Prentice-Hall.

Whitner, R. B. and O. Balci. 1989. Guidelines for Selecting and Using Simulation Model Verification Techniques, *Proc. of 1989 Winter Simulation Conf.*, Washington, D.C., pp. 559–568.

Wood, D. O. 1986. MIT Model Analysis Program: What We Have Learned About Policy Model Review, *Proc. of the 1986 Winter Simulation Conf.*, Washington, D.C., pp. 248–252.

Zeigler, B. P. 1976. *Theory of Modelling and Simulation,* John Wiley and Sons, Inc., New York.

AUTHOR BIOGRAPHY

ROBERT G. SARGENT is a professor at Syracuse University. He received his education at The University of Michigan. Professor Sargent has served his profession in numerous ways and has been awarded the TIMS (INFORMS) College on Simulation Distinguished Service Award for longstanding exceptional service to the simulation community. His research interests include the methodology areas of modeling and discrete event simulation, model validation, and performance evaluation. Professor Sargent is listed in *Who's Who in America.*