VALIDATION AND VERIFICATION OF SIMULATION MODELS

Robert G. Sargent

Department of Electrical Engineering and Computer Science
L.C. Smith College of Engineering and Computer Science
Syracuse University
Syracuse, NY 13244, U.S.A

ABSTRACT

In this paper we discuss validation and verfication of simulation models. Four different approaches to deciding model validity are described; two different paradigms that relate validation and verification to the model development process are presented; various validation techniques are defined; conceptual model validity, model verification, operational validity, and data validity are discussed; a way to document results is given; a recommended procedure for model validation is presented; and accreditation is briefly discussed.

1 INTRODUCTION

Simulation models are increasingly being used in problem solving and to aid in decision-making. The developers and users of these models, the decision makers using information obtained from the results of these models, and the individuals 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 validation and verification. 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 specified model accreditation criteria according to a specified process. A related topic is model credibility. Model credibility is concerned with developing in (potential) users the confidence they require in order to use a model and in the information derived from that model.

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. Numerous 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 the model's accuracy is within its acceptable range, which is the amount of accuracy required for the model's intended purpose. This usually 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., performing model validation and verification, is generally considered to be a process and is usually part of the (total) 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). If a test determines that a model does not have sufficient accuracy for any one of the sets of experimental conditions, then the model is invalid. However, determining that a model has sufficient accuracy for numerous experimental conditions does *not guarantee* that a model is valid everywhere in its applicable domain. The relationships of (a) cost (a similar relationship holds for the amount of time) of performing model validation and (b) the value of the model to a user as a function of model confidence are shown in Figure 1. The cost of model validation is usually quite significant, especially when extremely high model confidence is required.
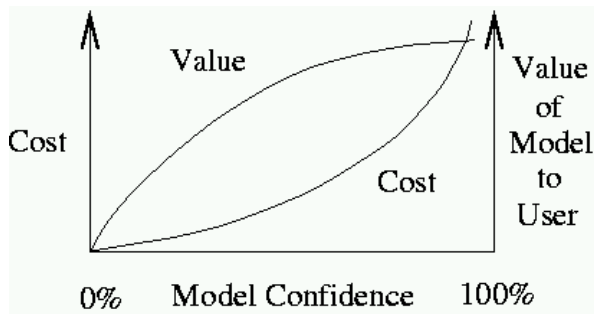
Figure 1: Model Confidence

The remainder of this paper is organized as follows: Section 2 presents the basic approaches used in deciding model validity; Section 3 describes two different paradigms used in verification and validation; Section 4 defines validation techniques; Sections 5, 6, 7, and 8 discuss data validity, conceptual model validity, computerized model verification, and operational validity, respectively; Section 9 describes a way of documenting results; Section 10 gives a recommended validation procedure; Section 11 contains a brief description of accreditation; and Section 12 presents the summary.

2 BASIC APPROACHES

There are four basic approaches for deciding whether a simulation model is valid. Each of the approaches requires the model development team to conduct validation and verification as part of the model development process, which is discussed below. One approach, and a frequently used one, is for the model development team itself to make the decision as to whether a simulation model is valid. A subjective decision is made based on the results of the various tests and evaluations conducted as part of the model development process. However, it is usually better to use one of the next two approaches, depending on which situation applies.

If the size of the simulation team developing the model is not large, a better approach than the one above is to have the user(s) of the model heavily involved with the model development team in determining the validity of the simulation model. In this approach the focus of who determines the validity of the simulation model moves from the model developers to the model users. Also, this approach aids in model credibility.

Another approach, usually called "independent verification and validation" (IV&V), uses a third (independent) party to decide whether the simulation model is valid. The third party is independent of both the simulation development team(s) and the model sponsor/user(s). The IV&V approach should be used when developing large-scale simulation models, whose developments usually involve several teams. This approach is also used to help in model credibility, especially when the problem the simulation model is associated with has a high cost. The third party needs to have a *thorough* understanding of the intended purpose(s) of the simulation model in order to conduct

IV&V. There are two common ways that the third party conducts IV&V: (a) IV&V is conducted concurrently with the development of the simulation model and (b) IV&V is conducted after the simulation model has been developed.

In the concurrent way of conducting IV&V, the model development team(s) receives input from the IV&V team regarding verification and validation as the model is being developed. When conducting IV&V this way, the development of a simulation model should not progress to the next stage of development until the model has satisfied the verification and validation requirements in its current stage. It is the author's opinion that this is the better of the two ways to conduct IV&V. When IV&V is conducted after the model has been completely developed, the evaluation performed can range from simply evaluating the verification and validation conducted by the model development team to performing 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 performing a complete IV&V effort after the simulation has been completely developed is both extremely costly and time consuming, especially for what is obtained. This author's view is that if IV&V is going to be conducted on a completed simulation model then it is usually best to *only* evaluate 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 Balci (1989), Gass (1993), and Gass and Joel (1987) for examples of scoring models.) 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 seldom used in practice.

This author does not believe in the use of scoring models for determining validity because (1) a model may receive a passing score and yet have a defect that needs to be corrected, (2) the subjectiveness of this approach tends to be hidden and thus this approach appears to be objective, (3) the passing scores must be decided in some (usually) subjective way, (4) the score(s) may cause over confidence in a model, and (5) the scores can be used to argue that one model is better than another.

3 PARADIGMS

In this section we present and discuss paradigms that relate validation and verification to the model development process. There are two common ways to view this relationship. One way uses a simple view and the other uses a complex view. Banks, Gerstein, and Searles (1988) reviewed work using both of these ways and concluded that the simple way more clearly illuminates model verification and validation. We present a paradigm for each way, both devel-

oped by this author. The paradigm of the simple way is presented first and is this author's preferred paradigm.

Consider the simplified version of the model development process in Figure 2 (Sargent 1981). 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*.
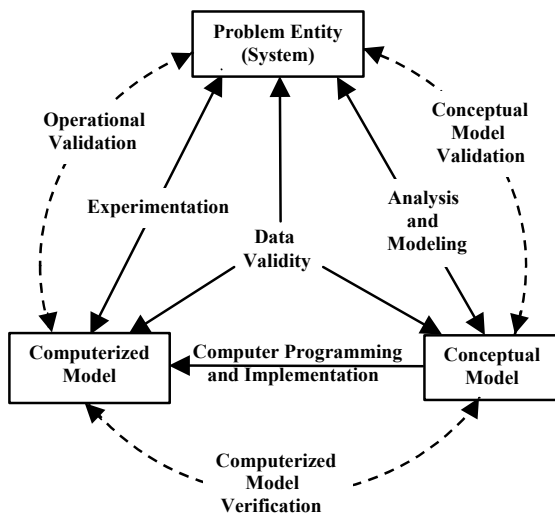

Figure 2: Simplified Version of the Modeling Process

We now relate model validation and verification to this simplified version of the modeling process. (See Figure 2.) *Conceptual model validation* 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 assuring that the computer programming and implementation of the conceptual model is correct. *Operational validation* 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.

In using this paradigm to develop a valid simulation model, several versions of a model are usually developed during the modeling process prior to obtaining a satisfactory valid model. During each model iteration, model validation and verification are performed (Sargent 1984). A variety of (validation) techniques are used, which are given 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).

A detailed way of relating verification and validation to developing simulation models and system theories is shown in Figure 3. This paradigm shows the processes of developing system theories and simulation models and relates verification and validation to both of these processes.

This paradigm (Sargent 2001b) shows a Real World and a Simulation World. We first discuss the Real World. There exist some *system or problem entity* in the real world of which an understanding of is desired. *System theories* describe the characteristics of the system (or problem entity) and possibly its behavior (including data). *System data and results* are obtained by conducting experiments (*experimenting*) on the system. System theories are developed by *abstracting* what has been observed from the system and by *hypothesizing* from the system data and results. If a simulation model exists of this system, then *hypothesizing* of system theories can also be done from simulation data and results. System theories are validated by performing *theory validation*. Theory validation involves the comparison of system theories against system data and results over the domain the theory is applicable for to determine if there is agreement. This process requires numerous experiments to be conducted on the real system.

We now discuss the Simulation World, which shows a slightly more complicated model development process than the other paradigm. A simulation model should only be developed for a set of well-defined objectives. The *conceptual model* is the mathematical/logical/verbal representation (mimic) of the system developed for the objectives of a particular study. The *simulation model specification* is a written detailed description of the software design and specification for programming and implementing the conceptual model on a particular computer system. The *simulation model* is the conceptual model running on a computer system such that experiments can be conducted on the model. The *simulation model data and results* are the data and results from experiments conducted (*experimenting*) on the simulation model. The conceptual model is developed by *modeling* the system for the objectives of the simulation study using the understanding of the system contained in the system theories. The simulation model is obtained by *implementing* the model on the specified computer system, which includes programming the conceptual model whose specifications are contained in the simulation model specification. Inferences about the system are obtained by conducting computer experiments (*experimenting*) on the simulation model. *Conceptual model validation* is defined as determining that the theories and assumptions underlying the conceptual model are consistent with those in the system theories and that the model representation of the system is "reasonable" for the intended purpose of the simulation model. *Specification verification* is defined as assuring that the software design
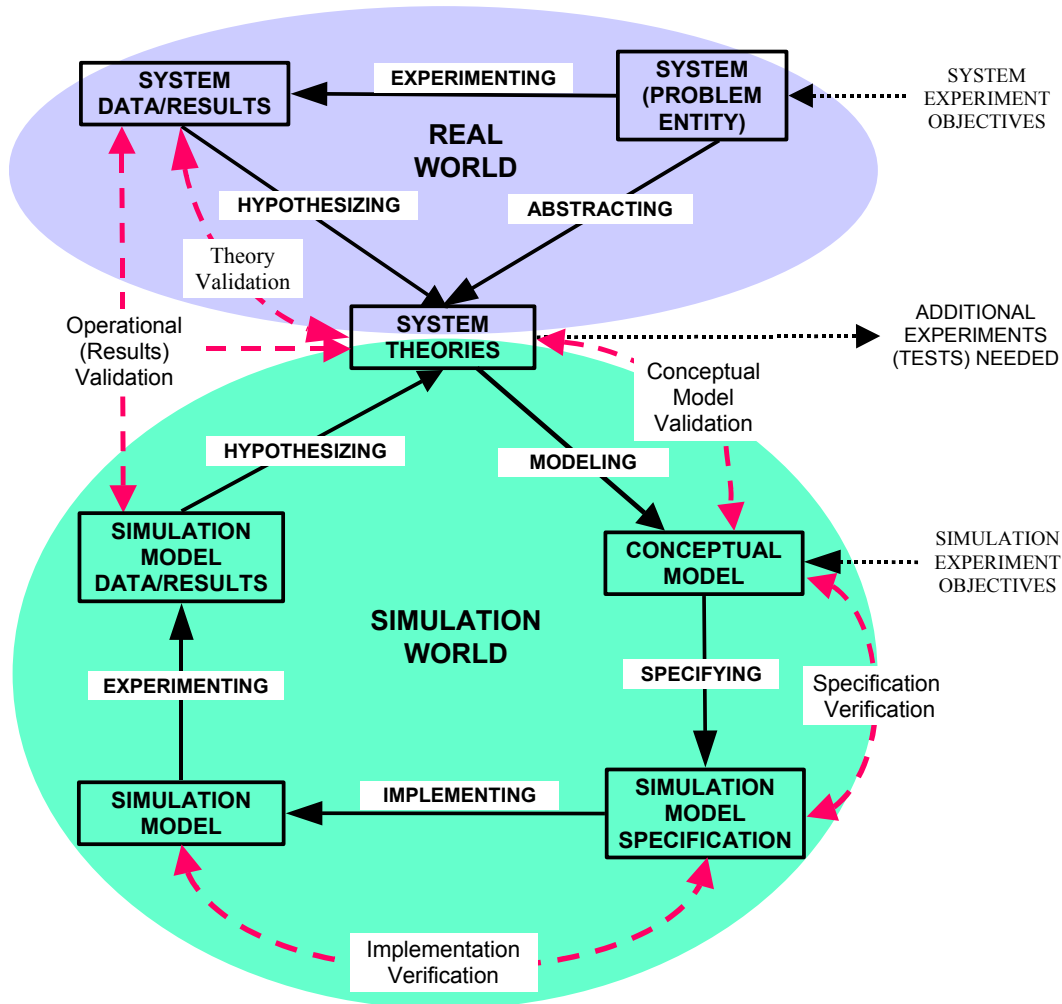
Figure 3: Real World and Simulation World Relationship with Verification and Validation

and the specification for programming and implementing the conceptual model on the specified computer system is satisfactory. *Implementation verification* is defined as assuring that the simulation model has been implemented according to the simulation model specification. *Operational validation* 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.

This paradigm shows processes for both developing valid system theories and valid simulation models. Both are accomplished through iterative processes. To develop valid system theories, which are usually for a specific purpose, the system is first observed and then abstraction is performed from what has been observed to develop proposed system theories. These theories are tested for correctness by conducting experiments on the system to obtain data and results to compare against the proposed system theories. New proposed system theories may be hypothesized from the data and comparisons made, and also possibly from abstraction performed on additional system ob-

servation. These new proposed theories will require new experiments to be conducted on the system to obtain data to evaluate the correctness of these proposed system theories. This process repeats itself until a satisfactory set of validated system theories has been obtained. To develop a valid simulation model, several versions of a model are usually developed prior to obtaining a satisfactory valid simulation model. During every model iteration, model verification and validation are performed. (This process is similar to the one for the other paradigm except there is more detail given in this paradigm.)

4   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, 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 or confi-

dence 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 run 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 are compared to known results of analytic models, and (2) the simulation model is 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 over 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. For example, compare the number of fires in a fire department simulation.

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

Face Validity: Asking individuals knowledgeable about the system whether the model and/or its behavior are reasonable. For example, is the logic in the conceptual model correct and are the model's input-output relationships reasonable.

Historical Data Validation: If historical data exist (or if data are collected on a system for building or testing a 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 samples from 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 mechanisms).

Internal Validity: Several replication (runs) of a stochastic model are made to determine the amount of (internal) stochastic variability in the model. A large 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, and general knowledge, (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., the number in queue and percentage of servers busy, are shown graphically as the model runs through time; i.e., the dynamical behaviors of performance indicators are visually displayed as the simulation model runs through time to ensure they are correct.

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 or 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's 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 be obtained by conducting experiments on the system, e.g., field tests.

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

5 DATA VALIDITY

We discuss data validity, even though it is often not considered to be part of model validation, because it is usually difficult, time consuming, and costly to obtain appropriate, accurate, and sufficient data, and is often the reason that attempts to valid 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 usually concerned only with data for the first two purposes.

To build a conceptual model we must have sufficient data on the problem entity to develop theories that can be

used to build the model, to develop mathematical and logical relationships for use in the model that will allow the model to adequately represent the problem entity for its intended purpose, and to test the model's underlying assumptions. In additional, behavioral data are 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, this data are system input/output data.) If behavior 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. One should develop good procedures for collecting and maintaining data, test the collected data using techniques such as internal consistency checks, and screen the data for outliers and determine if they are correct. If the amount of data is large, a database should be developed and maintained.

6 CONCEPTUAL MODEL VALIDATION

Conceptual model validity is determining that (1) the theories and assumptions underlying the conceptual model are correct and (2) the model's 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 of data, and arrivals are Poisson. Examples of applicable statistical methods are fitting distributions to data, estimating parameter values from the data, and plotting data to determine if the data 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, every 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 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 (Sargent 1986), 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.

7 COMPUTERIZED MODEL VERIFICATION

Computerized model verification ensures that the computer programming and implementation of the conceptual model are correct. The major factor affecting verification is whether a simulation language or a higher level programming language such as FORTRAN, C, or C++ is used. 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 level programming language is used. (The use of a simulation language also usually reduces both the programming time required and the amount of flexibility.)

When a simulation language is used, verification is primarily concerned with ensuring that an error free simulation language has been used, that the simulation language has been properly implemented on the computer, that a tested (for correctness) pseudo random number generator has been properly implemented, and the model has been programmed correctly in the simulation language. The primary techniques used to determine that the model has been programmed correctly are structured walk-throughs and traces.

If a higher level programming language has been used, then the computer program should have been designed, developed, and implemented using techniques found in software engineering. (These include such techniques as object-oriented design, structured programming, and program modularity.) In this case verification is primarily concerned with determining that the simulation functions (e.g., the time-flow mechanism, pseudo random number generator, and random variate generators) *and* the computer model have been programmed and implemented correctly.

There are two basic approaches for testing simulation software: static testing and dynamic testing (Fairley 1976). In static testing the computer program is analyzed to determine if it is correct by using such techniques as structured walk-throughs, correctness proofs, and examining the structure properties of the program. In dynamic testing the computer program is executed under different conditions and the values obtained (including those generated *during* the execution) are used to determine if the computer program and its implementations are correct. 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).

It is necessary to be aware while checking the correctness of the computer program and its implementation that errors found may be caused by the data, the conceptual model, the computer program, or the computer implementation. (For a detailed discussion on model verification, see Whitner and Balci (1989).)

8   OPERATIONAL VALIDITY

Operational validation is determining whether the simulation model's output behavior has the accuracy required for the model's intended purpose over the domain of the model's intended applicability. This is where much of the validation testing and evaluation take place. Since the simulation model is used in operational validation, any deficiencies found may be caused by what was developed in any of the steps that are involved in developing the simulation model including developing the system's theories or having invalid data.

All of the validation techniques discussed in Section 4 are applicable to operational validity. Which techniques and whether to use them objectively or subjectively must be decided by the model development team and the 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 problem entity. Table 1 gives a classification of the validation approaches in operational validity. "Comparison" means comparing/testing the model and system input-output behaviors, and "explore model behavior" means to examine the output behavior of the model using appropriate validation techniques, including 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.

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 Graphical Displays	• Comparison To Other Models Using Statistical Tests

To obtain a *high* degree of confidence in a model and its results, comparisons of the model's and system's input-output behaviors for several different sets of experimental conditions are usually required. There are three basic approaches used: (1) the use of graphs of model and system behavior data to make a subjective decision, (2) the use of

confidence intervals, and (3) the use of hypothesis tests. It is preferable to use confidence intervals or hypothesis tests for the comparisons because these allow for objective decisions. However, it is frequently not possible in practice to use either one of these two approaches because (a) the statistical assumptions required cannot be satisfied or only with great difficulty (assumptions usually necessary are data independence and normality) and/or (b) there is insufficient quantity of system data available, which causes the statistical results not to be "meaningful" (e.g., the length of a confidence interval developed in the comparison of the system and model means is to large for any practical usefulness). As a result, the use of graphs is the most commonly used approach for operational validity. Each of these three approaches is discussed in the following subsections.

8.1   Graphical Comparisons 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 the model's intended purpose. Three types of graphs are used: histograms, box (and whisker) plots, and behavior graphs using scatter plots. (See Sargent (1996a, 2001b) for a thorough discussion on the use of these for model validation.) Examples of a histogram and a box plot are given in Figures 4 and 5, respectively, both taken from Lowery (1996). Examples of
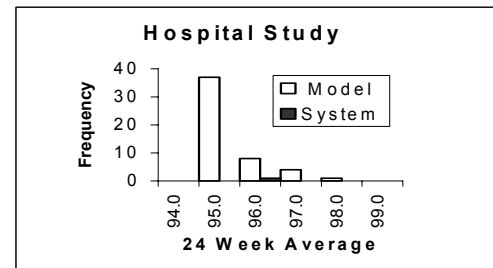

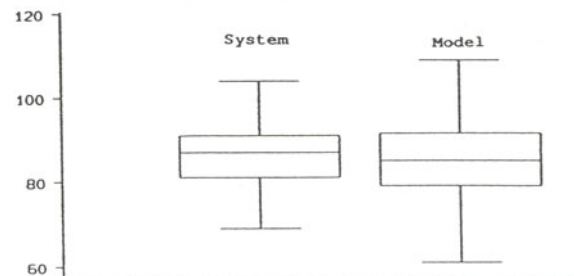Figure 4: Histogram of Hospital Data

Figure 5: Box Plot of Hospital Data

behavior graphs, taken from Anderson and Sargent (1974), are given in Figures 6 and 7. A variety of graphs using different types of (1) measures such as the mean, variance, maximum, distribution, and times series of a variable, and (2) relationships between (a) two measures of a single variable (see Figure 6) and (b) measures of two variables (see Figure

7) 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) and Lowery (1996) for examples of sets of graphs used in the validation of two different simulation models.
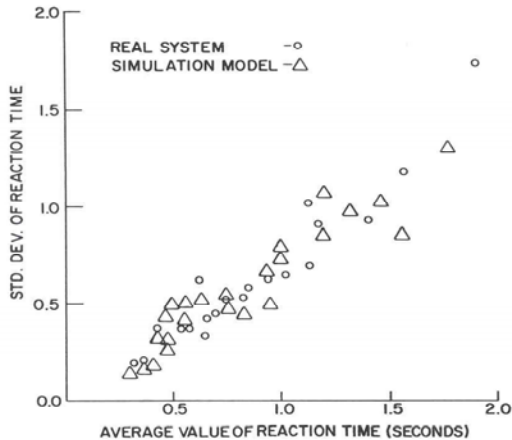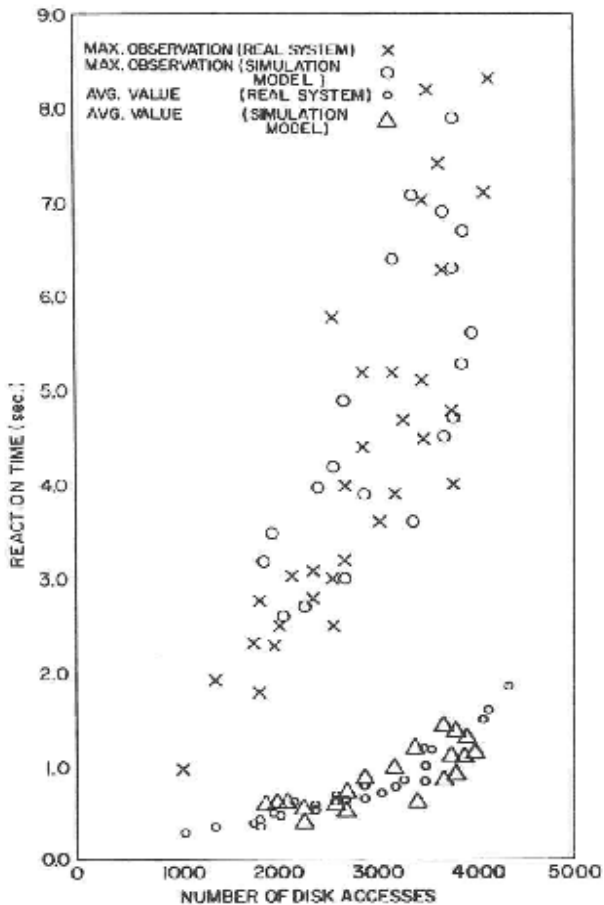

Figure 6: Reaction Time

Figure 7: Disk Access

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 posses sufficient accuracy for its intend 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 in Turing tests. Fourth, the graphs can be used in different ways in IV&V. We note that the data in these graphs do *not* need to be independence nor satisfy any statistical distribution requirement such as normality of the data (Sargent 1996a, 2001a, 2001b).

8.2 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 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 2000) 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 et al. 2000, Law and Kelton 2000) could 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 output 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 con-

tained in Balci and Sargent (1984b). A brief discussion on the use of c.i. for model validation is also contained in Law and Kelton (2000).

8.3 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 I error, α, is called *model builder's risk*, and the probability of 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 the amount of agreement between the model and the system decrease 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 8 contains three different operating characteristic
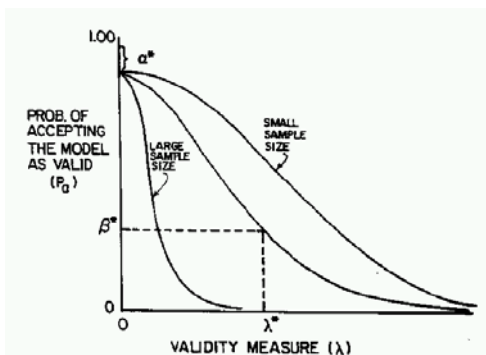

Figure 8: Operating Characteristic Curves

curves to illustrate how the sample size of observations affects 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 builder's risk α can be determined for a given λ^* as follows:

$$\alpha^* \leq \text{model builder's risk } \alpha \leq (1 - \beta^*)$$
$$0 \leq \text{model user's risk } \beta(\lambda) \leq \beta^*.$$

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 hypotheses 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).

9   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. (For examples of two other evaluation tables, see Sargent (1994, 1996b).) The columns of Table 2 are self-explanatory except for the last column, which refers to the confidence the evaluators have in the results or conclusions. These are often expressed as low, medium, or high.

10   RECOMMENDED PROCEDURE

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

1. An agreement be made prior to developing the model between (a) the model development team

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

and (b) the model sponsors and (if possible) the users that specifies 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 a few sets of experimental conditions, and preferably for several sets.
7. Develop validation documentation for inclusion in the 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.

Some models are developed for repeated use. A procedure for reviewing the validity of these models over their life cycles needs to be developed, as specified in 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 the last validation.

11 ACCREDITATION

The U. S. A. Department of Defense (DoD) has moved to accrediting simulation models. They define accreditation as the "official certification that a model, simulation, or federation of models and simulations and its associated data are acceptable for use for a specific application" (DoDI 5000.61). The evaluation for accreditation is usually conducted by a third (independent) party, is subjective, and often includes not only verification and validation but items such as documentation and how user friendly the simulation is. The acronym VV&A is used for Verification, Validation, and Accreditation. (Other areas and fields often use the term "Certification" to certify that a model (or product) conforms to a specified set of characteristics (see Balci (2003) for further discussion).)

12 SUMMARY

Model validation and verification 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 a model. Furthermore, no algorithm exists to determine what techniques or procedures to use. Every simulation project presents a new and unique challenge to the model development team.

There is considerable literature on validation and verification; see e.g. Balci and Sargent (1984a). Articles given under references, including several not referenced above, can be used to further your knowledge on model validation and verification. Research continues on these topics.

REFERENCES

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): 125-137.

Balci, O. 1989. How to assess the acceptability and credibility of simulation results. In *Proc. 1989 Winter Simulation Conf.*, ed. E. A. MacNair, K. J. Musselman, and P. Heidelberger, 62-71. IEEE.

Balci, O. 2003. Validation, verification, and certification of modeling and simulation applications. In *Proc. 2003*

Winter Simulation Conf., ed. S. Chick, P. J. Sanchez, E. Ferrin, and D. J. Morrice, 150-158. IEEE.

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 (6): 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): 185-192.

Balci, O. and R. G. Sargent. 1982b. Some examples of simulation model validation using hypothesis testing. In *Proc. 1982 Winter Simulation Conf.*, ed. H. J. Highland, Y. W. Chao, and O. S. Madrigal, 620-629. IEEE.

Balci, O. and R. G. Sargent. 1983. Validation of multivariate response trace-driven simulation models. In *Performance 83*, ed. Agrawada and Tripathi, 309-323. North Holland.

Balci, O. and R. G. Sargent. 1984a. A bibliography on the credibility assessment and validation of simulation and mathematical models. *Simuletter* 15 (3): 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): 375-406.

Banks, J., J. S. Carson II, B. L. Nelson, and D. Nicol. 2000. Discrete-event system simulation. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.

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

Carson II, J. S. 2002. Model Verification and Validation. In *Proc. 2002 Winter Simulation Conf.*, ed. E. Yucesan, C.-H. Chen, J.L. Snowdon, J. M. Charnes, 52-58.

DoDI 5000.61. 2003. DoD modeling and simulation verification, validation, and accreditation, May 13, 2003.

Fairley, R. E. 1976. Dynamic testing of simulation software. In *Proc. 1976 Summer Computer Simulation Conf.*, 40-46, Washington, D.C.

Gass, S. I. 1983. Decision-aiding models: validation, assessment, and related issues for policy analysis. *Operations Research* 31 (4): 601-663.

Gass, S. I. 1984. Documenting a computer-based model. *Interfaces* 14 (3): 84-93.

Gass, S. I. 1993. Model Accreditation: A Rationale and Process for Determining a Numerical Rating. *European Journal of Operational Research* 66 (2): 250-258.

Gass, S. I. and L. Joel. 1987. Concepts of Model Confidence. *Computers and Operations Research* 8 (4): 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): 431-479.

Johnson, R. A. 1994. *Miller and Freund's Probability and Statistics for Engineers*, 5th ed. Englewood Cliffs, NJ: Prentice-Hall.

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

Kleijnen, J. P. C. 1999. Validation of models: statistical techniques and data availability. In *Proc. 1999 Winter Simulation Conf.*, ed. P. A. Farrington, H. Black Nembhard, D. T. Sturrock, and G. W. Evans, 647-654. IEEE.

Kleindorfer, G. B. and R. Ganeshan. 1993. The philosophy of science and validation in simulation. In *Proc. 1993 Winter Simulation Conf.*, ed. G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles, 50-57. IEEE.

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. 2000. *Simulation modeling and analysis*. 3rd ed. McGraw-Hill.

Law, A. M. and M. G. McComas. 2001. How to Build Valid and Credible Simulation Models. In *Proc. 2001 Winter Simulation Conf.*, ed. B.A. Peters, J. S. Smith, D. J Medeiros, and M. W. Rohrer, 22-29. IEEE.

Lowery, J. 1996. Design of Hospital Admissions Scheduling System Using Simulation. In *Proc. 1996 Winter Simulation Conf.*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 1199-1204. IEEE.

Naylor, T. H. and J. M. Finger. 1967. Verification of computer simulation models. *Management Science* 14 (2): B92-B101.

Oren, T. 1981. Concepts and criteria to assess acceptability of simulation studies: a frame of reference. *Comm. of the ACM* 24 (4): 180-189.

Rao, M. J. and R. G. Sargent. 1988. An advisory system for operational validity. In *Artificial Intelligence and Simulation: the Diversity of Applications*, ed. T. Hensen, 245-250. San Diego, CA: Society for Computer Simulation.

Sargent, R. G. 1979. Validation of simulation models. In: *Proc. 1979 Winter Simulation Conf.*, ed. H. J. Highland, M. F. Spiegel, and R. E. Shannon, 497-503. IEEE.

Sargent, R. G. 1981. An assessment procedure and a set of criteria for use in the evaluation of computerized models and computer-based modeling tools. Final Technical Report RADC-TR-80-409, U. S. Air Force.

Sargent, R. G. 1982. Verification and validation of simulation Models. Chapter IX in *Progress in Modelling and Simulation*, ed. F. E. Cellier, 159-169. London: Academic Press.

Sargent, R. G. 1984. Simulation model validation. Chapter 19 in *Simulation and Model-Based Methodologies: An Integrative View*, ed. T. I. Oren, B. P. Zeigler, and M. S. Elzas, 537-555. Heidelberg, Germany: Springer-Verlag.

Sargent, R. G. 1985. An expository on verification and validation of simulation models. In *Proc. 1985 Winter*

Simulation Conf., ed. D. T. Gantz, G. C. Blais, and S. L. Solomon, 15-22. Piscataway, New Jersey: IEEE.

Sargent, R. G. 1986. The use of graphical models in model validation. In *Proc. 1986 Winter Simulation Conf.*, ed. J. R., Wilson, J. O. Henriksen, and S. D. Roberts, 237-241. Piscataway, New Jersey: IEEE.

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

Sargent, R. G. 1994. Verification and validation of simulation models. In *Proc. 1994 Winter Simulation Conf.*, ed. J. D. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, 77-87. Piscataway, New Jersey: IEEE.

Sargent, R. G. 1996a. Some subjective validation methods using graphical displays of data. In *Proc. 1996 Winter Simulation Conf.*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 345-351. Piscataway, New Jersey: IEEE.

Sargent, R. G. 1996b. Verifying and validating simulation models. In *Proc. of 1996 Winter Simulation Conf.*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 55-64. Piscataway, New Jersey: IEEE.

Sargent, R. G. 2000. Verification, validation, and accreditation of simulation models. In *Proc. 2000 Winter Simulation Conf.*, ed. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 50-59. Piscataway, New Jersey: IEEE.

Sargent, R. G. 2001a. Graphical displays of simulation model data as statistical references. In *Simulation 2001 (Proc. of the 4th St. Petersburg Workshop on Simulation)*, ed. S. M. Ermakor, Yu. N. Kashtanov, and V. B. Melas, 109-118. Publisher: Chemistry Research Institute of St. Petersburg University.

Sargent, R. G. 2001b. Some approaches and paradigms for verifying and validating simulation models. In *Proc. 2001 Winter Simulation Conf.*, ed. B. A. Peters, J. S. Smith, D. J Medeiros, and M. W. Rohrer, 106-114. Piscataway, New Jersey: IEEE.

Sargent, R. G. 2003. Verification and validation of simulation models. In *Proc. 2003 Winter Simulation Conf.*, ed. S. Chick, P. J. Sanchez, E. Ferrin, and D. J. Morrice, 37-48. Piscataway, New Jersey: IEEE.

Schlesinger, et al. 1979. Terminology for model credibility. *Simulation* 32 (3): 103-104.

Schruben, L.W. 1980. Establishing the credibility of simulation models. *Simulation* 34 (3): 101-105.

U. S. General Accounting Office, PEMD-88-3. 1987. DOD simulations: improved assessment procedures would increase the credibility of results.

Whitner, R. G. and O. Balci. 1989. Guideline for selecting and using simulation model verification techniques. In *Proc. 1989 Winter Simulation Conf.*, ed. E. A. MacNair, K. J. Musselman, and P. Heidelberger, 559-568. Piscataway, New Jersey: IEEE.

Wood, D. O. 1986. MIT model analysis program: what we have learned about policy model review. In *Proc. 1986 Winter Simulation Conf.*, ed. J. R. Wilson, J. O. Henriksen, and S. D. Roberts, 248-252. Piscataway, New Jersey: IEEE.

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

AUTHOR BIOGRAPHY

ROBERT G. SARGENT is a Professor Emeritus of Syracuse University. He received his education at The University of Michigan. Dr. Sargent has served his profession in numerous ways including being the General Chair of the 1977 Winter Simulation Conference, serving on the WSC Board of Directors for ten years, chairing the Board for two years, being a Department Editor for the *Communications of the ACM*, and currently holding the presidency of the WSC Foundation. He has received several awards including the INFORMS-College on Simulation Lifetime Professional Achievement Award and their Distinguished Service Award. His current research interests include the methodology areas of modeling and of discrete event simulation, model validation, and performance evaluation. Professor Sargent has published extensively and is listed in Who's Who in America. His web and e-mail addresses are <www.cis.syr.edu/srg/rsargent/> and <rsargent@syr.edu>.