VERIFICATION AND VALIDATION OF SIMULATION MODELS

Robert G. Sargent

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

ABSTRACT

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

1   INTRODUCTION

Simulation models are increasingly being used to solve problems 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 verification and validation. 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. 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. 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 identifying 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) and specifying the required acceptable range of accuracy for each variable. The acceptable range of accuracy for each model variable of interest is usually specified as the range that the difference between that model variable and the corresponding system variable can have for the model to be valid. 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 verification and validation, 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, 1984a). 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. Figure 1 shows the relationships between model confidence and (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. The cost of model validation is usually quite significant, especially when extremely high model confidence is required.
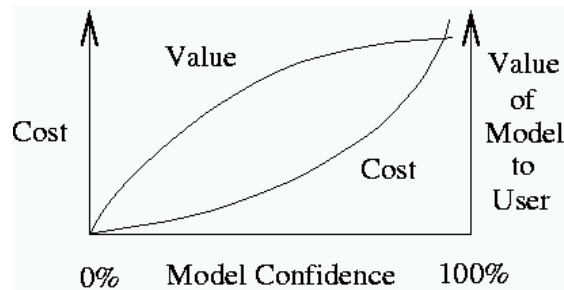

Figure 1: Model confidence

The remainder of this paper is organized as follows: Section 2 presents the basic approaches used in deciding model validity, Section 3 describes a graphical paradigm used in verification and validation, and Section 4 defines validation techniques. Sections 5, 6, 7, and 8 discuss data validity, conceptual model validity, computerized model verification, and operational validity, respectively and 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 decision-making approaches for deciding whether a simulation model is valid. Each of the approaches requires the model development team to conduct verification and validation as part of the model development process, which is discussed in Section 3. 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. It is usually better, however, to use one of the next two approaches for determining model validity.

If the size of the simulation team developing the model is small, a better approach than the one above is to have the user(s) of the model heavily involved with the model development team in deciding the validity of the simulation model. In this approach the focus of determining 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 useful 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 inputs 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 simulation 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 model 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; e.g., by the model development team.

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 resulting in this approach appearing 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

There are two common ways to view how verification and validation relate to the model development process. One way uses a simple view and the other uses a complex view. Banks, Gerstein, and Searles (1988) reviewed paradigms using both of these ways and concluded that the simple view more clearly illuminates model verification and validation. In this section we present a simple graphical paradigm developed by this author. A more complex paradigm developed by this author that includes both the "Simulation World" and the "Real World" is contained in Sargent (2001b). (A brief description of this more complex paradigm is contained in Sargent (2003, 2005, 2007, 2010b).)

Consider the simplified version of the model development process shown 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*.

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 as-

suring 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.
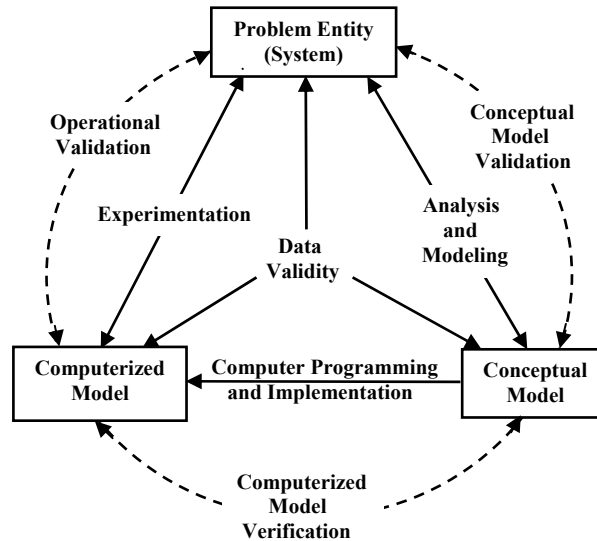

Figure 2: Simplified version of the modeling process

An iterative process is used to develop a valid simulation model (Sargent 1984a). A conceptual model is developed followed by conceptual model validation. This process is repeated until the conceptual model is satisfactory. Next the computerized model is developed from the conceptual model followed by computerized model verification. This process is repeated until the computerized model is satisfactory. Next operational validity is conducted on the computerized model. Model changes required by conducting operational validity can be in either the conceptual model or in the computerized model. Verification and validation must be performed again when any model change is made. Several models are usually developed prior to obtaining a valid simulation model.

4   VALIDATION TECHNIQUES

This section describes validation techniques and tests commonly 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 mathematical procedure or statistical test, e.g., hypothesis tests or confidence intervals. A combination of techniques is generally used. These techniques are used for verifying and validating the submodels and the overall model.

Animation: The model's operational behavior is displayed graphically as the model moves through time, e.g., 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 to the actual number of fires.

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 usually be zero.

Face Validity: Individuals knowledgeable about the system are asked 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 (e.g., data collected on a system specifically for building and 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 1984b).)

Historical Methods: The three historical methods of validation are *rationalism*, *empiricism*, and *positive economics*. Rationalism requires that the assumptions underlying a model be clearly stated and that they are readily accepted. 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's outcome(s) be correct and is not concerned with a model's assumptions or structure (causal relationships or mechanisms).

Internal Validity: Several replications (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 behave correctly.

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. This technique can be used qualitatively—directions only of outputs—and quantitatively—both directions and (precise) magnitudes of outputs. 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 behaviors 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 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 data problems are often 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 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 addition, 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 concerns with data are that appropriate, accurate, and sufficient data are available, and all data transformations, such as data disaggregation, are made correctly. Unfortunately, there is not much that can be done to ensure that the data are correct. One should develop good procedures for (1) collecting and maintaining data, (2) testing the collected data using techniques such as internal consistency checks, and (3) screening the data for outliers *and* determining if the outliers are correct. If the amount of data is large, a database of the data 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 follow a Poisson process. 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?

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 also 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 lan-

guage also usually reduces both the programming time required and the amount of flexibility, and increases the model execution times.)

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 that 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 walkthroughs 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 computerized (simulation) 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 walkthroughs, 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 the numerical values of 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. (See Whitner and Balci (1989) for a detailed discussion on model verification.)

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 techniques used in operational validity based on the decision approach and system observability. "Comparison" means comparing the simulation model output behavior to either the system output behavior or another model output behavior using graphical displays and/or statistical tests and procedures. "Explore model behavior" means to examine the output behavior of the simulation 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.

To obtain a *high* degree of confidence in a simulation model and its results, comparisons of the model's and system's output behaviors for several different sets of experimental conditions are usually required. Thus if a system is not observable, which is often the case, it is usually not possible to obtain a

high degree of confidence in the model. In this situation the model output behavior(s) should be explored as thoroughly as possible and comparisons made to other valid models whenever possible.

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

8.1   Explore Model Behavior

The simulation model output behavior can be explored either qualitatively or quantitatively. In qualitative analysis the directions of the output behaviors are examined and also possibly whether the magnitudes are "reasonable." In quantitative analysis both the directions and the precise magnitudes of the output behaviors are examined. Experts on the system usually know the directions and often know the "general values" of the magnitudes of the output behaviors. Many of the validation techniques given in Section 4 can be used for model exploration. Parameter variability-sensitivity analysis should usually be used. Graphs of the output data discussed in Subsection 8.2.1 below can be used to display the simulation model output behavior. A variety of statistical approaches can be used in performing model exploration including metamodeling and design of experiments. (See Kleijnen (1999) for further discussion on the use of statistical approaches.) Numerous sets of experimental frames should be used in performing model exploration.

8.2   Comparisons of Output Behaviors

There are three basic approaches used in comparing the simulation model output behavior to either the system output behavior or another model output behavior: (1) the use of graphs to make a subjective decision, (2) the use of confidence intervals to make an objective decision, and (3) the use of hypothesis tests to make an objective decision. It is preferable to use confidence intervals or hypothesis tests for the comparisons because these allow for objective decisions. However, it is often 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 required are data independence and normality) and/or (b) there is an insufficient quantity of system data available, which causes the statistical results to be "meaningless" (e.g., the length of a confidence interval developed in the comparison of the system and simulation model means is too large for any practical usefulness). As a result, the use of graphs is the most commonly used approach for operational validity. Extreme care must be used in using this approach. Each of these three approaches is discussed below using system output data (Note: these same approaches can also use with output data from a validated model instead of system output data when appropriate).

8.2.1 Graphical Comparisons of Data

The behavior data of the simulation 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 3 and 4, respectively; both taken from

Lowery (1996). Examples of behavior graphs, taken from Anderson and Sargent (1974), are given in Figures 5 and 6. A variety of graphs are required that use different types of (1) measures such as the mean, variance, maximum, distribution, and times series of the variables, and (2) relationships between (a) two measures of a single variable (see Figure 5) and (b) measures of two variables (see Figure 6). It is important that appropriate measures and relationships be used in validating a simulation 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.
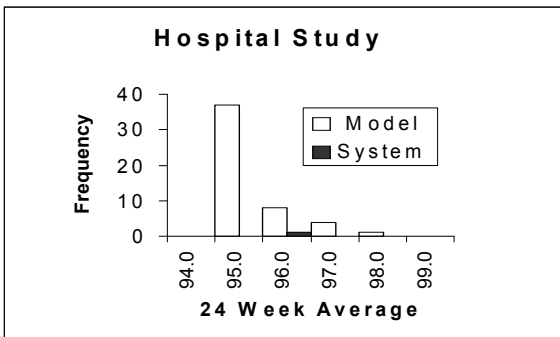

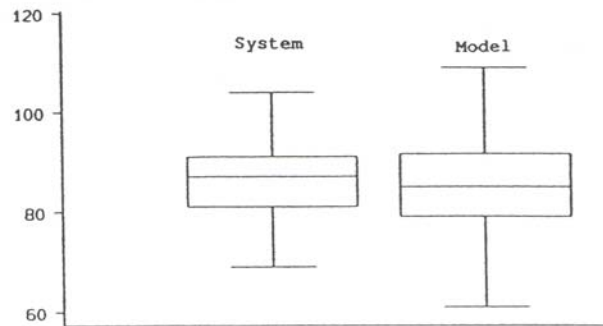Figure 3: Histogram of hospital data

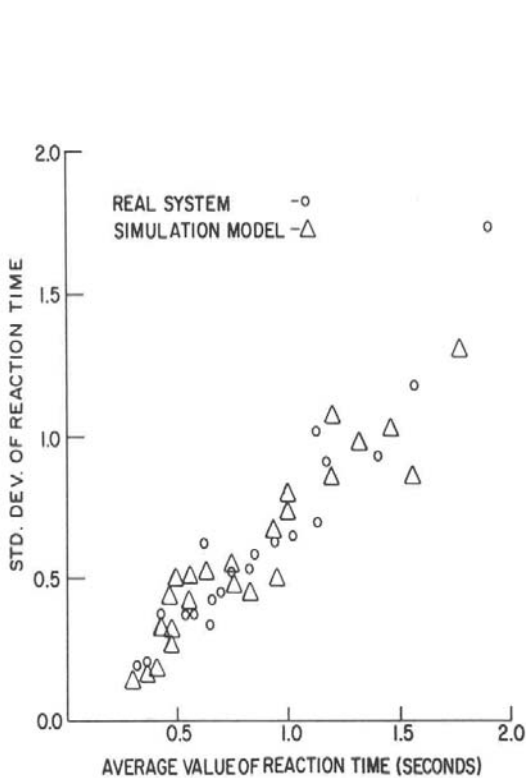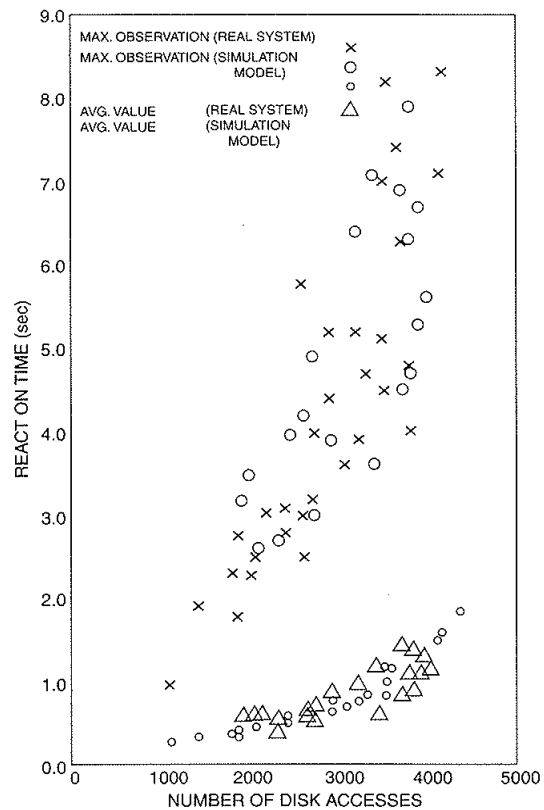Figure 4: Box Plot of hospital data

Figure 5: Reaction time

Figure 6: Disk

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 simulation 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 simulation 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 independent nor satisfy any statistical distribution requirement such as normality of the data (Sargent 1996a, 2001a, 2001b).

8.2.2 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 simulation 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 range of accuracy of a model 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 2007) 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. 2010, Law 2007) can be used in 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 contained in Balci and Sargent (1984b).

8.2.3 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 simulation 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 and is usually specified for each model variable of interest as a range for the difference between that model variable and the corresponding system variable.

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.

Statistical hypothesis tests usually test for a single point. Since the acceptable range of accuracy for each model variable of interest is usually specified as a range, a hypothesis test that uses a range is desired. Recently, a new statistical procedure has been developed for comparisons of model and system outputs using hypothesis tests when the amount of model accuracy is specified as a *range* (Sargent 2010a). This new statistical procedure is applied at each experimental condition to determine if the model is valid for that experimental condition. Both type I and II errors are considered through the use of the operating characteristic curve (Johnson, Miller, and Freund 2010; Hines et al. 2003) . Furthermore the model builder's and the model user's risk curves can be developed using the procedure. This procedure allows a trade-off to be made between the two risks for fixed sample sizes and for trade-offs among the two risks and variable sample sizes. See Sargent (2010a) for details of performing this new procedure.

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. (See Gass (1984) for a general discussion on documentation of computer-based models.) 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.

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

10   RECOMMENDED PROCEDURE

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

1. An agreement be made prior to developing the model between (a) the model development team 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 simulation 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 simulation model.
4. In each model iteration, perform at least face validity on the conceptual model.
5. In each model iteration, at least explore the simulation model's behavior using the computerized model.
6. In at least the last model iteration, make comparisons, if possible, between the simulation 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 simulation model documentation.
8. If the simulation model is to be used over a period of time, develop a schedule for periodic review of the model's validity.

Some simulation 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 simulation 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" (DoD 2003). The evaluation for accreditation is usually conducted by a third (independent) party, is subjective, and often includes not only verification and validation but such items as documentation and how user friendly the simulation is. The acronym VV&A is used for Verification, Validation, and Accreditation. (Other areas and fields sometimes 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 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 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.

In this paper we discussed 'practical approaches' to verification and validation of simulation models. For a discussion on the philosophy of model validation, see Kleindorfer and Ganeshan (1993).

There is considerable literature on model verification and validation; see, e.g., Balci and Sargent (1984a). Beyond the references already cited above, there are conference tutorials and papers (e.g., Sargent (1979, 1984b, 1990, 2000)), journal articles (e.g., Gass (1983), Landry, Malouin, and Oral (1983)), discussions in textbooks (e.g., Banks et al. (2010), Law (2007), Robinson (2004), Zeigler (1976)), U.S.A. Government Reports (e.g., U. S. General Accounting Office (1987)), and a book by Knepell and Arangno (1993) that can be used to further your knowledge on model verification and validation.

Research continues on these topics. This includes such items as advisory systems (e.g. Balci (2001) and Rao and Sargent (1988)), new approaches and procedures (e.g. Balci 2004, Sargent 2010b), and new

techniques (e.g. Balci et al. (2002), Rueβ and de Moura (2003)). See Sargent et al. (2000) for a discussion on research directions.

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.*, edited by E. A. MacNair, K. J. Musselman, and P. Heidelberger, 62-71. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Balci, O. 2001. "A Methodology for Certification of Modeling and Simulation Applications." *ACM Trans. on Modeling and Computer Simulation* 11(4):352-377.

Balci, O. 2003. "Validation, Verification, and Certification of Modeling And Simulation Applications." In *Proc. 2003 Winter Simulation Conf.*, edited by S. Chick, P. J. Sanchez, E. Ferrin, and D. J. Morrice, 150-158. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Balci, O. 2004. "Quality Assessment, Verification, and Validation of Modeling and Simulation Applications." In *Proc. 2004 Winter Simulation Conf.*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and H. A. Peters, 122-129. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Balci, O., R. E. Nance, J. D. Arthur, and W. F. Ormsby. 2002. "Expanding Our Horizons in Verification, Validation, and Accreditation Research and Practice." In *Proc. 2001 Winter Simulation Conf.,* edited by B. A. Peters, J. S. Smith, D. J Medeiros, and M. W. Rohrer, 653-663. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

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. 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. 2010. *Discrete-Event System Simulation*. 5th 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, 19(1):13-18. San Diego, CA: Society for Modeling and Simulation International.

DoD. 2003. *DoDI 5000.61: DoD Modeling And Simulation Verification, Validation, And Accreditation.* U.S. Department of Defense. May 13, 2003.

Fairley, R. E. 1976. "Dynamic Testing of Simulation Software." In *Proc. 1976 Summer Computer Simulation Conf.*, 40-46. San Diego, CA: Society for Modeling and Simulation International.

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.

Hines, W. W., D. C. Montgomery, D. M. Goldsman, and C. M. Borror. 2003. *Probability and Statistics for Engineers*. 4th ed. New York: John Wiley.

Johnson, R. A., I. Miller, and J. E. Freund. 2010. *Miller and Freund's Probability And Statistics For Engineers*. 8th 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 Simulation Conf.*, edited by P. A. Farrington, H. Black Nembhard, D. T. Sturrock, and G. W. Evans, 647-654. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Kleindorfer, G. B., and R. Ganeshan. 1993. "The Philosophy of Science and Validation in Simulation." In *Proc. 1993 Winter Simulation Conf.*, edited by G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles, 50-57. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

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

Landry, M., J. L. Malouin, and M. Oral. 1983. "Model Validation in Operations Research." *European Journal of Operational Research* 14: 207-220.

Law, A. M. 2007. *Simulation Modeling And Analysis*. 4th ed. New York: McGraw-Hill.

Lowery, J. 1996. "Design of Hospital Admissions Scheduling System Using Simulation." In *Proc. 1996 Winter Simulation Conf.*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 1199-1204. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Naylor, T. H., and J. M. Finger. 1967. "Verification of Computer Simulation Models." *Management Science* 14(2):B92-B101.

Rao, M. J., and R. G. Sargent. 1988. "An Advisory System For Operational Validity." In *Artificial Intelligence and Simulation: the Diversity of Applications*, edited by T. Hensen, 245-250. San Diego, CA: Society for Computer Simulation.

Robinson, S. 2004. *Simulation: The Practice of Model Development and Use*. Chichester, West Sussex, England: John Wiley.

Rueβ, H., and L. de Moura. 2003. "From Simulation to Verification (and Back)." In *Proc. 2003 Winter Simulation Conf.*, edited by S. Chick, P. J. Sanchez, E. Ferrin, and D. J. Morrice, 888-896. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 1979. "Validation of Simulation Models." In: *Proc. 1979 Winter Simulation Conf.*, edited by H. J. Highland, M. G. Spiegel, and R. E. Shannon, 497-503. Piscataway, NJ: Institute of Electrical and Electronic Engineers Inc.

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*, edited by F. E. Cellier, 159-169. London: Academic Press.

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

Sargent, R. G. 1984b. "A Tutorial on Verification and Validation of Simulation Models." In *Proc. 1984 Winter Simulation Conf.*, edited by S. Sheppard, U. W. Pooch, and C. D. Pegden, 114-121. Piscataway, NJ: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 1986. "The Use of Graphical Models in Model Validation." In *Proc. 1986 Winter Simulation Conf.*, edited by J. R. Wilson, J. O. Henriksen, and S. D. Roberts, 237-241. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

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.*, edited by J. D. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, 77-87. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 1996a. "Some Subjective Validation Methods Using Graphical Displays of Data." In *Proc. 1996 Winter Simulation Conf.*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 345-351. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 1996b. "Verifying and Validating Simulation Models." In *Proc. 1996 Winter Simulation Conf.*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 55-64. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 2000. "Verification, Validation, and Accreditation of Simulation Models." In *Proc. 2000 Winter Simulation Conf.*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 50-59. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

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)*, edited by 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.*, edited by B. A. Peters, J. S. Smith, D. J Medeiros, and M. W. Rohrer, 106-114. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 2003. "Verification and Validation of Simulation Models." In *Proc. 2003 Winter Simulation Conf.*, edited by S. Chick, P. J. Sanchez, E. Ferrin, and D. J. Morrice, 37-48. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 2005. "Verification and Validation of Simulation Models." In *Proc. 2005 Winter Simulation Conf.*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 130-143. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 2007. "Verification and Validation of Simulation Models." In *Proc. 2007 Winter Simulation Conf.*, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. D. Tew, and R. R. Barton, 124-137. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 2010a. "A New Statistical Procedure for Validation of Simulation and Stochastic Models." Technical Report SYR-EECS-2010-06, Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, New York.

Sargent, R. G. 2010b. "Verification and Validation of Simulation Models." In *Proc. 2010 Winter Simulation Conf.*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 166-183. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G., P. A. Glasow, J. P. C. Kleijnen, A. M. Law, I. McGregor, and S. Youngblood. 2000. "Strategic Directions in Verification, Validation, and Accreditation Research." In *Proc. 2000 Winter Simulation Conf.*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 909-916. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Schlesinger, S., R. E. Crosbie, R. E. Gagné, G. S. Innis, C. S. Lalwani, J. Loch, R. J. Sylvester, R. D. Wright, N. Kheir, and D. Bartos. 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. 1987. "DOD Simulations: Improved Assessment Procedures would Increase the Credibility of Results." Report ID: PEMD-88-3.

Whitner, R. G., and O. Balci. 1989. "Guideline for Selecting and Using Simulation Model Verification Techniques." In *Proc. 1989 Winter Simulation Conf.*, edited by E. A. MacNair, K. J. Musselman, and P. Heidelberger, 559-568. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Wood, D. O. 1986. "MIT Model Analysis Program: What We have Learned About Policy Model Review." In *Proc. 1986 Winter Simulation Conf.*, edited by J. R. Wilson, J. O. Henriksen, and S. D. Roberts, 248-252. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

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 and chairing the Board for two years, being a Department Editor for the *Communications of the ACM*, holding the Presidency and other offices of what is now the INFORMS Simulation Society, and serving as Founding President of the WSC Foundation. He has received several awards and honors for his professional contributions including the INFORMS Simulation Society's Lifetime Professional Achievement Award and also their Distinguished Service Award, a WSC 40th anniversary landmark paper award, the WSC Wilson Award, service awards from ACM and IIE, and becoming a Fellow of INFORMS. 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 and in Who's Who in the World. His e-mail is rsargent@syr.edu .