*Proceedings of the 1991 Winter Simulation Conference*
*Barry L. Nelson, W. David Kelton, Gordon M. Clark (eds.)*

# SIMULATION MODEL VERIFICATION AND VALIDATION*

Robert G. Sargent

Simulation Research Group
449 Link Hall
Syracuse University
Syracuse, New York 13244

## ABSTRACT

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

## 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 derived from the results of the models, and people effected by decisions based on such models are all rightly concerned with whether a model and its results are "correct". This concern is addressed through model verification and validation. Model validation is usually defined to mean "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" (Schlesinger, et al. 1979) and is the definition used here. Model verification is often defined as "ensuring that the computer program of the computerized model and its implementation is correct", and is the definition adopted here. A related topic is model credibility (or acceptability) which is developing in the (potential) users of information from the models (e.g., decision-makers) sufficient confidence in the information that they are willing to use it.

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. (Different models of the same system are sometimes developed for different purposes.) Several sets of experimental conditions are usually required to define the domain of a model's intended applicability. A model may be valid for one set of experimental conditions and be invalid in another. A model is considered valid for a set of experimental conditions if its accuracy is within its acceptable range of accuracy which is the amount of accuracy required for the model's intended purpose. This generally requires that the variables of interest, i.e. the variables used in answering the questions in the purpose of the model, be identified and their required accuracy determined. If the variables of interest are random variables, then properties and functions of the random variables such as their means and variances are frequently what is of primary interest and are what are used in determining model validity. Several versions of a model are usually developed prior to obtaining a satisfactory valid model. The substantiation that a model is valid, i.e. model and verification validation, is generally considered to be a process and is usually part of the model development process.

It is often too costly and time consuming to determine that a model is **absolutely** valid over the complete domain of its intended applicability. Instead, tests and evaluations are conducted until sufficient confidence is obtained that a model can be considered valid for its intended application (Sargent 1982, 1984 and Shannon 1975, 1981). The relationships of cost (and a similar relationship holds for the amount of time) of performing model validation and the value of the model to the user as a function of model confidence are illustrated in Figure 1. The cost of model validation is usually quite significant; in particular where extremely high confidence is required because of the consequence of an invalid model.

* This paper is an updated version of "A Tutorial on Validation and Verification of Simulation Models", *Proceedings of the 1988 Winter Simulation Conference Conference*, pp 33-39.

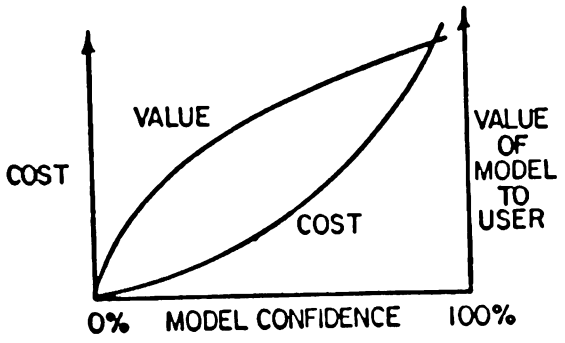


Figure 1: Model Confidence

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

## 2 VALIDATION PROCESS

There are three basic decision-making approaches used in determining that a simulation model is valid. Each of these approaches require the model development team to conduct verification and validation as part of the model development process and this is discussed below in some detail. The most common decision-making approach is for the model development team to make the decision that the model is valid. This decision is a subjective decision based on the results of the various tests and evaluations conducted as part of the model development process.

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

The evaluation used in the IV&V approach can be as simple as reviewing the verification and validation performed by the model development team or it may involve 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 (1986) makes is that a complete IV&V evaluation is extremely costly and time consuming for what is gained. This author's view is that if a third party is to be used, they should be used and involved **during** the model development process. If the model has already been developed, this author believes that a third party should usually only evaluate what verification and validation has already been performed and not repeat earlier work. (Also see Davis (1986) for an approach that simultaneously specifies and validates a model).

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

This author does not believe in the use of a scoring model for determine validity. One reason is that the subjectiveness of this approach tends to be hidden and thus it appears to be objective. A second reason is "how are passing scores" to be determined. A third reason is that a model may receive a passing score and yet have a defect that needs correction. A fourth reason is that the score(s) may cause over confidence in a model or be used to argue one model is better than another.

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

Consider the simplified version of the modelling process in Figure 2. The *problem entity* is the system (real or proposed), idea, situation, policy, or phenomena to be modelled; 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 modelling 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*.

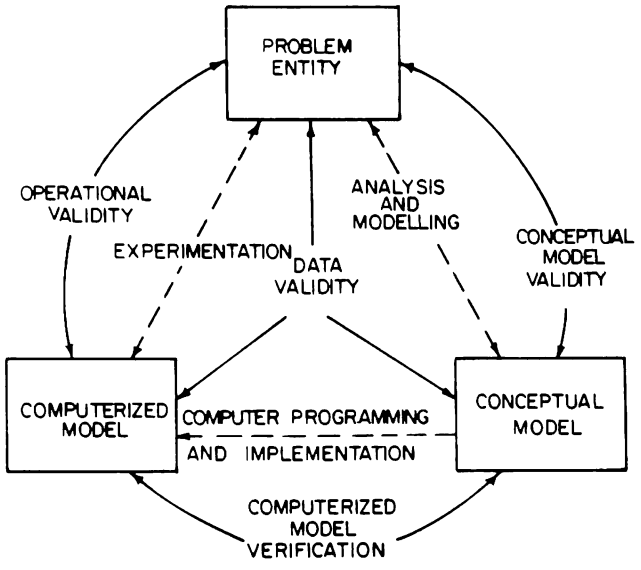


Figure 2: Simplified Version of the Modelling Process

We will now relate model validation and verification to this simplified version of the modelling process (See Figure 2). *Conceptual model validity* is defined as determining that the theories and assumptions underlying the conceptual model are correct and that the model representation of the problem entity is "reasonable" for the intended purpose of the model. *Computerized model verification* is defined as ensuring that the computer programming and implementation of the conceptual model is correct. *Operational validity* is defined as determining that the model's output behavior has sufficient accuracy for its intended purpose over the domain of the model's intended applicability. *Data validity* is defined as ensuring that the data necessary for model building, model evaluation and testing, and conducting the model experiments to solve the problem are adequate and correct.

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

## 3 VALIDATION TECHNIQUES

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

*Animation (Operational Graphics)*: The model's operational behavior is displayed graphically as the model moves through time. Examples are (i) the graphical plot of the status of a server over time, e.g., is it busy, idle, or blocked, and (ii) the graphical display of parts moving through a factory.

*Comparison to Other Models*: Various results (e.g., outputs) of the simulation model being validated are compared to results of other (valid) models. Examples are (i) simple cases of a simulation model may be compared to known results of analytic models, and (ii) the model may be compared to other (simpler) models that have been validated. (Sometimes the simulation model being validated requires modification to allow comparisons to be made.)

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

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

*Extreme-Condition Tests:* The model structure and output should be plausible for any extreme and unlikely combination of levels of factors in the system, e.g., if in-process inventories are zero, production output should be zero. Also, the model should be bound to restrict the behavior outside of normal operating ranges.

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

*Fixed Values:* Fixed values are used for all model input and internal variables. This should allow checking the model results against hand calculated values.

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

1984b).)

*Historical Methods:* The three historical methods of validation are *Rationalism, Empiricism, and Positive Economics*. Rationalism assumes that everyone knows whether the underlying assumptions of a model are true. Then 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 replications (runs) of a stochastic model are made to determine the amount of internal stochastic variability in the model. A high amount of variability (lack of consistency) may cause the model's results to be questionable, and, if typical of the problem entity, may question the appropriateness of the policy or system being investigated.

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

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

*Predictive Validation:* The model is used to predict (forecast) the system behavior and comparisons are made to determine if the system behavior and the model's forecast are the same. The system data may come from an operational system or from experiments performed 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*: People who are knowledgeable about the operations of a system are asked if they can discriminate between system and model outputs. (Schruben (1980) contains statistical procedures for Turing tests.)

## 4 DATA VALIDITY

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

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

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

## 5 CONCEPTUAL MODEL VALIDATION

Conceptual model validity is determining that the theories and assumptions underlying the conceptual model are correct, and that the model representation of the problem entity and the model's structure, logic, and mathematical and causal relationships are "reasonable" for the intended purpose of the model. The theories and assumptions underlying the model should be tested using mathematical analysis and statistical methods on problem entity data. Examples of theories and assumptions are linearity, independence, stationary, and Poisson arrivals. Examples of applicable statistical methods are fitting distributions to data; estimating parameter values, mean, variance, and correlations among data observations; and plotting data to see if it is stationary. In addition, all theories used should be

reviewed to ensure they were applied correctly; for example, if a Markov chain is used, does the system have the Markov property and are the states and transition probabilities correct?

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

## 6 COMPUTERIZED MODEL VERIFICATION

Computerized model verification is ensuring that the computer programming and implementation of the conceptual model is correct. To help ensure that a correct computer program is obtained, program design and development procedures found in the field of Software Engineering should be used in developing and implementing the computer program. These include such techniques as top-down design, structured programming, and program modularity. A separate program module should be used for each submodel, the overall model, and for each simulation function (e.g., time-flow mechanism, random number and random variate generators, and integration routines) when using general purpose higher order languages, e.g., FORTRAN or PASCAL, and where possible when using simulation languages (Chattergy and Pooch 1977). (See Whitner and Balci (1989) for a more detailed discussion on model verification. )

One should be aware that the use of different types of computer languages effects the probability of having a correct program. The use of a special purpose simulation language, if appropriate, generally will result in having less errors than if a general purpose simulation language is used, and using a general purpose simulation language will generally result in having less errors than if a general purpose higher order language is used. Not only does the use of simulation languages increase the probability of having a correct program, they usually reduce programming time.

After the computer program has been developed, implemented, and hopefully most of the programming "bugs" removed, the program must be tested for correctness and accuracy. First, the simulation functions should be tested to see if they are correct. Usually straightforward tests can be used here to determine if they are working properly. Next, each submodel and the overall model should be tested to see if they are correct. Here the testing is much more difficult. There are two basic approaches to testing: static and dynamic testing (analysis) (Fairley 1976). In static testing the computer program of the computerized model is analyzed to determine if it is correct by using such techniques as correctness proofs, structured walk-through, and examining the structure properties of the program. The commonly used structured walk-through technique consists of each program developer explaining their computer program code statement by statement to other members of the modelling team until all are convinced it is correct (or incorrect).

In dynamic testing, the computerized model is executed under different conditions, and the values obtained are used to determine if the computer program and its implementations are correct. This includes both the values obtained during the program execution and the final values obtained. There are three different strategies to use in dynamic testing: bottom-up testing which means, e.g., testing the submodels first and then the overall model; top-down testing which means, e.g., testing the overall model first using programming stubs (sets of data) for each of the submodels and then testing the submodels; and mixed testing, which is using a combination of bottom-up and top-down testing (Fairley 1976). The techniques commonly used in dynamic testing are traces, investigations of input-output relations using the 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 to reduce the number of tests needed or use certain types of design of experiments (Kleijnen 1987), e.g., factor screening experiments (Smith and Mauro 1982) to identify the key variables, in order to reduce the number of experimental conditions that need to be tested.

One must continuously be aware while checking the correctness of the computer program and its implementation, that errors may be caused by the data, the conceptual model, the computer program, or the computer implementation.

## 7 OPERATIONAL VALIDITY

Operational validity is primarily concerned with

determining that the model's output behavior has the accuracy required for the model's intended purpose over the domain of its intended applicability. This is where most of the validation testing and evaluation takes place. The computerized model is used in operational validity and thus any deficiencies found may be due to an inadequate conceptual model, an improperly programmed or implemented conceptual model (e.g., due to programming errors or insufficient numerical accuracy), or due to invalid data.

All of the validation techniques discussed in Section 3 are applicable to operational validity. Which techniques and whether to use them objectively or subjectively must be decided by the model development team and other interested parties. The major attribute effecting operational validity is whether the problem entity (or system) is observable or not, where observable means it is possible to collect data on the operational behavior of the program entity. Figure 3 gives one classification of the validation approaches for operational validity. The "explore model behavior" means to examine the behavior of the model using appropriate validation techniques for various sets of experimental conditions from the domain of the model's intended applicability and usually includes parameter variability-sensitivity analysis.

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

| | OBSERVABLE SYSTEM | NON-OBSERVABLE SYSTEM |
|---|---|---|
| SUBJECTIVE APPROACH | * COMPARISON OF DATA USING GRAPHICAL DISPLAYS<br>* EXPLORE MODEL BEHAVIOR | * EXPLORE MODEL BEHAVIOR<br>* COMPARISON TO OTHER MODELS |
| OBJECTIVE APPROACH | * COMPARISON OF DATA USING STATISTICAL TESTS AND PROCEDURES | * COMPARISON TO OTHER MODELS USING STATISTICAL TESTS AND PROCEDURES |

Figure 3: Operational Validity Classification

## 7.1 Graphical Comparison of Data

The model's and system's behavior data are plotted on graphs for various sets of experimental conditions to determine if the model's output behavior has sufficient accuracy for its intended purpose. (See Figures 4 and 5 for examples of such graphs.) A variety of graphs using different types of measures and relationships are required. Examples of measures and relationships are (i) time series, means, variances, and maximums of each output variable, (ii) relationships between parameters of each output variable, e.g., means and standard deviations, and (iii) relationships between different output variables. 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. As an example of a set of graphs used in the validation of a model, see Anderson and Sargent (1974).

These graphs can be used in model validation in three ways. First, the model development team can use the graphs in the model development process to make a subjective judgement on whether the model does or does not possess sufficient accuracy for its intended purpose.

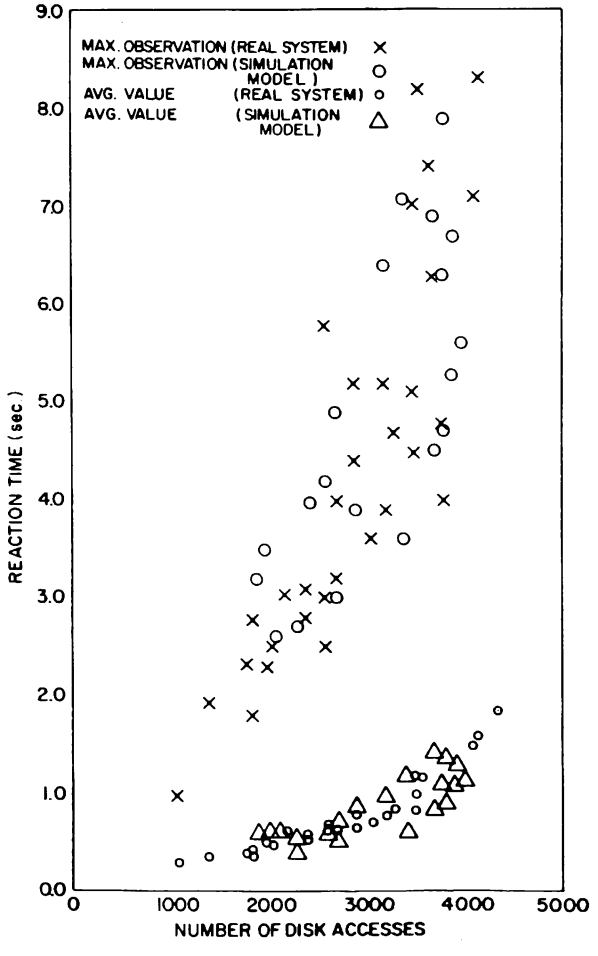


Figure 4: Disk Access

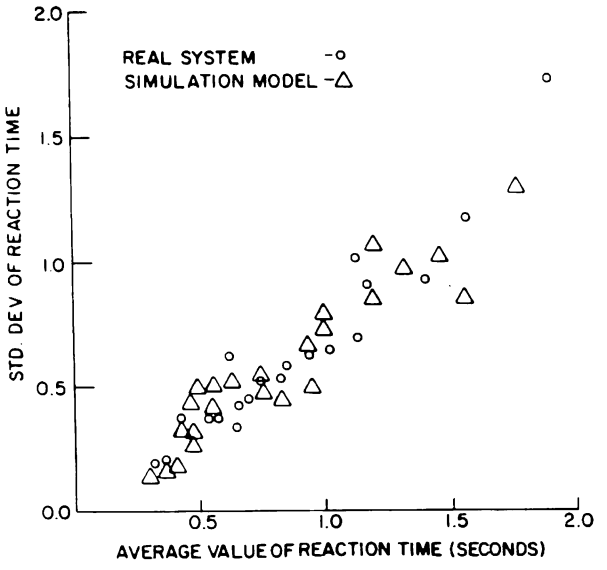


Figure 5: Reaction Time

Secondly, they can be used in the face validity technique where experts are asked to make subjective judgements on whether a model does or does not possess sufficient accuracy for its intended purpose.

The third way the graphs can be used is in Turing Tests. Sets of data from the model and from the system are plotted on separate graphs. The graphs are shuffled and then experts are asked to determine which graphs are from the system and which are from the model. The results for each measure and relationship can be evaluated either subjectively or statistically. The subjective method requires that a subjective decision be made whether the results are satisfactory or not. The statistical method requires that the results be analyzed statistically. See Schruben (1980) for a variety of statistical methods for analyzing the results of Turing Tests and examples of their use.

## 7.2 Confidence Intervals

Confidence intervals (c.i.), simultaneous confidence intervals (s.c.i.), and joint confidence regions (j.c.r.) can be obtained for the differences between the population parameters, e.g., means and variances, and distributions of the 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 parameter of interest. The statistical techniques used can be divided into two groups: (A) univariate statistical techniques and (B) multivariate statistical techniques. The univariate techniques can be used to develop c.i. and with the use of the Bonferroni inequality (Law and Kelton 1991) s.c.i. The multivariate techniques can be used to develop s.c.i. and j.c.r. Both parametric and nonparametric techniques can be used.

The method of data collection must satisfy the underlying assumptions of the statistical technique being used. The standard statistical techniques and data collection methods used in simulation output analysis can be used for developing the model range of accuracy; namely (1) replication, (2) batch means, (3) regenerative, (4) spectral, (5) time series, and (6) standardized time series (Banks and Carson 1984, Law and Kelton 1991).

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 specification of the model range of accuracy will usually be. The lengths and the sizes of the joint confidence regions are affected by the values of confidence levels, variances of the model and system response variables, and sample sizes. The lengths can be shortened or sizes made smaller by decreasing the confidence levels. Variance reduction techniques (Law and Kelton 1991) can be used when collecting observations from a simulation model to decrease the variability and thus obtain a smaller range of accuracy. The lengths can also be shortened or the size decreased by 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. In those cases where the cost of data collection is significant for either the model or system, the data collection cost should also be considered in the tradeoff analysis. Tradeoff curves can be constructed to aid in the tradeoff analysis. Figure 6 is an example of a set of tradeoff curves which contain the relationship between the significance level, $\gamma$, estimated half lengths of the confidence interval, and cost of data collection.

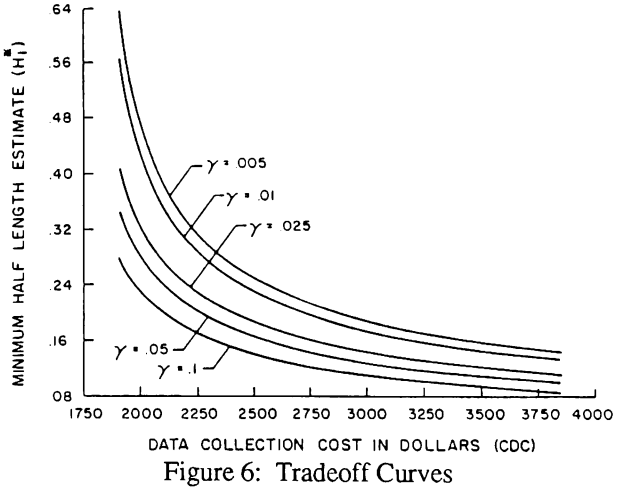


Figure 6: Tradeoff Curves

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

### 7.3. Hypothesis Tests

Hypothesis tests can be used in the comparison of parameters, distributions, and time series of the output data 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. (1)

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

Two types of errors are possible in testing the hypotheses in (1). The first or type I error is rejecting the validity of a valid model; the second or type II error is accepting the validity of an invalid model. The probability of a type error I is called *model builder's risk* ($\alpha$) and the probability of type II error is called *model user's risk* ($\beta$). In model validation, model user's risk is extremely important and must be kept small. Thus *both* type I and type II errors must be considered in using hypothesis testing for model validation.

The amount of agreement between a model and a system can be measured by a validity measure. The validity measure is chosen such that the model accuracy or the amount of agreement between the model and the system decreases as the value of the validity measure increases. The acceptable range of accuracy can be used to determine an acceptable validity range, $0 \le \lambda \le \lambda^*$.

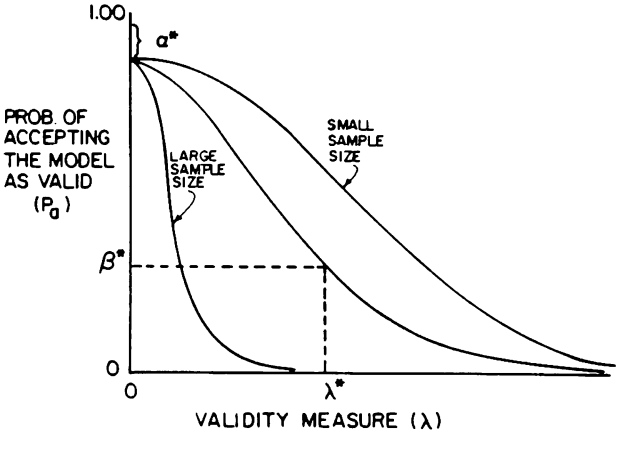


Figure 7: Operating Characteristic Curves

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 (Miller and Freund 1985). Figure 7 contains three different operating characteristic curves to illustrate how the sample size of observations affect $P_a$ as a function of $\lambda$. As can be seen, an inaccurate model has a high probability of being accepted if a small sample size of observations are used and an accurate model has a low probability of being accepted if a large sample size of observations are used. The location and shape of the operating characteristic curves is a function of the statistical technique being used, the value of $\alpha$ chosen for $\lambda$=0, $\alpha^*$, 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 $\alpha$ can be determined for a given $\lambda$ as follows:

$$\alpha \le \text{model builder's risk } \alpha \le (1-\beta^*) \quad (2)$$
$$0 \le \text{model user's risk } \beta(\lambda) \le \beta^*.$$

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

In those cases where the data collection cost is significant for either the model or system, the data collection cost should also be considered in performing the tradeoff analysis. A cost model for data collection should be developed as a function of the sample sizes of observations for the model and the system. An optimization problem can be formulated and solved to determine the optimum sample sizes for a given data collection budget and statistical test to minimize the model user's risk.

Data can be generated for different values of the tradeoff parameters and placed in schedules (tables) which can be used to generate two different types of curves to

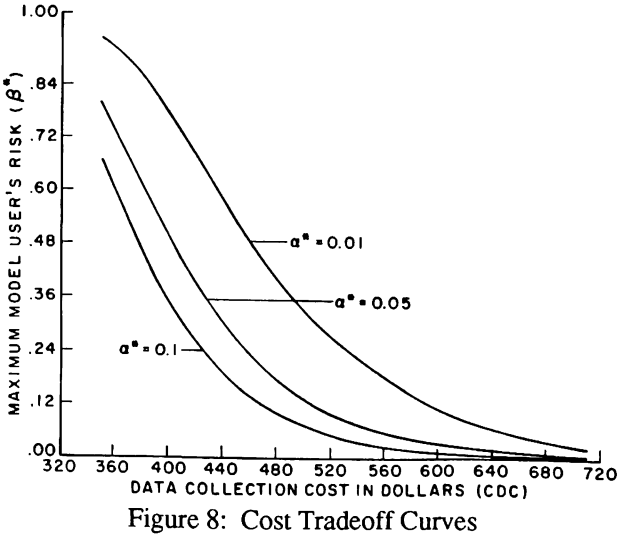


Figure 8: Cost Tradeoff Curves

be used in the tradeoff analysis. One type of curve is the operating characteristic curve shown in Figure 7. The other type of curve is shown in Figure 8. The latter curve shows the relationships among $\alpha^*$, $\beta^*$, and the data collection budget (or sample sizes if data collection cost is not considered) for a given model accuracy. These curves can be used by the model development team, model sponsor, or both to aid in making judgement decisions in regard to determining the appropriate values to use in testing the validity of a model for a given set of experimental conditions.

Details of the methodology of using Hypothesis Tests in comparing 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) and in Banks and Carson (1984).

## 8 DOCUMENTATION

Documentation on model verification and validation is usually critical in convincing users of the "correctness" of a model and its results, and should be included in the simulation model documentation (For a general discussion on documentation of computer-based models, see Gass (1984).) Both detailed and summary documentation is desired. The detailed documentation should include specifies on the tests, evaluations used, 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 Figure 9 for an example of one of the first four tables and Figure 10 for an example of the overall summary. The columns of the tables are self explanatory except for the last column which refers to the confidence the evaluators have in the results or conclusions and is usually expressed as low, medium, or high.

## 9 RECOMMENDED MODEL VALIDATION PROCEDURE

There are currently no algorithms or procedures available to identify specific validation techniques, statistical tests, etc., to use in the validation process. Various authors suggest, e.g., Shannon (1975), that as a minimum the three steps of (1) Face Validity, (2) Testing of the Model Assumptions, and (3) Testing of Input-Output Transformations be made. These recommendations are made in general and are not related to the steps of the model development process discussed in Section 2.

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

(1)An agreement be made between (i) the model development team and (ii) the model sponsors and users (if possible) on the basic validation approach and on a minimum set of specific validation techniques to be used in the validation process ***prior*** to developing the model.

(2) The assumptions and theories underlying the model be tested, when possible.

(3) In each model iteration, at least face validity be performed on the conceptual model.

| Category/Item | Technique(s) Used | Justification for technique used | Reference to supporting report | Result/ conclusion | Confidence in result |
|---|---|---|---|---|---|
| • *Theories*<br>• *Assumptions*<br>• *Model representation* | • *Face validity*<br>• *Historical*<br>• *Accepted approach*<br>• *Derived from empirical data*<br>• *Theoretical derivation* | | | | |

| | |
|---|---|
| **Strengths** | |
| **Weaknesses** | |

| **Overall evaluation for Conceptual Model Validity** | *Overall conclusion* | *Justification for conclusion* | *Confidence in conclusion* |
|---|---|---|---|

Figure 9: Evaluation Table for Conceptual Model Validity

| EVALUATION AREA | OVERALL CONCLUSION | JUSTIFICATION FOR OVERALL CONCLUSION | CONFIDENCE IN CONCLUSION |
|---|---|---|---|
| • *Conceptual Model Validity*<br>• *Computer Model Verification*<br>• *Operational Validity*<br>• *Data Validity* | | | |

| | |
|---|---|
| **Overall Strengths** | |
| **Overall Weaknesses** | |

| **Overall Summary Simulation Evaluation** | *Summary conclusion* | *Justification for summary conclusion* | *Confidence in conclusion* |
|---|---|---|---|

Figure 10: Evaluation Table for Overall Summary

(4) In each model iteration, explore the model's behavior using the computerized model.
(5) In at least the last model iteration, comparisons be made between the model and system behavior (output) data for at *least* two sets of experimental conditions, when possible.
(6) Validation described in the model documentation.
(7) If the model is to be used over a period of time, a schedule for periodic review (and possible revalidation) of it be made and followed.

Models occasionally are developed to be used more than once. A procedure for reviewing the validity of these models over their life cycles needs to be developed as specified by step (7). No general procedure can be given as each situation is different. For example, if no data were available on the problem entity (e.g., a system) when the model was initially developed and validated, then revalidations of it should take place prior to each time the model is used if additional data or system understanding has occurred since its last validation.

## 10 SUMMARY

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

There is considerable literature on verification and validation (Balci and Sargent (1984a)). Articles given in the limited bibliography can be used as a starting point for furthering your knowledge on verification and validation.

## LIMITED BIBLIOGRAPHY

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

Balci, O. 1989. How to Assess the acceptability and Credibility of Simulation Results, *Proceedings of the 1989 Winter Simulation Conference*, edited by MacNair, Musselman, and Heidelberger, Washington, D.C. pp. 62-71.

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

Balci, O. and R.G. Sargent. 1982a. Validation of Multivariate Response Simulation Models by Using Hotelling's Two-Sample $T^2$ Test, *Simulation*, Vol. 39, No. 6, pp. 185-192.

Balci, O. and R.G. Sargent. 1982b. Some Examples of Simulation Model Validation Using Hypothesis Testing, *Proceedings of the 1982 Winter Simulation Conference,* edited by Highland, Chao, and Madrigal, pp. 620-629.

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

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

Balci, O. and R.G. Sargent. 1984b. Validation of Simulation Models via Simultaneous Confidence Intervals, *American Journal of Mathematical and Management Science*, 4, No. 3 and 4, pp. 375-406.

Banks, J. and J.S. Carson II 1984. *Discrete-Event System Simulation*, Prentice-Hall, Englewood Cliffs, N.J.

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

Chattergy, R. and V.W. Pooch. 1977. Integrated Design and Verification of Simulation Programs, *Computer*, Vol. 10, No. 4, pp. 40-46.

Davis, E.A. 1986. Use of Seminar Gaming to Specify and Validate Simulation Models, *Proceedings of the 1986 Winter Simulation Conference*, edited by Wilson, Henriksen, and Roberts, Washington, D.C., pp. 242-247.

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

Fairley, R.E. 1976. Dynamic Testing of Simulation Software, *Proceedings of the 1976 Summer Computer Simulation Conference*, Washington, D.C., pp. 40-46.

Gass, S.I. 1977. A Procedure for Evaluation of Complex Models, *Proceedings of the First International Conference on Mathematical Modeling*, University of Missouri, pp. 247-257.

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

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

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

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

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

Miller, I. and J.E. Feund. 1985. *Probability and Statistics for Engineers*, Prentice-Hall, Englewood Cliffs, N.J.

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

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

Rao, M.J. and R.G. Sargent. 1988. An Advisory System for Operational Validity, *Artificial Intelligence and Simulation: The Diversity of Applications*, edited by T. Hensen, SCS, San Diego, CA, pp. 245-250.

Sargent, R.G. 1979. Validation of Simulation Models, *Proceedings of the 1979 Winter Simulation Conference*, edited by Highland, H.J., et al., San Diego, California, pp. 497-503.

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

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

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

Sargent, R.G. 1985. An Expository on Verification and Validation of Simulation Models, *Proceedings of the 1985 Winter Simulation Conference*, edited by Gantz, Blais, and Solomon, Dallas, Texas, pp. 15-22.

Sargent, R.G. 1986. The Use of Graphic Models in Model Validation, *Proceedings of the 1986 Winter Simulation Conference*, edited by Wilson, Henriksen, and Roberts, Washington, D.C., pp. 237-241.

Sargent, R.G. 1988. A Tutorial on Validation and Verification of Simulation Models, *Proceedings of 1988 Winter Simulation Conference*, edited by Abrams, Haigh, and Comfort, San Diego, CA, pp. 33-39.

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

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

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

Shannon, R.E. 1981. Tests for the Verification and Validation of Computer Simulation Models, *Proceedings of the Winter Simulation Conference*, edited by Oren, Delfosse, and Shub, pp. 573-577.

Smith, D.E. and C.A. Mauro. 1982. Factor Screening in Computer Simulation, *Simulation*, 38, 2, pp. 49-54.

Whitner, R.B. and O. Balci. 1989. Guidelines for Selecting and Using Simulation Model Verification Techniques, *Proceedings of 1989 Winter Simulation Conference*, edited by MacNair, Musselman, and Heidelberger, Washington, D.C., pp. 559-568.

Wood, D.O.. 1986. MIT Model Analysis Program: What We Have Learned About Policy Model Review, *Proceedings of the 1986 Winter Simulation Conference*, edited by Wilson, Henriksen, and Roberts, Washington, D.C., pp. 248-252.

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