ROBERT G. SARGENT is a Professor of Industrial Engineering and Operations Research and a member of the Computer and Information Science Faculty at Syracuse University. He currently is Department Editor of Simulation Modeling and Statistical Computing for the Communications of the ACM, is a member of the Board of Directors of the Winter Simulation Conference, and is a Director of the Society for Computer Simulation. He has served as chairman of the TIMS College on Simulation and Gaming, as chairman of the Board of Directors of the Winter Simulation Conference and was the General Chairman of the 1977 Winter Simulation Conference. Dr. Sargent was also a coeditor of the Proceedings of the 1976 and the 1977 Winter Simulation Conference. His current research interests are model validation, simulation, performance evaluation, and applied operations research. Professor Sargent received his education at the University of Michigan and is a member of AHM, New York Academy of Sciences, Sigma Xi, ACM, IIE, ORSA, SCS, and TIMS.

Professor Robert G. Sargent
Department of Industrial Engineering
and Operations Research
Syracuse University
Syracuse, N.Y. 13210
(315) 423-4348

Proceedings of the 1984
Winter Simulation Conference
S. Sheppard, U. Pooch, D. Pegden (eds.)

A TUTORIAL ON VERIFICATION AND VALIDATION OF SIMULATION MODELS

Robert G. Sargent
Department of Industrial Engineering
and Operations Research
Syracuse University
Syracuse, New York 13210

ABSTRACT

In this tutorial paper we give a general introduction to verification and validation of simulation models, define the various validation techniques, and present a recommended model validation procedure.

1. INTRODUCTION

Simulation models are often used to aid in decision making and problem-solving. The users of these models are rightly concerned with whether the models and information derived from them can be used with confidence. Model developers address this concern 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 frequently defined as ensuring that the computer program of the computerized model (i.e. the simulator) and its implementation is correct and will be the definition used 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 use and its validity determined with respect to that purpose. Several sets of experimental conditions are usually required to define the domain of the model's intended application. 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 the acceptable range of accuracy which is defined as the amount of accuracy required for the model's intended purpose.

The substantiation that a model is valid, i.e. model validation, is part of the total model development process and is itself a process. This process consists of performing tests and evaluations within the model development process to determine whether a model is valid or not. Usually it is not feasible to determine that a model is absolutely valid over the complete domain of its intended application. 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)].

Recent research [Gass and Thompson (1983), Sargent (1981, 1982, 1984), and Schlesinger et al. (1979)] has related model validation and verification to specific steps of the model development process. We will follow the development of [Sargent (1982, 1984)] and use Figure 1. 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*.

We relate validation and verification to this simplified version of the modelling process as shown in Figure 1. *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 "reasonble" for the intended use 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 or use over the domain of the model's intended application. *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 models are usually developed in the modelling process prior to obtaining a satisfactory valid model. During each model iteration, model validation and verification are performed [Sargent (1984)]. A variety of (validation) techniques are used, which are described below. Unfortunately, no algorithm or procedure exists to select which techniques to use. Some of their attributes are discussed in Sargent (1984).

2. 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 (1980, 1984) and Sargent (1982) for a detailed bibliography], although they 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, goodness-of-fit tests, and confidence intervals.) A combination of techniques is usually used. These techniques are used for verifying and validating both the submodels and the overall model.

Comparison to Other Models: Various results (e.g. outputs) of the simulation model being validated are compared to results of other (valid) models. For

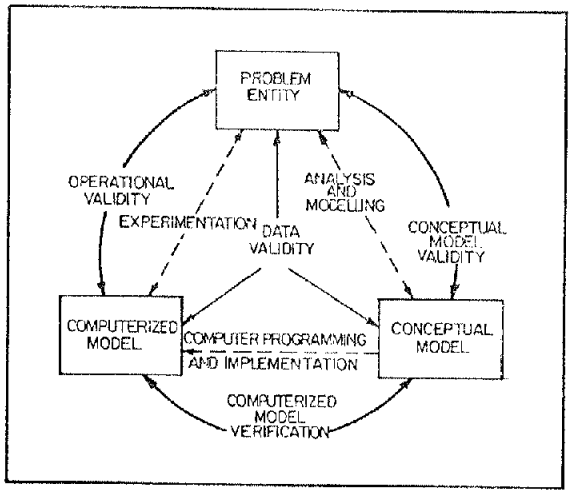

Figure 1: Simplified Version of Modelling Process

example, simple cases of a simulation model may be compared to known results of analytic models. (In some cases, the model being validated may require minor modifications to allow comparisons to be made between it and analytic models.)

Degenerate Tests: The degeneracy of the model's behavior is tested by removing portions of the model or by appropriate selection of the values of the input 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 the 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 bound and 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 model flowchart 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, 1983a)]).

Historical Methods: 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 its 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 intuition, (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 (animation): 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 as the model moves through time, i.e., is it busy, idle, or blocked, and (ii) the graphical display of parts moving through a factory.

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 and its output. The same relationships should occur in the model as in the real system. Those parameters which 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 specific experiments may be performed, 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. (See [Schruben (1980)] for a statistical procedure for Turing Tests.)

3. DATA VALIDITY

Even though data validity is usually not considered

part of model verification and 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 initial attempts to validate a model fail. Basically, data is needed for these purposes: for building the conceptual model, for validating the model, and for performing experiments with the validated model. In model verification and validation, we are concerned only with the first two types of data.

To build a conceptual model, we must have sufficient data about the problem entity in order to develop the mathematical and logical relationships in the model for it to adequately represent the problem entity for its intended use. It is also highly desirable to have data to develop theories that can be used in building the model and to test the model's underlying assumptions. The second type of data desired 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 data; test the collected data using such techniques as internal consistency checks and screening for outliers and determine if any outliers found are correct; and develop good procedures to properly maintain the collected data. If the amount of data is large, a data base should be developed and maintained.

4. 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 use of the model. The theories and assumptions underlying the model should be tested, if possible, 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 date, estimating parameter values, e.g., mean, variance, and the correlation between 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, are the states and transition probabilities correct?

Next, each submodel and the overall model must be evaluated to determine if their abstractions are reasonable and correct for the intended use 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 validiation and traces. Face validation is having an expert or experts of 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 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.

5. 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, and where possible when using simulation languages [Chattergy and Pooch (1977)].

One should be aware that the use of different types of

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

Figure 2: Operational Validity Classification

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 also 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 straight forward 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 analysed 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 execution and the final values. There are three different strategies to use in dynamic testing: bottom-up testing which means testing the submodels first and then the overall model; top-down testing which means 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 (1974, 1982)], 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 in 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.

6. 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 application. 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 can be due to an inadequate conceptual model, an improperly programmed or implemented conceptual model on the computer (e.g., due to programming errors or insufficient numerical accuracy), or due to invalid data.

All of the validation techniques discussed in section 2 are applicable to operational validity. Which techniques and whether to use them objectively or subjectively, must be decided by the model developer 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 2 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

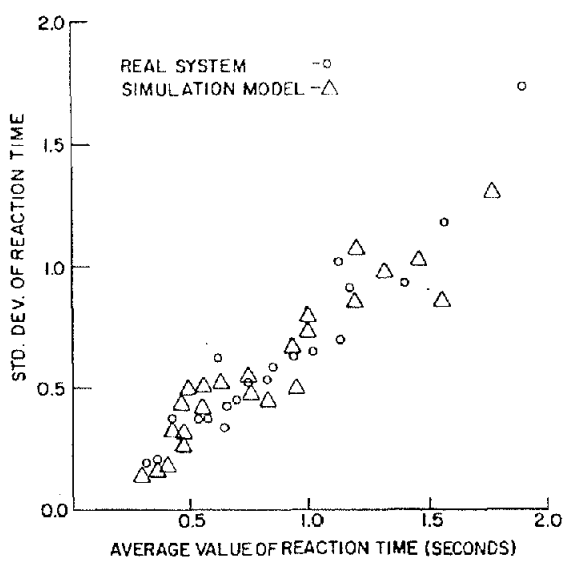

Figure 3: Reaction Time

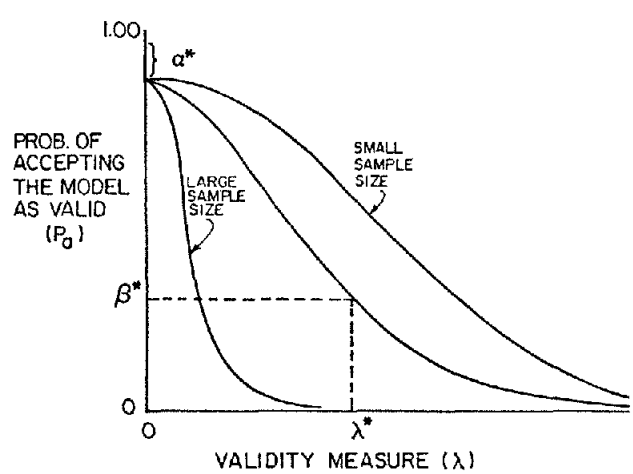

Figure 4: Operating Characteristic Curves

conditions from the domain of the model's intended use.

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) hypothesis tests, and (iii) confidence intervals.

6.1. 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. A variety of graphs showing 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. (See Figure 3 for an example of a graph.) It is important that the 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 are used in model validation in three ways. First, the model developer 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. Secondly, they can be used in the face validity technique where experts are asked to make subjective judgments 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.

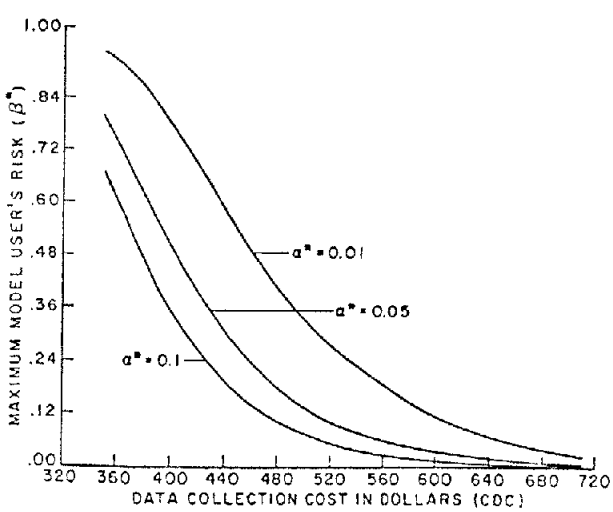

Figure 5: Cost Tradeoff Curves

6.2. 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 hypothesis 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* (α) and the probability of type II error is called *model user's risk* (β). 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*, λ. 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 increase. The acceptable range of accuracy can be used to determine an acceptable validity range, $0 \leq \lambda \leq \lambda^*$.

The probability of acceptance of a model being valid, P_a, can be examined as a function of the validity measure λ by using an Operating Characteristic Curve (Miller and Freund (1977)). Figure 4 contains three different operating characteristic curves to illustrate how the sample size of observations affect P_a as a function of λ. As can be seen, an inaccurate model has a high probability of being accepted if a small sample size of observations 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 α chosen for $\lambda=0$, α^*, and the sample size of observations. Once the operating characteristic curves are constructed, the intervals for the model user's risk $\beta(\lambda)$ and the model builder's risk α can be determined for a given λ^* as follows:

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

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 which can be used to generate two different types of curves to be used in the tradeoff analysis. One type of curve is the operating characteristic curve shown in Figure 4. The other type of curve is shown in Figure 5. The latter curve shows the relationships among α^*, β^*, 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 developer, model sponsor, or both to aid in making judgment 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 (1981a). Examples of the application of this methodology in the testing of output means for model validation are given in Balci and Sargent (1982a, 1982b, 1983a) and in Banks and Carson (1984).

6.3. 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 type of parameters 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 (1982)] 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 used 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 (1982), Law (1983)].

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 (1982)] can be used in some cases 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 analysis 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. Details of a methodology for performing a tradeoff analysis and for using c.i., s.c.i., and j.c.r. can be found in Balci and Sargent (1981b, 1983b). Law and Kelton (1982) also discuss the use of c.i. for model validation.

7. 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 (for example, see Shannon [1975, p. 29] 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 modelling process discussed in the Introduction.

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

(1) An agreement be made between (i) the modelling team and (ii) the model sponsors and users (if possible) on the basic validation approaches 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, if possible.

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

(4) In each model iteration, exploration of the model's behavior be made 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, if possible.

(6) Validation discussed in the model documentation.

8. SUMMARY

Model validation is one of the most critical issues faced by the simulationist. Unfortunately, there is no set of specific tests that can be easily applied to determine the validity of the model. Furthermore, no algorithm exist to determine what techniques or

procedures to use. Every new simulation project presents a new and unique challenge.

There is a considerable literature on verification and validation [Balci and Sargent (1980, 1984)]. Articles given in the references can be used as a starting point for furthering your knowledge on verification and validation.

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, pp. 125-137.

Balci, O. and R.G. Sargent (1980), "Bibliography on Validation of Simulation Models," Newsletter of TIMS College on Simulation and Gaming, 4, 2, pp. 11-15.

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

Balci, O. and R.G. Sargent (1981b), A Methodology for Validating Multivariate Response Simulation Models by Using Simultaneous Confidence Intervals, Working Paper No. 81-009, Dept. of Industrial Engineering and Operations Research, Syracuse University.

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," Proc. Winter Simulation Conference, edited by Highland, Chao and Madrigal, pp. 620-629.

Balci, O. and R.G. Sargent (1983a), "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 (1983b), "Sample Size Selection in Confidence Interval Approach to Simulation Model Validation," Working Paper No. 83-003, Dept. of Industrial Engineering and Operations Research, Syracuse University.

Balci, O. and R.G. Sargent (1984), "A Bibliography on the Credibility, Assessment and Validation of Simulation and Mathematical Models," Simuletter, Vol. 15, No. 3, pp. 15-27.

Banks, J. and J.S. Carson II (1984), Discrete-Event System Simulation, Prentice-Hall, N.J.

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

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

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

Gass, S.I. (1983), "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis," Operaions Research, 31, 4, pp. 601-63.

Kleijnen, J.P.C. (1975), Statistical Techniques in Simulation, Part II, Marcel Dekker, New York.

Kleijnen, J.P.C. (1982), "Experimentation with Models: Statistical Design and Analysis Techniques," Progress in Modelling & Simulation, F.E. Cellier, Editor, Academic Press, London.

Law, A.M. (1983) "Statistical Analysis of Simulation Output Data," Operations Research, Vol. 31, No. 6, pp. 983-1029.

Law, A.M. and W.D. Kelton (1982), Simulation Modeling Analysis, McGraw-Hill Book Company.

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.

Sargent, R.G. (1979), "Validation of Simulation Models" Proceedings of 1979 Winter Simulation Conference, edited by Highland, H.J., et al., San Diego, CA, pp. 497-503.

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

Sargent, R.G. (1982), "Verification and Validation of Simulation Models," Chapter IX in Progress in Modelling and Simulation, 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.

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

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 Winter Simulation Conference, edited by Oren, Delfosse, and Shub, pp. 573-577.

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

Smith, D.E. and C.A. Mauro (1982), "Factor Screening in Computer Simulation," Simulation, Vol. 38, No. 2, pp. 49-54.

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