Proceedings of the 1983
Winter Simulation Conference
S. Roberts, J. Banks, B. Schmeiser (eds.)

VALIDATING SIMULATION MODELS

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

In this paper we give a general introduction to model validation, define the various validation techniques, discuss conceptual and operational validity, and present a recommended model validation procedure.

1. INTRODUCTION

Simulation models are used routinely 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 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. Related to model validation are model verification, which is discussed below, and model credibility or acceptability, which is developing in the (potential) users of information from 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 invalid. 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, 1983) and Shannon (1975, 1981)].

Recent research [Gass and Thompson (1980), Sargent (1981, 1982, 1983), 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, 1983) 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 "reasonable" 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

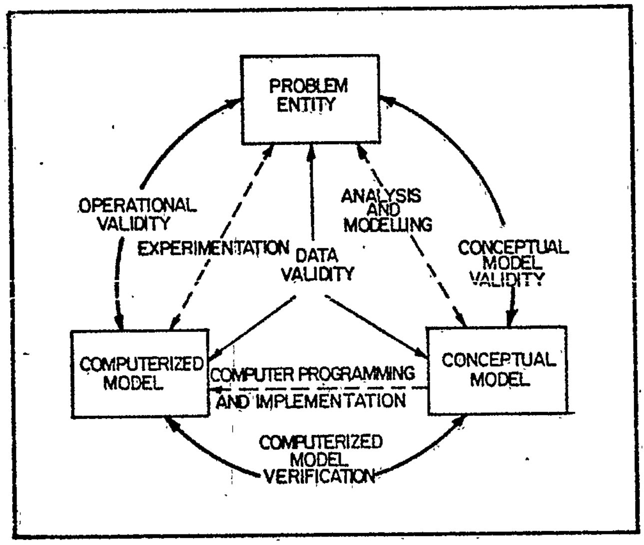

FIGURE 1: Simplified Version of the Modelling Process

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 (1983)]. A variety of techniques are used, which are described below, in the validation process. Unfortunately, no algorithm or procedure exists to select which techniques to use. Some of their attributes are discussed in Sargent (1983).

We are going to be concerned here with only conceptual model validity and operational validity. See Law and Kelton (1982), Oren (1981), Sargent (1982) and Shannon (1981) for discussion of computerized model verification and Sargent (1982) for discussion of data validity.

2. VALIDATION TECHNIQUES

This section describes various validation techniques (and tests) used in model validation. Most of the techniques described here are found in the literature (see Balci and Sargent (1980) 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 mathematical or statistical test or procedure, e.g., hypothesis tests, goodness-of-fit tests, and confidence intervals. A combination of these techniques is usually used in the validation process. It should be noted that these techniques are used for validating submodels as well as the overall model and that they are also frequently used in computerized model verification.

Comparison to Other Models: Various results, e.g., outputs, of the model being validated are compared to results of other (valid) models. For example, simple cases of a simulation model may be compared to known results of 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 increase as the model's run time increases 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. Examples of events might be the number of deaths in a given fire department simulation or number of fires having a given amount of fire damage.

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 - 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 or its behavior is reasonable. For example, it is applied to the model flowchart to determine if the logic is correct. Also, face validity is used to determine if a model's input-output relationships and its internal behavior (e.g., queue lengths) are reasonable.

Historical Data Validation: If historical data exist (or if data is collected on a system prior to building the model), use part of the data to build the model and the remaining data to determine (test) if the model behaves the same as the system does.

Historical Methods: Three historical methods of validation are *Rationalism*, *Empiricism*, and *Positive Economics* [Naylor and Finger (1967)]. 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 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, system, or issue 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: 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, (ii) the graphical display of the flow of traffic moving through an intersection in a traffic simulator, and (iii) 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, or output, should be made sufficiently accurate prior to using the model. (This may require iterations in model development.)

Predictive Validation: The model is used to predict (forecast) the system behavior and 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 statisical procedure for Turing Tests).

3. 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 linearilty, independence, stationary, and Poisson arrivals. Examples of applicable statistical methods are fitting distributions to data, estimating parameter values, e.g., mean and variance, calculation of 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 validation 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.

4. OPERATIONAL VALIDITY

Operational validity is primarily concerned with determining that the model's output behavior has the accuracy required for the model's purpose

	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

over the domain of the model's 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 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 non-observable, where observable means that data can be collected on the operational behavior of the system. Figure 2 gives one classification of the validation approaches for operational validity.

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. This comparison can be made either subjectively or objectively. This author believes that in many cases the subjective approach in comparing the model's and system's input-output behavior is appropriate and satisfactory. However, the use of this approach requires the model developer (validator) to be resourceful in choosing the appropriate comparisons to be made graphically. Comparison of mean values by themselves are usually insufficient. Measures such as variances and maximums are usually required. Figure 3 gives an example of what is required. In addition to these graphs being used by the model developer to determine if the model's output behavior has the required accuracy, they can also be used in Turing Tests. A real world application of

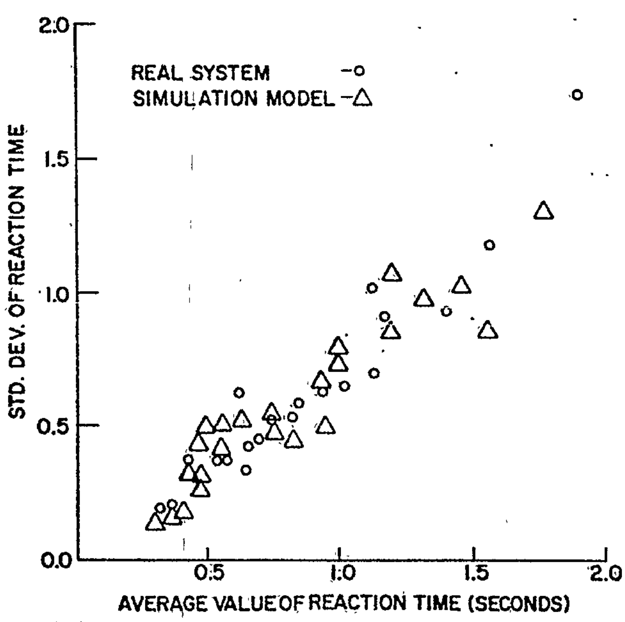

Figure 3: The Standard Deviation of Reaction Time Versus the Average Value.

using the subjective approach and these types of graphs to validate a simulation model can be found in Anderson and Sargent (1974).

There are two basic objective approaches given in the literature for comparison of model and system output behavior data: Hypothesis Testing and Confidence Intervals. These two approaches are statistically related. For a discussion on using hypothesis tests, including the necessity of considering Type II error, see Balci and Sargent (1981a, 1982a, 1982b, 1983). For discussions on using the confidence interval approach, see Balci and Sargent (1981b) and Law and Kelton (1982).

5. 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, if possible, comparisons be made between the model and system output behavior data for at least two sets of experimental conditions.
(6) Validation discussed in the model documentation.

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

There is a considerable literature on validation. A bibliography prepared by Balci and Sargent in 1980 listed 125 articles on validation and numerous articles have been written since then. Articles given in the references can be used as a starting point for furthering your knowledge on validation. This paper gives only a general introduction to the topic and does not discuss the statistical techniques and procedures commonly used.

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. 195-192.

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 (1983), "Validation of Multivariate Response Trace-Driven Simulation Models", Performance 83, edited by Agrawada and Tripathi, North Holland, pp. 309-323.

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.

Law, A.M. and W.D. Kelton (1982), Simulation Modeling and 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 Modelling and Simulation, edited by F. E. Cellier, Academic Press (London), pp. 159-169.

Sargent, R.G. (1983), "Simulation Model Validation", forthcoming in 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

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