Proceedings of the 1989 Winter Simulation Conference
E.A. MacNair, K.J. Musselman, P. Heidelberger (eds.)

VERIFICATION AND VALIDATION: A CONSULTANT'S PERSPECTIVE

John S. Carson, II
Carson/Banks & Associates
Atlanta, GA 30342, U.S.A.

ABSTRACT

The purpose of this paper is to look at the questions and issues regarding verification and validation of simulation models from the industrial and/or consultant's perspective. We will discuss some of the practical as well as conceptual issues involved. In addition, this paper and presentation will briefly describe several simulation projects/case studies and the verification and validation techniques used in them.

1. DEFINITIONS

To clarify the issues involved, a distinction is made between verification, validation, and credibility:

1. Verification is the process whereby a modeler asks the questions: Is the model performing as it should? Does the computer code correctly represent the model assumptions and system data?

   Verification is the process of comparing the conceptual model with the computer code that implements that conception.

2. Validation is the process whereby the modeler and end user ask the questions: How accurately does the model represent reality? Can the model be used in place of the real system for the purpose of making decisions concerning the real system?

   A valid model is one that is sufficiently accurate for the purpose at hand and which can be used as a substitute for the real system for the purpose of addressing the "what if ..." questions of interest.

3. A credible model is one that is accepted by the client as being sufficiently accurate to be used as an aid in making decisions.

Validation and verification should not be regarded as steps that can be tacked onto the end of a project. Although conceptually we may treat verification and validation as separate issues, in practice they are, and must be, completely integrated into the model building process.

2. THE IMPORTANCE OF CLIENT INVOLVEMENT IN MODEL VALIDATION

Verification and validation cannot be carried out by the modeler alone. By involving the client in a team effort, it is much more likely that the model will be accurate, and will be accepted and used in the decision-making process.

For the ultimate credibility of the model, it is very important from the outset that the modeler not raise expectations too highly. Simulation alone will not solve problems, but it provides a valuable tool for identifying problem areas, testing alternative solutions, and gaining insight into complex system operations.

The modeler should make it clear that a team effort involving the client's staff, especially those engineers and managers knowledgeable about system operations, is crucial for testing the validity of the model. In addition, the client or model user usually will provide most or all of the required input data, the accuracy of which is also crucial for model validity.

Management should be aware of the level of commitment needed to support the simulation effort. Besides the modeler's time and effort, a successful model may require the knowledge of numerous plant and manufacturing engineers, supervisors and foremen and workers, and industrial and systems engineers. It is they who will provide the modeler with the necessary assumptions regarding real system operation; and it is they who will ultimately pass judgment on the validity of the model, and pass that judgment up the line to management.

Thus, communication with the client plays a large role both in building a valid model and in establishing its credibility.

3. CONSTRAINTS

Time and cost always play a constraining role in the amount of effort available for model verification and validation. Two of the most frequently encountered constraints are (1) the (un)willingness of the client to expend the time and/or expense for data collection; and (2) a time deadline imposed by upper management, usually because of some urgency such as simulation being a part of a much larger design project, or the need to rapidly solve a manufacuring problem. It is my belief that the practical constraints of time and cost, data availability, and communications with the client, are the real problems and difficulties with verifying and validating models. Most of the conceptual/philosophical issues appear to be resolved, at least so far as they need to be to provide practical guidance to the modeler.

4. THE FIRST STEPS IN MODEL BUILDING

Validation and verification begin at the onset of the model building process and continue throughout the project. A model is usually validated with respect to a specified objective. Indeed, a model may be valid for one purpose but not valid (or at least not tested) for another. In addition, the degree of validity built into a model depends to a large extent on the modeler's ability to ask the right questions and to find the people among the client's staff who know the answers.

When beginning a new simulation project, the first two general questions that the modeler needs to address are:

What is the objective of this study?

For a model based on an existing system, how does the system work? For a model based on a new system design, how is it supposed to work?

Together with management, the modeler should establish realistic objectives for the simulation project. Typically, management and engineers have many questions that they would like to see addressed. These may regard plant capacity, possible bottlenecks for a given design and a given load, the effect of unscheduled downtime, the feasibility of a particular scheduling policy, the adequacy of the planned material handling system, and so on. With the objectives in mind, the modeler and client should identify the most appropriate system performance measures for determining how closely a design alternative meets those objectives. Examples of system performance measures are throughput, mean work in process, and machine utilizations. If at all possible, the model should be validated with respect to the identified performance measures.

The next part of the modeler's job is to learn as much as possible about how the real system works, or if the system is a newly designed one, how the new system is supposed to work. This effort requires a continuing dialogue between the modeler and the client's engineer or system design team and other personnel knowledgeable about real system operation. The simulation team must define the boundary of the system: What is inside the system and what is outside? What external factors impinge on the system? An ongoing effort involves deciding the appropriate level of detail in the model. This depends on both the purpose of the model and the available knowledge. Broadly speaking, the purpose may be classified into one of three categories which require increasingly detailed models:

1. Determine the design basis of a new system.
2. Evaluate operational aspects of an existing system.
3. Test control strategies.

The design of a new system may necessitate the least detail, due to a lack of knowledge and data. Testing control strategies (such as production schedules, job release rules, AGV control) requires a greater amount of detail. The level of model detail in turn affects the level of objectivity versus subjectivity that is possible in the validation process.

For large-scale real systems, it is seldom the case that one individual understands how the system works in sufficient enough detail to build an accurate simulation model. The modeler must be willing to be a bit of a detective to ferret out the necessary knowledge. It is of invaluable help to the modeler to have an advocate among the client's staff who will dig out the answers to detailed questions concerning system operation.

Asking the right questions of the right people is more an art than a science, and more a function of experience than of education. Example: A simulation team (not from C/B & A) is modeling the sheet metal facility of an aircraft manufacturer. Working with the client's group that deals with building new planes, the team accesses a mainframe database for the planes currently being built, and builds a model database on part attributes, part routing and processing times, etc. It is only toward the end of the project that a review by another group reveals that 50% of sheet metal production is for spare parts for existing planes, and the mainframe database is not representative of overall production in the sheet metal facility. Much of validation requires getting the right information in the first place.

It is good practice for the modeler to provide a list of questions in writing concerning plant operations and the assumptions that will be built into the model. It is also good practice, as does a good detective, to ask these questions of a number of client's personnel. Quite often, the answers to a given question will be different! The purpose is not to trip up the client, but for all involved to build a common understanding concerning the assumptions that will be built into the model. This collection of assumptions is known as the conceptual model. Building good simulation models requires the combined skills and knowledge of the modeler and the client's engineers. Rarely do these skills and knowledge occur in one individual.

5. DATA VALIDITY

In addition to a thorough understanding of how the system works, the modeler will need good input data. In addition, system "output" data is needed for comparison to model output in later stages of the validation process.

By data, we mean processing times, travel and conveyance times, time to failure of various machines, repair times, and any other system parameters or variables of importance. In some sense, the model itself defines the input data that is needed to drive the model, while, on the other hand, the availability of data may place restrictions or limitations on the model.

Good data in the form needed for simulation purposes is often not immediately available. Although most modern manufacturing plants generate masses of data, much of it is either out of date, in the wrong form, or simply the wrong data. Quite often, some recent change to the manufacturing process makes past data of questionable value. In many cases, data is reported as averages; but averages hide important facts regarding the variations in system operation. For example, processing times may be reported as averages over long periods of time, or averages over a wide variety of different products, while the data the simulation needs is individual processing times, not averages. For a specific tool, the simulation may need the time until a tool-related failure occurs, but due to an existing policy, tools are replaced at scheduled intervals and when other failures occur. The only data available is the time between replacements (and no record was kept of the cause of the replacement). To use the time between replacements (under the existing maintenance policy) as a substitute for the time to failure would bias any simulation study designed to test a variety of maintenance policies.

There are a variety of ways to get the data needed:

- Time studies or automated data collection facilities
- Historical records
- Vendor's claims
- The client's best guess
- The modeler's best guess

Of course, data validity may vary widely from source to source.

A time study of the existing system, if designed properly, will usually provide the most accurate data. If not personally conducting the time study, the modeler should clearly explain the types and accuracy of data required. If the project involves a design of a new system, a time study of a similar system may be of value, if the new system will contain some components identical to those in the existing system.

Historical records, such as production and downtime reports and past time studies, should be carefully scrutinized for accuracy and relevance. For example, it should not be assumed that a long machine idle time that began with a downtime can be totally attributed to repair time. Perhaps when the downtime occurred, the production schedule was re-arranged and the down machine was not going to be needed until the following week, so repair was postponed. Due to this lack of completeness of many historical data records, such data should be used with caution, and preferably only after conferring with the people who gathered the data. For similar reasons, data collected by automated means may be equally unreliable or unusable. Was reported machine idle time due to actual repair, waiting for repair, lack of raw materials, an operator on break, or a change in the production schedule? Machine downtime may be accurately recorded, but the modeler needs to know why the machine was down. Without such knowledge, the data may well be useless for simulation purposes.

For new machines or material handling equipment, the vendor may supply claimed processing times, conveyance times, and/or mean times to failures. This information is useful as a starting point, but may be modified by the client's experience with similar equipment.

When time studies are impossible, or in the early stages of a project before a time study can be conducted, subjective estimates provided by knowledgeable engineers can be very useful.

When the modeler is so fortunate, or management is so committed, that a random sample of valid data is available, then the modeler can use all his statistical skills (or consult a statistician) to decide the best way to use this data in the model, whether directly or by fitting a statistical distribution to the data.

The availability of good data is just as important for developing a valid simulation as the correctness of the built-in assumptions. The old cliche, garbage-in-garbage-out, applies to simulation software just as much as to other software.

6. VERIFICATION TECHNIQUES

Verification refers to the process of testing and checking the computer code to assure that it truly represents the assumptions and data accurately. Conceptually, we view verification as the modeler's job, and a necessary first step before validation can begin.

Many of the techniques for verifying a simulation model's computer implementation are identical to those for debugging any piece of software. Other techniques are specific to simulation. These techniques include:

- Use of structured programming techniques, such as modularity and top-down design (although some existing languages have built-in limitations in this regard).

- Extensive program testing under a wide variety of input parameters (that is, a sensitivity analysis).

- Collection and display of numerous statistics, and a thorough examination of those statistics for reasonableness.

- Use of a trace.

- Use of an interactive debugger.

- Use of an animated display of system operations.

Use of statistical output for verification

Languages that automatically collect and print a large volume of output data make the verification task easier. Extensive output statistics can often be of help in identifying errors in the model. For example, a utilization of zero may indicate that no loads are getting to a specific machine. A utilization of 100% may indicate a misspecified capacity, a misspecified service time, or an error in routing of loads.

Use of a trace for verification

A trace consists of detailed output representing the step-by-step progress of the simulation model over (simulated) time. A trace can be of especial value for detecting the cause of subtle errors or verifying that the model handles exceptional situations correctly. The client's engineer may be much more interested in how the model handles exceptional situations such as unscheduled downtimes, full or empty buffers, and running out of raw material, than in the long-run statistical behavior of the model. A trace can provide the modeler with a valuable tool for verifying that the model handles these exceptions according to agreed upon assumptions.

Use of an interactive debugger

An interactive debugger allows the user to step through the simulation, speed-up or slow down system operation, view system state or statistics at any point in time or upon the occurrence of a specified event, and view how the model handles the occurrence of exceptional situations.

Use of animation in verification

Graphical animation of simulation output is one of the newest features of many simulation languages. Animation essentially provides a visual equivalent of the trace. With animation, the modeler can visually verify whether the model handles normal and exceptional circumstances correctly.

7. VALIDATION TECHNIQUES

We recommend a three-step approach to validation:

1. Develop a model with high face validity.
2. Validate model assumptions.
3. Validate model output.

A model with high face validity is one which is reasonable to the end users of the model. Using all the techniques previously discussed will help to assure high face validity and valid model assumptions. Another valuable technique is the "structured walk-through" of the model's flowchart, covering all model assumptions, and a graphical animation of system operations before an audience of all knowledgeable personnel. Animation removes many of the "black box" aspects of a simulation model and allows users to "see" model assumptions in action rather than depending on the modeler's assurance and long-run statistical output to verify model correctness.

Animation of system operation allows the validation of models involving entities moving through space to a degree that previously was virtually impossible to attain. For example, animation allows a visual check that AGVs on crossing paths do not collide, that waiting AGVs queue correctly behind each other and not on "top" of each other, and that loads enter a conveyor correctly spaced into an empty window and not on top of one another. While not impossible, many of these spatial relationships are difficult to verify without either an interactive debugger or animator. Example: We recently revived an old GPSS model of a pallet conveyor system. Using the GPSS/H interactive debugger, we identified an error in conveyor logic that would allow two pallets to be in adjacent load positions, which violates the "window" assumptions. The error would occur fairly rarely and was never noticed by looking at the "Block Counts" printed at the end of each 8-hour shift within a run.

Although credibility can be increased by animation, a possible danger is that hasty judgments about real system operation will be made based on too short a runlength and before validation is complete. This "danger" is probably overplayed, as most manufacturing engineers will use the animation to gain insight into system operations, but will want to see the output data of numerous long runs before making any final judgments.

The validation of model output by comparing it to similar data from the real system is the most objective and scientific method of validation. This technique requires that a real system exists that corresponds to one setting of the model parameters. Thus, the technique cannot be applied to a model of a completely new system until after the system is constructed and in operation.

The main idea is to match as closely as possible the "inputs" to the real system and the "inputs" to the model. Then for the model to be considered valid, the system "outputs" should approximate closely the model "outputs". By "inputs" we mean the particular settings of model parameters and variables which correspond to real system behavior in some past period of time. Examples of inputs include:

- Number and capacity of machines at each workstation
- Processing times for each machine
- Time to random failure, and repair time, for each piece of equipment
- Times and durations of preventive maintenance
- Percent of pieces requiring rework

There are two distinct ways to match those "inputs" that are random in nature:

1. Use statistical distributions based on past data.
2 Use actual historical values.

The first technique will provide a test of the conceptual model plus the data assumptions, while the second provides a test of the conceptual model alone.

8. A CASE STUDY

The case study is based on a simulation model of the cigarette fabrication process at one of the larger cigarette manufacturers in the U.S. The cigarette fabrication process consists of a maker/filter tip attachment (FTA) machine combination, reservoir, packer, pack downdrop and pack conveyor, and wrapper/cartoner. What follows is a simplified description of the system, the model, and the validation effort. (For further reading, see the article by Carson, Wysowski, Carroll, and Wilson [1981]).

The maker/FTA fabricates a continuous rod of tobacco, wraps paper around the rod, and cuts it into the desired length. The FTA applies a filter to the tobacco section. The cigarette is conveyed through a reservoir to a packer, which places foil and a label around groups of 20 cigarettes. Packs are then conveyed to the wrapper/cartoner.

Modern makers operate at speeds from about 4000 to over 7000 cigarettes per minute. Makers and packers are subject to fairly frequent product-induced failures (for example, the cigarette paper tears, or glue in the packer gets in the wrong place). These occur from one to perhaps 15 times per hour, with downtimes lasting from a few seconds to a few minutes. The reservoir acts as a buffer between the maker and the packer. Most modern makers and packers have a multi-speed capability and adjust speed according to the current contents of the reservoir. As stopping and starting the packer tends to increase quality problems and reduce yield, continuous operation of the packer is most desirable. Of course, to maximize production, the maker must be kept running as much as possible.

The objective of the modeling project was to determine the "optimum" capacity of the reservoir, given a particular configuration of machines, their processing rates and downtime characteristics. The performance measure of greatest interest was production yield over an appropriate period of time.

Table 2: A List of Verification Techniques

Documentation(*) - This is the recorded information concerning a model. The documentation related to verification is that which details the model design and the code which implements that design.

Operational Graphics(*) - The model's operational behavior is displayed graphically as the model moves through time. This helps the modeler to visualize the progress of the simulation during execution and aids in the detection of errors.

Structured Programming - This refers to the use of specific techniques that enable one to more easily understand programs and troubleshoot existing code. Techniques include program modularity and top-down design. Program modularity is defined as the decomposition of the model into sub-models which have well defined functions and interfaces. Top-down design refers to the notion of creating a detailed plan of the model before writing the code, in order to avoid revising the original structure of the model.

Structured Walkthrough - This refers to assembling a group of programmers to participate in a line-by-line evaluation of the model or a module of the model. A structured walkthrough is utilized to detect errors early in the model development cycle, to provide an informal review, and to encourage technical exchange in a constructive, non-fault finding atmosphere.

Traces(*) - The behavior of different types of specific entities are followed through the model's execution to determine if the coding is correct and if the necessary accuracy is obtained. The trace is accomplished by printing out the state of the simulated system just after each event occurs.

* This technique is also a validation technique.

Table 3: A List of Validation Techniques

Documentation(*) - This is the recorded information concerning a model. Documentation can be sub-divided into two parts; descriptive and technical. Descriptive documentation is general information about the model's capabilities, limitations, and assumptions. Technical documentation is that detailed information that describes how the simulation model works and the exact mechanics of the model.

Event Validity - The events of the simulation model are compared to those of the real system to determine if they are the same.

Extreme Condition Test - The model structure and output should be plausible for any extreme and unlikely combination of levels of factors in the system.

Face Validity - This technique refers to asking people knowledgeable about the system whether the model and/or its behavior is reasonable. Face validity can be used in all facets of validating the system model.

Table 3: A List of Validation Techniques (Cont'd)

Historical Data - In using this technique, part of the historical data is used to build the model and the remaining historical data is used to determine if the model behaves as the system does. This technique can only be used if historical data exist.

Input-Output Transformations - This refers to the model's ability to predict the future behavior of the real system when the model input data match the real inputs and when a policy implemented in the model is implemented at some point in the system. The structure of the model should be accurate enough for the model to make good predictions for the range of data sets that are of interest.

Operational Graphics(*) - The model's operational behavior is displayed graphically as the model moves through time. This helps to visualize the progress of the simulation during execution and aids in the validation process.

Sensitivity Analysis - This validation technique consists of changing values of the input parameters of a model to determine the effect upon its output. The same relationships should occur in the model as in the real system.

Statistical Tests - This refers to statistical procedures used to determine input data validity as well as for comparing real-world observations and simulation output data.

Traces(*) - The behavior of different types of specific entities in the model are traced through the model to determine if the model's logic is correct and if the necessary accuracy is obtained. The trace is accomplished by printing out the state of the simulated system just after each event occurs.

Turing Tests - People who are knowledgeable about the operations of a system are asked if they can discriminate between system and model outputs.

* This technique is also a verification technique.

REFERENCES

Banks, J. and J. S. Carson, Chapter 10 of Discrete-Event System Simulation, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

Carson, J. S., N. Wilson, D. Carroll, and C.H. Wysowski, "A Discrete Simulation of a Cigarette Fabrication Process", Proceedings of the Twelfth Modeling and Simulation Conference, University of Pittsburgh, April 30-May 1, 1981.

Feltner, Charles E. and Steven A. Weiner, "Models, Myths and Mysteries in Manufacturing", Industrial Engineering, Vol. 17, No. 7, July, 1985.

Law, A. M. and W. D. Kelton, Chapter 10 of Simulation Modeling and Analysis, McGraw-Hill, New York, 1982.

Measures used in the validation effort included production yield, number of times and duration of reservoir filling, and number of times and duration of reservoir emptying.

Discussions with plant engineers, maintenance personnel, mechanics and operators led to development of a process flow diagram and a list of detailed assumptions on process operation. Meetings and actual observation of the equipment were used to resolve differences of opinion on process operation.

As this was the first simulation model developed for use by the company, many of the engineers and management were skeptical of the model's accuracy and its ability to duplicate the same types of downtimes and interaction problems that occurred in the real system. Therefore, the model developers quickly realized that considerable accuracy would have to be demonstrated before acceptance and use of the model would result.

Many of the techniques for model verification and validation discussed in this article were used on the cigarette fabrication model. Here we will restrict the discussion to the input-output validation effort.

The first part of the input-output validation effort consisted of generating three sets of sample input-output data from the simulation model as well as obtaining three sets of the same data from observations of the real system. This data was placed onto six sheets of paper in identical formats. Individuals familiar with the process were then asked if they could distinguish between the real data and the "fake" data. This procedure is called a Turing test. It proved to be extremely successful; no one could distinguish the real from the simulated data. The Turing test was especially useful in convincing those skeptical of simulation modeling.

The second part of the input-output validation effort involved conducting a detailed time study over a four hour period of an existing fabrication group. Individual runtimes-until-failure and downtime durations were recorded for each machine and piece of equipment. These historical runtimes and downtimes were used to drive the simulation model (in place of statistical distributions which had been based on a longer time study of 24 hours). The main criteria for comparison were production yield and interaction downtime due to a full or empty reservoir. It is important to realize that production yield and interaction downtimes are not inputs to the model, but rather are model predictions. As shown in Table 1, this approach resulted in model predictions that differed by less than 1.1% from the actual observations taken during the four hour time study. These results were instrumental in convincing management that an accurate model had, in fact, been developed.

With periodic updates of the most important data, this model has been used to size the reservoir (and for other purposes) a number of times whenever makers and packers with different processing speeds and downtime characteristics were being considered for purchase. The model-generated data is combined with accounting and cost data to aid in choosing the most economic combination of equipment. Time studies on fabrication groups with the "optimum" reservoir capacity have appeared to confirm model predictions, further increasing model credibility.

Table 1: Validation Results for the Cigarette Fabrication Model

	Actual System (*)	GPSS Model (*)	% Diff. (**)
Maker production (cig)	825,300	834,540	1.1%
Packer production (packs)	41,132	41,089	0.2%
Full reservoir:			
Total duration (min)	20.34	20.12	1.1%
No. of occurrences	4	4	0
Packer random downtime:			
Total duration (min)	42.25	42.25	0
No. of occurrences	15	15	0

(*) Figures are coded for reasons of confidentiality.

(**) Percent difference is the actual attained.

9. CONCLUSIONS

At best, an input-output validation effort can provide an indication that the model worked correctly only for a particular setting of the parameters and for some particular past historical period. Typically, it is desired to use models to experiment with system configurations beyond the current experience of the client's engineers. Thus, the importance of judgment and experience cannot be minimized.

No single validation technique is adequate alone to prove model validity, but a variety of techniques should be used. Tables 2 and 3 provide lists of the most important verification and validation techniques, respectively, that were discussed in this article, plus a few additional ones.

If a model is used for any period of time, validation is never really finished. Data needs to be updated and the model periodically tested. It is a real danger that after the model is "validated," the end user will suspend judgment and accept anything the model says, even though the system data and the system itself have changed over time. When a model is used again a long time after its initial development, it should be remembered that the system has probably changed and the data is probably out-of-date or on the verge of being out-of-date. A periodic re-validation may provide cheap insurance against erroneous conclusions from an invalid model.

The manager of a simulation project should never forget that a simulation project requires a team effort. The modeler(s) bring modeling, computer programming and statistical data analysis skills to the effort. The client's engineers, managers, and other personnel bring knowledge of how the system really works and a great interest in solving real problems. Bringing these two groups together and committing the proper resources will maximize the chances of building an accurate and valid model.

AUTHOR'S BIOGRAPHY

JOHN S. CARSON, II is a partner in Carson/Banks & Associates, a simulation consulting firm. He has been a simulation consultant to a number of manufacturing and service industries since 1978. He has served on the faculties of the Georgia Institute of Technology and the University of Florida. He is co-author, with Jerry Banks, of the text Discrete-Event Systems Simulation. He is a member of IIE, ORSA and TIMS. Dr. Carson earned his BA in mathematics from the University of Virginia, his MS in mathematics and his Ph.D (1978) in engineering from the University of Wisconsin at Madison.

John S. Carson, II
Carson/Banks & Associates
4279 Roswell Road
Suite 604-173
Atlanta, Georgia 30342, U.S.A.
(404) 971-1256