Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

GUIDELINES FOR SUCCESSFUL SIMULATION STUDIES†

Osman Balci

Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061-0106

ABSTRACT

The purpose of this paper is to provide guidelines for conducting a successful simulation study. The guidelines are provided throughout the entire life cycle of a simulation study that is composed of 10 processes, 10 phases, and 13 credibility assessment stages. The practitioners can follow the guidelines presented herein and significantly increase their chance of being successful in conducting a simulation study.

1. INTRODUCTION

In a simulation study, we work with a model of the problem rather than directly working with the problem itself. If the model does not possess a sufficiently accurate representation, we can easily have "junk input" and "junk output." It is no challenge to write a computer program which accepts a set of inputs and produces a set of outputs to do simulation. The challenge is to do it right. Multifaceted and multidisciplinary knowledge and experience are required for a successful simulation study. In order to gain the basic knowledge for using simulation correctly, Shannon [1986] predicts that a practitioner is required to have about 720 hours of formal classroom instruction plus another 1440 hours of outside study (more than 1 man-year of effort).

Assessing the acceptability and credibility of simulation results is *not* something that is done after the simulation results are obtained. Assessment of accuracy (i.e., verification, validation, and testing) must be done right after completing each phase of a simulation study.

The key to success in a simulation study is to follow a comprehensive life cycle in an organized and well-managed manner. The ever-increasing complexity of systems being simulated can only be managed by following a structured approach to conducting the simulation study.

The purpose of this paper is to present a comprehensive life cycle of a simulation study and guide the simulationist in conducting 10 processes, 10 phases, and 13 credibility assessment stages of the life cycle. The paper begins by defining the life cycle in Section 2. This is followed, in Section 3, by guidelines for each of the 10 processes and 10 phases of the life cycle. It goes on, in Section 4, to discuss each of the 13 credibility assessment stages of the life cycle and proposes an overall evaluation scheme for assessing the acceptability and credibility of simulation results. Concluding remarks are given in Section 5.

2. THE LIFE CYCLE OF A SIMULATION STUDY

The life cycle is composed of ten phases as depicted in Figure 1 [Balci 1987; Nance 1981]. The ten phases are shown by oval symbols. The dashed arrows describe the processes which relate the phases to each other. The solid arrows refer to the credibility assessment stages (CASs). The life cycle should not be interpreted as strictly sequential. The sequential representation of the dashed arrows is intended to show the direction of development throughout the life cycle. The life cycle is iterative in nature and reverse transitions are expected.

† This paper is a modified version of: Balci, O. (1989), "How to Assess the Acceptability and Credibility of Simulation Results," In *Proceedings of the 1989 Winter Simulation Conference*, pp. 62-71.

3. PROCESSES OF THE LIFE CYCLE

The ten processes of the life cycle are shown by the dashed arrows in Figure 1. Although each process is executed in the order indicated by the dashed arrows, an error identified may necessitate returning to an earlier process and starting all over again. Some guidelines are provided for each of the ten processes below.

3.1 Problem Formulation

When a problem is recognized, a decision maker (a client or sponsor group) initiates a study by communicating the problem to an analyst (a problem-solver, or a consultant/research group). The communicated problem is rarely clear, specific, or organized. Hence, an essential study to formulate the *actual* problem usually follows. *Problem Formulation* (problem structuring or problem definition) is the process by which the initially communicated problem is translated into a formulated problem sufficiently well defined to enable specific research action [Woolley and Pidd 1981].

Balci and Nance [1985] present a high-level procedure that: (1) guides the analyst during problem formulation, (2) structures the verification of the formulated problem, and (3) seeks to increase the likelihood that the study results are utilized by decision makers. The reader is referred to [Balci and Nance 1985] for details of the procedure.

3.2 Investigation of Solution Techniques

All alternative techniques that can be used in solving the formulated problem should be identified. A technique whose solution is estimated to be too costly or is judged to be not sufficiently beneficial with respect to the study objectives should be disregarded. Among the qualified ones, the technique with the highest expected benefits/cost ratio should be selected.

The statement "when all else fails, use simulation" is misleading if not invalid. The question is not to bring a solution to the problem, but to bring a sufficiently credible one which will be accepted and used by the decision maker(s). A technique other than simulation may provide a less costly solution, but it may not be as useful.

Sometimes, the communicated problem is formulated under the influence of a solution technique in mind. Occasionally, simulation is chosen without considering any other technique just because it is the only one the analyst(s) can handle. Skipping the investigation process may result in unnecessarily expensive solutions, sometimes to the wrong problems.

As a result of the investigation process, we assume that simulation is chosen as the most appropriate solution technique. At this point, the simulation project team should be activated and be made responsible for the verification of the formulated problem and feasibility assessment of simulation before proceeding in the life cycle.

3.3 System Investigation

Characteristics of the system that contains the formulated problem should be investigated for consideration in system definition and modeling. Shannon [1975] indicates six major system characteristics: (1) change, (2) environment, (3) counterintuitive behavior, (4) drift to low performance, (5) interdependency, and (6) organization. Each characteristic should be examined with respect to the study objectives that are identified with the formulation of the problem.

In simulation, we mostly deal with stochastic and dynamic real

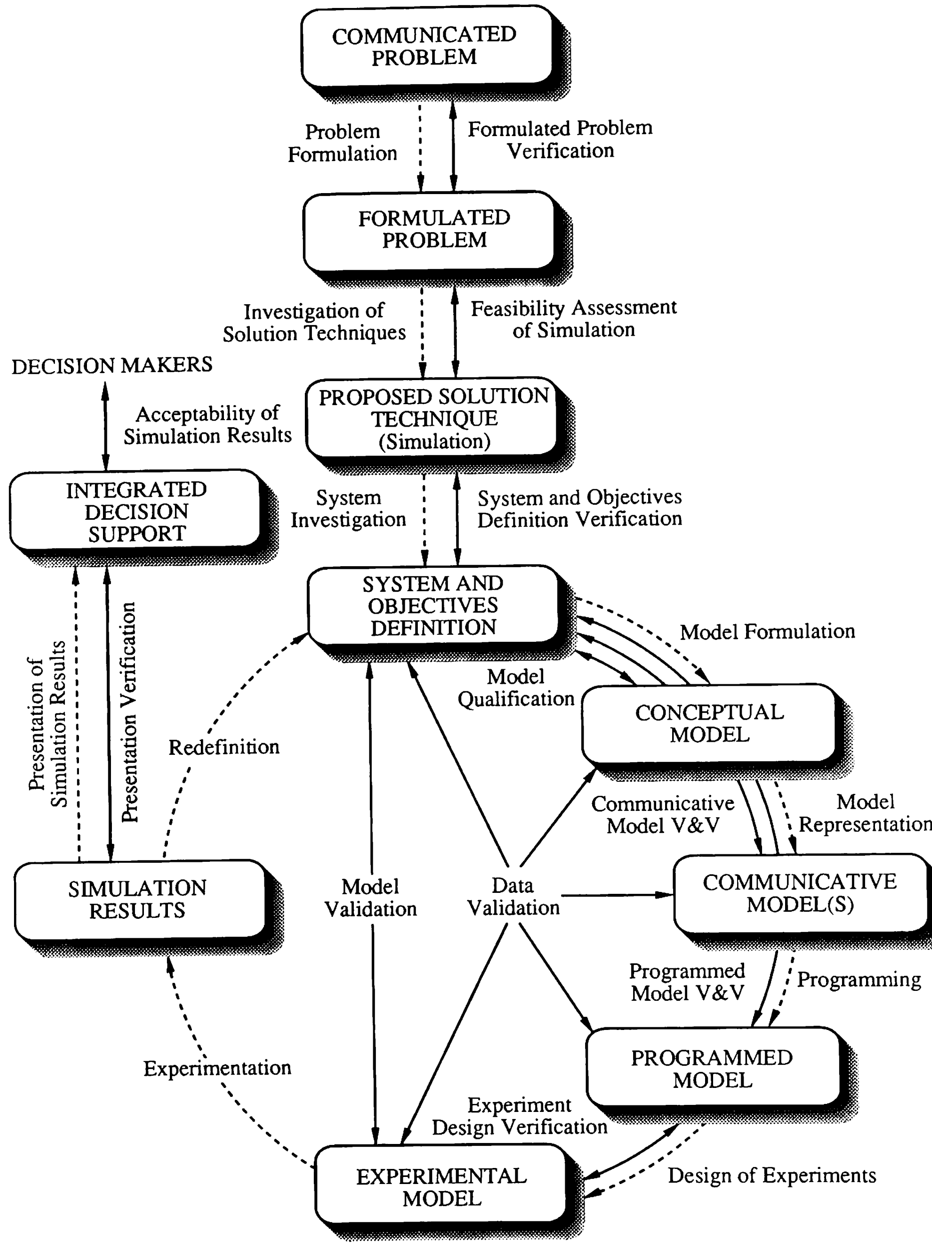

Figure 1. The Life Cycle of a Simulation Study

systems that *change* over a period of time. How often and how much the system will change during the course of a simulation study should be estimated so that the model representation can be updated accordingly. Changes in the system may also change the study objectives.

A system's *environment* consists of all input variables that can significantly affect its state. The input variables are identified by assessing the significance of their influence on the system's state with regard to the study objectives. For example, for a traffic intersection system, the interarrival time of vehicles can be identified as an input variable making up the environment, whereas pedestrian arrivals may be omitted due to their negligible effect on the system's state. Underestimating the influence of an input variable may result in inaccurate environment definition.

Some complex systems may show *counterintuitive behavior* which we should try to identify for consideration in defining the system. However, this is not an easy task, especially for those systems containing many subjective elements (e.g., social systems). Cause and effect are often not closely related in time or space. Symptoms may appear long after the primary causes [Shannon 1975]. To be able to identify counterintuitive behavior, it is essential that the simulation project employs people who have expert knowledge about the system under study.

A system may show a *drift to low performance* due to the deterioration of its components (e.g., machines in a manufacturing system) over a period of time. If this characteristic exists, it should be incorporated within the model representation especially if the model's intended use is forecasting.

Before we start abstracting the real system for the purpose of modeling, we should examine the *interdependency* and *organization* characteristics of the system. In a complex stochastic system, many activities or events take place simultaneously and influence each other. The system complexity can be overcome by way of decomposing the system into subsystems and subsystems into other subsystems. This decomposition can be carried out by examining how system elements or components are organized.

Once the system is decomposed into subsystems the complexity of which is manageable and the system characteristics are documented, model formulation process can be started following the system and objectives definition verification.

3.4 Model Formulation

Model formulation is the process by which a conceptual model is envisioned to represent the system under study. The *Conceptual Model* is the model which is formulated in the mind of the modeler [Nance 1981]. Model formulation and model representation constitute the process of model design.

Input data analysis and modeling [Law and Kelton 1982] is a subprocess of Model Formulation and is conducted with respect to the way the model is driven. Simulation models are classified as self-driven or trace-driven. A *self-driven* (distribution-driven or probabilistic) simulation model is the one which is driven by input values obtained via sampling from probability distributions using random numbers. A *trace-driven* (or retrospective) simulation model, on the other hand, is driven by input sequences derived from trace data obtained through measurement of the real system.

Under some study objectives (e.g., evaluation, comparison, determination of functional relations) and for model validation, input data model(s) are built to represent the system's input process. In a self-driven simulation (e.g., of a computer system), we collect data on an input random variable (e.g., interarrival time of jobs), identify the distribution, estimate its parameters, and conclude upon a probability distribution as the input data model to sample from in driving the simulation model [Law and Kelton 1982]. In a trace-driven simulation, we trace the system (e.g., using hardware and software monitors) and utilize the refined trace data as the input data model to use in driving the simulation model.

3.5 Model Representation

This is the process of translating the conceptual model into a communicative model. A *Communicative Model* is "a model representation which can be communicated to other humans, can be judged or compared against the system and the study objectives by more than one human" [Nance 1981]. A communicative model (i.e., a simulation model design) may be represented in any of the following forms: (1) structured, computer-assisted graphs, (2) flowcharts, (3) structured English and pseudocode, (4) entity-cycle (or activity-cycle) diagrams, (5) condition specification [Overstreet and Nance 1985], and (6) more than a dozen diagramming techniques described in [Martin and McClure 1985].

Several communicative models may be developed; one in the form of Structured English intended for nontechnical people, another in the form of a micro flowchart intended for a programmer. Different representation forms may also be integrated in a stratified manner.

The representation forms should be selected based upon: (1) their applicability for describing the system under study, (2) the technical background of the people to whom the model is to be communicated, (3) how much they lend themselves to formal analysis and verification, (4) their support for model documentation, (5) their maintainability, and (6) their automated translatability into a programmed model.

3.6 Programming

Translation of the communicative model into a programmed model constitutes the process of programming. A *Programmed Model* is an executable simulation model representation in a simulation programming language (e.g., GPSS, SIMAN, SIMSCRIPT, SIMULA, SLAM, etc.) or in a high-level programming language (e.g., C, Fortran, Pascal, etc.) that do not incorporate an experiment design. There is an abundance of literature on simulation programming languages. Balci [1988] describes how to conduct the programming process in high-level languages.

3.7 Design of Experiments

This is the process of formulating a plan to gather the desired information at minimal cost and to enable the analyst to draw valid inferences [Shannon 1975]. An *Experimental Model* is the programmed model incorporating an executable description of operations presented in such a plan.

A variety of techniques are available for the design of experiments. *Response-surface methodologies* can be used to find the optimal combination of parameter values which maximize or minimize the value of a response variable [Law and Kelton 1982]. *Factorial designs* can be employed to determine the effect of various input variables on a response variable [Fishman 1978]. *Variance reduction techniques* can be implemented to obtain greater statistical accuracy for the same amount of simulation [Law and Kelton 1982]. *Ranking and selection techniques* can be utilized for comparing alternative systems [Law and Kelton 1982; Banks and Carson 1984]. Several methods (e.g., replication, batch means, regenerative) can be used for statistical analysis of simulation output data.

3.8 Experimentation

This is the process of experimenting with the simulation model for a specific purpose. Some purposes of experimentation are [Shannon 1975]: (1) comparison of different operating policies, (2) evaluation of system behavior, (3) sensitivity analysis, (4) forecasting, (5) optimization, and (6) determination of functional relations. The process of experimentation produces the Simulation Results.

3.9 Redefinition

This is the process of: (1) updating the experimental model so that it represents the current form of the system, (2) altering it for obtaining another set of results, (3) changing it for the purpose of maintenance, (4) modifying it for other use(s), or (5) redefining a new system to model for studying an alternative solution to the problem.

3.10 Presentation of Simulation Results

In this process, simulation results are interpreted and presented to the decision makers for their acceptance and implementation. Since all simulation models are descriptive, concluding upon a solution to the problem requires rigorous analysis and interpretation of the results.

The presentation should be made with respect to the intended use of the model. If the model is used in a "what if" environment, the results should be integrated to support the decision maker in the decision-making process. Complex simulation results may also necessitate such an integration. The report documenting the study and its results together with its presentation also constitutes a form of supporting the decision maker.

4. CREDIBILITY ASSESSMENT STAGES OF THE LIFE CYCLE

In Elmaghraby's words, "It is well to remember the dictum that nobody solves *the* problem. Rather, everybody solves the model that he [or she] has constructed of the problem" [Elmaghraby 1968, p. 305]. Thus it is crucial that we assess the credibility of each process as we progress in the life cycle.

Simulation Model Verification is substantiating that the simulation model has been transformed from one form into another as intended with sufficient accuracy. *Simulation Model Validation* is substantiating that the simulation model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives. Verification deals with building the simulation model right. Validation deals with building the right simulation model. *Testing* is demonstrating that bugs exist or revealing the existence of errors (bugs).

Since a model is an abstraction of the reality, we cannot talk about its *absolute* accuracy. Credibility, quality, validity, and verity are measures that are assessed with respect to the study objectives for which the model is intended. In some cases, a 60% level of confidence in the credibility of simulation results may very well serve for the purpose; in another, 90% may be required.

Three types of errors may be committed in conducting a simulation study. *Type I Error* is committed when the study results are rejected when in fact they are sufficiently credible. *Type II Error* is committed when the study results are accepted when in fact they are *not* sufficiently credible. The definitions of type I and type II errors can be extended to apply for every CAS. In the case of model validation, we called the probability of committing type I error as *model builder's risk* and the probability of committing type II error as *model user's risk* [Balci and Sargent 1981]. *Type III Error* is committed when the formulated problem does not completely contain the *actual* problem [Balci and Nance 1985]. Committing type III error corresponds to solving the wrong problem.

A simulation project team should possess at least four areas of knowledge and experience to be successful: (1) project leadership, (2) modeling, (3) programming, and (4) knowledge of the system under study [Annino and Russell 1979]. Lacking the necessary level of expertise in any area may result in a failure of the project or an error of type I, II or III.

Subjectivity is and will always be part of the credibility assessment for a reasonably complex simulation study. The reason for subjectivity is two-fold: modeling is an art and credibility assessment is situation dependent.

A subjective, yet quite effective method for evaluating the acceptability of simulation results is *peer assessment*, the assessment of the acceptability by a panel of expert peers. This panel should be composed of: (1) people who have expert knowledge of the system under study, (2) expert modelers, (3) expert simulation analysts, and (4) people with extensive experience with simulation projects.

The panel examines the overall study based upon the project team's presentation and detailed study of documentation. Working together and sharing their knowledge among each other, panel members measure the indicators shown by the leaves of the tree in Figure 2 which assists in explaining the hierarchy of CASs. A branch of the tree represents a CAS except the branch of "other indicators" of experimental model quality.

An *indicator* is an indirect measure of a concept. It can be decomposed into other indicators. The ones at the base level (i.e., the ones that are not decomposed further) must be directly measurable.

The indicators (leaves) are presented in the following subsections. The kth indicator out of N_{ij} ones corresponding to the jth branch (from the top of Figure 2) at level i is measured with a score, S_{ijk}, out of 100 and is weighted with W_{ijk}, a fractional value between 0 and 1, according to its importance. The following constraint must be satisfied:

$$\sum_{k=1}^{N_{ij}} W_{ijk} = 1 \qquad (i, j) = \{(1,2), (2,1), (2,2), (2,3), (3,1), (3,2), \ldots, (3,7)\}$$

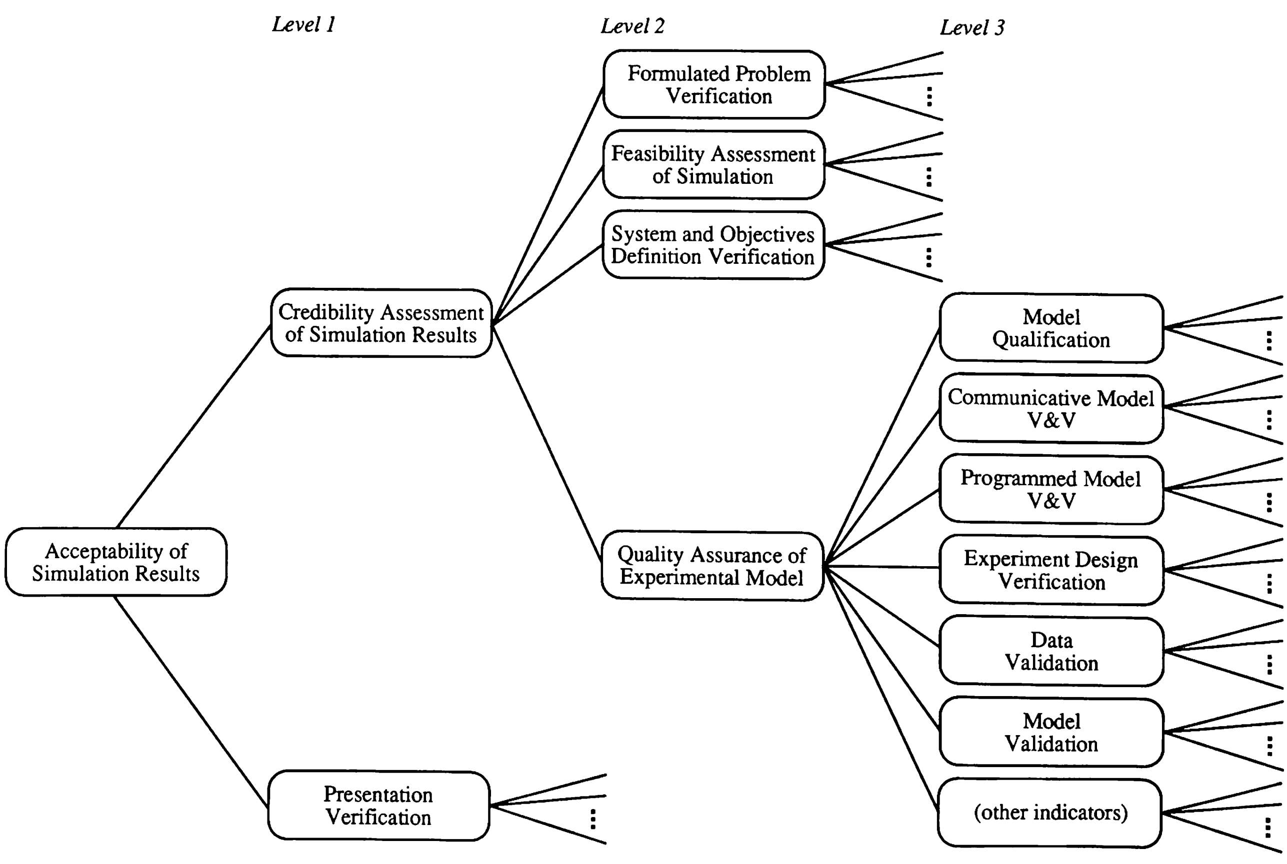

Figure 2. A Hierarchy of Credibility Assessment Stages for Evaluating the Acceptability of Simulation Results

Thus, a score for the jth branch at level i, S_{ij}, is calculated on a scale from 0 to 100 where 0 represents "not credible at all" and 100 means "sufficiently credible."

$$S_{ij} = \sum_{k=1}^{N_{ij}} W_{ijk} S_{ijk} \qquad (i, j) = \{(1,2), (2,1), (2,2), (2,3) \\ (3,1), (3,2), ..., (3,7)\}$$

In addition to weighting the indicators, panel members can also weight the branches based on experience and training. For example, model validation branch should be given higher weight than the other branches at level 3, if it is possible to validate the model objectively using the real system data under all experimental conditions of interest. On the other hand, if the model represents a nonexistent system or a future-oriented situation in which the past is not a good predictor of the future, higher weight should be given to other branches at level 3.

Assume that W_{ij} denotes the weight for the jth branch at level i. W_{ij} is a fractional value between 0 and 1 where 0 represents "not critical at all" and 1 indicates "extremely critical."

W_{ij}'s are specified with the following constraints:

$$\sum_{j=1}^{2i} W_{ij} = 1 \qquad i = 1, 2 \text{ and} \sum_{j=1}^{7} W_{3j} = 1$$

Thus, a credibility score for the quality assurance branch is calculated as

$$S_{24} = \sum_{j=1}^{7} W_{3j} S_{3j}$$

Similarly, S_{11} is computed and an overall score, S ($= W_{11}S_{11} + W_{12}S_{12}$), is obtained as a value on a scale from 0 to 100.

The higher the overall score the more confidence we gain for the acceptability of simulation results. However, even a perfect score would not guarantee that the results will be accepted and used by the decision makers; because, acceptability is an attribute of the decision maker not an attribute of the simulation study. Perfect results may be rejected due to the lack of credibility of the institution performing the study or due to a political reason. Nevertheless, the objective of the simulation project management should be to increase the confidence as much as possible. A higher overall score may not guarantee the acceptance of results but a lower overall score can easily result in their rejection or an error of type II.

4.1 Formulated Problem Verification

"Substantiation that the formulated problem contains the *actual* problem in its entirety and is sufficiently well structured to permit the derivation of a sufficiently credible solution" is called *formulated problem verification* [Balci and Nance 1985]. For this substantiation, a Questionnaire is developed in [Balci and Nance 1985] with 38 indicators. People who are intimately knowledgeable of the problem(s) based on experience and training verify the formulated problem by measuring the indicators. The reader is referred to [Balci and Nance 1985] for the details of the verification.

4.2 Feasibility Assessment of Simulation

Are the benefits and cost of simulation solution estimated correctly? Do the potential benefits of simulation solution justify the estimated cost of obtaining it? Is it possible to solve the problem using simulation within the time limit specified? Can all of the resources required by the simulation project be secured? Can all of the specific requirements (e.g., access to pertinent classified information) of the simulation project be satisfied? These questions are the indicators of the feasibility of simulation.

4.3 System and Objectives Definition Verification

We should justify that the system characteristics are identified and the study objectives are explicitly defined with sufficient accuracy. An error made here may not be caught until very late in the life cycle resulting in a high cost of correction or an error of type II or III.

Since systems and objectives may change over a period of time, will we have the same system and objectives definition at the conclusion of the simulation study (which may last from 6 months to several years)? Is the system's environment (boundary) identified correctly? What counterintuitive behavior may be caused within the system and its environment? Will the system significantly drift to low performance requiring a periodic update of the system definition? Are the interdependency and organization of the system characterized accurately?

4.4 Model Qualification

A model should be conceptualized under the guidance of a structured approach such as the Conical Methodology [Nance 1987, 1981]. One key idea behind the use of a structured approach is to control the model complexity so that we can successfully verify and validate the model. The use of a structured approach is an important factor determining the success of a simulation project, especially for large-scale and complex models.

During the conceptualization of the model, one makes many assumptions in abstracting the reality. Each assumption should be explicitly specified. Model Qualification deals with the justification that all assumptions made are appropriate and the conceptual model provides an adequate representation of the system with respect to the study objectives.

4.5 Communicative Model Verification and Validation

In this stage, we confirm the adequacy of the communicative model to provide an acceptable level of agreement for the domain of intended application. *Domain of Intended Application* [Schlesinger et al. 1979] is the prescribed conditions for which the model is intended to match the system under study. *Level of Agreement* [Schlesinger et al. 1979] is the required correspondence between the model and the system, consistent with the domain of intended application and the study objectives.

Communicative Model Verification and Validation can be conducted by using informal and static analysis techniques shown in Figure 3 [Whitner and Balci 1989].

4.6 Programmed Model Verification and Validation

Whitner and Balci [1989] describe software testing techniques applicable for Programmed Model Verification and Validation. They provide a taxonomy of testing techniques in six categories as shown in Figure 3. The reader is referred to [Whitner and Balci 1989] for the details of the testing techniques.

4.7 Experiment Design Verification

The design of experiments can be verified by measuring the following indicators: (1) Are the algorithms used for random variate generation theoretically accurate?; (2) Are the random variate generation algorithms translated into executable code accurately? (Error may be induced by computer arithmetic or by truncation due to machine accuracy, especially with order statistics (e.g., $X = -\log_e(1-U)$) [Schmeiser 1981].); (3) How well is the random number generator tested? (Using a generator which is not rigorously shown to produce uniformly distributed independent numbers with sufficiently large period may invalidate the whole experiment design.); (4) Are *appropriate* statistical techniques implemented to design and analyze the simulation experiments? How well are the underlying assumptions satisfied? (See [Law 1983] for several reasons why output data analyses have not been conducted in an appropriate manner.); (5) Is the problem of the initial transient (or the start-up problem) [Wilson and Pritsker 1978] appropriately addressed?; and (6) For comparison studies, are identical

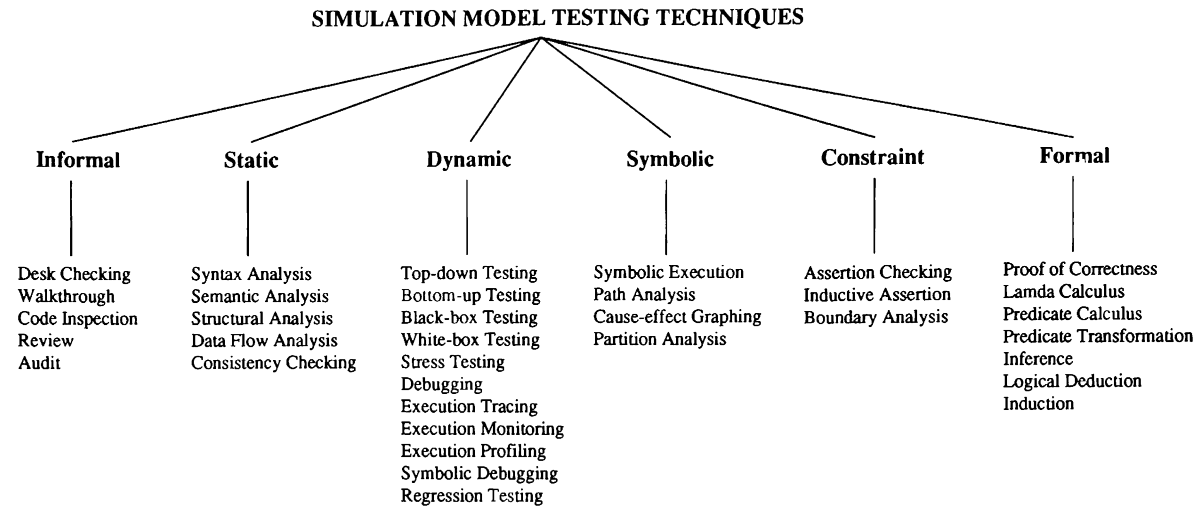

Figure 3. A Taxonomy for Simulation Model Testing Techniques

experimental conditions replicated correctly for each of the alternative operating policies compared?

4.8 Data Validation

In this stage, we confirm that the data used throughout the model development phases are accurate, complete, unbiased, and appropriate in their original and transformed forms. The data used can be classified as (1) model input data and (2) model parameters data.

Data validation deals with the substantiation that each input data model used possesses satisfactory accuracy consistent with the study objectives, and that the parameter values are accurately identified and used. Here are some indicators to measure data validity: (1) Does each input data model possess a sufficiently accurate representation?; (2) Are the parameter values identified, measured, or estimated with sufficient accuracy?; (3) How reliable are the instruments used for data collection and measurement?; (4) Are all data transformations done accurately? (e.g., are all data transformed correctly into the same time unit of the model?); (5) Is the dependence between the input variables, if any, represented by the input data model(s) with sufficient accuracy? (Blindly modeling bivariate relationships using only correlation to measure dependence is cited as a common error by Schmeiser [1981].); and (6) Are all data up-to-date?

4.9 Model Validation

Substantiating that the simulation model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives is called *Simulation Model Validation*. The *Domain of Applicability* is the set of prescribed conditions for which the experimental model has been tested, compared against the system to the extent possible, and judged suitable for use [Schlesinger et al. 1979].

Model validation is performed by comparing model behavior with system behavior when both model and system are driven under identical input conditions. Only under those input conditions we can claim model validity, because a model which is sufficiently valid under one set of input conditions can be completely absurd under another. If a model is used in a "what if" environment or if it is a forecasting model, the possible input conditions may form a very large domain over which model validation may become infeasible.

The existing literature on simulation model validation [Balci and Sargent 1984] generally falls into two broad areas: subjective validation techniques and statistical techniques proposed for validation. Tables 1 and 2 list these techniques.

The applicability of the techniques in Tables 1 and 2 depends upon the following cases where the system being modeled is: (1) completely observable – all data required for validation can be collected from system, (2) partially observable – some required data can be collected, and (3) nonexistent or completely unobservable. The techniques in Table 1 are described below. The statistical techniques in Table 2 are applicable only for case 1.

Event Validation employs identifiable events or event patterns as criteria against which to compare model and system behaviors. Events should be identified at a level of generality appropriate with the study objectives. Observing the same events in both the model and system outputs does not justify event validity if their temporal sequence differs. It may be necessary to weight events since a particular event may be more important for the model to replicate than another.

Face Validation is useful as a preliminary approach to validation. The project team members, potential users of the model, people knowledgeable about the system under study, based upon their estimates and intuition, subjectively compare model and system behaviors to judge whether the model and its results are reasonable.

Field Tests place the model in an operational situation for the purpose of collecting as much information as possible for model validation. These tests are especially useful for validating models of military combat systems. Although it is usually difficult, expensive, and sometimes impossible to devise meaningful field tests for complex systems, their use wherever possible helps both the project team and decision makers to develop confidence in the model.

Graphical Comparisons is a subjective, inelegant, and heuristic, yet quite practical approach especially useful as a preliminary approach. The graphs of values of model variables over time are compared with the graphs of values of system variables to investigate similarities in periodicities, skewness, number and location of inflection points, logarithmic rise and linearity, phase shift, trend lines, exponential growth constants, etc.

Historical Methods are quite philosophical in nature and explore three major methodological positions concerning the validation of especially economic models: rationalism, empiricism, and positive economics.

Hypothesis Validation deals with comparing hypothesized relationships among system variables with simulated ones. If X_s is identified to bear a given relationship to Y_s in the system, then X_m should bear a corresponding relationship to Y_m in a valid model.

Internal Validation is another preliminary approach to valida-

Table 1. Subjective Validation Techniques

Event Validation
Face Validation
Field Tests
Graphical Comparisons
Historical Methods
Hypothesis Validation
Internal Validation
Multistage Validation
Predictive Validation
Schellenberger's Criteria
Sensitivity Analysis
Submodel Testing
Turing Test

Table 2. Statistical Techniques Proposed for Validation

Analysis of Variance
Confidence Intervals/Regions
Factor Analysis
Hotelling's T^2 Tests
Multivariate Analysis of Variance
• Standard MANOVA
• Permutation Methods
• Nonparametric Ranking Methods
Nonparametric Goodness-of-fit Tests
• Kolmogorov-Smirnov Test
• Cramer-Von Mises Test
• Chi-square Test
Nonparametric Tests of Means
• Mann-Whitney-Wilcoxon Test
• Analysis of Paired Observations
Regression Analysis
Theil's Inequality Coefficient
Time Series Analysis
• Spectral Analysis
• Correlation Analysis
• Error Analysis
t-Test

tion. Holding all exogenous inputs constant, several replications of a stochastic model are made to determine the amount of stochastic variability in the model. The unexplained variance between these replications would provide a measure of internal validity.

Multistage Validation combines the three historical methods into a three-stage approach. In stage 1, a set of hypotheses or postulates are formulated using all available information such as observations, general knowledge, relevant theory, and intuition. In stage 2, it is attempted to justify model's assumptions where possible by empirically testing them. The third stage consists of testing the model's ability to predict system behavior.

Predictive Validation requires past data. The model is driven by past system input data and its forecasts are compared with the corresponding past system output data to test the predictive ability of the model.

Schellenberger's Criteria include (1) technical validation which involves the identification of all divergences between model assumptions and perceived reality as well as validity of the data used, (2) operational validation which addresses the question of how important the divergences are, and (3) dynamic validation which insures that the model will continue to be valid during its lifetime.

Sensitivity Analysis is performed by systematically changing the values of model input variables and parameters over some range of interest and observing the effect upon model behavior. Unexpected effects may reveal invalidity. The input values can also be changed to induce errors to determine the sensitivity of model behavior to such errors. Sensitivity analysis can identify those input variables and parameters to the values of which model behavior is very sensitive. Then, model validity can be enhanced by assuring that those values are specified with sufficient accuracy.

Submodel Testing refers to both submodel verification and submodel validation, and requires a top-down model decomposition in terms of submodels. The experimental model is instrumented to collect data on all input and output variables of a submodel. The system is similarly instrumented (if possible) to collect similar data. Then, each submodel behavior is compared with corresponding subsystem behavior. If a subsystem can be modeled analytically (e.g., as an M/M/1 model), its exact solution can be compared against the simulation solution to assess validity quantitatively.

Turing Test is based upon the expert knowledge of people about the system under study. These people are presented with two sets of output data obtained, one from the model and one from the system, under the same input conditions. Without identifying which one is which, the people are asked to differentiate between the two. If they succeed, they are asked how they were able to do it. Their response provides valuable feedback for correcting model representation. If they cannot differentiate, our confidence in model validity is increased.

4.10 Quality Assurance of Experimental Model

There are other indicators in addition to the ones presented in Sections 4.4 – 4.9 for assuring the quality of experimental model. These indicators correspond to the 7th branch at level 3 in Figure 2 and are derived from software quality characteristics. Execution efficiency, maintainability, portability, reusability are some important simulation model quality characteristics.

The dependence among the seven branches at level 3 in Figure 2 and the dependence among the corresponding indicators should be observed for assuring the quality of experimental model. An indicator with a low score may result in an unacceptable score for an indicator later in the life cycle. For example, failure to justify the reasonableness of an assumption in Model Qualification may result in poor model validity. Due to this dependence, it is important to detect errors as early as possible in the model development phases.

4.11 Credibility Assessment of Simulation Results

Concluding upon sufficient quality of the experimental model is a necessary but not a sufficient requirement for the credibility of simulation results. The experimental model quality is assured with respect to the definition of system and study objectives. An error made in defining the system or a study objective or failing to identify the *real* problem may cause unacceptable simulation results or an error of type II.

4.12 Presentation Verification

The simulation project management should verify the presentation of simulation results before they are presented to the panel of expert peers for acceptability assessment. There are four major indicators that need to be measured for the verification: (1) interpretation of simulation results, (2) documentation of simulation study, (3) communication of simulation results, and (4) presentation technique.

4.12.1 Interpretation of Simulation Results

Since all simulation models are descriptive, simulation results must be interpreted. In the simulation of an interactive computer system, for example, the model may produce a value of 20 seconds for the average response time; but, it does not indicate whether the value 20 is a "good" result or a "bad" result. Such a judgment is made by the simulation analyst depending upon the study objectives. Under one set of study objectives the value 20 may be too high; under another, it may be reasonable.

Naturally, if simulation results are not sufficiently credible the interpretation will be in error. For example, when comparing different operating policies, if experiments are not conducted under identical conditions, a policy cannot be claimed as the best. If the analyst fails to satisfy this requirement, the simulation results will be erroneous resulting in inaccurate interpretation. Therefore, interpretation accuracy is directly dependent on the credibility of simulation results.

The project team should review the way the results are interpreted in every detail to evaluate interpretation accuracy. Errors may be induced due to the complexity of simulation results, especially for large scale and complex models.

4.12.2 Documentation of a Simulation Study

Gass [1983] points out that "we do not know of any model assessment or modeling project review that indicated satisfaction with the available documentation." This problem should be attributed to the lack of automated support for documentation generation integrated with model development continuously throughout the entire life cycle [Balci 1986; Balci and Nance 1987].

A simulation study should be documented with respect to the phases, processes, and CASs of the life cycle.

4.12.3 Communication of Simulation Results

The simulation project team must devote sufficient effort in communicating technical simulation results to decision makers in a language they will understand. They must pay more attention to translating from the specialized jargon of the discipline into a form that is meaningful to the nonsimulationist and nonmodeler.

4.12.4 Presentation Technique

Simulation results may be presented to the decision makers as integrated within a Decision Support System (DSS). With the help of a DSS, a decision maker can understand and utilize the results much better. The integration accuracy of simulation results within the DSS must be verified.

If results are directly presented to the decision makers, the presentation technique (e.g., overheads, slides, films, etc.) must be ensured to be effective enough. The project management must make sure that the team members are trained and possess sufficient presentation skills.

4.13 Acceptability of Simulation Results

Acceptability of simulation study results is an attribute of the decision maker(s) or the sponsor. The management of a simulation project cannot control this attribute; however, they can significantly influence it by following the guidelines provided herein.

It is assumed that the acceptance of the study results implies their implementation. If the results are not implemented, they are considered to be rejected even if they are accepted by the sponsor.

5. CONCLUDING REMARKS

Although the life cycle of a simulation study is characterized with 13 CASs, a literature review [Balci and Sargent 1984] reveals that most work has concentrated on model validation and very little has been published on the other CASs. Model validity is a necessary but not a sufficient requirement for the credibility of simulation results. Sufficient attention must be devoted to every CAS in order for a simulation study to be successful.

A simulation study is multifaceted and multidisciplinary as illustrated by the life cycle presented herein. Sufficient effort must be devoted to every process of the life cycle. Inadequate coverage of a process, due to insufficient knowledge or time, may result in unacceptable results or an error of type II.

The list of indicators for the CASs is not intended to be exhaustive. Additional indicators that are specific to the area of application should be employed whenever possible. There is also the issue of assessing the validity and reliability of these indicators which is extremely difficult if not impossible for the broad scope adopted herein; however, for a specific area of application it should be achievable.

REFERENCES

Annino, J.S. and E.C. Russell (1979), "The Ten Most Frequent Causes of Simulation Analysis Failure – And How to Avoid Them!," *Simulation 32*, 6, 137-140.

Balci, O. (1986), "Requirements for Model Development Environments," *Computers & Operations Research 13*, 1, 53-67.

Balci, O. (1987), "Guidelines for Successful Simulation Studies," Technical Report TR-85-2, Department of Computer Science, VPI&SU, Blacksburg, VA.

Balci, O. (1988), "The Implementation of Four Conceptual Frameworks for Simulation Modeling in High-level Languages," In *Proceedings of the 1988 Winter Simulation Conference*, M.A. Abrams, P.L. Haigh, and J.C. Comfort, Eds. IEEE, Piscataway, NJ, 287-295.

Balci, O. and R.E. Nance (1985), "Formulated Problem Verification as an Explicit Requirement of Model Credibility," *Simulation 45*, 2, 76-86.

Balci, O. and R.E. Nance (1987), "Simulation Model Development Environments: A Research Prototype," *Journal of Operational Research Society 38*, 8, 753-763.

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

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

Banks, J. and J.S. Carson (1984), *Discrete-Event System Simulation*, Prentice-Hall, Englewood Cliffs, NJ.

Elmaghraby, S.E. (1968), "The Role of Modeling in IE Design," *Industrial Engineering 19*, 6, 292-305.

Fishman, G.S. (1978), *Principles of Discrete Event Simulation*, Wiley-Interscience, New York, NY.

Gass, S.I. (1983), "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis," *Operations Research 31*, 4, 603-631.

Law, A.M. (1983), "Statistical Analysis of Simulation Output Data," *Operations Research 31*, 6, 983-1029.

Law, A.M. and W.D. Kelton (1982), *Simulation Modeling and Analysis*, McGraw-Hill, New York, NY.

Martin, J. and C. McClure (1985), *Diagramming Techniques for Analysts and Programmers*, Prentice-Hall, Englewood Cliffs, NJ.

Nance, R.E. (1981), "Model Representation in Discrete Event Simulation: The Conical Methodology," Technical Report CS81003-R, Department of Computer Science, VPI&SU, Blacksburg, VA.

Nance, R.E. (1987), "The Conical Methodology: A Framework for Simulation Model Development," In *Methodology and Validation*, O. Balci, Ed. SCS, San Diego, CA, 38-43.

Overstreet, C.M. and R.E. Nance (1985), "A Specification Language to Assist in Analysis of Discrete Event Simulation Models," *Communications of the ACM 28*, 2, 190-201.

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

Schmeiser, B. (1981), "Random Variate Generation," In *Proceedings of the 1981 Winter Simulation Conference*, T.I. Ören, C.M. Delfosse, and C.M. Shub, Eds. IEEE, Piscataway, NJ, 227-242.

Shannon, R.E. (1975), *Systems Simulation: The Art and Science*, Prentice-Hall, Englewood Cliffs, NJ.

Shannon, R.E. (1986), "Intelligent Simulation Environments," In *Intelligent Simulation Environments*, P.A. Luker and H.H. Adelsberger, Eds. SCS, San Diego, CA, 150-156.

Whitner, R.B. and O. Balci (1989), "Guidelines for Selecting and Using Simulation Model Verification Techniques," In *Proceedings of the 1989 Winter Simulation Conference*, E.A. MacNair, K.J. Musselman, and P. Heidelberger, Eds. IEEE, Piscataway, NJ, 559-568.

Wilson, J.R. and A.A.B. Pritsker (1978), "A Survey of Research on the Simulation Startup Problem," *Simulation 31*, 2, 55-58.

Woolley, R.N. and M. Pidd (1981), "Problem Structuring – A Literature Review," *Journal of Operational Research Society 32*, 3, 197-206.