Simulation Modeling and Statistical Computing
N. Adam
Guest Editor

Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference

Tuncer I. Ören
University of Ottawa

The existing trend of application of computerized simulation studies to large and complex systems necessitates the development of an assessment methology for simulation studies. The basic concepts and criteria necessary for such an assessment methodology are presented in a systematic way. The proposed framework permits discussion of the concepts and criteria related to the acceptability of the following components of a simulation study: Simulation results, real world and simulated data, parametric model and the values of the model parameters, specification of the experimentation, representation and execution of the computer program, and modeling, experimentation, simulation, and programming methodologies or techniques used. The acceptability of the components of a simulation study are discussed with respect to the goal of the simulation study, the structure and data of the real system, the parametric model, the model parameter set, the specification of the experimentation, and the existing or conceivable norms of modeling methodology, experimentation technique, simulation methodology, and software engineering.

Key Words and Phrases: assessment of simulation studies, model validation, model certification, model robustness, model calibration, model fitting, figure of merit, design of experiments, software testing, software validation, program referability, program efficiency, code verification, software reliability, software robustness, software certification

CR Categories: 4.6, 8.1

Author's present address: Tuncer I. Ören, Department of Computer Science, University of Ottawa, Ottawa, Ontario, Canada K1N 9B4.

I. Introduction

Credibility of simulation studies, validation of models, and verification of simulation programs have long attracted the attention of both practitioners and theoreticians involved in simulation studies. There are several large-scale system studies such as environmental systems [29] or nuclear fuel waste management programs [7], which task the abilities of the traditional model-based methodologies.

As reported by Roth, Gass, and Lemoine, "The U.S. Government is the largest sponsor and consumer of models in the world. Estimates have indicated that over a half-billion dollars are being spent annually on developing, using, and maintaining computer models in the Federal Government" [63, p. 214]. Due to this fact, several symposia have already been organized by the U.S. National Bureau of Standards to discuss the problems of assessing different types of models. Several guidelines have already been made public by or for different agencies of the U.S. Federal Government [23, 24, 47]. However, the assessment of the acceptability of simulation studies has numerous other facets [52, 56]. A comprehensive and systematic view of the concepts and criteria related to the assessment of the acceptability of simulation studies is presented in this paper.

It is assumed that the reader is familiar with several simulation techniques and basic software engineering concepts. Simulation is experimentation with models. One of its characteristics is that it is extremely versatile. Hence there are different types of simulation studies based on the goal of the experimentation, the nature of the simuland, the nature and characteristics of the models, the nature and degree of involvement of the computer, and the type of application area [51]. Simulation can be applied to all three types of system problems, i.e., analysis, synthesis (design), and control (instrumentation) [34]. As a decision-making tool, simulation can be used in all five types of decisions, i.e., the descriptive, explanative, predictive, evaluative, and prescriptive decisions [61]. Simulation can be used to satisfy different goals such as product design, performance prediction, action prediction, experimentation with control strategies, testing theories, gaining insight, and arousing public interest [18]. Owing to this versatility, different criteria have to be considered in assessing the acceptability of different types of simulation studies.

A systematic consideration of the concepts and criteria that can be used in assessing the acceptability of simulation studies can be beneficial in comprehending the many facets of the problem and in thorough assessment. A relevant framework to systematize computer-aided modeling and model processing has recently been prepared [57].

Table I. Possibilities for the Assessment of the Acceptability of the Components of a Simulation Study.

Acceptability of: \ With respect to:		Goal of the Study	Real System: Structure	Real System: Data	Specific Model: Parametric model	Specific Model: Model parameter set	Another model	Experimentation Specification	Norms of: Modeling methodology	Norms of: Experimentation technique	Norms of: Simulation methodology	Norms of: Software Engineering
Simulation results		√										
Data	Real world data	√								√		
	Simulated data			√				√				
Model	Parametric model	√	√				√		√			
	Values of parameters		√	√								
Experimentation Specification	Experimental frame	√			√	√				√		
	Runs (number, length)	√								√		
Program	Representation Execution				√	√		√				√
Methodology/ Technique	Modeling								√			√
	Experimentation									√		
	Simulation										√	
	Programming											√

II. Partitioning the Problem

The problem of acceptability of a simulation study can be partitioned into the acceptability of different components of a simulation study such as simulation results, data, model, experimentation specification, the program, and methodologies or techniques used. Most of these components consist of more than one element: *Data* has to be considered as real system or simulated data. A *specific model* used in a simulation study consists basically of a parametric model where the values of the parameters need not be specified and a particular parameter set which consists of the values of the parameters of the model. *Specification of an experimentation* has two components: experimental frame(s) and simulation run(s). An *experimental frame* defines a limited set of circumstances under which the system (or the model) is to be observed or subjected to experimentation. It requires specification of the observational variables, input schedules, initialization, and termination conditions and collection, compression, and display of data [60]. A *simulation run* is the observation of the behavior of a specific model under an experimental frame. A *program* has both representation and execution aspects. Moreover, one has to consider different groups of methodology or technique-oriented issues of modeling, experimentation, simulation, and programming.

The acceptability of different components of a simulation study can be assessed with respect to the goal of the study, the structure and data of the real system, the parametric model, the model parameter set, the experimentation specification, and the conceived norms of the modeling methodology, experimentation technique, simulation methodology, and software engineering.

Table I depicts the possibilities for the assessment of the acceptability of the components of a simulation study. The components are listed in rows and the criteria to be used for assessing their acceptability are given in columns. Check marks relate the criteria to the components.

Groups of acceptability problems are discussed in the following order: acceptability of simulation results, data, models, experimentation specifications, programs, and methodologies/techniques used in the study. For every group a list of the subproblems are given, followed by definitions of selected terms.

III. Acceptability of Simulation Results

The problems relevant to the acceptability of simulation results are listed in Table II.

The results of a simulation study are evaluated with respect to the goal of the study. The evaluation of the recommendations of the simulation study consists of the

Table II. Acceptability of Simulation Results With Respect to the Goal of the Study.

Acceptability of the study
- Credibility of the study
- Cost-effectiveness of the study
- Timeliness of the recommendations
- Comprehensibility of the recommendations
  - Documentation of the study

Applicability of the recommendations
- Scope of the study
- Sensitivity of the recommendations

Utility of the study
(Usefulness of the study)
Certification of the simulation study

acceptability of the study, the applicability of the recommendations, the utility of the study, and the certification of the simulation study. Acceptability involves credibility, cost-effectiveness, timeliness, and comprehensibility of the study. Watt pointed out yet another possibility: the vested-interest of the decision-maker, especially in public systems [79]. The term, "...credibility of the study," implies the credibility of the people and/or institution responsible for the development of the study. The use of simulation techniques, especially in large-scale systems, increased the awareness of the necessity for better documentation of simulation studies. Other authors who have elaborated on the credibility of simulation results are Elzas [17, 18], Kahne [33], Lee [38], Mihram et al. [44], and Schruben [72].

Comprehensibility of the recommendations is only part of a much larger set of problems associated with the comprehensibility of the goal of the study: the model, the values of the parameters, the experimental conditions, and the set of questions poseable to a given model. As pointed out by Toffler, "Under conditions of high-speed change a democracy without the ability to anticipate condemns itself to death. But an anticipatory government without effective citizen participation and, indeed, control may be no less lethal. The future must neither be ignored nor captured by an elite. Only anticipatory democracy can provide a way out of the contradiction in which we find ourselves" [76, p. 104]. Ören expressed the similar wish that, "Simulation models will be comprehensible, especially to those who will be affected by the implications of particular models. Comprehensibility is paramount in the rational selection of models in a participatory democracy" [51, p. 183].

Applicability of the recommendations implies the scope of the study and the sensitivity of the recommendations. The utility (or usefulness) of the study can be assessed before or after the implementation of the recommendations depending on the nature of the simulation problem. In an analysis problem, the goal may simply be to obtain insight into a complicated system. The simulation study therefore may be useful without even considering any implementation. However, if it is a design (synthesis) or control problem, one must implement the results to assess the utility of the recommendations of a simulation study. A list of some of the causes of failure is given by Annino and Russell [2].

Certification of the simulation study, a concept not yet in current use, should be done by an agency independent from the developer and the user. Certification should embrace all the different components of the study.

IV. Acceptability of Data

The problems relevant to the acceptability of data are listed in Table III.

There are two types of data, i.e., real system data and simulated data; therefore, the acceptability of data has to be discussed as the acceptability of real system data and the acceptability of simulated data. Acceptability of real system data can be considered with respect to the goal of the study or with respect to the norms of experimentation techniques. Only the real world data which is relevant to the goal of the study has to be taken into consideration. Variables not readily quantifiable such as the ones representing human feelings should be treated carefully. An important problem is distinguishing measurement noise from data representing an unknown event or an event with a very low probability of occurrence.

Table III. Acceptability of Data.

- Acceptability of real-system data
  - Acceptability of real-system data with respect to the goal of the study
    - Relevance of real-system data
    - Existence of qualitative data
  - Acceptability of real-system data with respect to the norms of experimentation technique
    - Assessing the measurement noise
- Acceptability of simulated data
  - Acceptability of simulated data with respect to the real-system data collected under similar conditions
    - Applicability of validation
    - Model validity
      - —Replicative validity of a model
      - —Predictive validity of a model
  - Acceptability of simulated data with respect to experimentation specification (without reference to real-system data):
    - Internal validity of a model

The acceptability of simulated data has to be considered with respect to real system data collected under similar conditions. If the simulated data is acceptable, then the model used is said to be valid. However, a model which is valid under certain experimental conditions may not necessarily be valid under different experimental conditions. The literature on model validity abounds [4, 14, 21, 22, 25, 26, 27, 28, 30, 35, 40, 41, 42, 43, 44, 47, 62, 64, 65, 66, 67, 69, 70, 71, 73, 74, 75, 80]. Specialized support software for verification/validation is already in the literature [1] and several model evaluation programs exist [77, 78]. Elzas points out that validation is not always applicable [18, p. 182]. One must consider different types of systems, i.e., repeatable systems, recurrent systems, and unique systems.

> Repeatable systems can be the subjects of controlled experiments (e.g., chemical reactions, industrial processes, technical objects). Recurrent systems can generally not be experimented with, but at least present themselves to us with some repetition frequency in their discernable states (e.g., weather changes, stellar formations, specimen of a species). Unique systems can be characterized by a state history that occurs only once (e.g., a climate, the world, evolution). It is clear that validation is readily feasible for the first class of systems, can be considered for the second kind of systems (if enough time and data are available) and is out of the question for unique systems. So models of unique systems can only be speculative, cannot be verified or validated, but only 'trusted' to some extent if the modeling techniques used have led to useful results when applied to repeatable or recurrent systems. The main dilemma in the simulation field today is that we can provide some reliable information about a growing number of repeatable and recurrent systems while the society demands trustworthy predictive data on the development trends of unique systems (e.g., demography, energy, politics, etc.) [18, p. 182].

Validity terminology also abounds. For example, the terms, face validity, internal validity, multistage validity, event validity (or time-series validity), historical validity (replicative validity), and predictive validity are all used in the literature. A model is *replicatively* valid if it generates data which matches historical data collected from the real system under similar experimental conditions. A model is *predictively* valid if it generates data which matches the real system data to be generated at a later time under similar experimental conditions. A stochastic model is *internally* valid if the simulated data has a low variance when replicated with all exogenous inputs held constant [20, p. 204]. Therefore, one is concerned with the acceptability of simulated data with respect to the experimentation specification and without reference to real system data.

V. Acceptability of Models and Parameters

The problems relevant to the acceptability of models are listed in Table IV.

Table IV. Acceptability of Models and Parameters.

- Acceptability of a model
  - Model certification
  - Acceptability of a model with respect to the goal of the study
    - Model relevance
      - —Domain of intended application
      - —Range of applicability of a model
    - Model usability
      - —Model referability
      - —Model modifiability
    - Model comprehensibility
      - —Model documentation
    - Acceptability of a model with respect to its technical system specification
  - Acceptability of a model with respect to the real-system structure
    - Structure identification
    - Model validity
      - —Structural validity of a model
    - Algorithmic analysis of the concordance of units
  - Acceptability of a model with respect to another model
    - In model transformation
      - —Validity of model simplification
      - —Approximation in model simplification
      - —Validity of model elaboration
      - —Approximation in model elaboration
      - —Validity of renaming descriptive variables
    - In model comparison
      - —Model homomorphism
      - —Model isomorphism
      - —Model equivalencing
  - Acceptability of a model with respect to the modeling formalism
    - Model consistency
    - (mathematical correctness of the specification of a model)
    - Model robustness
- Acceptability of the parameters of a model with respect to the real system
  - Model fitting
  - Model calibration
  - Model identification

Acceptability of a model can be discussed on two levels: acceptability of a parametric model and acceptability of the values of the parameters of a model. At the highest level of acceptability one can consider model certification. Acceptability of a parametric model can be assessed with respect to the goal of the study, the real system structure, another model, and with respect to the modeling formalism.

Acceptability of a model with respect to the goal of the study is decomposed to relevance, usability, and comprehensibility of the model, and its acceptability with respect to its technical system specification. Model relevance refers to the domain of intended application and the range of applicability of a model. Model usability depends on model referability and model modifiability. Model referability can be enhanced by computerized model files [60]. A different approach is taken by Dickhoven [16] who uses software interfaces to tie together different model-oriented programs. In a computerized model file, component models may be connected to define large resultant models. Some of the component models can be modified before specifying the coupling. Model usability can be tremendously improved if the computer language used is not one of the traditional simulation languages. A model should not be cluttered with the specification of the experimentation. The values and ranges of the parameters should be specified separately [50, 53]. Some guidelines for model documentation already exist [48, 49]. Model comprehensibility should stem from the descriptive power of a model-oriented language and therefore model documentation should not depend solely on off-line or on-line commentaries about models. Computerized model documentation can answer different queries on the static and dynamic structure of a model. After a simulation run, more advanced documentation needs can also be satisfied. Traditionally the result of a simulation study is the time-trajectories of several descriptive variables of the model, a statistically compressed form of them, or a figure of merit summarizing how well the model satisfied its goal of existence in an operating condition. In more advanced simulation studies where, for example, the model is an adaptive one, the model configuration may change throughout simulation run. The output of the simulation run may then include, in addition to the traditional numerical results, a documentation of the change of the model composition to display its adaptive behavior [56, p. 45]. Thissen provides in-depth treatment of the model comprehensibility problem [74, 75].

In a design problem, the technical system specifications are given, permitting assessment of the simulation model. Acceptability of a model with respect to the real system structure involves structure identification, structural validity of a model, and algorithmic analysis of the concordance of the units. Zeigler proposed that structural validity of a model implies the model is not only replicatively and predictively valid but its internal operation mimics the operation of the real system [85]. Algorithmic analysis of the concordance of units is not yet implemented in simulation languages. It is hoped that it will become part of the algorithmic consistency check procedures for model-oriented languages [9, 56].

In traditional simulation the sole interest is generation of behavior, so algorithmic model manipulation is not a common practice. However, as was pointed out elsewhere, once a mathematical model is specified in a model-oriented system, the model can be used for several purposes other than simulation [52, 53, 57, 60, 85]. In model transformation, validity and approximation of simplification and elaboration of models can be considered. In model comparison, homomorphism, isomorphism, and several types of equivalencing of models can be considered.

If a well-defined modeling formalism is used as a framework for specifying models, then the acceptability of a model can be assessed with respect to the modeling formalism used. For example, if a Moore machine formalism is used to specify a finite state machine, then the modeler has to specify input, state, and output alphabets as well as the state-transition and output functions. Once the three alphabets are specified, then one could have, for example, the following consistency checks: The state-transition matrix should consist of elements of the state-alphabet; similarly, the output values specified as part of the output function should be valid output symbols of the output alphabet. If several Moore machines are coupled, then for every pair (M_1, M_2) of coupled Moore machines, the following consistency check can be done algorithmically: If M_1 provides inputs to M_2, then the output alphabet of M_1 should be a subset of the input alphabet of M_2 [58]. Similarly, formalisms for other types of mathematical models can be defined where algorithmic consistency checks can be performed to assess the acceptability of a model with respect to modeling formalism. For example, GEST language which identifies inputs, states, outputs, constants, parameters, tabular functions as well as derivatives and output functions for every component system of a coupled system provides similar possibilities for the assessment of the acceptability of models describing systems expressed by ordinary differential equations with or without discontinuities [54, 59, 19].

As pointed out by Hunt, systems which are not robust, "...do not degrade softly, or take proper default actions, when unpredictable input occurs. They fail abruptly, leaving stranded and bewildered the bureaucracts who serve them and the citizens ensnared by them ... To make the class of admissible inputs all-encompassing is a defeat; no system accepts *any* value as an admissible input." [32, p. 147]. Unfortunately, owing to the wrong emphasis of old simulation methodologies, model robustness has not been raised as an issue until recently [19, 51]. Practically all existing simulation models are not robust. They do not check admissibility of input values and therefore are ready to generate meaningless results should the input be erroneous. Acceptability of the parameters of a model with respect to real systems include assessment of the fitting, calibration, and identification of models [68, 80, 81].

VI. Acceptability of Experimentation Specifications

The problems relevant to the acceptability of experimentation specifications are listed in Table V.

Experimentation specification consists of the specification of the experimental frames and their application to specific models. There are three basic groups of problems related to the acceptability of experimentation specification: assessments with respect to the goal of the study, with respect to a specific model, and with respect to the norms of experimentation techniques.

Assessment of the experimentation specification with respect to the goal of the study includes assessing the relevance of the objective function of the study, determination of the data collection time period, determination of the length of a simulation run, and determination of the number of simulation runs.

A fundamental question is the selection of proper metric.[1] In *Platform for Change*, Beer elucidates the fact that, "The metric of money is too useful to abandon but it ought to be viewed essentially as the metric of constraint." [5]. This concept is especially important for large human systems [55].

At least two distinct types of performance measure and figure of merit of systems have to be distinguished. In one group, the computation necessitates a given scenario which is the case of a simulation run. In another group of performance measure computation, a set of, if not all, possible operating conditions have to be taken into consideration.

Determination of the number of simulation runs depends on the nature of the utilization of simulation as a decision-making technique. In simulation studies with discrete models, simulation is normally not considered to be part of an optimization scheme. Only a few alter-

Table V. Acceptability of Experimentation Specifications.

Acceptability of experimentation specification with respect to the goal of the study
Relevance of the objective function of the study
Measure of performance
Figure of merit
Determination of the data collection time period
to study transient behavior
to study steady-state behavior
Determination of the run length
Determination of number of simulation runs
Acceptability of experimentation specification with respect to a specific model
Applicability of an experimental frame to a model
Comparison of experimental frames
Acceptability of experimentation specification with respect to the norms of experimentation techniques
Evaluation of experimental design

[1] Some references on performance measures are [3, 83].

native situations are simulated and hence, the number of simulation runs is known in advance. The number of simulation runs is selected based on the analysis of the statistical design of experiments or as is the common practice, the number of scenarios to be evaluated. In simulation studies with continuous models, simulation can be used as part of an optimization scheme. In this case, at every simulation run, the behavior of the system under investigation is generated and an overall optimization algorithm may change the values of some of the parameters. The total number of simulation runs may even be unknown at the beginning of the simulation study.

Acceptability of experimentation specification with respect to a specific model involves assessment of the applicability of an experimental frame to a model and comparison of experimental frames. The first concept, i.e., applicability of an experimental frame to a model, is a systematic way of assessing whether or not a given question can be posed meaningfully to a specific model under well-defined experimental constraints [86, 60].

Algorithmic comparisons of experimental frames, to determine whether or not a frame is contained in another one, can save a simulation run if the results of the run with larger scope are available in a file. If a specific experimental design technique is used in a simulation study, whether or not the design is consistent with respect to the experimentation techniques, it can and should be evaluated. Some authors who have elaborated on the application of statistical techniques to simulation studies are Kleijnen [36, 37] and Mihram [41].

VII. Acceptability of Programs

The problems relevant to the acceptability of simulation programs are listed in Table VI(A), VI(B), VI(C), and VI(D).

A simulation program can be assessed with respect to two groups of criteria, i.e., the simulation study and the norms of software engineering. Assessing a program with respect to a simulation study consists of checking whether or not the program represents correctly and robustly the simulation model and the simulation experimentation. Software engineering techniques on software testing and validation can and should be applied for the assessment of simulation programs. The possible assessments are referability, reliability, and efficiency of programs.

Program referability has several dimensions such as modularity, comprehensibility, modifiability, and transportability of programs. Program modularity should be assessed under two major headings, i.e., modularity of user programs and modularity of model-oriented utility libraries. If the criteria listed under acceptability of program with respect to the simulation study are met, then the simulation program will already have well-defined modules to be assessed separately.

As can be seen in Table VI(B), in a model-oriented software system, several groups of software modules can exist as part of a library of programs. They can be considered under different headings: (1) Tutorial modules for the specifications of models and experimentations, (2) Modules to handle files of models and data, (3) Modules for algorithmic manipulation of models and experimentation specification, and (4) Simulation runtime library modules for structure-specification, generation, collection, compression, and display of data as well as for time advance and integration of continuous, piecewise continuous (or discontinuous), and stiff, differential equations.

Program comprehensibility should not depend on explicit commentary lines inserted but on the expressive power of a model-oriented language. Along this line, Nance promoted the idea of SMSDL (Simulation Model Specification and Documentation Language) [47]. The GEST (General System Theory Implementor) language, specified by Ören, serves a similar purpose [19, 54, 59].

Program modifiability can best be achieved if a model-oriented language allows the analyst to specify model and experimentation segments separately along the framework summarized in Table VI(B). The "simulation program" should then be generated with the assistance of the computer.[2]

Program transportability has two aspects, depending on whether or not the program is required to run on similar or different types of machines. In the first case, idiosyncrasies (such as different word lengths) of basically similar machines or the compilers may cause different results from the same program run on different machines. If the program is not transportable, it will not run on a similar machine. In the second case, as advocated by Elzas, a given simulation program should run on such different types of machines as digital or hybrid computers.

Table VI(A). Acceptability of Programs.

Acceptability of program with respect to the simulation study
- Specification of a specific model
  - Parametric model
    - Component models
      - Descriptive variables
        - —Input, state, output
        - —Constant, parameter, auxiliary variable
      - Functions
        - —State transtion, output
    - Coupling of component models
  - Set of values of model parameters
- Specification of the experimentation
  - experimental frame
  - simulation run
  - simulation study

Acceptability of program with respect to the norms of software engineering
- Program referability [see Table VI(B)]
- Program reliability [see Table VI(C)]
- Program efficiency [see Table VI(D)]

[2] A reference for program modifiability is Belady and Leavenworth [6].

Table VI(B). Program Referability.

- Program referability
  - Program modularity
    - Modularity of user program
      - Specific model
        - Parametric model
          - Component models
            - Descriptive variables
              - —Input, state, output
              - —Constant, parameter auxiliary variable
            - Functions
              - —State transition, output
          - Input/output relationship (coupling) of component models
        - Set of values of model parameters
      - Specification of the experimentation
        - Experimental frame
          - Observational variables
          - Input schedules (forcing functions)
          - Initialization
          - Termination conditions
          - Data collection/compression
        - Simulation run—figure of merit
        - Simulation study—overall figure of merit
    - Modularity of model-oriented utility library
      - Tutorial modules
        - for model specification
        - for experimentation specification
      - Modules to handle files
        - of models
        - of data—real-system data
          - —simulated data
      - Modules for algorithmic manipulation of
        - models—for consistency checks
          - —for comparisons
          - —for computerized documentation
        - experimentation specification
      - Simulation run-time library for
        - Special data-structures
          - sets, queues
        - data generation
          - special variable type
            - —declared random variables
          - input generators
        - data collection
        - data compression—statistical
          - —analytical
        - data display
        - time advance—event scheduling
        - integration
  - Program comprehensibility
    - Program documentation
  - Program modifiability
  - Program transportability
    - Same type of machine (different installations)
    - Different type of machine (i.e., digital or hybrid computer)

Table VI(C). Program Reliability.

- Program reliability
  - Program correctness
    - For already existing programs
      - Detection (and elimination) of errors (program verification)
        - —Program testing
        - —Program debugging
      - Proving lack of errors
        - Program analysis
          - —Static analysis
          - —Dynamic analysis
        - Program proving
          - —Inductive program proof
          - —Proof utilizing system specification morphism
    - For program generation (Techniques to realize correct programs)
      - Program generators from specifications
      - Increased use of well-tested (or proved) library routines
  - Program robustness
    - Defensive programming techniques
      - —Static checks
      - —Dynamic checks
  - Program certification

Program reliability is an important research area in software engineering [39, 45, 46, 84]. Some specific algorithms for simulation software have already been reported in the literature [9, 10, 11]. The results of most of these studies can and should be used to increase the reliability of simulation programs.

The assessment of the reliability of a program has three components, i.e., assessing its correctness, assessing its robustness and certifying its reliability. Program correctness techniques can be applied either for already existing programs or for program generation. In the case of already existing programs, there are two groups of assessments: detecting and eliminating errors (i.e., program verification) and proving lack of errors. Program verification is divided further into program testing and program debugging. Proving lack of errors is currently more challenging and consists of program analysis and program proving. A majority of program analysis techniques (static or dynamic) are already well-known from compiling techniques and optimization techniques applied to language processors [15, 31]. Program proving is the more challenging problem. There are already special

Table VI(D). Program Efficiency.

- Program efficiency
  - with respect to time:
    - Specification time
      - —model
      - —experimentation
    - Programming time
    - Compilation time
    - Execution time
    - File manipulation time
      - Simulation specification
        - —Model files
        - —Experimental frames
      - Data
        - —Real-system data
        - —Simulated data
    - Program modification
    - Simulation documentation time
    - Problem solution time
  - with respect to memory:
    - For compilation
    - For execution
    - For file manipulation
      - Simulation specification
        - —Model files
        - —Experimental frames
      - Data
        - —Real-system data
        - —Simulated data

proofs developed for simulation software such as the proof utilizing system specification morphism given by Cutler [12, 13].

Specification of simulation studies should be based on a well-structured framework as elaborated in the section on program modularity. The next generation of simulation software, based on advanced modeling and experimentation concepts, [60] will be generated using program synthesizers. In this way, several algorithms can be used to ensure correctness of specifications of different modules and the generated program. Since run-time simulation library routines are widely used in simulation studies, it is highly desirable to establish certification procedures for such routines.

Robustness of a program ensures erroneous data is not accepted that may generate meaningless results.

Program efficiency is an important factor although it should be stressed that program reliability is much more important. Program efficiency can be considered under two major categories: efficiency with respect to time and with respect to memory. When considered with respect to time, program efficiency encompasses the following types of time considerations: time for specification (model and experimentation), programming, compilation, execution, file manipulation, program modification, simulation documentation, and problem solution. With respect to memory, program efficiency must be evaluated for compilation, execution, and file manipulation.

VIII. Acceptability of Methodologies/Techniques Used

The problems relevant to the assessment of the acceptability of methodologies or techniques used in a simulation study are listed in Table VII.

In a simulation study, there are four technique or methodology groups: modeling, experimentation, simulation, and programming. Therefore, there are basically four groups of assessment problems.

Assessing the acceptability of the used experimentation technique with respect to the norms of experimentation techniques is different than assessing acceptability of a particular experimentation specification with respect to the specific experimentation technique adopted. In the first case, one has to evaluate the rationale of selecting a particular experimentation technique. In the second case, one has to evaluate whether or not the selected experimentation technique is used correctly.

In a simulation study, not only does coding of the simulation technique have to be done accurately, but a correct simulation technique has to be used to start with. Some of the possible problem areas are time advance, random variables, interpolation, and integration.

As far as time advance is concerned, in stochastic system simulation the time increment should be chosen carefully or the stochastic model may not hold. For example, if the time increment chosen is too large, information may disappear and assumptions may no longer be valid [20, p. 42].

Table VII. Acceptability of Methodologies/Techniques Used.

Acceptability of modeling methodology used
With respect to the norms of modeling methodology
suitability of modeling methodology
—to represent piecewise continuous models (discontinuous models)
—to represent adaptive systems where the structure of the system is time-varying
With respect to the norms of software engineering
—use of computerized model files
—computer-assisted modeling
—computerized consistency checks
Acceptability of experimentation technique used
With respect to the norms of experimentation techniques
Acceptability of simulation techniques used
With respect to the norms of simulation methodology
—Time advance
Choosing appropriate time increment for stochastic system simulation
—Random variables
Adequacy of the random variable generators
—Interpolation
Appropriate interpolation technique for continuous or discontinuous tabular functions
—Integration
Selection of proper integration method
—Automated selection of integration algorithm
—To handle discontinuities
Jump discontinuity
Derivative discontinuity
—To handle stiff system
—To handle real-time simulation
Correct use of the selected integration method
—Proper selection of error criteria
—Proper selection of stability criteria
—Proper application of backcasting (backward integration, or retrodiction)
Acceptability of programming technique used
With respect to the norms of software engineering
—Selection of appropriate language
—Programming
Systematic program construction
—By programmer
—By program synthesizer (program generator)

The stability criteria must also be selected properly [11] for integration. Backcasting (backward integration) was once believed to be a useful validation test [82] but Britting has shown that it can not be used for testing the validity of systems having negative feedback. Furthermore Britting states that,

> "Retrodictive tests draw conclusions from the model's behavior outside of its design range. The model builder is under no obligation to ensure reasonable behavior outside of the model's time span. As models are constructed to encompass longer time spans, the structure must generally be modified to account for the changing relationships between the variables and parameters as the system evolves with time. Thus it is clear that retrodictive testing cannot be a meaningful procedure for assessing the validity of models." [8, p. 148].

IX. Conclusion

Computerized simulation studies have many aspects. The acceptability and hence, the improvement possibilities of a simulation study depend on various factors.

The existing trend of application of computerized simulation studies to large and complex systems which task the abilities of analytical decision-making techniques necessitates the development of an assessment methodology. Basic concepts and criteria useful for such an assessment methodology were presented in a systematic way.

Acknowledgment. I wish to express my appreciation and thanks to the reviewers of this article for their constructive and valuable comments.

Received 4/80; revised 10/80; accepted 11/80

References

1. Allen, B. and Clema, J.K. Independent Verification/Validation Support Software. Proc. Nat'l Aerospace and Electron. Conf., Dayton, Ohio, March 1976, pp. 276–281.
2. Annino, J.S. and Russell, E.C. The ten most frequent causes of simulation analysis failure-and how to avoid them. *Simulation Today, 60,* (June 1979), pp. 137–140.
3. Bahr, D. A note on figures of merit. *ACM Performance Evaluations Rev., 4,* 3, (July 1975), pp. 1–3.
4. Balci, O. and Sargent, R.G. Bibliography on Validation of Simulation Models. *Newsletter–TIMS College on Simulation and Gaming,* Spring 1980, pp. 11–15.
5. Beer, S. *Platform for Change.* Wiley, New York, 1975.
6. Belady, L.A. and Leavenworth, B. Program Modifiability. IBM, Thomas J. Watson Research Center, Res. Rept. RC 8147, 1980.
7. Boulton, J. and Gibson, A.R. First Ann. Rept. Canadian Nuclear Fuel Waste Management Program, Rept AECL-6443, Atomic Energy of Canada Ltd., Pinawa, Manitoba, 1979.
8. Britting, K.R. Backward integration of system dynamics models–A useful validation test? In *World Modeling: A Dialogue.* C.W. Churchman and R.O. Mason (Eds.) Elsevier, New York, 1976, pp. 141–149.
9. Cellier, F.E. Combined continuous/discrete system simulation by use of digital computers: Techniques and tools, Ph.D. Th.,Diss ETH No. 6483, Zurich, Switzerland. 1979, (Chapter 6–Software Robustness, pp. 121–159).
10. Cellier, F.E. How to enhance the robustness of simulation software. In *Progress in Modelling and Simulation.* F.E. Cellier (Ed.) Academic (London), England, (In press).
11. Cellier, F.E. and Moebius, P.J. Towards robust general purpose simulation software. *ACM SIGNUM 14,* 2, (June 1979).
12. Cutler, M.M. Proving properties of simulation programs for system verification and validation. Proc., Summer Comptr Simulation Conf. Toronto, Canada, 1979, pp. 610–616.
13. Cutler, M.M. A formal program model for discrete event simulations and its use in the mathematical verification and validation of system models and implementations. Ph.D. Th., UCLA, Calif., 1979.
14. Deeter, C.R. and Hoffman, A.A.J. *An Annotated Bibliography of Mathematical Models Classified by Validation and Error Analysis Methods.* Texas Christian University, Fort Worth, TX 1978.
15. De Millo, R.A., Lipton, R.J., Perlis, A.J. Social processes and proofs of theorems and programs. *Comm. ACM, 22,* 5, (May, 1979), pp. 271–280.
16. Dickhoven, S. Software requirements for an improvement of transfer and adaptability of models. In *Utility and Use of Large-Scale Mathematical Models,* S. Gass (Ed.). Nat'l Bur Stand. Special Publication No. 534, U.S. Dept. of Commerce, Washington, D.C., May 1979, pp. 195–205.
17. Elzas, M.S. Whither simulation? Proc., UKSC Conf. on Comptr Simulation, IPC Science and Technology Press, Chester, England, April 1978, pp. 494–501.
18. Elzas, M.S. Results of a simulation: Can they be "trusted"? Proc. Seminar on Modelling in Business of IFIP Applied Information Processing Group, Amsterdam, The Netherlands, Jan. 1978, pp. 170–183.
19. Elzas, M.S. What is needed for robust simulation? In *Methodology in Systems Modelling and Simulation,* B.P. Zeigler et al. (Eds.) North-Holland, Amsterdam, 1979, pp. 57–91.
20. Emshoff, J.R. and Sisson, R.L. *Design and Use of Computer Simulation Models,* Collier–MacMillan, New York, 1970.
21. Gass, S.I. Evaluation of complex models. *Comptrs and Oper. Res. 4,* 1 (March, 1977), pp. 27–35.
22. Gass, S.I. A procedure for the evaluation of complex models. Proc. First Int. Conf. in Mathematical Models, Univ. of Missouri, 1977, pp. 247–258,
23. Gass, S.I. Computer model documentation: A review and an approach. Nat. Bur. Stand. Special Publication No. 500–39, U.S. Dept. of Commerce, Washington, D.C. Feb. 1979.
24. Gass, S.I. (Ed.). Utility and use of large-scale mathematical models. Proc. of a Workshop held at the Nat'l. Bur. Stand., Gaithersburg, Maryland, Apr. 28–29, 1977, National Bureau of Standards, Special Publication 534, Washington, D.C., U.S. Dept. of Commerce, May 1979.
25. Gass, S.I. Bibliography—Model evaluation and assessment. In *Utility and Use of Large-Scale Mathematical Models,* S.I. Gass (Ed.). Nat'l. Bur. Stand. Special Publication No. 534, U.S. Dept. of Commerce, Washington, D.C., May 1979, pp. 215–217.
26. Gass, S.I. *Validation and Assessment Issues of Energy Models.* Prepared for the NATO Advanced Research Institute, The Application of Systems Science to Energy Policy Planning. Nov. 12–16, Brookhaven National Laboratory, Upton, New York, 1979.
27. Greenberger, M. and Richels, R. Assessing energy policy models: Current state and future directions. *Ann. Rev. Energy, 4,* (1979).
28. Henize, J. Evaluating policy models. Proc. Summer Comptr Simulation Conf., Toronto, Canada, 1979, pp. 550–556.
29. Holling, C.S. *Adaptive Environmental Assessment and Management,* John Wiley, New Yok, 1978.
30. House, P.W. and McLeod, J. *Large-Scale Models for Policy Evaluation,* Wiley-Interscience, New York, 1977.
31. Howden, W.E. Applicability of software validation techniques to scientific programs. *ACM Trans. Programming Languages and Systems. 2,* 3, (July 1980), pp. 307–320.
32. Hunt, R. Large-scale systems theories: Some paradigms from software systems engineering. In *Cybernetics and Modelling and Simulation of Large-Scale Systems.* T.I. Ören (Ed.) Int. Assoc. for Cybernetics, Namur, Belgium, 1978, pp. 141–150.
33. Kahne, S. Model credibility for large-scale systems. *IEEE Trans. SMC,* (Aug. 1976), pp. 586–590.
34. Karplus, W.J. The spectrum of mathematical modelling and systems simulation. In *Simulation of Systems,* L. Dekker (Ed.) North-Holland, Amsterdam, 1976, pp. 5–13.
35. Kheir, N.A. and W.M. Homes. On validating simulation models of missile systems. *Simulation, 30,* 4, (Apr. 1978), pp. 117–128.
36. Kleijnen, J.P.C. Statistical design and analysis of simulation experiments. Proc. Simulation '75, Zürich, Switzerland Acta Press, 1975.
37. Kleijnen, J.P.C. The role of statistical methodology in simulation. In *Methodology in Systems Modelling and Simulation,* B.P. Zeigler, M.S. Elzas, G.J. Klir, and T.I. Ören (Eds.) North-Holland, Amsterdam, 1979, pp. 425–445.
38. Lee, D.B. Requiem for large-scale models. *J. Am. Inst. of Planners, 39,* 3 (May 1973), pp. 163–178.
39. Marmor-Squires, A.B. Software validation: An annotated bibliography. In *Current Trends in Programming Methodology. Vol. II. Program Validation,* R.T. Yeh (Ed.) Prentice-Hall, Englewood Cliffs, N.J., 1977, pp. 269–319.
40. McLeod, J. Terms and definitions–'Play it Again, Stew'. *Simulation Today, 29,* 4 (Oct. 1977) pp. vii–ix.
41. Mihram, G.A. *Simulation—Statistical Foundations and Methodology.* Academic Press, New York, 1972.
42. Mihram, G.A. Some practical aspects of the verification and validation of simulation models. *Operational Res. Quart., 23,* (1972) pp. 17–29.
43. Mihram, G.A. Four questions regarding the credibility of simulation. Proc. 7th Ann. Pittsburgh Conf., Part 2–Modeling and Simulation, Pittsburgh, PA, Instrum. Soc. Am. Apr 1976, pp. 1225–1230.
44. Mihram, G.A., G. Innis, E.J. Mastascusa, S. Schlesinger, and T.J. Teorey. Bibliography on simular credibility. *Simulation, 23,* 5, (Dec. 1974) p. vi.
45. Miller, E., and Howden, W.E. *Tutorial: Software Testing and Validation Techniques.* IEEE Comptr Soc., New York, NY, 1978.
46. Myers, G.J. *The Art of Software Testing.* Wiley, New York, 1979.
47. Nance, R.E. The feasibility of and methodology for developing federal documentation standards for simulation models. Rept. Dept. of Comptr Sci. VPI and State University, Blacksburg, VA, 1977.

48. Newton, O.L. and Weatherbee, J.E. Guidelines for documenting computer simulation models. *Project Memorandum* MV-068-060-USAF, Federal Computer Performance Evaluation and Simulation Center, Washington, DC, 1979.
49. Newton, O.L. and Whetherbee, J.E. Guidelines for documenting simulation models: A review and procedures. In *Simulation with Discrete Models: A State-of-the-Art View*, T.I. Ören, C.M. Shub, and P.F. Roth (Eds.). IEEE, New York, 1980 pp. 243–258.
50. Ören, T.I. Deductive general systems theories and simulation of large-scale systems. Proc., Summer Comptr Simulation Conf., Houston, Texas, July 1974 pp. 13–16.
51. Ören, T.I. Simulation–as it has been, is, and should be. *Simulation, 29*, 5 (Nov. 1977) pp. 182–183.
52. Ören, T.I. Modelling, model manipulation and programming concepts in simulation. In *Modelling, Identification, and Control in Environmental Systems*, G.C. van Steenkiste (Ed.) North-Holland, Amsterdam, 1978 pp. 833–842.
53. Ören, T.I. Rationale for large-scale systems simulation software based on cybernetics and general system theories. In *Cybernetics and Modelling of Large-Scale Systems*, T.I. Ören (Ed.) Int. Assoc. for Cybernetics, Namur, Belgium, 1978 pp. 151–179.
54. Ören, T.I. Reference manual of GEST 78–Level 1. (A modelling and simulation language for combined systems). Tech. Rept. 78-02, Univ. of Ottawa, Comptr Sci. Dept., Ottawa, Ontario, Canada, 1978.
55. Ören, T.I. Epilogue. In *Cybernetics and Modelling of Large-Scale Systems*. T.I. Ören (Ed.) Int. Assoc. for Cybernetics, Namur, Belgium, 1978, pp. 181–189.
56. Ören, T.I. Concepts for advanced computer-assisted modelling. In *Methodology in Systems Modelling and Simulation*, B.P. Zeigler, M.S. Elzas, G.J. Klir, and T.I. Ören (Eds.). North-Holland, Amsterdam, 1979, pp. 29–55.
57. Ören, T.I. Computer-aided modelling systems. In *Progress in Modelling and Simulation* F.E. Cellier (Ed.), Academic (London), England (In press).
58. Ören, T.I. and Collie, B. Design of SEMA: A software system for computer-aided modelling and simulation of SEquential MAchines. In Proc. 1980 Winter Simulation Conf., T.I. Ören, C.M. Shub, and P.F. Roth (Eds.), 1980, pp 113–123.
59. Ören, T.I. and den Dulk, J. Ecological models expressed in GEST 78, Rept prepared for the Dept. of Theoretical Production Ecology, Dutch Agricultural Univ., Wageningen, the Netherlands, 1978.
60. Ören, T.I. and Zeigler, B.P. Concepts for advanced simulation methodologies. *Simulation, 32*, 3 (March 1979) pp. 69–82.
61. O'Shaugnessy, J. *Inquiry and Decisions*. Allen & Unwin, London, 1972.
62. Pugh, R.E. *Evaluation of Policy Simulation Models*. Information Resources Press, Washington, DC, 1977.
63. Roth, P. F., Gass, S.I., and Lemoine, A.J. Some considerations for improving federal modelling. Proc. 1978 Winter Simulation Conf., Dec. 1978 pp. 213–217.
64. Salaway, G. and Siler, K.F. Validation considerations during parameter development for simulation models. Proc., Summer Comptr Simulation Conf., Toronto, Canada, July 1979 pp. 624–628.
65. Sargent, R.G. Validation of discrete-event simulation models. Proc. NSF Workshop on Validation of Mathematical Models in Energy Related Research and Development. Forth Worth, Texas, 1978.
66. Sargent, R.G. Validation of simulation models. Proc., Winter Simulation Conf., Dec. 1979 pp. 497–503.
67. Sargent, R.G. Verification and validation of simulation models. In *Progress in Modelling and Simulation*. F.E. Cellier (Ed.), Academic (London), England, (In Press).
68. Saridis, G.N. Comparison of six on-line identification algorithms. *Automatica, 10*, 1 (Jan. 1974) pp. 69–79.
69. Schellenberger, R.E. Criteria for assessing model validity for managerial purposes. *Decision Science, 5*, 4 (Oct. 1974) pp. 644–653.
70. Schlesinger, S., J.R. Buyan, E.D. Collender, W.K. Clarkson, and F.M. Perkins. Developing standard procedures for simulation validation and verification. Proc. Summer Comptr Simulation Conf., July 1974, pp. 927–933.
71. Schlesinger, S. et al. Terminology for model credibility. *Simulation, 32* 3, (1979) pp. 103–104.
72. Schruben, L.W. Establishing the credibility of simulations. *Simulation, 34*, 3 (March 1980) pp. 101–105.
73. Teorey, T.J. Validation criteria for computer systems simulation. Proc. Symp. Simulation of Computer Systems, Boulder, Colorado, Nat'l Bur Stand, Aug. 1975, pp. 161–173.
74. Thissen, W. Investigations into the Club of Rome's World 3 Model – Lessons for understanding complicated models. Ph.D. Th., Dept. of Technical Physics, Eindhoven University of Technology, The Netherlands, 1978.
75. Thissen, W. Investigations into the World 3 Model: Lessons for understanding complicated models. *IEEE Trans. on Systems, Man, and Cybernetics, SMC-8*, 3 (March 1978) pp. 183–193.
76. Toffler, A. *The Eco-spasm Report*. Bantam Books, New York, 1975.
77. U.S. Dep't Justice. Descriptions and assessments of the model evaluation program projects, Stock Number 27-000-00797-0, U.S. Govt Printing Office, Washington, DC., 1979.
78. Waller, J.D., Scanlon, J.W., Kemp, D.M., and Nalley, P.G. Developing useful evaluation capability: Lessons from the model evaluation program, Stock Number 27-000-00798-8, U.S. Govt Printing Office, Washington, DC., 1979.
79. Watt, K.F. Why won't anyone believe us? *Simulation, 28* (Jan. 1977), pp. 1–3.
80. Wigan, M.R. The fitting, calibration, and validation of simulation models. *Simulation, 18*, 5, (May 1972) pp. 188–192.
81. Wong. C.S.Y. Criteria sensitivity analysis: A new approach to parameter sensitivity. *Applied Mathematical Modelling. 4*, (Feb. 1980) pp. 7–15.
82. Wright, R.D. Retrodictive tests of dynamic models. Proc. Summer Comptr Simulation Conf., Montreal, Canada, July 1973.
83. Wymore, A.W. *Systems Engineering Methodology for Interdisciplinary Teams*. Wiley, New York, 1974.
84. Yeh, R.T. (Ed.) *Current Trends in Programming Methodoloy, Vol. II—Program Validation*. Prentice-Hall, Englewood Cliffs, NJ, 1977.
85. Zeigler, B.P. *Theory of Modelling and Simulation*. Wiley-Interscience, New York, 1976.
86. Zeigler, B.P. Structuring the Organization of Partial Models. In *Cybernetics and Modelling of Large-Scale Systems*. T.I. Ören (Ed.). Int. Assoc. for Cybernetics, Namur, Belgium, 1978 pp. 127–139.