Validation, verification, and testing techniques throughout the life cycle of a simulation study

Osman Balci

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

Life cycle validation, verification, and testing (VV&T) is extremely important for the success of a simulation study. This paper surveys current software VV&T techniques and current simulation model VV&T techniques and describes how they can all be applied throughout the life cycle of a simulation study. The processes and credibility assessment stages of the life cycle are described and the applicability of the VV&T techniques for each stage is stated. A glossary is provided to explicitly define important terms and VV&T techniques.

1. Introduction

Simulation is the process of constructing a model of a system which contains a problem and conducting experiments with the model on a computer for a specific purpose of experimentation to solve the problem. Credibility of simulation results not only depends on model correctness, but also is significantly influenced by accurate formulation of the problem. Therefore, validation, verification, and testing (VV&T) techniques must be employed throughout the life cycle of a simulation study starting with problem formulation and culminating with presentation of simulation results.

A *model* is a representation and an abstraction of anything such as a system, concept, problem, or phenomena. It can have inputs, parameters, and outputs as illustrated in figure 1. The term "system" is used to refer to the entity that contains the problem to be solved.

Model Validation is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives. Model validation deals with building the *right* model. It is conducted by running the model under the "same" input conditions that drive the system and by comparing model behavior with the system behavior. The comparison of model and system behaviors should *not* be made one output variable at a time, i.e., O_1^m versus O_1^s, O_2^m versus O_2^s, etc. A multivariate comparison should be carried out to incorporate the correlations among the output variables.

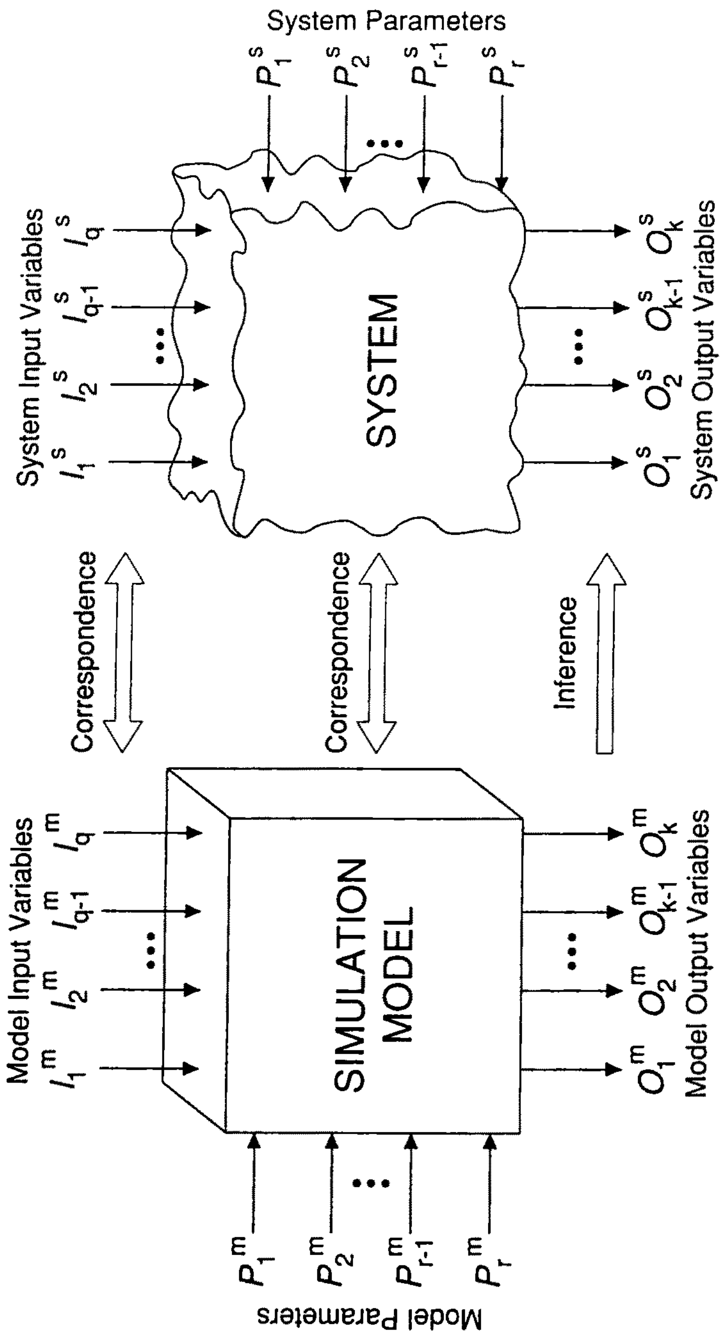
System Parameters
P_1^s P_2^s … P_{r-1}^s P_r^s
System Input Variables
I_1^s I_2^s … I_{q-1}^s I_q^s
SYSTEM
O_1^s O_2^s … O_{k-1}^s O_k^s
System Output Variables
Correspondence
Correspondence
Inference
Model Input Variables
I_1^m I_2^m … I_{q-1}^m I_q^m
SIMULATION MODEL
O_1^m O_2^m … O_{k-1}^m O_k^m
Model Output Variables
P_1^m P_2^m … P_{r-1}^m P_r^m
Model Parameters

Figure 1. Model and System Characteristics.

Model Verification is substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy. Model verification deals with building the model *right*. The accuracy of transforming a problem formulation into a model specification or the accuracy of converting a model representation in micro flowchart into an executable computer program is evaluated in model verification.

Model Testing is demonstrating that inaccuracies exist or revealing the existence of errors in the model. In model testing, we subject the model to test data or test cases to see if it functions properly. "Test failed" implies the failure of the model, not the test. Testing is conducted to perform validation and verification. Some tests are devised to evaluate the behavioral accuracy (i.e., validity) of the model, and some tests are intended to judge the accuracy of model transformation from one form into another (verification). Therefore, we commonly refer to the whole process as *model VV&T*.

Model VV&T is employed to prevent the occurrence of three major types of errors in conducting simulation studies [8]: *Type I Error* is the error of rejecting the model credibility when in fact the model is sufficiently credible. *Type II Error* is the error of accepting the model credibility when in fact the model is *not* sufficiently credible. *Type III Error* is the error of solving the wrong problem. Probability of committing the Type I Error is called *Model Builder's Risk* and probability of committing the Type II Error is called *Model User's Risk*. Committing the Type I error increases the cost of model development. The consequences of committing the Type II and Type III errors may be catastrophic. Therefore, a cost risk analysis should be conducted in those cases where data can be collected from the system under study [12].

Significant differences exist between simulation software engineering and other types of software engineering. First, simulation software engineering corresponds to simulation modeling and as such the art of modeling should be applied. Second, the results are obtained by experimenting with the simulation program (experimental model) as opposed to just executing it once like other types of programs. Third, the results are descriptive and must be carefully interpreted to come up with a solution to the problem. Fourth, software requirements specification corresponds to problem (system) description. Validation is conducted by comparing the model (computer program) with the system description as opposed to with the requirements specification. In spite of these differences, all software VV&T techniques are directly applicable for simulation models.

Every organization conducting a substantial simulation study should have a department or group called Simulation Quality Assurance (SQA). The SQA group is responsible for total quality management and closely works with the simulation project managers in planning, preparing test cases, and administering some of the VV&T activities throughout the simulation study. The SQA is a managerial approach which is critically essential for the success of a simulation study. Ören [71–73] presents concepts, criteria, and paradigms which can be used in establishing an SQA program within an organization.

The purpose of this paper is to survey current software VV&T techniques and current model VV&T techniques and describe how they can all be applied throughout the life cycle of a simulation study. Section 2 presents the life cycle of a simulation study and provides guidelines for conducting its ten processes. The VV&T techniques are briefly described under a taxonomy in section 3. Section 4 describes the credibility assessment stages of the life cycle and shows the applicability of the VV&T techniques for each stage. Concluding remarks and research directions are given in section 5. A glossary is provided in section 6 to explicitly define important terms and VV&T techniques.

2. The life cycle of a simulation study

The life cycle of a simulation study is presented in figure 2 [8, 64]. The phases are shown by shaded oval symbols. The dashed arrows describe the processes which relate the phases to each other. The solid arrows refer to the credibility assessment stages. Banks et al. [18] and Knepell and Arangno [51] review other modeling processes for developing simulations.

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. Every phase of the life cycle has an associated VV&T activity. Deficiencies identified by a VV&T activity may necessitate returning to an earlier process and starting all over again.

The VV&T is not a phase or step in the life cycle, but a continuous activity throughout the entire life cycle. Conducting the VV&T for the first time in the life cycle when the experimental model is complete is analogous to the teacher who gives only a final examination [40]. No opportunity is provided throughout the semester to notify the student that he or she has serious deficiencies. Severe problems may go undetected until it is too late to do anything but fail the student. Frequent tests and homeworks throughout the semester are intended to inform the students about their deficiencies so that they can study more to improve their knowledge as the course progresses.

The situation in the VV&T is exactly analogous. The VV&T activities throughout the entire life cycle are intended to reveal any quality deficiencies that might be present as the simulation study progresses from the communication of the problem until the implementation of the simulation results. This allows us to identify and rectify quality deficiencies during the life cycle phase in which they occur.

The ten processes of the life cycle are shown by the dashed arrows in figure 2. 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 below for each of the ten processes.

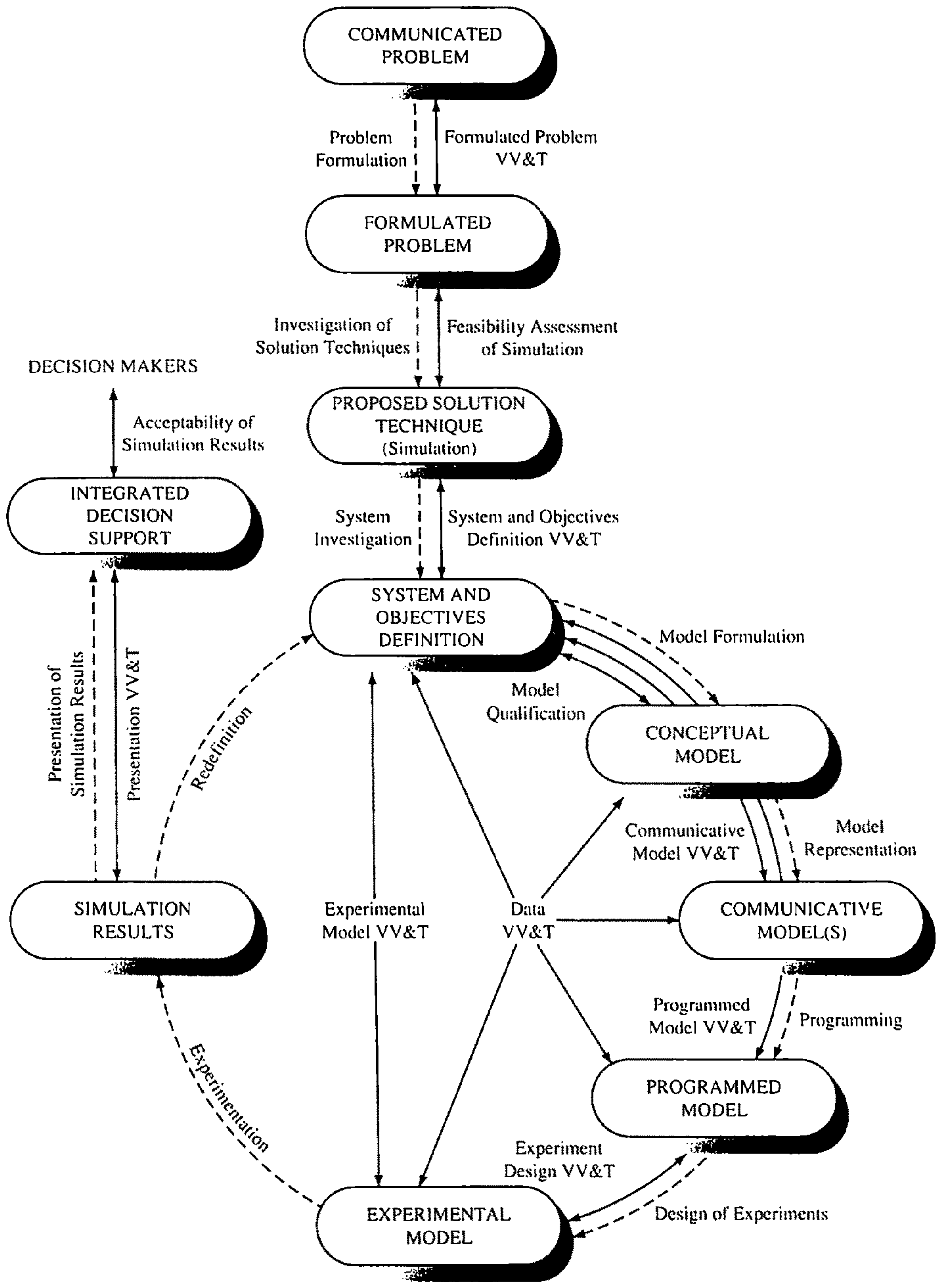

Figure 2. The life cycle of a simulation study.

2.1. 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 must follow. *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 [98].

Balci and Nance [10] present a high-level procedure that: (1) guides the analyst during problem formulation, (2) structures the formulated problem VV&T, and (3) seeks to increase the likelihood that the study results are utilized by decision makers.

2.2. 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 formulated problem VV&T and feasibility assessment of simulation before proceeding in the life cycle.

2.3. SYSTEM INVESTIGATION

Characteristics of the system that contains the formulated problem should be investigated for consideration in system definition and modeling. Shannon [87] identifies 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 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 [87]. 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 VV&T.

2.4. 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 [64]. Model formulation and model representation constitute the process of model design.

Input data analysis and modeling [54] 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 [54]. 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.

2.5. 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" [64]. 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 [70], and (6) more than a dozen diagramming techniques described in [56].

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.

2.6. 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 [7] describes how to conduct the programming process in high-level languages.

2.7. 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 [87]. 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 [54]. *Factorial designs* can be employed to determine the effect of various input variables on a response variable [32]. *Variance reduction techniques* can be implemented to obtain greater statistical accuracy for the same amount of simulation [54]. *Ranking and selection techniques* can be utilized for comparing alternative systems [54, 17]. Several methods (e.g., replication, batch means, regenerative) can be used for statistical analysis of simulation output data.

2.8. EXPERIMENTATION

This is the process of experimenting with the simulation model for a specific purpose. Some purposes of experimentation are [87]: (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.

2.9. 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.

2.10. 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.

3. Validation, verification, and testing techniques

Figure 3 shows a taxonomy which categorizes the VV&T techniques into six distinct credibility assessment perspectives: informal, static, dynamic, symbolic, constraint, and formal. The level of mathematical formality of each category increases from very informal on the far left to very formal on the far right. Likewise, the complexity also increases as the category becomes more formal [96].

It should be noted that some of the categories presented in figure 3 possess similar characteristics and in fact have techniques which overlap from one category to another. However, a distinct difference between each classification exists, as will be evident in the discussion of each in this section.

3.1. INFORMAL VV&T TECHNIQUES

Informal techniques are among the most commonly used ones. They are called informal because the tools and approaches used rely heavily on human reasoning and subjectivity without stringent mathematical formalism. The "informal" label does not imply any lack of structure or formal guidelines for the use of the techniques.

3.1.1. Audit

The audit is undertaken by a single person to investigate how adequately the simulation study is conducted with respect to established practices, standards, and guidelines. The audit also seeks to establish traceability within the simulation study. When an error is identified, it should be traceable to its source via its audit trail. Auditing is carried out on a periodic basis through a mixture of meetings, observations, and examinations [41].

3.1.2. Desk Checking

Desk Checking is the process of thoroughly examining one's work to ensure correctness, completeness, consistency, and unambiguity. It is considered to be the very first step in VV&T and is particularly useful for the early stages of development. To be effective, Desk Checking should be conducted carefully and thoroughly, preferably by another person since it is usually difficult to see one's own errors [2].

3.1.3. Face Validation

The project team members, potential users of the model, people knowledgeable about the system under study, based on their estimates and intuition, subjectively

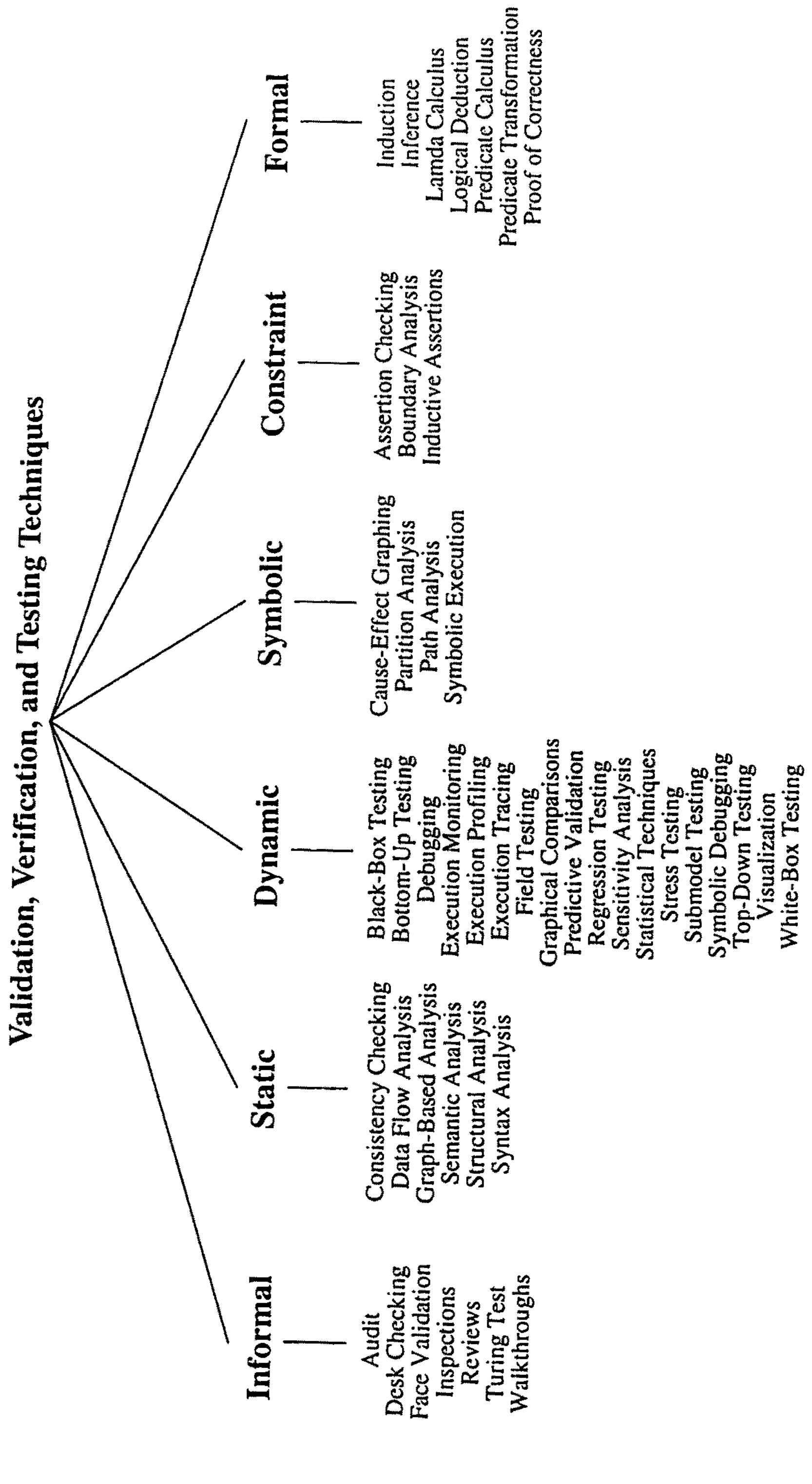

Figure 3. A taxonomy of validation, verification, and testing techniques.

compare model and system behaviors to judge whether the model and its results are reasonable. Face Validation is useful as a preliminary approach to validation [39].

3.1.4. Inspections

Inspections are conducted by a team of four to six members. For example, in the case of a design inspection, the team consists of: (1) Moderator: manages the inspection team and provides leadership; (2) Reader: narrates the model design (communicative model) and leads the team through it; (3) Recorder: produces a written report of detected faults; (4) Designer: is the creator of the model design; (5) Implementer: translates the model design into code (programmed model); and (6) Tester: SQA group representative.

An inspection goes through five distinct phases: overview, preparation, inspection, rework, and follow-up [82]. In phase I, the designer gives an overview of the (sub)model design to be inspected. The (sub)model characteristics such as purpose, logic, and interfaces are introduced and related documentation is distributed to all participants to study. In phase II, the team members prepare individually for the inspection by examining the documents in detail. The moderator arranges the inspection meeting with an established agenda and chairs it in phase III. The reader narrates the (sub)model design documentation and leads the team through it. The inspection team is aided by a checklist of queries during the fault finding process. The objective is to find and document the faults, not to correct them. The recorder prepares a report of detected faults immediately after the meeting. Phase IV is the rework in which the designer resolves all faults and problems specified in the written report. In the final phase, the moderator ensures that all faults and problems have been resolved satisfactorily. All changes must be examined carefully to ensure that no new errors have been introduced as a result of a fix.

A disadvantage of the inspection technique is that, like the walkthrough, it might be used for performance appraisal of the development team. Major differences exist between inspections and walkthroughs. An inspection is a five-step process, but walkthroughs consist of only two steps. The inspection team uses the checklist approach for uncovering errors. The procedure used in each phase of the inspection technique is formalized. The inspection process takes much longer than a walkthrough; however, the extra time is justified because an inspection is a powerful and cost-effective way of detecting faults early in the model development life cycle [1, 26, 52, 82].

3.1.5. Reviews

The review is conducted in a similar manner as the inspection and walkthrough except in the way the team members are selected. The review team also involves managers. The review is intended to give management and study sponsors evidence that the development process is being carried out according to stated study objectives and evaluate the model in light of development standards, guidelines, and specifications. As such, the review is a higher level technique than the inspection and walkthrough.

Each review team member examines the model documentation prior to the review. The team then meets to evaluate the model relative to specifications and standards, recording defects and deficiencies. The review team may be given a set of indicators to measure such as: (1) appropriateness of the definition of system and study objectives, (2) adequacy of all underlying assumptions, (3) adherence to standards, (4) modeling methodology used, (5) model representation quality, (6) model structuredness, (7) model consistency, (8) model completeness, and (9) documentation. The result of the review is a document portraying the events of the meeting, deficiencies identified, and review team recommendations. Appropriate action may then be taken to correct any deficiencies.

As opposed to inspections and walkthroughs, which concentrate on correctness assessment, reviews seek to ascertain that tolerable levels of quality are being attained. The review team is more concerned with model design deficiencies and deviations from stated model development policy than it is with the intricate line-by-line details of the implementation. This does not imply that the review team is not concerned with discovering technical flaws in the model, only that the review process is oriented towards the early stages of the model development life cycle [41, 96].

3.1.6. Turing Test

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 [86, 91, 94].

3.1.7. Walkthroughs

Walkthroughs are conducted by a team composed of a coordinator, model developer, and three to six other members. Except the model developer, all other members should not be directly involved in the development effort. A typical structured walkthrough team consists of: (1) Coordinator: most often is the SQA group representative who organizes, moderates, and follows up the walkthrough activities; (2) Presenter: most often is the model developer; (3) Scribe: documents the events of the walkthrough meetings; (4) Maintenance Oracle: considers long-term implications; (5) Standards Bearer: concerned with adherence to standards; (6) Client Representative: reflects the needs and concerns of the client; and (7) Other reviewers such as simulation project manager and auditors.

The main thrust of the walkthrough technique is to detect and document faults; it is not performance appraisal of the development team. This point must be made clear to everyone involved so that full cooperation is achieved in discovering errors.

The coordinator schedules the walkthrough meeting, distributes the walkthrough material to all participants well in advance of the meeting in order to allow for careful preparation, and chairs the meeting. During the meeting, the presenter walks the other members through the walkthrough documents. The coordinator encourages questions and discussion so as to uncover any faults [2, 23, 61, 62, 101].

3.2. STATIC VV&T TECHNIQUES

Static VV&T techniques are concerned with accuracy assessment on the basis of characteristics of the static model source code. Static techniques do not require machine execution of the model, but mental execution may be used. The techniques are very popular and widely used, with many automated tools available to assist the VV&T. The simulation language compiler is itself a static VV&T tool.

Static VV&T techniques can obtain a variety of information about the structure of the model, coding techniques and practices employed, data and control flow within the model, syntactical accuracy, and internal as well as global consistency and completeness of implementation [96].

3.2.1. Consistency checking

Consistency checking deals with substantiating that: (a) the model representation does not contain contradictions, (b) the cosmetic style with which language elements (e.g., naming conventions, use of upper, lower, and mixed case, etc.) are applied is used consistently, and (c) the data elements are manipulated uniformly (e.g., data assignment to variables, data use within computations, data passing among submodels, data representation and use during model input and output).

3.2.2. Data flow analysis

Data flow analysis is used to assess model accuracy with respect to the use of model variables. This assessment is classified according to the definition, referencing, and unreferencing of variables [2], i.e., when variable space is allocated, accessed, and deallocated. A data flow graph is constructed to aid in the data flow analysis. The nodes of the graph represent statements and corresponding variables. The edges represent control flow.

Data flow analysis can be used to detect undefined or unreferenced variables (much as in static analysis) and, when aided by model instrumentation, can track minimum and maximum variable values, data dependencies, and data transformations during model execution. It is also useful in detecting inconsistencies in data structure declaration and improper linkages among submodels [4, 96].

3.2.3. Graph-based analysis

Data-flow and control-flow are both graph-based analysis techniques which are similar in many ways [2]. Data-flow analysis is described in section 3.2.2. In control-flow analysis, a node of the model graph usually represents a logical junction where the flow of control changes, while an edge represents towards which junction it changes. This technique examines sequences of control transfers and is useful for identifying incorrect or inefficient constructs within model representation.

Nance and Overstreet [67] proposed several diagnostics which are based on analysis of graphs constructed from a particular form of model specification called Condition Specification [60, 70]. The diagnostic assistance is categorized into three: (1) analytical–determination of the existence of a property, (2) comparative–measures of differences among multiple model representations, and (3) informative–characteristics extracted or derived from model representations. Action cluster attribute graph, action cluster incidence graph, and run-time graph constitute the basis for the diagnosis.

The analytical diagnosis is conducted by measuring the following indicators: attribute utilization, attribute initialization, action cluster completeness, attribute consistency, connectedness, accessibility, out-complete, and revision consistency. The comparative diagnosis is done by measuring attribute cohesion, action cluster cohesion, and complexity. The following indicators are measured for the informative diagnosis: attribute classification, precedence structure, decomposition, and run-time graph [67].

3.2.4. Semantic analysis

Semantic analysis is conducted by the simulation programming language compiler and attempts to determine the modeler's intent in writing the code. The compiler informs the modeler about what is specified in the source code so that the modeler can verify that the true intent is accurately reflected.

The compiler generates a wealth of information to help the modeler determine if the true intent is accurately translated into the executable code: (1) *Symbol Tables* which describe: the elements or symbols that are manipulated in the model, function declarations, type and variable declarations, scoping relationships, interfaces, dependencies, etc.; (2) *Cross-reference Tables* which describe: called versus calling submodels (where each data element is declared, referenced and altered), duplicate data declarations (how often and where occurring), and unreferenced source code; (3) *Subroutine Interface Tables* which describe the actual interfaces of the caller and the called; (4) *Maps* which relate the generated runtime code to the original source code; and (5) "*Pretty Printers*" or *Source Code Formatters* which provide: reformatted source listing on the basis of its syntax and semantics, clean pagination, highlighting of data elements, and marking of nested control structures.

3.2.5. Structural analysis

Structural analysis is used to examine the model structure and to determine if it adheres to structured principles. It is conducted by constructing a control flow graph of the model structure and examining the graph for anomalies, such as multiple entry and exit points, excessive levels of nesting within a structure, and questionable practices such as the use of unconditional branches (i.e., GOTOs).

Yücesan and Jacobson [102] and Jacobson and Yücesan [46] apply the theory of computational complexity and show that the problem of verifying structural properties of simulation models is intractable. They illustrate that modeling issues such as accessibility of states, ordering of events, ambiguity of model specifications, and execution stalling are NP-complete decision problems.

3.2.6. Syntax analysis

Syntax analysis is carried on by the simulation programming language compiler to assure that the mechanics of the language are applied correctly.

3.3. DYNAMIC VV&T TECHNIQUES

Dynamic VV&T techniques require model execution and are intended for evaluating the model based on its execution behavior. Most dynamic VV&T techniques require model instrumentation.

The insertion of additional code (probes) into the executable model for the purpose of collecting information about model behavior during execution is called *model instrumentation*. Probe locations are determined manually or automatically based on static analysis of model structure. Automated instrumentation is accomplished by a preprocessor which analyzes the model static structure (usually via graph-based analysis) and inserts probes at appropriate places.

The simplest probe type is a counter and is illustrated in figure 4. In this example, computer jobs arrive according to a Poisson process. A fundamental assumption underlying a Poisson arrival process is that the interarrival times must be nonzero. Therefore, to find out what percentage of time this assumption is violated, the model in figure 4 is instrumented by inserting Blocks 3–6. Running the model for 50,000 jobs reveals that the percentage of violation (100*COUNT/Total number of entries into Block 4) is 18% (see [85, p. 165] for an explanation). To decrease this unacceptable frequency of violation, the time unit can be changed to 10 milliseconds; thus, changing 5 and 4 in the INITIAL Block to 50 and 40 reduces the percentage of violation to 2%. If 100 milliseconds were chosen as the time unit, the frequency of violation would only be 0.19%. Consequently, the higher the value of MIAT the lower the frequency of violation.

Dynamic VV&T techniques are usually applied using the following three steps. In step 1, the programmed or experimental model is instrumented. In step 2,

```
*************************************************************************
*  This is a GPSS/H programmed model of a single Central Processing Unit
*  (CPU) serving computer jobs arriving according to a Poisson process
*  with First-Come First-Served (FCFS) queue discipline. Job processing
*  times follow an exponential probability distribution.
*
*      MIAT        =  Mean InterArrival Time (1/arrival rate)
*      MPT         =  Mean Processing Time
*      NEWAT       =  New Arrival Time
*      OLDAT       =  Old Arrival Time
*      COUNT       =  Counter for zero interarrival times
*      Time Unit   =  Milliseconds
*************************************************************************
       SIMULATE
*
*      Initializations
*
       INITIAL      XH$MIAT,5/XH$MPT,4
*
*      Function Definition
*
 XPDIS FUNCTION     RN1,C24   Exponential Probability Distribution Function
0,0/.1,.104/.2,.222/.3,.355/.4,.509/.5,.69/.6,.915/.7,1.2/.75,1.38
.8,1.6/.84,1.83/.88,2.12/.9,2.3/.92,2.52/.94,2.81/.95,2.99/.96,3.2
.97,3.5/.98,3.9/.99,4.6/.995,5.3/.998,6.2/.999,7/.9998,8
*
*      Model Segment
*
       GENERATE     XH$MIAT,FN$XPDIS,,,,1PF     Job enters computer system
       ASSIGN       1,XH$MPT,XPDIS,PF           Record job processing time
       SAVEVALUE    NEWAT,AC1,XF                Record new arrival time (AT)
 TSTB  TEST E       XF$NEWAT,XF$OLDAT,OKAY      New AT = Old AT ?
       SAVEVALUE    COUNT+,1,XF                 Count zero interarrival time
 OKAY  SAVEVALUE    OLDAT,XF$NEWAT,XF           Set old AT = new AT
*
       QUEUE        SYSTEM                      Collect statistics
       SEIZE        CPU                         Job captures the CPU
       ADVANCE      PF1                         CPU processes the job
       RELEASE      CPU                         Job frees the CPU
       DEPART       SYSTEM                      Record statistics
       TERMINATE    1                           Job leaves computer system
*
*      Control Statements
*
       START        50000                       Run for 50,000 jobs
       END                                      End simulation
```

Figure 4. An instrumented GPSS/H programmed model.

the instrumented model is executed, and in step 3, the model output is analyzed and dynamic model behavior is evaluated.

3.3.1. Black-box testing

Black-box testing, also called functional testing, is used to assess the accuracy of model input output transformation. It is applied by feeding inputs (test data) to the model and evaluating the corresponding outputs. The concern is how accurately the model transforms a given set of input data into a set of output data.

It is virtually impossible to test all input-output transformation paths for a reasonably large and complex simulation model since the number of those paths could be in millions. Therefore, the objective of black-box testing is to increase our confidence in model input-output transformation accuracy as much as possible rather than trying to claim absolute correctness.

Generation of test data is a crucially important but a very difficult task. The law of large numbers does not apply here. Successfully testing the model under 1,000 input values (test data) does not imply high confidence in model input-output transformation accuracy just because 1,000 is a large number. Instead, we should compare the 1,000 with the number of allowable input values and determine how much of the model input domain is covered in testing. The more the model input domain is covered in testing, the more confidence we gain in the accuracy of the model input-output transformation [43, 62].

3.3.2. Bottom-up testing

Bottom-up testing is used in conjunction with bottom-up model development strategy. In bottom-up development, model construction starts with the submodels at the base level (i.e., the ones that are not decomposed further) and culminates with the submodels at the highest level. As each submodel is completed, it is thoroughly tested. When submodels belonging to the same parent have been developed and tested, the submodels are integrated and integration testing is performed. This process is repeated in a bottom-up manner until the whole model has been integrated and tested. The integration of completed submodels need not wait for all "same level" submodels to be completed. Submodel integration and testing can be, and often is, performed incrementally.

Some of the advantages of bottom-up testing are: (1) it encourages extensive testing at the submodel level; (2) since most well-structured models consist of a hierarchy of submodels, there is much to be gained by bottom-up testing; (3) the smaller the submodel and more cohesion it has, the easier and more complete its testing will be; and (4) it is particularly attractive for testing distributed simulation models.

Major disadvantages of bottom-up testing include: (1) individual submodel testing requires drivers, more commonly called test harnesses, which simulate the calling of the submodel and passing test data necessary to execute the submodel; (2) developing harnesses for every submodel can be quite complex and difficult; (3) the harnesses may themselves contain errors; and (4) faces the same cost and complexity problems as does top-down testing.

3.3.3. Debugging

Debugging is an iterative process the purpose of which is to uncover errors or misconceptions that cause the model's failure and to define and carry out the model changes that correct the errors. This iterative process consists of four steps. In step 1, the model is tested revealing the existence of errors (bugs). Given the detected errors, the cause of each error is determined in step 2. In step 3, the model changes believed to be required for correcting the detected errors are identified. The identified model changes are carried out in step 4. Step 1 is re-executed right after step 4 to ensure successful modification because a change correcting an error may create another one. This iterative process continues until no errors are identified in step 1 after sufficient testing [27].

3.3.4. Execution monitoring

Execution monitoring is used to reveal errors by examining low-level information about activities and events which take place during model execution. It requires the instrumentation of a programmed or experimental model for the purpose of gathering data to provide activity–or event-oriented information about the model's dynamic behavior. For example, the instrumented model in figure 4 monitors the number of zero interarrival times which is important to know for judging the input model validity. The model can also be instrumented to provide other low-level information such as number of jobs with zero processing time, average arrival rate, and average processing time.

3.3.5. Execution profiling

Execution profiling is used to reveal errors by examining high-level information (profiles) about activities and events which take place during model execution. It requires the instrumentation of a programmed or experimental model for the purpose of gathering data to present profiles about the model's dynamic behavior. For example, the model in figure 4 can be instrumented to produce the following profiles to assist in model VV&T: (1) a histogram of job interarrival times, (2) a histogram of job processing times, and (3) a histogram of job waiting times in the queue.

3.3.6. Execution tracing

Execution tracing is used to reveal errors by "watching" the line-by-line execution of a simulation model. It requires the instrumentation of a programmed or experimental model for the purpose of tracing the model's line-by-line dynamic behavior. For example, the model in figure 4 can be instrumented to record the values of NEWAT in a file during the model execution. Then, the interarrival times can be extracted from the trace data and can be statistically tested to see if they have an exponential probability distribution with mean MIAT *as intended.*

The major disadvantage of the tracing technique is that execution of the instrumented model may produce a large volume of trace data that may be too complex to analyze. To overcome this problem, the trace data can be stored in a database and the modeler can analyze it using a query language [30, 31].

3.3.7. Field Testing

Field Testing places the model in an operational situation for the purpose of collecting as much information as possible for model validation. It is 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 [87, 94].

3.3.8. Graphical Comparisons

Graphical Comparisons is a subjective, inelegant, and heuristic, yet quite practical approach especially useful as a preliminary approach to model VV&T. The graphs of values of model variables over time are compared with the graphs of values of system variables to investigate characteristics such as similarities in periodicities, skewness, number and location of inflection points, logarithmic rise and linearity, phase shift, trend lines, and exponential growth constants [21, 34, 57, 99].

3.3.9. Predictive Validation

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 [29].

3.3.10. Regression testing

Regression testing is used to substantiate that correcting errors and/or making changes in the model do not create other errors and adverse side-effects. It is usually accomplished by retesting the modified model with the previous test data sets used. Successful regression testing requires planning throughout the model development life cycle. Retaining and managing old test data sets are essential for the success of regression testing.

3.3.11. Sensitivity analysis

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 [87]. 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 [39, 58, 59, 94].

3.3.12. Statistical techniques

Much research has been conducted in applying statistical techniques for model validation. Table 1 presents the statistical techniques proposed for model validation and lists related references.

The statistical techniques listed in table 1 require that the system being modeled is completely observable, i.e., all data required for model validation can be collected from the system. Model validation is conducted by using the statistical

Table 1

Statistical techniques proposed for validation.

Analysis of variance	[68]
Confidence intervals/regions	[16, 54, 87]
Factor analysis	[21]
Hotelling's T^2 tests	[12–15, 87]
Multivariate analysis of variance	[37]
–Standard MANOVA	
–Permutation methods	
–Nonparametric ranking methods	
Nonparametric goodness-of-fit tests	[35, 68]
–Kolmogorov–Smirnov test	
–Cramer–Von Mises test	
–Chi-square test	
Nonparametric tests of means	[87]
–Mann–Whitney–Wilcoxon test	
–Analysis of paired observations	
Regression analysis	[3, 21, 44]
Theil's inequality coefficient	[48, 79, 90]
Time series analysis	
–Spectral analysis	[33, 36, 44, 45, 94, 95]
–Correlation analysis	[95]
–Error analysis	[22, 92]
t-test	[87, 89]

techniques to compare the model output data with the corresponding system output data when the model is run with the "same" input data that derive the real system. Due to the multiple response problem [87], the comparison of model and system outputs must be carried out by using a multivariate statistical technique to incorporate the correlations among the output variables.

A validation procedure based on the use of simultaneous confidence intervals is presented below.

3.3.12.1. A validation procedure using simultaneous confidence intervals

The behavioral accuracy (validity) of a simulation model with multiple outputs can be expressed in terms of the differences between the corresponding model and system output variables when the model is run with the "same" input data and operational conditions that drive the real system. The range of accuracy of the *j*th model output variable can be represented by the *j*th confidence interval (c.i.) for the differences between the means of the *j*th model and system output variables. The simultaneous confidence intervals (s.c.i.) formed by these c.i.'s are called the model range of accuracy (m.r.a.) [16].

Assume that there are k output variables from the model and k output variables from the system as shown in figure 1. Let $(\boldsymbol{\mu}^m)' = [\mu_1^m, \mu_2^m, \ldots, \mu_k^m]$ and $(\boldsymbol{\mu}^s)' = [\mu_1^s, \mu_2^s, \ldots, \mu_k^s]$ be the k-dimensional vectors of the population means of the model and system output variables, respectively. Basically, there are three approaches for constructing the s.c.i. to express the m.r.a. for the mean behavior.

In approach I, the m.r.a. is determined by the $100(1-\gamma)\%$ s.c.i. for $\boldsymbol{\mu}^m - \boldsymbol{\mu}^s$ as

$$[\boldsymbol{\delta}, \boldsymbol{\tau}], \tag{1}$$

where $\boldsymbol{\delta}' = [\delta_1, \delta_2, \ldots, \delta_k]$ representing lower bounds and $\boldsymbol{\tau}' = [\tau_1, \tau_2, \ldots, \tau_k]$ representing upperbounds of the s.c.i. We can be $100(1-\gamma)\%$ confident that the true differences between the population means of the model and system output variables are simultaneously contained within (1).

In approach II, the $100(1-\gamma^m)\%$ s.c.i. are first constructed for $\boldsymbol{\mu}^m$ as

$$[\boldsymbol{\delta}^m, \boldsymbol{\tau}^m], \tag{2}$$

where $(\boldsymbol{\delta}^m)' = [\delta_1^m, \delta_2^m, \ldots, \delta_k^m]$ and $(\boldsymbol{\tau}^m)' = [\tau_1^m, \tau_2^m, \ldots, \tau_k^m]$. Then, the $100(1-\gamma^s)\%$ s.c.i. are constructed for $\boldsymbol{\mu}^s$ as

$$[\boldsymbol{\delta}^s, \boldsymbol{\tau}^s], \tag{3}$$

where $(\boldsymbol{\delta}^s)' = [\delta_1^s, \delta_2^s, \ldots, \delta_k^s]$ and $(\boldsymbol{\tau}^s)' = [\tau_1^s, \tau_2^s, \ldots, \tau_k^s]$. Finally, using the Bonferroni inequality, the m.r.a. is determined by the following s.c.i. for $\boldsymbol{\mu}^m - \boldsymbol{\mu}^s$ with a confidence level of at least $(1-\gamma^m-\gamma^s)$ when the model and system outputs are dependent and with a level of at least $(1-\gamma^m-\gamma^s+\gamma^m\gamma^s)$ when the outputs are independent [50]:

$$[\boldsymbol{\delta}^m - \boldsymbol{\tau}^s, \;\; \boldsymbol{\tau}^m - \boldsymbol{\delta}^s]. \tag{4}$$

In approach III, the model and system output variables are observed in pairs and the m.r.a. is determined by the $100(1-\gamma)\%$ s.c.i. for $\boldsymbol{\mu}^d$, the population means of the differences of paired observations, as

$$\left[\boldsymbol{\delta}^d, \boldsymbol{\tau}^d\right], \tag{5}$$

where $(\boldsymbol{\delta}^d)' = [\delta_1^d, \delta_2^d, \ldots, \delta_k^d]$ and $(\boldsymbol{\tau}^d)' = [\tau_1^d, \tau_2^d, \ldots, \tau_k^d]$.

The approach for constructing the m.r.a. should be chosen with respect to the way the model is driven. The m.r.a. is constructed by using the observations collected from the model and system output variables by running the model with the "same" input data and operational conditions that drive the real system. If the simulation model is self-driven, then the "same" indicates that the model input data are coming independently from the same populations or stochastic process of the system input data. Since the model and system input data are independent of each other, but coming from the same populations, the model and system output data are expected to be independent and identically distributed. Hence, approach I or II can be used. The use of approach III in this case would be less efficient. If the simulation model is trace-driven, the "same" indicates that the model input data are exactly the same as the system input data. In this case, the model and system output data are expected to be dependent and identical. Therefore, approach II or III should be used.

Sometimes, the model sponsor, model user, or a third party may specify an acceptable range of accuracy for a specific simulation study. This specification can be made for the mean behavior of a stochastic simulation model as

$$\boldsymbol{L} \leq \boldsymbol{\mu}^m - \boldsymbol{\mu}^s \leq \boldsymbol{U}, \tag{6}$$

where $\boldsymbol{L}' = [L_1, L_2, \ldots, L_k]$ and $\boldsymbol{U}' = [U_1, U_2, \ldots, U_k]$ are the lower and upper bounds of the acceptable differences between the population means of the model and system output variables. In this case, the m.r.a. should be compared against (6) to evaluate model validity.

The shorter the lengths of the m.r.a., the more meaningful is the information they provide. The lengths can be decreased by increasing the sample sizes or by decreasing the confidence level. However, such increases in sample sizes may increase the cost of data collection. Thus, a trade-off analysis may be necessary among the sample sizes, confidence levels, half length estimates of the m.r.a, data collection method, and cost of data collection. For details of performing the trade-off analysis see [16].

The confidence interval validation procedure is presented in figure 5.

3.3.13. Stress testing

Stress testing is intended to test the model validity under extreme workload conditions. This is usually accomplished by increasing the congestion in the model.

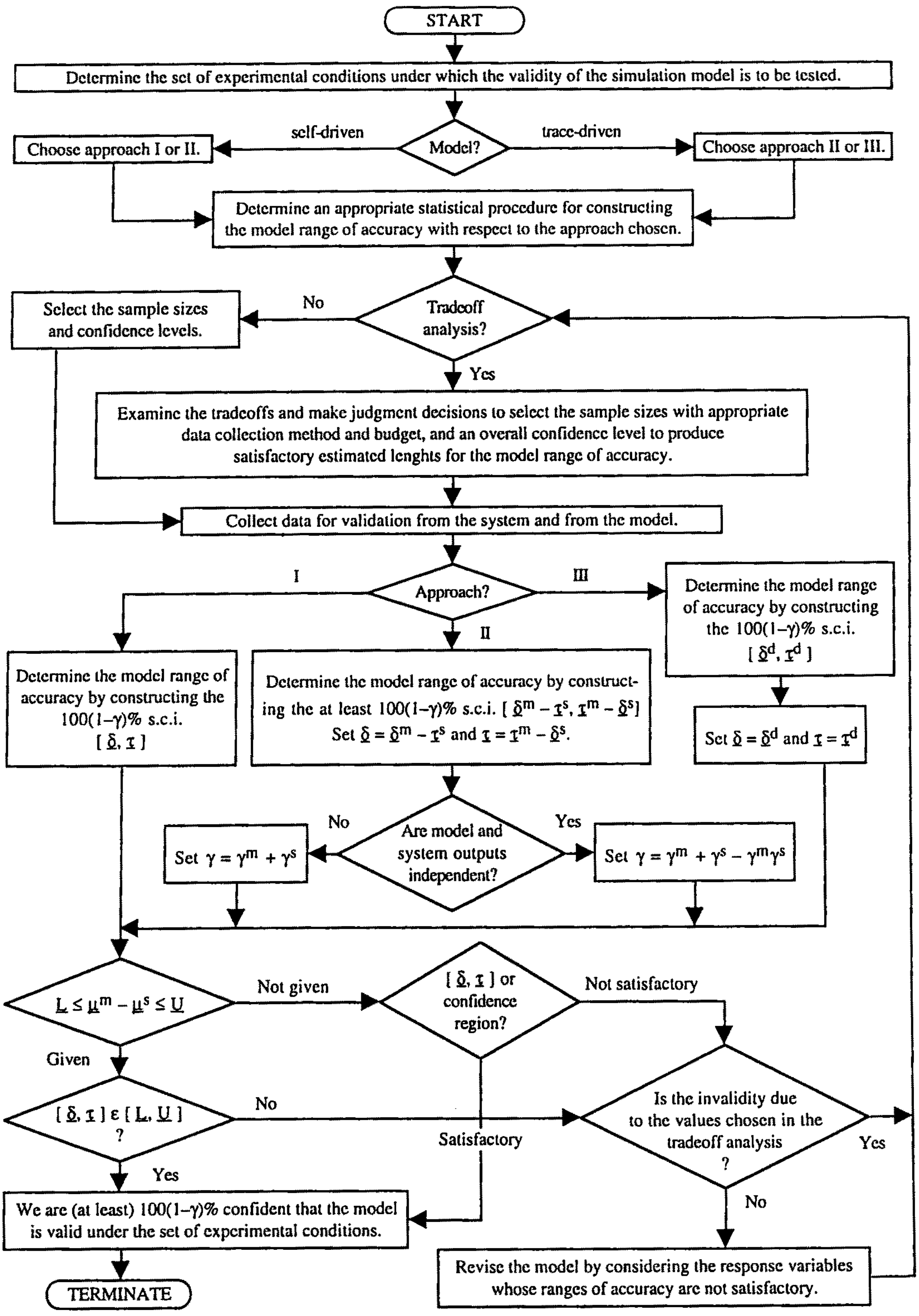
START
Determine the set of experimental conditions under which the validity of the simulation model is to be tested.
Model?
self-driven
trace-driven
Choose approach I or II.
Choose approach II or III.
Determine an appropriate statistical procedure for constructing the model range of accuracy with respect to the approach chosen.
Tradeoff analysis?
No
Yes
Select the sample sizes and confidence levels.
Examine the tradeoffs and make judgment decisions to select the sample sizes with appropriate data collection method and budget, and an overall confidence level to produce satisfactory estimated lenghts for the model range of accuracy.
Collect data for validation from the system and from the model.
Approach?
I
II
III
Determine the model range of accuracy by constructing the $100(1-\gamma)\%$ s.c.i. $[\underline{\delta}, \underline{\tau}]$
Determine the model range of accuracy by constructing the at least $100(1-\gamma)\%$ s.c.i. $[\underline{\delta}^m - \underline{\tau}^s, \underline{\tau}^m - \underline{\delta}^s]$ Set $\underline{\delta} = \underline{\delta}^m - \underline{\tau}^s$ and $\underline{\tau} = \underline{\tau}^m - \underline{\delta}^s$.
Determine the model range of accuracy by constructing the $100(1-\gamma)\%$ s.c.i. $[\underline{\delta}^d, \underline{\tau}^d]$
Set $\underline{\delta} = \underline{\delta}^d$ and $\underline{\tau} = \underline{\tau}^d$
Are model and system outputs independent?
No
Yes
Set $\gamma = \gamma^m + \gamma^s$
Set $\gamma = \gamma^m + \gamma^s - \gamma^m\gamma^s$
$\underline{L} \le \underline{\mu}^m - \underline{\mu}^s \le \underline{U}$
Not given
Given
$[\underline{\delta}, \underline{\tau}]$ or confidence region?
Not satisfactory
Satisfactory
$[\underline{\delta}, \underline{\tau}] \in [\underline{L}, \underline{U}]$?
No
Yes
Is the invalidity due to the values chosen in the tradeoff analysis ?
Yes
No
We are (at least) $100(1-\gamma)\%$ confident that the model is valid under the set of experimental conditions.
TERMINATE
Revise the model by considering the response variables whose ranges of accuracy are not satisfactory.

Figure 5. A validation procedure using simultaneous confidence intervals.

For example, the model in figure 4 can be stress tested by decreasing the mean interarrival time (MIAT), i.e., bringing in more jobs per unit time. Such increase in workload may create high congestion and result in an unacceptably high number of zero interarrival times causing incorrect representation of the input Poisson arrival process. Under stress testing, the model may exhibit invalid behavior; however, such behavior should be as expected and meaningfully documented. Without the instrumentation, the model in figure 4 could produce inaccurate results under such stress testing [27, 62].

3.3.14. Submodel testing

Submodel testing 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 to judge submodel validity. 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.

Validating each submodel individually does not imply sufficient validity for the whole model, because each submodel validity is passed with some error tolerance and the allowable errors can accumulate to make the whole model invalid. Therefore, after individually validating each submodel, the whole model itself must be subjected to testing.

3.3.15. Symbolic debugging

Symbolic debugging assists in model VV&T by employing a debugging tool that allows the modeler to manipulate model execution while viewing the model at the source code level. By setting "breakpoints", the modeler can interact with the entire model one step at a time, at predetermined locations, or under specified conditions. While using a symbolic debugger, the modeler may alter model data values or cause a portion of the model to be "replayed", i.e., executed again under the same conditions (if possible). Typically, the modeler utilizes the information from execution history generation techniques, such as tracing, monitoring, and profiling, to isolate a problem or its proximity. He then proceeds with the debugger to understand how and why the error occurs.

Current state-of-the-art debuggers allow viewing the runtime code as it appears in the source listing, setting "watch" variables to monitor data flow, viewing complex data structures, and even communicating with asynchronous I/O channels. The use of symbolic debugging can greatly reduce the debugging effort while increasing its effectiveness. Symbolic debugging allows the modeler to locate errors and check numerous circumstances which lead up to the errors [96].

3.3.16. Top-down testing

Top-down testing is used in conjunction with top-down model development strategy. In top-down development, model construction starts with the submodels at the highest level and culminates with the submodels at the base level (i.e., the ones that are not decomposed further). As each submodel is completed, it is thoroughly tested. When submodels belonging to the same parent have been developed and tested, the submodels are integrated and integration testing is performed. This process is repeated in a top-down manner until the whole model has been integrated and tested. The integration of completed submodels need not wait for all "same level" submodels to be completed. Submodel integration and testing can be, and often is, performed incrementally.

Top-down testing begins with testing the global model at the highest level. When testing a given level, calls to submodels at lower levels are simulated using submodel "stubs". A stub is a dummy submodel which has no other function than to let its caller complete the call. Fairley [31] lists the following advantages of top-down testing: (1) model integration testing is minimized, (2) early existence of a working model results, (3) higher level interfaces are tested first, (4) a natural environment for testing lower levels is provided, and (5) errors are localized to new submodels and interfaces.

Some of the disadvantages of top-down testing are: (1) thorough submodel testing is discouraged (the entire model must be executed to perform testing), (2) testing can be expensive (since the whole model must be executed for each test), (3) adequate input data is difficult to obtain (because of the complexity of the data paths and control predicates), and (4) integration testing is hampered (again, because of the size and complexity induced by testing the whole model) [31].

3.3.17. Visualization

Visualization (animation) of a simulation model greatly assists in model VV&T [80]. Displaying graphical images of internal and external dynamic behavior of a model during execution enables us to discover errors by seeing. For example, in visual simulation of a traffic intersection, we can observe the arrivals of vehicles in different lanes and their movements through the intersection as the traffic light changes. Seeing the visualization of the model as it executes and comparing it with the operations of the real traffic intersection can help us identify discrepancies between the model and the system.

Seeing the model in action is very useful for uncovering errors; however, seeing is not believing here [74]. Seeing model behavior during execution does not guarantee model correctness. Therefore, visualization should be used with caution in model VV&T.

3.3.18. White-box testing

White-box testing is used to evaluate the model based on its internal structure (how it is built) whereas black-box testing is intended for assessing the input-output transformation accuracy of the model. White-box testing employs data flow and control flow diagrams to assess the accuracy of internal model structure by examining model elements such as internal logic, internal data representations, submodel interfaces, and model execution paths. White-box testing is quite effective for detecting redundant code, faulty model structure, and special case errors.

3.4. SYMBOLIC VV&T TECHNIQUES

Symbolic VV&T techniques, like dynamic VV&T techniques, are used to evaluate the dynamic behavior of the model during execution.

3.4.1. Cause–effect graphing

Cause–effect graphing assists model VV&T by addressing the question of "what causes what in the model representation?" It is performed by first identifying causes and effects in the system being modeled and by examining if they are accurately reflected in the model specification. For example, in the simulation of a traffic intersection, the following causes and effects may be identified: (1) the change of lane 1 light to red immediately causes the vehicles in lane 1 to stop, (2) an increase in the duration of lane 1 green light causes a decrease in the average waiting time of vehicles in lane 1, and (3) an increase in the arrival rate of lane 1 vehicles causes an increase in the average number of vehicles at the intersection.

As many causes and effects as possible are listed, and the semantics are expressed in a cause–effect graph. The graph is annotated to describe special conditions or impossible situations. Once the cause–effect graph has been constructed, a decision table is created by tracing back through the graph to determine combinations of causes which result in each effect. The decision table is then converted into test cases with which the model is tested [62, 96].

3.4.2. Partition analysis

Partition analysis is used for testing the model with the test data generated by analyzing the model's functional representatives (partitions). It is accomplished by: (1) decomposing both model specification and implementation into functional representatives (partitions), (2) comparing the elements and prescribed functionality of each partition specification with the elements and actual functionality of corresponding partition implementation, (3) deriving test data to extensively test the functional behavior of each partition, and (4) testing the model by using the generated test data.

The model decomposition into functional representatives (partitions) is derived through the use of symbolic evaluation techniques which maintain algebraic expressions of model elements and show model execution paths. These functional representations are the model computations. Two computations are equivalent if they are defined for the same subset of the input domain which causes a set of model paths to be executed, and if the result of the computations is the same for each element within the subset of the input domain [42]. Standard proof techniques are used to show equivalence over a domain. When equivalence cannot be shown, partition testing is performed to locate errors, or as Richardson and Clarke [78] state, to increase confidence in the equality of the computations due to the lack of error manifestation. By involving both model specification and implementation, partition analysis is capable of providing more comprehensive test data coverage than other test data generation techniques.

3.4.3. Path analysis

Path analysis [42] attempts to assess model correctness on the basis of complete testing of all model control paths. It is performed in three steps. In step 1, the model control structure (e.g., through structural analysis) is determined and represented in a control flow diagram. In step 2, test data is generated to cause selected model logical paths to be executed. Symbolic execution can be used to identify and group together classes of input data based on the symbolic representation of the model. The test data is generated in such a way as to: (1) cover all statements in the path, (2) encounter all nodes in the path, (3) cover all branches from a node in the path, (4) achieve all decision combinations at each branch point in the path, and (5) traverse all paths [75]. In step 3, by using the generated test data, the model is forced to proceed through each path in its execution structure, thereby providing comprehensive testing.

In practice, only a subset of all possible model paths are selected for testing due to the budgetary constraints. Recent work has sought to increase the amount of coverage per test case or to improve the effectiveness of the testing by selecting the most critical areas to test. The path prefix strategy is an "adaptive" strategy that uses previous paths tested as a guide in the selection of subsequent test paths. Prather and Myers [75] prove that the path prefix strategy achieves total branch coverage.

The identification of essential paths is a strategy which reduces the path coverage required by nearly 40 percent [20]. The basis for the reduction is the elimination of non-essential paths. Paths which are overlapped by other paths are non-essential. The model control flow graph is transformed into a directed graph whose arcs (called primitive arcs) correspond to the essential paths of the model. Non-essential arcs are called inheritor arcs because they inherit information from the primitive arcs. The graph produced during the transformation is called an inheritor-reduced graph. Chusho [20] presents algorithms for efficiently identifying non-essential paths and reducing the control graph into an inheritor-reduced graph, and for applying the concept of essential paths to the selection of effective test data.

3.4.4. *Symbolic execution*

Symbolic execution is used to assess model accuracy by executing the model using symbolic values rather than actual data values for input. It is performed by feeding symbolic inputs into the (sub)model and producing expressions for the output which are derived from the transformation of the symbolic data along model execution paths. Consider, for example, the following function:

```
function jobArrivalTime(arrivalRate,currentClock,randomNumber)
  lag = -10
  Y = lag * currentClock
  Z = 3 * Y
  if Z < 0 then
    arrivalTime = currentClock - log(randomNumber) / arrivalRate
  else
    arrivalTime = Z - log(randomNumber) / arrivalRate
  end if
  return arrivalTime
end jobArrivalTime
```

In symbolic execution, lag is substituted in Y resulting in Y = –10*currentClock. Substituting again, we find z = –30*currentClock. Since currentClock is always zero or positive, an error is detected that z will never be greater than zero.

When unresolved conditional branches are encountered, a decision must be made which path to traverse. Once a path is selected, execution continues down the new path. At some point in time, the execution evaluation will return to the branch point and the previously unselected branch will be traversed. All paths eventually are taken.

The result of the execution can be represented graphically as a symbolic execution tree [2,49]. The branches of the tree correspond to the paths of the model. Each node of the tree represents a decision point in the model and is labeled with the symbolic values of data at that juncture. The leaves of the tree are complete paths through the model and depict the symbolic output produced.

Symbolic execution assists in showing path correctness for all computations regardless of test data and is also a great source of documentation. However, it has the following disadvantages: (1) the execution tree can explode in size and become too complex as the model grows, (2) loops cause difficulties although inductive reasoning and constraint analysis may help, (3) loops make thorough execution impossible since all paths must be traversed, and (4) complex data structures may have to be excluded because of difficulties in symbolically representing particular data elements within the structure [25,49,76].

3.5. CONSTRAINT VV&T TECHNIQUES

Constraint VV&T techniques are employed to assess model correctness using assertion checking, boundary analysis, and inductive assertions.

3.5.1. Assertion checking

An *assertion* is a statement that should hold true as the simulation model executes. Assertion checking is a verification technique used to check what *is* happening against what the modeler *assumes* is happening so as to guard model execution against potential errors. The assertions are placed in various parts of the model to monitor model execution. They can be inserted to hold true *globally* – or the whole model; *regionally* – for some submodels; *locally* – within a submodel; or at *entry* and *exit* of a submodel. The assertions are similar in structure and the general format for a local assertion is [88]:

```
ASSERT LOCAL (extended-logical-expression) [optional-qualifiers]
                                           [control-options]
```

The "optional-qualifiers" may be chosen as all, some, after jth job, before time t, etc. The "control-options" may have the following example syntax [88]:

$$\cdots \text{LIMIT n [VIOLATIONS]} \left[\left\{ \begin{array}{l} \underline{\text{HALT}} \\ \text{EXIT [VIA] proc - name} \end{array} \right\} \right] \cdots$$

For example, the programmed model in figure 4 can be instrumented by inserting the following local assertion right after the 5th Block to stop the simulation if the frequency of violation exceeds an allowable value of 1%:

```
ASSERT LOCAL((100*XF$COUNT/N$TSTB)'LE'1) LIMIT 1
```

where N$TSTB is a GPSS/H Standard Numerical Attribute giving the total number of entries into the Block labeled as TSTB.

Consider, for example, the following pseudo-code [96]:

```
Base := Hours * PayRate;
Gross := Base * (1 + BonusRate);
```

In just these two simple statements, several assumptions are being made. It is assumed that Hours, PayRate, Base, BonusRate, and Gross are all non-negative. The following asserted code can be used to prevent execution errors due to incorrect values inputted by the user:

```
Assert Local (Hours ≥ 0 and PayRate ≥ 0 and BonusRate ≥ 0);
Base := Hours * PayRate;
Gross := Base * (1 + BonusRate);
```

Clearly, the assertion checking serves two important needs: (1) it verifies that the model is functioning within its acceptable domain, and (2) the assertion statement documents the intentions of the modeler. However, the assertion checking degrades the model execution performance. If the execution performance is critical, the assertions should be labeled as comments and should be kept permanently to provide both documentation and means for maintenance testing [2].

Usually, a language should be defined for assertion specification [43] and a preprocessor is needed to translate assertions into the simulation programming language. Unfortunately, most of the current simulation programming languages do not provide assertion checking although this notion is not at all new, dating back to 1972 when Satterthwaite included an ASSERT statement in his version of Algol W [81].

3.5.2. Boundary analysis

Boundary analysis is employed to test model accuracy by using test cases on the boundaries of input equivalence classes. A model's input domain can usually be divided into classes of input data (known as equivalence classes) which cause the model to function the same way. For example, a traffic intersection model might specify the probability of left turn in a three-way turning lane as 0.2, the probability of right turn as 0.35, and the probability of travelling straight as 0.45. This probabilistic branching can be implemented by using a uniform random number generator that produces numbers in the range $0 \leq rn \leq 1$. Thus, three equivalence classes are identified: $0 \leq rn \leq 0.2$, $0.2 < rn \leq 0.55$, and $0.55 < rn \leq 1$. Each test case from within a given equivalence class has the same effect on the model behavior, i.e., produces the same direction of turn.

In boundary analysis, test cases are generated just within, on top of, and just outside of the equivalence classes [62]. In the example above, the following test cases are selected for the left turn: 0.0, ± 0.000001, 0.199999, 0.2, and 0.200001. In addition to generating test data on the basis of input equivalence classes, it is also useful to generate test data which will cause the model to produce values on the boundaries of output equivalence classes [62]. The underlying rationale for this technique as a whole is that the most error-prone test cases lie along the boundaries [69]. Notice that invalid test cases used in the example above will cause the model execution to fail; however, this failure should be as expected and meaningfully documented.

3.5.3. Inductive assertions

Inductive assertions are used to assess model correctness based on an approach that is very close to formal proof of model correctness. It is conducted in three steps. In step 1, input-to-output relations for all model variables are identified. In step 2, these relations are converted into assertion statements and are placed along the model execution paths in such a way as to divide the model into a finite number of "assertion-bound" paths, i.e., an assertion statement lies at the beginning and end of each model execution path. In step 3, verification is achieved by proving that for each path: if the assertion at the beginning of the path is true, and all statements along the path are executed, then the assertion at the end of the path is true. If all paths plus model termination can be proved, by induction, the model is proved to be correct [55, 77].

3.6. FORMAL VV&T TECHNIQUES

Formal VV&T techniques are based on formal mathematical proof of correctness. If attainable, formal proof of correctness is the most effective means of model VV&T. Unfortunately, "if attainable" is the overriding point with regard to formal VV&T techniques. Current state-of-the-art formal proof of correctness techniques are simply not capable of being applied to even a reasonably complex simulation model. However, formal techniques serve as the foundation for other VV&T techniques and the most commonly known seven techniques are covered below: (1) induction, (2) inference, (3) λ-calculus, (4) logical deduction, (5) predicate calculus, (6) predicate transformation, and (7) proof of correctness [47, 96].

Induction, *inference*, and *logical deduction* are simply acts of justifying conclusions on the basis of premises given. An argument is valid if the steps used to progress from the premises to the conclusion conform to established *rules of inference*. Inductive reasoning is based on invariant properties of a set of observations (assertions are invariants since their value is defined to be true). A typical inductive argument would be one similar to the one given in section 3.5.3 for inductive assertions: given that the initial model assertion is correct, it stands to reason that if each path progressing from that assertion can be shown to be correct, and subsequently each path progressing from the previous assertion is correct, etc., then the model must be correct if it terminates. Formal induction proof techniques exist for the intuitive explanation just given.

The *λ-calculus* [19] is a system for transforming the model into formal expressions. It is a string-rewriting system and the model itself can be considered as a large string. The λ-calculus specifies rules for rewriting strings, i.e., transforming the model into λ-calculus expressions. Using the λ-calculus, the modeler can formally express the model so that mathematical proof of correctness techniques can be applied.

The *predicate calculus* provides rules for manipulating predicates. A predicate is a combination of simple relations, such as *completed_jobs > steady_state length.* A predicate will either be true or false. The model can be defined in terms of predicates and manipulated using the rules of the predicate calculus. The predicate calculus forms the basis of all formal specification languages [86]. *Predicate transformation* [24, 100] provides a basis for verifying model correctness by formally defining the semantics of the model with a mapping which transforms model output states to all possible model input states. This representation provides the basis for proving model correctness.

Formal *proof of correctness* corresponds to expressing the model in a precise notation and then mathematically proving that the executed model terminates and it satisfies the requirements specification with sufficient accuracy [5]. Attaining proof of correctness in a realistic sense is not possible under the current state of the art. However, the advantage of realizing proof of correctness is so great that when the capability is realized, it will revolutionize the model VV&T.

4. Credibility assessment stages

It is very important to understand the principles of simulation model VV&T when applying the VV&T techniques throughout the entire life cycle of a simulation study. Balci [9] presents 15 principles that help the researchers, practitioners and managers better understand what model VV&T is all about. These principles serve to provide the underpinnings for the VV&T techniques described in section 3. Understanding and applying these principles is crucially important for the success of a simulation study.

Table 2 marks the VV&T techniques that are applicable for each of the ten credibility assessment stages described below. The more of these techniques we apply the more confidence we gain in the credibility of a life cycle phase. The VV&T activities should continue until a sufficient level of confidence is achieved.

4.1. FORMULATED PROBLEM VV&T

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. Consequently, 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 [98].

Formulated problem VV&T deals with substantiating 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 [10]. Failure to formulate the actual problem results in the Type III error. Once the Type III error is committed, regardless of how well the problem is solved, the simulation study will either end unsuccessfully or with the Type II error. Therefore, the accuracy of the formulated problem greatly affects the credibility and acceptability of simulation results.

Audit, cause–effect graphing, consistency checking, desk checking, face validation, inspections, reviews, structural analysis, and walkthroughs can be applied for conducting formulated problem VV&T. In applying cause–effect graphing, a causality network is created to analyze the potential root causes of the communicated problem [10]. The questionnaire developed by Balci and Nance [10] with 38 indicators can be used in applying audit, inspections, reviews, and walkthroughs.

Table 2

Applicability of the VV&T techniques for the credibility assessment stages.

	FP VV&T	FA of Sim.	S&OD VV&T	Model Qual.	CM VV&T	PM VV&T	ED VV&T	Data VV&T	EM VV&T	Pres. VV&T
Assertion checking						✓	✓	✓	✓	
Audit	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Black-box testing						✓	✓		✓	
Bottom-up testing						✓			✓	
Boundary analysis						✓			✓	
Cause–effect graphing	✓				✓	✓			✓	
Consistency checking	✓		✓	✓	✓	✓	✓	✓	✓	✓
Data flow analysis					✓	✓	✓	✓	✓	
Debugging						✓	✓		✓	
Desk Checking	✓		✓	✓	✓	✓	✓	✓	✓	✓
Execution Monitoring						✓	✓		✓	
Execution Profiling						✓	✓		✓	
Execution Tracing						✓	✓		✓	
Face Validation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Field Testing									✓	
Graph-based analysis					✓	✓	✓		✓	
Graphical comparisons						✓	✓		✓	
Induction						✓			✓	
Inductive assertions						✓			✓	
Inference						✓			✓	
Inspections	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lambda calculus						✓			✓	
Logical deduction						✓			✓	
Partition analysis						✓			✓	
Path analysis						✓	✓		✓	
Predicate calculus						✓			✓	
Predicate transf.						✓			✓	
Predictive validation									✓	
Proof of correctness						✓			✓	
Regression testing						✓			✓	
Reviews	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Semantic analysis						✓	✓		✓	
Sensitive analysis						✓	✓		✓	
Statistical techniques								✓	✓	
Stress testing						✓			✓	
Structural analysis	✓		✓		✓	✓	✓		✓	✓
Submodel testing						✓			✓	
Symbolic debugging						✓	✓		✓	
Symbolic execution						✓	✓		✓	
Syntax analysis						✓	✓		✓	
Top-down testing						✓			✓	
Turing test									✓	
Visualization						✓			✓	
Walkthroughs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
White-box testing						✓			✓	

4.2. FEASIBILITY ASSESSMENT OF SIMULATION

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 use. Skipping the investigation of solution techniques process may result in unnecessarily expensive solutions, sometimes to the wrong problems.

As a result of the investigation of solution techniques 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 formulated problem VV&T and feasibility assessment of simulation before proceeding in the life cycle.

Audit, face validation, inspections, reviews, and walkthroughs can be applied for assessing the feasibility of simulation with the use of indicators such as: (1) Are the benefits and cost of simulation solution estimated correctly?; (2) Do the potential benefits of simulation solution justify the estimated cost of obtaining it?; (3) Is it possible to solve the problem using simulation within the time limit specified?; (4) Can all of the resources required by the simulation project be secured?; and (5) Can all of the specific requirements (e.g., access to pertinent classified information) of the simulation project be satisfied?

4.3. SYSTEM AND OBJECTIVES DEFINITION VV&T

For the purpose of generality, the term "system" is used to refer to the entity that contains the formulated problem. System and objectives definition VV&T deals with assessing the credibility of the system investigation process in which system characteristics are explored for consideration in system definition and modeling. Shannon [87] identifies 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 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 behavior of interest (e.g., average waiting time of vehicles). 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 [87]. 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.

Audit, consistency checking, desk checking, face validation, inspections, reviews, structural analysis, and walkthroughs can be applied for conducting system and objectives definition VV&T by using indicators such as: (1) 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 six months to several years)?; (2) Is the system's environment (boundary) identified correctly?; (3) What counterintuitive behavior may be caused within the system and its environment?; (4) Will the system significantly drift to low performance requiring a periodic update of the system definition?; and (5) Are the interdependency and organization of the system characterized accurately? The objective here is to substantiate 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.

4.4. MODEL QUALIFICATION

Model qualification is intended for assessing the credibility of the model formulation process. 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 [64]. Model formulation and model representation constitute the process of model design. *Input data analysis and modeling* [54] is a subprocess of model formulation process and is conducted with respect to the way the model is driven, self- or trace-driven.

Under some study objectives (e.g., evaluation, comparison, determination of functional relations) and for model validation, input data model is 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 [54]. In a trace-driven simulation, we trace the system (e.g., using hardware or software monitors) and utilize the refined trace data as the input data model to use in driving the simulation model.

A model should be conceptualized under the guidance of a structured approach such as the Conical Methodology [64–66]. 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. Audit, consistency checking, desk checking, face validation, inspections, reviews, and walkthroughs can be applied for conducting model qualification.

4.5. COMMUNICATIVE MODEL VV&T

Communicative model VV&T is concerned with substantiating the sufficient accuracy of the model representation process which 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" [64]. 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.

Communicative model VV&T deals with confirming the adequacy of the communicative model to provide an acceptable level of agreement for the domain of intended application. *Domain of Intended Application* [83] is the prescribed conditions for which the model is intended to match the system under study. *Level of Agreement* [83] is the required correspondence between the model and the system, consistent with the domain of intended application and the study objectives.

Audit, cause–effect graphing, consistency checking, data flow analysis, desk checking, face validation, graph-based analysis, inspections, reviews, structural analysis, and walkthroughs can be applied for conducting communicative model VV&T.

4.6. PROGRAMMED MODEL VV&T

Programmed model VV&T deals with the assessment of the process of programming which is the process of translating the communicative model into a programmed model. A *Programmed Model* is an executable simulation model representation in a simulation programming language such as GPSS, SIMAN, SIMSCRIPT, SIMULA, and SLAM or in a high-level programming language such as C++, Fortran, and Pascal. The programmed model does not incorporate an experiment design.

The techniques applicable for conducting communicative model VV&T are marked in table 2.

4.7. EXPERIMENT DESIGN VV&T

Experiment design VV&T deals with substantiating the sufficient accuracy of the process of design of experiments which is the process of formulating a plan to gather the desired information at minimal cost and to enable the analyst to draw valid inferences [87]. 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 [54]. *Factorial designs* can be employed to determine the effect of various input variables on a response variable [32]. *Variance reduction techniques* can be implemented to obtain greater statistical accuracy for the same amount of simulation [54]. *Ranking and selection techniques* can be utilized for comparing alternative systems [54, 17]. Several methods (e.g., replication, batch means, regenerative) can be used for statistical analysis of simulation output data.

The techniques marked in table 2 can be applied for conducting experiment design VV&T with the use of indicators such as: (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., $X = -\log_e(1 - U)$) [84].); (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 [53] 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) [97] appropriately addressed?; and (6) For comparison studies, are identical experimental conditions replicated correctly for each of the alternative operating policies compared?

4.8. DATA VV&T

Data VV&T involves input data model VV&T and deals with substantiating that all data used throughout the model development phases of the life cycle in figure 2 are accurate, complete, unbiased, and appropriate in their original and transformed forms. An input data model is the characterization of an input process (e.g., characterization of an arrival process by Poisson probability distribution). U.S. GAO [93] emphasizes the importance of input data model validation in credibility assessment of simulations.

In those cases where data cannot be collected, data values may be determined through calibration. *Calibration* is an iterative process in which a probabilistic characterization for an input variable or a fixed value for a parameter is tried until the model is found to be sufficiently valid.

Assertion checking, audit, consistency checking, data flow analysis, desk checking, face validation, inspections, reviews, statistical techniques, and walkthroughs can be applied for conducting data VV&T with the use of indicators such as: (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 dependency is cited as a common error by Schmeiser [84].); and (6) Are all data up-to-date?

4.9. EXPERIMENTAL MODEL VV&T

Experimental Model VV&T deals with substantiating that the experimental model has sufficient accuracy in representing the system as defined under the study objectives. All of the 45 techniques listed in table 2 can be applied for conducting model VV&T. The applicability of the 45 techniques depends upon the following cases where the system being modeled is: (1) completely observable – all data required for model VV&T can be collected from the system, (2) partially observable – some required data can be collected, or (3) nonexistent or completely unobservable. The statistical techniques in table 1 are applicable only for case 1.

4.10. PRESENTATION VV&T

Presentation VV&T deals with justifying that the simulation results are interpreted, documented, and communicated with sufficient accuracy.

Since all simulation models are descriptive, simulation results must be interpreted. (A descriptive model describes the behavior of a system without any value judgment on the "goodness" or "badness" of such behavior [28].) 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" one. 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. 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.

Gass [38] points out that "we do not know of any model assessment or modeling project review that indicated satisfaction with the available documentation." Nance [63] advocates the use of standards in simulation documentation. The documentation problem should be attributed to the lack of automated support for documentation generation integrated with model development continuously throughout the entire life cycle. The model development environment [6, 11] provides such computer-aided assistance for documenting a simulation study with respect to the phases, processes, and credibility assessment stages of the life cycle in figure 2.

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. 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.

Audit, consistency checking, desk checking, face validation, inspections, reviews, structural analysis, and walkthroughs can be applied for conducting presentation VV&T.

5. Concluding remarks and research directions

The life cycle application of VV&T is extremely important for successful completion of complex and large-scale simulation studies. This point must be clearly understood by the sponsor of the simulation study and the organization conducting the simulation study. The sponsor must furnish funds under the contractual agreement and require the contractor to apply VV&T throughout the entire life cycle.

Assessing credibility throughout the life cycle of a simulation study is an onerous task. Applying the VV&T techniques throughout the life cycle is time consuming and costly. In practice, under time pressure to complete a simulation study, the VV&T and documentation are sacrificed first. Computer-aided assistance for the VV&T is required to alleviate these problems. More research is needed to bring automation to the application of the VV&T techniques.

Integration of VV&T with model development is crucial. This integration is best achieved within a computer-aided simulation software engineering environment [6, 11]. More research is needed for this integration.

How much to test or when to stop testing depends on the study objectives. The testing should continue until we achieve sufficient confidence in credibility and acceptability of simulation results. The sufficiency of the confidence is dictated by the study objectives.

Establishing a simulation quality assurance (SQA) program within the organization conducting the simulation study is extremely important for successful credibility assessment. The SQA management structure goes beyond VV&T and is also responsible for assessing other model quality characteristics such as maintainability, reusability, and usability (human-computer interface). The management of the SQA program and the management of the simulation project must be independent of each other and neither should be able to overrule the other [82].

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 unifying approach based on the use of indicators measuring qualitative as well as quantitative aspects of a simulation study should be developed.

6. Glossary

Assertion Checking: A technique for examining what *is* happening against what the modeler *assumes* is happening so as to guard model execution against potential errors.

Audit: A technique for investigating how adequately the simulation study is conducted with respect to established practices, standards, and guidelines and for establishing traceability within the simulation study.

Black-Box Testing: A technique for assessing the accuracy of model input–output transformation. It is applied by feeding inputs (test data) to the model and evaluating the corresponding outputs. The concern is how accurtately the model transforms a given set of input data into a set of output data.

Bottom-Up Testing: A technique, used in conjunction with bottom-up model development strategy, in which testing starts with the submodels at the base level (i.e., the ones that are not decomposed further) and culminates with the submodels at the highest level. As each submodel is completed, it is thoroughly tested. When submodels belonging to the same parent have been developed and tested, the submodels are integrated and integration testing is performed. This process is repeated in a bottom-up manner until the whole model has been integrated and tested.

Boundary Analysis: A technique for assessing model accuracy by using test cases on the boundaries of input equivalence classes which are classes of input data that cause the model to function the same way.

Calibration: An iterative process in which a probabilistic characterization for an input variable or a fixed value for a parameter is tried until the model is found to be sufficiently valid.

Cause–Effect Graphing: A technique conducted by addressing the question of "what causes what in the model representation?" It is performed by first identifying causes and effects in the system being modeled and by examining if they are accurately reflected in the model specification. Then, the semantics are expressed in a cause–effect graph. A decision table is created by tracing back through the graph to determine combinations of causes which result in each effect. The decision table is then converted into test cases with which the model is tested.

Communicative Model: A model representation which can be communicated to other humans and can be judged or compared against the system and the study objectives by more than one human [64].

Communicative Model VV&T: Confirming the adequacy of the communicative model to provide an acceptable level of agreement for the domain of intended application.

Conceptual Model: The model which is formulated in the mind of the modeler [64].

Consistency Checking: A technique which deals with substantiating that: (a) the model representation does not contain contradictions, (b) the cosmetic style with which language elements (e.g., naming conventions, use of upper, lower, and mixed case, etc.) are applied is used consistently, and (c) the data elements are manipulated uniformly (e.g., data assignment to variables, data use within computations, data passing among submodels, data representation and use during model input and output).

Data Flow Analysis: A technique for assessing model accuracy with respect to the use of model variables. A data flow graph is constructed to aid in the data flow analysis. The nodes of the graph represent statements and corresponding variables. The edges represent control flow.

Data VV&T: Substantiating that each input data model has sufficient accuracy in representing the simulation model's input process (e.g., assessing the accuracy of characterizing an arrival process by Poisson probability distribution) and that all data used throughout the model development phases are accurate, complete, unbiased, and appropriate in their original and transformed forms.

Debugging: An iterative process the purpose of which is to uncover errors or misconceptions that cause the model's failure and to define and carry out the model changes that correct the errors.

Design of Experiments: The process of formulating a plan to gather the desired information at minimal cost and to enable the analyst to draw valid inferences [87].

Desk Checking: A technique for thoroughly examining one's work to ensure correctness, completeness, consistency, and unambiguity.

Domain of Applicability: 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 [83].

Domain of Intended Application: The prescribed conditions for which the model is intended to match the system under study [83].

Execution Monitoring: A technique for revealing errors by examining low-level information about activities and events which take place during model execution. It requires model instrumentation.

Execution Profiling: A technique for revealing errors by examining high-level information (profiles) about activities and events which take place during model execution. It requires model instrumentation.

Execution Tracing: A technique for revealing errors by "watching" the line-by-line execution of a simulation model. It requires model instrumentation.

Experiment Design VV&T: Substantiating that the experiments are designed and implemented with sufficient accuracy.

Experimental Model: The programmed model incorporating an executable description of an experiment design.

Face Validation: A technique in which the project team members, potential users of the model, people knowledgeable about the system under study, based on their estimates and intuition, subjectively compare model and system behaviors to judge whether the model and its results are reasonable.

Field Testing: A technique in which the model is placed in an operational situation for the purpose of collecting as much information as possible for model validation. It is especially useful for validating models of military combat systems.

Formulated Problem VV&T: Substantiating 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.

Functional Testing: See Black-Box Testing.

Graph-Based Analysis: A ~~technique for assessing model credibility based on the~~ use of a graphical representation of the model (e.g., data flow graph, control flow graph).

Graphical Comparisons: A technique by which the graphs of values of model variables over time are compared with the graphs of values of system variables to investigate characteristics such as similarities in periodicities, skewness, number and location of inflection points, logarithmic rise and linearity, phase shift, trend lines, and exponential growth constants.

Indicator: An indirect measure of a concept which is decomposed into other indicators until the ones at the base level (i.e., the ones that are not decomposed further) are directly measurable.

Induction: A formal proof of correctness technique for estimating the validity of the whole set of observations based on the validity of a subset of observations as evidence.

Inductive Assertions: A technique for assessing model correctness based on an approach that is very close to formal proof of model correctness. It is conducted in three steps: (1) input-to-output relations for all model variables are identified; (2) these relations are converted into assertion statements and are placed along the model execution paths in such a way as to divide the model into a finite number of "assertion-bound" paths, i.e., an assertion statement lies at the beginning and end of each model execution path; and (3) verification is achieved by proving that for each path: if the assertion at the beginning of the path is true, and all statements along the path are executed, then the assertion at the end of the path is true. If all paths plus model termination can be proved, by induction, the model is proved to be correct.

Inference: A formal proof of correctness technique for deriving logical conclusions from the premises given.

Input Data Model VV&T: Substantiating that the input data model has sufficient accuracy in representing the simulation model's input process (e.g., assessing the accuracy of characterizing an arrival process by Poisson probability distribution).

Inspections: A technique conducted by a team of four to six members with the objective of finding and documenting faults. It consists of five distinct phases: overview, preparation, inspection, rework, and follow-up.

Lambda Calculus: A technique for transforming the model into formal expressions by which mathematical proof of correctness techniques can be applied for the purpose of VV&T.

Level of Agreement: The required correspondence between the model and the system, consistent with the domain of intended application and the study objectives [83].

Logical Deduction: A formal proof of correctness technique for reasoning in which a conclusion follows necessarily from the premises given.

Model: Representation and abstraction of anything such as a system, concept, problem, or phenomena.

Model Builder's Risk: The probability of committing Type I error.

Model Certification: Confirmation (usually by a third party) that a simulation model, within its domain of applicability, can produce results which are sufficiently credible with respect to the study objectives.

Model Instrumentation: The insertion of additional code (probes) into the executable model for the purpose of collecting information about model behavior during execution. Probe locations are determined manually or automatically based on static analysis of model structure. Automated instrumentation is accomplished by a preprocessor which analyzes the model static structure (usually via graph-based analysis) and inserts probes at appropriate places.

Model Qualification: Justifying that all assumptions underlying the conceptual model are appropriate and the conceptual model provides an adequate representation of the system under study with respect to the study objectives.

Model Range of Accuracy: Simultaneous confidence intervals for the differences between the means of corresponding model and system output variables obtained by running the model with the "same" input data that drives the real system.

Model Testing: Demonstrating that inaccuracies exist or revealing the existence of errors in the model. In model testing, we subject the model to test data or test cases to see if it functions properly. "Test failed" implies the failure of the model, not the test.

Model User's Risk: The probability of committing Type II error.

Model Validation: Substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives. Model validation deals with building the *right* model. It is conducted by running the model under the "same" input conditions that drive the system and by comparing model behavior with the system behavior.

Model Verification: Substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy. Model verification deals with building the model *right*. The accuracy of transforming a problem formulation into a model specification or the accuracy of converting a model representation in micro flowchart into an executable computer program is evaluated in model verification.

Model VV&T: Model testing is conducted to perform model validation and verification. Some tests are devised to evaluate the behavioral accuracy (i.e., validity) of the model, and some tests are intended to judge the accuracy of model transformation from one form into another (verification). Therefore, we commonly refer to the whole process as model VV&T.

Partition Analysis: A technique for testing the model with the test data generated by analyzing the model's functional representatives (partitions). It is accomplished by: (1) decomposing both model specification and implementation into functional representatives (partitions), (2) comparing the elements and prescribed functionality of each partition specification with the elements and actual functionality of corresponding partition implementation, (3) deriving test data to extensively test the functional behavior of each partition, and (4) testing the model by using the generated test data.

Path Analysis: A technique for assessing model correctness on the basis of complete testing of all model control paths. It is performed in three steps: (1) the model control structure is determined and represented in a control flow diagram, (2) test data is generated to cause selected model logical paths to be executed, and (3) by using the generated test data, the model is forced to proceed through each path in its execution structure, thereby providing comprehensive testing.

Predicate Calculus: A technique that provides rules for manipulating predicates. A predicate is a combination of simple relations which will either be true or false. The model can be defined in terms of predicates and manipulated using the rules of the predicate calculus for the purpose of VV&T.

Predicate Transformation: A technique that provides a basis for verifying model correctness by formally defining the semantics of the model with a mapping which transforms model output states to all possible model input states. This representation provides the basis for proving model correctness.

Predictive Validation: A technique by which model validation is conducted using past system 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.

Presentation VV&T: Substantiating that the simulation results are interpreted, documented, and communicated with sufficient accuracy.

Programmed Model: A model representation that admits execution by a computer to produce results [64].

Programmed Model VV&T: Substantiating that the programmed model possesses sufficient accuracy in representing the system under study.

Proof of Correctness: A technique for expressing the model in a precise notation and then mathematically proving that the executed model terminates and it satisfies the requirements specification with sufficient accuracy.

Regression Testing: A technique for substantiating that correcting errors and/or making changes in the model do not create other errors and adverse side-effects. It is usually accomplished by retesting the modified model with the previous test data sets used.

Reviews: A technique conducted by a team of experts and managers with the objective of finding and documenting faults.

Self-Driven Simulation Model: A model driven by input values obtained via sampling from probability distributions using random numbers.

Semantic Analysis: A technique by which the simulation programming language compiler generates a wealth of information to help the modeler determine if the true intent is accurately translated into the executable code.

Sensitivity Analysis: A technique for systematically changing the values of model input variables and parameters over some range of interest and observing the effect upon model behavior [87]. Unexpected effects may reveal invalidity.

Simulation: The process of constructing a model of a system which contains a problem and conducting experiments with the model on a computer for a specific purpose of experimentation to solve the problem.

Simulation Quality Assurance: Refers to the management structure responsible for planning, preparing test cases, and administering VV&T activities throughout the life cycle of a simulation study to assure sufficient credibility of simulation study results.

Stress Testing: A technique for assessing model validity under extreme workload conditions. This is usually accomplished by increasing the congestion in the model.

Structural Analysis: A technique for examining the model structure and determining if it adheres to structured principles. It is conducted by constructing a control flow graph of the model structure and examining the graph for anomalies, such as multiple entry and exit points, excessive levels of nesting within a structure, and questionable practices such as the use of unconditional branches (i.e., GOTOs).

Submodel Testing: 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 to judge submodel validity. 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.

Symbolic Debugging: A technique which allows the modeler to locate errors and check numerous circumstances which lead up to the errors by employing a debugging tool that allows the modeler to manipulate model execution while viewing the model at the source code level. By setting "breakpoints", the modeler can interact with the entire model one step at a time, at predetermined locations, or under specified conditions.

Symbolic Execution: A technique for assessing model accuracy by executing the model using symbolic values rather than actual data values for input. It is performed by feeding symbolic inputs into the (sub)model and producing expressions for the output which are derived from the transformation of the symbolic data along model execution paths.

Syntax Analysis: A technique carried on by the simulation programming language compiler to assure that the mechanics of the language are applied correctly.

System and Objectives Definition VV&T: Substantiating that the system characteristics are identified and the study objectives are explicitly defined with sufficient accuracy.

Top-Down Testing: A technique, used in conjunction with top-down model development strategy, in which testing starts with the submodels at the highest level and culminates with the submodels at the base level (i.e., the ones that are not decomposed further). As each submodel is completed, it is thoroughly tested. When submodels belonging to the same parent have been developed and tested, the submodels are integrated and integration testing is performed. This process is repeated in a top-down manner until the whole model has been integrated and tested.

Trace-Driven Simulation Model: A model driven by input sequences extracted from trace data obtained through measurement of the real system.

Turing Test: A technique in which people with expert knowledge about the system under study 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.

Type I Error: The error of rejecting the model credibility when in fact the model is sufficiently credible. Occurrence of Type I error may increase the cost of model development or may cause the simulation study to end unsuccessfully.

Type II Error: The error of accepting the model credibility when in fact the model is *not* sufficiently credible. Consequences of committing Type II error can be catastrophic.

Type III Error: The error of solving the wrong problem. Once the Type III error is committed, regardless of how well the problem is solved, the simulation study will either end unsuccessfully or with the Type II error.

Visualization: Displaying graphical images of internal and external dynamic behavior of a simulation model during execution.

Walkthroughs: A technique conducted by a team composed of a coordinator, model developer, and three to six other members with the objective of discovering and documenting faults.

White-Box Testing: A technique which employs data flow and control flow diagrams to assess the accuracy of internal model structure by examining model elements such as internal logic, internal data representations, submodel interfaces, and model execution paths. White-box testing is quite effective for detecting redundant code, faulty model structure, and special case errors.

References

[1] A.F. Ackerman, P.J. Fowler and R.G. Ebenau, Software inspections and the industrial production of software, in: *Software Validation: Inspection, Testing, Verification, Alternatives*, Proc. Symp. on Software Validation, Darmstadt, Germany, ed. H.-L. Hausen (1983) pp. 13–40.

[2] W.R. Adrion, M.A. Branstad and J.C. Cherniavsky, Validation, verification, and testing of computer software, Comp. Surveys 14(1982)159–192.

[3] D.J. Aigner, A note on verification of computer simulation models, Manag. Sci. 18(1972)615–619.

[4] F.E. Allen and J. Cocke, A program data flow analysis procedure, Commun. ACM 19(1976) 137–147.

[5] R.C. Backhouse, *Program Construction and Verification* (Prentice–Hall, London, 1986).

[6] O. Balci, Requirements for model development environments, Comp. Oper. Res. 13(1986)53–67.

[7] O. Balci, The implementation of four conceptual frameworks for simulation modeling in high-level languages, in: *Proc. 1988 Winter Simulation Conf.*, ed. M.A. Abrams, P.L. Haigh, and J.C. Comfort (IEEE, Piscataway, NJ, 1988) pp. 287–295.

[8] O. Balci, Guidelines for successful simulation studies, in: *Proc. 1990 Winter Simulation Conf.*, ed. O. Balci, R.P. Sadowski, and R.E. Nance (IEEE, Piscataway, NJ, 1990) pp. 25–32.

[9] O. Balci, Principles of simulation model validation, verification, and testing, Technical Report TR-94-24 Department of Computer Science. Virginia Tech, Blacksburg, VA (1994).

[10] O. Balci and R.E. Nance, Formulated problem verification as an explicit requirement of model credibility, Simulation 45(1985)76–86.

[11] O. Balci and R.E. Nance, Simulation model development environments: A research prototype, J. Oper. Res. Soc. 38(1987)753–763.

[12] O. Balci and R.G. Sargent, A methodology for cost-risk analysis in the statistical validation of simulation models, Commun. ACM 24(1981)190–197.

[13] O. Balci and R.G. Sargent, Some examples of simulation model validation using hypothesis testing, in: *Proc. 1982 Winter Simulation Conf.*, ed. H.J. Highland, Y.W. Chao and O.S. Madrigal (IEEE, Piscataway, NJ, 1982) pp. 620–629.

[14] O. Balci and R.G. Sargent, Validation of multivariate response models using Hotelling's two-sample T^2 test, Simulation 39(1982)185–192.

[15] O. Balci and R.G. Sargent, Validation of multivariate response trace-driven simulation models, in: *Performance '83*, ed. A. K. Agrawala and S.K. Tripathi (North-Holland, Amsterdam, 1983) 309–323.

[16] O. Balci and R.G. Sargent, Validation of simulation models via simultaneous confidence intervals, Amer. J. Math. Manag. Sci. 4(1984)375–406.

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

[18] J. Banks, D. Gerstein and S.P. Searles, Modeling processes, validation, and verification of complex simulations: A survey, in: *Methodology and Validation*, ed. O. Balci (SCS, San Diego, CA, 1987) pp. 13–18.

[19] H.P. Barendregt, *The Lambda Calculus: Its Syntax and Semantics* (North-Holland, New York, 1981).

[20] T. Chusho, Test data selection and quality estimation based on the concept of essential branches for path testing, IEEE Trans. Software Eng. SE-13(1987)509–517.

[21] K.J. Cohen and R.M. Cyert, Computer models in dynamic economics, Quarterly J. Econ. 75(1961) 112–127.

[22] M.J. Damborg and L.F. Fuller, Model validation using time and frequency domain error measures, ERDA Report 76-152, NTIS, Springfield, VA (1976).

[23] M.S. Deutsch, *Software Verification and Validation: Realistic Project Approaches* (Prentice–Hall, Englewood Cliffs, NJ, 1982).

[24] E.W. Dijkstra, Guarded commands, non-determinacy and a calculus for the derivation of programs, Commun. ACM 18(1975)453–457.

[25] L.K. Dillon, Using symbolic execution for verification of Ada tasking programs, ACM Trans. Progr. Languages Syst. 12(1990)643–669.

[26] J.H. Dobbins, Inspections as an up-front quality technique, in: *Handbook of Software Quality Assurance*, ed. G.G. Schulmeyer and J.I. McManus (Van Nostrand–Reinhold, New York, NY, 1987) pp. 137–177.

[27] R.H. Dunn, The quest for software reliability, in: *Handbook of Software Quality Assurance*, ed. G.G. Schulmeyer and J.I. McManus (Van Nostrand–Reinhold, New York, NY, 1987) pp. 342–384.

[28] S.E. Elmaghraby, The role of modeling in IE design, Ind. Eng. 19(1968)292–305.

[29] J.R. Emshoff and R.L. Sisson, *Design and Use of Computer Simulation Models* (MacMillan, New York, NY, 1970).

[30] R.E. Fairley, An experimental program-testing facility, IEEE Trans. Software Eng. SE-1(1975) 350–357.

[31] R.E. Fairley, Dynamic testing of simulation software, in: *Proc. 1976 Summer Computer Simulation Conf.*, Washington, DC (Simulation Councils, La Jolla, CA, 1976) pp. 708–710.

[32] G.S. Fishman, *Principles of Discrete Event Simulation* (Wiley–Interscience, New York, NY, 1978).

[33] G.S. Fishman and P.J. Kiviat, The analysis of simulation generated time series, Manag. Sci. 13(1967) 525–557.

[34] J.W. Forrester, *Industrial Dynamics* (MIT Press, Cambridge, MA, 1961).

[35] A.V. Gafarian and J.E. Walsh, Statistical approach for validating simulation models by comparison with operational systems, in: *Proc. 4th Int. Conf. on Operations Research* (Wiley, New York, NY, 1969) pp. 702–705.

[36] A.R. Gallant, T.M. Gerig and J.W. Evans, Time series realizations obtained according to an experimental design, J. Amer. Statist. Assoc. 69(1974)639–645.

[37] M. Garratt, Statistical validation of simulation models, in: *Proc. 1974 Summer Computer Simulation Conf*, Houston, TX (Simulation Councils, La Jolla, CA, 1974) pp. 915–926.

[38] S.I. Gass, Decision-aiding models: Validation, assessment, and related issues for policy analysis, Oper. Res. 31(1983)603–631.

[39] C.F. Hermann, Validation problems in games and simulations with special reference to models of international politics, Behav. Sci. 12(1967)216–231.

[40] W. Hetzel, *The Complete Guide to Software Testing* (QED Information Sciences, Wellesley, MA, 1984).

[41] C.P. Hollocker, The standardization of software reviews and audits, in: *Handbook of Software Quality Assurance*, ed. G.G. Schulmeyer and J.I. McManus (Van Nostrand-Reinhold, New York, NY, 1987) pp. 211–266.

[42] W.E. Howden, Reliability of the path analysis testing strategy, IEEE Trans. Software Eng. SE-2(1976)208–214.

[43] W.E. Howden, Functional program testing, IEEE Trans. Software Eng. SE-6(1980)162–169.

[44] P. Howrey and H.H. Kelejian, Simulation versus analytical solutions, in: *The Design of Computer Simulation Experiments*, ed. T.H. Naylor (Duke University Press, Durham, NC, 1969) pp. 207–231.

[45] A.W. Hunt, Statistical evaluation and verification of digital simulation models through spectral analysis, Ph.D. Dissertation, University of Texas at Austin, Austin, TX (1970).

[46] S.H. Jacobson and E. Yücesan, On the NP-completeness of verifying structural properties of discrete event simulation models, Technical Report, Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA (1993).

[47] S. Khanna, Logic programming for software verification and testing, Comp. J. 34(1991)350–357.

[48] N.A. Kheir and W.M. Holmes, On validating simulation models of missile systems, Simulation 30(1978)117–128.

[49] J.C. King, Symbolic execution and program testing, Commun. ACM 19(1976)385–394.

[50] J.P.C. Kleijnen, *Statistical Techniques in Simulation*, Vol. 2 (Marcel Dekker, New York, NY, 1975).

[51] P.L. Knepell and D.C. Arangno, *Simulation Validation: A Confidence Assessment Methodology*, Monograph 3512-04 (IEEE Computer Society Press, Los Alamitos, CA, 1993).

[52] J.C. Knight and E.A. Myers, An improved inspection technique, Commun. ACM 36(1993)51–61.

[53] A.M. Law, Statistical analysis of simulation output data, Oper. Res. 31(1983)983–1029.

[54] A.M. Law and W.D. Kelton, *Simulation Modeling and Analysis*, 2nd ed. (McGraw–Hill, New York, NY, 1991).

[55] Z. Manna, S. Ness and J. Vuillemin, Inductive methods for proving properties of programs, Commun. ACM 16(1973)491-502.

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

[57] D.K. Miller, Validation of computer simulations in the social sciences, in: *Proc. 6th Annual Conf. on Modeling and Simulation* (Pittsburg, PA, 1975) pp. 743–746.

[58] D.R. Miller, Model validation through sensitivity analysis, in: *Proc. 1974 Summer Computer Simulation Conf.*, Houston, TX (Simulation Councils, La Jolla, CA, 1974) pp. 911–914.

[59] D.R. Miller, Sensitivity analysis and validation of simulation models, J. Theor. Biol. 48(1974)345–360.

[60] R.L. Moose and R.E. Nance, The design and development of an analyzer for discrete event model specifications, in: *Impacts of Recent Computer Advances on Operations Research*, ed. R. Sharda, B.L. Golden, E. Wasil, O. Balci and W. Stewart (Elsevier, New York, NY, 1989) pp. 407–421.

[61] G.J. Myers, A controlled experiment in program testing and code walkthroughs/inspections, Commun. ACM 21(1978)760–768.

[62] G.J. Myers, *The Art of Software Testing* (Wiley, New York, NY, 1979).

[63] R.E. Nance, The feasibility of and methodology for developing federal documentation standards for simulation models: Final report to the National Bureau of Standards, Department of Computer Science, VPI&SU, Blacksburg, VA (1977).

[64] R.E. Nance, Model representation in discrete event simulation: The conical methodology, Technical Report CS81003-R, Department of Computer Science, VPI&SU, Blacksburg, VA (1981).

[65] R.E. Nance, The conical methodology: A framework for simulation model development, in: *Methodology and Validation*, ed. O. Balci, (SCS, San Diego, CA, 1987) pp. 38–43.

[66] R.E. Nance, Conical methodology: An evolutionary convergence of systems and software engineering, Ann. Oper. Res. 53(1994), this volume.

[67] R.E. Nance and C.M. Overstreet, Diagnostic assistance using digraph representations of discrete event simulation model specifications. Trans. SCS 4(1987)33–57.

[68] T.H. Naylor and J.M. Finger, Verification of computer simulation models, Manag. Sci. 14(1967) B92–B101.

[69] M.A. Ould and C. Unwin, *Testing in Software Development* (Cambridge University Press, Cambridge, 1986).

[70] C.M. Overstreet and R.E. Nance, A specification language to assist in analysis of discrete event simulation models, Commun. ACM 28(1985)190–201.

[71] T.I. Ören, Concepts and criteria to assess acceptability of simulation studies: A frame of reference, Commun. ACM 24(1981)180–189.

[72] T.I. Ören, Artificial intelligence in quality assurance of simulation studies, in: *Modelling and Simulation Methodology in the Artificial Intelligence Era*, ed. M.S. Elzas, T.I. Ören and B.P. Zeigler (North-Holland, Amsterdam, 1986) pp. 267–278.

[73] T.I. Ören, Quality assurance paradigms for artificial intelligence in modelling and simulation, Simulation 48(1987)149–151.

[74] R.J. Paul, Visual simulation: Seeing is believing?, in: *Impaçts of Recent ComputerAdvances on Operations Research*, ed. R. Sharda, B.L. Golden, E. Wasil, O. Balci, and W. Stewart (Elsevier, New York, NY, 1989) pp. 422–432.

[75] R.E. Prather and J.P. Myers, Jr., The path prefix software testing strategy, IEEE Trans. Software Eng. SE-13(1987)761–766.

[76] C.V. Ramamoorthy, S.F. Ho and W.T. Chen, On the automated generation of program test data, IEEE Trans. Software Eng. SE-2(1976)293–300.

[77] C. Reynolds and R.T. Yeh, Induction as the basis for program verification, IEEE Trans. Software Eng. SE-2(1976)244–252.

[78] D.J. Richardson and L.A. Clarke, Partition analysis: A method combining testing and verification, IEEE Trans. Software Eng. SE-11(1985)1477–1490.

[79] J.R. Rowland and W.M. Holmes, Simulation validation with sparse random data, Comp. Elect. Eng. 5(1978)37–49.

[80] R.G. Sargent, Validation and verification of simulation models, in: *Proc. 1992 Winter Simulation Conf.*, ed. J.J. Swain, D. Goldsman, R.C. Crain, and J.R. Wilson (IEEE, Piscataway, NJ, 1992) pp. 104–114.

[81] E. Satterthwaite, Debugging tools for high level languages, Software – Practice and Experience 2(1972)197–217.

[82] S.R. Schach, *Software Engineering*, 2nd ed. (Irwin, Homewood, IL 1993).

[83] S. Schlesinger et al., Terminology for model credibility, Simulation 32(1979)103–104.

[84] B. Schmeiser, Random variate generation, in: *Proc. 1981 Winter Simulation Conf.*, ed. T.I. Ören, C.M. Delfosse and C.M. Shub (IEEE, Piscataway, NJ, 1981) pp. 227–242.

[85] T.J. Schriber, *Simulation Using GPSS* (Wiley, New York, NY, 1974).

[86] L.W. Schruben, Establishing the credibility of simulations, Simulation 34(1980)101–105.

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

[88] L.G. Stucki, New directions in automated tools for improving software quality, in: *Current Trends in Programming Methodology*, Vol. 2, ed. R. Yeh (Prentice–Hall, Englewood Cliffs, NJ, 1977) pp. 80–111.

[89] T.J. Teorey, Validation criteria for computer system simulations, Simuletter 6(1975)9–20.

[90] H. Theil, *Economic Forecasts and Policy* (North-Holland, Amsterdam, 1961).

[91] A.M. Turing, Computing machinery and intelligence, in: *Computers and Thought*, ed. E.A. Feigenbaum and J. Feldman (McGraw–Hill, New York, NY, 1963) pp. 11–15.

[92] T.P. Tytula, A method for validating missile system simulation models, Technical Report E-78-11, U.S. Army Missile R&D Command, Redstone Arsenal, AL (1978).

[93] U.S. GAO, *DOD Simulations: Improved Assessment Procedures Would Increase the Credibility of Results*, U.S. General Accounting Office GAO/PEMD-88-3, Washington, DC (1987).

[94] R.L. Van Horn, Validation of simulation results, Manag. Sci. 17(1971)247–258.

[95] D. Watts, Time series analysis, in: *The Design of Computer Simulation Experiments*, ed. T.H. Naylor (Duke University Press, Durham, NC, 1969) pp. 165–179.

[96] R.B. Whitner and O. Balci, Guidelines for selecting and using simulation model verification techniques, in: *Proc. 1989 Winter Simulation Conf.*, ed. E.A. MacNair, K.J. Musselman and P. Heidelberger (IEEE, Piscataway, NJ, 1989) pp. 559–568.

[97] J.R. Wilson and A.A.B. Pritsker, A survey of research on the simulation startup problem, Simulation 31(1978)55–58.

[98] R.N. Woolley and M. Pidd, Problem structuring – A literature review, J. Oper. Res. Soc. 32(1981) 197–206.

[99] R.D. Wright, Validating dynamic models: An evaluation of tests of predictive power, in: *Proc. 1972 Summer Computer Simulation Conf.*, San Diego, CA, (Simulation Councils, La Jolla, CA, 1972) pp. 1286–1296.

[100] R.T. Yeh, Verification of programs by predicate transformation, in: *Current Trends in Programming Methodology*, Vol. 2, ed. R. Yeh (Prentice-Hall, Englewood Cliffs, NJ, 1977) pp. 228–247.

[101] E. Yourdon, *Structured Walkthroughs*, 3rd ed. (Yourdon Press, New York, NY, 1985).

[102] E. Yücesan and S.H. Jacobson, Building correct simulation models is difficult, in: *Proc. 1992 Winter Simulation Conf.*, ed. J.J. Swain, D. Goldsman, R.C. Crain, and J.R. Wilson (IEEE, Piscataway, NJ, 1992) pp. 783–790.