1. INTRODUCTION

The U.S. General Accounting Office (GAO) prepared three reports due to concerns about the credibility of federally funded computerized models. The first report, presented to the U.S. congress in 1976, identified the need for improved management and model development activities [U.S. GAO 1976]. The second one described the modeling process and presented a very high-level approach to model evaluation [U.S. GAO 1979]. The third one, presented to the National Security Subcommittee of the House of Representatives, described improved assessment procedures for the credibility of Department of Defense simulation studies [U.S. GAO 1987]. All three reports described simulation modeling problems that have emerged due to the lack of understanding the principles of simulation model validation, verification, and testing (VV&T).

According to the Webster's dictionary, a principle is defined as "**1.** an accepted or professed rule of action or conduct. **2.** a fundamental, primary, or general law or truth from which others are derived. **3.** a fundamental doctrine or tenet; a distinctive ruling opinion." All three definitions above apply to the way the term "principle" is used in this paper.

Principles are important to understand the foundations of simulation model VV&T. The principles help the researchers, practitioners and managers better comprehend what model VV&T is all about. They serve to provide the underpinnings for the 45 VV&T techniques, described by Balci [1994], that can be used throughout the life cycle of a simulation study. Understanding and applying these principles is crucially important for the success of a simulation study.

Model Validation is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives. In other words, in model validation, we substantiate that the input-output transformation of the model has sufficient accuracy in representing the input-output transformation of the system. 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 corresponding system behavior.

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). We commonly refer to the whole process as *model VV&T* or *model testing*.

The purpose of this paper is to present discrete-event simulation model VV&T principles. Although the principles are presented for discrete-event simulation studies, many of them are also applicable for continuous and combined simulation studies and for other types of modeling studies. Section 2 describes the principles and Section 3 contains concluding remarks.

2. PRINCIPLES OF SIMULATION MODEL VV&T

The fifteen principles presented herein are established based on the experience described in the published literature and the author's experience during his research in simulation model VV&T since 1978. The principles are listed below in no particular order.

2.1 Principle 1: The VV&T must be conducted throughout the entire life cycle of a simulation study

The VV&T is not a phase or step in the life cycle of a simulation study, but a continuous activity throughout the entire life cycle presented in Figure 1 [Balci 1994, Nance 1994]. The life cycle is composed of ten phases (shaded oval symbols), ten processes (dashed arrows), and 13 VV&T stages (solid arrows; quality assurance of experimental model and credibility assessment of simulation results are not shown). Balci [1994] presents 45 VV&T techniques and describes how they can all be applied throughout the life cycle of a simulation study.

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.

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 [Hetzel 1984]. 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 13 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.

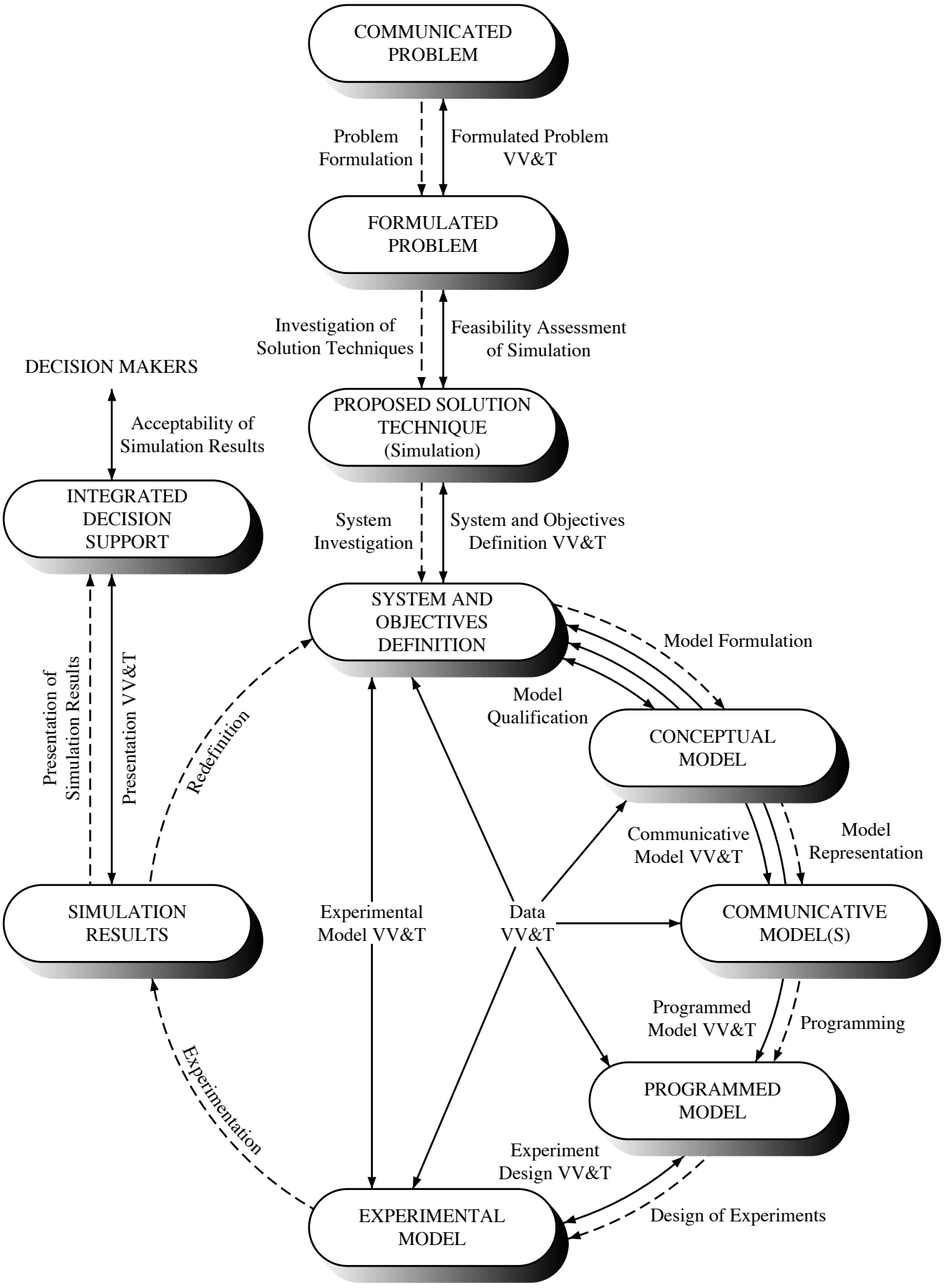

Figure 1. The Life Cycle of a Simulation Study

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 [1981, 1986, 1987] presents concepts, criteria, and paradigms which can be used in establishing an SQA program within an organization.

A simulation model goes through five levels of testing during its life cycle as described below:

- *Level 1: Private Testing*. Performed by the modeler in private with no documentation. Although informal, private testing is strongly encouraged prior to formal submodel testing [Beizer 1984].
- *Level 2: Submodel (Module) Testing*. Performed by the modeler to ensure the credibility of each submodel. Submodel testing treats each submodel as a stand-alone unit, with its own input and output variables, that can be tested without other submodels. Submodel testing should be planned by the simulation project managers and should be well documented so that another modeler can redo them if necessary.
- *Level 3: Integration Testing*. Planned, performed, and documented independently by the SQA group. Its objective is to substantiate that no inconsistencies in interfaces and communications between the submodels exist when the submodels are combined to form the model. It is assumed that each submodel has passed the submodel testing prior to integration testing.
- *Level 4: Model (Product) Testing*. Planned, performed, and documented independently by the SQA group. Its objective is to assess the validity of the overall model behavior.
- *Level 5: Acceptance Testing*. Planned, performed, and documented independently by the sponsor of the simulation study or the independent third party hired by the sponsor. Its objective is to establish the sufficient credibility of the simulation model so that its results can be accepted and used by the sponsor.

2.2 Principle 2: The outcome of simulation model VV&T should not be considered as a binary variable where the model is absolutely correct or absolutely incorrect

Since a model is an abstraction of a system, perfect representation is never expected. Shannon [1975] indicates that "it is not at all certain that it is ever theoretically possible to establish if we have an absolutely valid model; even if we could, few managers would be willing to pay the price."

The outcome of model VV&T should be considered as a degree of credibility on a scale from 0 to 100, where 0 represents absolutely incorrect and 100 represents absolutely correct. As depicted in Figure 2 [Shannon 1975; Sargent 1992], as the degree of model credibility increases, so will the model development cost. At the same time, the model utility will also increase, but most likely at a decreasing rate. The point of intersection of two curves changes from one model to another.

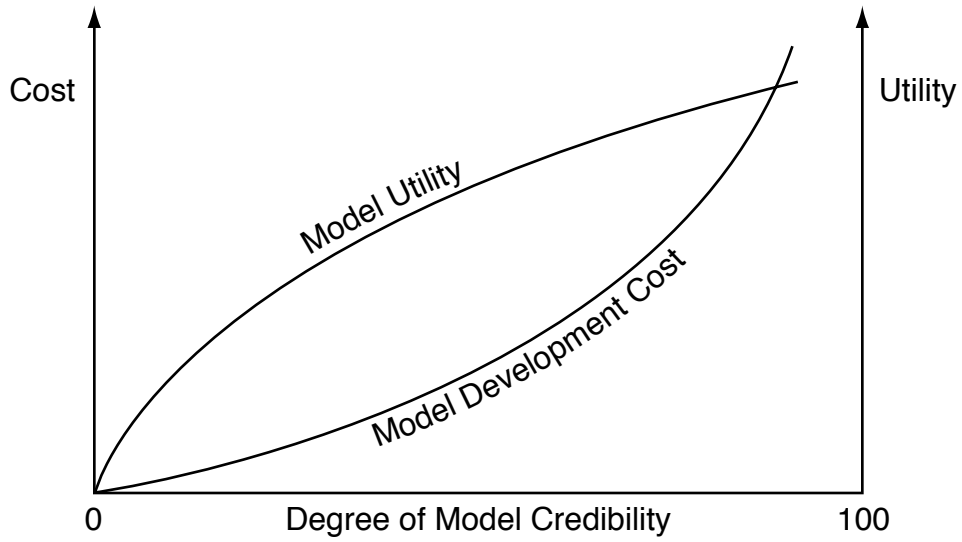

Figure 2. Model Credibility versus Cost and Utility

2.3 Principle 3: A simulation model is built with respect to the study objectives and its credibility is judged with respect to those objectives

The objectives of a simulation study are identified in the Formulated Problem phase, and explicitly and clearly specified in the System and Objectives Definition phase of the life cycle shown in Figure 1. Accurate specification of the study objectives is crucial for the success of a simulation study. The model is developed with respect to the study objectives within the circular portion of the life cycle in Figure 1.

The study objectives dictate how representative the model should be. Sometimes, 60% representation accuracy may be sufficient; sometimes, 95% accuracy may be required depending on the importance of the decisions that will be made based on the simulation results. Therefore, model credibility must be judged with respect to the study objectives. The adjective "sufficient" must be used in front of the terms such as model credibility, model validity, and model accuracy to indicate that the judgment is made with respect to the study objectives. It is more appropriate to say "the model is sufficiently valid" than saying "the model is valid." Here "sufficiently valid" implies that the validity is judged with respect to the study objectives and found to be sufficient.

2.4 Principle 4: Simulation model VV&T requires independence to prevent developer's bias

The model testing is meaningful when conducted in an independent manner by an unbiased person. The model developer with the most knowledge of the model may be the least independent when it comes to testing. The developers are often biased because they fear that negative testing results may be used for their performance appraisal. Similarly, the organization which is contracted to conduct the simulation study is also often biased because negative testing results can damage the credibility of the organization and may lead to the loss of future contracts.

The independence in model testing can be achieved in two ways: (1) establishing an SQA group within the organization conducting the simulation study, and (2) using an independent third party hired by the sponsor of the simulation study. The first one is required to achieve independence in level 3 and 4 testing within the organization as described in Section 2.1. The SQA group must be independent from the department in charge of conducting the simulation study and should report to the top management. The SQA group is responsible for planning, performing, and documenting all level 3 and 4 tests in an unbiased manner. It should be made clear to the simulation project team that the main thrust of testing is to detect and document faults; it is *not* performance appraisal of the project team. This point must be persuasively communicated to everyone involved so that full cooperation is achieved in discovering and documenting errors.

The second one is required to achieve independence in level 5 (acceptance) testing as described in Section 2.1. The requirements for acceptance testing must be specified in the contract by the sponsor. The test cases to be used must be well documented. Although the sponsor can perform the acceptance testing, it is recommended that an independent third party contracted by the sponsor be responsible for the testing. That way the organization conducting the simulation study cannot claim that the sponsor is biased.

2.5 Principle 5: Simulation model VV&T is difficult and requires creativity and insight

One must thoroughly understand the whole simulation model so as to design and implement effective tests and identify adequate test cases. Knowledge of the problem domain, expertise in the modeling methodology, and prior modeling and VV&T experience are required for successful testing.

However, it is not possible for one person to fully understand all aspects of a large and complex model especially if the model is a stochastic one containing hundreds of concurrent activities. The fundamental human limitation, called *Hrair Limit,* indicates that a human being cannot process more than 7 ± 2 entities simultaneously [Miller 1956]. Hence, testing a complex simulation model is a very difficult task and it requires creativity and insight.

The model developers are usually the most qualified to show the creativity and insight required for successful testing since they are intimately knowledgeable about the internals of a model. However, they are usually biased when it comes to model testing and they cannot be fully utilized. Therefore, the inability to use model developers effectively for testing increases the difficulty of testing.

False beliefs exist about testing as indicated by Hetzel [1984]: “testing is easy; anyone can do testing; no training or prior experience is required.” The difficulty of model VV&T must not be underestimated. The model testing must be well planned and administered by the SQA group.

2.6 Principle 6: Simulation model credibility can be claimed only for the prescribed conditions for which the model is tested

The accuracy of the input-output transformation of a simulation model is affected by the characteristics of the input conditions. The transformation that works for one set of input conditions may produce absurd output when conducted under another set of input conditions.

In the simulation of a traffic intersection, for example, a stationary simulation model can be built assuming constant arrival rate of vehicles during the evening rush hour and its credibility may be judged sufficient with respect to the evening rush hour input conditions. However, the simulation model will show invalid behavior when run under the input conditions of the same traffic intersection between 7:00 a.m. and 6:00 p.m. During this time period, the arrival rate of vehicles is not constant and a non-stationary simulation model is required. Hence, establishing sufficient model credibility for the evening rush hour conditions does not imply sufficient model credibility for input conditions during other times.

The prescribed conditions for which the model credibility has been established is called the *domain of applicability* of the experimental simulation model [Schlesinger *et al*. 1979]. Model credibility can be claimed only for the domain of applicability of the model.

2.7 Principle 7: Complete simulation model testing is not possible

Exhaustive (complete) testing requires testing the model under *all* possible inputs. Combinations of feasible values of model input variables can generate millions of logical paths in the model execution. Due to time and budgetary constraints, it is impossible to test the accuracy of millions of logical paths. Therefore, in model testing, the purpose is to increase our confidence in model credibility as much as dictated by the study objectives rather than trying to test the model completely.

How much to test or when to stop testing is dependent on the desired domain of applicability of the experimental model. The larger the domain the more the testing is required. The domain of applicability is determined with respect to the study objectives.

Hundreds of logical paths may need to be tested so as to substantiate model credibility under a set of prescribed conditions. Due to budgetary and time constraints we may not be able to test all logical paths. Test data or test cases are prepared to test the logical paths in a random manner. Test data can be generated by using: (1) random values, (2) deterministic values, (3) minimum values for all input variables, (4) maximum values for all input variables, (5) a mixture of minimum and maximum values for all input variables, (6) invalid values, and (7) simulated values.

In model testing using test data, it must be noted that the law of large numbers does simply not apply. The question is not how much test data is used, but what percentage of the valid input domain is covered by the test data. The higher the percentage of coverage the higher the confidence we can have in model credibility.

2.8 Principle 8: Simulation model VV&T must be planned and documented

Testing is not a phase or step in model development life cycle; it is a continuous activity *throughout* the entire life cycle. The tests should be identified, test data or cases should be prepared, tests should be scheduled, and the whole testing process should be documented.

Ad hoc or haphazard testing does not provide reasonable measurement of model accuracy. Hetzel [1984] points out that "such testing may even be harmful in leading us to a false sense of security." Careful planning is required for successful testing.

Planning and documenting model testing involves at least three groups of people: (1) sponsor of the simulation study, (2) SQA group of the organization conducting the simulation study, and (3) simulation project management. The sponsor is responsible for documenting the tests and specifying the test cases or data with which the acceptance testing will be performed. It is recommended that a plan for acceptance testing be made part of the legal contract. The SQA group is responsible for planning, performing, and documenting level 3 (integration) and level 4 (model) testing. The project management is responsible for planning and documenting level 2 (submodel) testing.

A *test plan* is a document describing what is selected for testing, test data base and code, test specifications, standards and conventions, test control, test configuration, test tools, and the results expected. An acceptance test plan is presented by Beizer [1984].

2.9 Principle 9: Type I, II, and III errors must be prevented

Three types of errors may be committed in conducting a simulation study as depicted in Figure 3 [Balci and Nance 1985]. *Type I Error* is committed when the simulation results are rejected when in fact they are sufficiently credible. *Type II Error* occurs when invalid simulation results are accepted as if they are sufficiently valid. *Type III Error* is the error of solving the wrong problem and committed when the formulated problem does not completely contain the actual problem.

Committing Type I Error unnecessarily increases the cost of model development. The consequences of Type II and Type III Errors can be catastrophic especially when critical decisions are made on the basis of simulation results. Type III Error implies the solution of the wrong problem and the simulation study results become irrelevant when it is committed.

The probability of committing Type I Error is called *Model Builder's Risk* and the probability of committing Type II Error is called *Model User's Risk* [Balci and Sargent 1981]. The VV&T activities must focus on minimizing these risks as much as possible. Balci and Sargent [1981] show how to quantify these risks when using hypothesis testing for the validation of a simulation model with two or more output variables.

Figure 3 illustrates the occurrence of three types of errors assuming that the simulation study

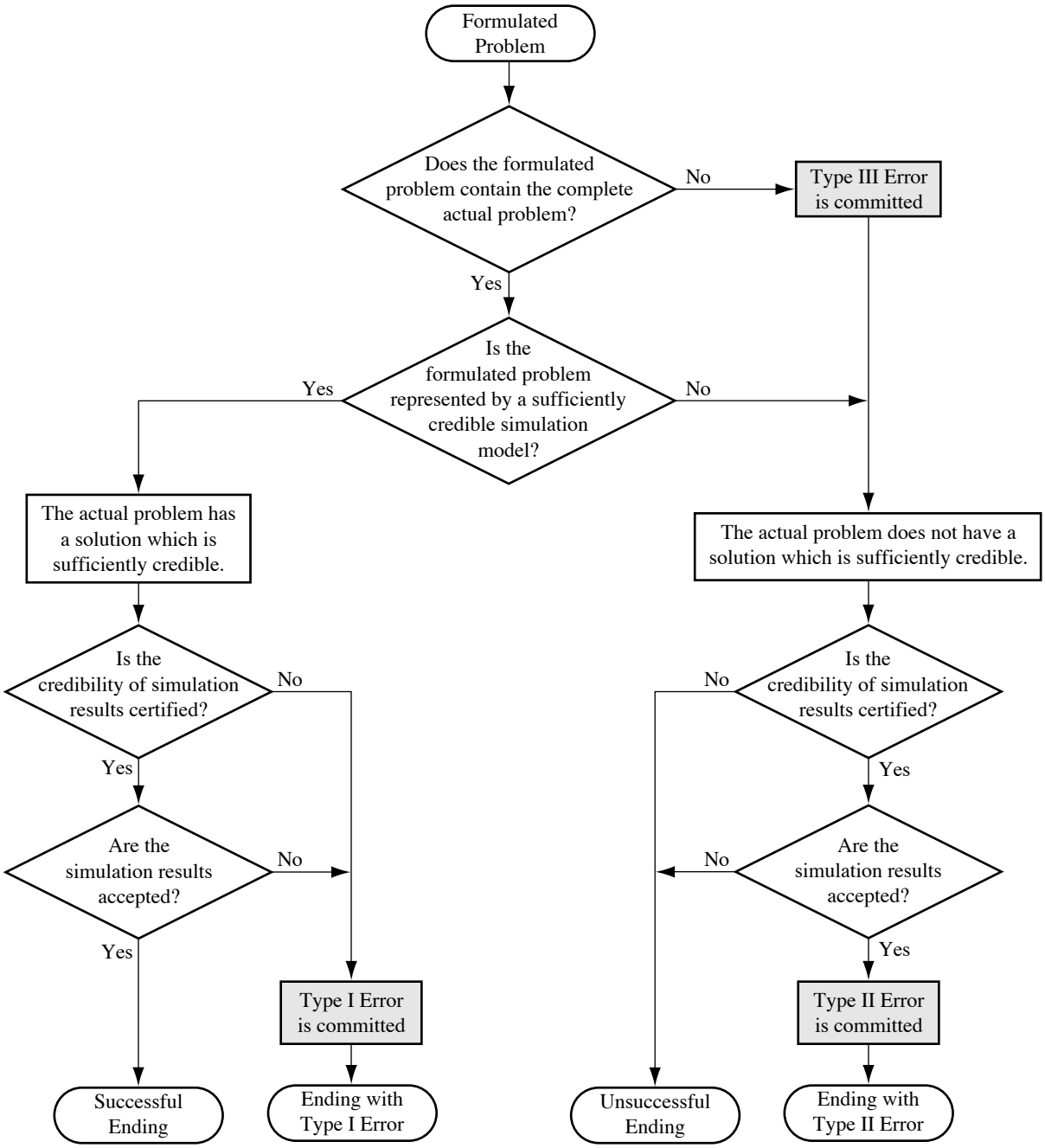

Figure 3. Type I, II, and III Errors in a Simulation Study

results are certified by an independent organization. Although not commonly implemented in practice, the idea of certification is an important one. Certification is a means of removing the developer's bias and promotes *independent* VV&T (See Principle 4).

2.10 Principle 10: Errors should be detected as early as possible in the life cycle of a simulation study

A rush to model implementation is a common problem in simulation studies. Sometimes simulation models are built by direct implementation in a (simulation) programming language with no or very little formal model specification. As a result of this harmful build-and-fix approach, experimental model VV&T becomes the only main credibility assessment stage. On the other hand, detecting and correcting major modeling errors at this stage is very time consuming, complex, and expensive.

Detection and correction of errors as early as possible in the life cycle of a simulation study must be the primary objective. Sufficient time and energy must be expended for each of the 13 credibility assessment stages shown in Figure 1. Correcting errors detected in later phases of the life cycle is much more time consuming and expensive. Some vital errors may not be detectable in later phases resulting in the occurrence of Type II or III error.

Nance and Overstreet [1987] advocate this principle and provide diagnostic testing techniques for models represented in the form of condition specification. A model analyzer software tool is included in the definition of a simulation model development environment so as to provide effective early detection of model specification errors [Balci and Nance 1992].

2.11 Principle 11: Multiple response problem must be recognized and resolved properly

Due to the multiple response problem described by Shannon [1975], the validity of a simulation model with two or more output variables cannot be tested by comparing the corresponding model and system output variables one at a time (i.e., O_1^m versus O_1^s, O_2^m versus O_2^s, ..., O_k^m versus O_k^s as shown in Figure 4) using a univariate statistical procedure. A multivariate statistical procedure must be used to incorporate the correlations among the output variables in the comparison.

Balci and Sargent [1984] present two multivariate statistical techniques using the Hotelling's T^2 statistic for constructing ellipsoidal joint confidence regions to assess model validity. The first technique requires independence between the model and system output data and is intended for self-driven simulation models of observable systems. Assume that the model and system each has k output variables as depicted in Figure 4 with n observations on each model output variable and N observations on each system output variable. Let $(\underline{\mu}^m)' = [\mu_1^m, \mu_2^m, ..., \mu_k^m]$ and $(\underline{\mu}^s)' = [\mu_1^s, \mu_2^s, ..., \mu_k^s]$ be the k dimensional vectors of population means of model and system output variables, respectively. Let $(\underline{\bar{o}}^m)' = [\bar{o}_1^m, \bar{o}_2^m, ..., \bar{o}_k^m]$ and $(\underline{\bar{o}}^s)' = [\bar{o}_1^s, \bar{o}_2^s, ..., \bar{o}_k^s]$ be the k dimensional vectors of sample means of observations on

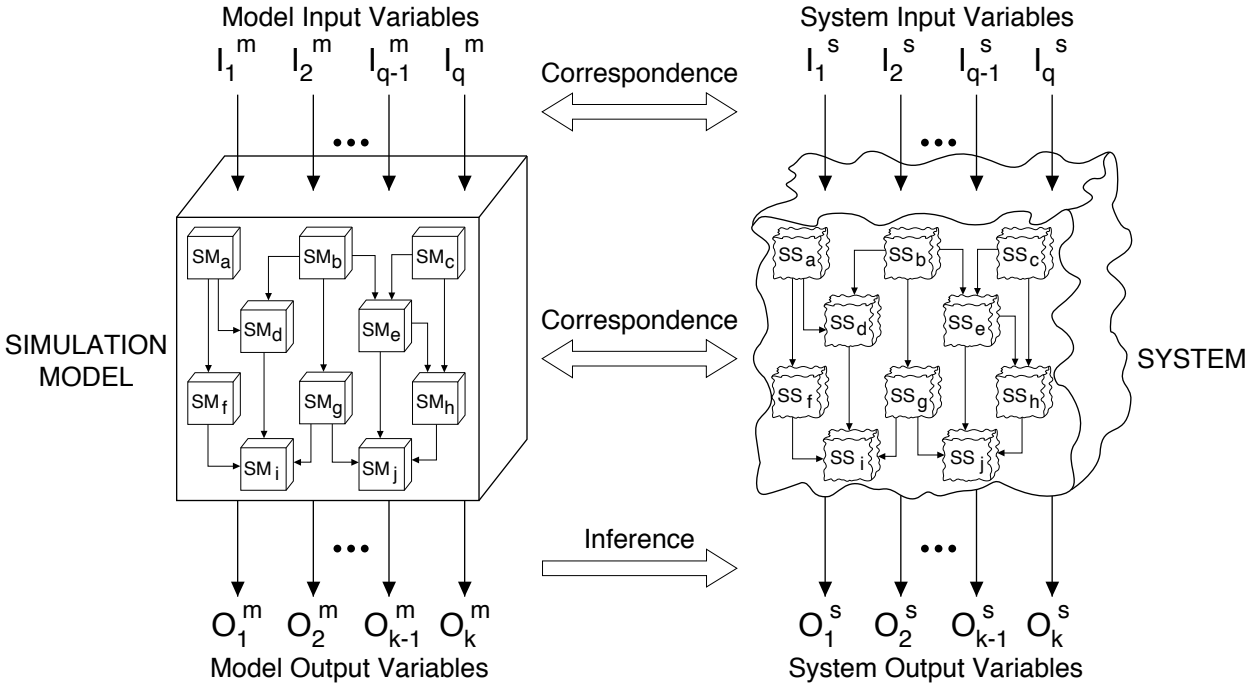

Figure 4. Model and System Characteristics

model and system output variables, respectively. Then, the 100(1 – γ)% joint confidence region is specified by those vectors $\underline{\delta} = (\underline{\mu}^m - \underline{\mu}^s)$ satisfying the inequality

$$(\underline{\bar{o}}^m - \underline{\bar{o}}^s - \underline{\delta})' \, S^{-1} \, (\underline{\bar{o}}^m - \underline{\bar{o}}^s - \underline{\delta}) \leq \frac{n+N}{nN} \, T^2_{\gamma;\, k,\, n+N-k-1}$$

where S is the pooled variance-covariance matrix and $T^2_{\gamma;\, k,\, n+N-k-1}$ is the upper γ percentage point of Hotelling's T^2 distribution with degrees of freedom k and n+N–k–1.

The second technique requires paired observations between the model and system output variables and is intended for trace-driven simulation models. Let $\underline{\bar{d}}' = [\bar{d}_1, \bar{d}_2, ..., \bar{d}_k]$ be the k dimensional vector of sample means of differences between the paired observations on the model and system output variables with a sample size of N. The 100(1 – γ)% joint confidence region consists of the vectors $\underline{\mu}^d$ satisfying the inequality

$$N \, (\underline{\bar{d}} - \underline{\mu}^d)' \, S_d^{-1} \, (\underline{\bar{d}} - \underline{\mu}^d) \;\leq\; T^2_{\gamma;\, k,\, N-k}$$

where S_d is the variance-covariance matrix of the differences.

When k = 3, the joint confidence region can be presented visually as illustrated in Figure 5 and can be used to assess the model accuracy with an exact level of 100(1 – γ)% confidence. We can conclude that we are 100(1 – γ)% confident that the differences between the population means of corresponding

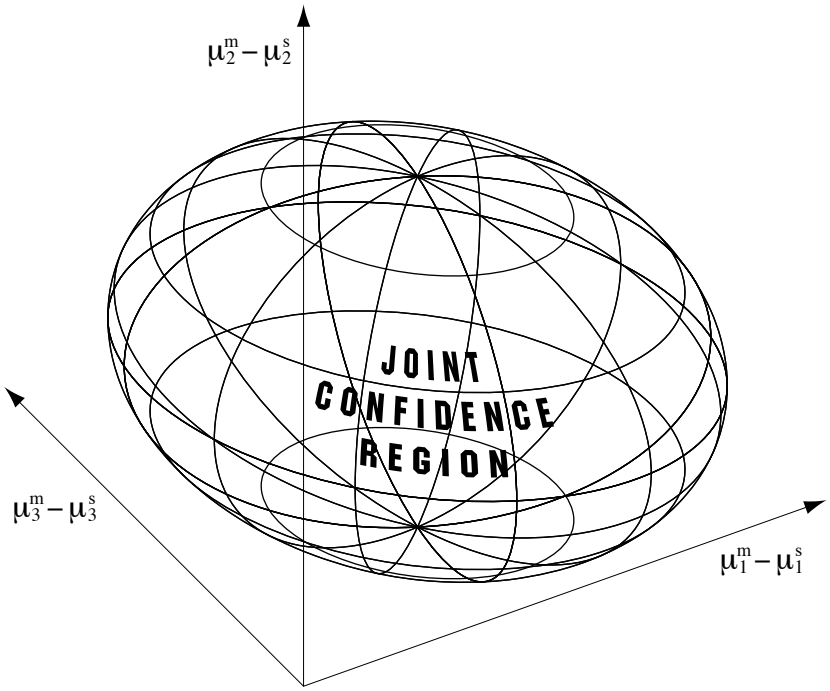

Figure 5. Joint Confidence Region Representing the Validity of a Simulation Model with Three Output Variables

model and system output variables are contained within the joint confidence region shown in Figure 5. Ideally, the joint confidence region contains zero at its center and the smaller the size it has, the better it is. Any deviation from the idealistic case is an indication of the degree of invalidity.

As k increases, interpretation of the joint confidence region becomes difficult, but not impossible. With the use of computer-aided assistance, the model validity can be examined.

2.12 Principle 12: Successfully testing each submodel does not imply overall model credibility

Suppose that a simulation model is composed of submodels (SM_x) representing subsystems (SS_x) respectively as depicted in Figure 4. Submodel *x* can be tested individually by comparison to subsystem *x*, where *x* = a, b, ..., j, using many of the 45 VV&T techniques described by Balci [1994].

The credibility of each submodel is judged to be sufficient with some error that is acceptable with respect to the study objectives. We may find each submodel to be sufficiently credible, but this does not imply that the whole model is sufficiently credible. The allowable errors for the submodels may accumulate to be unacceptable for the whole model. Therefore, the whole model must be tested even if each submodel is found to be sufficiently credible.

2.13 Principle 13: Double validation problem must be recognized and resolved properly

If data can be collected on both system input and output, model validation can be conducted by comparing model and system outputs obtained by running the model with the "same" input data that drives the system. Determination of the "same" is yet another validation problem within model validation. Therefore, this is called the double validation problem.

This is an important problem often overlooked. It greatly affects the accuracy of model validation. If invalid input data models are used, we may still find the model and system outputs sufficiently matching each other and conclude incorrectly on the sufficient validity of the model.

The "same" is determined by validating the input data models. We must substantiate that the input data models have sufficient accuracy in representing the system input process. Input data modeling deals with characterization of the system input data [Johnson and Mollaghasemi 1994]. Simulation models are categorized into two with respect to the way they are driven: trace-driven and self-driven. In trace-driven simulation, the system input is characterized by the trace data collected from the instrumented system. The trace data becomes the input data model which should be validated against the actual system input process.

In self-driven simulations, simulation model is driven by randomly sampling from the probabilistic models developed to represent the data collected on the system input process. Usually, input data modeling is achieved by fitting standard probability distributions to observed data. The input data models should be constructed using a multivariate statistical approach if the input variables are correlated. Individually building a probabilistic model for each input variable does not incorporate the correlations among the input variables; therefore, a multivariate probabilistic approach should be used.

2.14 Principle 14: Simulation model validity does not guarantee the credibility and acceptability of simulation results

Model validity is a necessary but not a sufficient condition for the credibility and acceptability of simulation results. We assess model validity with respect to the study objectives by comparing the model with the system as it is defined. If the study objectives are incorrectly identified and/or the system is improperly defined, the simulation results will be invalid; however, we may still find the model to be sufficiently valid by comparing it with the improperly defined system and with respect to the incorrectly identified objectives.

A distinct difference exists between the model credibility and the credibility of simulation results. Model credibility is judged with respect to the system definition and the study objectives, whereas the credibility of simulation results is judged with respect to the actual problem definition and involves the assessment of system definition and identification of study objectives. Therefore, model credibility assessment is a subset of credibility assessment of simulation results.

2.15 Principle 15: Formulated problem accuracy greatly affects the acceptability and credibility of simulation results

It has been said that a problem correctly formulated is half solved [Watson 1976]. Albert Einstein once indicated that the correct formulation of a problem was even more crucial than its solution. The ultimate goal of a simulation study should not be just to produce a solution to a problem but to provide one that is sufficiently credible and accepted and implemented by the decision makers. We cannot claim that we conducted an excellent simulation study but the decision makers did not accept our results and we cannot do anything about it. Ultimately we are responsible for the acceptability and usability of our simulation solutions although in some cases we cannot affect or control the acceptability.

The accuracy of the formulated problem assessed by the Formulated Problem VV&T (Figure 1) greatly affects the credibility and acceptability of simulation results. Insufficient problem definition and inadequate user participation in defining the problem are identified as two important problems in the management of computer-based models [U.S. GAO 1976]. The importance of the Formulated Problem VV&T has not been recognized sufficiently. The problem is mostly educational. In higher education, students are not trained on how to formulate a problem and how to assess its accuracy. Generally, the interest lies in solving problems; therefore, typically an instructor starts a lecture by saying "here is the problem, let's see how we can solve it." As a result, the students do not gain the much needed knowledge in problem formulation and its VV&T. Consequently, when a real-life problem is encountered, people usually jump into a solution based on the communicated problem without spending sufficient time and energy in properly formulating the problem.

It must be recognized that if the Formulated Problem VV&T is poorly conducted resulting in an incorrect problem formulation, no matter how fantastically we solve the problem, the simulation study results will be irrelevant. The actual story below perhaps places the Formulated Problem VV&T in the correct perspective [Balci and Nance 1985]:

> In a country, the name of which is not important, the Department of Energy (DoE) gave a project to a research institution to determine the best location for building a nuclear power plant. Following an extensive study, location *x* was proposed and accepted by the DoE. However, during the implementation phase, the Department of Defense (DoD) rejected the location for the reason that it can easily be attacked by the enemy. The study was re-initiated.

In the above scenario fault must be attributed to the research institution for failing to identify the DoD as a key influential decision maker and incorporating its objectives in the formulated problem. If it had been done, the research institution would have identified location *x* as infeasible right at the beginning of the study.

Balci and Nance [1985] present an approach to problem formulation and 38 indicators to use in assessing the formulated problem. Balci [1994] describes the following techniques for the Formulated

Problem VV&T: audit, cause-effect graphing, consistency checking, desk checking, face validation, inspections, reviews, structural analysis, and walkthroughs.

3. CONCLUDING REMARKS

Assessing the acceptability and credibility of complex simulation study results poses significant technical challenges to researchers, practitioners, and managers. Nevertheless, we are confident that the challenge can be met by understanding and properly applying the principles of simulation model VV&T presented in this paper. Computer-aided assistance, visualization, knowledge-based approach, and an effective management style can contribute to meeting the challenge.

Automated support is essential to reduce the time and cost of testing since tests are repeatedly used during development and maintenance. Visualization or animation greatly facilitates model VV&T; however, seeing is *not* believing in visual simulation [Paul 1989]. The use of a knowledge base may help improve the quality of testing and provides a means for automation. Management aspects of model VV&T are as important as its technical aspects. An effective SQA program within an organization can make a big difference for the success of model VV&T.

ACKNOWLEDGMENTS

This research was sponsored in part by the U.S. Navy under Contract No. N60921-89-D-A239-34 through the Systems Research Center at VPI&SU. Stimulating discussions with Richard E. Nance are gratefully acknowledged.

REFERENCES

Balci, O. (1994), "Validation, Verification, and Testing Techniques Throughout the Life Cycle of a Simulation Study," *Annals of Operations Research*, to appear.

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

Balci, O. and R.E. Nance (1992), "The Simulation Model Development Environment: An Overview," In *Proceedings of the 1992 Winter Simulation Conference* (Arlington, VA, Dec. 13-16). IEEE, Piscataway, NJ, pp. 726-736.

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

Balci, O. and R.G. Sargent (1984), "Validation of Simulation Models via Simultaneous Confidence Intervals," *American Journal of Mathematical and Management Sciences 4*, 3&4, 375-406.

Beizer, B. (1984), *Software System Testing and Quality Assurance*, Van Nostrand Reinhold Company, New York, NY.

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

Johnson, M.E. and M. Mollaghasemi (1994), "Simulation Input Data Modeling," *Annals of Operations Research*, to appear.

Miller, G.A. (1956), "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *The Psychological Review 63*, 2 (Mar.), 81-97.

Nance, R.E. (1994), "The Conical Methodology and the Evolution of Simulation Model Development" *Annals of Operations Research*, to appear.

Nance, R.E. and C.M. Overstreet (1987), "Diagnostic Assistance Using Digraph Representations of Discrete Event Simulation Model Specifications," *Transactions of the SCS 4*, 1, 33-57.

Ören, T.I. (1981), "Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference," *Communications of the ACM 24*, 4, 180-189.

Ören, T.I. (1986), "Artificial Intelligence in Quality Assurance of Simulation Studies," In *Modelling and Simulation Methodology in the Artificial Intelligence Era*, M.S. Elzas, T.I. Ören, and B.P. Zeigler, Eds. North Holland, Amsterdam, pp. 267-278.

Ören, T.I. (1987), "Quality Assurance Paradigms for Artificial Intelligence in Modelling and Simulation," *Simulation 48*, 4 (Apr.), 149-151.

Paul, R.J. (1989), "Visual Simulation: Seeing is Believing?" In *Impacts of Recent Computer Advances on Operations Research*, R. Sharda, B.L. Golden, E. Wasil, O. Balci, and W. Stewart, Eds. Elsevier Science Publishing, New York, NY, 422-432.

Sargent, R.G. (1992), "Validation and Verification of Simulation Models," In *Proceedings of the 1992 Winter Simulation Conference*, J.J. Swain, D. Goldsman, R.C. Crain, and J.R. Wilson, Eds. IEEE, Piscataway, NJ, 104-114.

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

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

U.S. GAO (1976), "Report to the Congress: Ways to Improve Management of Federally Funded Computerized Models," LCD-75-111, U.S. General Accounting Office, Washington, DC.

U.S. GAO (1979), "Guidelines for Model Evaluation," PAD-79-17, U.S. General Accounting Office, Washington, DC.

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

Watson, C.E. (1976), "The Problems of Problem Solving," *Business Horizons 19*, 4, 88-94.

TR-94-24

PRINCIPLES OF SIMULATION MODEL VALIDATION, VERIFICATION, AND TESTING

OSMAN BALCI

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

balci@vt.edu

10 June 1994

ABSTRACT

Sufficient experience has been gained over the last decade in simulation model validation, verification, and testing (VV&T) to establish basic principles about its characteristics. This paper presents 15 principles of simulation model VV&T. These principles help the researchers, practitioners and managers better understand what model VV&T is all about. They serve to provide the underpinnings for the VV&T techniques that can be used throughout the life cycle of a simulation study. Understanding and applying these principles is crucially important for the success of a simulation study.

CR Categories and Subject Descriptors: I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis; I.6.8 [**Simulation and Modeling**]: Types of Simulation—*Discrete event;* D.2.4 [**Software Engineering**]: Program Verification—*Validation;* D.2.5 [**Software Engineering**]: Testing and Debugging

Additional Key Words and Phrases: Credibility assessment of simulation studies, model testing, model validation, model verification, principles of model testing, simulation quality assurance.