Chapter 1

Introduction

CONCEPTS - Background and Rudiments.
Recommended for all readers.

The problem of simulation validation came to a head in early 1988. The General Accounting Office had just issued a report to Congress, titled "DOD Simulations - Improved Assessment Procedures Would Increase the Credibility of Results."[4] The report stated that "while DOD [Department of Defense] officials agree that credibility is important, DOD generally has not in fact established the credibility of its simulations systematically and uniformly." In the Department of Defense, simulations are a multibillion dollar business. Later that spring, the Computer Professionals for Social Responsibility (CPSR) expressed a major concern of the technical community. They issued a report criticizing a one billion dollar investment in a Strategic Defense Initiative project that involved a vast network of computer facilities and simulation centers which, if ever completed, "will be the world's largest simulation network."[5] The objective of the Strategic Defense Initiative is to build a trustworthy defense against ballistic missiles. The CPSR group's major statement was that "we cannot realistically simulate the conditions of such a conflict" thus invalidating any decisions based on simulation results.[5]

These concerns are not limited to the government sector. Simulations in private industry help managers and engineers design or operate complex systems. In his book, *The Day the Phones Stopped,* Leonard Lee provides graphic illustrations on how software failures can and have affected segments of our society.[6] His examples include the 1990 failure of AT&T's entire long-distance network and crashes of fly-by-wire aircraft. These sophisticated fly-by-wire aircraft include the Air Force's F-16, Boeing's 767, Airbus' A320, the Navy's F/A-18, and Sweden's Gripen jet fighter. The latter two experienced catastrophic failures due to software design problems. The design and test of the fly-by-wire systems are highly dependent on simulations. After the crash of the Gripen, the developer's program manager said, "We never encountered that situation in the simulation."[6]

If the risks and costs are high, why do we continue to build and use simulations? Perhaps the primary reason is that simulations allow us to investigate and understand systems that either do not exist or cannot be used for experimentation. The simulation of a system that does not exist will provide information on the system's probable performance under a variety of conditions. This will support decisions on basic concepts, system design, and feasibility of operation without going to the expense of developing prototypes or test models. This is especially important if testing a system may result in its destruction. It may be prudent to simulate the operation of existing systems because of the costs of running experiments, the inability to create or apply realistic test conditions, the difficulties in accessing the system, the lack of adequate testing equipment, or concerns about safety.

On the other hand, it is important to understand the risks involved with simulation. A simulation may not adequately represent the real-world system. The data used to drive it may be inaccurate. It may be too difficult to model the operational environment or all the interactions that affect the real system. Output data may be flawed or subject to misinterpretation. Despite all their potential for saving money, simulations can be costly in terms of human effort and computer resource requirements. And of course, there are always questions about the credibility of the simulation tool and its output.

Objectives

This book provides a systematic, procedural, and practical approach toward the evaluation of simulations. The process describes tools and techniques that should lead to an efficient, credible assessment of a simulation model. As such, the specific goals are to

- highlight key points concerned with establishing a model's credibility,

- provide a practical, systematic process and not a philosophical treatise on simulation issues and pitfalls,
- make the assessment less a matter of faith and more the result of the methodology, and
- establish the credibility of simulations used to support decisions, thus increasing confidence in those decisions.

Simulations are attempts to model the real world. They must be used carefully within their domain of application. Typically, models are developed for specific uses and have limitations on their application. When used for experimentation, variation of simulation parameters must be done wisely to prevent misinterpretation of output data or, consequently, of the system being modeled. In general, we will not explicitly cover the design, development, or application of simulations. However, cognizance of the assessment methodology will mitigate risks involved with these areas.

Definition of terms

The difficulty in establishing clear terminology in this field is legendary. A comprehensive bibliography search reported by Balci and Sargent encountered the following 16 common terms: acceptability, accuracy, analysis, assessment, calibration, certification, confidence, credibility, evaluation, performance, qualification, quality assurance, reliability, testing, validation, and verification.[7] One can even argue about the correctness of the title of this book since there are certain situations where simulations can never be validated given a strict definition of the term. Regardless, careful consideration of terminology is helpful in understanding the tasks at hand in the confidence assessment process. Implicit in the terminology are the limitations on simulation assessment as well. For example, if a simulation models a real-world system that does not yet exist, then we technically cannot validate the simulation. In their text on simulation modeling, Law and Kelton compares real-world data and simulation output. They state: "If there is reasonable agreement, we have increased confidence in the 'validity' of the model."[8] This statement shows the authors' preference for using levels of confidence to express model validity. It illustrates the difficulty in applying strict terminology in the field of simulation assessment.

In 1979, the Society for Computer Simulation Technical Committee on Model Credibility provided a framework for assessing simulations, as illustrated in Figure 1-1.[9] The definitions provided later are in the context of this illustration.

This figure illustrates a convenient decomposition of a simulation into three basic elements. The inner triangle shows the interrelationship of these elements. The outer arrows in the cycle refer to the procedures employed to establish credibility of a simulation. The scheme was further expanded by Robert Sargent and his model (described in Chapter 2) is the cornerstone of the confidence assessment methodology.

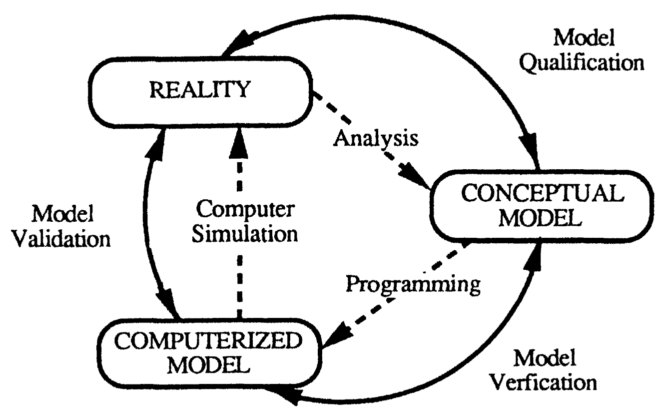


Figure 1-1. Simulation model.

The committee gave a set of definitions that describe the basic elements and their interrelationships.[9] This book is written in the context of these definitions, although for variety, alternative terminology is occasionally substituted. (These and other definitions are summarized in the Glossary in Appendix A.)

- *Simulation.* Modeling of systems and their operations using various means of representation. (Occasionally referred to as a model, tool, simulation model, or toolset.)
- *Reality.* An entity, situation, or system selected for analysis. (Also referred to as real-world system or real-world entity.)
- *Domain of applicability.* Prescribed conditions for which the computerized model has been tested, compared against reality to the extent possible, and judged suitable for use.
- *Range of accuracy.* Demonstrated agreement between the computerized model and reality within a stipulated domain of applicability.
- *Verification.* Substantiation that the computer program implementation of a conceptual model is correct and performs as intended.
- *Validation.* Substantiation that a computer model, within its domain of applicability, possesses a satisfactory range of accuracy consistent with the intended application of the model.
- *Confidence assessment.* The process of assessing the credibility of a simulation by means of the methodology as described in this book. (Occasionally referred to as assessment, simulation evaluation, or model evaluation.)

The terms model, tool, simulation model, and toolset are often substituted for the word simulation, as noted in the definition above. In the strictest sense, these terms have different meanings. In fact, a simulation may refer to the application of a simulator to an input data set and the collection of the output data. In the field, this is sometimes called a simulation run. A model is defined as a physical or mathematical abstraction of a real world process, device, or concept. A simulation model can be defined as the representation of a model in computer code. Simulators are tools to support analyses. A simulation toolset can be thought of as a simulator and the entire toolkit that makes it work (that is, host computer, input data preprocessor, output data post-processors, and so on). In this book, these terms are most often used correctly in context. From time to time, the words are interchanged for variety.

Historical background

The literature on simulations and simulation validation is rich and fascinating due to the breadth of application. Credit is given to Conway, Johnson, and Maxwell for the earliest discussion of simulation methodology in their 1959 paper.[10] Conway continued his vanguard work by providing the earliest documented listing of simulation assessment techniques in his 1963 article.[11] Naylor and Finger provided a very comprehensive article on the subject in 1967.[12] Yet Naylor, in his 1971 text on simulation lamented on the difficulty in establishing universally acceptable criteria for accepting a simulation model as a valid representation.[13]

In 1978, the Society for Computer Simulation formed a Technical Committee on Model Credibility. Their 1979 report to the general membership provided the first framework for simulation assessment.[9] In 1984, Balci and Sargent compiled over 300 references on the credibility assessment and validation of simulation models.[7] Sargent codified much of the previous literature in his landmark work.[14-18] Gass made a number of contributions to the field[19-23] and most recently coauthored a case study based on an assessment procedure.[24] Two other leaders in the field, Averill Law and David Kelton, provided guidance on building valid simulation models by devoting an entire chapter of their text to the subject.[8] Law and Kelton's list of techniques and three-step approach provide a practical methodology for evaluating simulations. This book contains all their techniques.

Simulation validation was a major concern for managers and users of the National Test Bed (NTB). The NTB was established by the Strategic Defense Initiative Organization to provide hardware and software to support simulations and experiments conducted on strategic defense concepts. In recognition of the GAO report mentioned earlier, the NTB formed a group of experts to advise their managers on how to systematically evaluate simulation models. The Simulation Evaluation Methodology Technical Group recommended creation of a confidence methodology guide.[1] The NTB commissioned the production of this guide,[3] which was subsequently applied to the assessment of many large-scale simulations. This guide provided much of the inspiration and material in this book.

Applying the confidence methodology guide

This book provides a complete methodology. It will apply for the worst-case scenario of assessing a fully developed tool where no previous assessment was performed. It gives procedures to take advantage of testing, verification, or validation previously performed on the model. Guidance is also provided on applying the methodology to evolving models. The methodology given in this book may appear overwhelming to apply; it can be tailored for particular situations and objectives. Numerous assessment aids and tools are provided. In fact, an entire chapter is devoted to providing an extensive example of aids for a formal assessment.

Figure 1-2 describes the flow of this guide. It illustrates that after this introduction, the basic concepts of simulation assessment are covered in depth in Chapters 2 and 3. Given the practical advice and tools provided in these chapters, the practitioner must decide on the type of assessment to perform, either a formal or limited assessment. This is illustrated by the branch point on the flow chart for Chapters 4 and 5. There are two special topics areas in this guide to help assess simulations that involve man-in-the-loop or hardware-in-the-loop. For example, a simulation that supports an interactive tool for training operators of a power generation plant may need to apply both special procedures. Finally, Chapter 8 goes through a mock formal assessment to further familiarize the practitioner with all of the planning and organizing aids given in this guide.

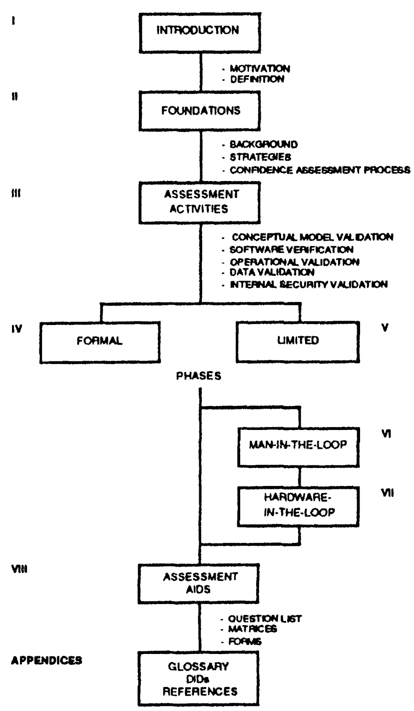


Figure 1-2. Confidence methodology guide flow chart.

This book is organized into three parts, as shown in the table below.

Part I - Concepts	**Part II - Methodologies**	**Part III - Special Topics**
Chap. 1 - Introduction	Chap. 4 - Formal Assessments	Chap. 6 - Man-in-the-Loop Models
Chap. 2 - Foundations	Chap. 5 - Limited/Maintenance Assessments	Chap. 7 - Hardware-in-the-Loop Models
Chap. 3 - Assessment Activities		Chap. 8 - Assessment Aids

Table 1-1. Organization.

Part I, *Concepts*, introduces the fundamentals of confidence assessments at a survey level. This part should be of special interest to managers who want to know the rudiments of confidence assessment without too much detail on programmatic issues.

- Chapter 1, *Introduction*, motivates this subject from a pragmatic point of view. After defining some common terminology, the subject is put into perspective with an anecdotal example. A brief recount of historical background is provided for researchers in the field.

- Chapter 2, *Foundations*, lays out the practical issues involved with implementing the confidence assessment methodology. It introduces the team concept, developer involvement, user involvement, and general approach. A structure is described for viewing simulations which accordingly provides a natural basis for prescribing assessment activities. We also present issues governing model certification and recommend some alternatives.

- Chapter 3, *Assessment Activities*, is the guts of confidence assessment. Assessment processes and their associated activities are explicitly discussed. The processes provide five different perspectives on the model being assessed. This is a very thorough set of processes – probably the most complete listing in the industry – which can be disorienting to the practitioner. For easy reference, a comprehensive listing of the assessment processes is provided at the end of the book.

Part II, *Methodologies*, provides structured schema for the confidence assessment of simulations. It is addressed to the members of the confidence assessment team that should be comprised of technical experts, software engineers, software quality assurance personnel, and a team director. The director will use the information in Part II and, possibly, Part III to plan, organize, and monitor the assessment effort.

- Chapter 4, *Formal Assessment*, is a guide for the complete assessment of a simulation model. The methodology provided involves the greatest depth and scope, and should be applied when the intended use of the simulation is very well defined. It will result in assurances that the development of the model was thoroughly investigated, that it was rigorously tested, and that the risks involved in using the simulation are exposed.

- Chapter 5, *Limited/Maintenance Assessment*, describes a lower level of confidence assessment that provides an indication of model credibility when a formal assessment cannot be undertaken because of time or resource constraints, or when a quick look is adequate. This methodology can be applied during model development, while model discrepancies can easily be corrected. Maintenance assessments are performed on previously assessed models which were either modified or applied to significantly different conditions.

Part III, *Special Topics*, examines aids for assessments. The first two chapters cover procedures for unique concerns. Here the assessment of the simulation will require different procedures and modifications to the assessment methodology.

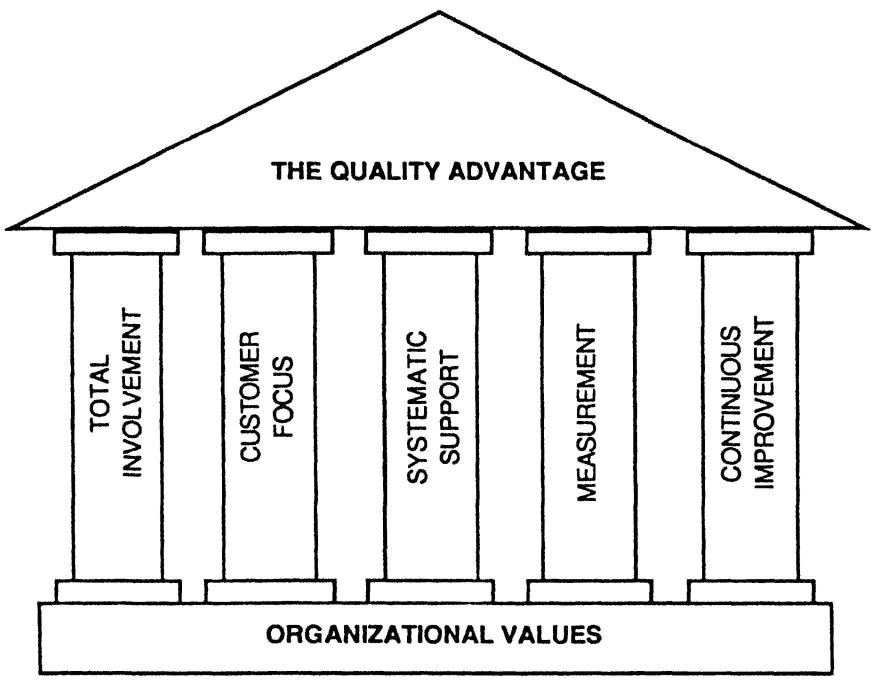
- Chapter 6, *Man-in-the-Loop Models*, provides guidance for assessment of models that involve human interface in the simulation process. These simulations model man-machine interfaces and must include algorithms to provide information to humans, accept their responses, and generate actions based on the human response. These simulations pose unique evaluation criteria.

- Chapter 7, *Hardware-in-the-Loop Models*, discusses the special concerns when the simulation must interface with hardware. Specifically, these models may be linked to machine emulators, real hardware devices, or integrated portions of the system being modeled.

- Chapter 8, *Assessment Aids*, contains specialized tools to assist in planning, organizing, and conducting a systematic assessment. These include a question list for formal assessments, a model characterization matrix, cross-reference matrices, and a typical assessment schedule. An anecdotal example illustrates application of these tools.

Total Quality Management (TQM)

The TQM process is reinforced with the application of the confidence assessment methodology. A quality organization is concerned with the people and products that establish organizational values, forming a foundation for an organizational approach toward quality. Figure 1-3 illustrates the TQM pillars of quality, as defined by Organizational Dynamics, Inc.[25]



Reprinted, with permission, from ODI, 25 Mall Road, Burlington, MA 01803.

Figure 1-3. Five TQM pillars of quality.

The association between the confidence assessment methodology and these pillars is described below.

- *Total involvement* gets everyone in the organization involved in achieving quality. The team concept, described in the next chapter, includes the assessment team, management, developers, users, and a review group. Someone in this team also represents the sponsor of the assessment.

- *Customer focus* is the link between the organization and its customer. The customer can be the intended user of the simulation or the sponsor of the assessment. The team concept and periodic reviews of the assessment process always include the user or sponsor. This will ensure that customer expectations and requirements are satisfied.

- *Systematic support* is the support of all systems in the organization toward a quality effort. In an organization that either develops or uses a simulation, the entire support structure must work together. For example, operations personnel must allow for simulation operations by providing facilities, data, and equipment. It is also a by-product of the team approach.

- *Measurement monitors quality.* Assessment involves the characterization of the simulation tool and a set of procedures to evaluate selected technical areas. This decomposes the problem into measurable components.

- *Continuous improvement* requires constant vigilance to correct problems, prevent problems, and make improvements. The Confidence Assessment Methodology is built on this principle. Limited and Maintenance Assessments are made on simulations as they evolve. Discrepancies are noted and reported to the simulation developers. All assessments encourage good configuration control and correction of discrepancies. The final products of an assessment are a characterization of the simulation and identification of any risk areas. This is an organized effort to prevent problems with the ultimate application of the simulation.

Illustrative example

Several chapters introduce or mention examples of real-world systems that are excellent candidates for simulation studies. These examples illustrate the complexity involved in adequately representing these systems in a simulation and the corresponding challenge in assessing those models. As a means of introduction to this guide, consider the following power distribution problem.

The Goodbytes Company was contacted to perform a confidence assessment of Spark State Power Company's system simulation. Their model is 10 years old and was informally reviewed by an in-house team two years ago. Since then, the company hired Loosebits Company to create a mock-up control room and modify the simulation to drive the displays. The mock-up control room will be used for training employees. The system being simulated is a network of four power generation plants supplying a customer base in the Midwest. They are also linked to their neighbor, the Surge Power Company, in case of sudden changes in demand.

Given a broad problem statement like this, Goodbytes' assessment team may have difficulty defining their effort. Chapter 2 discusses the practical aspects of starting an assessment. In the early stages the team needs to ask the following questions.

What is the intended use of the simulation?

Here it seems fairly clear that the simulation will interact with human input in order to provide realistic training scenarios.

Who is the user and who are the model developers? Are there any other interested parties? Who is ultimately interested in the outcome of the assessment?

Here they can clearly determine that Spark State is the user of the simulation and Loosebits are the developers. But, will the public utilities commission or some other regulatory agency be interested in their assessment? If so, they must anticipate the political atmosphere.

How important are the decisions resulting from the use of this simulation model?

The decisions may significantly affect the revenues of the company. Public safety may be a factor. The potential impact of these decisions should influence the extent and cost of the confidence assessment effort.

How extensive is the simulation?

This will help them judge the scope of the assessment and decide on the expertise that they will need on their team. No doubt, the actual system is quite complex with the potential for several scenarios. The inputs, like changes in weather or power-generating capacity, are time varying. Is the simulation capable of integrating real-world data inputs?

What has been done in the past?

The previous review mentioned could be very helpful as a spring board into the assessment or it could be entirely worthless. The team can judge that by comparing the procedures used with those recommended in this guide, in Chapters 3, 4, and 5. If the previous review was credible, they may want to employ a limited/maintenance assessment as described in Chapter 5 instead of the more extensive formal assessment described in Chapter 4.

What are the users' expectations? What is their budget?

This could be a very complex question. For example, the assessment may be requested because of pressure from a regulatory agency. Then the user's expectations are that they will make the agency happy. Then they need to establish the agency's expectations and not the user's! The answers here will also help them determine if they should do a limited or formal sssessment.

Are there any special considerations in this assessment? (For example, are the developers done? Does the assessment team have access to the model? Are there man- or hardware-in-the-loop models?)

This is especially important because the answers will tell the team about the extent of and constraints on their assessment efforts and the work atmosphere that they will encounter. Clearly, this problem will involve "interactive gaming" which is another term for man-in-the-loop models. Chapter 6 will be their guide for that consideration. Does the simulation interact with power plants or distribution centers? If so, then this is a hardware-in-the-loop model and they will need to consult Chapter 7.

As you can see, the evaluation of a simulation can be almost as much fun as designing the simulation itself. Now that you are sufficiently prepared and motivated, dig into the guide and happy hunting!