Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

WELL-DEFINED INTENDED USES: AN EXPLICIT REQUIREMENT FOR ACCREDITATION OF MODELING AND SIMULATION APPLICATIONS

Osman Balci

Department of Computer Science
660 McBryde Hall, MC 0106
Virginia Tech
Blacksburg, VA 24061, U.S.A.

William F. Ormsby

Naval Surface Warfare Center
Dahlgren Division, Code T12
17320 Dahlgren Road
Dahlgren, VA 22448-5100, U.S.A.

ABSTRACT

A modeling and simulation (M&S) application is built for a specific purpose and its acceptability assessment is carried out with respect to that purpose. The accreditation decision for an M&S application is also made with respect to that purpose. The purpose is commonly expressed in terms of "intended uses." The quality of expressing the intended uses significantly affects the quality of the acceptability assessment as well as the quality of making the accreditation decision. The purpose of this paper is to provide guidance in proper definition of the intended uses.

1 INTRODUCTION

A *model* is a representation or abstraction of something such as an entity, a system or an idea. *Simulation* is the act of experimenting with or exercising a model or a number of models under diverse objectives including acquisition, analysis, and training. For example, if the analysis objective is to predict the performance of a complex system design, we *experiment* with a model or a distributed set of models representing the system design. If the predicted performance is used in making an acquisition decision, the process is called *simulation-based acquisition.* If the training objective is to teach military commanders how to make decisions under a combat scenario, we *exercise* a model or a distributed set of models in an interactive manner by using the trainees as part of the simulation. We refer to a specific simulation created for a particular objective as a modeling and simulation (M&S) application.

The U.S. Department of Defense (DoD) is the largest sponsor and user of M&S applications in the world. DoD uses many different types of M&S applications such as continuous, discrete-event, distributed, hardware-in-the-loop, human-in-the-loop, Monte Carlo, parallel, and synthetic environments bringing together simulations and real-world systems. The DoD Instruction 5000.61 states that "it is the DoD policy that: … models and simulations used to support major DoD decision-making organizations and processes … shall be accredited for that use by the DoD component sponsoring the application" (DoDI 1996).

The DoD Instruction 5000.61 defines accreditation as "the official certification that a model, simulation, or federation of models and simulations is acceptable for use for a specific purpose" (DoDI 1996).

This paper focuses on the explicit definition of the "specific purpose" for which the accreditation recommendation is formulated. Section 2 describes the importance of an M&S application's purpose, which is commonly expressed in terms of intended uses. Section 3 illustrates different levels of the intended uses by using an analogy. Section 4 presents a hierarchical definition of the intended uses for an example M&S application. Section 5 discusses the design of experiments. Concluding remarks are given in Section 6.

2 IMPORTANCE OF THE INTENDED USES

Figure 1 shows some basic model and system concepts. We build a model to represent a system. However, the representation is never intended to be perfect because a model, by definition, is an abstraction or approximation of a system. Therefore, we try to create a representation so that the correspondence between the model and the system is good enough. How good is good enough is judged with respect to the intended uses.

In judging how good is good enough, we assess the fidelity of the model. *Model fidelity* is the degree of representativeness of a model and is judged with respect to the intended uses.

Modeling is an art. It requires artful balancing of opposites. On the one hand, we strive not to include unnecessary details of the system in the model representation and make the model unnecessarily complex and difficult to analyze. On the other hand, we strive to include all essential elements of the system. The artful balancing of what-to-include and what-not-to-include is carried out with respect to the intended uses.

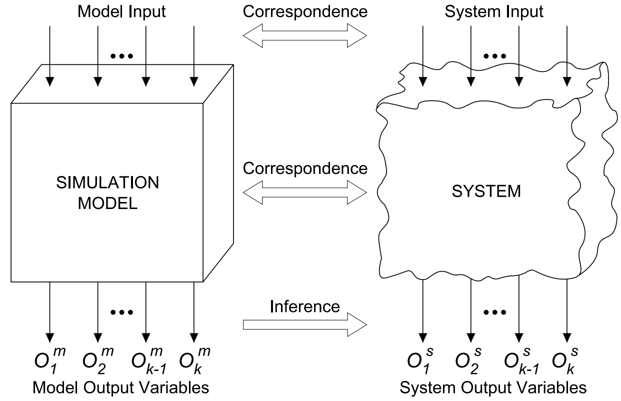

Figure 1: Basic Model and System Concepts

Consider a simulation model being built to represent a traffic intersection. Should the model representation include pedestrians, bicycles, emergency vehicles, right turn on red, or automatic sensors? If any of these is included, at what level of detail should each be represented? These types of questions refer to the assumptions we make in abstracting a real or imaginary system and the answers depend on the intended uses.

A typical test for model validity (the degree to which the model accurately mimics the system behavior) is carried out by running or exercising the simulation model with the "same" input data and conditions that drive the system, and comparing the model and system outputs. This test assumes that the system is real and data can be collected on its input and output. The "same" is assessed based on the degree of correspondence between the model and system inputs. For stochastic system inputs (e.g., interarrival times of vehicles to a traffic intersection lane), probabilistic models (e.g., exponential probability distribution with a given mean) are built. These models are called *input data models*. Thus, in performing this test, we face a double validation problem (Balci 1997). We need to first validate the input data models in validating the simulation model itself. Both validations are also carried out with respect to the intended uses.

3 ILLUSTRATION OF DIFFERENT LEVELS OF THE INTENDED USES

In this section, we use an M&S application for simulating the U.S. National Missile Defense (NMD) system design as an example to illustrate the definition of the intended uses. We refer to this M&S application as the NMDMSA.

Figure 2 illustrates by analogy that different levels of intended uses exist. At the highest level, the intended use given in Figure 2 describes the purpose of the M&S application broadly. The next level indicates that NMDMSA is intended for supporting the Deployment Readiness Review (DRR) by way of simulating the entire NMD system design. This specification is, by analogy, as clear and meaningful as the purpose stated as "develop a vehicle for the intended use of land transportation." However, it is more detailed than the specification given at the previous level.

On the third level, the intended use specification is given with more detail. We should assess whether the detailedness of the intended use specification is sufficient to guide the development of NMDMSA and to perform experiments with NMDMSA to produce the required results. Certainly, the detailedness is insufficient since no specification is made about *how* NMDMSA is intended to support the assessment of the DRR system functions 6 through 10. No information is given about how the

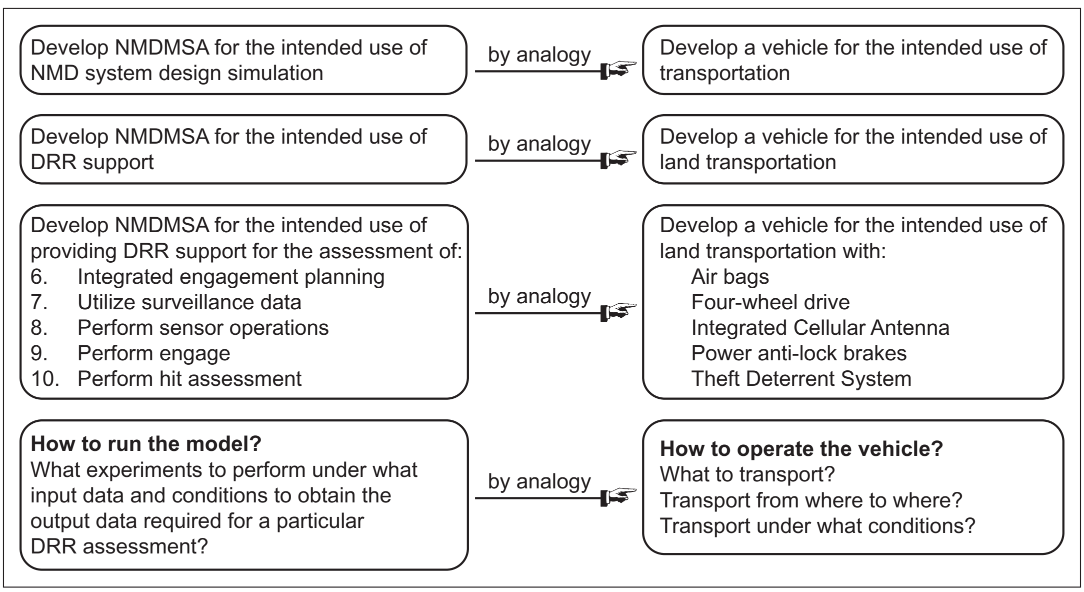

Figure 2: Illustration of Different Levels of the Intended Uses by Analogy

experiments will be conducted to produce the results required for the assessment of the DRR system functions.

The intended use specification on the third level should be decomposed further. The decomposition should continue until the specification is sufficiently detailed to guide the NMDMSA development and to design the experiments for producing the desired results.

4 HIERARCHICAL DEFINITION OF THE INTENDED USES

In this section, we present a hierarchy of decomposition of the NMDMSA intended uses. At the highest level, the *NMDMSA purpose* is decomposed into:

1. NMD system performance assessment
2. NMD ground and flight test prediction, planning, and design
3. NMD system integration support
4. NMD DRR support
5. Operational Test Agency (OTA) analysis

We call the sub-purposes at the highest level as the NMDMSA *Domains of Applicability*. Each domain of applicability should be decomposed further. Table 1 depicts the decomposition of the “NMD DRR support” domain of applicability.

The NMD Detailed Analysis Plan (DAP) describes how NMDMSA will be used to support the DRR decision-making process. DAP presents the big eleven DRR system functions and numerous technical performance measures (TPMs) as depicted in Table 1. We call the DRR system functions corresponding to the “X”s in the rightmost column of Table 1 as the NMDMSA DRR support *sub-domains of applicability*.

The system TPMs marked with an “X” in the rightmost column of Table 1 represent the NMDMSA output variables or response variables (see Figure 1). The marked TPMs are estimated by conducting experiments with NMDMSA. The NMDMSA intended uses are partially defined under each sub-domain of applicability by the marked TPMs. To complete the definition of each intended use, we must specify a design of experiments under which each TPM will be estimated. For example, an intended use can be specified as:

> “We intend to use NMDMSA to estimate the probability of integrated system effectiveness P_{ISE} by conducting experiments under the design <design id> in support of the NMD DRR decision-making process.”

NMDMSA can be fully or partially used in estimating a TPM value. Other tools and techniques may also contribute to the estimation process. The acceptability of NMDMSA should be assessed only with respect to its contribution in the estimation process.

5 DESIGN OF EXPERIMENTS

Models are classified into two categories: Prescriptive (normative) models and descriptive models. A *prescriptive model* is the one that describes the behavior of a system <u>with</u> a value judgment on the “goodness” or “badness” of such behavior. For example, Operations Research models such as linear programming, integer programming, or mixed integer linear programming produce a solution with a value judgment such as “feasible”, “infeasible”, or “optimal.”

A *descriptive model* is the one that describes the behavior of a system <u>without</u> a value judgment on the “goodness” or “badness” of such behavior. *All simulation models are descriptive models*. A simulation model produces a large amount of data that need to be structured and interpreted to reach a meaningful conclusion. The requirements for such structuring and enabling the analyst to interpret the simulation output data demand the design of experiments.

One major difference between a typical software product and a simulation model is that we execute the software product once to obtain the results, but we run the simulation model many times for thousands or millions of observations (i.e., experiment with it) to obtain the results. Therefore, we need to create an experimental design.

An *experimental design* is the process of formulating a plan under which the simulation model is executed to produce the required information at minimal cost in a suitable form to enable the analyst to draw valid inferences (Shannon 1975). A simulation model incorporates an executable description of the operations presented in such a plan.

An experimental design typically specifies the:

- number of times to run the simulation model.
- length of each simulation run.
- length of model warm-up (transient) period.
- input data.
- experimental conditions and scenarios.
- generation of random variates to represent stochastic input and system conditions.
- strategy to collect data during each simulation run.
- strategy to achieve the experimentation purpose such as comparison, evaluation, optimization, prediction, ranking and selection, and sensitivity analysis.
- statistical techniques to be used for presenting the simulation output data (e.g., confidence intervals).

Table 1: Decomposition of the "NMD DRR Support" Domain of Applicability

DRR System Function *NMDMSA Sub-Domain of Applicability*	NMD System Technical Performance Measure	NMDMSA Intended Use
1. System Operations Activation		
2. Readiness Operations		
3. System Status	• Health and Status Reporting Time	
	• Critical Data & Display Accuracy with NMD & ITW/AA	
	• Display Concurrency of NMD / BMC2	
4. Collateral Mission		
5. Control of Defense	• System Activation Time	
	• Situational Awareness	X
	• DEA Granting Time	X
	• Management by Exception	X
6. Integrated Engagement Planning	• Probability of Engagement Planning	X
	• Engagement Planning Time	X
7. Surveillance Data	• Probability of Warning	X
	• Reported Position Accuracy	X
	• Reported Velocity Accuracy	X
8. Sensor Operations	• XBR Probability of Acquisition	X
	• UEWR Probability of Acquisition	X
	• XBR Probability of Sensor Track Reporting	X
	• UEWR Probability of Sensor Track Reporting	X
	• XBR Position Track Accuracy	X
	• UEWR Position Track Accuracy	X
	• XBR Velocity Track Accuracy	X
	• UEWR Velocity Track Accuracy	X
	• XBR Probability of Sensor Discrimination	X
	• UEWR Probability of Sensor Discrimination	X
9. Engagement	• Probability of Kill Single Shot – P_{KSS}	X
10. Hit / Kill Assessment	• Probability of Hit – P_{HIT}	X
11. Launch Essential Maintenance		
Overall System Wrap-Up	• Probability of Integrated System Effectiveness – P_{ISE}	X
	• Engagement Timing Margin	X
	• Graceful Degradation	
	• Integration of Operator Interfaces	
	• NMD and ITW/AA Defended Area	

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. *Factorial designs* can be employed to determine the effect of various input variables on a response variable. *Variance reduction techniques* can be implemented to obtain greater statistical accuracy for the same amount of simulation. *Ranking and selection techniques* can be utilized for comparing alternative systems. Several methods such as replication, batch means, and regenerative can be used for statistical analysis of simulation output data. (Banks 1998; Banks, Carson, and Nelson 1996; Law and Kelton 2000)

An experiment is designed for a specific intended use or a set of intended uses. An M&S application must be developed in such a way that a set of experiments can be designed and performed to meet the needs of its intended uses. For example, by analogy, we cannot use a vehicle for sea transportation if it is not built for that purpose. Similarly, we cannot conduct an experiment with an M&S application unless it is built in such a way that it lends itself for such experimentation. Hence, the intended uses and experimental designs should be defined very early in the M&S application development life cycle.

6 CONCLUDING REMARKS

The credibility of an M&S application can be claimed only for the prescribed conditions for which the M&S application is subjected to acceptability assessment. The accuracy of the input-output transformation of an M&S application 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.

Definition of an intended use of an M&S application includes the description of the input data and experimental scenarios under which the M&S application is subjected to acceptability assessment. The accreditation recommendation can be made only for those input data and experimental scenarios underlying an intended use. Interpretation of the recommendation to imply that the M&S application is acceptable for estimating the value of a particular TPM under any input conditions must be avoided.

Delaying the explicit and proper definition of the intended uses to later stages of the M&S application development life cycle may result in the creation of an unacceptable application. Since the M&S acceptability is assessed and M&S development is carried out with respect to the intended uses, improper definition of intended uses creates an inappropriate point of reference for the development and verification and validation (V&V) activities throughout the life cycle.

The V&V of the design of experiments should not be underestimated. The M&S application can be created to be acceptable, but its results can be useless or the credibility of its results can be seriously damaged due to erroneous experimental design. Experimental design V&V should assess indicators such as:

- How accurate are the random variate generation algorithms theoretically?
- How accurately are the random variate generation algorithms translated into executable code?
- How appropriate are the statistical techniques used?
- How well are the experimental scenarios specified and implemented?
- How well are the input data specified and used?
- How well are the statistical techniques specified and implemented?
- How well is the problem of the initial transient (or the start-up problem) resolved?
- How well is the random number generator tested?
- How well is the strategy to achieve the experimentation purpose specified and implemented?

ACKNOWLEDGMENT

The research leading to the development of this paper has been funded by the Naval Surface Warfare Center Dahlgren Division, Dahlgren, Virginia.

REFERENCES

Balci, O. 1997. Principles of simulation model validation, verification, and testing. *Transactions of the Society for Computer Simulation International* 14(1): 3-12.

Banks, J. ed. 1998. *The handbook of simulation*. John Wiley & Sons, New York, NY.

Banks, J., J.S. Carson, and B.L. Nelson. 1996. Discrete-event system simulation. 2nd ed. Prentice-Hall, Upper Saddle River, NJ.

DoDI 1996. *DoD modeling and simulation verification, validation, and accreditation*. Department of Defense Instruction 5000.61, Apr.

Law, A.M. and W.D. Kelton. 2000. Simulation modeling and analysis. 3rd ed. McGraw-Hill, New York, NY.

Shannon, R.E. 1975. Systems simulation: the art and science. Prentice-Hall, Englewood Cliffs, NJ.

AUTHOR BIOGRAPHIES

OSMAN BALCI is Professor of Computer Science at Virginia Tech and President of Orca Computer, Inc. He received his Ph.D. degree from Syracuse University in 1981. Dr. Balci is the Editor-in-Chief of two international journals: *Annals of Software Engineering* and *World Wide Web*; Verification, Validation and Accreditation (VV&A) Area Editor of *ACM Transactions on Modeling and Computer Simulation*; and Modeling and Simulation (M&S) Category Editor of *ACM Computing Reviews*. He serves as a member of the Winter Simulation Conference Board of Directors representing the Society for Computer Simulation. Most of Dr. Balci's research has been funded by DoD since 1983. Recently, he has provided technical services for the National Missile Defense (NMD) program in the areas of NMD system design M&S VV&A and NMD system design IV&V. His current research interests center on VV&A, IV&V, certification, quality assurance, and credibility assessment of (a) M&S applications, (b) software systems, and (c) complex software / hardware / humanware system designs. His e-mail and web addresses are `<balci@vt.edu>` and `<http://manta.cs.vt.edu/balci>`

WILLIAM F. ORMSBY is a Project/Program Manager for the Naval Surface Warfare Center Dahlgren Division (NSWCDD). He is currently the technical lead for the Ballistic Missile Defense Organization (BMDO) National Missile Defense (NMD) Independent Verification and Validation (IV&V) Program. In this capacity, Mr. Ormsby serves as Co-Chair of the NMD M&S VV&A Subgroup, which is chartered to ensure M&S VV&A processes are in place and executed to increase the confidence and reduce risk in the NMD system acquisition. He has produced software/hardware systems since 1982 when he received his M.S. degree in Engineering from Virginia Tech. He has developed concepts, written requirements, designed, implemented and delivered products for space systems supporting U.S. Navy mission objectives. Prior to his NMD assignment he led the development of models and simulations used to verify tactical guidance software performance for the TRIDENT II missile system. His e-mail address is `<OrmsbyWF@nswc.navy.mil>`