Proceedings of the 1997 Winter Simulation Conference
ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson

OPTIMIZING THE SELECTION OF VV&A ACTIVITIES
A RISK / BENEFIT APPROACH

Paul R. Muessig

Naval Air Warfare Center, Weapons Division
Code 418000D
China Lake, California 93555, U.S.A.

Dennis R. Laack
John W. Wrobleski, Jr.

Computer Sciences Corporation
711 Daily Drive
Camarillo, California 93010, U.S.A

ABSTRACT

This paper discusses a risk/benefit analysis approach to the selection of an optimal set of VV&A activities. The approach is an adaptation of MIL-STD-882C (System Safety Program Requirements) to the requirements of establishing the credibility of models and simulations (M&S). Risks and benefits of M&S use are quantified in terms of impact level and probability of occurrence within the context of specific applications. These assessments are then used to determine the level of M&S credibility required to minimize risk and/or maximize benefit. The required level of credibility is then used to select the most appropriate mix of VV&A activities from a standard set of activities that is calibrated to varying levels of required credibility. The end result is an ability to objectively justify the chosen set of VV&A activities, and to monitor the impact of risk/benefit changes on VV&A requirements. (This paper was originally presented at the Summer Computer Simulation Conference in Arlington, Virginia in July, 1997. It is presented here (in modified form) as part of WSC '97 IT-101 (Establishing Simulation Credibility Through VV&A)).

1 INTRODUCTION

M&S are being used with ever increasing frequency throughout society to address an ever-widening variety of questions. Many of the questions and issues being addressed with M&S have potentially significant consequences attached to the outcome or decision. Hence, there is a growing emphasis on demonstrating that the information generated through use of M&S is credible and reliable. Typically, the credibility of M&S results is demonstrated through verification, validation, and accreditation (VV&A) of the model or simulation, and through verification, validation and certification (VV&C) of the data that are used as inputs to the model or simulation. Until now, the determination of what VV&A and/or VV&C tasks are necessary or appropriate to demonstrate M&S credibility has been a matter of subjective judgment, typically left up to the organization charged with VV&A or VV&C responsibilities.

Frequently, the extent of VV&A activities is limited primarily by budgetary restrictions, with less consideration given to objective requirements for M&S credibility. This practice often results in disagreement over the adequacy of the V&V efforts relative to the requirements to accredit the model or simulation.

In an ideal world, the selection of the appropriate type and level of V&V activities would be based primarily on technical rationale and requirements for credibility, vice fiscal constraints. It is intuitively obvious that the level of M&S credibility needed in a particular situation (and hence the amount of V&V that should be performed), is related to the scope and the severity of any negative consequences that might arise from basing an important decision on the outputs of the M&S in question. Consider, for example, a situation in which the effectiveness of newly designed automobile safety features are to be evaluated using analysis based primarily on the outputs of M&S. The eventual decisions regarding safety design can have a significant impact on whatever portion of the population drives that automobile. Therefore, it would seem appropriate to do extensive V&V on any model (and its associated data) used in this application, so as to guarantee a high level of M&S credibility, and to reduce the risk of injury to passengers (and of lawsuits to the manufacturer). On the other hand, if M&S are to be used to predict the power requirements of a small new household appliance, the impact of an erroneous prediction is probably not very serious, and the level of M&S credibility required is much less. Hence, one would not expect that the same amount and level of V&V would be required for M&S used in this application.

To generalize from these hypothetical examples, it can be inferred that any decision which entails substantial risks and/or benefits requires an appropriate level of credibility for the information used to make the decision. If the information to be used to support the decision comes primarily or substantially from M&S, then the credibility required of the M&S predictions should be related to these potential risks and/or benefits,

and in turn, the scope and depth of the V&V done should be related to these credibility requirements. Figure 1 is a conceptual flow diagram illustrating the relationships between M&S predictions, decisions, risks and benefits. In this diagram, M&S generate information that contributes to the decision-making process. Each decision has associated with it a variety of risks and benefits, and the decision-making objective is to maximize the benefits while minimizing risks.

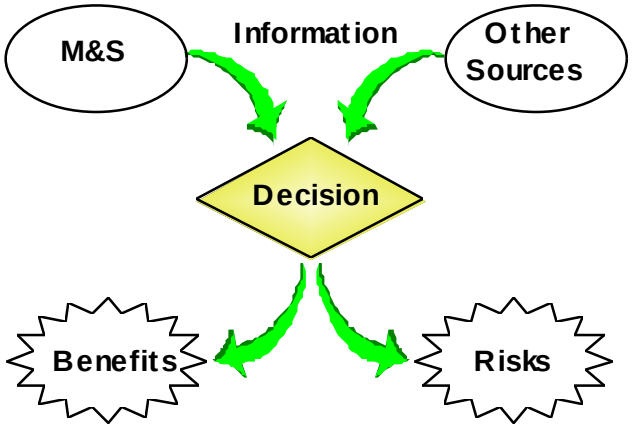

Figure 1: Generic Problem Paradigm

It stands to reason that the information used to make a decision must have a higher level of credibility the greater the potential benefits and/or risks. Since V&V results are used to demonstrate M&S credibility, the appropriate level and type of V&V required depends on two factors: the degree of potential benefits and risks, and the availability of other information to corroborate M&S results. This paper describes an approach to defining V&V requirements based on an assessment of the potential benefits and risks associated with decisions that use M&S outputs as inputs to the decision making process. The paper is written from a Department of Defense (DoD) perspective, since DoD has been one of the leaders in emphasizing the need for M&S credibility. The approach begins with techniques for quantifying decision-based risks and benefits, and concludes with a method for selecting the appropriate V&V tasks based on identified risks and benefits.

2 RISK ASSESSMENT

Risk is made up of two components: the impact (or consequences) of an event, and the probability or frequency of the event's occurrence. If each of these components could be quantified, the level of risk could be expressed using the formula:

Risk = (Impact Level) x (Probability of Occurrence)

In most cases the factors in this equation cannot be quantified absolutely, but can be subjectively quantified using the principles described in MIL-STD-882C, "System Safety Program Requirements." The general process for determining the overall level of risk first requires quantification of impact severity and probability for each separately identified risk factor. Using these two elements, an overall level of risk is assigned. This process is repeated for each individual risk factor, and the highest level of risk associated with the decision is selected as the level that drives the credibility requirement. The subjective criteria used in each step of the process are all explicitly stated, and can be tailored to the specifics of individual problems. A primary advantage of this methodology is that these subjective criteria are stated explicitly, so they can be easily discussed and understood, forming the basis for consensus-driven decision-making. The details of the process for determining credibility requirements are described in the following paragraphs.

2.1 Quantifying Impact Severity

The approach presented in MIL-STD-882C groups impact severity into four bands: catastrophic, critical, marginal, and negligible. (A risk factor is a specific type of outcome or result. For example, one risk factor might be injury or death of personnel; another might be damage to equipment; a third might be damage to a particular part of the environment.) The criteria for assigning one of these impact bands to a particular risk or benefit depends on the category of that risk or benefit. The impact categories that are discussed in 882C are personnel and equipment safety, environmental damage, and occupational illness. Depending on the particular use of M&S being considered, some of these impact categories might not apply, and additional categories might be added, for example: impact on end-user capability or effectiveness, cost, performance, schedule, and political or public reaction. A set of criteria for determining the level of impact for each of the different impact categories is given in
Table 1. This table uses the criteria found in MIL-STD-882C for the safety categories, and adds parallel criteria for the additional categories.

2.2 Quantifying Probability of Occurrence

The other element affecting risk is the probability that the impact will occur or be experienced. The probability of occurrence of a given risk factor can be described in four different ways, depending on the type of risk factor being considered. These four different ways are the expected number of occurrences over the life of a system; per items in a population; per unit of time; or per number of events. MIL-STD-882C divides the

probability continuum into five bands, and gives guidelines for selecting the appropriate band.
Table 2, extracted from the standard, provides these guidelines in terms of the number of occurrences over a lifetime, and per number of items in a population. These guidelines can be extrapolated by M&S users to other types of impacts that would be experienced over time or over a number of events.

Table 1: Criteria for Determining Impact Severity

IMPACT	**IMPACT LEVELS**			
CATEGORIES	CATASTROPHIC	CRITICAL	MARGINAL	NEGLIGIBLE
Personnel Safety	Death	Severe Injury	Minor Injury	Less than minor injury
Equipment Safety	Major equipment loss; Broad scale major damage	Small scale major damage	Broad scale minor damage	Small scale minor damage
Environmental Damage	Severe	Major	Minor	Some trivial
Occupational Illness	Severe & broad scale	Severe or broad scale	Minor & small scale	Minor or small scale
Cost	Loss of program funds; 100% cost growth	Funds reductions; 50-100% cost growth	20-50% cost growth	< 20% cost growth
Performance	Design does not meet critical thresholds	Severe design deficiencies but thresholds met	Minor design flaws, but fixable	Some trivial “out of spec” design elements
Schedule	Slip reduces overall DoD capabilities	Slip has major cost impacts	Slip causes internal turmoil	Republish schedules
Political or Public Impact	Widespread (Watergate)	Significant (Tailhook ‘91)	Embarrassment ($200 hammer)	Local

Table 2: Probability Levels

PROBABILITY DESCRIPTION	**LIKELIHOOD OF OCCURRENCE OVER LIFETIME OF AN ITEM**	**LIKELIHOOD OF OCCURRENCE PER NUMBER OF ITEMS****
Frequent	Likely to occur frequently	Widely experienced
Probable	Will occur several times in life of item	Will occur frequently
Occasional	Likely to occur some time in life of item	Will occur in several items
Remote	Unlikely but possible to occur in life of item	Unlikely but can reasonably be expected to occur
Improbable	So unlikely, it can be assumed occurrence may not be experienced	Unlikely to occur but possible

Table 3: Risk Assessment Matrix

FREQUENCY	**LEVEL OF IMPACT**			
	CATASTROPHIC	CRITICAL	MARGINAL	NEGLIGIBLE
FREQUENT	High	High	Medium	Low
PROBABLE	High	High	Medium	Low
OCCASIONAL	High	Medium	Low	Low
REMOTE	Medium	Medium	Low	Low
IMPROBABLE	Medium	Low	Low	Low

2.3 Assigning Risk Levels

As mentioned previously, the level of risk depends on both impact severity and probability (or frequency of occurrence). MIL-STD-882C presents several sample tables that quantify the level of risk based on the different levels of impact and probabilities. These tables are termed "risk assessment matrices." Each problem or application might have a unique risk assessment matrix. Table 3 is a typical risk assessment matrix.

The point of this discussion is that, for each M&S application, if risks associated with the expected decision or problem outcome can be identified, then impacts and probabilities can be banded using explicit criteria. Using these inputs, it is possible to quantify the level of risk, using explicitly stated criteria.

3 BENEFITS ASSESSMENT

"Benefits" are sort of the flip side of "Risks"; positive risks, if you will. Some decisions or problem outcomes will lead to very definite benefits, which can be evaluated and "quantified" with techniques that are similar to the risk assessment techniques described in the previous section. Within DoD, the term "benefits" almost always refers to operational benefits, which means increased warfighting or management capabilities (e.g. administrative communications, personnel management, morale & welfare improvements, reduction in operating costs, etc.).

3.1 Quantifying Beneficial Impacts

Regardless of the type of benefit, it is important to define some criteria by which the benefits can be ranked or categorized according to their level of impact. Table 4 provides suggested criteria for assigning a beneficial impact level to the different types of benefits. This table is similar in construct to the Impact Severity Criteria table in the risk assessment section. The criteria are based on the principles and value judgments similar to those used to construct
Table 1.

Table 4: Beneficial Impact Levels

BENEFIT	**BENEFICIAL IMPACT**			
CATEGORIES	REVOLUTIONARY	SIGNIFICANT	MARGINAL	NEGLIGIBLE
EFFECTIVENESS OF MAJOR SYSTEMS	Order of magnitude improvements in capability	Greater than 50% improvements	20 - 50% improvements	Improvements of less than 20%
MORALE & WELFARE	Impacts = to introduction of indoor plumbing	Impacts = to 10% pay raise	Impacts = to an extra vacation day	Impacts = to discounts on recreation activities
OPERATING COST REDUCTIONS	> 50%	25% - 50%	10% - 25%	< 10%

3.2 Quantifying Frequency of Occurrence

The other factor that is important in quantifying benefits is the probability that they will actually be realized. This factor is equivalent to the probability or frequency of occurrence in the risk analysis process, and is determined using a similar table. Benefits are typically experienced by a percentage of a population. In order to quantify this factor, namely the benefit realization frequency, it is necessary to specify three parameters: first, the population or "universe" over which the benefit might apply; second, the percentage of that population or universe that might be expected to realize the particular benefit; and third, a common period of time for measuring this realization frequency. For our purposes, namely development of a V&V tailoring technique for use within the DoD acquisition community, the population base might be some well-defined

organizational unit with similar equipment, capabilities and missions within a branch of the Service (a carrier air wing, for example, or a mechanized infantry battalion). The expected period of time wherein benefits might accrue might be arbitrarily set at one year in peace time, for example, or one month during combat operations. Table 5 is a subjectively developed example of quantifying the probability of realizing a benefit, which could be tailored to individual programs seeking to use this approach to determine the appropriate level of V&V to be conducted on M&S used to support program decisions. The important issue to be addressed in tailoring this approach to individual applications is that the criteria must be explicitly defined, so that the rationale for quantifying the benefits can be clearly

reviewed and understood during the accreditation process.

3.3 Assigning Benefit Levels

Knowing the probability that a benefit might be realized, and the degree of beneficial impact, a table can be developed that assigns an overall benefit rating based on these two factors. Table 6 is a benefits assessment matrix that was developed using values and principles similar to those used in constructing the risk assessment matrices. Just as with the Risk Assessment Matrix (Table 3), individual programs may wish to construct their own benefits matrix using similar principles.

4 CREDIBILITY VS RISK/BENEFIT

Having identified major risks and/or benefits, it is now possible to determine the level of credibility needed for the information that will be used to make or support a decision. The principle that governs the determination of information credibility requirements is that high levels of risk or benefit demand high levels of confidence in the information used to make the controlling decision(s). Lower levels of risk or benefit will lead to lower credibility requirements.

Two factors influence the determination of information credibility requirements. The first is that there can be multiple risks or benefits associated with a given decision or problem. Each risk or benefit may be quantified at a different level. The second factor is the availability of information from other sources that can be used to corroborate key information inputs to the decision making process (e.g., M&S results).

The first factor (multiple risks and benefits) can be addressed by quantifying each risk or benefit separately. The risk or benefit that is assigned the highest level determines the most critical information credibility requirement. The second factor (effects of corroborating information) can be addressed by reducing the level of credibility needed in proportion to the amount (and credibility) of the available corroborating information. A suggested rule of thumb, for example, might be that the required level of credibility can be reduced by one level if appropriate corroborating information is available and used to support the decision. The amount of reduction, if any, is dependent on the quality and quantity of corroborating information. In every case, the amount of reduction should be determined explicitly, and that determination should be reviewed as part of the accreditation decision.

Table 5: Criteria for Quantifying the Frequency of Benefits Realization

PERCENTAGE CATEGORY	**CRITERIA FOR SELECTION (Over a period of one year)**
Very High	> 20% of all DoD
High	10% - 20% of all DoD or major segments of a Theater Force Major segments of a type of force (e.g. amphibious, air, sea, etc.)
Medium	3% - 10% of all DoD or major segments of a Task Force Moderate segments of a type of force (e.g. amphibious, air, sea, etc.)
Low	1% - 3% of all DoD or several individual units (e.g. aircraft, tank, Platoon, Small ship etc.)
Minimal	< 1% of all DoD or few individual units

Table 6: Benefits Assessment Matrix

PERCENTAGE AFFECTED	**BENEFIT IMPACT**			
	REVOLUTIONARY	SIGNIFICANT	MARGINAL	NEGLIGIBLE
VERY HIGH	High	High	Medium	Low
HIGH	High	High	Medium	Low
MEDIUM	Medium	Medium	Medium	Low
LOW	Medium	Medium	Low	Low
VERY LOW	Medium	Low	Low	Low

5 DETERMINING APPROPRIATE V&V TASKS

Since one of the primary uses of V&V data is to establish the level of credibility that can be attached to a model or simulation when used for a particular application, and to reduce the uncertainty surrounding model or simulation outputs, the credibility requirements determined as above can be used to identify the V&V information that best provides the necessary credibility. To determine the appropriate V&V tasks there are two prerequisites. The first is a well-defined menu of V&V tasks that are recognized within the community as being appropriate for M&S typically used to solve problems within that community. (As an example, the Distributed Interactive Simulation (DIS) 9-Step VV&A process had 72 specific tasks associated with the VV&A of DIS applications. Each of these tasks was defined and articulated within the context of DIS application development. The DIS has now been replaced by the Simulation Interoperability Standards Organization, SISO.) The second is a well-defined information product that normally results from the completion of each task, and which provides a well-defined contribution to M&S credibility. With these two prerequisites it is a fairly simple task to select those V&V tasks that are best suited to generate the types of information that will provide the needed level of credibility.

As an example of this task selection process, consider the example cited earlier, where a model is being used to predict the power requirements for a small new electrical appliance. Assume that an analysis of the application led to the conclusion that "nominal" credibility was required for the models that are used in generating this power requirement. If the electrical appliance producers used a set of V&V techniques similar to those employed by the Joint Accreditation Support Activity (JASA), the appropriate set of tasks that would give "nominal" credibility to the model(s) could be selected from Table 7. (This table was developed based on M&S experience with several acquisition programs, and with V&V on selected M&S used in airborne weapon system acquisition analyses. It is intended only as an example of how to apply the methodology described in the article, and should be modified to suit individual applications.) From the table it is clear that only the first five tasks are necessary for this level of credibility. This table can be modified by adding or deleting tasks for different credibility levels.

The judgment as to what tasks are most appropriate to provide the indicated credibility level is normally made by the decision-maker who will make the final decision based on M&S outputs, or by a body of experts designated by that manager. A table such as this can be generated for any list of V&V tasks provided that the products of each task are well defined, and the contribution of these products to M&S credibility are clearly understood.

6. SUMMARY

The process described herein is a logical means of determining credibility requirements for a model or simulation that will be used to address a specific problem, and to determining which V&V tasks are most appropriate to provide this level of credibility. The process allows an analyst to start with a basic description of the problem, especially the potential risks and benefits associated with that particular problem, and to derive credibility requirements using these risk and benefit elements. Once the credibility requirements have been identified and agreed upon (i.e. the intermediate subjective guides have been reviewed and accepted) they are used in turn to select the most appropriate V&V information elements that will yield the required credibility. To perform this last step, it is necessary to have a standard menu of candidate V&V tasks, each of which provide a known contribution to establishing M&S credibility.

In addition to generating a clearly defined set of V&V tasks that are appropriate for a particular application, this approach also provides a logical rationale that can be used to justify budgetary requirements to complete the V&V efforts. Without such a rationale, budget justifications are often difficult, since there is typically no clear linkage between the planned V&V efforts and the credibility requirements of the application. Thus, risk/benefit analysis can be an effective means of developing cost-effective and objectively justifiable VV&A plans.

AUTHOR BIOGRAPHIES

DR. PAUL R. MUESSIG is currently Director of the Joint Accreditation Support Activity (JASA) at the Naval Air Warfare Center, Weapons Division, China Lake, CA. JASA's goal is to support airborne weapons system acquisition programs (and others) with credible integration of M&S into program objectives. JASA services include M&S application analysis, VV&A requirements definition, and VV&A planning, execution and reporting. (For more information, visit the JASA Home Page at http://www.nawcwpns.navy.mil/~jasa). Dr. Muessig has extensive experience in VV&A methodology development, planning and execution in support of numerous M&S applications in acquisition and testing. He has also worked as a defense analyst at the Center for Naval Analyses in Washington, D.C., contributing to survivability analyses for advanced technology aircraft, and leading model validation efforts for the Advanced Low Altitude Radar Model (ALARM). Dr. Muessig holds a B.S. in Chemistry from St. Joseph's University and a doctorate in Physical Chemistry from Brown University.

DENNIS R. LAACK is a retired combat Naval Aviator with extensive experience in directing and managing research, development, acquisition, and testing of naval aircraft, weapons and related systems. For the last five years he has provided management and technical assistance to the Joint Accreditation Support Activity (JASA), including original contributions in the areas of accreditation requirements determination, M&S acceptance criteria development, and V&V requirements derivation. He holds a B.E.E. from Marquette University, an M.S. in Systems Acquisition Management, and both an M.S. and an Ae.E. in Aeronautical Engineering from the Naval Postgraduate School. Mr. Laack is employed by Computer Sciences Corporation in Camarillo, California.

JOHN J. WROBLESKI, JR. is a retired Naval Flight Officer with extensive experience in directing and managing research, development, acquisition, and testing of naval aircraft, weapons and related systems. For the last two years he has provided management and technical assistance to the Joint Accreditation Support Activity (JASA), including original contributions in the areas of accreditation requirements determination, M&S acceptance criteria development, and V&V requirements derivation. He holds a Bachelor of Aerospace Engineering and Mechanics from the University of Minnesota, and both an M.S. and Ae.E. in Aeronautical Engineering from the Naval Postgraduate School. Mr. Wrobleski is employed by Computer Sciences Corporation in Camarillo, California.

Table 7: Guide for Selection of V&V Tasks

	Credibility Level		
	High	Moderate	Nominal
Baseline Definition	X	X	X
Determine C/M Attributes	X	X	X
Assess M&S Documentation	X	X	X
Establish VV&A Status, Usage History	X	X	X
ID Assumptions & Limits	X	X	X
S/W Quality Assessment	X		
Produce Design Documentation	X	X	
Perform Logical Verification	X	X	
Detailed Code Verification	X		
Sensitivity Analysis (Model Level)	X	X	
Sensitivity Analysis (Function Level)	X		
Face Validation		X	
Results Validation (Model Level)	X		
Results Validation (Function Level)	X		